https://doi.org/10.7840/kics.2023.48.10.1238

실시간 계층적 심층강화학습 기반 드론 궤적 생성 알고리즘 파라미터 제어

지 창 훈•, 한 연 희*, 문 성 태°

Real-Time Hierarchical Deep Reinforcement Learning-Based Drone Trajectory Generation Algorithm Parameter Control

Chang-Hun Ji•, Youn-Hee Han*, Sung-Tae Moon°

요 약

최근 드론의 사용이 증가하면서 드론 궤적 생성 알고리즘 연구가 활발히 진행되고 있다. 드론 궤적 생성 알고리즘은 장애물 회피를 고려하면서 실시간 궤적 생성하는 것을 목표로 한다. 드론 궤적 생성 알고리즘 연구는 최근 숲과 같은 복잡한 동적 환경에서 안전하고 효율적인 궤적을 생성하고, 여러 대의 드론을 동시에 제어하는 등 큰 진전을 보여줬다. 하지만, 드론의 안정성을 위해 대부분의 드론 궤적 생성 알고리즘은 드론의 최대 속도와 최대 가속도를 파라미터화 하여 비교적 낮은 수치로 제한한다. 이러한 속도 관련 파라미터 제한은 드론의 효율성과 실용성을 저하한다. 본 논문에서는 계층적 강화학습 기반 실시간 환경을 고려한 드론의 최대 속도 파라미터와 최대 가속도 파라미터를 설정하는 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 제안한다. 제안하는 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘에서 계층적 강화학습의 상위 계층 에이전트는 실시간 환경을 고려하여 최대 속도 파라미터와 최대 가속도 파라미터를 설정하고 하위 계층 에이전트는 이를 활용하여 실시간 궤적을 생성한다. 또한, 최대 속도와 최대 가속도를 설정하는 모든 드론 궤적 생성 알고리즘에 적용 가능하다. 시뮬레이션을 통해 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용한 드론 궤적 생성 알고리즘이 기존보다 뛰어난 성능의 속도, 경로 길이, 경로 부드러움을 가지고 있음을 보여준다.

Key Words : Drone Autonomous Flight, Drone Trajectory Generation Algorithm, Hierarchical Deep Reinforcement Learning

ABSTRACT

With the increasing use of drones, research on drone trajectory generation algorithms has gained significant momentum. These algorithms aim to generate real-time trajectories while considering obstacle avoidance. Recent advancements have shown promising results in generating safe and efficient trajectories in complex dynamic environments, such as forests, as well as controlling multiple drones simultaneously. However, most existing

※ 이 논문은 2023년도 한국기술교육대학교 교수 교육연구진흥과제 지원에 의한 연구결과이며, 또한 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. NRF-2023R1A2C1003143 & RS-2022-00166347)의 연구결과임.
◆ First Author : Future Convergence Engineering, Korea University of Technology and Education, koir5660@koreatech.ac.kr, 학생회원
° Corresponding Author: Department of Computer Science & Engineering, Korea University of Technology Education, stmoon@koreatech.ac.kr, 정회원
* Future Convergence Engineering, Korea University of Technology and Education, yhhan@koreatech.ac.kr, 종신회원
논문번호 : 202308-024-C-RU, Received August 1, 2023; Revised August 11, 2023; Accepted August 11, 2023

drone trajectory generation algorithms impose limitations on the maximum speed and acceleration parameters to ensure drone stability. These restrictions on speed-related parameters hinder the efficiency and practicality of drones. In this paper, we propose a novel approach called “Hierarchical Deep Reinforcement Learning-Based Active Parameter Control Algorithm” that addresses this limitation. This algorithm dynamically sets the maximum speed and acceleration parameters of a drone based on the real-time environment using a hierarchical reinforcement learning framework. The upper layer agent in the hierarchy is responsible for adjusting the maximum speed and acceleration parameters considering the current environmental conditions. The lower layer agent then utilizes these parameters to generate a real-time trajectory. Notably, this approach can be applied to all drone trajectory generation algorithms that involve setting maximum speed and maximum acceleration. Through extensive simulations, we demonstrate that applying the proposed algorithm to drone trajectory generation algorithms results in superior performance in terms of speed, path length, and path smoothness. These improvements showcase the potential of our approach in enhancing the efficiency and overall capabilities of drones operating in complex and dynamic environments.

Ⅰ. 서 론

최근 다양한 분야에서 드론의 사용이 증가하면서, 드론 궤적 생성에 관한 연구가 활발히 이루어지고 있다. 드론 궤적 생성 알고리즘은 장애물을 회피하는 실시간 경로 생성과 동시에 드론의 충돌을 최소화하는 것이 목적이다. 대부분의 드론 궤적 생성 알고리즘들은 위에서 언급한 조건들을 만족하기 위해 드론의 최대 속도와 최대 가속도를 파라미터화 하여 제한하고 있다. 특히, 실시간 드론 궤적 생성 알고리즘들은 비교적 낮은 수치의 속도 관련 파라미터들을 가지고 있다. 예를 들어 실제 동적 환경에서 드론 자율 비행에 성공하여 많은 주목을 받은 EGO-Planner[1]는 최대 속도와 최대 가속도를 낮은 수치로 제한하였다. 이 수치는 숙련자가 직접 조종하는 드론에 비해 매우 낮은 수치임을 알 수 있다. 이런 낮은 수치의 속도 관련 파라미터 설정은 드론의 비행 성능을 상당 부분 저해할 뿐만 아니라, 효율적인 경로의 생성도 방해한다. 하지만 속도 및 가속도를 적절한 수치로 제한하지 않는다면, 드론은 장애물 충돌이나, 기체의 급격한 선회 등 다양한 요소로 인해 비행 안정성을 보장할 수 없다. 따라서 실시간 환경을 고려하여, 드론의 비행 안정성과 비행 성능을 동시에 고려하는 최대 속도 및 가속도 파라미터 설정이 필요하다.

본 논문에서는 계층적 심층강화학습 프레임워크[2, 3]를 활용하여 실시간 바뀌는 환경에 따라 드론 자율 비행 알고리즘들의 최대 속도 파라미터와 최대 가속도 파라미터를 새롭게 설정하는 방법인 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 제안한다. 계층적 강화학습의 상위 계층 에이전트는 SAC (Soft-Actor-Critic)[4] 알고리즘을 활용하여 실시간 바뀌는 환경을 고려해 최대 속도 파라미터와 최대 가속도 파라미터를 설정한다. 하위 계층 에이전트인 EGO-Planner는 설정한 파라미터들을 활용해 드론 궤적 생성 알고리즘으로부터 효율적인 실시간 궤적을 생성한다.

본 논문에서 제안하는 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘은 드론의 최대 속도와 최대 가속도를 제한하고 있는 모든 드론 궤적 생성 알고리즘에 적용할 수 있다. 본 논문에서는 EGO-Planner에 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용할 경우 기존 EGO-Planner보다 속도, 경로 길이, 경로의 부드러움에서 더 뛰어난 것을 보여준다.

본 논문에서 제안하는 방법의 기여점은 다음과 같다:

- 드론 궤적 생성 알고리즘들의 비교적 낮은 수치의 최대 속도와 최대 가속도 제한으로 인한 실용성 및 효율성 감소를 방지하기 위해, 제안한 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 활용하여 실시간 환경에서 안정성과 효율성을 모두 고려한 드론의 최대 속도와 최대 가속도를 설정한다.
- 복잡한 문제를 여러 단계로 나누어 해결하는 강화학습 프레임워크인 계층적 심층강화학습을 활용한다. 전통적인 드론 궤적 생성 알고리즘을 하위 계층 에이전트로 설정하여, 드론 제어의 안정성과 학습의 성능을 동시에 향상시켰다.

2장에서는 관련 연구에 대해 기술하고, 3장에서 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘에 관해 설명한다. 그 후, 4장에서 실험을 통해 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘의 실용성 및 효율성을 보여준다. 마지막 5장에서 결론

을 통해 본 논문을 마무리한다.

Ⅱ. 관련 연구

드론은 사용 초기 군사적 목적으로 시작하였으나, 최근에는 군사적 목적뿐 아니라, 민간에서 농업, 방송 혹은 취미와 같은 여러 분야에서 다양하게 이용되고 있다[6]. 드론의 활용 증가는 자연스럽게 드론 자율 비행을 위한 드론 궤적 생성 연구를 활성화했다. FASTER[7]는 알려진 공간과 알려지지 않은 공간 모두에서 최적화를 진행함으로 효율적인 궤적을 생성하였다. 특히, 안전한 궤적과 효율적인 궤적을 동시에 추출하여, 효율적인 궤적의 안정성이 확보되지 않았을 경우를 대비하였다. EGO-Planner는 유클리드 부호 거리 필드(Euclidean Signed Distance Field)를 고려하지 않은 경사도 기반 최적화를 활용하여 궤적 생성 시간을 크게 단축했다. 또한, A* 알고리즘을 활용하여 생성한 안전한 경로를 활용하여 필요한 장애물 정보만 추출하면서 안전한 경로를 생성한다. EGO-Planner 연구진들은 EGO-Planner를 기반으로 여러 대의 드론, 즉 군집 드론을 제어하는 알고리즘인 EGO-Swarm[8]을 발표하였다. MADER[9]는 군집 드론 자율 비행 알고리즘으로 드론 경로의 각 구간의 외부 다면체 표현을 사용하여 다른 동적 장애물이나 에이전트와 실시간 충돌 회피를 수행한다. 또한, MADER는 최적화에서 다른 드론들의 경로를 제약 조건으로 포함하는 것으로 군집 드론 자율 비행 알고리즘의 안정성을 확보하였다. 앞서 언급한 드론 궤적 생성 알고리즘들은 모두 고정된 하이퍼파라미터가 존재한다. 하이퍼파라미터가 환경에 영향이 크게 있을 경우 하이퍼파라미터의 고정으로 인해, 환경의 일반화 성능이 저하될 수 있다.

강화학습은 학습 동안 피드백을 통해 환경에 적응하며 최적의 행동을 학습하는 기계 학습의 한 종류이다[10]. 강화학습은 지도학습과 다르게 정답 데이터가 필요 없고, 해당 분야에 전문적인 지식이 없어도 최적의 행동을 학습할 수 있다는 장점이 있다. 강화학습의 이러한 장점을 활용하여 드론을 직접 제어하려는 연구가 활발히 이뤄지고 있다[11]. 대표적인 강화학습 기반 드론 자율 비행 알고리즘인 MAPPER[12]는 분산 네트워크를 활용한 군집 드론 제어 및 강화 학습 방법을 제안하여 동적 환경에서 효과적인 로컬 경로 계획 정책을 학습한다. 하지만 강화학습은 시행착오를 통해 학습이 이루어지므로, 초기 학습 단계에서는 안정성이 낮을 수 있다는 단점이 있다. 이러한 강화학습의 단점으로 인해, MAPPER를 포함한 대부분의 강화학습 기반 드론 제어 연구들은 주로 간단한 그리드 환경에서 검증되었다[13-15].

본 논문에서는 앞서 언급한 드론 궤적 생성 알고리즘의 고정된 하이퍼파라미터로 인한 환경의 일반화 성능 저하와 강화학습의 초반 낮은 학습 안정성을 해결하는 새로운 알고리즘을 제안한다. 제안하는 알고리즘은 계층적 심층강화학습 프레임워크를 활용하여 강화학습의 높은 환경의 일반화 성능과 드론 궤적 생성 알고리즘의 높은 안정성을 동시에 달성하였다.

Ⅲ. 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘

3.1 제안하는 계층적 심층강화학습 프레임워크

드론 궤적 생성 알고리즘들은 동적 환경에서 궤적 생성을 위해 드론의 최대 속도와 최대 가속도를 파라미터화하여 설정한다. 기존 드론 궤적 생성 알고리즘들은 실시간 환경이 변함에도 불구하고 드론의 최대 속도와 최대 가속도의 초기 설정값을 변경하지 않고 경로를 생성한다.

특히, 실제 동적 환경에서 드론 자율 비행이 가능한 EGO-Planner의 최대 속도 파라미터를 확인해보면 우리가 일반적으로 기대하는 드론의 속도에 크게 못 미친다는 것을 확인할 수 있다. EGO-Planner뿐만 아니라 대부분의 드론 궤적 생성 알고리즘들은 드론의 최대 속도와 최대 가속도를 파라미터화 하여 고정된 값으로 설정하였다. 드론의 효율성과 실용성을 높이기 위해서는 드론의 최대 속도와 최대 가속도를 실시간 환경에 맞춰 변경해줄 필요가 있다. 하지만, 규칙 기반 알고리즘으로 드론의 최대 속도와 최대 가속도를 설정하는 것은 동적 환경의 모든 예외를 처리할 수 없다.

본 논문에서는 EGO-Planner의 실시간 속도 관련 파라미터 조정을 위해, 그림 1에서 제시된 계층적 강화학습 프레임워크을 활용한다. 계층적 강화학습 프레임워크는 최종적인 문제를 해결하기 위해 여러 단계의 서브

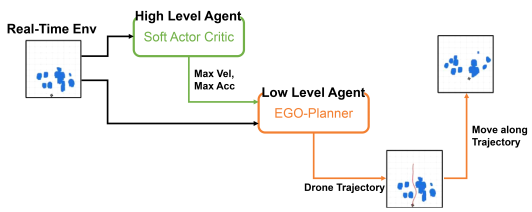


그림 1. 제안하는 계층적 심층강화학습 기반 알고리즘 프레임워크
Fig. 1. The proposed algorithm framework based on hierarchical deep reinforcement learning

문제들을 해결하는 강화학습 프레임워크를 말한다. 일반적으로 계층적 강화학습의 상위 계층 에이전트는 문제를 크게 분류하여 기본적인 행동을 선택하게 되고, 하위 계층 에이전트는 상위 계층 에이전트에서 선택된 행동에서 파생된 서브 문제를 해결하기 위해 학습한다. 계층적 강화학습에서 서로 다른 계층의 두 에이전트는 서로 다른 시간 스케일에서 작동하며, 상위 에이전트는 더 높은 시간 스케일에서 동작하는 반면 하위 에이전트는 더 낮은 시간 스케일에서 동작한다. 계층적 강화학습은 복잡한 작업을 서브 문제의 계층 구조로 분해할 수 있으므로 비교적 쉬운 일반화가 가능하고, 학습 효율성이 높다는 장점이 있다.

계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘에서 상위 계층 에이전트는 심층강화학습 SAC 알고리즘을 활용하여, 동적으로 변화하는 실시간 환경을 고려한 드론의 최대 속도와 최대 가속도를 산출한다. 산출된 드론의 최대 속도 및 최대 가속도는 실시간 동적 환경에 따라 안정성과 속도, 경로의 부드러움을 모두 고려한 최적 최대 속도와 최대 가속도이다. 하위 계층 에이전트인 EGO-Planner 알고리즘은 상위 계층 에이전트에서 산출된 최대 속도 및 최대 가속도를 활용하여 드론의 궤적을 생성한다. 즉, 상위 계층 에이전트에서는 심층강화학습을 사용하고 하위 계층에서는 EGO-Planner이 제시하는 전통적인 제어 알고리즘을 사용한다. 이러한 프레임워크는 기존 강화학습 기반 드론 제어 연구들이 가지고 있던 불안정성을 크게 개선하였다.

Algorithm 1: Proposed Algorithm

G: global target
O: environment surrounding the drone
P: position of drone
π_ϕ: actor of SAC
$\mathcal{Q}_\theta$: critic of SAC (state action value function)
$\mathcal{B}$: replay memory

inialize $\mathcal{B}$
Set G
repeat
  foreach *each episode* **do**
    foreach *each episode step* **do**
      $MaxVel_t, MaxAcc_t \sim \pi_\phi(O_t)$; // high level agent
      /* low level agent starts */
      $\Gamma = \text{EgoSwarm}(MaxVel_t, MaxAcc_t, G, O_t)$
      foreach *t second* **do**
        $P \sim MoveDroneAlongTraj(\Gamma)$
        CheckCollision(P)
      /* low level agent ends */
      $O_{t+1} \sim DroneSensor(P)$
      $\mathcal{B} \leftarrow \mathcal{B} \cup (O_t, (MaxVel_t, MaxAcc_t), r_t, O_{t+1})$
    foreach *the number of training* **do**
      Training SAC agent (i.e. π_ϕ and $\mathcal{Q}_\theta$)
until *Drone Reach G*;

그림 2. 제안하는 알고리즘의 의사 코드
Fig. 2. Pseudocode of proposed algorithm

계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘의 의사 코드는 그림 2와 같다. 먼저 드론의 최종 목적지를 설정한 이후, 현재 드론의 위치에서 드론의 센서에 감지되는 환경 정보를 상위 계층 에이전트에게 입력값으로 준다. 상위 계층 에이전트가 입력받은 주변 환경 정보를 고려하여 드론의 최대 속도와 최대 가속도를 추출하게 되면, 하위 계층 에이전트가 시작된다. 하위 계층 에이전트는 상위 계층에서 추출한 드론의 최대 속도와 최대 가속도, 최종 목적지, 실시간 환경을 고려한 EGO-Planner에서 경로를 생성한다. 생성한 경로를 따라 드론은 t초 동안 이동한다. 에피소드 종료 시, 상위 계층 에이전트인 SAC 알고리즘을 훈련 시킨다. 위에서 설명한 일련의 과정을 드론이 목적지에 도달할 때까지 반복하게 된다.

3.2 상위 계층 에이전트

계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘의 상위 계층 에이전트는 심층강화학습 SAC 알고리즘을 활용한다. 강화학습 알고리즘 중 하나인 SAC는 특정 상태에 대해 최적의 행동 산출을 목적으로 하는 Actor와 산출된 행동을 평가하는 Critic으로 구성된 Actor-Critic 알고리즘의 구조를 기반으로 한다. SAC 알고리즘은 탐험을 장려하기 위해 Actor 학습의 목표인 목적 함수에 엔트로피 항을 추가한다. 엔트로피를 최대화함으로써 현재 Actor가 판단하는 최적 정책뿐만 아니라 다른 근사 최적 정책들도 모두 고려할 수 있게 된다. 이로 인해 탐험을 더 효과적으로 수행할 수 있으며, 다양한 근사 최적 정책을 찾을 수 있어 학습이 강건해진다.

SAC 알고리즘은 off-policy 알고리즘이기 때문에 사용한 학습 데이터를 다시 사용할 수 있어 학습 데이터 샘플링이 효과적이라는 장점이 있다[10]. SAC 알고리즘에서 Critic의 손실 함수는 다음과 같이 정의된다[4].

$$J_Q(\theta) = E_{(s_t,a_t)}[\frac{1}{2}(Q_\theta(s_t,a_t) - (r(s_t,a_t) + \gamma(Q_\theta((s_{t+1},a_{t+1}) - \alpha log(\pi_\phi(a_{t+1}|s_{t+1}))))^2] \quad (1)$$

Q는 상태-행동 가치 함수인 Critic을 뜻하며 θ는 Q의 학습 가능한 파라미터이다. r은 강화학습 환경의 보상 함수를 뜻한다. 또한, a와 s는 각각 행동과 상태를 뜻한다. 마지막으로 π는 Actor를 정의하며 ϕ는 π의 학습 가능한 파라미터이다. π는 특정 s를 입력으로 받을 때, 최적의 a의 추출을 목표로 학습이 이루어진다. Critic의 평가를 통해 학습되는 Actor의 목적 함수는 다음과 같

이 정의된다.

$$J_\pi(\phi) = \sum_t E_{s_t}[E_{a_t \sim \pi_\phi}[Q_\theta(s_t, a_t)] + \alpha H(\pi_\phi(\bullet|s_t))] \quad (2)$$

여기서 H는 일반적으로 사용되는 엔트로피 식이고, α는 엔트로피와 상태-행동 가치 간의 중요도를 결정하는 파라미터이다. α는 아래 목적 함수를 통해 학습이 진행된다.

$$J(\alpha) = E_{a_t \sim \pi_t}[-\alpha log \pi_t(a_t|s_t) - \alpha \overline{H}] \quad (3)$$

SAC 알고리즘을 사용하는 상위 계층 에이전트의 MDP (Markov Decision Process)[16] 구성 요소인 상태(State)는 드론의 센서로 감지되는 주변 장애물 및 목적지까지의 방향 정보로 구성되며, 행동(Action)은 하위 계층의 EGO-Planner 알고리즘에서 활용할 드론의 최대 속도 및 최대 가속도이며, 보상은 목적지 도착 시에는 +1.0, 장애물 충돌 시에는 -1.0, 그 외 상태에서는 -0.01를 할당한다. 설정한 MDP는 충돌하지 않고, 목적지에 더 빠른 속도로 도달하도록 학습을 유도한다.

3.3 하위 계층 에이전트

계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘의 하위 계층 에이전트는 EGO-Planner를 활용한다. EGO-Planner는 숲과 같이 사전 정보가 없는 복잡한 환경에서 드론의 자율 비행을 위한 궤적 생성 알고리즘이다. EGO-Planner은 외부 위치 정보나 비행 환경에 대한 사전 정보 없이 드론에 탑재된 시스템만을 활용해 경로를 생성한다.

EGO-Planner은 시간-공간 동시 최적화를 통해 경로를 생성하고, 각 드론의 시간적인 부분을 조정한다. 이를 통해 EGO-Planner은 사전 정보가 없는 복잡한 환경에서도 ms 단위로 경로를 생성할 수 있다. 또한, EGO-Planner은 다중 목적 최적화 함수를 통해 물체 추적, 대형 유지와 같은 특정 임무를 추가할 수 있으며, 군집 드론들이 서로의 궤적을 공유하고 이를 통해 데이터의 전송을 최소화하고 신뢰성이 낮은 통신 네트워크에서도 작동할 수 있다.

EGO-Planner는 드론의 최대 속도와 최대 가속도를 파라미터화하여 사전에 설정한다. 생성된 궤적에서 드론의 속도와 가속도가 사전에 설정한 최대 파라미터보다 높다면 궤적의 시간을 재할당하여 드론의 속도와 가속도가 최대 파라미터보다 낮아지도록 궤적을 조정한다. EGO-Planner에서 생성한 궤적의 속도 및 가속도를 최대 속도 파라미터 및 최대 가속도 파라미터와 비교는 다음 식으로 진행된다.

$$r_e = \max(|V/v_{\max}|, \sqrt{|A/a_{\max}|}, \sqrt[3]{|J/j_{\max}|}, 1) \quad (4)$$

여기서 V, A, J는 각각 궤적의 속도와 가속도 그리고 가가속도를 의미한다. 그리고 $v_{\max}, a_{\max}, j_{\max}$은 사전에 설정한 최대 속도, 최대 가속도 그리고 최대 가가속도를 의미한다. 수식 (4)에 의해 생성한 궤적의 속도나 가속도 그리고 가가속도가 미리 설정한 최대 파라미터들을 얼마나 초과했는지에 대한 비율 r_e이 나온다. r_e는 수식 (5)를 통해 기존에 할당된 시간 Δt를 재할당한다. 그리고 궤적은 재할당된 시간 $\Delta t'$에 의해 조정된다. 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘은 가가속도는 고려하지 않는다.

$$\Delta t' = r_e \Delta t \quad (5)$$

한편, 하위 계층 에이전트에서 드론 경로 생성 후 t초 동안 경로를 따라 드론이 이동하며, t초 이후 상위 계층에서 SAC 에이전트 타임 스텝이 1회 진행된다. 하위 계층에서 드론이 장애물과 충돌되거나 목적지에 도달하면 상위 계층에서 SAC 알고리즘의 에피소드는 종료된다.

Ⅳ. 실 험

4.1 실험 환경

계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘의 훈련은 매번 임의로 생성되는 장애물이 있는 실험 환경에서 진행한다. 상위 계층의 심층강화학습 에이전트는 python으로 구현이 되어있으며, EGO-Planner는 C++로 구현이 되어있다. EGO-Planner에서 생성된 실시간 환경 정보는 ROS 통신을 이용해서 심층강화학습 에이전트에게 전달된다. 전달된 실시간 환경 정보를 활용하여 심층강화학습 에이전트는 보상을 계산하고, 학습을 시작한다.

본 실험에서는 SAC 알고리즘의 Critic 학습률 $\lambda_c = 0.0003$, Actor 학습률 $\lambda_a = 0.0003$, Entropy 학습률 $\lambda_e = 0.0001$, 감가율 $\gamma = 0.99$, 배치 사이즈 256, 리플레이 버퍼 크기 10^6, 그리고 총 학습 에피소드 횟수로 50,000을 사용하였다. 마지막으로, 실험에 활용된 컴퓨터 재원은 우분투 20.04 LTS 운영체제에서 i9-9900K CPU와 64GB RAM, RTX 3090 GPU이다.

4.2 경로의 효율성 비교 실험

그림 3은 장애물이 없는 환경과 복잡한 환경에서 기존 EGO-Planner와 최대 속도를 3배 증가시킨 Fast EGO-Planner 그리고 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용한 EGO-Planner와 비교한다. 2개의 환경 모두 드론은 출발점에서 시작하여 경유지를 지나 도착점으로 향한다. 복잡한 환경의 경우, 드론은 파란색의 장애물과의 충돌을 피해 비행이 이루어져야 한다. 두 환경에서 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘 적용되었을 때, 가장 효율적인 경로를 생성함을 알 수 있다. 특히 Fast-EGO-Planner는 장애물이 없는 환경에서는 기존 EGO-Planner보다 비교적 효율적인 궤적을 생성하였으나, 복잡한 환경에서는 장애물과 충돌한 것을 확인할 수 있다. 하지만 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용하였을 때, 더 효율적인 경로를 생성함과 동시에 복잡한 환경에서도 안정성을 확보하였다. 또한, 복잡한 환경에서 비교 실험 시 시간마다 각 알고리즘의 현재 속도, 최대 속도, 최대 가속도를 표시한 표 1에서 확인할 수 있듯이, 속도 성능에서도 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘이 기존 EGO-Planner보다 좋다는 것을 보여준다.

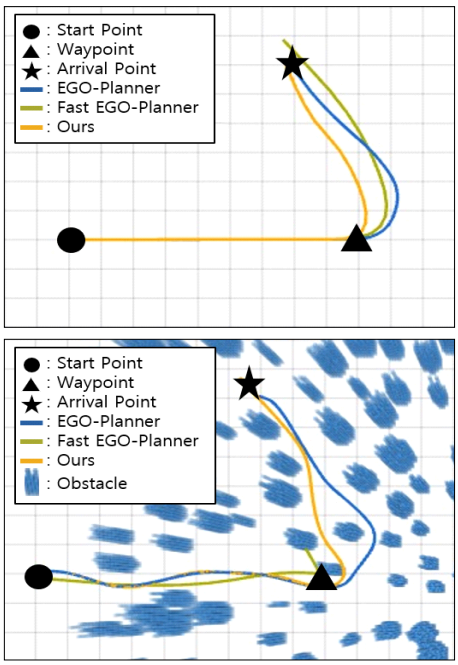


그림 3. 장애물이 없는 환경에서 비교 실험(위)과 복잡한 환경에서 비교 실험(아래)
Fig. 3. Comparison experiment in an obstacle-free environment (above) and a complex environment (below).

표 1. 복잡한 환경에서 시간에 따른 각 알고리즘의 현재 속도, 최대 속도, 현재 가속도, 최대 가속도
Table 1. Current speed, maximum speed, and maximum acceleration of each algorithm over time in a complex environment.

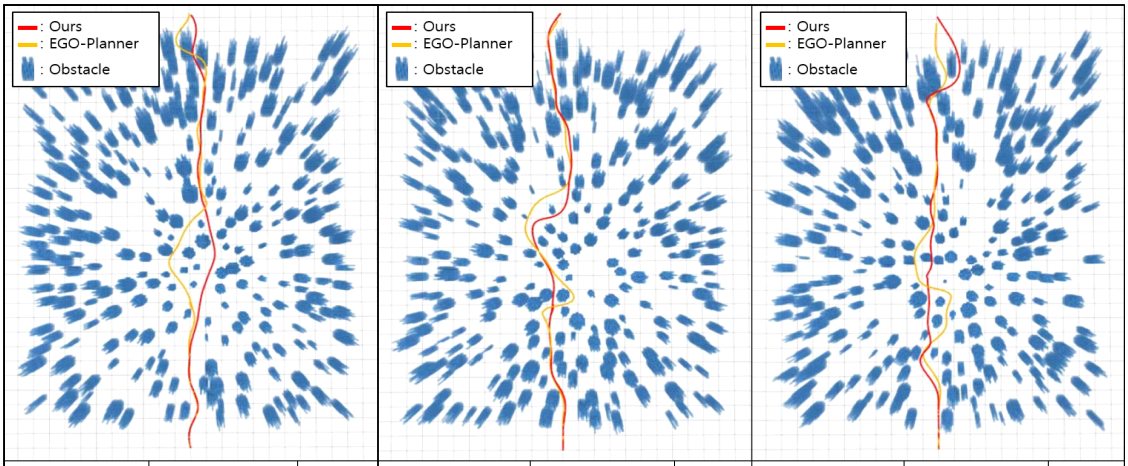
		1sec	2sec	3sec	4sec	5sec
EGO-Planner	current vel(m/s)	0.9	0.62	0.53	0.78	0.52
	max vel(m/s)	2.0	2.0	2.0	2.0	2.0
	current acc(m/s^2)	0.2	-0.1	2.6	-0.7	0.0
	max acc(m/s^2)	6.0	6.0	6.0	6.0	6.0
Fast EGO-Planner	current vel(m/s)	2.1	2.3	1.5	-	-
	max vel(m/s)	6.0	6.0	6.0	6.0	6.0
	current acc(m/s^2)	0.2	-0.2	2.3	-	-
	max acc(m/s^2)	6.0	6.0	6.0	6.0	6.0
Ours	current vel(m/s)	**2.9**	**0.85**	**0.77**	**2.1**	**1.0**
	max vel(m/s)	**6.0**	**3.6**	**3.4**	**5.2**	**4.3**
	current acc(m/s^2)	**0.9**	**-1.7**	**2.9**	**-0.5**	**0.3**
	max acc(m/s^2)	**1.3**	**2.8**	**3.1**	**2.1**	**1.6**

표 1에 표시된 각 알고리즘의 현재 속도는 x와 y축의 평균 속도이다.

4.3 복잡한 환경에서 경로의 효율성 비교 실험

복잡한 환경에서 검증하기 위해 총 3개의 임의의 환경에서 경로의 효율성 비교 실험을 진행하였다. 드론들은 각 환경에서 기존 EGO-Planner과 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용한 EGO-Planner에서 생성되는 경로를 따라 출발지에서 목적지로 이동하게 된다.

그림 4의 실험 결과에서, 빨간색 경로는 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 활용하여 생성한 경로이고, 주황색 경로는 EGO-Planner을 활용해 생성한 경로이다. 각 환경에서 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용한 EGO-Planner가 기존 EGO-Planner보다 평균 속도가 더 빠른 것을 확인할 수 있다. 그리고 평균 속도에 따라 목적지까지 도달하는 시간(traj time)이 평균적으로 약 25% 감소한 것을 확인할 수 있다. 또한, 그림에 표현된 것과 같이 전체적으로 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘을 적용할 때 경로의 부드러



	EGO-Planner	**Ours**		EGO-Planner	**Ours**		EGO-Planner	**Ours**
average vel (m/s)	1.41	**1.99**	average vel (m/s)	1.45	**2.16**	average vel (m/s)	1.35	**1.90**
traj time (s)	26.26	**19.54**	traj time (s)	25.91	**18.44**	traj time (s)	26.82	**18.86**
traj length (m)	30.76	**30.56**	traj length (m)	30.27	**30.16**	traj length (m)	30.94	**29.57**

그림 4. 시뮬레이션 환경에서 제안하는 알고리즘과 EGO-Planner 알고리즘의 성능 비교 실험 결과
Fig. 4. Experimental results between the proposed algorithm and the EGO-Planner algorithm in the ROS simulation environment

움이 향상된 것을 확인할 수 있다. 이에 따라 목적지까지 도달하는 경로 길이(traj length)가 감소한다.

V. 결 론

본 논문에서는 최근 주목받고 있는 드론 자율 비행 알고리즘을 보유한 EGO-Planner의 최대 속도와 최대 가속도를 실시간으로 변경해주는 새로운 계층적 심층강화학습 알고리즘을 제안한다. 본 논문에서 제안하는 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘은 시뮬레이션에서 비교 실험을 통해 기존 알고리즘보다 속도, 경로의 길이, 경로의 부드러움 면에서 성능이 뛰어남을 확인할 수 있었다. 특히, 환경을 고려하지 않고 최대 속도를 올릴 경우, 복잡한 환경에서 충돌함을 보여줌과 동시에 계층적 심층강화학습 기반 능동적 파라미터 제어 알고리즘은 환경을 고려하여 하위 계층 에이전트의 최대 속도와 최대 가속도를 설정해 기존 알고리즘의 안정성을 동시에 고려하는 것을 보여줬다. 향후에는 한 대의 드론이 아닌 군집 드론을 활용할 때의 실시간 궤적 생성 성능 향상 연구와 실제 환경에서 드론에 적용하는 연구를 진행할 예정이다.

References

[1] X. Zhou, et al., "Ego-planner: An esdf-free gradient-based local planner for quadrotors," *IEEE Robotics and Automat. Lett.*, vol. 6, no. 2, pp. 478-485, 2020. (https://doi.org/10.1109/LRA.2020.3047728)

[2] A. G. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," *Discrete event Dynamic Syst.*, vol. 13, no. 1-2, pp. 41-77, 2003. (https://doi.org/10.1023/A:1022140919877)

[3] S. Pateria, et al., "Hierarchical reinforcement learning: A comprehensive survey," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1-35, 2021. (https://doi.org/10.1145/3453160)

[4] T. Haarnoja, et al., "Soft actor-critic algorithms and applications," *arXiv preprint arXiv:1812.05905*, 2018. (https://doi.org/10.48550/arXiv.1812.05905)

[5] M. Quigley, et al., "ROS: An open-source robot operating system," *ICRA Wkshp. Open*

Source Softw., vol. 3, no. 3.2, 2009.

[6] S. Lee and Y. Choi, "Reviews of unmanned aerial vehicle (drone) technology trends and its applications in the mining industry," *Geosyst. Eng.*, vol. 19, no. 4, pp. 197-204, 2016. (https://doi.org/10.1080/12269328.2016.1162115)

[7] J. Tordesillas, B. T. Lopez, and J. P. How, "Faster: Fast and safe trajectory planner for flights in unknown environments," *2019 IEEE/RSJ Int. Conf. IROS IEEE*, pp. 1934-1940, 2019. (https://doi.org/10.1109/IROS40897.2019.8968021)

[8] X. Zhou, et al., "Ego-swarm: A Fully autonomous and decentralized quadrotor swarm system in cluttered environments," *2021 IEEE ICRA IEEE*, pp. 4101-4107, 2021. (https://doi.org/10.48550/arXiv.2011.04183)

[9] J. Tordesillas and J. P. How, "MADER: Trajectory planner in multiagent and dynamic environments," *IEEE Trans. Robotics*, vol. 38, no. 1, pp. 463-476, 2021. (https://doi.org/10.1109/TRO.2021.3080235)

[10] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, MIT Press, 2018. (https://doi.org/10.1016/S1364-6613(99)01331-5)

[11] A. T. Azar, et al., "Drone deep reinforcement learning: A review," *Electronics*, vol. 10, no. 9, p. 999, 2021. (https://doi.org/10.3390/electronics10090999)

[12] Z. Liu, et al., "Mapper: Multi-agent path planning with evolutionary reinforcement learning in mixed dynamic environments," *2020 IEEE/RSJ Int. Conf. IROS IEEE*, pp. 11748-11754, 2020. (https://doi.org/10.1109/IROS45743.2020.9340876)

[13] Y. Chen, et al., "Efficient drone mobility support using reinforcement learning," *2020 IEEE WCNC IEEE*, pp. 1-6, 2020. (https://doi.org/10.1109/WCNC45663.2020.9120595)

[14] W. J. Yun, et al., "Distributed deep reinforcement learning for autonomous aerial eVTOL mobility in drone taxi applications," *ICT Express*, vol. 7, no. 1, pp. 1-4, 2021. (https://doi.org/10.1016/j.icte.2021.01.005)

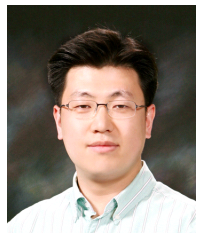
[15] U. Challita, W. Saad, and C. Bettstetter, "Deep reinforcement learning for interference-aware path planning of cellular-connected UAVs," *2018 IEEE ICC, IEEE*, pp. 1-7, 2018. (https://doi.org/10.1109/ICC.2018.8422706)

[16] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*, John Wiley & Sons, 2014. (https://doi.org/10.1002/9780470316887)

지 창 훈 (Chang-hun Ji)

2020년 8월 : 한국기술교육대학교 메카트로닉스공학과 졸업
2023년 2월 : 한국기술교육대학교 컴퓨터공학과 석사
2023년 3월~현재 : 한국기술교육대학교 컴퓨터공학과 박사과정
<관심분야> 사물인터넷, 정밀 제어, 강화학습
[ORCID:0009-0003-4372-8675]

한 연 희 (Youn-hee Han)



1996년 2월 : 고려대학교 수학과 졸업
1998년 2월 : 고려대학교 컴퓨터공학과 석사
2002년 2월 : 고려대학교 컴퓨터공학과 박사
2002년 3월~2006년 2월 : 삼성종합기술원 전문연구원
2013년 9월~2014년 8월 : SUNY at Albany, Department of Computer Science 방문교수
2006년~현재 : 한국기술교육대학교 컴퓨터공학부 교수
<관심분야> 사물인터넷, 5G/6G, 딥러닝, 강화학습, 조합최적화
[ORCID:0000-0002-5835-7972]

문 성 태 (Sung-Tae Moon)

2005년 2월 : 전남대 컴퓨터정보학부 졸업
2007년 2월 : 광주과학기술원 석사
2007년~2010년 : 국방과학연구소 연구원
2010년~2011년 : 국가보안기술연구소 연구원
2012년~2022년 : 한국항공우주연구원 선임연구원
2022년~2023년 : 한국기술교육대학교 조교수
2023년~현재 : 충북대학교 조교수
<관심분야> 무인 항공기 위치 예측, 군집비행, 딥러닝
[ORCID:0000-0002-1638-6898]