

드론의 경사면 착륙을 위한 PPO 기반 도메인 랜덤화 강화 학습 환경 및 보상함수 설계

이호성*¹, 안희석*², 김건우**

Domain Randomization Reinforcement Learning Environment and Reward Function Design for Drone Slope Landing based on PPO

Ho-Seong Lee*¹, Hui-Seok Ahn*², and Gun-Woo Kim**

본 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (NRF-2021R1G1A1006381) 및 2025년도 산업통상자원부 및 한국산업기술기술평가원(KEIT)의 연구비 지원(RS-2025-02633048)에 의한 연구결과임

요약

드론이 경사면에 착륙하는 것은 재난 구조나 군사 작전과 같은 예측 불가능한 환경에서 중요한 과제다. 전통적인 제어 방법은 경사면의 비선형 동역학을 다루는 데 한계가 있다. 본 연구는 PPO 알고리즘을 활용해 드론의 경사면 착륙을 훈련시키고, 위치, 속도, 방향 정렬을 위한 희소-밀집 결합 보상 함수와 도메인 랜덤화 기법을 적용한 0도에서 30도의 무작위 경사각 생성 훈련 환경을 구축했다. 실험 결과, 제안된 방법은 0도에서 30도의 경사면에서 99.67%의 일관된 착륙 성공률을 달성해 특정 각도에 과적합 되는 커리큘럼 러닝 방법보다 개선되었으며, PPO, SAC, A2C 알고리즘 비교에서도 PPO가 가장 우수했다. 본 연구는 커리큘럼 러닝 없이 일반화된 드론 착륙 학습 방법을 제시하며, 향후 3차원 환경 확장과 실제 조건 검증을 할 예정이다.

Abstract

Landing a drone on a slope is a critical task in unpredictable environments such as disaster relief or military operations. Traditional control methods have limitations handling nonlinear slope dynamics. This study utilized PPO to train slope landing, designing a combined sparse-dense reward function for position, velocity, and orientation alignment, and applying domain randomization with random slope angles ranging from 0 to 30 degrees. Results showed 99.67% landing success rate across the range of 0 to 30 degrees, outperforming curriculum learning methods that exhibited overfitting, with PPO superior to SAC and A2C algorithms. This research presents a generalized drone landing approach without curriculum learning, with future work on 3D environment extension and real-world validation.

Keywords

reinforcement learning, drone, PPO, autonomous landing, UAV, domain randomization

* 경상국립대학교 AI융합공학과
- ORCID¹: <https://orcid.org/0009-0005-6168-6450>
- ORCID²: <https://orcid.org/0009-0008-2668-5542>
** 경상국립대학교 컴퓨터공학과 교수(교신저자)
- ORCID3: <https://orcid.org/0000-0001-5643-4797>

· Received: Apr. 29, 2025, Revised: May 19, 2025, Accepted: May 22, 2025
· Corresponding Author: Gun-Woo Kim
Dept. of Computer Science and Engineering, College of IT Engineering,
Gyeongsang National University, Jinju, Korea
Tel.: +82-55-772-3323, Email: gunwoo.kim@gnu.ac.kr

I. 서 론

최근 드론 기술의 급격한 발전과 함께 다양한 환경에서 자율 비행 및 착륙 기술에 대한 수요가 증가하고 있다. 특히 재난 구조, 군사작전과 같이 복잡하고 예측 불가능한 환경에서 임무를 수행해야 하는 드론은 다양한 지형과 기상 조건에 적응하여 안정적으로 착륙할 수 있는 제어 기법이 요구된다 [1][2]. 전통적인 PID(Proportional-Integral-Derivate) 제어는 선형 시스템에서는 양호한 성능을 보이나, 경사면과 같은 각도, 재질, 풍속 등 다양한 비선형 요소가 작용하는 환경에서는 제어기 설계가 복잡해지는 한계가 있다[2]. 이러한 한계를 극복하기 위해 유연하고 적응력 있는 제어 전략이 필요하다. 강화 학습을 이용한 드론 제어는 기존 제어의 한계를 극복하기 위해 연구되고 있다[2]-[4].

기존 연구에서는 희소 보상 함수와 커리큘럼 러닝을 활용한 PPO(Proximal Policy Optimization) 알고리즘으로 경사면 착륙을 하였는데 다양한 각도가 아닌 특정 각도의 경사면에 대하여 검증이 되었다 [2][5]. 이 접근법은 초기에는 낮은 난이도에서 시작하여 점진적으로 난이도를 높이는 방식으로, 특정 경사각에서는 높은 성공률을 보였으나 새로운 각도에 대한 일반화 성능이 저하되는 과적합 문제가 확인되었다.

본 논문은 기존 연구[2]에 적용된 커리큘럼 러닝의 한계를 개선하고자, PPO 알고리즘 기반[6] 도메인 랜덤화 기법을 적용한 경사면 생성 환경 및 경사면 착륙을 위한 효과적인 보상 함수 설계로 다양한 각도에서 높은 정확도의 착륙 정책을 제안한다. 또한, 동일한 환경과 보상 함수 하에서 SAC[7] 및 A2C[8] 알고리즘과의 비교 실험을 통해 제안하는 방법의 알고리즘 선택이 중요함을 확인하였다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 드론 경사면 착륙에 관한 관련 연구를 소개한다. 3장에서는 제안하는 강화학습 기반 드론 자동 착륙 기법의 구체적 설계(경사면 학습 환경 및 보상 함수)에 대해 상세히 기술하고, 4장에서는 다양한 실험 결과를 바탕으로 기법의 성능을 평가 및 분석한다. 마지막으로 5장에서는 연구 결과를 종합하고 향후 연구 방향에 대해 논의한다.

II. 관련 연구

기존 드론 착륙 제어는 주로 PID 컨트롤러와 비전 기반 자세 추정 기술에 의존한다[1][2][9]. 이러한 방법은 평평한 표면에서의 착륙에 효과적이지만 경사면과 같은 복잡한 환경에서는 한계가 있다. 특히, 지형 정보가 부족하거나 동역학적 변화가 큰 경우, 이러한 기존 방법은 안정적인 착륙을 보장하기 어렵다[1][2].

강화학습은 드론 제어에서 많이 적용되고 있다 [2]-[4][10][11]. 이는 복잡한 환경에서 적응할 수 있어 경로계획, 내비게이션, 자세 제어 등 다양한 작업에 적용된다. DDPG(Deep Deterministic Policy Gradient)는 3D 도시 환경에서의 장애물 회피에서 충돌 0회를 달성했으며, PPO는 고정익 및 하이브리드 UAV의 자세 제어에 사용되었다[5][11][12]. 다른 연구에서는 DDQN(Double Deep Q-Network)과 JNN(Joint Neural Network)를 사용해 현실적인 이웃 환경(나무, 전선, 차량, 집 등 장애물 포함)에서 드론 배달을 훈련시켰다[4]. E. Kaufmann et al.[12]은 PPO 알고리즘을 사용해 드론 경주에서 인간 챔피언을 능가하는 성능을 달성하였다.

d'Apolito et al.[10]은 강화학습을 사용해 드론의 롤, 피치, 요, 스로틀을 제어하도록 훈련시켰다. Airsim 시뮬레이션 환경에서 안정적인 비행 조건을 달성하고, 정해진 경로를 따라 비행할 수 있음을 입증했다. 착륙에 대한 강화학습 제어 연구 이루어졌다[2][9][10]. 다른 연구에서는 움직이는 플랫폼에 대한 드론의 착륙에 강화학습을 적용하였다[13].

드론의 경사면 착륙을 위한 연구는 다양한 방법으로 진행되고 있다. J. Bass et al.[14][15]은 역추진(Reverse thrust)을 사용하여 경사면 착륙 성능을 개선하는 방법을 제안했다. 이 연구에서는 역추진을 통해 경사면에서의 착륙 가능 각도를 크게 향상시켰다.

J. E. Kooi et al.[2]은 PPO 알고리즘과 희소 보상, 커리큘럼 학습을 이용해 경사면 착륙 컨트롤러를 설계했다. 이 연구에서는 시뮬레이션에서 90분 이내에 훈련이 가능했으며, 실제 Crazyfly 2.1 드론으로 실험 시 성공률은 초기 위치에 따라 90%~100%로 나타났다.

Isaac Lab[16]-[18]은 로봇 시뮬레이션 연구에 사용되는 NVIDIA 플랫폼으로, 드론 시뮬레이션 연구에도 활발히 이용되고 있다[12]. Pegasus Simulator [19]는 Isaac Lab의 시뮬레이션 도구인 Isaac Sim 기반으로 개발되어 시뮬레이션과 실제 간의 간극을 줄이는 데 유용하다.

OmniDrones[20]은 Isaac Sim의 드론 시뮬레이터로, 실제 환경과의 비교를 통해 드론 시뮬레이터로서의 가능성을 입증하였다.

강화 학습 방법에서 커리큘럼 러닝은 모델을 쉬운 작업부터 시작해 점진적으로 어려운 작업으로 이동시키는 훈련 전략이다[5]. 이 접근법은 학습 효율성을 높일 수 있어 강화 학습에서 많이 사용된다. 기존 드론 경사면 착륙 연구에서도 경사면 착륙을 위해 커리큘럼 러닝을 활용하여 훈련 시간과 정책의 질을 개선했다[2].

본 연구는 기존 방법[2]으로는 일관적이지 않은 착륙 성공률을 개선하기 위해 새로운 방법을 제안한다. 경사면 착륙을 위한 보상 함수를 설계하고, 도메인 랜덤화 기법을 적용한 경사면 학습 환경을 구축하여 다양한 각도에서의 일관된 성공률을 달성한다. 기존 연구의 커리큘럼 러닝과 보상 함수를 구현하고 본 연구의 방법과 비교하여, 제안된 방법이 더 다양한 각도에서 일정한 성공률을 보이며 더 높은 정확도를 달성함을 증명한다.

III. 드론 경사면 착륙을 위한 도메인 랜덤화 적용 환경 구축 및 보상함수 설계

3.1 강화학습 문제 정의

본 연구에서 상태 공간(s)은 12차원 벡터로 구성되며, 드론 좌표계에서의 선속도(v_x, v_y, v_z), 각속도(w_x, w_y, w_z), 드론 좌표계 기준 중력 벡터(g_x, g_y, g_z), 그리고 목표 위치까지의 상대적 거리(d_x, d_y, d_z)를 포함한다. 2차원 환경에서의 학습을 위하여 y 축 방향 이동과 롤(roll), 요(yaw) 회전이 제한되었고 다음과 같다.

$$s_{2D} = [v_x, 0, v_z, 0, w_y, 0, g_x, 0, g_z, d_x, 0, d_z] \quad (1)$$

행동 공간(a)은 추력 명령($a_{T_{cmd}}$), 롤($a_{R_{cmd}}$), 피치($a_{P_{cmd}}$), 요의 회전 모멘트 명령어로 구성된다. 2차원 환경에서의 비교를 위해, 요 회전 모멘트 명령을 0으로 제한하였다.

$$a = [a_{T_{cmd}}, a_{R_{cmd}}, a_{P_{cmd}}, a_{Y_{cmd}}] \quad (2)$$

$$a_{2D} = [a_{T_{cmd}}, 0, a_{P_{cmd}}, 0] \quad (3)$$

여기서 각 요소는 드론의 수직 추력 명령($a_{T_{cmd}}$), 피치 회전 모멘트 명령($a_{P_{cmd}}$)을 나타내며, 범위는 $[-1, 1]$ 이다. 추력(T)과 모멘트(τ)는 다음 수식으로 계산된다.

$$T = k_T \cdot m \cdot g \frac{a_{T_{cmd}} + 1}{2} \quad (4)$$

$$\tau = k_\tau [a_{P_{cmd}}]^T \quad (5)$$

여기서 k_T 는 추력 대 중량비(1.9), m 은 질량, g 는 중력가속도이다. 계산된 추력과 모멘트는 Isaac Sim 내 PhysX 엔진에 적용되어 드론의 움직임을 시뮬레이션한다.

착륙 성공 판정을 위한 기준(δ_{goal})은 다음과 같다.

$$\delta_{goal} = [\delta_x, \delta_z, \delta_{v_x}, \delta_{v_z}, \delta_\theta] \\ = [0.1, 0.1, 1.5, 1.5, 0.025] \quad (6)$$

각 요소는 위치(m), 속도(m/s), 각도(rad)에 대한 허용 오차를 나타낸다. 드론이 이 기준 내에 위치하면 착륙 성공으로 판정한다.

3.2 보상함수 설계

본 연구에서는 희소 보상함수와 밀집 보상함수를 결합한 보상함수를 설계하였다. 이는 학습 효율성과 정확한 착륙 성능을 동시에 확보하기 위함이다.

밀집 보상 함수를 구성하는 보상 요소는 위치 오차($p - p_{goal}$)에 대한 보상($R_{position-error}$), 속도(v)에

대한 보상 ($R_{velocity-error}$), 드론의 자세와 경사면에 대한 각도 차이(θ_{error})에 대한 보상($R_{orient-error}$)으로 구성하였다. 매 스텝(Δt) 마다 드론의 위치(p)가 착륙 위치(p_{goal})에 가까울수록, 속도(v)가 작을수록, 경사면과 드론이 이루는 각도 차이(θ_{error})가 작을수록 보상을 주어 착륙 요구도를 만족하게 할 수 있도록 유도하였고 식은 다음과 같다:

$$R_{position-error} = -1.5 \| p - p_{goal} \| \Delta t \quad (7)$$

$$R_{velocity-error} = -0.05 \| v \| \Delta t \quad (8)$$

$$R_{orient-error} = -0.01 \theta_{error} \Delta t \quad (9)$$

최소 보상 요소로는 드론 상태(s_{drone})의 착륙 성공(S_{goal}) 여부에 따른 보상($R_{success}$), 경계 이탈(S_{out}) 여부에 따른 패널티(R_{out}), 충돌($S_{collision}$) 여부에 따른 패널티($R_{collision}$)로 구성된다.

$$R_{success} = \begin{cases} 5 & \text{if } s_{drone} \in S_{goal} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$R_{out} = \begin{cases} -5 & \text{if } s_{drone} \in S_{out} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$R_{collision} = \begin{cases} -5 & \text{if } s_{drone} \in S_{collision} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

보상 함수에 대한 각 요소 가중치 -1.5, -0.05, -0.01 및 최소 보상 값 5, -5는 시행착오를 통한 실험적 튜닝 과정을 통해 결정하였다.

모든 요소를 결합한 최종 보상 함수(R_{total})는 다음과 같다:

$$R_{total} = R_{position-error} + R_{velocity-error} + R_{orient-error} + R_{success} + R_{out} + R_{collision} \quad (13)$$

이러한 보상함수 설계는 드론이 단순히 착륙 성공 여부에만 의존하지 않고, 착륙 과정 전반에 걸쳐 지속적인 피드백을 받을 수 있게 한다. 특히 도메인

랜덤화 환경에서는 초기 착륙 성공 경험을 얻기 어려울 수 있으므로, 이러한 밀집 보상 요소가 학습 초기부터 의미 있는 정책 발전을 가능하게 한다. 또한 성공, 경계 이탈, 충돌에 대한 최소 보상을 통해 최종 목표를 명확히 인식할 수 있게 함으로써, 보다 효율적이고 정확한 착륙 정책을 학습할 수 있도록 하였다.

3.3 강화학습 알고리즘 적용

본 연구에서는 주 알고리즘으로 PPO를 사용하였다. PPO 알고리즘의 주요 파라미터는 다음과 같이 설정하였다:

표 1. PPO 학습의 하이퍼 파라미터

Table 1. Hyperparameters for Training PPO

Hyperparameters	Value
Mini batch size	245746
Hidden layer	[64, 64], [64, 64]
Discount factor	0.99
Clipping	0.2

제안하는 PPO 기반 도메인 랜덤화 환경 및 보상 함수 설정 하에서 강화 학습 알고리즘 선택의 영향성을 확인하고자, SAC(Soft Actor-Critic)[7]와 A2C (Advantage Actor-Critic)[8] 알고리즘의 학습 성능을 동일 조건에서 함께 평가하였다.

3.4 학습 및 평가 방법

학습은 도메인 랜덤화가 적용된 환경에서 진행되었으며, 커리큘럼 러닝과의 비교를 위해 기존 연구 [2]의 방법론도 동일한 환경에서 구현하였다. 평가는 0도부터 30도까지의 다양한 경사각에서 각 알고리즘과 학습 방법의 착륙 성공률을 측정하는 방식으로 진행되었다.

3.4.1 도메인 랜덤화 학습 방법

도메인 랜덤화 학습 과정에서는 각 에피소드마다 드론 및 경사각에 대한 요소를 무작위화하였다. 구

체적으로 경사각은 0도에서 30도 범위 내에서 무작위로 설정하였으며, 드론의 초기 위치는 x 축으로 $-1.0m \sim 1.0m$, y 축으로 $0.41m \sim 1.0m$ 범위 내에서 무작위로 배치하였다.

이러한 랜덤화 과정을 통해 드론이 학습 초기 단계부터 다양한 경사각과 초기 위치 조건에 지속적으로 노출됨으로써, 특정 환경에 과적합되지 않고 다양한 상황에 적응 가능한 일반화된 정책을 학습할 수 있도록 하였다.

3.4.2 평가 지표

학습된 모델의 성능 평가를 위해 각 에피소드 진행 시 착륙 요구도를 만족 여부를 기록하여 착륙 성공률을 계산하였다. 에피소드는 1×10^5 회 이상 수행하여 충분히 성공률이 수렴할 수 있도록 하였다.

IV. 실험 및 결과

본 실험에서는 다양한 강화학습 방법을 사용한 드론의 경사면 착륙 성능을 평가하고 비교하였다. 모든 실험은 동일한 환경 설정에서 진행되었다. 각 알고리즘 및 학습 방법의 성능은 착륙 성공률을 기준으로 평가하였다.

4.1 시스템 구성 및 환경 설계

본 연구에서는 드론의 경사면 착륙 문제를 강화학습으로 해결하기 위해 Isaac Lab 기반의 시뮬레이션 환경을 구축하였다[16]-[18]. Isaac Lab은 PhysX 물리엔진 기반으로 로봇 시뮬레이션 및 강화학습에 이용되는 프레임워크다[18]-[20]. 기존 연구[2]에서 사용된 커리큘럼 러닝 방식과 달리, 다양한 경사각에 일반화된 착륙 성능을 보장하기 위해 도메인 랜덤화 기법을 시뮬레이션에서 각 에피소드 시작 시 적용시켰다.

4.1.1 시뮬레이션 환경

본 연구에서 사용된 시뮬레이션 소프트웨어 및 하드웨어 사양은 표 2, 표 3과 같다.

표 2. 소프트웨어 사양

Table 2. Software specifications

Category	Specifications
Simulation platform	Isaac Sim[4.5]
RL Tool-Kit	Isaac Lab[2.0]
RL Library	rl-games[1.6.1], skrl[1.4.1]

표 3. 하드웨어 사양

Table 3. Hardware specifications

Category	Specifications
CPU	Intel i9-13900K
GPU	RTX 4090
Memory	64GB

시뮬레이션 환경은 Isaac Lab의 Multi Environment 기능을 활용하여 병렬 학습이 가능하도록 설계하였다. 각 환경은 드론과 경사면으로 구성되며, 드론 모델은 기존 연구에서도 사용된 Crazyflie 모델(Isaac Sim에서 제공하는 cfx.usd)을 사용하였다. 경사면은 Isaac Lab API를 통해 생성하고 중앙에 붉은색 마커를 배치하여 착륙 목표 점으로 활용하였다. 그림 1은 연구를 위해 개발된 경사면 착륙 시뮬레이터이며 다양한 경사면 조건과 드론 위치에 대하여 학습하는 장면이다.

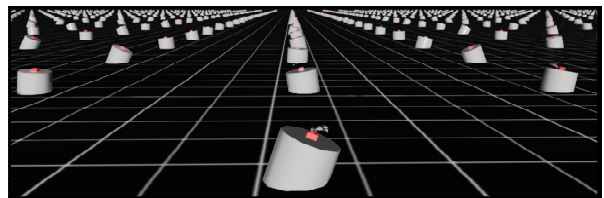


그림 1. 경사면 착륙 강화학습을 위한 시뮬레이터
Fig. 1. Simulator for reinforcement learning of slope landing

4.1.2 도메인 랜덤화 환경 설계

본 연구에서는 두 가지 도메인 랜덤화 기법을 적용했다. 첫째, 각 에피소드 시작 시 경사각을 0도에서 30도 범위에서 무작위로 생성하였다. 둘째, 각 에피소드 시작 시 드론의 초기 위치를 x 축 방향으로 $-1.0m \sim 1.0m$, z 축 방향으로 $0.41m \sim 1m$ 범위에서 무작위로 생성 되도록 하였다.

4.2 밀집 및 희소 보상 함수를 이용한 커리큘럼 러닝 기반 착륙 학습

본 연구에서 제안하는 밀집 및 희소 보상 함수에 기존 연구 방법[2]을 적용하여 기존 방법의 한계점을 확인하기 위한 실험을 진행하였다. 기존 연구[2]는 커리큘럼 러닝 기반으로 경사각은 0도에서 시작하여 점진적으로 30도까지 증가시켰다. 드론의 초기 위치는 착륙 목표 지점 근처에서 시작하여 훈련이 진행됨에 따라 탐색 범위를 점차 확대하였다. 성공 기준 또한 초기에는 상대적으로 완화된, 달성하기 쉬운 조건을 적용한 후 점진적으로 엄격하게 조정하여 학습의 난이도를 조절하였다. 그림 2는 학습 진행 후 테스트 시 각 경사각별 착륙 성공률을 보여준다. 30도 경사각에서 가장 높은 성공률을 보였으나, 전체적으로 고르지 않은 성공률 분포가 확인되었다.

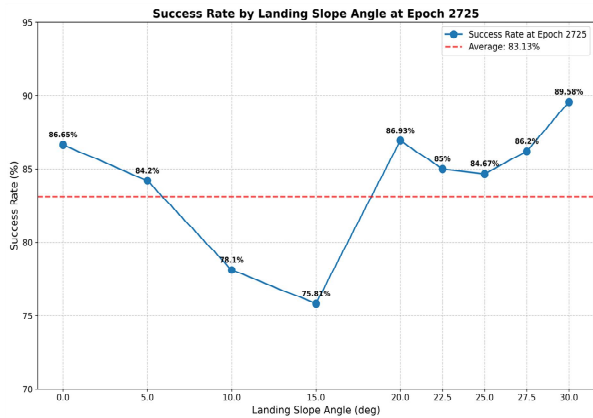


그림 2. 밀집, 희소 조합 보상 함수를 적용한 경사각에 따른 커리큘럼 러닝 학습의 착륙 성공률
 Fig. 2. Landing success rate of curriculum learning training according to slope angle with combined dense and sparse reward functions

4.3 도메인 랜덤화를 이용한 착륙 학습

본 연구의 제안인 도메인 랜덤화 접근법은 각 에피소드마다 0도에서 30도 범위의 무작위 경사각과 다양한 초기 위치를 생성하여 학습을 진행하였다. 그림 3은 학습 진행에 따른 보상 값과 착륙 성공률의 변화를 보여준다. 주목할 점은 기존 연구에서 고정되지 않은 착륙 조건 때문에 학습 성공의 지표로 사용될 수 없었던 점과 달리 도메인 랜덤화 환경

에서는 보상 값이 학습 지표로 효과적으로 사용될 수 있음을 확인하였다.

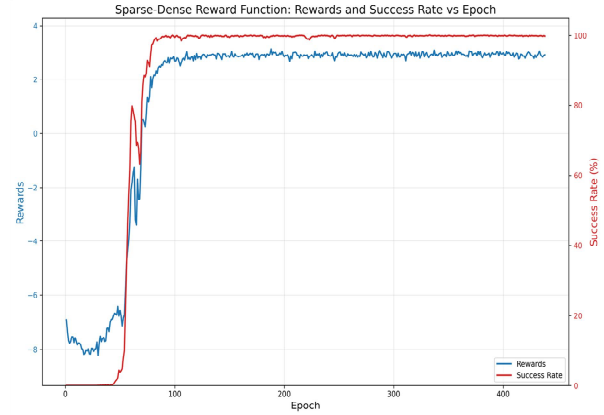


그림 3. 학습에 따른 보상 및 착륙 성공률 변화
 Fig. 3. Changes in reward and landing success rate according to training

그림 4은 학습된 정책의 각 경사각별 성공률을 보여주는데, 평균 99.67%의 착륙 성공률을 달성하며 모든 경사각에서 일관되게 높은 성능을 보였다. 이는 도메인 랜덤화 접근법이 다양한 경사각에 대한 일반화 성능을 효과적으로 향상시켰음을 입증한다.

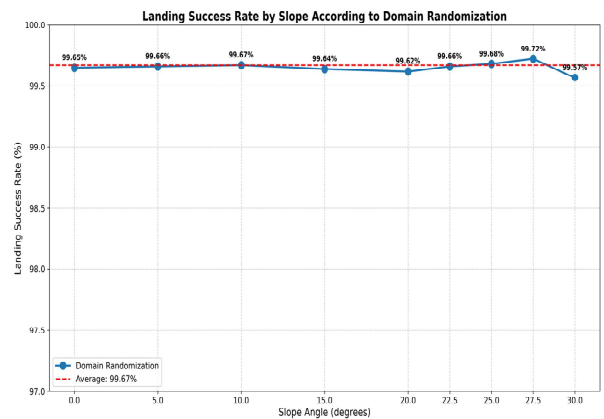


그림 4. 도메인 랜덤화 적용 시 착륙 경사각 별 성공률
 Fig. 4. Success rate by landing slope angle when applying domain randomization

그림 5은 본 논문에서 제안하는 동일한 보상 함수 적용한 커리큘럼 러닝 기법과 도메인 랜덤화 기법의 경사각별 착륙 성공률이다.

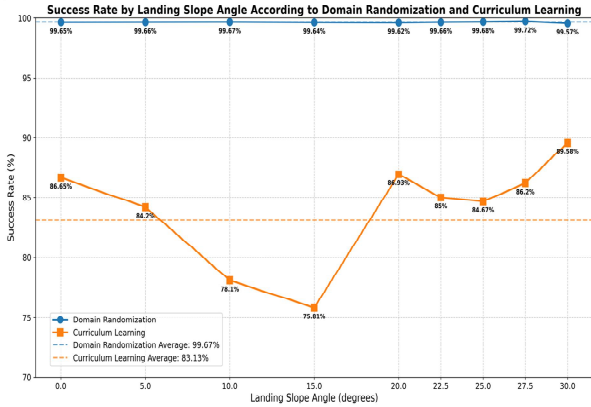


그림 5. 도메인 랜덤화, 커리큘럼 러닝에 따른 착륙 경사각 별 성공률

Fig. 5. Success rate by landing slope angle according to domain randomization and curriculum learning

같은 보상 함수를 적용했음에도 불구하고 커리큘럼 러닝 기법에서 망각 현상과 낮은 성공률을 보여준다. 커리큘럼 학습은 학습 스케줄, 학습 난이도 설정 등을 조절하면 보다 높은 성공률을 기록 할 수 있겠으나 도메인 랜덤화 대비 많은 시행착오를 필요로 할 것이다.

4.4 희소 보상 함수의 도메인 랜덤화 적용 학습

기존 연구[2]의 희소 보상 함수만을 도메인 랜덤화 환경에 적용한 실험도 진행하였다. 그림 6는 이 설정에서의 학습 진행에 따른 보상 값과 착륙 성공률 변화를 보여준다. 학습이 진행됨에 따라 에이전트가 얻는 보상 값은 점진적으로 증가했으나 최종 착륙 성공률은 0%로 나타났다. 이는 복잡한 제어 작업에서 희소 보상 함수만으로는 효과적인 학습이 어렵다는 것을 확인시켜주며, 특히 도메인 랜덤화 환경에서는 다양한 초기 조건으로 인해 성공 경험 획득이 더욱 어렵기 때문이다.

표 4는 0도에서 30도 범위의 경사면에 대한 학습 방법 및 보상 함수별 평균 착륙 성공률을 요약하여 보여준다. 도메인 랜덤화와 희소-밀집 보상 함수를 이용한 본 연구의 제안 방법이 99.67%로 가장 높은 성공률을 달성했으며, 커리큘럼 러닝을 적용한 희소-밀집 보상함수의 도메인 랜덤화 학습에서도 83.13%의 준수한 성공률을 보였다. 반면 커리큘럼 러닝 방식을 적용한 희소보상 함수의 학습은 4.41%의 저조한 성공률을 기록했다.

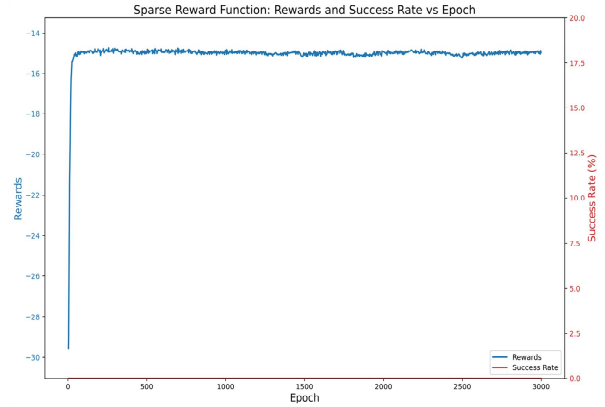


그림 6. 기존 연구의 희소 보상 함수 및 도메인 랜덤화 적용 시 착륙 성공률 변화

Fig. 6. Changes in landing success rate when applying sparse reward function and domain randomization from previous research

표 4. 0에서 30도 범위의 경사면에 대한 학습 방법 및 보상 함수에 대한 평균 착륙 성공률

Table 4. Average landing success rate for learning methods and reward functions on slopes from 0 to 30 degrees

Learning methods \ Reward fuctions	Sparse	Sparse - Dense
	Curriculum learning	4.41%
Domain randomization	0%	99.67%

이러한 결과는 도메인 랜덤화를 통한 일반화 능력 향상과 효과적인 학습을 동시에 달성하기 위해서는 밀집 보상과 희소 보상의 결합이 효과적임을 알 수 있다.

4.5 강화학습 알고리즘 별 성능 평가

본 연구에서는 제안하는 학습 환경과 보상 함수가 다양한 강화학습 알고리즘에 미치는 영향도 평가하였다.

PPO, SAC, A2C 알고리즘을 동일한 조건에서 비교한 결과, 그림 7에서 볼 수 있듯이 PPO가 모든 경사각에서 일관되게 높은 성능을 보였다. 표 5에 정리된 바와 같이, PPO는 99.67%의 평균 성공률을 달성하여 SAC(41.32%)와 A2C(6.64%)보다 높은 성능을 보였다.

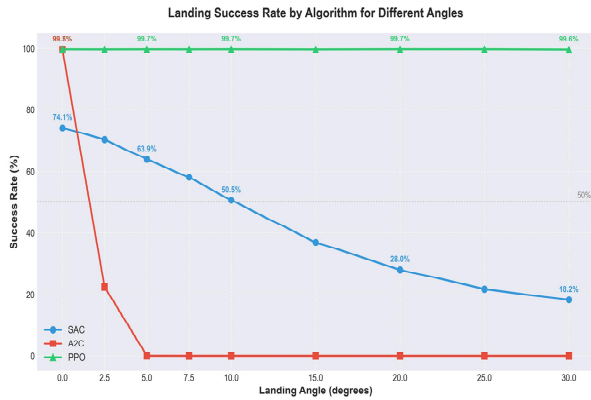


그림 7. PPO, SAC, A2C 별 경사각에 따른 착륙 성공률
Fig. 7. Landing success rate by slope angle for PPO, SAC, and A2C

표 5. 알고리즘 별 평균 착륙 성공률
Table 5. Average landing success rate by algorithm

Algorithm	Average landing success rate
PPO	99.67%
SAC	41.32%
A2C	6.64%

이는 PPO의 정책 최적화 방식이 드론 경사면 착륙 문제의 복잡한 동역학을 효과적으로 학습하는데 더 적합함을 시사한다.

종합적으로, 본 실험 결과는 도메인 랜덤화 기법과 복합 보상 함수의 결합이 다양한 경사각에서 일관되고 우수한 착륙 성능을 달성하는 데 효과적임을 입증하였다. 또한 PPO 알고리즘이 이러한 학습 환경에서 적합한 선택임을 확인하였다. 이는 경사면 착륙에서 커리큘럼 러닝과 같은 복잡한 학습 스케줄링 없이도 희소-밀집 보상함수를 적용한 도메인 랜덤화만으로 효과적인 일반화 성능을 달성할 수 있음을 보여 준다.

V. 결론 및 향후 과제

본 연구에서는 PPO 기반 도메인 랜덤화 강화 학습을 활용하여 드론의 경사면 착륙 문제를 해결하는 새로운 방법을 제안하였다. 연구 결과를 종합하면 다음과 같다.

첫째, 위치, 속도, 방향 정렬을 유도하기 위한 희소 보상과 연속 보상을 결합한 새로운 보상 함수를

설계하였다. 이 보상 함수는 드론이 착륙 과정 전반에 걸쳐 지속적인 피드백을 받을 수 있게 하여 희소 보상만으로는 달성하기 어려운 복잡한 제어 작업에서도 안정적인 학습이 가능함을 입증하였다.

둘째, 일반화 성능을 강화하기 위해 매 에피소드마다 0도에서 30도 범위의 무작위 경사각을 생성하는 도메인 랜덤화 학습 환경을 구축하였다. 이는 기존 커리큘럼 러닝 방식에서 나타난 특정 각도에 과적합되는 문제를 효과적으로 해결하였다. 실험 결과 제안된 방법은 0도에서 30도 범위의 경사면에서 99.67%의 일관된 착륙 성공률을 달성하여, 커리큘럼 러닝 기반 방법보다 향상된 성능을 보였다.

셋째, 다양한 강화학습 알고리즘(PPO, SAC, A2C)을 동일한 환경에서 비교 평가한 결과, PPO 알고리즘이 99.67%의 성공률로 다른 알고리즘들보다 드론 경사면 착륙 문제에 더 적합함을 확인하였다.

본 연구는 커리큘럼 러닝과 같은 복잡한 학습 스케줄링 없이도 도메인 랜덤화와 효과적인 보상 함수 설계만으로 다양한 경사각에서 일반화된 드론 착륙 정책을 학습할 수 있는 방법을 제시한다.

그러나 본 연구에는 몇 가지 한계점이 존재한다. 첫째, 현재 연구는 시뮬레이션 환경에 국한되어 있어 실제 환경의 불확실성과 외부 요인(바람, 센서 오차 등)이 충분히 반영되지 않았다. 둘째, 비교를 위해 2차원 환경에서 실험을 진행하여 3차원 공간에서의 드론 운동 역학이 완전히 고려되지 않았다. 마지막으로, 특정 드론 모델(Crazyflie)에 최적화된 설계로 다양한 드론 플랫폼에 대한 일반화 가능성은 추가 검증이 필요하다.

향후 연구에서는 다음과 같은 방향으로 연구를 확장할 계획이다.

첫째, 현재의 2차원 환경을 3차원으로 확장하여 보다 현실적인 드론 운동 모델을 적용할 것이다.

둘째, 실제 드론 하드웨어에서의 검증을 통해 시뮬레이션-현실 간 격차(Sim-to-real gap)를 분석하고 해결 방안을 모색할 것이다.

셋째, 바람이나 센서 노이즈와 같은 외부 요인을 도메인 랜덤화에 추가하여 보다 강건한 착륙 정책을 개발할 것이다.

결론적으로, 본 연구는 도메인 랜덤화와 효과적

인 보상 함수 설계를 통해 드론의 경사면 착륙 문제에 대한 강화학습 적용 가능성을 보였으며 이는 향후 더 복잡한 환경에서의 자율 드론 제어 연구의 기반을 마련할 것으로 기대된다.

References

- [1] T. Iizuka, T. Sasatani, T. Nakamura, N. Kosaka, M. Hisada, and Y. Kawahara, "An Inclination Estimation Method for UAV Landing Surfaces Using Millimeterwave Radar", Proc. IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, pp. 2049-2052, Jul. 2023. <https://doi.org/10.1109/IGARSS52108.2023.10282437>.
- [2] J. E. Kooi and R. Babuska, "Inclined Quadrotor Landing using Deep Reinforcement Learning", Proc. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, pp. 2361-2368, Sep. 2021. <https://doi.org/10.1109/IROS51168.2021.9636096>.
- [3] A. T. Azar, A. Koubaa, N. A. Mohamed, H. A. Ibrahim, Z. F. Ibrahim, M. A. Kazim, A. Ammar, B. Benjdira, A. M. Khamis, I. A. Hameed, and G. Casalino, "Drone Deep Reinforcement Learning: A Review", Electronics, Vol. 10, No. 9, pp. 999-1028, May 2021. <https://doi.org/10.3390/electronics10090999>.
- [4] G. Muñoz, C. Barrado, E. Çetin, and E. Salami, "Deep Reinforcement Learning for Drone Delivery", Drones, Vol. 3, No. 3, pp. 72-90, Sep. 2019. <https://doi.org/10.3390/drones3030072>
- [5] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning", Proc. 26th International Conference On Machine Learning (ICML 2009), Montreal Quebec, Canada, pp. 41-48, Jun. 2009. <https://doi.org/10.48550/arXiv.2101.10382>.
- [6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms", arXiv:1707.06347, Jul. 2017. <https://doi.org/10.48550/arXiv.1707.06347>.
- [7] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor", Proc. International Conference on Machine Learning (ICML), Stockholm, Sweden, Jul. 2018. <https://doi.org/10.48550/arXiv.1801.01290>.
- [8] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous Methods for Deep Reinforcement Learning", Proc. International Conference on Machine Learning (ICML), New York, USA, Jun. 2016. <https://doi.org/10.48550/arXiv.1602.01783>.
- [9] L. Mu, Q. Li, X. Xue, W. Sun, B. Wang, Y. Zhang, and N. Feng, "A Vision-Based Autonomous Landing Guidance Strategy for a Micro-UAV by the Modified Camera View", Drones, Vol. 7, No. 6, Jun. 2023. <https://doi.org/10.3390/drones7060400>.
- [10] F. d'Apolito and C. Sulzbachner, "Flight Control of a Multicopter using Reinforcement Learning", IFAC PapersOnLine, Vol. 54, No. 13, pp. 251-255, Sep. 2021. <https://doi.org/10.1016/j.ifacol.2021.10.454>.
- [11] S. G. Park and D. H. Kim, "Autonomous Flight of Drone using PPO-based Reinforcement Learning", Journal of Institute of Control, Robotics and Systems, Vol. 26, No. 11, pp. 955-963, Nov. 2020. <https://doi.org/10.5302/J.ICROS.2020.20.0125>.
- [12] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, "Champion-level drone racing using deep reinforcement learning", Nature, Vol. 620, No. 7976, pp. 982-987, Aug. 2023. <https://doi.org/10.1038/s41586-023-06419-4>.
- [13] P. Goldschmid and A. Ahmad, "Reinforcement Learning based Autonomous Multi-Rotor Landing on Moving Platforms", arXiv:2302.13192, Feb.

2023. <https://doi.org/10.48550/arXiv.2302.13192>.
- [14] J. Bass and A. L. Desbiens, "Improving Multirotor Landing Performance on Inclined Surfaces Using Reverse Thrust", IEEE Robotics and Automation Letters, Vol. 5, No. 4, pp. 5850-5857, Oct. 2020. <https://doi.org/10.1109/LRA.2020.3010208>.
- [15] J. Bass, I. Tunney, and A. L. Desbiens, "Adaptative Friction Shock Absorbers and Reverse Thrust for Fast Multirotor Landing on Inclined Surfaces", IEEE Robotics and Automation Letters, Vol. 7, No. 3, pp. 6701-6708, Jul. 2022. <https://doi.org/10.1109/LRA.2022.3176102>.
- [16] NVIDIA, "Isaac Lab: A Unified Framework for Robot Learning", <https://developer.nvidia.com/blog/fast-track-robot-learning-in-simulation-using-nvidia-isaac-lab>. [accessed: Mar. 03, 2025]
- [17] N. Rudin, D. Hoeller, M. Hutter, P. Reist, and M. Hutter, "Learning to Walk in Minutes Using Massively Parallel Deep Reinforcement Learning", 5th Conference on Robot Learning, CoRL 2021, London, UK, pp. 91-100, Nov. 2021. <https://doi.org/10.48550/arXiv.2109.11978>.
- [18] S. H. Jang, J. H. Park, and S. H. Ok, "Research on Parallelization of Reinforcement Learning for Robots using Isaac Gym", Proc. of KIIT Conference, Jeju, Korea, pp. 624-627, Nov. 2023.
- [19] M. Jacinto, J. Pinto, J. Patrikar, J. Keller, R. Cunha, S. Scherer, and A. Pascoal, "Pegasus Simulator: An Isaac Sim Framework for Multiple Aerial Vehicles Simulation", Proc. 2024 International Conference on Unmanned Aircraft Systems (ICUAS), Chania - Crete, Greece, pp. 917-922, Jun. 2024. <https://doi.org/10.1109/ICUAS60882.2024.10556959>.
- [20] B. Xu, F. Gao, C. Yu, R. Zhang, Y. Wu, and Y. Wang, "OmniDrones: An Efficient and Flexible Platform for Reinforcement Learning in Drone Control", IEEE Robotics and Automation Letters, Vol. 9, No. 3, pp. 2838-2844, Mar. 2024. <https://doi.org/10.1109/LRA.2024.3356168>.

저자소개

이 호 성 (Ho-Seong Lee)



2017년 2월 : 경상대학교
항공우주시스템공학과(공학사)
2017년 1월 ~ 현재 :
(주)한국항공우주산업 디지털
엔지니어링팀 선임연구원
관심분야 : Aerospace Digital
Twin, Model-Base System
Engineering, Reinforcement Learning

안 희 석 (Hui-Seok Ahn)



2009년 2월 : 경상대학교
기계항공공학부(공학사)
2018년 6월 ~ 현재 :
(주)한국항공우주산업 디지털
엔지니어링팀 책임연구원
관심분야 : 항공우주 분야 AI응용,
산업용 지능형 시스템 설계, AI
기반 항공기 설계 및 운용 최적화

김 건 우 (Gun-Woo Kim)



2006년 12월 : 호주뉴캐슬대학교
컴퓨터공학과(공학사)
2007년 9월 : 호주뉴캐슬대학교
정보공학과(공학석사)
2017년 8월 : 한양대학교
컴퓨터공학과(공학박사)
2021년 9월 ~ 현재 :
경상국립대학교 컴퓨터공학과 부교수
관심분야 : 인공지능, 시멘틱 헬스케어, 데이터마이닝