실외 환경에서 Eye-in-Hand 로봇 매니퓰레이터의 Pose-based Visual Servoing 수행을 위한 강화학습 방법론

Reinforcement Learning Approach for the Eye-in-Hand Robot Manipulator to Perform Pose-based Visual Servoing in Outdoor Environments

이정우[1] · 김송우[1] · 정호진[1] · 윤한얼[2]
Jungwoo Lee, Songwoo Kim, Ho-Jin Jung and Han Ul Yoon

[1]연세대학교 일반대학원 전산학과
E-mail: {jw.lee, swkim, hojinj}@yonsei.ac.kr
[2]연세대학교 소프트웨어학부
E-mail: huyoon@yonsei.ac.kr

요 약

본 논문은 실외 환경에서 eye-in-hand 로봇 매니퓰레이터의 pose based visual servoing(PBVS) 수행을 위한 강화학습 방법론에 대해 논한다. 일반적으로 PBVS 방법론은 엔드 이펙터의 자세(pose)를 물체의 목표 자세(desired pose)로 옮기기 위해서 사용된다. 그러나, 실외 노지와 같은 비정형 환경에서는 바람, 광원 위치 등과 같은 불확실성 요소들로 인해 PBVS으로 성공적인 작업 수행이 어려울 수 있다. 따라서, 본 연구에서는 PBVS로 푸는 태스크 중 목표 물체 접근(target object reaching) 태스크를 강화학습으로 해결하는 방법론을 제안한다. 먼저, PBVS 시스템 아키텍처를 소개한다. 다음으로, 일반적인 PBVS 아키텍처에서 visual servo controller와 robot controller에 의해 수행되는 목표 물체 접근 태스크를 MDP 문제로 정의한다. 이후, 앞서 정의된 MDP 문제를 해결하는 강화학습 에이전트를 훈련한다. 최종적으로, 학습된 강화학습 에이전트를 통해, 바람과 광원 위치가 반영된 시뮬레이션 환경에서 PBVS가 잘 동작하는 것을 확인한다.

키워드 : Pose based Visual Servoing, 강화학습, Target Object Reaching, Eye-in-Hand 로봇 매니퓰레이터

1. 서 론

비주얼 서보잉(visual servoing)이란, 카메라를 통해 목표 이미지 특징에 도달할 때까지, 카메라가 달린 매니퓰레이터를 조작하는 것을 말한다. Pose based visual servoing(PBVS)에서는 목표 이미지 특징이 6D 자세가 되어, 목표 6D 자세 도달을 목표로 하는 물체 접근 태스크(reaching task)로 간주할 수 있다[1].

비정형 실외 환경에서는, 바람, 광원 위치 등의 요소들이 목표 물체 6D 자세 파악과 로봇 매니퓰레이터 조작 수행에 있어 불확실성 요소로 작용한다. 강화학습은 탐험(exploration)과 탐색(exploitation) 과정을 통해 획득한 정보를 바탕으로 축적되는 보상의 기대값을 최대화하는 최적 정책을 생성한다. 따라서 불확실성 요소에 강인한 액션을 생성하며, 비정형 환경에서의 목표 물체 접근 수행에 매우 유용하다.

He et al.은 DenseFusion이 적용된 PVN3D를 백본 아키텍처로 하여, RGB 및 depth 이미지로부터 특징을 구하고 목표 물체의 6D 자세를 추정하는 FFB6D 네트워크를 제안하였다[2]. Schulman et al.은 clipped surrogate objective function을 적용하여 적은 데이터로 연속공간 문제를 안정적으로 풀이하는 PPO 강화학습 알고리즘을 보였다[3]. 이 두 연구에서 검증된 결과를 바탕으로, 본 연구에서는 다음과 같이 PBVS를 강화학습으로 풀기 위한 아키텍처를 제안한다. 먼저, 6-DoF 로봇 매니퓰레이터의 손목 카메라로부터 획득된 영상에 FFB6D를 적용하여 목표 물체 6D 자세를 추정한다. 다음으로, 목표 물체 접근 태스크의 markov decision process(MDP)를 정의하여 PPO 기반 강화학습 에이전트를 학습한다. 최종적으로, 학습된 에이전트를 통해 물체 접근 태스크 수행 결과를 확인해본다.

감사의 글: 이 성과는 정부(산업통상자원부)의 재원으로 한국산업기술기획평가원의 알키미스트 프로젝트 프로그램(No. RS-2024-00423702) 및 로봇산업기술개발사업No. 20023014) 지원에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

2. 제안하는 방법론

2.1 PBVS 프레임워크

그림 1은 물체 접근 수행을 위한 PBVS 프레임워크를 보여준다. 먼저, 로봇 매니퓰레이터(UR10e, Universal Robots, Odense, Denmark)의 손목에 장착된 카메라로 부

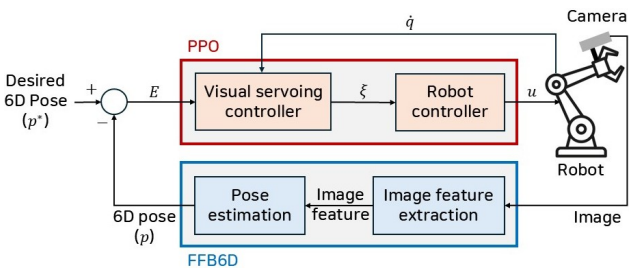

그림 1. 물체 접근 태스크 수행을 위한 PBVS 블록다이어그램

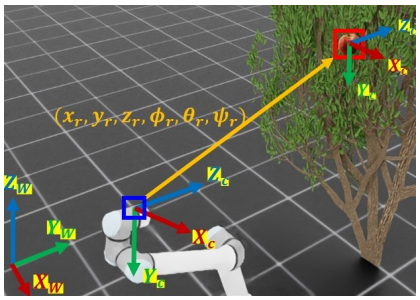

그림 2. 실험 환경의 coordinate 설정

터 이미지를 획득한다. FFB6D 네트워크는 획득된 이미지를 입력으로 받아, 특징 추출을 거쳐 목표 물체의 6D 자세 추정 벡터 p를 출력한다. 현재 FFB6D의 출력으로 나온 p와 물체의 목표 6D 자세 p^*의 차이를 PBVS에서 최소화해야 하는 에러 $E = \|p^* - p\|$로 정의한다.

PPO는 입력 E에 대해 로봇 매니퓰레이터를 조작하는 액션 u를 출력하여, 기존 PBVS의 visual servoing controller와 robot controller 역할을 수행한다. Visual servoing controller는 입력 E를 감소시키기 위한 $\xi := (v, \omega)$를 생성하며, 이는 control law에 따라 카메라를 조작하는 선속도 v와 각속도 ω로 구성된다. 이를 로봇 시스템에 적용하기 위해, robot controller는 ξ를 입력으로 받아 action $u = \ddot{q}_{1:6}$를 생성한다. 최종적으로 이를 이용해 6 DoF 로봇 매니퓰레이터의 조인트 $q_{1:6}$를 조작하고, 업데이트된 카메라 6D 자세로부터 이미지를 획득한다. 이 과정을 반복하여, E를 최소화하도록 PBVS를 수행한다.

2.2 강화학습기반 PBVS 수행을 위한 MDP 정의

강화학습의 에이전트가 주어진 환경에서 최적의 액션 정책을 찾기 위해 MDP를 정의한다. 따라서, E를 최소화하는 u를 출력하기 위한 MDP $M := (S, A, R, P_a, \gamma)$를 정의한다. 각각은 state space S, action space A, reward R, probability transition P_a, discount factor γ로 아래와 같이 정의한다.

$$S = [x_r, y_r, z_r, \phi_r, \theta_r, \psi_r, \dot{q}_{1:6}]^T, \quad (1)$$

$$A = [\ddot{q}_{1:6}]^T, \quad (2)$$

$$R(S,A) = c_p\|P_p - P_p^*\|^2 + c_o\|P_o - P_o^*\|^2 + c_v\sum_{i=1}^{6}\dot{q}^2 + c_a\sum_{i=1}^{6}\ddot{q}_i^2 \quad (3)$$

S는 목표 물체와 엔드이펙터의 상대 6D 자세와 조인트의 각속도, A는 조인트의 각가속도로 정의된다. R은 4개의 term을 가지며, 다음과 같이 정의된다. 엔드 이펙터 position P_p와 목표 물체 position P_p* 차이, 엔드 이펙터 orientation P_o와 목표 물체 orientation P_o^* 차이, 조인트 속도 제어 및 목표 지점에서 정지를 위한 velocity term, 조인트 속도가 급변하지 않도록 하는 control effort term을 가진다. $c_{p,o,v,a}$은 coefficient로, 0 미만의 값을 가진다. P는 Isaac Lab 물리엔진을 따르며, $\gamma \in [0,1)$의 값을 가진다.

3. 실험 및 결과

3.1 실험 소개

학습을 위해 Isaac Lab 환경을 이용하였다. 그림 2는 실험 환경에서 설정한 coordinate을 나타낸다. UR10e와 사과나무를 scene에 배치하고, 목표 6D 자세를 나무 위에 랜덤하게 생성하여 정의한 M을 바탕으로 학습을 진행하였다. 이때 비정형 환경에서의 물체 접근 수행을 위해 S에 노이즈를 추가하여 학습하였다.

실험에 사용한 MDP tuple $c_{p,o,v,a}$, γ값은 각각 c_p=-1, c_o=-0.8, c_v=-0.001, c_a,=-00001 γ=0.99로 설정하였다. 6D 자세 차이보다 조인트의 움직임에 따른 reward 감소가 커지지 않도록, $c_{v,a}$의 크기를$c_{p,o}$보다 작게 설정했다.

3.2 실험 결과

그림 3은 로봇 매니퓰레이터가 목표 물체까지 접근하는 모습을 순서대로 나타낸다. 실험에는 ZED mini 카메라를 사용했으며, 672*376 해상도로 진행하였다. 목표 물체 6D 자세는 카메라에 보이는 범위 내에 랜덤하게 생성했으며, FFB6D로 목표 물체의 6D 자세를 추정하고, PPO를 통해 엔드 이펙터가 목표 물체까지 도달하는 모습을 볼 수 있다.

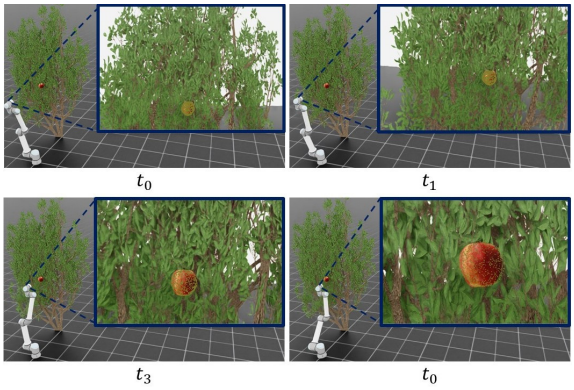

그림 3. 강화학습에서 생성한 액션을 따라 UR10e가 목표 물체(사과) 접근을 수행하는 과정

4. 결론 및 향후 연구

본 논문에서는 실외 환경에서 로봇 매니퓰레이터를 사용하여 PBVS 작업을 수행하는 아키텍처를 제안하였다. 먼저, PBVS는 목표 물체 접근 태스크로 간주할 수 있기 때문에, PBVS를 MDP로 정의하였다. 이후, 정의된 MDP 기반으로 강화학습 에이전트를 훈련하여 가상 실외 환경에서 PBVS 작업을 성공적으로 수행하는 모습을 확인하였다. 향후에는 훈련된 강화학습 에이전트를 실제 로봇에 이식하여 적용해보고자 한다. 또한, PBVS 작업은 최종적으로 수행하고자 하는 행동(behavior)의 첫 번째 과정이므로, 이후 이어지는 다양한 작업에 대해 수행해보고자 한다.

참 고 문 헌

[1] F. Chaumette et al., *Visual Servoing.* Hand-book of Robotics, 2nd edition, Springer, pp.841-866, 2016.

[2] Y. He et al., "Ffb6d: a full flow bidirectional fusion network for 6d pose estimation," in Proc. of *the IEEE/CVF Conf. on CVPR*, online, June. 19-25, 2021, pp. 3003-3013.

[3] J. Schulman et al. "Proximal policy optimization algorithms," *arXiv preprint* arXiv:1707.06347, 2017.