1 Motivation

Conventional disaster exploration robots often have fixed modes of movement. For instance, many disaster rescue robots cannot switch between different modes of movement when faced with complex terrain, which limits their practicality in disaster relief scenarios. Fortunately, the new Multi-Modal Mobility Morphobot (M4) provides a more flexible solution. M4 can perform various types of locomotion, such as rolling, walking, crouching, and flying, by repurposing its multi-functional components: wheels, legs, and thrusters, to suit the situation [1]. This project aims to design and train a decision-making system that helps M4 to select the best mode of movement to overcome various terrains and obstacles.

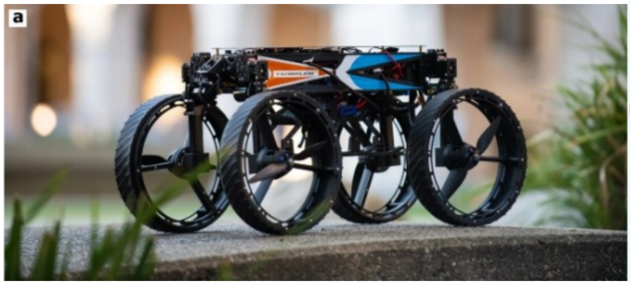


Figure 1: M4 [1]

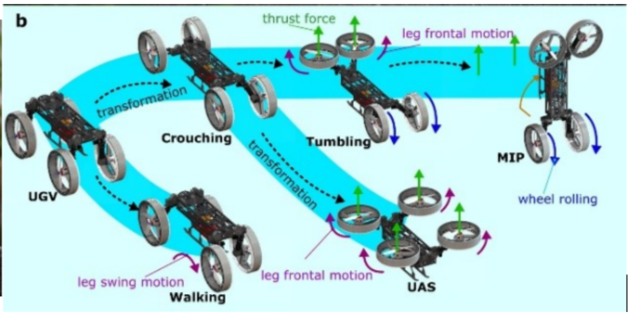


Figure 2: M4 transformation [1]

2 Conventional study

In the current M4 project, the MM-PRM and A^* algorithms have shown promising results in path planning and navigating through various environments. However, the scenarios these algorithms address are relatively straightforward, which may limit their flexibility in handling more complex situations. Since these algorithms rely on predefined heuristics, M4 might not always be able to select the most optimal mode of action when presented with multiple options. For example, when M4 operates in UAV mode and encounters a long downhill, it decides: should it remain in UAV mode to descend, or switch to UGV mode to complete the task more energy-efficiently? These more intricate environments pose challenges that extend beyond the current capabilities of the algorithms. As such, exploring deep reinforcement learning approaches that can learn and predict optimal strategies in response to complex and dynamic environments may provide a more robust solution for M4.

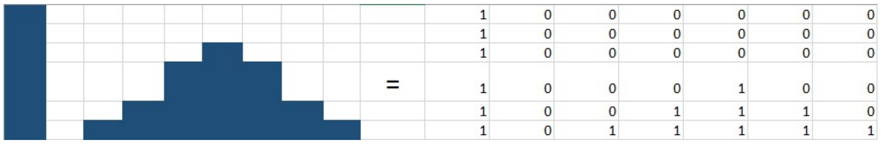
3 Problem statement

Model the world as an $N \times M$ matrix, where N represents the height and M represents distance, with a manually designed environment that includes high obstacles, slopes, stairs, and low-clearance openings. M4 will move from 0 to M while prioritizing energy efficiency. It can only obtain depth information in the form of an $N \times 1$ matrix, where each number in N ranges from 0 to d, representing the distance to obstacles at various heights. This simplified depth input allows M4 to make real-time decisions about the best movement mode. For this training purpose, this project considers using Soft Actor-Critic (SAC) as a reinforcement learning algorithm.

3.1 RL cast

- Learning method: PPO.
- State space:
  1. One matrix: the agent's distance from the obstacle is represented by an $(N \times 1)$ Matrix, where each element ranges from 0 to d. This matrix represents the agent's perception capability in the real world, such as LiDAR.
  2. Four scalar values:
     (a) Agent's field of view distance: d;
     (b) Position of agents: pos(x, y);
     (c) Remaining power: power.
- Action space:

  Each of the two actions – running, walking, crawling, and flying – switches to three other states respectively.
- Rewards structure:
  1. Positive rewards:
     (a) Pass an obstacle (+2 to +5);
     (b) Complete the whole task (+8 to +10).
  2. Negative rewards:
     (a) Collision (-4 to -6), whose reward must have an absolute value greater than the reward for a successful pass;
     (b) Transformation to another mode will get different negative rewards. In different movement modes, the M4 robot will consume different amounts of energy per step, according to Appendix Table 1;
- Environment: The environment for this project is the disaster scene, which can be expressed by, for example, a 6 by N matrix. Digit 1 in the matrix is the obstacle, and 0 means the free space, as figure 3 shows.



1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	1	0	0
1	0	0	1	1	1	0
1	0	1	1	1	1	1

Figure 3: Environment example

References

[1] E. Sihite, A. Kalantari, R. Nemovi *et al.*, “Multi-modal mobility morphobot (m4) with appendage repurposing for locomotion plasticity enhancement,” *Nature Communications*, vol. 14, p. 3323, 2023. [Online]. Available: https://doi.org/10.1038/s41467-023-39018-y

A Appendix

Power consumption expression:
The power consumption of each state is in positive proportion to the moving distance in the corresponding state as the equation 1 shows, the y is the power consumption in the state, x is the moving distance in the state and A is the proportionality coefficient. Different states have different proportionality coefficients. b indicates the power cost of the actions.

$$y = Ax + b \tag{1}$$

new mode pre mode	fly	crawl	run	walk
fly		1 2	1 3	2 4
crawl	4 2		1 1	2 2
run	4 3	1 1		2 1
walk	4 4	1 2	1 1	

Table 1: energy consumption index table