RL-SAR을 활용한 여러 강화학습 프레임워크 정책 실행 및 성능 비교

경희대학교 기계공학과
석사 3기
윤지원

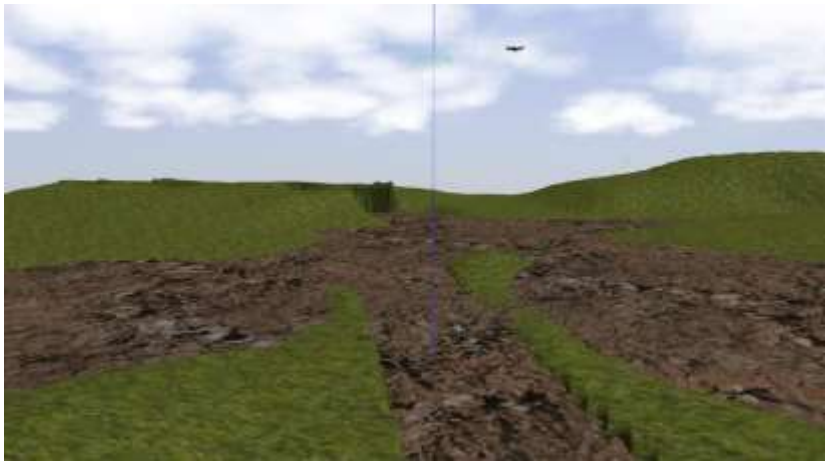
2025. 11. 00

rl_sar 소개

배경

기존의 바퀴 달린 로봇이나 전통적인 내비게이션(경로 계획) 방식은 평평하고 정돈된 실내 환경에서는 잘 작동합니다.

하지만 지진 현장, 무너진 건물, 산악 지형 등 비정형 환경에서는 한계가 명확합니다. 지면이 고르지 않고, 장애물이 불규칙하며, 지도가 없기 때문입니다.



필요성

이런 복잡한 환경에서는 미리 모든 경로를 계획하는 것이 거의 불가능합니다.

대신, 로봇이 스스로 넘어지지 않고 걷는 법을 학습을 통해 터득하게 만듭니다. -> Policy 활용

시뮬레이션에서 학습을 하고 실제환경에서 테스트하는 과정은 상당히 복잡합니다.
-> 이 강의에서는 시뮬레이션으로만 테스트

rl_sar(Simulation And Real) 소개

핵심 개념

- 여러 학습 프레임워크에서 학습된 정책(policy)을 시뮬레이션 환경에 Deployment 하여 잘 작동하는지 검증하고, 실제 로봇 하드웨어에 최종 Deployment 합니다.

Support List:

Robot Name (rname:=)	Pre-Trained Policy	Gazebo	Mujoco	Real
Unitree-A1 (a1)	legged_gym (IsaacGym)	■	✕	■
Unitree-Go2 (go2)	himloco (IsaacGym) robot_lab (IsaacSim)	■	■	■ ■
Unitree-Go2W (go2w)	robot_lab (IsaacSim)	■	■	■
Unitree-B2 (b2)	robot_lab (IsaacSim)	■	■	●
Unitree-B2W (b2w)	robot_lab (IsaacSim)	■	■	●
Unitree-G1 (g1)	robomimic/locomotion (IsaacGym) robomimic/charleston (IsaacGym) whole_body_tracking/dance_102 (IsaacSim) whole_body_tracking/gangnam_style (IsaacSim)	■	■	■
FFTAI-GR1T1 (gr1t1) (Only available on Ubuntu20.04)	legged_gym (IsaacGym)	■	✕	●
FFTAI-GR1T2 (gr1t2) (Only available on Ubuntu20.04)	legged_gym (IsaacGym)	■	✕	●
zhinao-L4W4 (l4w4)	legged_gym (IsaacGym)	■	✕	■
Deeprobotics-Lite3 (lite3)	himloco (IsaacGym)	■	✕	■
DDTRobot-Tita (tita)	robot_lab (IsaacSim)	■	✕	●

핵심 원리

- 여러 학습 프레임 워크로 학습을 완료하면 학습된 .pt 파일이 생성됩니다.
- rl_sar은 .pt 파일을 load 하게 됩니다.
- 이 정책을 ROS2시스템에 연결합니다.
- 로봇의 실제 센서(IMU, 관절 센서 등) 값을 ROS 토픽에서 받아와 학습된 정책 신경망에 입력합니다.
- 정책 신경망이 계산한 출력값(모터 제어 명령)을 다시 ROS 토픽으로 발행하여 실제 로봇이나 Gazebo 시뮬레이션의 로봇을 움직이게 합니다.

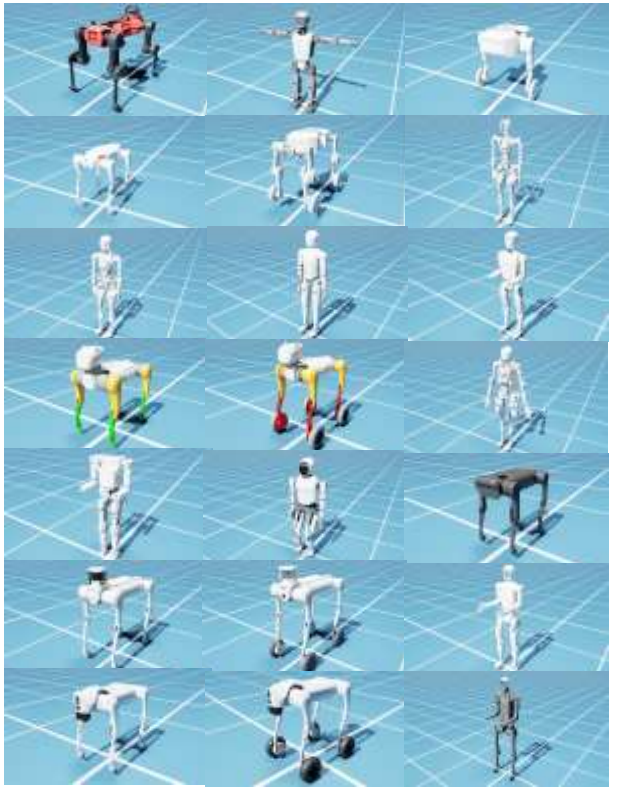
목표

- 성능 테스트를 위한 rl_sar 패키지 실행(policy 성능 테스트를 위한 gazebo 환경 셋팅)
- 여러 학습 프레임워크 튜토리얼
- 각종 학습 프레임워크 Policy 적용 , rl_sar를 활용하여 결과비교

Robot_lab

핵심 개념

- IsaacLab을 기반으로 하는 확장된 라이브러리로, 로봇 강화학습 알고리즘을 개발하고 테스트할 수 있게 해줍니다.
- 다양한 확장 기능 라이브러리 - IsaacLab의 안정성을 해치지 않으면서 자신만의 새로운 로봇이나 알고리즘을 쉽게 추가하고 관리할 수 있습니다.



다양한 로봇 지원

- Anymal D, Unitree (Go2, B2, A1), Deeprobotics Lite3, Zsibot ZSL1, Magiclab MagicDog(사족보행)
- Unitree (Go2W, B2W), Deeprobotics M20, DDTRobot Tita, Zsibot ZSL1W, Magiclab MagicDog-W(바퀴형 로봇)
- Unitree (G1, H1), FFTAI (GR1T1, GR1T2), Booster T1, RobotEra Xbot, Openloong Loong, RoboParty ATOM01, Magiclab (MagicBot-Gen1, MagicBot-Z1)(휴머노이드)

목표

- RSL-rl 기본 학습
- 학습된 정책 테스트
- RL_sar 연계 테스트

HIMloco

핵심 개념

- 복잡한 휴머노이드 로봇을 여러 하위 에이전트 (예: 상체, 왼다리, 오른다리)로 분할하여 제어 문제를 단순화하였습니다.
- 상위 레벨 정책 (High-level policy)이 전반적인 목표를 설정하고, 하위 레벨 정책 (Low-level policies)이 각 신체 부위의 세부 관절을 제어하는 계층적 구조를 가집니다.

주요 특징

- 휴머노이드의 높은 자유도(DoF)와 복잡한 동역학 문제를 여러 개의 작은 문제로 나누어 학습 효율을 높입니다.
- 각 에이전트(신체 부위)를 모듈처럼 다룰 수 있어, 특정 부위의 정책만 수정하거나 새로운 동작을 추가하기 용이합니다.
- 상위 정책이 하위 에이전트 간의 협력을 조율하여, 전체 로봇이 안정적이고 일관된 동작을 생성하도록 유도합니다.

목표

- HIMloco a1 학습
- RL_sar 연계 테스트

HYBRID INTERNAL MODEL: LEARNING AGILE LEGGED LOCOMOTION WITH SIMULATED ROBOT RESPONSE

Junfeng Long[1]*, Zirui Wang[1,2]*, Quanyi Li[1], Jiawei Gao[1,3], Liu Cao[1,3], Jiangmiao Pang[1]✉
[1]OpenRobotLab, Shanghai AI Laboratory, [2]Zhejiang University, [3]Tsinghua University

Figure 1: Our locomotion policy can drive robots to walk across any terrain under any disturbances. Key insight lies in alternatively estimating environmental dynamics with the response of the robot.

ABSTRACT

Robust locomotion control depends on accurate state estimations. However, the sensors of most legged robots can only provide partial and noisy observations, making the estimation particularly challenging, especially for external states like terrain frictions and elevation maps. Inspired by the classical Internal Model Control principle, we consider these external states as disturbances and introduce Hybrid Internal Model (HIM) to estimate them according to the response of the robot. The response, which we refer to as the hybrid internal embedding, contains the robot's explicit velocity and implicit stability representation, corresponding to two primary goals for locomotion tasks: explicitly tracking velocity and implicitly maintaining stability. We use contrastive learning to optimize the embedding to be close to the robot's successor state, in which the response is naturally embedded. HIM has several appealing benefits: It only needs the robot's proprioceptions, i.e., those from joint encoders and IMU as observations. It innovatively maintains consistent observations between simulation reference and reality that avoids information loss in mimicking learning. It exploits batch-level information that is more robust to noises and keeps better sample efficiency. It only requires 1 hour of training on an RTX 4090 to enable a quadruped robot to traverse any terrain under any disturbances. A wealth of real-world experiments demonstrates its agility, even in high-difficulty tasks and cases never occurred during the training process, revealing remarkable open-world generalizability.

*Equal Contributions. ✉ Corresponding Author. Project page at this URL.