Sensor management in mobile robotics

Mikko Lauri

Abstract—My forthcoming thesis work considers the active control of a information gathering process, known as sensor management or active sensing. Methods from sequential decision-making under uncertainty are applied to find control policies that maximize the informativeness of measurement data. The main application area considered is mobile robots, which are sensor platforms with operational degrees of freedom.

I. INTRODUCTION

Mobile robots may be viewed as sensor platforms with operational degrees of freedom; either by being able to move or by carrying sensors that have multiple possible operating modes. As availability of sensor data has increased, questions on how to manage the collection of the data have become increasingly important. Where and when should a robot move to best identify some spatio-temporal process of interest? Where should a sensor such as a camera focus its attention to best help achieve an operational goal? Analysing and answering such questions has the potential to reduce energy consumption, e.g. by reducing the distance a mobile sensor platform travels, and to improve efficiency of executing tasks by reducing the required time, while simultaneously maintaining or even improving the quality of information obtained. The aim of my thesis work is to provide answers to such questions.

In the following sections, we review related work, present an overview of our current results and conclude with a brief description of future work.

II. RELATED WORK

The active control of a sensing process is referred to as sensor management or active sensing. Sensor management is typically formulated as a sequential decision-making problem under uncertainty [1]. The objective is to find a control policy that maximizes the expected utility of future measurement data over a specified horizon of time, possibly infinite. The general framework handling uncertainty in both action effects and observations is the partially observable Markov decision process, or POMDP.

The state $s \in S$ of the system is hidden and may only be observed via observations $z \in Z$. The state and the observation process may be affected by sequentially executing actions $a \in A$. A stochastic state transition model $T : S \times A \times S \rightarrow [0, 1]$ describes the effect of actions on the state, and a probabilistic observation model $O : Z \times S \times A \rightarrow [0, 1]$ describes the interdependence between states, actions and observations. In a Markovian system, a belief state or state estimate which summarizes information about the state given the history of actions and observations is maintained by sequential Bayesian filtering. The belief state $b \in B$ is a probability density function over $S$. The new belief state which results after executing action $a$ in belief state $b$ and then perceiving observation $z$ is denoted $b' = \tau(b, a, z)$.

A bounded reward function $\rho : B \times A \rightarrow \mathbb{R}$ gives the immediate reward when action $a$ is executed in belief state $b$. An optimal value function $V^*_t : B \rightarrow \mathbb{R}$ gives the maximum expected total reward when $t$ decisions are remaining. Dynamic programming is applied to recursively compute the value functions for any finite horizon as follows:

$$Q_t(b, a) = \rho(b, a) + \mathbb{E} \left[ V^*_{t+1}(\tau(b, a, z)) \right]$$

$$V^*_t(b) = \max_{a \in A} Q_t(b, a),$$

starting with $Q_1(b, a) = \rho(b, a)$, taking the expectation under the prior probability $p(z | b, a)$ of the observations. The optimal action at belief state $b$ when $t$ decisions are remaining until end of the horizon is given by $\arg\max_{a \in A} Q_t(b, a)$.

Classical approaches to solving POMDPs consider reward functions linear in $B$, defined as the expectation $\rho(h, a) = \mathbb{E} \left[ R(s, a) \right]$ of a purely state- and action-dependent reward function $R$ under $b$. This leads to a finite-horizon value functions that are piecewise linear and convex (PWLC), thus having a finite representation as a convex hull of a set of so-called $\alpha$-vectors [2]. Although this representation has lead to a great number of efficient exact and approximate (see e.g. [3], [4]) solution algorithms, the requirement for reward functions is restrictive from a sensor management point of view.

The goal of many sensor management problems is to minimize the uncertainty related to the state estimate $b$. For instance, if the sensor manager cannot affect the state, this leads to an uncontrolled state transition model $T : S \times S \rightarrow [0, 1]$. In such a case the usefulness of a state-dependent reward function $R(s, a)$ is dubious. Instead, in a sensor management problem quantities such as the variance or information entropy of the state estimate, or the expected informativeness of choosing a particular sensing action are more appropriate descriptions of the objective. Such a reward function $\rho(b, a)$ is non-linear in $B$, the PWLC property does not hold, and application of most of the powerful approximation algorithms developed for POMDPs are ruled out.

Feasible methods for solving POMDPs with general reward functions $\rho(b, a)$ include point-based methods not employing the PWLC representation [3], and online forward tree search methods based e.g. on branch-and-bound pruning [4].
or applying Monte Carlo simulations to evaluate the rewards [5]. Some special methods for sensor management problems have also been studied, e.g. [6], however limited to problems with a small finite number of states (in the order of 10). For $\rho(b, a)$ which is a convex function of $b$, it was shown in [7] that standard methods may be applied with bounded error.

Problems in mobile robotics sometimes have properties that may be exploited to solve sequential decision-making tasks more efficiently. Such properties include e.g. mixed observability, where part of the state space is completely observable. This was exploited in [8] to solve robotic navigation problems.

III. CURRENT RESULTS

The envisaged contribution of my forthcoming thesis is twofold. First, it aims to contribute to theoretical knowledge on solving POMDPs where the instantaneous reward is given by a general belief-dependent function. The emphasis is on studying properties such as mixed observability that are found in mobile robotics domains. Secondly, the aim is to study the applicability of the developed and currently known methods for solving sensor management problems formulated as POMDPs. Both belief-dependent and state-dependent reward functions have been considered in this part. In particular, applications in mobile robotics with real-world demonstrations are considered.

A. Theoretical results

In [9] we studied a mixed observability POMDP, where the reward function was the mutual information (MI) between the state and observation. The objective was to find a sensing policy maximizing the expected cumulative MI. The action space in the problem was constrained such that only local action were possible — some sensing actions were only possible after a sequence of such local actions. We showed that when the action constraints are relaxed, the problem is equivalent to a multi-armed bandit (MAB) problem. The optimal solution of an MAB is an index policy, typically easier to compute than dynamic programming solutions to POMDPs. A comparison with a state-of-the-art Monte Carlo planning method [5] showed competitive performance and better worst case behaviour.

B. Applications

We studied automated robotic exploration in [10], [11], with the objective of maximizing MI over a finite horizon. Monte Carlo methods from [5] and [12] were applied to solve an open loop approximation of the underlying POMDP problem. Preliminary work [13], [14] applied a similar solution method, although the reward function there was state-dependent.

A new approximation for MI especially suitable for robotic exploration problems was derived. Compared to the state-of-the-art, we were able to handle general models for the environment and robot state with non-Gaussian noise. Combining the open loop approximation with the receding horizon control principle was shown to improve performance over earlier myopic (one-step) exploration approaches. Performance was studied in both real-world and simulation domains.

Earlier work [15] studied sensor management problems in a continuous-state POMDP model with a state-dependent reward. A choice between a low-noise and high-noise sensor had to be made at each time step. The formulation was found feasible, although numerical stability when the difference in value between the measurement options was small was noted as a possible problem.

IV. FUTURE WORK

Remaining future work includes studying whether the results of [9] may be expanded further. Additionally, a comparison of existing POMDP solvers and their applicability to problems with a belief-dependent reward will be considered in form of a literature survey.

REFERENCES