From Reactive to Active Sensing: a Survey on Information Gathering in Decision-Theoretic Planning

TIAGO VEIGA*, Department of Computer Science, Norwegian University of Science and Technology, Norway

JENNIFER RENOUX*, Center for Applied Autonomous Sensor Systems, Örebro University, Sweden

In traditional decision-theoretic planning, information gathering is a means to a goal. The agent receives information about its environment (state or observation) and uses it as a way to optimize a state-based reward function. Recent works, however, have focused on application domains in which information gathering is not only the mean but the goal itself. The agent must optimize its knowledge of the environment. However, traditional Markov-based decision-theoretic models cannot account for rewarding the agent based on its knowledge, which leads to the development of many approaches to overcome this limitation. We survey recent approaches for using decision-theoretic models in information-gathering scenarios, highlighting common practices and existing generic models and show that existing methods can be categorized into three classes: *reactive sensing*, *single-agent active sensing*, and *multi-agent active sensing*. Finally, we highlight potential research gaps and suggest directions for future research.

CCS Concepts: • **Computing methodologies** → **Planning under uncertainty**; Multi-agent systems.

Additional Key Words and Phrases: decision-theoretic planning, information gathering, active sensing

1 INTRODUCTION

Sequential decision-making under uncertainty, i.e., the problem of optimizing a sequence of actions given a stochastic or partially observable environment, includes an information-gathering component inherently. The agent acts based on its world model and observes the result of its actions to decide on the next one. In most scenarios, usually modeled with decision-theoretic models, information gathering is a by-product of a reward defined on states (e.g., using the appropriate treatment, reaching a certain point in space, finding a specific object). However, some scenarios consider information gathering as their primary goal, for instance, target tracking or patrolling applications. This sub-field of decision-theoretic planning has gathered the attention of several researchers in the past years. These are the scenarios we consider in this paper, which follows from identifying the need to compile the current state of the art for this topic.

Therefore, in this survey, we analyze the state of the art about decision-theoretic approaches for information gathering. We observe that many branches in the literature claim to do information gathering with many different approaches, some more empirical and others more comprehensive.

We categorize the literature into three main branches, which follow different levels of systematization, agents' self-awareness, and capability of analyzing the environment and other actors' intentions. We term these categories as *reactive sensing*, *single-agent active sensing*, and *multi-agent active sensing*, which separate agents which perform information gathering without explicitly reasoning about their internal knowledge state from agents which

---

*Both authors contributed equally to this research.

Authors' addresses: Tiago Veiga, tiago.veiga@ntnu.no, Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway; Jennifer Renoux, Center for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden, jennifer.renoux@oru.se.

decide based on this knowledge state, and finally agents which incorporate in their knowledge state information about the other actors in the environment. After analyzing these categories of approaches, we discuss our main findings while reviewing those articles. Among others, we argue that the current state of the art points out that decision-making processes that explicitly depend on knowledge are more generalizable and make it possible to target more types of problems. However, there is a lack of literature on real applications using these general methods.

1.1 Contributions

This paper aims to bring a renewed overview of the information gathering and epistemic planning frameworks with decision-theoretic planning methods. The main contributions of this paper are:

- A principled formulation of the relation between information gathering and epistemic planning, where we identify the intersection between epistemic planning and information gathering fields.
- A comprehensive literature review on sequential decision-making with information-gathering goals, where we propose a grouping between different approaches within each sub-field we identified.
- The identification of future research steps in this topic, based on our literature review analysis.

1.2 Problem definition

The term *information gathering* usually refers to problems in which collecting information about the environment is the goal of the system in itself. Over the years, many methods have been developed for information gathering, which differ in how they model the knowledge acquisition process and the kind of reasoning the planning agent is capable of.

Active sensing has been first defined in [8] as "the problem of an intelligent data acquisition process", more specifically a "problem of controlling strategies applied to the data acquisition process which will depend on the current state of the data interpretation and the goal or the task of the process". Even though it initially focused on computer vision systems, this definition still holds for other active perception systems developed over the years. Bajcsy et al. [9] later revisited this definition, emphasizing that the most essential aspect of the *active perceiver* compared to other artificial agents is that the active perceiver *knows* why it is sensing, *chooses* what to perceive and determine when, where, and how to perceive. This specificity implies that the active perceiver is not only able to decide how to act to perceive better but also when and how *not* to act, for instance, if it infers that the action is not worth the information gain. By being able to reason over its own knowledge and plan for the best way to acquire new knowledge, the active perceiver is inherently performing *Epistemic Planning*.

The term *Epistemic Planning* has been first introduced in [14] as the approach that allows planning on *epistemic models* and specifically using *Dynamic Epistemic Logic* (DEL). DEL considers lack of knowledge as an input for reasoning, the same way knowledge is. It also considers that partial observability and lack of knowledge are related but different problems. Therefore, the DEL-based epistemic planning allows for actions (also called *event models*) to be either factual (i.e., changing the world itself, also referred to as *ontic*), epistemic (changing the agent's knowledge about the world but not the world itself), or a combination of thereof.

More recently, the meaning of the term *epistemic planning* has broadened to encompass any approach that allows planning with epistemic states, actions, and goals [10]. In its broader sense, the epistemic planning problem encompasses the active sensing problem but is not limited to it. Indeed, epistemic planning can be performed without the goal of information gathering, e.g., to coordinate with a team or influence another agent's knowledge state. However, any problem that reason on knowledge acquisition, i.e. active sensing, is, in essence, an epistemic planning problem.

However, when referring to information-gathering problems, it is not strictly necessary that the agent is an active perceiver and can decide not to act. In many cases, especially in patrolling and surveillance problems, the

agent's model is designed so that the agent can reach a desirable level of knowledge, but without any reasoning on the impact of the sensing action per se. We will refer to these models in the remainder of this paper as *reactive sensing*. We define reactive sensing as the process by which an agent gathers information without explicitly optimizing for information gain but by reacting to external stimuli (e.g., being in the same location as a moving target).

Figure 1 summarizes the relationship between these different concepts. We claim that active sensing combines epistemic planning in its broadest sense and information gathering. Reactive sensing is a set of methods to perform information gathering which does not optimize for information gain per se.
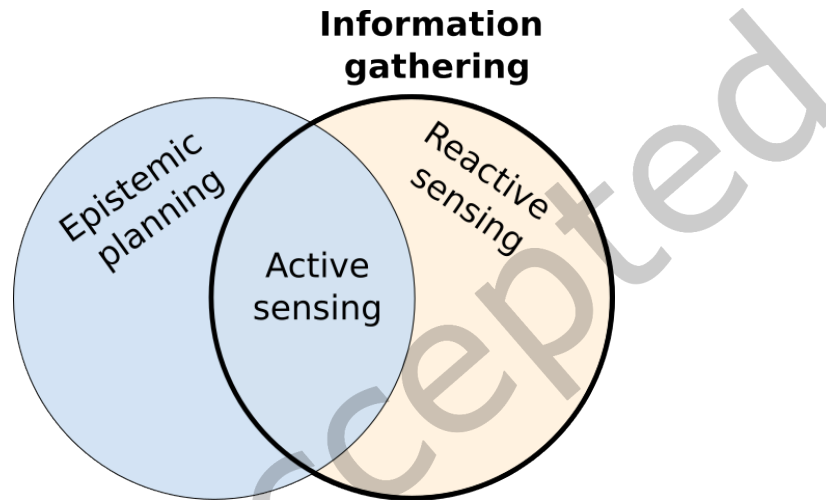

Fig. 1. Information gathering can be performed through active or reactive sensing. Active sensing uses epistemic planning approaches to optimize for information gain.

1.3 Illustrative Example: Target Tracking

Throughout this survey, we will use a target tracking scenario as an example to illustrate the concepts. This is a generic but common scenario found as a testing environment for many planning methods. In this scenario, depicted in Figure 2, a mobile agent must patrol a given scenario environment containing several nodes to track a target. In this context, tracking might have different meanings, such as triggering an alarm whether or not a target (e.g., an intruder) is found in the environment or being able to follow the target accurately, given its uncertain movements. These differences might also influence the planning method and will be further discussed throughout the paper.

We first consider a simple scenario with one agent and one target and will extend this scenario in Section 5 to multi-agent problems.

1.4 Methodology

We build the set of papers we surveyed iteratively using specialized search tools (Google Scholar, Semantic Scholar, Connected Papers) and combinations of the following keywords: information-gathering, (partially observable) Markov decision processes (including their acronyms), decision-theoretic planning, active sensing, multi-agent systems. We considered three inclusion criteria :
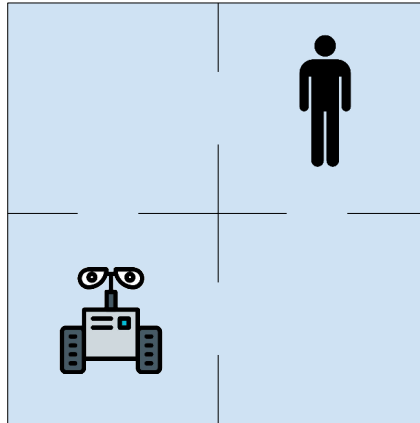
Fig. 2. Illustration of the *Target Tracking* toy problem example. This scenario models an autonomous robot that must patrol a discretized environment by moving freely between the different rooms and is rewarded for correctly identifying the position of the human target.

(1) **Information-gathering as a goal:** Many papers consider information-gathering actions as a way to improve the planning agent's knowledge about its environment, e.g., in [49]. However, the main goal of these systems is still to reach a given state, and information gathering is only a means to reach such a state. In the current review, we only included papers in which information-gathering is the main problem being addressed.
(2) **Decision-theoretic formalization:** We only included papers that formulated their information-gathering problem as a decision-theoretic problem (regardless of how the problem is solved)
(3) **Peer-reviewed:** We only included papers that have been accepted in a peer-reviewed venue before the 1st of April 2022.

These criteria resulted in 51 collected papers, which we classified into three categories: reactive sensing (Section 3), single-agent active sensing (Section 4), and multi-agent active sensing (Section 5). We decided to separate single and multi-agent active sensing as the multi-agent aspect brings additional challenges that ought to be considered separately and will be reviewed in Section 5. Table 1 presents a summary of how the papers in this survey have been organized.

	Reactive Sensing	Active Sensing
Single Agent	[34] [32] [33] [36] [35] [7] [31] [70] [73] [72] [39] [51] [53] [21] [50] [67] [46] [25] [26] [4] [61]	[65] [3] [74] [64] [18] [23] [63] [17] [76] [24] [29] [77] [41] [19] [55] [52] [16] [20] [78] (application) [79] (application) [80] (application)
Multi-Agent	[38]	[22] [58] [59] [60] [42] [44] [45] [43] [12] [13] [68]

Table 1. Summary of the classification of each surveyed paper

1.5 Outline

The remainder of this paper will be organized as follows. Section 2 will present some of the required background knowledge on decision-theoretic planning. Then Section 3 will give our definition of *reactive sensing* and review papers falling in this category. Section 4 will review papers proposing models to perform *active sensing*. Section 5 will focus on models specifically designed to tackle *multi-agent aspects*. Finally, Section 6 will summarize and discuss our findings and propose leads for future research areas.

2 BACKGROUND ON DECISION-THEORETIC PLANNING

This section will review some of the basics of decision-theoretic planning, limited to the concepts needed to understand the remainder of this paper. For a more detailed introduction to decision-theoretic planning, we refer the reader to [40].

Decision-theoretic planning can be defined as "an extension of the classical AI planning paradigm because it allows one to model problems in which actions have uncertain effects, the decision maker has incomplete information about the world" [15]. It roots in decision theory to weigh the value of different courses of action and extends toward the problem of constructing plans to optimize those outcomes. A general model that captures this problem is given by Markov Decision Processes (MDP) and their extension to state uncertainty, Partially Observable Markov Decision Processes (POMDP) [37, 71]. Here, we provide an overview of decision-theoretic planning under uncertainty in stochastic environments modeled as a POMDP.

2.1 Definition

A POMDP models the interaction of an agent with a stochastic and partially observable environment, and it provides a rich mathematical framework for acting optimally in such environments. It can be represented by a tuple $\langle S, A, O, T, \Omega, R, h, \gamma \rangle$. At any time step, the environment is in a state $s \in S$, the agent takes an action $a \in A$ and receives a reward $R(s, a)$ from the environment as a result of this action, while the environment switches to a new state s' according to a known stochastic transition model $T : p(s'|s, a)$. After transitioning to a new state, the agent perceives an observation $o \in O$, that may be conditional on its action, which provides information about the state s' through a known stochastic observation model $\Omega : p(o|s', a)$. The agent's task is defined by the reward it receives each time step t, and its goal is to maximize its expected long-term reward $E[\sum_{t=0}^{h} \gamma^t R(s_t, a_t)]$, where h is the planning horizon, and γ is a discount rate, $0 \le \gamma < 1$.

Given the transition and observation model, the POMDP can be transformed to a belief-state MDP: the agent summarizes all information about its past using a belief vector $b(s)$. The initial state of the system is drawn from the initial belief b_0, and every time the agent takes an action a and observes o, its belief is updated through Bayes' rule:

$$b^{ao}(s') = \frac{p(o|s', a)}{p(o|a, b)} \sum_{s \in S} p(s'|s, a)b(s), \tag{1}$$

where

$$p(o|a, b) = \sum_{s' \in S} p(o|s', a) \sum_{s \in S} p(s'|s, a)b(s) \tag{2}$$

is a normalizing constant.

2.2 Solutions

Solving a POMDP means finding the best possible policy that maps the belief state to agent actions. Solving POMDPs optimally is hard, and thus algorithms that compute approximate solutions are often used [34, 56, 75].

There are two approaches to compute policies, dubbed offline and online solving. As the names indicate, offline solving pre-computes the policy given a description of the world model [66]. Then, at runtime, looking up the

best action for the current belief is efficient. These methods typically exploit the fact that POMDP solutions are piecewise linear and convex (PWLC), allowing for their compact representation, but are limited to linear rewards over the belief space. On the other hand, online methods attempt to compute a good local policy for the current belief state by interleaving planning and execution [62]. Generally, they use heuristics to expand a tree of reachable belief states and compute the best possible action path down the tree. Online methods do not have restrictions on the form of rewards, but heuristics are often derived from offline approximations, which have this restriction.

2.3 Multi Agent Planning

In multi-agent settings, the model chosen to represent the agents in the system will depend on the relation between the agents in the system. The two most extreme types of relations are fully-cooperative and fully-competitive. In fully-cooperative settings, the agents share the same reward function and work for a system-wide goal. In fully competitive settings, an agent can only win when another loses. Between these two extremes, agents may share some goals but also have personal objectives that may or may not compete with other agents' objectives. Many extensions of existing decision-theoretic models have been proposed to tackle the multi-agent planning problem, such as the Decentralized POMDP [11] (Dec-POMDP), the Multi-agent Team Decision Problem (MTDP)[57], or the Interactive POMDP (I-POMDP)[28].

3 REACTIVE SENSING

In this section, we will tackle the literature review on what we named *reactive sensing*. Roughly, this includes a broad set of approaches that tackle the problem of finding information about some feature in the agent's environment but without directly tackling information measures. We claim this set of methods to be reactive, in the sense that the agent's decisions are not actively directed towards information gain but, instead, guided as reactions to external stimuli. We start by framing an example in our target tracking problem to illustrate the type of behavior exhibited by these methods and then classify them in the literature review in five different sub-categories based on similarity.

3.1 Illustrative Example

Figure 3 illustrates the reactive sensing strategy in the target tracking example. In this approach, the agent must obtain some information from the environment (the position of the target) but is driven by physical events, namely, being in the same node as the target at the same time (or detecting the target in its field of view). The side effect of this approach is that negative past observations are not as valuable as they could be for the agent. For instance, even if the agent has visited all but one node in the system and theoretically "knows" that the target is in the non visited one, it is not rewarded for this knowledge and must visit the last node to receive its reward.

3.2 Literature Review

Our literature analysis revealed that works dealing with reactive sensing problems present some commonalities in how they formulate and solve the problem. Based on these commonalities, we identified five strategies used to tackle reactive sensing problems: (1) the target tracking strategy, (2) the classification action strategy, (3) the external information metrics strategy, (4) the sensing actions strategy, and (5) the oracular observation providers strategy . Subsections 3.2.1 to 3.2.5 present each of these strategies. Each section starts with a generic presentation of the strategy, with mathematical formulations when applicable, followed by a review of the works that follow this strategy.

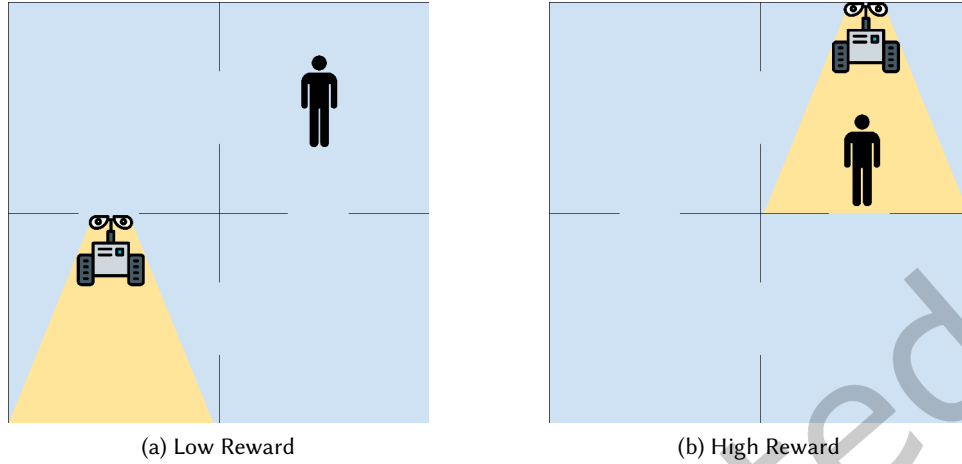(a) Low Reward   (b) High Reward

Fig. 3. Illustrative *Target Tracking* toy problem example in a reactive sensing setting. Here, the objective is formulated in terms of the current physical situation. I.e., the robot reacts to the current observation available. It gets a low reward when the target is not directly visible, subfigure 3a, and a high reward otherwise, subfigure 3b.

3.2.1 Target tracking. One strategy often taken to gather information is to reward the agent when the gatherer (or gatherer's field of view) coincides with the "target"'s position. This is the strategy usually chosen by patrolling and target tracking applications. The reward function for such cases generally looks like the following:

$$R(s) = \begin{cases} a & \text{if } X = Y \\ b & \text{if } X \neq Y \end{cases} \tag{3}$$

where X can be the position of the agent and Y a generic representation of the target's position or any observable feature, and $b \leq 0 < a$. Note that, in this case, the planning solver weights the reward function with the current belief to determine the current belief-based reward:

$$R(b, a) = \sum_{s \in S} R(s, a)b(s) \tag{4}$$

In practice, this means that an agent will try uncertainty reduction measures if that helps perform a given task, but its goals are formulated to achieve a determined world configuration and not with respect to its internal knowledge state.

A typical scenario in which perception plays an important role is target tracking, either with mobile agents or sensor networks. The objective formulation, in this case, can be formulated by giving a fixed reward when two desired values of state variables collide or a fixed cost otherwise. Then, this triggers the behavior of trying to find the information of interest in the environment to receive a positive reward. This formulation was implemented in [34], which uses a target tracking scenario as the use case to test a POMDP solver. Here, the agent is awarded a positive reward for being in the same cell as the target, and the only source of uncertainty is the non-deterministic target movement. In the same spirit, a few authors exploited decision theory to extend classic classifiers with a Bayesian cost formulation. Hitchings and Castañón [32, 33] tackle the problem of, for each of a set of available sensors, selecting the location to observe and the sensor mode. Deriving from a control-theoretic formulation but decoupling into several POMDPs, the objective function is obtained from a Bayes cost formulation, i.e., the cost

of selecting an estimate v_i when the actual state is x_i. A similar approach is found in [36]. Also, Ji and Carin [35] implements a classification system (i.e., given a set of classes and feature acquisition actions, declare which class is the real one for the problem) with a POMDP-like structure. In [7], a sensor scheduling problem is formulated as a POMDP. There is an energy cost for sensor usage and a fixed tracking cost for each time unit that the target is not observed. Kartal et al. [38] presents an extension of MCTS for multi-agent continuous patrolling of an environment. The reward is defined as a positive reward if one member of the patrolling team is in the same location as the intruder.

3.2.2 Classification actions.
Historically, classification rewards can be seen as an ancient version of active sensing models. The agent still does not actively reason about its knowledge state but indirectly through an internal variable that represents the current knowledge of the agent. This is implemented by extending the world model with classifying actions C, which declare that a given state variable is of a given type. When the classification is correct, the reward is positive, and if it is incorrect, then the reward is negative. Additionally, the agent has the option to not classify in case its knowledge is not sufficient for such a decision. Formally, the reward function takes the following form:

$$R(s, a_c^{X=y}) = \begin{cases} a & \text{if } X = y \, \cap \, C = yes \\ b & \text{if } X \neq y \cap \, C = yes \\ c & \text{if } C = no \end{cases} \qquad (5)$$

where X is the state variable of interest, y any of its possible values, $a_c^{X=y}$ is a binary classifying action declaring that the environment feature X takes the value y, and $b \leq 0 \leq c < a$.

This subset of methods is arguably borderline between reactive and active sensing in that they are designed with the information-gathering aspect at their core. In practice, the goal is to maximize the information available to the agent. However, we decided to include them in the reactive sensing section of our analysis for two main reasons. First, the reward system is still constructed on top of state-based rewards with an extra bookkeeping variable to force the model to improve its knowledge level. Second, the model designer defines the reward levels empirically with no guiding rules on how to design them. This contrasts with the methods presented later under the active sensing umbrella, in Section 4, which directly reward the level of information and have guidelines for designing rewards according to the minimal threshold of knowledge level desired by the system designer.

Guo [31] presents the first formulation for this approach and sets the equivalence of the problem modeled as a Bayesian network classifier and a POMDP. Then, it describes a POMDP framework for information gathering in which the actions use a particular sensor when enough information has been gathered, outputting a particular classification label. Then, Spaan [70] and Spaan et al. [73] generalize this idea for factored models, which allows focusing on specific features of the environment, and present a practical implementation with POMDPs in surveillance problems with networks of cameras.

3.2.3 External information metrics.
A different approach within this strategy is to reflect external information metrics in the state-based reward function. For instance, Spaan and Lima [72] models a dynamic sensor selection problem as a POMDP to select a subset of cameras to observe a given area. Here, the reward is obtained through the covariance matrix of observing a spot through a selection of cameras. Although formulated as an MDP and therefore not incorporating any uncertainty tracking, Kent and Chernova [39] tackles the problem of visual coverage of targets, in which the objective is formulated as a measure of the coverage of a region of interest by the robot's field of view, with added costs for collisions with humans, the degree of intrusion in the human's space and power consumption. Liu and Williams [51] mixes classification actions with costs based on covariances in a sensor selection setting, exploiting the cost function's submodular properties. This approach is also followed with reinforcement learning, given the difficulty of explicitly maintaining internal information measures. Murad et al.

[53] implements a reinforcement-learning-based method to decide on sensing actions in which the state-based reward is based on information from an external inference process. Eck and Soh [21] uses reinforcement learning to compute policies for Observer Effect POMDPs, which balance the use of sensing resources that deteriorate over usage with rewards based on externally measured knowledge refinement. Linke et al. [50] surveys several other approaches under this topic.

3.2.4 Sensing actions. The advance of factored models allowed the exploitation of statistical independence between model variables, either state or action variables. This allows a better, easier design of world models and the development of more efficient optimizers [67]. Factored state spaces allow algorithms to focus on features of interest in the environment and, therefore, the system designer to specify which of these features the perception system should highlight. In addition, pure sensing actions do not physically affect the environment but influence the agent's knowledge state. Thus, the actions space can be redefined as $A = A_d \times A_s$, where A_d is the set of non-sensing actions that affect the environment and A_s is the set of sensing actions that do not have an effect on the environment but affect how the agent observes it. Lauri and Ritala [46] exploits this to define sets of sensing and non-sensing actions, which allows the agent to select which sensor to use at each timestep, independently of the actions which interact with the environment. This idea will be explored later by more systematic approaches to information gathering. Ghasemi and Topcu [25, 26] systematize this idea into an online algorithm and an extension of point-based value iteration methods, respectively. The main point is to extend the regular POMDP decision loop with a greedy perception scheme that chooses the best perception action for that timestep.

3.2.5 Oracular observation providers. Similar to the inclusions of sensing and non-sensing sets of actions, we find a line of work that extends the POMDP model with sets of *oracular* actions, which extends the actions space with a query action. Therefore, $A = A_d \cup a_{ask}$, where A_d is the previous set of actions available to the agent and a_{ask} the query action, allowing the agent to ask an external observer, typically a human, about the true nature of a given feature of the environment [4]. This decision can be based on the uncertainty levels [61], but the proposed formulation resorts to solving the underlying MDP and then, at execution time, adding the current value of information and the possible gain for asking. This results in a myopic policy that does not directly consider the effects of the agent's actions on the value of its information.

4 SINGLE-AGENT ACTIVE SENSING

Contrasting to the approaches summarized in the previous section, we define *Active Sensing* as the category of methods which actively reason about the features of interest in the environment and proactively take actions to improve the level of information about them. Although they might have similarities in spirit, there are key differences in the inner motivations for the agent's actions. An illustrative example is, for instance, the value of negative observations with which an agent can infer that, by not observing something, there is a higher chance somewhere else. Reactive approaches, in this case, are not able to make such inferences. Consequently, active methods tend to be more generalizable as they focus less on specific problems but rather on the general problem of how to plan for information gain.

4.1 Illustrative Example

We now go back to the illustrative target tracking example and describe what an active sensing approach looks like, illustrated by Figure 4. Again, the general goal of the agent is to detect the target in the environment. Unlike in the information-gathering approach, the agent incorporates previous observations to generate a (typically probabilistic) state of knowledge that also learns from negative observations. For instance, let us say that the agent maintains a probabilistic distribution of the possible target locations. It starts as a uniform distribution (Figure 4a) and is updated based on the output of the observations of the agent and its (known a priori) probabilities.
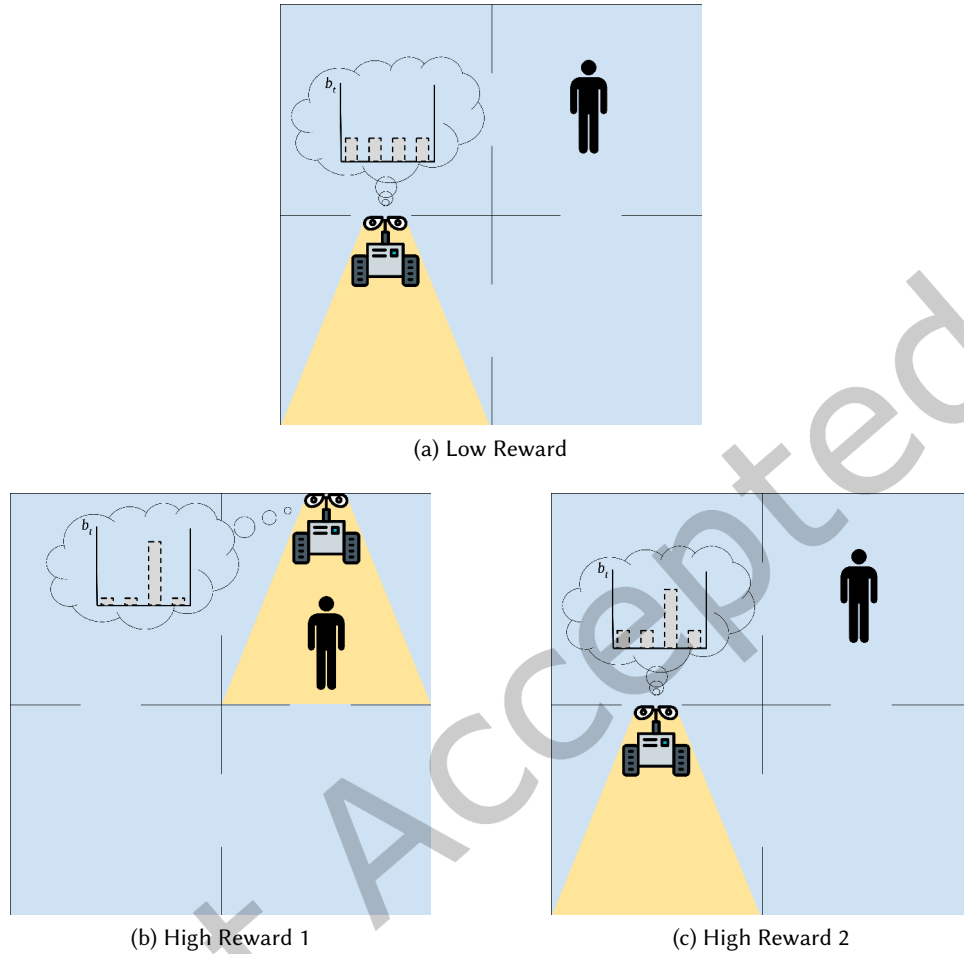
(a) Low Reward

(b) High Reward 1

(c) High Reward 2

Fig. 4. Illustrative *Target Tracking* toy problem example for active sensing, where b_t is the belief (or knowledge state) of the agent at a given timestep t. Here, the reward is based on the probabilistic information of the target's possible location. The agent is rewarded for having low-uncertain information about the target's position. This implies a low reward when the information is ambiguous (Subfigure 4a). It also implies that the agent can get a high reward when the target is directly visible (Subfigure 4b) but also when the history of observations leads to a highly accurate belief even when never receiving a positive observation (Subfigure 4c).

A negative observation at a given node will reduce the probability of the target being there and increase the probability of being somewhere else. Therefore, there are now two different ways that the agent can localize the target. First, it can directly observe the target (Figure 4b) with a search that can be guided by the higher probability level for different nodes. Alternatively, the agent might infer the target's position without directly observing it but rather collecting a set of negative observations at all other locations (Figure 4c).

4.2 Literature Review

The most straightforward and studied way to switch from reactive to active sensing is to change the reward function of decision-theoretic models to reward the agent on a belief state instead of an environmental state. Such a belief-based reward, usually noted $\rho(b,a)$[1], may introduce information quantifying measures such as in [65], which seems to be the first formal proposal in the direction of active sensing, with a scenario of choosing sensor modes according to an entropy-based measure. Commonly used information quantifying measures are the entropy of the belief, Eq.6 or the distance to the center of the belief simplex, Eq.7.

$$\rho(b,a) = \begin{cases} log_s(|S|) + \displaystyle\sum_{s \in S} b(s)log_2(b(s)) & \text{if entropy} \\ \|b - c\|_p & \text{if distance to center simplex} \end{cases} \tag{6} \tag{7}$$

This approach is followed by the first general model dedicated to information gathering: the ρ-POMDP[3]. The ρ-POMDP is a modified POMDP in which the state-based reward function $R(S, a)$ is replaced by a general belief-based reward $\rho(b, a)$.

However, switching from a state-based reward to a belief-based reward brings many challenges in solving the resulting model. Indeed, much of the literature on POMDP solving relies on the PWLC property of the value function, which in turn, depends on a state-based formulation of the reward function. In the remainder of this section, we will explore how contributions in active sensing dealt with this issue.

4.2.1 Linear rewards. Early active sensing works, like the one already mentioned [3, 65] rely on linearizations of convex uncertainty functions (entropy and distance to the center of the belief simplex). They note that the solving methods existing at the time, in particular Incremental Pruning, require piecewise linear rewards and, therefore, propose a linearization with tangents of a non-linear reward function.

Araya-López et al. [3] also replace the classical reward representation with a set of vectors representing a PWLC reward function and show that if these vectors are tangents of convex information measures, then the error in the value function is bounded relative to using the true reward function. The alternative POMDP-IR approach [74] extends the action space such that, at each decision step, the agent decides on a so-called domain action and predictions actions. Domain actions physically affect the environment and influence transitions and observations. In contrast, prediction actions only affect the reward, such that a positive reward is given if the predictions match the real value of the state features or negative otherwise. The reward function is still piecewise linear, but the balance between these predictions rewards will is such that the agent only gets a positive reward if the belief for that particular state feature is above a pre-defined threshold, giving the system designer the possibility to define priorities between gathering information on different features. Later, Satsangi et al. [64] proved the equivalence between both formulations such that a problem formulated as a ρPOMDP can be reduced to a POMDP-IR and vice-versa. Dressel and Kochenderfer [18] extends another point-based POMDP solver, SARSOP, to use belief-based rewards. These belief-based rewards are linearizations of information measures to avoid losing the PWLC property. Fehr et al. [23] claims that there are situations in which the desired information-based reward is non-convex and replace the PWLC requirement of the reward function to Lipschitz continuity, demonstrating that, for finite horizons, Lipschitz reward functions lead to Lipschitz continuous optimal value function. In practice, the approach is also to linearize the reward function using the Lipschitz continuity properties, and, therefore, it can still be solved with adapted point-based methods. An extension of the information reward concept to reinforcement learning is proposed in [63], which introduced deep anticipatory networks that drive the agent to take actions that help it predict the true value of an unknown variable.

---

[1]Notation may differ, from using $\rho(b, a)$ to represent any function or to represent only non-linear function such as entropy.

4.2.2 Non-linear rewards. Alternative approaches attempt to circumvent the PWLC requirement. For instance, Chong et al. [17] discusses POMDPs for adaptive sensing with references to information gain metrics as the rewards with approximations to Q function computation which do not restrict the reward function, although with limitations. One common approach is the use of online methods. This class of algorithms constructs a tree with possible future outcomes with the current belief as the root. It does not rely on an explicit closed-form representation of the value function but on forward simulations from the current situation. Therefore, in theory, any kind of reward is acceptable in this approach. However, there is a distinction between the case when the full belief is used and propagated down the tree or when particle filter approaches are used to approximate the belief state. In the latter, at any stage, one can compute any belief-based reward at the cost of computing belief updates for every node. Arora et al. [5, 6] and Lauri and Ritala [48] use this approach in the field of autonomous robots, one to tackle the problem of autonomously deciding whether observed events are of scientific interest when exploring remote environments while the other maximizes the amount of information collected about the environment. Furthermore, these approaches used mutual information between state and observation as the reward function, which roughly indicates the information gain between current and future knowledge conditioned on the received observation. Similar approaches are found in [54] and [47] with, respectively, greedy search algorithms for teams of agents and planning for inference of random variables by relaxing the problem to a multi-armed bandit equivalent. When computing exact beliefs is not feasible or is costly, approximations such as POMCP approximate a search in the belief space by performing many simulations in the state space, sampling the root particle from the root belief. Usually, the tree construction process is guided by meaningless rollouts for information-oriented problems, as one cannot estimate the best paths down the tree with a few simulations. Expanding on this concept, the ρPOMCP [76] extends the POMCP algorithm by generating a set of additional particles, kept in parallel and propagated down the tree such that it offers an initial estimate for the belief (and subsequently, a measure of the information quality) as a node is initialized. A similar approach is followed in [24] for continuous state and observation domains. Similar to the approach of defining parallel sensing actions, Greigarn et al. [29] presents an approach to compute the best sensing actions with information-based rewards. It proposes a myopic, online procedure that, at each timestep, computes the best sensing action to minimize the entropy of future task actions. Välimäki and Ritala [77] formulates the problem of a robot deciding where to turn its cameras with a POMDP formulation that assumes gaussian dynamics. However, it considers only myopic optimization of the immediate information gain. A hierarchical belief space planning is proposed in [41], where different levels deal with different kinds of uncertainty. Planning is based on a forward search for *n* steps into the future, allowing directly belief-based rewards.

4.2.3 Belief-MDP. Some approaches circumvent the problem of solving POMDPs by reducing the formulation to a belief MDP. This definition of belief MDP implies that the belief is one of the features included in the state space of an MDP, along with other fully observable state features. Typically, this implies that the belief is maintained externally and is not directly part of the planning process. Dressel and Kochenderfer [19] uses this approach in a drone tracking scenario, where the cost includes the entropy of the belief, and Oh and Powell [55] similarly models a contamination cleaning problem with this approach. Eck and Soh [22] presents a multi-agent approach to reason when to sense the environment, request, or share information. An information-gathering POMDP formulation is reduced to a Knowledge State MDP, with fully observable knowledge variables in the state space (e.g., a discretization of the belief entropy), although physical actions (e.g., where to sense?) is left out of the model. How the authors deal with the multi-agent aspect is reviewed in Section 5.

4.2.4 Hybrid rewards. So far, we have distinguished state-based and belief-based rewards as two opposing approaches to formulating information gain objectives. However, most of the reviewed approaches may accommodate a mix of both, which is dubbed in the literature as multi-objective performance criterion [52], mixed criterion [16], or hybrid rewards [20]. Generally, those represent the total agent's reward as a weighted sum

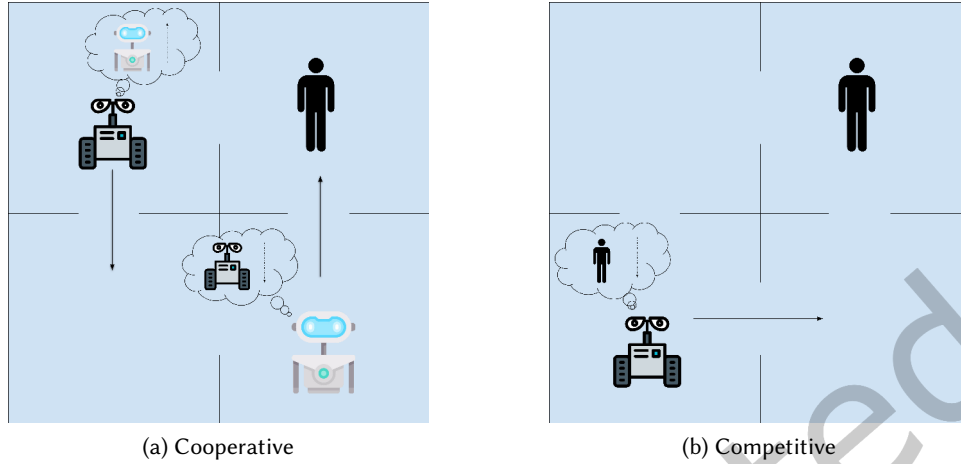(a) Cooperative   (b) Competitive

Fig. 5. Illustrative *Target Tracking* toy problem example for multi-agent settings, meaning a planning agent takes into account the intentions or plans of the other agents or targets. Subfigure 5a illustrates the case when multiple agents act cooperatively to identify the target; therefore, agents improve their behavior by avoiding actions that do not return added value. Subfigure 5b illustrates the case when the planning agent considers the target's possible intentions to drive its actions.

between task utilities and cost with information measures. This is typically represented with a mix of equations 4 and one of the possible representations for $\rho(b, a)$, such as Eqs.6, 7: $R(b, a) = \omega \sum_{s \in S} R(s, a)b(s) + (1 - \omega)\rho(b, a)$, where ω is the optional weight factor. In practice, this approach is the most suitable for real scenarios as one typically wants to balance different costs (energy, maintenance, communication, etc.) with information gains.

5 MULTI-AGENT ACTIVE SENSING

In Section 4, we covered studies for single-agent active sensing. Even though multi-agent problems could be represented in such models (through a centralized planning process), the models are not specifically designed to address multi-agent-specific challenges. This section focuses on inherently multi-agent models, both in cooperative and competitive settings.

5.1 Illustrative Example

We will now extend our illustrative example to multi-agent settings, as seen in Figure 5. We consider the case in which two patrolling agents must locate the target. This scenario presents both cooperative and competitive multi-agent settings. In the cooperative setting (Figure 5a), the agents' goal is to coordinate to detect the target faster. One robot communicates that it intends to go down. The second agent can then reason on this fact and decide to go up to avoid overlapping. In the competitive setting (Figure 5b), the target is considered an agent in the system instead of a moving feature of the environment, and the planning agent can incorporate one additional level of reasoning, which is the internal state of the target. Therefore the planning agent does not only reason about the target's "environmental" properties (location, speed, etc.) but also its possible beliefs, intentions, goals, etc., which allows for more efficient action planning. In our example, the robot knows that the target intends to go down (from previous sensing actions, such as intercepting a message, observing the target's past behavior, or inferring from the target's type) and therefore plans to go right itself to intercept the target.

5.2 Literature Review

Multi-agent active sensing development is relatively recent. Models in the literature can address cooperative, self-interested, or adversarial settings. Cooperative settings usually focus on coordination to efficiently cover the environment or to maintain a common belief state. In self-interested and adversarial settings, the agents typically focus on gathering information to anticipate the other agents' actions and react appropriately to avoid an undesirable state such as a collision or counter an adversarial attack. In both settings, however, it is interesting to note that the information-gathering process targets not only environmental factors (including a possible opponent location, speed, etc.) but also the other agents' internal states (their intention, goals, beliefs, etc.).

5.2.1 Cooperative Settings. To our knowledge, the first models to address multi-agent active perception are proposed by Renoux et al. [58] and Eck and Soh [22], who develop two different models focusing on cases in which agents cannot collaborate in advance to create a joint policy. Renoux et al. [58] represents each agent in the system with a POMDP that reasons over extended belief states, which includes the agent's own beliefs over the environmental factors but also a level-1 nested belief over the other agents in the systems. The reward function is based on a model of agent-based relevance, defined in [59] and presented in Equation 8.

$$\text{rel}_i\left(o\right) = \alpha D_{KL}\left(b_t \| b_{t+1}\right) + \beta\left(H\left(b_t\right) - H\left(b_{t+1}\right)\right) + \delta \tag{8}$$

where b_t is the belief state of the agent before receiving the observation o, b_{t+1} is the belief state of the agent after receiving the observation o, D_{KL} is the Kullback-Leibler distance, and H is the negative entropy. This formulation assumes that an observation is relevant for a given agent if it is new (i.e., changes the agent's belief significantly, measured by the Kullback-Leibler distance) or allows to render the agent's beliefs more precise (measured by the difference in entropy). The parameters α and β allow balancing these two (sometimes contradictory) aspects, and the parameter δ ensures that the relevance remains positive. Using this formulation and the extended belief state, the agents can proactively weigh the expected impact an observation for another agent and optimize to send the most impactful one. This decentralized information sharing allows decentralized cooperation. To solve the extended POMDP, Renoux et al. [58] transform it into a Belief-MDP, similar to what has been described in the single-agent case (Section 4.2.3). Later, Renoux et al. [60] use similar extended belief states to extend the POMDP-IR framework into a Communicative POMDP-IR (Com-POMDP-IR). The Com-POMDP-IR extends the prediction actions from the POMDP-IR to incorporate predictions about a human operator's belief state and uses these prediction actions to optimize a communication strategy. In this case, the resulting POMDP is solved using standard POMDP approaches, similar to what has been described in the single-agent case (Section 4.2.1).

Eck and Soh [22] also represents each agent as a POMDP and considers that each agent could request information from their neighbors. When they share information, agents share their entire belief space, and the receiving agent updates its own beliefs according to Equation 9.

$$b'(x) = \frac{b(x)\left[w \cdot b_{Sh}(x) + (1-w)\left(1 - b_{Sh}(x)\right)\right]}{\sum_{x' \in DOM(X)} b\left(j, x'\right)\left[w \cdot b_{Sh}\left(j, x'\right) + (1-w)\left(1 - b_{Sh}\left(x'\right)\right)\right]} \tag{9}$$

where b is the belief state of the agent receiving the information, b_{Sh} the shared belief space, $x, x' \in DOM(X)$ are the possible values for a partially observable phenomenon X, and w is a constant weight that dampens shared information, as is commonly used in information fusion literature [27]. The resulting model is solved by transforming it into a Knowledge POMDP, according to what has already been described in 4.2.3. The main difference between [58] and [22] lies in the fact that [58] exchanges single observations based on their estimated relevance and update the planning agent's belief state using standard POMDPs' Bayes rule, while [22] exchanges entire belief states and use an information fusion approach to merge the two belief states. However, both approaches focus on "on-the-go" multi-agent cooperation, where previous synchronization is unnecessary.

Lauri et al. [42] propose an alternative approach to multi-agent active perception by extending the ρ-POMDP approach to decentralized systems. The resulting ρDec-POMDP applies an entropy measure to the reward function to perform active sensing. The authors use an exact algorithm and assume periodic explicit communication to compute a joint belief estimate. Lauri et al. [44] relax the explicit communication hypothesis and present a heuristic for the ρDec-POMDP, making it possible to solve bigger problems. This heuristic uses a final reward definition, i.e., a reward granted at the end of the finite horizon. This reward is also based on the Shannon entropy of the joint belief, as in [42]. Because of this final reward, the algorithm in [44] and [45] still requires the explicit computation of the joint belief estimate, requiring a large memory and computation overhead. Lauri and Oliehoek [43] show that the ρDec-POMDP can be converted into a Dec-POMDP with linear rewards, similar to the the ρ-POMDP and POMDP-IR equivalence. To do so, the authors introduce individual prediction actions in the ρDec-POMDP model, similar to the prediction actions of the POMDP-IR model. This equivalence then allows to use standard Dec-POMDP solvers and therefore does not require the computation of joint state estimates. Finally, Best et al. [12] introduced the Decentralized Monte-Carlo Tree Search (Dec-MCTS), an online algorithm for asynchronous multi-robot coordination. Even though the Dec-MCTS is not limited to active perception scenarios, the authors illustrate its efficiency in an active perception scenario. Coordination is achieved by considering the plans of the other robots, which are communicated during a dedicated communication phase.

In the approaches above, the active perception process only concerns environmental factors, and mechanisms are implemented to allow the agents to reach a common belief over these factors. However, very little work has been done to include active perception of the other agents' internal states. To our knowledge, only Best et al. [13] chose this approach and extended the Dec-MCTS model with a communication planning algorithm to optimize a communication requests sequence. Agents evaluate other agents' belief evolution and send a request for other agents' to send them their plan. In doing so, the agents introduce active perception actions over other agents' internal states by only requesting their plan when needed.

5.2.2 Non-cooperative Settings. Non-cooperative settings focus on predicting the external agents' behavior by performing actions that will give the most information on their internal state. The external agents can be self-interested, i.e., they share the same environment, but their goals do not conflict with the planning agents' goals; or adversarial, i.e., their goals conflict with the planning agents' goals. One could relate the problem of active sensing in adversarial settings to the Adversarial Intention Recognition (AIR) problem. The AIR problem is concerned with recognizing an adversarial's goal or policy to prevent or counter its attack. However, existing decision-theoretic approaches to solve the AIR problem focus on identifying behavior based on observation of the adversary's action but do not act explicitly to gather evidence and therefore do not include active perception in their framework [2, 30].

Shen and How [68] explicitly consider an active perception scenario in which an opponent is either neutral, in which case its behavior is modeled as a reactive policy, or adversary, in which case its behavior is modeled with an MDP and bounded rationality. They then use a belief-space reward and a soft Q-Learning approach to optimize the policy. So far, this is the only framework that tackles the problem of active perception in a non-cooperative multi-agent setting.

6 DISCUSSION

6.1 Reactive sensing

In our literature review, we grouped our findings into five major sub-categories: *target tracking*, *classification actions*, *external information metrics*, *sensing actions* and *oracular observation providers*. In common, all of them attempt to give agents tools to plan their actions to improve the available information about the environment or some features of the environment. However, one of our major findings is that they differ in formulating that goal. In other words, it lacks a common modeling framework that can systematically be used across different

application scenarios, although, of course, one might be able to reproduce most of them by adapting to a specific scenario.

Also interesting to note that part of these approaches lay the foundations for the *active sensing* category. For instance, classification rewards are a primitive form of defining the value of information, although not flexible in formalizing the exact values of desired information. Extending the set of actions with extra actions which do not directly influence the environment but rather the agent's internal state is also commonly seen to influence the agent's decisions toward information gain. However, they are not coupled with explicit measures of the value of information. Finally, the outsourcing of belief inference and use of external value of information metrics for a state-based designed reward also shows hints of what is directly incorporated in *active sensing* methods.

Traditionally, for classic decision-theoretic formulations, information gain is a mean and not a goal in itself. In other words, there is some physical objective, and sensing actions are an intermediate step toward achieving the goal. As a consequence of not defining the reward based on information value metrics, *reactive sensing* approaches follow this assumption but, in one way or another, tricks that formulation to perform some information gathering task, e.g., the *target tracking* positive reward if the agent is in the same state as the target. Thus we claim that these methods lack generality, for their models need scenario-dependent tweaks to achieve information-gathering behaviors.

6.2 Single agent Active sensing

Contrary to our findings with *reactive sensing* methods, most of the literature found here follows systematic approaches, i.e., that attempt to generally formalize the theoretical problem instead of focusing on solving a specific scenario. We see a distinction between offline and online solvers. Offline solvers are mostly variations of the POMDP-IR and ρPOMDP models. They are formulated around linearizations of the reward function, making them compatible with the formulation of classic state-of-the-art solvers, in particular, point-based. On the other hand, online solvers are more flexible, given that there are generally no restrictions to the format of the reward. The efficiency of these approaches is still limited by how they handle key issues, such as the ability to choose the most promising paths down the search tree. Typically, this can be achieved by maintaining lower and upper bounds or computing rollouts from leaf nodes with particle-based methods. However, lower bounds are based on rough offline approximations (which are PWLC-restricted), and rollouts are performed from state particles. Solutions such as the bag of particles by ρPOMCP are promising, but their complexity is dependent on the inner structure of the observation dynamics and, therefore, might not be feasible for complex problems.

We also note a remarkable contrast between the generality of *active sensing* approaches and their usage in real-world scenarios. Although they present a systematic approach to the problem, we found only a few papers reporting their application. Within those, there are information gathering problems for localization of objects in domestic scenarios through an information-guided construction of semantic maps [78, 79], and the use of ρPOMDPs on smart grid management associated with privacy concerns [80]. However, this ability to transfer between scenarios suggests that this framework generalizes better.

6.3 Multi agent Active sensing

We found that research on multi-agent active sensing is still extremely recent, and a vast majority of models target cooperative settings and focus on gathering *environmental information*. We noted two distinct approaches. The first one focuses on ad-hoc configurations, in which it is not possible or desirable for the agents to coordinate a plan a priori and integrate measures on what or when to share information in an "extended" single-agent POMDP. The second approach focuses on extending existing multi-agent decision-theoretic models, namely, Dec-POMDPs, to belief-based reward functions.

Research on non-cooperative settings and non-environmental information gathering is still extremely recent and limited.

6.4 Future research directions

This survey allowed us to identify four directions in which future work would be beneficial.

Real-world applications of general frameworks. In our literature review, we found a surprising unbalance between the number of tailored solutions for specific problems compared to the use of more general frameworks applied to real-world applications. This indicates that, unlike what is commonly observed in other fields in which theoretical advances are followed by the transfer of such knowledge to real-world problems, practitioners in information gathering do not resort as much to the existing literature.

Model-free information gathering. By definition, frameworks that depend on the belief as the internal knowledge state are model-based since the model of the world is needed to update the belief state. However, model-free methods that learn policies for autonomous agents directly from the interaction with the environment are continuously improving with impressive results by reinforcement learning agents in, for example, competitions against humans [69]. Furthermore, online planning algorithms based on Monte Carlo tree search have shown state-of-the-art results without needing the full environment model but access to a generative model (similar to reinforcement learning). This indicates possible research opportunities to scale up active sensing methods, with a couple of approaches cited in this paper already trying to alleviate the need for a full belief representation.

Non-cooperative multi-agent information gathering. The problem of multi-agent information gathering in non-cooperative settings is still very unexplored in the decision-theoretic community. The possibility for agents to reason about other agents' internal states and act to perform active sensing on such internal states would be a valuable step towards more generic and applicable models. To our knowledge, the most complete model allowing an agent to reason on other agents' models is the Interactive POMDP (I-POMDP) [28]. However, it has not yet been extended or applied to active sensing scenarios.

Human-centered information gathering. All of the works reviewed in this survey consider only artificial planning agents. When present, humans are merely receivers of the information (as in [60]). However, Hybrid Human-Artificial Intelligence (HHAI), i.e., the development of artificially intelligent agents that can interact with humans to leverage their strengths and empower them [1], is a rapidly growing field, and collaborative human-agent information gathering scenarios are an important research domain. Future works could investigate how existing models could account for human teammates and how to develop new approaches for *hybrid active sensing*.

7 ACKNOWLEDGMENTS

REFERENCES

[1] Zeynep Akata, Dan Balliet, Maarten de Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen V. Hindriks, Holger H. Hoos, Hayley Hung, Catholijn M. Jonker, Christof Monz, Mark A. Neerincx, Frans A. Oliehoek, Henry Prakken, Stefan Schlobach, Linda C. van der Gaag, Frank van Harmelen, Herke van Hoof, Birna van Riemsdijk, Aimee van Wynsberghe, Rineke Verbrugge, Bart Verheij, Piek Vossen, and Max Welling. 2020. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* 53, 8 (2020), 18–28. https://doi.org/10.1109/MC.2020.2996587

[2] Samuel Ang, Hau Chan, Albert Xin Jiang, and William Yeoh. 2017. Game-Theoretic Goal Recognition Models with Applications to Security Domains. In *Decision and Game Theory for Security - 8th International Conference, GameSec 2017, Vienna, Austria, October 23-25,*

2017, Proceedings (Lecture Notes in Computer Science, Vol. 10575), Stefan Rass, Bo An, Christopher Kiekintveld, Fei Fang, and Stefan Schauer (Eds.). Springer, 256–272. https://doi.org/10.1007/978-3-319-68711-7_14

[3] Mauricio Araya-López, Olivier Buffet, Vincent Thomas, and François Charpillet. 2010. A POMDP Extension with Belief-dependent Rewards. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta (Eds.). Curran Associates, Inc., 64–72. https://proceedings.neurips.cc/paper/2010/hash/68053af2923e00204c3ca7c6a3150cf7-Abstract.html

[4] Nicholas Armstrong-Crews and Manuela M. Veloso. 2007. Oracular Partially Observable Markov Decision Processes: A Very Special Case. In *2007 IEEE International Conference on Robotics and Automation, ICRA 2007, 10-14 April 2007, Roma, Italy*. IEEE, 2477–2482. https://doi.org/10.1109/ROBOT.2007.363691

[5] Akash Arora, Robert Fitch, and Salah Sukkarieh. 2017. An approach to autonomous science by modeling geological knowledge in a Bayesian framework. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2017, Vancouver, BC, Canada, September 24-28, 2017*. IEEE, 3803–3810. https://doi.org/10.1109/IROS.2017.8206230

[6] Akash Arora, P. Michael Furlong, Robert Fitch, Salah Sukkarieh, and Terrence Fong. 2019. Multi-modal active perception for information gathering in science missions. *Auton. Robots* 43, 7 (2019), 1827–1853. https://doi.org/10.1007/s10514-019-09836-5

[7] George K. Atia, Venugopal V. Veeravalli, and Jason A. Fuemmeler. 2011. Sensor Scheduling for Energy-Efficient Target Tracking in Sensor Networks. *IEEE Trans. Signal Process.* 59, 10 (2011), 4923–4937. https://doi.org/10.1109/TSP.2011.2160055

[8] Ruzena Bajcsy. 1988. Active perception. *Proc. IEEE* 76, 8 (1988), 966–1005. https://doi.org/10.1109/5.5968

[9] Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. 2018. Revisiting active perception. *Auton. Robots* 42, 2 (2018), 177–196. https://doi.org/10.1007/s10514-017-9615-3

[10] Chitta Baral, Thomas Bolander, Hans van Ditmarsch, and Sheila McIlrath. 2017. Epistemic Planning (Dagstuhl Seminar 17231). *Dagstuhl Reports* 7, 6 (2017), 1–47. https://doi.org/10.4230/DagRep.7.6.1

[11] Daniel S. Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The Complexity of Decentralized Control of Markov Decision Processes. *Math. Oper. Res.* 27, 4 (2002), 819–840. https://doi.org/10.1287/moor.27.4.819.297

[12] Graeme Best, Oliver M. Cliff, Timothy Patten, Ramgopal R. Mettu, and Robert Fitch. 2019. Dec-MCTS: Decentralized planning for multi-robot active perception. *Int. J. Robotics Res.* 38, 2-3 (2019). https://doi.org/10.1177/0278364918755924

[13] Graeme Best, Michael Forrai, Ramgopal R. Mettu, and Robert Fitch. 2018. Planning-Aware Communication for Decentralised Multi-Robot Coordination. In *2018 IEEE International Conference on Robotics and Automation, ICRA 2018, Brisbane, Australia, May 21-25, 2018*. IEEE, 1050–1057. https://doi.org/10.1109/ICRA.2018.8460617

[14] Thomas Bolander and Mikkel Birkegaard Andersen. 2011. Epistemic planning for single and multi-agent systems. *J. Appl. Non Class. Logics* 21, 1 (2011), 9–34. https://doi.org/10.3166/jancl.21.9-34

[15] Craig Boutilier, Thomas L. Dean, and Steve Hanks. 1999. Decision-Theoretic Planning: Structural Assumptions and Computational Leverage. *J. Artif. Intell. Res.* 11 (1999), 1–94. https://doi.org/10.1613/jair.575

[16] Caroline Ponzoni Carvalho Chanel, Florent Teichteil-Königsbuch, and Charles Lesire. 2012. POMDP-based online target detection and recognition for autonomous UAVs. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012 (Frontiers in Artificial Intelligence and Applications, Vol. 242)*, Luc De Raedt, Christian Bessiere, Didier Dubois, Patrick Doherty, Paolo Frasconi, Fredrik Heintz, and Peter J. F. Lucas (Eds.). IOS Press, 955–960. https://doi.org/10.3233/978-1-61499-098-7-955

[17] Edwin K. P. Chong, Christopher M. Kreucher, and Alfred O. Hero III. 2009. Partially Observable Markov Decision Process Approximations for Adaptive Sensing. *Discret. Event Dyn. Syst.* 19, 3 (2009), 377–422. https://doi.org/10.1007/s10626-009-0071-x

[18] Louis Dressel and Mykel J. Kochenderfer. 2017. Efficient Decision-Theoretic Target Localization. In *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS 2017, Pittsburgh, Pennsylvania, USA, June 18-23, 2017*, Laura Barbulescu, Jeremy Frank, Mausam, and Stephen F. Smith (Eds.). AAAI Press, 70–78. https://aaai.org/ocs/index.php/ICAPS/ICAPS17/paper/view/15761

[19] Louis Dressel and Mykel J. Kochenderfer. 2019. Hunting Drones with Other Drones: Tracking a Moving Radio Target. In *International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*. IEEE, 1905–1912. https://doi.org/10.1109/ICRA.2019.8794243

[20] Adam Eck and Leen-Kiat Soh. 2012. Evaluating POMDP rewards for active perception. In *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4-8, 2012 (3 Volumes)*, Wiebe van der Hoek, Lin Padgham, Vincent Conitzer, and Michael Winikoff (Eds.). IFAAMAS, 1221–1222. http://dl.acm.org/citation.cfm?id=2343932

[21] Adam Eck and Leen-Kiat Soh. 2013. Observer effect from stateful resources in agent sensing. *Auton. Agents Multi Agent Syst.* 26, 2 (2013), 202–244. https://doi.org/10.1007/s10458-011-9189-y

[22] Adam Eck and Leen-Kiat Soh. 2015. To Ask, Sense, or Share: Ad Hoc Information Gathering. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, May 4-8, 2015*, Gerhard Weiss, Pinar Yolum, Rafael H. Bordini, and Edith Elkind (Eds.). ACM, 367–376. http://dl.acm.org/citation.cfm?id=2772928

[23] Mathieu Fehr, Olivier Buffet, Vincent Thomas, and Jilles Steeve Dibangoye. 2018. rho-POMDPs have Lipschitz-Continuous epsilon-Optimal Value Functions. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (Eds.). 6933–6943. https://proceedings.neurips.cc/paper/2018/hash/de7f47e09c8e05e6021ababdf6bc58e7-Abstract.html

[24] Johannes Fischer and Ömer Sahin Tas. 2020. Information Particle Filter Tree: An Online Algorithm for POMDPs with Belief-Based Rewards on Continuous Domains. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 3177–3187. http://proceedings.mlr.press/v119/fischer20a.html

[25] Mahsa Ghasemi and Ufuk Topcu. 2019. Online Active Perception for Partially Observable Markov Decision Processes with Limited Budget. In *58th IEEE Conference on Decision and Control, CDC 2019, Nice, France, December 11-13, 2019*. IEEE, 6169–6174. https://doi.org/10.1109/CDC40024.2019.9029762

[26] Mahsa Ghasemi and Ufuk Topcu. 2019. Perception-Aware Point-Based Value Iteration for Partially Observable Markov Decision Processes. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, Sarit Kraus (Ed.). ijcai.org, 2371–2377. https://doi.org/10.24963/ijcai.2019/329

[27] Robin Glinton, Paul Scerri, and Katia P. Sycara. 2010. Exploiting scale invariant dynamics for efficient information propagation in large teams. In *9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), Toronto, Canada, May 10-14, 2010, Volume 1-3*, Wiebe van der Hoek, Gal A. Kaminka, Yves Lespérance, Michael Luck, and Sandip Sen (Eds.). IFAAMAS, 21–30. https://dl.acm.org/citation.cfm?id=1838210

[28] Piotr J. Gmytrasiewicz and Prashant Doshi. 2005. A Framework for Sequential Planning in Multi-Agent Settings. *J. Artif. Intell. Res.* 24 (2005), 49–79. https://doi.org/10.1613/jair.1579

[29] Tipakorn Greigarn, Michael S. Branicky, and Murat Cenk Cavusoglu. 2019. Task-Oriented Active Sensing via Action Entropy Minimization. *IEEE Access* 7 (2019), 135413–135426. https://doi.org/10.1109/ACCESS.2019.2941706

[30] Nicolas Le Guillarme, Abdel-Illah Mouaddib, Sylvain Gatepaille, and Amandine Bellenger. 2016. Adversarial Intention Recognition as Inverse Game-Theoretic Planning for Threat Assessment. In *28th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2016, San Jose, CA, USA, November 6-8, 2016*. IEEE Computer Society, 698–705. https://doi.org/10.1109/ICTAI.2016.0111

[31] AnYuan Guo. 2003. Decision-theoretic active sensing for autonomous agents. In *The Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2003, July 14-18, 2003, Melbourne, Victoria, Australia, Proceedings*. ACM, 1002–1003. https://doi.org/10.1145/860575.860766

[32] Darin C. Hitchings and David A. Castañón. 2010. Receding horizon stochastic control algorithms for sensor management. In *American Control Conference, ACC 2010, Baltimore, Maryland, USA, June 30 - July 2, 2010*. IEEE, 6809–6815. https://doi.org/10.1109/ACC.2010.5531634

[33] Darin C. Hitchings and David A. Castañón. 2011. Sensor control for search and identification of Markov objects. In *50th IEEE Conference on Decision and Control and European Control Conference, 11th European Control Conference, CDC/ECC 2011, Orlando, FL, USA, December 12-15, 2011*. IEEE, 6272–6277. https://doi.org/10.1109/CDC.2011.6160670

[34] David Hsu, Wee Sun Lee, and Nan Rong. 2008. A point-based POMDP planner for target tracking. In *2008 IEEE International Conference on Robotics and Automation, ICRA 2008, May 19-23, 2008, Pasadena, California, USA*. IEEE, 2644–2650. https://doi.org/10.1109/ROBOT.2008.4543611

[35] Shihao Ji and Lawrence Carin. 2007. Cost-sensitive feature acquisition and classification. *Pattern Recognit.* 40, 5 (2007), 1474–1485. https://doi.org/10.1016/j.patcog.2006.11.008

[36] Shihao Ji, Ronald Parr, and Lawrence Carin. 2007. Nonmyopic Multiaspect Sensing With Partially Observable Markov Decision Processes. *IEEE Trans. Signal Process.* 55, 6-1 (2007), 2720–2730. https://doi.org/10.1109/TSP.2007.893747

[37] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. 1998. Planning and Acting in Partially Observable Stochastic Domains. *Artif. Intell.* 101, 1-2 (1998), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

[38] Bilal Kartal, Julio Godoy, Ioannis Karamouzas, and Stephen J. Guy. 2015. Stochastic Tree Search with Useful Cycles for patrolling problems. In *IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015*. IEEE, 1289–1294. https://doi.org/10.1109/ICRA.2015.7139357

[39] David Kent and Sonia Chernova. 2020. Human-Centric Active Perception for Autonomous Observation. In *2020 IEEE International Conference on Robotics and Automation, ICRA 2020, Paris, France, May 31 - August 31, 2020*. IEEE, 1785–1791. https://doi.org/10.1109/ICRA40945.2020.9197201

[40] Mykel J Kochenderfer. 2015. *Decision making under uncertainty: theory and application*. MIT press.

[41] Michael William Lanighan, Takeshi Takahashi, and Roderic A. Grupen. 2018. Planning Robust Manual Tasks in Hierarchical Belief Spaces. In *Proceedings of the Twenty-Eighth International Conference on Automated Planning and Scheduling, ICAPS 2018, Delft, The Netherlands, June 24-29, 2018*. 459–467. https://aaai.org/ocs/index.php/ICAPS/ICAPS18/paper/view/17780

[42] Mikko Lauri, Eero Heinänen, and Simone Frintrop. 2017. Multi-robot active information gathering with periodic communication. In *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*. IEEE, 851–856. https://doi.org/10.1109/ICRA.2017.7989104

[43] Mikko Lauri and Frans Oliehoek. 2020. Multi-agent active perception with prediction rewards. In *Advances in Neural Information Processing Systems*, Vol. 33. 13651–13661. https://proceedings.neurips.cc/paper/2020/file/9db6faeef387dc789777227a8bed4d52-Paper.pdf

[44] Mikko Lauri, Joni Pajarinen, and Jan Peters. 2019. Information Gathering in Decentralized POMDPs by Policy Graph Improvement. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, Edith Elkind, Manuela Veloso, Noa Agmon, and Matthew E. Taylor (Eds.). International Foundation for Autonomous Agents and Multiagent Systems, 1143–1151. http://dl.acm.org/citation.cfm?id=3331815

[45] Mikko Lauri, Joni Pajarinen, and Jan Peters. 2020. Multi-agent active information gathering in discrete and continuous-state decentralized POMDPs by policy graph improvement. *Auton. Agents Multi Agent Syst.* 34, 2 (2020), 42. https://doi.org/10.1007/s10458-020-09467-6

[46] Mikko Lauri and Risto Ritala. 2013. Planning for multiple measurement channels in a continuous-state POMDP. *Ann. Math. Artif. Intell.* 67, 3-4 (2013), 283–317. https://doi.org/10.1007/s10472-013-9361-y

[47] Mikko Lauri and Risto Ritala. 2015. Optimal sensing via multi-armed bandit relaxations in mixed observability domains. In *IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015*. IEEE, 4807–4812. https://doi.org/10.1109/ICRA.2015.7139867

[48] Mikko Lauri and Risto Ritala. 2016. Planning for robotic exploration based on forward simulation. *Robotics Auton. Syst.* 83 (2016), 15–31. https://doi.org/10.1016/j.robot.2016.06.008

[49] Mikko Lauri, Aino Ropponen, and Risto Ritala. 2017. Meeting a deadline: shortest paths on stochastic directed acyclic graphs with information gathering. *Ann. Math. Artif. Intell.* 79, 4 (2017), 337–370. https://doi.org/10.1007/s10472-016-9527-5

[50] Cam Linke, Nadia M. Ady, Martha White, Thomas Degris, and Adam White. 2020. Adapting Behavior via Intrinsic Reward: A Survey and Empirical Study. *J. Artif. Intell. Res.* 69 (2020), 1287–1332. https://doi.org/10.1613/jair.1.12087

[51] Jun Liu and Ryan K. Williams. 2018. Optimal Intermittent Deployment and Sensor Selection for Environmental Sensing with Multi-Robot Teams. In *2018 IEEE International Conference on Robotics and Automation, ICRA 2018, Brisbane, Australia, May 21-25, 2018*. IEEE, 1078–1083. https://doi.org/10.1109/ICRA.2018.8460215

[52] L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, and J. De Schutter. 2002. Active Sensing for Robotics - A Survey. In *in Proc. 5 th International Conference on Numerical Methods and Applications*. 316–324.

[53] Abdulmajid Murad, Frank Alexander Kraemer, Kerstin Bach, and Gavin Taylor. 2020. Information-driven adaptive sensing based on deep reinforcement learning. In *IoT '20: 10th International Conference on the Internet of Things, Malmö, Sweden, October 6-9, 2020*, Paul Davidsson and Marc Langheinrich (Eds.). ACM, 2:1–2:8. https://doi.org/10.1145/3410992.3411001

[54] Hoa Van Nguyen, Hamid Rezatofighi, Ba-Ngu Vo, and Damith Chinthana Ranasinghe. 2020. Multi-Objective Multi-Agent Planning for Jointly Discovering and Tracking Mobile Objects. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 7227–7235. https://ojs.aaai.org/index.php/AAAI/article/view/6213

[55] Yura Oh and Warren B. Powell. 2019. Optimal Policies for Routing Multi Sensor-Effector Combined Autonomous Devices. In *2019 International Symposium on Multi-Robot and Multi-Agent Systems, MRS 2019, New Brunswick, NJ, USA, August 22-23, 2019*. IEEE, 182–184. https://doi.org/10.1109/MRS.2019.8901060

[56] Pascal Poupart. 2005. *Exploiting structure to efficiently solve large scale partially observable Markov decision processes.* Ph. D. Dissertation. University of Toronto, Canada. https://librarysearch.library.utoronto.ca/permalink/01UTORONTO_INST/14bjeso/alma991106619319006196

[57] David V. Pynadath and Milind Tambe. 2002. The Communicative Multiagent Team Decision Problem: Analyzing Teamwork Theories and Models. *J. Artif. Intell. Res.* 16 (2002), 389–423. https://doi.org/10.1613/jair.1024

[58] Jennifer Renoux, Abdel-Illah Mouaddib, and Simon Le Gloannec. 2015. A decision-theoretic planning approach for multi-robot exploration and event search. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2015, Hamburg, Germany, September 28 - October 2, 2015*. IEEE, 5287–5293. https://doi.org/10.1109/IROS.2015.7354123

[59] Jennifer Renoux, Abdel-Illah Mouaddib, and Simon Le Gloannec. 2014. A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering. In *Workshop on Multiagent Sequential Decision Making Under Uncertainty*. Paris, France.

[60] Jennifer Renoux, Tiago S. Veiga, Pedro U. Lima, and Matthijs T. J. Spaan. 2020. A Unified Decision-Theoretic Model for Information Gathering and Communication Planning. In *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020, Naples, Italy, August 31 - September 4, 2020*. IEEE, 67–74. https://doi.org/10.1109/RO-MAN47096.2020.9223597

[61] Stephanie Rosenthal and Manuela M. Veloso. 2011. Modeling humans as observation providers using POMDPs. In *20th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2011, Atlanta, Georgia, USA, July 31 - August 3, 2011*, Henrik I. Christensen (Ed.). IEEE, 53–58. https://doi.org/10.1109/ROMAN.2011.6005272

[62] Stéphane Ross, Joelle Pineau, Sébastien Paquet, and Brahim Chaib-draa. 2008. Online Planning Algorithms for POMDPs. *J. Artif. Intell. Res.* 32 (2008), 663–704. https://doi.org/10.1613/jair.2567

[63] Yash Satsangi, Sungsu Lim, Shimon Whiteson, Frans A. Oliehoek, and Martha White. 2020. Maximizing Information Gain in Partially Observable Environments via Prediction Rewards. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent*

Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020, Amal El Fallah Seghrouchni, Gita Sukthankar, Bo An, and Neil Yorke-Smith (Eds.). International Foundation for Autonomous Agents and Multiagent Systems, 1215–1223. https://doi.org/10.5555/3398761.3398902

[64] Yash Satsangi, Shimon Whiteson, Frans A. Oliehoek, and Matthijs T. J. Spaan. 2018. Exploiting submodular value functions for scaling up active perception. *Auton. Robots* 42, 2 (2018), 209–233. https://doi.org/10.1007/s10514-017-9666-5

[65] BL Scala, Mohammad Rezaeian, and Bill Moran. 2005. Optimal adaptive waveform selection for target tracking. In *7th International Conference on Information Fusion*, Vol. 1. IEEE, 552–557.

[66] Guy Shani, Joelle Pineau, and Robert Kaplow. 2013. A Survey of Point-based POMDP Solvers. *Autonomous Agents and Multi-Agent Systems* 27, 1 (July 2013), 1–51.

[67] Guy Shani, Pascal Poupart, Ronen I. Brafman, and Solomon Eyal Shimony. 2008. Efficient ADD Operations for Point-Based Algorithms. In *Proceedings of the Eighteenth International Conference on Automated Planning and Scheduling, ICAPS 2008, Sydney, Australia, September 14-18, 2008*, Jussi Rintanen, Bernhard Nebel, J. Christopher Beck, and Eric A. Hansen (Eds.). AAAI, 330–337. http://www.aaai.org/Library/ICAPS/2008/icaps08-041.php

[68] Macheng Shen and Jonathan P. How. 2019. Active Perception in Adversarial Scenarios using Maximum Entropy Deep Reinforcement Learning. In *International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019*. IEEE, 3384–3390. https://doi.org/10.1109/ICRA.2019.8794389

[69] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Vedavyas Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy P. Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nat.* 529, 7587 (2016), 484–489. https://doi.org/10.1038/nature16961

[70] Matthijs T. J. Spaan. 2008. Cooperative active perception using POMDPs. In *AAAI 2008 workshop on advancements in POMDP solvers*.

[71] Matthijs T. J. Spaan. 2012. Partially Observable Markov Decision Processes. In *Reinforcement Learning: State of the Art*, Marco Wiering and Martijn van Otterlo (Eds.). Springer Verlag.

[72] Matthijs T. J. Spaan and Pedro U. Lima. 2009. A Decision-Theoretic Approach to Dynamic Sensor Selection in Camera Networks. In *Proceedings of the 19th International Conference on Automated Planning and Scheduling, ICAPS 2009, Thessaloniki, Greece, September 19-23, 2009*. AAAI. http://aaai.org/ocs/index.php/ICAPS/ICAPS09/paper/view/758

[73] Matthijs T. J. Spaan, Tiago Veiga, and Pedro U. Lima. 2010. Active cooperative perception in network robot systems using POMDPs. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 18-22, 2010, Taipei, Taiwan*. IEEE, 4800–4805. https://doi.org/10.1109/IROS.2010.5648856

[74] Matthijs T. J. Spaan, Tiago S. Veiga, and Pedro U. Lima. 2015. Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *Auton. Agents Multi Agent Syst.* 29, 6 (2015), 1157–1185. https://doi.org/10.1007/s10458-014-9279-8

[75] Matthijs T. J. Spaan and Nikos Vlassis. 2005. Perseus: Randomized Point-based Value Iteration for POMDPs. *J. Artif. Intell. Res.* 24 (2005), 195–220. https://doi.org/10.1613/jair.1659

[76] Vincent Thomas, Gérémy Hutin, and Olivier Buffet. 2020. Monte Carlo Information-Oriented Planning. In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020) (Frontiers in Artificial Intelligence and Applications, Vol. 325)*, Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang (Eds.). IOS Press, 2378–2385. https://doi.org/10.3233/FAIA200368

[77] Tuomas Välimäki and Risto Ritala. 2016. Optimizing gaze direction in a visual navigation task. In *2016 IEEE International Conference on Robotics and Automation, ICRA 2016, Stockholm, Sweden, May 16-21, 2016*, Danica Kragic, Antonio Bicchi, and Alessandro De Luca (Eds.). IEEE, 1427–1432. https://doi.org/10.1109/ICRA.2016.7487276

[78] Tiago S. Veiga, Pedro Miraldo, Rodrigo Ventura, and Pedro U. Lima. 2016. Efficient object search for mobile robots in dynamic environments: Semantic map as an input for the decision maker. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2016, Daejeon, South Korea, October 9-14, 2016*. IEEE, 2745–2750. https://doi.org/10.1109/IROS.2016.7759426

[79] Tiago S. Veiga, Miguel Silva, Rodrigo Ventura, and Pedro U. Lima. 2019. A Hierarchical Approach to Active Semantic Mapping Using Probabilistic Logic and Information Reward POMDPs. In *Proceedings of the Twenty-Ninth International Conference on Automated Planning and Scheduling, ICAPS 2018, Berkeley, CA, USA, July 11-15, 2019*, J. Benton, Nir Lipovetzky, Eva Onaindia, David E. Smith, and Siddharth Srivastava (Eds.). AAAI Press, 773–781. https://ojs.aaai.org/index.php/ICAPS/article/view/3546

[80] Jiyun Yao and Parv Venkitasubramaniam. 2013. On the privacy-cost tradeoff of an in-home power storage mechanism. In *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*. IEEE, 115–122. https://doi.org/10.1109/Allerton.2013.6736513