심층강화학습 기반 로봇 네비게이션 연구 동향

한양대학교 ▌권 준·노강찬·박상백·오도경·고병진·박태준

1. 서 론

현대 사회에서 모바일 로봇은 인간의 일상 생활과 다양한 산업 분야에서 핵심적인 역할을 수행한다. 특히, 환경 탐색, 자료 수집, 보안, 의료 지원 등의 임무에서 모바일 로봇의 활용이 급증하고 있다. 이러한 임무들은 자율적으로 이뤄져야 하며, 그 중심에는 자율주행 기술이 놓여 있다.

강화학습은 인공지능 분야에서 학습과 의사결정을 수행하는 방법 중 하나로, 최근에 그 성능이 빠르게 발전하고 있다. 에이전트가 환경과 상호작용하며 보상을 최대화하는 방향으로 학습하게 함으로써, 복잡한 문제를 해결할 수 있는 강력한 방법론이다. 이러한 강화학습의 발전은 모바일 로봇의 자율주행 분야에서도 큰 주목을 받았다.

이전의 연구들에서 강화학습은 모바일 로봇의 자율주행에 적용되어왔다. 예를 들어, 주행 경로 최적화, 장애물 회피, 환경 변화에 대한 적응 등에서 강화학습은 놀라운 결과를 보여주었으며, 특히 모바일 로봇의 효율적이고 유연한 작동을 가능케 하였다. 그러나 강화학습 기반의 모바일 로봇 자율주행은 아직 해결해야할 여러 과제와 한계가 존재한다. 대표적으로, 로봇의 소셜 에티켓 부재와 강화학습 알고리즘의 높은 샘플 요구량 , 안정성 보장의 부재 그리고 희소 보상 학습 문제가 있다.

이에 대한 해결책을 모색하기 위해, 최근 연구 동향을 살펴보면, 샘플 효율성 문제를 해결하기 위한 또 다른 학습 기법의 도입이나, 강화학습 알고리즘에 소셜 에티켓 개념을 이용, 기존 알고리즘과 강화학습의 통합, 다른 알고리즘이 생성한 데이터를 사용하여 희소 보상 환경에서 학습하는 등의 다양한 시도가 진행중이다. 이러한 연구들은 모바일 로봇의 자율주행 분야에서 더 나은 학습 효율성과 안전성을 추구하는 데 있어 중요한 전환점을 제시하고 있다.

† 본 연구는 과학기술정보통신부 및 정보통신기획 평가원의 Grand ICT 연구센터지원사업의 연구결과로 수행되었음(ITP-2023-2020-0-01741).

2. 강화학습

강화학습은 기계 학습의 한 종류로, 에이전트가 환경과 상호작용하며 특정 작업을 수행하는 방법을 학습하는 알고리즘이다. 에이전트(Agent)는 환경과 상호작용하는 주체로, 정책(Policy)에 따라 특정 상태에서 특정 행동을 선택한다. 환경(Environment)은 에이전트가 상호작용하는 외부 시스템으로, 에이전트의 행동에 따라 상태가 변하며 보상을 제공한다. 상태(State)는 에이전트가 환경과 상호작용할 때 현재의 상황을 나타낸다. 시간 t에서의 상태는 S_t로 표시한다. 행동(Action)는 에이전트가 특정 상태에서 선택하는 행위이다. 보상(Reward)는 환경으로부터 에이전트가 얻는 피드백으로, 에이전트의 행동이 얼마나 좋은 지를 나타낸다. 시간 t에서의 보상은 R_t로 표시한다.

강화학습의 목표는 보상을 최대화하는 최적의 정책을 찾는 것이다. 이를 위해 가치함수나 행동가치함수 등의 개념을 사용한다. 가치함수(Value Function)은 상태나 상태-행동 쌍에 대해 특정 정책을 따라 기대되는 미래 보상의 합을 나타낸다. 미래 보상의 합은 G_t로 표시한다. $V(s)$는 상태 s에서의 가치를 나타낸다. 행동가치함수(Action-Value Function)은 특정 상태에서 특정 행동을 선택했을 때의 기대되는 미래 보상의 합을 나타낸다. $Q(s,a)$는 상태 s에서 행동 a를 선택했을 때의 가치를 나타낸다.

$$G_t = R_{t+1} + \gamma R_{t+2} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

$$V(s) = E[G_t \mid S_t = s]$$

$$Q(s,a) = E[G_t \mid S_t = s, A_t = a]$$

가치함수나 행동가치함수는 벨만 방정식(Bellman Equation)을 통해 업데이트된다. 벨만 방정식은 현재

가치를 이전 가치와 새로운 정보로 업데이트하는 방법을 제시한다. 다양한 강화학습 알고리즘에서 벨만 방정식을 활용하여 강화학습의 목표를 달성한다.

$$V(s) = E[R_{t+1} + \gamma V(S_{t+1}) \mid S_t = s]$$
$$Q(s,a) = E[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) \mid S_t = s, A_t = a]$$

3. Human Aware Navigation 연구

강화학습을 통한 모바일 로봇의 자율주행은 사람이 적은 지역에선 원활하게 주행할 수 있지만, 특히나 소셜 에티켓이 필요한 공항과 쇼핑몰과 사람이 많은 지역에선 성능이 저하되는 경향이 있다. 따라서 이러한 성능 저하를 개선하기 위한 방법론이 Human Aware Navigation이다. Human Aware Navigation은 사람과 로봇의 상호작용 및 사람과 사람간의 상호작용을 고려하여 주행하는 방법이며, 여기서 상호작용이란 좌/우측통행 과 같은 보이지 않는 행동 방식을 의미한다. 이러한 행동 방식을 강화학습에 적용하기 위해 센서로부터 받은 정보를 Hand made feature로 변환 후 Socal Norm, Danger Zone과 같은 새로운 reward function을 설계하거나, Recurrent Neural Network(RNN), Attention mechanism 그리고 Graph Convolution Network(GCN)을 state extractor로써 강화학습 알고리즘에 적용하는 방법론들이 있다.

3.1 Reward를 활용한 방법론

먼저 Passing, Crossing, Overtaking과 같은 Social Norm을 활용한 방법론 [1]이 있다. 이는 강화학습 자율주행에서 처음 제안되는 Human aware 방법이며, 본 논문은 자율 주행에 있어서 사람의 행동과 유사해야 안전한 주행이 가능하다는 점을 제안한다. 따라서 Social Norm과 관련된 reward function 설계하였다. 그 결과로 로봇의 주행이 Social Norm 과 유사하게 행동이 가능하며, 주행의 성능이 향상되는 결과를 보여주었다.

이와 유사하게 Danger Zone의 경우, 사람의 다음 행동을 예측할 수 있어야 로봇이 군중 속에서 안전하게 이동가능하다는 점을 제안한다. 여기서 Danger Zone이란, 달리기, 걷기 등과 같은 사람의 상태에 따른 관측된 속도 값을 활용하여 정의된 구역을 의미한다[2,3]. 여기서 더 나아가 사람을 어린이, 성인 등과 같이 객체화 하여 이에 따른 Danger Zone을 설계한 방법[4]이 있다. 이러한 방법들 또한 [1]과 동일하게 Danger Zone이라는 새로운 reward function을 설계하여, 로봇의 주행에 있어서 향상된 성능을 보여주었다.

3.2 딥러닝을 활용한 아키텍처 모듈화

기존에 Hand made feature를 사용하게 되는 경우 강화학습 모델에 넣을 수 있는 데이터의 개수가 고정되어 있다. 그로 인해 일반적으로 로봇이 관측하는 사람 수가 많은 환경에서 주행의 성능이 떨어지게 되며, 이를 보완하기 위해 제시된 논문이 [5]이다. 본 논문에서는 LSTM에 hand made feature를 넣어줄 때 로봇과 사물 거리가 먼 것부터 먼저 hidden cell에 넣어주며, 가장 가까운 것을 최근 hidden cell에 넣어준다. 이와 같이 RNN 계열을 사용함으로써, 로봇이 관측하는 사람 수가 증가함에도 일정한 성능을 보장할 수 있다. 이와 유사하게 [6]의 경우 Attention mechanism을 사용하였다. Attention mechanism이란 데이터를 서로 다른 3가지의 Linear layer를 거쳐 Key, Value, Query를 얻은 후에, 아래와 같이 Query와 Key에 Softmax연산을 거쳐 Attention weight를 얻은 다음 Value에 곱하는 방법이다. 아래 식에서 d는 Query의 벡터 차원을 나타낸다.

$$A(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

따라서 이러한 특징을 본 논문에서 이용하여, 데이터의 해 거리에 따른 가중치 고려가 아닌, network가 학습을 통해 가중치를 고려하도록 하였다.

위에 제시된 방법론들은 사람에 대해 가중치를 고려하여 우선순위를 생각할 뿐, 데이터의 관계성을 직접적으로 파악하거나 활용하지 않아 복잡한 환경에서 주행의 성능이 떨어지는 점이 있다. 이를 해결하기 위해서 Structure data 분석에 용이한 GCN을 활용한 방법론[7]이 제시되었다. GCN은 node와 edge로 구성된 Graph를 분석하는 방법론으로, node의 latent feature matrix X와 adjacency matrix A를 활용하여 node끼리의 관계성을 message passing이라 불리는 아래의 수식을 통해 파악한다.

$$H^{(l+1)} = \sigma(AH^{(l)}W^{(l)}) + H^{(l)} \;\; where \;\; H^0 = X$$

위 식에서 W는 학습가능한 weight matrix, l은 node feature level, σ은 activation function을 뜻한다.

이 외에도 최근 많은 Human Aware Navigation 방법론들이 앞서 소개한 reward, RNN, Attention Mechanism, GCN을 활용하고 있다. reward의 경우 사람의 개개인에 대한 것만 아니라 집단까지 고려한 [8,9]이 있으며, RNN과 Attention mechanism의 경우 공간적 [10], 마지막으로 GCN의 경우 사람의 다양한

행동 양상을 고려하기 위해 Cross-Modal Transformer를 적용한 [11], 사람과 사람, 로봇과 사람의 상호작용에서 더 나아가 사람의 행동까지 고려한 [12], 그리고 동적인 사물까지 고려한 [13]이 있다.

4. 샘플 효율성 관련 강화학습 연구 및 모바일로봇 자율주행 연구

강화학습 에이전트는 사람이 배우는 방식과는 다르게 각기 다른 작업에 대해 처음부터 학습한다. 학습 초기에는 의미가 적은 전환들이 많아 에이전트가 효과적으로 학습하고 업데이트하기 어렵다. 픽셀과 같은 고차원 관측에서의 강화학습에서 양상이 더 심해진다. 실례로, Kalashnikov et al.은 로봇 그랩 가치 함수와 정책을 개발하기 위해 수개월 동안 대규모 로봇 상호작용 데이터를 수집한 로봇 팔 농장이 필요하였다 [14]. 또한 강화학습 기반 모바일 로봇 자율주행에서도 샘플 효율성 문제가 존재하며 특히 센서 데이터를 관측으로 사용 시 에이전트 수준의 관측을 사용한 경우보다 더욱 더 샘플 효율성이 떨어진다. 그래서 가치 있는 전략을 더 잘 탐색하고 가치 있는 경험의 샘플링 효율성을 향상하는 것이 강화학습이 적용된 연구에서도 중요한 방향성이며 모바일 로봇 자율주행 분야에서도 이를 해결하는 연구가 최신 흐름 중 하나이다. 본 단락에서는 강화학습 기반 모바일로봇 자율주행에 적용될 여지가 높은 샘플 효율성을 향상한 강화학습 연구와 강화학습 기반 모바일로봇 자율주행에서 샘플 효율성 문제를 완화시킨 연구에 대해서 간단히 정리하였다.

4.1 샘플 효율성 관련 강화학습 연구

강화학습 샘플 효율성을 향상시키기 위해서 명시적으로 세계의 예측 모델(model)을 구축하고, 그 모델을 사용하여 가치함수를 구하는 방법이 있다. 구축된 모델을 통하여 가치함수를 계획(plan)하거나 model-free 알고리즘으로 학습하여 가치함수를 학습한다 [15-18]. 이는 강화학습의 model-based 알고리즘이다.

Schaul et al.은 prioritized experience replay(PER) 방법을 제안했다 [19]. 해당 연구는 학습 효율성을 향상시키기 위해 샘플의 중요도(즉, temporal-difference)에 따라 확률을 계산했다. PER은 off-policy 강화학습 방법의 효율성을 향상시키는 효과적인 방법이다. 이미 일부 연구에서는 PER을 강화 학습 기반의 모션 플래닝 방법에 적용한 작업도 진행되고 있다 [20, 21].

Laskin et al.은 비지도 대조 학습(Unsupervised Contrastive Learning)을 보조 작업(auxiliary task)으로 활용하여 샘플 효율성을 높였다 [22]. 그들은 센서 데이터에서 유효한 특징을 비교 학습 과정을 통해 추출한 다음, 이러한 특징을 강화 학습 모듈에 입력했다.

Kostrikov et al.은 데이터 정규화된 Q(DrQ) 방법을 제안했다 [23]. 그들은 샘플링 훈련 과정을 시작하기 전에 관측치 입력에 대해 augmentation을 수행하고 목표-Q 및 현재 Q를 동시에 계산했다. 또한 정규화 기술과 결합하여 원시 입력 데이터의 샘플 효율성을 크게 향상시켰다.

Schwarzer et al.은 자기 예측적 표현(Self-predictive Representation) 접근법을 제안했다 [24]. 에이전트가 잠재적 미래 상태의 다단계 표현을 예측하도록 훈련시킴으로써 샘플 효율성을 개선했다. 이러한 작업은 에이전트가 다양한 환경 관측 하에서 시간적으로 예측 가능하고 일관된 표현을 학습할 수 있게 했다.

4.2 샘플 효율성 관련 모바일로봇 자율주행 연구

Barzin Moridian et al.은 모바일로봇의 pose를 기반으로 관측치를 augmentation을 수행해 샘플을 늘려 샘플 효율성을 증가시켰다.[25] 그리고 Mark Pfeiffer et al.은 IL(Imitation Learning)으로 expert demonstration을 수행해 강화학습의 샘플 효율성 문제를 완화하였다.[26]

5. 안정성 보장 관련 연구

학습 기반 알고리즘들은 특히나 상태(state) 하나하나 마다 결과가 다르기 때문에 안정성이 보장이 되지 않는다. 여기서 말하는 안정성이란 로봇의 비현실적인 속도와 관련이 되어있다. 즉, non-smooth trajectory가 발생함을 의미하며, 이러한 문제점을 해결하기 위해 제안된 방법은 크게 2가지로 첫 번째로 reward를 통한 방법론과 기존 자율주행 (e.g. Dynamic Window Approach)와 같은 알고리즘과 병합한 방법이 있다.

4.1 Reward를 통한 Kinematic guarantee

[27]에서 제안된 방법론으로 로봇의 불안전한 주행을 reward로 통해 컨트롤하는 방법이다. 아래 수식과 같이 각속도(angular velocity)에 대한 간단한 reward function을 정의하였다.

$$r_{osc} = -0.1|w| \text{ if } |w| > 0.3$$

이를 통해 기존의 강화학습 기반 주행의 Trajectory보다 더 부드러운 성능을 이끌어냈다.

4.2 기존 알고리즘과의 병합 연구

[28]의 경우 Dynamic window approach (DWA)의 특징 중 하나인 Kinematic constraints를 사용한 방법론이다. 여기서 DWA란 로봇의 Kinematic constraints가 고려된 방법이다. 여기서 Trajectories에 대해 cost를 계산하여, 하단에 나와 있는 Object function을 최대로 하는 Trajectory를 선택, 주행하는 방법론이다.

$$G(v,w) = \alpha\, head(v,w) + \beta dist(v,w) + \gamma vel(v,w)$$

위 식의 head항은 목표지점과 로봇 간의 거리, dist 항은 사물과의 거리, vel항은 로봇의 속도이고 α,β,γ는 해당 항의 하이퍼파라미터이다.

본 논문은 여기서 Trajectory의 cost 계산 결과를 DRL의 상태로 주었으며, agent의 행동을 Trajectory에 있는 velocities를 선택하게 함으로써 로봇의 주행의 성능을 향상시켰다.

이와 유사하게 [29]의 경우 Navigation function(NF)을 사용한 방법론으로 NF이란 Artificial Potential Field(APF)에서 척력(repulsive force), 인력(attractive force)를 이용하여 로봇의 주행을 이끌어 내는데 여기서 척력과 인력이 0이 되는 지점에 로봇이 멈춰버리는 문제를 해결하기 위해 제안된 수식이다.

$$\psi(x) = \frac{\gamma_d(x)}{[\gamma_d^K(x) + \beta(x)]^{\frac{1}{K}}}$$

위 수식은 *K*에 따라 결과가 상이하게 되는데 본 논문은 이 *K*와 subgoal point를 agent의 행동으로 하게 함으로써, Trajectory의 불안정성을 낮추었다.

6. 희소 보상 관련 네비게이션 연구

강화학습에서 희소 보상을 가진 환경에서는 원하는 동작에 대한 보상이 거의 주어지지 않기 때문에 에이전트는 원하는 행동을 학습하는 데 많은 시간이 소요될 수 있다. 이로 인해 학습 속도가 느려지고, 특히 실제 환경에서 에이전트가 적절한 정책을 학습하는 데 어려움이 생길 수 있다. 또한 에이전트가 새로운 행동을 시도하고 다양한 경로를 탐험하는 것을 어렵게 만들 수 있다.

Wu et al.은 희소 보상 환경의 문제를 해결하고 RL 에이전트를 훈련하기 위해 이중 소스 체계(double-source scheme)를 사용을 제안하였다 [30]. 로봇에 Depth camera를 사용하여 네비게이션을 TEB local path planner 알고리즘과 RL 모델(SAC)에서 경험 샘플을 수집하여 학습을 진행하였다.

Dawood et al.은 희소 보상 환경에서 RL 에이전트를 훈련하기 위한 경험 소스로 모델 예측 제어(MPC)를 사용할 것을 제안하였다 [31]. MPC 모델(경험 소스)과 강화학습 모델(SAC)을 병렬로 선택하여 학습함. MPC rate는 RL보다 MPC를 선택할 확률을 의미하고, 학습이 진행됨에 따라 MPC rate는 감소한다.

7. 결론 및 향후 연구

지금까지 강화학습을 활용한 모바일 로봇의 자율주행에서 연구가 되고 있는 분야 대해 알아보았다. 간단하게 정리하자면, 로봇과 사람의 상호작용을 고려한 Human Aware navigation, Raw data의 샘플 효율성을 고려한 주행, 강화학습 기반 방법론의 비현실적인 주행을 막기위한 방법, 그리고 희소 보상 관련 네비게이션에 대해서 알아보았다. 그러나 이러한 방법들 모두 시뮬레이션 상에서 진행되기 때문에 현실과의 괴리감을 무시할 순 없다.

따라서 최근 향후 연구는 주로 “sim2real” 및 더 현실적인 시뮬레이션에 중점을 두고 있다. “Sim2real”은 시뮬레이션에서 얻은 결과를 실제 환경으로 전이시키는 것을 의미한다. 이는 로봇이 시뮬레이션에서 효과적으로 학습하고, 그 학습된 정책을 실제 환경에서 성공적으로 적용할 수 있도록 한다.

시뮬레이션은 로봇의 학습을 위한 중요한 도구이지만, 현실 세계와의 차이로 인해 전이 학습이 어려울 수 있다. 따라서 더 정확하고 현실적인 시뮬레이션 환경을 구축하는 연구가 필요하다.

또한 로봇이 한 도메인에서 학습한 내용을 다른 도메인으로 전이시키는 능력은 매우 중요하다. 시뮬레이션과 실제 세계 사이의 차이를 극복하고, 로봇이 다양한 환경에서 효과적으로 작동할 수 있도록 하는 연구가 필요하다.

마지막으로 로봇이 특정 장소나 환경에 국한되지 않고 다양한 곳에서 작동할 수 있도록 하는 연구가 중요하다. 이를 위해 다양한 지형, 조명 조건, 날씨 등 다양한 환경에서의 성능을 일반화할 수 있는 방법을 연구해야한다.

참고문헌

[1] Chen, Yu Fan, et al., “Socially aware motion planning with deep reinforcement learning,” *2017 IEEE/RSJ Inter-*

national Conference on Intelligent Robots and Systems (IROS), IEEE, 2017.

[2] S. S. Samsani and M. S. Muhammad, "Socially Compliant Robot Navigation in Crowded Environment by Human Behavior Resemblance Using Deep Reinforcement Learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5223-5230, July 2021.

[3] Montero, E.E., Mutahira, H., Pico, N. et al., "Dynamic warning zone and a short-distance goal for autonomous robot navigation using deep reinforcement learning," Complex Intell. Syst. (2023).

[4] L. Kästner, J. Lil, Z. Shen and J. Lambrecht, "Enhancing Navigational Safety in Crowded Environments using Semantic-Deep-Reinforcement-Learning-based Navigation," *2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Sevilla, Spain, 2022, pp. 87-93.

[5] Everett, Michael, Yu Fan Chen, and Jonathan P. How. "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," IEEE Access, Vol. 9, pp. 10357-10377, 2021.

[6] Chen, Changan, et al., "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," *2019 international conference on robotics and automation (ICRA)*, IEEE, 2019.

[7] Chen, Changan, et al., "Relational graph learning for crowd navigation," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020.

[8] Katyal, Kapil, et al., "Learning a group-aware policy for robot navigation," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022.

[9] Kim, Mincheul, Youngsun Kwon, and Sung-Eui Yoon. "Group Estimation for Social Robot Navigation in Crowded Environments," *2022 22nd International Conference on Control, Automation and Systems (ICCAS)*, IEEE, 2022.

[10] Shi, Weixian, et al., "Enhanced spatial attention graph for motion planning in crowded, partially observable environments," *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022.

[11] Wang, Weizheng, et al., "NaviSTAR: Socially Aware Robot Navigation with Hybrid Spatio-Temporal Graph Transformer and Preference Learning," *arXiv preprint arXiv:2304.05979*, 2023.

[12] Prakash, Varun Ganjigunte, "Behavior-Aware Robot Navigation with Deep Reinforcement Learning," *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, 2022.

[13] Zhou, Zhiqian, et al., "Navigating Robots in Dynamic Environment With Deep Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 12, pp. 25201-25211, 2022.

[14] Kalashnikov, D., Irpan, A., Pastor, P., Ibarz, J., Herzog, A., Jang, E., Quillen, D., Holly, E., Kalakrishnan, M., Vanhoucke, V., et al., "Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation," *arXiv preprint arXiv:1806.10293*, 2018.

[15] Sutton, R. S., "Integrated architectures for learning, planning, and reacting based on approximating dynamic programming," *Machine Learning Proceedings 1990*, pp. 216-224. Elsevier, 1990.

[16] Ha, D. and Schmidhuber, J., "World models," *arXiv preprint arXiv:1803.10122*, 2018.

[17] Kaiser, L., Babaeizadeh, M., Milos, P., Osinski, B., Campbell, R. H., Czechowski, K., Erhan, D., Finn, C., Kozakowski, P., Levine, S., et al., "Model-based reinforcement learning for atari," *arXiv preprint arXiv:1903.00374*, 2019.

[18] Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., et al., "Mastering atari, go, chess and shogi by planning with a learned model," *arXiv preprint arXiv:1911.08265*, 2019.

[19] Schaul T, Quan J, Antonoglou I, et al., "Prioritized experience replay," https://arxiv.org/pdf/1511.05952.pdf.

[20] Hu Z J, Gao X G, Wan K F, et al., "Relevant experience learning: a deep reinforcement learning method for UAV autonomous motion planning in complex unknown environments," *Chinese Journal of Aeronautics*, Vol. 34, No. 12, pp. 187-204, 2021.

[21] He Z C, Dong L, Sun C Y, et al., "Reinforcement learning based multi-robot formation control under separation bearing orientation scheme," *Proc. of the Chinese Automation Congress*, 3792-3797, 2020.

[22] Askin M, Srinivas A, Abbeel P., "CURL: contrastive unsupervised representations for reinforcement learning," *Proc. of the International Conference on Machine Learning*, pp. 5639-5650, 2020.

[23] Kostrikov I, Yarats D, Ferhus R., "Image augmentation

is all you need: regularizing deep reinforcement learning from pixels," https://arxiv.org/abs/2004.13649.

[24] Chwarzer M, Anand A, Goel R, et al., "Data-efficient reinforcement learning with self-predictive representations," https://arxiv.org/abs/2007.05929.

[25] Moridian, Barzin, Brian R. Page, and Nina Mahmoudian. "Sample efficient reinforcement learning for navigation in complex environments," *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, IEEE, 2019.

[26] Pfeiffer, Mark, et al., "Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations," *IEEE Robotics and Automation Letters*, Vol. 3, No. 4, pp. 4423-4430, 2018.

[27] Liang, Jing, et al., "Crowd-steer: Realtime smooth and collision-free robot navigation in densely crowded scenarios trained using high-fidelity simulation," *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021.

[28] Patel, Utsav, et al., "Dwa-rl: Dynamically feasible deep reinforcement learning policy for robot navigation among mobile obstacles," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021.

[29] Bektaş, Kemal, and H. Işıl Bozma, "Apf-rl: Safe mapless navigation in unknown environments," *2022 International Conference on Robotics and Automation (ICRA)*, IEEE, 2022.

[30] K. Wu, H. Wang, M. A. Esfahani and S. Yuan, "Learn to Navigate Autonomously Through Deep Reinforcement Learning," *IEEE Transactions on Industrial Electronics*, Vol. 69, No. 5, pp. 5342-5352, May 2022.

[31] M. Dawood, N. Dengler, J. de Heuvel and M. Bennewitz, "Handling Sparse Rewards in Reinforcement Learning Using Model Predictive Control," *2023 IEEE International Conference on Robotics and Automation (ICRA)*, London, United Kingdom, pp. 879-885, 2023.

약 력

권 준

2023 한양대학교 에리카캠퍼스 로봇공학과 졸업(학사)
2023~현재 한양대학교 인공지능융합학과 석사과정
Email : kjun998@hanyang.ac.kr

노 강 찬

2023 한양대학교 에리카캠퍼스 봇공학과 졸업(학사)
2023~현재 한양대학교 인공지능융합학과 석사과정
Email : rohgarie2000@hanyang.ac.kr

박 상 백

2023 한양대학교 에리카캠퍼스 건설환경공학과 졸업(학사)
2023~현재 한양대학교 인공지능융합학과 석사과정
Email : sbp0783@hanyang.ac.kr

오 도 경

2020 한양대학교 에리카캠퍼스 로봇공학과 졸업 (학사)
2022 시스콘 자율주행 SW 1팀 책임 연 구원(최종)
2023~현재 한양대학교 인공지능융합 학과 석사과정
Email : zepplim0@hanyang.ac.kr

고 병 진

2021 대구경북과학기술원 박사과정
2022 한양대학교 에리카캠퍼스 박사후 연구원
2022~현재 한양대학교 에리카캠퍼스 스마트 융합공학부 조교수
Email : byungjinko@hanyang.ac.kr

박 태 준

1994~2000 LG전자 우면동연구소 선임연구원
2001~2005 University of Michigan Ann Arbor 박사과정
2005~2008 삼성전자 삼성종합기술원 수석연구원
2008~2011 한국항공대학교 조교수
2011~2015 DGIST 정보통신융합공학전공 부교수
2015~현재 한양대학교 로봇공학과 교수
한양대학교 AI 협동로봇사업단장
교육부 지능형로봇 혁신공유대학사업 총괄사업단장
Email : taejoon@hanyang.ac.kr