

멀티모달 인지 기반 모바일 매니퓰레이터의 자율 승강기 인터랙션 연구

정도영*, 홍혁주*, 주승현*, 최재훈*, 김정주*
호서대학교*

A Study on Autonomous Elevator Interaction of Mobile Manipulators Based on Multimodal Perception

Do-Young Jeong*, Hyeok-Joo Hong*, Seung-Hyeon Joo*, Jae-Hoon Choi*, Jung-Ju Kim*[†]
Hoseo University*

Abstract - 본 논문은 휴머노이드 및 이동형 서비스 로봇의 자율 승강기 이용을 위한 멀티모달 인지 기반 자율 승강기 인터랙션 구조를 제안한다. 기존 건물의 승강기는 대부분 인간 사용자를 기준으로 설계되어 있으며, 로봇과 직접 연동 가능한 표준 제어 API(Application Programming Interface)나 통신 인터페이스를 제공하지 않는 경우가 많다. 이에 따라 로봇이 기존 승강기 인프라를 개조하지 않고 다층 이동을 수행하기 위해서는 승강기가 인간에게 제공하는 버튼 패널, 층수 정보, 음성 안내와 같은 시각·청각 정보를 직접 인식하고 이를 조작 행동으로 연결해야 한다. 선행연구에서도 기존 건물 환경에서 로봇이 인간용 인터페이스로 설계된 승강기를 조작할 필요성이 제시되었으며, 모바일 매니퓰레이터(Mobile Manipulator)의 다층 이동을 위해 승강기 버튼 조작 능력이 중요한 요소로 보고되었다. 제안 구조는 컴퓨터 비전 기반 버튼 검출, 광학 문자 인식(Optical Character Recognition, OCR) 기반 층수 인식, 목표 버튼 판단, 청각 기반 승강기 상태 인식, 카메라-로봇 좌표 변환, 로봇팔의 접근 - 정렬 - 누름 동작 생성 과정으로 구성된다. 실험에서는 목표 버튼 인식 결과, 안내 음성 인식 결과 및 버튼 누름 결과를 중심으로 제안 구조의 적용 가능성을 평가하였다. 이를 통해 승강기 버튼 인식 및 상태 인식 결과가 실제 로봇 조작 행동으로 연결될 수 있음을 확인하고, 기존 승강기 인프라 비개조 조건에서 이동형 서비스 로봇의 자율 승강기 이용 가능성을 제시한다.

1. 서 론

승강기는 다층 건물에서 사람과 물류의 수직 이동을 담당하는 핵심 설비이며, 국내 전기공학 분야에서도 승강기 안전성 및 검사 기법에 관한 연구가 지속적으로 수행되어 왔다[1]. 또한 서비스 로봇, 순찰 로봇, 안내 로봇 및 물류 보조 로봇은 병원, 공공기관, 연구시설, 산업 현장과 같은 실내 다층 건물 환경에서 활용 범위가 확대되고 있다. 이러한 로봇이 단일 층에 제한되지 않고 연속적인 임무를 수행하기 위해서는 층간 이동 능력이 필요하며, 승강기 이용은 다층 건물에서 가장 현실적인 이동 수단이다. 그러나 기존 승강기는 대부분 인간 사용자를 기준으로 설계되어 있어, 로봇과 직접 연동 가능한 표준 제어 API나 통신 인터페이스를 제공하지 않는 경우가 많다. 기존 건물 환경에서 로봇이 인간용 인터페이스로 설계된 문과 승강기를 직접 다룰 필요성이 보고되었으며[2], 모바일 매니퓰레이터의 다층 운동을 위해 승강기 버튼 조작 능력이 중요한 기능으로 제시되었다[3]. 또한 로봇이 승강기와 직접 통신할 수 없는 경우에는 인간과 유사하게 승강기 버튼을 인식하고 조작해야 한다[4]. 따라서 기존 승강기 인프라를 개조하지 않고 로봇이 승강기를 이용하기 위해서는 승강기가 제공하는 시각·청각 정보를 직접 인식하고 이를 행동으로 연결할 수 있어야 한다. 시각 정보는 버튼 패널 검출, 개별 버튼 인식, OCR 기반 층수 인식 및 목표 버튼 판단에 활용될 수 있으며, 기존 연구에서도 승강기 버튼 인식과 YOLO(You Only Look Once)-OCR 기반 문자 인식 방법이 제안되어 왔다[5-7]. 그러나 이러한 연구들은 주로 버튼 검출 또는 문자 인식 성능에 초점을 두며, 인식 결과를 로봇 기준 좌표 변환과 실제 버튼 누름 동작으로 연결하는 전체 인터랙션 구조를

충분히 다루지 않는다. 따라서 본 연구는 기존 승강기 인프라를 개조하지 않는 조건에서 로봇이 목표 층 버튼을 인식하고 조작하기 위한 멀티모달 인지 기반 자율 승강기 인터랙션 구조를 제안한다. 제안 구조는 시각 기반 버튼 검출, OCR 기반 층수 인식, 목표 버튼 판단, 청각 기반 승강기 상태 인식, 카메라-로봇 좌표 변환, 로봇팔의 접근 - 정렬 - 누름 - 후퇴 동작 생성을 하나의 인지 - 판단 - 행동 흐름으로 연결한다. 본 논문에서는 버튼 검출 결과, 안내 음성 인식 결과, 버튼 누름 결과를 중심으로 제안 구조의 적용 가능성을 실험적으로 확인한다.

2. 본 론

본 연구는 AMR(Autonomous Mobile Robot) 위에 로봇팔을 장착한 모바일 매니퓰레이터 시스템을 기반으로 자율 승강기 인터랙션 기능을 구현하였다. 로봇이 건물 내 다층 환경에서 자율적으로 이동하기 위해서는 승강기 호출, 목표 층 판단, 승강기 안내 음성 인식, 버튼 누름 동작이 순차적으로 수행되어야 한다. 이에 본 연구에서는 승강기 이용 과정에서 요구되는 인식 및 조작기능을 핵심 논리로 적용하여 시스템을 설계하고, 각 기능의 수행 가능성을 실험적으로 확인하였다.

2.1 멀티모달 인지 기반 자율 승강기 인터랙션 구조

제안 구조는 로봇이 승강기 버튼 패널을 인식하고 목표 버튼을 조작하기 위한 인지 - 판단 - 행동 연결 구조이다. 전체 시스템은 시각 인지 모듈, 청각 인지 모듈, 목표 층 판단 모듈, 좌표 변환 모듈, 로봇 행동 생성 모듈 및 통신 모듈로 구성된다. 시각 인지 모듈은 카메라 영상으로부터 승강기 버튼을 검출하고, OCR을 통해 층수 정보를 인식한다. 목표 층 판단 모듈은 사용자 또는 상위 제어 시스템으로부터 입력된 목표 층 정보와 OCR 결과를 비교하여 조작 대상 버튼을 선택한다. 선택된 목표 버튼의 영상 좌표는 좌표 변환 모듈을 통해 로봇 기준 좌표계의 3차원 위치로 변환된다. 이후 로봇 행동 생성 모듈은 변환된 목표 위치를 기반으로 접근, 정렬, 누름 및 후퇴 동작을 생성하며, 생성된 제어 명령은 TCP/IP 기반 통신을 통해 로봇 제어기로 전달된다. 청각 인지 모듈은 승강기 음성 안내를 기반으로 문 개폐 상태, 이동 방향 및 도착 층 정보를 판단하기 위한 확장 구성 요소로 정의하였다.

2.2 시각 인지 : 버튼 검출 및 목표 버튼 판단

버튼 검출 단계는 RGB-D 카메라가 수집한 승강기 버튼 패널 영상으로부터 조작 가능한 버튼을 검출하고, 그중 목표 층에 해당하는 버튼을 선택하는 과정이다. 본 연구에서는 객체 검출 기반 알고리즘을 사용하여 버튼 후보 영역, 클래스 정보 및 신뢰도 값을 획득한다. 검출된 버튼 후보 중 층수 버튼 영역에는 OCR을 적용하여 숫자 또는 문자 정보를 인식한다. 승강기 버튼 패널은 조명 반사, 촬영 각도, 버튼 재질, 글자 크기 및 패널 색상에 따라 인식 난이도가 달라질 수 있다. 따라서 본 연구에서는 패널 외곽을 단일 객체로 판단하기보다 개별 버튼 검출 결과와 버튼들의 공간적 배열을 중심으로 목표 버튼을 판단한다. 이후 입력된 목표 층 정보와 OCR 결과를 비교하여 최종 조작 대

상 버튼을 선택한다. 본 단계의 출력은 로봇 조작을 위한 목표 버튼의 영상좌표 (u, v) 이다. 여기서 목표 버튼의 중심 좌표는 관심 영역(Region of Interest, ROI)의 중심점으로 정의되며, 이후 액션 변환 모듈에서 카메라 좌표계 및 로봇 기준 좌표계로 변환되는 입력값으로 사용된다. 따라서 버튼 검출 결과는 단순한 객체 인식 결과가 아니라, 이후 좌표 변환 및 로봇 누름 동작 생성을 위한 기준 입력으로 사용된다. OCR 신뢰도가 낮거나 인식이 실패한 버튼에 대해서는 버튼 배열 기반 층수 추론을 적용하였다. 검출된 버튼 중심 좌표 (u, v) 를 이용하여 승강기 패널의 2차원 행·열 구조를 파악하고, OCR 신뢰도가 높은 버튼을 기준점으로 설정하여 OCR 인식 실패 버튼의 층 번호를 추론하고 이를 통해 목표 버튼 판단을 도와주는 후처리 과정이다.

2.3 액션 변환 : 목표 버튼 좌표 변환 및 누름 동작 생성

액션 변환은 시각 인지에서 선택된 목표 버튼 정보를 로봇팔의 조작 동작으로 변환하는 역할을 수행한다. 본 연구에서는 RGB-D 카메라로부터 RGB 영상과 깊이(Depth) 데이터를 획득한다. 이때 RGB 영상에서 목표 버튼의 중심 픽셀 좌표 (u, v) 를 산출하고 수집된 깊이 결과에서 해당 픽셀의 깊이값 Z_c 를 획득한다. 따라서 목표 버튼은 카메라에서 수집된 3차원 데이터 (u, v, Z_c) 로 표현되며, 이를 로봇 기준 좌표계의 3차원 위치로 변환해야 한다. 이 과정에서 카메라 내부 파라미터 행렬 K 는 다음과 같이 정의된다.

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

여기서 f_x, f_y 는 픽셀 단위의 초점거리이며 c_x, c_y 는 영상 평면의 주점 좌표이다. 카메라 좌표계에서의 목표 버튼 P_c 는 목표 버튼 중심 픽셀 (u, v) 에 대응하는 깊이 값 Z_c 에 따라 식(2)와 같이 계산된다.

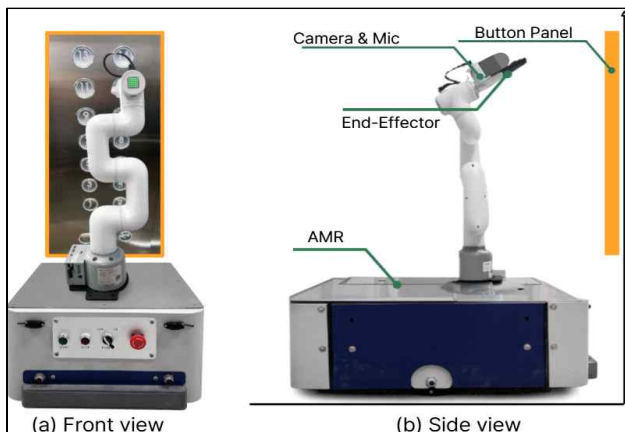
$$P_c = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = Z_c K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (2)$$

이후 핸드-아이 캘리브레이션(hand-eye calibration) 통해 획득한 변환 행렬을 이용하여 카메라 좌표계의 목표 버튼 위치를 로봇 베이스 좌표계로 변환한다.

$$\begin{bmatrix} X_b \\ Y_b \\ Z_b \\ 1 \end{bmatrix} = T_b^e(q) T_e^c \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (3)$$

여기서 $T_b^e(q)$ 는 현재 관절각 q 에 따른 로봇 베이스 좌표계에서 로봇팔 엔드이펙터 좌표계까지의 변환 행렬이고, T_e^c 는 로봇팔 엔드이펙터에서 카메라 좌표계까지의 고정 변환 행렬이다.

3. 실험



〈그림 1〉 모바일 매니퓰레이터

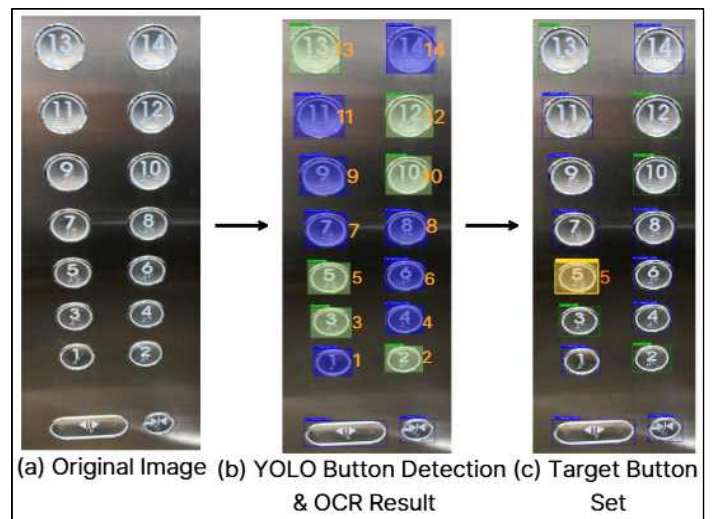
본 실험에서는 제안한 멀티모달 인지 기반 모바일 매니퓰레이터의 자율 승강기 인터랙션 구조의 적용 가능성을 검증하기 위해 시각 인지, 음성 인식, 액션 변환 실험을 수행하였다. 3.1절에서는 WLAN을 통해 목표 층 정보를 전송하고, RGB-D 카메라 영상 기반 버튼 검출 및 OCR을 통해 목표 버튼을 판단하였다. 3.2절에서는 승강기 안내 음성을 STT로 변환하여 이동 방향, 문개폐 상태 및 도착 층 정보를 인식하였다. 3.3절에서는 목표 버튼 좌표를 로봇 기준 좌표계로 변환한 후 접근-정렬-누름 동작의 수행 결과를 평가하였다. 이를 통해 인지 결과가 실제 로봇 조작 행동으로 연결되는 전체 과정을 확인하였다.

〈표 1〉 모바일 매니퓰레이터 주요 사양

항목	사양
AMR	크기(mm)
	740 x 513 x 316
	주행 방식
	SLAM
	적재 하중
매니퓰레이터	140kg
	구동 방식
	차동
	정지 정확도
	± 5mm
매니퓰레이터	최대 속도
	1m/s
	크기(mm)
	151 x 160 x 498
	자유도
매니퓰레이터	6축
	작업 반경
	320mm
매니퓰레이터	가반 하중
	1kg
매니퓰레이터	엔드이펙터
매니퓰레이터	myGripper F100

3.1 시각 인지

시각인지의 목표 버튼 검출 성능 평가를 위해 원격 PC에서 WLAN(Wireless Local Area Network)을 통해 로봇의 임베디드 컴퓨터로 목표 층 정보를 전송하였다. 자율주행 이동로봇에 탑재된 로봇팔은 승강기 버튼 패널 전방에서 RGB-D 카메라 영상을 획득하였으며, 획득된 영상에서 YOLO 기반 버튼 검출과 OCR 기반 층수 인식을 수행하였다. 실험은 1층부터 14층까지의 버튼을 대상으로 각 층별 50회씩 수행하여 총 700회의 검출 결과를 산출하였다. 검출 성공 여부는 목표 층 ROI의 중심점이 실제 목표 버튼 내부에 위치하는지를 기준으로 판단하였으며, 중심점이 버튼 외부에 위치하거나 다른 버튼으로 판단된 경우는 실패로 간주하였다. 이를 바탕으로 YOLO 단독 인식 결과와 버튼 배열 추론을 포함한 인식 결과를 비교하였다.



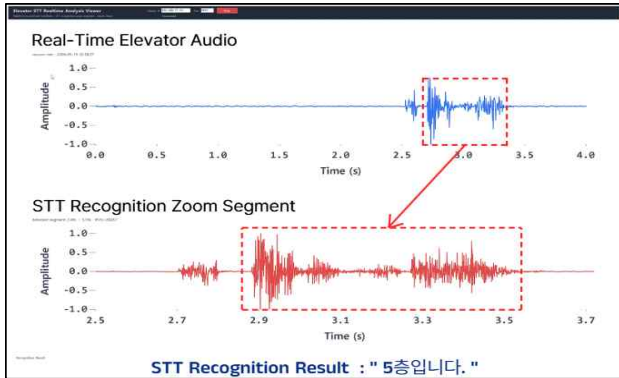
〈그림 2〉 타겟 버튼 인식

〈표 2〉 목표 버튼 인식 결과

실험 구분	시행 횟수	YOLO 단독 인식		YOLO + 배열 추론 인식	
		인식 횟수	성공률(%)	인식 횟수	성공률(%)
버튼 인식	700	612	87.43	687	98.14

3.2 음성 인식

안내 음성 인식 성능을 평가하기 위해 실험 참여자 2인의 육성 음성과 실험 대상 승강기에서 실제 출력되는 안내 멘트를 녹음한 파일을 사용하였다. 입력 음성은 이동 안내, 개폐 안내, 층수 안내로 구분하였으며, 각 수집된 음성 파일에 대해 시간영역 신호 파형을 수집하여 신호처리 STT(Speech-to-Text)를 통해 안내 문장을 문자 정보로 변환하였고 이후 변환된 문장에서 이동 방향, 문 개폐 상태, 도착 층 정보를 나타내는 핵심 키워드를 추출하였다. 본 실험을 통해 음성 안내 정보가 시각 인지 결과를 보완할 수 있으며, 시각-음성 멀티모달 구조가 승강기의 도착 층 판단 신뢰도를 향상시키는데 활용될 수 있음을 확인하였다.



〈그림 3〉 음성 신호처리 및 인식

〈표 3〉 안내 음성 인식 결과

구분	평가 음성	시행 횟수	인식 횟수	인식 성공률 (%)	비고
이동	“올라갑니다.”	30	30	100	-
	“내려갑니다.”	30	30	100	-
개폐	“문이 열립니다.”	30	30	100	-
	“문이 닫힙니다.”	30	30	100	-
층수	“□층입니다.”	420	402	95.71	층별 30회

3.3 액션 변환

액션 변환의 성능을 평가하기 위해 자율주행 이동로봇은 원격 PC가 송신한 목표 층 정보를 기반으로 RGB-D 카메라를 이용해 목표 층 버튼을 인식하고, 산출된 목표 버튼 좌표를 로봇 기준 좌표계로 변환한 뒤 접근-정렬-누름 동작을 순차적으로 수행하였다.



(a) Approach Motion (b) Target Alignment (c) Press Motion

〈그림 4〉 접근 - 정렬 - 누름

〈표 4〉 버튼 누름 결과

평가 항목	평가 기준	시행 횟수	성공 횟수	액션 성공률
접근	작업공간 도달	100	97	97.00%
정렬	버튼-엔드이펙터 중심정렬	100	98	98.00%
누름	버튼 누름	100	91	91.00%
통합	접근-정렬-누름	100	87	87.00%

실험은 총 400회 수행하였으며, 작업공간 도달, 핸드아이-버튼 중심 정렬, 버튼 누름 여부, 전체 통합 액션을 기준으로 평가하였다. 전체 통합 액션 실험 결과 목표 층 버튼을 실제로 누른 경우는 87회로 나타났으며, 액션 실패는 주로 정렬 오차 또는 접촉 깊이 부족에 의해 발생한 것으로 판단된다. 이는 향후 엔드이펙터에 촉각 피드백 장치를 추가하면 버튼 접촉 여부를 실시간으로 판단하여 누름 성공률을 향상시킬 수 있을 것으로 기대된다.

4. 결 론

본 논문에서는 기존 승강기 인프라를 개조하지 않고 휴머노이드 및 이동형 서비스 로봇이 승강기를 이용하기 위한 멀티모달 인지 기반 자율 승강기 인터랙션 구조를 제안하였다. 제안 구조는 시각 기반 버튼 검출, OCR 기반 층수 인식, 버튼 배열 기반 목표 층 추론, 청각 기반 안내 음성 인식, RGB-D 기반 목표 버튼 좌표 변환, 그리고 액션 변환을 통한 접근-정렬-누름 동작 생성으로 구성된다. 실험 결과, 버튼 배열 추론을 포함한 목표 버튼 검출 성공률은 98.14%로 나타났으며, 안내 음성 인식에서는 전체 시행 평균 96.67%의 인식 성능을 확인하였다. 또한 버튼 누름 통합 평가에서 87.00%의 성공률을 보였으며, 이를 통해 인식된 목표 버튼 정보가 실제 로봇팔의 조작 행동으로 연결될 수 있음을 확인하였다. 다만 시각 및 청각 인식 성능에 비해 실제 버튼 누름 성공률은 상대적으로 낮게 나타났으며, 이는 정렬 오차와 접촉 깊이 부족이 전체 시스템 신뢰도의 주요 제한 요소를 의미한다. 따라서 향후 연구에서는 촉각 피드백 기반 접촉 판별, 실시간 위치 보정 및 누름 깊이 제어를 적용하여 버튼 조작 안정성을 향상시킬 필요가 있다. 또한 다양한 승강기 패널 환경에서의 데이터셋 확장, 실시간 음성 인식, AMR 주행 연동을 통해 실제 다층 건물 환경에서의 현장 적용성을 검증할 계획이다.

감사의 글

본 연구는 산업통상부의 재원으로 한국산업기술진흥원(KIAT)의 지원을 받아 수행된 산학 융합 지구 조성 사업의 연구 결과입니다. (과제번호: P0029954, 연구과제명: 충남당진 산학융합지구 기업밀착형 산학융합촉진지원)

[참 고 문 헌]

- [1] 허윤섭, 엄용기, “엘리베이터 하중시험 대체 검사기법 개발,” 전기학회논문지 P권, 제62권, 제3호, pp. 127 - 132, 2013
- [2] P. R. Schulze, S. Müller, B. Stephan, T. Müller, and H.-M. Gross, “On realizing autonomous transport services in multi story buildings with doors and elevators,” *Frontiers in Robotics and AI*, vol. 12, 2025, doi: 10.3389/frobt.2025.1546894.
- [3] T. V. Nguyen, J. Jeong, and J. Jo, “An efficient approach for the elevator button manipulation using the visual-based self-driving mobile manipulator,” *Industrial Robot: The International Journal of Robotics Research and Application*, vol. 50, no. 1, pp. 84 - 93, 2023, doi: 10.1108/IR-03-2022-0063.
- [4] Z. Zhao, J. Zhao, and Y. Lou, “Autonomous Operation of Elevator Buttons for Multi-floor Navigation,” *Lecture Notes in Electrical Engineering*, vol. 706, pp. 161 - 170, Springer, 2021, doi: 10.1007/978-981-15-8458-9_18.
- [5] D. Zhu, Y. Fang, Z. Min, D. Ho, and M. Q.-H. Meng, “OCR-RCNN: An Accurate and Efficient Framework for Elevator Button Recognition,” *IEEE Transactions on Industrial Electronics*, vol. 69, no. 1, pp. 582 - 591, 2022, doi: 10.1109/TIE.2021.3050357.
- [6] X. Tang, C. Wang, J. Su, and C. Taylor, “An Elevator Button Recognition Method Combining YOLOv5 and OCR,” *Computers, Materials & Continua*, vol. 75, no. 1, pp. 117 - 131, 2023, doi: 10.32604/cmc.2023.033327.
- [7] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137 - 1149, 2017, doi: 10.1109/TPAMI.2016.2577031.