Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A collaborative target tracking algorithm for multiple UAVs with inferior tracking capabilities*

Zhi ZHENG†‡, Shuncheng CAI

College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117, China

†E-mail: zhengz@fjnu.edu.cn

Received July 20, 2020; Revision accepted Mar. 31, 2021; Crosschecked Aug. 16, 2021

Abstract: Target tracking is one of the hottest topics in the field of drone research. In this paper, we study the multiple unmanned aerial vehicles (multi-UAV) collaborative target tracking problem. We propose a novel tracking method based on intention estimation and effective cooperation for UAVs with inferior tracking capabilities to track the targets that may have agile, uncertain, and intelligent motion. For three classic target motion modes, we first design a novel trajectory feature extraction method with the least dimension and maximum coverage constraints, and propose an intention estimation mechanism based on the environment and target trajectory features. We propose a novel Voronoi diagram, called MDA-Voronoi, which divides the area with obstacles according to the minimum reachable distance and the minimum steering angle of each UAV. In each MDA-Voronoi region, the maximum reachable region of each UAV is defined, the upper and lower bounds of the trajectory coverage probability are analyzed, and the tracking strategies of the UAVs are designed to effectively reduce the tracking gaps to improve the target sensing time. Then, we use the Nash Q-learning method to design the UAVs' collaborative tracking strategy, considering factors such as collision avoidance, maneuvering constraints, tracking cost, sensing performance, and path overlap. By designing the reward mechanism, the optimal action strategies are obtained as the control input of the UAVs. Finally, simulation analyses are provided to validate our method, and the results demonstrate that the algorithm can improve the collaborative target tracking performance for multiple UAVs with inferior tracking capabilities.

Key words: Collaborative target tracking; Intent estimation; MDA-Voronoi diagram; Multi-UAV; Inferior tracking capability

https://doi.org/10.1631/FITEE.2000362 **CLC number:** TP391

1 Introduction

Unmanned aerial vehicles (UAVs) are one of the main types of unmanned systems, and in recent years they have been widely used in military and civilian fields, such as data collection, environmental monitoring, military strikes, and urban counter-terrorism. Effective tracking of agile targets in a complex and uncertain environment is one of the hottest and most difficult problems in the field of drone research. Because a single UAV is restricted by structure, load capability, and motor constraints, it cannot undertake complex tasks. Multi-UAV coordination can overcome these shortcomings to a great extent. Therefore, multi-UAV collaborative target tracking and motion planning have received much attention (Ruan and Duan, 2020; Shao et al., 2020; Skorobogatov et al., 2020).

Many scholars have made outstanding contributions to UAV target tracking. A new tracker called the "tracking learning detection and kernelized correlation filter algorithm" was presented by Liu

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (No. 61873033), the Science Foundation of Fujian Normal University (No. Z0210553), and the Natural Science Foundation of Fujian Province, China (No. 2020H0012)

ORCID: Zhi ZHENG, https://orcid.org/0000-0002-9455-0059; Shuncheng CAI, https://orcid.org/0000-0002-5750-7326

et al. (2019), in which a conditional scale-adaptive algorithm was adopted to improve the tracking performance of a quadrotor UAV in cluttered outdoor environments. An efficient, complete, and off-line algorithm, named the "auction-based spanning tree coverage" algorithm (Gao and Xin, 2019), was proposed to deal with the multi-robot coverage motion planning problem. A path planning algorithm (Ragi and Chong, 2012) was designed to guide UAVs in tracking multiple ground targets based on the theory of partially observable Markov decision processes. A three-dimensional (3D) real-time path planning method (Yao et al., 2015) was proposed which combines the improved Lyapunov guidance vector field, the interfered fluid dynamical system, and the strategy of varying receding-horizon optimization from the model predictive control. A decentralized planning algorithm that relies on an auction scheme (Yu et al., 2015) was designed to plan finite look-ahead paths. This algorithm maximizes the sum of the joint probability of sensing over all vehicles' tracking. The problem of communication-aware UAV placement and motion planning for target localization and tracking (Di et al., 2016) was investigated. A nominal belief-state optimization method (Ragi and Chong, 2013) was proposed to track multiple targets with an average target location error and average communication cost. A tracking strategy (Meng et al., 2017) concerned with integrated autonomous takeoff, target search, task assignment, and tracking using multiple fixed-wing UAVs in urban environments was proposed. An online autonomous UAV path planning method (Jiang and Liang, 2018) was proposed for bearing-only standoff multi-target following in a threat environment. A multi-UAV cooperative path planning method (Wang DB et al., 2015) was proposed for ground target tracking via chemical reaction optimization. Zollars et al. (2019) proposed an optimal path planning algorithm to reduce computation time and increase solution accuracy for small UAVs in 3D constrained environments. Some approaches involving obstacle velocity (Douthwaite et al., 2019) have been presented for multi-agent collision avoidance in communication-denied environments.

In recent years, reinforcement learning has also been successfully applied in target tracking and monitoring. Cai et al. (2012) proposed a multi-sensor cooperative target tracking method based on distributed Nash Q-learning, in which a probability statistics method based on Bayesian inference was presented to update the Q function. Wang T et al. (2019) proposed an online distributed algorithm for tracking and searching, while considering energy refueling. Adepegba et al. (2016) proposed an area coverage control law in cooperation with reinforcement learning techniques for deploying multiple autonomous agents in a two-dimensional (2D) planar area. Vanegas et al. (2016) described target finding and tracking in Global Positioning System (GPS) denied and cluttered environments for UAVs. Goldhoorn et al. (2018) presented a unified method for searching and tracking a person using a group of mobile robots in a large continuous urban environment with dynamic obstacles.

Most of the aforementioned studies have some common features as follows: the UAVs have wide enough sensing ranges, the target motion is simple, and the UAVs move faster than the targets. In addition, there is less emphasis on target intelligence and the trade-off between sensing performance and the number of UAVs. However, in practice, the UAV sensing ranges are limited and the targets may have a certain intelligence. In addition, the number of UAVs is limited, and targets may move faster than the UAVs. There are two main objectives for tracking: one is to maximize the sensing performance, and the other is to minimize the tracking loss. In some research (Quintero et al., 2015), the tracking strategy is to make the UAVs move along the target's trajectory, constantly getting closer to achieve optimal tracking; this is a passive tracking strategy. If the target moves faster than the tracker or the target motion trajectory has a certain decoy, passive tracking is difficult to execute well, and it is easy to lose the target or take more detours. Thus, the active tracking method based on effective prediction and cooperative preparation in advance has better performance. As shown in Fig. 1, a target moves along the black solid line from left to right, and the positions of the target at three different time points are marked in turn from left to right. There are two UAVs in different locations, named UAV1 and UAV2. In Fig. 1a, UAV1 follows the target trajectory and UAV2 flies directly to the target; they are trying to get closer to the target. Because of the lower speed, both of them may have no chance of keeping the target in their sensing range in this way, while in

Fig. 1b, by analyzing the trajectory characteristics of the target and predicting the trend of the trajectory, the two UAVs can compensate for their inferior tracking capabilities and greatly improve the sensing time for the target through cooperation. How UAVs achieve effective tracking through cooperation without superior tracking capabilities is a difficult problem.

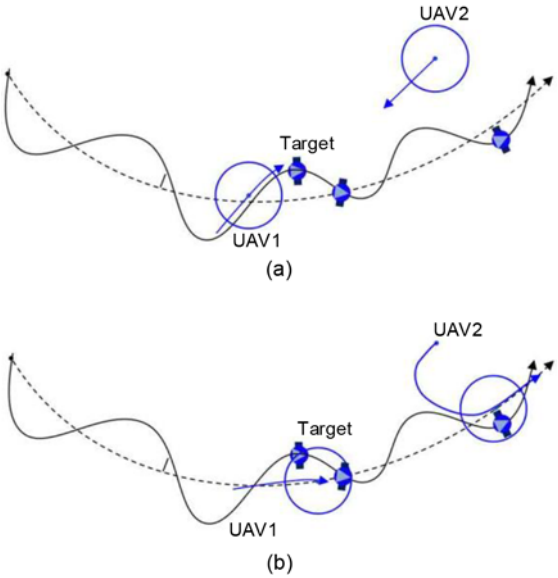

Fig. 1 Two tracking mechanisms: (a) passive tracking, in which the UAVs follow the actual trajectory of the target or fly directly to the target; (b) active tracking, in which the UAVs track based on the characteristics of the target's trajectory

The black solid curves denote the true motion trajectories of the target. The black dashed curve is the tracking route extracted from the target's motion features, and is the route that the UAVs expect to follow. The blue curves are the UAV trajectories and the circles are their sensing range. References to color refer to the online version of this figure

In this study, we study the multi-UAV collaborative target tracking problem, and propose a novel tracking method based on intent estimation and effective cooperation in an environment with obstacles, where the moving targets have agile, uncertain, and decoy motions. The UAVs' sensing ranges are limited and the maximum tracking speed is lower than those of the targets. First, the motion models of the UAV and the target, and the sensor sensing model, are established. Then, within the observation time window, a novel trajectory feature extraction method with least dimension and maximum coverage constraints is designed, and the changes of the target's velocity and the obstacle distribution around the target are obtained according to the trunk trajectory. The intention of the target is estimated using the Sigmoid function, the target's intended trajectory change within the next time window is predicted, and based on this information, we adjust the prediction time window and sampling number for the observation of unexpected situations. Then we propose a novel Voronoi diagram, called MDA-Voronoi, which divides the area with obstacles according to the minimum reachable distance and the minimum steering angle of each UAV. In each MDA-Voronoi region, the maximum reachable region of each UAV is defined and used to discover the extreme coverage holes (ECOs). Under the premise of the optimal tracking distance and angle, the target's escape probability related to the number, distribution, speed, and sensing radius of the UAVs is analyzed. Furthermore, the upper and lower bounds of the trajectory coverage probability are proved, and the UAV tracking strategies are designed to effectively reduce the tracking gaps and thus increase the sensing time for the targets. Using Nash Q-learning, the learning strategies are used as the control inputs of the UAVs, and the multiple targets, collision avoidance, tracking contribution, sensing performance, and path overlap are taken into account. We also design a reward mechanism to track multiple targets with the expected sensing performance and the minimum number of UAVs. Finally, simulation results demonstrate the effectiveness of our strategies.

2 Mathematical models

2.1 UAV motion model

Define $\mathbb{R}^2$ as a 2D mission area, and $\boldsymbol{q} \in \mathbb{R}^2$ a point in the mission area. Assume that there are N UAVs, which are flying at a fixed speed V and a fixed altitude, controlled only by the direction angle ψ. The i^{th} UAV is denoted by $\boldsymbol{U}_i$, $i \in \{1, 2, \ldots, N\}$, and the simplified motion model of $\boldsymbol{U}_i$ is

$$\begin{cases} x_i(k) = x_i(k-1) + V\cos(\psi_i(k))\Delta t, \\ y_i(k) = y_i(k-1) + V\sin(\psi_i(k))\Delta t, \\ \psi_i(k) = \psi_i(k-1) + \Delta\psi_i, \end{cases} \tag{1}$$

where $[x_i(k), y_i(k)]$ are the coordinates $\boldsymbol{P}_i(k)$ of U_i at time k, $i \in \{1, 2, \ldots, N\}$. Δt is the sampling time. $\psi_i(k)$ is the direction angle of U_i at time k, and $\Delta\psi_i$ is the change in the direction angle. Due to motorized

constraints, we assume that $\Delta\psi_i \in [\Delta\psi_{\min}, \Delta\psi_{\max}]$, where $\Delta\psi_{\min}$ and $\Delta\psi_{\max}$ are the lower and upper bounds of $\Delta\psi_i$, respectively.

2.2 UAV sensing model

Assume that the ground station can observe the targets and broadcast the estimated information to the UAV group. The UAV's tracking keeps targets within the sensing range of itself or others for certain sensing requirements, such as monitoring, gathering specific information, and sensing can be viewed as a sensor coverage problem. Suppose that the on-board sensor is located at the UAV's center of mass, and that the detection center shaft is perpendicular to the plane where the body is located and points to the ground. Therefore, when the UAV is flying at a fixed altitude, the sensing range of the on-board sensor on the ground can be approximated as a circle with radius r and the center at the position of the UAV. Without loss of generality, we simply assume that the sensing performance model of the on-board sensor is a sensing probability function that is a measure of the UAV's sensing system. The sensing performance, also called the coverage performance, is closely related to the distance between the sensor and the target to be covered. The longer the distance, the lower the sensing performance. However, it may be blocked by obstacles; i.e., behind the obstacle, there is no sensing. Define the sensing performance of U_i to point $\boldsymbol{q}$ at time k as

$$D_{iq}(k) = \begin{cases} \mathrm{e}^{-\|\boldsymbol{P}_i(k)-\boldsymbol{q}\|}, & \|\boldsymbol{P}_i(k)-\boldsymbol{q}\| \in [0, r], \\ 0, & \|\boldsymbol{P}_i(k)-\boldsymbol{q}\| > r, \end{cases} \tag{2}$$

where r is the detection range of the sensor. Eq. (2) indicates that the sensor's sensing performance decreases as the distance from the target to the UAV increases, and there is no sensing out of the detection range. The sensing performance can describe the sensing quality in specific applications.

2.3 Target motion and observation models

Suppose that there are M agile targets and $\mathcal{Z}$ ground stations that can observe the targets. The motion model of target i ($i \in \{1, 2, \ldots, M\}$) is

$$\begin{cases} (\boldsymbol{X}_i^{\mathrm{t}}(k))^{\mathrm{T}} = \boldsymbol{\Gamma}(\boldsymbol{X}_i^{\mathrm{t}}(k-1))^{\mathrm{T}} + \boldsymbol{\omega}_X(k-1), \\ \boldsymbol{\mathcal{Z}}_{ij}(k) = \boldsymbol{H}_{ij}(\boldsymbol{X}_i^{\mathrm{t}}(k))^{\mathrm{T}} + \boldsymbol{\omega}_z(k), \end{cases} \tag{3}$$

where $\boldsymbol{X}_i^{\mathrm{t}}(k) = [x_i^{\mathrm{t}}(k), y_i^{\mathrm{t}}(k), \dot{x}_i^{\mathrm{t}}(k), \dot{y}_i^{\mathrm{t}}(k)]$ is the state vector of target i at time k, $\boldsymbol{p}_i^{\mathrm{t}}(k) = [x_i^{\mathrm{t}}(k), y_i^{\mathrm{t}}(k)]$ is the coordinate vector of target i at time k, and $\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is the state transfer matrix. $\boldsymbol{\mathcal{Z}}_{ij}(k)$ is the observation of target i by ground station j ($j \in \{1, 2, \ldots, \mathcal{Z}\}$) at time k. $\boldsymbol{H}_{ij} = \mathrm{diag}(1, 1, 1, 1)$ is the observation coefficient matrix. $\boldsymbol{\omega}_X(k-1)$ and $\boldsymbol{\omega}_z(k)$ are uncorrelated Gaussian noises.

We use $\hat{\boldsymbol{X}}_i(k) = \frac{1}{\mathcal{Z}} \sum_{j=1}^{\mathcal{Z}} \hat{\boldsymbol{X}}_{ij}(k)$ ($i = 1, 2, \ldots, M$) to denote the estimated fusion for target i at time k, where $\hat{\boldsymbol{X}}_{ij}(k)$ is the state estimate of target i by ground station j ($j = 1, 2, \ldots, \mathcal{Z}$), and we use Kalman filtering to estimate the state vector $\hat{\boldsymbol{X}}_{ij}(k)$ by the observation $\boldsymbol{\mathcal{Z}}_{ij}(k)$.

Assuming that the target has agile, uncertain, and intelligent motion, we construct the variable velocity model of target i ($i \in \{1, 2, \ldots, M\}$) at time k as

$$v_i(k) = V + \Re(\underline{r}, \bar{r}) \exp\left(\frac{-\|\boldsymbol{P}_j(k) - \boldsymbol{p}_i^{\mathrm{t}}(k)\|}{\varpi(k)}\right), \tag{4}$$

where $\Re(\underline{r}, \bar{r})$ is a random number in $[\underline{r}, \bar{r}]$, and $\|\boldsymbol{P}_j(k) - \boldsymbol{p}_i^{\mathrm{t}}(k)\|$ is the distance between U_j ($j \in \{1, 2, \ldots, N\}$) and target i. $\varpi(k)$ represents the extent to which the environment around the target is favorable for its cover at time k, and we use the distance between target i and the nearest obstacle to define it as

$$\varpi(k) = \frac{1}{\|\boldsymbol{p}_i^{\mathrm{t}}(k) - \boldsymbol{p}^{\mathrm{o}}(k)\|}, \tag{5}$$

where $\boldsymbol{p}^{\mathrm{o}}(k)$ are the coordinates of the obstacle that is nearest to the target. Without loss of generality, the closer the target is to the UAV, and the more conducive the environment is to being used as cover, the faster the target will move. The velocity of the target is always greater than those of the UAVs, and it reflects the intelligence and uncertainty of the target. We assume that Eq. (4) is unknown to the UAVs, but the target can observe the UAVs' coordinates and adjust its velocity according to the UAVs and its intention.

2.4 UAV collision constraint

In the mission environment, there are obstacles and multiple UAVs. For safety, each UAV is required

to maintain a minimum safe distance $d_{\min}$ from obstacles and other UAVs to avoid collisions.

3 Intention estimation and tracking strategies based on the MDA-Voronoi diagram

Due to the uncertainty of the target movement, prediction is effective in improving tracking performance. There are some excellent tracking strategies based on trajectory prediction, such as receding-horizon control (Wang L et al., 2011) and extended Kalman filtering (Bordonaro et al., 2019; Song et al., 2019; Khalkhali et al., 2020). These methods make predictions mainly by observing the historical trajectory, with less consideration of the environment and the target's intentions. For decoy behaviors of intelligent targets, there may be large deviations in predictions. Although the trajectories of intelligent targets are complex and varied, they contain intent information. If the intention can be correctly identified, it can improve prediction, and make an effective decision in advance to improve tracking. As shown in Fig. 2, although the target appears to move irregularly, it can be seen from the trajectory feature that the target clearly intends to approach the destination. By analyzing the target trajectory within the observation time window, UAVs can extract the main characteristics of the trajectory, combine them with the environmental information, obtain the short-term intention of the target by a probability function, and estimate the target's trajectory features in the next time window.

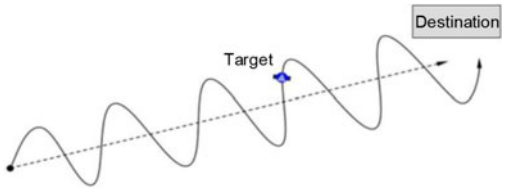

Fig. 2 Target movement with intent

The black solid curve denotes the true motion trajectory of the target, and the black dashed line reflects the growth direction of the trajectory. The black dot and the arrow in the solid curve denote the starting point and the motion direction of the target, respectively

3.1 Trajectory feature extraction

Definition 1 The trunk trajectory refers to a curve that can reflect the overall change trend, curvature, and other major characteristics of the trajectory.

Ignoring some unimportant details properly and retaining the trunk trajectory, which reflects the main trajectory features within a certain time window, is conducive to subsequent analysis of the trajectory feature and intent prediction.

Fitting approaches work well in tracking targets that move in smooth courses, such as passenger aircraft and ships (Li, 2019; Li et al., 2017, 2019). Because the velocity of the target is larger than those of the UAVs, we design a novel trajectory feature extraction method with least dimension and maximum coverage constraints (LCE), which not only reflects the trajectory features, but also enables the UAVs to have the maximum target coverage in less time. In our work, the trajectory within the observation time window is discretized into h sampling points at a fixed time interval, and we have the discrete trajectory sequence $\boldsymbol{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_h]$, where $\boldsymbol{s}_i$ $(i\in\{1,2,\ldots,h\})$ are the coordinates of the sampling point i. The trunk trajectory, denoted by $\boldsymbol{T}$, is composed of two parts: one is the trajectory obtained according to the sampling points within the observation time window by the LCE method, and the other is the prediction part, which is the extension of the trajectory extracted within the observation time window and according to the velocity of the target within the prediction time window. To retain most of the features of the sampling points and enable the UAVs to maximize the coverage of the sampling points with the least time cost when moving along the trunk trajectory, we use the least dimension to represent the trajectory characteristics of the target and maximize the variance of the projection of the sampling points onto the trunk trajectory to reduce the loss of trajectory characteristics. Furthermore, if there are obstacles in the prediction time window, the obstacles should be taken into account in the prediction. Taking the intersection of the predicted trajectory and the obstacle as the starting point, the target may have a variety of possible path options when encountering obstacles, and these trajectories have different angles, denoted by θ_i $(i=1,2,\ldots,k)$, with the surface of the obstacle.

The initial probabilities that these trajectories are chosen can be given by the angles. The smaller the angle, the greater the probability. Then the probability of path i is dynamically adjusted by

$$\rho_i=\frac{|\theta_i|}{\complement(\theta)}, \tag{6}$$

where $|\theta_i|$ represents how many times path i has been chosen by the target, and $\complement(\theta)$ is the total prediction time so far. This reflects the historical experience of the target's movement and the way by which the target is most likely to move when encountering obstacles.

The optimal trunk trajectory can be constructed as a constrained optimization problem:

$$\begin{aligned} &\boldsymbol{T}=\boldsymbol{L}_{\mathrm{o}}\cup\boldsymbol{L}_{\mathrm{p}},\\ &\text{s.t. } \boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}\notin\boldsymbol{p}^{\mathrm{o}},\\ &\quad \max \mathrm{Var}(\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}),\\ &\max \mho(\boldsymbol{L}_{\mathrm{o}})\exp\left(-\sum_{\boldsymbol{S}}\|\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}\|\right), \end{aligned} \tag{7}$$

where $\boldsymbol{L}_{\mathrm{o}}$ is the vector within the observation time window, and $\boldsymbol{L}_{\mathrm{p}}$ is a vector within the prediction time window. $\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}$ is the set of projections of the sampling points onto the vector $\boldsymbol{L}_{\mathrm{o}}$. $\boldsymbol{p}^{\mathrm{o}}$ are the coordinates of the obstacles. $\mathrm{Var}(\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}})$ is the variance of $\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}$. $\mho(\boldsymbol{L}_{\mathrm{o}})$ is the number of sampling points covered by the drone when it moves along $\boldsymbol{L}_{\mathrm{o}}$. $\sum_{\boldsymbol{S}}\|\boldsymbol{P}_{\boldsymbol{L}_{\mathrm{o}}}\|$ is the sum of the distances from the sampling points to $\boldsymbol{L}_{\mathrm{o}}$.

The trunk trajectory also contains important intention information, as shown in Fig. 2. Although the movement of the target appears to be irregular, it can be seen from the growth direction that the target has a clear destination.

3.2 Intent estimation

The target intention is reflected by its own trajectory and surrounding environment. Using the relationship between the target trajectory and its surrounding environment, from the perspective of whether there is a destination demand, we identify the three classic intentional movements within the observation time window: maintaining motion, trend motion, and reverse motion. Other intentions can be represented by combining the three types. The maintaining motion is a movement without purpose; there is no intention of moving toward or away from the destination. Thus, the trajectory within the next time window can be predicted more accurately according to the current trunk trajectory. The trend motion is a purposeful movement trend toward a destination, and it is possible to change its state suddenly with the help of the environment near the destination, such as hiding and escaping. Therefore, the surrounding environment needs to be taken into account when predicting the trajectory. Reverse motion is the movement that deviates from the destination, as opposed to the trend motion. Without loss of generality, in this study, we take barriers as examples to characterize the environment.

We use two characteristics of the target's intention: the changes in the speed at which the target's trunk trajectory approaches obstacles and the density distribution of surrounding obstacles. The smaller the distance between the target and the obstacle, and the faster the approaching speed, the more obvious the intention of the target approaching the obstacle. When the target chooses to approach from an open area where it is easy to move to a dense obstacle area, the intention of the target escaping or hiding itself using the shielding of obstacles becomes more obvious. Conversely, when the target moves from a dense obstacle area to an open area, the intention of shielding with the help of obstacles is weakened, and the intention of the reverse movement is strengthened. When the target has been moving in an open area or the relative position between the target and the obstacles remains unchanged over a certain time, the intention of the maintaining motion is obvious. We denote the intent probability function by the Sigmoid function:

$$f(x)=\frac{1}{1+\mathrm{e}^{-x}}, \tag{8}$$

$$x=\alpha(\tilde{v}-\bar{v})+\beta(\tilde{f}-\bar{f}), \tag{9}$$

$$\bar{f}=\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{1}{\|\boldsymbol{p}^{\mathrm{t}}(k)-\boldsymbol{p}^{\mathrm{o}}(k)\|}, \tag{10}$$

where α and β are the weight coefficients, and $\tilde{v}$ and $\bar{v}$ are the average speeds of the target within the current observation time window and the previous observation time window, respectively. Let δ be a neighborhood radius of the target. $\tilde{f}$ and $\bar{f}$ are the average density distributions of obstacles in δ within the current observation time window and the previous observation time window, respectively. $\bar{f}$ can be represented by the average potential function (10), where $\boldsymbol{p}^{\mathrm{t}}(k)$ and $\boldsymbol{p}^{\mathrm{o}}(k)$ are the coordinates of the target and the nearest obstacle within δ at time k, respectively. κ is the sampling number within the observation time window. The two terms in Eq. (9) represent the changes in the velocity of the target approaching obstacles and in the density distribution of

the obstacles around it within the observation time window.

Combining the above two terms as the input of the Sigmoid function, we have the output as the probability of intention. If $f(x) \in (0.5, 1)$, it is the trend motion; if $f(x) \in (0, 0.5)$, it is the reverse motion; if $f(x)$=0.5, it is the maintaining motion.

Define Δt_{o} as the observation time window. Let η be a variable coefficient, which can be adjusted according to the intention of the target:

$$\eta = 2.5 - f(x) - |f(x) - 0.5|. \tag{11}$$

For the maintaining motion and the reverse motion, obstacles have little influence on tracking. For the trend motion, the adverse effect of obstacles on tracking should be considered. The short-term predicted trajectory should be on the same side of the obstacle to avoid obstructing observation. To improve prediction accuracy, the sampling number in the observation window is adjusted by

$$h(\eta) = \frac{1}{\eta}\bar{h}, \tag{12}$$

where $\bar{h}$ is the preset upper bound of the sampling number.

Define Δt_{p} as the prediction time window:

$$\Delta t_{\mathrm{p}} = \eta \Delta \bar{t}_{\mathrm{p}}, \tag{13}$$

where $\Delta \bar{t}_{\mathrm{p}}$ is a preset constant of the time window.

When the intention of trend motion is more obvious, the probability of the target using the obstacle and making a state saltation is increased. At this time, η becomes smaller and shortens the prediction window to adapt to the observation for emergencies. Meanwhile, the sampling number in the observation window is increased to improve the accuracy of curve fitting.

Fig. 3 shows the process of target escaping using obstacles.

According to Eq. (3), we can obtain the estimated trajectory of target j, for all $j \in \{1, 2, \ldots, M\}$, within an observation window denoted by

$$\boldsymbol{\Omega}^{j}_{\Delta t_{\mathrm{o}}} = \{\hat{\boldsymbol{X}}_j(k)\}, \quad \forall k \in \Delta t_{\mathrm{o}}. \tag{14}$$

Based on $\boldsymbol{\Omega}^{j}_{\Delta t_{\mathrm{o}}}$ we use the curve fitting mentioned in Section 3.1 to predict the trajectory $\boldsymbol{\Omega}^{j}_{\Delta t_{\mathrm{p}}}$ of target j in the future prediction time window Δt_{p}.

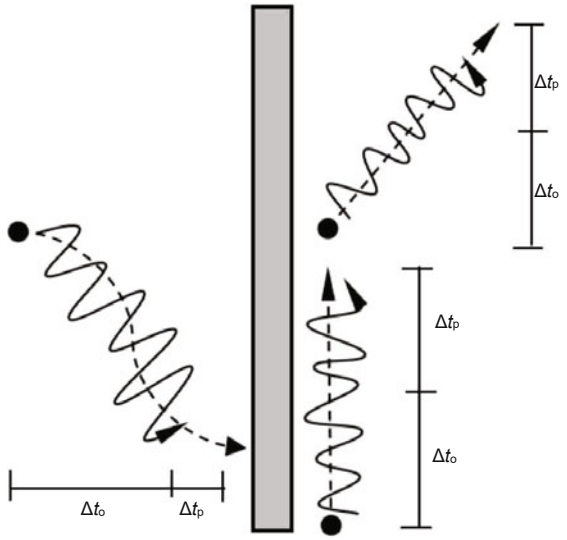

Fig. 3 The target escapes with a barrier cover
The solid curves and the dashed curves are the trajectories and the trunk trajectories of the target, respectively. The black rectangle is an obstacle that the target uses as cover

3.3 MDA-Voronoi diagram

Due to the inferior UAV tracking abilities and target intelligence, it is very difficult to accurately predict the trajectory coordinates at a future time point and to maintain sensing of the targets. We propose a geometric segmentation method and define the maximum reachable region of each UAV, which is used to discover the extreme coverage holes. Then we design a tracking method to effectively reduce the tracking gaps to increase the target monitoring time.

We design a novel Voronoi diagram called the MDA-Voronoi diagram. This diagram divides the region according to the minimum reachable distance and the minimum steering angle, and the linear combination of the distance and the angle is called the MDA distance from the UAVs to a point $\boldsymbol{q} \in \mathbb{R}^2$.

A group of N UAVs $\boldsymbol{U} = \{\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_N\}$ is located at $\boldsymbol{P} = \{\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_N\}$. Define the MDA distance from $\boldsymbol{U}_i$ to a point $\boldsymbol{q} \in \mathbb{R}^2$ as

$$d(\boldsymbol{U}_i, \boldsymbol{q}) = \tilde{\alpha} d_{\min}(\boldsymbol{U}_i, \boldsymbol{q}) + \tilde{\beta} \theta_{\min}(\boldsymbol{U}_i, \boldsymbol{q}), \tag{15}$$

where $\tilde{\alpha}$, $\tilde{\beta}$ are the weight coefficients. $d_{\min}(\boldsymbol{U}_i, \boldsymbol{q}) = \min\|\boldsymbol{U}_i - \boldsymbol{q}\|$ is the minimum reachable distance from $\boldsymbol{U}_i$ to $\boldsymbol{q}$, $\theta(\boldsymbol{U}_i, \boldsymbol{q})$ is the angle at which $\boldsymbol{U}_i$ should turn when it moves from point $\boldsymbol{P}_i$ to point $\boldsymbol{q}$, and $\theta_{\min}(\boldsymbol{U}_i, \boldsymbol{q}) = \min\theta(\boldsymbol{U}_i, \boldsymbol{q})$ is the minimum angle between the speed direction of $\boldsymbol{U}_i$ and the line connecting it with $\boldsymbol{q}$. If there is an obstacle between $\boldsymbol{U}_i$ and

$\boldsymbol{q}$, $\theta_{\min}(\boldsymbol{U}_i, \boldsymbol{q})$ is the sum of the required corners of $\boldsymbol{U}_i$ along the shortest route from its location to $\boldsymbol{q}$. As can be seen from Fig. 4, assume that $\boldsymbol{U}_i$ is located at point $\boldsymbol{A}$, moving from $\boldsymbol{A}$ to $\boldsymbol{E}$. There exist countless lines from point $\boldsymbol{A}$ to point $\boldsymbol{Q}$, such as curves $\widehat{\boldsymbol{ABQ}}$, $\widehat{\boldsymbol{ACQ}}$, and $\widehat{\boldsymbol{ADQ}}$, from which it is easy to find the shortest one $\widehat{\boldsymbol{ABQ}}$, so $d_{\min}(\boldsymbol{U}_i, \boldsymbol{Q}) = L(\widehat{\boldsymbol{ABQ}})$, where $L(\widehat{\boldsymbol{ABQ}})$ is the length of curve $\widehat{\boldsymbol{ABQ}}$, and in this case, $L(\widehat{\boldsymbol{ABQ}}) = \|\boldsymbol{A} - \boldsymbol{B}\| + \|\boldsymbol{B} - \boldsymbol{Q}\|$, $\theta_{\min}(\boldsymbol{U}_i, \boldsymbol{Q}) = \angle \boldsymbol{EAB} + \angle \boldsymbol{FBQ}$.

$d(\boldsymbol{U}_i, \boldsymbol{q})$ reflects the advantage of the UAV's position and angle relative to $\boldsymbol{q}$ in an environment with obstacles. Each region in the MDA-Voronoi contains only one UAV, which is the closest one in terms of the MDA distance to any point within that region. This partition also represents the optimal area division for each UAV in terms of the MDA distance at a time point. It can be seen that the MDA-Voronoi changes dynamically according to the positions and angles of the UAVs relative to the targets. Analogous to the conventional Voronoi partition, each MDA-Voronoi region $\boldsymbol{\Pi}_i$ ($\forall i \in \{1, 2, \ldots, N\}$), in this diagram can be characterized as

$$\boldsymbol{\Pi}_i = \{\boldsymbol{q} \in \mathbb{R}^2 | d(\boldsymbol{U}_i, \boldsymbol{q}) \le d(\boldsymbol{U}_j, \boldsymbol{q}),\ \forall j \in \mathbb{N} \setminus \{i\}\}. \tag{16}$$

Definition 2 Consider $\boldsymbol{U}_i$ with sensing radius r and the corresponding MDA-Voronoi region $\boldsymbol{\Pi}_i, \forall i \in \{1, 2, \ldots, N\}$. The i^{th} sensing region with respect to $\boldsymbol{U}_i$ is defined as the intersection of $\boldsymbol{\Pi}_i$ and a circle of radius r centered at $\boldsymbol{P}_i$, and is denoted by $\boldsymbol{S}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$, i.e.,

$$\boldsymbol{S}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i} = \boldsymbol{\Pi}_i \cap \boldsymbol{C}(\boldsymbol{P}_i, r), \tag{17}$$

where $\boldsymbol{C}(\boldsymbol{P}_i, r)$ is a circle of radius r centered at $\boldsymbol{P}_i$. Denote the total sensing region of UAVs for the field by $\boldsymbol{S}_{\text{total}} = \sum_{i=1}^{N} \boldsymbol{S}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$.

Definition 3 Assume that $\boldsymbol{U}_i$ is located at $\boldsymbol{P}_i$ at time t_0. Then, due to mobility, the maximum sensing range that it can reach within $[t_0, t_0 + \Delta t_{\text{p}}]$ is a circle of radius $r + \Delta t_{\text{p}} V$ centered at $\boldsymbol{P}_i$, defined as $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$, and is expressed as

$$\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i} = \boldsymbol{C}(\boldsymbol{P}_i, (r + \Delta t_{\text{p}} V)). \tag{18}$$

Definition 4 For any $\boldsymbol{\Pi}_i$ and $\boldsymbol{U}_i$, the ECO with respect to $\boldsymbol{U}_i$ is defined as the intersection of $\boldsymbol{\Pi}_i$ and the exterior of $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$, denoted by $\bar{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i} = \boldsymbol{\Pi}_i - \tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$. Furthermore, the union of all ECOs in an MDA-Voronoi diagram is called the total extreme coverage holes (TECOs), and its area is represented by $\bar{\boldsymbol{S}} = \sum_{i=1}^{N} \bar{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$ (Fig. 5).

Definition 5 The escape trajectory of target j, $\forall j = \{1, 2, \ldots, M\}$, is defined as the part of its trajectory that cannot be covered by any UAV. Furthermore, the ratio of the undetected trajectory to the total trajectory is referred to as the escape probability, i.e., the ratio of its trajectory length outside the UAVs' sensing range to its total trajectory length. The intersection of target j's prediction trajectory and $\boldsymbol{\Pi}_i$, and $\boldsymbol{C}(\boldsymbol{P}_i, r)$, are denoted by $\boldsymbol{\Omega}_{\boldsymbol{\Pi}_i}^{j}$

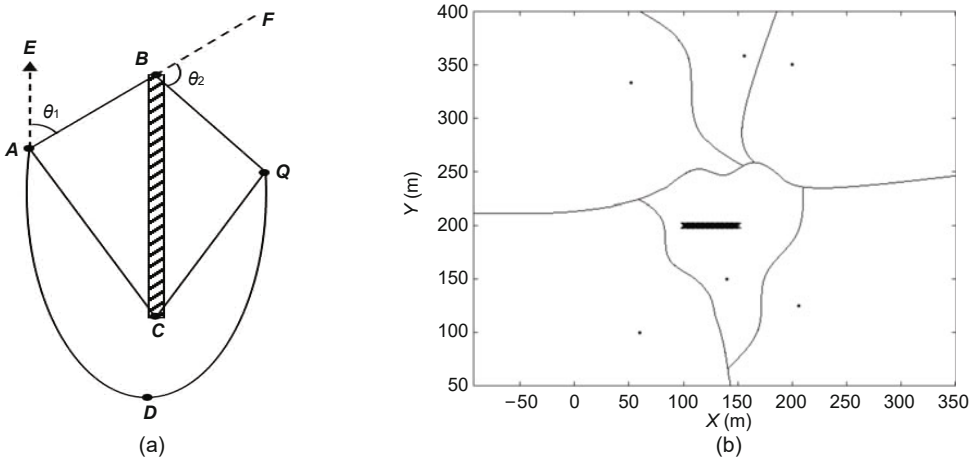

Fig. 4 MDA-Voronoi diagram: (a) minimum reachable distance and the minimum steering angle; (b) an example of an MDA-Voronoi diagram for a group of six UAVs in a 2D plane with an obstacle

The black rectangle is an obstacle. In (b), the positions of the UAVs are [52, 333], [156, 358], [200, 350], [206, 125], [140, 150], and [60, 100]. The angles relative to the right horizontal direction are −45, −90, −135, 135, 45, and 45 degrees, respectively

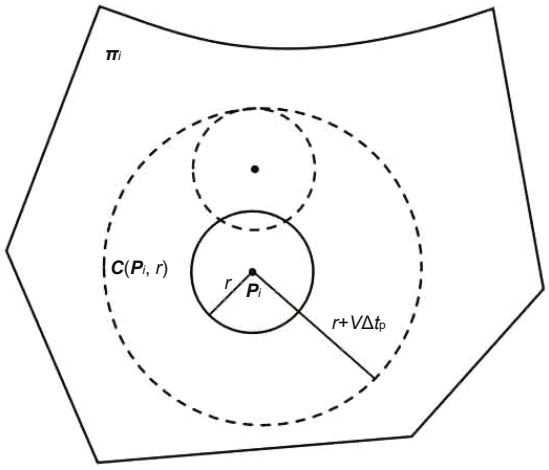

Fig. 5 Extreme coverage hole with respect to U_i

and $\boldsymbol{\Omega}_{\boldsymbol{C}_i}^j$, respectively. The trajectory curve of target $j, \forall j = \{1, 2, \ldots, M\}$, is denoted by $\boldsymbol{\Omega}_j = \bigcup\limits_{i \in \mathbb{N}} (\boldsymbol{\Omega}_{\boldsymbol{\Pi}_i}^j)$, and the length is denoted by $L(\boldsymbol{\Omega}_j)$.

Theorem 1 Under the premise of the optimal tracking distance and angle, the target's escape probability is related to the number, distribution, speed, and sensing radius of the UAVs, i.e., the more uniform the distribution, the larger the number, the higher the speed, and the larger the sensing radius of UAVs, the smaller the target's escape probability.

Proof Consider a set of N UAVs randomly located in a 2D field. The number of intersections of $\boldsymbol{\Omega}_j$ $(\forall j = \{1, 2, \ldots, M\})$ and all UAVs' sensing areas is denoted by

$$\mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C}) = \sum_{i=1}^{N} \mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C}(\boldsymbol{P}_i, r)), \tag{19}$$

where

$$\mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C}(\boldsymbol{P}_i, r)) = \begin{cases} 0, & \boldsymbol{\Omega}_j \cap \boldsymbol{C}(\boldsymbol{P}_i, r) = \varnothing, \\ 1, & \boldsymbol{\Omega}_j \cap \boldsymbol{C}(\boldsymbol{P}_i, r) \neq \varnothing. \end{cases} \tag{20}$$

The probability of target j being covered is

$$\Theta_j = \frac{1}{L(\boldsymbol{\Omega}_j)} \sum_{i=1}^{\mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C})} (\max(t_c) \cdot v_j) \cap \boldsymbol{\Pi}_i, \tag{21}$$

where t_c is the length of time when the target is within the sensing range of $\boldsymbol{U}_i$, $\forall i \in \{1, 2, \ldots, N\}$, and $\max(t_c)$ represents the longest time the target can be in the sensing range of $\boldsymbol{U}_i$. v_j is the velocity of target j. $v_j - V$ is positive and represents the velocity difference between target j and $\boldsymbol{U}_i$.

It is clear that, in a bounded region, for a fixed trajectory, $\mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C})$ increases as $\boldsymbol{S}_{\text{total}}$ increases. With the small number, uneven distribution, and the small sensing range of the UAVs, sensing holes exist. When the target is always in the sensing holes, i.e., $\boldsymbol{\Omega}_j \in \mathbb{R}^2 - \boldsymbol{S}_{\text{total}}$, $t_c = 0$, and the probability of the target being covered is 0. When the target trajectory passes through $\boldsymbol{C}(\boldsymbol{P}_i, r)$, $\forall i \in \{1, 2, \ldots, N\}$, $\boldsymbol{U}_i$ tracking along the direction of target speed until the target leaves $\boldsymbol{C}(\boldsymbol{P}_i, r)$ is the best way to increase the sensing time, and in this case $\text{minmax}(t_c) = \frac{2r}{v_j - V}$. If the target is always stationary or is moving in $\boldsymbol{C}(\boldsymbol{P}_i, r)$, i.e., $t_c \to \infty$, then $\lim\limits_{t_c \to \infty} (\max(t_c) v_j) = L(\boldsymbol{\Omega}_j \cap \boldsymbol{\Pi}_i)$, and in this case, $\Theta_j = 1$.

The smaller the velocity difference, the larger the sensing radius and $\mathbb{N}(\boldsymbol{\Omega}_j, \boldsymbol{C})$, the larger Θ_j, and the probability of target j being covered by UAVs is in the range $[0, \Theta_j]$.

3.4 Tracking strategies for UAVs

Definition 6 For any $\boldsymbol{\Pi}_i$ $(\forall i \in \{1, 2, \ldots, N\})$ and $\boldsymbol{\Omega}_j$ that satisfy $\boldsymbol{\Omega}_{\boldsymbol{\Pi}_i}^j \cap \tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i} \neq \varnothing, \forall j = \{1, 2, \ldots, M\}$, we have the following definitions. The two intersection points of $\boldsymbol{\Omega}_j$ with $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$ are defined as the entry point and exit point, respectively. The time–space intersection point (TSIP) of $\boldsymbol{\Omega}_j$ is defined as the coordinates at which $\boldsymbol{U}_i$ can first cover the target on $\boldsymbol{\Omega}_j$ within $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$. The maximum time-space intersection curve (MTSIC) of $\boldsymbol{\Omega}_j$ is defined as the part on $\boldsymbol{\Omega}_j$ from the TSIP point to the exit point within $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$. For example, for the case in Fig. 6b, the entry point and exit point of $\boldsymbol{\Omega}_1$ are $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Without loss of generality, for a better description, in Fig. 6 we assume that the TSIP point is the entry point. The minimum distance between two curves is defined as the shortest distance between them within $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$, and the maximum distance between two curves is defined as the longest distance between them within $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$. The neighbor curves of $\boldsymbol{\Omega}_j$ are referred to as the curves whose maximum distances from $\boldsymbol{\Omega}_j$ are within $2r$, denoted as Ξ_j. The farthest neighbor of $\boldsymbol{\Omega}_j$ is defined as the farthest curve within Ξ_j. The longest curve in $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$ is defined as the curve with the largest length in $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$.

For any $\boldsymbol{\Pi}_i$, $\forall i \in \{1, 2, \ldots, N\}$, and $\boldsymbol{\Omega}_j$, $\forall j = \{1, 2, \ldots, M\}$, if $\boldsymbol{\Omega}_{\boldsymbol{\Pi}_i}^j \cap \tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i} = \varnothing$, i.e., $\boldsymbol{\Omega}_j$ is not within the range of the maximum reachable region of

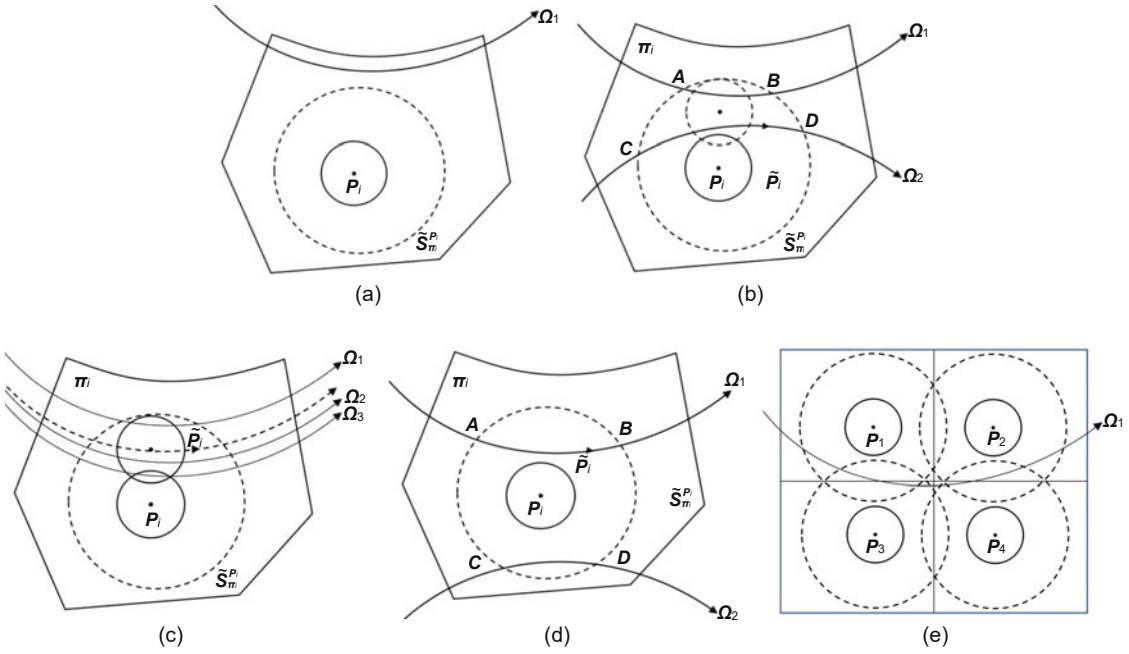

Fig. 6 Different tracking strategies: (a) a UAV does not need to be involved in tracking; (b) a UAV will choose to cover the longest curve based on the maximum distance between the trajectories; (c) a UAV will cover as many trajectories as possible; (d) a UAV will choose to cover the longest one based on the minimum distance between the trajectories; (e) more than one UAV will participate in tracking

$\boldsymbol{U}_i$ during Δt_{p}, then $\boldsymbol{U}_i$ does not need to be involved in tracking because it cannot cover the target anyway (Fig. 6a). With the assistance of the ground stations, the UAV can still obtain the motion coordinates of the target, which guide it to move along the direction of the target.

If more than one prediction trajectory passes through $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$ and the maximum distances between them are greater than $2r$, $\boldsymbol{U}_i$ will choose to cover the longest MTSIC curve, and move to a tracking point $\tilde{\boldsymbol{P}}_i$. Let $\tilde{\boldsymbol{P}}_i$ be a virtual point moving at velocity V along the longest curve from the TSIP point to the exit point (Fig. 6b). If the maximum distances between them are less than $2r$, then we select two curves with the maximum distance and set the virtual tracking point $\tilde{\boldsymbol{P}}_i$ to move along their middle curve, which is the curve at the same distance from the two curves; in this way, $\boldsymbol{U}_i$ can cover as many trajectories as possible when tracking (Fig. 6c).

If more than one trajectory passes through $\tilde{\boldsymbol{S}}_{\boldsymbol{\Pi}_i}^{\boldsymbol{P}_i}$ and the minimum distance between them is larger than $2r$, $\boldsymbol{U}_i$ will choose to cover the longest one (Fig. 6d).

If the trajectory is chosen by more than one UAV at the same time, then the UAVs can also participate in tracking (Fig. 6e).

4 Collaborative tracking based on Nash Q-learning

Reinforcement learning (Sutton and Barto, 1998) is based on the Markov decision process, does not rely on prior knowledge, and is suitable for the decision-making problems in unknown or uncertain complex environments.

Conventional Q-learning is a single-agent reinforcement learning algorithm, which uses state-action rewards during learning iterations to seek the optimal strategy π^*. In state $\boldsymbol{s}_t$, action $\boldsymbol{a}_t$ is chosen according to the ε-greedy strategy, the immediate reward r_t and the next state $\boldsymbol{s}_{t+1}$ are obtained, and then Q values are updated as

$$\begin{aligned} &Q(\boldsymbol{s}_t,\boldsymbol{a}_t)\leftarrow Q(\boldsymbol{s}_t,\boldsymbol{a}_t)\\ &+\sigma[r_t+\gamma\max_{\boldsymbol{a}\in A}Q(\boldsymbol{s}_{t+1},\boldsymbol{a})-Q(\boldsymbol{s}_t,\boldsymbol{a}_t)], \end{aligned} \tag{22}$$

where $\sigma\in[0,1]$ is the learning rate, A is the action space, and $\gamma\in[0,1]$ is the discount factor. Because the conventional Q-learning algorithm considers only

the actions of a single agent, without considering the influence of other agents, we use the Nash Q-learning algorithm (Hu and Wellman, 2003), which considers the collaboration among multiple agents and extends the state-action Q function to the joint state-action Q function. The Q value of $\boldsymbol{U}_i$ is updated as

$$\begin{aligned} &Q^i(\boldsymbol{s}_t,\boldsymbol{a}_t)\\ &\leftarrow Q^i(\boldsymbol{s}_t,\boldsymbol{a}_t)+\sigma[r_t^i+\gamma \mathrm{Nash}Q^i(\boldsymbol{s}_{t+1})-Q^i(\boldsymbol{s}_t,\boldsymbol{a}_t)], \end{aligned} \tag{23}$$

$$\mathrm{Nash}Q^i(\boldsymbol{s}_{t+1})=\pi^1(\boldsymbol{s}_{t+1})\cdots\pi^N(\boldsymbol{s}_{t+1})Q^i(\boldsymbol{s}_{t+1}), \tag{24}$$

where $(\boldsymbol{s}_t,\boldsymbol{a}_t)$ is a joint state-action pair, r_t^i is the immediate reward of $\boldsymbol{U}_i$, and $\pi^1(\boldsymbol{s}_{t+1}),\ldots,\pi^N(\boldsymbol{s}_{t+1})$ is the Nash equilibrium for multiple UAVs under joint state $\boldsymbol{s}_{t+1}$. $\mathrm{Nash}Q^i(\boldsymbol{s}_{t+1})$ is the payoff of $\boldsymbol{U}_i$ in state $\boldsymbol{s}_{t+1}$ for the selected equilibrium. Because we use a mixed-strategy game, the Nash equilibrium exists.

4.1 State-action space

Because a high-dimensional continuous state-action space leads to huge numbers of states and actions, and slow convergence, we separate the state-action space without affecting the overall effect, and design a smaller state space S and the action space A. For any $\boldsymbol{U}_i$, $i\in\{1,2,\ldots,N\}$, S is divided into H sections based on the distance between $\boldsymbol{U}_i$ and tracking point $\tilde{\boldsymbol{P}}_i$, i.e.,

$$S=\begin{cases}1, & 0<d_{ij}(k)\le \bar{d}_1,\\ 2, & \bar{d}_1<d_{ij}(k)\le \bar{d}_2,\\ \vdots & \vdots\\ H, & \bar{d}_{H-1}<d_{ij}(k)\le \bar{d}_H,\end{cases} \tag{25}$$

where $d_{ij}(k)$ is the distance between $\boldsymbol{U}_i$ and $\tilde{\boldsymbol{P}}_i$ at time k, $\bar{d}_i-\bar{d}_{i-1}=\bar{d}$, and $i\in\{2,3,\ldots,H\}$ is the dividing threshold. Based on the UAV heading angle constraint, $\Delta\psi_i$ is discretized into 13 parts, which we define as the action space A:

$$A=\{0^\circ,\pm15^\circ,\pm30^\circ,\pm45^\circ,\pm60^\circ,\pm75^\circ,\pm90^\circ\}. \tag{26}$$

4.2 Reward function design

In the reward function, we consider factors including collision avoidance between UAVs, and between UAVs and obstacles, sensing performance, tracking contribution, and path overlap, described as follows.

Let $d_{\mathrm{uo}}(k)$ and $d_{\mathrm{ut}}(k)$ be the distances between the UAV and the obstacle and between the UAV and the tracking point at time k, respectively. $d_{\mathrm{uo}}(k+1)-d_{\mathrm{uo}}(k)$ is the difference of the distance between the UAV and the obstacle at time $k+1$ and time k, denoted as Δ_1, which indicates whether the UAV is approaching or moving away from the obstacle, and we assume that the obstacle can be a fixed barrier or other UAVs closest to the UAV. $d_{\mathrm{ut}}(k+1)-d_{\mathrm{ut}}(k)$ is the difference of the distance between the UAV and the tracking point at time $k+1$ and time k, denoted as Δ_2, reflecting whether the UAV is approaching or moving away from the tracking point.

$$\mu=\left|\frac{d_{\mathrm{uo}}(k+1)-d_{\mathrm{uo}}(k)}{d_{\mathrm{ut}}(k+1)-d_{\mathrm{ut}}(k)}\right| \tag{27}$$

is the speed ratio of the UAV approaching the obstacles and the tracking point, reflecting the trade-off between tracking and avoiding collisions, denoted as Δ_3. Define

$$d_{\mathrm{s}}(k)=\sum_{j\ne i}\|\boldsymbol{P}_i(k)-\boldsymbol{P}_j(k)\| \tag{28}$$

as the sum of the distances between $\boldsymbol{U}_i$ and others at time k, and $d_{\mathrm{s}}(k+1)-d_{\mathrm{s}}(k)$, denoted as Δ_4, reflects the aggregation degree between the UAVs, avoiding the path overlap in the tracking and increasing the dispersion of the UAVs, which is conducive to reducing the TECOs. In addition, we consider the rewards for both UAV safety and tracking cost. Based on the four indicators and constraints above, we can describe the different states that the UAVs may encounter during tracking, and set different reward values to define the reward function, as given in Eq. (29) (see the next page). In Eq. (29), $\Phi=\sum_{k=1}^{E}\sum_{i=1}^{N}D_{it}(k)$ represents the total sensing performance of the target by the UAVs during a training cycle, E is the total number of steps in a cycle, and $D_{it}(k)$ represents the sensing performance of $\boldsymbol{U}_i$ to the target at time k. When Φ reaches a small neighborhood ξ of the expected value $\bar{D}$, which can be set according to experience, i.e., $\Phi\in[\bar{D}-\xi,\bar{D}+\xi]$, the cooperation achieves the expected tracking performance, and a large positive reward is given. When $\Phi>\bar{D}+\xi$, the sensing has been exceeded, and some UAVs may not need to participate in tracking to avoid wasting

$$r(\boldsymbol{s},\boldsymbol{a})=\begin{cases}-200, & \Delta_1<0,\Delta_2>0,\Delta_3<1,\Delta_4<0,\\ -175, & \Delta_1<0,\Delta_2>0,\Delta_3<1,\Delta_4>0,\\ -150, & \Delta_1<0,\Delta_2>0,\Delta_3>1,\Delta_4<0,\\ -125, & \Delta_1<0,\Delta_2>0,\Delta_3>1,\Delta_4>0,\\ -100, & \Delta_1>0,\Delta_2>0,\Delta_3<1,\Delta_4<0,\\ -75, & \Delta_1>0,\Delta_2>0,\Delta_3<1,\Delta_4>0,\\ -50, & \Delta_1>0,\Delta_2>0,\Delta_3>1,\Delta_4<0,\\ -25, & \Delta_1>0,\Delta_2>0,\Delta_3>1,\Delta_4>0,\\ 25, & \Delta_1<0,\Delta_2<0,\Delta_3<1,\Delta_4<0,\\ 50, & \Delta_1<0,\Delta_2<0,\Delta_3<1,\Delta_4>0,\\ 75, & \Delta_1<0,\Delta_2<0,\Delta_3>1,\Delta_4<0,\\ 100, & \Delta_1<0,\Delta_2<0,\Delta_3>1,\Delta_4>0,\\ 125, & \Delta_1>0,\Delta_2<0,\Delta_3<1,\Delta_4<0,\\ 150, & \Delta_1>0,\Delta_2<0,\Delta_3<1,\Delta_4>0,\\ 175, & \Delta_1>0,\Delta_2<0,\Delta_3>1,\Delta_4<0,\\ 200, & \Delta_1>0,\Delta_2<0,\Delta_3>1,\Delta_4>0,\\ -500, & d_{\mathrm{uo}}(k)<5,\\ 500, & d_{\mathrm{ut}}(k)<5,\\ 100, & \Phi\in[\bar{D}-\xi,\bar{D}+\xi],\\ -100, & \Phi\notin[\bar{D}-\xi,\bar{D}+\xi].\end{cases}\tag{29}$$

resources, and a negative reward is given. When $\Phi<\bar{D}-\xi$, the UAVs' tracking does not meet the sensing requirement, and a negative reward is given. Φ reflects the collaboration performance of the UAVs, and enables as few UAVs as possible to achieve the expected sensing, to improve the tracking efficiency. Because the total reward belongs to all drones involved in tracking in a learning cycle, it needs to be assigned to each individual. In the case of meeting the sensing, we define the ratio of the sensing contribution to the flight distance as the tracking contribution of $\boldsymbol{U}_i$, $i\in\{1,2,\ldots,N\}$, i.e.,

$$\varsigma_i=\frac{\sum\limits_{k=1}^{E}D_{it}(k)}{\sum\limits_{k=1}^{E}\mathrm{dis}_i(k)},\tag{30}$$

where $\mathrm{dis}_i(k)$ is the flight distance of $\boldsymbol{U}_i$ at step k. All UAVs are ranked by their tracking contribution. For those with long flight distance and small sensing contribution, it is not necessary to participate in tracking. Therefore, it is necessary to give a larger reward for the actions that give up participation, so as to avoid their participation in tracking. Thus, the number of individuals participating in tracking is minimized, which can reduce the total energy consumption and maintain dispersion of the UAVs to a certain extent, facilitating subsequent deployment.

5 Simulation verifications

In the simulations, the observation time window and the preset constant of the time window are Δt_{o}=2.5 s and Δt_{p}=1.5 s, respectively. The learning cycle is 500 and the learning step is 80. We use different colors to represent the tracking state of a UAV, in which green indicates the state in which the UAV can cover a target, while red indicates not, and yellow indicates the state in which the UAV should not be involved in tracking. Two scenarios are used to demonstrate the proposed strategies. In scenario 1, we consider four UAVs and only one target. In scenario 2, the situation is more complex, and two targets, two obstacles, and five UAVs are included. The targets' movements are more deceptive. The simulation parameters are shown in Table 1.

5.1 Scenario 1

In Fig. 7a, the target is trying to escape with the cover of the oval obstacle. The true motion trajectory of the target is represented by the black solid line. The black circle denotes the starting point of the target and the black triangle denotes the end point. The cross symbols denote the boundary of the obstacle. The four black squares denote ground stations, located at [80, 60], [80, 380], [150, 60], and [150, 380]. The blue curves are estimates of the Kalman filtering, and the black curve is the average of the estimates.

In Fig. 7b, the red circles and red triangles denote the starting and end points of the UAVs, respectively. The black circle and black triangle denote the starting and end points of the target, respectively. The trunk trajectory, represented by the black solid line, shows that in the first stage, the target quickly approaches the upper edge of the obstacle, and the intention of taking the obstacle as cover is obvious. In the second stage, the target suddenly changes the direction and approaches the edge of the obstacle again, and the intention to escape with the help of the obstacle's cover increases. From the UAVs'

Table 1 Simulation parameters

Parameter	Value	
	Scenario 1	Scenario 2
Task environment	250 m × 400 m	250 m× 400 m
UAV number	N=4	N=5
UAV initial positions	U_1=[52, 333], U_2=[206, 125], U_3=[156, 358], U_4=[212, 225]	U_1=[36, 135], U_2=[86, 30], U_3=[135, 27], U_4=[192, 83], U_5=[115, 303]
Initial flight angles ψ	[−45°, 180°, 180°, 180°]	[45°, 60°, −45°, 90°, 135°]
Probability of the initial action selection of the UAVs	1/8	1/8
UAV velocity	V=1.5 m/s	V=1.3 m/s
Target number	M=1	M=2
Target initial position	[65, 325]	[60, 75], [185, 50]
Center of the obstacle	[125, 200]	[90, 80], [165, 120]
Sensing threshold	$\bar{D}$=200	$\bar{D}$=360
Sensing radius	r=65 m	r=50 m
Other parameters	ξ=10, $\bar{d}$=100 m, $d_{\min}$=5 m, α=1, β=3, $\tilde{\alpha}=2/3$, $\tilde{\beta}=1/3$, Δt=1 s, $\bar{h}$=12, δ=20, σ=0.6, γ=0.9, $\underline{r}$=1, $\bar{r}$=2	

trajectories, in the early stage, the target trajectory is closer to $\boldsymbol{U}_1$ and is in the MDA-Voronoi region of $\boldsymbol{U}_1$, rather than in that of others. Therefore, at this time, $\boldsymbol{U}_1$ tracks the target, but others cannot sense the target and do not participate in tracking. However, after a short time, due to the larger target speed, $\boldsymbol{U}_1$ is gradually evaded by the target and cannot cover it. However, the target gradually approaches $\boldsymbol{U}_3$ and is within the MDA-Voronoi region of U_3. At this time, $\boldsymbol{U}_3$ participates in tracking. After a while, $\boldsymbol{U}_3$ is also gradually evaded by the target. However, the target gradually enters the region of $\boldsymbol{U}_2$, and then $\boldsymbol{U}_2$ is involved in tracking. Then, the predicted trajectory of the target enters the region of $\boldsymbol{U}_1$ again, and $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ cooperate to track it. In the initial stage, the target is not within the maximum sensing range that $\boldsymbol{U}_4$ can reach, so $\boldsymbol{U}_4$ does not participate in tracking and is in a random patrol mode. At about step 50, the target is within its maximum sensing range, and $\boldsymbol{U}_4$ participates in tracking. At about step 79, because the UAVs' sensing of the target has met the demand, $\boldsymbol{U}_4$ is no longer involved in the tracking, and keeps the last action to continue patrolling. As can be seen, our method makes full use of the UAVs' position advantages using the dynamic MDA-Voronoi diagram to achieve effective collaboration. In terms of avoiding collisions, no UAVs collide with the obstacle or other UAVs.

Fig. 7c shows the average Q value during training. It can be seen that in the early training stage, the UAVs are in exploratory learning and the reward is low. As training progresses, the UAVs' strategies are gradually improved, the reward gradually increases, and the average Q value converges to a steady state.

As shown in Fig. 7d, due to the inferior tracking abilities, the UAVs' sensing of the target will drop or even be zero, but according to the dynamic MDA-Voronoi diagram and trajectory prediction, the UAVs can effectively track the target, reduce tracking gaps to a large extent, and improve the sensing performance. For example, at about step 58, $\boldsymbol{U}_1$'s sensing of the target ends, but at this time, the target sensing of $\boldsymbol{U}_3$ and $\boldsymbol{U}_4$ increases. At about step 77, $\boldsymbol{U}_3$'s sensing of the target begins to drop, but at the same time, $\boldsymbol{U}_2$ maintains the coverage of the target. At about step 138, $\boldsymbol{U}_3$'s sensing of the target ends, but $\boldsymbol{U}_1$'s sensing of the target starts to increase significantly.

We compare our strategies with a passive tracking method in which the UAVs move directly to the target. We define the tracking loss as the number of time steps when the target is not within the sensing range of any UAV. The tracking loss and the tracking trajectories are shown in Figs. 7e–7g. The total tracking loss of the proposed method is 0, and that of the passive method is 29. It can also be seen from Fig. 7f that the target is not within the sensing range of the UAVs for a long time during tracking.

5.2 Scenario 2

The two targets, named targets 1 and 2, are trying to make curve flights to the top right and top

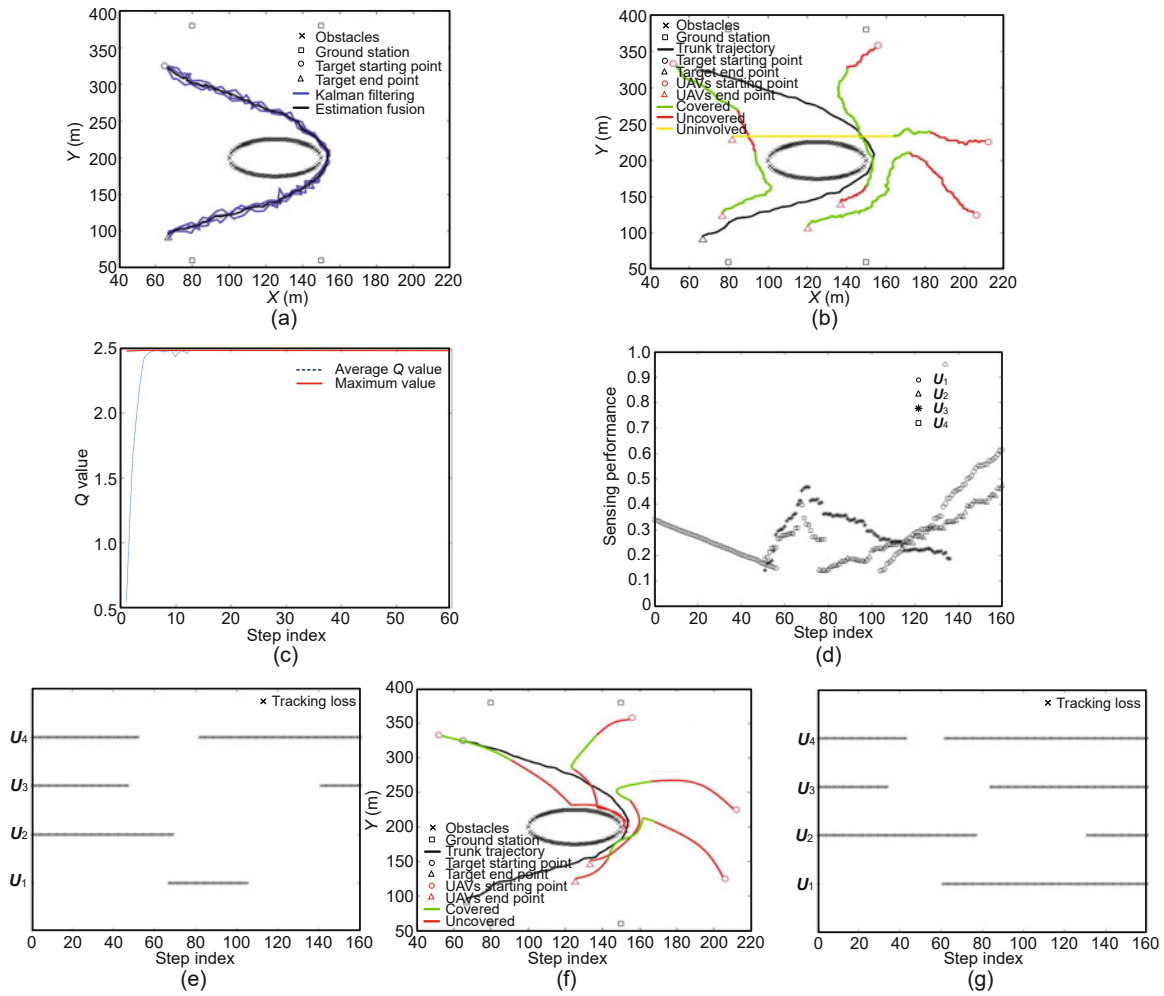

Fig. 7 Simulation results of scenario 1: (a) target trajectory; (b) UAVs' tracking trajectories; (c) average Q value during training; (d) UAVs' sensing performance in tracking; (e) tracking loss of the proposed method; (f) passive tracking trajectories of the UAVs; (g) tracking loss of the passive method

References to color refer to the online version of this figure

left, respectively, and the actual trajectories are as presented in Fig. 8a.

In Fig. 8b, it can be seen that, in the early stage of the first prediction window, $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are tracking target 1, and $\boldsymbol{U}_3$ and $\boldsymbol{U}_4$ are tracking target 2. Because $\boldsymbol{U}_5$ is far away from targets 1 and 2, it does not participate in tracking at this time. As the targets' trajectories change, $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are gradually evaded by target 1, while their tracking advantages for target 2 increase, so they start tracking target 2. Similarly, the tracking advantages of $\boldsymbol{U}_3$ and $\boldsymbol{U}_4$ for target 2 are diminishing, and even target 2 completely leaves $\boldsymbol{U}_4$'s sensing range. After a short time, the tracking advantages of $\boldsymbol{U}_3$, $\boldsymbol{U}_4$, and $\boldsymbol{U}_5$ on target 1 increase gradually, so they start to track target 1. It can be seen that all UAVs have successfully avoided collision with fixed obstacles and each other.

Fig. 8c shows the average Q value during training. Fig. 8d shows the sensing performance of each UAV during tracking. It can be seen that there is also effective collaboration between the UAVs; e.g., at about step 60, although $\boldsymbol{U}_4$'s sensing of a target ends, other UAVs effectively compensate for $\boldsymbol{U}_4$'s sensing loss at this time. After a short time, $\boldsymbol{U}_4$ is also involved in the effective tracking of another target. Figs. 8e and 8f show the tracking loss concerning the two targets during the whole tracking.

We also compare our strategies with a passive tracking method in which the UAVs move directly to the targets. The tracking loss and tracking

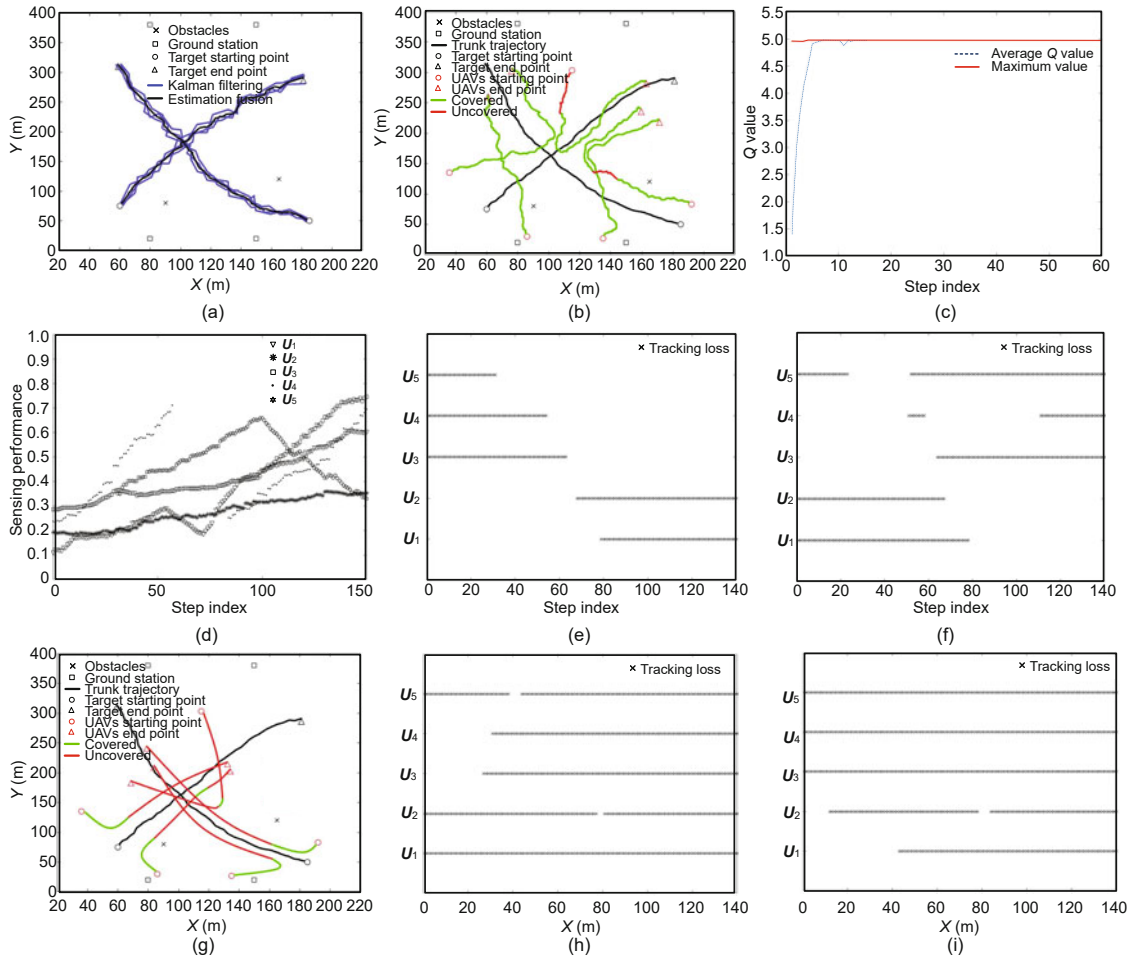

Fig. 8 Simulation results of scenario 2: (a) target trajectory; (b) UAVs' tracking trajectories; (c) average Q value during training; (d) UAVs' sensing performance in tracking; (e) tracking loss of the proposed method about target 1; (f) tracking loss of the proposed method about target 2; (g) passive tracking trajectories of the UAVs; (h) tracking loss of the passive method about target 1; (i) tracking loss of the passive method about target 2

References to color refer to the online version of this figure

trajectories are shown in Figs. 8g–8i. The total tracking loss of the proposed method about target 1 is 0, and that of the passive method is 112. The total tracking loss of the proposed method about target 2 is 3, and that of the passive method is 106. It can also be seen from Fig. 8g that the targets are not within the sensing range of the UAVs for a long time during tracking. Compared with scenario 1, our method has more advantages in this scenario, which indicates that the more complex the scenario is, the more advantages our method has.

6 Conclusions

In this study, we propose a novel multi-UAV collaborative target tracking method for UAVs with a limited detection range, inferior tracking capabilities, and targets with agile mobility, trajectory uncertainty, and decoy behavior. We first build the motion models of the UAV, the target, and the sensor detection model with a limited detection range. After discretizing the target trajectory within the observation time window, a novel trajectory feature extraction method with least dimension and maximum coverage constraints is designed. The extracted trajectory characteristics and the trunk trajectory trend approaching the obstacle are taken as inputs of a Sigmoid function to estimate the target's intent. Based on this estimate, we adjust the prediction time window and sampling number for the observation of emergencies. Then we propose a novel Voronoi

diagram, called MDA-Voronoi, which divides the area with obstacles according to the minimum reachable distance and the minimum steering angle of each UAV. In each MDA-Voronoi region, the maximum reachable region of each UAV is defined and used to discover the ECOs based on the distribution, sensing radius, and velocity of the UAVs. Furthermore, the upper and lower bounds of the trajectory coverage probability are analyzed, and the tracking strategies of the UAVs are designed to effectively reduce the tracking gaps. Based on Nash Q-learning, the learning strategies are taken as the UAV control inputs, and collision avoidance with UAVs and obstacles, sensing performance, and path overlap are also considered to reduce the number of UAVs involved in tracking while achieving the expected sensing performance to improve tracking efficiency. Finally, we give two scenarios to validate the theoretical results that the multiple UAVs can monitor multiple intelligent targets with effective cooperation in complex and uncertain environments.

Contributors

Shuncheng CAI processed the data, did the simulation verifications, and drafted the paper. Zhi ZHENG formulated the overarching research goals, helped organize the paper, and revised and finalized the paper.

Compliance with ethics guidelines

Zhi ZHENG and Shuncheng CAI declare that they have no conflict of interest.

References

Adepegba A, Miah MS, Spinello D, 2016. Multi-agent area coverage control using reinforcement learning. Proc 29th Int Florida Artificial Intelligence Research Society Conf, p.368-373.

Bordonaro SV, Willett P, Bar-Shalom Y, et al., 2019. Converted measurement sigma point Kalman filter for bistatic sonar and radar tracking. *IEEE Trans Aerosp Electron Syst*, 55(1):147-159. https://doi.org/10.1109/TAES.2018.2849179

Cai J, Huang CQ, Guo HF, 2012. Multi-sensor cooperative tracking using distributed Nash Q-learning. *Adv Mater Res*, 591-593:1475-1478. https://doi.org/10.4028/www.scientific.net/AMR.591-593.1475

Di B, Zhou R, Dong ZN, 2016. Cooperative localization and tracking of multiple targets with the communication-aware unmanned aerial vehicle system. *Contr Dec*, 31(4):616-622 (in Chinese). https://doi.org/10.13195/j.kzyjc.2015.0145

Douthwaite JA, Zhao SY, Mihaylova LS, 2019. Velocity obstacle approaches for multi-agent collision avoidance. *Unmann Syst*, 7(1):55-64. https://doi.org/10.1142/S2301385019400065

Gao GQ, Xin B, 2019. A-STC: auction-based spanning tree coverage algorithm for motion planning of cooperative robots. *Front Inform Technol Electron Eng*, 20(1):18-31. https://doi.org/10.1631/FITEE.1800551

Goldhoorn A, Garrell A, Alquézar R, et al., 2018. Searching and tracking people with cooperative mobile robots. *Auton Robot*, 42(4):739-759. https://doi.org/10.1007/s10514-017-9681-6

Hu JL, Wellman MP, 2003. Nash Q-learning for general-sum stochastic games. *J Mach Learn Res*, 4:1039-1069.

Jiang H, Liang YQ, 2018. Online path planning of autonomous UAVs for bearing-only standoff multi-target following in threat environment. *IEEE Access*, 6:22531-22544. https://doi.org/10.1109/ACCESS.2018.2824849

Khalkhali MB, Vahedian A, Yazdi HS, 2020. Multi-target state estimation using interactive Kalman filter for multi-vehicle tracking. *IEEE Trans Intell Transp Syst*, 21(3):1131-1144. https://doi.org/10.1109/TITS.2019.2902664

Li TC, 2019. Single-road-constrained positioning based on deterministic trajectory geometry. *IEEE Commun Lett*, 23(1):80-83. https://doi.org/10.1109/LCOMM.2018.2879478

Li TC, Su JY, Liu W, et al., 2017. Approximate Gaussian conjugacy: parametric recursive filtering under nonlinearity, multimodality, uncertainty, and constraint, and beyond. *Front Inform Technol Electron Eng*, 18(12):1913-1939. https://doi.org/10.1631/FITEE.1700379

Li TC, Chen HM, Sun SD, et al., 2019. Joint smoothing and tracking based on continuous-time target trajectory function fitting. *IEEE Trans Autom Sci Eng*, 16(3):1476-1483. https://doi.org/10.1109/TASE.2018.2882641

Liu YS, Wang QX, Hu HS, et al., 2019. A novel real-time moving target tracking and path planning system for a quadrotor UAV in unknown unstructured outdoor scenes. *IEEE Trans Syst Man Cybern Syst*, 49(11):2362-2372. https://doi.org/10.1109/TSMC.2018.2808471

Meng W, He ZR, Su R, et al., 2017. Decentralized multi-UAV flight autonomy for moving convoys search and track. *IEEE Trans Contr Syst Technol*, 25(4):1480-1487. https://doi.org/10.1109/TCST.2016.2601287

Quintero SAP, Copp DA, Hespanha JP, 2015. Robust UAV coordination for target tracking using output-feedback model predictive control with moving horizon estimation. American Control Conf, p.3758-3764. https://doi.org/10.1109/ACC.2015.7171914

Ragi S, Chong EKP, 2012. Dynamic UAV path planning for multitarget tracking. American Control Conf, p.3845-3850. https://doi.org/10.1109/ACC.2012.6314653

Ragi S, Chong EKP, 2013. Decentralized control of unmanned aerial vehicles for multitarget tracking. Int Conf on Unmanned Aircraft Systems, p.260-268. https://doi.org/10.1109/ICUAS.2013.6564698

Ruan WY, Duan HB, 2020. Multi-UAV obstacle avoidance control via multi-objective social learning pigeon-inspired optimization. *Front Inform Technol Electron Eng*, 21(5):740-748. https://doi.org/10.1631/FITEE.2000066

Shao Y, Zhao ZF, Li RP, et al., 2020. Target detection for multi-UAVs via digital pheromones and navigation algorithm in unknown environments. *Front Inform Technol Electron Eng*, 21(5):796-808. https://doi.org/10.1631/FITEE.1900659

Skorobogatov G, Barrado C, Salamí E, 2020. Multiple UAV systems: a survey. *Unmann Syst*, 8(2):149-169. https://doi.org/10.1142/S2301385020500090

Song WH, Wang JA, Zhao SY, et al., 2019. Event-triggered cooperative unscented Kalman filtering and its application in multi-UAV systems. *Automatica*, 105(3):264-273. https://doi.org/10.1016/j.automatica.2019.03.029

Sutton RS, Barto AG, 1998. Introduction to Reinforcement Learning. MIT Press, Cambridge, MA, USA.

Vanegas F, Campbell D, Eich M, et al., 2016. UAV based target finding and tracking in GPS-denied and cluttered environments. IEEE/RSJ Int Conf on Intelligent Robots and Systems, p.2307-2313. https://doi.org/10.1109/IROS.2016.7759360

Wang DB, Wang Y, Jiang WY, et al., 2015. Unmanned aerial vehicles cooperative path planning for ground target tracking via chemical reaction optimization. *Sci Sin Technol*, 45(6):583-594. https://doi.org/10.1360/N092015-00126

Wang L, Peng H, Zhu HY, et al., 2011. Cooperative tracking of ground moving target using unmanned aerial vehicles in cluttered environment. *Contr Theor Appl*, 28(3):300-308 (in Chinese).

Wang T, Qin RX, Chen Y, et al., 2019. A reinforcement learning approach for UAV target searching and tracking. *Multim Tools Appl*, 78(4):4347-4364. https://doi.org/10.1007/s11042-018-5739-5

Yao P, Wang HL, Su ZK, 2015. Real-time path planning of unmanned aerial vehicle for target tracking and obstacle avoidance in complex dynamic environment. *Aerosp Sci Technol*, 47(6):269-279. https://doi.org/10.1016/j.ast.2015.09.037

Yu HL, Meier K, Argyle M, et al., 2015. Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles. *IEEE/ASME Trans Mech*, 20(2):541-552. https://doi.org/10.1109/TMECH.2014.2301459

Zollars MD, Cobb RG, Grymin DJ, 2019. Optimal SUAS path planning in three-dimensional constrained environments. *Unmann Syst*, 7(2):105-118. https://doi.org/10.1142/S2301385019500031