

SMA-NBO: A Sequential Multi-Agent Planning with Nominal Belief-State Optimization in Target Tracking

Tianqi Li¹, *Student Member, IEEE*, Lucas W. Krakow² and Swaminathan Gopalswamy¹

Abstract—In target tracking with mobile multi-sensor systems, sensor deployment impacts the observation capabilities and the resulting state estimation quality. Based on a partially observable Markov decision process (POMDP) formulation comprised of the observable sensor dynamics, unobservable target states, and accompanying observation laws, we present a distributed information-driven solution approach to the multi-agent target tracking problem, namely, *sequential multi-agent nominal belief-state optimization* (SMA-NBO). SMA-NBO seeks to minimize the expected tracking error via receding horizon control including a heuristic expected cost-to-go (HECTG). SMA-NBO incorporates a computationally efficient approximation of the target belief-state over the horizon. The agent-by-agent decision-making is capable of leveraging on-board (edge) compute for selecting (sub-optimal) target-tracking maneuvers exhibiting non-myopic cooperative fleet behavior. The optimization problem explicitly incorporates semantic information defining target occlusions from a world model. To illustrate the efficacy of our approach, a random occlusion forest environment is simulated. SMA-NBO is compared to other baseline approaches. The simulation results show SMA-NBO 1) maintains tracking performance and reduces the computational cost by replacing the calculation of the expected target trajectory with a single sample trajectory based on maximum a posteriori estimation; 2) generates cooperative fleet decision by sequentially optimizing single-agent policy with efficient usage of other agents' policy of intent; 3) aptly incorporates the multiple weighted trace penalty (MWTP) HECTG, which improves tracking performance with a computationally efficient heuristic.

Index Terms—Reactive and sensor-based planning, cooperating robots, policy of intent, random occlusion forest, heterogeneous fleet

I. INTRODUCTION

Thanks to the advances in wireless technology, we are witnessing a revolution in information acquisition driven by the convenience and advantages of wireless communications. A direct impact is on tracking of targets using a network of robots equipped with sensors, which has many practical applications such as in smart cities, autonomous driving, rescue tasks, surveillance monitoring, etc. In this context, many studies have been performed on planning for enhancing and

maintaining future information gathering, such as *information driven planning* [1], *sensor-driven control* [2], or *active sensing acquisition* [3]. While the concept of information-driven control is self-explanatory in its approach to improve target tracking, the realization of this approach needs to address two fundamental issues: the uncertainty of the target trajectories, and the coordination of the sensors in the network.

Information-driven control has been studied based on different target tracking objectives. A natural and greedy objective is the maximal coverage of the sensors' field-of-view (FoV) area for richer observation [4]. However, considering the limited resources of the sensory network and the dynamic behavior of the target, maximal coverage does not guarantee the probability of detection; the Bayesian experiment design, on the other hand, plays the role of increasing such information gain from a probabilistic perspective [5]. In information theory, entropy is one solid mathematical quantification of information gathering and is treated as the objective to minimize via observation selection and action optimization [2], [5]. Another approach to target tracking seeks to plan the robot motion towards reduction of the mean squared error (MSE) of associated targets, which leads to the trace of covariance matrix $tr[\mathbf{P}]$ as the indicator of planning [6], [7], [8], [9].

The first of the aforementioned challenges is the dynamic unknown motion of the targets. Tracking a target entails the planning and consequent movement of the sensors (the robots carrying the sensor) such that they attempt to maintain the targets in the FoV to generate an accurate representation of the movement of the targets. Since the motion of the targets is unknown apriori to the sensors, it is necessary for the planner to construct future trajectories of the targets. Typically a partially observable Markov decision process (POMDP) model is utilized in studying this problem, the observable states being those of the sensors and the beliefs on the states of the targets, the unobservable states being the actual states of the targets. In order to plan the motion of the sensors, the optimization over all possible future trajectories up to a time horizon needs to be performed. To make such an optimization computationally tractable, future trajectories need to be approximated. The Monte Carlo method is studied as one solution to such approximation, which draws target trajectory samples from beliefs of the current target state and assumptions on the dynamic behavior of the target [9], [10]. The disadvantage of such an approach is the computational issue for general Monte Carlo methods since a significant number of samples are required for a good approximation. Recently this has been addressed

¹Tianqi Li and Swaminathan Gopalswamy are with the Department of Mechanical Engineering, Texas A&M University, College Station, TX 77840, USA {tianqili, sgopalswamy}@tamu.edu

²Lucas W. Krakow is with the Bush Combat Development Complex (BCDC), Texas A&M University, Bryan, TX 77807, USA lwkrakow@tamu.edu

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

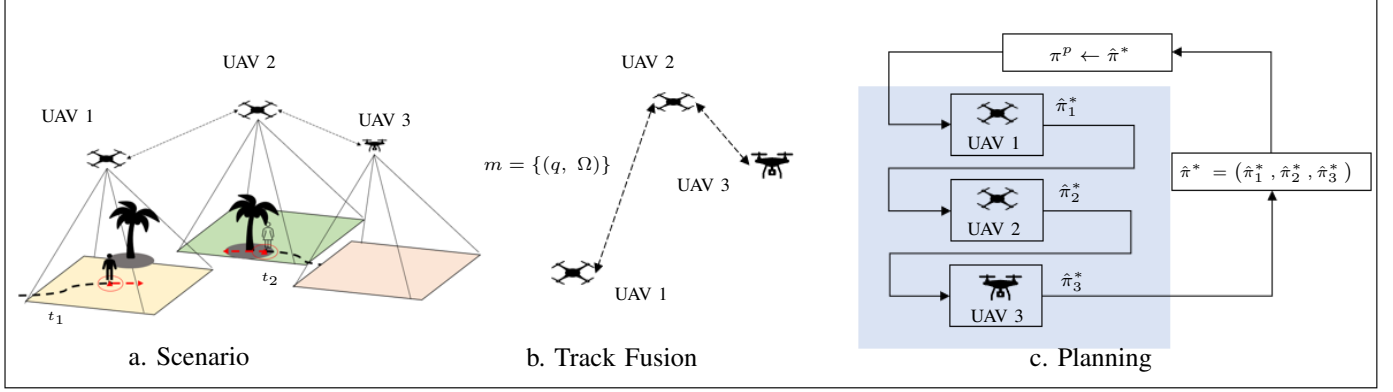


Fig. 1: The overall scenario of multi-sensor target tracking. (a) group of UAVs are tracking the OOIs (people) in an environment with occluded areas (tree shadows), the black dashed lines of targets t_1 and t_2 show the observed trajectories and the red dashed lines are nominal future trajectories for these two tracks; (b) the distributed information exchange in the team; (c) proposed SMA information flow diagram, Algorithm 3 sequentially generates $\hat{\pi}^*$ based on the intention of other agents π^P of previous decision epoch.

by generating the nominal trajectory of the targets directly from the estimated pose and dynamic behavioral models of the targets, leading to the nominal belief-state optimization (NBO) approach [6]. Our proposed approach builds on this method.

The second of the aforementioned challenges is the coordination between the sensors. In particular when the number of sensors n increases there is a corresponding exponential growth in the action space for the control over which optimization has to be performed. To avoid directly solving combinatorial decomposition, the greedy selection algorithm is studied with submodular objective functions [11], [12] which has $1 - e^{-1}$ best performance approximation. [8] studies resilience to sensor deterioration by sensor reconfiguration. The solution is a centralized mixed integer programming, and is shown to be NP-complete. A decentralized optimization approach in Dec-POMDP is designed for mobile sensory networks without communication [13]. Even though such Dec-POMDP frees the sensors from communication load, the decentralized optimization requires each sensor to optimize the joint action, which increases the total computation.

Following the logic of the discussions above, we propose a sequential method of planning that we call SMA-NBO in this paper. To address the challenges in multi-target tracking, SMA-NBO applies NBO as the target trajectory approximation method, optimizes agents' planning in a sequential method, and utilizes the decision of the previous epoch as other agents' intention. Such sequential decision making in multi-agent system has a computation complexity linear in n [14]. Specifically, the problem of target tracking by a fleet of unmanned aerial vehicles (UAV) team is addressed, using a Kalman Filter estimator based on assumptions on the dynamics of the targets, an information-driven POMDP, coordination facilitated through the SMA-NBO approach.

The primary contributions of this paper are: (i) A computationally efficient multi-agent tracking algorithm that utilizes the *policy of intent* in sequential planning, extending works from [9] that uses belief state optimization on multi-agent

tracking problem and [15] that uses NBO on a single agent tracking problem. (ii) The introduction of the concept of a random forest to automatically simulate a large number of occluded environments with different quality of occlusions, and use of the random forest to statistically demonstrate the superior performance of the proposed algorithm. (iii) The use of a heuristic *expected cost-to-go* (HECTG) that captures the expected costs beyond the fixed time used in the optimization, and demonstration of its effectiveness through the simulations in occluded environments.

II. PROBLEM DEFINITION

Consider a team of UAVs labeled by $\{1, 2, \dots, n\}$ with each UAV carrying a nadir camera as its sensor, shown as Fig. 1a. This fleet of UAVs is tasked to track objects inside an area of interest (AOI). Term *agent* is used in the remainder of the paper to describe the individual UAV in this robot team. The robot team leverages communications capability to operate in a distributed manner, with each agent i processing local observation $Z_{i,k}$ at time instant k and generating local information, then making *average-consensus* on target estimation with its neighbors in a fixed number L of *consensus steps*. The data association and consensus component is consistent with the description in [9]. After the agents' local processing and consensus steps, the fleet provides the estimation of target in set Θ_k at a constant sampling rate, Δt (frequency $f_1 = 1/\Delta t$ Hz). The message m being passed is the set of tuples of information filter, $\{(q, \Omega)\}$, seen in Fig. 1b.

In this paper, we use letter χ as the true state of object of interest (OOI), ξ as the mean value of estimated state of OOI with letter t denotes target ID; s stands for the state of UAV and i is an agent ID. We have the following assumptions in this paper.

Assumption 1: Absolute detection. We assume we have a perfect detection algorithm, such that no false alarms or missed detections occur. Even with perfect detection, observation data is still noisy (based on common sensor models). Thus, estimation and data association is still required. Specifically

we apply Joint Probabilistic Data Association which handles the target IDs, initialization and deletion of targets.

Assumption 2: Nearly constant velocity (NCV) motion model for targets. Without loss of generality, the linear Kalman filter is applied in the data association algorithm with NCV motion model, i.e., for object with state $\xi_k = (p_k^x, p_k^y, v_k^x, v_k^y) \in \mathbb{R}^4$, its dynamics is

$$\xi_{k+1} = \mathbf{F}_k \xi_k + \mathbf{w}_k, \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \quad (1)$$

with disturbance white noise on acceleration covariance σ_a in \mathbf{Q} , and the motion model \mathbf{F}_k defined as

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{Q} = \sigma_a^2 \begin{bmatrix} \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} & 0 \\ 0 & \frac{\Delta t^4}{4} & 0 & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & 0 & \Delta t^2 & 0 \\ 0 & \frac{\Delta t^3}{2} & 0 & \Delta t^2 \end{bmatrix}$$

Here p_x, p_y and v_x, v_y are the position and velocity in the x and y dimensions respectively.

Assumption 3: We assume that the appearance of our controlled agents has no impact on OOI maneuvers, i.e., $P(\chi|s) = P(\chi)$.

Assumption 4: Fixed-time consensus on information. We assume $L_0 = 5$ consensus time-steps during simulation for sufficient convergence [16]. Given sufficient consensus steps $L > L_0$, $\forall t \in \Theta_k$, the track estimation difference between any two agents $\forall i, j \in [n]$ is small enough, i.e., for the posterior Gaussian distribution $(\xi_{i,k|k}, \mathbf{P}_{i,k|k})$ and $(\xi_{j,k|k}, \mathbf{P}_{j,k|k})$ maintained by agent i and j , \exists small $\epsilon_1, \epsilon_2 \in \mathbb{R}_+$, s.t. $|\xi_{i,k|k} - \xi_{j,k|k}| \leq \epsilon_1$ and $|tr[\mathbf{P}_{i,k|k}] - tr[\mathbf{P}_{j,k|k}]| \leq \epsilon_2$.

A. An Information-Driven POMDP

Similar to the work in [9], a POMDP model is formulated in this multi-agent scenario. Define a POMDP model as a tuple $\mathcal{P} = (\mathcal{X}, \mathcal{O}, \mathcal{U}, \mathcal{T}, C)$.

State \mathcal{X} : The state of POMDP contains the state of agents \mathcal{S} , state of OOIs χ and state of filter \mathcal{F} . The state of all n agents is defined as $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_n$, with agent $i \in [n]$ time instance k the state $s_{i,k} = (p_{i,k}^x, p_{i,k}^y, \psi_{i,k}, v_{i,k}^x, v_{i,k}^y)^T \in \mathbb{R}^5$ in a 2-dimensional horizontal plane including position, yaw angle and velocity. Each agent contains its field of view (FoV) $\phi_{i,k}(s)$ as a geometric variable parametrized by its state and sensor, thus the team's FoV is described as $\phi_k(s_k) = (\phi_{1,k}, \dots, \phi_{n,k})$. Additionally, with a semantic map W which contains the information of occlusion area in the AOI where no observation can be acquired by any agent, the semantic FoV is denoted as $\phi_k(s; W)$. This FoV information $\phi_k(s; W)$ with semantic map is implicitly contained in the agent state. OOIs' state χ contains all position and velocities of current OOIs. The filter state $\mathcal{F}_k = (\xi_{k|k}, \mathbf{P}_{k|k})$ maintains tracks represented by Gaussian distributions with posterior mean $\xi_{k|k}$ and covariance matrix $\mathbf{P}_{k|k}$. To avoid the redundancy of terminology, vector mean ξ and covariance matrix \mathbf{P} contains all targets, i.e., $\xi_{k|k} = (\xi_{1,k|k}, \dots, \xi_{m_k,k|k})$ and block-diagonal matrix $\mathbf{P}_{k|k} = \text{diag}(\mathbf{P}_{1,k|k}, \dots, \mathbf{P}_{m_k,k|k})$ for for total m_k targets. The dynamics of the target is based on the linear constant velocity motion (1). The POMDP state is summarized as $x_k = (s_k, \chi_k, \mathcal{F}_k)$.

Observation and Observation Law \mathcal{O} : Observation data of each agent is a set of 2-dimensional positions captured as targets. Given the state estimate of a target t at time instance k , $\xi_{t,k} \in \mathbb{R}^4$, the observation model is

$$\mathbf{z}_{t,k} = \mathbf{H}_k \xi_{t,k} + \mathbf{v}_{t,k}, \mathbf{v}_{t,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad (2)$$

\mathbf{v}_k represents the measurement noise. The *range-bearing sensor* model is applied as the observation model, for agent i with state $s_{i,k}$, let r_k and ρ_k denote the estimated range and bearing of the target t , $r_k = \max(d(s_{i,k}, \xi_{t,k}), r_0)$ with Euclidean distance $d(x, y) = \|x - y\|$, r_0 is defined as the minimal effective range threshold, and ρ_k is the angular measurement between the sensor and target. The observation covariance matrix $\mathbf{R}(s_{i,k}, \xi_{t,k})$ for target t is

$$\mathbf{R}(s_{i,k}, \xi_{t,k}) = \alpha_i \mathbf{G}(\rho_k) \begin{bmatrix} 0.1r_k & 0 \\ 0 & 0.1\pi r_k \end{bmatrix} \mathbf{G}(\rho_k)^T \quad (3)$$

where $\mathbf{G}(\rho)$ is the rotation matrix of angle ρ , and sensing quality factor α_i is a scalar in this uncertainty matrix that varies by agent. The observation model (3) for a given sensor can be extended for all the targets it observes to obtain the block diagonal \mathbf{H} and \mathbf{R} matrices for each agent. This observation setup imposes the spatially varying measurement error [7]. With observation law defined, the observation of n agents at each time step is $Z_k \in \mathcal{O}$, $Z_k = \{Z_{i,k} | i \in [n]\}$.

Action \mathcal{U} : The action of an agent i is the command of horizontal velocity, i.e. $u_i = (u_i^x, u_i^y) \in \mathbb{R}^2$ with maximum velocity constraint $|u| \leq v_{max}$. We assume the motion of agent is deterministic by

$$s_{i,k+1} = f(s_{i,k}, u_{i,k}) = \begin{bmatrix} p_{i,k}^x + u_{i,k}^x \Delta t \\ p_{i,k}^y + u_{i,k}^y \Delta t \\ \arctan(u_{i,k}^y / u_{i,k}^x) \\ u_{i,k}^x \\ u_{i,k}^y \end{bmatrix}. \quad (4)$$

The joint action domain over n agent is $\mathcal{U} = \mathcal{U}_1 \times \dots \times \mathcal{U}_n$.

State Transition \mathcal{T} : State transition law is defined as the mapping $\mathcal{T} : \mathcal{X} \times \mathcal{U} \rightarrow \mathcal{X}$. In the state $x_k \in \mathcal{X}$, $x_k = (s_k, \chi_k, \mathcal{F}_k)$, $u_k \in \mathcal{U}$, the state transition is decomposed into the following:

- the agents' deterministic transitions (4);
- OOI states $\chi_{t,k} \in \chi_k$ are stochastic, independent on sensor state s_k and control variable u_k ;
- the filter state \mathcal{F}_k is dependent on both agent state s_k , FoV of sensors $\phi_k(s_k; W)$, observation law \mathcal{O} and target state χ_k .

Cost C : The cost function is a mapping $\mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}_{\geq 0}$, with the overall optimization objective of minimizing the total expected Mean Square Error (MSE). Correspondingly, the one-step cost for the system is

$$C(x_k, u_k) = \mathbb{E}_{w_k, v_{k+1}} [\|\chi_k - \xi_k\|^2 | x_k, u_k] \quad (5)$$

B. Belief-State MDP

Because of the partial observability in the state, the unobserved variable is depicted by the probability distribution. Applying the MDP solution to POMDP requires the definition

of belief state [17]: $b_k(x) := b(x_k) \in \mathcal{B}(\mathcal{X})$, $b_k(x) = P_{x_k}(x|Z_0, Z_1, \dots, Z_k, u_0, \dots, u_k) = (b_k^s, b_k^x, b_k^\xi, b_k^P)$. All belief states of observable state components (b_k^s, b_k^ξ, b_k^P) can be represented by Dirac delta function, e.g. $b_k^s = \delta(s - s_k)$; the target belief state is represented by filter's posterior estimate $b_k^x \sim \mathcal{N}(\xi_{k|k}, \mathbf{P}_{k|k})$. Utilizing the belief state as the system state describes a fully observable system with information sufficiently characterized by the above defined distributions. Now standard MDP theory can be applied to the resulting MDP $\mathcal{M} = (\mathcal{B}, \mathcal{O}, \mathcal{U}, \tilde{\mathcal{T}}, \tilde{C})$ which includes the adaptation of the cost \tilde{C} and state transition laws $\tilde{\mathcal{T}}$.

Belief-State Transition $\tilde{\mathcal{T}}$: Belief-state transition law is the mapping $\tilde{\mathcal{T}} : \mathcal{B} \times \mathcal{U} \rightarrow \mathcal{B}$. Following (4) for each agent, the action to a belief state makes deterministic transition to agents state $s \in \mathcal{S}$; the filter state \mathcal{F} depends on sensors FoVs $\phi_k(s_k; W)$ and OOI state χ_k , which is interpreted by the state belief $b_k(x)$; OOI state is action-independent.

Belief Cost \tilde{C} : Based on state transition law $\tilde{\mathcal{T}}$, action $u \in \mathcal{U}$ makes one-step cost over the belief state b_k with the expected value of $\tilde{C}(b_k, u)$.

$$\begin{aligned} \tilde{C}(b_k, u) &= \int C(x_k, u) b_k(x) dx \\ &= \int \mathbb{E}_{w_k, v_{k+1}} [||\chi_{k+1} - \xi_{k+1}||^2 | s_k, \xi_k, u] b_k^x(x) dx \\ &= \text{tr}[\mathbf{P}_{k+1|k+1}] \end{aligned} \quad (6)$$

The third equality is derived based on the Gaussian noise assumption [15].

Consistent with previous work [9], we seek a solution from receding horizon control over fixed time horizon H in belief-state MDP defined. Denote the policy over belief state as $\pi : \mathcal{B} \rightarrow \mathcal{U}$, we define the cumulative cost over horizon H as

$$J_H^\pi(b_k) = \mathbb{E} \left[\sum_{l=k}^{k+H-1} \tilde{C}(b_l, \pi(b_l)) \middle| b_k \right] \quad (7)$$

then the optimal policy π^* follows the Bellman's principle

$$\pi^*(b_k) \in \arg \min_u (\tilde{C}(b_k, u) + \mathbb{E}[J_{H-1}^*(b_{k+1}) | b_k, u]) \quad (8)$$

and J_{H-1}^* is the minimized objective cost over subsequent horizon $H - 1$. Define the Q -value function over horizon H as $Q_H : \mathcal{B} \times \mathcal{U} \rightarrow \mathbb{R}$, which is the expected cost of executing u in belief state b_0

$$Q_H(b_k, u) = \tilde{C}(b_k, u) + \mathbb{E}[J_{H-1}^*(b_{k+1}) | b_k, u] \quad (9)$$

and optimal policy $\pi^*(b_k) \in \arg \min_{u \in \mathcal{U}} Q_H(b_k, u)$.

III. SMA-NBO

Given the problem setup, we introduce our SMA-NBO as the planning algorithm of multi-sensor target tracking. Since the non-fixed number of targets and redundant filter state dimension, methods like Q-learning that directly explores the whole state \mathcal{B} are unrealistic in computation. Our algorithm SMA-NBO seeks the proper approximation of optimization in (9) from two aspects: given current belief state b_0 , the approximation of future belief state b_k (NBO) and joint action optimization in the multi-agent team (SMA).

Algorithm 1 Nominal Belief Optimization

Require: Initial belief state $b_k = (b_k^s, b_k^x, b_k^\xi, b_k^P)$, target mean $\hat{\xi}_k = \xi_k$.

1. Generate nominal trajectory $\{\hat{b}^x\}_{k+1:k+H}$
for l in $k+1$ to $k+H$ **do**

$$\hat{\xi}_l = \mathbf{F}_{l-1} \hat{\xi}_{l-1}$$

$$\hat{b}_l^x = \delta(\chi - \hat{\xi}_l)$$
end for
2. Policy optimization

Optimize $\tilde{J}_H^\pi(b_k)$ by policy π^* .
- return** π^*

A. Nominal Belief Optimization

To obtain the optimal policy $\pi^*(b_k)$ given belief state b_k , the subsequent *expected cost-to-go* (ECTG) J_{H-1}^* is essential in (9). Denote the future belief of target trajectory over horizon H as $\{b\}_{k+1:k+H}$. However, the future target trajectory belief $\{b^x\}_{k+1:k+H}$ is stochastic and independent from action selection by Assumption 2. The approach of NBO is introduced for action optimization by approximation to future target belief $\{\hat{b}^x\}_{k+1:k+H}$.

Algorithm 1 is a description of NBO with linear dynamic target assumption (1). Step 1 generates the nominal trajectory, which is maximum a posterior (MAP) estimate of the belief-state distribution by ignoring the disturbance term w_k in the NCV transition law. The approximate cost over horizon is

$$\tilde{J}_H^\pi(b_k) = E \left[\sum_{l=k}^{k+H-1} \tilde{C}(\hat{b}_l, \pi(\hat{b}_l)) \middle| b_k, \hat{b}^x \right] \quad (10)$$

In Step 2, the action optimization is obtained from approximation function \tilde{J}_H^π . The output of NBO $\pi^*(b_k) = (u_k, \dots, u_{k+H-1})$ is the joint fleet plan for H future steps, and the agents only execute the first action u_k as their immediate control command. The approximate objective \tilde{J}_H^π is typically called truncated version since it only calculates the cost over the fixed-length horizon [14], [15].

Receding horizon control, optimizing actions over a fixed horizon of H , garners computational savings by considering only the initial portion of the infinite horizon. Additionally, the accuracy of the nominal trajectory approximation decreases with the extended horizon lengths due to compounding stochasticity resulting from the target motion model. In contrast, accounting for impacts of actions beyond horizon H in the planning objective further emphasizes the non-myopic behavior of the multi-agent system. For this reason, we implement a HECTG term $\hat{J}(\hat{b}_{k+H})$ added at the end of horizon. The HECTG not only emphasizes non-myopic planning but also avoids the detrimental computational impacts of over-extended horizons. In this way we better approximate the infinite horizon control with NBO

$$\tilde{J}_\infty^\pi(b_k) = E \left[\sum_{l=k}^{k+H-1} \tilde{C}(\hat{b}_l, \pi(\hat{b}_l)) + \hat{J}(\hat{b}_{k+H}) \middle| b_k, \hat{b}^x \right] \quad (11)$$

For multi-sensor multi-target tracking, [15] proposed MWTP as such a HECTG. We have adapted MWTP for FoV

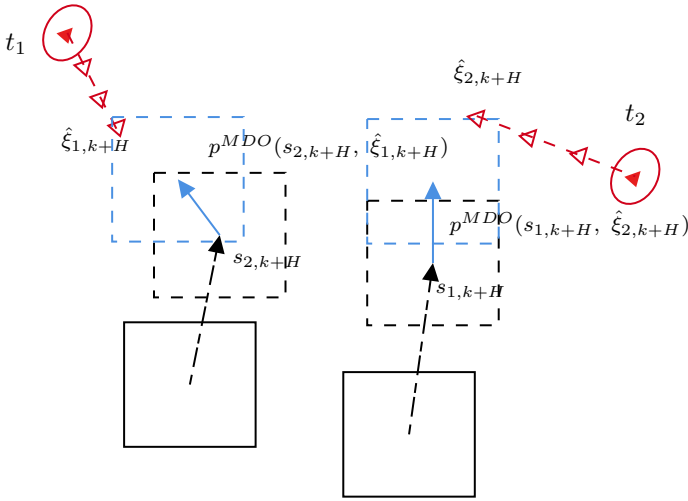


Fig. 2: An example of MWTP. Sensors' current FoVs positions are rectangles of solid line, and the targets in set T defined are red solid triangles and ellipses as means and covariances. The nominal trajectories of targets in the horizon ($H = 3$ in this case) are dashed lines with mean in hollow triangles. Rectangles of black dashed line are the final positions of sensors FoVs, and rectangles of blue dashed line are the adjusted positions of FoVs $p^{MDO}(s_{i,k+H}, \hat{\xi}_{t,k+H})$ when matching target t to sensor i .

considerations. Denote the set of targets at end of planning horizon that are outside sensors FoVs as set $T = \{\hat{\xi}_{t,k+H} : \hat{\xi}_{t,k+H} \notin \phi(\hat{b}_{k+H}^s)\}$, and sensors set at end of horizon as $S = \{\hat{b}_{k+H}^{s_i}\}$. MWTP is detailed in Algorithm 2, which takes the targets with higher uncertainty ($tr[\hat{\mathbf{P}}_{t,k+H}]$) to match the nearest sensors first. The if statement enforces a bipartite matching such that each sensor accrues a penalty based on at most one unobservable target. In Step 2 of Algorithm 2, the matching of targets to sensors will generate the adjusted position $p^{MDO}(s_{i,k+H}, \hat{\xi}_{t,k+H})$, which is the position of minimum distance to observation (MDO) for sensor i that its FoV $\phi(s_{i,k+H})$ just covers the targets nominal mean position $\hat{\xi}_{t,k+H}$ with minimal displacement. An example is shown in Fig. 2, targets not covered in sensors FoVs are in set $T = \{t_1, t_2\}$, and we have sensors' positions at end of horizon $S = \{s_{1,k+H}, s_{2,k+H}\}$. The MWTP matches sensor 1 to target 2, and sensor 2 to target 1.

B. Sequential Multi-Agent Decision Making

In the multi-agent system, the optimization problem is typically constrained by the resource limit, for example, the exponential complexity growth with the number of agents in the team. Such issue exists in NBO which defines the joint action space as \mathcal{U} and optimizes the team's policy π^* in Step 2 of Algorithm 1. An alternative but suboptimal way is to consider a sequence in optimizing single agents' actions based on the intention of other agents. We denote the *policy of intent* as $\bar{\pi}$. Define the policy of agent i as $\pi_i(b_k) = (u_{i,k}, \dots, u_{i,k+H-1})$ (i.e., the sequence of actions over the planning horizon) and the policy for a sequence of

Algorithm 2 Multiple Weighted Trace Penalty

Require: Targets outside sensors FoVs at end of horizon T , sensors set at end of horizon S .

1. Sort targets T in decreasing order of $tr[\hat{\mathbf{P}}_{t,k+H}]$
2. Initialize $\hat{J} = 0$ and all sensors $\forall i \in [n], D_i = 0$

for target t in T **do**

Get sensor $i \in \arg \min_{i \in [n]} D_i + d(s_{i,k+H}, \hat{\xi}_{t,k+H})$

if $D_i = 0$ **then**

$\hat{J} = \hat{J} + \beta d(s_{i,k+H}, \hat{\xi}_{t,k+H}) tr[\hat{\mathbf{P}}_{t,k+H}]$

end if

$D_i = D_i + d(s_{i,k+H}, \hat{\xi}_{t,k+H})$

$s_{i,k+H} = p^{MDO}(s_{i,k+H}, \hat{\xi}_{t,k+H})$

end for

return \hat{J}

Algorithm 3 SMA-NBO

Require: Initial belief state $b_k = (b_k^s, b_k^x, b_k^\xi, b_k^p)$, target mean $\hat{\xi}_k = \xi_k$, previous decision π^p

1. Generate nominal trajectory $\{\hat{b}^x\}_{k+1:k+H}$
2. Generate policy of intent $\forall i \in [n], \bar{\pi}_i = (\pi_i^p, \bar{u}_i)$
3. Sequential multi-agent decision making

for agent i from 1 to n **do**

$\hat{\pi}_i^* \in \arg \min_{\pi_i} \hat{J}_H^{(\hat{\pi}_{1:i-1}^*, \pi_i, \bar{\pi}_{i+1:n})}(b_k)$

end for

return $\hat{\pi}^*$

agents from i to $i+j$ as $\pi_{i:i+j} = (\pi_i, \pi_{i+1}, \dots, \pi_{i+j})$. Then at each decision epoch, the action optimization starts with first agent by

$$\hat{\pi}_1^* \in \arg \min_{\pi_1} J_H^{(\pi_1, \bar{\pi}_{2:n})}(b_k) \quad (12)$$

This decision is passed to the next agent in the optimization sequence. The general optimization of agent i is

$$\hat{\pi}_i^* \in \arg \min_{\pi_i} J_H^{(\hat{\pi}_{1:i-1}^*, \pi_i, \bar{\pi}_{i+1:n})}(b_k) \quad (13)$$

However, the policy of intent $\bar{\pi}$ is a key factor in optimization. After agent j executes action $u_{j,k}$ at time k , there is a remaining action sequence $\pi_j^p = (u_{j,k+1}, \dots, u_{j,k+H})$. This remainder provides insight to the j th agent's *future intentions* and can be used at time $k+1$ to inform the control decisions of agents $1:i$ in (13) s.t. $i < j$ by simply extending π_j^p with an additional action \bar{u}_j generated by a heuristic base single-agent base policy. This yields an approximate H -step policy $\bar{\pi}_j = (\pi_j^p, \bar{u}_j)$ containing agent j 's future intent.

Fig. 1c shows the proposed SMA in block diagram and Algorithm 3 details the SMA-NBO. The reduction of computation is the main advantage of SMA-NBO. For the general searching method, the time complexity of the optimization in (13) is the planning domain of agent i over horizon H , denoted as $O(|\mathcal{U}_i^H|)$, and SMA-NBO has complexity increase linear to agent number, $O(n|\mathcal{U}_i^H|)$; while for joint optimization of agents the complexity grows exponentially, $O(|\mathcal{U}_i^H|^n)$. Also, such sequential decision-making makes distributed computations possible rather than a single agent making and commanding fleet-wide decisions.

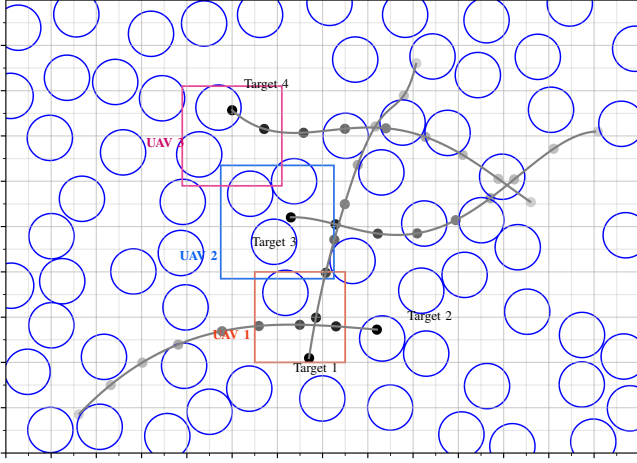


Fig. 3: Multi-sensor multi-target tracking environment, which contains 3 UAVs with their FoVs in rectangles, 4 targets and trajectory crossovers and divergence.

IV. SIMULATION RESULTS

This section presents the simulation setup, performance results and conclusions drawn from the SMA-NBO experiments. Fig. 3 shows the general simulation environment of a 3-heterogeneous UAV tracking team. These UAVs are detecting people as targets in a 150×100 m rectangle AOI with sensing frequency $f_1 = 5$ Hz, and making decisions at frequency of 1 Hz. The FoVs are squares with edge lengths of $\{20, 25, 22\}$ m, and sensing quality factors of agents α_i are $\{0.1, 0.15, 0.12\}$. The trajectories of OOI are designed with step lengths obeying the Levy walk in the speed interval of $[1.0, 3.0]$ m/s, describing human gaits. Such curves in Fig. 3 contain the general tracking challenges such as crossover, non-linearity and divergent trajectories, which requires sufficient coordination to maintain good tracking performance. Each agent has the speed limit of $v_{max} = 5$ m/s. The Optimal Subpattern Assignment (OSPA) [18] metric is applied as the evaluation of multi-target tracking performance. For two finite sets $X = \{x_1, \dots, x_a\}$ and $Y = \{y_1, \dots, y_b\}$ without loss of generality, $b \geq a$, let Π_k be the set of permutation on $\{1, 2, \dots, k\}$. The OSPA metrics for these two sets is

$$\bar{d}_p^{(c)}(X, Y) = \left[\frac{1}{b} \left(\min_{\pi \in \Pi_b} \sum_{i=1}^a d^{(c)}(x_i, y_{\pi(i)}) + c^p(b-a) \right) \right]^{\frac{1}{p}} \quad (14)$$

with $d^{(c)}(x, y) := \min(c, d(x, y))$ for metrics d . In target tracking evaluation, we can regard tracking set Θ_k and OOI true state χ_k as the sets X, Y . In (14), the first part addresses the tracking error with metrics d , which is Euclidean distance in our problem; second part is the penalty of cardinality, either missing or false alarm in target tracking. The parameters $c = 50$ m, $p = 2$ are selected in OSPA, with c the cut-off value of target tracking error, and $p = 2$ returns L-2 norm.

A. A Random Occlusion Forest Environment

One interest of SMA-NBO algorithm is to achieve non-myopic performance indicated by (7), which provides the

ability to overcome or compensate for occluded areas during tracking. Also, the behavior of exchanging duty of tracking occluded targets between agents demonstrates fleet coordination. However, there are few studies in target tracking of the performance evaluation with random occlusion area. Inspired by studies about collision avoidance in a random obstacle environment [19], we create a random forest of occlusion objects shown in Fig. 3. Imagine trees of fixed radius generate shadow areas in Fig. 1a, the naive but direct factors impacting the tracking performance are tree density λ and shadow radius R . We choose three densities, and the expected numbers of trees in this 150×100 m AOI are $\lambda = \{15, 45, 75\}$, which follow Poisson distribution in the random forest [19]; two sizes of tree are selected $R = \{1, 5\}$ m. For each combination of (λ, R) , we generate 50 maps that randomizes tree positions without overlapping. These maps are utilized for statistical performance evaluation of algorithms in the following report.

B. SMA-NBO Tracking Performance

First, the tracking performance of the proposed SMA-NBO is explored in all density and radius random forests; horizon length H is varied since it plays an important role in featuring the coordinated and non-myopic behavior of the agents, with the trade-off in computation. Longer horizons H avoid greedy behavior in general, specifically when targets are in occlusion areas, and enables agents to realize the targets will exit the occlusion in the future. Fig. 4 reports the tracking performance in the format of accumulated frequency (empirical cumulative distribution function, ECDF) plots over OSPA values, resulting from the simulations over 50 random forests. It should be noted that same batch of random maps are used in testing different algorithms in same parameter specifications (λ, R) . Examination of Fig. 4 shows the benefits of longer planning horizons in higher density forests with larger occlusions. SMA-NBO achieves 95%+ OSPA values lower than 1 m in small occlusion size, $R = 1$, with horizon $H > 1$; in $R = 1$, SMA-NBO with $H = 5$ reaches 90%, 70%, and 60% OSPA values lower than 1 m in all densities, showing promising tracking performance of SMA-NBO.

Larger horizons in SMA-NBO display benefits that are visible from the frequency plot. In cases of $R = 1$, i.e., small occlusions, the performance of $H = 3$ and $H = 5$ is almost identical; while in the case of $R = 5$, $\lambda = 75$, longer horizons

λ	H	SMA-NBO	MWTP	MCR	Dec-POMDP
15	1	616	643	1517	-
	3	1401	1457	3309	2192
	5	2094	2033	5160	2641
45	1	600	628	1639	-
	3	1386	1393	3673	-
	5	2028	2023	5883	-
75	1	574	604	1662	-
	3	1307	1319	3822	1789
	5	1914	1924	6291	2551

TABLE I: Average runtime per simulation trial of different algorithms: truncated SMA-NBO (SMA-NBO), SMA-NBO with MWTP (MWTP), Monte Carlo Multi-agent Rollout (MCR) and Dec-POMDP, unit in second.

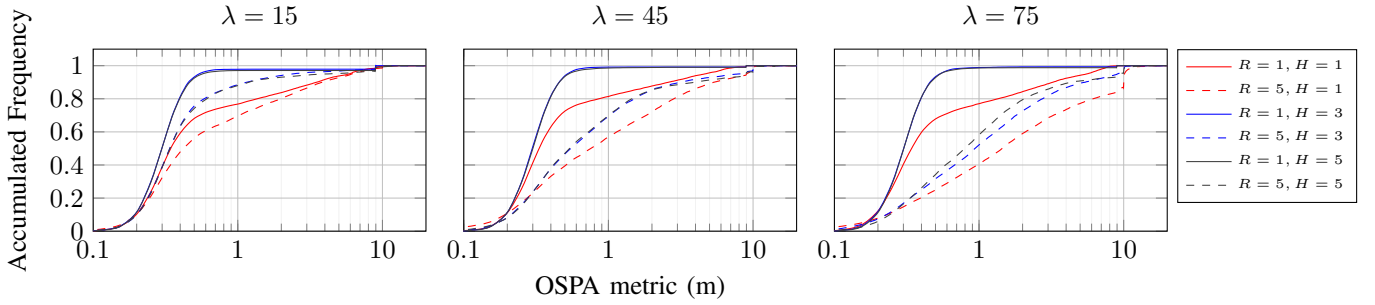


Fig. 4: Tracking performance of truncated SMA-NBO in 50 random maps of different density.

highlight performance improvement when compensating dense occlusion coverage in the forest (39%). Although density has some impact on tracking performance, the advantages of horizon length are tightly linked with occlusion radii.

The runtime statistics are recorded in Table. I. For SMA-NBO, the general rule is that simulation time relative to planning horizon length due to the extension of action space via time dimension. For an intuitive illustration of coordination in SMA-NBO, in the **video attachment** the behaviors different horizons are compared in the map of $R = 5, \lambda = 45$. From 17sec-25sec with $H = 5$, the red UAV chases the bottom-left target which is in diverging from others, 25sec-35sec the blue and pink UAVs deal with 3 targets with hand-offs of the center targets. Such behavior is lacking in $H = 1, 3$ because of the limit in the horizon; longer horizons implicitly inspire cooperation.

C. HECTG SMA-NBO

As described in (11), the horizon can be further approximated with a HECTG function, and MWTP is one candidate detailed in Algorithm 2. We test this SMA-NBO + MWTP in random forests of $R = 5$ with results shown in Fig. 5. From the simulation result of Fig. 5, the major performance improvement by MWTP can be seen in $H = 1$, which helps generating non-myopic behavior with greedy planning cases; in low density forests ($\lambda = 15, 45$) MWTP $H = 1$ even achieves close performance to longer horizon $H = 3, 5$. On the other hand, no significant computation load is introduced from the MWTP term by comparing the runtime of SMA-NBO and MWTP in Table. I. Thus, significant improvement in performance is seen to acknowledge MWTP as a HECTG without detrimental computational effects.

D. Different Approximation Approaches

The Monte Carlo method is one of the most popular methods in handling unobservable variables, and it is considered as another approximation approach to future target trajectories [10], [9]. Typically it requires a set of samples drawn from the belief distribution b^x , then plans as an optimal control problem accordingly based on the approximate expected objective function with samples. The number of trajectory samples should be sufficient to obtain a good value estimate. However, the NBO approach only uses the nominal trajectory as the MAP estimation, which brings an advantage in computation

cost. To study the difference between these two methods, we compared the SMA-NBO with Monte Carlo multi-agent Rollout (MCR) method [9] in random forests of $R = 5$. The latter runs with 50 Monte Carlo trajectory samples in simulation. From the performance perspective in Fig. 5, MCR provides better tracking than truncated SMA-NBO in low horizon ($H = 1$), but SMA-NBO with MWTP achieves same level of performance as MCR; when $H > 1$, the advantage of MCR is no longer obvious. However, the computation in MCR is much more intense reflected by Table. I, which requires over 2.2 times runtime than SMA-NBO.

E. Comparison of Decision-Making Architecture

Dec-POMDP is a decentralized approach with no decision communication in the fleet. In Fig. 5, we selectively pick the scenarios of horizon length $H = 3, 5$ and the extreme densities $\lambda = 15, 75$ for the comparison of Dec-POMDP (brown lines) and sequential algorithms. It is worth mentioning that NBO is also applied to solve Dec-POMDP, similar to [6]. In all four scenarios, there is no significant difference in OSPA distribution between Dec-POMDP and SMA-NBO series algorithms. Even though each agent in Dec-POMDP generates the fleet-wise optimal planning, the sequential method of SMA-NBO can also achieve similar tracking performance. On the other hand, from Table I, the runtime difference between SMA-NBO and Dec-POMDP shows the computational advantage of sequential decision making and leveraging the policy of intent. Given the time complexity's exponential increase by agent number n in Dec-POMDP, the runtime difference will be even larger when more UAVs are in the system.

V. CONCLUSION

This paper introduces the SMA-NBO algorithm as the information-driven planning algorithm of multiple mobile sensors in the task of target tracking. Nominal belief-state optimization is adapted for sequential multi-agent decision making, admitting a distributed system architecture. Specifically, SMA-NBO recycles optimized single-agent policies from the prior decision epoch, constructing a *policy of intent* to inform future agent action selections. Additionally, this sequential methodology retains tracking performance and reduces the computational load when compared to contemporary distributed methods, e.g., Dec-POMDP and MCR.

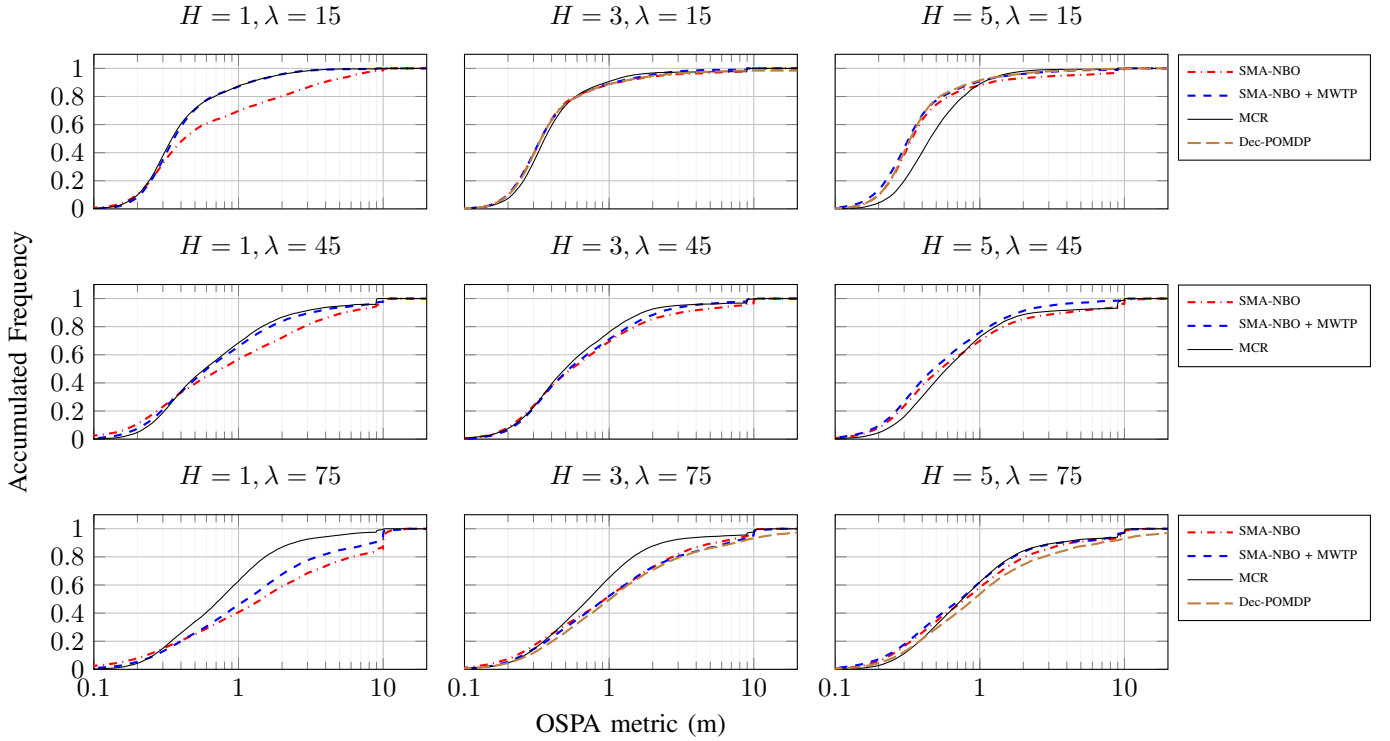


Fig. 5: Tracking performance comparison of truncated SMA-NBO, SMA-NBO+MWTP and Monte Carlo Multi-agent Rollout (MCR) and Dec-POMDP in random maps of size $R = 5$.

We evaluate SMA-NBO, Dec-POMDP and MCR performance in a random occlusion forest parameterized by occlusion size R and density λ . This environment exemplifies the impacts of look-ahead horizon lengths, showing strong correlation to the occlusion size R . Realizing the benefit of non-myopic decisions, we augment SMA-NBO with a HECTG, namely an adapted MWTP [15]. The statistical result shows SMA-NBO is capable of multi-sensor coordination to track targets in significantly occluded environments.

Based on our investigation of SMA-NBO, the horizon selection of SMA-NBO based on the size and density of the occlusions leads to practical implementation. Also, theoretic performance boundary of the sequential information-driven planning is worth studying.

REFERENCES

- [1] S. Ferrari and T. A. Wettergren, *Information-Driven Planning and Control*. MIT Press, 2021.
- [2] L. Paull, S. Saeedi, M. Seto, and H. Li, "Sensor-driven online coverage planning for autonomous underwater vehicles," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 6, pp. 1827–1838, 2012.
- [3] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Decentralized active information acquisition: Theory and application to multi-robot slam," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 4775–4782.
- [4] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Trans. Robot. Autom.*, vol. 20, no. 2, pp. 243–255, 2004.
- [5] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, no. 2, 2008.
- [6] S. Ragi and E. K. Chong, "Uav path planning in a dynamic environment via partially observable markov decision process," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 4, pp. 2397–2412, 2013.
- [7] L. W. Krakow, C. M. Eaton, and E. K. Chong, "Simultaneous non-myopic optimization of uav guidance and camera gimbal control for target tracking," in *Proc. IEEE Conf. Control Techn. Appl.*, 2018, pp. 349–354.
- [8] R. K. Ramachandran, N. Fronda, and G. Sukhatme, "Resilience in multi-robot multi-target tracking with unknown number of targets through reconfiguration," *IEEE Control Netw. Syst.*, 2021.
- [9] T. Li, L. W. Krakow, and S. Gopalswamy, "Optimizing consensus-based multi-target tracking with multiagent rollout control policies," in *Proc. IEEE Conf. Control Techn. Appl.*, 2021, pp. 131–137.
- [10] S. Ragi and H. D. Mittelmann, "Random-sampling monte-carlo tree search methods for cost approximation in long-horizon optimal control," *IEEE Control Syst. Lett.*, vol. 5, no. 5, pp. 1759–1764, 2020.
- [11] M. Corah and N. Michael, "Distributed matroid-constrained submodular maximization for multi-robot exploration: Theory and practice," *Auton. Robots.*, vol. 43, no. 2, pp. 485–501, 2019.
- [12] T. Friedrich, A. Göbel, F. Neumann, F. Quinzan, and R. Rothenberger, "Greedy maximization of functions with bounded curvature under partition matroid constraints," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, 2019, pp. 2272–2279.
- [13] S. Ragi and E. K. Chong, "Decentralized guidance control of uavs with explicit optimization of communication," *J. Intell. Robot. Syst.*, vol. 73, no. 1, pp. 811–822, 2014.
- [14] D. Bertsekas, "Multiagent reinforcement learning: Rollout and policy iteration," *IEEE/CAA J. Autom. Sin.*, vol. 8, no. 2, pp. 249–272, 2021.
- [15] S. A. Miller, Z. A. Harris, and E. K. Chong, "Coordinated guidance of autonomous uavs via nominal belief-state optimization," in *Proc. Amer. Control Conf.*, 2009, pp. 2811–2818.
- [16] G. Battistelli and L. Chisci, "Kullback–leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.
- [17] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. intell.*, vol. 101, no. 1–2, pp. 99–134, 1998.
- [18] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, 2008.
- [19] S. Karaman and E. Frazzoli, "High-speed flight in an ergodic forest," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2012, pp. 2899–2906.