Journal of Korea Robotics Society (2026) 21(1):026-032
https://doi.org/10.7746/jkros.2026.21.1.026 ISSN: 1975-6291 / eISSN: 2287-3961

불확실성 추정을 통한 강화학습 기반 실내 탐색 개선

Improving Reinforcement Learning-Based Indoor Exploration Through Uncertainty Estimation

하 준 수[1] · 이 병 호[2] · 김 영 훈[1] · 심 성 준[3] · 서 덕 현[3] · 박 종 우†
Junsu Ha[1], Byeongho Lee[2], Younghun Kim[1], Sungjun Shim[3],
Duckhyun Suh[3], Frank Chongwoo Park†

Abstract: Efficient exploration of indoor environments, such as office buildings, is essential for applications in robotics, surveillance, and emergency response. Reinforcement learning (RL) based approaches, such as Proximal Policy Optimization (PPO), often fail to achieve sufficient area coverage due to the complexity of the environment or inappropriate reward design. To address this limitation, we propose a novel algorithm that combines a generative model to create uncertainty maps of unexplored regions. When the RL policy underperforms, the additional model guides the agent toward areas with the highest uncertainty, thereby reducing the overall uncertainty of the map. Experimental results show that our method significantly improves exploration efficiency, highlighting the potential of combining reinforcement learning with uncertainty estimation to achieve superior performance in complex indoor environments.

Keywords: Reinforcement Learning, Indoor Exploration, Uncertainty Estimation

1. 서 론

현대 사회에서 모바일 로봇의 효율적인 실내 탐색 능력은 다양한 응용 분야에서 중요한 요소로 부각되고 있다. 오피스 빌딩, 병원, 쇼핑몰 등 실내 공간에서의 탐색은 로봇을 이용한 재난 대응, 보안 그리고 군사 작전 등 여러 분야에서 필수적인 기술이 되었다. 이러한 실내 탐색 작업은 공간의 구조적 복잡성, 제한된 탐색 시간 등 다양한 도전 과제를 수반하며, 이를 효과적으로 해결하기 위한 고도화된 탐색 알고리즘의 개발이 요구된다.

특히 최근에는 기계학습(Machine learning)의 발전에 힘입어 심층 강화학습(Deep Reinforcement Learning, DRL)이 최적의 탐색 알고리즘을 학습하는 강력한 도구로 자리매김하고 있다[1-3]. 더 나아가, SLAM (Simultaneous Localization and Mapping)과의 결합을 통해 지도를 작성함과 동시에 미지의 환경을 능동적으로 탐색하는 방법에 대한 연구 역시 활발히 이루어지고 있다[4]. 하지만 DRL 기반의 방법이 항상 적절한 action을 생성하는 것은 아니며, 높은 성능을 얻기 위해 주어진 환경과 목표에 적절한 보상 함수가 설계되어야 한다[1].

이러한 본질적인 불안정성으로 인해 강화학습을 이용할 경우 벽에 충돌하는 등의 지양해야 할 action이 생성 될 수 있다. 이를 해결하기 위해 특정 constraint를 만족함으로써 충돌을 회피하는 등의 안정성을 보장하는 policy를 학습하는 연구들 역시 활발히 진행중이다[5].

본 논문에서는 강화학습을 이용한 실내 탐색을 더욱 효율적으로 할 수 있도록 개선하는 알고리즘을 새롭게 제안한다. 만약 강화학습을 이용한 실내 탐색이 모종의 이유로 실패할 경우, 탐색하지 않은 영역을 추정하고, 불확실성을 수치화 하여 그 수치가 가장 큰 방향으로 추가 탐색을 진행한다. 불확실성 추정과 이를 이용한 탐색 알고리즘의 구체적인 방법은 추후에 2장과 3장에서 다룬다.

Received : Sep. 4. 2025; Revised : Oct. 6. 2025; Accepted : Nov. 3. 2025
※ This work was supported by Korea Research Institute for defense Technology planning and advancement (KRIT) grant funded by the Korea government (DAPA (Defense Acquisition Program Administration)) since 2022 (No.KRIT-CT-22-006-005, Control technology for collective operation of military ultra-small ground robots (Contribution rate: 100%)).
1. Ph.D. Candidate, Mechanical Engineering, Seoul National University, Seoul, Korea (hajunsu, yhun@robotics.snu.ac.kr)
2. Staff Engineer, Samsung Electronics, Seoul, Korea (bhlee@robotics.snu.ac.kr)
3. Researcher, Intelligent/Autonomous Control SW Team of LIG Nex1, Seongnam, Korea (sungjun.shim, duckhyun.suh@lignex1.com)
† Associate Professor, Corresponding author: Mechanical Engineering, Seoul National University, Seoul, Korea (fcp@snu.ac.kr)

본 논문은 다음과 같은 구조로 구성된다. 우선, 본 연구에서의 불확실성이란 무엇인지 정의하고, 생성 모델을 이용해 이를 정량화 하는 방법에 대해 알아본다. 이어서, 정량화된 불확실성을 이용해 강화학습 탐색 알고리즘이 실패할 경우 이를 개선할 수 있는 방법을 제안한다. 이후, 다양한 실내 환경을 생성할 수 있는 시뮬레이션 환경에서 실험을 통해 제안된 방법의 유효성을 평가 및 분석한다. 마지막으로, 결론과 함께 연구의 한계점과 개선 방안에 대하여 고찰한다.

2. 불확실성 추정 알고리즘

본 논문에서 사용하는 지도는 다음의 공통 색상 규칙을 따른다: 검은색=미탐색, 빨간색=벽, 파란색=복도, 노란색=방, 초록색=문. 이 규칙은 모든 Figure의 지도에 적용된다.

이 장에서는 실내 탐색 과정 중 탐색하지 않은 영역의 불확실성을 추정하고, 이를 기반으로 어떤 영역을 우선적으로 탐색해야 하는지 결정하는 방법에 대하여 알아본다. 특히, 다양한 불확실성 측정 방법 중, 생성 모델을 이용한 샘플링 기반 방법에 대하여 자세히 알아본다.

2.1 샘플링 기반 불확실성 추정 방법

딥러닝에서의 불확실성은 데이터 자체의 본질적인 변동성과 노이즈로 인해 발생하는 알레아토릭 불확실성(Aleatoric Uncertainty)과 데이터의 한계에 의해 발생하는 에피스테믹 불확실성(Epistemic Uncertainty)으로 나뉜다. 본 논문에서는 실내를 충분히 탐색하지 않음으로써 발생하는 에피스테믹 불확실성을 줄이는 방향으로 추가적인 탐색을 진행하는 적응적 탐색 전략(Adaptive Exploration Strategy)를 사용할 것이다. 이를 위해서는 불확실성 추정(Uncertainty Estimation)이 필수적이며, 본 논문에서는 생성 모델을 이용한 샘플링 기반 접근법을 사용한다.

본 논문에서 사용하는 조건부 생성 모델 G는 다음과 같이 조건 c와 노이즈 z를 입력으로 받아 새로운 이미지 x를 생성하는 모델이다.

$$x = G(z; c) \tag{1}$$

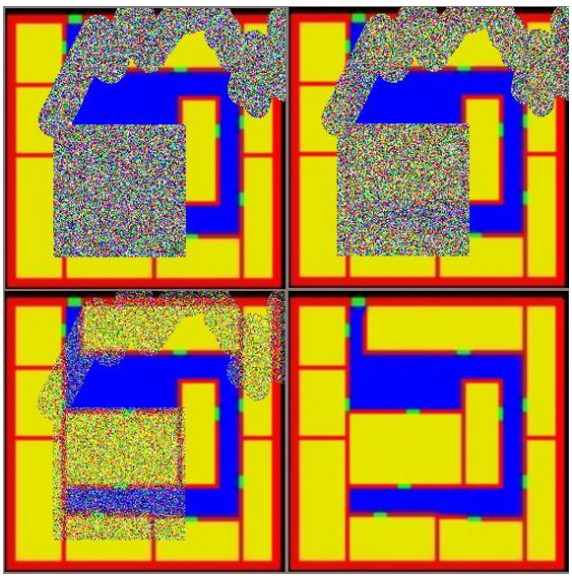
주어진 조건 c에 대하여 정규분포로부터 샘플링한 노이즈 $z_i \sim \mathcal{N}(0, I)$를 입력으로 사용하여 다양한 이미지 x_i를 생성할 수 있다.

$$x_i = G(z_i; c) \tag{2}$$

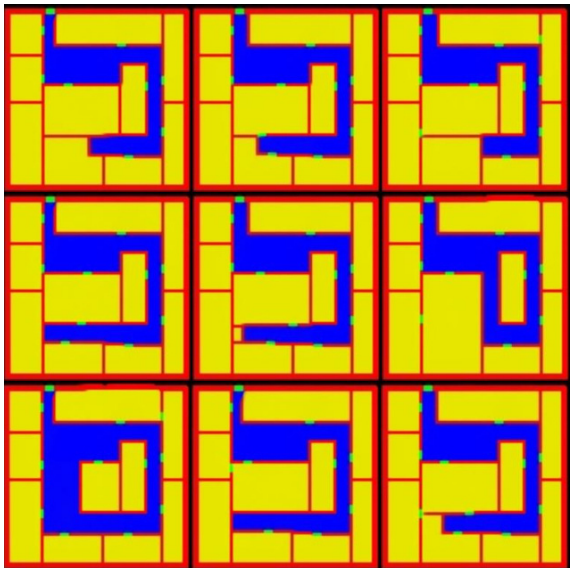
이렇게 생성된 이미지들의 분산을 구함으로써 각 픽셀의 불확성을 다음과 같이 수치화 할 수 있다.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3}$$

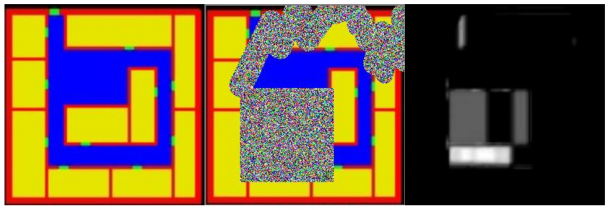
$$\mathrm{Var}(x) = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2$$



[Fig. 1] Image generation process of the diffusion model; starting from a partially obscured map with noise in the upper-left quadrant, the diffusion model performs denoising to generate the complete map image in the lower-right quadrant

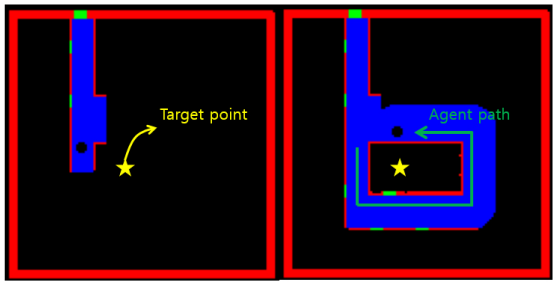


[Fig. 2] Sample images generated by the diffusion model; various examples of images produced using the diffusion model, demonstrating its capability to generate diverse and high-quality map



(a) (b) (c)

[Fig. 3] Uncertainty map derived from diffusion model outputs; (a) ground truth image, (b) partially observed image, (c) corresponding uncertainty map calculated from the diverse outputs of the diffusion model



(a) (b)

[Fig. 4] Determining unreachable areas using the A* algorithm; (a) a scenario where the target point is set inside an inaccessible room, making it impossible to reach, (b) The final shortest path solution generated by the A* algorithm

2.2 생성 모델 학습

앞서 알아보았던 샘플링 기반 불확실성 추정 방법은 우리가 다루는 데이터가 완전히 랜덤이 아닌 어느 정도의 규칙성을 가진다는 가정을 전제로 한다. 지도 생성의 경우를 예로 들어보면, 미탐색 영역이 있더라도 '벽은 서로 수직이다.' 또는 '복도는 서로 연결되어 있다.'와 같은 숨겨진 규칙이 있을 수 있다. 만약 생성 모델이 이러한 규칙을 학습하지 못하고, 완전히 무작위의 이미지를 생성한다면, 미탐색 영역 전체의 불확실성이 높게 추정되어 효율적인 탐색이 불가능하다.

불확실성 추정을 위해 사용 가능한 대표적인 생성 모델은 크게 세가지로 변분 오토인코더(VAE, Variational Autoencoder)[6], 생성적 적대 신경망(GAN, Generative Adversarial Networks)[7], 그리고 확산 모델(Diffusion Model)[8] 등이 있다. 생성되는 이미지가 실제 지도에 가까울수록 불확실성 추정 성능이 향상되기 때문에, 본 논문에서는 최근 각광 받고 있는 확산 모델 기반의 이미지 복원 모델인 Palette[9]을 사용한다.

확산 모델 학습을 위해선 전체 지도와 부분 지도 데이터가 필요하다. 전체 지도의 경우 랜덤 오피스 빌딩 지도 생성 알고리즘을 통해 생성하고[10], 부분 지도의 경우 전제 지도에서 랜덤한 영역을 삭제하여 생성한다. 확산 모델의 인풋으로는 [Fig. 1]의 좌상단 이미지와 같이 랜덤으로 삭제된 영역에 노이즈를 추가한 이미지를 사용하고, 이를 이용해 전체 지도를 추측하는 과정은 [Fig. 1]에서 확인 가능하다. 최종적으로 학습된 확산 모델은 생성 모델이므로 [Fig. 2]와 같이 다양한 형태의 추측 지도를 생성한다. 마지막으로 다양한 생성 결과를 통해 계산된 불확실성 지도는 [Fig. 3]의 (c)에서 확인 가능하다.

2.3 불확실성 기반 탐색 알고리즘

확산모델을 통해 얻은 불확실성 지도가 있을 경우, 불확실성이 최대인 지점으로 탐색을 진행하여 불확실성을 낮추는 것이 최적의 탐색 방법 중 하나이다. 이를 위해 현재 로봇의 위치와 지금까지 관측된 장애물 정보를 바탕으로 원하는 위치로 가는 경로를 생성할 수 있는 알고리즘이 필수적이다.

본 연구에서는 최적성과 완전성을 보장한다는 장점을 갖는 휴리스틱 기반의 탐색 알고리즘 A*[11]를 사용한다. 이를 이용해 현재 위치와 불확실성이 가장 높은 지점의 위치, 그리고 현재 부분적으로 탐색한 지도 정보를 통해 최적의 경로를 찾는다. 특히 A* 알고리즘은 해가 존재할 경우 반드시 찾아주는 완전성과 최단 경로를 보장하는 최적성을 가지므로 강건한 경로 탐색 성능을 보장한다. 완전성의 경우 장애물로 둘러싸여 탐색이 불가능한 영역을 판별하는데 유용하다. [Fig. 4]의 (a)와 같이 방안이 목표 지점으로 설정된 경우를 생각해 보자. 부분적으로 관측된 지도를 이용해 경로를 생성하고 새로 지도가 관측될 때 마다 이를 최신화 한다면 [Fig. 4]의 (b)와 같이 방 주위를 회전하게 된다. 이 때 A*의 완전성으로 인해 해당 지점으로 가는 경로가 존재하지 않음을 알 수 있다. 이를 통해 해당 지점은 장애물로 분류되어 불확실성이 높더라도 탐색 후보 지역에서 제외할 수 있다.

하지만 불확실성 기반 탐색 알고리즘에는 두가지 한계점이 존재한다. 첫째는 불확실성 추정을 정확히 하기 위해 생성하는 이미지 수를 늘림에 따라 계산 시간이 증가한다는 점이다. 둘째는 A* 알고리즘이 2D 기반 grid map에서 셀 수가 N인 경우 최대 시간 복잡도 $O(N\,log\,N)$을 가져 큰 지도에서 경로 생성이 오래 걸린다는 것이다. 이에 대한 계산 속도 실험은 실험 파트에서 추가로 다룬다. 따라서 이러한 단점을 보완하기 위해 강화학습 기반의 탐색 알고리즘과 결합하여 서로의 단점을 보완한다.

3. 개선된 탐색 알고리즘

강화학습을 통한 탐색을 진행할 경우, 빠른 action 생성이 가능하지만 학습의 불완전성으로 인해 일정 영역에 갇혀 탐색을

더 이상 진행하지 못하는 문제가 발생할 수 있다. 반면, 불확실성 기반의 탐색의 경우 강건하게 작동하지만 특정 상황에서 속도가 느려질 수 있다. 이번 장에서는 두 모델을 상호 보완적으로 사용함으로써 우수한 탐색 성능을 보이는 새로운 모델을 제안한다. 우선 기준으로 사용된 강화학습 모델에 대해 알아보고, 불확실성 기반 알고리즘과의 결합 방법에 대해 알아본다.

3.1 강화학습 모델

본 논문은 사용하는 강화학습 모델의 경우, 실내 탐색 문제를 풀기 위해 이미지로 된 state space로부터 적절한 action(상, 하, 좌, 우)을 생성하도록 학습한다. 강화학습 모델의 경우 세가지를 사용한다. 우선 표준적인 actor-critic 알고리즘인 A2C[12], 그리고 안정적인 정책 업데이트를 통해 학습에 안정성을 더한 알고리즘인 TRPO (Trust Region Policy Optimization)[13]와 PPO (Proximal Policy Optimization)[14]를 사용한다. 학습에 사용된 reward는 다음과 같이 설계한다.

- 새로운 영역 탐색: +1
- 벽과 충돌: -1
- 탐색 영역 재방문: -0.5

강화학습만을 사용한 학습의 결과는 불확실성 기반탐색과의 성능 비교를 위한 baseline으로 사용되었으며 그 결과는 4장에서 확인 할 수 있다.

3.2 강화학습과 불확실성을 이용한 탐색 알고리즘

본 논문에서는 강화학습을 이용한 탐색과 불확실성을 기반으로 하는 탐색을 번갈아 가며 진행하는 방식을 제안한다. 우선 탐색이 전혀 진행되지 않은 초기에는 강화학습을 이용해 탐색을 시작한다. 그 후 특정 조건에 따라 강화학습과 불확실성 기반의 탐색을 변경하며 탐색한다.

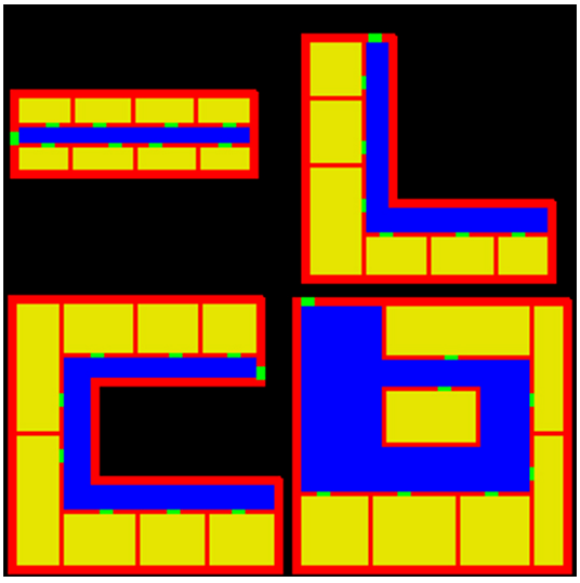
강화학습기반 탐색의 경우 벽에 충돌하거나 한 영역을 계속 반복하여 순환하는 상황을 실패로 정의한다. 이를 위해 로봇의 최근 탐색 경로를 일정 기간 저장하여 이를 다시 방문할 경우 불확실성 기반의 탐색으로 전환한다. 불확실성 기반 탐색의 경우 이를 종료하고 강화학습으로 전환하는 두가지 기준이 존재한다. 첫째는 목표로 했던 불확실성이 가장 높은 지점에 도달했을 경우이다. 이 경우 원하는 지점을 탐색 하였으므로 강화학습 알고리즘으로 전환한다. 둘째는 A* 알고리즘이 경로를 찾지 못할 경우이다. 이 경우 원하는 지점을 탐색하지는 못하였지만, 그 지점이 장애물로 가려진 탐색 불가 영역이라는 정보를 얻었기 때문에 다시 강화학습 기반의 탐색으로 전환한다. 이 때, 벽으로 둘러싸인 영역은 모두 탐색 불가 지역이므로 다시 방문하지 않도록 그 정보를 저장하여 반영한다.

4. 실험 및 결과

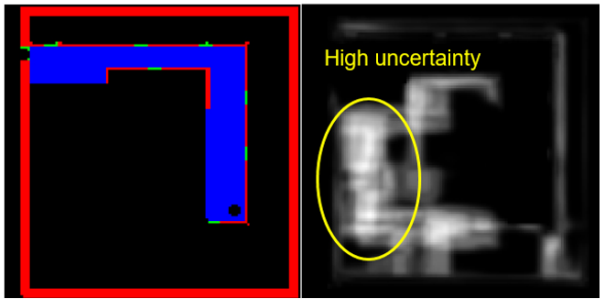
이 장에서는 본 논문에서 제안하는 알고리즘의 효용성을 평가하기 위한 구체적인 실험 방법과 기준, 그리고 결과에 대해 알아본다. 추가적으로 지도의 크기에 따른 A*의 계산 시간에 대해 알아본다.

4.1 실험 조건

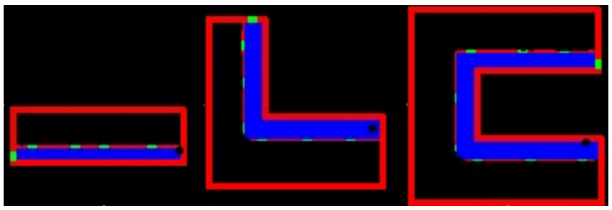
실험 환경은 [Fig. 5]에 있는 4가지 종류의 오피스 빌딩 환경을 사용한다. 같은 종류의 환경이라도 매번 새로운 환경이 생성될 수 있는 랜덤 생성 알고리즘을 사용하고, 각 종류에 대하여 100개의 지도를 이용해 실험을 진행한다. 각 방의 문이 닫혀 복도만 탐색하는 상황을 가정하고, 빠른 시간 안에 주어진 영역을 얼마나 탐사 가능한지 비교하기 위하여 탐색 시간을 제한한다. 또한 본 논문에서는 건물의 외벽의 형상을 이미 알고 있다고 가정하여 [Fig. 6]와 같이 부분 탐색 지도를 기반으로 불확실성을 추정할 경우 외벽과 건물 외부 영역은 그 대상에서 제외한다. 따라서 나머지 영역 중 불확실성이 가장 높은 영역을 우선적으로 탐색하게 되며, A* 알고리즘을 사용할 때에 외벽의 정보도 사용한다.



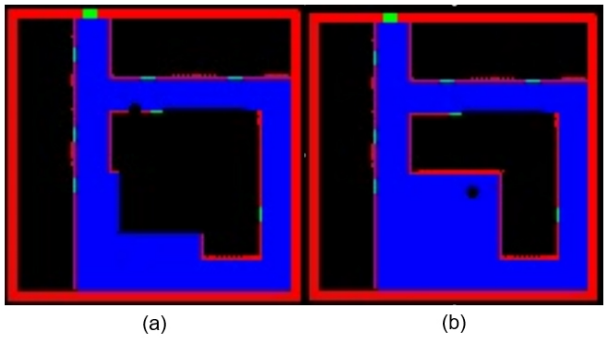
[Fig. 5] Examples of the four environments used in the experiments; balck areas represent unexploared area, red areas represent walls, blue areas represent corridors, yellow areas represent rooms, and green areas represent doors. This color scheme is used across all figures



[Fig. 6] Examples of partial maps and uncertainty maps generated during indoor exploration



[Fig. 7] Exploration results using the PPO Algorithm only in simple map environments



[Fig. 8] Comparison of exploration results in complex environments; PPO Alone versus PPO with uncertainty estimation. (a) Exploration using only the PPO algorithm fails to fully cover the entire area. (b) Exploration incorporating uncertainty estimation successfully covers the entire area

4.2 실험 결과 및 분석

강화학습 알고리즘만을 이용하여 탐색을 진행한 결과 [Fig. 7]과 같이 간단한 환경에서는 전체 탐색을 완료한 반면, [Fig. 8]의 (a)와 같이 복잡한 환경에서는 중간에 벽에 충돌하며 전체 영역 탐색에 실패한다. 반면, 불확실성을 함께 이용할 경우 [Fig. 8]의 (b)와 같이 전체 영역을 성공적으로 탐색함을 확인 할 수 있다.

[Table 1]을 통해 제한된 시간 내에 각 강화학습 알고리즘의 기본 탐색 성능과 불확실성을 이용해 향상된 성능을 비교할 수 있다. 우선 강화학습 모델의 기본 성능의 경우 모두 50% 아래로 저조한 성능을 보인다. 이는 불확실성 추정과 A* 알고리즘만을 이용한 결과인 69% 보다 낮은 성능이다. 강화학습과 불확실성 추정을 함께 이용할 경우 강화학습만을 이용한 경우와 비교해 아주 큰 성능 향상을 보인다. 이는 세가지 강화학습 모델 모두에서 확인 가능하여 본 논문에서 제안하는 방식이 강화학습 모델에 제한 받지 않는 방법임을 보여준다. 하지만 불확실성 추정만 사용한 경우와 비교할 경우 A2C 알고리즘을 함께 사용하는 것은 유의미한 성능향상을 보이지 않는다. 이는 함께 사용하는 강화학습의 성능이 최종 탐색 성능에 끼치는 영향을 보여준다. 불확실성 기반의 탐색을 추가로 사용해도 100%의 탐색률을 보이지 않는 이유는 제한된 시간 내에 탐색한 영역의 비율을 측정했기 때문으로 분석된다.

[Table 1] Exploration rate comparison for A2C, PPO, and TRPO with and without Uncertainty Estimation (UE) under a predefined time constraint

Model	RL w/o UE	RL w/ UE (ours)
A2C [12]	21.5%	70.0%
TRPO [13]	40.7%	87.5%
PPO [14]	38.2%	87.3%

[Table 2] Computation time of A* on grid maps of varying sizes at 30% obstacle density

Map size	Computation time [ms]
128 × 128	42
256 × 256	206
512 × 512	770
1024 × 1024	3275

추가적으로 grid map의 크기에 따른 A* 알고리즘의 계산 시간 변화를 i7-770K에서 측정한다. 그 결과 값은 [Table 2]에 나타나 있으며, 지도의 크기가 커짐에 따라 계산 시간이 급격히 증가함을 확인할 수 있다. 따라서 이를 통해 지도가 커질 경우 불확실성 추정을 통한 탐색 만이 아닌 강화학습을 추가로 사용해야 하는 이유를 다시 한번 확인 할 수 있다.

5. 결론 및 고찰

본 논문에서는 불확실성 추정을 통해 강화학습을 이용한 실내 탐색 알고리즘의 성능을 향상시킬 수 있는 새로운 방법을 제안하였다. 생성 모델을 활용하여 탐색하지 않은 영역의 불확실성을 정량화하고, 이를 바탕으로 로봇이 높은 불확실성을 가진 영역을 우선적으로 탐색하도록 유도한다. 이를 통해 전체 영역의 불확실성을 낮추게 되고, 실험을 통해 탐색 성능이 크게 향상됨을 입증하였다. 이러한 불확실성 기반 방법은 2D

지도 환경에 국한되지 않으며 다양한 task에 적용 가능할 것으로 기대된다.

그러나 본 연구에는 몇 가지 한계점이 존재한다. 첫째, 불확실성을 추정하는 과정에서 사용한 생성 모델의 경우 정확한 추정을 위해 생성하는 지도의 수를 늘릴수록 추정에 소요되는 시간이 증가한다. 이는 알고리즘을 사용하는 환경에 따라 실행 속도에 부정적인 영향을 끼칠 수 있다. 따라서 향후 연구에서는 생성 모델의 추론 시간과 환경의 복잡성을 고려하여 적절한 수의 지도를 생성하도록 최적화할 필요가 있다. 예를 들어, 특정 환경 조건 하에서 최적의 생성 지도 수를 동적으로 조절하는 알고리즘을 개발함으로써, 추론 시간을 단축하고 불확실성 추정의 정확성을 유지할 수 있을 것이다. 둘째, 실내 탐색 상황이 아닌 다양한 환경에서 적용 가능한 탐색 방법 전환의 기준이 필요하다. 본 논문에서는 2D 건물 실내 탐색 문제에서 적용 가능한 하나의 기준만을 제안 하였기 때문에, 환경이 바뀔 경우 그에 적합한 기준을 사용자가 따로 정의해야 한다.

References

[1] K. Zhu and T. Zhang, “Deep reinforcement learning based mobile robot navigation: A review,” *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 674-691, Oct., 2021, DOI: 10.26599/TST.2021.9010012.

[2] T. Chen, S. Gupta, and A. Gupta, “Learning exploration policies for navigation,” *International Conference on Learning Representations (ICLR)*, *arXiv.1903.01959*, 2019, DOI: 10.48550/arXiv.19 03.01959.

[3] Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi, “Target-driven visual navigation in indoor scenes using deep reinforcement learning,” *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, pp. 3357-3364, 2017, DOI: 10.1109/ ICRA.2017.7989381.

[4] D. S. Chaplot, D. Gandhi, S. Gupta, A. Gupta, and R. Salakhutdinov, “Learning to explore using active neural SLAM,” *International Conference on Learning Representations (ICLR)*, *arXiv.2004.05 155*, 2020, [Online], https://arxiv.org/abs/2004.05155.

[5] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, “Safe exploration in continuous action spaces,” *arXiv:1801.08757*, 2018, DOI: 10.48550/arXiv.1801.08757.

[6] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv:1312.6114*, 2013, DOI: 10.48550/arXiv.1312.6114.

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*, pp. 2672-2680, 2014, DOI: 10.48550/arXiv.1406.2661.

[8] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 6840-6851, 2020, DOI: 10.48550/arXiv.2006.11239.

[9] C. Saharia, W. Chan, H. Chang, C. A. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi, “Palette: Image-to-image diffusion models,” *ACM SIGGRAPH 2022 Conference Proceedings*, Vancouver, Canada, pp. 1-10, 2022, DOI: 10.1145/3528233.3530757.

[10] Y. H. Kim, B. H. Lee, J. S. Ha, S. J. Shim, D. H. Suh, and F. C. Park, “A random 2D environment generation framework for learning-based navigation in office buildings,” *The Journal of Korea Robotics Society*, vol. 20, no. 3, pp. 464-470, Aug., 2025, DOI: 10.7746/jkros.2025.20.3.46 4.

[11] P. E. Hart, N. J. Nilsson, and B. Raphael, “A formal basis for the heuristic determination of minimum cost paths,” *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100-107, Jul., 1968, DOI: 10.1109/TSSC.1968.300136.

[12] 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,” *International Conference on Machine Learning (ICML)*, 2016, DOI: 10.48550/arXiv.1602.01783.

[13] J. Schulman, S. Levine, P. Abbeel, M. I. Jordan, and P. Moritz, “Trust region policy optimization,” *International Conference on Machine Learning (ICML)*, pp. 1889-1897, 2015, [Online], https://proceedings.mlr.press/v37/schulman15.html.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv:1707.06347*, 2017, DOI: 10.48 550/arXiv.1707.06347.

하 준 수

2022 서울대학교 기계공학부(학사)
2024 서울대학교 기계공학부(석사)
2024~현재 서울대학교 기계공학부(박사과정)

관심분야: 머신 러닝, 로봇 러닝

심 성 준

2002 경희대학교 기계공학부(학사)
2004 경희대학교 기계공학부(석사)
2013~현재 LIG넥스원 수석연구원

관심분야: 자율주행, 군집제어, 임무계획

이 병 호

2019 연세대학교 기계공학과(학사)
2021 서울대학교 기계공학부(석사)
2025 서울대학교 기계공학부(박사)
2025~현재 Samsung Electronics

관심분야: 로봇 러닝, Learning from Demonstration

서 덕 현

2018 광운대학교 로봇학부 정보제어전공(학사)
2022 고려대학교 전기전자공학부 제어, 로봇, 시스템 전공(석사)
2022~현재 LIG넥스원 연구원

관심분야: 인공지능, 제어 및 계측 시스템, 자율주행

김 영 훈

2019 서울대학교 기계공학부(학사)
2019~현재 서울대학교 기계공학부(박사과정)

관심분야: 머신 러닝, 로봇 러닝, 컴퓨터 비전

박 종 우

1985 MIT Elec. Eng. Comp. Sci 학사
1991 Harvard Univ. 응용수학 박사
1995~현재 서울대학교 기계공학부 교수

관심분야: 리만 기하학, 로봇 제어, 로봇 러닝, 컴퓨터 비전