

자연어 기반 로봇 제어 시스템 구현: 인형뽑기 기계 응용 사례

윤희종¹, 조현성¹, 윤재현¹, 김지환¹, 김태국^{2*}

¹국립부경대학교 컴퓨터·인공지능공학부 학생, ²국립부경대학교 컴퓨터·인공지능공학부 교수

Implementation of a Natural Language-Based Robot Control System: A Case Study on Claw Machine Application

Hee-Jong Yoon¹, Hyeon-Seong Jo¹, Jae-Hyeon Yoon¹, Ji-Hwan Kim¹, Tae-Kook Kim^{2*}

¹Student, Division of Computer and Artificial Intelligence Engineering, Pukyong National University

²Professor, Division of Computer and Artificial Intelligence Engineering, Pukyong National University

요약 본 논문은 자연어 명령을 통해 로봇을 제어하는 시스템을 설계하고 구현한 연구를 다룬다. 제안 시스템은 인형뽑기 기계를 대상으로, GPT-4 언어 모델, YOLOv5 객체 탐지, Raspberry Pi 제어 시스템을 통합하여 사용자 친화적인 인터페이스를 구축했다. 사용자가 자연어로 원하는 인형을 묘사하면, 시스템은 YOLOv5로 객체를 탐지 및 분석하고, GPT-4가 해당 정보를 해석하여 로봇 팔을 움직이는 정밀 제어 명령을 생성한다. 이 명령은 Raspberry Pi를 통해 인형뽑기 장치에 전달되어 작업을 수행한다. 또한, Server-Sent Events (SSE) 기술을 적용하여 객체 탐지 현황, 명령 실행 상태, 예외 상황 등을 사용자에게 실시간으로 알린다. 실험을 통해 사용자가 묘사한 대상 인형을 정확하게 인식하고 성공적으로 집어내는 과정을 검증했다. 본 연구는 복잡한 로봇 제어를 직관적인 자연어로 대체할 수 있는 가능성을 제시하며, 향후 제조 및 물류 현장에서의 효율적인 로봇 운영에 기여할 수 있을 것으로 기대된다.

주제어 : 사물인터넷, YOLO(You Only Look Once), 라즈베리 파이, 자연어, 로봇, 인공지능

Abstract This paper presents the design and implementation of a robot control system based on natural language commands. Using a claw machine as a testbed, the proposed system integrates the GPT-4 language model, YOLOv5 object detection, and a Raspberry Pi-based control system to establish an intuitive user interface. When a user describes a target object in natural language, the system performs detection and analysis via YOLOv5, while GPT-4 interprets the results to generate precise control commands for the robotic arm. These commands are then transmitted to the claw mechanism through the Raspberry Pi to execute the task. Furthermore, Server-Sent Events (SSE) technology is implemented to provide real-time updates, including object detection status, execution progress, and exception handling. Experimental results verify that the system accurately identifies and successfully retrieves the user-described targets. This study demonstrates the potential of replacing complex robot control with intuitive natural language interaction, contributing to more efficient robotic operations in future manufacturing and logistics sectors.

Key Words : Internet of Things (IoT), YOLO(You Only Look Once), Raspberry Pi, Natural language, Robot, Artificial intelligence (AI)

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2023-00242528).

*교신저자 : 김태국(king@pknu.ac.kr)

접수일 2025년 12월 22일 수정일 2026년 01월 15일 심사완료일 2026년 02월 03일

1. 서론

최근 산업 현장에서는 자동화 기술의 발전과 함께 로봇 활용이 빠르게 확산되고 있다. 특히 제조업과 물류 분야에서는 반복적이고 정밀한 작업을 수행할 수 있는 로봇이 인력을 대체하며, 산업 구조의 혁신적인 변화를 주도하고 있다. 한국의 공장 자동화 시장 또한 이러한 흐름 속에서 급속히 성장하고 있으며, 2024년 기준 약 8.56억 달러 규모를 형성하였고, 연평균 11.77%의 성장률을 기록하며 2029년에는 약 15억 달러에 이를 것으로 전망된다[1].

이와 같은 기술 확산은 산업 전반의 생산성과 효율성을 향상시키는 긍정적인 효과를 가져오고 있으나, 이를 효과적으로 운영할 수 있는 숙련된 전문 인력의 부족은 특히 중소기업을 중심으로 주요한 과제로 부각되고 있다 [2,3]. 한국로봇산업진흥원의 자료에 따르면, 국내 로봇 관련 종사자 중 기술 전문가의 비중이 낮아, 도입된 첨단 기술이 현장에서 충분히 활용되지 못하는 경우가 빈번하다. 이로 인해 누구나 쉽게 접근하고 조작할 수 있는 사용자 친화적인 로봇 제어 시스템의 필요성이 더욱 절실해지고 있다.

한편, 인공지능(AI)과 빅데이터 기술의 비약적인 발전은 이러한 문제를 해결할 수 있는 새로운 가능성을 제시하고 있다. 특히 자연어 처리(Natural Language Processing, NLP) 분야에서 GPT-4와 같은 대규모 언어 모델(Large Language Model, LLM)의 등장은 인간과 기계 간의 자연스러운 상호작용을 가능하게 하며, 복잡한 명령 입력 과정을 대체할 수 있는 대안으로 주목받고 있다[4-8].

본 연구는 이러한 기술적 흐름에 주목하여, 사용자가 일상적인 언어로 로봇을 제어할 수 있는 직관적이고 효율적인 시스템의 설계 및 구현을 목표로 한다. 또한, 변화하는 물리적 환경을 인식하고 적절히 대응하기 위해 실시간 객체 인식 기술인 YOLO(You Only Look Once) 알고리즘을 통합하였다[9-12]. YOLO는 빠른 연산 속도와 높은 인식 정확도를 바탕으로, 복잡한 환경에서도 다수의 객체를 신속하게 식별할 수 있어 산업 현장에서의 적용 가능성이 높다.

본 논문은 위와 같은 기술 요소들을 인형뽑기 기계에 통합 적용한 사례를 중심으로, 자연어 기반 로봇 제어 시스템의 실제 구현 가능성과 효용성을 실험적으로 검증하고자 한다.

2. 관련 연구 동향

인공지능(AI)을 로봇 제어에 적용하기 위한 연구는 다양한 분야에서 활발히 이루어지고 있다. 특히, 스마트 제조 시스템을 위한 AI PaaS(Platform as a Service) 기술은 인공지능 기반 플랫폼이 복잡한 제조 공정을 간소화하고 자동화를 가속화함으로써 산업 전반의 효율성을 높일 수 있음을 보여주었다. 이러한 기술은 사용자 편의성을 제고하는 방향으로 발전하고 있다[13].

또한, 객체 인식 기술과 로봇 제어의 융합을 시도한 연구로는 YOLO 알고리즘을 기반으로 휠체어 로봇팔(WMRA)을 제어한 사례가 있다. 이 연구는 객체 인식 결과를 바탕으로 정밀한 로봇 제어가 가능함을 입증하였으며, 다양한 환경 변화에 실시간으로 대응할 수 있는 가능성을 확인하였다[14].

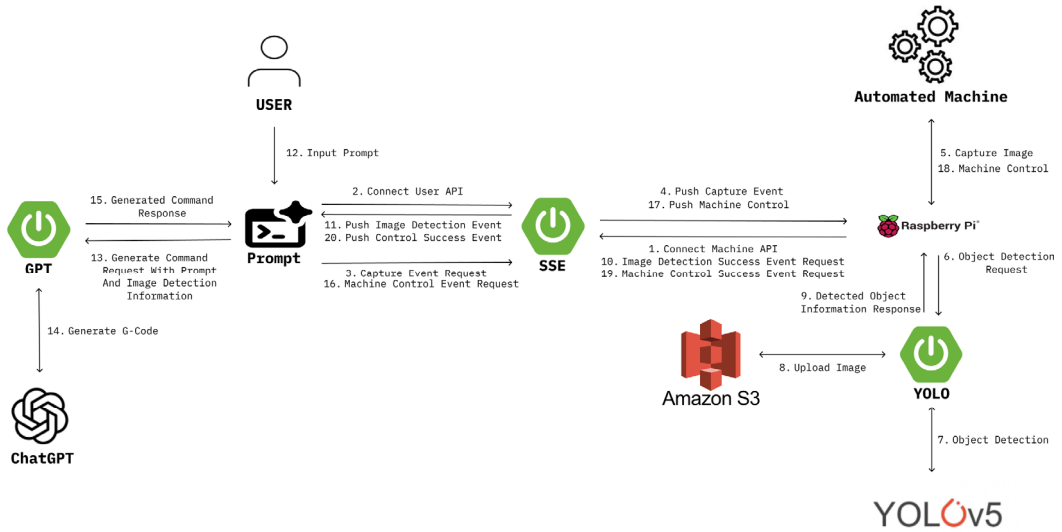
자연어를 활용한 로봇 제어 분야에서도 주목할 만한 연구가 이루어지고 있다. 예를 들어, 대규모 언어 모델(Large Language Model, LLM)과 음성 인터페이스를 결합한 ROS(Robot Operating System) 기반 모바일 로봇 제어 기술은 인간과 로봇 간의 상호작용을 자연스럽게 만들어 사용자 경험을 크게 향상시킬 수 있음을 보여주었다[15,16].

또 다른 연구에서는 LLM을 기반으로 한 일상생활 지원 로봇 시스템을 제안하여, 단순한 산업 영역을 넘어 개인화된 서비스 환경에서도 자연어 기반 로봇 제어의 활용 가능성을 제시하였다. 이 시스템은 자연어 명령을 통해 가사 및 일상 업무를 수행할 수 있어, 향후 생활 밀착형 서비스 로봇 개발에 실질적인 기여를 할 수 있음을 시사한다[17].

이와 같은 선행 연구들은 본 논문이 제안하는 자연어 기반 로봇 제어 시스템의 설계 방향과 밀접한 관련성을 가지며, 특히 YOLO 알고리즘과 LLM 기술의 융합이 실시간 객체 인식과 자연어 명령 해석을 통합한 고도화된 로봇 제어 방식의 실현 가능성을 뒷받침한다. 이러한 기술적 융합은 다양한 산업 및 생활 환경에서 로봇의 상황 인식 능력과 반응성을 높이는 동시에, 사용자 경험을 획기적으로 개선할 수 있는 기반이 될 수 있다.

3. 구현

본 연구에서는 자연어 기반 로봇 제어 시스템의 작동 원리를 실험적으로 검증하기 위해 인형뽑기 기계를 대상



[Fig. 1] System Architecture

으로 한 시제품을 구현하였다. 사용자는 그래픽 사용자 인터페이스(GUI)를 통해 원하는 인형의 특성을 자연어로 입력하며, 시스템은 이를 해석하여 해당 인형을 자동으로 선택하고 작동하도록 설계되었다.

이후 시스템은 SSE 기술을 활용하여 이벤트를 비동기적으로 처리하고, 객체 탐지 요청, 명령 실행 여부, 예외 상황 등을 실시간으로 사용자에게 전달한다. [Fig. 2]는 사용자 인터페이스의 구성 예시를 보여준다.

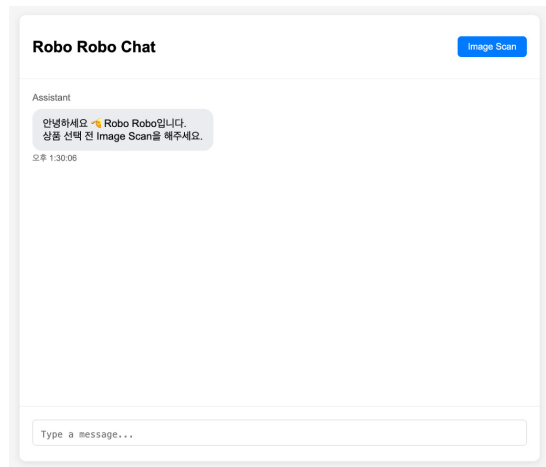
3.1 시스템 구조 및 주요 구성 요소

제안된 시스템은 사용자로부터 입력된 자연어 정보를 기반으로 이미지 내 객체를 인식하고, 그 결과를 기계 제어로 연계하는 통합 구조를 갖는다. [Fig. 1]은 시스템의 전체 구조를 개략적으로 나타낸다. 주요 구성 요소는 다음과 같다. 이벤트를 관리하는 SSE(Server-Sent Events)가 포함된다.

- GPT 언어 모델: 자연어 입력 해석 및 명령어 생성
- YOLOv5 객체 탐지 모델: 이미지 내 대상 객체 인식
- Raspberry Pi: 로봇 제어용 마이크로컨트롤러
- Amazon S3: 이미지 데이터의 저장 및 처리 중계
- SSE (Server-Sent Events): 실시간 이벤트 통신 및 상태 전달

3.2 사용자 인터페이스 및 이벤트 관리

사용자는 GUI 환경에서 두 가지 주요 입력을 수행한다. 첫째, 'Image Scan' 버튼을 눌러 이미지를 촬영하고 객체 인식을 요청하며, 둘째, 자연어로 인형의 특성을 프롬프트에 입력한다.



[Fig. 2] User Interface

3.3 객체 탐지 및 명령어 생성

객체 탐지 절차는 다음과 같은 흐름으로 진행된다.

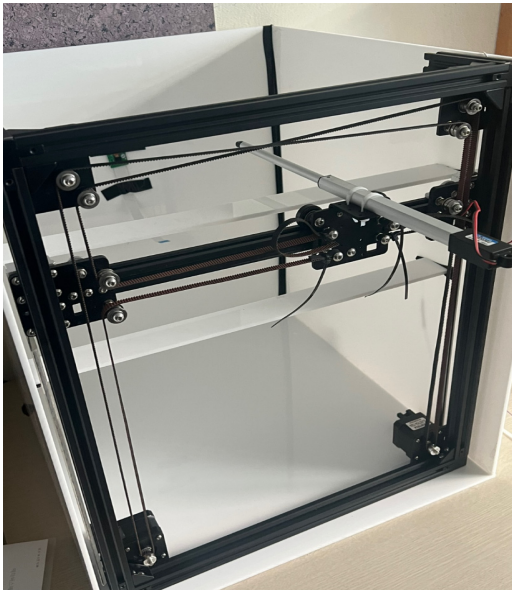
- ① Raspberry Pi에 연결된 카메라로 이미지를 촬영
- ② 촬영된 이미지를 Amazon S3에 업로드
- ③ 업로드된 이미지를 YOLOv5로 분석하여 객체 탐지 수행

탐지된 결과(객체 클래스, 바운딩 박스, 신뢰도 점수)는 SSE를 통해 반환되며, 이와 사용자 자연어 입력이 함께 GPT 모델에 전달된다. GPT는 이를 통합적으로 분석하여 로봇 제어를 위한 명령어를 생성하며, 해당 명령은 Raspberry Pi에 전달되어 기계 제어가 수행된다.

3.4 뽑기 장치 제어

인형뽑기 장치는 Raspberry Pi를 통해 직접 제어되며, 제어 결과는 다시 SSE를 통해 사용자에게 전달된다. [Fig. 3]은 Raspberry Pi 연결 전의 장치 모습을 보여준다. 장치 구성에는 다음과 같은 주요 하드웨어 요소가 포함된다.

- 카메라 센서: 이미지 수집
- 리니어 액추에이터: 인형을 미는 동작 수행
- 2축 모터 레일: 액추에이터의 위치 조절



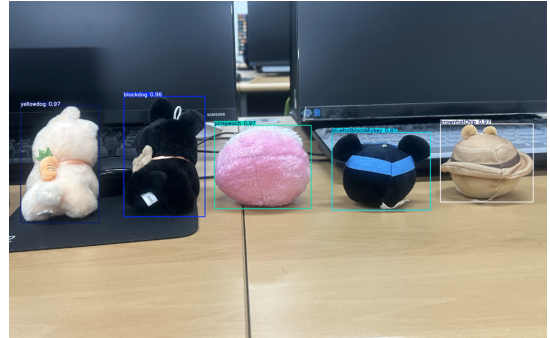
[Fig. 3] Claw Machine

3.5 구현 세부 사항

성능 및 정확도 측면에서 YOLOv3 대비 YOLOv5가 더 우수한 평가 결과를 보여주었기에, 본 연구에서는 YOLOv5를 채택하였다[18]. 모델 학습을 위해 인형뽑기 기계에서 사용되는 다양한 인형을 대상으로, 객체당 약 500장의 이미지를 수집하여 학습시킴으로써 인식률을 높였다.

YOLOv5를 통해 도출된 객체 정보는 클래스(label), 바운딩 박스 좌표, 신뢰도 점수로 구성되며, 이 정보는

GPT 모델이 적절한 제어 명령을 생성하는 데 활용된다.

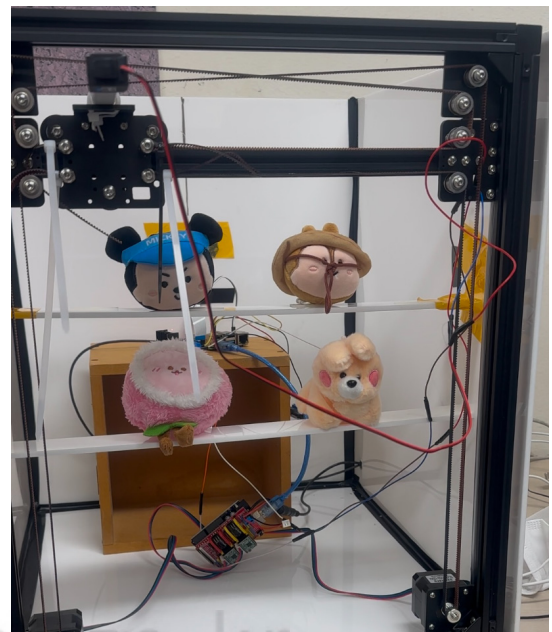


[Fig. 4] Object Detection

[Fig. 4]는 학습된 모델이 실제 객체를 인식한 결과를 시각화한 이미지로, 각 객체의 클래스, 위치, 신뢰도를 확인할 수 있다.

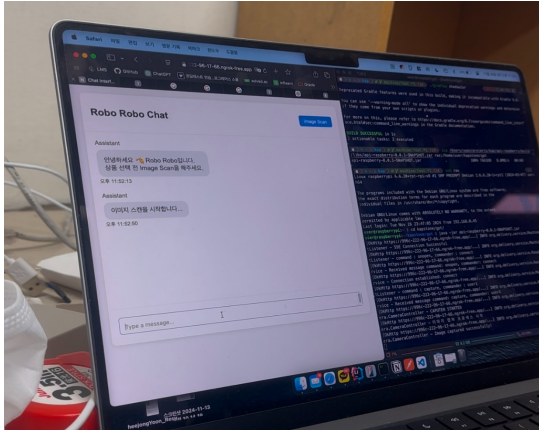
4. 구현 결과

본 장에서는 자연어 기반 로봇 제어 시스템의 작동 과정을 단계별로 설명하고, 구현된 핵심 기능의 결과를 시각 자료를 통해 검증한다. 제어시스템은 Raspberry Pi 4 (2GB), 1080p 카메라를 사용하였다.



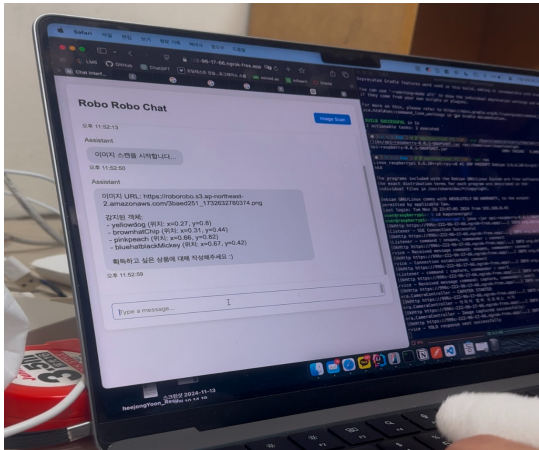
[Fig. 5] Initial State

[Fig. 5]는 시스템의 초기 상태로, 인형뽑기 기계 내 다양한 인형들이 전시되어 있으며 사용자가 선택을 수행하기 전의 대기 화면을 보여준다. 사용자는 우선 GUI에서 'Image Scan' 버튼을 선택하면 객체 인식이 시작한다. 이 과정을 실행한 화면은 [Fig. 6]과 같다.



[Fig. 6] Click Image Scan Button

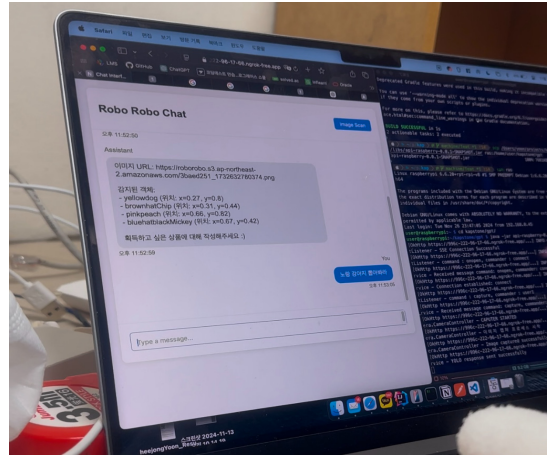
객체 인식이 완료되면, YOLOv5 모델에 의해 탐지된 인형들의 클래스명과 중심 좌표가 반환되며, 이는 시스템 내부 처리와 사용자 피드백에 활용된다. [Fig. 7]은 탐지 결과를 시각적으로 보여주는 화면으로, 객체의 위치와 식별 정보가 포함되어 있다.



[Fig. 7] Scan Complete

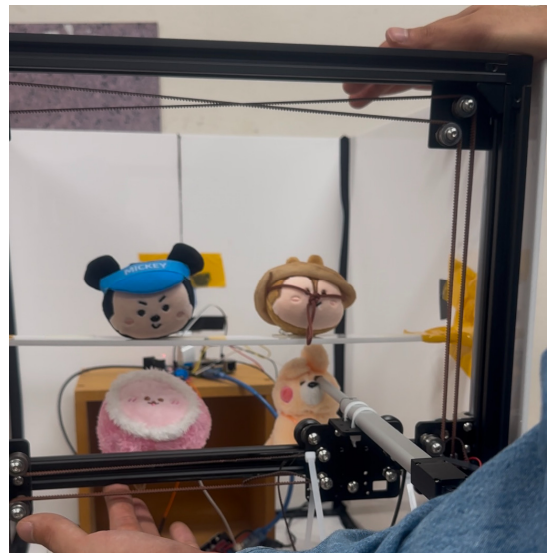
다음 단계로, 사용자는 자연어 프롬프트 입력란에 원하는 인형의 특성을 묘사하여 입력한다. 예를 들어 "노란색 병아리 모양 인형"과 같은 명령어를 통해 제어가 수행

된다. [Fig. 8]은 해당 프롬프트가 입력된 상태를 보여준다.



[Fig. 8] Prompt input status

최종적으로, 시스템은 객체 탐지 결과와 사용자 입력을 GPT 모델을 통해 통합적으로 분석한 뒤, 가장 부합하는 인형의 위치로 로봇 팔을 이동시켜 인형을 뽑는 작업을 수행한다. [Fig. 9]는 올바른 인형을 선택한 동작 결과를 나타낸다. 사용자의 명령에 정확히 대응하여 인형을 성공적으로 집어내는 장면은 실시간으로 GUI를 통해 확인할 수 있다.



[Fig. 9] Pick the right doll

또한 예외 상황에 대한 대응 기능을 구현하였다. 예를 들어, 사용자가 전시되지 않은 인형을 묘사하거나 탐지

결과와 일치하지 않는 입력을 제공할 경우, 시스템은 이를 인식하여 GUI에 적절한 경고 메시지를 출력하도록 설계되었다. 이를 통해 사용자 혼란을 최소화하고 시스템의 신뢰성을 향상시켰다.

5. 결론

본 연구에서는 사용자가 일상적인 자연어 명령을 통해 로봇을 제어할 수 있는 시스템을 설계하고, 이를 인형뽑기 기계에 적용하여 그 가능성을 실험적으로 검증하였다. 제안한 시스템은 GPT-4 기반의 자연어 처리 모듈, YOLOv5 객체 탐지 모델, Raspberry Pi 제어 시스템, 그리고 실시간 이벤트 통신을 위한 SSE 기술을 통합하여 구성하였다.

이와 같은 방식은 향후 제조 및 물류 산업 현장에서 사용되는 로봇 제어 시스템에 직접적으로 응용될 수 있다. 로봇을 생산하거나 운영하는 기업은 로봇 관리자와 같은 사용자가 시스템에 접속할 수 있도록 지원하는 Client SSE Connection API, 로봇 제어 로직이 구현된 장비를 시스템과 연동하는 Machine SSE Connection API, 자연어 명령을 처리하는 LLM 기반 API, 그리고 이미지 및 영상을 처리하는 객체 인식 API를 함께 제공함으로써 직관적인 로봇 제어 환경을 구축할 수 있다.

본 연구는 이러한 구성 요소들을 결합함으로써, 복잡한 명령어 입력 없이도 자연스러운 언어를 통해 로봇을 제어할 수 있는 시스템 개발이 가능하다는 점을 입증하였다. 특히 비전문가도 쉽게 조작할 수 있는 사용자 중심의 제어 환경은, 전문 인력 부족 문제를 겪고 있는 중소 제조 및 물류 기업에 실질적인 도움을 줄 수 있는 기술적 해법이 될 수 있다.

향후 연구에서는 다양한 자연어 표현을 정교하게 해석하기 위한 프롬프트 최적화 기법의 고도화, 다중 객체가 혼재된 환경에서의 상황 인지 및 판단 알고리즘의 개선, 그리고 산업용 로봇을 포함한 보다 복잡한 응용 분야의 확장 가능성에 대한 체계적인 연구가 필요하다.

REFERENCES

- [1] Mordor Intelligence, "South Korea Factory Automation and Industrial Controls Market Size & Share Analysis - Growth Trends & Forecasts (2024-2029)," Mordor Intelligence Industry Reports, 2024. [Online]. Available: <https://www.mordorintelligence.com/industry-reports/south-korea-factory-automation-and-industrial-controls-market>.
- [2] K.S.Ko, H.J.Jun, J.I.Oh, "A Study on the Factors that Affect the Adoption of a Smart Factory - Focusing on the Comparison between Customers and Suppliers -," *Korea Business Review*, Vol.25, No.3, pp.129-151, 2021.
- [3] Y.J.Son, H.Y.Choi, "A Study on Smart Factory Introduction Cases and Sustainable Effect," *Journal of Practical Engineering Education (JPPE)*, Vol.14, No.1, pp.127-136, 2022.
- [4] T.Kim, T.K.Kim, "Systemic Analysis of the QS International Research Network Indicator Using Big Data: Regional Inequalities and Recommendations for Improved University Rankings," *IEEE Access*, Vol.13, pp.111335-111353, 2025.
- [5] T.Kim, T.K.Kim, "Global Inequality in Research: A Quantitative Analysis of 1628 Institutions in THE World University Rankings 2025," *IEEE Access*, Vol.13, pp.161262-161278, 2025.
- [6] I.R.Numonov, B.Peng, Y.Li, Y.I.H.Ugli, T.O.Lee, T.K.Kim, "Development of an AI-Based Energy Management System for Factory Power Saving," *Journal of Internet of Things Convergence*, Vol.10, No.6, pp.49-55, 2024.
- [7] J.W.Kim, B.S.Geum, T.K.Kim, "Implementing a Lecture Summary Application Using STT API and ChatGPT API," *Journal of Internet of Things Convergence*, Vol.11, No.3, pp.167-174, 2025.
- [8] Y.Li, I.N.R.Ugli, Y.I.H.Ugli, T.O.Lee, T.K.Kim, "Optimizing Models and Data Denoising Algorithms for Power Load Forecasting," *Energies*, Vol.17, No.21, pp.5513, 2024.
- [9] S.W.Jeon, D.S.Kim, H.K.Jung, "YOLO-based lanedetection system," *Journal of the Korea Institute of Information and Communication Engineering*, Vol.25, No.3, pp.464-470, 2021.
- [10] B.Peng, T.K.Kim, "YOLO-HF: Early Detection of Home Fires Using YOLO," *IEEE Access*, Vol.13, pp.79451-79466, 2025.
- [11] S.I.Kim, "A model to secure storage space for CCTV video files using YOLO v3," *Journal of the Korea Society of Computer and Information*, Vol.28, No.1, pp.65-70, 2023.
- [12] G.H.Jo, K.M.Hyun, Y.J.Song, "Parallel U-Net Based Semantic Segmentation Method Using Generated Data from YOLO V5," *The Journal of Korean Institute of Communications and Information Sciences*, Vol.48, No.3, pp.319-326, 2023.
- [13] J.J.Yu, J.P.Lee, W.K.Hyun, "A real-time AI-based control system of a mobile robot in an environment with complex obstacles," *The Korean Institute of Electrical Engineers Conference*, 2019.
- [14] G.T.Park, J.G.Yun, J.H.Park, C.J.Kim, D.W.Lim,

"YOLO-based Control for the Wheelchair Mountable Robot Arm (WMRA)," *The Korean Society of Mechanical Engineers IT Convergence Division Spring Conference*, p.102, 2023.

- [15] S.H.Jang, C.U.Jeun, S.H.Ok, "ROS based Mobile Robot Control using LLM and Voice Interface," *Proceedings of the 2024 Korea Information Technology Society Summer Conference*, pp.532-534, 2024.
- [16] S.H.Lee, A.E.Kwak, S.H.Lee, T.K.Kim, "Indoor autonomous driving system based on Internet of Things," *Journal of Internet of Things Convergence*, Vol.10, No.2, pp.69-75, 2024.
- [17] S.U.Choi, D.I.Choi, "LLM-based daily life task robot system," *Journal of the KSME*, Vol.64, No.10, pp.44-52, 2024.
- [18] M.H.Park, J.H.Choi, W.J.Lee, "Object detection for various types of vessels using the YOLO algorithm", *Journal of the Korean Society of Marine Engineering*, Vol.48, No.2, pp.81-88, 2024.

윤 희 종(Hee-Jong Yoon) [준회원]



■ 2019년 3월 ~ 2025년 2월 :
국립부경대학교 컴퓨터·인공지능
공학부

<관심분야>
사물인터넷, 웹 개발

조 현 성(Hyeon-Seong Jo) [준회원]



■ 2019년 3월 ~ 2025년 2월 :
국립부경대학교 컴퓨터·인공지능
공학부

<관심분야>
사물인터넷, 웹 개발

윤 재 현(Jae-Hyeon Yoon) [준회원]



■ 2019년 3월 ~ 2025년 2월 :
국립부경대학교 컴퓨터·인공지능
공학부

<관심분야>
사물인터넷, 웹 개발

김 지 환(Ji-Hwan Kim) [준회원]



■ 2019년 3월 ~ 2025년 2월 :
국립부경대학교 컴퓨터·인공지능
공학부

<관심분야>
사물인터넷, 웹 개발

김 태 국(Tae-Kook Kim) [종신회원]



■ 2004년 8월 : 고려대학교
전기전자전파공학부(공학사)
■ 2006년 8월 : 고려대학교
메카트로닉스학과(공학석사)
■ 2014년 8월 : 고려대학교
모바일솔루션학과(공학박사)
■ 2016년 3월 ~ 2022년 2월 :
동명대학교 AI학부 교수
■ 2022년 3월 ~ 현재 : 국립부경대학교 컴퓨터·인공지능
공학부 교수

<관심분야>
사물인터넷(IoT), 콘텐츠 전송 네트워크(CDN), 이동성,
인공지능(AI), 빅데이터(big data), 모바일 서비스