

가상 환경에서의 강화학습 기반 긴급 회피 조향 제어 Reinforcement Learning based Autonomous Emergency Steering Control in Virtual Environments

이훈기¹ · 김태윤¹ · 김효빈¹ · 황성호^{1*}

Hunki Lee¹, Taeyun Kim¹, Hyobin Kim¹ and Sung-Ho Hwang^{1*}

Received: 08 Nov. 2022, Revised: 28 Nov. 2022, Accepted: 28 Nov. 2022

Key Words : Autonomous Driving(자율주행), Reinforcement Learning(강화학습), Virtual Environment(가상환경), Autonomous Emergency Steering(자율긴급조향), Autonomous Emergency Braking(자율긴급제동)

Abstract: Recently, various studies have been conducted to apply deep learning and AI to various fields of autonomous driving, such as recognition, sensor processing, decision-making, and control. This paper proposes a controller applicable to path following, static obstacle avoidance, and pedestrian avoidance situations by utilizing reinforcement learning in autonomous vehicles. For repetitive driving simulation, a reinforcement learning environment was constructed using virtual environments. After learning path following scenarios, we compared control performance with Pure-Pursuit controllers and Stanley controllers, which are widely used due to their good performance and simplicity. Based on the test case of the KNCAP test and assessment protocol, autonomous emergency steering scenarios and autonomous emergency braking scenarios were created and used for learning. Experimental results from zero collisions demonstrated that the reinforcement learning controller was successful in the stationary obstacle avoidance scenario and pedestrian collision scenario under a given condition.

1. 서 론

자율주행 기술의 발전은 최근 몇 년간 빠른 속도로 가속화되고 있으며, 안전, 교통 혼잡, 에너지, 환경 등 도로 위의 문제들을 해결하는 데 기여할 것으로 전망되고 있다.¹⁾ 연구의 영역뿐만 아니라 세계 각국의 자동차 제조사에서는 상용화를 위해 개발 경쟁에 돌입하고 있다. 그러나 자율 주행 차량과 인간 운전자가 혼재하는 도로 환경에서는 예상치 못한 위험이 존재하고 특히 도심지와 같은 복잡한 주행 환경에서는 규칙 기반의 정책의 한계가 존재한다.²⁾ 따라서 자율주행의 센서 처리, 인지, 판단과 제어 등

여러 분야에서 인공지능과 딥러닝 기술의 중요성이 높아지고 있다.³⁻⁵⁾ 이와 같은 기계 학습 기반의 방법론들은 많은 양의 데이터와 주행을 요구하므로 주행 가상환경 시뮬레이션을 활용하는 연구가 진행되고 있다.⁶⁻⁸⁾

강화학습은 인공지능의 한 분야로 주어진 상황에서 선택할 수 있는 행동을 점차 개선해나가는 방향으로 에이전트를 학습한다. 강화학습은 다량의 샘플링이 필요하고 학습 초기 단계에서 반복적인 실패를 통해 학습하는 특징이 있어 특히 가상환경에서의 시뮬레이션을 통한 학습의 이점이 부각된다. 강화학습을 자율차량 제어에 적용하고자 하는 연구는 적응형 순항 제어와 같은 운전자 보조 시스템⁹⁾이나 차선 변경 판단, 보행자 회피 경로 생성과 같은 다양한 분야에서 진행되었다. 강화학습을 차량의 차선 변경, 경로 생성, 경로 추종 제어 등 다양한 분야에 적용한 모델이 제시되었지만 정적 장애물이나 동적 장애물이 포함되지 않은 시나리오에서 단순 경로 추종 제

* Corresponding author: hsh0818@skku.edu

¹ Department of Mechanical Engineering, Sungkyunkwan University, Suwon 16419, Korea

Copyright © 2022, KSFC

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

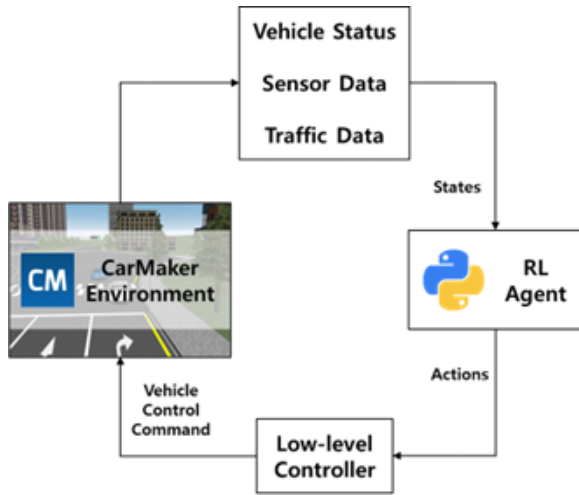


Fig. 1 Reinforcement learning framework

어만을 목표로 하거나,¹⁰⁻¹¹⁾ 차선 변경 시점 및 보행자 회피 경로 생성에 그쳐 경로를 따라가기 위한 추가적인 제어가 필요한 한계점이 존재한다.¹²⁻¹³⁾ 보행자, 차량과 같은 동적 장애물을 고려하는 모델에서도 강화학습의 상태 공간을 적용하고자 하는 시나리오에 한정적으로 구성하여 범용적인 주행상황에 적용하기 어려운 문제점 또한 존재한다.

본 논문에서는 이와 같이 자율주행 차량의 제어에 강화학습을 적용하여 정적 장애물과 보행자 회피 등의 시나리오에 적용하고자 한다. 앞서 기술한 한계점을 극복하기 위해 상태 공간을 차량의 상태와 주변 차량의 상태로 구성하고 모델의 출력을 차량의 목표 가속도 및 조향 각으로 설정하였다. 학습을 위한 가상 환경은 IPG CarMaker 소프트웨어를 사용하여 구성하였고, Python 환경의 강화학습 에이전트와 통신을 위해 MMF(Memory Mapped File) 방법을 사용하여 통신 환경을 구성하였다. 곡선 도로에서 차량의 기본적인 경로 추종을 위한 학습을 진행하고 Stanley, Pure Pursuit과 같은 대표적인 차량 횡 제어 알고리즘과 성능을 비교하였다. 정지 장애물 회피 시험, 보행자 추돌 시험 두 가지의 시나리오를 생성하여 학습하고 시험 결과를 분석하였다.

2. 학습 환경 구성

강화학습은 의사결정의 수학적 모델인 마르코프 결정 과정(MDP)으로 정의되는 문제를 푸는 방법론이다. MDP는 환경이 가질 수 있는 관측 가능한 상태의 집합 S , 에이전트가 선택할 수 있는 행동의 집합 A , 상태 s 에서 특정 행동을 취했을 때 상태 s' 로 전

이할 확률을 나타내는 P , 특정 상태에서 행동에 대해 받는 보상 R , 그리고 미래의 보상에 대한 감쇠인자 γ 의 5중쌍으로 표현된다. 도로 위의 차량의 주행 제어 상황에 강화학습을 적용하기 위해 다음과 같이 MDP를 정의했다.

2.1 상태 공간 정의

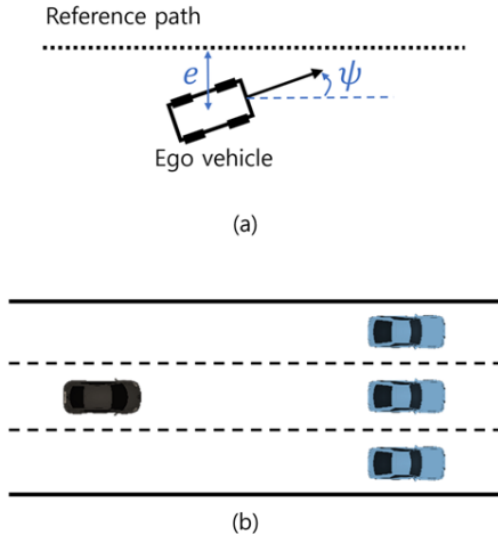
상태 공간은 차량의 목표 속도, 종 방향 속도, 가속도, 헤딩 오차, 횡 방향 오프셋, 조향 각, 도로 곡률로 구성하였다. 목표 속도를 상태 공간에 포함시키고 네트워크의 입력으로 전달함으로써 추가적인 학습 없이 동일한 모델로 목표 속도를 선택하여 주행이 가능하도록 구성하였다. 여기에 차차 기준으로 현재 차로, 왼쪽 차로, 오른쪽 차로 전방 세 대의 주변 객체에 대해 각각 상대 위치, 상대 속도, 상대 가속도 정보를 추가하여 최종 MDP의 상태 공간은 총 25가지 입력으로 Table 1과 같다.

Table 1 State space definition

State ($i = 1, 2, 3$)	Meaning
s_1	차차 목표 속도, v_{target}
s_2	차차 속도, v_x
s_3	차차 가속도, a_x
s_4	차차 헤딩 오차, ψ
s_5	차차 횡방향 오프셋, e
s_6	차차 스티어링 휠 각도, δ_{steer}
s_7	도로 곡률, κ
s_{6i+2}, s_{6i+3}	주변 객체 상대 거리, d_x, d_y
s_{6i+4}, s_{6i+5}	주변 객체 상대 속도, v_x, v_y
s_{6i+6}, s_{6i+7}	주변 객체 상대 가속도, a_x, a_y

2.2 행동 공간 정의

행동 공간은 Table 2와 같이 차량의 목표 가속도와 스티어링 휠의 목표 각속도로 설정하여 종-횡 방향 제어가 동시에 가능하도록 구성했다. 목표 가속도의 경우 최소 $-10m/s^2$ 에서 최대 $10m/s^2$ 의 연속적인 범위로 설정했고, CarMaker의 차량 제어를 통해 목표 속도를 추종하기 위한 가감속 입력으로 변환된다. 횡 방향 제어를 위한 스티어링 휠의 각속도는 최대 $\pm 150deg/s$ 의 연속적인 값으로 설정했으며, 스티어링 휠의 최대 범위는 $-3rad$ 에서 $3rad$ 으로 제한했다.



(a) Ego vehicle (b) Surrounding vehicles
Fig. 2 State space schematic diagram

Table 2 Action space definition

Action	Meaning
a_1	목표 가속도, $a_{desired}$
a_2	스티어링 휠 각속도, δ_{steer}

2.3 보상 함수 정의

보상 함수는 목표 속도 유지를 위한 항, 차선 유지를 위한 항, 전방 차량과의 안전거리 유지를 위한 항, 횡 방향 진동에 대한 페널티 항, 그리고 충돌에 대한 페널티 항으로 구성했으며 다음 수식과 같이 정의하였다.

$$r = r_v + r_{lane} + r_{distance} + r_{oscillation} + r_{collision} \quad (1)$$

보상 함수의 각 항이 속도에 비례하도록 각각에 곱해지는 목표 속도에 대한 현재 차량 속도의 비율을 정의하였다. 세부적인 보상의 정의는 Table 3과 같다.

$$\tilde{v} = \min \left[\frac{v_x}{v_{target}}, 1 \right] \quad (2)$$

Table 3 Reward function definition

Reward	Definition
r_v	$- v_{target} - v_x / v_{target}$
r_{lane}	$\tilde{v} \cdot (\cos\psi - \sin\psi - e)$
$r_{distance}$	if $d_x < 2cdotv_x$: $(d_x - 3.6cdotv_x) / v_{target}$
$r_{oscillation}$	$-\Delta\delta_{steer} / 20$
$r_{collision}$	if collision: $-200 \cdot \tilde{v}$

2.4 네트워크 구성

이처럼 에이전트가 선택할 수 있는 행동의 범위가 연속적인 실수일 때에는 일반적으로 DQN (Deep Q-Network)와 같은 가치 기반 방법론보다 정책 기반 방법론이 적합하다.¹⁴⁾ 그 중에서도 본 연구에서는 대부분의 문제에서 높은 성능을 보장한다고 알려진 PPO(Proximal Policy Optimization) 알고리즘을 사용하였다.¹⁵⁾ PPO는 하나의 샘플을 여러 번의 업데이트에 사용하므로 수렴 속도가 빠르지만, 하이퍼 파라미터에 민감하여 여러 번의 학습의 결과를 바탕으로 최적의 수렴성을 보장했던 하이퍼 파라미터를 사용하였고 그 결과는 Table 4와 같다.

액터 및 크리틱의 네트워크는 25개의 입력과 128x64 크기의 은닉층, 그리고 2개의 출력으로 구성하였고 그 구조는 Fig. 3과 같다. 은닉층의 활성화함수는 ReLU를 사용했으며 최종 출력 레이어에는 하이퍼볼릭 탄젠트를 사용함으로써 행동의 값을 설정한 행동 공간의 범위와 같아지도록 제한하였다.

Table 4 Hyper parameters of learning

Parameter	Meaning	Value
K_{epoch}	num of epochs in 1 update	80
lr_{actor}	learning rate of actor	3e-4
lr_{critic}	learning rate of critic	1e-3
t_{max}	maximum training timesteps	1e6
f_{update}	update frequency in timestep	3e3
$\sigma_{initial}$	initial standard deviation for action distribution	0.3
σ_{min}	minimum std	0.05
$\Delta\sigma$	std decaying rate in timestep	2.5e-2
f_{σ}	std decaying frequency	6e4

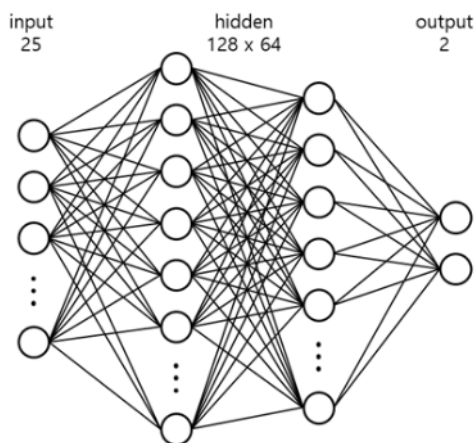


Fig. 3 Network structure

3. 시나리오 구성

학습 시나리오는 크게 세 가지로 구분하여 구성했다. 가장 먼저 앞서 구성한 학습 환경에서 차량이 차로 중앙을 유지하는 제어가 가능함을 검증하고자 장애물 없는 곡선 도로에서 학습을 진행하였다. 학습된 에이전트를 기반으로 정지 장애물 회피 및 보행자 정면 충돌 시나리오를 학습하여 일반적인 주행 상황과 긴급 제동, 회피 조향 상황에 범용적인 모델을 학습하였다.

3.1 곡선로 주행

강화학습 에이전트의 경로 추종 제어 성능 검증을 위한 곡선로 주행 학습을 진행하였다. 이 때, 도로는 매 에피소드마다 선회 반경 60m에서 240m의 범위에서, 회전 각도는 60도에서 120도 범위에서 랜덤하게 생성하였다. 각 선회 사이에는 최대 50m까지의 직진 도로를 추가하고 직진 도로와 선회 도로 사이에는 곡률이 연속적인 값을 갖도록 클로소이드 곡선 형상의 도로를 생성했다. 임의로 생성된 도로의 예시는 Fig. 4와 같다.

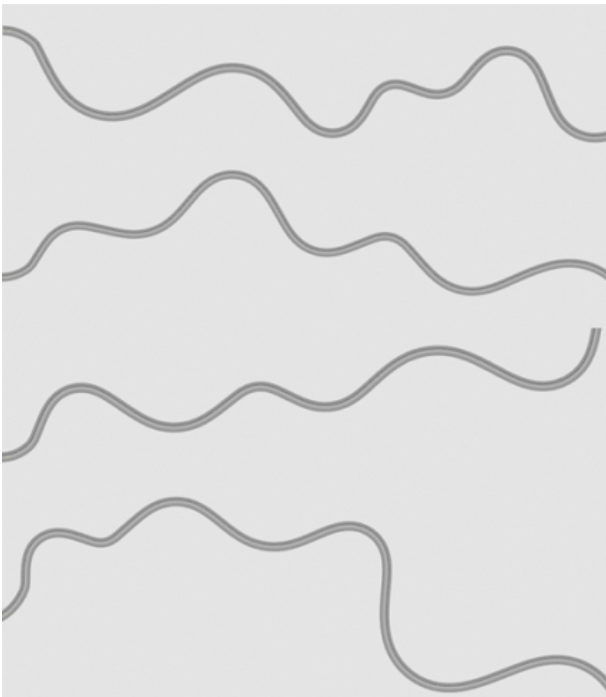


Fig. 4 Randomly generated road examples

3.2 정지 장애물 회피

KNCAP의 긴급조향기능장치 시험방법 중 정지자동차 회피 시험 조건을 참고하여 CarMaker 시나리오

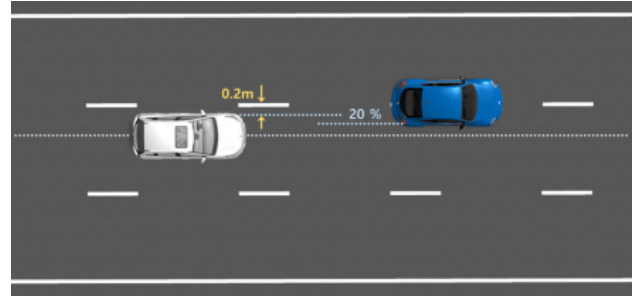


Fig. 5 Stationary car avoiding scenario

를 생성하고 학습을 진행하였다. 자차는 한쪽 차선과 0.2m 간격을 유지하여 주행하고, 목표 자동차는 자차 너비의 20%만큼 오버랩 된 상태로 위치시켰다. 목표 속도는 50km/h에서부터 80km/h의 범위에서 5km/h 단위로 랜덤하게 부여하였다.

3.3 보행자 정면 충돌

KNCAP의 시가지모드 비상자동제동장치 시험방법 중 보행자 정면 충돌 시험 조건을 참고하여 CarMaker 시나리오를 생성하고 학습을 진행했다. 목표보행자의 이동속도는 출발지점부터 1.5m 구간까지 가속하여 8km/h를 유지하고, 차로 반대편까지 총 15m를 이동하여 횡단하도록 설정했다. 자차의 목표속도를 기준으로 계산하여 충돌 예상 지점을 자차의 중심선과 일치할 수 있도록 횡단 시점을 설정하였다.

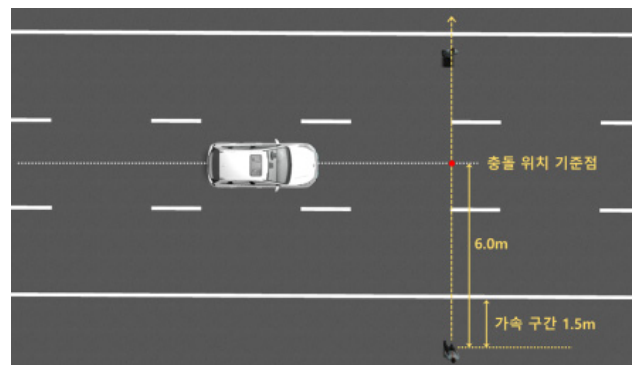
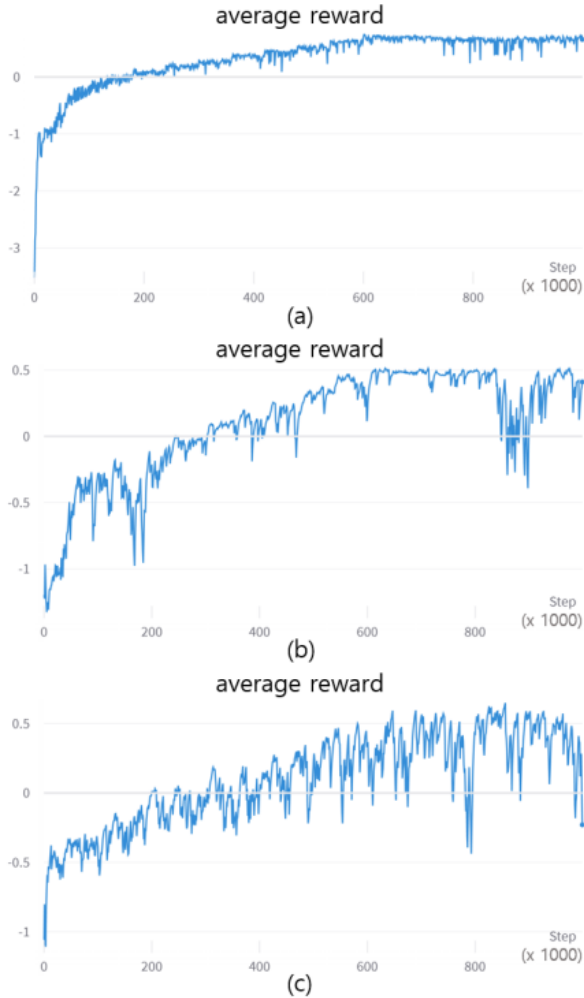


Fig. 6 Pedestrian front collision scenario

4. 학습 및 실험 결과

앞 장에서 설명한 세 가지의 시나리오에 대해 각각 학습을 진행하고 결과를 검증했다. 세 시나리오에서의 학습 타임스텝 별 평균 보상의 결과는 Fig. 7과 같다.

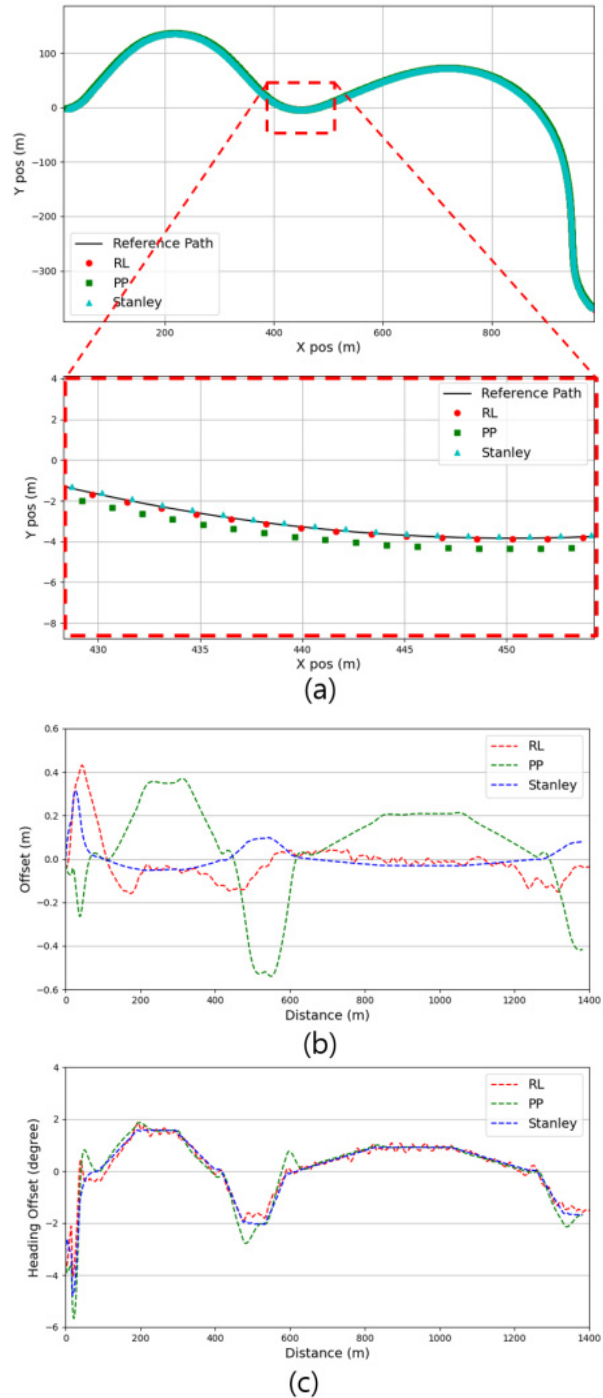


(a) curve road driving (b) stationary car avoiding
(c) pedestrian front collision

Fig. 7 Training reward results

4.1 곡선로 주행 실험 결과

강화학습 에이전트의 경로 추종 성능을 검증하기 위해 곡선로 주행 시나리오를 학습하고 Pure Pursuit, Stanley 제어기와 결과를 비교했다.¹⁶⁾ 시나리오 단계에서와 동일하게 임의의 곡률로 생성된 테스트 도로에서 제어기를 변경해가며 실험 후 횡 방향 오프셋 및 헤딩 오차를 비교했다. Pure Pursuit과 Stanley 제어기를 실험할 때의 종 방향 제어는 Intelligent Driver Model을 기반으로 설정된 목표 속도로 주행할 수 있도록 하였다.¹⁷⁾ 각 제어기의 횡 방향 오차 결과의 제곱평균제곱근과 평균은 Table 5와 같다. Stanley 제어기의 횡 방향 오프셋이 더 작게 나타난 이유는 실험의 목표 속도가 도로 곡률에 비해 빠르지 않고 곡률 변화량이 작은 클로소이드 곡선의 도로 조건이었기 때문에 Stanley 제어기 사용에 최적화된 조건이었던 것으로 분석하였다.



(a) trajectory (b) lateral offset (c) heading offset

Fig. 8 Training reward results

Table 5 RMS and mean value of path offset

Controller	RMS [m]	Mean [m]
RL	0.0975	0.1564
PP	0.2194	0.4229
Stanley	0.0661	0.0090

4.2 정지 장애물 회피 실험 결과

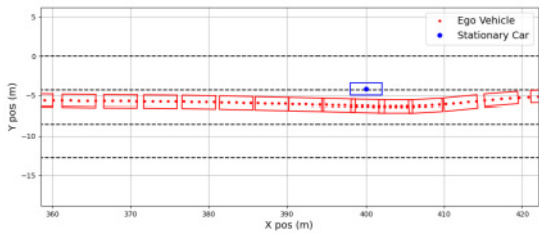


Fig. 9 Ego vehicle trajectory and pose

정지 장애물 회피 시험 시나리오를 학습하고 결과를 분석하였다. 시험 시나리오는 65km/h 목표 속도 조건에서 자차와 목표 자동차의 횡 방향 오버랩 20%, 차선까지의 거리 0.2m 조건에서 진행하였으며 대상 자동차의 위치가 좌측, 우측인 경우에 대해 각각 다섯 번, 총 10회 시뮬레이션을 진행하였다. 10회의 실험 중 충돌 횟수는 0회이고 차선 침범 횟수 역시 0회로 학습된 환경 내에서 정지 장애물 회피 성능을 검증했다. 좌측 정지자동차 회피 시나리오에서 자차의 경로 및 자세를 나타낸 결과는 Fig. 9와 같다.

4.3 보행자 정면 충돌 실험 결과

보행자 정면 충돌 시나리오의 경우 목표 속도를 50km/h에서 80km/h까지 10km/h 단위로 증가시키며 실험을 진행하였다. 보행자의 속도는 8km/h로 설정하고 자차가 목표 속도로 주행한다고 가정했을 때 차로 중앙에서 충돌하도록 출발 시점을 결정하였다. 자차 목표 속도에 따라 4회의 시나리오를 보행자 진행 방향이 좌측과 우측인 경우에 대해 총 8회의 시뮬레이션을 진행하였고 네 가지 속도 조건 모두에서 감

속 및 조향을 통해 충돌하지 않음을 검증했다. 보행자의 진행 방향이 좌측이고 목표속도가 60km/h와 70km/h 조건인 시나리오에서 자차와 보행자의 경로 및 자세를 나타낸 결과는 Fig. 10과 같다.

4.4 강화학습 제어기 성능 분석

세 가지의 실험을 통해 강화학습 제어기는 곡선 도로에서 안정적인 주행이 가능하고 정지 장애물 회피와 보행자 횡단 시나리오에 대응이 가능함을 확인하였다. 또한, 강화학습 제어기는 정지 장애물과 보행자 회피 상황에서 기존의 자율주행 알고리즘에서와 달리 장애물 회피를 위한 경로를 계획하거나 충돌하지 않기 위한 목표 속도를 계산하는 과정 없이 감속 또는 조향을 통해 충돌을 회피하는 결과를 확인하였다.

5. 결론

본 논문에서는 강화학습을 활용하여 자율주행 차량의 경로를 추종하기 위한 제어 방법을 제안하였다. 차량의 상태와 경로 정보를 상태 입력으로 받아 종횡 방향 차량 제어 명령을 출력하는 강화학습 에이전트를 학습하고 제어 성능을 검증하였다.

또한, 경로 상의 정적 장애물이 있는 경우와 무단 횡단하는 보행자를 마주한 상황에서 긴급 제동 및 회피 조향을 통해 충돌 회피가 가능한 제어 방법을 제시하였다. KNCAP의 자동차안전평가시험 등에 관한 기술규정에 근거하여 정지 장애물과 보행자 회피 시나리오를 구성하고 학습 및 검증에 활용하였다. 본 연구에서 제안한 강화학습 모델은 주어진 경로 상에 장애물과의 충돌 위험이 감지되었을 때 추가적인 지역 경로 계획 없이 경로 오차를 감수하면서 감속 및 조향을 통해 충돌을 회피하는 제어 명령을 출력할 수 있다.

추후 연구에서는 복합적인 주행 상황이 포함된 시나리오를 구성하고 현재까지 학습이 완료된 모델을 기반으로 학습을 진행하여 일반적인 주행 상황과 다양한 긴급 상황에 범용적으로 적용할 수 있는 강화학습 기반 자율주행 제어기를 개발할 예정이다. 또한, 본 논문에서 정의했던 상태 공간을 확장하여 도심지 도로에도 적용 가능한 모델을 개발할 계획이다.

후 기

본 연구는 국토교통부/국토교통과학기술진흥원 교통물류연구사업의 연구비지원 (22TLRP-C152478-04)

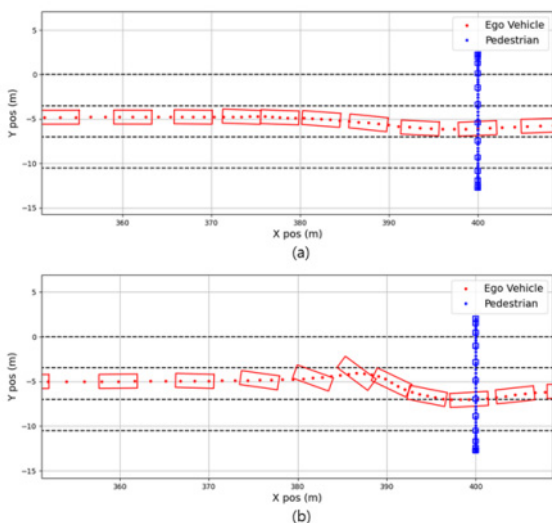


Fig. 10 Ego vehicle and pedestrian trajectory and pose (a) 60km/h (b) 70km/h

과 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행된 결과물입니다. (IITP-2022-2018-0-01426)

이해관계(CONFLICT OF INTEREST)

저자는 이 논문과 관련하여 이해관계 충돌의 여지가 없음을 명시합니다.

References

- 1) C. Y. Chan, "Advancements, prospects, and impacts of automated driving systems," *International journal of transportation science and technology*, Vol.6, No.3, pp.208-216, 2017.
- 2) L. Liangzhi, K. Ota and M. Dong, "Humanlike driving: Empirical decision-making system for autonomous vehicles," *IEEE Transactions on Vehicular Technology*, Vol.67, No.8, pp.6814-6823, 2018.
- 3) Y. Chen, C. Dong, P. Palanisamy, P. Mudalige, K. Muelling and J. M. Dolan, "Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- 4) X. Liang, T. Wang, L. Yang and E. Xing, "Cirl: Controllable imitative reinforcement learning for vision-based self-driving," *Proceedings of the European conference on computer vision (ECCV)*, pp-584-599, 2018.
- 5) K. S. Kim, J. I. Lee, S. W. Gwak, W. Y. Kang, D. Y. Shin and S. H. Hwang, "Construction of Database for Deep Learning-based Occlusion Area Detection in the Virtual Environment," *Journal of Drive and Control*, Vol.19, No.3, pp.9-15, 2022.
- 6) J. I. Lee, G. S. Gwak, K. S. Kim, W. Y. Kang, D. Y. Shin and S. H. Hwang, "Development of Virtual Simulator and Database for Deep Learning-based Object Detection," *Journal of Drive and Control*, Vol.18, No.4, pp.9-18, 2021.
- 7) S. Wang, D. Jia and X. Weng, "Deep reinforcement learning for autonomous driving," *arXiv:1811.11329*, 2018.
- 8) J. Chen, B. Yuan and M. Tomizuka, "Model-free deep reinforcement learning for urban autonomous driving," 2019 *IEEE intelligent transportation systems conference (ITSC)*, 2019.
- 9) C. Desjardins and B. Chaib-draa. "Cooperative adaptive cruise control: A reinforcement learning approach," *IEEE Transactions on intelligent transportation systems*, Vol.12, No.4, pp.1248-1260, 2021.
- 10) A. Folkers, M. Rick and C. Büskens, "Controlling an autonomous vehicle with deep reinforcement learning," 2019 *IEEE Intelligent Vehicles Symposium (IV)*, 2019.
- 11) O. P. Gil, R. Barea, E. L. Guillen, L. M. Bergasa, C. G. Huelamo, R. Gutierrez and A. D. Diaz, "Deep reinforcement learning based control for autonomous vehicles in carla," *Multimedia Tools and Applications*, Vol.81, No.3, pp.3553-3576, 2022.
- 12) A. Fehér, S. Aradi and T. Bécsi, "Online Trajectory Planning with Reinforcement Learning for Pedestrian Avoidance," *Electronics*, Vol.11, No.15, 2022.
- 13) M. Yoshimura, G. Fujimoto, A. Kaushik, B. K. Padi, M. Dennison, I. Sood, K. Sarkar, A. Muneer, "Autonomous Emergency Steering Using Deep Reinforcement Learning For Advanced Driver Assistance System," 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2020.
- 14) R. S. Sutton, D. McAllester, S. Singh, Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in neural information processing systems*, Vol.12, 1999.
- 15) J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, "Proximal policy optimization algorithms," *arXiv:1707.06347*, 2017.
- 16) D. Y. Yu, D. G. Kim, H. S. Choi and S. H. Hwang, "Hybrid Control Strategy for Autonomous Driving System using HD Map Information," *Journal of Drive and Control*, Vol.17, No.4, pp.80-86, 2020.
- 17) A. Kesting, M. Treiber and D. Helbing. "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol.368, No.1928, pp-4585-4605, 2010.