Multi-Agent Reinforcement Learning for Control of Heterogeneous AMRs in Smart Factories

Jongin Lee[†] and Soyi Jung[‡]
[†]*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
[‡]*Dept. Electrical and Computer Engineering, Ajou University, Suwon, 16499, South Korea*
{jongcgs, sjung}@ajou.ac.kr

***Abstract*—The advancement of smart factories and multi-stage logistics has accelerated the need for efficient and adaptive coordination of autonomous mobile robots (AMRs). While most existing research has focused on homogeneous automated guided vehicles (AGVs) systems, the control of heterogeneous AMRs remains underexplored. This paper proposes a multi-agent reinforcement learning (MARL)-based control framework tailored for heterogeneous AMRs in smart factory environments. The framework incorporates AMR-specific characteristics such as field of view (FOV) into the observation space and models decision-making using a multi-discrete action structure with action masking to ensure feasibility. The reward function promotes efficient task execution by encouraging successful pickup, delivery, and goal-directed movement. Simulation results demonstrate that the proposed approach achieves stable learning, improves delivery completion rates, and reduces task execution time, validating its effectiveness in heterogeneous and complex factory settings.**

***Index Terms*—Autonomous mobile robots, heterogeneous robot systems, smart factory, multi-agent reinforcement learning.**

I. Introduction

In modern manufacturing, the expansion of smart factories and the growth of multi-stage logistics have increased the demand for adaptability in diverse operational settings. Autonomous mobile robots (AMRs), offering greater flexibility in navigation compared to automated guided vehicles (AGVs), have emerged as a suitable solution for such complex environments.

Path planning has been a central topic in multi-agent systems [1]. Recent studies have applied reinforcement learning to multi-AGV control [2], [3], yet most assume homogeneous AGVs. In practice, smart factories demand coordination among heterogeneous AMRs with varying sizes, speeds, and task volumes. However, research in this direction remains insufficient. To address this, this paper proposes a multi-agent reinforcement learning (MARL)-based control framework for heterogeneous AMRs in smart factory environments. The proposed approach explicitly incorporates AMR-specific features such as field of view (FOV) into the control process, while leveraging a multi-discrete action space to represent the complexity of decision-making [4]. This design enables the framework to more accurately reflect AMR characteristics and enhance coordination performance in heterogeneous settings.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00359330)
(Corresponding authors: Soyi Jung)

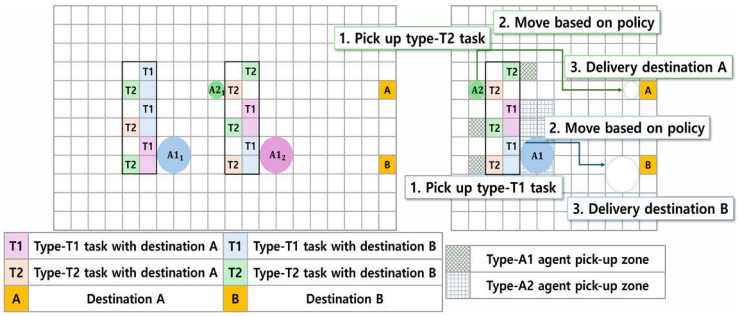


Fig. 1: Smart factory simulation initial environment and scenario example.

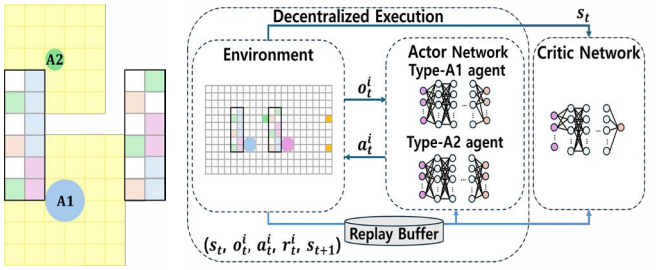
II. System Model

In this study, we consider a 12×20 grid-based smart factory environment as illustrated in Fig. 1. The agent set is defined as $\mathcal{N} = \{n_i \mid i = 1,2,\ldots,N_a\}$ with $N_a = 3$, consisting of two type-A1 agents ($A1_1$, $A1_2$) and one type-A2 agent ($A2_1$). The task set is given as $\mathcal{T} = \{t_j \mid j = 1,2,\ldots,N_t\}$ with $N_t = 12$, where type-T1 tasks (size 2×1) can only be executed by type-A1 agents, while type-T2 tasks (size 1×1) can be executed by both type-A1 and type-A2 agents. Each task has a designated destination $\mathcal{D} = \{A, B\}$, and agents transport tasks from predefined pick-up zones to their destinations.

The global state at time t is defined as $\mathbf{s}_t = \{f_0, f_1, \ldots, f_{N_s-1}\}$, where each feature plane f_k encodes agent positions, type-T1 and type-T2 tasks with statuses pending or assigned, and destinations A and B. This global state $\mathbf{s}_t$, with $N_s = 14$ feature planes, is provided to the critic network for centralized training.

The local observation of agent i is $\mathbf{o}_t^i = \{f_0^i, f_1^i, \ldots, f_{N_o-1}^i\}$, where each f_k^i represents the same categories of information restricted to its field of view (Fig. 2(a)). The $N_o = 13$ feature planes include the agent's position, movable cells, nearby agents, and task and destination locations. Type-A1 and Type-A2 agents observe 6×6 and 5×5 windows, respectively, and perform decentralized execution, while the critic utilizes $\mathbf{s}_t$ (Fig. 2(b)).

The action space is formulated as a multi-discrete structure. The action of agent i at time t is denoted by $a_t^i = (a_t^{i,1}, a_t^{i,2}, \ldots, a_t^{i,m_i})$, where $a_t^{i,k}$ represents the k-th action com-



(a) (b)

Fig. 2: Local observation size and framework. (a) AMRs with FOV. (b) Heterogeneous MARL framework in smart factory.

TABLE I: Type-A1 agent multi-discrete action.

Component	Range	Description
Action type	[0–4]	Stop, Move 1 cell, Pick up, Deliver, Queue
Move 1 cell direction	[0–3]	East, West, South, North
Pickup type	[0–1]	Pick up type-T1 task, type-T2 task
Delivery type	[0–1]	Deliver to destination A or B
Queue direction	[0–4]	Stop, East, West, South, North

TABLE II: Type-A2 agent multi-discrete action.

Component	Range	Description
Action type	[0–5]	Stop, Move 1 cell, Move 2 cells, Pick up, Deliver, Queue
Move 1 cell direction	[0–3]	East, West, South, North
Move 2 cell direction	[0–3]	East, West, South, North
Pickup type	[0–1]	Pick up type-T1 task or type-T2 task
Delivery type	[0–1]	Deliver to destination A or B
Queue direction	[0–4]	Stop, East, West, South, North

ponent and m_i denotes the total number of components for agent i. For type-A1 agents, the action space is defined as $m_i = 5$ with dimensions $[5,4,2,2,5]$, while for type-A2 agents it is $m_i = 6$ with $[6,4,4,2,2,5]$, as shown in Table I and Table II. To ensure feasibility, action masking restricts type-A1 agents to prioritize type-T1 task, whereas type-A2 agents are limited to type-T2 task.

The reward is defined as follows:

$$R(t) = R^{\text{pickup}}(t) + R^{\text{delivery}}(t) + R^{\text{move}}(t) \tag{1}$$

$$R^x(t) = \begin{cases} r_x^{A1}(t) - \beta\Delta t & \text{if a type-A1 agent succeeds in } x, \\ r_x^{A2}(t) - \beta\Delta t & \text{if a type-A2 agent succeeds in } x, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$R^{\text{move}}(t) = \begin{cases} \gamma & \text{if the agent moves closer to pickup/delivery,} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

where $R(t)$ is the immediate reward at time t, composed of pickup, delivery, and movement terms. For $x \in \{\text{pickup}, \text{delivery}\}$, $R^x(t)$ provides success rewards $r_x^{A1}(t)$ or $r_x^{A2}(t)$ depending on the agent type, penalized by elapsed time Δt with coefficient $\beta > 0$. In pickup, Δt denotes waiting time until success; in delivery, it represents duration from pickup to completion. $R^{\text{move}}(t)$ grants a shaping reward $\gamma \geq 0$ when the agent approaches its target.

III. Performance Evaluation

The proposed heterogeneous AMR control model was trained over 500 episodes. As shown in Fig. 3, both episode and team rewards exhibited a gradual upward trend, indicating stable policy optimization. From around the 400th episode, all tasks were consistently completed, as confirmed by the convergence of the delivery completion rate. Moreover, the decreasing episode length shows that tasks were executed more efficiently as training progressed. These results demonstrate that the proposed MARL-based framework effectively derives robust control strategies in complex smart factory environments.

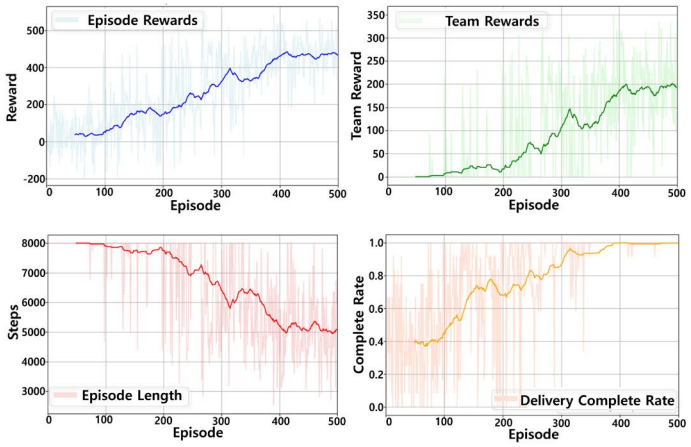


Fig. 3: Training performance results.

IV. Conclusion

This paper proposed a MARL-based control framework for heterogeneous AMRs in smart factory environments. By incorporating AMR-specific features such as FOV and adopting a multi-discrete action space, the framework was designed to better reflect the diverse characteristics of AMRs and enhance coordination performance. Experimental results confirmed that the proposed approach enables efficient task completion and improved adaptability in complex environments.

Future work will focus on extending the framework to more diverse factory layouts and operational conditions, as well as comparing its performance with alternative MARL algorithms to further validate its effectiveness.

References

[1] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-agent systems: A survey," *IEEE Access*, vol. 6, pp. 28 573–28 593, 2018.

[2] H.-B. Choi, J.-B. Kim, C.-H. Ji, U. Ihsan, Y.-H. Han, S.-W. Oh, K.-H. Kim, and C.-S. Pyo, "MARL-based optimal route control in multi-AGV warehouses," in *2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2022, pp. 333–338.

[3] H. Lee and J. Jeong, "Mobile robot path optimization technique based on reinforcement learning algorithm in warehouse environment," *Applied Sciences*, vol. 11, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/3/1209

[4] S. Huang and S. Ontañón, "A closer look at invalid action masking in policy gradient algorithms," *CoRR*, vol. abs/2006.14171, 2020. [Online]. Available: https://arxiv.org/abs/2006.14171