Spatio-temporal coverage planning algorithm for multi-UAV and multi-sensor cooperative search

Deyan Zheng\textsuperscript{1a*}, Chunhui Liu\textsuperscript{2b*}, Lizhen Huang\textsuperscript{3c*}

\textsuperscript{1} Science and Technology on Avionics Integration Laboratory, Shanghai 200233, China
\textsuperscript{2} Institute of Unmanned system, Beihang University, Beijing 100191, China
\textsuperscript{3} School of Electronic and Information Engineering, Beihang University, Beijing 100191, China
\textsuperscript{a*}deyanzheng\textsuperscript{@}163.com, \textsuperscript{b*}liuchunhui\textsuperscript{2134@}buaa.edu.cn, \textsuperscript{c*}huanglizhen\textsuperscript{361@}163.com

Abstract—A spatio-temporal coverage planning algorithm for multi-UAV and multi-sensor cooperative search is proposed in this paper, named STC-MARL, which is based on multi-agent reinforcement learning and allows several UAVs equipped with two types of sensors to learn to complete full view of a field of interest (FOI). The proposed STC-MARL algorithm can make participating UAVs to learn from the environment to complete the full coverage of a specific FOI while minimizing field of views (FOVs) interacted with each other. The experimental results show in detail with simulation that the UAVs in the mission can effectively accomplish the task of spatio-temporal coverage planning.

1. Introduction
Unmanned aerial vehicles (UAV) UAV is playing a more and more important role in various military and civilian fields in recent years. Spatio-temporal coverage planning for multi-UAV is an active research branch, and this technology completes the optimal coverage of multiple UAVs to a particular field of interest (FOI) which is target area on the ground [1]. In most of the previous work, authors always have to build the accurate mathematical model of the environment [2-4]. However, it is difficult to build a completely accurate model in the actual task environment, because environmental data is usually incomplete or unavailable. Many reinforcement learning methods are proposed to solve this problem, such as tracking/following [5-7]. For the applications of space-time coverage planning for multi-UAV, Multi-Agent Reinforcement Learning (MARL) is an effective means. A study conducted by Pham et al. [8] proposes a distributed algorithm which is shown to optimally cover a defined FOI. However, in their study, the fields of view (FOV) of each UAV are the same, that means each UAV in the team is equipped with the same sensors, which limits the application range of the system. In this paper, a spatio-temporal coverage planning algorithm named STC-MARL for multi-UAV and multi-sensor cooperative search is proposed, which allows a team of UAV equipped with two types of cameras to learn to provide full view of a FOI while minimizing the overlap.

2. Problem formulation
As shown in Figure 1, in the given field coverage problem, suppose that two UAVs is sent out to perform the task of multi-UAV and multi-sensor cooperative search. Each UAV node is independent, and the system is distributed. Fig. 1 show the FOV of each UAV with half-angles $\theta^f = [\theta_1, \theta_2]^T$.
If a point \( q \) on the ground is covered by the \( i \) th UAV, it can satisfy the following equations:

\[
\frac{|q - c_i|}{z_i} \leq \tan \theta^T
\]

(1)

where \( c_i \) is the projected position, and \( z_i \) is the height of the UAV \( i \) respectively.

In practical missions, different UAV nodes often have fixed heights and carry different, at least two or more sensor types.

3. Algorithm

3.1. Multi-Agent Reinforcement Learning

The multi-agent systems (MAS) can effectively solve the problems in the scenario in Fig.1. The MAS strive for the greatest expected rewards by selecting which action to take that will obtain more higher reward over a long period of time. In Multi-Agent Reinforcement Learning (MARL), the action-state function is represented as:

\[
Q(s_{i,k}, a_{i,k}, s_{-i,k}, a_{-i,k}) = Q(S_k, A_k) = E \left( \sum_{k} \gamma r_{i,k+1} \right)
\]

(2)

The function above is called Q-function, where \( S_k \) is the joint state space and \( A_k \) is the joint action space at time step \( k \), \( s_{i,k} \) and \( s_{-i,k} \) denotes the individual state of agent \( i \) and the rest of the state of other agents, \( a_{i,k} \) and \( a_{-i,k} \) denotes the individual action of agent \( i \) and the rest of the action of other agents, \( 0 < \gamma \leq 1 \) is the discount factor of the learning process.

In our paper, an approximation method named Fixed Sparse Representation (FSR) [4] is employed to reduce the memory requirement of the original MARL. The original Q is mapped by FSR to a parameter vector based on a group of basis functions \( \phi : [S] \times [A] \rightarrow \mathbb{R} \). A column vector \( \phi(S, A) \) of the size \( D \cdot ||A|| \) is used by FSR, and \( D \) denotes the sum of dimensions of the corresponding state space.

The elements in \( \phi \) are defined as:

\[
\phi(x, y) = \begin{cases} 
1, & \text{if } x = S_k, \ y = A_k \\
0, & \text{otherwise}
\end{cases}
\]

(3)

The global reward of the UAVs is defined as follow:

\[
GR(S_k, A_k) = \begin{cases} 
r, & \text{if } \sum_{i} f_i(S_k) \geq f_b, \sum_{i} \alpha_i(S_k) \leq 0 \\
0, & \text{otherwise}
\end{cases}
\]

(4)

where \( f_b \in \mathbb{R} \) is an man-made bound of the covered field. A linear programming (LP) problem is formulated to optimize the stable action as follow:
The learning process is proposed as Algorithm 1.

**Algorithm 1:** Space-time coverage planning algorithm

**Input:** Parameter setting: Discount factor $\gamma$, learning rate $\alpha$, schedule $\varepsilon$, number of step $L$.

**Input:** Basis Function vector $\phi(S,A), \forall s_{i,0} \in S, \forall s_{i,0} \in A_i$

1: Initialize $\theta_{i,0} \leftarrow 0$, $i=1,...,m$

2: for episode = 1, 2, ... do:

3: Randomly initialize state $s_{i,0}, \forall i$

4: for $k = 0, 1, 2, ...$ do:

5: Obtain the other UAVs’ $s_{j,k}$ and parameters $\theta_j, j \neq i, j = 1...m$

6: $\pi(A_k) = \begin{cases} 
\text{Find an optimal joint-action by solving LP(5), with probability } 1-\varepsilon \\
\text{Take a random joint action, otherwise} 
\end{cases}$

7: Select optimal joint action $A_k$, and take corresponding joint action.

8: Obtain other UAVs’ new states $s_{j,k+1}, j \neq i, j = 1,...,m$

9: Obtain global reward $r_{k+1} = GR(S_k, A_k)$

10: Update:

$$\theta_{i,k+1} \leftarrow \theta_{i,k} + \alpha \left[ GR(S_k, A_k) + \gamma \max_{A' \in A} \left( \phi^T(S_{k+1}, A') \theta_{i,k} \right) - \phi^T(S_{k+1}, A_k) \theta_{i,k} \right] \phi(S_k, A_k)$$

**Output:** parameter vector $\theta_i, i = 1...m$ and action policy $\pi$

4. Results and Discussions

A simulation on Pycharm and Matlab [10] environment is set up to verify the effectiveness of the proposed algorithm. The mission space is set as a 3-D space of $7 \times 7 \times 5$, and a FOI with an irregular shape (Fig. 2). The system has 3 UAVs, each UAV is at a fixed height which simulate the low altitude, medium and high altitude UAVs. The FOV angles of the 3 UAVs are $15^\circ, 15^\circ, 30^\circ$, respectively. The learning rate was $\alpha = 0.1$, and the discount rate $\gamma = 0.9$, $\varepsilon = 0.9$, which was diminished over time.
Fig. 2 shows the convergence curve of the proposed algorithm. As can be seen from Figure 2, the algorithm gradually converges after about 100 episodes. Fig. 3 shows the coverage results of 3 UAVs in a stable state of algorithm convergence. It can be seen that after 12 steps, three UAVs complete the optimal coverage of the FOI.

5. Conclusion
Based on the results and discussions presented above, the conclusions are obtained as below:
(1) The proposed spatio-temporal coverage planning algorithm for multi-UAV and multi-sensor cooperative search is effective, which allows several UAVs to complete a full coverage of a FOI.
(2) The FSR technology can effectively reduce the size of state space, reduce the computational complexity of the algorithm, and maintain the accuracy of the model.

Acknowledgments
This work was financially supported by the Aviation Science Foundation (20185551019).

References
[1] Dong Z.L., Liu C.H. (2021). Collaborative Coverage Path Planning of UAV Cluster based on Deep Reinforcement Learning. IEEE International Conference on Frontiers Technology of Information and Computer, 2021.
[2] Chakravarthy, A., Ghose, D. (1998) Obstacle avoidance in a dynamic environment: a collision cone approach. IEEE Transactions on Systems Man & Cybernetics, 28(5):562-574.
[3] Huang L.Z., Liu C.H., Dong Z.L. (2021) Deep Reinforcement Learning Based Collaborative Optimization of Communication Resource and Route for UAV Cluster, IEEE International Conference on Unmanned Systems, 2021.
[4] Sutton, R. S. and Barto, A. G. (1998) Reinforcement Learning: An Introduction, Cambridge, MA, USA: MIT Press.
[5] Borrajo, D., Parker, L. E., et al., (2005) A reinforcement learning algorithm in cooperative multi-robot domains, Journal of Intelligent and Robotic Systems, 43(2-4):161–174.
[6] Rowell, R. X., Wei, C. Ding. (2019) UAV intelligent route planning method based on DQN deep reinforcement learning. Proceedings of China Automation Conference (CAC2019): 6.
[7] Papadimitriou, C. H., Roughgarden, T. (2008) Computing correlated equilibria in multi-player games. J. ACM, 55(3).
[8] Pham, H. X., La, H. M., Feil-Seifer, D., Nefian, A. (2018) Cooperative and distributed reinforcement learning of drones for field coverage. arXiv:1803.07250.