

# 에너지 효율적인 UAV 자율주행을 위한 심층강화학습 프레임워크 및 시각화

이구상, 김수현, 심병호

서울대학교

{gslee, soohyunkim, bshim}@islab.snu.ac.kr

## Energy Efficient UAV Autonomous Framework via Deep Reinforcement Learning and Visualization

Gusang Lee, Soohyun Kim, and Byonghyo Shim

Seoul National University

### 요약

본 논문은 재난 구조 등 극한 환경에서의 UAV 자율제어를 위해 에너지 기반 모델과 심층강화학습(Deep Reinforcement Learning, DRL)을 결합한 프레임워크를 제안한다. 제안한 프레임워크는 시뮬레이션 기반 도시환경에서 UAV의 자율경로를 학습하고, 에너지 맵 상에서 효율적인 웨이포인트를 최소화하여 통신량과 에너지 소비를 줄인다. 학습된 경로는 Unity 기반 3D 시각화 환경에서 사용자 제어 경로와 비교되며, 실험 결과 PPO(Proximal Policy Optimization) 기반 DRL 모델이 사용자보다 더 효율적인 경로를 산출함을 확인하였다.

### I. 서론

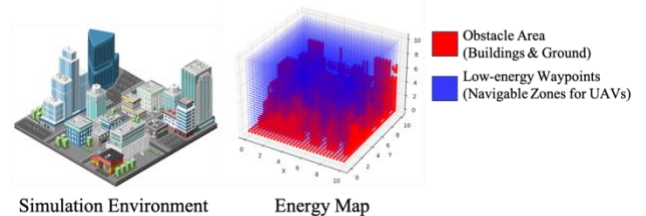
최근 무인항공기(UAV)는 재난 구조[1], 항공 촬영[2], 물체 추적[3] 등 다양한 분야에서 활용되고 있으며, 특히 인간의 접근이 어려운 극한 환경에서 실시간 모니터링과 제어에 효과적인 수단으로 주목받고 있다. 하지만 UAV는 통신 불안정, 에너지 제약, GPS 오차 등으로 인해 자율주행이 어려운 문제가 존재한다[4]. 이러한 문제를 해결하기 위해 기존에는 최적화 기반의 제어 기법이 사용되었으나, 실시간 환경 적응성이 부족하다는 한계가 있다. 이에 따라 최근에는 강화학습 기반의 접근이 주목받고 있으며, 특히 연속적인 행동 공간을 다루는 데 강점이 있는 PPO(Proximal Policy Optimization)[5] 알고리즘이 UAV 제어에 적합한 대안으로 떠오르고 있다. 본 논문에서는 Unity 기반 3D 도시 환경에서 에너지 모델을 적용한 시뮬레이션을 구축하고, PPO를 통해 UAV가 충돌 없이 효율적인 경로를 학습할 수 있도록 설계하였다. 더불어 GPS 오차와 통신 제약 조건을 반영하여 실제 상황과 유사한 조건에서 자율비행 성능을 검증하였다. 본 논문이 기여하는 바는 다음과 같다.

1. 실제 도시 재난 시나리오를 반영한 에너지 기반 3D 시뮬레이션 환경을 구현하였다.
2. 강화학습(PPO) 기반의 UAV 자율경로 학습 프레임워크를 설계하였다.
3. 학습된 경로와 사용자 경로를 시각적으로 비교하여 성능을 정량·정성적으로 분석하였다.

### II. 시스템 및 학습 구성

제안하는 시스템은 Unity 엔진 기반의 도시 환경 시뮬레이션 내에서 구현되며, UAV는 강화학습 알고리즘의 학습 주체로 설정된다. 시뮬레이션 환경은 고층 건물, 좁은 골목, 장애물, 바람 등 복잡한 도시 조건을 반영하여 실제 재난 구조 상황을 가정한 고차원

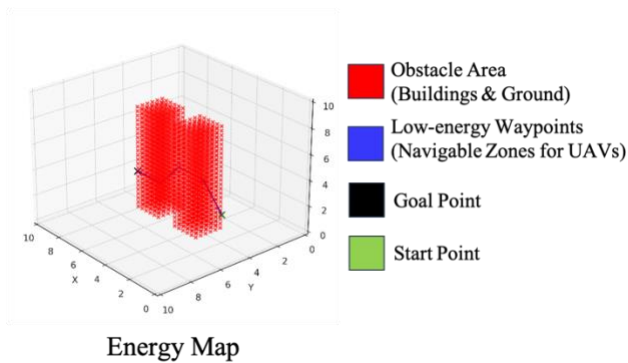
환경으로 구성된다. UAV에는 Unity 내장 센서를 탑재하여 실시간으로 장애물 위치, 자신의 좌표( $x, y, z$ ), 충돌 여부, 목표지점과의 거리 등의 정보를 감지하고 이를 학습에 필요한 state로 사용한다. 또한 학습된 정책은 Unity의 ML-Agents Toolkit을 통해 UAV의 행동(action)에 직접 반영되며, PPO 알고리즘 기반으로 업데이트된다. 경로 최적화를 위해 제안 시스템은 먼저 3차원 공간상에서 \*\*에너지 맵(Energy Map)\*\*을 구축한다. 에너지 값은 각 좌표에서의 충돌 확률, 거리, 환경 난이도(바람, 장애물 밀도 등)를 기반으로 계산되며, 이는 UAV가 해당 위치를 통과할 때 소비하는 비용으로 해석된다. 아래 [그림 1]과 같이, 이 에너지 맵을 기반으로 출발지와 목적지 사이의 전 공간을 3D Grid 형태로 분할하고, 그 중 일정 임계값 이하의 에너지를 가진 지점만 웨이포인트 후보로 샘플링한다.



[그림 1] 도시 시뮬레이션 환경에서의 에너지 기반 웨이포인트 후보 분포

다음 단계에서는 아래 [그림 2]와 같이 샘플링된 웨이포인트들에 대해 목적지까지의 방향 벡터 및 거리 연관성을 기준으로 필터링 과정을 거친다. 이는 UAV가 실제로 가야 할 방향과 무관한 불필요한 지점을 제거하기 위함이며, 최종적으로는 UAV의 이동을 최소한의 웨이포인트로 구성함으로써 통신 부담을 줄이고 효율적인 경로 학습이 가능하도록 한다. 이와 같은 방식은 UAV가 자율적으로 저에너지-저위험 경로를 선택하면서, 불필요한 명령 횟수를 줄이고,

제한된 통신 환경에서도 안정적으로 임무를 수행할 수 있도록 설계되어 있다.



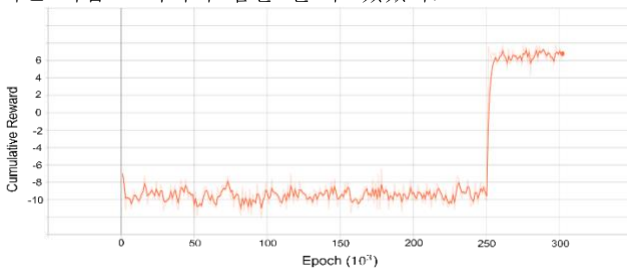
[그림 2] 도시 건물 사이를 자율적으로 통과하는 UAV의 경로 시각화

### III. 실험 구성 및 결과

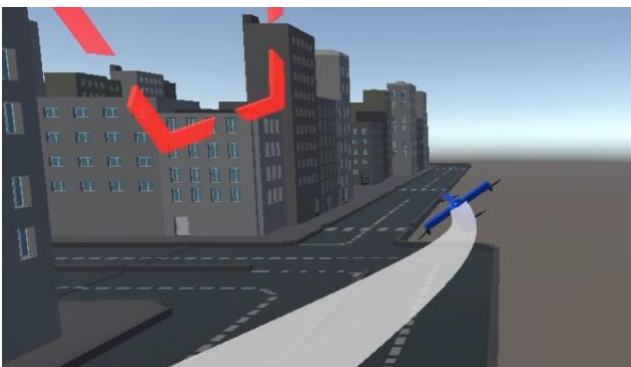
강화학습은 PPO 알고리즘을 기반으로 하며, UAV의 상태(state)는 현재 위치, 장애물 거리, 에너지 값 등으로 구성된다. Action은 연속적인 3D 이동 벡터이며, Reward는 충돌 회피, 에너지 소비 최소화, 도착 시간 단축 요소를 포함한다. 학습된 정책은 Unity ML-Agent의 Brain에 저장되어 시뮬레이션에서 사용된다.

Unity-RL을 이용하여 환경을 불러오고, 또한 상태, 행동, 보상을 정의하여 학습하는 방법에 대해 소개하였다.

PPO를 사용하여 분산을 줄이고 학습의 효율을 높였다. 그 결과 [그림 3]에서 알 수 있듯이, 학습 25만회만에 보상이 수렴함을 알 수 있었고, 또한 에피소드 내 시간에 따른 학습도 최적화됨을 알 수 있었다.



[그림 3] 에피소드에 따른 학습결과



[그림 4] 유니티 환경에서의 학습 결과

유니티에서의 학습의 시각화 또한 진행되었으며 실제로 복잡한 UAM 환경에서 학습이 진행되는 과정과 학습 완료 후 드론의 움직임을 볼 수 있다.

### IV. 결론 및 향후 연구 방향

본 논문은 극한 환경에서 에너지 효율적인 UAV 자율주

행을 위한 DRL 프레임워크를 제안하였다. 에너지 기반 경로 모델링과 웨이포인트 최적화, PPO 기반 경로 학습을 통해 UAV가 안전하고 효율적으로 목표 지점에 도달하는 것이 가능함을 보였다. 향후에는 다양한 환경 변화(예: 건물 밀도, 새와 같은 움직이는 장애물)에 강건한 정책 학습과 사용자 피드백을 반영한 학습 개선을 통해 실제 환경에 적용 가능한 범용 프레임워크로 발전시킬 예정이다.

### 참 고 문 헌

- [1] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the Sky: Leveraging UAVs for Disaster Management," IEEE Network, 2017.
- [2] M. Asadpour, et al., "A survey on UAV applications in civil engineering," Automation in Construction, 2021.
- [3] Y. Zhou, J. Wang, Y. Xu, "UAV path planning based on deep reinforcement learning," Sensors, 2019.
- [4] J. Schulman, et al., "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.
- [5] K. P. Valavanis and G. J. Vachtsevanos, \*Handbook of Unmanned Aerial Vehicles\*, Springer, 2015.