

# Multi-Objective Multi-Agent Planning for Jointly Discovering and Tracking Mobile Objects

Hoa Van Nguyen,<sup>1</sup> Hamid Rezaatofghi,<sup>1</sup> Ba-Ngu Vo,<sup>2</sup> Damith C. Ranasinghe<sup>1</sup>

<sup>1</sup>The University of Adelaide

<sup>2</sup>Curtin University

{hoavan.nguyen, hamid.rezaatofghi, damith.ranasinghe}@adelaide.edu.au, ba-ngu.vo@curtin.edu.au

## Abstract

We consider the challenging problem of online planning for a team of agents to autonomously search and track a time-varying number of mobile objects under the practical constraint of detection range limited onboard sensors. A standard POMDP with a value function that either encourages discovery or accurate tracking of mobile objects is inadequate to *simultaneously* meet the *conflicting goals* of searching for undiscovered mobile objects whilst keeping track of discovered objects. The planning problem is further complicated by misdetections or false detections of objects caused by range limited sensors and noise inherent to sensor measurements. We formulate a *novel* multi-objective POMDP based on information theoretic criteria, and an online multi-object tracking filter for the problem. Since controlling multi-agent is a well known combinatorial optimization problem, assigning control actions to agents necessitates a greedy algorithm. We prove that our proposed multi-objective value function is a monotone submodular set function; consequently, the greedy algorithm can achieve a  $(1 - 1/e)$  approximation for maximizing the submodular multi-objective function.

## Introduction

We study the problem of controlling a team of agents to jointly track discovered mobile objects and explore the environment to search for undiscovered mobile objects of interest. Such problems are ubiquitous in wildlife tracking (Cliff et al. 2015; Kays et al. 2011; Nguyen et al. 2019a; Thomas, Holland, and Minot 2012), search and rescue missions (Gerasenko et al. 2001; Murphy et al. 2008). For instance, a team of unmanned aerial vehicles (UAVs) can be deployed to monitor activities of endangered radio-tagged wildlife in a survey scene, or to search for victims in a disaster response (Beck et al. 2018). Hence, it is critical to not only search for undiscovered objects but also track the movements of discovered objects of interest. Consequently, the overall team’s objectives arise as a natural multi-objective optimization problem, where several pertinent goals (*i.e.*, tracking and discovering) need to be simultaneously achieved.

Intrinsically, searching for undiscovered objects whilst simultaneously tracking visible objects are *competing* goals

because, in practice, agent sensor systems, such as cameras, have limited detection range. A single agent may only observe a small region of space and a decision to leave a visible object to explore hitherto unseen regions will lead to losing track of visible objects. Therefore, an agent observing a small region of the search area needs to collaboratively interact with other agents to *plan* its course of actions to collectively maximize the overall team’s objectives of *tracking* and *discovering* multiple objects.

Multi-agent planning to achieve *multiple competing objectives* remains a challenging problem because of the complex interactions between agents leading to combinatorial optimization problems (Wai et al. 2018). In practice, the problem is *further* complicated because: *i*) the agent sensors are not only limited in range but also sensitivity, and measurements are always subjected to environmental noise. Consequently, object detectors suffer from both missing detections of objects and false detections; and *ii*) the number of objects of interest is often unknown, and varies with time since mobile objects can enter and leave the scene anytime (Vo et al. 2012). Most critically, the computation of optimal planning actions must be timely for real-world applications.

We propose a framework for multiple agents to jointly plan, search and track a time-varying number of objects using a novel multi-objective information-based value function formulation. Our multi-objective value function captures the competing objectives of planning for *tracking* and *discovery*. We adopt the random finite set (RFS) model for the collection of objects of interest to account for the random appearance and disappearance of objects and their dynamics. Our proposed multi-objective value function maximizes information gain over a look-ahead horizon for both discovered and undiscovered objects. Most importantly, our multi-objective value function is proven to be a monotone submodular set function; thus, we can cope with the intractability of the multi-objective optimization problem (MOP) by employing a greedy algorithm. Our ability to use a greedy algorithm facilitates the computation of approximately optimal control actions with linear complexity in the number of agents for realizing an online planning method.

**Related Work:** Multi-agent path planning in partially observable environments is a difficult problem for which the

Partially Observable Markov Decision Processes (POMDP) approach has recently gained significant interest (Silver and Veness 2010; MacDermed and Isbell 2013; Messias, Spaan, and Lima 2011). Although the cooperation problem can be formulated as a decentralized POMDP (Dec-POMDP), its exact solutions are NEXP-hard (Bernstein et al. 2002). This is especially problematic for multi-agent POMDPs since the action and observation space grows exponentially with the number of agents (Amato and Oliehoek 2015). To cope with this intractability, we adopt the MPOMDP centralized approach (Messias, Spaan, and Lima 2011) for controlling multiple agents (Dames and Kumar 2015; Dames, Tokekar, and Kumar 2017; Wang et al. 2018).

POMDP has also been employed for sensor selection problems, *e.g.*, (Spaan, Veiga, and Lima 2015; Satsangi et al. 2018) proposed using the  $\rho$ POMDP (Araya et al. 2010) for a mobile agent to select  $K$  in  $N$  available sensors to search and track multiple objects. In particular, (Spaan, Veiga, and Lima 2015) proposed a method that always assumes the existence of one extra object in the scene to encourage discovery. However, biasing the cardinality estimate generates sub optimal planing decisions at the cost of tracking performance.

Our study focuses on the problem of controlling a team of agents for the task of tracking and discovering mobile targets. The task requires a suitable tracking framework. Studies on tracking objects have employed approaches such as Multiple Hypotheses Tracking (MHT) (Reid 1979) or Joint Probabilistic Data Association (JPDA) (Blackman and Popoli 1999). The complex nature of our problem requires a framework that has the notion of probability of a random collection due to a time-varying and random number of objects where the states of objects are random vectors. The random finite set (RFS) (Mahler 2007) is the only framework that has the notion of probability density of a random set. Hence, we adopt RFS as our tracking framework.

Information-based path planning under the RFS framework for a single agent has been studied in several works (Beard et al. 2017; Hoang and Vo 2014; Ristic and Vo 2010). Most studies on multi-agent path planning using an RFS framework, is based on the Generalized Covariance Intersection (GCI) methods with the assumption that agents have a consensus view of all objects (Gostar, Hoseinnezhad, and Bab-Hadiashar 2016; Wang et al. 2018) and using only a single look-ahead horizon. (Dames, Tokekar, and Kumar 2017) proposed to control multiple fixed-wing UAVs to localize mobile taxis with a single objective value function. For localizing and searching objects simultaneously, (Dames and Kumar 2015) and (Charrow, Michael, and Kumar 2015) considered a similar scenario, but only for stationary objects. Planning using multi-objective optimization (MOP) has not been explored yet, except for single sensor selection (Zhu, Wang, and Liang 2019) or using the weighted sum method presented in (Charrow, Michael, and Kumar 2015) where the weighting parameters are difficult to define without prior knowledge. In contrast, we focus on optimizing all value functions (*i.e.*, tracking and discovering) simultaneously using MOP. In particular, our proposed tracking and discovering value functions are based on information criteria. The tracking value function maximizes the mutual information between future mea-

surements and discovered object states under a multi-sensor Bernoulli filter; the discovering value function maximizes the mutual information between empty measurements and undiscovered object states under a grid occupancy filter.

**Our contributions:** The main contributions of our work are: *(i)* We formulate a multi-agent planning problem with competing objectives and propose a planning algorithm for searching and tracking multiple mobile objects; *(ii)* We unify tracking and planning algorithms under a Bernoulli-based model; *(iii)* We prove that our proposed multi-objective value function is submodular; hence, the greedy algorithm can be used to rapidly determine the approximately optimal control actions with a bounded performance guarantee at  $(1 - 1/e)$ OPT.

## Problem Formulation

First, we introduce assumptions to help define our problem and introduce the notations we adopt in our work. Second, we provide a brief overview of the multi-sensor Bernoulli filter which unifies the tracking and discovering formulation. Next, we formulate our MPOMDP multi-agent planning approach for controlling the multi-agent team.

## Assumptions and Notations

**Assumptions:** We consider a team of  $S$  agents surveying a large area to detect and track an unknown and time-varying number of mobile objects using detection-based measurements. We assume that each agent can localize itself (*e.g.*, using an onboard GPS for UAVs) and that all agents can communicate to a central node to enables us to adopt the centralized approach for MPOMDP. Consequently, we assume that all of the measurements are transferred to a central node that analyzes received information and subsequently sends control actions to all of the agents. Here, we employ a discrete control action space to reduce the computational load (Beard et al. 2017; Dames, Tokekar, and Kumar 2017). We further assume that the measurements from an object collected by the agents are conditionally independent given the object’s state (Charrow, Michael, and Kumar 2015; Thrun, Burgard, and Fox 2005).

**Notations:** We use the convention that lower-case letters (*e.g.*,  $x$ ) represent single-object states, upper-case letters (*e.g.*,  $X$ ) represent multi-object (finite-set) states, and blackboard bold letters (*e.g.*,  $\mathbb{X}, \mathbb{Z}$ ) represent spaces. We denote the inner product  $\int f(x)g(x)dx = \langle f, g \rangle$ .

## Multi-sensor Bernoulli filter (MS-BF)

In practice, an object can randomly enter and leave the surveillance region, hence the number of objects of interest is unknown and time-varying. Further, it is important to consider the existence of objects of interest to allow the agents to discover new objects when they enter the scene and to prevent agents following false-positives. This can be addressed by the random finite sets (RFSs) approach, first proposed by (Mahler 2007). RFSs are finite-set valued random variables. We assume that each measurement is uniquely identified, *e.g.*, transmit frequencies from radio beacons (Cliff et al. 2015;

Kays et al. 2011; Nguyen et al. 2019a; 2019b; Thomas, Holland, and Minot 2012) or MAC address (Beck et al. 2018; Charrow, Michael, and Kumar 2015), which is typical for wildlife tracking or search and rescue missions. Since each object is uniquely identified, we propose using a the multi-sensor Bernoulli filter (MS-BF) (Vo et al. 2012), where each object's state is a Bernoulli RFS, and run multiple MS-BF filters parallelly to track multiple objects. A Bernoulli RFS  $X$  on  $\mathbb{X}$  has at most one element with probability  $r$  for being singleton or  $1 - r$  for being empty. Its probability density  $\pi(\cdot) = (r, p(\cdot))$  given by

$$\pi(X) = \begin{cases} 1 - r, & X = \emptyset, \\ r \cdot p(x), & X = \{x\}. \end{cases}$$

**Object tracking with MS-BF:** We model each object's state at time  $k$  by  $X_k$  as a Bernoulli RFS. The MS-BF propagates the two quantities: the existence probability  $r$  and spatial density  $p(\cdot)$ . If the posterior density is  $\pi_{k-1} = (r_{k-1}, p_{k-1})$ , then the predicted density  $\pi_{k|k-1} = (r_{k|k-1}, p_{k|k-1})$  is also a Bernoulli RFS, with  $r_{k|k-1} = r_{B,k}(1 - r_{k-1}) + r_{k-1} \langle p_{S,k}, p_{k-1} \rangle$ ;  $p_{k|k-1}(x_k) = [r_{B,k}(1 - r_{k-1})b_k(x_k) + r_{k-1} \langle f_{k|k-1}(x_k|\cdot), p_{S,k}(\cdot)p_{k-1}(\cdot) \rangle] / r_{k|k-1}$ . Here,  $r_{B,k}$  and  $p_{S,k}$  are the probabilities of object birth and object survival,  $b_k(\cdot)$  is the object birth density. Further, the updated density  $\pi_k$  is also a Bernoulli RFS, given by  $\pi_k = (r_k, p_k)$  with  $r_k = \Psi_k^{(S)} \circ \dots \circ \Psi_k^{(1)}(r_{k|k-1})$ ;  $p_k = \Psi_k^{(S)} \circ \dots \circ \Psi_k^{(1)}(p_{k|k-1})$ . Here,  $\circ$  denotes composition (of operators),  $\Psi_k^{(s)}$  is an update operator for agent  $s$ , *i.e.*:

$$[\Psi_k^{(s)}(r)] = \langle \eta^{(s)}(Z^{(s)}|\cdot), p(\cdot) \rangle r / [(1 - r)e^{-\lambda^{(s)}} + r \langle \eta^{(s)}(Z^{(s)}|\cdot), p(\cdot) \rangle], \quad (1)$$

$$[\Psi_k^{(s)}(p)](x) = \eta^{(s)}(Z^{(s)}|x)p(x) / \langle \eta^{(s)}(Z^{(s)}|\cdot), p(\cdot) \rangle \quad (2)$$

where the superscript  $(s)$  denotes the parameters of agent  $s$ ,  $\lambda^{(s)}$  is the clutter rate, and  $\eta^{(s)}(Z^{(s)}|x)$  denotes the likelihood of measurement set  $Z^{(s)}$  from agent  $s$  given the object's state  $x$ .  $\eta^{(s)}(Z^{(s)}|x)$  is also a Bernoulli RFS, given by

$$\eta^{(s)}(Z^{(s)}|x) = \begin{cases} 1 - p_d^{(s)}(x), & Z^{(s)} = \emptyset, \\ p_d^{(s)}(x)g^{(s)}(z|x), & Z^{(s)} = \{z\}. \end{cases}$$

Here,  $p_d^{(s)}(x)$  is the probability that agent  $s$  detects object  $x$ , and  $g^{(s)}(z|x)$  is the (conventional) likelihood function of measurement  $z$  given object's state  $x$ .

## Planning

At time  $k$ , the team of  $S$  agents needs to plan how they manoeuvre over the time interval  $k + 1 : k + H$  to improve its estimation of the states of multiple objects  $X_k$ , where  $H$  denotes the look-ahead horizon length. Let  $\mathbb{A} \subseteq \mathbb{R}^N$  be all possible set of control actions for a given agent. When the control action  $a_k^i \in \mathbb{A}$  is applied to an agent  $i$ , it follows a trajectory comprised of sequence of the discrete poses  $u_{k+1:k+H}^i(a_k^i) = [u_{k+1}^i, \dots, u_{k+H}^i]^T$  with corresponding measurements  $Z_{k+1:k+H}^i(a_k^i) = [Z_{k+1}^i, \dots, Z_{k+H}^i]^T$  (for notational compactness, we omit the dependence on  $X_k$

here). Let  $A_k = [a_k^1, \dots, a_k^S]^T \in \mathbb{A}^S$  be the control actions where  $\mathbb{A}^S = \mathbb{A} \times \dots \times \mathbb{A}$  is the control action space for  $S$  agents, and the corresponding measurement set is  $Z_{k+1:k+H}(A_k) = [Z_{k+1:k+H}^1(a_k^1), \dots, Z_{k+1:k+H}^S(a_k^S)]^T$ .

The objective of path planning is to find the optimal action  $A_k^* \in \mathbb{A}^S$  that maximizes the value function, *i.e.*,

$$A_k^* = \arg \max_{A_k \in \mathbb{A}^S} V(X_{k+1:k+H}, Z_{k+1:k+H}(A_k)). \quad (3)$$

where  $V(X_{k+1:k+H}, Z_{k+1:k+H}(A_k)) = \mathbf{E} \left[ \sum_{j=1}^H \mathcal{R}(X_{k+j}, Z_{k+j}(A_k)) \right]$  is the value function or the expected sum of immediate rewards  $\mathcal{R}(\cdot)$  over a finite horizon  $H$ . Since an analytic solution does not exist for the expected reward, we use the predicted ideal measurement set (PIMS) (Mahler 2004)—a computationally low-cost approach. The value function is calculated:

$$V(X_{k+1:k+H}, \hat{Z}_{k+1:k+H}(A_k)) = \sum_{j=1}^H \mathcal{R}(X_{k+j}, \hat{Z}_{k+j}(A_k))$$

where  $\hat{Z}_{k+j}(A_k)$  denotes the ideal measurement set of  $Z_{k+j}(A_k)$  calculated using the measurement model and the estimated states of objects without measurement noise. For notational compactness, we write the value function  $V(X_{k+1:k+H}, \hat{Z}_{k+1:k+H}(A_k))$  as  $V(A_k)$ .

## Planning for Tracking and Discovering Multiple Objects

### Planning for tracking discovered mobile objects

In this problem, we consider maximizing an information-based reward function to reduce the overall uncertainty of the discovered mobile objects because more information naturally implies less uncertainty. In particular, we propose using the mutual information  $I(X; Z)$  between the object's state  $X$  and measurement state  $Z$  as the immediate reward function, and the long-term sum of rewards over a finite horizon  $H$ , or so-called the value function is given by

$$V_1(A_k) = \sum_{j=1}^H I(X_{k+j}; \hat{Z}_{k+j}(A_k)) \quad (4)$$

where  $I(X; Z) = h(X) - h(X|Z)$ , with  $h(X)$  is the generalization of differential entropy for a finite set  $X \subseteq \mathbb{X}$  with density  $f(X)$  defined as  $h(X) = - \int_{\mathbb{X}} f(X) \log f(X) \delta X$ ; here  $\int_{\mathbb{X}} \cdot \delta X$  is the set integral (Mahler 2007). For Bernoulli RFS, this integration is simplified to  $h(X) = - [f(X = \emptyset) \log f(X = \emptyset) + \int f(X = x) \log f(X = x) dx]$ . We have the following theorem.

**Theorem 1.** *The mutual information  $I(X; Z)$  between the object state  $X$  and measurement state  $Z$  is a monotone submodular set function of  $Z$ .*

*Proof.* We want to prove that this mutual information  $I(X; Z)$  is a monotone submodular set function, *i.e.*, for  $Z_1 \subseteq Z_2 \subseteq \mathbb{Z}$ , and  $z \in \mathbb{Z} \setminus Z_2$  independent of  $Z_1$  and  $Z_2$ :

$$I(X; Z_2, \{z\}) - I(X; Z_2) \leq I(X; Z_1, \{z\}) - I(X; Z_1).$$

Since  $Z_1 \subseteq Z_2 \subseteq \mathbb{Z}$ , using mutual information inequalities (Cover and Thomas 2012, p.50), we have:

$$\begin{aligned} I(Z_2; \{z\}) &\geq I(Z_1; \{z\}), \\ \Leftrightarrow h(z) - h(z|Z_2) &\geq h(z) - h(z|Z_1), \\ \Leftrightarrow h(z|Z_1) &\geq h(z|Z_2), \\ \Leftrightarrow h(Z_1, \{z\}) - h(Z_1) &\geq h(Z_2, \{z\}) - h(Z_2). \end{aligned} \quad (5)$$

Further, since  $I(Z_2; \{z\}|X) = I(Z_1; \{z\}|X) = 0$  is due to  $z$  is independent of  $Z_1$  and  $Z_2$  given  $X$ , we have:

$$\begin{aligned} h(\{z\}|X) &= h(\{z\}|X, Z_2) + I(Z_2; \{z\}|X) = h(\{z\}|X, Z_2) \\ &= h(X, Z_2, \{z\}) - h(X, Z_2), \\ h(\{z\}|X) &= h(\{z\}|X, Z_1) + I(Z_1; \{z\}|X) \\ &= h(X, Z_1, \{z\}) - h(X, Z_1). \end{aligned}$$

Hence,

$$h(X, Z_2, \{z\}) - h(X, Z_2) = h(X, Z_1, \{z\}) - h(X, Z_1). \quad (6)$$

Subtracting (5) from (6), we have:

$$\begin{aligned} [h(X, Z_2, \{z\}) - h(X, Z_2)] - [h(Z_2, \{z\}) - h(Z_2)] \\ \geq [h(X, Z_1, \{z\}) - h(X, Z_1)] - [h(Z_1, \{z\}) - h(Z_1)] \end{aligned}$$

Using differential entropy chain rules (Cover and Thomas 2012, p.253), we have that  $h(X|Z_2, \{z\}) = h(X, Z_2, \{z\}) - h(Z_2, \{z\})$  and  $h(X|Z_2) = h(X, Z_2) - h(Z_2)$ , thus the above equation is equivalent to

$$\begin{aligned} h(X|Z_2, \{z\}) - h(X|Z_2) &\geq h(X|Z_1, \{z\}) - h(X|Z_1) \\ \Leftrightarrow [h(X) - h(X|Z_2, \{z\})] - [h(X) - h(X|Z_2)] \\ &\leq [h(X) - h(X|Z_1, \{z\})] - [h(X) - h(X|Z_1)], \\ \Leftrightarrow I(X; Z_2, \{z\}) - I(X; Z_2) &\leq I(X; Z_1, \{z\}) - I(X; Z_1). \end{aligned}$$

Thus,  $I(X; Z)$  is a submodular set function. Further, using the chain rule we have:

$$I(X; Z_2, \{z\}) - I(X; Z_2) = I(X; Z_2|\{z\}) \geq 0$$

Therefore,  $I(X; Z)$  is a monotone submodular set function.  $\square$

**Remark:** Our mutual information formulation is different to that in (Krause, Singh, and Guestrin 2008) used for sensor selection problems. Krause et al. showed that for  $Z \subseteq \mathbb{Z}$ , the mutual information  $I(Z; \mathbb{Z} \setminus Z)$  is a submodular set function. In other words, the mutual information  $I(Z_1; Z_2)$  is submodular with the *property* that  $Z_1 \cup Z_2 = \mathbb{Z}$  and  $|\mathbb{Z}|$  is fixed. In contrast, we measure the mutual information between the random set object state  $X$  and the random set measurement state  $Z$  and prove  $I(X; Z)$  is also a submodular set function of  $Z$  without the aforementioned *property*.

**Corollary 2.** *The value function  $V_1(A_k)$  in (4) is a monotone submodular set function.*

*Proof.* Since  $I(X_k; \hat{Z}_{k+j}(A_k))$  is a monotone submodular set function and  $V_1(A_k)$  is a positive linear combination of it, according to (Nemhauser, Wolsey, and Fisher 1978, p.272),  $V_1(A_k)$  is a monotone submodular set function.  $\square$

**Mutual Information Calculation based on MS-BF:** Assume that each object  $i$  is associated with a Bernoulli distribution  $\pi(X_i) = (r_i, p_i)$ . Let  $p_i(x)$  be approximated by a set of  $N_s$  particles, such that  $p_i(x) \approx \sum_{m=1}^{N_s} w_i^{(m)} \delta(x^{(m)} - x)$

with  $\sum_{m=1}^{N_s} w_i^{(m)} = 1$  and  $\delta(\cdot)$  is the Kronecker delta function;  $X = X_1 \cup \dots \cup X_n$  be the state of multiple objects. Since each object is uniquely identified by its label and estimated by an individual Bernoulli filter, we have

$$\begin{aligned} h(X) &= \sum_i^n h(X_i) \approx \sum_i^n \left[ - (1 - r_i) \log(1 - r_i) \right. \\ &\quad \left. - r_i \sum_{m=1}^{N_s} [w_i^{(m)} \log(r_i w_i^{(m)})] \right]. \end{aligned} \quad (7)$$

According to the definition of the mutual information  $I(X; Z) = h(X) - h(X|Z)$ , thus the tracking value function  $V_1(A_k)$  can be calculated as  $V_1(A_k) = \sum_{j=1}^H [h(X_{k+j}) - h(X_{k+j}|\hat{Z}_{k+j}(A_k))]$ , where  $h(X_{k+1})$  is calculated directly in (7). For  $h(X_{k+j}|\hat{Z}_{k+j}(A_k))$ , it has the same form as in (7); however,  $r_{k+j,i}$  and  $w_{k+j,i}^{(m)}$  are calculated by propagating  $r_{k,i}$  and  $w_{k,i}^{(m)}$  from time  $k$  to  $k+j$  using (1) and (2) respectively with the ideal measurements  $\hat{Z}_{k+j}(A_k)$ .

## Planning to search for undiscovered mobile objects

**Occupancy Grid Filter:** Since an agent is equipped with a sensor with a limited detection range, we propose using an occupancy grid to represent the probability of any undiscovered objects (Elfes 1989). We extend the static grid approach in (Charrow, Michael, and Kumar 2015; Thrun, Burgard, and Fox 2005) by incorporating the birth probability into each occupancy cell to account for the possibilities of mobile objects entering and leaving the survey area, anytime. The survey area is divided into an occupancy grid  $G = \{g^1, \dots, g^{N_g}\} \subset \mathbb{R}^N$ , where each cell  $g^i \in G$  associated with a Bernoulli random variable  $r^i$ . Here,  $r^i$  is the probability that cell  $g^i$  contains at least one undiscovered object. For initialization, we set  $r_0^i = r_B$  such that every cell has the same prior. Each cell  $i$  is propagated through MS-BF over time using the predict and update equations. In particular, let  $r_{k-1}^i$  be the probability of cell  $g^i$  containing at least one undiscovered object, then its predict and update probabilities at time  $k$  are (8) and (9). Note that since these objects are yet to be discovered, we use empty measurements for all agents (denoted as  $Z^0$  to update).

$$r_{k|k-1}^i = r_B(1 - r_{k-1}^i) + r_{k-1}^i p_S, \quad (8)$$

$$r_k^i = \Psi_k^{(S)} \circ \dots \circ \Psi_k^{(1)}(r_{k|k-1}^i), \quad (9)$$

where  $[\Psi_k^{(s)}(r^i)] = (1 - p_d^{(s)}(g^i))r^i / [1 - r^i + r^i(1 - p_d^{(s)}(g^i))]$ .

**Searching for undiscovered objects:** As before, we propose using mutual information as the immediate reward function. We want to maximize the mutual information between the

estimated occupancy grid  $G$  and the ideal empty future measurement  $\hat{Z}_{k+1:k+H}^\emptyset(A_k)$ , i.e.,

$$V_2(A_k) = \sum_{j=1}^H I(G_k; \hat{Z}_{k+j}^\emptyset(A_k)), \quad (10)$$

where  $I(G_k; \hat{Z}_{k+j}^\emptyset(A_k)) = \mathcal{H}(G_k) - \mathcal{H}(G_k | \hat{Z}_{k+j}^\emptyset(A_k))$  and  $\mathcal{H}(G_k)$  is the Shannon entropy of  $G_k$ :

$$\mathcal{H}(G_k) = - \sum_{i=1}^{N_g} [r_k^i \log(r_k^i) + (1 - r_k^i) \log(1 - r_k^i)], \quad (11)$$

and  $\mathcal{H}(G_{k+j} | \hat{Z}_{k+j}^\emptyset(A_k))$  has the same form as in (11) with  $r_{k+j}^i$  is calculated by propagating  $r_{k+j}^i$  from  $k$  to  $k+j$  using the update step in (9) with empty measurements  $\hat{Z}_{k+j}^\emptyset(A_k)$ .

**Theorem 3.** *The value function  $V_2$  in (10) is a monotone submodular set function.*

*Proof.* We can apply a similar strategy as per Theorem 1 to prove that  $I(G_k; \hat{Z}_{k+j}^\emptyset(A_k))$  is a monotone submodular set function, note that  $\mathcal{H}(\cdot)$  is the Shannon entropy (a discrete version of differential entropy  $h(\cdot)$ ). Further, since  $V_2(A_k)$  is a positive linear combination of  $I(G_k; \hat{Z}_{k+j}^\emptyset(A_k))$ , according to (Nemhauser, Wolsey, and Fisher 1978, p.272),  $V_2(A_k)$  is a monotone submodular set function.  $\square$

## Multi-objective value function for tracking and discovering

In this problem, we want to control the team of agents to perform both tracking and discovering; this naturally leads to a multi-objective problem. Specifically, we want to maximize

$$V(A_k) = [V_1(A_k), V_2(A_k)]^T$$

subject to  $A_k \in \mathbb{A}^S$  where  $V_1$  and  $V_2$  are defined in (4) and (10), respectively. Multi-objective optimization provides a meaningful notion of multi-objective optimality such as the Pareto-set, which represents trade-offs between the objectives such that there is no other solution that can improve one objective without degrading any remaining objectives (Whitson and Roijers 2016). Online planning necessitates selecting one compromised solution from the Pareto-set on-the-fly. One approach is Robust Submodular Observation Selection (RSOS) (Krause et al. 2008), which is robust against the worst possible objective; however, even if each  $V_i$  is submodular,  $V_{min} = \min_i V_i$  is generally not submodular. Other approaches include Weighted Sum (WS) and Global Criterion Method (GCM); simplicity of these methods are not only attractive for meeting the demands of online planning but also result in a submodular value function. In this work, we adopt GCM to select the compromised solution considering the distance from the ideal solution. Inspired by (Koski 1993), we define the value function  $V_{mo}$  (with  $V_{mo}(\emptyset) = 0$ ) as:

$$V_{mo}(A_k) = \sum_{i=1}^2 \frac{V_i(A_k) - \min_{A_k \in \mathbb{A}^S} V_i(A_k)}{\max_{A_k \in \mathbb{A}^S} V_i(A_k) - \min_{A_k \in \mathbb{A}^S} V_i(A_k)}. \quad (12)$$

The global criterion method admits a unique optimal solution from the Pareto-set (Coello et al. 2007). Hence, the multi-objective problem becomes

$$A_k^* = \arg \max_{A_k \in \mathbb{A}^S} V_{mo}(A_k). \quad (13)$$

Since finding the optimal control action  $A^* \in \mathbb{A}^S$  is a combinatorial optimization problem, we want to show that the multi-objective value function  $V_{mo}(A)$  in (12) is also a monotone submodular set function on  $Z$ . This enables us to use the greedy algorithm to find the optimal action that approximately maximize this multi-objective value function.

**Corollary 4.** *The multi-objective value function  $V_{mo}$  in (12) is a monotone submodular set function.*

*Proof.* Since  $V_i(A_k)$  is a monotone submodular set function and  $V_{mo}(A_k)$  is a positive linear combination of it, according to (Nemhauser, Wolsey, and Fisher 1978, p.272),  $V_{mo}(A_k)$  is a monotone submodular set function.  $\square$

## Greedy search algorithm

We proved our multi-objective value function  $V_{mo}(\cdot)$  is a monotone submodular set function—see **Corollary 4**. For submodular functions, (Nemhauser, Wolsey, and Fisher 1978) proved the greedy search algorithm guarantees a performance bound at  $(1 - 1/e)$ OPT, where OPT is the optimal value of the submodular function. Therefore, if the optimal value of our value function is  $V_{mo}(A^*)$ , we can simply state the following fundamental performance bound for our submodular value function:

**Theorem 5.** *From (Nemhauser, Wolsey, and Fisher 1978). Let  $A_G$  be the output greedy control action and  $A^*$  be the optimal control action evaluated using brute-force method of (13). Then*

$$V_{mo}(A_G) \geq (1 - 1/e)V_{mo}(A^*) \quad (14)$$

where  $e = 2.718 \dots$  is the base of the natural logarithm.

Hence, we propose using the greedy search algorithm—see the work by (Dames, Tokekar, and Kumar 2017)—by simply adding agents sequentially and picking the next agent which provides the maximum value function  $V_{mo}(\cdot)$ .

## Experiments

We evaluate the proposed value function using a series of comprehensive synthetic experiments since we can control all of the parameters of the problem, especially with a time-varying number of agents and objects. We compare three planning algorithm formulations: **(i)** using the single objective value function  $V_1(\cdot)$  in (4) for tracking. **(ii)** using a single objective value function based on our *new* discovery value function  $V_2(\cdot)$  in (10). **(iii)** using our proposed multi-objective value function  $V_{mo}(\cdot)$ .

We use optimal sub-pattern assignment (OSPA) (Schuhmacher, Vo, and Vo 2008) to measure performance. We report **OSPA Dist** as the *main* metric to evaluate the *overall* team performance since it incorporates both tracking and discovery indicators. For further insights into our planning formulations, we also report: **(i) OSPA Loc** as a localization

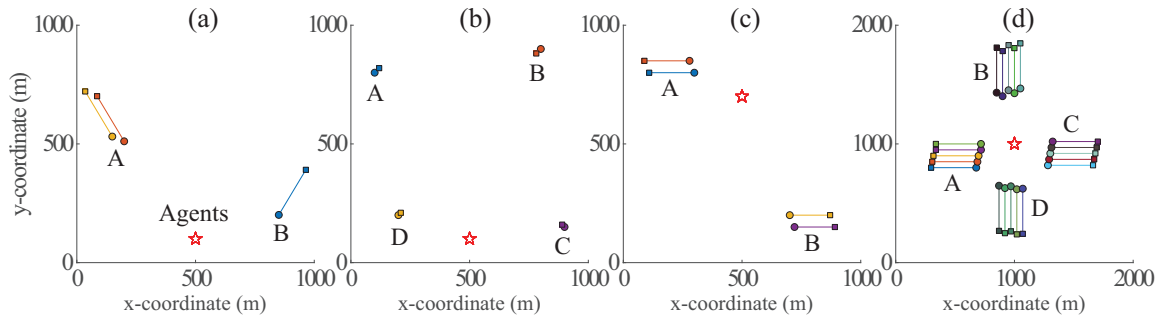


Figure 1: Setup for four scenarios: (a) Scenario 1; (b) Scenario 2; (c) Scenario 3; (d) Scenario 4. Start/Stop locations for each object are denoted by  $\circ/\square$ . Start locations for agents are denoted by  $\star$ .

accuracy measure, (ii) **OSPA Card** as an object discovery performance measure; and (iii) **Search Area Entropy** as the average entropy of the occupancy grid to measure the coverage area of the team. For demonstration, a team of quad-copter UAVs flying at different altitudes is considered. The detailed parameter settings are provided in the appendix, while scenario setups are shown in Figure 1. Our experiments considered four different scenarios and two different detection-based sensors subject to noisy measurements.

**Scenario 1 (FastMoving):** *Three fast moving objects in two groups travelling in the same direction.* A team of agents starts at  $[500, 100]^T$  m as depicted in Figure 1a.

**Scenario 2 (LateBirth):** *Late birth objects.* We investigate a searching and tracking scenario in Figure 1b) with four slow-moving mobile objects using a team of agents. Here, the groups of objects  $D$  and  $C$  enter the scene when the agents are out of their detection range—late birth. This scenario favours agent planning with the discovery value function encouraging exploration and demonstrates the effectiveness of our multi-objective value function with its competing tracking and discovery objectives.

**Scenario 3 (Opposite):** *Four objects in two groups (A and B) moving rapidly in opposing directions.* Figure 1c illustrates the scenario. We use this setting to confirm the effectiveness of our multi-objective value function. Now, the possibility to discover group  $B$  out of the sensor detection range must be achieved through exploration while planning to track group  $A$  in the vicinity of the agents is immediately rewarded by the tracking objective.

**Scenario 4 (Explosion):** *Multiple groups of fast moving objects in opposing directions.* Here, we consider a team of agents to search and track 20 fast moving mobile objects as shown in Figure 1d.

**Detection-based sensors:** (i) We considered agents equipped with a range and bearing based sensor—common in wildlife tracking (Cliff et al. 2015) for example. Let  $u^s = [p_x^s, p_y^s, p_z^s]^T$  be the position of agent  $s$ ,  $x_p = [p_x, p_y, 1]^T$  be the position of object  $x$ , each detected object  $x$  leads to a noisy measurement  $z$  of range and bearing given by:  $z = [\arctan[(p_y - p_y^s)/(p_x - p_x^s)], \|x_p - u^s\|]^T + v$ . Here,  $\|\cdot\|$  is the Euclidean norm;  $v \sim \mathcal{N}(0, R)$  with  $R = \text{diag}(\sigma_\phi^2, \sigma_\rho^2)$  where  $\sigma_\phi = \sigma_{0,\phi} + \beta_\phi \|x_p - u^s\|$ ,  $\sigma_\rho = \sigma_{0,\rho} + \beta_\rho \|x_p - u^s\|$ . (ii) To demonstrate the sensor-agnostic nature of our ap-

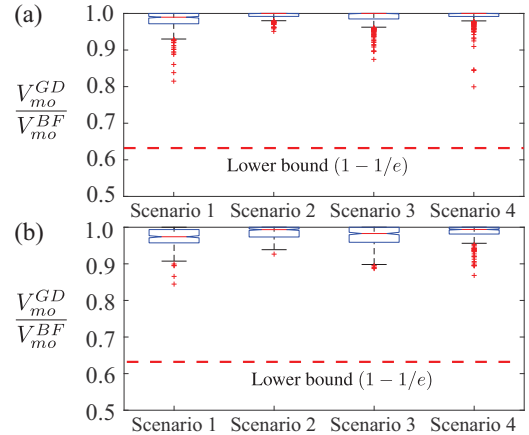


Figure 2: Multi-objective value function ratio between the greedy  $V_{mo}^{GD}$  and brute-force  $V_{mo}^{BF}$  algorithms with agents (a)  $S = 2$  and (b)  $S = 3$  (20 MC runs, range and bearing based sensor with  $r_d = 200$  m).

proach, we consider agents equipped with a vision-based sensor. Each detected object  $x$  leads to a measurement  $z$  of noisy  $xy$  positions, given by:  $z = [p_x, p_y]^T + v$ . Here,  $v \sim \mathcal{N}(0, R)$  with  $R = \text{diag}(\sigma_x^2, \sigma_y^2)$ .

**Experiment 1 Comparing greedy and brute force algorithm results for our submodular multi-objective value function:** Figure 2 depicts the ratio of our multi-objective value function obtained from greedy and brute-force algorithms for the four scenarios. The result obtained from 20 Monte-Carlo (MC) runs for each scenario agrees with the performance guarantee of the greedy algorithm to yield an approximately optimal solution with a bounded performance guarantee at  $(1 - 1/e)$  OPT.

**Experiment 2: Comparing multi-objective multi-agent planning with single objective multi-agent planning.** Table 1 compares results for scenario 1, 2, 3 and 4 collected from 20 MC runs for agents with *range and bearing* based sensors. It is expected that the average search area entropy is smallest for  $V_2$  since it encourages agents to explore the search area. Consequently,  $V_2$  can also be seen to generate the best performance in term so of OSPA cardinality—**OSPA Card**. In contrast, we can see that the multi-agent planning with

Table 1: Comparing multi-agent planning for tracking mobile objects using our multi-objective value function  $V_{mo}$  across Scenarios 1, 2, 3 and 4 with detection range  $r_d = 200$  m.  $V_1$  and  $V_2$  are baselines and the results are averaged over 20 MC trials.

		Scenario 1 (FastMoving)				Scenario 2 (LateBirth)			
Indicators		Overall Performance	Tracking Performance	Discovery Performance		Overall Performance	Tracking Performance	Discovery Performance	
Agents	Value Functions	OSPA Dist (m)	OSPA Loc (m)	OSPA Card (m)	Search Area Entropy (nats)	OSPA Dist (m)	OSPA Loc (m)	OSPA Card (m)	Search Area Entropy (nats)
	$S = 3$	$V_1$	33.9	4.4	29.5	0.23	57.0	4.0	53.0
$V_2$		21.2	9.7	11.5	0.12	<b>41.1</b>	10.3	30.8	0.12
$V_{mo}$		<b>17.7</b>	6.1	11.6	0.17	52.1	5.2	46.9	0.17
$S = 5$	$V_1$	25.4	5.1	20.3	0.2	53.4	3.6	49.8	0.17
	$V_2$	20.3	9.2	11.1	0.09	43.9	9.5	34.4	0.09
	$V_{mo}$	<b>16.8</b>	5.7	11.1	0.13	<b>38.8</b>	5.1	33.7	0.11
		Scenario 3 (Opposite)				Scenario 4 (Explosion)			
$S = 3$	$V_1$	51.7	3.0	48.7	0.24	53.7	3.1	50.6	0.32
	$V_2$	18.3	12.8	5.5	0.12	55.0	29.8	25.2	0.28
	$V_{mo}$	<b>11.1</b>	5.9	5.2	0.18	<b>40.7</b>	9.7	31.0	0.32
$S = 5$	$V_1$	51.2	2.9	48.3	0.24	36.9	5.0	31.9	0.30
	$V_2$	<b>10.5</b>	6.9	3.6	0.09	31.2	19.5	11.7	0.25
	$V_{mo}$	<b>10.5</b>	5.9	4.1	0.15	<b>17.4</b>	6.4	11.1	0.29

the single value function (to encourage only tracking accuracy)  $V_1$ , achieves improved results for object localization accuracy only (low **OSPA Loc** results) but at the expense of missing objects often out of the range of the sensors (as seen by significantly large **OSPA Card** results). Most importantly, our results verify that  $V_{mo}$  performs best in term of overall tracking and cardinality accuracy (reported by **OSPA Dist**) since  $V_{mo}$  not only rewards agents for undertaking the discovery of new objects but also rewards agents for accurately tracking discovered objects.

Figure 3 shows the grid occupancy probability and the trajectories of the agents for scenario 3. The results demonstrate the effectiveness of our proposed planning method, where agents not only track but discover distant mobile objects.

**Experiment 3:** Explore the asymptotic behavior of tracking performance with an increasing number of agents for our planning formulation. Figure 4 depicts the overall mean tracking accuracy from 20 MC runs for agent teams with each detection-based sensor. It confirms that planning with  $V_{mo}$  consistently performs better than  $V_1$  or  $V_2$  alone. As expected, when the number of agents increases,  $V_1$  and  $V_2$  tracking performances improve and approach that of  $V_{mo}$ . Interestingly, multi-agent planning with a single exploration objective closely approaches the tracking performance of the multi-objective value function when the team of agents is large enough to cover the survey area with its range limited sensors and all objects become visible to the agents.

## Conclusion

In this paper, we have formulated a multi-objective planning approach for multi-agent tracking and searching for mobile objects. We have established that our formulation results in a value function that is monotone and submodular. We presented a series of extensive experimental results to demonstrate the effectiveness of our method and performance guarantees when using the low-cost greedy algorithm to determine control actions for the multi-agent.

We consider a centralized MPOMDP approach where scalability can be a limitation. Factored-POMDP (Oliehoek et

al. 2008) can be employed to achieve further system scalability. We require reliable and fast communications between all agents and their centralized controller. If there are any delays in communications, the problem can be formulated as an MPOMDP with delayed communications. It is extremely challenging to plan and track mobile objects in an online manner without any communications among agents as in Dec-POMDP, hence it is an open question for future work.

## Acknowledgments

This work was jointly supported by the Western Australia Parks and Wildlife (WA Parks), the Australian Research Council (LP160101177, DP160104662), the Defense Science and Technology Group (DSTG), and The Shultz Foundation.

## Appendix

### Parameter settings for experiments

The search areas for the first three scenarios and scenario 4 are  $1000 \text{ m} \times 1000 \text{ m}$  and  $2000 \text{ m} \times 2000 \text{ m}$ , respectively. Each agent is controlled to fly at a fixed and different altitude (*i.e.*, 5 m altitude gap between each agent) to prevent collisions with other team members. The minimum altitude starts at 30 m for the first agent and increases 5 m for each additional agent. Further, all objects are assumed exist on a horizontal ground plane to speed up the numerical experiments by tracking in 2D. Each object state  $\mathbf{x} = (x, l)$  is uniquely identified by its label  $l$ , while its motion state  $x = [p_x, \dot{p}_x, p_y, \dot{p}_y]^T$  comprises of object's position and velocity in Cartesian coordinates. Each object moves in accordance with the constant velocity (CV) model given by  $x_k = F^{CV} x_{k-1} + q_{k-1}^{CV}$ . Here,  $F^{CV} = [1, T_0; 0, T_0] \otimes I_2$ ,  $T_0$  is the sampling interval ( $T_0 = 1$  s for our experiments),  $\otimes$  denotes for the Kronecker tensor product;  $I_2$  is the  $2 \times 2$  identity matrix;  $q_{k-1}^{CV} \sim \mathcal{N}(0, Q^{CV})$  is a  $4 \times 1$  zero mean Gaussian process noise, with co-variance  $Q^{CV} = \sigma_{CV}^2 [T_0^3/3, T_0^2/2; T_0^2/2, T_0] \otimes I_2$ . The detection probability is  $p_D(u_s, x_p) =$

$$\begin{cases} 0.98 & \|x_p - u^s\| \leq r_d \\ \max(0, 0.98 - (\|x_p - u^s\| - r_d)\hbar) & \text{otherwise;} \end{cases}$$

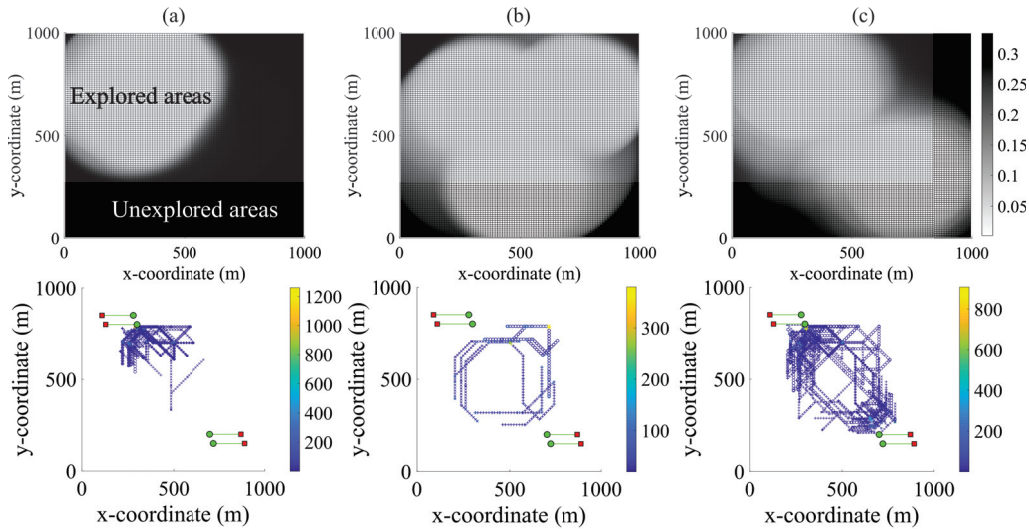


Figure 3: Scenario 3. Grid occupancy probability (top) and heat map (bottom) of trajectories for 3 agents over 20 MC runs with  $r_d = 200$  m using (a)  $V_1$ . Late birth group  $B$  never discovered, (b)  $V_2$ . Extensive exploration, and (c)  $V_{mo}$ . Discovers the late birth group  $B$  whilst tracking both groups.

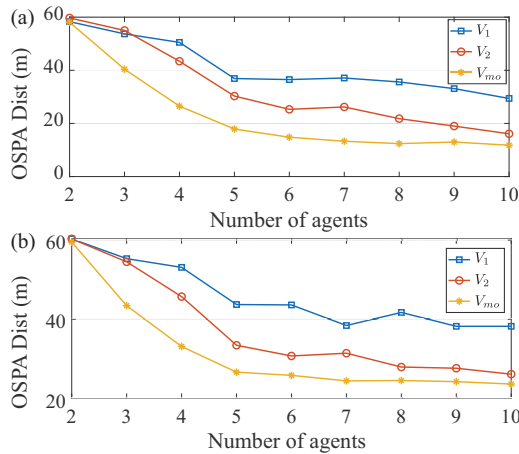


Figure 4: Overall tracking performance over 20 MC runs based on multi-agent planning with our multi-objective value function  $V_{mo}$  compare with the single objective value functions  $V_1$  and  $V_2$  when the number of agents are increased from 2 to 10 for Scenario 4 (**Explosion**) with  $r_d = 200$  m using (a) agents with *range and bearing* based sensors, (b) agents with *vision* based sensors.

where  $r_d$  is the sensor detection range and  $\hbar = 0.008 \text{ m}^{-1}$ . The sensor reports *false detections* or false-alarm measurements following a Poisson RFS with a clutter rate  $\lambda = 0.2$ , where each agent collects at most one measurement per time step for each object, either from the real objects, clutters (false detections) or the measurement is empty (missed detections). For sensor noise, the *range and bearing* based measurement is corrupted with a zero mean Gaussian process noise that depends on the distance between objects and agents, *i.e.*,  $v \sim \mathcal{N}(0, R)$  with  $R = \text{diag}(\sigma_\phi^2, \sigma_\rho^2)$  where

$\sigma_\phi = \sigma_{0,\phi} + \beta_\phi \|x_p - u^s\|$ ,  $\sigma_\rho = \sigma_{0,\rho} + \beta_\rho \|x_p - u^s\|$ ;  $\sigma_{0,\phi} = 2\pi/180$  rad,  $\beta_\phi = 1.7 \cdot 10^{-5}$  rad/m,  $\sigma_{0,\rho} = 10$  m, and  $\beta_\rho = 5 \cdot 10^{-3}$ . Similarly, for *vision-based* sensor, each detected object  $x$  leads to a measurement  $z$  of noisy  $x - y$  positions, given by:  $z = [p_x, p_y]^T + v$ . Here,  $v \sim \mathcal{N}(0, R)$  with  $R = \text{diag}(\sigma_x^2, \sigma_y^2)$  where  $\sigma_x = \sigma_y = \sigma_{0,xy} + \beta_{xy} \|x_p - u^s\|$  with  $\sigma_{0,xy} = 10$  m, and  $\beta_{xy} = 1 \cdot 10^{-2}$ . The grid size is  $100 \times 100$  across four scenarios. This corresponds to a grid cell of  $10 \text{ m} \times 10 \text{ m}$  for scenario 1,2 and 3 and a grid cell of  $20 \text{ m} \times 20 \text{ m}$  for scenario 4. The total time is 200 s. The agent does not have any prior knowledge about object's state, thus it uses the initial birth probability  $r_B = 0.005$ , and a Gaussian density  $p_B = \mathcal{N}(x; m_B, Q_B)$  with  $m_B = [500, 0, 500, 0]^T$  and  $Q_B = \text{diag}([500, 10, 500, 10])$ .

## References

- Amato, C., and Oliehoek, F. 2015. Scalable planning and learning for Multiagent POMDPs. In *AAAI*, 1995–2002.
- Araya, M.; Buffet, O.; Thomas, V.; and Charpillet, F. 2010. A POMDP extension with belief-dependent rewards. In *NeurIPS 23*, 64–72.
- Beard, M.; Vo, B.-T.; Vo, B.-N.; and Arulampalam, S. 2017. Void probabilities and Cauchy-Schwarz divergence for generalized labeled multi-Bernoulli models. *IEEE Trans. on Sig. Proc.* 5047–5061.
- Beck, Z.; Teacy, W. L.; Rogers, A.; and Jennings, N. R. 2018. Collaborative online planning for automated victim search in disaster response. *RAS* 100:251–266.
- Bernstein, D. S.; Givan, R.; Immerman, N.; and Zilberstein, S. 2002. The complexity of decentralized control of Markov decision processes. *MOS* 27(4):819–840.
- Blackman, S., and Popoli, R. 1999. *Design and analysis of modern tracking systems*. Artech House.



- Charrow, B.; Michael, N.; and Kumar, V. 2015. Active control strategies for discovering and localizing devices with range-only sensors. In *Algo. Found. of Rob.*, 55–71. Springer.
- Cliff, O. M.; Fitch, R.; Sukkariéh, S.; Saunders, D.; and Heinsohn, R. 2015. Online localization of radio-tagged wildlife with an autonomous aerial robot system. In *RSS*.
- Coello, C. A. C.; Lamont, G. B.; Van Veldhuizen, D. A.; et al. 2007. *Evolutionary algorithms for solving multi-objective problems*, volume 5. Springer.
- Cover, T. M., and Thomas, J. A. 2012. *Elements of information theory*. John Wiley & Sons.
- Dames, P., and Kumar, V. 2015. Autonomous localization of an unknown number of targets without data association using teams of mobile sensors. *IEEE Trans. on Auto. Sci. and Eng.* 12(3):850–864.
- Dames, P.; Tokekar, P.; and Kumar, V. 2017. Detecting, localizing, and tracking an unknown number of moving targets using a team of mobile robots. *IJRR* 36(13-14):1540–1553.
- Elfes, A. 1989. Using occupancy grids for mobile robot perception and navigation. *Computer* 22(6):46–57.
- Gerasenko, S.; Joshi, A.; Rayaprolu, S.; Ponnaivaikko, K.; and Agrawal, D. P. 2001. Beacon signals: what, why, how, and where? *Computer* 34(10):108–110.
- Gostar, A. K.; Hoseinnezhad, R.; and Bab-Hadiashar, A. 2016. Multi-Bernoulli sensor-selection for multi-target tracking with unknown clutter and detection profiles. *Signal Processing* 119:28 – 42.
- Hoang, H. G., and Vo, B. T. 2014. Sensor management for multi-target tracking via multi-Bernoulli filtering. *Automatica* 50(4):1135–1142.
- Kays, R.; Tilak, S.; Crofoot, M.; Fountain, T.; Obando, D.; and et al. 2011. Tracking animal location and activity with an automated radio telemetry system in a tropical rainforest. *The Computer Journal* 1931–1948.
- Koski, J. 1993. *Multicriterion Structural Optimization*. Dordrecht: Springer Netherlands. 793–809.
- Krause, A.; McMahan, H. B.; Guestrin, C.; and Gupta, A. 2008. Robust submodular observation selection. *JMLR* 9(Dec):2761–2801.
- Krause, A.; Singh, A.; and Guestrin, C. 2008. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *JMLR* 9(Feb):235–284.
- MacDermed, L. C., and Isbell, C. L. 2013. Point based value iteration with optimal belief compression for Dec-POMDPs. In *NeurIPS* 26, 100–108.
- Mahler, R. 2004. Multitarget sensor management of dispersed mobile sensors. In *Theory and algorithms for cooperative systems*, 239–310. World Scientific.
- Mahler, R. P. 2007. *Statistical Multisource-Multitarget Inf. Fusion*. Artech House, Inc.
- Messias, J. V.; Spaan, M.; and Lima, P. U. 2011. Efficient offline communication policies for Factored Multiagent POMDPs. In *NeurIPS* 24, 1917–1925.
- Murphy, R. R.; Tadokoro, S.; Nardi, D.; Jacoff, A.; Fiorini, P.; Choset, H.; and Erkmen, A. M. 2008. *Search and Rescue Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg. 1151–1173.
- Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming* 14(1):265–294.
- Nguyen, H. V.; Chesser, M.; Koh, L. P.; Rezatofghi, H.; and Ranasinghe, D. C. 2019a. TrackerBots: Autonomous unmanned aerial vehicle for real-time localization and tracking of multiple radio-tagged animals. *Journal of Field Robotics* 36(3):617–635.
- Nguyen, H. V.; Rezatofghi, H.; Vo, B.-N.; and Ranasinghe, D. C. 2019b. Online UAV path planning for joint detection and tracking of multiple radio-tagged objects. *IEEE Trans. on Sig. Proc.* 67(20):5365–5379.
- Oliehoek, F. A.; Spaan, M. T.; Whiteson, S.; and Vlassis, N. 2008. Exploiting locality of interaction in factored decpmdps. In *AAMAS*, 517–524.
- Reid, D. 1979. An algorithm for tracking multiple targets. *IEEE Trans, on Auto. Control* 24(6):843–854.
- Ristic, B., and Vo, B.-N. 2010. Sensor control for multi-object state-space estimation using random finite sets. *Automatica* 46(11):1812–1818.
- Satsangi, Y.; Whiteson, S.; Oliehoek, F. A.; and Spaan, M. T. 2018. Exploiting submodular value functions for scaling up active perception. *Autonomous Robots* 42(2):209–233.
- Schuhmacher, D.; Vo, B.-T.; and Vo, B.-N. 2008. A consistent metric for performance evaluation of multi-object filters. *IEEE Trans. on Sig. Proc.* 56(8):3447–3457.
- Silver, D., and Veness, J. 2010. Monte-carlo planning in large POMDPs. In *NeurIPS* 23, 2164–2172.
- Spaan, M. T.; Veiga, T. S.; and Lima, P. U. 2015. Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *AAMAS* 1157–1185.
- Thomas, B.; Holland, J. D.; and Minot, E. O. 2012. Wildlife tracking technology options and cost considerations. *Wildlife Research* 38(8):653–663.
- Thrun, S.; Burgard, W.; and Fox, D. 2005. *Probabilistic robotics*. MIT press.
- Vo, B. T.; See, C. M.; Ma, N.; and Ng, W. T. 2012. Multi-sensor joint detection and tracking with the Bernoulli filter. *IEEE Trans. on Aero. and Elect. Syst.* 48(2):1385–1402.
- Wai, H.-T.; Yang, Z.; Wang, P. Z.; and Hong, M. 2018. Multi-agent reinforcement learning via double averaging primal-dual optimization. In *NeurIPS* 32, 9649–9660.
- Wang, X.; Hoseinnezhad, R.; Gostar, A. K.; Rathnayake, T.; Xu, B.; and Bab-Hadiashar, A. 2018. Multi-sensor control for multi-object Bayes filters. *Signal Processing* 142:260–270.
- Whiteson, S., and Roijers, D. M. 2016. Tutorial: Multi-objective planning under uncertainty. *ICAPS*.
- Zhu, Y.; Wang, J.; and Liang, S. 2019. Multi-objective optimization based multi-Bernoulli sensor selection for multi-target tracking. *Sensors* 19(4):980.