Autonomous Agents and Multi-Agent Systems (2019) 33:132–158
https://doi.org/10.1007/s10458-019-09401-5

Systemic design of distributed multi-UAV cooperative decision-making for multi-target tracking

Yunyun Zhao[1] · Xiangke Wang[1] · Chang Wang[1] · Yirui Cong[1] · Lincheng Shen[1]

Published online: 14 January 2019

Abstract
In this paper, we consider the cooperative decision-making problem for multi-target tracking in multi-unmanned aerial vehicle (UAV) systems. The multi-UAV decision-making problem is modeled in the framework of distributed multi-agent partially observable Markov decision processes (MPOMDPs). Specifically, the state of the targets is represented by the joint multi-target probability distribution (JMTPD), which is estimated by a distributed information fusion strategy. In the information fusion process, the most accurate estimation is selected to propagate through the whole network in finite time. We propose a max-consensus protocol to guarantee the consistency of the JMTPD. It is proven that the max-consensus can be achieved in the connected communication graph after a limited number of iterations. Based on the consistent JMTPD, the distributed partially observable Markov decision algorithm is used to make tracking decisions. The proposed method uses the Fisher information to bid for targets in a distributed auction. The bid is based upon the reward value of the individual UAV's POMDPs, thereby removing the need to optimize the global reward in the MPOMDPs. Finally, the cooperative decision-making approach is deployed in a simulation of a multi-target tracking problem. We compare our proposed algorithm with the centralized method

This research is sponsored by the National Key Laboratory of Science and Technology on UAV, Northwestern Polytechnical University, under the Grant Number 614230110080817.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10458-019-09401-5) contains supplementary material, which is available to authorized users.

✉ Xiangke Wang
xkwang@nudt.edu.cn

Yunyun Zhao
zhaoyycloudy@gmail.com

Chang Wang
wangchang07@nudt.edu.cn

Yirui Cong
congyirui11@nudt.edu.cn

Lincheng Shen
lcshen@nudt.edu.cn

[1] College of Intelligence Science and Technology, National University of Defense Technology, Changsha, China

and the greedy approach. The simulation results show that the proposed distributed method has a similar performance to the centralized method, and outperforms the greedy approach.

Keywords Multi-UAV · Decision-making · Multi-target tracking · Distributed information fusion · Max-consensus

1 Introduction

Multiple Unmanned Aerial Vehicles (UAVs) have been widely employed in active information gathering missions, such as target tracking [1], exploration [2], surveillance [3,4], monitoring [5,6] and search-and-rescue [7]. There are many challenges associated with these tasks, due to limited communication range, uncertain and dynamic environment, and dynamic communication network [8].

Multiple targets tracking by multiple UAVs has attracted attention in the multi-agent community. In this paper, we are particularly interested in the cooperative decision-making problem for this task. The work is motivated by the application of UAVs in both military and civilian surveillance systems. For instance, the multi-target tracking can be employed in a traffic mapping application [8], where the locations of ground vehicles are estimated with limited communication ranges and bandwidths. In a surveillance application [4], a team of UAVs keeps tracking several targets such as cars and people with uncertainties caused by occlusion or motion blur.

Active target tracking [9] typically includes two subtasks: estimating the motion of the target from the sensory data, and actively controlling the agent to gather more data. For multi-target tracking, it is more challenging that the joint multi-target probability distribution (JMTPD) estimation and the cooperative tracking decision-making [9] have to be done under the conditions of limited communication ranges and noisy sensor measurements. In this paper, we propose a novel distributed multi-target tracking system based on distributed information fusion and decision-making. The tracking system is based on a distributed multi-UAV multi-target tracking framework, which integrates the JMTPD estimation, multi-target assignment and multi-UAV action selection. The key contributions are as follows:

1. A distributed information fusion strategy with an individual Kalman filter as well as the max-consensus protocol is provided to estimate the JMTPD of the UAVs. With a modified distributed data propagation mechanism, the most accurate estimation of the JMTPD can be obtained in finite time through the whole network. Also, a terminal condition with respect to the local structure of the network is explored to reduce the number of iterations.
2. A time-efficient auction algorithm is presented to solve the distributed multi-UAV POMDPs based on the theory of information geometry, which takes full advantage of the Fisher information matrix (FIM) with less computation.

The rest of the paper is organized as follows. Section 2 presents related work. Section 3 describes the task scenario and formulates the problem. Section 4 presents the distributed estimation algorithm based on max-consensus and constrained communication. Section 5 introduces information geometry and the distributed auction-based decision-making method for target assignment as well as action selection. The comprehensive algorithm is designed in Sect. 6. Section 7 discusses the results of our experiments, and Sect. 8 concludes the paper and outlines plans for future work.

2 Related work

The multi-target tracking problem has been extensively studied in the literature [9–12]. The simplest form of multi-target tracking is to track a known number of static targets [11]. Tracking dynamic targets has been of greater interest because it is more realistic for real world problems, but it makes the problem more challenging due to the uncertainty of the target states. Using multiple UAVs for the tracking task is even more difficult because real-time path planning of the multiple UAVs will become time-consuming when the numbers of UAVs or targets increase dramatically. This problem can be mitigated by using approximation algorithms [10] or an anytime planning algorithm [13]. However, the above-mentioned literature focuses on the fast solution of a planning algorithm and fails to fundamentally solve this real-time problem, such as problem decomposition or modeling. In this paper, we divide the multi-UAV tracking multi-target problem into several sub-problems and deal with them separately.

In multi-UAV multi-target tracking, each UAV runs a local filter to estimate the state of each target, and the UAV must be able to share information across the team. In the case where UAVs have limited communication range and limited field of view, it can be difficult to maintain information consistency across the team [14], which means the estimations of the JMTPD among the UAVs are similar. The methods of estimating the JMTPD can be centralized or distributed, depending on the communication and computation capabilities of the UAVs. The centralized algorithms use a fusion centre to receive data from each UAV and compute a joint estimation of the target states, which depends on high frequency data transmission. In addition, as the number of the UAVs increases, the scalability of the centralized algorithms is not good. In contrast, the distributed algorithms scale better [15]. These methods use only local information which is exchanged among the neighbors. For example, Capitan et al. [16] used a decentralized data fusion (DDF) approach [17] to efficiently maintain a joint belief state in a multi-robot system. Adamey and Ozguner [18] proposed a decentralized particle filtering approach based on Gaussian Mixture Models and covariance intersection to estimate target locations. Chagas and Waldmann [19] derived a novel methodology from the linear minimum mean square error estimator to fuse the delayed measurements in a distributed sensor network. Nevertheless, these methods usually require complex calculations to maintain the information consistency among the agents.

Thus, it is necessary to reduce the computational complexity in keeping information consistency/consensus. Consensus algorithms, especially the max-consensus algorithm, have been widely used for task assignment [20,21], formation control [22,23] and distributed estimation [24,25]. Palacios-Gasós et al. [26] combined distributed max-consensus with additive inputs to estimate the coverage of the environment using only local information. Di Paola et al. [27] proposed a distributed Kalman filtering strategy based on the max-consensus agreement protocol, which provided the opportunity for estimating the JMTPD. However, the termination conditions of the max-consensus algorithm in the process of distributed information fusion have not been discussed in the literature.

Based on the estimated JMTPD, the multi-UAV system has to develop an optimal tracking strategy. Partially observable Markov decision processes (POMDPs) provide a mathematical framework for decision-making in realistic, uncertain and partially observable environments [28,29]. A variety of extended POMDPs have been proposed for multi-agent systems, including multi-agent POMDPs (MPOMDPs) [30] and decentralized POMDPs (Dec-POMDPs) [31]. MPOMDPs are typically centralized, while Dec-POMDPs are distributed [16]. There are also alternative decision-making frameworks where limited communication is assumed

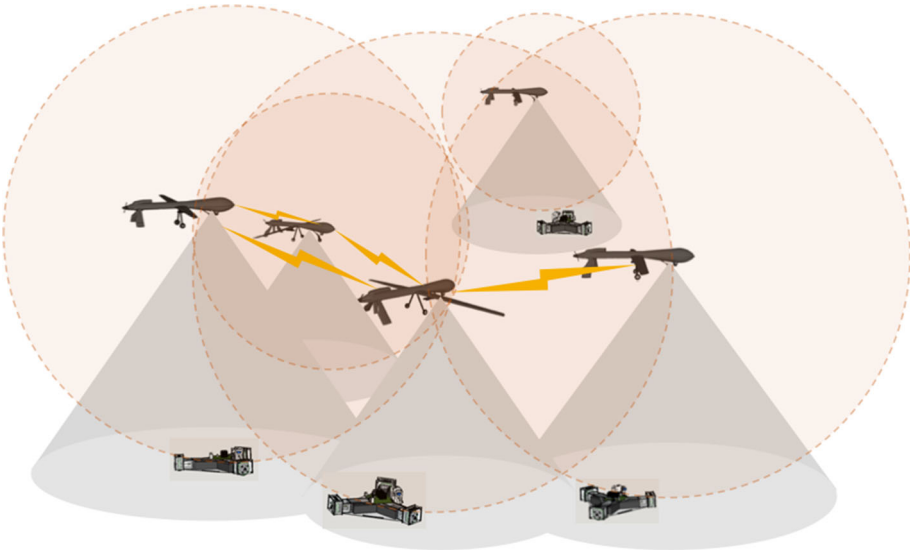

Fig. 1 Illustration of the multi-UAV system tracking multiple targets under constraints of observation and communication. The gray areas illustrate the observation range of the UAV and the orange areas are the communication range

[32–34] to balance the communication cost and the cooperation performance. For this purpose, distributed suboptimal approaches have been proposed. For instance, Dames et al. [9] presented a greedy algorithm to assign trajectories for active target tracking. Decentralized planning methods have also been proposed by combining POMDPs with policy auctions [35,36] to optimize the performance of the overall system. For example, Capitan et al [4] generalized a centralized POMDP auction to a distributed version that assigns policies to the individuals in swarms. As auction algorithms have been widely used in optimal multi-robot task allocation [37,38], distributed POMDPs based on the auction mechanism are an interesting research direction in distributed decision-making [16]. In this paper, we have chosen the auction mechanism for multi-target distributed tracking decision-making.

For decision-making problems under uncertain conditions, the decision theory based on POMDP provides a principled approach to optimize task performance directly [39]. The reward function can depend on different sets of state variables, which allows us to model different goals. In our previous work [40], we have used information gain from information geometry as the reward criterion of the POMDP. The simulation results have shown the advantage in terms of computation time. As an extension of our previous work, we will use Fisher information to quantify the bid at auction.

3 Problem formulation

In this paper, a multi-UAV system performs the task of tracking multiple ground vehicles as shown in Fig. 1. Specifically, N^u cooperative UAVs track N^t moving targets simultaneously. These UAVs are assumed to fly at different altitudes to avoid collisions. We use $\mathcal{U}$ and $\mathcal{T}$ to represent the set of UAVs and targets, respectively. The variables related to the UAVs and the

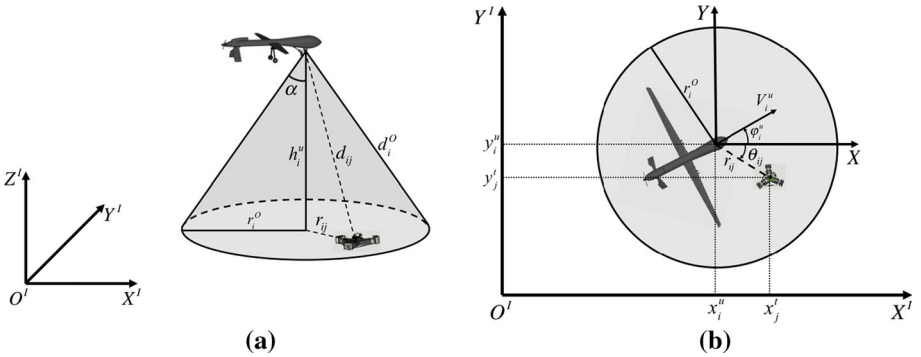

Fig. 2 **a** Three-dimensional side-view of a single UAV observing a single target. **b** Top view of a single UAV observing a single target. The observation range r_i^O is the projection of d_i^O on the ground

targets are distinguished by superscripts u and t. We also assume that each UAV can track only one target in each decision cycle.

The UAVs are deployed in a three-dimensional space. The state of the ith UAV is denoted by $\mathbf{s}_i^u = (x_i^u, y_i^u, h_i^u, V_i^u, \varphi_i^u)^{\mathrm{T}}$. Assume that the ith UAV flies at a constant altitude h_i^u and a constant speed V_i^u. We also assume that the ith UAV knows its horizontal position $\mathbf{p}_i^u = (x_i^u, y_i^u)^{\mathrm{T}}$ and the heading angle φ_i^u in the inertial coordinate system accurately. The kinematics of fixed-wing UAVs are high-dimensional, nonlinear, and dynamic. In this paper, we choose the unicycle model at a constant speed to model the kinematics [41]. The model is fully capable of capturing the kinematic characteristics of the closed-loop behaviors of fixed-winged UAVs. The UAV autopilot controls the bank angle ϕ_i^u directly, which is selected from a discrete finite action set $\mathcal{A}$. The discrete kinematic model from the time step k to $k+1$ is as follows:

$$\begin{cases} x_i^u(k+1) = x_i^u(k) + V_i^u \Delta T \cos\left(\varphi_i^u(k)\right), \\ y_i^u(k+1) = y_i^u(k) + V_i^u \Delta T \sin\left(\varphi_i^u(k)\right), \\ h_i^u = const, \\ V_i^u = const, \\ \varphi_i^u(k+1) = \varphi_i^u(k) + g\Delta T \tan\left(\phi_i^u(k)\right)/V_i^u, \end{cases} \tag{1}$$

where g is the gravitational acceleration, and ΔT is the length of the time step.

The undirected communication graph among the UAVs is represented by $\mathcal{G} = \{\mathcal{I}, \mathcal{E}\}$, where $\mathcal{I} = \{1, \ldots, N^u\}$ is the set of nodes and $\mathcal{E}$ is the set of edges. $\mathcal{E}$ depends on the relative location between UAVs and the communication range r_i^C of each UAV, which is simplified as r^C. We note that the whole graph $\mathcal{G}$ is dynamic and can be non-connected. We use $\mathcal{G}_i^{un} = \{\mathcal{I}_i^{un}, \mathcal{E}_i^{un}\}$ to represent the communication subgraph containing the ith UAV and its neighbors, where $\mathcal{I}_i^{un} = \{q \in \mathcal{U} | \sqrt{\left(x_i^u - x_q^u\right)^2 + \left(y_i^u - y_q^u\right)^2 + \left(h_i^u - h_q^u\right)^2} \leqslant r^C\}$ and $\mathcal{E}_i^{un} = \{(i,q) | q \in \mathcal{I}_i^{un}, q \neq i\}$. $\mathcal{G}_i^{un}$ can change with time, because the distances between UAVs are changing. We note that this is the the precondition of information fusion and cooperative decision-making.

In this paper, we assume that the airborne gimbal measurement units are homogeneous range-bearing sensors with limited sensing capability and fixed field of view (FOV). The geometric relation of the measurement model is shown in the perspective view and top view in Fig. 2a, b, respectively. The ith UAV can observe all the targets within its measurement

range r_i^O, which is determined by its altitude h_i^u and FOV 2α. At the same time, the ith UAV can communicate with all the neighbors within its communication range r^C.

The state of the target j is defined by $\mathbf{s}_j^t = (x_j^t, y_j^t, \dot{x}_j^t, \dot{y}_j^t)^{\mathrm{T}}$, where $\mathbf{p}_j^t = (x_j^t, y_j^t)^{\mathrm{T}}$ is the position of the target j and $(\dot{x}_j^t, \dot{y}_j^t)^{\mathrm{T}}$ is the velocity vector. At time step k, the sensor of the ith UAV can measure the relative distance $d_{ij}(k)$ and the relative bearing $\theta_{ij}(k)$ of the target in the UAV coordinate system. Assume that the measurement noise is white Gaussian with zero mean and covariance matrix $\mathbf{C}_{ij}(k)$. The related covariance matrix is given by:

$$\mathbf{C}_{ij}(k) = \begin{bmatrix} \sigma_{d_{ij}}^2(k) & 0 \\ 0 & \sigma_{\theta_{ij}}^2(k) \end{bmatrix}, \tag{2}$$

where $\sigma_{d_{ij}}(k)$ and $\sigma_{\theta_{ij}}(k)$ are the standard deviations of the distance and the bearing measurements at the time step k. The noise of the range-bearing measurement grows with the distance between the target and the UAV, and the standard deviation increases exponentially when approaching the maximum measurement value [42]. Accordingly, the interpolation functions are used in [27] to approximate the experimental data reported in [42]:

$$\begin{cases} \sigma_{d_{ij}}(k) = k_d \left(1 + e^{k_r \left(\frac{d_{ij}(k)}{d_i^O} - 1 \right)} \right), \\ \sigma_{\theta_{ij}}(k) = k_\theta \frac{d_{ij}(k)}{d_i^O}, \end{cases} \tag{3}$$

where k_d, k_r and k_θ are the fitting coefficients associated with the sensor. The measurement value of the relative position of the target j relative to the ith UAV is:

$$\mathbf{z}_{ij}^P(k) = \begin{bmatrix} d_{ij}(k) \\ \theta_{ij}(k) \end{bmatrix} + \boldsymbol{\omega}_{ij}(k), \tag{4}$$

where the measurement noise is $\boldsymbol{\omega}_{ij}(k) \sim \mathcal{N}\big(0, \mathbf{C}_{ij}(k)\big)$.

We transform the measurement of the target position from the polar coordinate system to the inertial Cartesian coordinate system for the calculation of the Kalman filter. The transformation function is as follows:

$$\begin{cases} x_j^t(k) = x_i^u(k) + \sqrt{d_{ij}^2(k) - (h_i^u)^2} \cos\big(\theta_{ij}(k)\big), \\ y_j^t(k) = y_i^u(k) + \sqrt{d_{ij}^2(k) - (h_i^u)^2} \sin\big(\theta_{ij}(k)\big), \end{cases} \tag{5}$$

Then, we formalize the measurement model of the UAV i to the target j in the Cartesian coordinate system as:

$$\mathbf{z}_{ij}^C(k) = \mathbf{H}\mathbf{s}_j^t(k) + \boldsymbol{\upsilon}_{ij}(k), \tag{6}$$

where $\mathbf{z}_{ij}^C(k)$ is the measurement value and the measurement matrix $\mathbf{H}$ is defined as:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

Define $\boldsymbol{\varphi}(\cdot)$ as the transition function from the polar coordinate system to the 2-D Cartesian coordinate system, i.e.,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \boldsymbol{\varphi}\left(\begin{bmatrix} d \\ \theta \end{bmatrix} \right) = \begin{bmatrix} \sqrt{d^2 - h^2} \cos\theta \\ \sqrt{d^2 - h^2} \sin\theta \end{bmatrix}, \tag{7}$$

where (x, y) and (d, θ) are the coordinates in the 2-D Cartesian coordinate system and the polar coordinate system, respectively.

Then, the relative position of the target to the UAV in the Cartesian coordinate system, defined as $\mathbf{Hs}_j^t(k)$, has the following form

$$\mathbf{Hs}_j^t(k) = \boldsymbol{\varphi}\left(\mathbf{z}_{ij}^P(k) - \boldsymbol{\omega}_{ij}(k)\right). \tag{8}$$

By using the Taylor series expansion of $\boldsymbol{\varphi}\left(\mathbf{z}_{ij}^P(k) - \boldsymbol{\omega}_{ij}(k)\right)$ around the noisy measurement $\mathbf{z}_{ij}^P(k)$, we obtain:

$$\mathbf{Hs}_j^t(k) = \boldsymbol{\varphi}\left(\mathbf{z}_{ij}^P(k)\right) - \mathbf{J}_{ij}(k)\boldsymbol{\omega}_{ij}(k) + \mathcal{O}\left(\boldsymbol{\omega}_{ij}(k)\right), \tag{9}$$

where $\mathcal{O}\left(\boldsymbol{\omega}_{ij}(k)\right)$ stands for the higher order (≥ 2) terms, and the Jacobian $\mathbf{J}_{ij}(k)$ is evaluated at the noisy measurement $\mathbf{z}_{ij}^P(k)$:

$$\mathbf{J}_{ij}(k) = \begin{bmatrix} \frac{d_{ij}(k)}{r_{ij}(k)}\cos\left(\theta_{ij}(k)\right) & -r_{ij}\sin\left(\theta_{ij}(k)\right) \\ \frac{d_{ij}(k)}{r_{ij}(k)}\sin\left(\theta_{ij}(k)\right) & r_{ij}\cos\left(\theta_{ij}(k)\right) \end{bmatrix}, \tag{10}$$

where r_{ij} is the projection of the relative distance between the target j and the UAV i in the ground coordinate system.

Considering $\boldsymbol{\varphi}\left(\mathbf{z}_{ij}^P(k)\right) = \mathbf{z}_{ij}^C(k)$ and the Eq. (9), the exactly converted measurement model (6) can be written as:

$$\mathbf{z}_{ij}^C(k) = \mathbf{Hs}_j^t(k) + \underbrace{\mathbf{J}_{ij}(k)\boldsymbol{\omega}_{ij}(k) - \mathcal{O}\left(\boldsymbol{\omega}_{ij}(k)\right)}_{\boldsymbol{\upsilon}_{ij}(k)}. \tag{11}$$

The standard approach usually ignores $\mathcal{O}\left(\boldsymbol{\omega}_{ij}(k)\right)$ in the exact model (11), and treats $\boldsymbol{\upsilon}_{ij}(k)$ approximately as zero-mean with covariance:

$$\mathbf{R}_{ij}(k) = \mathbf{J}_{ij}(k)\mathbf{C}_{ij}(k)\mathbf{J}_{ij}^{\mathrm{T}}(k). \tag{12}$$

The kinematic model of the target motion embedded in the local Kalman filter of each UAV [27] is given as:

$$\mathbf{s}_j^t(k+1) = \mathbf{A}\mathbf{s}_j^t(k) + \boldsymbol{\nu}_j(k), \tag{13}$$

where $\mathbf{A}$ is the state transition matrix and $\boldsymbol{\nu}_j(k)$ is a noise term with the covariance matrix $\mathbf{Q}$. The specific forms of $\mathbf{A}$ and $\mathbf{Q}$ are:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{14}$$

and

$$\mathbf{Q} = \begin{bmatrix} \frac{\Delta T^2}{4}\sigma_1^2 & \frac{\Delta T}{2}\sigma_1^2 & 0 & 0 \\ \frac{\Delta T}{2}\sigma_1^2 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \frac{\Delta T^2}{4}\sigma_2^2 & \frac{\Delta T}{2}\sigma_2^2 \\ 0 & 0 & \frac{\Delta T}{2}\sigma_2^2 & \sigma_2^2 \end{bmatrix}, \tag{15}$$

where σ_1^2 and σ_2^2 are the variances.

We assume that the state transition of the tracking system is Markovian, and it can be carried out in the POMDPs framework. Here we briefly describe the multi-agent POMDPs model. A discrete-time POMDP is a 8-tuple:

$$\langle \mathcal{S}, \mathcal{A}, \mathcal{Z}, T, O, R, b, \lambda \rangle, \tag{16}$$

where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions, $\mathcal{Z}$ is the set of observations, $T(\mathbf{s}', \mathbf{s}, \mathbf{a}) = p(\mathbf{s}'|\mathbf{s}, \mathbf{a})$ is the conditional transition probability between states, $O(\mathbf{z}, \mathbf{s}', \mathbf{a}) = p(\mathbf{z}|\mathbf{s}', \mathbf{a})$ is the conditional observation probability, b is a probability distribution over the state space as the states are partially observable, and λ is the discount factor. The reward obtained at each step is $R(\mathbf{s}, \mathbf{a})$, which is determined by the current state $\mathbf{s} \in \mathcal{S}$ and the selected action $\mathbf{a} \in \mathcal{A}$. The goal is to maximize the sum of expected rewards weighted by a discount factor $\lambda \in [0, 1]$ during h time steps. The true state is unknown and the belief state $b(\mathbf{s})$ is maintained as a probability density function over the state space. The objective of POMDPs is to obtain a policy $\pi(b) \rightarrow \mathbf{a}$ that maps belief states into actions, so that the total expected reward is maximized. This expected reward gathered by following π determined by belief b is called the value function. The value function at horizon h can be constructed iteratively from the value function at horizon $h-1$ (value iteration):

$$\boldsymbol{V}^h(b) = \max_{\mathbf{a}} \left\{ \sum_{\mathbf{s}\in\mathcal{S}} R(\mathbf{s}, \mathbf{a}) b(\mathbf{s}) + \lambda \sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{z}|b, \mathbf{a}) \boldsymbol{V}^{(h-1)}(b_{\mathbf{a}}^{\mathbf{z}}) \right\}, \tag{17}$$

where $p(\mathbf{z}|b, \mathbf{a})$ is the conditional observation probability based on the belief state and $b_{\mathbf{a}}^{\mathbf{z}}$ is the updated belief according to a Bayesian filter:

$$b_{\mathbf{a}}^{\mathbf{z}} = \eta O(\mathbf{z}, \mathbf{s}', \mathbf{a}) \sum_{\mathbf{s}\in\mathcal{S}} T(\mathbf{s}', \mathbf{s}, \mathbf{a}) b(\mathbf{s}), \tag{18}$$

where η is a normalizing constant.

For a multi-UAV system, this process should be expanded to the multi-agent POMDPs. Specifically, the transition function is defined over the set of joint actions $\mathcal{A}_{joint} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_{N^u}$, the reward of the whole system defined over the joint set of states $\mathcal{S}_{joint} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_{N^u}$ and the joint actions: $\mathcal{S}_{joint} \times \mathcal{A}_{joint} \rightarrow R_{joint}$. Without loss of generality, the reward can be decomposed into the local factors R_i^{local} from the UAV i and the joint factor R^{joint}. Thus, the global reward can be expressed as

$$R_{joint}\left(\mathbf{s}_{joint}, \mathbf{a}_{joint}\right) = R_1^{local}\left(\mathbf{s}_1, \mathbf{a}_1\right) + \cdots + R_{N^u}^{local}\left(\mathbf{s}_{N^u}, \mathbf{a}_{N^u}\right) + R^{joint}\left(\mathbf{s}_{joint}, \mathbf{a}_{joint}\right), \tag{19}$$

where $\mathbf{s}_i (i \in \mathcal{U})$ includes the state of the ith UAV and the state of the target tracked by this UAV.

We define $\{c_i\}_{i=1,\ldots,N^u} \in \{1, \ldots, N^t\}$ as the targets for the corresponding UAVs. Then the local rewards in (19) are determined by the target to be tracked and the selected action. The value function in (17) is modified as follows:

$$\begin{aligned} \boldsymbol{V}^k(b) = \max_{a_{joint}, c_{joint}} \Bigg\{ & \sum_{\mathbf{s}_{joint}\in\mathcal{S}_{joint}} \left[R_1^{c_1}\left(\mathbf{s}_1, a_1\right) + \cdots + R_{N^u}^{c_{N^u}}\left(\mathbf{s}_{N^u}, a_{N^u}\right) \right. \\ & \left. + R^{joint}\left(\mathbf{s}_{joint}, a_{joint}, c_{joint}\right) \right] b(\mathbf{s}_{joint}) \\ & + \lambda \sum_{\mathbf{z}_{joint}\in\mathcal{Z}_{joint}} p(\mathbf{z}_{joint}|b, a_{joint}) \boldsymbol{V}^{(k-1)}(b_{a_{joint}}^{\mathbf{z}_{joint}}) \Bigg\}. \end{aligned} \tag{20}$$

We remove the central control unit of the centralized multi-agent POMDPs, and the goal of the distributed POMDPs is to compute an optimal joint policy $\pi^* = \{\pi_1, \ldots, \pi_{N^u}\}$ that maximizes the expected discount reward. Each UAV maintains a JMTPD using a distributed max-consensus-based information fusion algorithm. The value function (20) is maximized by selecting a policy and using a distributed auction-based decision-making algorithm. We will discuss the details of the algorithms in Sect. 4.

4 Distributed Kalman filtering based on max-consensus and constrained communication

In this section, we discuss a distributed approach to estimate the joint belief state of multiple targets among multiple UAVs. Each UAV maintains an information table. Taking the ith UAV as an example, its information table is represented by the tuple:

$$\left\langle \hat{\mathbf{s}}_i(k), \hat{\mathbf{P}}_i(k), \mathcal{X}_i(k), \boldsymbol{\gamma}_i(k), \bar{\mathbf{s}}_i(k), \bar{\mathbf{P}}_i(k) \right\rangle. \tag{21}$$

The definition of each variable is as follows

- $\hat{\mathbf{s}}_i(k) := [\hat{\mathbf{s}}^t_{i1}(k), \hat{\mathbf{s}}^t_{i2}(k), \ldots, \hat{\mathbf{s}}^t_{iN^t}(k)]$ is the local estimation of the multi-target states and $\hat{\mathbf{P}}_i(k) := [\hat{\mathbf{P}}^t_{i1}(k), \hat{\mathbf{P}}^t_{i2}(k), \ldots, \hat{\mathbf{P}}^t_{iN^t}(k)]$ is the error covariance matrix for the multi-target states estimation. These two variables are produced by the local Kalman filter of the ith UAV at the time step k. The detailed derivation process is shown in (23)–(25).
- The measurement area $\mathcal{X}_i(k) := \left\{ \mathbf{p} \,\middle|\, \left\| \mathbf{p}^u_i(k) - \mathbf{p} \right\| \leq r^O_i \right\}$ is used to model the limited sensing capability of the ith UAV. The ith UAV can perform a measurement of the target j if and only if the position of the target is within the measurement range, i.e., $\mathbf{p}^t_j(k) \in \mathcal{X}_i(k)$, $j \in \mathcal{T}$.
- The quantity $\boldsymbol{\gamma}_i(k) := [\gamma^t_{i1}(k), \gamma^t_{i2}(k), \ldots, \gamma^t_{iN^t}(k)]$ is the perception confidence and quantifies the estimated accuracy of the ith UAV's estimation of the multi-target states. We note that the perception confidence value is proportional to the accuracy of the estimation. Specifically, by considering two estimated states of UAV m and UAV n for the same target j, represented by $\hat{\mathbf{s}}^t_{mj}(k)$ and $\hat{\mathbf{s}}^t_{nj}(k)$, if $E\left[\left\| \hat{\mathbf{s}}^t_{mj}(k) - \mathbf{s}^t_j(k) \right\|^2 \right] < E\left[\left\| \hat{\mathbf{s}}^t_{nj}(k) - \mathbf{s}^t_j(k) \right\|^2 \right]$, then $\gamma_{mj}(k) > \gamma_{nj}(k)$ holds (where $\mathbf{s}^t_j$ is the real state of target j and $E[\cdot]$ indicates the statistical expectation). Since each UAV performs an individual estimation through a local Kalman filter, just the trace of the covariance matrix can be used to quantify the perception confidence value as follows:

$$\gamma^t_{ij}(k) := \frac{1}{Trace\left[\hat{\mathbf{P}}^t_{ij}(k)\right]}, \; j \in \mathcal{T}. \tag{22}$$

  Refer to "Appendix A" for the reason of using the matrix trace.
- The estimated states $\bar{\mathbf{s}}_i(k) := [\bar{\mathbf{s}}^t_{i1}(k), \bar{\mathbf{s}}^t_{i2}(k), \ldots, \bar{\mathbf{s}}^t_{iN^t}(k)]$ and the corresponding covariance matrix $\bar{\mathbf{P}}_i(k) := [\bar{\mathbf{P}}^t_{i1}(k), \bar{\mathbf{P}}^t_{i2}(k), \ldots, \bar{\mathbf{P}}^t_{iN^t}(k)]$ are the values after the information fusion based on the max-consensus protocol.

The information fusion consists of two phases, i.e., the individual estimation phase and the distributed fusion phase. Assuming that each UAV owns the information of the tuple (21) after the $(k-1)th$ iteration, we describe the two phases of the kth iteration from the perspective of the ith UAV as follows. In the individual estimation phase, the first four items of the information table have been updated from the value at the time step $k-1$ to the value

at the time step k by local Kalman filtering. In the distributed fusion phase, the initialization of the ith UAV's information table is: $\left\langle \hat{\mathbf{s}}_i(k), \hat{\mathbf{P}}_i(k), \mathcal{X}_i(k), \boldsymbol{\gamma}_i(k), \bar{\mathbf{s}}_i(k-1), \bar{\mathbf{P}}_i(k-1) \right\rangle$. The distributed fusion process based on the max-consensus protocol is used to update the value of the last two items.

In the individual estimation phase, each UAV uses the Kalman filter to estimate the states of all targets and updates its information table in iteration. Specifically, if the UAV does not observe the target, the filter just uses the estimated state from the previous time step to produce an state estimation at the current time step. If the UAV observes the target, the predicted state is combined with the current observation information to refine the state estimation.

Take the state estimation process of the target j by UAV i as an example. The ith UAV performs a Kalman filter with the aid of the posteriori state $\bar{\mathbf{s}}_i(k-1)$ and the error posteriori covariance matrix $\bar{\mathbf{P}}_i(k-1)$ at time step $k-1$. We use subscript ij to indicate the information about target j, which is stored by UAV i.

The ith UAV uses the existing information to predict the state of target j, as follows:

$$\begin{cases} \hat{\mathbf{s}}_{ij}^t(k|k-1) = \mathbf{A}\bar{\mathbf{s}}_{ij}^t(k-1), \\ \hat{\mathbf{P}}_{ij}^t(k|k-1) = \mathbf{A}\bar{\mathbf{P}}_{ij}^t(k-1)\mathbf{A}^{\mathrm{T}} + \mathbf{Q}, \end{cases} \tag{23}$$

where $\mathbf{A}$ and $\mathbf{Q}$ are related to the hypothetical kinematic model (13) of target j. $\hat{\mathbf{s}}_{ij}^t(k|k-1)$ and $\hat{\mathbf{P}}_{ij}^t(k|k-1)$ are the estimated state and estimated covariance, respectively.

If UAV i cannot observe target j in its FOV, only the prediction is performed. Then, we have

$$\begin{cases} \hat{\mathbf{s}}_{ij}^t(k) = \hat{\mathbf{s}}_{ij}^t(k|k-1), \\ \hat{\mathbf{P}}_{ij}^t(k) = \hat{\mathbf{P}}_{ij}^t(k|k-1). \end{cases} \tag{24}$$

If target i is observable, i.e., $\mathbf{z}_{ij}^C(k) \in \mathcal{X}_i(k)$, UAV i updates the prediction results with the measurements as follows

$$\begin{cases} \mathbf{K}_{ij}(k) = \hat{\mathbf{P}}_{ij}^t(k|k-1)\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\bar{\mathbf{P}}_{ij}^t(k-1)\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{ij}(k)\right)^{-1}, \\ \hat{\mathbf{s}}_{ij}^t(k) = \hat{\mathbf{s}}_{ij}^t(k|k-1) + \mathbf{K}_{ij}(k)\left(\mathbf{z}_{ij}^C - \mathbf{H}\hat{\mathbf{s}}_{ij}^t(k|k-1)\right), \\ \hat{\mathbf{P}}_{ij}^t(k) = \left(\mathbf{I} - \mathbf{K}_{ij}(k)\mathbf{H}\right)\hat{\mathbf{P}}_{ij}^t(k|k-1). \end{cases} \tag{25}$$

We note that $\mathbf{R}_{ij}(k)$ is approximated by the calculation of the target measurement. The ith UAV ends the individual estimation phase by computing the perception confidence value $\gamma_{ij}^t(k)$ based on (22). Eventually, each UAV estimates the state of all targets by the above filter algorithm and updates its information table at the time step k.

In the distributed fusion phase, each UAV exchanges information with its neighbors and updates the local information based on the max-consensus protocol. Then, the estimation of each target state with the highest perception confidence value in the connected network is selected, and the posteriori state and its posteriori error covariance matrix are propagated over the network in finite time. We note that due to the limitation of communication distance, the entire UAV network may be decomposed into multiple independent connected subgraphs in each tracking decision-making cycle. In order to make the information fusion phase converge before the subsequent decision-making phase, a new sampling time connected with the discrete time index t is defined to drive the max-consensus protocol running at a higher frequency. In order to express the process clearly in the new time index t, we define a set of temporary variables with t as a parameter, i.e., $\left\langle \hat{\mathbf{s}}_i'(t), \hat{\mathbf{P}}_i'(t), \boldsymbol{\gamma}_i'(t) \right\rangle$. The initial value of the

temporary variable, in which $t = 0$, is a part of the information table at the time step k:

$$\left\langle \hat{\mathbf{s}}_i'(t), \hat{\mathbf{P}}_i'(t), \boldsymbol{\gamma}_i'(t) \right\rangle_{t=0} := \left\langle \hat{\mathbf{s}}_i(k), \hat{\mathbf{P}}_i(k), \boldsymbol{\gamma}_i(k) \right\rangle. \tag{26}$$

The UAV sends information tuple $\left\langle \hat{\mathbf{s}}_i'(t-1), \hat{\mathbf{P}}_i'(t-1), \boldsymbol{\gamma}_i'(t-1) \right\rangle$ and receives the information tuple from its neighbors at the t-th iteration. Then, the ith UAV updates its temporary information tuple to $\left\langle \hat{\mathbf{s}}_i'(t), \hat{\mathbf{P}}_i'(t), \boldsymbol{\gamma}_i'(t) \right\rangle$ using the max-consensus protocol. The update rule based on the max-consensus is:

$$\gamma_{ij}'^{t}(t) = \max_{q \in \mathcal{I}_i^{u_n}} \left\{ \gamma_{qj}'^{t}(t-1) \right\}, \ j \in \mathcal{T}. \tag{27}$$

The ith UAV also updates the corresponding estimated state $\hat{\mathbf{s}}_{ij}'^{t}(t)$ and error covariance matrices $\hat{\mathbf{P}}_{ij}'^{t}(t)$, respectively. The above communication and update processes are repeated until a consensus on the confidence value is reached. In order to explore the convergence of the state estimation, a concept named the max-consensus for the target tracking system is defined, according to the general definition of max-consensus in [43].

Definition 1 (Max-Consensus): Note the connected communication subgraph $\mathcal{G}_i^u = \{\mathcal{I}_i^u, \mathcal{E}_i^u\}$, where the ith UAV is located. The subgraph is composed of N_i^u UAVs, and each UAV has an initial vector of the perception confidence value related to all targets $\boldsymbol{\gamma}_i'(0) = [\gamma_{i1}'^{t}(0), \gamma_{i2}'^{t}(0), \ldots, \gamma_{iN^t}'^{t}(0)]^{\mathrm{T}}$, $i \in \mathcal{I}_i^u$. The update rule is (27), then the max-consensus is said to be achieved, if $\exists l \in \mathbb{N}^0$ such that

$$\begin{aligned} \gamma_{ij}'^{t}(t) &= \gamma_{qj}'^{t}(t) \\ &= \max \left\{ \gamma_{1j}'^{t}(0), \gamma_{2j}'^{t}(0) \ldots, \gamma_{N_i^u j}'^{t}(0) \right\}, \ \forall j \in \mathcal{T}, \ \forall t \geq l, \ \forall i, q \in \mathcal{I}_i^u, \end{aligned} \tag{28}$$

where l is the lower bound of the number of iterations to make the information in the whole connected graph $\mathcal{G}_i^u$ achieve the max-consensus. However, the value of l is difficult to obtain because it is related not only to the number of UAVs, but also to the specific information values.

For the ith UAV, after a limited number of iterations, represented by D_i, the max-consensus is achieved within the connected subnetwork where UAV i is located. Then, the ith UAV updates its information table, such that $\bar{\mathbf{s}}_i(k) = \hat{\mathbf{s}}_i'(D_i)$ and $\bar{\mathbf{P}}_i(k) = \hat{\mathbf{P}}_i'(D_i)$. These two variables will be returned as the basis for multi-target tracking decision-making in the current cycle and the distributed Kalman filter in the next cycle. In order to obtain the value of D_i, we first present the following lemma:

Lemma 1 *Given an undirected graph $\mathcal{G}_i^u$, the max-consensus (28) is achieved if and only if $\mathcal{G}_i^u$ is connected. In addition, the max-consensus will be achieved in a number of iterations l, which satisfies $l \leq d(\mathcal{G}_i^u) \leq N_i^u - 1$, where $d(\mathcal{G}_i^u)$ is the graph diameter* [43].

In order to ensure the information consistency of all UAVs, based on Lemma 1, the iteration number of the max-consensus algorithm is usually set to the upper bound of the diameter of the graph [44], which is the total number of UAVs, i.e., $N^u - 1$. However, the number of iterations is usually redundant. Firstly, the UAVs only communicates directly or indirectly with other UAVs in the same communication subgraph and the max-consensus only holds for a subset of all UAV nodes. Secondly, the diameter of the graph (i.e., the subgraph where the ith UAV is located) is the upper bound of the number of iterations, which is $N_i^u - 1$. Hence, it is likely to be redundant for most UAVs local distributed information fusion algorithm

required. In addition, taking into account the dynamics of the system and the locality of the information, no "structure information" about the subgraph is assumed in our algorithm, including the size or the bound on the size of the network. Therefore, the UAV does not know the number of UAVs in its subgraph N_i^u.

Without loss of generality, we also study the minimum number of iterations D_i based on (28) about a distributed system from the viewpoint of the ith UAV. We introduce the following theorem:

Theorem 1 *In the distributed information fusion process based on the max-consensus protocol, if the update rule is as shown in Eq. (27), then for the ith UAV, the iteration number to achieve max-consensus is the diameter of the shortest path tree (SPT) of $\mathcal{G}_i^u$ rooted at node i.*

The detailed proof of Theorem 1 is given in "Appendix B".

Remark 1 – In each iteration, the ith UAV only exchanges information with the neighbors instead of all the UAVs connected to the graph $\mathcal{G}_i^u$, i.e., $\mathcal{I}_i^{un} \subseteq \mathcal{I}_i^u$.
– The minimum number of iterations D_i can ensure the ith UAV to get the maximum value, but for different UAVs in the network, the minimum numbers of iterations are different, because each UAV may have different relative topological relations in the network.
– The diameter of SPT of $\mathcal{G}_i^u$ rooted at node i can be obtained through a distributed algorithm [45], referred to as the all-pairs shortest paths (APSP) algorithm, which does not need to know the global information about the graph.

Clearly, if $\forall m \in \mathcal{I}_i^u$ operates the max-consensus algorithm (27) with the number of iterations set to $D_m(\mathcal{G}_i^u)$, the max-consensus of the whole graph $\mathcal{G}_i^u$ will be achieved, where $D_m(\mathcal{G}_i^u)$ is the diameter of SPT of $\mathcal{G}_i^u$ rooted at node m.

It is assumed that decision-making also occurs synchronously, like communications. After the entire network has the completed information fusion, decision-making instants may be defined either by a clock or by the occurrence of external events with global synchronization. The next section introduces the distributed decision-making process.

5 Distributed auction POMDPs based on Fisher information

In this section, we discuss how the decision-making process is done in a distributed manner. Based on the JMTPD obtained by the distributed information estimation and fusion algorithm, each UAV predicts the state of the targets and determines its optimal tracking policy by locally maximizing the value function (20). Note that the joint reward term R^{joint} in (20) is not considered in the distributed decision making process, since it is hard to be obtained in a distributed way. To narrow the global optimality gap caused by dropping R^{joint}, we recompute the policy at each time step. The specific flow of the distributed algorithm is as follows:

– At first, the optimal tracking action is selected from action set $\mathcal{A}$ on the basis of the prediction of the state of the target and its kinematic model (1), in which the prediction is based on the JMTPD obtained by the distributed information fusion.
– At the same time, the corresponding tracking reward, which is $R_{ij}(\mathbf{s}_i(k), a_i(k))$, is obtained. In order to decide which target to track, the UAV shares the tracking reward value as its own bids for each target to its neighbors.
– With the bids received from its neighbors, each UAV computes the policy by locally employing Hungarian algorithm [46], which guarantees the optimality (without considering R^{joint}) and the efficiency in deriving the tracking policy.

– Finally, each UAV carries out its own tracking policy to track the assigned target with the best action.

Remark 2 If the number of the targets is not smaller than that of the UAVs, each UAV can be assigned to a target within one assignment. In this case, some targets are not tracked by UAVs. If the number of the targets is smaller than that of the UAVs, the assignment algorithm will be repeated until each UAV is assigned to a target.

Each UAV in the multi-UAV system runs the same decision-making algorithm. Then the proposed distributed auction algorithm provides a set of action policies for the multi-UAV system ($\pi' = \{\pi'_1, \pi'_2, \ldots, \pi'_{N^u}\}$) that is suboptimal. Considering that the value function (20) is challenging to obtain in a distributed way because of R^{joint}, the value function used to obtain the policy is redefined as (29).

$$\begin{aligned} \boldsymbol{V}'^{(h)}(b) = & \sum_{\mathbf{s}_{joint} \in \mathcal{S}_{joint}} \left[R_1^{c_1}\left(\mathbf{s}_1, a_1(\pi'_1)\right) + \cdots + R_{N^u}^{c_{N^u}}\left(\mathbf{s}_{N^u}, a_{N^u}(\pi'_{a_{N^u}})\right) \right] b(\mathbf{s}_{joint}) \\ & + \lambda \sum_{\mathbf{z}_{joint} \in \mathcal{Z}_{joint}} p(\mathbf{z}_{joint} | a_{joint}, b) \boldsymbol{V}'^{(h-1)}(b_{a_{joint}}^{\mathbf{z}_{joint}}) \end{aligned} \tag{29}$$

with

$$c_i = \arg\max_{c_i} \left[\sum_{q \in \mathcal{I}_i^{un}} V_q^{(h)}(b_q) \right], i = 1, 2, \ldots, N^u,$$

where $V_q^{(h)}(b_q)$ is the local reward value of the jth UAV, derived from (17). Therefore, the algorithm is a local optimal solution as it does not optimize the original value function directly. Note that the target c_i to be tracked is computed by only considering the ith UAV's neighbors $\mathcal{I}_i^{un}$ and the local belief from JMTPD, which is obtained from its connected teammates $\mathcal{I}_i^u$. At the same time, the auction algorithm guarantees that the UAV and its neighbors track all targets as much as possible.

Remark 3 An offline planning phase is considered to calculate the optimal action strategy for each behavior in [16]. However, it cannot be employed in our problem to obtain the optimal action ahead of time for the state of each target at any time, as the state of the target is continuous.

All processes of decision-making are performed online, which requires us to design a reasonable and easy-to-calulate reward function. The objective of decision-making is to adjust the position/attitude of the UAV based on its predictions of the target state to obtain a better measurement. Therefore, the reward function is similar to the confidence value (22).

The determinant of the Fisher information matrix (FIM) is used as the reward value instead of the trace of the Kalman filter error covariance matrix. The FIM is from information geometry, which offers comprehensive results about statistical models simply by considering them as geometrical objects and the statistical structures as geometrical structures. We use this reward function for the following reasons. Firstly, the linearized Kalman filter is applied in the Cartesian coordinate system, while the real measurement is carried out in the sensor coordinate system. This linearized conversion ignores the bias in the converted measurements, which may lead to substantially degraded performance and even filtering divergence. On the contrary, the FIM in information geometry processes the original measurement data from the sensor. It is derived from the measurement model in the natural sensor polar coordinate system

directly and has a definite physical meaning, which reflects the volume of information of the measurement data. Besides, as is well known, the FIM defines the volume of information which is provided by the measurement in estimating the parameters. It is also the inverse of the Cramer Rao Lower Bound (CRLB), which is the lower bounds of the position error covariance, i.e.,

$$FIM^{-1} = CRLB \leq E\Big[\Big\|\hat{\mathbf{p}}_j^t - \mathbf{p}_j^t\Big\|^2\Big], j \in \mathcal{T}, \tag{30}$$

where $\hat{\mathbf{p}}_j^t$ and $\mathbf{p}_j^t$ are the estimated and true values of the target state, respectively. Therefore, maximizing the FIM can reduce the uncertainty. This line of reasoning is a main theme in several papers dealing with active sensing [42].

We define $\mathbf{G}_{ij}(k)$ to express the FIM of UAV i at time k, regarding $c_i = j(j \in \mathcal{T})$, then the reward of the POMDP is $R_{ij}(\mathbf{s}_i(k), a_i(k)) = \left|\mathbf{G}_{ij}(k+1)\right|$ ($|\cdot|$ denotes the determinant of the matrix). Then the following Algorithm 1 summarizes the distributed decision-making algorithm based on Fisher information, which is performed locally at the ith UAV.

Algorithm 1: The distributed POMDP based on auction for the ith UAV at time k

Input : $\bar{\mathbf{s}}_i(k), \bar{\mathbf{P}}_i(k)$
Output: The optimal action $\phi_i^u(k)_{opt}$

for $j = 1$ **to** N^t **do**
  $\phi_{ij}^u(k)_{opt} \leftarrow \arg\max_{\phi_{ij}^u(k),\ldots,\phi_{ij}^u(k+h-1)\in\mathcal{A}} \sum_{t=1}^{h} |\mathbf{G}_{ij}(k+t)|$;
  $V_{ij}(k)_{opt} \leftarrow \max_{\phi_{ij}^u(k),\ldots,\phi_{ij}^u(k+h-1)\in\mathcal{A}} \sum_{t=1}^{h} |\mathbf{G}_{ij}(k+t)|$;
end
Send $V_i(k)_{opt} := [V_{i1}(k)_{opt}, V_{i2}(k)_{opt}, \ldots, V_{iN^t}(k)_{opt}]$;
Receive $V_q(k)_{opt}, q \in \mathcal{I}_i^{un}, q \neq i$;
$\Phi(k) \leftarrow \{V_{qj}(k)_{opt}\}_{q,j}, q \in \mathcal{I}_i^{un}, j \in \mathcal{T}$;
The tracked target $c_i \leftarrow Hungarian(\Phi(k))$;
$\phi_i^u(k)_{opt} \leftarrow \phi_{ic_i}^u(k)_{opt}$.

In Algorithm 1, each UAV has updated their local information table, i.e., $\langle \bar{\mathbf{s}}_i(k), \bar{\mathbf{P}}_i(k) \rangle$ before making a decision at iteration k, through state estimation and information fusion. Then $\bar{\mathbf{s}}_i(k)$ is used to predict the positions of the targets. Based on the predictions, the UAV selects the optimal action for each target to maximize the corresponding value function with a horizon h, which is deduced from the sum of the determinant of the FIM at each time.

Remark 4

- In this paper, we assume that the UAVs do not switch their tracked targets within each time step (which would just be true for a horizon equal to 1). However, they can change their targets from one time step to the next.
- The derivation of FIM ($\mathbf{G}_{ij}(k+1)$) in the algorithm is shown in "Appendix C". The reward value is used as the bid for the distributed auction algorithm.

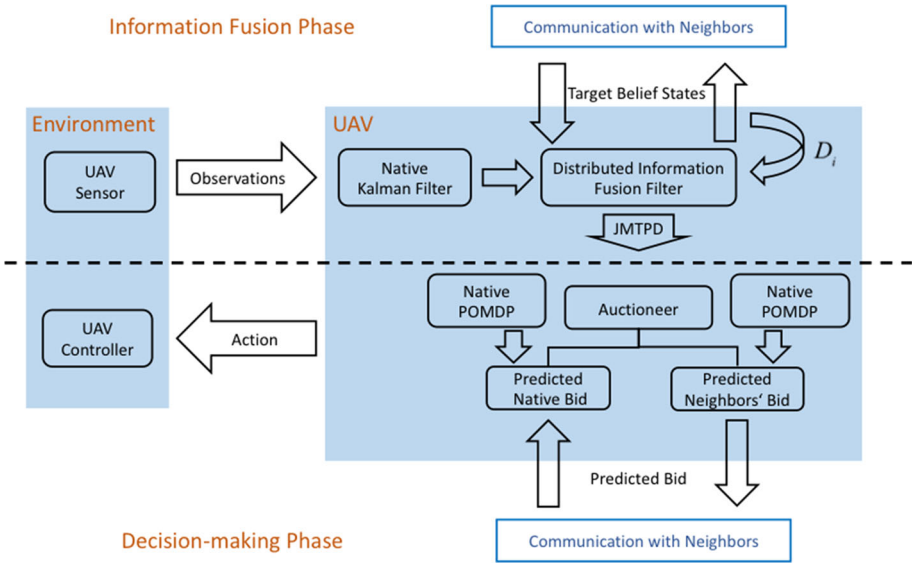

Fig. 3 The distributed multi-UAV POMDP framework for each UAV. Arrows indicate the flow of data. D_i is the number of times the ith UAV exchanges information with its neighbor(s)

6 Comprehensive design

The distributed information estimation and fusion algorithm, and the auction-based distributed decision algorithm are described independently, given in Sects. 4 and 5, respectively. As the state is partially observable and its transition is Markovian, the whole tracking algorithm can be unified under the distributed multi-UAV POMDP framework. As shown in Fig. 3, each UAV makes its own collaborative decision locally, which is a POMDP.

The framework shown in Fig. 3 includes two key phases (Information Fusion Phase and Decision-making Phase), where the information flows are indicated by the directed arrows. Due to the uncertainty of the sensor measurement, the real state of the target is not available for decision-making. The state of the target $\mathbf{s}_j^t (j \in \mathcal{T})$ needs to be approximated by belief state of the target, which is a part of the JMTPD. As shown in the "Information Fusion Phase", a local Kalman filter and a distributed information fusion strategy based on the max-consensus protocol are proposed to obtain and maintain the JMTPD among the UAVs. For the ith UAV, the information reaches the consensus in the connected communication graph after D_i iterations, which is the input of the multi-UAV POMDP. According to the JMTPD, each UAV operates a local POMDP firstly during the "Decision-making Phase". Then, each UAV exchanges the Fisher information reward value for each target with its neighbors, which acts as the auction bids. Finally, a distributed auction method is proposed to solve this multi-UAV POMDP approximately, in which the sub-optimal solution is able to reduce the computational complexity dramatically. In summary, the distributed tracking approach is built on two mechanisms: the distributed states estimation and information fusion, and the auction POMDP based on Fisher information. The former allows multiple UAVs to share information and obtain a common joint belief state, and the latter is used to assign the target to be tracked and to select the action policy in a cooperative way.

The following Algorithm 2 summarizes the distributed approach in which the estimation, fusion and decision-making are performed locally at the ith UAV.

Algorithm 2: Tracking decision-making for the ith UAV at time k

Input : $\bar{\mathbf{s}}_i(k-1), \bar{\mathbf{P}}_i(k-1)$
Output: The optimal action $\phi_i^u(k)_{opt}$

for $j = 1$ **to** N^t **do**
  $\hat{\mathbf{s}}^t_{ij}(k|k-1) \leftarrow \mathbf{A}\bar{\mathbf{s}}^t_{ij}(k-1)$;
  $\hat{\mathbf{P}}^t_{ij}(k|k-1) \leftarrow \mathbf{A}\bar{\mathbf{P}}^t_{ij}(k-1)\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$;
  if *Target j is measurable* **then**
    Obtain $\mathbf{z}^C_{ij}(k)$;
    Approximate $\mathbf{R}_{ij}(k)$, ref. (2), (3), (12) and (10);
    $\mathbf{K}_{ij}(k) \leftarrow \hat{\mathbf{P}}^t_{ij}(k|k-1)\mathbf{H}^{\mathrm{T}}\Big(\mathbf{H}\bar{\mathbf{P}}^t_{ij}(k-1)\mathbf{H}^{\mathrm{T}} + \mathbf{R}_{ij}(k)\Big)^{-1}$;
    $\hat{\mathbf{s}}^t_{ij}(k) \leftarrow \hat{\mathbf{s}}^t_{ij}(k|k-1) + \mathbf{K}_{ij}(k)\big(\mathbf{z}^C_{ij} - \mathbf{H}\hat{\mathbf{s}}^t_{ij}(k|k-1)\big)$;
    $\hat{\mathbf{P}}^t_{ij}(k) \leftarrow \big(\mathbf{I} - \mathbf{K}_{ij}(k)\mathbf{H}\big)\hat{\mathbf{P}}^t_{ij}(k|k-1)$;
  else
    $\langle \hat{\mathbf{s}}^t_{ij}(k), \hat{\mathbf{P}}^t_{ij}(k)\rangle \leftarrow \langle \hat{\mathbf{s}}^t_{ij}(k|k-1), \hat{\mathbf{P}}^t_{ij}(k|k-1)\rangle$;
  end
  $\gamma^t_{ij}(k) \leftarrow \big[Trace(\hat{\mathbf{P}}^t_{ij}(k))\big]^{-1}$;
end
$\langle \hat{\mathbf{s}}'_i(0), \hat{\mathbf{P}}'_i(0), \boldsymbol{\gamma}'_i(0)\rangle \leftarrow \langle \hat{\mathbf{s}}_i(k), \hat{\mathbf{P}}_i(k), \boldsymbol{\gamma}_i(k)\rangle$;
Compute D_i (from protocol APSP);
for $t = 1$ **to** D_i **do**
  Send $\langle \hat{\mathbf{s}}'_i(t-1), \hat{\mathbf{P}}'_i(t-1), \boldsymbol{\gamma}'_i(t-1)\rangle$;
  Receive $\langle \hat{\mathbf{s}}'_l(t-1), \hat{\mathbf{P}}'_l(t-1), \boldsymbol{\gamma}'_l(t-1)\rangle, l \in \mathcal{I}^{un}_i$;
  for $j = 1$ **to** N^t **do**
    $m \leftarrow \arg\max_{l \in \mathcal{I}^{un}_i} \gamma'^t_{lj}(t-1)$;
    $\langle \hat{\mathbf{s}}'^t_{ij}(t), \hat{\mathbf{P}}'^t_{ij}(t), \gamma'^t_{ij}(t)\rangle \leftarrow \langle \hat{\mathbf{s}}'^t_{mj}(t-1), \hat{\mathbf{P}}'^t_{mj}(t-1), \gamma'^t_{mj}(t-1)\rangle$;
  end
end
$\langle \bar{\mathbf{s}}_i(k), \bar{\mathbf{P}}_i(k)\rangle \leftarrow \langle \hat{\mathbf{s}}'_i(D_i), \hat{\mathbf{P}}'_i(D_i)\rangle$;
Run Algorithm 1.

In the absence of R^{joint}, our approach has lower computational complexity (w.r.t the number of robots and targets) than the centralized MPOMDP. The complexity of a single step of computing the policy bid for the MPOMDP is $\mathcal{O}(N^u N^t|\mathcal{A}|)$, where $|\mathcal{A}|$ is the cardinality of the action set. We can see that this complexity is dependent on the numbers of the UAVs and the targets. In our method, each UAV computes its own bid and the complexity is $\mathcal{O}(N^t|\mathcal{A}|)$, which is independent of the number of the UAVs. Apart from computing the policies in the MPOMDP, the auction phase entails a complexity $\mathcal{O}\big(\big[\max(N^u, N^t)\big]^3\big)$ [46], which is the worst case in our distributed method. In our auction Dec-POMDP, the complexity for each

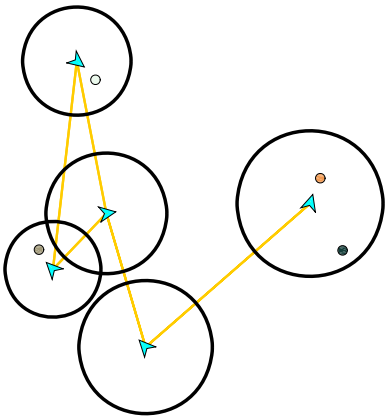

Fig. 4 The simulation scenario. The arrows represent the UAV, which indicates the heading direction, and the circles represent the randomly moving targets. The rings illustrate the observation range of the UAVs, and the straight lines illustrate the communication between the UAVs.

UAV is $\mathcal{O}\left(\left[\max(N_i^u, N^t)\right]^3\right)$, where N_i^u is the number of its neighbors and N^t is the number of the targets.

7 Simulation results

In this section, we apply the proposed Algorithm 2 to multi-target tracking by multiple UAVs. Due to the limited sensing ranges of the UAVs, only a subset of UAVs can observe parts of the targets during a time interval, and some uncovered areas of the space may still exist. The aim of the distributed targets tracking mission is to keep as many targets as possible in the FOV of multiple UAVs for as long as possible, which needs to estimate the states of the targets and make the action decisions of the UAVs based on the estimations. We assess the performance of the proposed approach through the observation rate, which is the ratio of the time duration when each target is observed to the total simulation time. To validate the performance of our method we carry out a series of simulation experiments. We compare the observation rates and the decision-making time of our method, the greedy policy selection algorithm and the centralized tracking algorithm. We also study the effect of the communication range and FOV on the performance of tracking.

The simulation is implemented in MATLAB R2016b (maci64) with 2.7 GHz Intel Core i5 Processor and 8 GB RAM, and one of the simulation scenarios is shown in Fig. 4. The performance of the tracking decision-making algorithms is analyzed by Monte Carlo simulations. The main parameters in the simulation system are set as follows: $V^u = 15\,\text{m/s}$, $k_d = 0.1056$, $k_r = 10.07$, $k_\theta = 0.01$, $\sigma_1 = \sigma_2 = 0.16$.

The targets are assumed to lay on a two-dimensional square field. Each UAV is flying at a specified constant altitude, and the altitude of each UAV is different. These UAVs are connected through an undirected communication graph with a limited communication range r_C. The position of the sensor is the same as that of the corresponding UAV with the same FOV 2α. Considering the randomness in the experiments, including the initial positions of the targets and the UAVs, and the acceleration of the targets movements, we performed 50 trials of each set of experiments with 500 simulation steps.

The UAV speed is a constant and the action set, which is the heading angle set, is $\{\pm\phi, 0\}, \phi > 0$. The action set is suitable for the angular velocity and speed of the UAV.

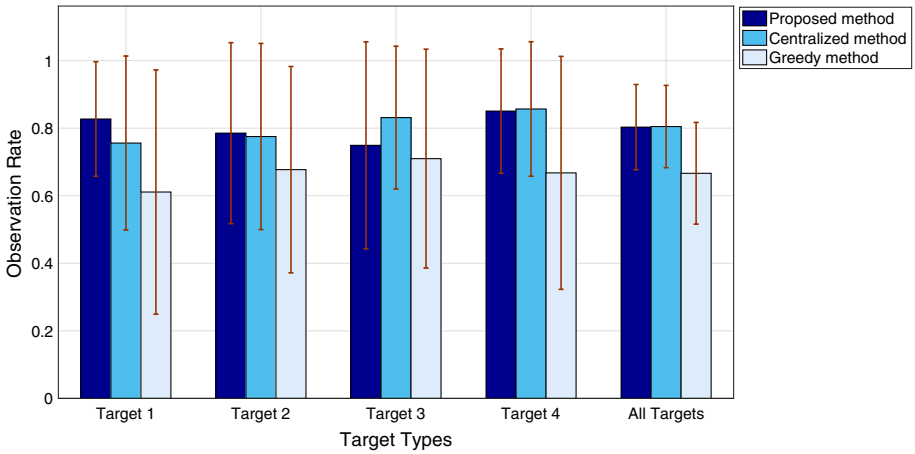

Fig. 5 Average simulation results ($\pm\sigma$) of 5 UAVs tracking 4 targets. The proposed distributed method is compared with the centralized and greedy methods. The bar graph illustrates the mean observation rate values for each method. The error bars represent the standard deviation

The parameters of the sensors are the same as those in the paper [27], except for the communication range and FOV.

7.1 Comparison with other algorithms

First of all, we compare our approach with the greedy policy selection algorithm and the centralized tracking algorithm. The greedy policy selection algorithm uses the distributed information fusion algorithm based on the max-consensus to estimate the states of the targets, which is the same as our method. But each UAV uses the greedy algorithm to select the target and the tracking action without any interaction, which is a completely decentralized POMDP. In the centralized tracking algorithm, there is a central entity to collect measurement data from all the UAVs, where the max-consensus protocol is employed for data fusion. Then the targets are assigned and the action policy at each UAV is derived accordingly.

In these three experiments, the algorithms are applied to the scene where 5 homogeneous UAVs track 4 ground moving targets. The communication range of the UAV is set to 50 m, and the FOV is 10° i.e., $\alpha = 5°$. The average results with standard deviation are presented in Figs. 5 and 6.

The average observation rate of each target and the overall average observation rate for all targets are shown in Fig. 5. The former reflects the tracking performance of this multi-UAV system for each target. Since the initial positions and motions of the targets are random, the overall average observation rate of all targets can be used to compare the performance of the algorithm directly. This value is 0.8031 ± 0.1262 for our proposed method, 0.8048 ± 0.1217 for the centralized method, and 0.6664 ± 0.1507 for the greedy method. The experimental results are consistent with the theoretical analysis, that is, the centralized algorithm is theoretically optimal and the greedy algorithm has the worst performance. It is worth noting that our distributed information fusion and tracking decision-making method performs nearly as well as the centralized approach, and outperforms the greedy method.

Furthermore, we sorted the observation rate of each target for each experiment, and averaged the results of 50 trials, which is shown in Fig. 6. The reason for looking at the minimum,

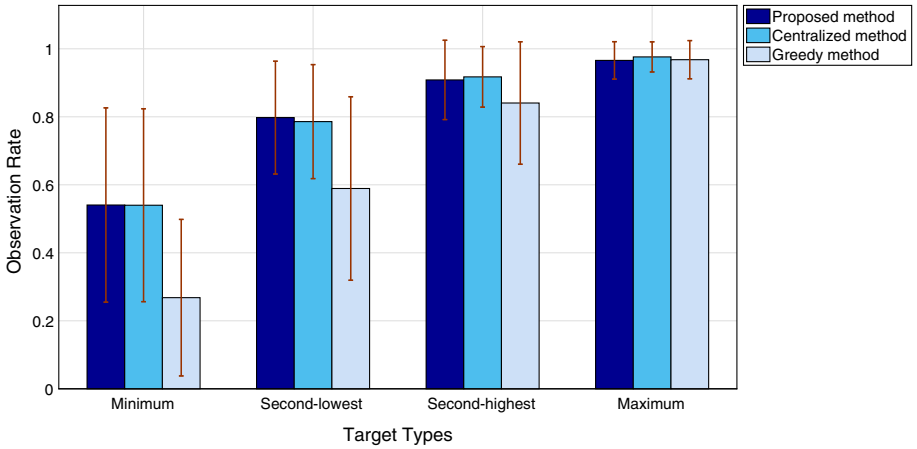

Fig. 6 Average simulation results ($\pm\sigma$) of 5 UAVs tracking 4 targets. The proposed distributed method is compared with the centralized and greedy methods. The observation rate ranges from the minimum to the maximum. The bar graph illustrates the mean observation rate values for each method. The error bars represent the standard deviation

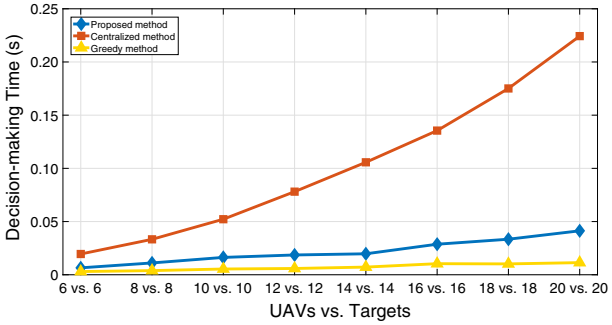

Fig. 7 Average simulation result of N UAVs tracking M targets (N vs. M). The proposed distributed method is compared with the centralized and greedy methods w.r.t. the decision-making time

second-lowest, etc. is to study the ranges of the observation rate, which can reflect the tracking ability of this multi-UAV system. Fig. 6 shows that our algorithm is similar to the centralized algorithm and superior to the greedy algorithm from another dimension. Although the initial positions and motions of the targets are random, the tracking algorithm can still keep the observation rate at a high level, which is greater than 0.5 in both our method and the centralized method.

In addition, in order to study the time performance of the algorithm, we compared the average decision-making time of different algorithms in different scenarios, as shown in Fig. 7. Even though our method takes much less time in computation compared with the centralized algorithm, the average rate for each target is still similar. The greedy algorithm is similar to our algorithm in this respect.

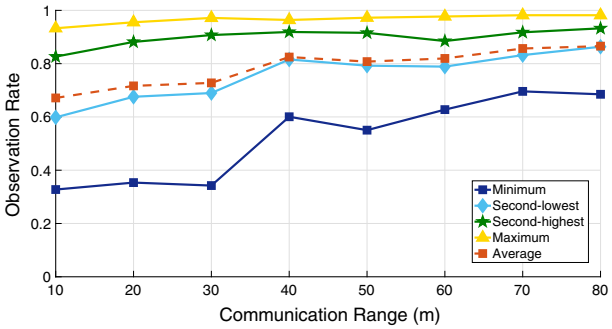

Fig. 8 The tracking performance of different communication ranges

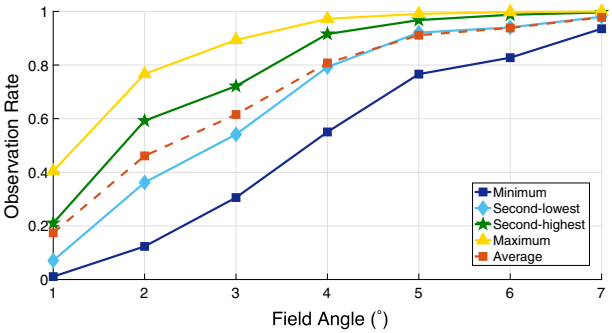

Fig. 9 The tracking performance of different FOVs. The x-coordinate is half of the FOV i.e., α

7.2 The effects of different parameters

Then we repeat the simulations with different communication ranges and FOVs to study the effect of these parameters on the performance of our distributed tracking decision-making method.

Similarly, we obtain the variations of the observation rate from lowest to highest and average the observation rate with these two parameters. When we change the communication range from 10 to 80 m, the FOV is set to be 10°. When we change the FOV, the communication range is 50 m (Figs. 8, 9).

As expected, from the minimum observation rate to the maximum observation rate, they become larger as the parameters increase. However, because of the randomness of other parameters and the limitation of the trials number, we cannot obtain a strict numerical relationship. Despite this, the results still show that the tracking performance can be enhanced by improving the communication and sensor capability. The results can be used as a reference to improve the observation rate in practical applications.

8 Conclusion

In this paper, we have proposed a distributed decision-making method for multiple UAVs tracking multiple targets. A distributed information fusion strategy based on the max-consensus protocol has been provided to estimate the JMTPD, and a decision-making

algorithm has been designed based on POMDPs and auctions to select the target to track and the tracking action. In the distributed information fusion phase, we have presented and proved the condition for achieving the max-consensus of the distributed system. Furthermore, we have provided a distributed auction method using fisher information from information geometry as a bid in the decision-making phase. The comparative simulations demonstrate the advantages of the algorithm in the observation rate over the greedy decision-making and centralized algorithms. We have also studied the effect of the communication range and FOV on tracking performance. The results show that improving the communication and observation capability of the UAV can significantly enhance the observation performance.

The kinematics model of the UAVs and targets in this paper is simplified, and the positioning errors of the UAVs and collision avoidance in the horizontal plane are not considered. In the future work, we will consider these issues. We also intend to investigate possible approaches to further reducing communication load and respond to the loss of targets or failure of communications more effectively. Another interesting extension is to consider the communication delay problem to improve the practical application of the decision-making algorithm.

A The properties of the trace of the covariance matrix

The estimations of multi-target states are produced by Kalman filter. The state of the filter is represented by two variables:

- $\hat{\mathbf{s}}^t(k)$, the posteriori state of the target at time k;
- $\hat{\mathbf{P}}^t(k)$, the posteriori error covariance matrix (a measure of the accuracy of the state estimation).

$$\hat{\mathbf{P}}^t(k) = \mathrm{var}\left(\mathbf{s}^t(k)\right). \tag{31}$$

Then covariance matrices accurately reflect the covariance of estimations.

One generalization to a scalar-valued covariance for vector-valued random variables can be obtained by interpreting the deviation as the Euclidean distance:

$$\mathrm{var}_s\left(\mathbf{s}^t(k)\right) = \mathrm{E}\left[\|\mathbf{s}^t(k) - \mathrm{E}[\mathbf{s}^t(k)]\|_2^2\right]. \tag{32}$$

The expression can be rewritten as

$$\begin{aligned}
\mathrm{var}_s\left(\mathbf{s}^t(k)\right) &= \mathrm{E}\left[(\mathbf{s}^t(k) - \hat{\mathbf{s}}^t(k)) \cdot (\mathbf{s}^t(k) - \hat{\mathbf{s}}^t(k))\right] \\
&= \mathrm{E}\left[\sum\nolimits_{i=1}^{n} (s_i^t(k) - \hat{s}_i^t(k))^2\right] \\
&= \sum\nolimits_{i=1}^{n} \mathrm{E}\left[(s_i^t(k) - \hat{s}_i^t(k))^2\right] \\
&= \sum\nolimits_{i=1}^{n} \mathrm{var}(s_i^t(k)) \\
&= \sum\nolimits_{i=1}^{n} \hat{P}_{ii}^t(k),
\end{aligned} \tag{33}$$

where $s_i^t(k)$ and $\hat{s}_i^t(k)$ are the ith element of the target state $\mathbf{s}^t(k)$ and its mean $\hat{\mathbf{s}}^t(k)$. The $(i, j)th$ element of the covariance matrix $\hat{\mathbf{P}}^t(k)$ is $\hat{P}_{ii}^t(k)$. Finally, this can be simplified to

$$\mathrm{var}_s\left(\mathbf{s}^t(k)\right) = \mathrm{tr}(\hat{\mathbf{P}}^t(k)), \tag{34}$$

which is the trace of the covariance matrix.

In conclusion, the trace of the matrix is the Euclidean distance deviation of the state estimation, which can be a scalar-valued measure of the accuracy of the state estimation.

B Proof of Theorem 1

The proof is based on the "Max-Plus Algebra", defined in [43]. It is a powerful tool for the timed cyclic discrete-event systems and allows for a compact representation of weighted graphs.

The max-plus algebra consists of two binary operations, $\oplus$ and $\otimes$, on the set $\mathbb{R}_{\max} := \mathbb{R} \cup \{-\infty\}$. The operations are defined as follows:

$$\begin{aligned} a \oplus b &:= \max(a, b), \\ a \otimes b &:= a + b. \end{aligned} \tag{35}$$

The neutral element of the max-plus addition $\oplus$ is $-\infty$, denoted by ε. The neutral element of multiplication $\otimes$ is 0, denoted by e. The elements ε and e are also referred to as the zero and one element of the max-plus algebra. Similar to conventional algebra, the associativity, commutativity, and distributivity of multiplication over addition also hold for the max-plus algebra. Both operations can be extended to matrices in a straightforward way. For $A, B \in \mathbb{R}_{\max}^{m\times n}$,

$$(A \oplus B)_{ij} := a_{ij} \oplus b_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n. \tag{36}$$

For $A \in \mathbb{R}_{\max}^{m\times n}$, $B \in \mathbb{R}_{\max}^{n\times q}$,

$$(A \otimes B)_{ij} := \bigoplus_{k=1}^{n} (a_{ik} \otimes b_{kj}) = \max_{k}(a_{ik} + b_{kj}),\ i = 1, \ldots, m,\ j = 1, \ldots, q. \tag{37}$$

Multiplication of a matrix $A \in \mathbb{R}_{\max}^{m\times n}$ and a scalar $\alpha \in \mathbb{R}_{\max}$ is defined by

$$(\alpha \otimes B)_{ij} := \alpha \otimes a_{ij} = \alpha + a_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n. \tag{38}$$

Note that, as in conventional algebra, the multiplication symbol $\otimes$ is often omitted.

In the sequel, we also need matrices of zero elements, denoted by N, and of one elements, denoted by E. The identity matrix I is a square matrix with

$$(I)_{ij} := \begin{cases} \mathrm{e} & \text{for } i = j; \\ \varepsilon & \text{else.} \end{cases} \tag{39}$$

For any matrix $A \in \mathbb{R}_{\max}^{n\times n}$, its *precedence graph* $\mathcal{G}(A)$ is defined in the following way: it has n nodes, denoted by $1, \ldots, n$, and (j, i) is an edge if and only if $a_{ij} \neq \varepsilon$. In this case a_{ij} is the weight of edge (j, i). Then

- A path in $\mathcal{G}(A)$ is a sequence of $p > 1$ nodes, denoted by $\rho := i_1, \ldots, i_p$, such that $a_{i_{k+1}i_k} \neq \varepsilon, k = 1, \ldots, p-1$.
- $(A^k)_{ij}$ represents the maximal weight of all paths of length k from node j to node i, where

$$A^k := \underbrace{A \otimes A \otimes \ldots \otimes A}_{(k-1)-\text{times multiplication}},\ k \geq 1 \tag{40}$$

and $A^0 = I$.

On this basis we define $A_i \in \mathbb{R}_{\max}^{N_i^u \times N_i^u}$ as a matrix, representing the communication topology of the connected undirected graph $\mathcal{G}_i^u$, where the ith UAV is located. There exists a path of length d from node m to node n if and only if $(A_i^d)_{mn} = \mathrm{e}$, where $m, n = 1, 2, \ldots, N_i^u$. In addition, E_i is defined as a particular class of matrix with one element: $(E_i)_{mn} := \mathrm{e}$ for $m = i$ or $n = i$. The other elements of E_i can be at any value.

As mentioned above, the system is dynamic and the communication range is limited. Therefore, the communication graph of all UAVs may be split into multiple independent subgraphs and the topology is changing in real time. Theorem 1 essentially gives the minimum number of iterations to ensure that the ith UAV achieves the maximum in the subgraph $\mathcal{G}_i^u$. When each UAV in the subgraph has iterated for a corresponding number of times, the states in the entire subgraph achieve the max-consensus.

Now, it is ready to complete the proof.

Proof A necessary and sufficient condition for max-consensus held in node i is that

$$A_i^l = E_i, \exists l \in \mathbb{N}^0, \tag{41}$$

First, the sufficiency. Given a undirected graph $\mathcal{G}_i^u$ composed of N_i^u nodes, we define an initial vector of the perception confidence value $\gamma_m'^t(0) := [\gamma_{1m}'^t(0), \gamma_{2m}'^t(0), \ldots, \gamma_{N_i^u m}'^t(0)]^{\mathrm{T}}$, $m = 1, 2, \ldots, N^t$.

The Eq. (41) implies $\gamma_{im}'^t(l) = \left(E_i \otimes \gamma_m'^t(0)\right)_i$, where $(\cdot)_i$ is the ith element of the column vector. i.e.,

$$\begin{aligned} \gamma_{im}'^t(l) &= \bigotimes_{j=1,\ldots,N_i^u} (\gamma_{jm}'^t(0)) \\ &= \max\{\gamma_{1m}'^t(0), \ldots, \gamma_{N_i^u m}'^t(0)\}. \end{aligned} \tag{42}$$

Applying the rules for multiplying matrices in the Max-Plus Algebra, we obtain

$$A_i^{d+l} = E_i, \ \forall d \in \mathbb{N}^0, \tag{43}$$

and hence

$$\gamma_{im}'^t(d + l) = \max\{\gamma_{1m}'^t(0), \ldots, \gamma_{N_i^u m}'^t(0)\}. \tag{44}$$

Necessity is obvious.

If $A_i^l \neq E_i$, $\forall l$; then $\forall l$, $\exists k$ s.t. $(A_i^l)_{ik} = \varepsilon$, i.e., $\gamma_{im}'^t(l)$ dose not depend on $x_{km}(0)$. If $x_{km}(0)$ is the maximum element of $\{\gamma_{1m}'^t(0), \ldots, \gamma_{N_i^u m}'^t(0)\}$, the max-consensus will not hold.

In (41), $A_i^l = E_i$ implies that there exists a path of length l from the node i to any nodes in $\mathcal{G}_i^u$. Take l as the maximum of the length of all minimum path from i to each node on graph $\mathcal{G}_i^u$, which is the diameter of the shortest paths tree (SPT) of $\mathcal{G}_i^u$ rooted at node i [45].

Therefore, in order to establish the result of Theorem 1, it is essentially proved that if there exists an l, such that $A_i^l = E_i$ and $A_i^d \neq E_i$ for all $d < l$, then

$$l = D_i(\mathcal{G}_i^u), \tag{45}$$

where $D_i(\mathcal{G}_i^u)$ is the diameter of SPT of $\mathcal{G}_i^u$ rooted at node i.

Recall the condition. $A^l = E_i$ and $A^d \neq E_i$ means that there is a shortest path from node i to each node in $\mathcal{G}_i$ whose length is less or equal to l. As the maximum length of these paths is the diameter of SPT of $\mathcal{G}_i$ rooted at node i, Eq. (45) is true. □

C The derivation of the FIM

The FIM [47] at time k is defined by

$$\mathbf{G}_{ij}(k) = E\Big[\big(\nabla_{\mathbf{s}_i(k)} \ln p(\mathbf{z}_{ij}^P(k)\,|\mathbf{s}_i(k))\big)\big(\nabla_{\mathbf{s}_i(k)} \ln p(\mathbf{z}_{ij}^P(k)\,|\mathbf{s}_i(k))\big)^{\mathrm{T}}\Big], \tag{46}$$

where $p(\mathbf{z}_{ij}^P(k)\,|\mathbf{s}_i(k))$ is the batch measurement likelihood. For the sake of simplicity, we omit the iterative step k in the following derivation. In this tracking scenario, as the measurement accuracy is determined by the position relationship between the UAV and the target, in which the covariance matrix is $\mathbf{C}_{ij}$ shown in (2), the FIM is only related to the position of the UAV and that of the target. Since the target is non-cooperative, it is able to change the FIM only by adjusting the position of the UAV. Therefore, it can be restated as:

$$\mathbf{G}_{ij} = E\Big[\big(\nabla_{\mathbf{p}_i^u} \ln p(\mathbf{z}_{ij}^P \,|\mathbf{p}_i^u, \mathbf{p}_j^t)\big)\big(\nabla_{\mathbf{p}_i^u} \ln p(\mathbf{z}_{ij}^P \,|\mathbf{p}_i^u, \mathbf{p}_j^t)\big)^{\mathrm{T}}\Big]. \tag{47}$$

The batch measurement likelihood in (47) is defined as (48):

$$p\left(\mathbf{z}_{ij}^P \,\middle|\mathbf{p}_i^u, \mathbf{p}_j^t\right) = \frac{1}{\sqrt{2\pi\,|\mathbf{C}_{ij}|}} \exp\Big(-\frac{1}{2}\big(\mathbf{z}_{ij}^P - \boldsymbol{\ell}(\mathbf{p}_i^u, \mathbf{p}_j^t)\big)^{\mathrm{T}}\mathbf{C}_{ij}^{-1}\big(\mathbf{z}_{ij}^P - \boldsymbol{\ell}(\mathbf{p}_i^u, \mathbf{p}_j^t)\big)\Big). \tag{48}$$

In the above formula, $\boldsymbol{\ell}(\mathbf{p}_i^u, \mathbf{p}_j^t)$ denotes the real value of range-bearing, which is given by

$$\begin{aligned}\boldsymbol{\ell}(\mathbf{p}_i^u, \mathbf{p}_j^t) &= \begin{bmatrix} d_{ij}(\mathbf{p}_i^u, \mathbf{p}_j^t) \\ \theta_{ij}(\mathbf{p}_i^u, \mathbf{p}_j^t) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{(x_j^t - x_i^u)^2 + (y_j^t - y_i^u)^2 + (h_i^u)^2} \\ \arctan\left(\frac{y_j^t - y_i^u}{x_j^t - x_i^u}\right) \end{bmatrix}.\end{aligned} \tag{49}$$

The first order derivative of the log-density function is given by

$$\frac{\partial \ln p(\mathbf{z}_{ij}^P \,|\mathbf{p}_i^u, \mathbf{p}_j^t)}{\partial \mathbf{p}_i^u} = -\frac{1}{2}\left[\frac{\partial \ln |\mathbf{C}_{ij}|}{\partial \mathbf{p}_i^u} + \frac{\partial}{\partial \mathbf{p}_i^u}\Big(\big(\mathbf{z}_{ij}^P - \boldsymbol{\ell}\big)^{\mathrm{T}}\mathbf{C}_{ij}^{-1}\big(\mathbf{z}_{ij}^P - \boldsymbol{\ell}\big)\Big)\right]. \tag{50}$$

Then the Fisher information matrix can be written as

$$\big[\mathbf{G}_{ij}\big]_{mn} = \left[\frac{\partial \boldsymbol{\ell}}{\partial \mathbf{p}_i^u(m)}\right]^t \mathbf{C}_{ij}^{-1}\left[\frac{\partial \boldsymbol{\ell}}{\partial \mathbf{p}_i^u(n)}\right] + \frac{1}{2} Trace\left(\mathbf{C}_{ij}^{-1}\frac{\partial \mathbf{C}_{ij}}{\partial \mathbf{p}_i^u(m)}\mathbf{C}_{ij}^{-1}\frac{\partial \mathbf{C}_{ij}}{\partial \mathbf{p}_i^u(n)}\right). \tag{51}$$

where $\mathbf{G}_{ij}$ is a square matrix of order 2; $m, n \in \{1, 2\}$ represents the row number and column number of each element in $\mathbf{G}_{ij}$; $\mathbf{p}_i^u(1) = x_i^u$, and $\mathbf{p}_i^u(2) = y_i^u$. The specific form of each element is as following:

$$[\mathbf{G}_{ij}]_{11} = \frac{\left(x_j^t - x_i^u\right)^2}{d_{ij}^2\sigma_r^2} + \frac{\left(y_j^t - y_i^u\right)^2}{r_{ij}^2\sigma_\theta^2} + \frac{2k_d^2k_r^2\exp\left(2k_r(\frac{d_{ij}}{d_i^O} - 1)\right)\left(x_j^t - x_i^u\right)^2}{d_{ij}^2(d_i^O)^2\sigma_r^2} + \frac{2k_\theta^2\left(x_j^t - x_i^u\right)^2}{d_{ij}^2(d_i^O)^2\sigma_\theta^2}, \tag{52}$$

$$[\mathbf{G}_{ij}]_{12} = [\mathbf{G}_{ij}]_{21} = \frac{\left(x_j^t - x_i^u\right)\left(y_j^t - y_i^u\right)}{d_{ij}^2\sigma_r^2} - \frac{\left(y_j^t - y_i^u\right)\left(y_j^t - y_i^u\right)}{r_{ij}^2\sigma_\theta^2} + \frac{2k_d^2k_r^2\exp\left(2k_r(\frac{d_{ij}}{d_i^O} - 1)\right)\left(x_j^t - x_i^u\right)\left(y_j^t - y_i^u\right)}{d_{ij}^2(d_i^O)^2\sigma_r^2} + \frac{2k_\theta^2\left(x_j^t - x_i^u\right)\left(y_j^t - y_i^u\right)}{d_{ij}^2(d_i^O)^2\sigma_\theta^2}, \tag{53}$$

$$[\mathbf{G}_{ij}]_{11} = \frac{\left(y_j^t - y_i^u\right)^2}{d_{ij}^2\sigma_r^2} + \frac{\left(x_j^t - x_i^u\right)^2}{r_{ij}^2\sigma_\theta^2} + \frac{2k_d^2k_r^2\exp\left(2k_r(\frac{d_{ij}}{d_i^O} - 1)\right)\left(y_j^t - y_i^u\right)^2}{d_{ij}^2(d_i^O)^2\sigma_r^2} + \frac{2k_\theta^2\left(y_j^t - y_i^u\right)^2}{d_{ij}^2(d_i^O)^2\sigma_\theta^2}. \tag{54}$$

References

1. Kassas, Z. M., & Özgüner, Ü. (2010). A nonlinear filter coupled with hospitability and synthetic inclination maps for in-surveillance and out-of-surveillance tracking. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(1), 87–97.
2. Cesare, K., Skeele, R., Yoo, S. H., Zhang, Y., & Hollinger, G. (2015). Multi-UAV exploration with limited communication and battery. In *2015 IEEE international conference on robotics and automation (ICRA)* (pp. 2230–2235). IEEE.
3. Darrah, M., Wilhelm, J., Munasinghe, T., Duling, K., Yokum, S., Sorton, E., et al. (2015). A flexible genetic algorithm system for multi-UAV surveillance: Algorithm and flight testing. *Unmanned Systems*, *3*(1), 49–62.
4. Capitán, J., Merino, L., & Ollero, A. (2016). Cooperative decision-making under uncertainties for multi-target surveillance with multiples UAVs. *Journal of Intelligent & Robotic Systems*, *84*(1–4), 371–386. Springer.
5. Caraballo, L., Acevedo, J., Díaz-Báñez, J., Arrue, B., Maza, I., & Ollero, A. (2014). The block-sharing strategy for area monitoring missions using a decentralized multi-UAV system. In *2014 International conference on unmanned aircraft systems (ICUAS)* (pp. 602–610). IEEE.
6. Mersheeva, V., & Friedrich, G. (2015). Multi-UAV monitoring with priorities and limited energy resources. In *Proceedings of the 25th international conference on automated planning and scheduling* (pp. 347–356).
7. Qi, J., Song, D., Shang, H., Wang, N., Hua, C., Wu, C., et al. (2016). Search and rescue rotary-wing UAV and its application to the lushan ms 7.0 earthquake. *Journal of Field Robotics*, *33*(3), 290–321. Wiley Online Library.
8. Adamey, E., & Ozguner, U. (2011). Cooperative multitarget tracking and surveillance with mobile sensing agents: A decentralized approach. In *2011 14th International IEEE conference on intelligent transportation systems (ITSC)* (pp. 1916–1922). IEEE.
9. Dames, P., Tokekar, P., & Kumar, V. (2017). Detecting, localizing, and tracking an unknown number of moving targets using a team of mobile robots. *The International Journal of Robotics Research*, *36*(13–14), 1540–1553. SAGE.

10. Tokekar, P., Isler, V., & Franchi, A. (2014). Multi-target visual tracking with aerial robots. In *2014 IEEE/RSJ international conference on intelligent robots and systems (IROS 2014)* (pp. 3067–3072). IEEE.
11. Bayram, H., Stefas, N., Engin, K. S., & Isler, V. (2017). Tracking wildlife with multiple UAVs: System design, safety and field experiments. In *2017 International symposium on multi-robot and multi-agent systems (MRS)* (pp. 97–103). IEEE.
12. Hausman, K., Müller, J., Hariharan, A., Ayanian, N., & Sukhatme, G. S. (2016). Cooperative control for target tracking with onboard sensing. In M. A. Hsieh, O. Khatib, & V. Kumar (Eds.), *Experimental robotics* (pp. 879–892). Berlin: Springer.
13. Schlotfeldt, B., Thakur, D., Atanasov, N., Kumar, V., & Pappas, G. J. (2018). Anytime planning for decentralized multirobot active information gathering. *IEEE Robotics and Automation Letters*, *3*(2), 1025–1032. IEEE.
14. Nestmeyer, T., Giordano, P. R., Bülthoff, H. H., & Franchi, A. (2017). Decentralized simultaneous multi-target exploration using a connected network of multiple robots. *Autonomous Robots*, *41*(4), 989–1011. Springer.
15. Mahmoud, M. S., & Khalid, H. M. (2013). Distributed Kalman filtering: A bibliographic review. *IET Control Theory & Applications*, *7*(4), 483–501. IET.
16. Capitan, J., Spaan, M. T., Merino, L., & Ollero, A. (2013). Decentralized multi-robot cooperation with auctioned POMDPs. *The International Journal of Robotics Research*, *32*(6), 650–671. SAGE.
17. Capitán, J., Merino, L., Caballero, F., & Ollero, A. (2011). Decentralized delayed-state information filter (DDSIF): A new approach for cooperative decentralized tracking. *Robotics and Autonomous Systems*, *59*(6), 376–388. Elsevier.
18. Adamey, E., & Ozguner, U. (2012). A decentralized approach for multi-UAV multitarget tracking and surveillance. In *SPIE Defense, Security, and Sensing* (Vol. 8389, pp. 838915-1–838915-6). International Society for Optics and Photonics.
19. Chagas, R. A. J., & Waldmann, J. (2015). A novel linear, unbiased estimator to fuse delayed measurements in distributed sensor networks with application to UAV fleet. In D. Choukroun, Y. Oshman, J. Thienel, & M. Idan (Eds.), *Advances in estimation, navigation, and spacecraft control* (pp. 135–157). Berlin: Springer.
20. Fanti, M. P., Mangini, A. M., & Ukovich, W. (2012). A quantized consensus algorithm for distributed task assignment. In *2012 IEEE 51st annual conference on decision and control (CDC)* (pp. 2040–2045). IEEE.
21. Luo, L., Chakraborty, N., & Sycara, K. (2015). Distributed algorithms for multirobot task assignment with task deadline constraints. *IEEE Transactions on Automation Science and Engineering*, *12*(3), 876–888. IEEE.
22. Peng, Z., Yang, S., Wen, G., & Rahmani, A. (2014). Distributed consensus-based robust adaptive formation control for nonholonomic mobile robots with partial known dynamics. In *Mathematical Problems in Engineering*. https://doi.org/10.1155/2014/670497.
23. Seo, J., Kim, Y., Kim, S., & Tsourdos, A. (2012). Consensus-based reconfigurable controller design for unmanned aerial vehicle formation flight. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *226*(7), 817–829. SAGE.
24. Maggs, M. K., O'Keefe, S. G., & Thiel, D. V. (2012). Consensus clock synchronization for wireless sensor networks. *IEEE Sensors Journal*, *12*(6), 2269–2277.
25. Battistelli, G., & Chisci, L. (2014). Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, *50*(3), 707–718. Elsevier.
26. Palacios-Gasós, J. M., Montijano, E., Sagüés, C., & Llorente, S. (2016). Distributed coverage estimation and control for multirobot persistent tasks. *IEEE Transactions on Robotics*, *32*(6), 1444–1460. IEEE.
27. Di Paola, D., Petitti, A., & Rizzo, A. (2015). Distributed Kalman filtering via node selection in heterogeneous sensor networks. *International Journal of Systems Science*, *46*(14), 2572–2583. Taylor & Francis.
28. Smallwood, R. D., & Sondik, E. J. (1973). The optimal control of partially observable markov processes over a finite horizon. *Operations Research*, *21*(5), 1071–1088. INFORMS.
29. Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1), 99–134. Elsevier.
30. Ponda, S. S., Johnson, L. B., Geramifard, A., & How, J. P. (2015). Cooperative mission planning for multi-UAV teams. In K. P. Valavanis & G. J. Vachtsevanos (Eds.), *Handbook of unmanned aerial vehicles* (pp. 1447–1490). Berlin: Springer.
31. Oliehoek, F. A. (2012). Decentralized POMDPs. In *Reinforcement Learning* (Vol. 12, pp. 471–503). Springer.

32. Wu, F., Zilberstein, S., & Chen, X. (2011). Online planning for multi-agent systems with bounded communication. *Artificial Intelligence*, *175*(2), 487–511. Elsevier.
33. Vaisenberg, R., Della Motta, A., Mehrotra, S., & Ramanan, D. (2014). Scheduling sensors for monitoring sentient spaces using an approximate POMDP policy. *Pervasive and Mobile Computing*, *10*, 83–103. Elsevier.
34. Panella, A., & Gmytrasiewicz, P. (2017). Interactive POMDPs with finite-state models of other agents. *Autonomous Agents and Multi-Agent Systems*, *31*(4), 861–904.
35. Yu, H., Meier, K., Argyle, M., & Beard, R. W. (2015). Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles. *IEEE/ASME Transactions on Mechatronics*, *20*(2), 541–552. IEEE.
36. Farmani, N., Sun, L., & Pack, D. (2015). Tracking multiple mobile targets using cooperative unmanned aerial vehicles. In *2015 International conference on unmanned aircraft systems (ICUAS)* (pp. 395–400). IEEE.
37. Zhang, K., Collins, E. G, Jr., & Shi, D. (2012). Centralized and distributed task allocation in multi-robot teams via a stochastic clustering auction. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, *7*(2), 21.
38. Edalat, N., Tham, C. K., & Xiao, W. (2012). An auction-based strategy for distributed task allocation in wireless sensor networks. *Computer Communications*, *35*(8), 916–928. Elsevier.
39. Spaan, M. T., Veiga, T. S., & Lima, P. U. (2015). Decision-theoretic planning under uncertainty with information rewards for active cooperative perception. *Autonomous Agents and Multi-Agent Systems*, *29*(6), 1157–1185. Springer.
40. Zhao, Y., Wang, X., Cong, Y., & Shen, L. (2018). Information geometry based action decision-making for target tracking by fixed-wing UAV: From algorithm design to theory analysis. *International Journal of Advanced Robotic Systems*, *15*(4), 1729881418787061.
41. Ragi, S., & Chong, E. K. (2013). Uav path planning in a dynamic environment via partially observable Markov decision process. *IEEE Transactions on Aerospace and Electronic Systems*, *49*(4), 2397–2412. IEEE.
42. Burguera, A., González, Y., & Oliver, G. (2009). Sonar sensor models and their application to mobile robot localization. *Sensors*, *9*(12), 10217–10243. Molecular Diversity Preservation International.
43. Nejad, B. M., Attia, S. A., & Raisch, J. (2009). Max-consensus in a max-plus algebraic setting: The case of fixed communication topologies. In *XXII international symposium on information, communication and automation technologies, 2009 (ICAT 2009)* (pp. 1–7). IEEE.
44. Petitti, A., Di Paola, D., Rizzo, A., & Cicirelli, G. (2011). Consensus-based distributed estimation for target tracking in heterogeneous sensor networks. In *2011 50th IEEE conference on decision and control and European control conference (CDC-ECC)* (pp. 6648–6653). IEEE.
45. Bui, M., Butelle, F., & Lavault, C. (2004). A distributed algorithm for constructing a minimum diameter spanning tree. *Journal of Parallel and Distributed Computing*, *64*(5), 571–577. Elsevier.
46. Burkard, R. E. (2002). Selected topics on assignment problems. *Discrete Applied Mathematics*, *123*(1), 257–302. Elsevier.
47. Rao, C. R. (1992). Information and the accuracy attainable in the estimation of statistical parameters. In N. L. Johnson, A. W. Kemp, & S. Kotz (Eds.), *Breakthroughs in statistics* (pp. 235–247). Berlin: Springer.