

# 궤적 생성 알고리즘 성능 향상을 위한 Zero-Shot Sim-to-Real 기반 강화학습 최대 속도 및 가속도 결정 알고리즘

지창훈, 최호빈, 최요한, 이재원, 한연희<sup>1)</sup>

한국기술교육대학교 컴퓨터공학과 미래융합공학전공

{koir5660, chb3350, yoweif, asd5742, yhhan}@koreatech.ac.kr

## Trajectory Generation Algorithm Performance Enhancement Using Zero-Shot Sim-to-Real Based Reinforcement Learning for Maximum Velocity and Acceleration Determination

Chang-Hun Ji, Ho-Bin Choi, Yo-Han Choi, Jae-Won Lee, Youn-Hee Han<sup>1)</sup>

Future Convergence Engineering, Dept. of Computer Science and Engineering, KOREATECH

### 요약

본 논문은 무인기의 궤적 생성을 위해 심층 강화학습을 활용한 새로운 알고리즘을 제안한다. 동적 환경에서 실시간으로 궤적의 최대 속도 및 가속도를 최적화하는 이 알고리즘은 Sim-to-Real 방식을 사용하여 가상 시뮬레이션 환경에서 학습한 후 현실 세계에서 검증한다. PX4-ROS2 기반 강화학습 프레임워크를 활용하여 시뮬레이션과 실제 환경에서의 무인기 구동 방식을 통일함으로써, 제로-샷 Sim-to-Real이 가능하다. 실제 환경에서 비교 실험 결과, 제안된 알고리즘을 적용하였을 경우 기존 궤적 생성 알고리즘 대비 궤적의 안정성과 효율성을 향상시키는 것으로 나타났다.

### I. 서론

무인기의 궤적 생성은 고도의 정밀성을 요구하는 작업으로, 최대 속도와 가속도와 같은 파라미터가 궤적의 효율성과 안정성을 결정한다. 이러한 파라미터들의 최적 값은 주변 환경이 바뀌는 동적 환경에서 실시간으로 변화한다. 하지만, 사람이 직접 실시간으로 주변의 환경을 고려하여 최적의 파라미터를 결정하는 것은 불가능하다. 따라서 본 논문은 심층 강화학습을 활용하여 동적 환경에서 실시간으로 궤적 생성 알고리즘의 최적의 최대 속도 및 가속도를 결정하는 새로운 알고리즘을 제안한다.

제안한 알고리즘은 Sim-to-Real 방식에 따라 가상의 시뮬레이션 환경에서 학습 후 현실 세계에서 검증한다. 본 논문은 안정적인 Sim-to-Real을 위해 PX4-ROS2 기반의 강화학습 프레임워크를 추가적으로 제안한다. PX4-ROS2 기반의 강화학습 프레임워크는 시뮬레이션과 실제 환경에서의 무인기 구동 방식을 통일한다. 이를 통해 우리는 실제 환경에서 무인기 구동 시 발생할 수 있는 다양한 변수들을 고려하여 심층 강화학습 모델의 학습을 진행하였다. 그 결과, 본 논문에서 제안한 알고리즘은 실제 환경에서 추가적인 학습이나 조정 없이도 실제 환경에 바로 적용 가능한 제로-샷 Sim-to-Real이 가능하다. 최종적으로 본 논문은 실제 환경에서 기존 궤적 생성 알고리즘에 제안하는 알고리즘을 적용할 경우, 궤적의 안정성과 효율성이 동시에 향상되는 것을 비교 실험을 통해 보여준다.

### II. 본론

본 논문에서는 심층강화학습을 활용하여 궤적 생성 알고리즘의 최대 속도 및 가속도 파라미터를 실시간으로 최적 결정하는 새로운 알고리즘을 제안한다. 또한 PX4-ROS2 기반 강화학습 프레임워크를 제안하여 안정적인 제

로-샷 Sim-to-Real을 가능하게 하였다.

#### 1. 실시간 파라미터 결정 심층 강화학습 알고리즘

강화학습은 행동을 결정하는 에이전트가 주어진 환경과의 상호작용을 통해 에이전트가 얻는 누적 보상 기댓값의 최대화를 위해 학습하는 머신러닝의 한 종류이다 [1]. 강화학습 에이전트는 반복된 시행착오를 통해 해당 분야의 전문 지식이 없이 복잡한 문제를 해결할 수 있다. 특히, 강화학습은 정답 데이터가 없어도 학습이 가능하고 일반화 성능이 높기 때문에 로봇 제어, 게임 등 많은 분야에서 뛰어난 성능을 보여준다.

본 논문에서는 로봇 제어 분야에서 널리 채택되어 사용되는 강화학습 알고리즘 중 하나인 Soft-Actor-Critic [2]에 기반하여 궤적 생성 알고리즘의 최대 속도 및 가속도 파라미터를 실시간으로 결정하는 새로운 알고리즘을 제안한다. 기존 궤적 생성 알고리즘의 최대 속도 및 가속도 파라미터는 궤적 생성이 진행되는 동안 고정된 값을 활용한다. 하지만 해당 파라미터들은 주변 환경과 매우 밀접한 관련이 있어, 주변 환경에 따라 최적 값이 변동된다. 제안하는 Soft-Actor-Critic 기반 실시간 파라미터 결정 알고리즘은 실시간 주변 환경에 따라 최대 속도 및 가속도 파라미터의 최적 값을 결정한다. 이를 통해 우리가 설계한 강화학습 모델에 따라 생성된 궤적의 안정성과 효율성을 향상시킬 수 있다. 환경의 상태는 주변 환경의 장애물 조감도 및 현재 위치부터 목적지까지 직선 경로를 포함하며, 강화학습 에이전트의 행동 결정에 입력으로 활용된다. 결정된 행동은 환경에 적용되며 새로운 상태 및 보상이 에이전트에게 주어진다. 보상은 에이전트에 관계된 다음 조건에 의해 결정된다: 1) 충돌 시 -1, 2) 목적지 도착 시 10, 3) 그 외 -0.1.

1) 한연희(Youn-Hee Han, [yhhan@koreatech.ac.kr](mailto:yhhan@koreatech.ac.kr)): 교신저자

## 2. PX4-ROS2 기반 강화학습 프레임워크

Sim-to-Real의 성능을 높이기 위해서는 매우 세밀하고 정밀하게 학습된 모델이 필요하다. 시뮬레이션과 실제 환경 간의 동작 방식이 다를 경우, 시뮬레이션에서 훈련한 모델은 실제 환경 적용 시 실제 환경의 다양한 변수들을 고려하지 못한다. 예를 들어, 정보 전달의 과정과 그로 인한 지연 시간을 시뮬레이션 상에서 고려하지 않을 경우, 실제 환경에서 시뮬레이션에서의 예측과 다르게 작동할 수 있다. 따라서 시뮬레이션과 실제 환경의 동작 방식을 통일하는 것은 안정적인 Sim-to-Real을 위해 매우 중요하다. 본 논문에서는 시뮬레이션과 실제 환경의 동작 방식을 통일하기 위해 PX4-ROS2 기반 강화학습 프레임워크를 제안한다. 제안하는 프레임워크를 통해 강화학습 에이전트가 훈련을 할 경우 실제 환경의 다양한 변수에 적용할 수 있어 안정적인 제로-샷 Sim-to-Real이 가능하다.

## III. 실험

본 논문에서는 제안된 알고리즘은 대표적인 궤적 생성 알고리즘인 EGO-Planner [3]에 적용하여 로버 자율 주행에 활용한다. 적용된 알고리즘은 PX4-ROS2 기반 강화학습 프레임워크를 통해 Gazebo 시뮬레이션 환경에서 학습을 마친 후 실제 환경에서 비교 실험을 수행하였다. 실제 환경에서 사용된 로버 시스템의 세부 사항은 다음과 같다

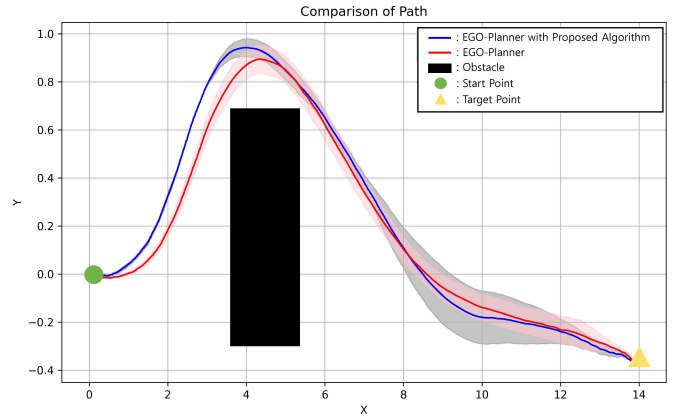
<표 1> 로버 시스템 세부 사항

시스템 구성 요소	모델 명
Antenna for GPS	HX-CH6601A
RC Receiver	Futaba r7008sb
RTKGPS	Piksi Multi
Flight Control Computer	PX4 autopilot
Mission Computer	NVIDIA Xavier NX
Antenna for DDS	CN9 2dbi
Camera	RealSense D455

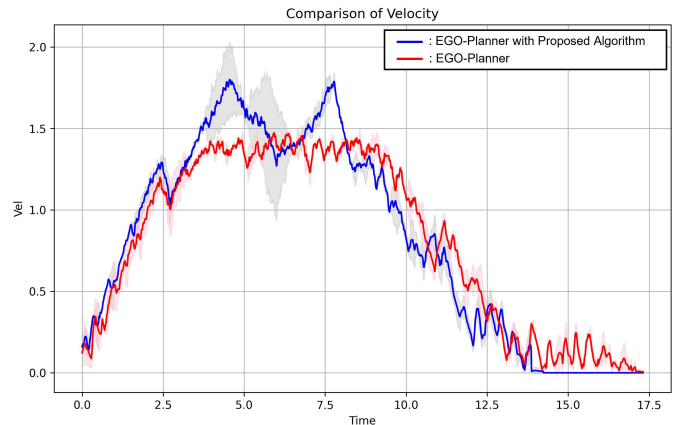
실제 환경에서 EGO-Planner와 제안하는 알고리즘을 적용한 EGO-Planner의 비교 실험을 진행하였다. 각 알고리즘마다 3번씩 실험이 진행되었다. 비교 실험 결과는 [그림 1]과 [그림 2]와 같다. [그림 1]은 실제 환경에서 특정 장애물을 회피하여 목적지에 도착할 때의 경로를 비교한다. 경로를 비교하였을 때, 제안하는 알고리즘을 적용하였을 경우 더 안전한 경로를 생성하는 것을 확인할 수 있다. [그림 2]는 시간에 따른 속도 비교 그래프이다. 제안하는 알고리즘을 적용하였을 경우 주변 환경에 따라 동적으로 속력이 변하는 것을 확인할 수 있다. 직선 구간에 기존 알고리즘보다 속력이 높으며 장애물을 회피할 때 속력이 낮아지는 것을 확인하였다. 또한 장애물 회피가 끝난 후 다시 속력이 증가해 효율적인 경로를 생성하는 것을 확인하였다. 마지막으로 경로의 길이는 제안하는 알고리즘을 적용하였을 경우 안전성을 향상시키기 위해 더 길어지지만, 속도 비교 그래프를 보면 제안하는 알고리즘을 적용하였을 때, 목적지에 더 빨리 도착하여 속력이 0이 된 것을 확인할 수 있다. 우리는 이를 통해 제안하는 알고리즘을 EGO-Planner에 적용하였을 때, 기존 EGO-Planner보다 더 안전하면서 효율적인 경로를 생성하는 것을 보여준다.

## IV. 결론

본 논문에서는 궤적 생성 알고리즘에서 무인기의 최대 속도 및 가속도를 실시간으로 결정할 수 있는 실시간 파라미터 결정 심층강화학습 알고리즘



[그림 1] 실제 환경에서 경로 비교



[그림 2] 실제 환경에서 시간에 따른 속도 비교

을 제안한다. 제안하는 알고리즘은 동적으로 변화하는 환경을 고려하여 무인기의 최대 속도와 최대 가속도를 실시간으로 조정한다. 이때, 결정된 무인기의 최대 속도 및 가속도는 궤적의 효율성뿐만 아니라 안정성도 고려한다. 또한 PX4-ROS2 기반 강화학습 프레임워크를 활용하여 안정적인 제로-샷 Sim-to-Real이 가능하게 하였다. 실제 환경에서 비교 실험을 통해 제안하는 알고리즘을 적용할 경우 기존 알고리즘 대비 더 안전할 뿐만 아니라 효율적인 궤적을 생성할 수 있다는 것을 확인했다.

## ACKNOWLEDGMENT

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2023R1A2C1003143 & No. 2018R1A6A1A03025526)

## 참고 문헌

- [1] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction, MIT Press, 2018
- [2] Haarnoja, Tuomas, et al. "Soft Actor-Critic Algorithms and Applications," arXiv preprint arXiv:1812.05905, 2018.
- [3] Zhou, Xin, et al. "Ego-planner: An esdf-free gradient-based local planner for quadrotors," IEEE Robotics and Automation Letters 6(2), pp.478-485, 2020.