Journal of Advanced Technology Research ISSN xxxx-xxxx

심층강화학습 기반 EGO-Swarm 파라미터 제어
Deep Reinforcement Learning-Based Control of EGO-Swarm Parameters

지 창 훈 (ChangHun Ji)
한 연 희 (YounHee Han)

Advanced Technology Research Center
Korea University of Technology and Education
{koir5660, yhhan}@koreatech.ac.kr

요 약

최근 EGO-Swarm 시스템은 복잡한 실제 환경에서 성공적인 원활한 드론 비행 경로 생성을 함으로써 주목받고 있다. EGO-Swarm 은 원활한 경로 생성을 위해 다양한 파라미터 값들을 설정하는데, 이러한 파라미터 중에서 드론의 최대 속도와 최대 가속도는 드론 비행 성능 향상에 중요한 역할을 한다. EGO-Swarm 은 처음 설정한 파라미터 값을 경로 생성 중에 고정되지만, 최대 속도와 최대 가속도의 경우 실시간으로 변하는 환경에 따라 최적의 값이 변하기 때문에 실시간으로 변화하는 환경에 따라 새롭게 변경하는 것이 바람직하다. 본 논문에서는 실시간 환경 변화에 따라 최대 속도와 최대 가속도를 계층적 심층강화학습을 통해 새롭게 설정하는 새로운 알고리즘을 제안한다. 제안 기법의 타당성을 검증하기 위해 ROS 시뮬레이션에서 기존 EGO-Swarm 알고리즘과 제안하는 알고리즘의 비교 실험을 진행하며, 실험 결과 제안하는 알고리즘이 평균 속도 향상 및 경로의 길이 측면에서 기존 EGO-Swarm 알고리즘보다 성능이 더 좋다는 것을 확인할 수 있다.

키워드: 드론 자율 비행, EGO-Swarm, 계층적 심층강화학습

Abstract

The EGO-Swarm system has recently gained attention for its successful generation of smooth drone flight paths in complex real-world environments. EGO-Swarm utilizes various parameter values for smooth path generation, and among these parameters, the maximum speed and maximum acceleration of the drone play a crucial role in improving flight performance. However, in EGO-Swarm, the initially set parameter values remain fixed during path creation, despite the fact that the optimal values for maximum speed and maximum acceleration may change in real-time dynamic environments. Therefore, it is desirable to dynamically adjust these values according to the changing real-time environment. In this paper, we propose a novel algorithm that dynamically sets the maximum speed and maximum acceleration using hierarchical deep reinforcement learning in response to real-time environmental changes. To validate the effectiveness of the proposed method, a comparative experiment between the existing EGO-Swarm algorithm and the proposed algorithm

is conducted in a ROS simulation. The experimental results demonstrate that the proposed algorithm outperforms the existing EGO-Swarm algorithm in terms of average speed improvement and path length.
Key words: Drone Autonomous Flight, EGO-Swarm, Hierarchical Deep Reinforcement Learning

1. 서 론

최근 다양한 분야에서 드론의 사용이 늘어나면서, 드론 자율 주행에 관한 연구가 활발히 이루어지고 있다. 드론 자율 주행 연구는 장애물을 회피하는 실시간 경로 생성을 함과 동시에 드론의 기계적 결함을 최소화하기 위해 동역학, 경로의 부드러움 및 안전성을 동시에 고려하는 것이 필수적이다. 이러한 드론 자율 주행의 요구사항은 복잡한 환경에서는 더욱 필수적지만, 해결하기 어려운 과제이다.

드론 자율 주행 알고리즘 중 하나인 EGO-Swarm [1]은 여러 대의 드론 경로를 실시간으로 생성하는 알고리즘을 지닌 시스템이다. 특히, 복잡한 실제 환경에서 드론의 실시간 경로 생성에 성공하면서 관련 연구자들로부터 많은 주목을 받고 있다. EGO-Swarm 은 원활한 경로 생성을 위해 많은 파라미터를 가지고 있는데, 그중 드론의 최대 속도와 최대 가속도가 있다. 최대 속도와 최대 가속도는 현재 경로가 생성되는 환경이 실시간으로 변함에 따라 최적의 값이 변하지만, EGO-Swarm 은 초기에 설정한 값을 유지한다.

따라서 본 논문에서는 계층적 심층강화학습 [2, 3]을 활용하여 실시간 바뀌는 환경에 따라 최대 속도와 최대 가속도를 지정한다. 상위 계층 에이전트는 Soft-Actor-Critic (SAC) [3] 알고리즘을 활용하여 실시간 바뀌는 환경을 고려해 최대 속도와 최대 가속도를 설정한다. 하위 계층 에이전트는 설정한 최대 속도와 최대 가속도를 활용해 EGO-Swarm 알고리즘으로부터 실시간 경로를 생성한다. 복잡한 환경을 나타내는 ROS 시뮬레이션에서 제안하는 알고리즘과 기존 EGO-Swarm 을 비교 실험하며, 제안하는 알고리즘이 속도, 경로 길이, 경로의 부드러움에서 더 뛰어난 것을 보여준다

2. 관련 연구 및 문제 제시

2.1. EGO-Swarm

EGO-Swarm 은 숲과 같이 사전 정보가 없는 복잡한 환경에서 여러 대 드론, 즉 군집 드론의 자율 주행을 위한 경로 생성 알고리즘이다. EGO-Swarm 은 외부 위치 정보나 주행 환경에 대한 사전 정보 없이 드론에 탑재된 온보드만을 활용해 경로를 생성한다.

EGO-Swarm 은 시간-공간 동시 최적화를 통해 경로를 생성하고, 각 드론의 시간적인 부분을 조정한다. 이를 통해 EGO-Swarm 은 사전 정보가 없는 복잡한 환경에서도 ms 단위로 경로를 생성할 수 있다. 또한, EGO-Swarm 은 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.

2.2. Soft-Actor-Critic

강화학습은 에이전트가 환경과 상호 작용하여 최적의 결정을 내리는 방법을 배우는 기계 학습의 한 분야이다. 에이전트는 자신이 추출한 행동에 따라 받는 보상을 최대화시키는 방향으로 학습이 이루어진다.

강화학습 알고리즘 중 하나인 Soft-Actor-Critic (SAC)은 Actor 와 Critic 으로 구성된 Actor-Critic 알고리즘의 구조를 기반으로 한다. SAC 알고리즘은 탐험을 장려하기 위해 Actor 학습의 목표인 objective 에 엔트로피 항을 추가한다. 엔트로피를 최대화함으로써 현재 Actor 가 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. SAC 알고리즘에서 Critic 의 objective 는 다음과 같이 정의된다.

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

Q는 상태-행동 가치를 뜻하며, r은 강화학습 환경의 보상 함수를 뜻한다. 또한 a와 s는 각각 행동과 상태를 뜻한다. 마지막으로 π는 Actor 를 정의하며, 특정 s를 입력으로 받을 때, 최적의 a의 추출을 목표로 학습이 이루어진다.

2.3. 계층적 심층강화학습

계층적 심층강화학습은 최종적인 문제를 해결하기 위해 여러 단계의 서브 문제들을 해결하는 심층강화학습 프레임워크를 말한다. 일반적으로 계층적 심층강화학습의 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 심층강화학습에서 서로 다른 계층의 두 에이전트는 서로 다른 수준의 타임 스텝 호라이즌(Time Step Horizen)에서 작동하며, 상위 에이전트는 더 높은 시간 척도에서 동작하는 반면 하위 에이전트는 더 낮은 시간 척도에서 동작한다.

계층적 심층강화학습 복잡한 작업을 서브 문제의 계층 구조로 분해할 수 있기 때문에 비교적 쉬운 일반화 가 가능하고, 학습 효율성이 높다는 장점이 있다.

2.4. 문제 제시

EGO-Swarm 은 원활한 경로 생성을 위해 다양한 파라미터들을 가지고 있는데 그중 드론의 최대 속도와 최대 가속도가 있다. 드론의 최대 속도와 최대 가속도는 현재 경로가 생성되는 드론의 주변 환경에 따라 최적의 값이 변한다. 하지만, 기존 EGO-Swarm 은 실시간 환경이 변함에도 불구하고 드론의 최대 속도와 최대 가속도의 초기 설정값을 변경하지 않고 경로를 생성한다. 따라서, 본 논문은 경로가 생성될 때마다 실시간 동적 환경을 고려하여 최적의 최대 속도와 최대 가속도를 설정하는 알고리즘을 제안한다. 제안하는 알고리즘은 계층적 심층강화학습 프레임워크를 활용하며, 기존 EGO-Swarm 보다 드론의 속도와 경로의 부드러움, 경로의 길이를 개선하였다.

3. 제안하는 계층적 심층강화학습 프레임워크

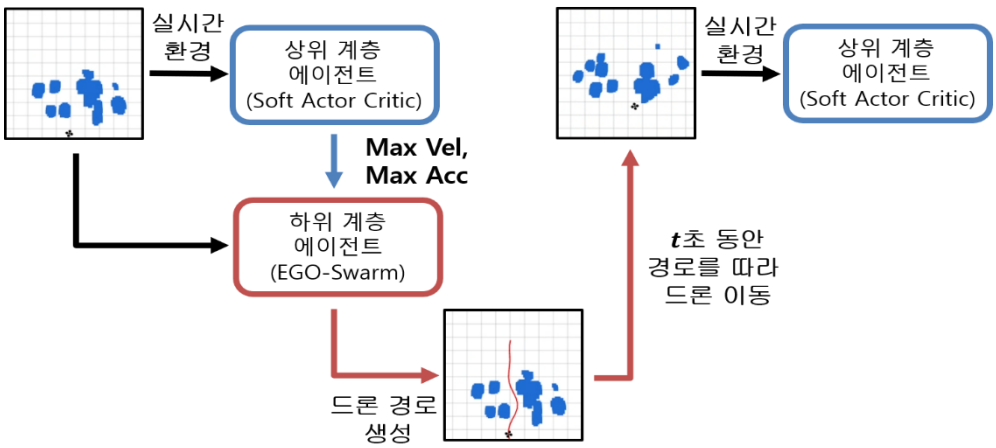


그림 1 제안하는 계층적 심층 강화학습 기반 알고리즘 프레임워크

본 논문에서는 EGO-Swarm 의 실시간 파라미터 조정을 위해, [그림 1]에서 제시된 계층적 심층강화학습을 활용한다. 제안하는 계층적 심층강화학습을 활용한 알고리즘에서 상위 계층 에이전트는 심층강화학습 SAC 알고리즘을 활용하여, 동적으로 변화하는 현재 환경을 고려하여 드론의 최대 속도와 최대 가속도를 산출한다. 이러한 드론의 최대 속도 및 최대 가속도를 활용하여 하위 계층에서는 EGO-Swarm 알고리즘이 드론 자율 비행을 위한 드론의 경로를 생성한다. 즉, 상위 계층에서는 심층강화학습을 사용하고 하위 계층에서는 EGO-Swarm 이 제시하는 전통적인 제어 알고리즘을 사용하는 방안을 제안한다.

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 알고리즘 의사코드

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

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

제안하는 알고리즘의 의사 코드는 [그림 2]와 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

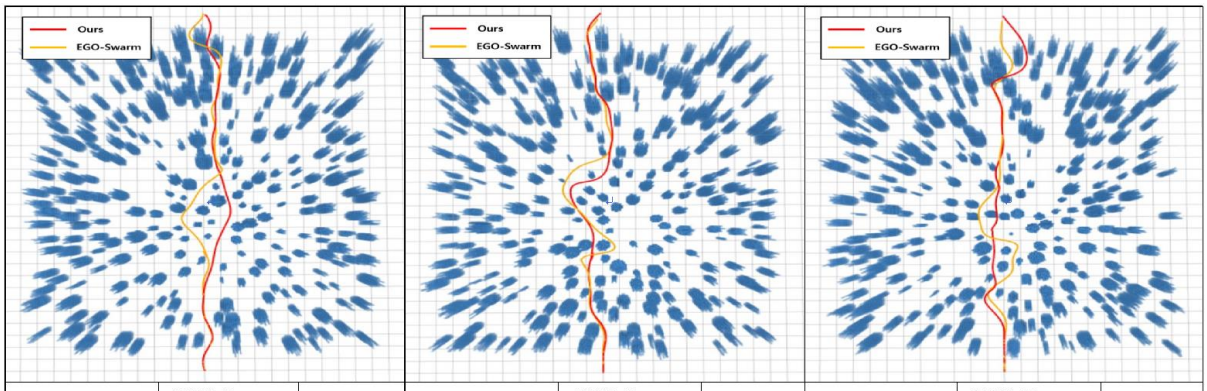
고려하여 드론의 최대 속도와 최대 가속도를 추출하게 되면, 하위 계층 에이전트가 시작된다. 하위 계층 에이전트는 상위 계층에서 추출한 드론의 최대 속도와 최대 가속도, 최종 목적지, 실시간 환경을 고려한 EGO-Swarm 에서 경로를 생성한다. 생성한 경로를 따라 드론은 초 동안 이동한다. 에피소드 종료 시, 상위 계층 에이전트인 SAC 알고리즘을 훈련시킨다. 위에서 설명한 일련의 과정을 드론이 목적지에 도달할 때까지 반복하게 된다.

4. 실험 평가

본 논문에서는 EGO-Swarm 에서 제공하는 ROS 기반 시뮬레이션을 활용하여 비교 실험을 진행한다. EGO-Swarm 은 복잡한 환경의 ROS 시뮬레이션을 제공한다. 총 3 개의 환경에서 비교 실험을 진행하였다. 드론들은 각 환경에서 EGO-Swarm 과 제안하는 알고리즘에서 생성되는 경로를 따라 출발지에서 목적지로 이동하게 된다.

[그림 3]의 실험 결과에서, 빨간색 경로는 제안하는 알고리즘을 활용하여 생성한 경로이고, 주황색 경로는 EGO-Swarm 을 활용해 생성한 경로이다. 각 환경에서 제안하는 알고리즘은 기존 EGO-Swarm 보다 평균 속도가 더 빠른 것을 확인할 수 있다. 그리고 평균 속도에 따라 목적지까지 도달하는 시간이 더 적게 걸리는 것을 확인할 수 있다. 또한, 그림에 표현된 것과 같이 전체적으로 제안하는 알고리즘에서 경로의 부드러움이 향상된 것을 확인할 수 있다. 이에 따라 목적지까지 도달하는 경로 길이가 감소한다.

…



	EGO-Swarm	ours
평균 속도(m/s)	1.41	1.99
걸린 시간(s)	26.26	19.54
경로 길이(m)	30.76	30.56

	EGO-Swarm	ours
평균 속도(m/s)	1.45	2.16
걸린 시간(s)	25.91	18.44
경로 길이(m)	30.27	30.16

	EGO-Swarm	ours
평균 속도(m/s)	1.35	1.90
걸린 시간(s)	26.82	18.86
경로 길이(m)	30.94	29.57

그림 2 ROS 시뮬레이션 환경에서 제안하는 알고리즘과 EGO-Swarm 알고리즘의 성능 비교 실험 결과

5. 결론

본 논문에서는 최근 주목받고 있는 드론 자율 비행 알고리즘을 보유한 EGO-Swarm 의 최대 속도와 최대 가속도를 실시간으로 변경해주는 새로운 계층적 심층강화학습 알고리즘을 제안한다. 제안하는 알고리즘은 ROS 시뮬레이션에서 비교 실험을 통해 그 효능을 검증한다. 다음 연구는 군집 드론을 활용할 때의 실시간 자율 비행 성능 향상 연구를 수행할 예정이다.

참고 문헌

[1] Zhou, Xin, et al., "Ego-swarm: A Fully Autonomous and Decentralized Quadrotor Swarm System in Cluttered Environments." 2021 IEEE international conference on robotics and automation (ICRA). IEEE, 2021.

[2] Barto, Andrew G., and Sridhar Mahadevan. "Recent advances in hierarchical reinforcement learning." Discrete event dynamic systems 13.1-2 (2003): 41-77.

[3] Pateria, Shubham, et al., "Hierarchical Reinforcement Learning: A Comprehensive Survey," ACM Comput. Surv. 54, 5, 2021.

[4] Haarnoja, Tuomas, et al. "Soft Actor-Critic Algorithms and Applications." arXiv preprint arXiv:1812.05905, 2018.