

YOLO v5와 ESP-32 카메라를 이용한 스마트 지팡이

김정현*, 전현진*, 임승훈**, 원치훈*, 신희원*, 최상호***

Smart Cane using YOLO version 5 and ESP-32 Camera

Jung Hyun Kim*, Hyun Jin Jeon*, Seung Hun Lim**, Chi Hun Won*, Hee Won Shin*, Sangho Choe***

*가톨릭대학교 정보통신전자공학부 & S&Y Corporation

**아주대학교 기계공학과 & S&Y Corporation

***가톨릭대학교 정보통신전자공학부 명예교수 & S&Y Corporation

요약

본 논문에서는 아두이노 초음파 센서로 장애물 탐지시 진동모터에 진동을 일으켜 사용자에게 알려줄 뿐만 아니라, 카메라로 촬영한 이미지를 서버로 전송 AI 기반 장애물 인식기능을 갖는 시각장애인용 스마트 지팡이를 설계한다. 기존 대부분의 시각장애인용 지팡이는 단순히 전방에 장애물 유무만 판별하는 탐지 기능만 있을 뿐 장애물 이름 및 종류를 파악(라벨링)하는, 즉, 장애물 인식기능은 제한된다. 제안 스마트 지팡이는 장착한 ESP-32 카메라로 촬영한 장애물 이미지를 BLE(Bluetooth Low Energy) 경유 서버로 전송하면 Yolo v5 라벨링 프로세스를 거쳐 음성안내(TTS) APP에서 시각장애인에게 안심보행 서비스를 제공한다.

I. 서론

시각 기관인 눈은 시각적 정보를 처리하고 처리한 정보를 뇌에 전달하는 삶에 있어서 필수적인 장기이지만 세계보건기구(WHO)의 보고서에 따르면, 전 세계 시각장애 인구는 22억 명이며 그 중 실명과 같은 중등도 또는 중증 이상의 시력장애 인구는 최소 10억 명에 달한다.[1] 시각장애 인의 이동 보조 수단으로 과거부터 지팡이가 많이 활용되어 왔지만 지팡이는 단순히 장애물이 있다는 것을 알려줄 뿐 장애물이 어떤 종류이고 또한 움직이는 물체에 대해서는 실시간 탐지가 어렵다는 단점이 있다. 본 논문에서는 시각장애인을 위한 ESP-32 카메라 기반의 스마트 지팡이를 설계 제안한다. 제안한 스마트 지팡이는 초음파 센서를 이용해 전방의 물체(보행 장애물)를 탐지하면 햅틱 센서 모듈을 이용해 진동을 발생시켜 이용자에게 알리고 이용자는 ESP-32 카메라를 이용해 사진을 촬영해 이를 BLE 경유 서버로 전송하면, 서버에서는 YOLO v5 모델을 이용해 장애물의 종류를 구별 후 해당 결과를 이용자의 핸드폰 APP으로 피드백하여 이를 이용자에게 알리게 된다.

II. 본론

1. 초음파센서(HC-SR04)와 원리

초음파 센서는 초음파를 이용해 초음파를 쏘주면 초음파가 물체에 부딪혀 되돌아오는 시간을 이용해 물체의 거리를 측정한다. 초음파 센서의 핵심은 초음파를 쏘주는 발사(Trigger), 물체에서의 반사(Reflection), 반사된 초음파의 수신(Echo), 그리고 반사된 시간을 계산해 거리로 환산하는 것이다. 음파의 속도는 매질에 따라 다르지만 일반적으로 공기 중에서 약 340m/s이다. 이를 이용해 거리(cm) = $\frac{\text{시간}(\mu\text{s}) \times 0.034}{2}$ 의 공식으로

거리를 측정할 수 있다. 측정한 왕복 거리를 2로 나누어 편도 거리를 구할 수 있다. 일반적으로 초음파는 금속, 비금속, 사람 등 소리 반사 특성을 가지는 거의 모든 물체를 감지할 수 있기 때문에 많이 사용되며 특히 HC-SR04 모델은 아두이노에서 범용적으로 사용되는 초음파 센서이기에 본 논문에서 설계 제안하는 스마트 지팡이에도 HC-SR04 모델을 사용했다.

2. ESP-32 카메라와 원리

ESP-32 CAM은 Wi-Fi와 BLE(Bluetooth Low Energy)기능이 내장된 ESP-32칩에 OV2640 카메라 모듈을 결합한 보드로 영상 스트리밍과 IoT(Internet On Things) 프로젝트에 많이 사용되는 보드이다. ESP-32 카메라는 OV2640 카메라를 이용해 사진을 촬영해 PSRAM에 데이터를 임시 저장하고 Wi-Fi 혹은 BLE를 이용해 사진을 전송하게 된다.

OV2640 카메라는 200만 화소(2MP)까지 지원하며 촬영한 사진을 서버로 보내 YOLO에서 분석하기에 문제가 없는 화소이다.

3. 전송 방식 및 코드

ESP-32 카메라로 촬영한 사진을 BLE를 이용해 서버로 보내기 위해 `#include <base64.h>` 라이브러리를 추가하였다.

```
base64Data = base64::encode(fb->buf, fb->len);
size_t fbLen = base64Data.length();
Serial.printf("Base64 String Size: %d bytes\n", fbLen);
```

해당 코드를 통해, 이미지를 Base 64 문자열로 변환하였다. BLE를 이용해 이미지를 보내는 것은 Wi-Fi를 이용해 이미지를 보내는 것보다 시간이 오래 걸리고 이미지 처리에 필요한 RAM이 부족해 이미지 분할(Chunking) 방식을 이용해 이미지를 서버로 전송한다. Chunk 사이즈를 500bytes로 세팅해 전송 속도를 높이고 마지막에 500bytes 보다 사이즈가 작은 Chunk 이미지를 remainder 변수에 저장 후 이 값이 0보다 크면 remainder 만큼 전송하여 이미지의 손실 없이 전송 가능하게 하였다. 또한 delay를 1ms로 줄여 실시간 이미지 전송이 가능하게 했다.

```
const size_t chunkSize = 500;
for (size_t i = 0; i < fullChunks; i++) {
    String chunk = base64Data.substring(currentIdx, currentIdx + chunkSize);
    pCharacteristic->setValue((char*)chunk.c_str());
    pCharacteristic->notify();
    currentIdx += chunkSize;
    delay(1);
```

4. YOLO의 이미지 탐지 방식 및 API

YOLO는 You Only Look Once라는 뜻의 실시간 객체 검출 시스템이다. YOLO는 CNN(Convolutional Neural Networks) 딥러닝 모델을 기반으로 특징을 추출한 뒤 이를 이용해 물체의 종류와 위치를 Bounding Box로 표시해 Label로 분류한다.[2] 이를 통해 실시간 객체 탐지가 가능하다. YOLO는 2016년 첫 출시된 이후 다양한 버전이 출시되어 왔는데 본 논문에서 설계 제안하는 시스템에서는 YOLO version 5을 사용한다.

YOLO v5은 PyTorch를 기반으로 구현하여 다양한 모델 크기를 제공하고 설치와 구현이 쉽기 때문에 범용적으로 사용되는 모델이기 때문에 본 논문에서는 YOLO v5을 사용했다. 본 논문에서 사용한 YOLO v5 모델은 Github 등 시중에 나와 있는 Open source를 사용했고 획단보도, 차량, 사람 등 시각 장애인이 보행 중 마주칠 수 있는 장애물 약 31,600장을 4주에 걸쳐 추가로 학습시켜 객체 탐지 모델을 개발했다.



fig 1. YOLO v5 실행 후 탐지 결과

YOLO를 이용해 탐지한 결과는 Bounding Box 위에 Label로 표시되고 이를 다시 휴대폰 APP으로 전송해 사용자에게 알리는 것이 본 논문에서 제안하는 스마트 지팡이 설계 구조이다.

5. 제안 스마트 지팡이 시스템

제안 시스템은 초음파 센서로 물체를 탐지하여 햅틱 센서 모듈로 보행자에게 알리는 방식은 기준에 직접 지팡이에 물체가 접촉해야만 위험을 인지할 수 있었던 방식과 구별된다. 즉, 비접촉식 사전 인지를 통해 미리 물체를 인식하여 안전성과 편의성을 개선한다. 또한, ESP-32칩에 있는 OV2640 카메라를 사용하여 앞에 있는 물체의 사진을 찍어 BLE를 경유 서버에 전송할 수 있다. 제안 시스템은 단순한 장애물 거리 측정을 넘어서, 딥러닝 기반의 객체 인식을 통해 장애물의 이름 및 종류를 실시간으로 판별(라벨링)하여 사용자에게 전달할 수 있다. 특히 하드웨어의 제한적인 메모리와 통신 대역폭을 고려하여 Base64 인코딩 및 데이터 Chunking 기법을 도입함으로써, 좁은 대역의 BLE 채널 환경에서도 이미지 손실 없이 안정적인 데이터 전송 체계를 구축한다.

III. 설계 과정에서의 난제

1. 서로 다른 개발 언어(Python-Dart) 간의 JSON 데이터 연동

객체 인식 AI 모델(YOLO)은 Python 환경에서 구동되는 반면, 앱은 Dart(Flutter) 언어로 개발되어 언어와 구동 환경이 다르다. 이 차이로 인해 서버의 정보를 앱으로 실시간 전송하고 처리하는 통신 체계를 구축하는 것이 설계 상 난제이다. 이를 해결하기 위해 표준적인 API 연동 방식을 채택하고 JSON 데이터 형식을 매개체로 활용하여, 언어 장벽 없이 서버와 모바일 앱 간 실시간 송수신할 수 있는 연결 통로를 구축해야 한다. 특히 실제 서버 연결 절. 가짜 심호(더미 데이터)를 생성하여 데이터

수신 여부를 교차 검증하며 연동을 진행해야 한다.

2. YOLO 모델의 연산 부하에 따른 미세 딜레이(Delay) 발생

YOLO AI 모델 자체의 연산량이 매우 방대하여 모든 데이터를 실시간으로 완벽하게 처리하기에 물리적 부담이 크다. 이로 인해 서버의 데이터 전송 과정에서 정보 전송 지연이 발생하는 문제가 있다. 모든 부하를 서버에 집중시키지 않고, 음성 안내(TTS) 처리를 플러터(앱) 측에서 담당하도록 설계해야 한다. 서버는 객체 감지에만 집중하게 하고 후속 처리는 앱이 수행하도록 하여 서버에 집중될 수 있는 연산 부담을 분담하도록 구성해야 한다.

IV. 결론

본 연구에서는 시각장애인의 지팡이가 실시간 탐지가 되지 않아 장애물에 대한 정확한 정보나 움직이는 물체를 알지 못한다는 단점을 인식하고 이를 보완하고자 ESP-32 카메라, 초음파 센서, 그리고 YOLO v5 기반의 딥러닝 기술을 통합한 스마트 지팡이 시스템을 제안하였다.

본 연구자들은 보행자 중심의 거시적인 관점에서 3가지 고려 요인을 선정해 제안 스마트 지팡이 시스템을 최적화하였다. 첫째, 보행자의 안정성과 편의성 보장이다. 단순히 장애물의 유무를 넘어 물체의 종류를 즉각적으로 보행자에게 알림으로써, 시각장애인이 보행 중 마주하는 도보 환경을 파악하는데 도움을 주었다. 이는 보행자의 의사결정 속도를 높여 사고 예방에 직접적으로 기여한다. 둘째, 기술적 보조를 통한 심리적 안정감 제공이다. 본 시스템의 실시간성을 통해 보행자가 시야가 좋지 않은 상황에서 느끼는 불안감을 해소시킨다. 이는 보행자의 심리적 안정감을 주어 시각장애인의 사회 활동 범위를 넓히는 긍정적인 효과를 기대할 수 있다. 마지막으로, 설계 및 구현 과정에서 발생한 서로 다른 언어 간의 데이터 연동 및 서버 연산 부하 문제는 해결하였다. Python 기반의 YOLO 서버와 Dart 기반의 모바일 앱 사이의 통신 규격을 JSON 형식으로 표준화하여 데이터 무결성을 확보하였고, 연산량이 많은 TTS(Text-to-Speech) 처리를 앱 측으로 분산시켜 시스템 전체의 응답 지연 시간을 최소화할 수 있었다. 약 31,600장의 대규모 학습 데이터를 통해 횡단보도, 차량 등 보행 시 필수적인 객체에 대한 높은 탐지 정확도를 확보한 것은 실용화 가능성을 입증한 것이 주요성과이다. 결론적으로, 효율적인 실시간 장애물 탐지/인식 기능을 내장한 임베디드 시스템 구현으로 시각장애인에게 안심 보행 서비스를 제공할 수 있음을 보여주었다. 다만, 현재 시스템의 이미지 전송 속도와 야간 및 악천후 상황에서의 인식 성능은 향후 개선이 필요한 과제이다. 차후 연구에서는 통신 알고리즘의 최적화와 데이터셋 고도화를 통해 시스템의 신뢰성을 더욱 향상시키고, 사용자 편의성을 극대화한 실질적인 보행 보조 기구로 발전시켜 나갈 계획이다.

감사의 글

본 논문은 경기도 기술개발사업의 사업비지원(과제번호 D2505032)에 의해 수행되었습니다.

참고문헌

- [1] <https://www.who.int/publications/item/9789241516570>
 - [2] 김형운, 오현진, 장민호, “시각장애인 악기 교육에서의 컴퓨터 비전 활용 연구 : YOLOv8s 모델을 통한 접근성 개선”, *한국시각장애인교육&체육학회*, 2025