

엣지 디바이스 기반 시각장애인을 위한 도보 보조 시스템

김장환, 이유진, *진윤희
국립한밭대학교

kjh@edu.hanbat.ac.kr, yujin@edu.hanbat.ac.kr, *yhjeon@hanbat.ac.kr

A Walking Assistance System for the Visually Impaired on Edge Device

Janghwan Kim, Yujin Lee and *Yunho Jeon

Hanbat National University

요약

본 논문은 시각장애인의 안전한 보행을 지원하기 위한 도보 보조 시스템을 제안한다. 저사양 On-device 환경에서도 원활한 실시간 처리를 가능하게 하기 위해 YOLOv5n 객체 탐지 모델과 Efficient Speech TTS(Text-to-Speech) 모델을 사용하였다. YOLOv5n은 FP16 양자화 및 ONNX 최적화를 통해 메모리 사용량과 연산 요구사항을 줄여 라즈베리 파이와 같은 저사양 장치에서도 효율적으로 작동할 수 있다. 객체 탐지 후, 거리 추정 모듈은 Bounding Box 면적 변화율과 중심점 이동 벡터를 분석하여 객체의 접근 여부를 판단한다. 이후 TTS 모듈은 한국어로 실시간 음성 안내를 제공하여 시각장애인이 주변 위험 요소를 인지하고 안전하게 이동할 수 있도록 돕는다. 제안된 시스템은 저사양 하드웨어 환경에서도 실시간 탐지 및 안내 성능을 검증하였으며, 시각장애인의 이동 안전성과 자율성을 크게 향상시킬 것으로 기대된다.

I. 서론

시각장애인은 보행 시 자전거가 길을 막거나 예기치 못한 장애물이 나타나는 등 다양한 어려움에 직면하게 되는데 본 논문에서는 이러한 문제들을 휴대 가능한 On-device 시스템을 통해 보다 안전하고 편리한 이동 환경을 지원하고자 한다. On-device 연산 환경에서는 메모리, 연산 속도, 전력 소비 등 제한된 하드웨어 자원을 효율적으로 활용하기 위해 모델의 경량화와 최적화가 필수적이다. 이를 위해 본 논문에서는 다양한 최적화 기법을 적용하여 제한된 자원 내에서도 높은 성능을 발휘할 수 있는 객체 인식 모델 및 TTS 모델을 사용함으로써, 실시간으로 장애물을 인식하고 처리할 수 있는 시스템을 구현하였다. 궁극적으로 본 시스템은 시각장애인들의 이동 안전성을 향상하고, 일상생활에서의 자율성과 독립성을 증진하는 데 기여할 것으로 기대된다.

II. 본론

도로 보행 시 자전거, 차량 등 다양한 물체에 의해 시각장애인들의 이동이 제한될 수 있다. 휴대가 가능한 저사양 기기에 객체 탐지 모델을 탑재하여 주변의 객체를 실시간으로 감지하고 이를 시스템에 전달하는 방안을 제시한다.

탐지 모델 YOLO[1]을 활용하여 다양한 객체를 인식하는 객체 탐지 모듈이다. 두 번째는 탐지된 객체의 거리를 추정하고, 이를 바탕으로 객체의 접근 여부를 판단하는 거리 추정 모듈이다. 마지막으로, 접근하는 물체에 대한 정보를 바탕으로 음성을 생성하여 경고를 제공하는 TTS (Text-to-Speech) 모듈을 포함하여 시스템을 설계하였다.

2.1 객체 탐지 모듈

본 논문에서는 실시간 처리가 필수적인 도보 보조 시스템을 구현하기 위해 Ultralytics에서 개발한 YOLOv5[2]를 선택하였다. YOLOv5는 최신 모델인 YOLOv8에 비해 계산량이 적고 처리 속도가 빨라 저사양 환경에서 실시간 처리에 적합하다. 이와 같은 특성으로 인해 저사양 기기인 라즈베리 파이(Raspberry Pi)에서도 더욱 신속하고 효율적인 객체 탐지 및 실시간 처리가 가능하다. 또한, 모델의 실행 환경을 최적화하기 위해 ONNX 형식을 활용하였으며, 모델의 효율성을 극대화하기 위해 FP16 양자화를 수행하였다. FP32는 높은 정밀도를 제공하지만 메모리 사용량이 많아 저사양 기기에서 실시간 처리에 부담이 될 수 있다. 반면, FP16은 메모리 사용량과 연산 요구사항을 절반으로 줄이면서도 충분한 정밀도를 유지한다. 이를 통해 YOLOv5n 모델을 FP32에서 FP16으로 변환하였으며, 그 결과 라즈베리 파이에서 객체 탐지 속도가 개선되어 실시간 처리가 가능해졌다.

이 과정을 통해 최적화된 YOLOv5n 모델을 기반으로, 본 논문에서 제안한 객체 탐지 모듈은 실시간으로 주변 환경을 인식하고 위험 요소를 감지하는 초기 핵심 역할을 담당한다. 먼저, 카메라를 통해 실시간으로 주변 환경을 촬영하여 지속적으로 영상 프레임들을 캡처한다. 캡처된 프레임은 사전에 최적화된 YOLOv5n 모델을 로드한 후 입력 데이터를 일관된 형식으로 전처리한다. 이후, 전처리된 이미지는 YOLOv5n 모델을 통해 객체가 탐지되며, 사람, 차량, 자전거 등 도로와 관련된 위험 요소를 필터링한 뒤 중요 객체들에 대해서만 거리 추정 및 위험도 평가를 수행한다.

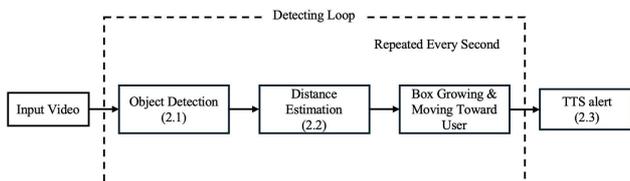


그림 1 시스템 구성도

본 논문에서 제안하는 시스템은 그림 1과 같이 카메라를 통해 입력 데이터를 수집하며, 이를 세 가지 주요 모듈로 처리한다. 첫 번째는 객체

2.2 거리 추정 모듈

앞서 YOLOv5n 모델을 사용하여 각 프레임에서 객체를 탐지하고, 해당 객체에 대한 Bounding Box의 좌표를 계산하게 된다. Bounding Box는 객체를 둘러싸는 직사각형으로, 좌상단 좌표 (x_1, y_1) 와 우하단 좌표 (x_2, y_2) 로 정의된다. YOLOv5n은 이러한 Bounding Box 좌표를 실시간으로 계산하여 객체의 위치를 지속해서 추적한다. 먼저 객체의 접근 여부를 정확하게 판단하기 위해 Bounding Box의 면적 변화율과 중심점의 이동 방향을 함께 분석한다. 각 Bounding Box의 면적을 다음과 같이 계산한다.

$$A_t = (x_2 - x_1) \times (y_2 - y_1)$$

수식 3. 면적 계산

객체가 사용자에게 접근할수록 Bounding Box의 면적은 증가하게 된다. 이에 따른 면적 변화율을 다음과 같이 계산할 수 있다.

$$\Delta A = \frac{A_t - A_{t-1}}{A_{t-1}}$$

수식 4. 변화율 계산

Bounding Box의 면적 변화율이 증가하는 것은 객체의 면적이 빠르게 커지고 있음을 의미하며, 이를 통해 객체가 사용자에게 접근 중인지 판단할 수 있다. 또한, Bounding Box 중심점의 이동 벡터의 크기와 방향을 분석함으로써 객체가 실제로 사용자 쪽으로 이동하고 있는지를 확인할 수 있다. 특히, 면적 변화율이 임계값을 초과하고 중심점의 이동 벡터의 방향이 카메라 쪽을 향할 때, 해당 객체는 사용자에게 접근 중인 위험 요소로 판단된다. 이러한 조건이 충족되면 TTS(Text-to-Speech) 모듈을 통해 음성 경고가 출력되는 단계로 이어진다.

2.3 TTS 모듈

본 논문에서 제안한 시스템에서는 실시간 경고 음성을 제공하기 위해 딥러닝 기반의 TTS 모델 Efficient Speech[3]를 사용하였다. Efficient Speech는 라즈베리 파이와 같은 저전력 기기에서도 효율적으로 구동되며, 시스템의 실시간성 요구사항을 충분히 충족할 수 있다. 또한, ONNX 최적화를 통해 처리 능력이 향상되었으며 이를 통해 제한된 하드웨어 환경에서도 신속하고 정확한 음성 경고를 제공한다. 기본적으로 Efficient Speech는 영어로 동작하지만, 본 시스템에서는 KSS(Korean Single Speaker Speech) 데이터셋을 활용하여 한국어 지원이 가능하도록 fine tuning을 수행하였다.

거리 추정 모듈에서 탐지된 위험 객체는 영어로 표현된 정보를 한국어로 변환한 후, 텍스트 문장을 생성하여 Efficient Speech 모델에 입력한다. 이후 Efficient Speech 모델은 입력된 텍스트를 기반으로 음성 신호를 생성하며, 이 음성 신호는 즉시 재생되어 사용자에게 전달한다.

III. 시스템 구현 및 결과

3.1 시스템 개발 환경

본 논문에서 제안한 시스템에서 사용할 On-device 장치는 라즈베리 파이5이다. 실시간 객체 인식을 위해 라즈베리 파이5에 존재하는 USB 포트에 카메라 모듈을 장착하여 객체 인식을 수행한다. 개발 도구는 python이며 ONNX를 통하여 배포하는 형식으로 개발한다.

3.2 시스템 구현

본 논문에서 제안한 시스템의 성능을 검증하기 위해 다양한 시나리오에서 시연 영상을 촬영하였다. 시연 영상에서 시야에 차량, 사람과 자전거가 동시에 포착되는 3가지 경우를 확인할 수 있다.

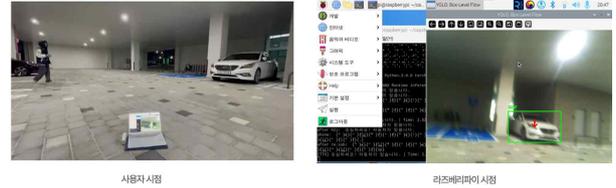


그림 2 시야에 차량이 포착될 경우

그림 2에서와 같이 시야에 차량이 포착될 경우, YOLOv5n 모델은 차량을 탐지하여 Bounding Box를 생성하고 면적 증가율을 계산하여 차량과 가까워지고 있음을 판단한다. 중심점 이동 방향을 통해 차량이 사용자 진행 방향에 존재하고 있음을 판단하면, 시스템은 경고 음성을 Efficient Speech를 통해 실시간으로 출력한다.



그림 3 시야에 사람과 자전거가 포착될 경우

그림 3에서는 시야에 사람과 자전거가 포착되는 복합적인 상황을 시연하였다. 두 객체가 모두 사용자에게 접근하고 있는 것으로 판단되면, 시스템은 복합적인 경고 음성을 생성하여 출력한다.

III. 결론

본 논문에서는 시각장애인을 위한 도보 보조 시스템을 개발함으로써, 그들의 일상생활에서의 자율성과 안전성을 크게 향상시키는 것을 목표로 하였다. YOLOv5n 모델의 FP16 양자화와 ONNX 최적화, Efficient Speech TTS 모델의 한국어 파인튜닝을 통해, 본 시스템은 저 사양 기기인 라즈베리 파이에서도 높은 성능을 발휘하며 실시간으로 정확한 경고를 제공할 수 있음을 입증하였다. 이를 통해 시각장애인이 위험 상황을 신속하게 인지하고 적절히 대응할 수 있도록 지원하였다. 본 논문에서 제안한 시스템은 기술적 측면에서의 성과뿐만 아니라, 사회적 측면에서도 큰 의미를 지닌다. 시각장애인이라는 사회적 약자를 대상으로 한 실용적인 보조 소프트웨어를 개발함으로써, 그들의 이동 안전성을 높이고 사회적 참여를 촉진하는 데 기여하였다. 향후, 본 시스템은 추가적인 장치를 통한 정확한 거리 추정 알고리즘 등의 기능을 도입함으로써 발전할 수 있을 것이다.

참고 문헌

- [1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.
- [2] Ultralytics. (n.d.). YOLOv5 Documentation. Retrieved December 24, 2024, from <https://docs.ultralytics.com/ko/models/yolov5/>.
- [3] Atienza, R. (2023). EfficientSpeech: An On-Device Text to Speech Model. arXiv preprint arXiv:2305.13905.