



보고서

시각장애인을 위한 지능형 장애물 탐지 및 간판 인식 시스템

과목명 : 캡스톤디자인

팀명 : 하얀조끼

팀장 : 20221052 서민경

팀원 : 20221044 김내경, 20221047 김윤희

지도교수 : 현장훈 교수님

1. 연구 배경 및 목적

시각장애인의 이동 보조 도구로 가장 널리 사용되는 '흰 지팡이'는 물리적인 장애물을 감지하는 데 유용하지만, 복잡한 도심 환경에서는 한계를 갖는다. 특히 변화가나 상점가와 같은 밀집된 공간에서는 고정된 장애물과 움직이는 장애물이 혼재되어 있어 경로를 정확히 파악하기 어렵다. 또한, 흰 지팡이는 주로 발 앞의 장애물 감지에 초점이 맞춰져 있어, 상체 높이에 있는 간판이나 돌출된 구조물과 같은 장애물은 효과적으로 탐지하기 어렵다. 이와 더불어, 시각장애인은 특정 목적지를 찾거나 새로운 경로를 탐색할 때 어려움을 겪는다. 예를 들어, 처음 방문하는 지역에서는 주변 환경을 직관적으로 인식하기 어렵고, 보행 중 가게나 건물의 위치를 파악하는 것이 쉽지 않다. 특히 간판과 같은 시각적 정보를 활용할 수 없는 환경에서는 특정 장소를 식별하는 것이 더욱 어렵다. 이는 시각장애인의 이동 효율성과 안전성을 저하시켜 사회적 활동의 제한을 초래할 수 있다. 최근 객체 탐지(Object Detection) 및 텍스트 인식(OCR, Optical Character Recognition) 기술이 발전함에 따라, 장애물을 인식하고 주변의 상점 정보를 제공하는 시스템을 개발할 수 있는 가능성이 열리고 있다. 이러한 기술을 활용하면 기존 흰 지팡이의 한계를 보완하고, 시각장애인의 이동 편의성을 대폭 향상시킬 수 있을 것으로 기대된다.

2. 시스템 설계 및 구현

2.1 상대 속도 추정 기반 동적 정적 장애물 분류 알고리즘

2.1.1 사용자 이동 거리 추정 (IMU 기반)

사용자의 움직임을 정확히 추정하기 위해 IMU 센서(가속도계)를 활용한다.

IMU로부터 얻은 가속도 데이터를 시간 간격(Δt) 기준으로 1차 적분하여 속도(v)를, 다시 2차 적분하여 위치(p)를 계산한다.

이를 통해 두 프레임 간의 물리적 위치 변화 Δs 를 추정할 수 있다:

- 속도:

$$v(t_n) = v(t_{n-1}) + \int_{t_{n-1}}^{t_n} a(\tau) d\tau$$

- 위치:

$$p(t_n) = p(t_{n-1}) + \int_{t_{n-1}}^{t_n} v(\tau) d\tau$$

- 이동 거리:

$$\Delta s_n = |p(t_n) - p(t_{n-1})|_2$$

2.1.2 객체 검출 및 트래킹 (YOLO + ByteTrack)

정확한 객체 추적을 위해 세그멘테이션 기반 검출과 바운딩 박스 트래킹을 결합한다.

- 객체 검출: YOLOv11-seg 모델을 사용하여 객체의 바운딩 박스 및 픽셀 단위 마스크를 추출한다.
- 중복 제거: Non-Maximum Suppression(NMS) 기법을 적용하여 중복 탐지를 제거 한다.
- 객체 추적: ByteTrack 알고리즘을 사용하여 프레임 간 객체를 연결하고 고유 ID를 유지한다.
(칼만 필터 기반의 예측 및 최장 경로 매칭을 통해 일시적 가림에도 안정적인 추적을 보장한다.)

2.1.3 객체별 거리 산출 (깊이 센서 기반)

깊이 정보를 통해 객체와 사용자의 거리를 추정한다.

- RGB-D 센서에서 얻은 깊이 영상을 객체의 마스크에 맞춰 필터링한다.
- 노이즈, 그림자 등으로 인한 이상치를 제거한 후, 중앙값(Median)을 객체 거리로 정의한다:

$$d_i = \text{median}\{ \text{depth}(x, y) \mid (x, y) \in \text{mask}_i \}$$

2.1.4 정적 및 동적 장애물 분류

객체의 이동 여부를 사용자의 이동과 비교하여 분류한다.

- 이전 프레임과 현재 프레임 사이의 거리 변화량을 계산한다.

$$\Delta d_i = d_{i,\text{prev}} - d_{i,\text{cur}}$$

- 분류 기준:

- 동적 장애물:

$$\Delta d_i > \Delta s_n \quad \text{and} \quad d_{i,\text{cur}} < 7 \text{ m}$$

- 정적 장애물:

$$\Delta d_i \leq \Delta s_n \quad \text{and} \quad d_{i,\text{cur}} < 5 \text{ m}$$

이 방식은 사용자의 실제 이동 거리를 기준으로 상대적 거리 변화를 분석함으로써, 고정된 장애물과 이동 중인 장애물을 효과적으로 구분할 수 있도록 한다.

2.2 방향 안내 시스템

2.2.1 방향 구역 분할 (Left–Center–Right Region Segmentation)

본 시스템은 사용자의 진행 방향 판단을 위해 카메라 시야를 기준으로 전체 화면을 좌·

중앙·우의 세 구역으로 분할한다. 화면의 가로폭을 3등분하여 구성하였으며, 각 객체의 중심 좌표(cx)를 기준으로 해당 객체가 어느 구역에 위치하는지를 판단한다. 이를 통해 객체의 공간적 분포를 효율적으로 해석하고, 각 방향별 위험도 계산의 기초 데이터를 제공한다.

2.2.2 방향별 위험도 계산 (Directional Risk Scoring)

객체의 위치뿐만 아니라 거리, 동적 여부, 객체 크기 변화(approaching 여부), 객체 종류를 반영하여 각 방향의 위험도를 산출한다. 위험도 계산식은 다음과 같다.

$$Risk_{dir} = \sum_{i \in dir} W_{class}(i) \cdot W_{state}(i) \cdot \frac{1}{distance(i)}$$

또한 바운딩 박스의 크기가 일정 비율 이상 증가하는 경우, 즉 사용자 방향으로 접근 중인 동적 객체로 판단하여 해당 객체가 속한 방향의 위험도를 추가로 증가시킨다.

2.2.3 최적 진행 방향 선택 (Navigation Direction Recommendation)

세 구역(좌·중앙·우)의 위험도를 비교하여 가장 위험도가 낮은 방향을 최적 이동 방향으로 사용자에게 안내한다.

예를 들어, 중앙 구역에 동적 객체가 접근 중이거나 장애물이 밀집해 있으며 좌측과 우측의 위험도가 상대적으로 낮을 경우, 시스템은 **"좌측으로 이동하세요" 혹은 "우측으로 이동하세요"**와 같은 안내 메시지를 제공한다.

안내 메시지는 다음과 같은 형태로 출력된다.

"정면 경로가 안전합니다."

"좌측에 장애물이 밀집되어 있습니다. 우측으로 이동하세요."

"우측에서 동적 객체가 접근 중입니다. 좌측 경로를 이용하세요."

이 과정은 매 프레임마다 반복되므로, 시스템은 사용자 이동 과정 전체에서 지속적으로 안전한 방향을 제시한다.

2.3 시야 정보를 활용한 시각 안내 시스템

2.3.1 정지 상태 감지 (User Stillness Detection)

IMU 센서 데이터를 분석하여 사용자가 일정 시간 이상 움직이지 않을 경우 “정지 상태”로 간주한다. 이 시점부터 주변의 안내 텍스트를 인식하여 정보 제공을 시작한다.

2.3.2 시야 영역 필터링 및 전처리

정지 상태가 되면 카메라 시야 내 객체 중 일정 거리 이내에 있는 부분만 필터링하며, OCR 성능 향상을 위해 밝기 정규화, 왜곡 보정 등의 전처리를 수행한다.

2.3.3 OCR 후처리 및 클러스터링

OCR 결과 중 중복되거나 깨진 텍스트는 제거하고, 위치 기반 클러스터링을 통해 중요 단어(예: 시설명, 안내문)를 추출한다. 필요 시 Depth 데이터를 이용해 거리 정보 기반 정렬을 수행한다.

2.3.4 TTS(Text-to-Speech) 변환 및 출력 (추가)

클러스터링된 텍스트 중 가장 중요한 내용을 TTS 엔진을 통해 음성으로 출력한다.

동시에 YOLO 등의 객체 탐지 결과를 활용하여 사용자의 이동 경로 상 장애물을 실시간으로 분석한다.

장애물의 위치(x축 중심 기준)와 거리 정보를 바탕으로 다음과 같이 경고음을 제공한다.

왼쪽 장애물: 낮은 톤의 삑삑 소리

정면 장애물: 높은 톤의 삑삑 소리

오른쪽 장애물: 중간 톤의 삑삑 소리

경고음은 장애물이 일정 거리(예: 1.5m) 이내에 접근했을 때만 출력하며, 이를 통해 사용자가 직관적으로 장애물 방향을 인지하고 안전하게 이동할 수 있다.

TTS 음성과 경고음은 동시에 제공될 수 있으며, 이를 통해 시각장애인은 안내문 정보와 주변 장애물 위험 정보를 통합적으로 인식할 수 있다.

2.3.5 TTS(Text-to-Speech) 변환 및 음성 안내 출력

클러스터링된 텍스트와 장애물 방향·거리 정보를 결합해, “오른쪽 5.2미터 거리에 ‘출구’ 표지판이 감지되었습니다.” 와 같은 형태의 자연스러운 안내 멘트를 생성한다.

또한 GPT 기반 문장 교정을 통해 OCR 오타를 자동 보정하여 사용자가 이해하기 쉬운 형태로 음성 안내를 제공한다.

본 시스템은 시각장애인의 자립적인 이동을 지원하는 데 중요한 역할을 수행한다. 장애물 충돌 방지를 위해 카메라 화면을 좌·정면·우 세 구역으로 분할하고, 각 구역 내 물체 위치를 분석하여 보행자에게 이동 방향을 안내한다. 알고리즘을 통해 탐지된 정보가 과도하게 많아 사용자가 혼란을 겪지 않도록, GPT 기반 필터링을 통해 중요 정보만 선별한다. 이후 필터링된 정보를 바탕으로 음성 안내를 제공하여 사용자가 장애물의 위치와 방향을 직관적으로 인지할 수 있도록 설계할 예정이다. 또한 간판이나 안내 표지판 등 시각 정보를 OCR로 인식하고, 중요 내용을 선별하여 음성으로 전달함으로써 사용자가 필요한 정보를 효율적으로 확인할 수 있도록 한다. 이러한 기능들은 시각장애인의 다양한 환경에서 안전하게 이동할 수 있도록 돋는 핵심 요소가 되며, 시각적 정보와 음성 안내를 결합한 통합 지원 도구로 자리매김할 수 있다.