Design and Implementation of AI-based Indoor Autonomous Guide Robot

Gyeong-Min Yu*† · Jeong Woo Choi** · Jae-Won Park* · Seung-Woo Nam* · Ui-Jun Baek** · Myung Sup Kim**
*Department of Computer Convergence software, Korea University
**Department of Computer and information Science, Korea University

AI 기반 실내 자율 안내 로봇 설계 및 구현

유경민*† · 최정우** · 박재원* · 남승우* · 백의준** · 김명섭**
*고려대학교 컴퓨터융합소프트웨어 · **컴퓨터정보학과

Abstract

In recent times, guide robots have taken on an important role in assisting users in public spaces, transportation facilities, commercial establishments, and more. Guide robot technologies have been continuously evolving, starting from simple voice guidance robots to advanced robots that exhibit human-like facial expressions and behaviors, incorporating technologies such as speech recognition, understanding, computer vision, and more. The paper integrates WAV2LIP and YOLOv7 while considering the current challenges faced by guide robots to design and implement a guide robot. This research presents a novel approach to the design and implementation of guidance and autonomous driving systems for individuals with hearing impairment, while exploring the potential for practical applications. By harnessing WAV2LIP and YOLOv7, the paper suggests the development potential of a support system for individuals with hearing impairment, enhancing their safety and convenience.

Keywords : Guide Robot, Wav2Lip, YOLOv7

1. 서 론

인간의 모방과 지능의 발전은 현대 기술의 중심 주제 중 하나이며, 이러한 발전은 우리 일상생활을 혁신적으로 변화시켰다. 특히 자율주행 기술과 인공지능(AI)은 혁신적인 기술 분야로 떠오르고 있으며, 다양한 분야에서 혜택을 제공하고 있다. 이러한 기술들이 결합되어 시청각장애인들에게 도움을 주는 AI 기반 안내 로봇의 설계와 구현에 관한 연구가 중요한 주제가 되

† To whom correspondence should be addressed.
Corresponding Author: tmskim@korea.ac.kr

Received 26 June 2023; Revised 30 June 2023; Accepted 30 June 2023

었다.

시청각장애는 많은 사람들이 직면하는 신체적 제약 중 하나이다. 이들은 소리를 듣거나 의사소통을 원활하게 할 수 있는 능력에 제한을 받기 때문에 일상생활에서 다양한 어려움을 겪는다. 특히, 건물 내부에서 길 안내에 대해 결여된 부분이 많아 어려움을 겪은 적이 많다.우리는 안내로봇 서비스에 WAV2LIP 기술을 사용함으로써 로봇과의 유대감을 증진시켜 사람에게 직접 안내받는 느낌을 준다. 또한 YOLOv7을 사용하여 화면을 사용자 중앙에 놓이게 하여 사용자의 어려움을 극복했다. 우리는 WAV2LIP과 YOLOv7 기술을 활용하여 시청각장애인들에게 혁신적인 안내 시스템을 제공하고자 한다.

우리의 연구에서는 WAV2LIP을 통해 청각장애인들이 화자의 입 모양을 시각적으로 파악할 수 있도록 도와준다. 이를 통해 화자의 발화 내용과 입 모양을 동기화하여 청각장애인들에게 명확한 정보 전달을 가능케 한다.

또한, YOLOv7을 사용하여 로봇이 사용자를 실시간으로 탐지한다.

이러한 기술들을 결합한 AI 기반 안내 로봇은 시청각장애인들에게 보다 독립적이고 자율적인 환경을 제공할 수 있다. 이를 통해 일상적인 활동에서 발생하는 어려움과 불편함을 최소화하고, 사회적인 참여와 자립성을 촉진할 수 있다.

본 논문에서는 현재 안내 로봇 기술이 직면하고 있는 여러 도전과제들을 고려하여 AI 기반 안내 로봇을 구현했다. 또한 이를 통해 시청각 장애인의 어려움을 해결하고자 했다. 환경 인식과 경로 계획, 자연어 이해와 대화 기능, 상호 작용과 사용자 경험, 그리고 안전과 신뢰성이 이러한 도전과제들 중 일부이다.

첫째로, 환경 인식과 경로 계획은 AI 기반 안내 로봇이 주어진 환경 내에서 정확한 위치를 파악하고 주변 환경의 변화를 인식할 수 있어야 한다. 이를 위해, 센서와 SLAM, NAVIGATION을 활용하여 로봇이 주변의 장애물 등을 인식하고 이를 기반으로 최적의 경로를 계획해야 한다.

둘째로, 자연어 이해와 대화 기능은 사용자와 로봇 사이의 자연스러운 의사소통을 위해 중요하다. 사용자는 로봇에게 자유롭게 질문하고 지시할 수 있어야 하며, 로봇은 사용자의 의도를 정확히 이해하고 적절한 응답을 제공해야 한다. 이를 위해 자연어 처리 기술과 대화 시스템을 통합하여 사용자와의 원활한 상호 작용을 가능케 하는 기능을 개발해야 한다.

셋째로, 상호 작용과 사용자 경험은 터치스크린, 음성 인식, 제스처 인식 등 다양한 상호 작용 방식을 활용하여 사용자의 요구에 맞는 편리한 경험을 제공해야 한다.

마지막으로, 안전과 신뢰성은 AI 기반 안내 로봇이 실시간 상황에 대처하고 사용자와 주변 사람들의 안전을 보장해야 하는 중요한 요소이다. 로봇은 사람과의 거리를 유지하고 충돌을 회피할 수 있는 기능을 갖추어야 한다.

이러한 도전과제들을 고려하여 우리는 AI 기반 안내 로봇의 구현에 대한 연구를 진행하고 있다. 이를 통해 시청각장애인들에게 보다 나은 생활환경을 제공하고, 일상생활에서의 어려움을 극복할 수 있는 도움을 주기 이위한 연구를 추진하고 있다.

본 논문에서는 이러한 도전과제들을 고려한 AI 기반 안내 로봇의 구현에 대한 설명과 결과를 제시하고, 그 성능과 잠재적인 응용 분야를 탐구하고자 한다. 이를 통해 시청각장애인들의 삶의 질을 향상시키고, 더 포괄적이고 포용적인 사회를 구축하는 데 일조하고자 한다.

2. 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 입술의 움직임을 음성 신호와 동기화하

여 자연스러운 입 모양을 생성하는 딥러닝 기반의 모델이다. 이 모델은 음성 신호를 텍스트로 변환하고, 변환된 텍스트를 WAV2LIP 모델에 입력한다. WAV2LIP 모델은 입력된 텍스트와 입술 움직임의 동기화를 수행하여 시간에 따라 변하는 입 모양을 생성한다. 이렇게 생성된 입 모양은 기존의 영상에 합성되어 음성에 맞춰 입술이 움직이는 영상을 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].

YOLO(You Only Look Once) 는 실시간 객체 탐지를 위한 컴퓨터 비전 알고리즘이며 타 객체 탐지 알고리즘과 달리, YOLO는 이미지를 한 번만 보고 객체의 위치와 클래스를 동시에 예측하는 특징을 가지며 본 논문에서는 YOLO를 개선한 v7 버전을 사용했다 [3].

본 연구의 선행 연구로는 논문 [4], [5]이 있다. 논문 [4]에서 실내 침입자 자율 추격 시스템을 제안했다. 해당 논문에서는 기존 보안 시스템의 문제점을 파악하고 실내 침입자 자율 추격 시스템을 제안하면서 인력의 안전과 업무의 효율성을 증대하는 시스템을 제안했다.

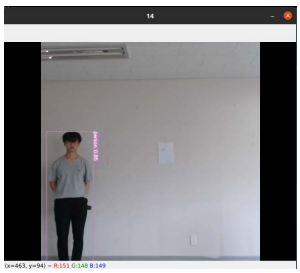
논문 [5]에서는 능동적으로 거수자와 로봇과의 거리를 파악하고 일정 거리를 유지하며 추적하는 딥러닝 기반의 사람 인식 및 위치 중앙 정렬 방법을 제안했다.

3. 본 론

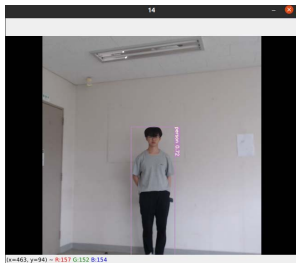
본 연구에서 설계 및 구현한 AI 기반 안내 로봇의 대표 기능은 크게 사용자 조우 시 상호 작용 과정과 사용자 안내 과정으로 구성된다.

3.1 사용자 조우 시 상호 작용 과정

사용자 조우 시 상호 작용 과정은 YOLO v7을 사용한 사람 탐지 대기, 사용자 위치 중심 회전, 시청각 안내 및 입력 대기로 구성된다. 사람 탐지 과정에서는 로봇의 카메라로 촬영한 이미지를 YOLO v7 모델에 입력하여 사람을 탐지한다. YOLO v7 모델은 감지한 사람의 경계에 해당하는 바운딩 박스(bounding box)를 생성한다. 바운딩 박스의 정보에는 왼쪽 위 모서리의 좌표(x_min, y_min)과 오른쪽 아래 모서리의 좌표(x_max, y_max), 신뢰도, 클래스 레이블로 구성이 되어있다. 바운딩 박스의 중심점을 계산하기 위해서 YOLO v7의 xyxy2xywh(x) 함수를 활용한다. xyxy2xywh(x) 함수 같은 경우, 왼쪽 위 모서리의 좌표(x_min, y_min) 과 오른쪽 아래 모서리의 좌표(x_max, y_max)를 활용해 바운딩 박스의 중심점의 x, y 좌표, 너비, 높이를 계산해주는 함수이다. 이 바운딩 박스는 감지한 사람의 공간적 경계를 표현한다. 사용자 위치 및 방향 조정 단계에서 로봇은 생성된 바운딩 박스 정보를 활용하여 사용자의 위치와 방향을 결정한다. 구체적으로는 바운딩 박스의 중심의 x 좌표를 로봇의 회전을 위한 기준으로 사용할 것이다. x 좌표는 0부터 1 사이의 실숫값으로 표현되며 0은 가장 왼쪽 위치를 나타낸다. 본 논문에서는 x 좌표 값이 0.4부터 0.6 범위에 있을 때 로봇이 사용자를 정면으로 바라보고 있다고 판단한다.



[Figure 1] x-coordinate of the detection box out of range

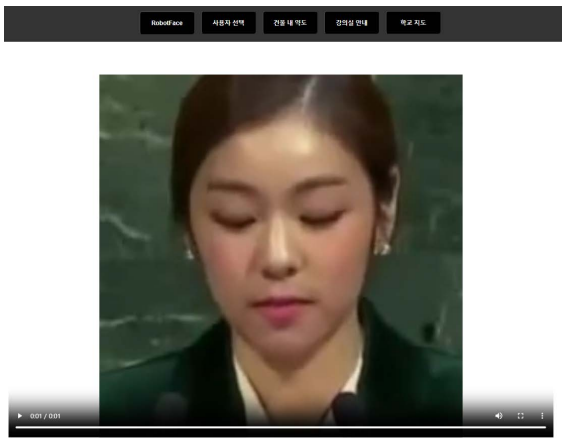


[Figure 2] x-coordinate of the detection box in range

로봇은 [Figure 1]과 같이 사용자와의 정면 위치를 유지하기 위해 이전에 언급한 x 좌표의 범위인 0.4에서 0.6 범위를 벗어날 경우, [Figure 2]와 같이 회전을 통해 조정한다. 또한, x 좌표 값을 모니터링하고 회전을

조정함으로써 로봇은 사용자의 중앙 시점에서 크게 벗어나지 않는 한 사용자를 정면으로 향하게 한다. 이후, 로봇이 사용자를 정면으로 바라보고 있다고 판단될 때 사용자에게 시청각으로 안내하며 미리 설정한 안내 문구를 Python3 텍스트 음성 변환 라이브러리인 pyttsx3 [6]을 사용하여 음성 파일(.wav)로 변환하고 재생한다.

시각 안내를 위해 WAV2LIP 모델을 활용한다. 우리 연구에서는 WAV2LIP 모델을 이용하여 텍스트에 따라 입술이 자연스럽게 움직이는 영상을 생성한다. [Figure 3]은 WAV2LIP을 통해 생성된 입 모양을 보여주는 예시이다. 이 예시에서는 텍스트에 따라 입술이 자연스럽게 움직이는 모습을 확인할 수 있다. 예를 들어, "안녕하세요"라는 텍스트에 대응하는 입 모양은 인사를 나타내는 입술의 움직임을 보여준다. 이러한 입 모양은 음성 메시지만으로는 전달하기 어려웠던 감정, 강세, 인텐션 등을 시각적으로 이해할 수 있는 기회를 제공한다.



[Figure 3] Visual Guidance Web UI

3.2 사용자 안내 과정

[Figure 5]는 AI 기반 안내 로봇의 안내 과정을 간략히 정리한 그래프이다. 사용자 안내 과정은 다음과 같이 구성된다.

1. 사용자 입력 분석 및 명령 생성:
   - 사용자 입력은 웹 UI를 통한 직접 조작 또는 음성 입력으로 구분된다.
   - 웹 UI를 사용하는 경우, 사용자는 건물 호수를 선택하기 위해 해당 건물 호수를 직접 클릭한다.
   - 사용자의 클릭 입력은 네비게이션 액션 클라이언트로 전달된다.
   - 음성입력을 사용하는 경우, WAV2LIP 모델을 통해 음성이 텍스트로 변환되고 텍스트에서 건물호수를 추출한다.
   - 텍스트 입력을 사용하는 경우, 텍스트가 텍스트에서 건물호수를 추출한다.
   - 추출된 호수가 네비게이션 액션 클라이언트로 전달된다.
   - 액션 클라이언트는 전달받은 건물 호수를 기반으로 navigate_to_pose 형식의 메시지를 생성하여 네비게이션 액션 서버로 전달한다.
   - navigate_to_pose 메시지는 강의실의 좌표(x, y)를 포함하여 목적지 위치를 명시한다.

[Figure 4] Building Address Selection Web UI

2. 로봇 명령 전달 및 음성 안내:
   - 네비게이션 액션 서버는 받은 목적지 위치로 이동을 담당하는 노드에 이동 명령을 전달한다.
   - 로봇은 SLAM (Simultaneous Localization And Mapping)과 Navigation을 기반으로 만들어진 지도 상에서 사용자가 입력한 위치를 확인하고 이동한다.
   - 로봇은 현재 위치와 강의실 도착 여부에 대한 피

드백을 주고받으며 이동 상태를 확인한다.

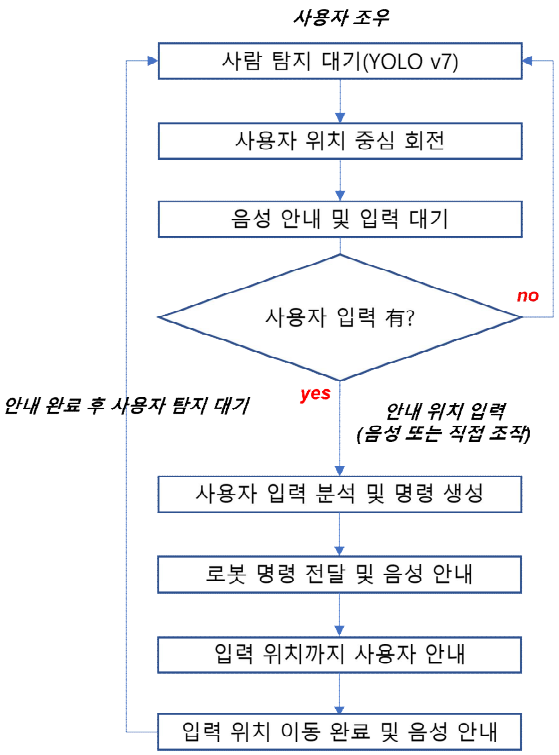
- 이동 중에는 음성 안내를 제공하여 사용자에게 이동 상태를 알려준다.

3. 입력 위치까지 사용자 안내:

- 로봇은 목적지에 접근하면 사용자가 입력한 위치에 도착하기 전까지 안내를 계속한다.
- 로봇은 위치 추정 기술을 사용하여 정확한 위치에 도달하기 위해 노력한다.
- 사용자는 로봇의 안내를 따라 입력 위치까지 안전하게 도달할 수 있다.

4. 이동 완료 시 음성 안내:

- 로봇이 사용자가 입력한 위치로 이동을 완료하면, 이를 사용자에게 알리기 위해 음성 안내를 제공한다.



[Figure 5] Guidance Process of AI-Based Guiding Robot

위와 같은 과정을 거치면 사용자는 직접 입력한 목적지로 안전하고 정확하게 이동할 수 있다. 이때, 로봇은 실시간으로 사용자와의 상호 작용을 유지하며, 안내 및 위치 파악 등의 역할을 수행한다. 이를 통해 일반인뿐만 아니라 시각 및 청각장애를 가진 사용자들에게도 편의성을 제공할 수 있다. 시각 및 청각 장애인들은 웹 UI를 통해 건물 호수를 선택하거나 음성 입력, 텍스트 입력을 통해 목적지를 지정할 수 있다. 웹 UI를 사용하는 경우, 화면에서 건물 호수를 선택할 수 있는 시각적 피드백을 제공한다. 이는 시각 장애인들이 목적지를 선택하는 데 도움을 줄 수 있다. 또한, 음성 입력을 통해 목적지를 지정하는 경우, 음성 안내를 통해 목적지 설정 과정을 시각 및 청각 장애인들에게 알려준다. 로봇은 입력된 목적지로 이동하기 위해 SLAM 기술 및 navigation을 활용하여 위치 파악 및 지도 상에서 경로 계획을 수행한다. 이를 통해 로봇은 사용자가 입력한 위치로 정확하게 이동할 수 있다. 로봇은 이동 과정에서 현재 위치와 목적지까지의 거리, 도착 여부 등에 대한 피드백을 사용자에게 제공한다. 이를 통해 사용자는 로봇의 이동 상태를 실시간으로 알 수 있고, 필요한 경우에는 로봇에게 추가적인 지시를 전달할 수 있다.

이동 완료 시, 로봇은 사용자에게 도착을 알리기 위해 음성 안내를 제공한다. 이 음성 안내는 도착 완료를 알리고 사용자의 관심과 주의를 요구할 수 있다. 이로써 사용자는 목적지에 정확하게 도착하고 필요한 작업을 수행할 수 있게 된다.

3.3 도전과제에 대한 정성적 결과

서론에서 제시한 도전과제에 대한 정성적 결과는 아래 <Table 1>과 같다.

<Table 1> Qualitative Results of Challenge Task

도전과제	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
상호 작용과 사용자 경험	로봇의 사용자 중앙 위치 정렬 기능 구현과 시청각 안내를 통한 사용자 경험 증진
안전과 신뢰성	ROS2 기반 자율주행 라이브러리에 포함된 장애물 회피 알고리즘 활성화와 사람인식을 통한 안전 및 신뢰성 확보

slam 및 ros2 패키지인 nav2을 적용하여 로봇이 주변 환경을 실시간으로 인식하고, 장애물이 있을 시 정지 또는 우회함을 통해 로봇은 안전한 경로를 계획하고 사용자를 정확하게 안내할 수 있다. 또한, WAV2LIP을 활용하여 로봇이 사용자의 요구를 정확하게 인식하고, 이를 음성으로 변환하여 처리할 수 있다. 이로써 로봇은 자연어에 대한 이해도가 향상되고, 사용자와 자연스러운 대화를 수행할 수 있게 된다. WAV2LIP을 활용하여 로봇이 입 모양을 생성함으로써, 사용자에게 상호 작용과 유대감을 느끼게 한다. 로봇의 입 모양을 통해 표현되는 표정과 감정은 사용자와의 상호 작용을 강화하고, 사용자 경험을 향상시킨다. 더불어 로봇의 사용자 중앙 위치 정렬 기능을 통해 사용자와 로봇 간의 상호 작용이 원활해진다. 사용자는 로봇이 자동으로 정면에 위치함으로써 로봇과의 상호 작용이 더 편리해지면서 사용자의 서비스 사용에 매우 효과적이다. 이는 사용자와 로봇 간 의사소통의 효율성을 높여 준다. 실시간으로 장애물을 감지하고, 이를 고려하여 안전한 경로를 계획한다. 자율주행 알고리즘을 개선하여 로봇은 정확한 위치 추정과 정밀한 이동을 수행하며, 사용자의 안전을 보장한다.

4. 결 론

결론적으로, 본 논문에서는 AI 기반 안내 로봇의 구현과 도전과제에 대해 다루었다. 하지만, 현재까지 구현된 시스템 내 여러 한계점이 존재한다. 특히, 자연어 이해와 대화 기능에서는 로봇 내의 소음으로 인해 음성 인식이 원활하지 못했다. 이러한 한계점을 해결하기 위해 음성 인식의 정확성을 향상시키는 기술 개발에 집중할 것이다.

또한, 사용자 경험을 개선하기 위해 다양한 인터페이스와 상호 작용 방법을 고려할 예정이다. 예를 들어, 터치스크린이나 제스처 인식 등을 활용하여 사용자가 더 편리하게 로봇과 소통할 수 있는 방안을 연구할 것이다.

뿐만 아니라, 안전과 신뢰성 측면에서도 로봇의 자율주행 능력을 강화하고, 환경 인식과 장애물 회피 기능을 개선할 예정이다. 이를 통해 로봇과 주변 환경 사이의 상호 작용을 더욱 원활하게 만들고, 사용자의 안전을 보장할 수 있다.

이러한 시스템은 건물 내부의 길 찾기와 같은 복잡한 작업을 로봇에게 맡김으로써 인간의 작업 부담을 줄여주고 효율성을 향상시킨다. 로봇은 사람보다 정확하고 일관된 경로를 제공할 수 있으며, 필요한 경우 실시간으로 경로를 조정하여 혼잡한 지역이나 장애물을 피할 수 있다.

총체적으로, 사용자 중앙 위치 정렬 및 음성 안내 시스템은 건물 내부에서의 이동성과 접근성을 개선하여 다양한 사용자들이 건물을 더 편리하게 이용할 수 있도록 도와준다. 이러한 기술의 발전은 장애를 가진 사람들의 일상생활 질을 향상시키고 사회적 포용을 증진시키는 데 기여할 수 있다.

향후 연구를 통해 이러한 도전과제를 극복하고, 사용자의 편의성과 안전을 고려한 최적화된 AI 기반 안내 로봇 시스템을 구현할 계획이다. 이를 통해 현실 세계에서의 실용적인 응용 가능성을 확장하고, 사용자들에게 더 큰 가치를 제공할 수 있을 것이다.

Acknowledgement

This work was supported by the Technology Innovation Program grant funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea) and the Korea Evaluation Institute of Industrial Technology (KEIT)
(No. 20008902, Development of SaaS SW Management Platform based on 5 Channel Discovery technology for IT Cost Saving) and supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-004).

References

[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] WANG, Chien-Yao; BOCHKOVSKIY, Alexey; LIAO, Hong-Yuan Mark. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022.

[4] Jeong-Woo Choi, Seung-Woo, Nam, Jae-Won Park, Gyeong-Min Yu, Ui-Jun Baek, Myung-Sup Kim, "Indoor intruder Autonomous chasing system design", KICS, pp. 1-2, 2023.

[5] Jae-Won Park, Gyeong-Min Yu, Seung-Woo, Nam, Ui-Jun Baek, Myung-Sup Kim, "A Deep Learning-based Human Detection and Position Centering Method." ,KICS, pp. 1-2, 2023.

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