삽 기반 조작 동작과 메타 정책을 통한 사족보행 로봇의 물체 수집 전략 학습

백찬우[0], 이윤상
한양대학교 컴퓨터·소프트웨어학과
{bcw0430, yoonsanglee}@hanyang.ac.kr

Learning Object Collection Strategies for a Quadruped Robot via Shovel-Based Behaviors and Meta-Policies

Chanwoo Baek[0], Yoonsang Lee
Dept. of Computer Software Engineering, Hanyang University

요약

본 연구는 사족보행 로봇에 삽(shovel)을 장착하여 바닥에 있는 물체를 짐칸에 수집하고, 이후 몸을 기울여 물체를 지정된 수집 통에 옮기는 연속 동작을 학습하는 것을 목표로 한다. 이전 연구에서는 삽을 통한 수집까지만 가능했으나[1], 본 연구에서는 로봇의 자세 조절을 통해 수집한 물체를 외부 공간으로 정확히 투하하는 동작까지 포함한 트랜스포팅 루틴(transporting routine)을 학습한다. 이를 위해 접근(π_approach), 수집(π_st), 투하(π_dump)의 세 정책을 메타 정책(π_meta, meta policy)를 통해 유기적으로 전환하며, 시뮬레이션 기반의 강화학습 환경에서 이를 실현한다.

1. 서론

기존의 로봇 기반 물체 수집 및 이송 작업은 주로 로봇 팔과 같은 정밀 조작 장비를 필요로 하며[2], 에너지 효율성과 기계적 복잡도 측면에서 한계가 있다. 본 연구에서는 이러한 한계를 극복하고자 사족보행 로봇에 삽 구조를 부착하고, 사지의 조작만으로 바닥에 있는 물체를 수집하고 외부 통에 이송하는 연속적 조작 루틴(Sequential Manipulation Routine)을 강화학습 기반으로 학습시킨다. 특히 이전 연구에서는 삽을 이용해 물체를 짐칸에 싣는 작업까지만 가능했으나[1], 이번 연구에서는 그 후속 동작으로 몸을 기울여 물체를 외부 수집 통에 붓고, 자세를 복원한 뒤 다시 다음 물체를 수집하러 이동하는 일련의 절차를 수행한다. 이를 위해 동작별로 구분된 세 가지 정책(접근, 수집, 투하)을 구성하고, 상황에 따라 적절한 정책의 동작을 수행하도록 하는 메타 정책 구조를 설계하였다.

* 구두(포스터) 발표논문
* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회 (논문지)에 제출 준비중임.
* 본 연구는 xxx 지원으로 수행되었음.

2. 제안하는 방법

2.1. 정책 구조

본 연구에서는 사족보행 로봇의 연속적 물체 운반 루틴(Sequential Transporting Routine)을 구현하기 위해, 세 가지 개별 정책을 구성하고 이들을 메타 정책 구조로 통합하는 방식을 설계하였다.

- π_approach: 로봇이 지정된 목표 지점에 접근하도록 학습된 정책으로, 이동 중 로봇의 정면이 진행 방향과 정렬되도록 요(yaw)를 조절하며 전진한다. 따라서 π_approach는 로봇이 이동 경로를 따라가며 자신의 방향(orientation)을 조정할 수 있도록 한다.
- π_st(scoop-toss): 물체를 쳐서 삽 안쪽으로 굴려 담은 뒤, 다리 관절의 빠른 움직임을 통해 물체를 짐칸에 던지는 정책이다. 물체 던지기 동작은 고정된 삽에 대해 로봇 관절의 움직임을 조정하여 수행되며, 이를 통해 물체가 짐칸 방향으로 정확히 투사되도록 유도한다.
- π_dump: 로봇의 상체를 들어올리고 뒷부분을 낮추는 동작을 통해 몸을 뒤로 기울여, 짐칸의 물체를 수집 통에 붓는 정책이다. 물체 낙하를 인식하면 로봇은 기본 자세로 복원하며 루틴을 반복한다.

이러한 세 정책은 향후 메타 정책 구조 하에서 상황에 따라 전환되며, 전체 운반 루틴을 수행하게 된다.

2.2. 단계별 학습 과정

본 연구에서는 전체 운반 루틴 중에서도 π_dump 정책

의 안정적인 학습을 우선적으로 구현하였으며, 이를 위해 2단계 학습 방식을 도입하였다. 이와 같은 학습 분할은 초기 단계에서의 동작 안정성 확보와 후속 동작의 효율적인 수렴을 유도하기 위한 것이다. 본 절에서는 현재까지 중점적으로 다룬 π_dump의 학습 과정과 이에 적용된 리워드 구조를 설명한다.

1단계에서는 로봇이 짐칸에 담긴 물체를 담은 채 후방으로 이동하는 동작을 먼저 학습한다. 이러한 후방 이동 동작의 학습 과정을 통해 피치(pitch)를 올리는 방향의 움직임을 유도하고, 기울어짐에 대한 기본 동작 감각을 익히도록 설계하였다.

2단계에서는 상체를 들어올려 피치를 증가시키는 동작을 학습하며, 실제 물체를 투하하기 위한 자세를 형성하는데 집중한다.

2.3. 보상(reward) 구성

본 절에서 설명하는 보상 구조는 현재까지 구현 및 학습이 진행된 π_dump 정책에 적용된 부상 구조를 중심으로 한다.

1단계 부상은 로봇이 후방으로 이동하는 동작을 먼저 학습하도록 유도하기 위해, 이동 거리 증가에 따른 선형 보상으로 구성되었다. 이 단계의 목적은 피치 자세 자체를 학습하는 것이 아니라, 2단계에서 피치 자세 탐색이 더 효과적으로 이루어질 수 있도록, 초기 자세에서의 물리적 진입 경로를 확보하는 것에 있다. 실제로 로봇이 평평한 자세에서 곧바로 피치를 크게 올리려 하면 실패 확률이 높고 탐색이 제한되지만, 후진 동작을 통해 로봇이 뒤로 이동하는 관성이 생긴 상태에서는 기울임 동작이 더 자연스럽고 안정적으로 시작될 수 있다.

2단계 보상은 1단계에서 학습된 후진 동작이 정책의 초기 행동에 영향을 주어, 피치 기울임에 유리한 탐색 경로를 자연스럽게 유도한다는 점을 활용하여 설계되었다. 이 단계에서는 로봇이 상체를 들어올려 피치 값을 크게 증가시키고, 일정 시간 동안 그 기울기를 유지하는 동작을 학습한다. 피치는 0°(수평)에서 −90°(수직)에 해당하는 범위를 0~1 사이로 정규화하고, 이를 45개의 구간으로 균등하게 분할하였다. 각 구간에 도달할 때마다 보너스(bonus) 보상이 점진적으로 증가하도록 설계되어, 더 깊은 기울기에 도달할수록 더 큰 보상이 주어진다. 보너스 보상은 구간의 인덱스(index)를 기반으로 선형으로 증가하며, 이를 통해 높은 피치 각도 도달을 강하게 유도하였다. 또한 목표 피치 구간에 도달한 이후에도 불안정한 자세로 인해 다시 범위 밖으로 이탈하는 행동을 억제하기 위해, 직전 스텝과 비교했을때 탈락하는 구간이 있을 경우 해당 구간에서 제공되던 보너스의 1.1배에 해당하는 값을 보너스 보상에서 빼는 패널티(penalty)를 적용하였다.

3. 실험

본 연구에서는 π_dump 정책의 학습 성능과 보상 구조의 효과를 검증하기 위해 Isaac Gym(아이작 짐) 기반 시뮬레이션 환경에서 실험을 수행하였다.

기존 연구[1]에서는 π_approach와 π_st를 연계한 메타 정책을 학습하여 반복적인 수집 루틴을 구성하였으나, 본 연구는 후속 정책인 π_dump 단일 동작의 안정적인 수행에 초점을 맞추었다. 학습된 π_dump 정책은 로봇이 정지 자세에서 상체를 기울여 물체를 외부 수집 통에 투하하는 동작을 성공적으로 수행하며, 이 과정을 시각화한 결과를 그림 1에 정리하였다.

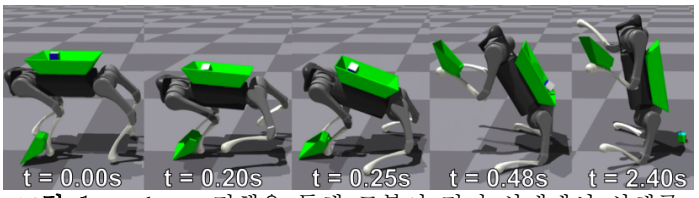

그림 1: π_dump 정책을 통해 로봇이 정지 상태에서 상체를 기울여 물체를 붓는 일련의 동작
(t는 기준 시점으로부터의 경과 시간(초))

4. 결론 및 향후 연구

본 연구에서는 사족보행 로봇이 물체를 짐칸에 담은 뒤 몸을 기울여 외부 수집 통에 투하하는 동작을 학습하는 π_dump 정책을 설계하고, 이를 Isaac Gym(아이작 짐) 기반 시뮬레이션 환경에서 구현하였다. π_dump는 기존 연구에서 제안된 π_approach 및 π_st 정책과 연계되어 전체 운반 루틴을 구성하는 메타 정책의 후속 동작으로 작동한다. 메타 정책을 통해 세 가지 하위 정책 간 전환이 자동으로 이루어지며, 이 중 π_dump는 루틴의 종단 동작으로서 작업의 완성도를 결정짓는 중요한 역할을 수행한다.

향후 연구에서는 우리의 시뮬레이션 기반 정책 전환을 실물 환경에서의 정책 전환으로 확장할 계획이다. 이를 위해 RGB 카메라와 비전 트랜스포머(ViT, Vision Transformer) 인코더를 활용하여 물체의 정보를 추정할 수 있도록 설계하기를 구상중이다. 이러한 후속 연구를 통해, 사족보행 기반 시스템의 활용도를 한층 더 향상시킬 수 있을 것으로 기대된다.

참고문헌

[1] M. Kang, C. Baek, and Y. Lee, Scoop-and-Toss: Dynamic Object Collection for Quadrupedal Systems, arXiv preprint arXiv:2406.12345, 2024.
[2] F. Zhang, L. Xiao, Y. Wang, and R. Xiong, Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion, in Proc. Conf. Robot Learning (CoRL), 2022.