

# 자율주행이 적용된 AI 기반 안내 로봇 설계 및 구현

유경민, 남승우, 박재원, 최정우, 김명섭

고려대학교

{rudals2710, nam131119, 2018270614, choigoya97, tmskim}@korea.ac.kr

## Implementing an AI-based Guide Robot with Autonomous Driving

Gyeong-Min Yu, Seung-Woo Nam, Jae-Won Park, Jeong-Woo Choi, Myung-sup Kim\*

Korea University

### 요약

안내 로봇은 현대 사회에서 공공 장소, 대중 교통 시설, 상업 시설 등에서 사용자들에게 필요한 정보를 제공하거나 경로 안내, 서비스 안내 등의 기능을 수행하는 중요한 역할을 담당하고 있다. 특히, AI 기술의 발전과 더불어 음성 인식 및 이해, 표정 및 행동 묘사, 컴퓨터 비전 기술 등이 적용된 최첨단 로봇들은 다양한 산업 분야에서 활용되고 있으며 활발히 연구 및 개발이 이루어지고 있다. 본 논문에서는 로봇 기술이 직면한 여러 도전과제를 고려하여 컴퓨터 비전 및 자연어 처리 기술이 적용된 AI 기반 안내 로봇을 구현하고 설명한다.

### I. 서론

안내 로봇은 현대 사회에서 공공 장소, 대중 교통 시설, 상업 시설 등에서 사용자들을 도와주는 중요한 역할을 담당하고 있으며 사용자들에게 필요한 정보를 제공하거나 경로 안내, 서비스 안내, 상품 안내 다양한 기능을 포함한다. 이러한 로봇은 음성, 시각, 로봇 기술의 통합을 통해 사용자와 소통하고 도움을 제공하고 있다. 안내 로봇 기술은 계속 발전해왔으며 초기 단순한 음성 안내 기능을 제공하는 로봇들부터 최근 AI 기술의 발전과 더불어 인간과 유사한 표정과 행동을 가진 로봇, 음성 인식 및 이해 기술, 컴퓨터 비전 기술 등이 적용된 최첨단 로봇에 관한 연구 및 개발이 수행되고 있다. 그런데도 현재 안내 로봇 기술이 직면한 여러 도전과제가 있으며 이는 다음과 같다:

- 환경 인식과 경로 계획**: 주어진 환경 내 정확한 위치 파악 및 주변 환경 업데이트 인식 필요
- 자연어 이해와 대화 기능**: 사용자와의 자연어 대화를 통해 적절한 응답을 제공해야 함
- 상호작용과 사용자 경험**: 로봇은 친근하고 사용자가 편안하게 상호 작용할 수 있는 디자인과 인터페이스를 통해 효과적인 안내와 도움을 제공해야 함
- 안전과 신뢰성**: 사람과의 거리 유지 및 충돌 회피 등의 안전 기능을 갖추고 신뢰성 있는 작동과 실시간 변화하는 상황에 대처하는 능력이 필요

본 논문에서는 이러한 도전과제들을 고려한 AI 기반 안내 로봇의 구현에 대하여 설명한다. 특히, 구현한 로봇은 YOLO v7을 이용한 컴퓨터 비전 기술과 SegWav2Lip 적용을 통해 시청각 정보를 적극 활용하여 효과적인 안내 서비스를 제공하며 사용자 경험을 향상시킬 수 있다.

### II. 관련 연구

ROS 2(Robot Operating System 2)는 분산 시스템을 위한 개방형 로봇 소프트웨어 플랫폼으로 신뢰성, 보안성, 확장성, 실시간성 등의 요구사항을 충족시키기 위해 개발되었으며, 다양한 로봇 애플리케이션 개발을 지원한다 [1]. Wav2Lip은 입술 동기화를 통해 음성과 시각의 상호작용을 강화하는 기술로 음성 신호와 입술 움직임을 연결시켜 실제 사람의 입모양을 생성한다. 안내 로봇에서 Wav2Lip을 적용하면 음성 안내와 함께 시각적인 안내를 제공할 수 있으며 사용자는 안내 로봇의 입술 모양을 통해 음성 안내를 시각적으로 확인할 수 있어 의사소통과 이해에 있어 더 큰 편의성과 효과를 얻을 수 있다 [2].

### III. AI 기반 안내 로봇

본 연구에서 설계 및 구현한 AI 기반 안내 로봇의 대표 기능은 크게 사용자 조우 시 상호작용 과정과 사용자 안내 과정으로 구성되며 간략한 과정은 그림 1과 같다.

사용자 조우 시 상호작용 과정은 YOLO v7을 사용한 사람 탐지 대기, 사용자 위치 중심 회전, 시청각 안내 및 입력 대기로 구성된다. 사람 탐지 대기 과정에서는 로봇의 카메라로 들어온 이미지를 YOLO v7에 입력하여 사람을 탐지하고 사람 경계에 해당하는 바운딩 박스를 생성한다. 사용자 위치 중심 회전 과정에서는 생성된 바운딩 박스의 중심점의 x값을 바탕으로 로봇을 회전한다. x값은 0과 1 사이의 실숫값으로 맨 왼쪽 값이 0이며 본 논문에서는 x값이 0.4부터 0.6 사이에 있을 때 로봇이 사용자를 정면으로 바라보고 있다고 판단하고 이 범위를 벗어나면 로봇을 회전시켰다. 이후, 로봇이 사용자를 정면으로 바라보고 있다고 판단될 때 사용자에게 시청각으로 안내하며 미리 설정한 안내 문구를 Python3 텍스트 음성 변환 라이브러리인 pyttsx3 [3]을 사용하여 음성 파일(.wav)로 변환하고 재생한다. 시각 안내의 경우 변환된 텍스트를 Wav2Lip 모델에 입력하여 텍스트에 따라 입술이 자연스럽게 움직이는 영상을 생성하고 이를 재생한다. 재생하는 영상 예시는 그림 3과 같다.

본 과제(결과물)는 2020년도 산업통상자원부 및 한국산업기술평가관리원(KEIT) 연구비 지원에 의한 연구(No. 20008902, IT비용 최소화를 위한 5세대 담지기술 기반 SaaS SW Management Platform(SMP) 개발)이며 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과 (2021RIS-004)입니다.

사용자 안내 과정은 사용자 입력 분석 및 명령 생성, 로봇 명령 전달 및 음성 안내, 입력 위치까지 사용자 안내, 이동 완료 시 음성 안내 과정으로 구성되어 있다. 사용자 입력은 사용자가 웹 UI를 통해 직접 조작하는 것과 음성 입력으로 구분할 수 있으며 이를 통해 일반인뿐만 아니라 시청각 장애인들에게도 편의성을 제공할 수 있다. 그림 2는 건물 호수를 선택할 수 있는 웹 UI를 나타내며 직접 조작 입력의 경우, 사용자가 안내를 원하는 건물 호수를 직접 클릭하면 해당 건물 호수가 네비게이션 액션 클라이언트로 전달된다. 액션 클라이언트는 전달받은 강의실의 호수를 바탕으로 *navigate\_to\_pose* 형식의 메시지를 생성하여 네비게이션 액션 서버로 전달한다. 이때, *navigate\_to\_pose* 메시지는 강의실의 좌표(x, y)를 포함한다. 네비게이션 액션 서버는 이동을 담당하는 노드에 해당 위치로 이동하는 명령을 내리고 로봇은 SLAM (Simultaneous Localization And Mapping)을 기반으로 만들어진 지도 상에서 사용자가 입력한 위치를 확인 후 이동한다. 로봇은 현재 위치와 강의실 도착 여부에 대한 피드백을 주고받으며 로봇이 현재 목적지에 도착했는지 판단하고 로봇이 해당 위치로 이동을 완료하면 시청각 안내 후 사람 탐지 대기 상태로 돌아간다.



그림 2. 건물 호수 직접 선택 웹 UI

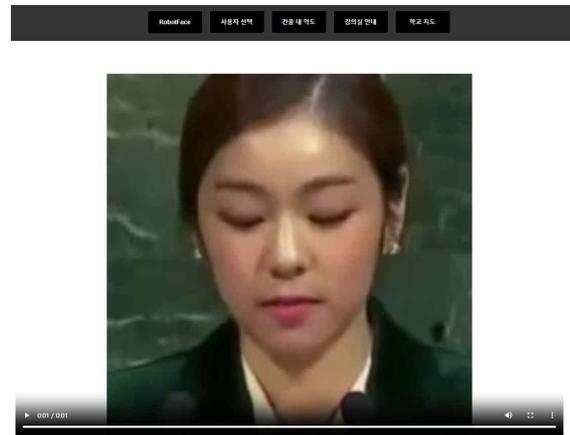


그림 3. 시각 안내 웹 UI

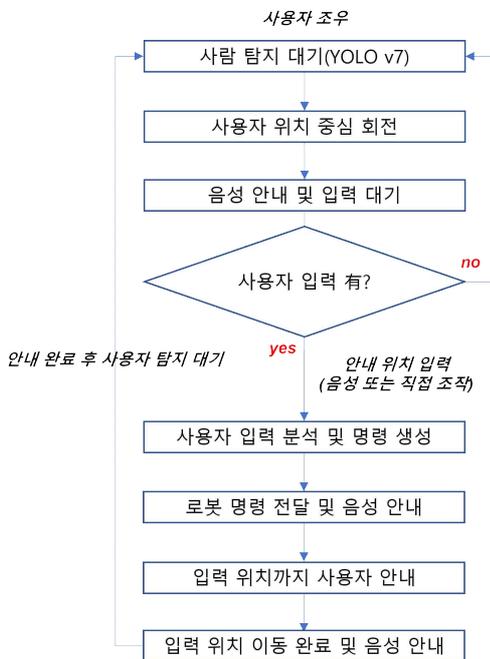


그림 1. AI 기반 안내 로봇의 안내 과정

도전 과제	설명
환경 인식 및 경로 계획	ROS2 기반의 자율 주행 라이브러리 적용을 통해 목표 위치가 주어졌을 때 자동으로 주행할 수 있고 장애물이 있을 경우 이를 자동으로 회피
자연어 이해와 대화 기능	오픈 프레임워크 STT 라이브러리 적용을 통해 자연어 변환
상호작용과 사용자 경험	로봇의 사용자 중앙 위치 정렬 기능구현과 시청각 안내를 통한 사용자 경험 증진
안전과 신뢰성	ROS2 기반 자율주행 라이브러리에 포함된 장애물 회피 알고리즘 활성화와 사람인식을 통한 안전 및 신뢰성 확보

표 1. 도전과제 고려에 대한 정성적 결과

서론에서 제시한 도전과제 고려에 대한 정성적 결과는 표1과 같으며 설계 및 구현한 로봇 안내 시스템은 도전과제를 모두 고려하였다.

#### IV. 결론 및 향후 연구

본 논문에서는 로봇 기술이 직면한 여러 도전과제를 고려하여 컴퓨터 비전 및 자연어 처리 기술이 적용된 AI 기반 안내 로봇을 구현하고 설명하였다. AI 기반 로봇의 도전과제를 정리하였고 구현한 AI 기반 안내 로봇은 이러한 도전과제를 모두 고려하여 구현하였다. 하지만, 현재까지 구현된 시스템 내 여러 한계점이 존재한다. 환경 인식 및 경로 계획에서 ROS2 기반의 로봇의 Localization 오차 누적 문제, 자연어 이해와 대화 기능에서는 음성을 로봇 명령으로 변환 시 오류 문제, 상호 작용과 사용자 경험을 위해 적용한 SegWav2Lip에서의 컴퓨팅 파워 부족 문제 등이 있으며 향후 연구로 이러한 한계점들을 해결할 예정이다.

#### 참고 문헌

[1] Open robotics (Ed.). (2023, May 17). ROS 2 Documentation: Foxy. ROS 2 Documentation. <https://docs.ros.org/en/foxy/index.html>

[2] PRAJWAL, K. R., et al. A lip sync expert is all you need for speech to lip generation in the wild. In: Proceedings of the 28th ACM International Conference on Multimedia. 2020. p. 484-492.

[3] Nateshmbhat. (2020, January 14). PyPI. Pyttsx3 2.90. PyPI. <https://pypi.org/project/pyttsx3/>