

대형 상용차량 종방향 속도추종 성능향상을 위한 LQR 페루프 참조모델 기반 잔차 강화학습 제어 방법론

변진규¹⁾ · 김지원¹⁾ · 김선희²⁾ · 유진우^{*3)}

국민대학교 자동차모빌리티대학원¹⁾ · 국민대학교 자동차공학과²⁾ · 국민대학교 자동차IT융합학과³⁾

LQR Closed-Loop Reference Model-Based Residual Reinforcement Learning Control Methodology for Improving Longitudinal Speed Tracking Performance of Heavy-Duty Vehicles

Jingyu Byeon¹⁾ · Jiwon Kim¹⁾ · Sunhee Kim²⁾ · Jinwoo Yoo^{*3)}

¹⁾Graduate School of Automobile and Mobility, Kookmin University, Seoul 02707, Korea,

²⁾Department of Automotive Engineering, Kookmin University, Seoul 02707, Korea

³⁾Department of Automobile and IT Convergence, Kookmin University, Seoul 02707, Korea

(Received 5 August 2026 / Revised 31 August 2026 / Accepted 31 August 2026)

Abstract : This paper proposes a model-reference-based reinforcement learning control method to improve longitudinal speed tracking of heavy-duty vehicles. Because of their large mass and inertia, such vehicles show significant variations in longitudinal dynamics due to payload, road grade, driving resistance, and external disturbances. These conditions make consistent tracking difficult with a single model-based controller. To address this, the study uses a Linear Quadratic Regulator (LQR) based on a linearized longitudinal model as the baseline controller. A residual control structure, in which a reinforcement learning agent compensates for the state deviation between the LQR closed-loop reference model and the actual vehicle, is also presented. A Twin Delayed Deep Deterministic Policy Gradient (TD3)-based residual policy generates additional compensation input according to this state deviation. Validated through IPG TruckMaker simulations under various driving conditions, the proposed method reduced speed-tracking error when compared with conventional controllers. Results showed improved longitudinal speed-tracking performance under model variations and disturbances.

Key words : Heavy-duty vehicle(대형 상용 차량), Longitudinal speed tracking control(종방향 속도 추종 제어), Reinforcement learning(강화학습), Longitudinal vehicle dynamics(종방향 차량 동역학), Model uncertainty(모델 불확실성), Adaptive control(적응 제어)

1. 서론

최근 자율주행 및 첨단 운전 보조 시스템의 적용 범위가 승용차 중심에서 대형 상용 차량으로 확대됨에 따라, 물류 운송, 장거리 주행, 군집 주행 등 다양한 운행 환경에서 안정적인 차량 제어 성능 확보가 중요해졌다.^{1,2)} 특히 대형 상용차량은 장시간 고속도로 주행과 반복적인 가감속 상황이 빈번하게 발생하므로, 목표 속도를 안정적으로 추종하는 종방향 제어 기술은 주행 안전성, 승차

감 및 에너지 효율 측면에서 핵심적인 요소로 작용한다.³⁻⁵⁾

그러나 대형 상용차량은 승용차 대비 큰 질량과 관성 특성을 가지며, 적재하중 변화에 따라 차량의 종방향 동특성이 크게 변화한다.^{6,7)} 또한 도로 구배, 주행 저항, 공기 저항, 외란 등의 영향으로 동일한 제어 입력에 대해서도 차량 응답이 달라질 수 있다.⁸⁾ 이러한 특성은 차량 모델의 불확실성을 증가시키며, 단일 모델에 기반한 기존 제어기만으로는 다양한 주행 조건에서 일관된 속도 추종

*Corresponding author, E-mail: jwyoo@kookmin.ac.kr

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.

성능을 확보하는 데 한계가 있다.^{9,10)}

종방향 차량 제어를 위해 PID(Proportional-Integral-Derivative) 제어와 같은 전통적인 제어 기법부터 LQR (Linear Quadratic Regulator), MPC(Model Predictive Control)와 같은 모델 기반 최적 제어 기법까지 다양한 방법들이 연구되어 왔다.¹¹⁻¹⁴⁾ 이러한 제어 기법은 수학적 모델을 기반으로 안정적인 제어 성능을 제공할 수 있다는 장점이 있다. 그러나 차량 파라미터 변화나 외란 조건이 설계 모델과 크게 달라지는 경우 제어 성능이 저하될 수 있으며, 이를 보완하기 위해서는 정확한 차량 모델링 또는 외란 추정 과정이 요구된다.¹⁵⁾

차량 시스템의 모델 불확실성과 외란에 대응하기 위해 적응 제어 기법이나 강인 제어와 같은 모델 기반 제어 기법들이 연구되어 왔다.^{16,17)} 이러한 기법들은 시스템 파라미터 변화나 외부 환경 변화에 대한 제어 성능을 향상시킬 수 있지만, 정확한 시스템 구조에 대한 가정이나 설계 조건에 따라 성능이 제한될 수 있다.

최근에는 이러한 한계를 보완하기 위해 강화학습 기반 제어 기법이 차량 제어 분야에 적용되고 있다.¹⁸⁻²⁰⁾ 강화학습은 시스템 모델을 명시적으로 정의하지 않더라도 상태와 보상 정보를 기반으로 제어 정책을 학습할 수 있다는 장점이 있다.²¹⁾ 그러나 강화학습 제어기를 단독으로 사용할 경우 학습 과정에서의 안정성 확보가 어렵고, 실제 차량 제어에 적용하기 위해서는 탐색 과정에서 발생할 수 있는 불안정한 제어 입력을 고려해야 한다.²²⁾

이를 보완하기 위해 모델 기반 제어기에 강화학습 기반 보정 입력을 더하는 잔차 강화학습 구조가 제안되어 왔다.^{23,24)} 잔차 강화학습은 기존 제어기가 보정하지 못하는 성분만을 학습하여 탐색 범위를 줄일 수 있다. 일반적인 목표 추종형 잔차 강화학습에서는 목표 명령과 실제 차량 상태 간의 오차를 학습 기준으로 사용할 수 있다. 그러나 목표 명령은 차량이 도달해야 할 값만을 나타낼 뿐 그 값에 도달하기까지의 응답 과정을 포함하지 않으므

로, 이 경우 기본 제어기가 설계한 응답 특성은 학습 기준에 반영되지 않는다.

본 논문에서는 대형 상용차량의 종방향 속도 추종 성능 향상을 위해 모델-참조 기반 강화학습 제어(Model-Reference Reinforcement Learning Control, MRRLC) 방법론을 제안한다. 제안 방법은 공칭 종방향 모델을 기반으로 설계된 LQR 제어기를 기본 제어기로 사용하고, LQR 페루프 참조 모델이 생성한 공칭 차량 동역학 및 기본 제어기의 페루프 특성이 반영된 상태 응답과 실제 상태 간의 편차를 잔차 강화학습 정책의 학습 기준으로 정의한다. 최종 제어 입력은 LQR 기본 입력과 강화학습 기반 잔차 입력의 합으로 구성되며, 잔차 정책은 기본 제어기를 대체하지 않고 실제 차량이 공칭 페루프 응답에서 벗어난 편차를 보정하는 추가 제어 입력만을 생성한다. 또한 외란이 없는 모델 불일치 조건에서 잔차 정책을 학습하고, 적재하중 변화와 맞바람이 추가된 조건에서 성능을 평가하여, 학습 조건에 포함되지 않은 모델 변화 및 외란 조건에서의 보정 성능을 검증하였다.

2. 제안하는 종방향 강화학습 제어 방법론

2.1 전체 아키텍처

본 절에서는 제안하는 종방향 강화학습 제어 방법론의 전체적인 구조에 대해서 설명한다. Fig. 1은 제안 구조의 전체 구성을 나타낸다. 제안 구조는 목표 속도 $r(t)$ 에 대해 공칭 차량 동역학과 LQR 페루프 특성이 반영된 응답을 생성하는 참조 모델(Reference model), 안정적인 기본 제어 입력 u_{LQR} 을 생성하는 LQR(Linear Quadratic Regulator) 기반 기본 제어기, 참조 모델과 실제 차량 간의 상태 오차 $e(t)$ 를 입력 받아 잔차 제어 입력 u_{RL} 을 생성하는 강화학습 에이전트, 그리고 실제 차량 플랜트로 구성된다. 최종 제어 입력은 기본 제어 입력과 잔차 제어 입력의 합 u_{tot} 으로 구성되어 차량 플랜트에 입력된다.

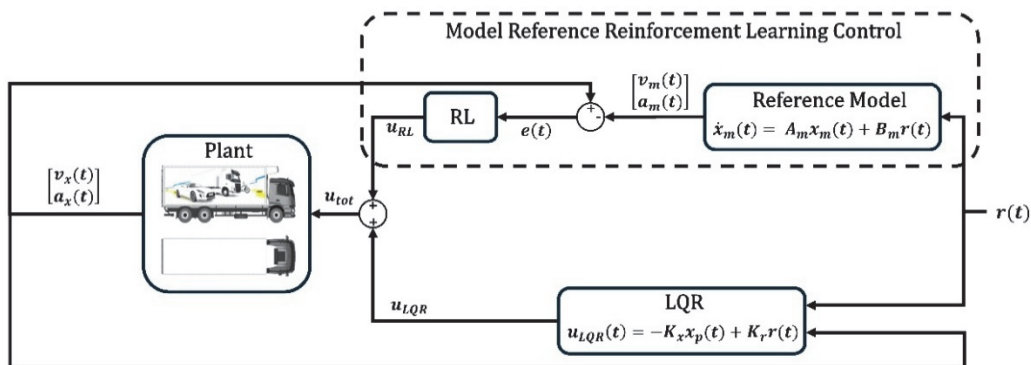


Fig. 1 Proposed RL control architecture

2.1.1 차량 종방향 동역학 모델

대형 상용차량은 적재하중, 도로 구배, 공기 저항 등의 영향으로 종방향 응답 특성이 크게 변화한다. 차량의 종방향 운동은 뉴턴의 제2법칙에 따라 구동력과 총 주행 저항력의 관계로 표현된다. Fig. 2는 경사로에서 대형 상용차량에 작용하는 종방향 힘을 나타낸다.

$$ma_x = F_T - F_R \quad (1)$$

여기서 m 은 차량 질량, a_x 는 종방향 가속도, F_T 는 구동력, F_R 은 총 주행 저항력이다. 식 (1)은 구동력과 총 주행 저항력의 차이에 따라 차량의 종방향 가속도가 결정됨을 나타낸다. 총 주행 저항력은 구배 저항 F_g , 구름 저항 F_r , 공기 저항 F_a 의 합으로 구성된다.

$$F_R = F_g + F_r + F_a \quad (2)$$

where $F_g = mg \sin \gamma$

$$F_r = C_r mg \cos \gamma$$

$$F_a = 0.5\rho C_d A_f (v_x - v_{air})^2$$

여기서 γ 는 도로 구배각, C_r 은 구름 저항 계수, ρ 는 공기 밀도, C_d 는 공기 저항 계수, A_f 는 전면 면적, v_{air} 는 풍속이다. 식 (2)에서 알 수 있듯이 종방향 가속도는 구동력뿐만 아니라 도로 구배와 풍속 등의 주행 조건에 따라 달라질 수 있다.

특히 대형 상용차량은 큰 질량과 관성 및 구동계 특성을 가지며, 적재하중 변화와 구동계 응답 지연에 따라 공칭 모델과 실제 차량 응답 간의 차이가 발생할 수 있다. 또한 실제 구동계는 제어 입력에 즉각 응답하지 않으므로, 구동계 응답을 1차 지연 모델로 근사하였다. 제어 입력을 종방향 가속도 명령 $a_{x,des}$, 실제 반영되는 가속도를 a_x 라 하면 식 (3)과 같다.

$$\tau_d \dot{a}_x + a_x = a_{x,des} \quad (3)$$

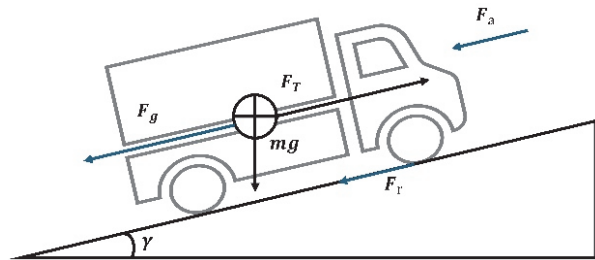


Fig. 2 Longitudinal dynamics of a vehicle

여기서 τ_d 는 구동계 응답 지연 시정수이다. 이를 바탕으로 차량 속도 v_x 와 가속도 a_x 를 상태 변수로, $a_{x,des}$ 를 제어 입력으로 하는 상태 공간 모델을 구성한다.

$$\dot{x}_p(t) = A_p x_p(t) + B_p u(t) \quad (4)$$

where $x_p(t) = [v_x(t), a_x(t)]^T$

$$u(t) = a_{x,des}(t)$$

$$A_p = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{\tau_d} \end{bmatrix}$$

$$B_p = \begin{bmatrix} 0 \\ \frac{1}{\tau_d} \end{bmatrix}$$

식 (4)의 상태 공간 모델은 구동계 응답 지연을 반영한 공칭 모델로, 이후 LQR 기본 제어기 설계와 LQR 페루프 참조 모델 구성에 활용된다.

2.1.2 LQR 기반 기본 제어기

앞 절에서 구성한 종방향 상태 공간 모델을 기반으로, 안정적인 속도 추종을 위한 기본 제어기로 LQR을 설계한다. LQR은 상태 변수의 추종 성능과 제어 입력의 크기를 동시에 고려하며, 다음의 2차 비용 함수를 최소화한다.

$$J = \int_0^\infty (x^T Q x + u^T R u) dt \quad (5)$$

여기서 Q 는 상태 가중치 행렬, R 은 제어 입력 가중치이다. Q 를 통해 차량 속도 및 가속도 상태의 추종 성능을 조정하고, R 을 통해 과도한 제어 입력을 억제한다. 비용 함수를 최소화하는 상태 피드백 이득 K_x 는 Riccati 방정식의 해로부터 계산되며, 참조 입력 이득 K_r 은 정상상태에서 목표 속도 추종이 이루어지도록 설정된다. 이를 이용한 LQR 기반 기본 제어 입력은 다음과 같이 정의된다.

$$u_{lqr}(t) = -K_x x_p(t) + K_r r(t) \quad (6)$$

여기서 $x_p(t)$ 는 실제 차량 상태, $r(t)$ 는 목표 속도 참조 입력이다. K_x 는 공칭 모델의 페루프 응답을 형성하는 상태 피드백 이득이고, K_r 는 참조 입력 이득이다.

본 연구에서는 공칭 조건에서의 속도 추종 응답과 제어 입력 크기를 주요 선정 기준으로 하여 Q 와 R 을 조정하였다. 먼저 속도와 가속도 상태를 모두 페루프 응답 설계에 반영하기 위해 두 상태에 동일한 수치의 초기 가중치를 부여하고, 상태 가중치를 조정하면서 목표 속도 추종 응답의 변화를 확인하였다. 이후 제어 입력 가중치 R

을 조정하여 과도한 가속도 명령을 억제하면서 추종 성능이 크게 저하되지 않는 범위를 선정하였다. 이러한 반복적인 페루프 응답 비교를 통해 최종적으로 $Q = \text{diag}(10,10)$, $R = 3$ 으로 선정하였다.

2.1.3 LQR 페루프 참조 모델 및 상태 오차 정의

앞서 설계한 LQR 기본 제어기와 공칭 종방향 모델을 이용하여 LQR 페루프 참조 모델을 구성한다. 참조 모델은 강화학습 기반 잔차 정책의 학습 기준을 정의하기 위한 것으로, 동일한 목표 속도 명령 $r(t)$ 에 대해 공칭 차량 동역학과 LQR 페루프 특성이 반영된 상태 응답을 생성한다. 참조모델의 상태벡터는 실제 차량과 동일하게 종방향 속도와 가속도로 구성하며, 식 (7)과 같이 정의한다.

$$x_m(t) = [v_m(t), a_m(t)]^T \quad (7)$$

여기서 $v_m(t)$ 와 $a_m(t)$ 는 각각 참조 모델의 종방향 속도와 가속도이다. 참조 모델은 실제 차량 상태 $x_p(t)$ 와 동일한 상태 구조를 가지므로, 실제 차량 응답과 공칭 페루프 응답간의 상태 편차를 구성할 수 있다. 참조 모델은 앞서 정의한 공칭 종방향 모델에 LQR 제어 입력을 적용하여 식 (8)과 같이 구성한다.

$$\dot{x}_m(t) = A_p x_m(t) + B_p u_{ref}(t) \quad (8)$$

$$\text{where } u_{ref}(t) = -K_x x_m(t) + K_r r(t)$$

여기서 $u_{ref}(t)$ 는 참조 모델에 적용되는 LQR 기반 참조 입력이다. 식 (8)에 $u_{ref}(t)$ 를 대입하면, 참조 모델의 페루프 상태 방정식은 다음과 같다.

$$\dot{x}_m(t) = A_m x_m(t) + B_m r(t) \quad (9)$$

$$\text{where } A_m = A_p - B_p K_x$$

$$B_m = B_p K_r$$

식 (9)에서 목표 속도 명령 $r(t)$ 는 공칭 차량 동역학과 LQR 페루프 특성이 반영된 참조 모델 상태 $x_m(t)$ 로 변환된다. 참조 모델은 실제 차량과 동일한 목표 속도 명령에 의해 구동되므로 기본 제어기와 동일한 속도 추종 목적을 갖는다. 그러나 목표 속도 명령을 직접 학습 기준으로 사용하는 대신, 목표 속도에 도달하기까지의 공칭 페루프 응답이 포함된 $x_m(t)$ 를 상태 참조로 사용한다.

이에 따라 기본 제어기가 형성하는 공칭 페루프 과도 응답과 실제 차량에서 모델 불일치 및 외란 등에 의해 추가적으로 발생하는 응답 편차를 구분할 수 있다.

$$e(t) = \begin{bmatrix} v_p(t) - v_m(t) \\ a_p(t) - a_m(t) \end{bmatrix} \quad (10)$$

여기서 식 (10)의 $e(t)$ 는 실제 차량이 공칭 LQR 페루프 응답으로부터 벗어난 상태 편차를 나타내며, 강화학습 기반 잔차 정책의 상태 입력으로 사용된다. 목표 속도와 실제 차량 속도의 차이를 직접 학습 기준으로 사용하는 경우에는 기본 제어기가 목표 속도로 수렴하는 과정에서 발생하는 과도응답 성분도 학습 오차에 포함될 수 있다. 반면 제안 방법은 실제 차량 상태와 참조 모델 상태 간의 편차를 이용함으로써, 잔차 정책이 공칭 페루프 응답에서 추가적으로 발생한 편차를 보정하도록 학습 문제를 구성한다. 또한 참조 모델 기반 상태 오차 $e(t)$ 는 편차를 발생시킨 모델 변화나 외란의 종류 및 크기를 별도의 입력으로 요구하지 않고, 실제 차량이 공칭 페루프 응답에서 벗어난 정도를 나타낸다. 이에 따라 본 연구에서는 외란 자체를 직접 식별하는 대신, 공칭 응답과 실제 차량 응답 간의 상태 편차를 기반으로 잔차 보정 정책을 학습하도록 구성한다.

2.1.4 보상 함수 설계

강화학습 에이전트의 학습 목표는 실제 차량 상태가 참조 모델의 응답을 추종하도록 하는 것이다. 이를 위해 보상 함수는 참조 모델과 실제 차량 간의 추종 오차를 최소화하는 동시에, 과도한 잔차 제어 입력이 발생하지 않도록 설계된다. 본 연구에서 사용한 보상 함수는 식 (11)과 같이 정의된다.

$$R_t = -e_t^T H_1 e_t - u_{rl,t}^T H_2 u_{rl,t} \quad (11)$$

여기서 H_1 은 상태 추종 오차에 대한 가중치 행렬, H_2 는 잔차 제어 입력에 대한 가중치 행렬이며, 두 행렬은 각각 $H_1 > 0, H_2 > 0$ 을 만족한다.

첫 번째 항은 실제 차량 상태가 참조 모델 상태를 추종하도록 유도하는 성능 항으로, 실제 차량과 공칭 페루프 참조 응답 간의 속도 및 가속도 편차를 함께 반영한다.

두 번째 항은 잔차 제어 입력의 크기에 대한 패널티 항이다. 대형 상용차량은 큰 질량과 관성으로 인해 종방향 가속도 시 하중 이동과 응답 지연이 크게 발생하므로, 추종 오차만을 최소화하도록 학습할 경우 에이전트가 과도한 잔차 입력을 생성하여 기본 제어기의 페루프 응답을 불필요하게 변화시키거나 제어 입력의 변동을 증가시킬 수 있다. 이를 방지하기 위해 잔차 입력에 대한 패널티를 부여하여 참조 모델 추종 오차와 잔차 제어 입력 크기 사이의 균형을 조절하였다.

따라서 본 보상함수는 실제 차량과 참조 모델 간의 상태 오차와 잔차 제어 입력의 크기를 동시에 고려하여, 공칭 페루프 응답에 대한 추종 오차를 줄이면서 과도한 잔차 입력을 억제하도록 구성하였다.

2.1.5 TD3 기반 잔차 정책 학습

본 연구에서는 전체 제어 입력을 강화학습 정책이 직접 생성하지 않고, LQR 기본 제어 입력에 추가되는 잔차 제어 입력만을 학습하도록 정책을 구성하였다. 잔차 정책은 앞 절에서 정의한 참조 모델 기반 상태 오차 $e(t)$ 를 입력으로 사용하여, 실제 차량이 공칭 페루프 응답에서 벗어난 정도에 따라 추가적인 보정 입력 $u_{rl}(t)$ 을 생성한다. 이에 따라 기본적인 목표 속도 추종은 LQR이 담당하고, 잔차 정책은 공칭 페루프 응답에서 추가적으로 발생한 편차를 보정하는 역할을 수행한다.

잔차 제어 입력은 연속적인 종방향 가속도 보정량으로 정의되므로, 본 연구에서는 연속 행동 공간에서 결정론적 정책을 학습하는 TD3(Twin Delayed Deep Deterministic Policy Gradient)를 적용하였다.²⁵⁾ TD3는 두 개의 Critic을 이용하여 가치함수의 과대추정 영향을 완화하고, Delayed policy update와 Target policy smoothing을 적용하여 안정적인 정책 업데이트를 유도한다.

본 연구에서는 강화학습 정책이 기본 제어 입력에 과도한 보정을 추가하지 않도록 Actor 출력에 Tanh 활성화 함수를 적용하여 잔차 제어 입력을 식 (12)와 같이 정의하였다.

$$u_{rl}(t) = u_{rl,max} \tanh(f_{\theta}(e(t))) \quad (12)$$

여기서 f_{θ} 는 학습된 Actor 네트워크의 출력이며, $u_{rl,max}$ 는 잔차 제어 입력의 최대 크기이다. Tanh 함수의 출력 범위에 따라 잔차 제어 입력은 $|u_{rl}(t)| \leq u_{rl,max}$ 로 제한되며, 이를 통해 잔차 정책이 LQR 기본 제어 입력에 추가할 수 있는 보정 범위를 제한하였다.

3. 실험

3.1 실험 환경

제안하는 모델-참조 기반 강화학습 제어기의 종방향 속도 추종 성능을 검증하기 위해 IPG TruckMaker와 MATLAB/Simulink 기반의 시뮬레이션 환경을 구성하였다. 실험 시나리오는 추가 적재 하중과 맞바람 조건에 따라 Table 1과 같이 4가지로 구성하였다.

Scenario 1은 추가 적재하중과 맞바람이 없는 조건이며, Scenario 2는 15,000 kg의 추가 적재하중, Scenario 3은 15,000 kg의 추가 적재하중과 맞바람, Scenario 4는 20,000 kg

Table 1 Scenario conditions

Scenario Number	Mass Loading	Sudden Headwinds [-30 -20 m/s]
Scenario 1	X	X
Scenario 2	15,000 kg	X
Scenario 3	15,000 kg	O
Scenario 4	20,000 kg	O

의 추가 적재 하중과 맞바람을 적용하였다. 이를 통해 모델 변화와 외란 조건에 따른 종방향 속도 추종 성능을 단계적으로 비교하였다.

강화학습 기반 잔차 정책은 Scenario 1에서만 학습하였다. 이 조건에서 발생하는 상태 오차는 공칭 선형 모델과 시뮬레이션에 사용된 비선형 차량 모델 간의 모델 불일치에 의해 발생한다. 이후 학습된 정책을 적재하중 변화와 맞바람이 추가된 Scenario 2~4에 적용하여, 학습에 직접 포함되지 않은 모델 변화 및 외란 조건에서의 보정 성능을 평가하였다.

도로 구배는 Fig. 3과 같이 주행 거리에 따라 도로 높이가 점진적으로 변화하도록 구성하였으며, 목표 속도는 Fig. 4와 같이 가속, 정속 및 감속 구간을 포함하도록 설정하였다. 맞바람은 가속 구간을 중심으로 -20, -25, -30 m/s로 적용하여 공기 저항 변화에 따른 종방향 응답을 평가하였다.

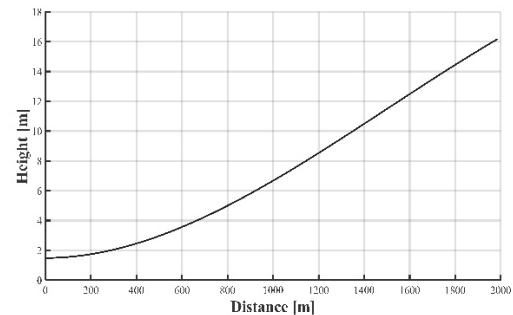


Fig. 3 Slope road environment

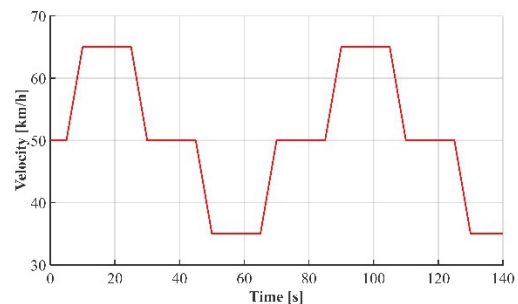


Fig. 4 Reference speed

3.2 평가 지표

제안 제어기의 종방향 속도 추종 성능을 정량적으로 평가하기 위해 RMSE(Root Mean Square Error), Peak Jerk, ETP(Error Tolerance Percentage)를 지표로 사용하였다.

RMSE는 실제 차량과 참조 모델 간의 속도 오차 크기를 나타내며, 값이 작을수록 공칭 페루프 참조 응답에 대한 추종 오차가 작음을 의미한다.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_{p,i} - v_{m,i})^2} \quad (13)$$

여기서 N 은 전체 데이터 개수, $v_{p,i}$ 와 $v_{m,i}$ 는 i 번째 시점의 실제 차량 속도와 참조 모델 속도이다.

Peak Jerk는 차량 가속도의 시간 변화율로, 제어 입력 및 가속 응답의 급격한 변화를 평가하는 지표이다. 값이 작을수록 종방향 가속도 변화가 작고, 보다 부드러운 종방향 응답을 의미한다.

$$\text{Peak Jerk} = \max_k \left| \frac{a[k] - a[k-1]}{\Delta t} \right| \quad (14)$$

여기서 $a[k]$ 는 k 번째 샘플의 종방향 가속도, Δt 는 샘플링 시간 간격이다.

ETP는 속도 추종 오차가 허용 범위 내에 포함되는 시간의 비율로, 값이 높을수록 목표 속도에 대한 오차가 허용 범위 내에서 유지된 시간이 길다는 것을 의미한다.

$$\text{ETP} = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(|v_{\text{ref},k} - v_{\text{ego},k}| \leq \epsilon) \times 100 \quad (15)$$

여기서 $v_{\text{ref},k}$ 와 $v_{\text{ego},k}$ 는 각각 k 번째 샘플의 목표 속도와 실제 차량 속도, $\mathbb{I}(\cdot)$ 는 조건을 만족하면 1, 아니면 0을 갖는 지시 함수이며, 허용 오차 범위는 $\epsilon = 0.3 \text{ m/s}$ 로 설정하였다.

RMSE는 평균오차 크기를, ETP는 허용 범위 내 유지 시간 비율을 나타내므로, 두 지표를 함께 사용하여 추종 정확도와 오차 유지 성능을 동시에 평가하였다. 또한 Peak Jerk를 통해 속도 추종 성능 향상 과정에서 최대 가속도 변화율을 함께 평가하였다.

3.3 실험 결과

3.3.1 참조 모델 유무에 따른 학습 기준 비교

LQR 페루프 참조 모델의 도입에 따른 학습 기준의 차이와 제어 성능에 미치는 영향을 분석하기 위해, 외란이 없고 모델 불일치 조건만 존재하는 Scenario 1에서 기존 Residual Reinforcement Learning(RRL)과 제안하는 MRRLC를 비교하였다. RRL은 목표 속도 $r(t)$ 와 실제 차량 속도

$v_p(t)$ 간의 오차를 잔차 정책의 학습 기준으로 사용하는 반면, MRRLC는 LQR 페루프 참조 모델의 속도 $v_m(t)$ 와 실제 차량 속도 $v_p(t)$ 간의 오차를 학습 기준으로 사용한다. 이를 통해 목표 명령을 직접 사용하는 경우와 공칭 차량 동역학 및 LQR 페루프 특성이 반영된 참조 응답을 사용하는 경우의 차이를 비교하였다.

Fig. 5(a)는 Scenario 1에서 목표 속도 $r(t)$, LQR 페루프 참조 모델의 속도 $v_m(t)$, 실제 차량 속도 $v_p(t)$ 의 응답을 나타낸다. 목표 속도가 변화하는 구간에서 참조 모델은 공칭 차량 동역학과 LQR 페루프 특성에 따라 목표 속도로 점진적으로 수렴하는 과도응답을 형성한다. 따라서 실제 차량이 기본 제어기가 의도한 공칭 응답을 추종하고 있더라도 목표 속도 $r(t)$ 와 직접 비교할 경우, 목표 속도에 도달하는 과정에서 발생하는 속도 차이가 잔차 정책의 학습 오차에 포함될 수 있다. 반면 참조 모델의 응답 $v_m(t)$ 을 학습 기준으로 사용하면 이러한 공칭 페루프 과도응답을 기준에 반영할 수 있으며, 실제 차량이 해당 응답으로부터 추가적으로 벗어난 편차를 구분할 수 있다. 이러한 학습 기준의 차이는 Fig. 5 (b)에 나타나며, RRL은 $v_p(t) - r(t)$ 를 학습 오차로 사용하므로 목표 속도 변화 구간에서 공칭 과도응답 성분이 함께 포함되는 반면, MRRLC는 $v_p(t) - v_m(t)$ 를 사용하여 공칭 페루프 응답으로부터 추가적으로 발생한 편차를 학습 대상으로 사용한다.

이러한 학습 기준의 차이가 제어 성능에 미치는 영향을 확인하기 위해 RRL과 MRRLC에 대해 서로 다른 random seed를 적용하여 각각 10회 반복 학습을 수행하였다. Table 2에서 MRRLC는 RRL 대비 평균 RMSE가 약 47.0 %, Peak Jerk가 약 25.3 % 감소하였으며, ETP 또한 증가하였다. 이는 목표 속도를 직접 학습 기준으로 사용하

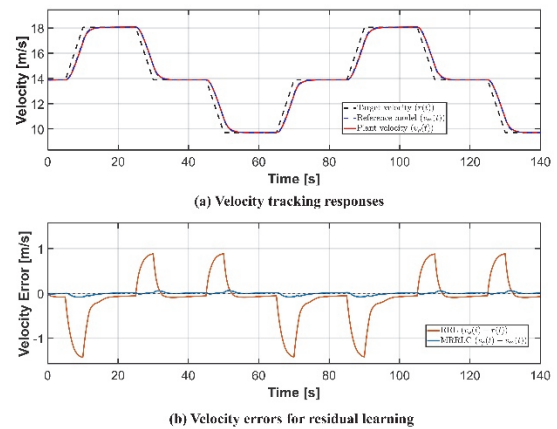


Fig. 5 Comparison of residual learning with/without reference model: (a) velocity responses; (b) velocity errors for residual learning

Table 2 Comparison of RRL and MRRLC over 10 training runs

Metric	LQR + RRL	LQR + MRRLC (Proposed)
Velocity RMSE [m/s]	0.1468 ± 0.0143	0.0778 ± 0.0484
Peak Jerk [m/s ³]	2.7894 ± 1.1933	2.0834 ± 0.0735
ETP [%]	99.2040 ± 1.4732	99.7300 ± 0.4739

는 RRL과 달리, MRRLC가 LQR 페루프 참조 응답으로부터 추가적으로 발생한 상태 편차를 잔차 학습 대상으로 구성함에 따라 모델 불일치에 대한 보정에 집중한 결과로 해석할 수 있다. 또한 Peak Jerk의 감소를 통해 추종 오차가 개선되는 과정에서 최대 종방향 가속도 변화율도 함께 감소하였음을 확인하였다.

3.3.2 잔차 제어 입력의 보정 특성

본 절에서는 모델 불일치 및 외란 조건에서 MRRLC의 잔차 정책이 실제 차량과 LQR 페루프 참조 모델 간의 응답 편차에 따라 어떠한 보정 입력을 생성하는지 분석하였다. 이를 위해 참조 모델 속도 오차 $v_p(t) - v_m(t)$ 와 잔차 제어 입력 $u_{r1}(t)$ 의 관계를 확인하고, 평균 절대 잔차 입력 Mean $|u_{r1}|$ 과 RMS u_{r1} 을 이용하여 보정 입력의 크기를 정량적으로 비교하였다.

Fig. 6은 모델 불일치와 외란의 영향이 가장 크게 적용된 Scenario 4에서 참조 모델 속도 오차 $v_{x,p} - v_{x,m}$ 와 잔차 제어 입력 $u_{r1}(t)$ 의 시간 응답을 나타낸다. 대표적으로 실제 차량 속도가 참조 모델 속도보다 낮아 오차가 0 미만인 구간에서 양의 잔차 제어 입력이 생성되며, 이는 LQR 기본 제어 입력에 추가되어 부족한 종방향 응답을 보정하는 방향으로 작용한다. 이를 통해 잔차 정책이 고정된 보정 입력을 생성하는 것이 아니라, 실제 차량이 공칭 페루프 참조 응답에서 벗어난 상태에 따라 추가적인 보정 입력을 생성함을 확인할 수 있다.

Table 3은 Scenario 1~4에서 참조 모델 속도 오차와 잔차 제어 입력의 크기를 정량적으로 비교한 결과이다. 여기서 Mean $|u_{r1}|$ 은 전체 주행 구간에서 발생한 잔차 제어 입력의 평균적인 크기를 나타내며, RMS u_{r1} 은 제곱 연산을 통해 상대적으로 큰 잔차 입력의 영향을 더 크게 반영한다. Scenario 1에서 Scenario 4로 조건이 가혹해짐에 따라 참조 모델 속도 오차의 RMSE는 약 26.3배 증가하였으며, 평균 잔차 입력의 크기 Mean $|u_{r1}|$ 과 RMS u_{r1} 은 각각 약 95.8%와 47.4% 증가하였다. 이러한 결과는 공칭 페루프 참조 응답과 실제 차량 간의 편차가 큰 조건에서 평균적인 잔차 제어 입력의 크기와 상대적으로 큰 입력의 영

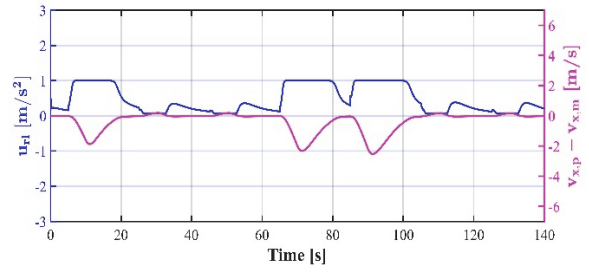


Fig. 6 Residual control input and reference-model velocity error in Scenario 4

Table 3 Analysis of reference-model error and residual input

Scenario	Ref.-Model RMSE [m/s]	Mean $ u_{r1} $ [m/s ²]	RMS u_{r1} [m/s ²]
Scenario 1	0.029547	0.23859	0.40723
Scenario 2	0.45526	0.39410	0.52908
Scenario 3	0.61356	0.42539	0.56262
Scenario 4	0.77786	0.46708	0.60008

향을 반영하는 RMS 값이 함께 증가하는 경향을 보여준다.

3.3.3 모델 불일치 및 외란 조건에 따른 추종 성능

모델 불일치 및 외란 조건에서 제안한 MRRLC의 종방향 추종 성능을 평가하기 위해, LQR 기본 제어기, LQR에 모델 참조 적응 제어(Model Reference Adaptive Control, MRAC)를 결합한 LQR+MRAC, 그리고 제안하는 LQR+MRRLC의 성능을 Scenario 1~4에서 비교하였다. 이를 통해 공칭 모델 기반 기본 제어, 적응 제어 및 참조 모델 기반 잔차 강화학습 제어의 추종 성능 차이를 분석하였다.

Fig. 7과 Fig. 8은 각각 Scenario 1과 Scenario 4에서 3개의 제어기의 종방향 응답을 비교한 결과이다. Scenario 1에서는 모델 불일치만 존재하므로 LQR 기본 제어기와 적응 보정을 적용한 LQR+MRAC 모두 목표 속도를 추종하지만, MRRLC는 공칭 페루프 참조 응답과 실제 차량 간의 편차를 잔차 보정 대상으로 사용함으로써 목표 속도 변화 구간의 추종 오차를 더욱 감소시켰다. 이러한 차이는 적재하중과 맞바람의 영향이 함께 증가한 Scenario 4에서 더욱 명확하게 나타난다. LQR은 공칭 모델과 실제 차량 응답 간의 차이가 커지면서 추종 오차가 증가하고, LQR+MRAC는 적응 보정을 통해 일부 오차를 감소시키지만, MRRLC는 참조 모델 기반 상태 편차에 따라 추가적인 잔차 입력을 생성하여 상대적으로 작은 추종 오차를 나타냈다. 또한 Scenario 4에서 MRRLC는 LQR+MRAC보다 낮은 Peak Jerk를 나타내어, 외란 조건에서 추종 오차 감소와 함께 종방향 응답의 급격한 변화도 완화되었다.

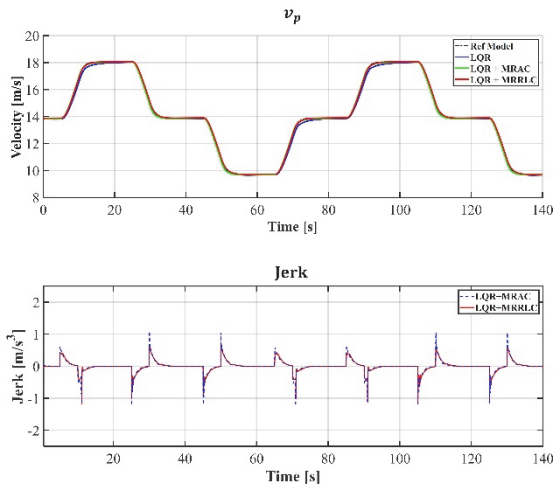


Fig. 7 Simulation results for Scenario 1

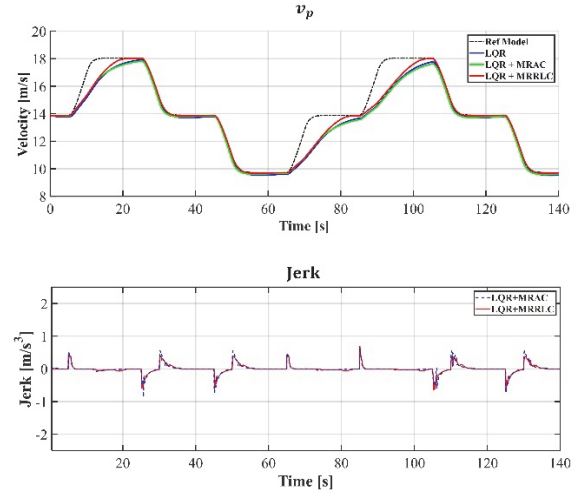


Fig. 8 Simulation results for Scenario 4

Table 4 Quantitative results across all scenarios (best results in bold)

Scenario	Controller	Velocity RMSE [m/s]	Peak Jerk [m/s ³]	ETP [%]
Scenario 1	LQR	0.130132	-	93.85
	LQR + MRAC	0.084641	2.115164	100.00
	LQR + MRRLC (Proposed)	0.029547 (0.0778 ± 0.0484)	2.115136 (2.0834 ± 0.0735)	100.00 (99.7300 ± 0.4739)
Scenario 2	LQR	0.544409	-	72.46
	LQR + MRAC	0.526330	0.818421	69.46
	LQR + MRRLC (Proposed)	0.455261 (0.5074 ± 0.0503)	0.760675 (0.7968 ± 0.0459)	80.32 (78.3699 ± 2.6727)
Scenario 3	LQR	0.711727	-	67.55
	LQR + MRAC	0.704777	0.848446	63.75
	LQR + MRRLC (Proposed)	0.613563 (0.6638 ± 0.0492)	0.828493 (0.8091 ± 0.0276)	76.60 (72.0622 ± 3.4273)
Scenario 4	LQR	0.882182	-	59.77
	LQR + MRAC	0.875711	0.858829	63.09
	LQR + MRRLC (Proposed)	0.777856 (0.8319 ± 0.0526)	0.691260 (0.7148 ± 0.0278)	72.36 (68.2079 ± 3.7726)

Table 4는 Scenario 1~4에서 세 제어기의 종방향 성능을 정량적으로 비교한 결과이다. 제안한 MRRLC는 서로 다른 Random seed를 적용하여 10회 반복 학습하였으며, 그 결과를 평균과 표준편차로 함께 제시하였다. 추가 적재 하중과 맞바람이 적용된 Scenario 2~4에서 LQR은 공칭 모델을 기반으로 고정된 상태 피드백을 사용하므로, 조건이 가혹해질수록 추종 오차가 증가하는 경향을 나타냈다. LQR+MRAC는 참조 모델과 실제 차량 간의 오차를 기반으로 적응 보정을 수행하여 LQR의 추종 오차를 일부

감소시켰으나, Scenario 2와 3에서는 RMSE 감소가 ETP 향상으로 이어지지 않았다. 반면 10회 반복학습의 평균값을 기준으로 MRRLC는 Scenario 2~4에서 LQR+MRAC 대비 RMSE를 각각 약 3.6 %, 5.8 %, 5.0 % 감소시키는 동시에 ETP를 각각 약 8.91 %, 8.31 %, 5.12 % 증가시켰다. 이는 MRRLC가 공칭 페루프 응답으로부터의 상태 편차를 잔차 정책의 입력으로 사용하고, 해당 편차에 따라 추가적인 보정 입력을 생성하도록 구성한 구조와 연관된 것으로 해석할 수 있다. 특히 앞 절에서 참조 모델 오차가

큰 조건일수록 잔차 입력의 평균 크기와 RMS가 증가하는 경향을 확인하였으므로, Scenario 2~4에서 나타난 성능 개선은 모델 변화와 외란으로 증가한 응답 편차에 따라 잔차 정책의 보정 입력이 조정된 결과로 해석할 수 있다. 또한 Scenario 4에서 MRRLC의 평균 Peak Jerk는 LQR+MRAC 대비 약 16.8 % 감소하여, 추종 성능 개선과 함께 최대 종방향 가속도 변화율도 감소하였음을 확인하였다.

3.3.4 강화학습 알고리즘 비교

동일한 MRRLC 구조에서 강화학습 알고리즘 선택에 따른 성능 차이를 확인하기 위해 DDPG, SAC 및 TD3를 비교하였다. 세 알고리즘은 동일한 제어 구조와 학습 조건에서 적용하였으며, Scenario 4에서 종방향 속도 추종 성능을 평가하였다.

Table 5 Comparison of RL algorithms for MRRLC in Scenario 4

Controller	Velocity RMSE [m/s]
LQR + MRAC	0.875711
LQR + MRRLC (DDPG)	0.862659
LQR + MRRLC (SAC)	0.870841
LQR + MRRLC (TD3)	0.777856

Table 5는 동일한 MRRLC 구조에서 DDPG, SAC, TD3를 적용한 결과를 비교한 것이다. TD3 기반 MRRLC는 DDPG 및 SAC 대비 Velocity RMSE를 각각 약 9.8 %와 10.7 % 감소시켜 가장 낮은 추종 오차를 나타냈다. 세 알고리즘은 동일한 참조 모델 기반 잔차 제어 구조를 사용하므로, 이러한 성능 차이는 잔차 정책의 학습 알고리즘에 따른 차이로 볼 수 있다. TD3는 Twin critic, Delayed policy update 및 Target policy smoothing을 적용하는 구조를 가지며, 본 실험 조건에서 가장 낮은 RMSE를 나타냄에 따라 MRRLC의 잔차 정책 학습 알고리즘으로 선정하였다.

4. 결론

본 논문에서는 대형 상용차량의 종방향 속도 추종 문제를 대상으로, LQR 기반 공칭 페루프 응답과 강화학습의 적응적 보정 성능을 결합한 모델-참조 기반 강화학습 제어(MRRLC) 방법론을 제안하였다. 제안 구조는 LQR 기반 기본 제어기가 공칭 조건에서의 기준 제어 입력과 페루프 응답을 형성하고, 강화학습 에이전트가 참조 모

델과 실제 차량 간의 상태 편차를 기반으로 추가적인 잔차 제어 입력을 생성하도록 구성하였다. 이를 통해 기본 LQR의 공칭 페루프 응답 특성을 기반으로 하면서, 모델 불일치 및 외란에 의해 추가적으로 발생하는 응답 편차를 보정하도록 하였다.

IPG TruckMaker 기반 시뮬레이션을 통해 참조 모델의 유무, 잔차 제어 입력의 보정 특성 및 모델 변화와 외란 조건에서의 추종 성능을 분석하였다. 참조 모델을 적용한 MRRLC는 목표 속도를 직접 학습 기준으로 사용하는 RRL에 비해 평균 추종 성능을 개선하였으며, 실제 차량과 공칭 페루프 참조 응답 간의 편차가 큰 조건에서 잔차 제어 입력의 크기도 증가하는 경향을 나타냈다. 이러한 결과는 MRRLC가 목표 속도 오차를 직접 보정하기 보다 공칭 페루프 응답으로부터 추가적으로 발생한 상태 편차를 잔차 보정 대상으로 구성하는 구조와 연관된 것으로 해석할 수 있다.

다만 본 연구의 검증은 시뮬레이션 환경에 한정되어 있으므로, 향후 연구에서는 보다 다양한 모델 변화 및 외란이 포함된 주행 시나리오로 검증 범위를 확대하고, 실차 기반 실험을 통해 제안 제어기의 실시간 적용 가능성과 실제 차량 환경에서의 제어 성능을 평가할 예정이다.

후 기

본 연구는 산업통상자원부 ‘자율주행개발혁신사업’의 지원을 받아 수행된 연구 결과입니다(20014476, 혼합 현실기반자율주행부품및시스템평가기술개발).

References

- 1) A. Alam, B. Besselink, V. Turri, J. Mårtensson and K. H. Johansson, “Heavy-Duty Vehicle Platooning for Sustainable Freight Transportation: A Cooperative Method to Enhance Safety and Efficiency,” IEEE Control Systems Magazine, Vol.35, No.6, pp.34-56, 2015.
- 2) World Economic Forum, Autonomous Trucks: An Opportunity to Make Road Freight Safer, Cleaner and More Efficient, White Paper, World Economic Forum, Geneva, Switzerland, 2021.
- 3) J. Yoo, Y. Ha, S. Moon, J. Kim and J. Yoo, “Real-Time Global Velocity Profile Calculation for Eco-Driving on Long-Distance Highways Using Variable-Step Spatial Segmentation,” Applied Sciences, Vol.15, No.19, Paper No.10811, 2025.
- 4) S. Tsugawa, S. Kato and K. Aoki, “An Automated Truck Platoon for Energy Saving,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), San Francisco, CA, USA, pp.4109-4114, 2011.

- 5) J. Lee, T. Oh and J. Yoo, "Adaptive Longitudinal Speed Control for Heavy-Duty Vehicles Considering Actuator Constraints and Disturbances Using Simulation Validation," *Applied Sciences*, Vol.15, No.13, Paper No.7327, 2025.
- 6) X. Y. Lu and J. K. Hedrick, "Heavy-Duty Vehicle Modelling and Longitudinal Control," *Vehicle System Dynamics*, Vol.43, No.9, pp.653-669, 2005.
- 7) R. Rajamani, *Vehicle Dynamics and Control*, 2nd Edn., Springer, New York, 2012.
- 8) T. D. Gillespie, *Fundamentals of Vehicle Dynamics*, Society of Automotive Engineers, Warrendale, 1992.
- 9) P. A. Ioannou and J. Sun, *Robust Adaptive Control*, Prentice Hall, Upper Saddle River, 1996.
- 10) A. Jo, H. Lee, D. Seo and K. Yi, "Model-Reference Adaptive Sliding Mode Control of Longitudinal Speed Tracking for Autonomous Vehicles," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, Vol.237, No.2-3, pp.493-515, 2023.
- 11) K. J. Åström and T. Hägglund, *PID Controllers: Theory, Design, and Tuning*, 2nd Edn., ISA, Research Triangle Park, 1995.
- 12) B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*, Prentice Hall, Englewood Cliffs, 1990.
- 13) M. A. S. Kamal, M. Mukai, J. Murata and T. Kawabe, "Model Predictive Control of Vehicles on Urban Roads for Improved Fuel Economy," *IEEE Transactions on Control Systems Technology*, Vol.21, No.3, pp.831-841, 2013.
- 14) T. Yuan and R. Zhao, "LQR-MPC-Based Trajectory-Tracking Controller of Autonomous Vehicle Subject to Coupling Effects and Driving State Uncertainties," *Sensors*, Vol.22, No.15, Paper No.5556, 2022.
- 15) J. Y. Lee, S. H. You, J. S. Jo, D. U. Sung and W. Cho, "A Study on Vehicle Parameter Estimation Using Tire Force Information," *Transactions of KSAE*, Vol.33, No.8, pp.617-626, 2025.
- 16) J. J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice Hall, Englewood Cliffs, 1991.
- 17) R. Sinha, J. Harrison, S. M. Richards and M. Pavone, "Adaptive Robust Model Predictive Control with Matched and Unmatched Uncertainty," *American Control Conference (ACC)*, Atlanta, GA, USA, pp.906-913, 2022.
- 18) J. Yao and Z. Ge, "Path-Tracking Control Strategy of Unmanned Vehicle Based on DDPG Algorithm," *Sensors*, Vol.22, No.20, Paper No.7881, 2022.
- 19) T. Kim, H. Lee, K. Kim and S. H. Hwang, "Path-Following Strategies for 4-Wheel Independent Steering EVs Using PPO Reinforcement Learning and Turning Radius Gain," *Transactions of KSAE*, Vol.31, No.8, pp.575-584, 2023.
- 20) J. Lee, C. Eom, D. Lee and M. Kwon, "Deep Reinforcement Learning-Based Autonomous Driving Strategy under On-Ramp Merge Scenario," *Transactions of KSAE*, Vol.32, No.7, pp.569-581, 2024.
- 21) R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd Edn., MIT Press, Cambridge, 2018.
- 22) J. García and F. Fernández, "A Comprehensive Survey on Safe Reinforcement Learning," *Journal of Machine Learning Research*, Vol.16, pp.1437-1480, 2015.
- 23) T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow and S. Levine, "Residual Reinforcement Learning for Robot Control," *IEEE International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, pp.6023-6029, 2019.
- 24) T. Silver, K. Allen, J. Tenenbaum and L. Kaelbling, "Residual Policy Learning," *arXiv Preprint arXiv:1812.06298*, 2018.
- 25) S. Fujimoto, H. Hoof and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," *International Conference on Machine Learning (ICML)*, Stockholm, Sweden, pp.1587-1596, 2018.