

강화학습 알고리즘을 이용한 초소형 전기자동차 현가장치의 제어 알고리즘 설계

Design of Control Algorithm for Micro Electric Vehicle Suspension System Using Reinforcement Learning Algorithm

김 영 준* · 홍 석 무** · 오 종 석†

Young-Jun Kim*, Seokmoo Hong** and Jong-Seok Oh†

(Received December 14, 2021 ; Revised February 7, 2022 ; Accepted February 10, 2022)

Key Words : Reinforcement Learning(강화 학습), Micro Electric Vehicle(소형 전기 자동차), Suspension(현가장치), Vibration Control(진동 제어), Ride Comfort(승차감)

ABSTRACT

In this study, a reinforcement learning-based controller was designed to enhance the vibration control performance of micro-electric vehicles. The dynamic model of a controllable suspension system was realized, and the corresponding mechanical system was established. Reinforcement learning algorithms such as the deep deterministic policy gradient (DDPG) and twin-delayed DDPG (TD3) agents were applied to the suspension system for the vibration controller design. The vibration control performances were evaluated for various road profiles. Moreover, the effective vibration control performance and ride comfort were assessed using computer simulation, and the uncontrolled and controlled results were compared. It was inferred that the vibration controller using DDPG and TD3 agents successfully reduced the pitching and rolling motions at the center of gravity point of the sprung mass.

1. 서 론

초소형 삼륜 자동차는 크기가 작다는 특성 덕분에 주차가 용이하고 회전반경이 작아 좁은 장소에서의 방향 전환이 유리하기 때문에 좁은 길 주행 또는 주차 시 운전의 편리성을 더해 주어 자동차 업계에 많은 관심을 받고 있으며 최근 자동차 시장에 전기 자동차의

형태로 개발이 이루어져 다양하게 출시되고 있다. 이러한 초소형 삼륜 전기자동차가 더욱 시장점유율을 높이기 위해서는 기본적으로 다양한 안전장치 기술 및 편의 기술이 개발 되어야할 필요가 있는데, 특히 삼륜 자동차는 일부 국가에서 비포장 도로에서의 인력 및 화물 수송의 주된 역할을 하는 만큼 원활한 운영을 위해 하나의 바퀴가 많은 무게를 지탱하면서도 비포장도로에서 발생하는 진동을 효과적으로 감쇠하여 최상의 승차감을 발휘하여야 하기 때문에 초소형 삼륜 자동차

† Corresponding Author; Member, Department of Future Convergence Engineering, Kongju National University, Professor
E-mail : Jongseok@kongju.ac.kr

* Department of Future Convergence Engineering, Kongju National University, Student

** Department of Future Convergence Engineering, Kongju National University, Professor

A part of this paper is presented at the KSNVE 2021 Annual Autumn Conference

‡ Recommended by Editor Heon Jun Yoon

© The Korean Society for Noise and Vibration Engineering

현가장치에 대한 기술개발이 매우 필요하다.

기존의 자동차 현가장치 제어는 다양한 제어를 통해 모델기반 제어를 수행하고 있으며⁽¹⁻³⁾, 또한 강화학습을 활용한 데이터기반 현가장치 제어 알고리즘 설계에 대한 연구들도 활발히 이루어지고 있다^(4,5). 그 중 영국 Howell 등⁽⁴⁾은 4wheels 차량을 대상으로 승차감을 개선하기 위해 차량의 반 능동 현가장치를 모델링한 후 강화학습을 이용하여 차체 진동의 제어를 수행하였다. 이에 반해 소형 삼륜 전기차에 대한 강화학습 제어 알고리즘의 연구사례는 정도가 미미하여 관련 연구가 필수적이다.

삼륜 자동차는 조향륜이 하나만 존재하는 구조적 특성으로 인해 회전 시 전도의 위험이 존재한다. 이를 위해 삼륜자동차의 현가장치는 차체가 과하게 기울어지지 않도록 버텨야 하면서 다양한 도로 파형에 따른 진동을 감쇄하여야 하기 때문에 우수한 성능을 갖는 진동 제어기에 대한 연구가 필요하다. 또한 기존 제어기 개발과정에서 발생하는 제어기의 튜닝에 들어가는 개발소요시간과 비용을 줄여야 할 필요가 있다. 하지만 수동으로 제어기를 튜닝하게 되면 상당히 많은 시행착오를 겪어야 하며 시간과 비용이 다수 소요된다. 이에 이 논문에서는 강화학습 알고리즘을 이용한 현가장치의 제어 알고리즘 설계에 관한 연구를 진행하여 해결하였다. 다양한 도로 조건에 따른 최적의 제어 게인 혹은 실시간 제어입력을 빠르게 도출할 수 있도록 하여 강화학습의 다양한 에이전트 알고리즘을 이용한 현가장치 제어기의 설계 및 개발에 대한 단순화의 가능성을 보였다.

이 논문의 2장에서는 강화학습 알고리즘과 계산 방식에 대해 설명한 뒤 이 연구에서 사용한 강화학습 에이전트 알고리즘과 hyper parameter를 설명하였다. 또한 초소형 삼륜 전기차의 제원을 설정하였으며 이를 바탕으로 1/3 모델링과 강화학습을 수행하여 결과를 도출하였다. 3장에서는 오프로드를 포함한 다양한 형상과 거칠기를 가진 도로에 따른 sprung mass의 수직 속도 결과를 도출하여 진동제어 성능을 확인하였으며, 승차감 평가를 위해 ISO 2631기준에 따라 탑승자가 느끼는 수직 진동의 정도를 수치적으로 평가하여 정리하였다. 또한 1/3 모델링을 기반으로 학습 시킨 제어기를 삼륜 소형 전기차 7자유도 전 차량 모델(7-dof. full car)에 적용시켜 결과를 도출하며 제어기 간의 간섭과 상호영향성이 없음을 확인하였다.

2. 연구 방법

2.1 강화학습 알고리즘

강화학습(reinforcement learning)이란 기계학습(machine learning)의 한 종류로 기계장치 등을 제어하기 위해 주어진 상황에서 취할 수 있는 최적의 정책(policy)을 선택하여 원하는 목표에 달성하는 최적 제어기를 구축하기 위한 학습 알고리즘이다. 강화학습은 시스템의 역할을 하는 환경(environment)과 해석장치 및 제어기의 역할을 하는 에이전트(agent)의 상호작용을 통해 학습이 진행되며 Fig. 1은 이러한 과정을 나타낸 개념도이다.

에이전트는 환경의 변화를 표현하는 상태(state)를 관측(observation)하며 관측된 정보를 기반으로 행동(action)을 수행하게 된다. 그러한 행동의 영향으로 변하는 환경상태에 따라 보상(reward)이 결정되는데, 이러한 보상은 시행(episode)별로 책정되며 에이전트는 누적된 총 보상을 최대화하는 것을 목표로 학습하여 최적의 행동을 결정한다.

강화학습에서는 에이전트의 종류에 따라 다양한 성격으로 해석이 진행되며 이는 비평가 네트워크(critic network)와 배우 네트워크(actor network)의 조합으로 다양하게 구성된다. 이러한 조합들은 3분류로 나뉘어지며⁽⁶⁾, 이 논문에서는 조합들 중 배우-비평가(actor-critic)방식으로 학습을 진행하였다.

배우-비평가 방식의 알고리즘은 DDPG와 TD3 알고리즘이 가장 기본적으로 사용되고 있는데, 이러한 DDPG(deep deterministic policy gradient)와 TD3(twin delayed deep deterministic policy gradient) 에이전트 알고리즘의 경우 연속적인 행동과 관찰 공간

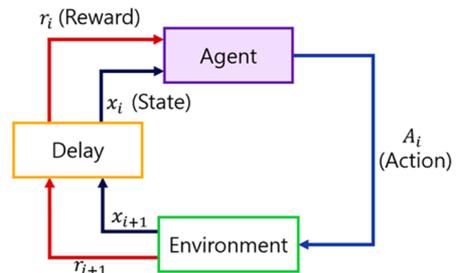


Fig. 1 Agent-environment interaction in reinforcement learning

이 모두 존재하는 환경에서 최적의 효율을 발휘하기 때문에⁽⁷⁾ 최근 자동차 현가장치 등 많은 응용분야에 적용되고 있다^(8,9). 이에 이 연구에서도 DDPG와 TD3 에이전트를 이용하여 소형 삼륜 전기차 현가장치 제어기의 제어 알고리즘 설계를 진행하였다.

DDPG는 다른 정책으로 만든 데이터를 이용해 현재의 정책을 업데이트하는 비 정책(off-policy) 방법을 이용하는 연속적인 행동 공간 환경에서의 비 모델기반 알고리즘이다. DDPG는 에이전트의 경험 샘플을 바로 학습에 적용하지 않고 메모리(memory 또는 버퍼(buffer))에 저장해둔 뒤 일정 샘플 이상이 모이면 샘플을 메모리에서 무작위로 도출하여 학습에 적용하는 경험 리플레이(experience replay)라는 방법을 이용하며 이는 비 정책 방법을 이용하는 알고리즘의 주된 특징이다. DDPG는 식 (1)을 통해 손실함수의 gradient를 계산하는데⁽¹⁰⁾ 여기서 ϕ 는 비평 매개변수이며 s_i 는 상태변수를, A 는 행동(action)을 의미한다.

$$\phi_{t+1} = \phi_t + \alpha_\phi \sum_i (y_i - Q^\mu(s_i, A_i | \phi)) \nabla_\phi Q_\phi(s_i, A_i | \phi) \tag{1}$$

식 (1)에서 y_i 는 시간차 타깃으로 식 (2)로 정의된다. DDPG 알고리즘에서는 DDPG 알고리즘 특유의 안정성에 관한 문제점을 해결하기 위하여 Q-네트워크와는 별도로 타깃(target) Q-네트워크를 도입하는데, 이것이 y_i 로 일정시간마다 Q-네트워크의 파라미터를 업데이트시키고 그 시간 동안의 파라미터를 고정시키는 역할을 한다.

Fig. 2 (a)와 같이 DDPG 알고리즘에서 비평가 네트워크의 손실함수($L(\phi)$)는 식 (3)을 따르게 된다. 여기서 M 은 미니배치의 크기이며 메모리에서 추출한 샘플들의 개수를 의미한다. 또한 미분의 연쇄법칙(chain rule)에 의한 배우 네트워크의 손실함수($L(\theta)$)는 식 (4)와 같으며 여기서 부호에 -를 두어서 손실함수를 최소화하면서 Q_ϕ 는 최대화하도록 돕는다.

$$y_i = r(s_i, u_i) + \gamma Q_\phi(s_{i+1}, \pi_\phi(s_{i+1})) \tag{2}$$

$$L(\phi) = \frac{1}{M} \sum_i M_i (y_i - Q_\phi(s_i, A_i | \phi))^2 \tag{3}$$

$$L(\theta) = - \sum_i (Q_\phi(s_i, \pi_\theta(s_i))) \tag{4}$$

DDPG에서는 식 (5)와 같이 파라미터가 원래 신경망의 파라미터를 천천히 따라가며 타깃 배우 네트워크와 타깃 비평가 네트워크를 업데이트하는데, 여기서 $\hat{\theta}$ 는 타깃 배우 네트워크를 의미하며 $\hat{\phi}$ 는 타깃 비평가 네트워크를 의미한다. τ 는 평활화 인자(smoothing factor)로 모든 시간 단계에서 평활화를 시킨다.

$$\begin{aligned} \hat{\theta} &\leftarrow \tau\theta + (1-\tau)\hat{\theta} \\ \hat{\phi} &\leftarrow \tau\phi + (1-\tau)\hat{\phi} \end{aligned} \tag{5}$$

이러한 DDPG 에이전트는 일반적으로 Q-값(Q-value)을 과대하게 평가한다는 문제점이 있다. 이에 TD3는 기존 DDPG에서 비평가 네트워크를 하나 더 추가하여 비평가 네트워크 간의 상호비교를 하고 정책과 대

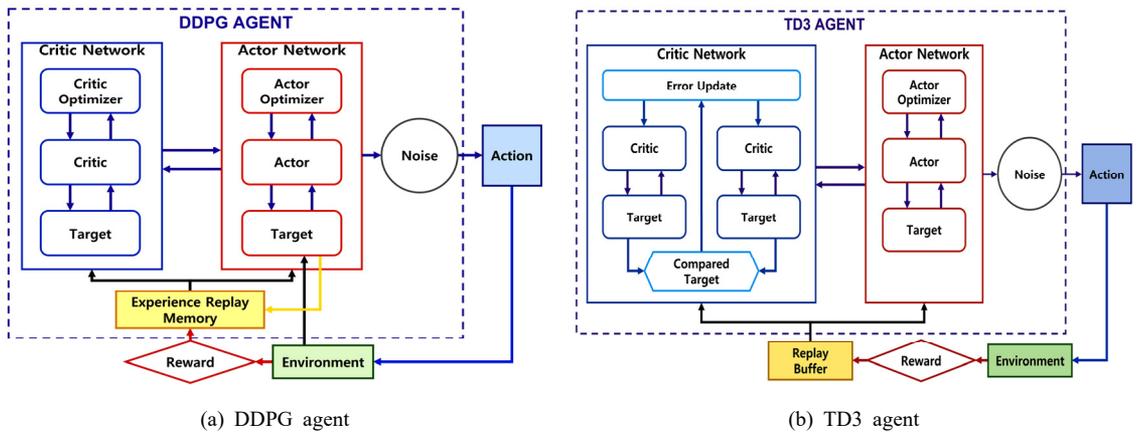


Fig. 2 Main concept of DDPG and TD3 agents

상의 업데이트 횟수를 줄여 문제점을 개선하고 발전 시킨 에이전트이다⁽¹¹⁾.

Fig. 2의 (a)는 DDPG 에이전트의 전개 과정이다. 환경에서 관측된 상태정보가 배우 네트워크로 전달되고, 전달된 상태정보 값에 기반하여 행동을 출력으로 내보내어 보상 값과 다음 상태 값(s_{i+1})을 관찰하며 이러한 데이터를 메모리에 저장하였다가 무작위로 선택하여 학습에 이용한다. 학습 시 배우 네트워크에서 도출된 행동의 결과에 대한 가치를 평가한 후 보상이 높은 행동을 반복할 수 있도록 정책을 업데이트하며 학습을 위해서 두 네트워크가 상호작용한다. 학습이 끝나게 되면 DDPG 에이전트는 각각의 신경망을 제어하면서 도출되는 제어입력 값으로 환경을 제어하게 된다.

또한 Fig. 2의 (b)는 TD3 에이전트의 전개 과정이며 이는 앞서 설명하였듯 DDPG 에이전트의 기존 비평가 네트워크에 비평가 네트워크를 하나 더 추가하여 상호 비교를 통해 Q-값의 과대평가를 방지한다는 것과 환경에서 바로 배우 네트워크로 상태정보가 바로 전달되지 않는다는 것이 Fig. 2의 (a)와의 가장 큰 차이점이다. TD3의 손실 및 업데이트 방식은 식 (2)~(5)와 유사하며 보다 자세한 내용은 참고문헌을 참고하기 바란다⁽¹²⁾.

이 논문에서 사용한 TD3 에이전트는 sprung mass의 절대속도(\dot{x})와 하나의 완전연결 레이어(fully connected layer)를 가진 신경망으로 모델링하여, 배우 신경망(actor neural network)의 결과물로 skyhook 제어를 위한 제어게인 값을 도출한다. 모델링 된 식은 식 (6)과 같으며 식에서 u 는 배우 신경망의 결과이며 K_p 는 제어기 신경망 가중치의 절대값을 의미한다. 식 (6)은 경사 하강 최적화(gradient descent optimization)를 통해 최적의 고정 제어 게인을 도출하도록 업데이트된다.

$$u = [-1 * \dot{x}] * [K_p] \tag{6}$$

따라서 앞서 전개하였듯이 DDPG 에이전트는 연속된 환경에서 연속된 제어 입력을 실시간으로 도출하며 TD3 에이전트는 최적의 제어 게인을 도출하게 된다.

이 논문에서는 이렇게 유사하지만 명확한 차이를 보이는 TD3와 DDPG 에이전트를 이용하여 능동 현가장치 제어를 위한 강화학습 알고리즘 설계를 진행하였다.

2.2 소형 전기차 제어 강화학습 알고리즘

앞서 전개한 강화학습의 이론적 내용을 바탕으로 이 논문의 목표인 삼륜 소형 전기차의 외란에 의한 진동을

효과적으로 제어하기 위해 강화학습을 이용한 시뮬레이션 알고리즘을 구현하였다. 에이전트는 환경 내에서 각각의 도로 별 외란을 시간에 따라 가진 받았으며 이를 현가장치 상태공간방정식에 대해 sprung mass 속도(velocity)의 값을 도출해 관측(observation)함수로 하여 보상(reward)함수와 함께 에이전트의 입력으로 설정하였다. 보상함수는 행동을 수행하면서 sprung mass의 속도, 기준 값과의 오차, 제어입력의 크기 제곱의 기대 값을 모두 이용하여 정의하였으며 목표에 도달할수록 보상이 커지도록 설계하였다. 따라서 가장 완벽하게 진동을 제어했을 때의 값인 0을 기준(reference)으로 하여 sprung mass의 속도와 오차를 -10의 보상을 하고 에이전트의 행동(action)에 -0.01의 보상을 하여 일정시간 지연(delay)을 통해 이전 시행과 구분을 하여 보상으로 제공하였다.

강화학습은 앞서 설명한 DDPG와 TD3의 복잡한 알고리즘과 여러 함수들을 단순화하여 설계를 진행하기 위해 MATLAB의 reinforcement learning toolbox와 MATLAB simulink를 이용하였다. 이들을 활용해 소형 삼륜 전기차의 sprung mass 속도를 제어하기 위한 다양한 도로에서의 에이전트에 따른 실시간 제어 입력과 최적의 제어 게인을 찾도록 설계하였으며 이 연구에서 사용한 DDPG와 TD3의 hyper parameter는 Table 1과 같이 각 에이전트의 특성에 따라 값에 약간의 차이를 두어 설정하였다.

2.3 소형 전기자동차 제어

이 연구의 대상인 삼륜형 소형 전기차는 소형 타이어

Table 1 Hyper-parameter for reinforcement learning

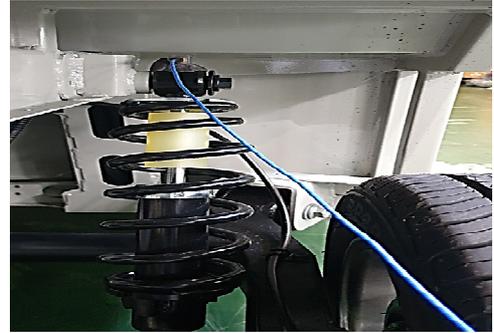
Hyper-parameter	DDPG_value	TD3_value
Max step	300	300
Actor learning rate	1e -05	1e -03
Critic learning rate	1e -04	1e -03
Gradient threshold	1	1
Sampling time	0.01	0.01
Mini batch size	64	128
Experience buffer length	1e 06	1e 06
Target smooth factor	1e -03	0.005
Discount factor	1.0	0.99
Score averaging window length	100	100

사용에 따른 운동 성능 및 승차감이 매우 낮으며 복잡한 현가장치 시스템을 설치할 공간 또한 대형차량보다 부족하여 단일 현가장치 제어기의 진동 제어력이 매우 중요하다. 따라서 이를 강화학습을 이용하여 제어하였으며 이러한 제어기의 개발을 위해선 소형 삼륜 전기차의 시뮬레이션 환경구축이 필요하다. 이에 이 연구자는 소형 삼륜 전기차의 한쪽 현가장치만을 고려하여 Fig. 3의 (a)와 같이 모델링하였다. 모델은 2 자유도 (2-dof) 모델의 수직방향 힘만을 고려하였으며 외력에 의한 1차적인 충격을 흡수하는 스프링과 스프링의 자유진동을 감쇠하여 승차감을 향상시키는 댐퍼를 Sprung mass(m_s)와 unsprung mass(m_u) 사이에 연결하여 설계하였다. Fig. 3(b)는 모델링에 사용된 후륜 현가장치의 실제 사진이다. 차량의 세부 제원은 Fig. 4와 같은 실제 소형 삼륜 전기 자동차의 제원 값을 따라 총 중량을 1500 kg으로 하여 이를 1/3로 나눈 500 kg을 sprung mass로 하였으며 unsprung mass는 29.5 kg으로 설정하였다. 또한 차량의 현가장치 제원은 sprung mass와 unsprung mass 사이에 위치해 있는 스프링의 상수 값(k_s)에 대해서는 19 600 N/m, 댐퍼 감쇠 값(c_s)은 1000 Nm, 스프링과 비슷한 역할을 수행하는 차량 바퀴의 스프링 상수 (k_t)값은 2 000 000 N/m로 각각 설정하여 설계하였다. 이러한 제원들을 대입하여 1/3 모델의 운동방정식을 유도하면 식 (7)과 같으며 인가되는 제어입력을 u 로 표현한 상태공간방정식은 식 (8)로 도출할 수 있다.

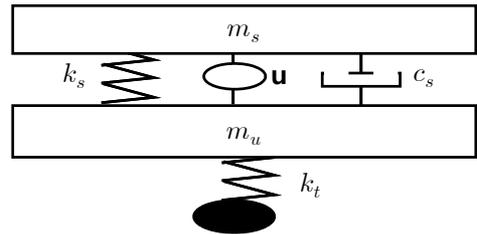
$$\begin{aligned}
 m_s \ddot{z}_s &= -k_s(z_s - z_u) - c_s(\dot{z}_s - \dot{z}_u) + u \\
 m_u \ddot{z}_u &= -k_s(z_u - z_s) - c_s(\dot{z}_u - \dot{z}_s) \\
 &\quad - k_t(z_u - z_r) - u
 \end{aligned} \tag{7}$$

$$\begin{aligned}
 \dot{x} &= Ax + Bu + Lz_r \\
 A &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-k_s}{m_s} & \frac{-c_s}{m_s} & \frac{k_s}{m_s} & \frac{c_s}{m_s} \\ 0 & 0 & 0 & 1 \\ \frac{k_s}{m_u} & \frac{c_s}{m_u} & -\frac{(k_s + k_t)}{m_u} & \frac{-c_s}{m_u} \end{bmatrix} \\
 B &= \begin{bmatrix} 0 & \frac{1}{m_s} & 0 & \frac{-1}{m_u} \end{bmatrix}^T \\
 L &= \begin{bmatrix} 0 & 0 & 0 & \frac{k_t}{m_u} \end{bmatrix}^T
 \end{aligned} \tag{8}$$

Fig. 3의 (a)에서 나타내었듯이 이 연구에서 설계한 현가장치는 제어가 가능한 댐퍼가 장착 되어있는 시스템이다. 이에 이 연구자는 이곳에 장착될 제어기를 강화학습의 DDPG와 TD3 에이전트를 이용하여 학습을 진행한 후 도출되는 제어 게인이 sprung mass의 절대속도(\dot{x})와 곱해져 제어기의 제어입력으로 계산되



(a) Suspension modeling



(b) Vehicle suspension system

Fig. 3 Three wheels micro electric vehicle suspension system



Fig. 4 Three wheels micro electric vehicle

어 제어를 수행하도록 하였다.

학습을 완료한 DDPG는 실시간으로 제어입력을 도출하여 연속적인 행동과 관찰에 기반한 연속적인 제어를 수행하는 만큼 별도로 하나의 고정된 제어 계인을 도출하진 않았지만 학습한 에이전트를 제어기에 적용하여 도출되는 실시간 제어입력을 도로의 파형 별 대입하며 비교하였다. 또한 학습을 마친 TD3는 최적의 제어 계인 값을 5.6724로 도출하였으며 이렇게 도출한 최적의 제어 계인 값을 제어기에 적용하여 제어 결과를 도로의 파형을 다양하게 바꾸며 비교하였다.

2.4 강화학습 결과 도출방법

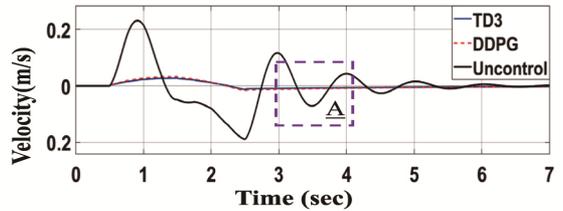
소형 전기자동차를 DDPG와 TD3 에이전트 알고리즘으로 학습하여 도출한 실시간 제어 입력과 최적 제어 계인은 도로 파형을 다르게 설정해 각각의 제어 전과 후의 sprung mass 속도를 비교하여 결과를 도출하였다. 도로는 총 4가지의 유형으로 설정하였으며 이는 과거 연구사례^(13, 14)에서 차량 현가장치의 수직 동적운동 및 승차감을 평가하기 위한 도로 입력으로 사용된 도로 유형으로 주변에서 쉽게 볼 수 있는 요철로(bump road)와 불규칙한 파형 및 거칠기를 가진 도로(B, C, D-class road)를 현가장치 도로입력으로 선정한 것이다. 도로 파형의 제원은 높이 0.1 m의 요철로와 도로 거칠기 계수 8.1×10^{-6} 과 파동계수 2.1의 B-class road, 도로 거칠기 계수 4.8×10^{-7} 과 파동계수 2.1의 C-class road, 도로 거칠기 계수 4.4×10^{-6} 과 파동계수 2.1의 D-class road이며 도로별로 각각 나누어 대입하여 검증하였다. 여기서 D-class road는 B, C, D-class road 중 가장 거친 도로이며 오프로드에 준하는 수준의 파형을 가진 도로이다. 또한 각각의 도로를 ISO 2631의 w_k 기준에 대입하여 제어 전/후 탑승자가 느끼는 진동에 대해서 감쇠 정도를 수치적으로 비교하여 결과를 분석하였다.

그리고 설계된 강화학습 기반 제어기의 성능 및 상호 영향성을 검증하기 위하여 삼륜 소형 전기차를 7 자유도 전 차량 모델(7-dof. full car)로 모델링하여 제어결과를 시뮬레이션 하였다. 전 차량 모델에 대해 12개의 상태변수를 가지는 상태공간 방정식은 앞서 전개한 식 (8)과 같으며 전 차량 모델에서의 A, B, L 행렬은 부록에 작성되어 있다.

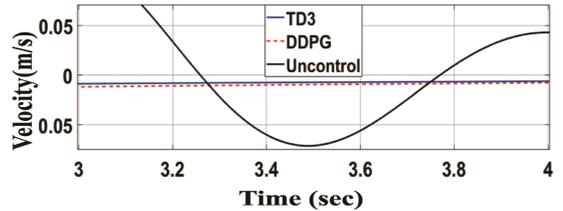
3. 결과 및 분석

3.1 제어 전후 결과 분석

소형 전기 자동차의 도로 별 제어기 도입 전후 결과 분석을 위해서는 도로 파형을 구분하여 에이전트 별로 비교를 진행하여야 한다. 이를 위해 도로 파형을 외란으로 7초간 가졌었으며 제어기에 TD3 에이전트의 최적 제어 계인을 대입한 결과와 DDPG 에이전트의 실시

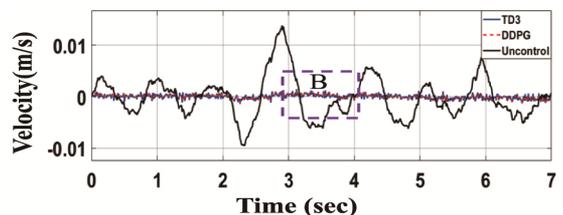


(a) Sprung mass velocity results

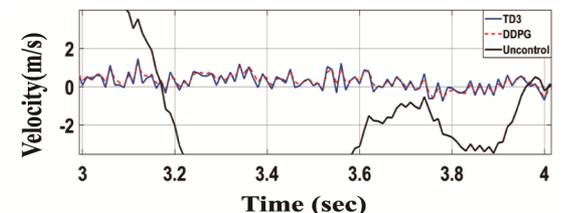


(b) Detail A

Fig. 5 Results of bump road



(a) Sprung mass velocity results



(b) Detail B

Fig. 6 Results of D-class road

Table 2 Results of ride comfort

Road	Uncontrol	DDPG	TD3
Bump road	0.333	0.029	0.024
B-class road	0.006	4.344e-04	3.747e-04
C-class road	0.011	8.682e-04	7.490e-04
D-class road	0.022	0.0017	0.0015

간 제어 입력을 대입한 결과, 대입하지 않은(uncontrol) 결과를 한 그래프에 표현하였다. 여기서 비제어(uncontrol) 상태란 4가지 도로유형을 주행하며 강화학습을 통해 도출한 제어 게인 및 실시간 제어 입력 값이 대입되지 않은 오직 현가장치의 댐퍼와 스프링에 의해서만 진동을 감쇠하는 상태로 소형 삼륜 자동차 현가장치 댐핑 값은 대입되어 있으나 제어기에 대한 게인은 대입되지 않은 상태이다. 결과는 Fig. 5와 Fig. 6에 나타내었다. Fig. 5에는 요철로에 대한 sprung mass의 속도를 시간(sec)영역에 대한 속도(m/s)로 나타내었으며 Fig. 6은 D-class road에 대한 sprung mass의 속도 결과를 나타낸 그래프이다. 각 결과 그래프는 7초까지의 전체그래프 (a)와 그 그래프의 부분 확대도 (b)로 나누어서 나타내었다.

결과는 DDPG와 TD3 모두 비제어 상태에 비해 제어가 우수하게 수행되었음을 보였으며 또한 B-class road와 C-class road에 대한 결과도 Table 2에 나타내었듯 비제어 상태 대비 제어 게인 및 실시간 제어입력 도입 후 약 93% 더 좋은 결과를 보이며 제어가 잘 수행되었음이 확인되었다.

3.2 승차감 평가 및 분석

소형 전기 자동차의 탑승자가 느끼는 수치적 결과 분석을 위해서 널리 활용되는 ISO 2631의 w_k 기준을 적용하였다. ISO 2631의 w_k 기준은 수직진동이 몸 전체(whole body)에 걸쳐 진동하는 것으로 앉아있는 경우에 대하여 탑승자가 진동을 느끼는 정도에 대하여 수치적으로 표현하기 위한 기준으로 차량의 승차감을 평가하는 지표이다. 해당 결과는 sprung mass의 가속도 데이터를 기반으로 도출하였으며 수치가 작을수록 좋은 결과이다. 해당 결과를 정리한 표는 Table 2와 같다. 비제어 결과에 비해 제어 후 결과들이 모두 우수한 결과를 보였으며 그중에서도 TD3의 결과가 가장 좋은 결과를 나타내었다. 하지만 TD3 에이전트는

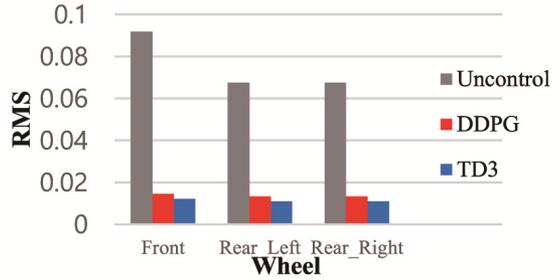


Fig. 7 RMS results of controllable full car model

고정된 게인을 도출하기 때문에 실시간으로 급변하는 상황에서의 대응이 어려울 수도 있다는 단점이 존재하며 DDPG 에이전트는 결과가 TD3보다는 좋긴 않지만 실시간 제어가 가능하여 다양한 상황별 대처가 가능하다는 장점이 있다.

이러한 결과들을 통해 강화학습을 이용하여 도출한 제어기의 제어 게인과 실시간 제어입력이 다양한 도로의 과형에 따른 진동을 효과적으로 감쇠시킴을 확인할 수 있었다.

3.3 전 차량 모델 제어 전후 결과 분석

앞서 강화학습을 이용하여 학습한 1/3모델기반 제어기 간의 간섭과 상호 영향성이 없음을 확인하기 위하여 소형 전기 자동차의 제어기 도입 전후 결과를 전차량모델(full car)에 대해서 도출하였다. 전차량모델은 식 (8)을 이용하였으며 상세 수식은 부록에 첨부하였다. 앞서 도출된 동일한 제어기를 적용하였을 때 제어 결과는 Fig. 7과 같으며 해당 결과는 sprung mass 속도의 제곱 평균 제곱근(RMS) 수치로 값이 낮을수록 우수한 결과이다. 그래프에서 볼 수 있듯 앞바퀴와 양쪽 뒷바퀴 모두 비제어 상태에 비해 제어기가 도입되었을 때의 결과가 월등히 좋으며 그중에서도 TD3의 결과가 가장 좋은 결과를 보였다. 이에 에이전트별 학습이 잘 진행되었으며 각각의 제어기 간의 간섭과 상호 영향성에 문제가 없음을 확인할 수 있었다.

4. 결 론

이 연구에서는 강화학습 알고리즘을 이용한 현가장치 제어기의 제어 게인과 실시간 제어입력 도출 알고리즘 설계를 수행하였다. 삼륜 전기 자동차를 기반으로 1/3모델을 구성하고 상태공간방정식을 도출하였으

며 이를 강화학습의 DDPG와 TD3 에이전트 알고리즘을 이용하여 실시간 제어입력과 제어 게인 도출을 위한 학습을 수행시켰다. 이에 대한 결과로 다양한 도로 파형에 따른 진동 제어 성능을 제어 전후 속도 그래프와 ISO 2631의 w_k 기준을 통해 평가하였으며 우수한 진동 감쇠 결과를 확인할 수 있었다. 또한 학습시켜 설계한 제어기를 삼륜 소형 전기차 전 차량 모델에 적용시켜 결과를 도출하여 제어기간의 간섭이 없음을 확인하여 이 논문의 제어 알고리즘 설계에 문제가 없음을 확인하였다.

이 연구를 이용하여 추후 다른 차량 모델을 모델링하여 강화학습을 진행할 것이며, 기타 다른 발전된 에이전트를 이용한 최적 제어 게인 및 실시간 제어입력 도출 알고리즘의 설계를 연구할 것이다. 또한 이 논문에서 도출한 결과를 시뮬레이션 프로그램(carsim)에 적용하여 다양한 운전 조건 및 상황에 대한 심층적인 결과를 도출할 예정이다.

후 기

이 과제는 2021년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다(2021RIS-004).

References

(1) Sung, K. G. and Choi, S. B., 2008, Vibration Control of Vehicle Suspension Featuring Magnetorheological Dampers: Road Test Evaluation, Proceedings of the KSNVE Annual Autumn Conference, pp. 980-985.

(2) Rao, L. G. and Narayanan, S., 2020, Optimal Response of Half Car Vehicle Model with Sky-hook Damper Based on LQR Control, International Journal of Dynamics and Control, Vol. 8, No. 2, pp. 488-496.

(3) Budireddi, G., 2021, Simulation Analysis of a Half Car Model with Sky-hook Damper using Matlab/Simulink, Design Engineering, pp. 2759-2771.

(4) Howell, M. N., Frost, G. P., Gordon, T. J. and Wu, Q. H., 1997, Continuous Action Reinforcement Learning Applied to Vehicle Suspension Control, Mechatronics, Vol. 7, No. 3, pp. 263-276.

(5) Fares, A. and Bani Younes, A., 2020, Online

Reinforcement Learning-based Control of an Active Suspension System Using the Actor Critic Approach, Applied Sciences, Vol. 10, No. 22, p. 8060.

(6) Grondman, I., Busoniu, L., Lopes, G. A. and Babuska, R., 2012, A Survey of Actor-critic Reinforcement Learning: Standard and Natural Policy Gradients, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 42, No. 6, pp. 1291-1307.

(7) Reinforcement Learning Agents, <https://kr.math-works.com/help/reinforcement-learning/ug/create-agents-for-reinforcement-learning.html> (accessed 11. 11. 2021).

(8) Kim, S., Kim, H. and Kang, D., 2018, Vibration Control of a Vehicle Active Suspension System Using a DDPG Algorithm, 18th International Conference on Control, Automation and Systems (ICCAS), pp. 1654-1656.

(9) Ming, L., Yibin, L., Xuewen, R., Shuashuai, Z. and Yanfang, Y., 2020, Semi-active Suspension Control Based on Deep Reinforcement Learning, IEEE Access, Vol. 8, pp. 9978-9986.

(10) Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D., 2015, Continuous Control with Deep Reinforcement Learning, arXiv preprint arXiv:1509.02971.

(11) Fujimoto, S., Hoof, H. and Meger, D., 2018, Addressing Function Approximation Error in Actor-critic Methods, Proceedings of the 35th International Conference on Machine Learning, arXiv:1802.09477, pp. 1587-1596.

(12) Zhang, F., Li, J. and Li, Z., 2020, A TD3-based Multi-agent Deep Reinforcement Learning Method in Mixed Cooperation-competition Environment, Neurocomputing, Vol. 411, pp. 206-215.

(13) Bashir, A. O., Rui, X. and Zhang, J., 2019, Ride Comfort Improvement of a Semi-active Vehicle Suspension Based on Hybrid Fuzzy and Fuzzy-PID Controller, Studies in Informatics and Control, Vol. 28, No. 4, pp. 421-430.

(14) Klockiewicz, Z., Ślaski, G. and Spadło, M., 2019, The Influence of the Conditions of Use and the Type of Model Used on the Vertical Dynamic Responses of a Car Suspension, Archiwum Motoryzacji, Vol. 85, No. 3, pp. 57-82.

부 록

Matrix for State Space Equation

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -k_f - 2k_r & -c_f - 2c_r & \frac{b_2k_r - b_1k_r}{M} & \frac{b_2c_r - b_1c_r}{M} & \frac{a_1k_f - 2a_2k_r}{M} & \frac{a_1c_f - 2a_2c_r}{M} & \frac{k_f}{M} & \frac{c_f}{M} & \frac{k_r}{M} & \frac{c_r}{M} & \frac{k_r}{M} & \frac{c_r}{M} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{b_2k_r - b_1k_r}{I_y} & \frac{b_2c_r - b_1c_r}{I_y} & -b_2^2k_r - b_1^2k_r & -b_2^2c_r - b_1^2c_r & \frac{a_2b_2k_r - a_2b_1k_r}{I_y} & \frac{a_2b_2c_r - a_2b_1c_r}{I_y} & 0 & 0 & -\frac{b_2k_r}{I_y} & -\frac{b_2c_r}{I_y} & \frac{b_1k_r}{I_y} & \frac{b_1c_r}{I_y} \\ 1_x & 1_x & 1_x & 1_x & 1_x & 1_x & 0 & 0 & 1_x & 1_x & 1_x & 1_x \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ a_1k_f - 2a_2k_r & a_1c_f - 2a_2c_r & a_2b_2k_r - a_2b_1k_r & a_2b_2c_r - a_2b_1c_r & -a_1^2k_f - 2a_2^2k_r & -a_1^2c_f - 2a_2^2c_r & -a_1k_f & -a_1c_f & a_2k_r & a_2c_r & a_2k_r & a_2c_r \\ I_y & I_y \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{k_f}{m_f} & \frac{c_f}{m_f} & 0 & 0 & -\frac{a_1k_f}{m_f} & -\frac{a_1c_f}{m_f} & -\frac{k_{tf}}{m_f} & -\frac{c_f}{m_f} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_r}{m_r} & \frac{c_r}{m_r} & -\frac{b_2k_r}{m_r} & -\frac{b_2c_r}{m_r} & \frac{a_2k_r}{m_r} & \frac{a_2c_r}{m_r} & 0 & 0 & -\frac{k_{tr} - k_r}{m_r} & -\frac{c_r}{m_r} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{m_r}{m_r} & \frac{m_r}{m_r} & 0 & 0 \\ \frac{k_r}{m_r} & \frac{c_r}{m_r} & \frac{b_1k_r}{m_r} & \frac{b_1c_r}{m_r} & \frac{a_2k_r}{m_r} & \frac{a_2c_r}{m_r} & 0 & 0 & 0 & 0 & -\frac{k_{tr} - k_r}{m_r} & -\frac{c_r}{m_r} \end{bmatrix}$$

$$B^T = \begin{bmatrix} 0 & \frac{1}{M} & 0 & 0 & 0 & -\frac{a_1}{I_y} & 0 & -\frac{1}{m_f} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{M} & 0 & -\frac{b_2}{I_x} & 0 & \frac{a_2}{I_y} & 0 & 0 & 0 & -\frac{1}{m_r} & 0 & 0 \\ 0 & \frac{1}{M} & 0 & \frac{b_1}{I_x} & 0 & \frac{a_2}{I_y} & 0 & 0 & 0 & 0 & 0 & -\frac{1}{m_r} \end{bmatrix}$$

$$L^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{k_{tf}}{m_f} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{k_{tr}}{m_r} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{k_{tr}}{m_r} & 0 \end{bmatrix}$$



Young-jun Kim is a master's student. He is currently conducting a master's degree program at Kongju National University Graduate School. His research interests are suspension device vibration control and vehicle steering control.



Jong-Seok Oh received the Ph.D. degrees in Mechanical Engineering from Inha University in 2015, respectively. He is currently assistant professor in Kongju National University. His research interests are robust controller design and control

of various systems using smart actuators such as magnetorheological (MR), electrorheological (ER) fluids, and piezo actuators.