

# 음성인식과 딥러닝 기반 객체 인식 기술이 접목된 모바일 매니퓰레이터 통합 시스템

## Integrated System of Mobile Manipulator with Speech Recognition and Deep Learning-based Object Detection

장동열<sup>1</sup>, 유승열<sup>†</sup>

Dongyeol Jang<sup>1</sup>, Seungryeol Yoo<sup>†</sup>

**Abstract:** Most of the initial forms of cooperative robots were intended to repeat simple tasks in a given space. So, they showed no significant difference from industrial robots. However, research for improving worker's productivity and supplementing human's limited working hours is expanding. Also, there have been active attempts to use it as a service robot by applying AI technology. In line with these social changes, we produced a mobile manipulator that can improve the worker's efficiency and completely replace one person. First, we combined cooperative robot with mobile robot. Second, we applied speech recognition technology and deep learning based object detection. Finally, we integrated all the systems by ROS (robot operating system). This system can communicate with workers by voice and drive autonomously and perform the Pick & Place task.

**Keywords:** Mobile Manipulator, Service Robot, Speech Recognition, Object Detection

### 1. 서 론

협동 로봇은 산업용 로봇보다 훨씬 공간을 차지하지 않는다는 장점이 있어 스마트공장에서 요구하는 모듈화나 분산화와 같은 유연성을 확보하는 데 중요한 역할을 하였다. 이로 인해 협동 로봇이 스마트공장의 한 축으로 자리 잡을 수 있었고 전 세계적으로 로봇 시장 또한 가파른 성장세를 보이고 있다. 우리나라 역시 인구가 감소하는 추세가 이어지고 주 52시간 근무제 도입 및 최저임금 상승 등에 대한 해결책으로 협동 로봇의 중요성이 확대되고 있다<sup>1)</sup>.

협동 로봇의 초기 형태는 지정된 공간에서 단순 작업을 반복하는 용도가 대부분이었기 때문에 산업용 로봇과 큰 차이점을 보이지 못했다. 하지만 최근 들어 제품의 생산성을 향상시키기 위한 목적이 아닌 작업자의 생산성을 향상시키거나 인간

의 제한된 근무시간을 보완하기 위한 목적으로 연구가 확대되고 있다. 대표적으로 Omron, Kuka, 한화정밀기계 등의 회사에서는 협동 로봇과 모바일 로봇을 결합한 모바일 매니퓰레이터를 제작하여 협동 로봇의 제한된 활동 영역을 보완하였다. 이를 통해 협동 로봇이 여러 위치에서 다양한 동작을 수행할 수 있게 되었으며 한 사람을 대체할 수 있는 형태의 로봇으로 발전하였다. 또한 AI기술 적용을 통해 전문적인 서비스 로봇으로 활용하려는 시도가 활발하게 이루어지고 있으며 시장의 규모 또한 크게 성장할 것으로 전망되고 있다. 기술적 요인뿐만 아니라 고령화, 저성장 시대 도래, 삶의 질 향상 추구와 같은 사회적 요인들도 서비스 로봇 시장의 성장을 촉진하고 있다<sup>2)</sup>.

이와 같은 사회적 변화에 맞춰 본 연구에서는 협동 로봇을 활용하여 작업자의 업무 효율을 향상시키고 한 사람을 완전히 대체할 수 있는 형태의 서비스 로봇을 제작하였다. 먼저 협동 로봇과 모바일 로봇을 결합한 모바일 매니퓰레이터를 제작하고 딥러닝 기반의 객체인식 기술을 통해 물체의 Pick & Place 기능을 구현하였다. 현재 상용화된 모바일 매니퓰레이터에 음성인식 기술이 접목된 제품은 존재하지 않는다. 본 논문에서 제작한 통합 시스템을 통해 음성인식이 접목된 모바일 매니퓰레이터의 활용성을 실험을 통해 증명하였다. 이를 통해 작업자가

Received : Apr. 5. 2021; Revised : May. 4. 2021; Accepted : Jul. 8. 2021

※ This project was funded by Korea Robotics Society (KROS), and is currently supported by the publication grant; This paper was supported by Education and Research promotion program of KOREATECH in 2021.

1. MS Student, Korea University of Technology and Education, Cheonan, Korea (wkdehdduf95@koreatech.ac.kr)

† Professor, Corresponding author: Mechanical Engineering, Korea University of Technology and Education, Cheonan, Korea (yoos@koreatech.ac.kr)

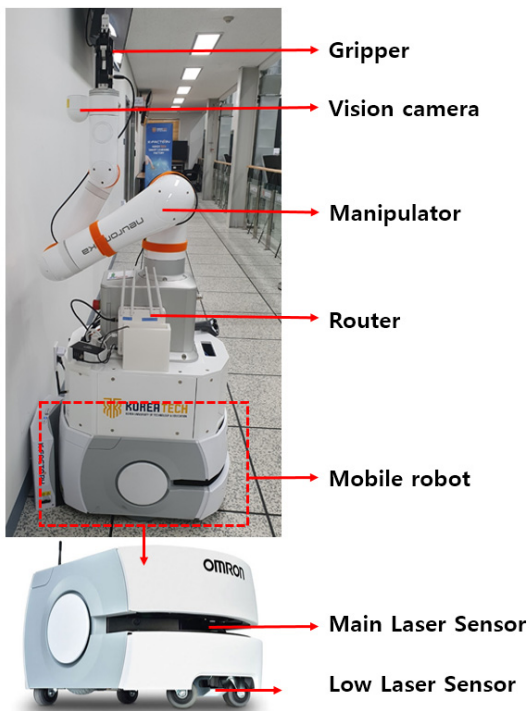
로봇과 의사소통 할 수 있는 기능을 구현함으로써 협업 효율을 증가시킬 수 있었다. 최종적으로 ROS (Robot Operating System)을 통해 시스템 통합을 수행하였으며, 이 시스템은 모바일 로봇 실행 노드, 매니퓰레이터 및 비전카메라 실행 노드, 음성인식 노드 총 3개의 실행 노드로 구성되어 있다.

## 2. 하드웨어 구성 및 통신 체계

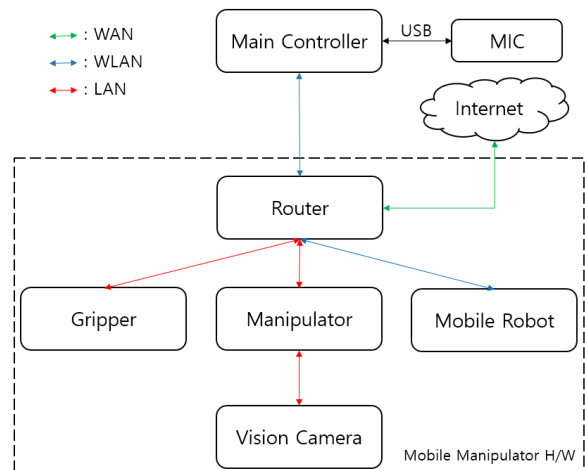
본 연구에서 제작한 모바일 매니퓰레이터는 [Table 1]과 같이 총 5개의 구성요소로 이루어져 있으며 시뮬레이션 환경이 아닌 실제 환경에서 검증이 가능하도록 [Fig. 1]과 같이 시스템을 통합하여 제작하였다. DC 12 V, DC 22-30 V 출력이 가능한 배터리를 통해 비전카메라를 제외한 모든 구성요소에 전원을 공급하며 비전카메라는 매니퓰레이터로부터 스탭다운 컨버터를 통해 전원을 공급받는 구조로 이루어져 있다.

[Table 1] Hardware components and spec of rated voltage

Components	Corporate / Product name	Rated voltage
Manipulator	Neuromeka / Indy7	24 V
Vision camera	Neuromeka / IndyEye	12 V
Mobile robot	Omron / LD-90	26 V
Gripper	DH robotics / AG-95	24 V
Router	ipTIME / A3008-MU	12 V



[Fig. 1] Mobile manipulator H/W



[Fig. 2] Communication System

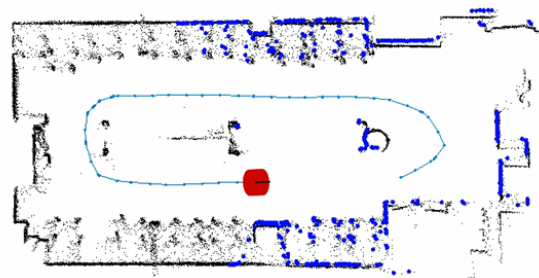
모바일 매니퓰레이터의 구성요소들은 [Fig. 2]와 같이 유선 혹은 무선으로 통신하는 구조로 이루어져 있다. 통합 제어기인 PC를 기반으로 외부 마이크를 통해 음성 데이터를 입력받고, 인식된 내용은 ROS service 통신으로 모바일 매니퓰레이터의 실행 노드로 전달된다.

## 3. 적용 기술

### 3.1 모바일 로봇

모바일 로봇은 협동 로봇을 이송하는 역할을 하며 레이저 센서와 후방 초음파 센서를 통해 장애물을 인지하고 새로운 경로를 생성하거나 비상정지 할 수 있다. 하지만 [Fig. 1]에서 확인할 수 있는 바와 같이 모바일 로봇의 중간 높이에 위치한 메인 레이저 센서는 작업 현장 바닥에 떨어져있는 작은 물체를 검출하지 못하는 경우가 발생할 수 있으므로 모바일 로봇 하단부에 보조 레이저를 설치하여 보다 안정적인 주행이 가능하도록 하였다.

모바일 로봇의 자율주행 이동을 위해서는 [Fig. 3]과 같이 SLAM (Simultaneous Localization And Mapping) 기법을 통해



[Fig. 3] Generate map by SLAM

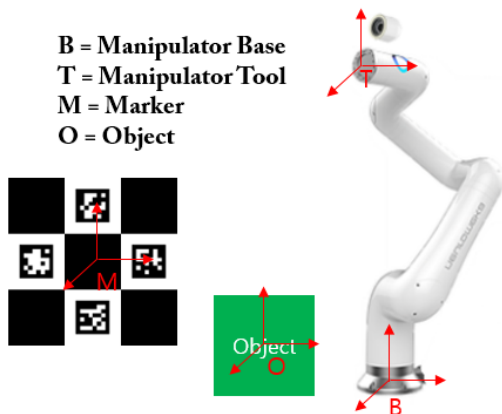
맵을 생성해야한다<sup>[3]</sup>. 맵 생성을 위해서는 제품의 전용 소프트웨어를 이용하여 작업 환경을 메인 레이저 센서로 스캔하고 Point cloud 데이터를 축적한다. 이후 맵 생성을 시작한 지점과 같은 위치에서 맵 생성을 종료하면 Loop closure 기법을 통해 누적된 측정 오차를 보정하고 최종 맵을 생성하게 된다. 최종 생성된 맵에서 특정 작업 위치를 좌표 지점과 모바일 로봇의 회전 각도로 설정할 수 있다. 이를 통해 주행의 안전성을 확보하고 원하는 지점으로 협동 로봇을 이동시킬 수 있게 된다.

### 3.2 매니플레이터 및 비전카메라

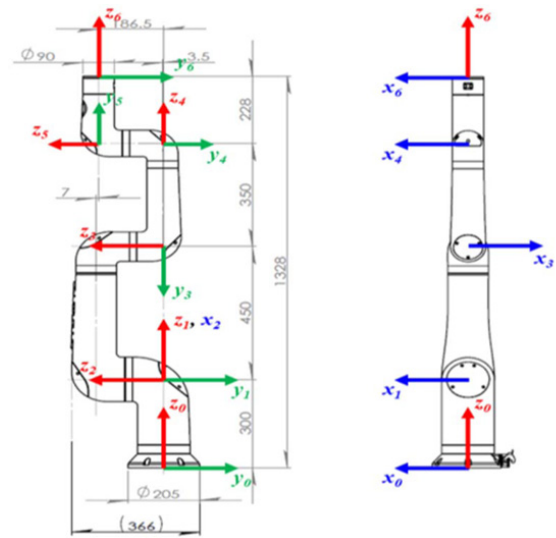
보편적인 물체 Pick up 동작은 매니플레이터와 물체가 항상 동일한 위치에 있기 때문에 좌표 Teaching만으로 구현할 수 있다. 하지만 모바일 매니플레이터는 자율주행으로 인하여 로봇의 작업 위치가 항상 동일하지 않기 때문에 위치 오차를 보정 해주어야 한다. 본 연구에서는 Pick up 동작 수행 시 Marker를 인식하여 자율주행으로 인해 발생하는 Teaching 좌표의 오차를 보정하였다. 여기서 Marker는 가로·세로 각각 90 mm 크기의 ChAruco marker를 사용하였다. ChAruco marker는 색상 구분이 뚜렷한 Checker board 환경에서 Aruco marker를 검출하여 자세를 추정하기 때문에 인식률이 뛰어나며, 코너점 보간법을 사용하여 카메라의 왜곡보정을 동시에 진행함으로써 보다 정확한 좌표축을 검출할 수 있는 장점이 있다<sup>[4]</sup>.

본 연구에서 구현한 로봇의 Pick & Place 동작은 사전 좌표 Teaching 단계, 수행 단계, Place 단계, 총 3단계로 구성되어 있으며 [Fig. 4]와 같이 좌표계를 설정하였다.

Teaching 단계와 수행 단계에서 사용한 HT (Homogeneous Transformation) matrix는 Denavit-Hartenberg<sup>[5]</sup> parameter 표현 방법과 동일하게 사용하였으며, [Fig. 5]와 같이 협동 로봇의 좌표계를 설정하였고 [Table 2]와 같이 DH parameter를 정의하였다.



[Fig. 4] Coordinate Settings



[Fig. 5] Manipulator with assigned link parameters

[Table 2] DH parameter of manipulator

n	$\alpha_{i-1}$	$a_{i-1}$	$d_i$	$\theta_i$
1	0	0	0.3	$\theta_1$
2	90°	0	0	$\theta_2 + 90^\circ$
3	0	0.45	0	$\theta_3 + 90^\circ$
4	90°	0	0.35	$\theta_4 + 180^\circ$
5	90°	0	0.1865	$\theta_5$
6	-90°	0	0.228	$\theta_6$

#### 3.2.1 Teaching 단계

먼저 비전카메라를 통해 Marker를 인식할 수 있는 Detect position으로 매니플레이터를 이동하고 각 관절 각도를 저장한다. 그리고 인식 위치에서 로봇의 Base 좌표 기준 Marker 좌표까지의 HT matrix  ${}^B T_M$ 을 계산한다. 동일한 위치에서 로봇의 Base 좌표 기준 로봇의 Tool 좌표까지의 HT matrix  ${}^B T_T$ 을 계산한다. 여기서 Marker 좌표 기준 Tool 좌표까지의 HT matrix를 다음 식으로 구할 수 있다.

$$({}^B T_M)^{-1} \times {}^B T_T = {}^M T_T \tag{1}$$

Marker 좌표를 기준으로 물체의 좌표가 항상 고정이라면 인식 위치에서 매니플레이터의 Tool 좌표 기준 물체의 좌표 또한 고정이다. 이 사실을 이용하여 Marker 인식 위치에서 매니플레이터의 Tool 좌표 기준 물체 좌표까지의 HT matrix는 식 (1)을 이용하여 다음과 같이 구할 수 있다.

$$({}^M T_T)^{-1} \times {}^M T_O = {}^T T_O \tag{2}$$

### 3.2.2 Pick up 단계

실제 자율주행을 통해 작업 위치로 이동한 모바일 매니플레이터는 Teaching과정에서 저장한 Detect position의 관절 각도로 이동한다. 그리고 현재 로봇의 Base 좌표 기준 Marker 좌표까지의 HT matrix  ${}^B T_M$ 을 계산한다. 여기서 식 (1)에서 구한  ${}^M T_T$ 을 곱해주면 현재 로봇의 Base 좌표 기준 Teaching 당시 Tool 좌표까지의 HT matrix를 구할 수 있다.

$${}^B T_M \times {}^M T_T = {}^B T_T \quad (3)$$

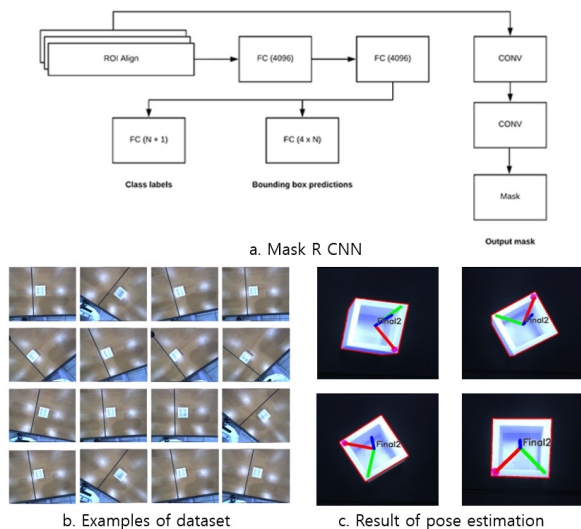
식 (3)에서 구한  ${}^B T_T$ 에 식 (2)에서 구한  ${}^T T_O$ 를 곱해주면, 바뀐 로봇의 Base 좌표 기준 물체 좌표까지의 HT matrix를 구할 수 있다.

$${}^B T_T \times {}^T T_O = {}^B T_O \quad (4)$$

이와 같은 방법을 통해 매번 매니플레이터의 작업 위치가 달라지더라도 Teaching 좌표를 보정할 수 있어 물체의 Pick up 동작을 반복적으로 수행할 수 있게 된다. 이러한 방식은 항상 정해진 위치에 존재하는 물체에 대해서만 Pick up 동작 수행이 가능하다. 또한, 각 물체마다 Marker가 추가되어야 하며 Teaching 작업이 필요하다. 하지만 임의의 위치에 존재하는 물체에 대한 Pick up 동작에 필요한 딥러닝 학습, 객체 인식을 위한 막대한 연산 등의 과정이 생략되기 때문에 제안한 방법은 한 번의 Teaching만으로 빠르고 정확한 Pick up 동작을 수행할 수 있다.

### 3.2.3 Place 단계

본 논문에서는 [Fig. 6]과 같이 딥러닝 기반 개체인식 Framework중 하나인 Mask R CNN (Convolutional Neural Network)<sup>[6]</sup>



[Fig. 6] Pose estimation of Container

을 활용하여 Object place 기능을 구현하였다. 먼저 물체를 배치할 수 있는 컨테이너를 STL파일 형식으로 모델링하고 제작한다. 이후 비전카메라를 이용하여 [Fig. 6]와 같이 다양한 각도에서 컨테이너 사진을 촬영한 후 모델링한 STL파일을 촬영한 사진에 투영한다. 이때 사영된 영역을 Mask로 선정하여 Data set을 생성하고 Mask R CNN을 통해 형상을 학습한다.

학습이 완료된 후 컨테이너를 인식하면 제작한 컨테이너가 대칭의 성질을 가지고 있기 때문에 좌표축의 원점을 기준으로 x축 방향이 랜덤하게 생성되는 현상이 발생한다. 본 연구에서는 x축의 기준점이 될 수 있는 붉은색 참조 형상을 컨테이너에 부착하여 일관적인 좌표축을 생성할 수 있도록 하였다.

먼저 붉은색 참조 형상 영역을 추출하여 Mask를 생성하고 생성된 Mask의 픽셀 값을 이용하여 Center of mass equation<sup>[7]</sup>을 적용하면 참조 형상의 중심점을 찾을 수 있게 된다. 이때 컨테이너의 원점으로부터 참조 형상의 중심점 방향으로 x축을 생성하면 [Fig. 6]와 같이 항상 일관된 좌표축을 생성할 수 있게 된다.

### 3.3 음성 인식

음성인식 오픈 API를 제공하는 회사는 대표적으로 Google, 카카오, NAVER 등이 있다. 그중 지원하는 컴퓨터 언어가 가장 다양하고 인식률이 뛰어나며 약 80여개의 언어를 지원하는 Google의 Cloud Speech Recognition API를 사용하여 음성 인식 기능을 구현하였다<sup>[8]</sup>.

Google API는 음성이 인식되면 오디오 데이터를 인터넷 환경을 통해 API를 거쳐 텍스트로 실시간 번역해주는 구조이기 때문에 인터넷 연결이 필수적이다. 따라서 본 연구에서는 무선 WAN을 통해 외부 인터넷 망과 통신하여 Google API를 사용할 수 있는 환경을 조성하였다.

오픈 API를 통해 음성이 텍스트로 변환되면 먼저 [Table 3]에 표기된 사용자 정의 문자가 변환된 텍스트에 포함되어 있는지를 판단한다. 이때 사용자 정의 문자가 포함되어 있으면 문자에 해당하는 숫자로 음성 인식의 결과를 변형한다. 이 숫자는 ROS service 통신을 통해 모바일 매니플레이터 실행 노드로 전달되며, 모바일 매니플레이터 실행 노드는 각 숫자에 해당하는 동작을 수행하게 된다.

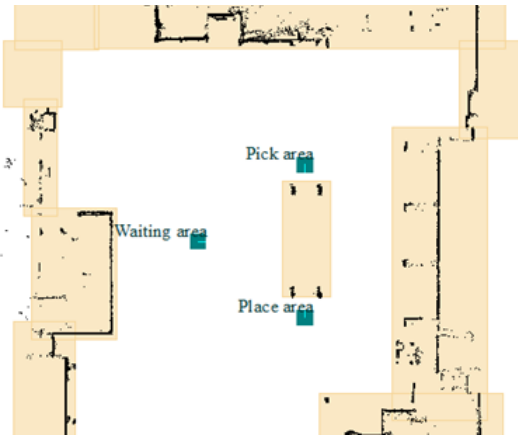
[Table 3] List of custom characters

Custom characters	Number	Motions
Hello	1	Wave robot arm
Start	2	Move to goal point and perform pick & place
Stop	3	Emergency stop
Termination	4	Program termination

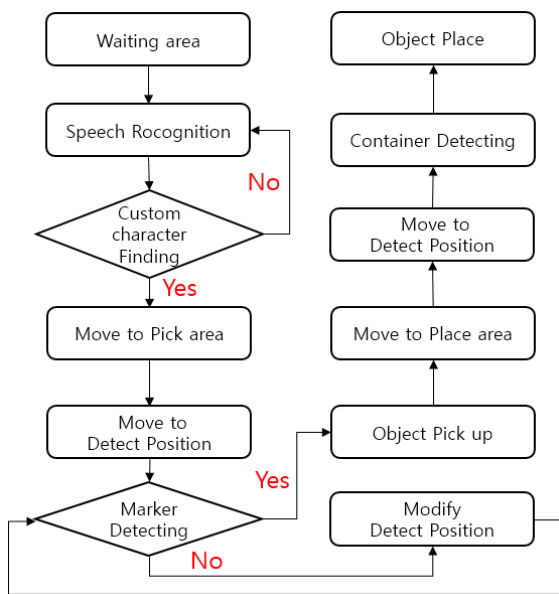
### 4. 서비스 작업 구현 및 성능 실험

제작한 통합 시스템의 활용성 검증을 위하여 음성 명령을 통한 물체 이송 서비스 작업을 구현하였고 성능 실험 환경을 다음과 같이 구성하였다. 먼저 [Fig. 7]과 같이 테이블의 양 끝 지점을 각각 작업 장소(Pick area)와 배치 장소(Place area)로 정의하고 작업 장소에 오차 보정을 위한 Charuco marker와 Pick up 동작 대상 물체인 큐브를 임의의 위치에 고정시킨다. 로봇이 작업 장소에 정확하게 위치하고 있는 상태에서 물체의 Pick up 동작을 수행할 수 있는 좌표 Teaching을 시행하고 배치 장소에는 큐브를 배치할 수 있는 크기의 컨테이너를 임의의 장소에 위치시킨다.

실험은 [Fig. 8]의 순서도와 같이 대기 장소에서 음성 명령을 전달받은 모바일 매니플레이터가 작업 장소로 이동하여 큐브의 Pick up 동작을 수행한 후 배치 장소로 이동하여 컨테이



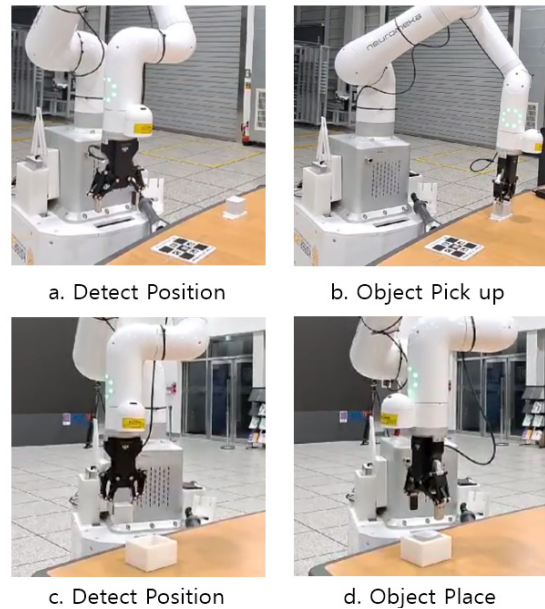
[Fig. 7] Map for transfer service



[Fig. 8] Transfer service flow chart

너에 물체의 Place 동작을 수행하는 과정을 총 20회 반복하여 성공 횟수를 측정한다. 이때 주행 거리는 약 6 m이며 [Fig. 9]와 같이 물체의 Pick & Place 동작을 완벽하게 수행했을 시 성공으로 간주한다.

실험 결과 [Table 4]와 같이 총 20번의 시도 중 16번을 성공하였다. 7번째와 17번째 시도에서는 빛의 영향으로 컨테이너의 좌표를 추출하지 못하였고, 4번째와 10번째 시도에서는 음성이 인식되지 않았다. 물체의 Pick up 동작 성공률은 100%를 보였으며 전체 서비스 작업 성공률은 80%를 보였다. 소요 시간은 [Table 5]와 같으며 이를 통해 제작한 시스템의 서비스 로봇으로의 활용 가능성을 확인할 수 있었다.



[Fig. 9] Implementation of Pick & Place

[Table 4] Results of experiment (S = success, F = Fail)

Trial	result	Trial	result	Trial	result	Trial	result
1	S	6	S	11	S	16	S
2	S	7	F	12	S	17	F
3	S	8	S	13	S	18	S
4	F	9	S	14	S	19	S
5	S	10	F	15	S	20	S

[Table 5] Time for complete the service (Second)

Trial	result	Trial	result	Trial	result	Trial	result
1	72	6	73	11	73	16	71
2	75	7	-	12	72	17	-
3	72	8	75	13	72	18	72
4	-	9	74	14	73	19	73
5	73	10	-	15	75	20	72

## 5. 결 론

본 연구에서는 음성으로 명령을 입력 받은 모바일 매니플레이터가 작업 장소로 이동하여 큐브를 Pick up하고 배치 장소로 이동하여 컨테이너에 Place하는 동작을 구현하였다. 또한 반복 실험을 통해 제작한 통합 시스템의 서비스 로봇으로의 활용할 수 있음을 검증하였다. 서비스 작업에 소요되는 시간은 사람이 수행했을 때보다 오래 걸리지만, 단순 작업을 로봇이 수행하도록 하고 그 시간 동안 인간은 창의적인 일에 집중함으로써 이상적인 생산 환경을 창출할 수 있을 것이다.

제작한 통합 시스템은 불편한 사람들에게 필요한 물건을 집어다 주는 로봇, 식당·카페 등에서 음식을 배달하는 로봇, 제조업 현장에서 작업 도구 혹은 제품을 이송해주는 로봇 등 매우 다양한 서비스 분야에서 활용이 가능하다는 점에 의의가 있다. 추후 음성인식과 객체인식 기술의 정확도를 높이는 연구를 통해 시스템의 성능을 향상시키고 얼굴 인식, 빈 피킹<sup>9)</sup> 등 다양한 AI기술을 추가적으로 적용하여 서비스 로봇으로의 활용성을 더욱 높이하고자 한다.

## References

- [1] H. J. Yoo and H. B. Kim, "Collaborative robot for manufacturing," *Korea Institute of S&T Evaluation and Planning*, Eumseong, Chungcheongbuk, Korea, [Online], <https://www.kistep.re.kr/>.
- [2] Y. G. Kim, "ICT Spot Issue (2017-06) The Next Big Thing, Trend and implication of Service Robot," *Institute for Information & Communication Technology Planning & Evaluation*, Daejeon, Korea, [Online], <https://www.iitp.kr>.
- [3] Omron, [Online], <http://asq.kr/rMZb6MZE1fugB>, Accessed: January 29, 2021.
- [4] D. Hu, D. DeTone, and T. Malisiewicz, "Deep ChArUco: Dark ChArUco Marker Pose Estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, California, USA, pp. 8436-8444, 2019, DOI: 10.1109/CVPR.2019.00863.
- [5] J. Denavit and R. S. Hartenberg, "A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices," *Journal of Applied Mechanics*, pp. 215-221, Jun., 1955, DOI: 10.1115/1.4011045.
- [6] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2961-2969, 2017, DOI: 10.1109/ICCV.2017.322.
- [7] Center of mass - Wikipedia, [Online], [https://en.wikipedia.org/wiki/Center\\_of\\_mass](https://en.wikipedia.org/wiki/Center_of_mass), Accessed: January 29, 2021.
- [8] S. Choi and J. Kim, "Comparison Analysis of Speech Recognition Open APIs' Accuracy," *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, vol. 7, no. 8, pp. 411-418, 2017, DOI : 10.35873/ajmahs.2017.7.8.038.
- [9] F. Spennath and A. Pott, "Gripping Point Determination for Bin Picking Using Heuristic Search," *10th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '16*, Ischia, Italy, pp. 606-611, 2017, DOI: 10.1016/j.procir.2016.06.015.



### 장 동 열

2019 한국기술교육대학교 기계공학과(학사)  
2019~현재 한국기술교육대학교 기계공학과  
(석사과정)

관심분야: 서비스 로봇, 영상처리, 인공지능



### 유 승 열

1990 KAIST 기계공학(학사)  
1992 POSTECH 기계공학(석사)  
2004 Rensselaer Polytechnic Institute  
기계공학(박사)  
2006~현재 한국기술교육대학교 교수

관심분야: 로봇, 자동화, 인공지능