시각장애인을 위한 실시간 엘리베이터 탐지 및 음성 안내 시스템 개발

이명우, 이성민, 최서현, 장인성, 박호민, 유대승†

한국전자통신연구원

dlauddn05@gmail.com, sungmin4123@gmail.com, vvvfkq1500@gmail.com, wkddlstjd2@gmail.com, hominpark@etri.re.kr, ooseyds@etri.re.kr†

Development of a Real-time Elevator Detection and Voice Guidance System for the Visually Impaired

Myeongwoo Lee, Seongmin Yi, Seohyeon Choi, Insung Jang, Ho-Min Park, Dae Seung Yoo†

Electronics and Telecommunications Research Institute

요 약

시각장애인에게 엘리베이터는 끼임 사고나 낙상 사고가 발생할 수 있는 에스컬레이터나 계단 대비 중요한 이동 수단 중 하나이다. 하지만 시각장애인들에겐 누군가의 도움 혹은 음성 안내 없이 엘리베이터의 위치를 찾고 엘리베이터의 층수를 판단해서 스스로 내리는 것이 매우 힘들다. 본 논문에서는 객체 탐지 모델인 YOLOv5를 사용하여 엘리베이터를 탐지한 후 MobileNetV2와 TTS를 사용하여 층수 판단과 음성 안내를 지원하는 시스템을 제안한다. 시스템 하드웨어 구성은 Jetson AGX Orin과 Depth Camera D435로 이루어지며 실시간 탐지 및 안내, 충돌 방지를 위한 알고리즘을 탑재하여 시연하였다. 이를 통해 제안된 시스템은 시각장애인이 엘리베이터를 안전하고 독립적으로 이용할 수 있도록 돕는 방안을 제시한다.

Ⅰ. 서 론

시각장애인에게 엘리베이터는 끼임 사고나 낙상 사고가 발생할 수 있는 에스컬레이터나 계단 대비 중요한 이동 수단 중 하나이다. 그러나 엘리베이터의 위치를 알려주는 시스템은 대부분 시각 정보가 필요하며 엘리베이터를 찾더라도 음성 안내가 나오지 않는다면 층수를 알 수 없다[1]. 그렇기 때문에 보조 동행인과 함께 움직이거나 어플리케이션을 사용해 엘리베이터를 찾아 이동해야 한다. 하지만 보조 동행인은 항상 옆에 있을 수 없으며 어플리케이션을 사용하는 것 또한 한계가 있다. 설리번+(SULLIVAN+) 혹은 비 마이 아이즈(Be My Eyes)와 같은 앱은 사진을 찍으면 사진에 담긴 모습이나 상황을 묘사해서 설명해 준다[2]. 하지만 이러한 기존 방식들은 실시간으로 사용하기에는 한계가 있다. 상기 두 앱 모두 서버-클라이언트 구조로 되어있어 별도의 서버가 필요하고, 이미지 처리 시간도 필요하므로 인터넷 환경과 속도에 큰 영향을 받는다.

본 논문은 시각장애인의 엘리베이터 접근성을 개선하기 위해 상용 하드웨어와 딥러닝 기술을 결합한 시스템을 제안한다. 제안하는 시스템은 장애물 탐지와 함께 YOLOv5[3]를 사용해 엘리베이터를 탐지하고, MobileNetV2[4]을 사용하여 층수를 인식한다. 또한 TTS(Text to Speech) 기술을 적용하여 엘리베이터와 층수 표시기 탐지 시 실시간으로 음성 안내가 출력되도록 한다. 제안된 시스템은 시각장애인이 엘리베이터를 안전하고 독립적으로 이용할 수 있도록 돕는 방안을 제시한다.

Ⅱ. 본론

1. 시스템 구성

본 논문에서는 [그림 1]과같이 데이터 수집, 연산, 송출을 위한 하드웨어 시스템에 YOLOv5, MobileNetV2, TTS 기술을 탑재하여 구성한다. 데이터 연산을 위한 하드웨어로는 NVIDIA Jetson AGX Orin과 데이터 수집 및 장애물 회피를 위한 하드웨어로 Intel Real Sense Depth Camera D435를 사용하였다.

1.1 YOLOv5

객체 탐지를 위한 데이터 셋은 울산과학기술진흥센터 1층과 4층, 지하 1층에서 촬영한 1,200개의 엘리베이터 내외부 이미지를 사용하였다. 열림, 중간, 닫힘, 바깥 버튼, 내부 층수 표시기까지 총 5가지의 class로 나누어 Labeling을 진행하였다. 사용한 모델은 YOLOv5s이며 Batch size=16, Image size=640, epoch=200으로 학습시켰다.

1.2 MobileNetV2

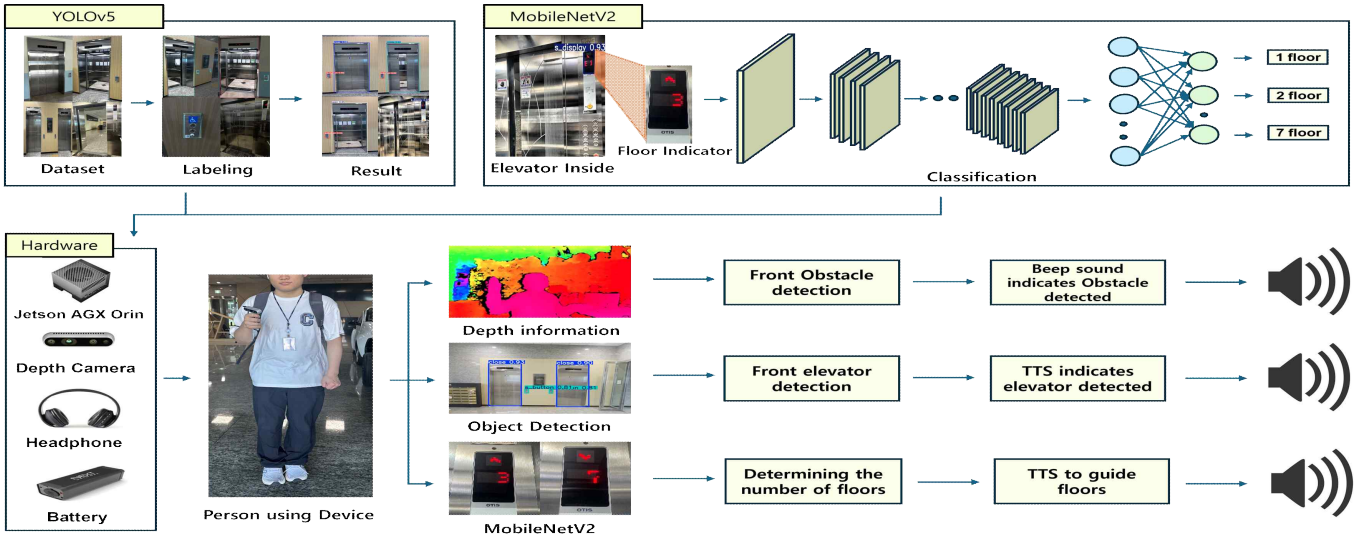
엘리베이터의 현재 층수를 인식하기 위해 먼저 YOLO로 층수 표시기를 탐지한 후 해당 영역의 이미지를 잘라 MobileNetV2을 이용하여 층수를 판단하는 2 stage 형태로 모델을 구성하였다. ImageNet으로 사전학습 된 MobileNetV2을 수집한 350개의 층수 표시기 이미지 데이터들로 미세 조정하였다. 학습 전, 전처리 과정에서는 빠른 이미지 처리를 위해 256*256으로 이미지 크기를 조정하였다.

1.3 TTS

시각장애인에게 엘리베이터 층수 등을 음성으로 안내하기 위해 TTS를 이용한 시스템을 제안한다. 비프음으로는 장애물 탐지와 충돌 방지 기능을 구현할 수 있지만, 엘리베이터 탐지와 층수 안내에는 TTS가 필수적이다. 본 시스템은 TTS와 비프음을 병행하여 사용자에게 효율적이고 편리한 엘리베이터 이용을 지원한다. KSS(Korean Single Speaker Speech) 데이터셋[5]을 활용하여 TTS 모델을 학습하였으며, 실시간 음성 안내 기능을 구현하였다.

2. 시연 시나리오 및 결과

[그림 2]는 제안하는 시스템이 동작하는 알고리즘의 순서도를 나타낸 것이다. 본 논문의 시연은 시연자가 익숙하지 않은 울산과학기술진흥센터 1층 정문에서 시작하였고 엘리베이터로 4층까지 이동하는 것으로 정하였다. 1층 정문에서 엘리베이터까지의 직선거리는 약 15M이고 1층 전체 면적은 약 20M×20M로 측정되었다. 시연은 총 2회 시행되었으며 실제 시각



[그림 1] 전체 구성도

장애인이 참가하여 Jetson AGX Orin과 연결된 Depth Camera를 손에 들고 엘리베이터 탐색을 수행하였다. 참가자는 Depth Camera를 활용하여 주변 장애물을 실시간으로 탐지하며 엘리베이터를 찾았다.

YOLOv5 모델이 '닫힘'와 '바깥 버튼' 클래스를 동시에 탐지하여 엘리베이터를 인식하면, '전방에 엘리베이터가 있습니다.'라는 음성 알림을 제공하였다. 이후 '열림' 클래스를 탐지하면 시스템은 엘리베이터가 열렸음을 인식하고 '엘리베이터가 열렸습니다.'라는 음성 안내를 제공하였다. 이 시점부터 사용자는 엘리베이터 내부에 있다고 가정하며, 탑재된 알고리즘은 사용자 상태를 '엘리베이터 내부'로 전환한다.

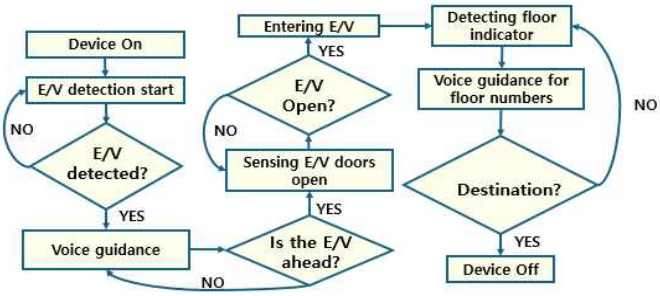
엘리베이터 내부에서 카메라가 층수 표시기를 탐지하면 시스템은 층수 표시기의 바운딩 박스를 추출하고, MobileNetV2 모델을 통해 1층부터 7층까지의 층수를 판별하여 음성으로 안내한다. 사용자는 원하는 층수에 도착하면 시스템을 종료할 수 있으며, 엘리베이터를 다시 찾고자 하는 경우 버튼을 눌러 시스템을 재시작할 수 있다.

시연 결과, 참가자는 1층에서 4층까지 성공적으로 이동하였으며, Depth Camera를 통해 장애물을 회피하며 엘리베이터를 탐지하였고 시연자가 엘리베이터 내부일 때, 층수 표시기가 탐지되면 MobileNetV2를 통해 층수를 도출했다. 엘리베이터, 층수 표시기 탐지 시 음성 안내는 실시간으로 제공했다. 제한적인 시연이었지만 인터넷 환경에 의존적인 서버를 이용하는 것 없이 Stand-alone 방식으로 시각장애인의 독립적인 엘리베이터 이용을 지원할 수 있음을 확인하였다.

Ⅲ. 결론

본 연구에서는 시각장애인의 독립적이고 안전한 이동을 지원하기 위해 딥러닝 기반의 객체 탐지 및 음성 안내 시스템을 개발하였다. 이 시스템은 YOLOv5 모델을 활용하여 엘리베이터의 문과 버튼을 실시간으로 탐지하고, MobileNetV2 모델을 이용하여 엘리베이터 층수 표시기를 인식한 후 시각장애인에게 음성으로 안내하여 엘리베이터 사용을 보조한다.

실험 결과, 제안된 시스템은 시연을 통해 엘리베이터 탐지와 층수 인식 기능, 음성 안내 기능을 동작 알고리즘에 따라 수행하였다. 시스템의 탐지 및 안내 과정은 3초 이내에 완료되었고 실시간 엘리베이터 이용 지원이 가능하도록 구성되었다. 향후 연구에서는 시연에 사용된 울산과학기술진흥센터뿐만이 아닌 범용성을 위해 다양한 건물에서도 적용할 수 있도록 추가적인 연구를 진행할 예정이다.



[그림 2] 제안하는 시스템의 시나리오 순서도

ACKNOWLEDGMENT

이 연구는 울산시-ETRI 2차 공동협력사업(24AB1610, 제조 혁신을 위한 주력산업 지능화 기술 개발 및 산업현장에서의 사람-이동체-공간 자율협업지능 기술 개발)의 연구결과로 수행되었음.

참 고 문 헌

[1] James H. Rimmer, Barth Riley, Edward Wang, Amy Rauworth, "Accessibility of Health Clubs for People with Mobility Disabilities and Visual Impairments", American Journal of Public Health 95, no. 11 (November 1, 2005): pp. 2022-2028.

[2] 강태희, 원예랑, 조희경. 2023, "시각장애인 이동권 개선을 위한GPS 기반 디지털 콘텐츠와 인터페이스 디자인", 디지털콘텐츠학회논문지, vol.24, no.6 pp.1151-1160.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection,", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.

[4] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520

[5] K. Park, Korean SingleSpeaker Speech Dataset, 2020, (https://www.kaggle.com/bryanpark/korean-single-speaker-speech-dataset)