2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Cooperative Aerial Robots Inspection Challenge Based on Hierarchical Reinforcement Learning

Jinwen Hu, Yifei Lei, Deteng Zhang, Zhao Xu, Kexin Guo, Chunhui Zhao, Yang Lu, Xiaolei Hou.

***Abstract*— The poster proposes a multi-UAV reinforcement learning strategy to train UAVs to get more points of interest scores. A reinforcement learning-based approach to a manoeuvre strategy for multi-drone hierarchical training is designed, where the exploration area is divided into multiple parts by the number of drones and explored separately in each area. The reward function of each intelligence is designed based on its total score and whether it collides with obstacles or not. The final goal is to get more scores.**

I. METHODS

A. Map Construction

Scanning and generating global maps using LIDAR-containing drones. The UAV carrying a LiDAR is used to move on a preset trajectory and scanned to obtain a global point cloud map for subsequent UAV obstacle avoidance work.

B. Regional Allocation

Divide the entire area into multiple sections for drones to explore one by one, as shown in Figure 1.

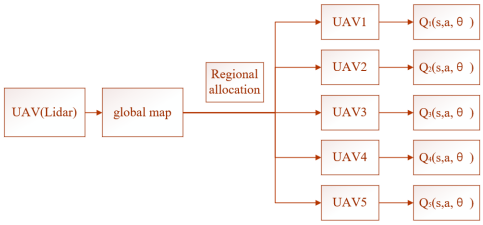

Fig. 1. Hierarchical Framework

Due to the large size of the map and the number of drones, a spatial division method is used to divide the entire area in terms of height, and each area is explored by a single drone individually in order to reduce the risk of drones colliding with each other. Layering by height is shown in Figure 2.

C. Reinforcement learning parameters

After the division of the region is over, each UAV only needs to explore the part within its own region, which we train separately using reinforcement learning algorithm.

Use angle and position to describe the state space of UAVs, being defined as $S = [x, y, z, \theta]$.

The action space is the UAV's manoeuvre library. The action output $a \in A$ of the UAV is the acceleration of the UAV in the three dimensions of space, and the action space can be represented as $A = [a_x, a_y, a_z]$.

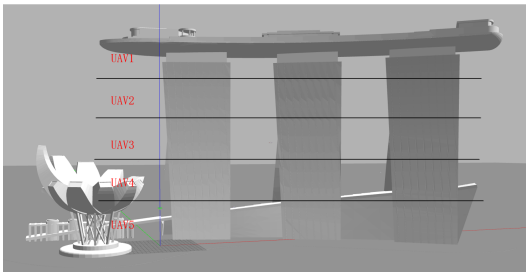

Fig. 2. Spatial Stratification

The reward function consists of both encouragement for the correct behaviour of the actor and punishment for the wrong behaviour. The reward for each time step is defined as a simple number based on the final goal and obstacle avoidance requirements. Our main goal is to get more rewards as well as to avoid colliding into obstacles, the reward function is expressed as

$$R = \begin{cases} Q_{k+1} & Q_{k+1} > Q_k \\ -Q_k & \text{collision} \end{cases}$$

where Q_k indicates the cumulative score obtained by the current drone at step k. and Q_{k+1} represents the cumulative score of the drone at the next moment after the drone has taken action.