Journal of Korea Robotics Society (2024) 19(4):347-353
https://doi.org/10.7746/jkros.2024.19.4.347 ISSN: 1975-6291 / eISSN: 2287-3961

음성 명령 및 객체 인식 기반 로봇 작업 계획

Robot Task Planning Based on Voice Command and Object Detection

박 무 열[1*] · 정 우 주[1*] · 김 민 조[1*] · 박 규 민†
Mu-Yeol Park[1*], Woo-Joo Jung[1*], Min-Jo Kim[1*], Kyu Min Park†

Abstract: Recently, studies on human-robot interaction (HRI) have been actively conducted in the field of robotics. For effective interaction, a robot should be equipped with an 'eye' for environmental perception, localization capabilities, and language skills to communicate with people. In this paper, we integrate vision-based object detection and speech recognition-based command delivery technologies to improve operational efficiency in line with user intent, and explore the implementation of a robot system for recycling. Our system utilizes an OpenMANIPULATOR-X robot arm and uses voice commands recognized by Python's SpeechRecognition library to issue control commands to the robot, where the speech recognition module classifies tasks by extracting keywords from the text converted from voice commands. A vision system equipped with an Intel D435i camera then identifies the status of transparent cups using YOLO and AlexNet models. Based on this information and the commanded task, the robot executes an appropriate action. Experimental results demonstrate our system's capability to effectively automate recycling, showing success rates of 91% and 95% for voice command recognition and object detection respectively.

Keywords: Human-Robot Interaction, Object Detection, Speech Recognition, Task Planning

1. 서 론

인공지능의 혁신적인 변화에 기반한 지능형 로봇 기술은 높은 가치를 창출하고 있다[1,2]. 과거 로봇이 정형화된 환경에서 반복적인 작업만을 수행하던 때와는 달리, 더욱 복잡하고 예측 불가능한 환경에서 작업을 수행해야 하는 오늘날 그 가치가 돋보인다. 대표적인 사례로 폐기물 재활용 산업에서의 로봇 기술 도입을 들 수 있다[3]. 수거, 선별, 재생 과정을 거치는 플라스틱 재활용의 경우 선별 작업의 문제를 해결하고자 인공지능 기술의 도입을 확대하였다.

이와 함께 발전하고 있는 인간-로봇 상호작용(HRI) 기술은 사용자와 로봇의 작업공간 공유를 통해 로봇의 활용도를 더욱 높이고 있다. 인공지능의 발전과 함께 접촉 기반, 비전 기반 등 다양한 HRI 기술이 개발되었으며[4,5], 특히 최근에는 로봇에 부착된 마이크로부터 취득한 음성정보에 기반한 언어적 상호작용에 대한 연구가 활발히 진행되고 있다. 이는 음성 인식(speech recognition), 화자인식(speaker recognition), 음원 추적(sound localization), 음원 분리(sound separation) 등을 포함한다[6]. 대부분 수작업으로 진행되는 복잡한 환경 속에서 음성(언어)을 통한 정확한 명령 전달의 필요성은 높아지고 있으며, 정확한 인식 및 높은 추적 성능을 위해 MP[7], 강화학습을 통한 자연어 생성의 최적화[8], 대규모 언어모델(Large Language Model)[9] 등의 다양한 연구가 진행되고 있다.

이러한 언어적 HRI 기술 개발 흐름에 발맞추어 본 연구에서는 음성 명령 및 객체 인식에 기반한 로봇 작업 계획(task planning) 기술을 제시하고, 투명 플라스틱 컵 재활용 시나리오를 통해 본

Received : Jul. 30. 2024; Revised : Aug. 13. 2024; Accepted : Aug. 14. 2024
※ This project was supported by the IITP (Institute of Information & Communications Technology Planning & Evaluation)-ICAN (ICT Challenge and Advanced Network of HRD) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2024-RS-2024-00436528).
* Mu-Yeol Park, Woo-Joo Jung, and Min-Jo Kim contributed equally to this work.
1. Undergraduate Student, Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Korea (parkanduf, universe-22, gk12fls34@naver.com)
† Assistant Professor, Corresponding author: Department of Artificial Intelligence and Robotics, Sejong University, Seoul, Korea (kyuminpark@sejong.ac.kr)

연구에서 제시하는 언어적 상호작용 기반 기술이 사용자 의도에 부합하는 작업 계획에 효과적으로 활용될 수 있음을 보인다.

본 연구의 로봇 작업 계획은 음성 명령 및 객체 인식 정보에 따라 분류되는 24가지의 작업 Case에 기반해 진행된다. 음성 인식 기반 명령 전달 단계에서는 Speech-to-Text (STT) 기술을 통해 텍스트로 변환된 음성의 키워드를 추출하여 작업 Case를 분류한다. 이후 객체 인식 모델을 통해, 학습된 객체 상태를 인식하여 물체의 상태 정보를 취득하고, 이에 따라 로봇이 각 Case에 적합한 작업을 수행하는 단계가 진행된다.

작업환경의 물체를 인식하고 이와 상호작용하며 작업을 수행하기 위해 1대의 카메라를 이용한 비전 시스템을 구현한다. 투명 플라스틱 컵의 상태(Empty/Full)를 검출하기 위해, YOLO v4 모델을 사용하여 라벨링 된 학습 데이터를 수집하고, AlexNet 모델을 전이 학습(transfer learning)하여 새로운 검출 모델을 생성한다. 생성된 모델을 기반으로 컵의 상태를 검출하여 분리수거 작업을 위한 정보를 획득한다.

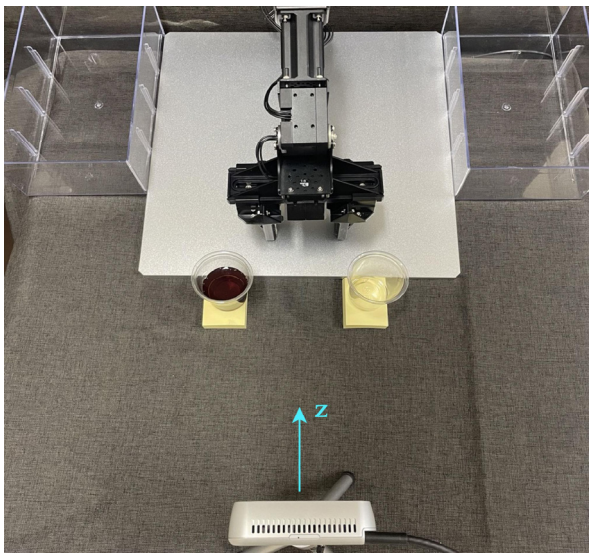
음성 명령과 객체 탐지를 통해 작업환경과 사용자의 의도에 대한 정보를 얻은 후에는, 로봇 팔을 이용해 컵을 파지 및 조작하여 내용물과 컵을 분류해 버리는 과정이 진행된다. 음성 명령과 객체 탐지를 통해 취득하는 정보에 따른 작업 Case를 미리 정의해 놓고, 정보를 받는 동시에 구현해 놓은 동작에 따라 로봇이 작동하도록 시스템을 구성한다. 본 연구에서는 5-자유도 로봇 머니퓰레이터 OpenMANIPULATOR-X를 활용한 작업 계획을 제시한다. 해당 로봇의 1-자유도 그리퍼의 개폐 동작을 통해 컵을 파지하여, 본 연구의 작업 시나리오가 고려하는 컵 재활용 동작을 효과적으로 수행할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 시스템의 전체적인 구성과 시스템을 이루는 전반적인 알고리즘을 소개한다. 3장에서는 시스템의 시작점인 음성 기반 명령 전달에 대한 자세한 설명을 진행하고, 4장에서는 객체 인식과 로봇 동작 계획을 설명한다. 5장에서는 제안된 작업 계획 시나리오의 검증을 위한 실험을 진행하며 결과를 설명한다. 마지막으로 6장에서 논문의 결론에 대한 논의와 향후 연구 방향을 제시한다.

2. 시스템 및 알고리즘

2.1 시스템 구성

[Fig. 1]은 시스템의 실험 세팅을 보여준다. 일반적으로 사용되는 음료수 컵의 축소 버전의 투명 컵을 사용한 재활용 시나리오를 가정하여 구현한다. 전체 시스템은 1) 노트북의 내장 마이크를 사용한 음성 명령 인식 시스템, 2) 단일 Intel D435i 카메라를 활용한 객체 인식 시스템(깊이 정보는 활용하지 않는다), 3) 정밀 조작을 위한 소형 로봇 OpenMANIPULATOR-X를 사용한 작업 자동화 시스템으로 구성된다.



[Fig. 1] Experimental setup overview

2.1.1 음성 명령 인식 시스템

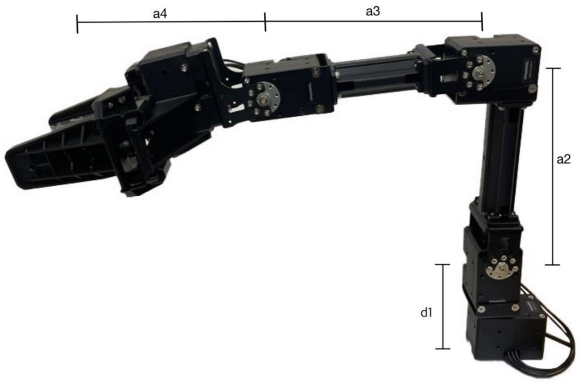
시스템의 첫 단계인 음성 기반 명령어 식별을 위해, 시스템과 화자에 관계없이 음성 신호를 인식할 수 있는 화자독립(speaker-independent) 인식 모듈을 Python의 음성 인식 라이브러리를 통해 구현한다. 실시간 음성 캡처를 통한 음성 정보 수집을 위해 라이브러리의 Recognizer와 Microphone 클래스를 사용하며, 음성 명령의 텍스트 변환을 위해 오디오 정보를 음성 정보로 변환한다. 변환된 텍스트에서 키워드 매칭에 기반한 명령어 분류를 통해, 사전 정의된 작업 Case에 부합하는 로봇 제어 명령을 수신한다.

2.1.2 객체 인식 시스템

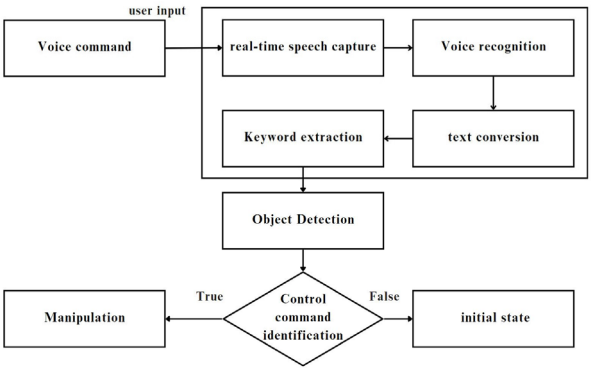
객체 인식 시스템은 하나의 카메라를 이용해 객체(컵)와 객체의 상태(컵의 내용물 여부)를 검출하여 로봇에게 전달하는 것을 목표로 한다. 카메라는 실험 대상인 컵과 z축 방향으로 20 cm 떨어진 곳에 고정되어 있으며, 카메라로부터 실시간 영상을 취득하고 컵의 내용물 여부를 검출하기 위한 비전 처리 과정이 수행된다. 실험은 실내 환경에서 진행되며, Intel D435i 카메라를 통해 안정적으로 영상 정보를 취득한다. 이와 같은 세팅을 통해 컵을 객체로 검출하고, 그 내용물을 분석하는 비전 시스템을 구축한다.

2.1.3 작업 자동화 시스템

본 연구에서는 무겁거나 모양이 다채로운 물체가 아닌 소형 컵에 대한 파지 및 조작이 필요하기에 1-자유도 그리퍼가 장착된 소형 머니퓰레이터 OpenMANIPULATOR-X를 사용한다.



[Fig. 2] Hardware configuration: OpenMANIPULATOR-X at its home (zero) position



[Fig. 3] Overall process of the system

해당 로봇은 [Fig. 2]와 같이 아래에서부터 베이스, 하단 팔, 상단 팔, 손목과 엔드 이펙터(그리퍼)로 구성되어 있다. 베이스에 부착된 1번 조인트는 수직축을 중심으로 팔을 회전시키며, 2번 조인트에 연결된 하단 팔은 베이스에서 수직으로 연장되어 있어 상부 구성 요소의 수직 이동과 위치 지정에 중요한 역할을 한다. 상단 팔은 3번 조인트를 통해 하단 팔과 연결되어 있으며 머니퓰레이터의 도달 범위와 기동성을 높여준다. 4번 조인트에 해당하는 손목은 엔드 이펙터를 조작해 정밀한 움직임을 가능하게 한다. 마지막으로 5번 조인트로 구동되는 엔드 이펙터를 통해 물체를 정밀하게 파지 및 조작한다.

베이스에서 2번 조인트까지의 거리는 7.7 cm (d1), 하단 팔 길이는 12.8 cm (a2), 상단 팔 길이는 12.4 cm (a3), 엔드 이펙터 길이는 12.6 cm (a4)이다. 링크 간 충돌과 링크와 모터 사이의 충돌을 방지하기 위해 2, 3, 4번 조인트는 -90°−90°의 범위에서 움직이도록 제한하며, 해당 구동 범위 내에서 컵 분리수거 작업을 위한 파지 및 조작을 안정적으로 수행할 수 있다.

2.2 알고리즘

전체 시스템의 알고리즘 실행 과정은 [Fig. 3]와 같이 음성 인식, 객체 인식, 동작 계획 순서로 구성하여 사용자 명령에 따라 자동으로 컵을 인식하고 처리하는 시스템을 구축한다. 사용자가 마이크를 통해 입력한 음성을 실시간으로 캡처하고 텍스트로 변환한 후, 변환된 텍스트에서 특정 명령어와 관련된 키워드를 추출한다. 이후 객체 인식 모델과 카메라를 사용해 투명 플라스틱 컵을 인식하고, 컵의 상태(내용물의 여부)를 판단한다. 동작 계획 단계에서는 추출된 키워드와 컵의 위치, 상태 정보를 사전에 정의된 24가지 작업 Case와 매칭하여 로봇이 수행할 작업을 결정하며, 로봇은 TCP 소켓 통신을 통해 명령을 전달받아 동작을 수행한다. 명령과 컵의 상태 정보를 바탕으로 컵 파지 후 내용물이 있는 컵은 내용물을 비우고 처리, 빈 컵은 바로 처리하며, 각 조인트에 부착된 DYNAMIXEL 모터의 각도를 직접 제어해 각 동작을 수행한다.

3. 음성 기반 명령 전달

3.1 음성 인식 모듈

음성 인식 모듈은 Python의 SpeechRecognition 라이브러리를 사용해 구현한다. 해당 모듈은 시스템의 기본 마이크를 통해 사용자의 음성 입력을 실시간으로 캡처하고 텍스트로 변환하며, Recognizer 클래스와 Microphone 클래스로 구성된다. Speech-Recognition 라이브러리의 모든 작업은 Recognizer 클래스에서 진행되며, 이는 다양한 API와 통합할 수 있는 메서드를 통해 높은 유연성을 제공한다. Microphone 클래스는 음성을 실시간으로 받아오는 역할을 수행하며 클래스 Instance 생성을 통해 시스템의 기본 마이크에 접근한다.

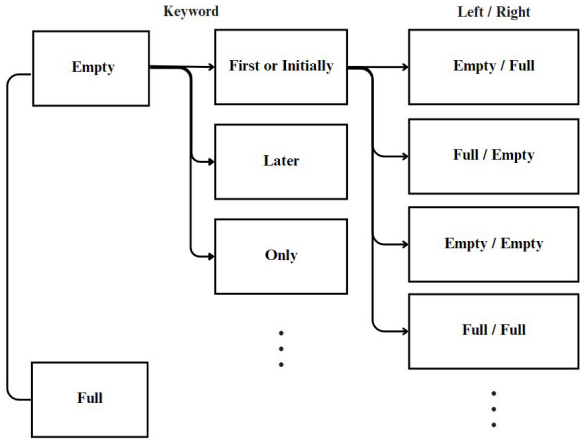
Recognizer 클래스의 listen 메서드를 통해 마이크에서의 입력을 실시간으로 캡처하며 침묵이 감지될 때까지의 입력을 audio에 기록한다. 음성을 입력하는 단계에서 음성 소음이 발생할 것을 대비해 adjust_for_ambient_noise 메서드를 통해 인식 임계값을 자동으로 조정한다. 이후 Google Web Speech API를 사용해 음성을 텍스트로 변환하는데, 이때 변환되는(인식) 언어는 매개변수를 통해 설정할 수 있으며 본 연구에서는 이를 한국어로 설정한다. 텍스트로 변환된 음성 인식 결과를 통해 사용자 의도 분석을 위한 키워드 추출 및 매칭을 진행한다.

3.2 키워드 추출 및 Case 분류

변환된 텍스트 데이터에서 명령어를 추출하기 위해 해당 단계에서는 키워드 매칭 기법을 사용한다. [Table 1]과 같이 특정 명령어와 관련된 키워드를 사전에 정의하고, 변환된 텍스트에서 해당 키워드를 탐색한다.

[Table 1] Keywords extracted from voice

Keyword		
Empty	First or Initially	Throw away or Clean up
	Later	
	Only	
Full	First or Initially	Throw away or Clean up
	Later	
	Only	



[Fig. 4] Overall case of manipulation

추출된 키워드를 기반으로 사용자의 작업 의도를 파악하고, Python 환경의 MATLAB Engine을 활용하여 컵의 상태(Empty/Full)를 감지해 동작을 결정한다. 예를 들어, 인식하는 카메라 기준 오른쪽에 Full 상태의 컵이, 왼쪽에 Empty 상태의 컵이 있는 경우 “빈 컵 나중에 치워”라는 명령을 입력받았을 때 오른쪽의 Full 상태 컵의 내용물을 비우고 빈 컵을 처리한 뒤 왼쪽의 빈 컵을 바로 처리한다.

작업 Case의 경우, 음성에서 추출된 키워드에 따른 6종류의 명령과 오른쪽/왼쪽 컵의 4가지 상태(Full/Full, Empty/Full, Full/Empty, Empty/Empty)의 조합을 통해 24가지의 수행 동작으로 분류된다. [Fig. 4]는 추출된 키워드와 물체의 위치에 따라 분류된 Case이다. 분류된 24가지 Case에 따라 지정된 최종 동작 명령은 C++로 구현된 로봇 제어 시스템에 전송된다.

3.3 명령 전송

Python의 socket 모듈을 사용하여 클라이언트와 서버 간의 통신을 구현한다. Python 클라이언트 서버에서 음성 인식과 객체 인식을 통해 결정된 동작 명령을 C++로 구현된 로봇 제어 서버로 전송한다. 사용자의 명령에 따른 정확하고 신속한 로봇 동작을 위해 실시간 데이터 전송에 적합한 TCP/IP 프로토콜 기반의 소켓 통신을 사용한다.

4. 객체 인식 및 동작 계획

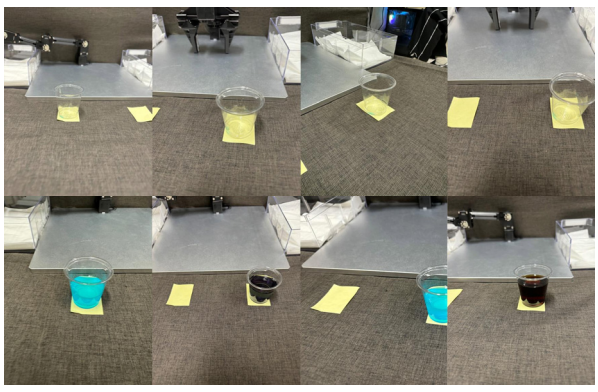
4.1 객체 인식

객체 인식 모델을 생성하기 위해 학습 데이터를 직접 촬영하여 수집한다. 인식하고자 하는 객체는 투명 소형 플라스틱 컵이며, 내용물이 없는 컵의 사진과 내용물이 담긴 컵의 사진을 1:1 비율로 총 500장 촬영한다. 컵이 투명하여 주변 환경(배경)의 영향으로 인해 인식 오류가 발생할 수 있으므로 이를 방지하기 위해 컵을 포스트잇 위에 위치시킨다. 내용물이 담긴 컵의 경우 내용물을 컵의 60%−80%까지 채운다. 객체 인식 모델의 일반화 성능을 위해 불규칙한 컵 배치와 다양한 카메라 시야각에서 촬영을 진행하며, [Fig. 5]는 학습 데이터의 일부이다. 촬영된 이미지는 YOLO v4 모델을 사용하여 라벨링 되며, 이후 라벨링된 이미지 데이터를 기반으로 전이 학습을 통해 컵 객체 탐지 모델을 생성한다. 생성된 모델은 컵의 객체 탐지와 함께, 컵의 내용물 여부를 확인할 수 있도록 설계된다.

생성된 모델을 기반으로 명령이 전달되면 카메라가 작동하여 객체를 인식하고 객체의 내용물 여부 정보를 로봇에게 전달하게 된다. 이를 통해 로봇이 사용자의 의도에 부합하는 동작을 수행할 수 있다.

4.2 동작 계획

컵이 놓일 위치와 내용물 수거 상자의 위치, 컵 수거 상자의 위치를 사전 설정하고, 두 컵 중 우선적으로 처리할 컵의 위치와 내용물 여부에 따른 Case를 총 5개의 코드로 분류해 놓는다. ‘0’: 내용물이 있는 컵이 오른쪽에 위치, ‘1’: 내용물이 없는 컵이 오른쪽에 위치, ‘2’: 내용물이 있는 컵이 왼쪽에 위치, ‘3’: 내용물이 없는 컵이 왼쪽에 위치, ‘4’: 음성 명령에 해당하는 경우가 없을 때, 즉 잘못된 명령일 때를 대비하여 초기 상태에서 아무 동작도 수행하지 않는 상태. 객체 인식을 통해 얻은 정보가



[Fig. 5] Model training image data

‘0’, ‘1’, ‘2’, ‘3’, ‘4’로 이루어진 문자열로 변수에 저장되어 TCP/IP 소켓 통신을 통해 로봇에게 전달되는 것을 시작으로 로봇 제어가 시작된다. 예를 들어 ‘02’ 문자열이 로봇에게 전달된다면 ‘0’과 ‘2’를 순차적으로 처리해야 하며, 오른쪽의 내용물이 있는 컵을 처리한 후 왼쪽의 내용물이 있는 컵을 처리하는 과정을 수행하게 된다. 이와 같은 두 자리 숫자 통신을 사용함으로써 컵 위치와 내용물 여부의 조합에 따른 다양한 경우의 수에 대해 간편하게 로봇을 제어할 수 있다.

DYNAMIXEL 모터는 엔코더 값을 입력받아 움직이기 때문에 모터 각도 제어 시 입력 각도를 엔코더 값으로 변환하는 과정을 적용한다. 각도 범위 0°−360°를 엔코더 값 범위 0−4095로 변환하기 위해 입력 각도에 일정 비율(4095/360)을 곱해 최종 입력 엔코더 값을 계산한다.

엔드 이펙터가 두 컵에 접근하기 용이한 자세(configuration)를 로봇의 초기 상태로 설정하여, 파지 및 조작의 시작과 종료 과정에 로봇이 초기 상태로 이동하도록 설정한다. 내용물이 있는 컵의 경우 내용물을 해당 상자에 비운 후 빈 컵을 컵 수거 상자에 넣고, 반대로 컵에 내용물이 없다면 바로 컵을 버리는 동작을 수행한다. 안정적인 파지와 컵의 내용물을 이동 중 쏟지 않도록(즉, 컵의 각도를 유지하도록) 접근 과정과 이동 과정 중 중간 자세들을 설정해 안정적인 조작을 가능하게 한다.

5. 실 험

5.1 음성 기반 명령 실행

먼저, 내장 마이크에 음성 명령을 전달하여 해당 명령의 음성 인식과 지정된 키워드의 추출이 정확히 진행되는지 확인하였다. 실내를 기준으로 소음 정도에 따라 약 20 dB의 조용한 환경, 약 60 dB의 생활 소음이 있는 환경, 약 70 dB의 야외 소음이 있는 환경을 무작위로 구성하였고, 키워드에 따라 분류된 6가지 상황에 적합한 명령어로 각 Case 별 30회의 명령을 전달해 음성 인식 및 키워드 추출의 정확도를 측정하였다. 키워드가 포함된 다양한 명령어를 사용하였고, 음성 인식 후 STT 과정을 거친 텍스트에서 추출된 키워드를 확인하였다. 이후 진행되는 객체 인식의 결과와 추출된 키워드에 따라 사전에 지정된 제어 명령을 식별하였다.

5.2 객체 인식

음성 명령의 키워드 추출 후, [Fig. 6]와 같이 전이 학습을 통해 훈련된 모델로부터 다양한 상태의 객체에 대한 인식 정확도를 측정하였다. [Table 2]의 인식 결과에서 볼 수 있듯이 높은 정확도의 객체 인식률을 보였으며, 모델 생성 당시 학습 데이터의 부족으로 인한 인식 오류 또한 확인할 수 있었다. 이후 카메라 기준 오른쪽부터 객체 정보를 2차원 배열로 지정하여 객체 정보를 추출하였다. 각각 배열에 해당하는 정보는 사전에 정의하였던 ‘Empty’는 0으로, ‘Full’은 1로 정의하였다.

[Fig. 6] Model output results

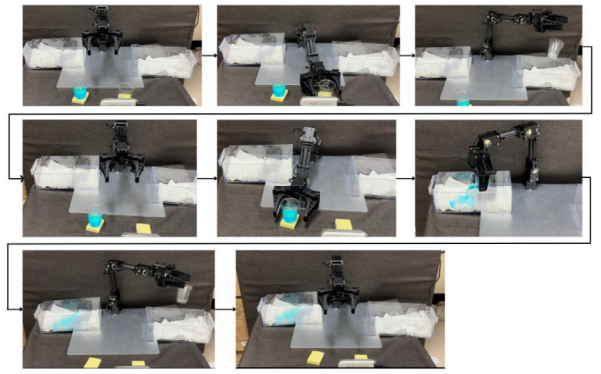
[Table 2] Object recognition success rate according to location and status of the cups

Left object	Right object	Success rate
Full	Empty	93% (14/15)
Empty	Full	93% (14/15)
Empty	Empty	100% (15/15)
Full	Full	93% (14/15)

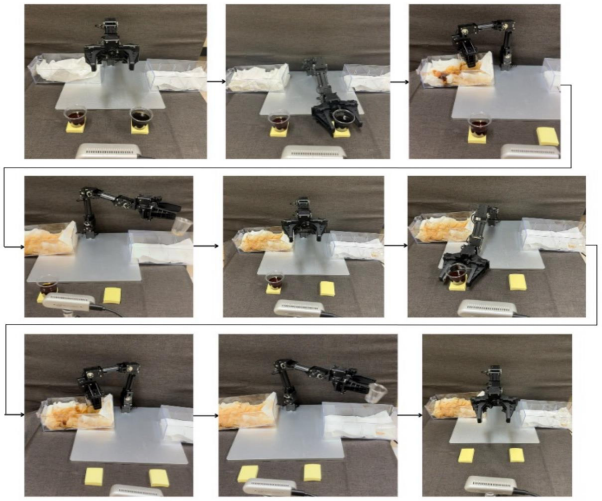
5.3 실험 및 결과

본 실험은 음성 기반 명령 실행부터 객체 인식, 동작 실행의 단계로 이루어졌으며, 키워드 기반의 6가지 Case에서 ‘빈’과 ‘채워진’을 기준으로 나누어 2가지 경우에 대해 진행하였다. [Fig. 7(a)]는 “빈 컵 먼저 치워”라는 명령에 대한 로봇 동작 과정이다. 음성 인식 모듈을 통해 해당 음성에서 ‘빈’, ‘먼저’, ‘치워’라는 키워드를 추출한다. 이후 객체 검출 모델을 통해 물체의 상태 인식이 진행된다. 2차원 배열로 구성된 위치별 물체의 상태 정보와 추출된 키워드를 기반으로 제어 명령을 식별하고, 식별된 값을 제어 서버에 전달한다. 제어 명령을 받은 로봇은 [Fig. 7(a)]와 같이 카메라 기준 오른쪽의 빈 컵을 먼저 처리하고 왼쪽의 채워진 컵을 처리하도록 동작한다. 작업 Case 분류에 따르면 ‘12’라는 값이 로봇에 전달되었음을 의미한다.

두 컵과 두 수거 상자의 위치는 고정되어 있으며, 엔드 이펙터가 각 위치에 도달하는 모터 각도를 로봇 역기구학을 사용



(a)



(b)

[Fig. 7] The whole process of robot manipulation in a recycling scenario: Keywords: (a) ‘Empty’, ‘First’, (b) ‘Full’, ‘Later’

해 계산하였다[10]. 각 위치에 해당하는 조인트 1−4의 모터 각도는 다음과 같다. 카메라 기준 왼쪽 상자(내용물): {-67°, 0°, 0°, 0°}, 오른쪽 상자(컵): {67°, 0°, 0°, 0°}, 왼쪽 컵: {-13°, 51°, -5°, -47°}, 오른쪽 컵: {13°, 52°, -3°, -49°}. 이후 내용물을 처리하기 위해 컵을 기울이는 자세에 해당하는 모터 각도는 {±67°, 0°, 0°, 105°}이다. 엔드 이펙터의 5번 모터 각도가 110°일 때는 그리퍼가 열린 상태, 185°일 때는 그리퍼가 닫힌 상태이다.

로봇이 작업 명령을 성공적으로 전달받아 우선 [Fig. 7(a)]의 첫 번째 사진과 같이 초기 상태로 이동하였다. 역기구학으로 계산된 각도를 통해 [Fig. 7(a)] 두 번째, 세 번째 사진과 같이 빈 컵을 잡고 버리는 과정을 수행할 수 있었다. 하나의 컵을 처리한 후 네 번째 사진과 같이 다시 초기 상태의 위치로 이동하고, 5, 6, 7번째 사진과 같이 채워진 컵을 비우고 버리는 과정을 거친 후 마지막 사진과 같이 다시 초기 상태로 돌아와 분리수거 작업을 마쳤다. [Fig. 7(b)]는 같은 메커니즘으로 “채워진 컵 우선 치워”

[Table 3] Voice recognition and keyword extraction accuracy

Keyword			Success rate
Empty	First or Initially	Throw away or Clean up	87% (26/30)
	Later		93% (28/30)
	Only		90% (27/30)
Full	First or Initially	Throw away or Clean up	90% (27/30)
	Later		97% (29/30)
	Only		90% (27/30)

명령을 인식하고 수행하는 과정을 나타낸다. 두 컵 모두 추출된 키워드와 일치하는 경우 카메라 기준 오른쪽 컵을 먼저 처리하도록 설정하였다.

해당 작업 계획 실험에서 키워드 매칭에 기반한 음성 명령의 인식 성공률은 [Table 3]와 같다. 또한 [Table 2]에서 객체 인식 성공률을 확인할 수 있다. 인식 성능 자체를 보았을 때 전체 실험에 대해 음성 인식의 경우 약 91% (164/180), 객체 인식의 경우 약 95% (57/60)의 높은 성공률을 확인할 수 있었다.

6. 결 론

본 연구에서는 투명 플라스틱 컵의 재활용 시나리오를 가정하여 음성 명령과 물체 인식에 기반한 로봇 작업 계획 방법을 제안하였다. 제안된 작업 계획은 음성(언어)에 기반하여 작동하도록, 음성의 키워드 매칭을 통해 로봇이 수행해야 할 작업을 Case로 분류하였으며, YOLO v4 모델을 통해 라벨링 된 데이터를 전이 학습하여 생성된 객체 탐지 모델을 이용해 컵의 상태를 인식하고 음성 명령에 맞는 동작을 수행하였다.

이러한 방법은 사용자의 의도에 부합하는 작업을 언어적 상호작용에 기반해 수행할 수 있다는 점에서 이점을 보인다. 하지만 어휘의 다양성, 작업환경의 복잡도 등에 따라 인식률의 한계가 명확하다. 따라서 음성 인식 측면의 향후 연구에서는, 고도화된 자연어 처리 기술을 사용하여 복잡한 음성 명령의 문장 구조를 파악하고 사용자의 감정을 분석해 더욱 효과적인 사용자 중심 명령 인식이 가능하도록 발전시킬 것이다.

객체 인식 측면에서는 물체의 위치를 고정했다는 가정하에 진행되었던 실험의 한계점을 극복하기 위해 삼각측량법을 사용하여 물체의 깊이 정보를 추정하는 방식으로 시스템을 개선해 나가는 노력이 필요하다. 또한 컵에 남아있는 내용물의 양을 인식해 이에 따라 로봇의 가속도를 조절하여 양이 적을 때는 빠른 동작을, 양이 많을 때는 내용물이 넘치지 않도록 느린 동작을 수행하도록 기술을 발전시킬 계획이다. 내용물 색상 인식을 위한 추가 학습을 진행한다면, 낮은 점도를 지닌 무색 액체의 경우 본 연구의 컵 처리 과정을 그대로 수행, 높은 점도를 지닌

유색 액체의 경우 추가 세척 동작을 수행하는 등 인식한 색에 따른 적절한 동작을 추가할 수 있을 것으로 기대된다.

References

[1] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajńik, "Artificial intelligence for long-term robot autonomy: A survey," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4023-4030, Oct., 2018, DOI: 10.1109/LRA.2018.2860628.

[2] M. Cho, J. Kim, D. Lee, and M. Jang, "An analysis of quality attributes and service satisfaction for artificial intelligence-based guide robot," *Journal of Korea Robotics Society*, vol. 18, no. 2, pp. 216-224, May, 2023, DOI: 10.7746/jkros.2023.18.2.216.

[3] B. Fang, J. Yu, Z. Chen, A. I. Osman, M. Farghali, I. Ihara, E. H. Hamza, D. W. Rooney, and P.-S. Yap, "Artificial intelligence for waste management in smart cities: A review," *Environmental Chemistry Letters*, vol. 21, no. 4, pp. 1959-1989, May, 2023, DOI: 10.1007/s10311-023-01604-3.

[4] A. Bonarini, "Communication in human-robot interaction," *Current Robotics Reports*, vol. 1, no. 4, pp. 279-285, Aug., 2020, DOI: 10.1007/s43154-020-00026-1.

[5] H. Jeon, J. Kang, and B.-Y. Kang, "Deep reinforcement learning-based cooperative robot using facial feedback," *Journal of Korea Robotics Society*, vol. 17, no. 3, pp. 264-272, Aug., 2022, DOI: 10.7746/jkros.2022.17.3.264.

[6] Z. Shi, L. Zhang, and D. Wang, "Audio-visual sound source localization and tracking based on mobile robot for the cocktail party problem," *Applied Sciences*, vol. 13, no. 10, pp. 1-14, May, 2023, DOI: 10.3390/app13106056.

[7] P. Haghighatkhah, A. Fokkens, P. Sommerauer, B. Speckmann, and K. Verbeek, "Better hit the nail on the head than beat around the bush: Removing protected attributes with a single projection," *2022 Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, United Arab Emirates, pp. 8395-8416, 2022, DOI: 10.18653/v1/2022.emnlp-main.575.

[8] A. Martin, G. Quispe, C. Ollion, S. Le Corff, F. Strub, and O. Pietquin, "Learning natural language generation with truncated reinforcement learning," *2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, USA, pp. 12-37, 2022, DOI: 10.18653/v1/2022.naacl-main.2.

[9] A. Mukanova, M. Milosz, A. Dauletkaliyeva, A. Nazyrova, G. Yelibayeva, D. Kuzin, and L. Kussepova, "LLM-powered natural language text processing for ontology enrichment," *Applied Sciences*, vol. 14, no. 13, pp. 1-14, Jul., 2024, DOI: 10.3390/app14135860.

[10] H. Z. Ting, M. H. M. Zaman, M. F. Ibrahim, and A. M. Moubark, "Kinematic analysis for trajectory planning of open-source 4-DoF robot arm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 768-775, Jun., 2021, DOI: 10.14569/IJACSA.2021.0120690.

박 무 열

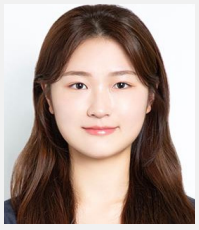
2019~현재 세종대학교 지능기전공학부 학사과정

관심분야: Human-Robot Interaction, Multimodal Interfaces

정 우 주

2021~현재 세종대학교 지능기전공학부 학사과정

관심분야: Multi-Spectral Sensor Fusion, Guidance and Control



김 민 조

2021~현재 세종대학교 지능기전공학부 학사과정

관심분야: Artificial Intelligence, Multimodal Interfaces

박 규 민

2016 서울대학교 전기정보공학(학사)

2022 서울대학교 기계항공공학(박사)

2022~2023 KIST 지능로봇연구단 Post-doc.

2023~현재 세종대학교 AI로봇학과 조교수

관심분야: Human-Robot Interaction, Artificial Intelligence, Robot Modeling, Planning and Control