Object Rearrangement in Clutter for Mobile Manipulator Using Hybrid Soft Actor-Critic Method

Jinhwi Lee[1,3], SeungHyun Kang[1,2], Changhwan Kim[1]

[1] Korea Institute of Science and Technology, Seoul, Korea
[jinhooi, ckim]@kist.re.kr

[2] Hyundai motor group, Uiwang, Korea
7328117@hyundai.com

[3] Department of Mechanical Engineering, Hanyang University, Seoul, Korea

EXTENDED ABSTRACT

1 Introduction

Bringing an object to people is an essential service in robot manipulation. A robot is often requested to plan an obstacle rearrangement task and execute it to grasp a target object when objects are placed in narrow spaces, such as cupboards, refrigerators, and shelves. Figure 3 shows an example environment with objects on the shelf and a mobile manipulator. The robot system has a multibody system as a combination of a manipulator with multiple DoF, a mobile device, and a vision system. In such environments, objects could be grasped from the side or front because the boards obstruct the manipulator from grasping from the top. Especially when objects stand around a target object in the environment, a robot must move obstacles aside and grasp the target object. In this kind of dense environment, a task and motion planner should determine which objects to move and where to relocate them before grasping a target.

2 Method

We employ one of the actor-critic methods because it deals with the continuous action space by learning the policy itself, not calculating the state or Q-value function. However, such a value-based method as DQN evaluates the state or Q-value function using the Bellman equation, which is hard to be calculated in the continuous space. In our problem, we apply *hybrid Soft Actor-Critic (hybrid SAC)* in [1] to simultaneously consider the two characteristics of the object rearrangement task and motion planning (OR-TAMP) problem. As shown in Fig. 1, the actor returns the mean and the standard deviation from a state s identical to the standard SAC algorithm in [2]. The actor also returns a discrete value by sharing the hidden layer, which is different from the standard SAC.
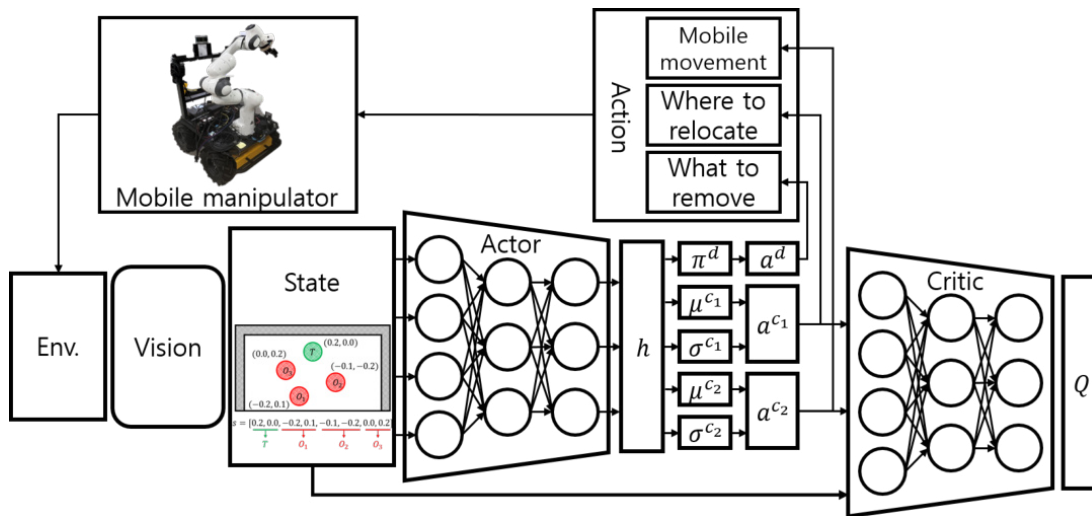

Figure 1: The structure of proposed method. The actor shares the hidden layers h to return one discrete distribution π^d, the means $\mu^{c_{1,2}}$ and the standard deviations $\sigma^{c_{1,2}}$ of the continuous component. The discrete action a^d (which obstacles to move in the OR-TAMP problem) is sampled from π^d. The continuous actions, a^{c_1} (where to relocate the obstalces) and a^{c_2} (mobile movements), are computed from their mean and standard deviation vectors by using standard normal noise and a tanh non-linearity. The critic takes continuous action results from the actor and the state representation. More details on hybrid SAC are referred into [1].

3 Experiments

We compare the simulation results of the proposed method with those of the method suggested in [3] as a baseline. For comparing the two methods, the same object configurations are used, and the outputs from both methods are the actions of rearrangement (i.e., which obstacles to move and where to relocate them). We perform the experiments with the 50 random instances for each of $N = 5, 10, 15$, where N is the number of objects. The results are shown in Fig. 2. The proposed method relocates fewer obstacles and shows higher success rates than the baseline method. Especially for a very dense environment due to many objects in a fixed workspace, the baseline method may often fail to find feasible relocation positions, and it could not grasp the target object finally.
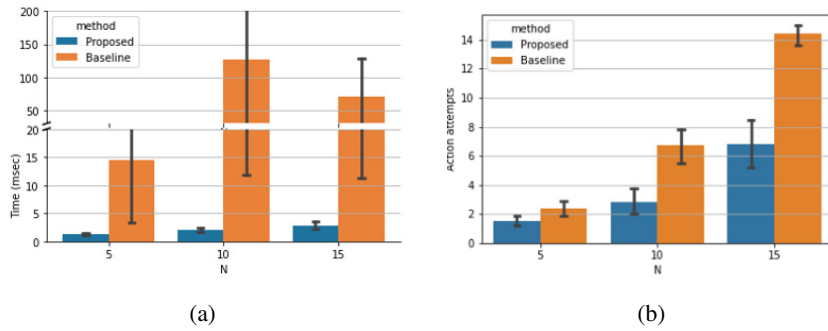
Figure 2: Simulation results of planning time (msec) when the episode is successful (a) and the number of action attempts until the episode ends (b).
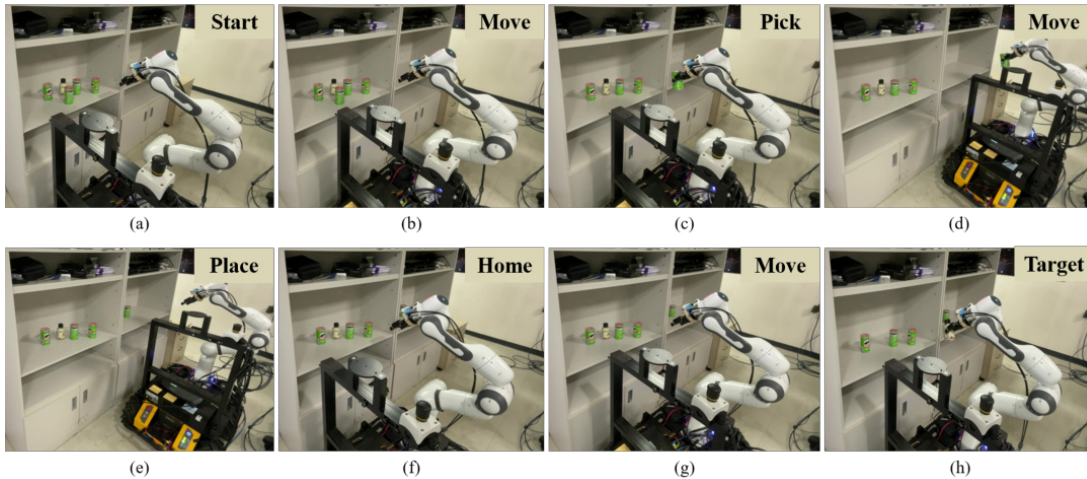

Figure 3: The action sequence from the proposed method in a real robot experiment. The robot starts in front of the shelf as (a). The robot follows the action sequence as (b) mobile movement for grasp the obstacle, (c) grasping the obstacle, (d) mobile movement for relocating the obstacle, (e) relocating the obstacle, (f) mobile movement to home, (g) mobile movement for grasping the target, and (h) grasping the target.

4 Conclusions

We propose a reinforcement learning model based on hybrid SAC to deal with the two characteristics of the OR-TAMP problem: One is to determine which obstacles to move among multiple objects, which is formulated in a discrete space, and another is to obtain where to relocate the obstacles in the continuous workspace. The method works faster than the baseline and plans more feasible actions. We observe that the method attempts more actions for rearrangement in denser object configurations as expected. The actions planned by the method are executed using a real robot within acceptable runtime.

Acknowledgments

References

[1] Delalleau, Olivier and Peter, Maxim and Alonso, Eloi and Logut, Adrien. Discrete and continuous action representation for practical rl in video games. arXiv preprint arXiv:1912.11077, 2019.

[2] Haarnoja, Tuomas and Zhou, Aurick and Abbeel, Pieter and Levine, Sergey. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. International conference on machine learning, pages 1861–1870, 2018.

[3] Lee, Jinhwi and Cho, Younggil and Nam, Changjoo and Park, Jonghyeon and Kim, Changhwan. Efficient obstacle rearrangement for object manipulation tasks in cluttered environments. International Conference on Robotics and Automation (ICRA), pages 183–189. 2019.