

# 강화 학습을 이용한 곡선 도로 외란 보상 기반 적응형 차선 유지 제어

이 정 현<sup>1)</sup> · 최 우 영<sup>\*2)</sup>

국립부경대학교 지능로봇공학과<sup>1)</sup> · 국립부경대학교 제어계측공학과<sup>2)</sup>

## Adaptive Lane-Keeping Control with Reinforcement Learning Compensation for Winding Road Disturbance

Jung Hyeon Lee<sup>1)</sup> · Woo Young Choi<sup>\*2)</sup>

<sup>1)</sup>Department of Intelligent Robot Engineering, Pukyong National University, Busan 48513, Korea

<sup>2)</sup>Department of Control and Instrumentation Engineering, Pukyong National University, Busan 48513, Korea

(Received 9 July 2025 / Revised 8 August 2025 / Accepted 12 August 2025)

**Abstract :** This paper proposed an adaptive lane-keeping controller that can help enhance a conventional Linear Quadratic Regulator (LQR) with a Reinforcement Learning (RL)-based compensator. While LQR performs well on straight road segments, its accuracy degrades on curved roads due to Winding Road Disturbance (WRD), which is absent in straight paths. Hence, to address this issue, the researchers integrated a Deep Deterministic Policy Gradient (DDPG) module that applies real-time steering corrections to compensate for WRD. The training process employs a two-stage Error-Characteristic Ornstein-Uhlenbeck (EC-OU) exploration scheme, in which noise covariance parameters are statistically retuned after the initial training stage to improve convergence in the second stage. The controller's performance was evaluated using a real vehicle in a simulation environment with road curvature data collected from the Gwangan Bridge in Busan. The experimental results demonstrate that the proposed method significantly reduces lateral position error and maintains closed-loop stability on curved roads, even under transient disturbances.

**Key words :** Reinforcement learning(강화 학습), Lane keeping system(차선 유지 시스템), Disturbance compensation(외란 보상), Lateral control(횡 방향 제어), Autonomous driving(자율주행)

### 1. Introduction

자율 주행 차량의 핵심 기술 중 하나인 자동 차선 유지 시스템은 차량이 도로 차선을 따라 안정적이고 차선을 유지하며 주행할 수 있도록 돕는 시스템이다. 이 시스템은 곡선 구간이나 다양한 주행 조건에서도 차량이 최적의 경로를 정확하게 유지하도록 제어가 되는 것을 목표로 한다. 일반적인 자동 차선 유지 시스템에서는 Proportional-Integral-Derivative(PID), Linear Quadratic Regulator(LQR)와 같은 제어 기법이 주로 사용되었으며, 이러한 기법은 일반적인 환경에서는 우수한 성능을 보였다.<sup>1,2)</sup>

그러나 곡선 도로에서는 도로 곡률에 의한 외부 외란으로 인해 일반적인 차량 횡 방향 제어를 통하여 차선을

정확하게 유지하는 것이 어렵다.<sup>1)</sup> 또한 차량이 사용하는 카메라 기반 비전 센서의 오작동으로 인하여 순간적인 외란이나 노이즈와 같은 문제가 발생하면, 차량의 횡 방향 제어 정확도가 떨어지고 차선 유지 시스템 성능이 급격히 저하되는 문제가 있다.<sup>2)</sup>

이러한 문제를 해결하기 위해 다양한 보상 기법들이 제안되어 왔다. 대표적인 방법으로는 곡선도로나 구불구불한 도로 곡률에 의해 발생하는 외부 외란인 Winding Road Disturbance(WRD)를 도로 곡률 외란으로 모델링하고, 이를 보상하기 위한 WRD 보상을 설계하여 곡선 구간에서의 제어 성능 저하를 효과적으로 완화하였다.<sup>1,3)</sup> 또한, 카메라의 낮은 샘플링 주기를 보완하기 위해 멀티

\*Corresponding author, E-mail: wychoi@pknu.ac.kr

<sup>1)</sup>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.

레이트 칼만 필터를 구동하여 차량 상태를 실시간으로 추정하고, 카메라가 차선을 순간적으로 놓치는 구간에는 예측 가상 차선을 생성함으로써 센서 지연이나 데이터 유실 환경에서도 안정적인 차선 유지 시스템을 실현할 수 있다.<sup>2,4)</sup> 추가적으로, 카메라와 레이더, 그리고 차량 자체 센서를 베이지안 방식으로 융합하고 정밀 클로소이드 도로 모델을 적용하여, 일부 센서 정보가 순간적으로 손실되더라도 고속도로에서 전방의 도로 기하를 안정적으로 추정할 수 있는 알고리즘을 제안하였다.<sup>5)</sup> 이외에도, 도로 곡률과 차량 속도에 따라 전방 예측 거리를 실시간으로 조정함으로써 기존의 고정 예측 거리 기반 제어에서 발생하는 추종 오차를 줄이고, 외란에 대한 민감도를 저감하는 방법도 제시되었다.<sup>6)</sup>

그러나 이러한 기법들은 복잡한 외란과 환경 변화에 대한 적응에 여전히 한계가 있으며, 이에 따라 강화 학습 기반 제어 전략이 대안으로 제시되고 있다. 예를 들어, 외란의 크기나 특성이 사전에 정의되지 않은 상황에서도 강인한 제어 성능을 확보하기 위해 외란 보상 구조를 포함한 강화 학습 기반 제어 방법이 제안되었으며, 이를 통해 기존의 모델 기반 보상기보다 더 유연한 대응이 가능해졌다.<sup>7)</sup> 또한, 복잡한 외란 조건 하에서의 안정적인 제어를 위해 강화 학습 기반 외란 제거 제어기가 제안되었으며,<sup>8)</sup> 외란이 존재하는 불확실한 환경에서도 최적의 성능을 유지할 수 있도록 하는 강화 학습 기법이 개발되고 있다.<sup>9)</sup> 또한 환경 및 모델의 불확실성을 반영하여 예측의 신뢰도를 기반으로 제어 정책을 학습하는 불확실성 모델 기반 강화 학습 기법도 제안되었다.<sup>10)</sup> 이러한 연구들은 복잡한 도로 환경 조건 하에서도 정밀한 제어 전략의 필요성을 시사한다.<sup>11)</sup>

본 논문에서는 곡선 도로에서의 곡률 변화에 의해 발생하는 WRD와 순간적인 외란에도 안정적으로 정확한 차선 유지 시스템 성능을 확보할 수 있는 Error-Characteristic based Ornstein-Uhlenbeck Noise를 고려한 Deep Deterministic Policy Gradient (DDPG) 기법을 적용한 보상 제어 기법을 제안한다. 제안된 방법은 LQR 제어기의 기본 추종 성능을 유지하면서 외란 항에 대한 보상 입력을 강화 학습 기반 보상기를 통해 보완한다. WRD와 센서 오류가 동시에 작용하는 상황에서도 안정적인 주행 성능을 확보할 수 있음을 시뮬레이션을 통해 검증하였다.

## 2. Road and Vehicle Model

### 2.1 Road Model

도로의 곡선은 차량의 주행 경로를 수학적으로 모델링하기 위해 3차 다항식을 사용하여 표현된다. 이 다항

식은 도로의 위치, 기울기, 곡률 및 곡률 변화율 등의 기하학적 특성을 포함하며, 차량 좌표계 기준에서 도로의 중심선을 다음과 같이 모델링 할 수 있다:

$$f(x) = c_0 + c_1 \cdot x + c_2 \cdot x^2 + c_3 \cdot x^3 \quad (1)$$

여기서  $c_0$ 는 횡 방향 위치 오차,  $c_1$ 은 Yaw 각도 오차,  $c_2$ 는 도로의 곡률,  $c_3$ 는 도로 곡률 변화율을 나타낸다. 이 다항식의 도함수를 이용하면 도로 곡선의 곡률  $\kappa(x)$ 를 계산할 수 있다. 실제 도로 설계에서는 직선과 곡선 구간 사이의 연속성과 부드러운 전환을 위해 클로소이드 곡선이 주로 사용된다. 클로소이드 곡선의 핵심 특징은 곡률이 호의 길이  $s$ 에 대해 선형적으로 변화한다는 점이다. 곡률의 변화가 완만한 경우, 종 방향 거리  $x$ 는 곡선 구간을 따라 이동한 호의 길이  $s$ 로 근사 될 수 있으며 다음과 같이 표현된다:

$$\kappa(s) = 6c_3s + 2c_2 \quad (2)$$

식 (2)는 클로소이드 모델에서 곡률이 호의 길이에 따라 선형적으로 변함을 보여주며, 이를 통해 직선과 곡선 사이의 부드러운 연결이 가능해진다. 따라서 클로소이드 기반 곡선 모델은 차량의 주행 안정성을 확보하는 데 효과적인 기하학적 표현 방식으로 활용된다.<sup>3)</sup>

### 2.2 Empirical Kinematic Lateral Motion Model

Fig. 1은 전방 주시 거리  $L$ 을 고려한 차량의 횡 방향 운동 모델을 나타낸다.<sup>2)</sup> 전방 주시 거리  $L$ 은 차량의 현재 위치를 기준으로 향후 주행 경로에 대한 정보를 반영함으로써, 도로 곡률 변화에 따라 발생할 수 있는 횡 방향 오차와 Yaw 각도 오차를 계산하는 데 활용된다. 이 모델은 전방 주시 거리와 현재 위치에서의 오차 및 상태 정보를 바탕으로 다음과 같은 상태 벡터  $\mathbf{x}$ 를 정의한다:

$$\mathbf{x} = [e_{yL}, \dot{e}_y, e_\psi, \dot{\psi}]^T \quad (3)$$

여기서  $e_{yL}$ 은 전방 주시 거리  $L$ 에서 측정된 횡 방향 오차,  $\dot{e}_y$ 는 현재 위치에서의 횡 방향 오차 변화율,  $e_\psi$ 는 Yaw 각도 오차,  $\dot{\psi}$ 는 Yaw 변화율을 의미한다. 제어 입력은  $\mathbf{u} = \delta$ 로 정의되며, 여기서  $\delta$ 는 조향 각도를 의미한다. 차량의 주행 경로에 따라 변화하는 도로 곡률이 Yaw 변화율 및 Yaw 각도 오차에 영향을 주는 점을 반영하여 외란 모델링이 필요하다. 이를 위해 전방 주시 거리에서 바라본 도로 기하와 차량 현재 상태 사이의 불일치로부터

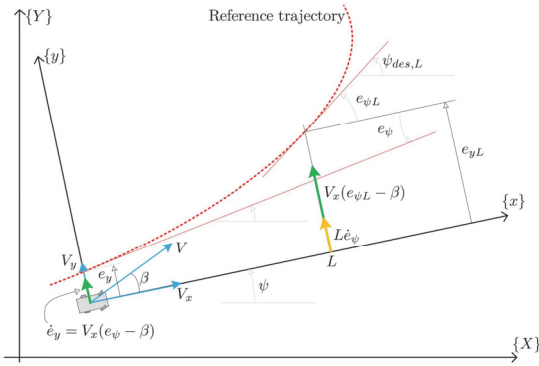


Fig. 1 Vehicle lateral motion with a look-ahead distance  $L$

발생하는 외란 WRD를 도로 곡률 계수  $c_2$ ,  $c_3$ 와 차량 모델의 동역학 모델에 따라 수학적으로 다음과 같이 정의한다:

$$\varphi = \begin{bmatrix} \dot{\psi}^d \\ e_{\psi L} - e_{\psi} \end{bmatrix} = \begin{bmatrix} 2V_x & 0 \\ 2L & 3L^2 \end{bmatrix} \begin{bmatrix} c_2 \\ c_3 \end{bmatrix} \quad (4)$$

여기서  $V_x$ 는 차량의 무게중심에서의 종 방향 속도이다.  $\dot{\psi}^d$ 는 목표 Yaw 각도 변화율이고,  $e_{\psi L}$ 는 전방 주시 거리  $L$ 에서의  $e_{\psi}$ 를 의미한다. 식 (4)에서 볼 수 있듯  $c_2$ ,  $c_3$ 와  $V_x$  그리고  $L$ 이 변화하면, WRD도 변화하여 제어 오차를 발생시키므로, 이를 보상하기 위한 제어 기법이 필요하다. 이러한 WRD를 모델에 통합해 곡률 변화가 주행 성능에 미치는 영향을 정밀히 반영할 수 있다. 이때 상태 벡터  $\mathbf{x}$ , 제어 입력  $\mathbf{u}$ , 외란  $\varphi$ , 그리고 전방 주시 거리  $L$ 을 고려한 차량 횡 방향 운동 모델 Empirical Kinematic Lateral Motion Model은 다음과 같은 이산 시스템 형태로 표현된다.<sup>6,12)</sup>

$$\mathbf{x}(k+1) = \Phi \mathbf{x}(k) + \Gamma \mathbf{u}(k) + \Gamma_{\varphi} \varphi(k) \quad (5)$$

여기서 각 행렬은 다음과 같다:

$$\Phi = \begin{bmatrix} 1 & T & 0 & -TL \\ 0 & -(1-\epsilon)^{-1}\epsilon & (1-\epsilon)^{-1}V_x & 0 \\ 0 & 0 & 1 & -T \\ 0 & 0 & 0 & 1 - \frac{T}{\tau} \end{bmatrix},$$

$$\Gamma = \begin{bmatrix} 0 \\ (1-\epsilon)^{-1} \frac{l_r}{l_{fr}} V_x \\ 0 \\ \frac{p}{\tau} \end{bmatrix}, \quad \Gamma_{\varphi} = \begin{bmatrix} TL & TV_x \\ 0 & 0 \\ T & 0 \\ 0 & 0 \end{bmatrix}$$

여기서  $T$ 는 제어 주기,  $l_f$ 와  $l_r$ 은 각각 앞, 뒷바퀴까지의 종 방향 거리,  $l_{fr}$ 은 휠 베이스,  $p = V_x/l_{fr}$ ,  $\tau$ 는 시간 상수,  $\epsilon$ 은 이완 계수로  $0 < \epsilon < 1$ 을 만족한다.

이와 같이 전방 주시 거리와 곡률 기반 외란 항을 통합한 횡 방향 운동 모델을 통해 실제 도로 주행 환경에서 발생하는 다양한 오차 요인을 보다 정밀하게 반영할 수 있다. 그러나 해당 모델에 선형 제어기인 LQR을 직접 적용하는 경우, 복잡한 도로 곡률이나 외란 조건에서 다음과 같은 구조적 한계가 존재한다. 다음 절에서는 이러한 한계를 중심으로 LQR 기반 Lane Keeping System (LKS) 제어기의 문제점을 살펴본다.

### 3. Problem statement for LKS

LKS는 도로의 곡률과 차량의 동적 상태를 기반으로 차량의 횡 방향 위치를 조정하여, 차선을 벗어나지 않도록 제어하는 기능을 수행한다. LKS 설계에서 일반적으로 채택되는 접근 방식은 LQR 기반의 제어 기법이다. LQR은 상태 오차와 제어 입력에 대한 가중합을 최소화하는 비용 함수에 따라 최적의 제어 이득 행렬  $\mathbf{K}$ 를 산출하며, 이에 따른 제어 입력은 다음과 같이 정의된다:

$$\mathbf{u}_{LQR}(k) = \mathbf{K}(\mathbf{x}^d(k) - \mathbf{x}(k)) \quad (6)$$

여기서  $\mathbf{x}^d(k)$ 는 주행 경로에서 차량의 목표 상태,  $\mathbf{x}(k)$ 는 차량의 현재 상태를 의미한다. LQR은 선형 시스템에서는 매우 효과적이지만, 곡선 도로와 같이 곡률이 변화하는 상황에서는 성능 저하가 발생할 수 있다.

곡선 구간의 일반적인 LQR 기반의 LKS는 곡률에 의해 발생하는 WRD에 대한 보상 능력이 제한적이라는 문제가 존재한다. WRD는 도로 곡률이 존재할 때 발생하는 외부 외란으로, 기존의 제어 입력 공간에 직접적으로 포함되지 않기 때문에 LQR 제어기만으로는 효과적인 상쇄가 어렵다. 이로 인해 횡 방향 위치 오차가 누적되고, 결과적으로 정확한 경로 추종 성능이 저해된다. 식 (4)와 같이 WRD는 입력 제어 공간 외부의 외란으로 작용하기 때문에 LQR 기반의 LKS는 한계가 존재한다. 이를 효과적으로 보상하기 위해서 강화 학습 기반 보상 제어기를 제안한다.

### 4. Adaptive Lane Keeping Control with RL Based WRD Compensation

#### 4.1 Reinforcement Learning Based WRD Compensator

앞서 언급한 LQR 기반 차선 유지 시스템의 구조적 한계를 극복하기 위해, Fig. 2와 같이 강화 학습 기반의 WRD

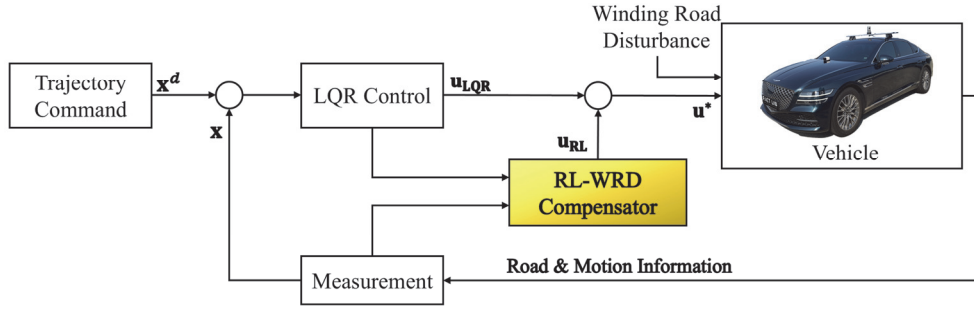


Fig. 2 LQR with Reinforcement Learning-Based WRD Compensator

보상기를 도입한 통합 제어 프레임워크를 제안한다. 해당 보상기는 WRD를 간접적으로 보상함으로써, 전체 시스템의 추적 오차를 줄이는 것을 목표로 한다. 제안된 최종 제어 입력은 기존의 LQR 제어 입력과 강화 학습 기반 보상 입력을 합산한 형태로 정의되며, 수식은 다음과 같다:

$$\mathbf{u}^*(k) = \mathbf{u}_{\text{LQR}}(k) + \mathbf{u}_{\text{RL}}(k) \quad (7)$$

여기서  $\mathbf{u}_{\text{RL}}(k)$ 은 외란 항  $\varphi$ 를 상쇄하기 위해 학습된 강화 학습 기반 보상기의 출력으로, 다음과 같은 관계를 만족하도록 설계된다:

$$\mathbf{u}_{\text{RL}}(k) \approx -\Gamma_{\varphi}\varphi(k) \quad (8)$$

이때  $\mathbf{u}_{\text{RL}}(k)$ 은 학습된 정책  $\mu$ 에 의해 결정된다. 정책 학습 시 에이전트는 차량의 추가 조향 각도를 생성한다. 정책은 관찰된 상태  $s(k)$ 를 기반으로 보상 신호를 최대화하는 최적 행동을 출력한다. 상태는 다음과 같이 정의된다:

$$s(k) = [e_y(k), e_{\psi}(k), e'_{\psi}(k), \mathbf{u}_{\text{LQR}}(k), c_2(k), c_3(k)] \quad (9)$$

이 상태 벡터는 식 (3)에서 정의된 오차항과 LQR의 제어 입력, 도로 곡률 계수  $c_2$ ,  $c_3$ 를 포함한다. 여기서  $c_2$ ,  $c_3$ 는 WRD를 유발하는 파라미터이므로, 이를 상태 벡터에 포함함으로써 강화 학습 정책이 외란의 추세를 사전에 인식할 수 있다. 정책  $\mu(s(k))$ 는 상태  $s(k)$ 를 입력으로 받아, 외란을 보상하는 제어 입력  $\mathbf{u}_{\text{RL}}(k)$ 을 생성한다. 이 과정은 보상 함수  $r(k)$ 를 최대화하는 방향으로 학습되며, 보상 함수는 시스템의 추적 성능을 정량적으로 평가하기 위해 다음과 같이 정의된다:

$$r(k) = -(w_0|e_y(k)| + w_1|e_{\psi}(k)| + w_2|e'_{\psi}(k)| + w_3|e'_{\psi}(k)|) \quad (10)$$

여기서  $w_0 \sim w_3$ 는 각 오차 항에 대한 가중치를 나타낸

다. 보상 함수는 주행 안정성과 정밀한 경로 추종 성능을 동시에 달성할 수 있도록 설계되며, 정책은 반복 학습을 통해 보상 값을 극대화하는 방향으로 수렴한다. 결과적으로, 강화 학습 기반 보상기는 외란의 영향을 줄이면서도 기존 제어기의 구조를 유지하는 유연하고 강건한 제어 성능을 제공한다.

#### 4.2 Policy Optimization Using DDPG Framework

강화 학습 기반 보상기의 학습은 연속 제어 환경에 적합한 DDPG 알고리즘을 기반으로 수행된다.<sup>13)</sup> DDPG는 Actor - Critic 구조를 기반으로 하며, 정책 함수와 가치 함수를 동시에 학습함으로써 연속 동작 공간에서의 최적 제어 입력을 추론할 수 있다. 앞서 정의한 바와 같이 보상기는 외란의 영향을 간접적으로 상쇄하는 것을 목표로 하며, 이를 위해 보상 함수는 외란을 간접적으로 반영한 추적 오차를 감소하도록 설계되었다. 이 보상 함수에 따라 강화 학습 정책  $\mu(s(k))$ 은 외란에 강인한 방향으로 최적화된다. DDPG에서는 정책  $\mu(s(k))$ 이 현재 상태와 Actor 네트워크의 가중치  $\theta^{\mu}$ 를 기반으로 결정론적으로 행동을 생성하며, 다음과 같이 정의된다:

$$a(k) = \mu(s(k)|\theta^{\mu}) + N(k) \quad (11)$$

여기서  $a(k)$ 는 학습 중 생성된 행동,  $\mu(s(k)|\theta^{\mu})$ 는 상태  $s(k)$ 에 대해 Actor 네트워크가 출력한 결정론적 행동이다.  $N(k)$ 는 탐색을 위한 노이즈이고, 자세한 설명은 4.3 절에서 설명한다. Critic 네트워크는 행동  $a(k)$ 에 대한 Q-value를 예측하며, Target Q-value는 다음과 같이 정의된다:

$$Q_{\text{target}}(k) = r(k) + \gamma Q(s(k+1), \mu(s(k+1), \theta^{\mu})|w') \quad (12)$$

여기서  $\gamma$ 는 할인 계수이고,  $\theta^{\mu}$ 와  $w'$ 는 각각 Target Actor 및 Target Critic 네트워크의 가중치이다. Critic 네트워크는

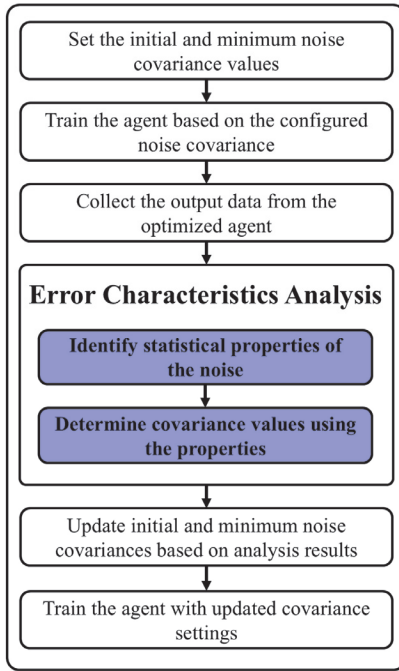


Fig. 3 Error-characteristic based noise covariance tuning process

예측된 Q-value와 Target Q-value의 차이를 최소화하는 방향으로 학습되며, 손실 함수는 다음과 같이 정의된다:

$$L(w) = \frac{1}{N} \sum_{k=1}^N (Q_{target}(k) - Q_w(s(k), a(k)|w))^2 \quad (13)$$

여기서  $N$ 은 미니 배치 크기이고,  $Q_w$ 는 현재 Critic 네트워크가 예측한 Q-value를 의미한다. Actor 네트워크는 Critic 네트워크의 Q-value를 최대화하는 방향으로 업데이트되며, 그 기울기는 다음과 같이 근사 된다:

$$\nabla_{\mu} \mathcal{J}(\mu) \approx \frac{1}{N} \sum_{k=1}^N \nabla_a Q(s(k), a(k)|w)|_{a(k)} \nabla_{\theta^{\mu}} \mu(s(k)|\theta^{\mu}). \quad (14)$$

학습의 안정성을 확보하기 위해, DDPG는 Target Network를 도입하며 지수 이동평균 방식으로 가중치를 다음과 같이 점진적으로 갱신한다:

$$\theta^Q \leftarrow \rho \theta^Q + (1 - \tau) \theta^{Q'}, \quad (15)$$

$$\theta^{\mu} \leftarrow \rho \theta^{\mu} + (1 - \tau) \theta^{\mu'}, \quad (16)$$

여기서  $\rho$ 는  $0 < \rho \leq 1$ 을 만족하는 업데이트 계수로

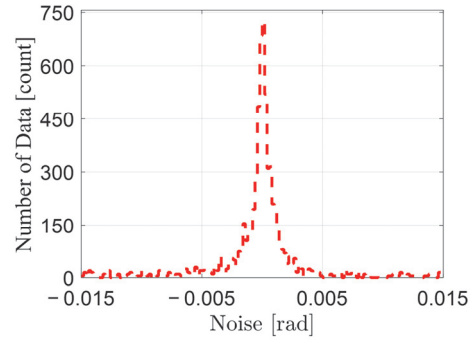


Fig. 4 Distribution of noise observed in the control outputs of the trained agent during evaluation

Target Network의 안정적인 수렴을 보장한다. 이와 같이 구성된 DDPG 프레임워크는 학습 시 탐색과 수렴의 균형을 유지하며, 강화 학습 기반 보상기가 외란 조건 하에서도 효과적으로 제어 입력을 생성할 수 있도록 한다. 식 (11)에서 사용되는 노이즈 항  $N(k)$ 는 학습 초기에는 충분한 탐색을 유도하고, 학습이 진행됨에 따라 그 분산을 점진적으로 줄임으로써 정밀한 정책 학습을 가능하게 한다. 이 노이즈 설계 방식은 다음 절에서 상세히 다룬다.

### 4.3 Error-Characteristic based Ornstein-Uhlenbeck

#### Noise

연속적인 제어 환경에서 강화 학습을 효과적으로 수행하기 위해 에피소드 단위로 노이즈 공분산을 점진적으로 감소시키는 Error-Characteristic based Ornstein-Uhlenbeck Noise를 제안한다. 해당 기법은 학습 초반에는 탐색을 강화하고, 학습이 진행됨에 따라 정밀한 정책 수렴을 유도하도록 설계되었다. 노이즈는 다음과 같은 업데이트 식을 따른다<sup>13)</sup>:

$$N_n(k+1) = N_n(k) + \alpha(\nu - N_n(k))\Delta k + \sigma_n \sqrt{\Delta k} \eta(k) \quad (17)$$

여기서  $N_n(k+1)$ 는  $n$ 번째 에피소드의  $k+1$ 번째 스텝에서의 노이즈 값이며,  $N_n(k)$ 는 해당 에피소드의  $k$ 번째 스텝에서의 노이즈를 의미한다.  $\nu$ 는 노이즈가 수렴하려는 목표 값이며,  $\alpha$ 는 노이즈가  $\nu$ 로 수렴하는 속도를 조절하는 계수이다.  $\eta(k)$ 는 평균 0, 분산 1의 가우시안 분포를 따르는 난수이며,  $\sigma_n$ 은  $n$ 번째 에피소드에서의 노이즈 공분산으로 다음과 같이 계산된다:

$$\sigma_n = \max(\sigma_{\min}, \sigma_0 \lambda^n) \quad (18)$$

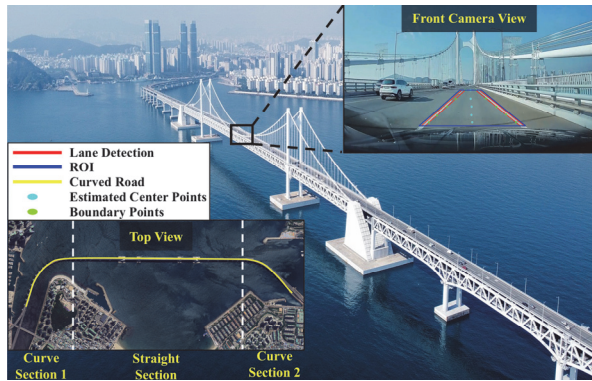


Fig. 5 Road curvature detection results and top view of the full curved path

여기서  $\lambda$ 는 공분산이 에피소드 진행에 따라 감소하는 비율을 의미한다. 이러한 방식은 학습 초기에 충분한 탐색을 보장하고, 이후 학습이 진행될수록 탐색을 줄이며 정밀한 정책 학습이 가능하도록 돕는다. 특히, 각 에피소드 내에서는 공분산이 일정하게 유지되므로 시간 스텝별 노이즈 특성이 안정적으로 유지될 수 있다.

Fig. 3은 이러한 공분산 업데이트 과정을 나타낸다. 초기에는 DDPG 프레임워크에서 일반적으로 사용되는 공분산 설정값을 기반으로 에이전트를 학습시켰으며,<sup>13)</sup> 1차 학습에서 수집된 에이전트의 행동 데이터를 분석하여 노이즈의 통계적 특성을 파악하였다. 이를 바탕으로 2차 학습부터는 데이터 기반의 공분산을 활용하여 초기 및 최소 노이즈 스케일을 재설정하고, 보다 안정적이고 효율적인 탐색이 가능하도록 구성하였다. Fig. 4는 MATLAB/Simulink 시뮬레이션 환경에서 평가된 강화 학습 기반 에이전트의 제어 출력 신호에 포함된 노이즈 분포를 보여준다. 이 그래프는 에이전트의 정책이 시간에 따라 생성하는 제어 명령의 변동 양상을 정량적으로 시각화한 것이며, 정규분포 형태의 노이즈 특성이 관찰되었다. 이러한 분포는 2차 학습에서 사용할 노이즈 공분산의 최솟값을 설정의 통계적 근거로 활용되었다.

노이즈 공분산의 점진적 감소와 데이터 기반 설계 방식은, 강화 학습 정책이 충분한 탐색과 수렴을 거쳐 안정적으로 수렴할 수 있도록 한다. 학습이 완료된 후 에이전트는 더 이상 노이즈를 포함하지 않은 결정론적 정책으로 제어 입력을 생성하며, 이는 다음과 같이 정의된다:

$$u_{RL}(k) \approx \mu(s(k)) \quad (19)$$

이 수식은 강화 학습 기반 보상이 실제 환경에서 생성하는 최종 정책 출력으로, 식 (8)을 만족한다.

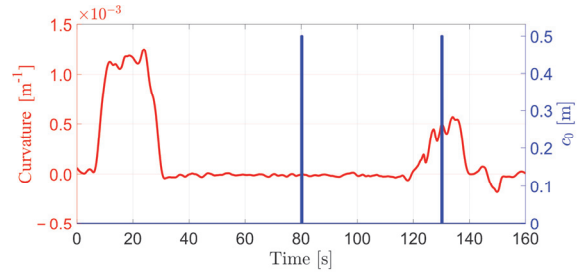


Fig. 6 Road curvature and transient disturbance profiles used in the simulation

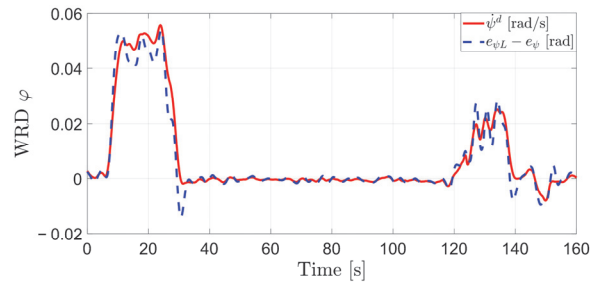


Fig. 7 Winding Road Disturbance (WRD) observed on the Gwangang Bridge

## 5. Application Results

### 5.1 Experimental Setup

자율주행 차량의 차선 유지 시스템 성능을 평가하기 위해 강화 학습 기반 제어 기법을 실제 도로 환경 데이터를 활용한 시뮬레이션에 적용하고, 외란이 작용하는 조건에서의 제어 안정성을 검증하였다. 이를 위해 부산 광안대교를 주행한 제네시스 G80 차량으로부터 내장 전방 카메라를 통해 주행 영상을 수집하였으며, 해당 데이터를 기반으로 MATLAB/Simulink 환경에서 차량의 횡 방향 동역학 시뮬레이션을 구축하였다. 차선 인식은 학습 기반 딥러닝 모델을 활용하였으며, YOLOv2를 적용하였다.<sup>14,15)</sup> 영상 내 차선 영역을 검출한 후, 좌우 차선으로부터 추출된 특징점을 기반으로 중앙 차선을 추정하였다. 이후 Bird's Eye View 투시 변환을 통해 중앙 차선 데이터를 평면화 하였으며, 최소 제곱법을 적용 하여 도로 곡률을 산출하였다. 추출된 곡률 값에는 Locally Weighted Regression(RLWR) 필터를 적용하여 고주파 노이즈를 제거하고,<sup>16)</sup> 곡률 변화율에는 저역 통과 필터를 추가로 적용하여 도로 형상의 연속성을 확보하였다.<sup>15)</sup>

Fig. 5는 부산 광안대교에서 수집된 주행 영상 데이터를 활용하여 수행된 차선 검출 및 도로 곡률 추정 과정을 시각적으로 나타낸다. 원본 전방 카메라 영상의 차선 인식 결과와 광안대교 구간 전체의 주행 경로를 Top View

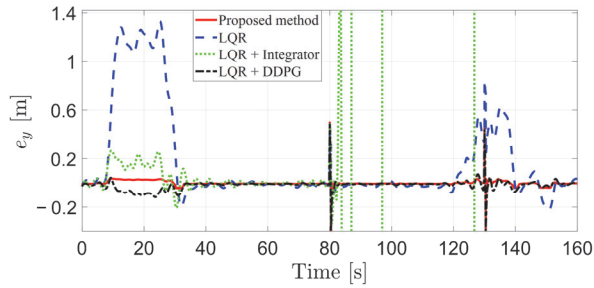


Fig. 8 Lateral position error

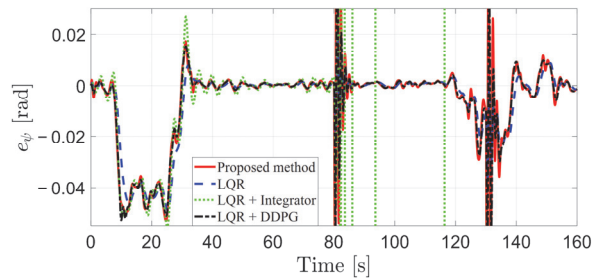


Fig. 9 Yaw angle error

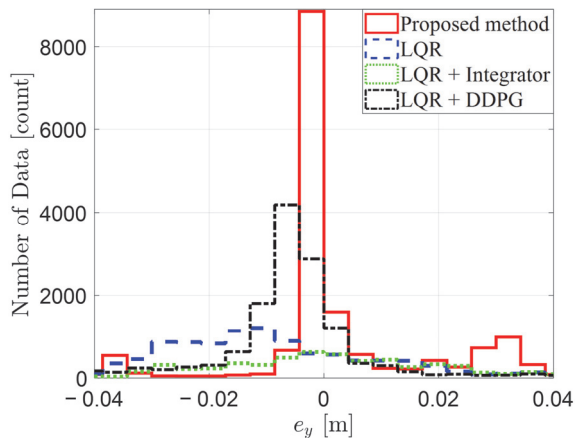


Fig. 10 Lateral position error distribution

로 실제 주행 궤적을 나타낸다. Fig. 6은 광안대교 구간에서 도출한 두 번의 우회전 구간과 140초 이후 작은 좌회전 구간의 도로 곡률 정보와 함께 일시적인 외란 구간을 시각화한 것이다. 외란 조건은 직선 구간인 80초 시점과 WRD가 발생 하는 곡선 구간인 130초에서 실제 모빌이 차량 전방 카메라에서 발생하는 0.1초 동안의 일시적인 외란 상황을 가정하여, 이 구간에 진폭 0.5의 노이즈를 삽입함으로써 구현하였다.<sup>1)</sup> Fig. 7은 추출된 광안대교 곡률에서 발생한 WRD를 보여준다.

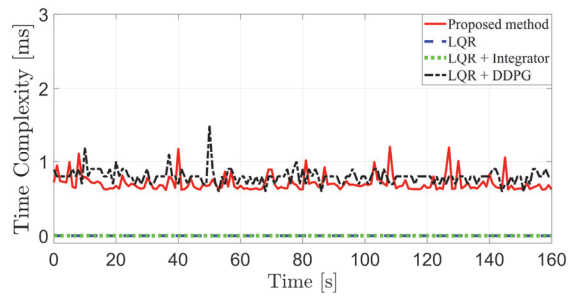


Fig. 11 Time complexity comparison of all controllers

### 5.2 Test Conditions

제안된 강화 학습 기반 제어 기법의 성능을 검증하기 위해, 기존에 사용되는 LQR, LQR+Integrator 제어 기법과 비교 분석을 진행하였으며, 이를 통해 횡 방향 위치 오차 변화를 비교하였다. 각 제어기의 상태 피드백 이득은 LQR을 통해 설계되었다. 다음과 같은 선형 이차 비용 함수를 최소화한다:

$$J = \sum_{k=0}^{\infty} (\mathbf{x}(k)^T \mathbf{Q}_{LQR} \mathbf{x}(k) + \mathbf{u}(k)^T \mathbf{R}_{LQR} \mathbf{u}(k)) \quad (20)$$

여기서  $\mathbf{Q}_{LQR}$ 과  $\mathbf{R}_{LQR}$ 은 각각 상태 벡터와 제어 입력 벡터의 가중치 행렬이다. 이 비용 함수는 Discrete Algebraic Riccati Equation(DARE)을 풀어 최적화한다. 모델의 파라미터의 경우  $T_s = 0.01 \text{ s}$ ,  $l_f = 1.016 \text{ m}$ ,  $l_r = 1.562 \text{ m}$ ,  $l_{fr} = 2.578 \text{ m}$ , 종 방향 속도  $V_x = 80 \text{ kph}$ , 전방 주시 거리  $L = 20 \text{ m}$ 로 설정되었다. 검증은 Fig. 6의 도로 데이터를 기반으로 시뮬레이션을 수행하였다.

### 5.3 Test result

Figs. 8, 10에서 제안된 강화 학습 기반 제어 기법, 기존 LQR, LQR+Integrator 그리고 일반적인 DDPG기반의 보상 제어 기법의 성능을 시뮬레이션 시나리오 전반에 걸쳐 횡 방향 위치 오차를 비교한 결과를 보여준다. 전체 시나리오에서 제안된 방법은 곡률이 변하는 구간에서 상대적으로 가장 안정적인 오차 범위를 유지하였으며, 일시적인 외란이 발생하는 80초 지점과 130초 지점 WRD와 일시적 외란이 동시에 발생하는 상황에서도 빠른 수렴 특성을 보였다. 반면, LQR+Integrator 방식은 곡선 구간에서 LQR보다 안정적인 오차 범위를 보인다. 하지만 외란 직후 일시적으로 오차를 줄이는 응답을 보였으나 추종 성능을 유지하지 못하였다. 이후 전체 시스템이 목표 궤적으로 복원을 효과적으로 수행하지 못하는 모습을 나타냈다. LQR의 경우 외란이 발생했을 때는 안정적인 수렴을 하였지만, 곡선 구간에서 비교 제어 기

법 중 가장 큰 진폭이 발생하는 모습을 보였다. 일반적인 DDPG 기반 보상기의 경우 곡선 구간에서 제안된 기법이 보다 더 큰 오차를 보였으며, 전체 시뮬레이션 시간동안 제안된 기법이 더 작은 오차를 유지했다. 이는 일반적인 DDPG가 학습과정에서 사용된 도로데이터 이외의 도로 곡률에 대해서는 최적의 결과를 보여주지 못한다는 것을 볼 수 있다. 이러한 결과는 제안된 기법이 일반적인 DDPG 기반 보상기보다 정밀한 수렴을 했기 때문이다. Fig. 9는 Yaw 각도 오차를 비교한 결과를 보여준다. 제안된 강화 학습 기반 제어 기법은 전체 시나리오 동안 상대적으로 작은 진폭과 빠른 감쇠 특성을 보이며, 외란이 발생하는 시점 이후에도 신속하게 오차를 복원하는 성능을 나타냈다. Fig. 11은 기존 제어 기법과 제안된 기법의 연산 실시간성 비교를 나타낸다. 제안된 기법의 처리 성능은 10ms 이하로, 실제 자율주행 제어 루프에 실시간으로 적용 가능한 수준임을 확인할 수 있다.<sup>17,18)</sup>

## 6. Conclusion

본 연구는 자율주행 차량의 차선 추적 성능 향상을 위해 기존 LQR 기반 제어에 EC-OU 기반의 DDPG 기법을 적용하였다. 시뮬레이션 결과, 제안된 방법은 곡선 구간과 일시적인 오류가 발생하는 상황에서도 최저 추종 오차와 높은 안정성을 보였다. 이는 제안된 기법이 일시적인 외란이 발생하는 상황에서 강인한 성능을 유지하기 때문이다. 결론적으로, 제안한 DDPG 기반 차선 추적 제어 시스템은 곡선 구간이나 일시적인 외란이 발생하는 상황에서도 기존 제어 방식보다 우수한 성능을 발휘하며, 자율주행 차량의 안정성과 주행 성능을 향상시킬 수 있음을 보여주었다.

## 후 기

이 논문은 국립부경대학교 자율창의학술연구비(2025년)에 의하여 연구되었음.

## References

- 1) W. Y. Choi, S. H. Lee and C. C. Chung, "Robust Vehicular Lane-Tracking Control with a Winding Road Disturbance Compensator," *IEEE Transactions on Industrial Informatics*, Vol.17, No.9, pp.6125–6133, 2020.
- 2) Y. S. Son, W. H. Kim, S. H. Lee and C. C. Chung, "Robust Multirate Control Scheme with Predictive Virtual Lanes for Lane Keeping System of Autonomous Highway Driving," *IEEE Transactions on Vehicular Technology*, Vol.64, No.8, pp.3378–3391, 2014.
- 3) W. Y. Choi, S. H. Lee and C. C. Chung, "Vehicular Lateral Tracking Control with Winding Road Disturbance Compensation," *IFAC PapersOnline*, Vol.53, No.2, pp.15699–15704, 2020.
- 4) B. A. Kim, Y. S. Son, S. H. Lee and C. C. Chung, "Development of a Multirate State Estimator Using Predictive Virtual Lane," *KSAE Spring Conference Proceedings*, pp.1003–1005, 2013.
- 5) L. Hammarstrand, M. Fatemi, A. F. García-Fernandez and L. Svensson, "Long-Range Road Geometry Estimation Using Moving Vehicles and Roadside Observations," *IEEE Transactions on Intelligent Transportation Systems*, Vol.17, No.8, pp.2144–2158, 2016.
- 6) W. Y. Choi and C. C. Chung, "Optimizing Look-Ahead Distance for Vehicle Lateral Control: A Controller-Adaptive Approach," *IEEE Transactions on Intelligent Vehicles*, 2024.
- 7) Y. Hu, J. Fu and G. Wen, "Safe Reinforcement Learning for Model-Reference Trajectory Tracking of Uncertain Autonomous Vehicles with Model-Based Acceleration," *IEEE Transactions on Intelligent Vehicles*, Vol.8, No.3, pp.2332–2346, 2023.
- 8) A. Coursey, A. Zhang, M. Quinones-Grueiro and G. Biswas, "Hybrid Control Framework of UAVs under Varying Wind and Payload Conditions," *2024 American Control Conference (ACC)*, pp.1543–1549, 2024.
- 9) O. Saleem, A. Tanveer and J. Iqbal, "Adaptive Reconfigurable Learning Algorithm for Robust Optimal Longitudinal Motion Control of Unmanned Aerial Vehicles," *Algorithms*, Vol.18, No.4, 2025.
- 10) J. Wu, Z. Huang and C. Lv, "Uncertainty-Aware Model-Based Reinforcement Learning: Methodology and Application in Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, Vol.8, No.1, pp.194–203, 2022.
- 11) J. H. Lee, Y. Lee, Y. S. Son and W. Y. Choi, "Vehicle Velocity Control System for Multi-Vehicle Collision Avoidance Based on Optimal Collision Point," *Transactions of KSAE*, Vol.32, No.11, pp.935–941, 2024.
- 12) W. Y. Choi, S. H. Lee and C. C. Chung, "Horizonwise Model-Predictive Control with Application to Autonomous Driving Vehicle," *IEEE Transactions on Industrial Informatics*, Vol.18, No.10, pp.6940–6949, 2021.
- 13) T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver and D. Wierstra, "Continuous Control with Deep Reinforcement Learning," *arXiv*

- preprint, arXiv:1509.02971, 2015.
- 14) C. Han, Q. Zhao, S. Zhang, Y. Chen, Z. Zhang and J. Yuan, "YOLOPv2: Better, Faster, Stronger for Panoptic Driving Perception," arXiv preprint, arXiv: 2208.11434, 2022.
  - 15) J. H. Yang, W. Y. Choi and C. C. Chung, "StrikeNet: Deep Convolutional LSTM-Based Road Lane Reconstruction With Spatiotemporal Inference for Lane Keeping Control," IEEE Access, 2024.
  - 16) Abdul Nurunnabi, Geoff West and David Belton, "Robust Locally Weighted Regression Techniques for Ground Surface Points Filtering in Mobile Laser Scanning Three Dimensional Point Cloud Data," IEEE Transactions on Geoscience and Remote Sensing, Vol.54, No.4, pp.2181–2193, 2015.
  - 17) D. H. Lee, "Lane-Keeping Control of Autonomous Vehicles through a Soft-Constrained Iterative LQR," IEEE Transactions on Vehicular Technology, Vol.74, No.4, pp.5610–5623, 2025.
  - 18) W. Y. Choi, D. J. Kim, C. M. Kang, S. H. Lee and C. C. Chung, "Autonomous Vehicle Lateral Maneuvering by Approximate Explicit Predictive Control," 2018 Annual American Control Conference (ACC), pp.4739–4744, 2018.