

## Active visual perception for mobile robot localization

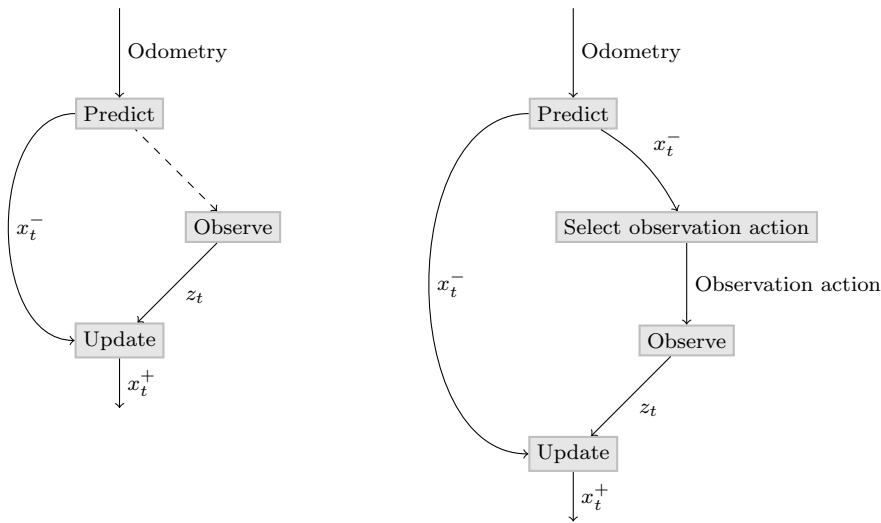
Javier Correa · Alvaro Soto

Received: date / Accepted: date

**Abstract** Localization is a key issue for a mobile robot, in particular in environments where a globally accurate positioning system, such as GPS, is not available. In these environments, accurate and efficient robot localization is not a trivial task, as an increase in accuracy usually leads to an impoverishment in efficiency and viceversa. Active perception appears as an appealing way to improve the localization process by increasing the richness of the information acquired from the environment. In this paper, we present an active perception strategy for a mobile robot provided with a visual sensor mounted on a pan-tilt mechanism. The visual sensor has a limited field of view, so the goal of the active perception strategy is to use the pan-tilt unit to direct the sensor to informative parts of the environment. To achieve this goal, we use a topological map of the environment and a Bayesian non-parametric estimation of robot position based on a particle filter. We slightly modify the regular implementation of this filter by including an additional step that selects the best perceptual action using Monte Carlo estimations. We understand the best perceptual action as the one that produces the greatest reduction in uncertainty about the robot position. We also consider in our optimization function a cost term that favors efficient perceptual actions. Previous works have proposed active perception strategies for robot localization, but mainly in the context of range sensors, grid representations of the environment, and parametric techniques, such as the extended Kalman filter. Accordingly, the main contributions of this work are: i) Development of a sound strategy for active selection of perceptual actions in the context of a visual sensor and a topological map; ii) Real time operation using a modified version of the particle filter and Monte Carlo based estimations; iii) Implementation and testing of these ideas using simulations and a real case scenario. Our results indicate that, in terms of accuracy of robot localization, the proposed approach decreases mean average error and standard deviation with respect to a pasive perception scheme. Furthermore, in terms of efficiency, the active scheme

---

Javier Correa, Alvaro Soto  
Escuela de Ingeniería  
Pontificia Universidad Católica de Chile  
Vicuña Mackenna 4860  
Santiago, Chile  
Tel. : +56-2-3542000



**Fig. 1** Steps of the localization process. a) Left diagram, main steps of robot localization using Bayesian techniques. b) Right diagram, proposed strategy.

is able to operate in real time without adding a relevant overhead to the regular robot operation.

**Keywords** robot localization · active perception · mobile robots

**Mathematics Subject Classification (2000)** 93C99

## 1 Introduction

The problem of obtaining an accurate estimation of the position of a mobile robot within its environment, task known as localization, is one of the key issues a mobile robot must solve. Achieving a robust and efficient localization method is not a trivial task, as an increase in robustness usually leads to an impoverishment in efficiency and viceversa. This stresses the need to focus the perceptual resources on the most informative parts of the environment.

Currently, state of the art solutions to indoor robot localization are mainly based on Bayesian techniques, such as the extended Kalman filter (EKF) or the particle filter (PF) [24]. These techniques consist of three main basic steps: i) A prediction step, where the robot uses its last position and self-motion information to predict its new position, ii) An observation step, where the robot senses the environment, and iii) A position refinement step, where the robot updates the estimation of its current position using the new observations from the environment. Figure 1-a shows a schematic view of these 3 steps.

In the previous Bayesian scheme, the observation step does not consider the possibility of actively controlling the sensors in order to direct them to the most informative parts of the environment. In effect, most current localization techniques consider passive sensors usually mounted on the robot. These sensors provide observations about

the part of the environment that is local to the current robot position. In this work, we use concepts from information theory and planning under uncertainty to develop a sound strategy to actively control the robot sensors.

As a testbed, we use a differential drive wheeled robot provided with an odometer on each wheel and a color camera. This camera has a limited field of view and is mounted on a pan-tilt mechanism. By using the odometer, the robot is able to track its motions. By using the video camera and suitable computer vision algorithms, the robot is able to distinguish a set of visual landmarks that represent the map of the environment. As our focus corresponds to the localization problem, we assume that the map of the environment is known in advance, see [24] for a description of relevant mapping techniques.

The basic idea of our localization approach resides on actively using the pan-tilt mechanism to direct the visual sensor to areas of the environment where the robot expects to obtain the most useful information to achieve an accurate localization. We accomplish this goal by adding a new step between the prediction and observation steps of Figure 1-a. In this step, the robot uses the prediction of its current position and the map of the environment to assign a score to each possible perceptual action. This score considers the expected utility and the cost of executing each action. Figure 1-b shows the new scheme.

The previous idea is closely related to recent applications of computational models of visual attention mechanisms that combine top-down information with bottom-up data [14] [9]. In our case, top-down information is provided by the map of the environment and the current estimation of the robot position, while bottom-up information is given by the current images acquired by the video camera. This allows the robot to bias its perception and maximize the expected information of the observation step. Our experiments indicate that by using this strategy, we are able to increase the efficiency and accuracy of the robot position estimation.

Previous works have proposed active perception strategies for robot localization, but mainly in the context of range sensors [5], evidence grid representations of the environment [22], and parametric techniques [7], such as, the extended Kalman filter. In our case, we develop our active perception strategy for the case of robot localization using a visual sensor, a topological map of the environment, and a Bayesian non-parametric estimation of robot position based on a particle filter. Accordingly, the main contributions of this work are: i) Development of a sound strategy for active selection of perceptual actions in the context of a visual sensor and a topological map; ii) Real time operation using a modified version of the particle filter and Monte Carlo based estimations; iii) Implementation and testing of these ideas using simulations and a real case scenario.

This document is organized as follows. Section 2 describes related previous work. Section 3 presents the proposed method and relevant implementation details. Section 4 presents our empirical results using simulations and a real case. Finally, Section 5 presents the main conclusions of this work and future avenues of research.

## 2 Previous work

First, we review previous works related to active visual perception mainly in the context of computer vision applications. Afterwards, we focus our review on relevant works related to using active perception in the context of mobile robots.

In terms of active vision some of the most relevant works are from the nineties [1] [3] [4] mainly motivated by the “where to look” and “how to look” problems. More recently, the “where to look” problem has gained interest due to the advantages provided by new visual attention mechanisms derived from visual saliency techniques and top-down feedback mechanisms [14] [11]. In computational terms, saliency is usually employed to focus processing resources in key parts of an image, in order to improve efficiency, performance, or both. Previous work in this area includes several models of visual attention, such as [25], [11], and [23]. In our case, it is possible to understand our approach as an instance of a top-down attention mechanism driven by information from a known topological map of the environment.

Besides the “where to look” and “how to look” problems, recent work in computer vision adds a new challenge: “what to look for” [17]. Here, seminal works, such as [15] and [13], have shown the relevance of using contextual information as top-down cues to boost efficient object recognition. Furthermore, in the context of robot localization, Siagian and Itti [20] and our own work [9] have shown the advantages of using high level place recognition as a top-down cue to efficiently recognize visual landmarks. As we discuss later, in this work we facilitate the landmark detection problem by using a set of easily detectable artificial landmarks.

In the context of mobile robot navigation, several works have proposed the use of active perception strategies to improve robot performance. Among these works, entropy based utility functions have been one of the most popular criteria. In [5], Burgard et al. present one of the earliest work using an entropy based utility function to improve robot localization through active perception. Their approach consists on a 2-step active localization strategy: i) Where to move, and ii) Where to look. The first step allows the robot to follow trajectories that lead to highly informative areas of the environment. The second step selects from an array of sonar beams the most suitable to sense the environment. In the context of path planning and localization, Roy and Thrun [19] propose an approach based on partially observable Markov decision processes (POMDP) and entropy minimization techniques. The goal is to find an informative route between two points in a map. The resulting robot behavior is a coastal navigation scheme, where the robot moves near informative structures of the environment, such as walls. In contrast to our approach, these previous works focus on path planning for localization using an occupancy grid representation and range sensors.

In the context of simultaneous localization and mapping (SLAM), Davison and Murray [7] use a Kalman filter approach to rate landmarks according to uncertainty. Using this information and a previously defined action policy, the robot decides where to point its camera in order to obtain informative readings from the environment. In [22], Stachniss et al. use information gain as the main score to develop an exploration plan for the SLAM process. The resulting plan balances accurate localization and exploration of new areas to improve map estimation. As in our work, they use a particle filter to keep track of the state, but in the context of an occupancy grid representation and a range sensor.

In a recent work closely related to our approach, Mitsunaga and Asada [12] present a strategy to actively select visual landmarks using a limited viewing angle camera that is mounted on a mobile robot. In contrast to our approach, their strategy focuses on selecting perceptual actions that facilitate the completion of the current goals of the robot instead of self-localization. The proposed strategy is based on decision tree classifiers and information gain to guide the action selection. The main drawbacks of the approach are the need to operate under constrained situations and to manually

train the robot to learn the action policy for each task. We believe that combining self-localization and goal completion in the selection of perceptual actions is an interesting idea, but a key issue is to devise strategies to scale this approach to general environments and tasks. In this regard, new efficient POMDPs based strategies are a promising research path, as shown in [16] [18]. Also, in the context of actively controlling a visual sensor onboard of a mobile robot, Soyer et al. [21] propose the use of history of visual fixations to achieve visual recognition. The history of fixations is given by the saccadic motions of an attentional vision system. In contrast to our approach, the main focus of that work is object recognition instead of robot localization.

### 3 Our approach

In this section we provide the main details of our approach. First, we briefly describe the main features of the particle filter, the technique that we use to keep a probabilistic estimation of the robot position. Then, we describe the main details of the motion and perception models used by our implementation of the particle filter. Afterwards, we describe our criterion to search the space of possible perceptual actions. Finally, we explain the details of implementing this criterion using a particle filter.

#### 3.1 Particle filter

The particle filter is a useful non-parametric technique to perform sequential state estimation via Bayesian inference. It provides great efficiency and extreme flexibility to approximate any functional non-linearity. The key idea is to use samples, also called particles, to represent the posterior distribution of the state given a sequence of sensor measurements. As new information arrives, these particles are constantly re-allocated to update the estimation of the state of the system.

In Bayesian terms, the posterior distribution of the state can be expressed by:

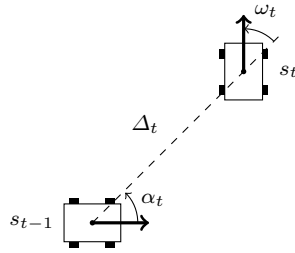
$$p(s_t|o_{1:t}) = \beta p(o_t|s_t) p(s_t|o_{1:t-1}), \quad (1)$$

where  $\beta$  is a normalization factor;  $s_t$  represents the state of the system at time  $t$ ; and  $o_{1:t}$  represents all the observations collected until time  $t$ . Equation (1) assumes that  $s_t$  totally explains the current observation  $o_t$ .

The particle filter provides an estimation of the posterior distribution in Equation (1) in 3 main steps: sampling, weighting, and re-sampling. The sampling step consists of taking samples (particles) from the so-called dynamic prior  $p(s_t|o_{1:t-1})$ . Next, in the weighting step, the resulting particles are weighted by the likelihood term  $p(o_t|s_t)$ . Finally, a re-sampling step is usually applied to avoid the degeneracy of the particle set [10]. The key issue that explains the efficiency of the filter comes from using a Markovian assumption to express the dynamic prior as:

$$p(s_t|o_{1:t-1}) = \int p(s_t|s_{t-1}) p(s_{t-1}|o_{1:t-1}) ds_{t-1}. \quad (2)$$

Equation (2) provides a recursive implementation that allows the filter to use the last state estimation  $p(s_{t-1}|o_{1:t-1})$  to select  $n$  particles  $s_{t-1}^i$ ,  $i \in [1..n]$ , for the next iteration. Each of these particles is then propagated by the dynamics of the process



**Fig. 2** Motion model. From initial position  $s_{t-1} = (x_{t-1}, y_{t-1}, \theta_{t-1})$ , the robot rotates an angle  $\alpha_t$ , then moves a distance  $\Delta_t$ , and finally rotates an angle  $\omega_t$  to reach its final state  $s_t = (x_t, y_t, \theta_t)$ .

$p(s_t | s_{t-1}^i)$  to obtain the new set of particles that estimate the state at time  $t$ . In this way, at all times, the state of the system is estimated by a set of weighted samples:

$$S_t = \{ \langle s_t^{(i)}, \omega_t^{(i)} \rangle | i \in [1..n] \}, \quad (3)$$

where the weights  $\omega_t^{(i)}$  form a set of normalized factors called importance weights.

In the area of robot localization, the particle filter was first used in [8]. Since then, it has become one of the favorite tools for robot localization and also for the related problem of building maps of the environment [24]. In the implementation of the particle filter, the key issues are the prediction step, given by  $p(s_t | s_{t-1}^i)$ , and the observation step, given by  $p(o_t | s_t)$ . In terms of robot localization, these steps are given by the so-called motion and perception models, respectively. We discuss next the motion and perception models used in this work.

### 3.2 Motion model

As usual in Mobile Robotics, we assume a point model for the robot [24]. Also, to simplify our analysis, we consider only the case of planar motions, although, the extension to the 3D case is straightforward. Accordingly, the position of the robot at time  $t$  is given by the state vector:  $s_t = (x_t, y_t, \theta_t)$ . Following [2], our motion model assumes that to move from  $s_{t-1}$  to  $s_t$ , the robot performs three independent actions, as represented in Figure 2. First, the robot rotates an angle  $\alpha_t$  to face the direction of its translation. Afterwards, it translates a distance  $\Delta_t$  to reach  $(x_t, y_t)$ . Finally, the robot rotates an angle  $\omega_t$  to face its final orientation  $\theta_t$ .

The set of actions  $\delta_t = (\alpha_t, \Delta_t, \omega_t)$  that moves the robot from  $s_{t-1}$  to  $s_t$  is estimated using odometry measurements. Let this estimation be  $\hat{\delta}_t = (\hat{\alpha}_t, \hat{\Delta}_t, \hat{\omega}_t)$ . We assume that the uncertainty in this estimation corresponds to unbiased Gaussian models with known variance. Therefore, the resulting probabilistic motion model  $p(s_t | s_{t-1})$  is given by [2]:

$$\begin{pmatrix} x_t \\ y_t \\ \theta_t \end{pmatrix} = \begin{pmatrix} x_{t-1} + \hat{\Delta}_t \cos(\theta_{t-1} + \hat{\alpha}_t) \\ y_{t-1} + \hat{\Delta}_t \sin(\theta_{t-1} + \hat{\alpha}_t) \\ \theta_{t-1} + \hat{\alpha}_t + \hat{\omega}_t \end{pmatrix} \quad (4)$$

where:

$$\begin{aligned}\hat{\alpha}_t &\sim N(\alpha_t; \phi_\alpha^2) \\ \hat{\Delta}_t &\sim N(\Delta_t; \phi_\Delta^2) \\ \hat{\omega}_t &\sim N(\omega_t; \phi_\omega^2)\end{aligned}\tag{5}$$

The values of the variances  $\phi_\alpha^2$ ,  $\phi_\Delta^2$ , and  $\phi_\omega^2$ , are estimated using training data obtained during a calibration period using the real sensor onboard our robot.

### 3.3 Perception model

Our perception model is based on the detection of visual landmarks that are modeled as 2D points. For each landmark, our visual detection system can sense range and bearing with respect to the robot's local frame of coordinates. Without loss of generality, we consider just a planar case, therefore, we limit our analysis to control only the panning motion of the robot camera. The extension to the 3D case, considering an azimuth angle, is straightforward, although, as we discuss below, there is an increase in computational complexity.

We use unbiased Gaussian models with known variance to quantify the uncertainty in the sensor measurements. We further assume that range and bearing are independent variables for each landmark. If we denote, respectively, as  $r_t^l$  and  $\beta_t^l$ , the range and bearing of a given landmark  $l$  sensed by the robot at position  $s_t$  at time  $t$ , the resulting probabilistic perception model  $p(r_t^l, \beta_t^l | s_t)$  is given by:

$$\begin{pmatrix} r_t^l \\ \beta_t^l \end{pmatrix} = \begin{pmatrix} \sqrt{(x_t - l_x)^2 + (y_t - l_y)^2} \\ \tan^{-1}(l_y - y_t, l_x - x_t) - \rho_t - \theta_t \end{pmatrix} + \begin{pmatrix} \xi_{\phi_r^2} \\ \xi_{\phi_\beta^2} \end{pmatrix}\tag{6}$$

where  $(l_x, l_y)$  is the position of landmark  $l$  according to the map of the environment and  $\rho_t$  is the panning angle of the camera.  $\xi_{\phi_r^2}$  and  $\xi_{\phi_\beta^2}$  are zero-mean Gaussian error variables with variances  $\phi_r^2$  and  $\phi_\beta^2$ , respectively. As in the case of the motion model, these variances are estimated using training data obtained during a calibration period using the real sensor onboard the robot.

### 3.4 Scoring of perceptual actions

To evaluate the expected benefit of each possible perceptual action, as in [22], we borrow from information theory the concept of information gain. Our intuition is to use a score that allows us to evaluate the level of uncertainty, or entropy, between the estimation of the robot position with and without considering the information provided by a given perceptual action. We can express this intuition in terms of information gain,  $IG$ , which can be expressed in terms of the entropy  $H(\cdot)$  of a random variable as [6]:

$$IG(o_t, a_t^i) = H(s_t | o_{1:t-1}, m) - H(s_t | o_{1:t}, m, a_t^i)\tag{7}$$

where  $o_t$  is the observation at time  $t$ ,  $a_t^i$  is the perceptual action  $i$  at time  $t$ ,  $m$  is the map of the environment, and  $s_t$  is the estimation of robot position at time  $t$ .

Equation (7) quantifies the expected reduction in the entropy of the state estimation  $s_t$  by executing perceptual action  $a_t^i$ . From now on, we omit  $m$  and  $o_{1:t-1}$  in the equations as they are constant to all terms.

In Equation (7) only the second term in the right hand side depends on the perceptual action  $a_t^i$ . Applying to this term the definition of entropy and Bayes theorem, we have:

$$\begin{aligned} H(s_t|o_t, a_t^i) &= - \int p(s_t|o_t, a_t^i) \log p(s_t|o_t, a_t^i) ds_t \\ &= - \frac{1}{p(o_t|a_t^i)} \left( \int p(o_t|a_t^i, s_t) p(s_t|a_t^i) \log (p(o_t|a_t^i, s_t) p(s_t|a_t^i)) ds_t \right. \\ &\quad \left. - \log p(o_t|a_t^i) \int p(o_t|a_t^i, s_t) p(s_t|a_t^i) ds_t \right) \end{aligned} \quad (8)$$

In terms of  $p(s_t|a_t^i)$ , each possible perceptual action  $a_t^i$  does not affect the knowledge that we have about the position  $s_t$ , therefore,  $p(s_t|a_t^i) = p(s_t)$ . In terms of  $p(o_t|s_t, a_t^i)$ , we know the map of the environment and the field of view (*FOV*) of the camera onboard the robot, therefore, we know deterministically which landmarks are visible at each robot position  $s_t$ . Using this fact, we estimate  $p(o_t|s_t, a_t^i)$  by assuming a constant probability  $P_{lm} > 0$  of correctly sensing each of the visible landmarks and a null probability of sensing landmarks that are out of the *FOV* of the robot. We further assume that landmarks are observed independently. In this way, we estimate (8) by:

$$\begin{aligned} H(s_t|o_t, a_t^i) &\approx - \frac{1}{p(o_t|a_t^i)} \left( \int_{FOV} P_{lm} p(s_t) \log (P_{lm} p(s_t)) ds_t - p(o_t|a_t^i) \log p(o_t|a_t^i) \right) \\ &= \log p(o_t|a_t^i) - \frac{P_{lm}}{p(o_t|a_t^i)} \int_{FOV} p(s_t) \log (P_{lm} p(s_t)) ds_t. \end{aligned} \quad (9)$$

where  $\int_{FOV}(\cdot) ds_t$  means integration over positions  $s_t$  with visible landmarks.

Replacing (9) in (7), we have:

$$IG(o_t, a_t^i) = H(s_t) - \log p(o_t|a_t^i) + \frac{P_{lm}}{p(o_t|a_t^i)} \int_{FOV} p(s_t) \log (P_{lm} p(s_t)) ds_t, \quad (10)$$

where:

$$p(o_t|a_t^i) = \int p(o_t|a_t^i, s_t) p(s_t|a_t^i) ds_t = P_{lm} \int_{FOV} p(s_t) ds_t \quad (11)$$

As we describe below, Equation (10) provides a way to search for the perceptual action associated to the highest expected *IG*. As a further refinement, we also consider in our model the cost of executing each perceptual actions. In general, this cost can be related to the complexity or time needed to complete each action. Considering the cost, the final expression to select the optimal perceptual action  $a_t^*$  is given by:

$$a_t^* = \arg \max_{a_t^i} \{ E[IG(o_t, a_t^i)]_{o_t|a_t^i} - \alpha \cdot cost(a_t^i) \} \quad (12)$$

where  $E[\cdot]$  denotes expected value with respect to  $p(o_t|a_t^i)$ , and  $\alpha$  is a weighting factor that trades-off the expected *IG* and the cost of each possible perceptual action. This factor also balances differences in units, in our case nats or bits for *IG* and seconds for the cost.

### 3.5 Implementation

Unfortunately, there is not a closed solution to Equation (12). In this work, we use a Monte Carlo approximation using the  $n$  samples provided by a particle filter and the corresponding set of normalized importance weights  $\omega_k$ . As in [22], we approximate  $H(s_t)$  by:

$$H(s_t) \approx - \sum_{k=1}^n \omega_k \log \omega_k \quad (13)$$

Similarly, using Equation (9) we approximate  $H(s_t|o_t, a_t^i)$  by:

$$H(s_t|o_t, a_t^i) \approx \log p(o_t|a_t^i) - \frac{P_{lm}}{p(o_t|a_t^i)} \sum_j \omega_j \log P_{lm} \omega_j \quad (14)$$

where  $p(o_t|a_t^i) = P_{lm} \sum_j \omega_j$ . It is important to note that the  $j$  index in the last term and in Equation (14) goes only over particles that represent positions where there are visible landmarks.

In terms of the cost function, in our case we have a pan-tilt unit to perform each perceptual action. We define a cost function proportional to the angle that the robot has to move the pan-tilt to observe the desired landmarks. We set empirically the weighting factor  $\alpha = 0.1$ , as a good compromise between rotational freedom of the pan-tilt unit and time to wait for the action to be completed.

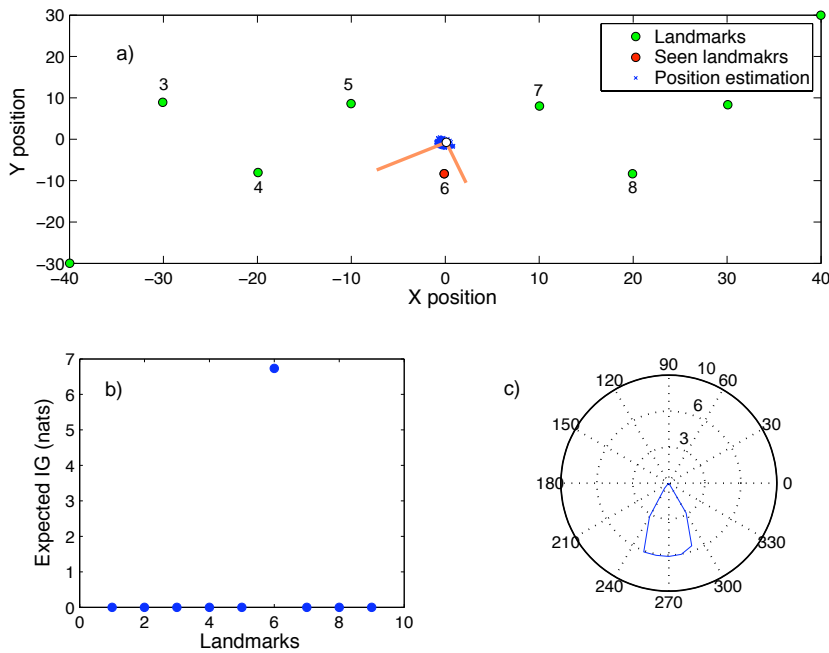
We estimate  $IG$  in Equation (10) using all the samples provided by the particle filter. Accordingly, the computational complexity of this estimation is  $O(l|A|n)$ , where  $l$  is the number of landmarks in the map,  $|A|$  is the cardinality of the action space, and  $n$  is the number of particles used in the estimation. This linear complexity allows us to operate in real time using a general purpose laptop machine. In effect, in Section 4.2 we show that, in a real case, the main overhead caused by the active perception capability is not given by the processing time, but by the time that the pan-tilt unit requires to reach the desired angle.

## 4 Results

In this section we show results of testing our approach using simulations and a real robot. In the case of simulations, we compare our strategy to non-adaptive schemes. In the case of a real robot, our testbed is a wheeled differential drive robot wandering in a hall of an office building environment.

### 4.1 Simulation

We use MATLAB<sup>®</sup> to simulate the behavior of a virtual robot. The robot is provided with a map of the environment given by  $(x, y)$  positions of a set of landmarks. The robot is also provided with a virtual odometer that generates noisy measurements according to the motion model described in Section 3.2. In all tests we use a standard deviation of 0.17 degrees for rotations and 0.4 meters for traveled distances. These values are obtained after a period of calibration using our real robot. In addition, the virtual robot is also equipped with a visual sensor that provides range and bearing

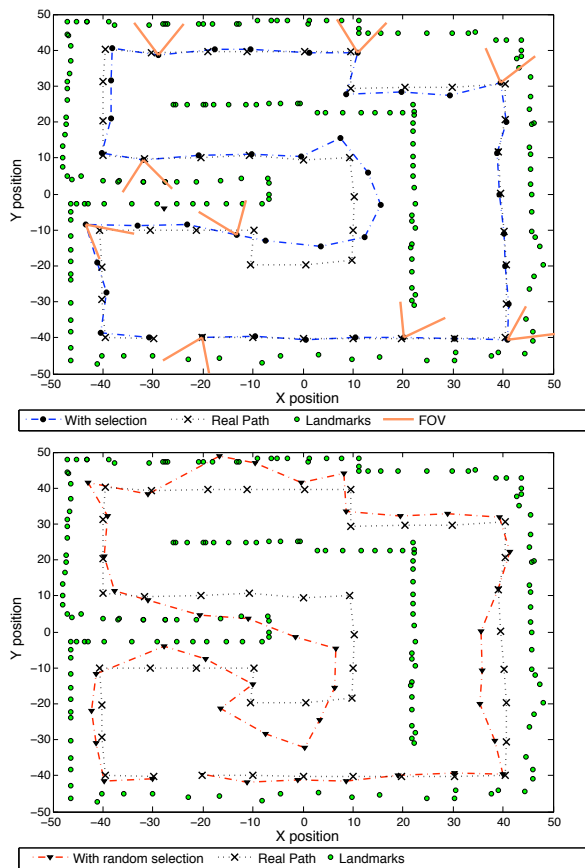


**Fig. 3** Graphical interface of the simulation. a) Particles are highly clustered around the true robot position. Only landmark number 6 is within the *FOV* of the robot. b) Expected *IG* of observing each landmark. c) Polar graph of expected *IG* for the panning angles in the action space (radial axis represents expected *IG* in nats units).

readings about all landmarks in its *FOV*. These readings are generated according to the perception model described in Section 3.3 using a standard deviation of 0.5 meters for range and 0.14 degrees for bearing. These values are obtained according to the level of noise observed in the real sensor. In this way, a simulation consists of generating a map and a path for the robot to follow. Odometry and visual perception readings are generated as the robot follows the path. Figure 3-a shows an instance of the simulated environment. In this case, there is just one landmark within the *FOV* of the robot. Accordingly, Figures 3-b and 3-c correctly show that the maximum information gain is reached for just one of the landmarks (landmark 6) when the camera has a panning angle of approximately  $270^\circ$  with respect to the X-Y axis.

To compare the effect of introducing an active perceptual strategy to the localization process, we carry out simulations under different perceptual schemes: i) A fixed camera always oriented according to the robot direction of motion, ii) An active sensor that, at each step, randomly chooses a panning angle according to a uniform distribution, and iii) The proposed strategy to actively select perceptual actions. Using the map and robot trajectory shown in Figure 4, we run 50 simulations for each perceptual strategy. In the simulation, the robot starts approximately at position  $(-20, -40)$ , and move counter-clockwise along the path. For each simulation, we calculate the mean square error (MSE) of the estimation of robot position along the path. Table 1 summarizes our results. In the case of a fixed camera, the position estimation is very poor given that the robot does not have a chance to sense most of the landmarks. In the case of randomly selecting a panning angle, the high number of landmarks around the sides

of the robot trajectory helps the robot to often sense some landmarks and keep a low error estimation. Finally under our strategy, we observe the best position estimation. Using the *IG* based action selection, the robot is always sensing one of its sides, thus, it is constantly sensing landmarks. In effect, the use of a cost term in Equation (12) keeps the camera pointing just to one of the sides to avoid delays related to the activation of the panning system. Figure 4 shows this behavior for some robot positions in the map (this figure looks better in color). It is important to note that the goal of this toy example is to illustrate the potential advantage of using an adaptive perceptual strategy. In general, such advantage depends on the configuration of the landmarks in the map. Clearly, an active perception scheme is more valuable in cases, such as the one in Figure 4, where most of the landmarks are not directly accessible by a fixed frontal sensor.



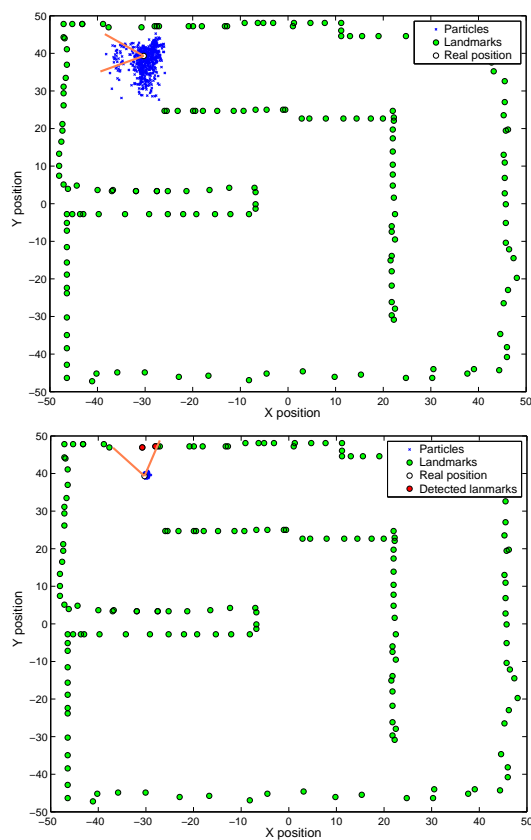
**Fig. 4** Example of robot path estimation using different perceptual schemes. a) Using our approach. b) Using a random selection of actions. In the case of our approach, we also show the *FOV* for some of the robot positions, which is always trying to observe informative landmarks.

Figure 5 shows a test run that highlights the level of uncertainty in the estimation of robot position with and without using an active selection of perceptual actions. In

	Averaged MSE	St. deviation
Fixed camera	17.54	9.62
Random action selection	6.09	3.06
Proposed strategy	4.23	3.47

**Table 1** Average MSE and standard deviation of error in position estimation, measured in pixels for a total of 50 simulations.

the figures, the blue clouds of points represent the estimation achieved by the particle filter. We observe that in the case of active sensing, there is a significant lower level of uncertainty in the position estimation, given by a greater density of particles around the true robot position.



**Fig. 5** Particles distribution using MonteCarlo localization after some steps following the robot path. a) Localization using a camera fixed to the robot orientation. b) Localization using our strategy for perceptual action selection.

## 4.2 Real robot

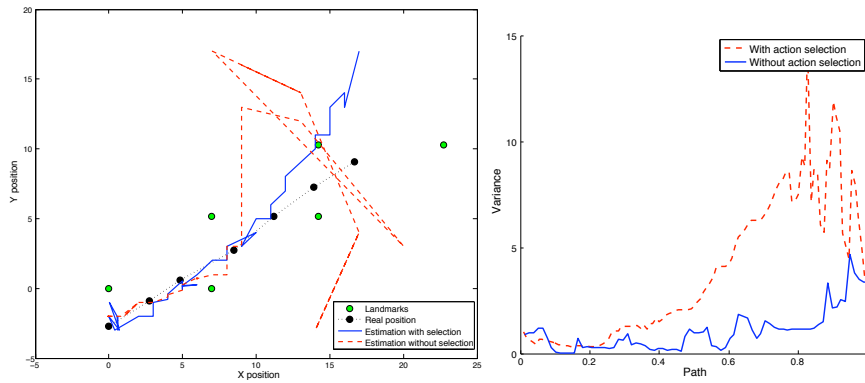
In this section we present results of testing our strategy on a real robot. We use a Pioneer-3 AT robot, equipped with a Directed Perception PTU-C46 pan-tilt unit and a Point Grey Dragonfly2 camera. Given that our main focus is an active localization strategy, we facilitate the landmark detection task by using a set of artificial visual landmarks. These landmarks consist of cylinders covered with two small patches of distinctive and easily detectable colors. Figure 6 shows pictures displaying the robot, the testing environment, and some of the artificial visual landmarks used in our test.



**Fig. 6** Pictures displaying the robot, the testing environment, and some of the artificial visual landmarks used in our test.

In our implementation we use 2000 particles. We discretize the  $360^\circ$  panning range into  $6^\circ$  steps, so the resulting action space is composed of 60 different panning angles. The values of the parameters of the motion and perception models are the same as those described in Section 4.1. In terms of landmark detection, as the size and shape of each landmark are known, we use this information to estimate range. We also estimate the bearing of each landmark using information from its horizontal position in the image. We use a detection probability  $P_{lm} = 0.7$ . This probability is estimated empirically by counting the times that landmarks are detected under different conditions. Landmarks are sometimes missed by the detector due to variations in lighting conditions. The map of the environment is given by the position of the artificial landmarks. In addition, in order to capture precise ground truth data, we manually marked on the floor waypoints that define the robot path for each test.

We test two strategies: i) A fixed camera always oriented according to the robot direction of motion, and ii) Our proposed strategy to select perceptual actions. Figure 7-a shows the results of one of our test. In the figure the robot moves from left to right. We observe that the active perceptual scheme achieves a better position estimation. This is explained by the fact that the robot is able to sense landmarks that are not available to the fixed camera case. Towards the end of the robot trajectory, the landmark sensing algorithm misses the two last landmarks closer to the robot path, and as a result, both approaches decrease the accuracy of the estimation. Figure 7-b shows that the active sensing scheme also provides an estimation with less variance than the fixed camera case.



**Fig. 7** a) Estimation of robot position. b) Variance of estimation of robot position.

## 5 Conclusions and Future Work

In this work, we present a robot localization strategy based on an active visual perception scheme. Using the machinery of a particle filter, and probabilistic motion and perception models, we are able to obtain an optimization function that provides our robot with an optimal perceptual action at each inference cycle of the localization process. The optimality criterion to select this action is based on expected information gain and the cost of performing the action.

In terms of implementation, the use of a non-parametric technique based on a particle filter allows us to achieve a linear computational complexity with respect to number of particles and landmarks. Also, by using Monte-Carlo approximations, we are able to calculate all the relevant terms using sums and log-sums of particle weights. In this sense, the main source of overhead introduced by the proposed strategy corresponds to the time to rotate the pan-tilt unit. In our implementation, we limit this problem by providing an adequate penalization cost to large rotations. It is important to note that, although we base our approach on a visual sensor with a limited *FOV*, the extension to other types of sensors or a more general visual sensor, such as an omnidirectional camera, is straightforward. In this last case, the selection of perceptual actions can be based on selecting relevant parts of the panoramic images to apply the algorithms that detect the landmarks.

In terms of results, all of our simulations and also the real test using a mobile robot show that the proposed strategy decreases the mean square error and uncertainty of the position estimation with respect to a passive perception scheme. The increase in performance by using the proposed strategy depends on the spatial distribution of the visual landmarks. Clearly, cases with few landmarks or where most of the landmarks are not directly visible in front of the robot, correspond to situations where the active perception approach contributes the most. In this regard, in the case of the real robot, we observe that, while navigating the environment, the robot shows a remarkable behavior by constantly turning his head to observe relevant landmarks that, otherwise, are missed. We believe that this is a highly desirable behavior for a mobile robot operating in a natural environment.

There are several research avenues to continue this work. One idea is to consider in the perceptual action selection not only the uncertainty in the localization process,

but also the uncertainty in the map estimation. The challenge will be to avoid the higher computational complexity given by planning in such a high dimensional space. Another path to follow is to expand the influence of top-down feedback to the level of visual landmark recognition. We believe that by using contextual information, such as the topological map used in this work, it is possible to obtain information not only about “where to look”, but also about “what to look for”. In this sense, the active perception strategy can be integrated with a perceptual planner that guides the landmark recognition process, for example by providing prior information about which perceptual algorithms are more suitable to detect each particular landmark.

## 6 ACKNOWLEDGMENTS

This work was partially funded by FONDECYT grant 1070760.

## References

1. Y. Aloimonos. *Active Perception, Vol. I of Advances in Computer Vision series*. Lawrence Erlbaum Associates, 1993.
2. A. Aranedá. *Statistical Inference in Mapping and Localization for Mobile Robots*. PhD thesis, Dept. of Statistics, Carnegie Mellon University, 2004.
3. R. Bajcsy and M. Campos. Active and exploratory perception. *CVGIP: Image Understanding*, 56(1):31–40, 1992.
4. D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.
5. W. Burgard, D. Fox, and S. Thrun. Active mobile robot localization. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1997.
6. T. Cover and J. Thomas. *Elements of information theory*. John Wiley, 1991.
7. A. Davison and D. Murray. Simultaneous localization and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):865–880, 2002.
8. F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo localization for mobile robots. In *Proc. of International Conference on Robotics and Automation (ICRA)*, 1999.
9. P. Espinace, D. Langdon, and A. Soto. Unsupervised identification of useful visual landmarks using multiple segmentations and top-down feedback. *Robotics and Autonomous Systems*, 56(6):538–548, 2008.
10. N. Gordon, D. Salmon, and A. Smith. A novel approach to nonlinear/non Gaussian Bayesian state estimation. In *IEE Proc. on Radar and Signal Processing*, pages 107–113, 1993.
11. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
12. N. Mitsunaga and M. Asada. How a mobile robot selects landmarks to make a decision based on an information criterion. *Autonomous Robots*, 21(1):3–14, 2006.
13. P. Murphy, A. Torralba, and W. T. Freeman. Using the forest to see the trees: a graphical model relating features, objects and scenes. In *Proc. of the 16th Conf. on Advances in Neural Information Processing Systems, NIPS*, 2003.
14. V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimal object detection. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2049–2056, 2006.
15. A. Oliva and A. Torralba. The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12):520–527.
16. J. Pineau, G. Gordon, and S. Thrun. Anytime point-based approximations for large POMDPs. *Journal of Artificial Intelligence Research*, 27:335–380, 2006.
17. A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *Proc. of Int. Conf. on Computer Vision (ICCV-07)*, pages 1–8, 2007.
18. N. Roy and G. Gordon. Exponential family PCA for belief compression in POMDPs. In *Advances in Neural Information Processing 15 (NIPS)*, pages 1043–1049, 2002.

19. N. Roy and S. Thrun. Coastal navigation with mobile robots. In *Proc. of Advances in Neural Processing Systems (NIPS)*, volume 12, pages 1043–1049, 1999.
20. C. Siagian and L. Itti. Biologically-inspired robotics vision Monte-Carlo localization in the outdoor environment. In *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS*, 2007.
21. Ç. Soyer, H. I. Bozma, and Y. I Stefanopoulos. Apes: Attentively perceiving robot. *Autonomous Robots*, (20):61–80, 2006.
22. C. Stachniss, G. Grisetti, and W. Burgard. Information gain-based exploration using rao-blackwellized particle filters. In *Proc. of Robotics: Science and Systems (RSS)*, 2005.
23. Y. Sun and R. Fisher. Object-based visual attention for computer vision. *Artificial Intelligence*, 146:77–123, 2003.
24. S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. Cambridge University Press, New York, 2006.
25. J. K. Tsotsos, S. Culhane, W. Wai, Y. Lai, N. Davis, and F. Nufo. Modeling visual-attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–545, 1995.