https://doi.org/10.7840/kics.2023.48.4.449

중앙집중형 DQN 기반 다중 로봇의 최적 이동 경로 제어 기법

전 지 민◆, 조 혜 빈◆◆, 이 승 민*, 이 호 원°

Centralized DQN-Based Optimal Path Control of Multiple Robots

Jimin Jeon◆, Hyebin Cho◆◆, Seungmin Lee*, Howon Lee°

요 약

최근 4차 산업혁명 시대의 도래와 함께, 스마트 공장 환경에서 자동화/지능화 로봇의 활용에 대한 연구가 매우 활발히 이루어지고 있다. 하지만 기존의 무인 운반 차량 (automatic guided vehicle, AGV)은 단일화된 경로에 국한된 이동성을 가지기 때문에, 높은 인프라 구축 비용과 경로 수정의 어려움 등 많은 문제점을 가지고 있다. 이를 위해, 본 논문에서는 다중 로봇들이 서로 간의 충돌을 최소화하며, 각자의 목적지까지 최단 경로로 이동할 수 있도록 중앙집중형 심층 큐 네트워크 (centralized deep Q-network, DQN) 기반 물류 로봇 최적 이동 경로 제어 방안을 제안한다. 구체적으로, 현실의 간이 공장 물류 창고와 유사한 시뮬레이션 환경을 구축하고, 해당 환경에서 물류 로봇들에 대한 오프라인 학습을 수행한다. 학습된 모델을 이용하여, 스마트 공장 환경에서 물류 로봇들을 각자의 최적의 경로로 다중 목적지에 도착하도록 실시간으로 로봇을 제어하는 데에 활용함으로써 그 성능의 우수성을 보인다. 하드웨어로 구현하고 이를 기반으로 한 테스트를 통해 다중 물류 로봇이 0.1% 미만의 낮은 충돌율을 가지면서 각자의 다중 목적지에 최적의 경로로 도착함을 보인다.

키워드 : 최적 경로 제어, 심층 큐 네트워크, 스마트 공장
Key Words : Optimal path control, deep Q-network, smart factory

ABSTRACT

With the recent advent of the 4th Industrial Revolution era, research on the use of autonomous and intelligent robots in a smart factory has been actively investigated. However, existing automatic guided vehicles (AGVs) still have many problems, such as the cost of infrastructure installation and difficulty in route modification, because of their significantly limited mobility. Therefore, we herein propose an optimal logistics robot control scheme based on a centralized deep Q-network (DQN) so that multiple robots can minimize collisions and get to their desired destinations in the shortest path. Specifically, a simulation environment similar to a simplified factory logistics warehouse is considered, and offline reinforcement learning of logistics robots is conducted in this environment. In addition, we demonstrate the performance excellency of the proposed DQN model so that the robots can arrive at their destinations by the optimal routes. Through the test based on hardware implementation, we show that multiple logistics robots can arrive at their desired destinations along optimal paths while minimizing collisions.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지원(IITP- 2020-0-01741, 지역지능화혁신인재양성(Grand ICT연구센터)사업, 50%)과 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2022R1A2C1010602, 50%).
◆ First Author : School of Electronic and Electrical Eng., Hankyong National University, jimin0516@hknu.ac.kr, 학생회원
◆◆ Co-First Author : School of Electronic and Electrical Eng., Hankyong National University, chhb1228@hknu.ac.kr, 학생회원
* School of Electronic and Electrical Eng., Hankyong National University
° Corresponding Author : Hankyong University Department of Electronic and Electrical Engineering, hwlee@hknu.ac.kr, 종신회원
논문번호 : KICS202208-177-C-RN, Received August 13 2022; Revised October 17, 2022; Accepted October 19, 2022

Ⅰ. 서 론

코로나19 팬데믹 이후, 이커머스 시장이 급격하게 성장하고 있으며[1], 이에 따라 늘어난 많은 양의 물류들을 현재의 시스템만으로는 지원하기가 어려워지면서 물류 자동화 및 지능화에 대한 필요성이 심각하게 대두되고 있다[2]. 현재의 물류 시스템에서는 주로 지정된 경로를 따라 이동하는 센서 기반의 물류 이송 장치인 무인 운반 차량 (automatic guided vehicle, AVG)이 사용되고 있다. 그러나 AGV는 경로를 설정하기 위해 부품들을 생산설비 전반에 포설해야 하고 설치 기간 동안 생산설비를 사용하지 못해 이에 따른 많은 기회비용이 발생하게 된다. 또한 한 번 설비 완료 이후 공정 및 경로 변경 시 대대적인 설비의 수정이 필요하게 된다. 이러한 문제점들을 해결하기 위해서 최근에는 지능형 AGV에 대한 연구가 활발히 이루어지고 있다[3,4]

본 논문에서는 중앙집중형 심층 큐 네트워크 (centralized deep Q-network, DQN)를 이용하여 다중 로봇들이 최적의 경로로 이동할 수 있도록 하는 최적 경로 제어 기법을 제안한다. 특히, 로봇별로 서로 다른 목적지가 존재하는 스마트 공장 환경에서 학습된 모델을 기반으로 다중 로봇들이 각자의 목적지로 이동하는 경우에 로봇 간 낮은 충돌율을 가지는 동시에 최적의 이동 경로로 목적지에 도달할 수 있음을 보인다.

[7]에서는 로봇의 현재 위치에 대한 최적의 이동 경로를 도출하여 상황 변화에 능동적이고 유연하게 대처하는 DQN 모델을 제안하였다. [8]에서는 드론 택시 시나리오에서 다중 에이전트 심층 강화학습을 통해 이동 경로를 최적화하는 연구를 진행하였으며, [9]에서는 수중 로봇이 복잡하고 예측 불가능한 환경에서 장애물을 회피하는 N-step Priority Double DQN 기반 경로 계획 알고리즘을 제안하였다. 또한, [10]에서는 단안 카메라가 장착된 UAV 쿼드로터의 DQN 기반 장애물 회피 기법을 제안하였다. 이에 본 연구에서는 기존 연구들에서 개별적으로 고려되었던 장애물 회피 기법과 최적 경로 설정 기법을 동시에 고려하여, 스마트 공장 환경에서 로봇들이 서로 충돌하지 않고 최적 경로를 찾을 수 있도록 한다.

본 논문의 Ⅱ장에서는 강화학습 기반 물류 인식 로봇 제어 방안의 시스템 모델을 제시하며, Ⅲ장에서는 다중 로봇이 다중 목적지로 최적으로 이동하는 중앙집중형 심층 큐 네트워크 알고리즘을 제안한다. Ⅳ장에서는 시뮬레이션을 통해 제안 방안의 성능을 검증하며, 마지막으로, Ⅴ장에서 결론을 맺는다.

Ⅱ. 시스템 모델

본 논문에서는, 중앙집중형 DQN 알고리즘의 적용을 통해 다중 로봇들이 개별 목적지까지 이동하는 최적 경로를 찾기 위한 기법을 제시하고, 하드웨어 구현을 통해 제안 방안의 현실적 적용 가능성을 검증한다. 그림 1은 제안 방안의 시스템 모델을 나타낸다.

그림 1(a)는 다중 로봇들의 최적 이동 경로 제어를 위한 DQN의 학습과정을 나타내며, 이 단계를 통해 학습된 모델은 검증 단계와 테스트 단계에서 사용된다. 또한, 그림 1(b)는 하드웨어에서의 테스트 단계를 표현한다. 구체적으로, 공장 내부에 설치된 카메라는 촬영된 사진을 서버로 보내고, 서버는 영상 처리를 통해 촬영된 사진으로부터 각 로봇의 현재 상태 $S(t)$를 파악한다. 그 후 사전에 학습된 심층 신경망 (deep neural network, DNN) 모델에 $S(t)$를 입력하여 모든 행동에 대한 Q-value를 반환받으며, epsilon-greedy 정책에 의해 각 로봇에 대한 행동 $A(t)$를 결정한다. 서버는 DNN 모델로부터 도출한 $A(t)$를 로봇에게 전달한다. 이후, 서버의 데이터를 수신한 로봇들은 $A(t)$에 따라 $S(t+1)$로 움직이며, 이러한 동작은 에피소드가 끝날 때까지 반복된다. 이 때, 카메라-서버간 통신은 정적인 환경에서 이루어지므로 통신 속도가 더 빠른 유선 통신을 사용하며, 이동성으로 인해 동적 환경에서 작동하는 로봇과 서버간의 통신은 Wi-Fi 기반 무선 통신을 이용한다.

테스트 단계에서 공장 내부 촬영은 Raspberry Pi에 연결된 카메라 모듈을 사용하며, 촬영된 사진은 서버로 전송된다. 서버는 사진으로부터 각 로봇의 현재 위치를 반환하기 위해 다음과 같은 영상 처리 과정을 진행한다.

1. 카메라 모듈로 찍은 사진을 Hue, Saturation, Value (HSV) 이미지로 불러온다.
2. 위치에 따른 조도 변화와 상관없이 모든 격자 위치에서의 빨간색과 파란색을 추출할 수 있는 HSV 범위를 설정하고, 객체를 구별하기 위해 빨간색과 파란색을 따로 추출하여 두개의 이미지를 생성한다.
3. 각각의 HSV 이미지를 그레이 스케일로 변환한 후, 다시 이진 이미지로 변환하여 객체 외곽선 검출을 유용하게 한다.
4. 객체의 외곽선을 검출하여 변환된 이미지에서 객체를 인식하고 각 로봇의 $S(t)$를 도출한다.

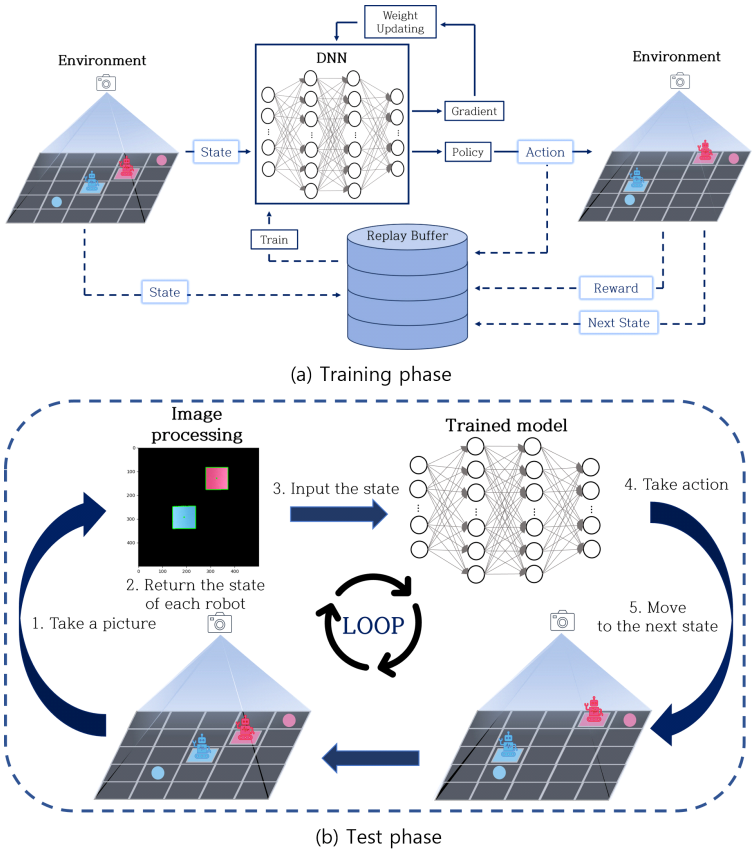


그림 1. 시스템 모델 (학습 단계 & 테스트 단계)
Fig. 1. System model (training phase & Test phase).

Ⅲ. 중앙집중형 DQN 기반 다중 로봇의 최적 이동 경로 제어 기법

본 논문에서 고려하고 있는 문제를 Markov Decision Process(MDP)로 정의하며, 다중 로봇들을 중앙 제어하는 서버가 에이전트로서 역할을 수행한다.

• State

제안 방안의 상태는 $X \times Y$ 그리드 환경에서 N개 로봇들의 정규화된 2차원 좌표로 정의된다. 타임스텝 t에서, 상태 $S(t)$는 다음과 같이 정의된다.

$$S(t) = [s_1(t), s_2(t), \cdots, s_N(t)], \quad (1)$$

$$s_i \in \{0 \le x_i \le X, 0 \le y_i \le Y\} \in \mathbb{Z}^2, \forall i, \quad (2)$$

수식 (2)에서, s_i는 에이전트 i의 2차원 위치 정보를 나타내며, x_i, y_i는 각각 s_i의 x축 좌표와 y축 좌표를 나타낸다. 또한, $\mathbb{Z}$는 정수 집합을 나타낸다.

• Action

에이전트는 각 로봇의 이동성을 제어하기 때문에 타임스텝 t에서, 에이전트가 선택한 행동 $A(t)$은 다음과 같이 정의된다.

$$A(t) = [a_1(t), a_2(t), \cdots, a_N(t)], \quad (3)$$

$$a_i \in \{\Delta x_i, -\Delta x_i, \Delta y_i, -\Delta y_i\}, \forall i, \quad (4)$$

수식 (4)에서, Δx_i, $-\Delta x_i$, Δy_i와 $-\Delta y_i$는 각각 로봇 i의 우측 이동, 좌측 이동, 상향 이동, 하향 이동을 나타낸다.

• Reward

제안 방안의 보상은 다음의 두 가지 목표를 고려하여 설계한다.

1. 각 로봇들이 각자의 목적지까지 최단경로로 이동하도록 한다.
2. 목적지까지 이동하는 동안 다른 로봇과의 충돌이 없도록 한다.

Algorithm 1: DQN Training Algorithm for Optimal Path Control

Initialize $S(\bullet)$, $Q(\bullet)$, and $\epsilon(\bullet)$;
Set randomly each robot's initial destination, $g_i(\bullet)$ for all i;
for episode $E = 1,2, \ldots$ **do**
  $\epsilon(E) = \epsilon(E-1) \times \epsilon_{decay}$;
  for $t = 1,2, \ldots$ **do**
    Select an action $A(t)$ considering $S(t)$ based on decayed $\epsilon-$greedy method $\epsilon(t)$;
    Move all Robots to the next state $S(t+1)$ by $A(t)$;
    Calculate a reward $R(t)$ according to Eq. (5);
    Store transition pair $(S(t), A(t), R(t), S(t+1))$ in replay buffer D;
    Select B batch samples of transitions $(S(t), A(t), R(t), S(t+1))$ randomly from D;
    Update Q-function as follows:
    $Q(S(t), A(t)) \leftarrow (1-\alpha) Q(S(t), A(t)) + \alpha(R(t) + \beta \times \max_{A(t+1) \in \mathbb{A}} Q(S(t+1), :)$;
    if $s_i(t+1) = g_i(t)$ **then**
      Randomly reset $g_i(t)$;
    end if
  end for
end for

따라서, 각 에이전트 i의 현재 위치$(x_i(t), y_i(t))$에서 각자의 목적지(x_i^{goal}, y_i^{goal})까지의 맨해튼 거리$(d_i(t))$를 계산하고, $d_i(t)$와 $d_i(t-1)$를 비교하여 현재 타임스텝 t에서 목적지까지 거리가 감소한 로봇의 개수를 $n_i(t)$라 정의한다. $n_i(t)$를 이용하여 다음과 같이 보상을 정의한다.

$$R(t) = \begin{cases} \dfrac{2n_i(t)}{N} - 1, & 0 \le n_i \le N \in \mathbb{Z}^2, \forall i \\ C, & \text{if } collision\ happens. \end{cases} \quad (5)$$

수식 (5)에서, 모든 에이전트가 목적지까지의 거리가 감소하도록 이동하는 경우에 가장 좋은 보상인 1의 값을 받으며, 서로 간의 충돌이 발생하였을 때 가장 나쁜 보상인 C의 값을 받도록 설정한다. 이 때, C는 충돌이 발생하였을 때의 보상이며, 에이전트가 충돌 회피를 잘 학습하도록 충돌이 일어나지 않았을 경우의 최소 보상인 -1보다 작은 값으로 C를 설정한다. 이러한 보상의 설정은 모든 에이전트가 서로 간의 충돌을 최소화하며 목적지까지 최소한의 타임스텝으로 이동할 수 있도록 한다. 알고리즘 1은 제안된 중앙집중형 DQN 기반 다중 로봇의 최적 알고리즘의 학습 단계를 보여준다.

학습 단계와 검증 단계는 로봇이 2개인 환경과 3개인 환경에 대해 진행하였다. 두 환경에서 학습에 이용된 파라미터와 신경망 모델은 각각 표 1, 2에 요약되어 있으며, 표 1에서 ϵ_{init}은 초기 epsilon 값을 의미한다. 또한, 학습 단계에서 한 번의 에피소드는 8000번의 타임스텝으로 구성된다.

학습에 대한 평가를 위해 에피소드마다 측정되는 다음의 3가지 성능 지표를 고려하였다. 첫째, 에피소드에 따른 평균 보상이며, 이는 학습의 수렴 과정을 나타낸다. 두 번째 성능 지표는 한 에피소드동안 모든 로봇이 목적지에 도달한 횟수의 합이며, 이는 본 논문에서의 목표 중 하나인 로봇의 최단 이동에 대한 학습 과정을 나타낸다. 마지막은 한 에피소드동안 로봇 간의 충돌이 발생한 누적 횟수이며, 이는 에피소드가 지남에 따라 모델이 충돌 회피를 학습하는 과정을 보여준다.

그림 2와 그림 3은 각각 로봇이 2개인 환경과 로봇이 3개인 환경에서 진행된 학습 단계에서의 3가지 성능지표를 보여준다. 로봇이 2개인 환경은 10×10의 그리드에서 충돌 보상 C를 -10으로 설정하여 학습을 진행하였으며, 로봇이 3개인 환경에서는 10×10 그리드에서 C를 -5로 설정하여 학습을 진행하였다. 그림 2와 그림 3의 (a)는 학습 단계에서 에피소드에 따른 평균 보상을 나타내며, 이를 통해 로봇이 2개인 환경과 로봇이 3개인 환경 모두 학습 단계에서 모델의 학습이

표 1. 학습 파라미터
Table 1. Training parameter.

Parameter	Value	Parameter	Value
# of Robots(N)	2, 3	$X \times Y$	10×10
Epsilon	1.0	Epsilon Decay	0.997
Discount Factor	0.95	Replay Buffer Size	50000
Batch Size	128	Learning Rate	0.0001
Episode	1000	Timestep	8000

표 2. 신경망 구조
Table 2. Neural network structure.

DNN			
Layer		Node	Rate
Input	FC Layer	8	-
Hidden	FC Layer	350	-
	Dropout	-	0.2
	FC Layer	250	-
	Dropout	-	0.2
Output	FC Layer	16	-
Optimizer		Adam	
Activaion Function		Relu	

수렴함을 알 수 있다. 본 논문에서의 목표 중 하나는 각 로봇이 각자의 목적지까지 최단 시간에 도달하는 것이다. 즉, 이는 제한된 타임스텝동안 최대한 많은 목적지로의 도달을 달성하는 것과 같다. 그림 2(b)와 그림 3(b)는 학습 단계에서 에피소드가 진행됨에 따라 각각 2개의 로봇과 3개의 로봇이 목적지에 도달한 누적 횟수가 증가함을 보여주며, 제한된 시간 동안 목적지에 도달하는 횟수의 증가는 각 로봇이 목적지까지 더 빠르게 이동하고 있음을 보여준다. 그림 2(c)와 그림 3(c)는 로봇이 2개인 환경과 3개인 환경에서 에피소드에 따라 로봇의 충돌 횟수가 감소함을 보여준다. 이를 통해 두 환경에서의 로봇이 충돌 회피를 잘 학습하고 있음을 확인하였다.

Ⅳ. 실험 결과 및 분석

4.1 검증 단계 결과

제안 방안의 목적은 각 로봇이 목적지까지 최단 경로로 이동하는 것과 서로 간의 충돌을 최소화하는 것이다. 본 논문에서는 DQN 모델의 최적 경로 학습 여부를 확인하기 위해 검증 단계를 진행한다. 검증 단계에서의 에피소드는 학습 단계에서와 다르게, 각 로봇이

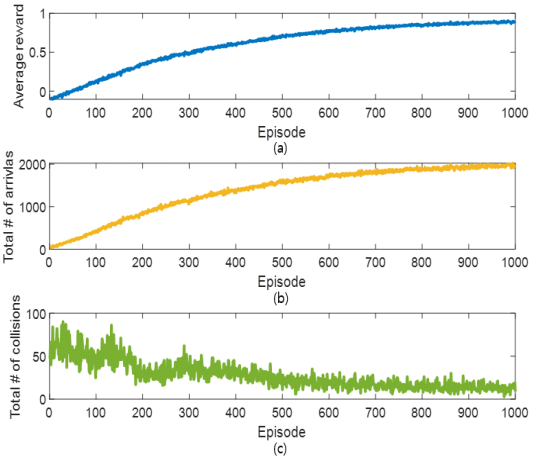


그림 2. 로봇이 2개인 환경에서의 학습 결과
Fig. 2. Training results when 2 robots exist.

각자의 목적지까지 도달하는 것으로 정의된다. 이 단계는 100,000번의 에피소드로 진행되었으며, 사용된 신경망 구조는 학습 단계에서와 같다. 또한, 각 로봇의 초기 상태와 목적지는 랜덤하게 발생한다.

검증 단계에서의 자세한 평가를 위해 에피소드마다 측정되는 오버스텝 횟수와 충돌 횟수를 성능 평가 지표로 고려하였다. 먼저, 모델에 의한 경로가 최단 경로인지를 확인하기 위해 맨해튼 거리와 DQN 모델에 의한 경로의 차를 오버스텝이라 정의한다. 이 때, 맨헤튼 거리는 타임스텝 t=1에서 각 로봇의 초기 위치와 목적지 사이의 최단 경로이며, 이 때 각 로봇 간의 충돌과 회피를 위한 우회를 고려하지 않는다. 또한, DQN 모델에 의한 경로는 로봇이 DQN에 의해 목적지까지 이동할 때까지 소요된 타임스텝 수로 표현된다. 두 번째 성능 지표는 검증 단계에서 에피소드당 발생하는 충돌 횟수이다. DQN 모델이 서로간의 충돌을 최소화하며 이동하는지 확인하기 위해 검증 단계에서의 충돌 횟수를 관찰하였다.

2개의 로봇의 경우, 에피소드 평균 오버스텝 횟수는 5×5 그리드 환경에서 충돌 보상이 -5인 경우 0.1번, 8×8 그리드 환경에서 충돌 보상이 -10일 때 0.35번으로 발생하였다. 또한 3개 로봇의 환경에서 에피소드 평균 오버스텝 횟수는 10×10 그리드 환경에서 -10의 충돌 보상을 가질 때 1.9번 발생하였음을 확인하였다. 이러한 결과가 충돌과 회피를 위한 우회를 고려하지 않은 맨해튼 거리와의 비교임을 고려하면, DQN 모델이 학습한 경로가 최단 경로와 매우 근접함을 알 수 있다. 반면, 100,000의 에피소드에서 충돌은 로봇이 2개인 환경에서 그리드 크기가 5×5 그리드 크기와 -

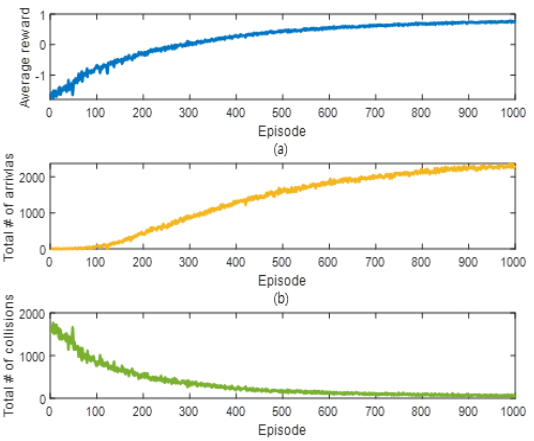
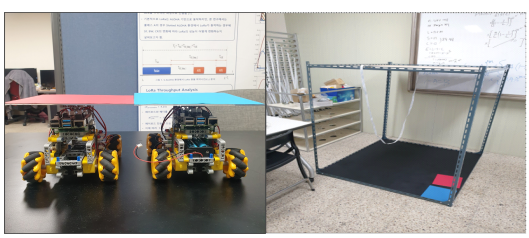


그림 3. 로봇이 3개인 환경에서의 학습 결과
Fig. 3. Training results when 3 robots exist.

5의 충돌 보상을 가질 때, 0.0103%의 확률로, 그리드 크기가 8×8이고 충돌 보상이 -10일 때 0.00176%의 확률로 발생하는 것을 확인하였다. 또한, 3개의 로봇 환경에서는 10×10 그리드 크기와 -10의 충돌 보상을 가질 때 0.05248%의 확률로 충돌이 발생하였다. 이러한 충돌 확률은 매우 적은 값이며, 이를 통해 학습된 DQN 모델은 다중 로봇이 최단 경로로 이동하도록 하는 것뿐만 아니라, 서로 간의 충돌 또한 최소화시키는 것 또한 알 수 있다.

4.2 하드웨어 구현 및 테스트 결과

학습된 모델의 성능을 입증하기 위해 현실의 간이 공장 물류 창고와 유사한 시뮬레이션 환경을 구축하였다. 그림 4는 구축한 시뮬레이션 환경을 보여준다. 그림 4(a)의 빨간색과 파란색의 종이가 붙여진 무선 조종 자동차들은 Raspberry Pi에 연결되며, 각 로봇의 역할을 수행한다. 그림 4(b)는 간이 공장 모형을 구현한 것이며, 간이 공장 모형은 밑판의 가로, 세로가 115cm, 높이가 135cm로 구성된다.



(a) Mobile robot (b) Simple factory model

그림 4.하드웨어 구현
Fig 4. Hardware implementation.

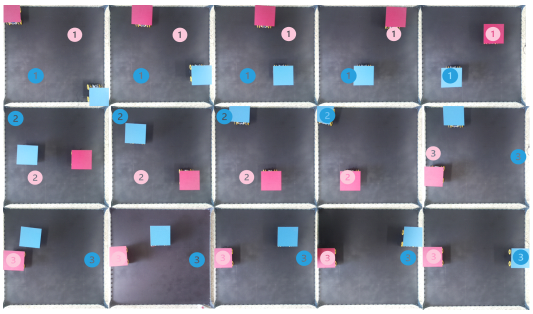


그림 5. 하드웨어 테스트 결과
Fig. 5. Hardware test results

테스트 단계는 학습 단계, 검증 단계와는 달리 그리드의 X×Y 크기가 5×5이며, 로봇의 수가 2개, 각 로봇별 목적지의 수가 3개인 환경에서 진행되었다. 이는, 본 논문에서 제안하는 DQN 모델이 다중 목적지 시스템으로 확장 가능함을 증명한다.

그림 5는 하드웨어로 구현한 실험 결과이며, 하나의 스냅샷은 매 타임스텝마다 로봇의 현재 위치를 나타낸다. 이를 통해 로봇의 이동 경로를 알 수 있다. 테스트 단계에서 각 로봇은 기본적으로 랜덤한 초기상태와 각 3개의 목적지를 가진다. 그림 5는 그 중 첫 번째 로봇(빨간색 로봇)의 초기상태가 (0,0), 목적지가 (3,1), (3,3), (0,2)이며, 두 번째 로봇(파란색 로봇)의 초기상태가 (4,4), 목적지가 (1,3), (0,0), (4,2)인 환경에서의 각 타임스텝마다의 이동을 보여준다. 이를 통해 최종적으로 현실의 간이 공장 물류 창고 환경에서 로봇이 각자의 목적지에 최적의 경로로 도달함을 알 수 있다.

V. 결 론

본 연구는 로봇들이 기존의 AGV와 같이 단일화된 경로에 국한되어 이동하지 않고 주어진 환경에서 최적 경로로 이동하게 하는 것을 목표로 한다. 구체적으로, 각 로봇 사이의 충돌을 최소화하면서, 각 로봇이 각자의 목적지까지 최적의 경로로 이동하는 문제를 포함한다. 이를 위해, 중앙집중형 DQN를 이용하여 다중 로봇의 최적 이동 경로 제어 모델을 학습한다. 시뮬레이션을 통해 다중 목적지 환경에서 0.1% 미만의 낮은 충돌율로 이동함을 확인하여, 제안 방안의 다중 목적지 시스템으로의 확장 가능성을 확인하였다. 또한 하드웨어를 기반으로 현실의 간이 공장 물류 창고 환경을 구현하여, 로봇이 각자의 목적지까지 최적의 경로로 이동함을 보였다.

제안 방안을 통해 기존 공장 자동화를 위한 시스템

인 AGV에서 요구되었던 많은 설치 부품들과 비용, 그리고 인력을 절감할 수 있으며, 물류 작업 속도가 향상될 것으로 기대된다. 또한, 본 시스템은 공장 자동화 분야뿐만 아니라, 드론, 통신, 항공, 자율 주행 같은 분분야에서도 이동 경로 최적화 측면에서 활용 가능하다. 하지만, 본 제안 방안이 중앙에서 모든 로봇을 제어하는 중앙집중형 DQN이기 때문에 로봇의 수가 증가함에 따라 액션 공간이 기하급수적으로 증가하므로, 추가적인 로봇의 증설이 어려울 수 있다. 이에 따라 분산적이고, 자율적인 강화학습 기반 로봇 제어 기법에 대한 연구를 수행하고자 한다.

References

[1] S. K. Lim, Y. E. Oh, and K. S. Park, "A study on the difference in consumers' perception of the e-commerce utilization factors according to COVID-19: Focused on fresh food," *The e-Business Stud.*, vol. 23, no 1, pp. 75-94, Feb. 2022. (https://doi.org/10.20462/tebs.2022.2.23.1.75)

[2] Grand View Research, *Industrial Automation And Control Systems Market Size, Share & Trends Analysis Report By Component (Industrial Robots, Control Valves), By Control System (DCS, PLC, SCADA), By Vertical, By Region, And Segment Forecasts, 2021-2028* (2021), Retrieved Oct. 10. 2021, from https://www.grandviewresearch.com/industry-analysis/industrial-automation-market

[3] J. Jia, W. Chen, and Y. Xi, "Design and implementation of an open autonomous mobile robot system," in *IEEE ICRA '04*, pp. 1726-1731, New Orleans, LA, USA, Apr. 2004. (https://doi.org/10.1109/ROBOT.2004.1308073)

[4] H. E. Stephanou, "Advanced automation in manufacturing and service industries," in *IEEE Int. Conf. Robotics and Automat.*, vol. 3, p. 3166, Nagoya, Japan, May 1995. (https://doi.org/10.1109/ROBOT.1995.525739)

[5] J. Liu, H. Lee, and H. Jin, "Multichannel S-ALOHA enabled autonomous self-healing in industrial IoT networks," *IEEE Trans. Ind. Informatics,* vol. 18, no. 12, pp. 8576-8585, Feb. 2022. (https://doi.org/10.1109/TII.2022.3149908)

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," in *NIPS Deep Learn. Wkshp. 2013*, Lake Tahoe, USA, Dec. 2013. (https://doi.org/10.48550/arxiv.1312.5602)

[7] K.-S. Park, J.-M. Park, W.-K. Yun, and S.-J. Yoo, "DQN reinforcement learning: The Robot's OptimumPath navigation in dynamic environments for smart factory," *J. KICS*, vol. 44, no. 12, pp. 2269-2279, Dec. 2019. (https://doi.org/10.7840/kics.2019.44.12.2269)

[8] W. J. Yun, S. Jung, J. Kim, and J.-H. Kim, "Optimal drone taxi path planning via multi agent deep reinforcement learning," in *Proc. Symp. KICS,* pp. 63-64, Nov. 2020.

[9] J. Yang, J. Ni, M. Xi, J. Wen, and Y. Li, "Intelligent path planning of underwater robot based on reinforcement learning," *IEEE Trans. Automat. Sci. and Eng.,* Jul. 2022. (https://doi.org/10.1109/TASE.2022.3190901)

[10] A. Singla, S. Padakandla, and S. Bhatnagar, "Memory-based deep reinforcement learning for obstacle avoidance in UAV with limited environment knowledge," *IEEE Trans. Intell. Transport. Syst.*, vol. 22, no. 1, pp. 107-118, Jan. 2021. (https://doi.org/10.1109/TITS.2019.2954952)

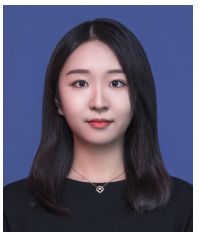
전 지 민 (Jimin Jeon)

2023년 2월 : 한경국립대학교 전기전자제어공학과 졸업

2023년 3월~현재 : 한경국립대학교 전자전기공학부 석사과정

<관심분야> 6G 무선 통신, 드론 통신, 강화학습기반 무선 자원 관리

조 혜 빈 (Hyebin Cho)



2023년 2월 : 한경국립대학교 전기전자제어공학과 졸업
2023년 3월~현재 : 한경국립대학교 전자전기공학부 석사과정
<관심분야> 6G 무선 통신, 위성 통신, 강화학습기반 무선자원관리

이 승 민 (Seungmin Lee)

2021년 2월 : 한경국립대학교 전기전자제어공학과 졸업
2023년 2월 : 한경국립대학교 전자전기공학부 석사
<관심분야> B5G/6G 모바일 네트워크, 무선자원관리, 고밀집 분산 네트워크, 강화학습기반 UAV 네트워킹

이 호 원 (Howon Lee)

2003년 2월 : KAIST 전자전산학과 졸업
2009년 8월 : KAIST 전기 및 전자공학과 박사 (석박사통합)
2009년 6월~2012년 2월 : KAIST ITC 연구조교수/팀장
2012 3월~2021년 2월 : KAIST 겸직교수
2012년 3월~2021년 2월 : KAIST 겸직교수
2012년 3월~현재 : 한경국립대학교 전자전기공학부 전자공학전공 교수
<관심분야> 5G/6G 모바일 네트워크, 무선자원관리, 드론 통신, 머신러닝기반 통신 네트워크
[ORCID:0000-0001-5509-9202]