ISSN 1975-8359 [Print] / ISSN 2287-4364 [Online]
The Transactions of the Korean Institute of Electrical Engineers, vol. 74, no. 8, pp. 1363~1372, 2025
https://doi.org/10.5370/KIEE.2025.74.8.1363

Implementation of 56 Transition Control of a Triple Inverted Pendulum Using Sim-to-Real Reinforcement Learning

Sim-to-Real 강화학습 기법을 활용한 직선형 3단 도립진자의 56가지 천이 제어 구현

Changseok Lim . Doyoon Ju . Young Sam Lee
임창석* . 주도윤* . 이영삼†

Abstract

This paper proposes the implementation of equilibrium-to-equilibrium transition control for a triple inverted pendulum system using Sim-to-Real reinforcement learning. Recently, multi-link inverted pendulum systems have introduced the new control challenge, extending beyond conventional swing-up and balancing controls toward equilibrium-to-equilibrium transition control. Transition control, which involves continuous transitions between multiple unstable equilibrium points, is particularly sensitive to disturbances. To address this, we apply the Sim-to-Real reinforcement learning technique, transferring control policies learned in simulation to the physical system. Furthermore a triple inverted pendulum system with high model consistency was designed and constructed to minimize the reality gap between simulation and physical environments. The proposed controller successfully achieved all 56 possible transitions among the eight defined equilibrium points. The results demonstrate that transition control based on Sim-to-Real reinforcement learning effectively resolves complex nonlinear control problems.

Key Words

Triple inverted pendulum, Reinforcement learning, Sim-to-Real Learning, Transition control

1. 서 론

도립진자 시스템은 비최소 위상 특성과 비선형적인 모델 방정식을 가지며 불안정한 동특성을 지닌 대표적인 부족 구동 시스템이다. 이러한 특성으로 인해 도립진자는 새로운 제어 이론이나 알고리즘의 유효성을 검증하기에 적합한 실험 모델로 널리 사용되어 왔다. 특히 시스템의 불안정성과 비선형적 특성을 효과적으로 제어하기 위해 진자를 수직 자세로 도달시키는 swing-up 제어나 해당 상태를 유지하는 균형 제어를 중심으로 연구가 진행되었다[1, 2, 3]. 최근 제어기의 성능을 평가하기 위한 고난도의 제어 대상을 필요로 함에 따라 진자의 수를 증가시킨 다단 도립진자 시스템 연구가 활발히 진행되고 있으며, 그중 3단 도립진자를 활용한 제어기 설계와 성능 검증 또한 수행되고 있다[4, 5]. 다단 도립진자 시스템은 링크가 추가됨에 따라 시스템의 상태 변수가 증가하며 이는 제어 난도를 크게 높일 뿐 아니라 기존의 제어 전략으로는 다루기 어려운 새로운 제어 문제를 제시한다. 특히 다단 도립진자 시스템에서는 단순히 진자를 세우거나 균형 상태를 유지하는 문제를 넘어 복수의 균형점(Equilibrium Point) 간 천이를 요구하는 천이 제어(Transition Control) 문제가 주요한 제어 문제로 확장된다.

천이 제어는 다단 도립진자 시스템에서 swing-up 제어와 유사한 특성을 가지면서도 더욱 확장된 개념을 포함한다. 일반적으로 도립진자 시스템의 균형점은 각 진자의 상태에 따라 진자가 위쪽을 향한 불안정한 균형점과 아래쪽을 향한 안정한 균형점으로 나누어진다. 단일 진자 시스템에서는 불안정한 균형점이 하나뿐이지만 다단 구조에서는 진자의 개수가 증가할수록 다양한 조합의 균형점들이 존재하게 된다. 이러한 다수의 균형점 간을 이동하는 천이 제어는 swing-up 제어가 주로 안정한 균형점에서 불안정한 균형점으로의 이동만을 목표로 하는 것과 달리 여러 불안정한 균형점 간의 천이를 포함하므로 더욱 복잡한 제어 전략을 요구한다. 천이 제어는 현재 균형점에서의 균형 제어, 목표 균형점으로의 천이 제어, 목표 균형점에서의 균형 제어의 순서로 구성되며 각 단계는 연속적인 제어 동작을 통해 수행된다. 이에 따라 천이 제어의 성공적인 구현을 위해 각각의 제어가 유기적으로 작동하는 제어 전략이 필요하다.

최근 도립진자 천이 제어 연구에서는 Direct collocation 기법

† Corresponding Author : Dept. of Electrical and Computer Engineering, Inha University Incheon, Republic of Korea.
E-mail : lys@inha.ac.kr https://orcid.org/0009-0003-0665-1464
* Dept. of Electrical and Computer Engineering, Inha University, Incheon, Republic of Korea.
https://orcid.org/0009-0008-8533-9164 https://orcid.org/0000-0001-7011-6779
Received : Apr. 16, 2025 **Revised** : Jun. 25, 2025 **Accepted** : Jul. 03, 2025

과 같은 최적 제어 기반의 방법을 활용하여 천이 궤적을 설계하였다[6, 7]. 그러나 사전에 계산된 Optimal trajectory는 외란이나 모델 불확실성에 대한 민감성이 높아 실제 시스템에 적용 시 안정적인 제어 성능을 확보하기 어렵다는 한계를 지닌다. 특히 천이 제어는 시스템이 다수의 불안정한 균형점 사이를 이동해야 하는 특성상 외란에 대한 민감도가 더욱 크게 나타난다. 최적 제어 기반의 천이 제어는 설계된 궤적을 정확히 추종해야 하므로 일정 수준 이상의 외란이 작용할 경우 목표 균형점으로의 안정적인 수렴이 어려울 수 있다[8]. 이러한 문제점을 극복하고자 본 논문에서는 강화학습의 기법 중 하나인 Sim-to-Real 기법을 사용해 3단 도립진자의 천이 제어를 수행한다. Sim-to-real 기법은 시뮬레이션 환경에서 학습한 데이터를 실물 시스템에 적용하는 기법이다[9]. 해당 기법은 학습 환경에서의 물리적 제약이 없어 임의의 초기 상태 설정이 가능하므로 다양한 상태에서의 학습을 통해 외란에 강건한 제어 정책을 수립할 수 있다.

그러나 Sim-to-Real 기법은 시뮬레이션 모델과 실제 하드웨어 간의 차이로 인해 발생하는 reality gap 문제를 동반한다[10]. 이 격차를 해소하지 못할 경우 시뮬레이션에서 학습된 제어 정책이 실물 시스템에서 원하는 성능을 보장하지 못할 수 있다. 본 연구에서는 저자들이 소속된 연구실에서 제작한 3단 도립진자 시스템을 사용하여 해당 문제점을 해결한다. 해당 시스템은 시뮬레이션 환경에서 사용할 모델 방정식과 실물 시스템 간에 높은 정합성을 지녀 reality gap을 최소화한다. 이를 통해 Sim-to-Real 강화학습 기법을 활용한 직선형 3단 도립진자의 56가지 천이 제어 구현을 목표로 한다.

본 논문의 구성은 다음과 같다. 2절에서는 Sim-to-Real 기법 및 강화학습 알고리즘에 대해 설명한다. 3절에서는 본 연구에서 활용하는 3단 도립진자 시스템의 기구적 설계 및 수학적 모델에 대해 설명한다. 4절에서는 강화학습 기반 제어기를 설계하고 실제 환경에서의 제어 결과를 분석한다. 끝으로 5절에서는 본 연구의 결론을 서술한다.

2. Sim-to-Real 학습 기반 제어기 및 알고리즘

2.1 Sim-to-Real 학습 기반 제어기

강화학습 기반 제어기는 전통적인 제어 방식에서 제어 연산을 수행하는 구성 요소를 강화학습 에이전트로 대체한 구조로 정의할 수 있다. 이때 강화학습 에이전트란 환경과의 반복적인 상호작용을 통해 최적의 제어 정책을 학습하는 시스템을 의미한다. 에이전트는 매 timestep에서 환경으로부터 관측된 상태를 입력으로 받아 현재의 정책에 맞춰 행동을 선택하고 그에 대한 보상을 통해 피드백을 받는다. 이러한 과정이 반복되며 에이전트는 경험을 축적하고 이를 바탕으로 정책을 점진적으로 개선해 나간다. 학습된 제어기는 주어진 상태 정보를 입력받아 학습된 정책에 맞춰 제어량을 출력하게 된다.

학습 및 평가 과정에서 에이전트와 상호작용이 이루어지는 환경은 크게 두 가지로 분류된다. 첫째는 실물 시스템을 기반으로 하는 물리적 환경, 둘째는 가상 시뮬레이션 기반의 가상 환경이다.

실물 시스템을 기반으로 학습을 진행할 경우 시스템의 수학적 모델이나 정확한 동역학 정보가 사전에 확보되지 않더라도 학습이 가능하다는 장점이 있다. 이는 모델 기반 제어기 설계에서 필수적인 파라미터 식별 과정이나 복잡한 비선형 모델링 없이도 환경과의 상호작용을 통해 최적의 정책을 자율적으로 학습할 수 있음을 의미한다. 특히 실제 환경에서 수집되는 데이터는 센서 노이즈, 마찰, 백래시, 하드웨어의 비선형성, 외란 등 다양한 비이상적 요소(non-idealities)를 자연스럽게 포함하고 있다. 따라서 이와 같은 환경에서 학습된 정책은 시뮬레이션 기반 학습 결과와 비교했을 때 더 높은 현실 적합성과 강건성을 갖는다는 특징이 있다.

하지만 실물 시스템을 대상으로 하는 제어기 학습에서는 다양한 제약과 위험 요소 또한 존재한다. 실제 환경에서 도립진자 시스템의 학습을 진행할 경우 모든 진자가 중력의 영향을 받아 바닥을 향한 상태에서 시작되며 연구자가 원하는 각도 및 각속도로 초기 상태를 설정하는 것이 어렵다. 또한 학습 속도 역시 현실의 물리적 시간에 의해 제한된다. 이러한 이유로 최근에는 시뮬레이션 기반의 가상 환경에서 충분한 학습을 수행한 후 이를 실제 환경에 이식하는 Sim-to-Real 학습 기법이 활발히 활용되고 있다[11, 12].

시뮬레이션 환경에서의 학습은 앞서 설명한 물리적 환경에서의 제약을 극복하고 반복적인 실험을 보다 안전하고 자유롭게 수행할 수 있다는 점에서 학습 효율성을 크게 향상시킨다. 특히 강화학습과 같이 수많은 시행착오를 통해 정책을 개선하는 방식에서는 시스템의 손상 가능성 없이 학습을 반복할 수 있다는 점이 큰 이점으로 작용한다. 또한 시뮬레이션 환경에서는 초기 상태를 임의로 설정할 수 있으며 실시간 학습이 아닌 가속화된 시뮬레이션을 통해 보다 짧은 시간 내에 대량의 데이터를 수집할 수 있다. 이를 통해 학습 속도를 효과적으로 향상시킬 뿐만 아니라 실제 환경에서 구현이 어려운 다양한 초기 조건에서도 학습을 수행할 수 있어 외란이 존재하는 환경에서도 강인한 제어 정책을 형성할 수 있다.

하지만 앞서 서론에서 언급했듯이 Sim-to-Real 기법은 reality gap이라는 근본적인 한계점이 존재한다. 시뮬레이션 환경은 실물 시스템의 모든 물리적 특성을 완벽하게 모사할 수 없기 때문에 시뮬레이션에서 학습된 정책이 실제 환경에서 그대로 적용되지 않거나 예기치 못한 동작을 유발할 수 있다. 이에 본 연구는 reality gap을 완화하기 위한 다양한 시도 중에서 실효성이 높은 두 가지 기법을 채택하여 Sim-to-Real 제어 성능을 향상시키고자 하였다.

먼저 소프트웨어적인 방법으로 시뮬레이션 내에서 적용 가능한 DR(Domain Randomization) 기법을 활용한다[13, 14]. DR 기법은 시뮬레이션 환경의 초기 조건을 무작위로 선정하여 학습을 진행시키는 기법이다. 이를 통해 강화학습 에이전트가 다양한 조건에서 학습을 진행할 수 있고 더욱 일반화된 제어 정책을 수립할 수 있다.

또한 하드웨어적인 방법으로 본 연구실에서 직접 제작한 3단 도립진자 시스템을 사용하여 물리적 정합성이 높은 시뮬레이션 환경을 구축한다. 이를 통해 시뮬레이션과 실제 환경의 모델 간 차이로 인한 reality gap을 효과적으로 완화할 수 있으며 이러한 Sim-to-Real 기반 학습 전략은 3단 도립진자 시스템과 같이 초기 조건의 제약이 크고 높은 비선형성을 가지는 천이 제어 문제를 해결하는 데 있어 효과적으로 활용될 수 있다.

2.2 강화학습 알고리즘

본 연구에서는 천이 제어와 같이 불안정한 균형점 간의 천이를 요구하는 고차 비선형 시스템의 제어 문제를 다루기 위해 Truncated Quantile Critics(TQC) 알고리즘을 적용하여 강화학습 기반 제어기를 구현하였다. 일반적인 강화학습 알고리즘은 극단적인 보상 예측으로 인해 정책이 불안정해지거나 수렴 속도가 저하되는 문제가 존재하며, 특히 도립진자와 같은 고차 비선형 시스템에서는 이러한 현상이 더욱 빈번하게 발생한다.

이를 해결하기 위해 Quantile Regression Deep Q-Network (QR-DQN)와 Soft Actor-Critic(SAC)의 장점을 결합한 TQC는 최신 분포 기반 강화학습 알고리즘으로 연속적인 행동 공간을 대상으로 하는 고성능 정책 학습에 적합하다. TQC의 핵심 전략은 예측된 보상 분포 중 상위 분위수를 제거함으로써 Q-값의 과대 평가를 억제하고 정책이 보다 현실적인 기대 보상을 기반으로 수렴할 수 있도록 유도하는 것이다. 이 과정은 강화학습 초기에 자주 발생하는 과도한 탐색(optimistic exploration)을 억제하여 학습 안정성을 높이고 실제 환경에 적용 시 안전성 측면에서도 유리하다.

표 1 강화학습 에이전트 구현에 사용된 하이퍼파라미터

Table 1 Hyperparameters used to implement reinforcement learning agents

Hyperparameter	Value
Optimizer	ADAM
Learning rate	0.0003
Discount factor (γ)	0.99
Replay buffer size	1e6
Number of critics (N)	3
Number of hidden layers in critic networks	3
Size of hidden layers in critic networks	512
Number of hidden layers in policy networks	2
Size of hidden layers in 1st policy networks	400
Size of hidden layers in 2nd policy networks	300
Minibatch size	256
Nonlinearity	ReLU
Target smoothing coefficient (β)	0.005
Target update interval	1
Gradient steps per iteration	1
Environment steps per iteration	1
Number of atoms (M)	25

특히 본 연구에서 다루는 3단 도립진자 시스템은 상태 공간이 고차원이며 초기 조건의 미세한 변화만으로도 동작이 급격히 불안정해질 수 있는 특성을 가진다. 이처럼 보상의 분산이 크고 실패 가능성이 높은 제어 환경에서는 보상의 tail 정보까지 고려하는 분포 기반 접근 방식이 효과적이며 TQC는 이러한 환경에 특화된 정책을 학습하는 데 있어 기존 방법보다 강인한 수렴 특성을 보인다. 또한 복수의 critic network를 활용하여 다양한 보상 분포를 학습하고 이를 통합하는 구조는 외란이나 모델 불확실성이 존재하는 실제 환경에서 정책의 일반화 성능과 안정성을 동시에 확보할 수 있다는 점에서 본 연구의 목적과 높은 부합성을 가진다.

본 연구에서는 천이 제어 과정에서 요구되는 정밀한 균형점 간의 천이와 초기 조건에 대한 강건성을 확보하고자 시스템의 특성에 맞춰 네트워크 구조와 하이퍼파라미터를 조정하였다. 구체적으로는 학습 속도와 연산 효율을 고려하여 critic network의 개수를 줄이고, 도립진자의 고차 모델 방정식을 반영하여 policy network의 크기를 조정하였다. 사용한 주요 하이퍼파라미터는 표 1에 정리되어 있으며, replay buffer size 등의 나머지 설정은 Kuznetsov[15]가 제안한 하이퍼파라미터를 사용했다.

3. 3단 도립진자 시스템 및 천이 제어

그림 1은 3단 도립진자의 기구적 개념도를 나타낸다. 본 논문에서는 국제 단위계(SI 단위계)를 사용하며 각 변수의 의미는 다음과 같다. M은 cart의 질량, m_1, m_2, m_3는 각 진자들의 질량을 의미한다. l_1, l_2, l_3는 각 진자들의 회전축으로부터 무게중심까지의 길이를 의미하고 L_1은 1단 진자의 회전축과 2단 진자의 회전축까지의 길이, L_2는 2단 진자의 회전축과 3단 진자의 회전축까지의 길이를 의미한다. u는 cart의 가속도, y는 cart의 초기위치로부터의 변위를 의미하고 c_1, c_2, c_3는 각 진자의 회전축에서 발생하는 마찰계수를 의미한다. θ_1은 1단 진자의

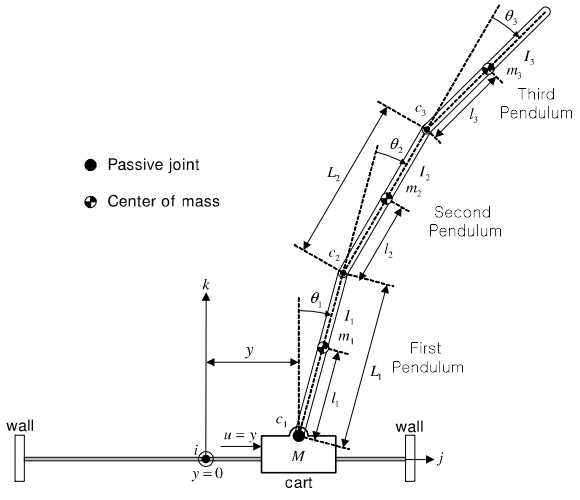


그림 1 3단 도립진자의 개념도

Fig. 1 The conceptual diagram of a triple inverted pendulum

회전 변위로써 지면에 대한 법선과 이루는 각이며, θ_2는 2단 진자가 1단 진자와 이루는 상대적인 회전 변위, θ_3는 3단 진자와 2단 진자가 이루는 상대적인 회전 변위이다. i, j, k는 각각 rail의 중심점을 원점으로 하는 직각좌표계의 좌표축을 나타낸다.

3.1 3단 도립진자의 모델 방정식

3단 도립진자의 수학적 모델은 Euler-Lagrange equation을 이용하여 유도하면 식 (1)과 같이 나타낼 수 있다.

$$\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \ddot{y} + \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ \star & m_{22} & m_{23} \\ \star & \star & m_{33} \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = 0. \tag{1}$$

여기서 ★는 대칭행렬의 대칭부 요소를 나타낸다. n_i, m_{ij}, r_i의 의미는 식 (2)와 같이 정리할 수 있다.

$$\begin{aligned}
n_1 &= h_6\cos(\theta_1) + h_7\cos(\theta_1+\theta_2) \\
&\quad + h_8\cos(\theta_1+\theta_2+\theta_3), \\
n_2 &= h_7\cos(\theta_1+\theta_2) + h_8\cos(\theta_1+\theta_2+\theta_3), \\
n_3 &= h_8\cos(\theta_1+\theta_2+\theta_3), \\
m_{11} &= h_1 + h_2 + h_3 + h_4 + h_5 + 2h_9\cos(\theta_2) \\
&\quad + 2h_{11}\cos(\theta_3) + 2h_{10}\cos(\theta_2+\theta_3), \\
m_{12} &= h_2 + h_3 + h_5 + h_9\cos(\theta_2) + 2h_{11}\cos(\theta_3) \\
&\quad + h_{10}\cos(\theta_2+\theta_3), \\
m_{13} &= h_3 + h_{11}\cos(\theta_3) + h_{10}\cos(\theta_2+\theta_3), \\
m_{22} &= h_2 + h_3 + h_5 + 2h_{11}\cos(\theta_3), \\
m_{23} &= h_3 + h_{11}\cos(\theta_3), \\
m_{33} &= h, \\
r_1 &= gh_8\sin(\theta_1+\theta_2+\theta_3) + gh_9\sin(\theta_1+\theta_2) \\
&\quad + gh_6\sin(\theta_1) + h_{10}d_1\sin(\theta_2+\theta_3) \\
&\quad + h_9d_2\sin(\theta_2) + h_{11}d_3\sin(\theta_3) - c_1\dot{\theta}_1, \\
r_2 &= gh_8\sin(\theta_1+\theta_2+\theta_3) + gh_9\sin(\theta_1+\theta_2) \\
&\quad - h_{10}d_4\sin(\theta_2+\theta_3) - h_9d_4\sin(\theta_2) \\
&\quad + h_{11}d_3\sin(\theta_3) - c_2\dot{\theta}_2, \\
r_3 &= gh_8\sin(\theta_1+\theta_2+\theta_3) - h_{10}d_4\sin(\theta_2+\theta_3) \\
&\quad - h_{11}d_5\sin(\theta_3) - c_3\dot{\theta}_3.
\end{aligned} \tag{2}$$

여기서 g는 중력가속도이며 h_i와 d_i의 의미는 식 (3)과 같다.

$$\begin{aligned}
h_1 &= m_1l_1^2 + I_1, & h_2 &= m_2l_2^2 + I_2, \\
h_3 &= m_3l_3^2 + I_3, & h_4 &= L_1^2m_2 + L_1^2m_3, \\
h_5 &= L_2^2m_3, & h_6 &= l_1m_1 + L_1m_2 + L_1m_3, \\
h_7 &= l_2m_2 + L_2m_3, & h_8 &= l_3m_3, \\
h_9 &= L_1l_2m_2 + L_1L_2m_3, & h_{10} &= L_1l_3m_3, \\
h_{11} &= L_2l_3m_3, \\
d_1 &= (2\dot{\theta}_1 + \dot{\theta}_2 + \dot{\theta}_3)(\dot{\theta}_2 + \dot{\theta}_3), \\
d_2 &= (2\dot{\theta}_1 + \dot{\theta}_2)\dot{\theta}_2, & d_3 &= (2\dot{\theta}_1 + 2\dot{\theta}_2 + \dot{\theta}_3)\dot{\theta}_3, \\
d_4 &= \dot{\theta}_1^2, & d_5 &= (\dot{\theta}_1 + \dot{\theta}_2)^2.
\end{aligned} \tag{3}$$

상태방정식 유도를 위해 식 (1)은 식 (4)로 정리할 수 있고

$$\begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \\ \ddot{\theta}_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ \star & m_{22} & m_{23} \\ \star & \star & m_{33} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \ddot{y} + \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \right\}. \tag{4}$$

식 (4)를 통해 $\ddot{\theta}_1$, $\ddot{\theta}_2$, $\ddot{\theta}_3$를 식 (5)처럼 표현할 수 있다.

$$\begin{aligned}
\ddot{\theta}_1 &= b_{11}r_1 + b_{12}r_2 + b_{13}r_3 - (b_{11}n_1 + b_{12}n_2 + b_{13}n_3)\ddot{y} \\
\ddot{\theta}_2 &= b_{12}r_1 + b_{22}r_2 + b_{23}r_3 - (b_{12}n_1 + b_{22}n_2 + b_{23}n_3)\ddot{y} \\
\ddot{\theta}_3 &= b_{13}r_1 + b_{23}r_2 + b_{33}r_3 - (b_{13}n_1 + b_{23}n_2 + b_{33}n_3)\ddot{y}
\end{aligned} \tag{5}$$

여기서 b_{ij}와 Φ는 식 (6)과 같다.

$$\begin{aligned}
b_{11} &= \frac{m_{22}m_{33} - m_{23}^2}{\Phi}, \\
b_{12} &= \frac{m_{13}m_{23} - m_{12}m_{33}}{\Phi}, \\
b_{13} &= \frac{m_{12}m_{23} - m_{13}m_{22}}{\Phi}, \\
b_{22} &= \frac{m_{11}m_{33} - m_{13}^2}{\Phi}, \\
b_{23} &= \frac{m_{12}m_{13} - m_{11}m_{23}}{\Phi}, \\
b_{33} &= \frac{m_{11}m_{22} - m_{12}^2}{\Phi}, \\
\Phi &= m_{11}m_{22}m_{33} + 2m_{12}m_{13}m_{23} \\
&\quad - m_{11}m_{23}^2 - m_{12}^2m_{33} - m_{13}^2m_{22}.
\end{aligned} \tag{6}$$

상태벡터를 $x_1 = y$, $x_2 = \theta_1$, $x_3 = \theta_2$, $x_4 = \theta_3$, $x_5 = \dot{y}$, $x_6 = \dot{\theta}_1$, $x_7 = \dot{\theta}_2$, $x_8 = \dot{\theta}_3$, $x_9 = \int_0^t y(\tau)d\tau$로 정의하고 $\ddot{y}$를 가속도 u로 나타내면 다음과 같이 3단 도립진자의 모델 방정식을 비선형 상태방정식으로 나타낼 수 있다.

$$\underbrace{\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{x}_7 \\ \dot{x}_8 \\ \dot{x}_9 \end{bmatrix}}_{=\dot{x}} = \underbrace{\begin{bmatrix} x_5 \\ x_6 \\ x_7 \\ x_8 \\ u \\ b_{11}r_1 + b_{12}r_2 + b_{13}r_3 - (b_{11}n_1 + b_{12}n_2 + b_{13}n_3)u \\ b_{12}r_1 + b_{22}r_2 + b_{23}r_3 - (b_{12}n_1 + b_{22}n_2 + b_{23}n_3)u \\ b_{13}r_1 + b_{23}r_2 + b_{33}r_3 - (b_{13}n_1 + b_{23}n_2 + b_{33}n_3)u \\ x_1 \end{bmatrix}}_{f(x,u)}. \tag{7}$$

상태변수의 마지막 요소인 $\int_0^t y(\tau)d\tau$는 cart의 위치 정상상태 오차를 없애기 위해 추가한 항이다. 3단 도립진자의 모델식은 cart에 j축 방향의 수평운동 외의 수평운동과 회전운동은 발생하지 않는다고 가정한다. 그리고 1단 진자와 2단 진자는 각 hinge에서 i축 방향의 회전축을 중심으로 한 회전만이 존재

한다고 가정한다. 해당 모델식은 속도에 선형적인 관계를 가지는 마찰만을 고려하며 비선형적 관계를 가지는 정지 마찰과 Coulomb 마찰은 고려하지 않는다. 유도된 모델 방정식을 이용해 Sim-to-Real 학습 기법을 사용하기 위해서는 이러한 가정에 최대한 부합하는 기구부 설계가 이루어져야 한다.

3.2 3단 도립진자의 기구부 및 구동부

실제 사용되는 시스템이 reality gap을 최소화하려면 이론적으로 유도된 모델 방정식과 높은 정합성을 유지해야 한다. 이를 위해 유도된 가정에 부합하는 동작만을 수행하도록 설계하는 것이 필수적이다. 만일 시스템이 가정과 다른 동작을 수행하면, 시뮬레이션 환경과 실물 시스템간의 동적 응답 차이가 발생하여 모델의 신뢰도가 저하될 수 있다. 따라서 본 연구에서 제안하는 3단 도립진자의 기구부 및 구동부 설계는 이론적/실험적 기준에 부합하도록 정합성을 극대화하는 것을 목표로 한다. 제안하는 3단 도립진자의 기구적 구조는 그림 2와 같다.

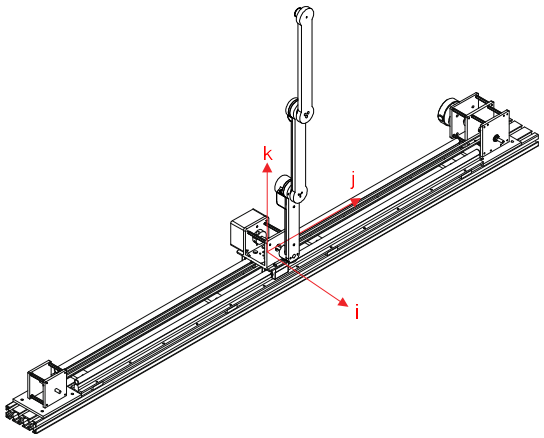


그림 2 3단 도립진자 시스템의 기구적 구조
Fig. 2 The mechanical structure of triple inverted pendulum system

제안된 3단 도립진자 시스템은 각 진자 간 연결 방식의 정밀도를 고려하여 설계되었다. 그림 3에서 확인할 수 있듯이, 각 진자를 연결하는 revolute joint는 단일 bearing이 아닌 복렬 bearing 구조를 적용하여 회전이 단일 축을 기준으로 안정적으로 이루어지도록 하였다. 이를 통해 불필요한 방향의 움직임을 최소화하고 정밀한 회전 성능을 확보할 수 있도록 하였다.

또한 3단 진자의 2단 진자에 대한 회전각 θ_3 및 2단 진자의 1단 진자에 대한 회전각 θ_2를 측정하기 위해 소형 자기식 엔코더를 장착하였다. 특히 θ_3를 측정하는 엔코더를 slip ring에 연결하기 위해서는 1단 진자와 2단 진자의 연결 부위를 관통해야 한다. 이를 위해 본 연구에서는 중공축(hollow shaft) revolute joint를 사용해 진자 간의 간섭을 줄이고 회전 정보를 원활히 받아올 수 있도록 설계하였다.

그림 4는 이전에 본 연구실에서 제작한 3단 도립진자 시스템의 rail 및 cart의 구조이다[5]. 해당 구조에서는 진자의 운동에 따라 카트에 비틀림(α)이 발생하는 문제가 관찰되었다. 이는 모델 방정식에서 고려되지 않은 요소이며 시뮬레이션 환경

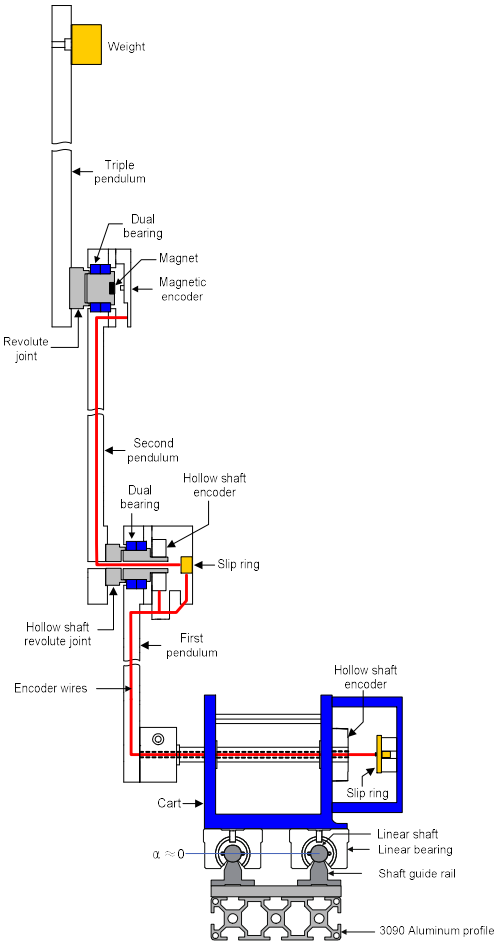


그림 3 제안되는 3단 도립진자의 단면도 및 엔코더 배선
Fig. 3 Cross-sectional view and encoder wiring of the proposed triple inverted pendulum

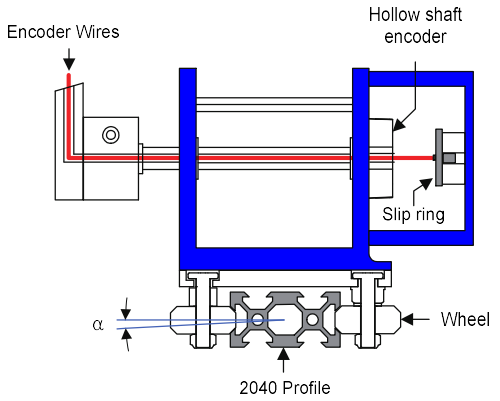


그림 4 2040 알루미늄 프로파일을 이용한 레일 및 카트 구조
Fig. 4 The structure of the rail and cart constructed using 2040 aluminum profile

과의 정합성을 저하시키는 원인 중 하나이다. 이를 해결하기 위해 그림 5와 같이 이중 샤프트 가이드 레일 구조를 적용하였다. 제안된 구조는 기존 구조 대비 더욱 견고한 고정을 제공하여 진자의 움직임으로 인한 비틀림을 완화할 수 있으며 벨트의 장력이 pulley를 회전시키는 축에만 전달되도록 유도할 수 있다.

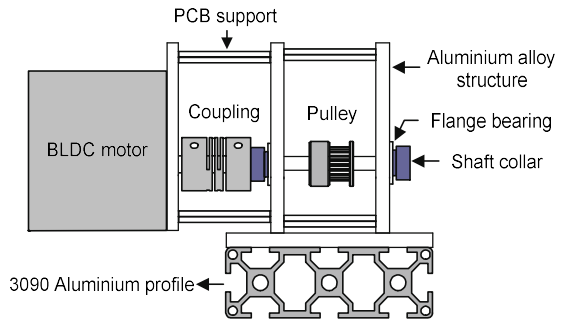


그림 5 제안되는 구동부 구조
Fig. 5 Proposed driving structure

본 연구에서는 그림 5와 같이 감속기를 사용하지 않은 BLDC 모터를 채택하여 pulley를 직접 구동하도록 설계하였다. 이러한 방식은 백래시를 제거하여 limit cycle 현상의 발생을 최소화하는 효과를 기대할 수 있다. 또한, BLDC 모터에서 동력을 전달하는 부분에 coupling을 사용해 불필요한 부하가 출력에 영향을 주는 것을 방지하였다.

제안되는 3단 도립진자 시스템에서는 cart의 이동부, 구동부, 그리고 각 진자에 bearing이 사용되며, 모델에서는 속도 및 회전각속도에 비례하는 점성 마찰만을 고려하였다. 정지 마찰이나 쿨롱 마찰 등은 포함하지 않으며 실제로 제작되는 도립진자 시스템 역시 이러한 모델링 특성을 반영하도록 설계되어야 한다.

공장에서 출고된 bearing은 장기간 사용을 고려하여 점성이 높은 grease가 도포된 상태이다. 그러나 이러한 bearing을 별도의 처리 없이 3단 도립진자에 적용할 경우 cart의 움직임과 진자 회전 시 불필요한 마찰이 발생하며 점성 마찰 성분 증가로 인해 원활한 구동을 방해할 가능성이 높다.

특히 revolute joint에 사용된 bearing에서 정지 마찰이 발생할 경우 도립진자가 초기 상태에서 움직이기 어려워지며 예기치 않은 초기 상태 편차가 발생할 가능성이 있다. 예를 들어 안정한 균형점에서 작은 편차가 생길 경우, 시스템이 초기 설정과 다른 상태로 이동할 수 있으며 이는 limit cycle 현상을 유발하는 주요 원인 중 하나로 작용할 수 있다. 이를 방지하기 위해 본 연구에서는 solvent를 사용하여 bearing의 그리스를 제거한 후 bearing 내부를 오일 처리하여 마찰을 최소화하는 방법을 적용하였다.

3.3 천이 제어

천이 제어는 다양한 균형점 간의 천이를 다루므로 시스템 내 균형점을 체계적으로 정의하고 이를 제어의 목표 상태로 명확히 설정하는 과정이 선행되어야 한다. 3단 도립진자의 균형점은 각 진자의 angle 값에 따라 총 8가지로 구분된다. 본 연구에서는 각 진자의 상태를 Down 또는 Up으로 표기하며 Down에 0, Up에 1을 대입하면 2진수 형식으로 표현이 가능하여 균형점의 순서를 구분하기 쉽게 나타낼 수 있다. 균형점은 EP(Equilibrium Point)로 표기하며 각 진자의 조합에 따라 EP0(Down, Down, Down), EP1(Down, Down, Up), EP2(Down, Up, Down), EP3(Down, Up, Up), EP4(Up, Down, Down), EP5(Up, Down, Up), EP6(Up, Up, Down), EP7(Up, Up, Up)과 같이 구분된다. 이러한 조합은 그림 6에 시각적으로 제시되어 있다.

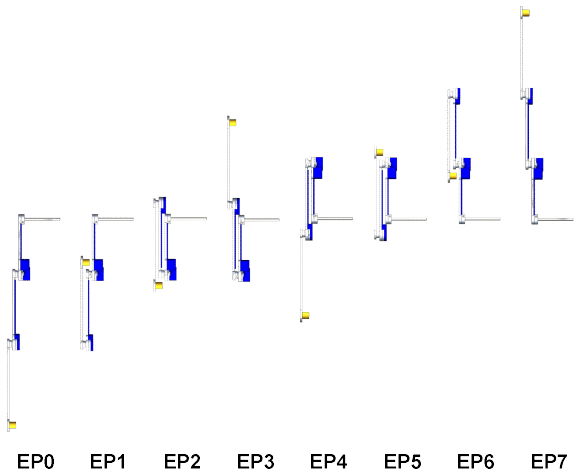


그림 6 3단 도립진자의 균형점
Fig. 6 Equilibrium point of triple inverted pendulum

천이 제어와 관련한 선행 연구는 각 균형점 간의 천이 궤적을 사전에 계산한 후 이를 추종하는 방식을 적용하였다[6, 7]. 이러한 방식은 궤적을 정확히 추종할 수 있는 환경에서는 효과적이지만 외란이 존재하는 경우에는 사전에 계산된 궤적을 따라가기 어려워 성능 저하가 발생할 수 있다. 반면 Sim-to-Real 방식은 천이 궤적을 직접 계산하지 않고 목표 균형점을 보상 함수의 최대값으로 설정하여 학습하는 방식을 사용한다. 즉 특정한 궤적을 사전에 정의하는 것이 아닌 균형점 자체를 최종 목표 상태로 설정함으로써 진자가 어떤 초기 상태에서 출발하든 주어진 목표 균형점으로 자연스럽게 수렴하도록 학습된다. 해당 방식은 3단 도립진자의 56가지 천이 궤적을 직접 구할 필요 없이 8가지 균형점에 대한 학습만으로도 천이 제어를 효과적으로 수행할 수 있다. 또한 Sim-to-Real 학습 기법을 통해 다양한 초기 조건과 환경 변화에도 강인한 제어 성능을 확보할 수 있고 천이 과정에서 발생할 수 있는 다양한 외란이나 모델 불확실성을 보다 효과적으로 극복할 수 있다.

4. 실험 및 결과

4.1 시뮬레이션 환경 설정

강화학습 에이전트가 학습 과정에서 직접 상호작용하는 환경은 3장에서 유도된 수학적 모델을 기반으로 Python을 이용하여 시뮬레이션 환경으로 구현하였다. 3단 도립진자 시스템의 물리적 파라미터를 반영하여 환경을 구축했으며 해당 파라미터는 표 2에 정리되어 있다. 또한 비선형 상미분 방정식의 해를 구하기 위해 ode4 Runge-Kutta 방법을 solver로 채택하였다.

표 2 3단 도립진자의 물리적 파라미터
Table 2 Physical parameters of triple inverted pendulum

Parameter	Link		
	$i=1$	$i=2$	$i=3$
m_i[kg]	0.2297	0.1345	0.1644
L_i[m]	0.1645	0.210	0.245
l_i[m]	0.0819	0.1239	0.1532
I_i[kgm²]	1.269e-03	9.371e-04	1.744e-03
c_i[Nms/rad]	1.293e-03	1.626e-06	3.305e-04

시뮬레이션 학습 환경에서 각 에피소드의 길이는 10초로 설정했으며, ODE solver는 1ms 간격으로 연산을 수행했고, 에이전트는 10ms마다 상태 정보를 관측하도록 구성하였다. 이러한 설정을 통해 에이전트는 에피소드당 최대 1000회 동안 환경과 상호작용하며 점진적으로 최적의 행동 정책을 학습할 수 있도록 설계되었다. 에피소드의 종료 조건은 timestep이 1000을 초과하는 경우 외에도 추가적으로 cart의 변위 y가 0.48[m]를 초과하거나 cart의 가속도 a가 2.5[m/s²] 보다 클 경우 조기 종료되도록 설정하였다. 이는 학습된 제어기가 실물 시스템에 적용될 때 cart가 레일의 한계를 벗어나거나 시스템에 손상이 갈 수 있는 상황을 방지하기 위한 사전적인 안전 조치이다.

4.2 보상함수 설계

강화학습 에이전트는 환경과 지속적으로 상호작용하며 매 시점에서 얻은 보상 값을 바탕으로 자신의 행동 정책을 점진적으로 최적화한다. 이때 보상 값을 산출하기 위한 보상 함수는 3단 도립진자 시스템에서 존재하는 8개의 균형점 중 어떤 균형점에 도달하기 위한 천이 제어를 수행하는지에 따라 달라지게 된다. 그림 7에서 정해둔 균형점에 맞춰 보상이 최대가 되는 target angle은 표 3과 같다.

표 3 균형점에 따른 각 진자의 목표 각도
Table 3 Target angle of each pendulum according to the equilibrium point

Equilibrium Point	Target Angle		
	θ_1^*	θ_2^*	θ_3^*
0	-π	-π	-π
1	-π	-π	0
2	-π	0	-π
3	-π	0	0
4	0	-π	-π
5	0	-π	0
6	0	0	-π
7	0	0	0

각 균형점에서 최대 보상이 되도록 설계한 보상 함수는 식 (8)과 같고 그래프로 표현하면 그림 7과 같다.

$$
\begin{aligned}
R_u &= \exp(-0.001 \cdot u^2),\\
R_y &= \exp(-0.3 \cdot |y|),\\
R_{\theta_1} &= 0.5+0.5 \cdot cos(\theta_1-\theta_1^*),\\
R_{\theta_2} &= 0.5+0.5 \cdot cos(\theta_1+\theta_2-\theta_2^*),\\
R_{\theta_3} &= 0.5+0.5 \cdot cos(\theta_1+\theta_2+\theta_3-\theta_3^*),\\
R_{\dot\theta_1} &= \exp(-0.015 \cdot |\dot\theta_1|),\\
R_{\dot\theta_2} &= \exp(-0.009 \cdot |\dot\theta_1+\dot\theta_2|),\\
R_{\dot\theta_3} &= \exp(-0.005 \cdot |\dot\theta_1+\dot\theta_2+\dot\theta_3|).
\end{aligned}
\tag{8}
$$

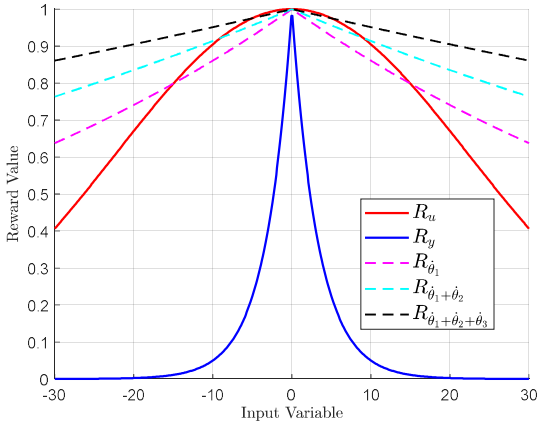


그림 7 보상 함수 그래프
Fig. 7 Reward function graph

최종적인 보상 함수는 모든 보상 값을 곱하여 최댓값이 1이 되는 형태로 식 (9)와 같이 표현할 수 있다.

$$
Reward = R_u \cdot R_y \cdot R_{\theta_1} \cdot R_{\theta_2} \cdot R_{\theta 3} \cdot R_{\dot\theta_1} \cdot R_{\dot\theta_2} \cdot R_{\dot\theta_3}. \tag{9}
$$

앞서 설계한 모든 보상 함수는 [0, 1]의 값으로 정규화가 이루어진 형태이며 각 에피소드는 최대 1000개의 timestep으로 구성되므로 단위 timestep마다 1의 보상을 얻는다고 가정할 경우 하나의 에피소드에서 획득할 수 있는 보상의 최댓값은 1000이 된다.

본 연구에서 사용된 보상 함수는 균형점에 대한 의존도에 따라 두 가지 유형으로 구분할 수 있다. 첫 번째 유형은 target angle에 종속적인 보상 함수로 이는 R_{θ_1}, R_{θ_2}, R_{θ_3}로 정의된다. 해당 보상 함수들은 목표 균형점과의 오차가 감소할수록 보상이 증가하는 형태를 가지며, 이를 통해 에이전트가 각 진자를 균형점으로 수렴시키는 행동 정책을 학습하도록 유도한다.

두 번째 유형은 R_u, R_y, R_{θ_1}, R_{θ_2}, R_{θ_3}로 구성되며 <u, y, $\dot\theta_1$, $\dot\theta_2$, $\dot\theta_3$>라는 각 매개변수의 값이 0에 가까워질수록 보상이 증가하는 방식으로 설계된다. 이는 에이전트가 제어 입력을 최소화하고 cart의 위치를 원점 부근으로 유지하며 진자의 불필요한 움직임을 억제할 수 있도록 학습하는데 도움을 준다.

4.3 학습 전략

시뮬레이션 환경을 설정한 후 각 균형점에 맞춰 target angle

을 변경해가며 총 8회에 걸쳐 학습을 진행하였다. 그 결과는 그림 8과 같으며 보상값이 약 700에서 800이라는 값에 도달한 후 일정한 수준을 유지하는 경향을 보였다. 또한 각 균형점마다 학습이 완료되는 시점이 다르게 나타났으며 이는 균형점 간의 제어 난이도 차이에 기인하는 것으로 분석된다. 이러한 차이는 각 균형점의 안정성 및 제어 난이도뿐만 아니라 보상 함수의 구조, 탐색 과정의 차이 등에 의해서도 영향을 받을 수 있다. 추가적으로 학습이 완료된 이후에도 보상이 일정한 값으로 완전히 수렴하지 않는 모습을 확인할 수 있었는데 이는 외란이 존재하는 환경에서도 강건한 제어 정책을 학습할 수 있도록 설계된 학습 조건 때문이다.

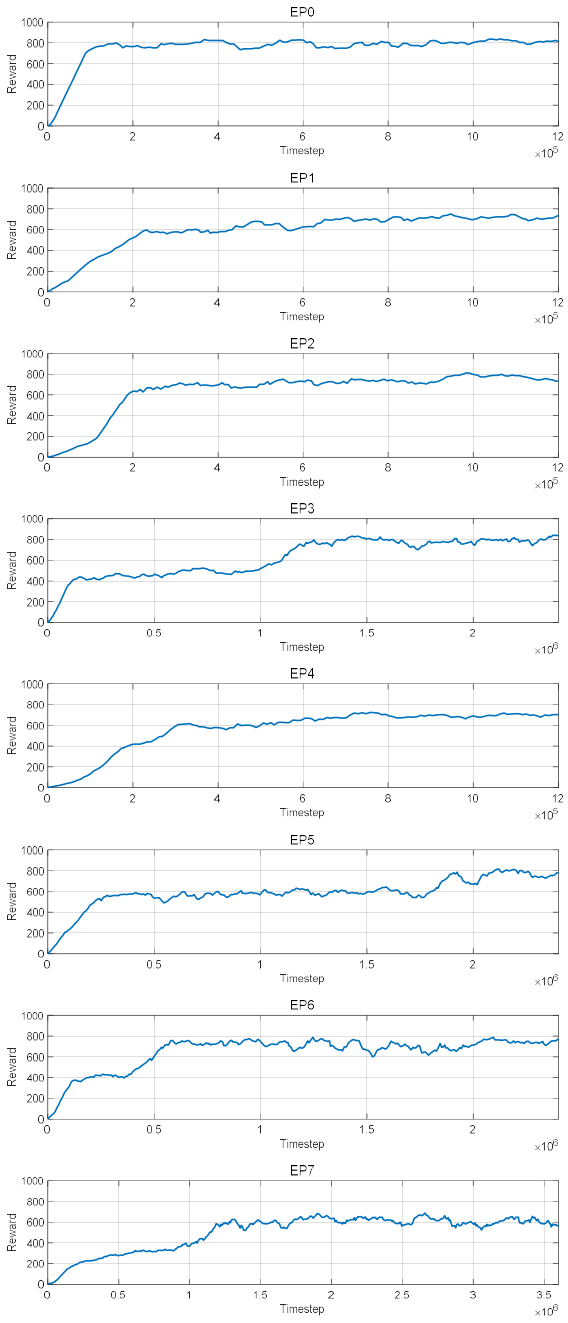


그림 8 각 균형점에 대한 학습 결과
Fig. 8 Result for learning about each equilibrium point

강화학습 에이전트가 보다 다양한 상태를 경험하고 일반화된 제어 정책을 학습할 수 있도록, 각 에피소드의 초기 상태는 무작위성을 가지도록 설정하였다. 이를 위해 시뮬레이션 환경의 비선형 상태방정식을 구성하는 상태 변수들을 난수로 초기화하여 에이전트가 광범위한 상태 공간을 탐색할 수 있도록 하였다. 다만 초기 상태 변수의 난수 범위는 실물 시스템의 물리적 한계를 고려하여 설정하였으며 그 범위는 식 (10)과 같이 정의된다.

$$\begin{aligned} y &\sim U(-0.3,0.3), & \dot{y} &\sim U(-1.2,1.2), \\ \theta_1 &\sim U(-\pi,\pi), & \dot{\theta}_1 &\sim U(-10,10), \\ \theta_2 &\sim U(-\pi,\pi), & \dot{\theta}_2 &\sim U(-20,20), \\ \theta_3 &\sim U(-\pi,\pi), & \dot{\theta}_3 &\sim U(-30,30). \end{aligned} \tag{10}$$

그러나 이러한 난수 기반 초기화 과정에서 일부 상태 변수 조합이 현실적인 물리 법칙을 위배하는 경우가 발생할 수 있다. 물리적으로 불가능한 초기 상태에서 학습이 시작되면 모델 방정식의 연산 결과 역시 비현실적인 값으로 이어질 수 있다. 강화학습 에이전트의 관점에서는 이전 학습 과정에서 한 번도 경험하지 못했던 불규칙한 상태를 입력받게 되며 학습된 행동 정책과 무관한 예측 불가능한 행동을 출력할 가능성이 증가한다. 이로 인해 제어 시스템의 동작이 비정상적으로 이루어지고 설정된 종료 조건을 조기에 충족시켜 학습이 조기 종료될 가능성이 증가한다. 이처럼 물리적으로 의미 없는 초기 상태가 특정 에피소드에서 발생할 경우 보상 값의 평균에도 변동성이 증가하는 경향을 보인다.

반면 학습된 제어기를 실물 시스템에 적용할 경우에는 이러한 문제가 발생하지 않는다. 실제 환경에서는 물리 법칙에 위배되는 상태가 자연적으로 발생할 수 없기 때문에 에이전트가 비현실적인 상태 정보를 관측할 가능성이 사라진다. 따라서 강화학습 과정에서 학습된 행동 정책이 정상적인 상태 정보에 기반하여 안정적으로 동작할 수 있으며 보다 신뢰성 높은 제어 성능을 기대할 수 있다.

4.4 실험 및 결과

그림 9는 3단 도립진자의 천이 제어 실험의 결과를 보여주는 Youtube 영상을 캡쳐한 그림이며 실제 영상의 주소는 https://youtu.be/vVx3ffGo2mk와 같다. (영상 제목 : World's first

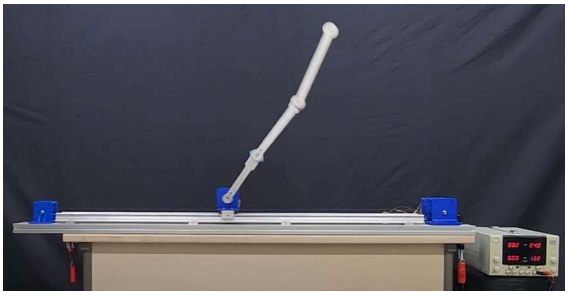


그림 9 천이 제어 실험 영상
Fig. 9 Experimental image of transition control

reinforcement learning-based transition control of a triple inverted pendulum, 채널명 : Embedded Control Lab.)

실험 결과 모든 균형점에서 제어가 성공적으로 이루어졌으며 그림 10은 그중 일부 천이 제어 결과를 시각적으로 제시한 것이다. 해당 그래프는 안정한 균형점인 EP0에서 시작해 각기 다른 균형점으로의 천이 제어 결과를 나타낸다. 천이 순서는 EP0을 시작으로 EP4, EP1, EP6, EP2, EP5, EP7 그리고 최종적으로 EP3로 이어진다. 그래프에서 확인할 수 있듯이 제어의 주요 대상인 θ_1, θ_2, θ_3는 모든 균형점에서 안정적으로 목표값에 수렴하는 양상을 보였다. 이는 각 균형점에 대한 학습이 성공적으로 이루어졌음을 의미하며 나아가 3단 도립진자의 56가지 천이 제어를 모두 성공적으로 수행할 수 있음을 실험적으로 입증한다.

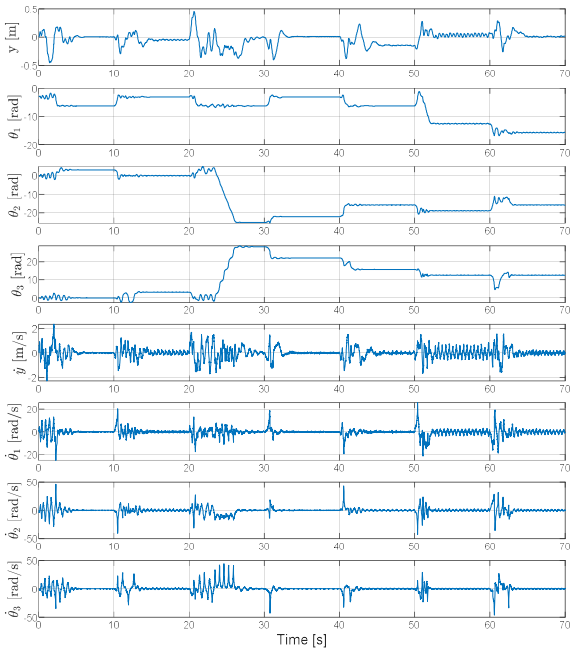


그림 10 천이 제어 결과
Fig. 10 Result of transition control

그러나 일부 균형점에서는 각도별로 약간의 진동이 관찰된다. θ_1은 EP3, EP7, θ_2는 EP1, EP3, EP5, EP7, θ_3는 EP1, EP5에서 진동이 발생하였다. 이러한 현상은 센서의 해상도에 따른 양자화 오차 영향으로 설명될 수 있으며, 이는 측정 정확도 저하와 직접적인 관련이 있는 것으로 분석된다. 본 연구에 사용된 실물 시스템은 모델 방정식과의 정합성을 고려하여 설계되었으나 진자의 각도를 측정하는 엔코더의 해상도는 한계가 존재한다. 도립진자 시스템의 1단 및 2단 진자부에는 8192 CPR(Counts Per Revolution), 3단 진자부에는 4096 CPR 해상도의 엔코더가 부착되어 있으며 각속도 산출 시 양자화 오차가 발생할 수 있다. 이러한 오차는 천이 제어 중 고속 구간에서는 영향이 미미하나 균형점 도달 이후 시스템이 저속 상태로 전환될 경우 관측된 상태 정보에 보다 큰 영향을 미치며 이로 인해 제어 입력이 반복적으로 미세하게 변동되며 리플이 발생할 수 있다.

또한 각 진자에서 진동이 발생한 균형점들의 공통 특성을 분석한 결과 θ_1의 경우 2단 및 3단 진자부가 모두 도립된 상태에서, θ_2는 3단 진자부가 도립된 상태에서, θ_3는 2단 진자부가 아래를 향하고 3단 진자부가 도립된 상태에서 진동이 발생하는 경향을 보였다. 세 경우 모두 공통적으로 3단 진자부가 도립된 상태라는 점에서 가장 복잡한 모델 특성을 가지는 3단 진자부의 제어 민감도가 진동 현상의 주요 원인으로 해석될 수 있다. 즉 3단 진자부는 양자화 오차에 따른 제어 입력의 작은 변동에도 민감하게 반응하며 이에 따른 반복적인 리플이 관찰된다. 이러한 안정화 이후의 리플 현상을 저감하기 위해서는 보다 고해상도의 엔코더를 적용하여 정밀한 각도 센싱을 수행하거나 모델 기반 필터링 기법을 통해 속도 정보를 소프트웨어적으로 보정하는 방식이 효과적일 것으로 기대된다.

5. 결 론

본 논문에서는 Sim-to-Real 강화학습 기법을 활용하여 직선형 3단 도립진자의 56가지 천이 제어를 구현하였다. 이를 위해 물리적 정합성이 우수한 기구부와 제어 환경을 설계하여 reality gap을 최소화하였다. 제안된 제어기는 목표 균형점에서의 보상이 최대가 되도록 보상 함수를 설정하고 각 균형점에 대한 개별 학습을 수행하여, 다양한 초기 조건 및 외란 상황에서도 강인하게 동작할 수 있도록 설계되었다. 시뮬레이션과 실물 시스템을 통한 실험 결과, 3단 도립진자의 모든 천이 제어에서 안정적인 수렴성과 높은 정합성을 확인하였다. 본 연구는 Sim-to-Real 강화학습 기법을 활용해 기존 방식으로 구현하기 어려웠던 복잡한 천이 제어 문제를 효과적으로 해결할 수 있음을 실증하였으며 향후 다양한 비선형 시스템의 실용적 제어기로서의 확장 가능성을 제시하였다.

Acknowledgements

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2024-00347193).

References

[1] Y. Otani, T. Kurokami, A. Inoue and Y. Hirashima, “A swingup control of an inverted pendulum with cart position control,” IFAC Proceedings Volumes, vol. 34, no. 22, pp. 13-22, 2001.
DOI:10.1016/S1474-6670(17)32971-3

[2] H. Li, Z. Nie, E. Zhu, W. He and Y. Zheng, “Double Loop DR-PID Control of A Rotary Inverted Pendulum,” 2021 IEEE International Conference on Networking, Sensing and Control(ICNSC), pp. 1-5, 2021.
DOI:10.1109/ICNSC52481.2021.9702192

[3] A. Dev, K. R. Chowdhury and M. P. Schoen, “Q-Learning Based Control for Swing-Up and Balancing of Inverted Pendulum,” 2024 Intermountain Engineering, Technology and Computing(IETC), pp. 209-214, 2024.
DOI:10.1109/IETC61393.2024.10564347

[4] T. Glück, A. Eder and A. Kugi, “Swing-up control of a triple pendulum on a cart with experimental validation,” Automatica, vol. 49, no. 3, pp. 801-808, 2013.
DOI:10.1016/j.automatica.2012.12.006

[5] C. Choi, D. Ju, J. Jeong and Y. S. Lee, “Structural Proposition for a Triple Inverted Pendulum and Implementation of Swing-up Control Using an LW-RCP02,” Journal of Institute of Control, Robotics and Systems (in Koreans), vol. 28, no. 10, pp. 916-925, 2022.
DOI:10.5302/J.ICROS.2022.22.0176

[6] J. Jeong, D. Ju, Y. Fujiyama and Y. S. Lee, “Transition Control of a Double Inverted Pendulum Using an LW-RCP,” Journal of Institute of Control, Robotics and Systems (in Koreans), vol. 29, no. 9, pp. 694-703, 2023.
DOI:10.5302/J.ICROS.2023.23.0100

[7] D. Ju, T. Lee and Y. S. Lee, “Transition Control of a Rotary Double Inverted Pendulum Using an Direct Collocation,” Mathematics, vol. 13, no. 4, Art. no. 640, 2025.
DOI:10.3390/math13040640

[8] L. R. E. Shead, K. R. Muske and J. A. Rossiter, “Conditions for which MPC fails to converge to the correct target,” IFAC Proceedings Volumes, vol. 41, no. 2, pp. 6968-6973, 2008.
DOI:10.3182/20080706-5-KR-1001.01181

[9] W. Zhu, X. Guo, D. Owaki, K. Kutsuzawa and M. Hayashibe, “A Survey of Sim-to-Real Transfer Techniques Applied to Reinforcement Learning for Bioinspired Robots,” IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 7, pp. 3444-3459, 2023.
DOI:10.1109/TNNLS.2021.3112718

[10] E. Salvato, G. Fenu, E. Medvet and F. A. Pellegrino, “Crossing the Reality Gap: A Survey on Sim-to-Real Transferability of Robot Controllers in Reinforcement Learning,” IEEE Access, vol. 9, pp. 153171-153187, 2021.
DOI:10.1109/ACCESS.2021.3126658

[11] B. Qin, Y. Gao and Y. Bai, “Sim-to-real: Six-legged Robot Control with Deep Reinforcement Learning and Curriculum Learning,” 2019 4th International Conference on Robotics and Automation Engineering(ICRAE), pp. 1-5, 2019.
DOI:10.1109/ICRAE48301.2019.9043822

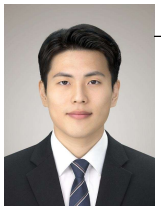
[12] G. Fang, Y. Tian, Z. Yang, J. M. P. Geraedts and C. C. L. Wang, “Efficient Jacobian-Based Inverse Kinematics With Sim-to-Real Transfer of Soft Robots by Learning,” IEEE/ASME Transactions on Mechatronics, vol. 27, no. 6, pp. 5296-5306, 2022.
DOI:10.1109/TMECH.2022.3178303

[13] M. Ranaweera and Q.H Mahmoud, “Bridging the Reality Gap Between Virtual and Physical Environments Through Reinforcement Learning,” IEEE Access, vol. 11, pp. 19914-19927, 2023.
DOI:10.1109/ACCESS.2023.3249572

[14] A. Pitkevich and I. Makarov, “A Survey on Sim-to-Real Transfer Methods for Robotic Manipulation,” 2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics(SISY), pp. 259-266, 2024.
DOI:10.1109/SISY62279.2024.10737545

[15] A. Kuznetsov, P. Shvechikov, A. Grishin and D. Vetrov, “Controlling Overestimation Bias with Truncated Mixture of Continuous Distributional Quantile Critics,” arXiv preprint arXiv:2005.04269, 2020.
DOI:10.48550/arXiv.2005.04269

저자소개



임창석(Changseok Lim)

He received B.S. degree in electrical engineering from Inha university in 2024. He is now a M.S. candidate in electrical and computer engineering at Inha university. His research interests include optimal control, reinforcement learning and embedded systems.

주도윤(Doyoon Ju)

He received M.S. degree in electrical and computer engineering from Inha university in 2023. He is now a Ph.D. candidate in electrical and computer engineering at Inha university. His research interests include optimal control, embedded systems and reinforcement learning.

이영삼(Young Sam Lee)

He received B.S. and M.S. degrees in electrical engineering from Inha University, Incheon, South Korea, in 1999, and the Ph.D. degree in electrical engineering from Seoul National University, South Korea, in 2003. From 2003 to 2004, he was a Senior Researcher with Samsung Electronics Co. Since 2004, he has been with the Department of Electrical and Computer Engineering, Inha University. He is the author of four books and more than 60 articles. His research interests include computer– aided control system designs, rapid control prototyping, control and instrumentation, robot engineering, and embedded systems.