Paper ID: 76241
interactive
student

IAF SPACE OPERATIONS SYMPOSIUM (B6)
Mission Operations, Validation, Simulation and Training (3)

Author: Dr. Siyao Lu
Beijing Institude of technology, China

Prof. Rui Xu
Beijing Institute of technology(BIT), China
Dr. Dengyun Yu
China Aerospace Science and Technology Corporation (CASC), China
Ms. Zhaoyu Li
Beijing Institude of technology, China
Dr. Ai Gao
Beijing Institute of Technology, China
Mr. Bang Wang
Beijing Institude of technology, China
Mr. Bo Pan
China

HIERARCHICAL REINFORCEMENT LEARNING BASED PLANNING METHOD WITH UNCERTAINTY IN LIMITED VISIONS FOR LUNAR ROVERS

Abstract

China and Russia will establish the International Lunar Research Station together in around 2035 when the resource acquisition from the lunar surface and the construction of lunar bases will be applied at the beginning period. However, the accuracy of the lunar surface digital elevation map (DEM) is only 10m currently, which cannot meet the needs of path planning or acting for lunar rovers. What's more, there are limitations to the vision of each rover compared to the wide moon surface, so rovers are required to foresee the obstacles and adjust movements for obstacle avoidance, power saving, and safety. Another problem is that the acquisition and construction are long-term tasks so pure path planning methods won't work properly. Therefore, we propose a new way of planning both the path and the task by hierarchical reinforcement learning, where hundreds of simulation environments in which the obstacles and places for acquisition, charging, blending, and construction are varied. Rovers can only obtain a vision of several meters and they will only know the approximate locations of targets. So, uncertainty occurs during the rovers' way to the targets, on which there are small and large obstacles. Targets will be given by the task level from which the guidance will be applied on the path level. However, data on the task level generated by the hierarchical environment is not enough for training the task policies so pre-generated data will be prepared for the pre-training of the task policies then the policies will be set on the task level with constraints while updating rather than joint training from the beginning. In our experiment, we intend that our way leads each rover to finish the long-term tasks without meeting large obstacles, trains the whole hierarchal policy more quickly than the traditional way, and generates a better result than pure path planning in the uncertainty environment for long-term tasks.