

시각장애인을 위한 보행 네비게이션 앱

SeeFar

물리적으로 멀리 보며 안내를 한다 + 시각장애인을 위한 더 나은 미래를 본다

ADAS&Vehicle Intelligence Project 주민범 연구원/ 리빙솔루션SW플랫폼개발Project 장재우 연구원
리빙솔루션SW플랫폼개발Project 이재원 연구원 / 리빙솔루션SW플랫폼개발Project 이예은 연구원

배경 및 목적

- 보행의 자유는 현대 사회에서 필수적인 권리이나, 시각장애인은 이에 대해 많은 어려움을 겪고 있음, 이에 본 어플리케이션은 시각장애인의 보행을 보다 안전하고 원활하게 지원하기 위해 개발 됨
- 실시간 네비게이션과 음성 안내 기능을 활용하고 장애물 회피 경로 안내 등의 정보를 직관적으로 제공하면서 보다 시각장애인이 독립적으로 이동할 수 있도록 돕는 것이 개발의 목적임

시스템 요구사항

- 본 시스템은 영상 및 음성 입력으로 받음

Image captioning & VQA:

주변 환경이 변경되면 사용자에게 음성으로 알려야 하며, 사용자 요청 시 시스템은 현재 주변 상황에 대하여 설명해야 함

Visual localization and navigating

현재 위치 추정 및 목적지 경로 탐색으로 사용자가 목적지로 안내 요청 시 목적지에 대한 경로 안내 음성 출력으로 수행해야 하며, 도착할 시 도착 여부를 안내해야 함

Object detection :

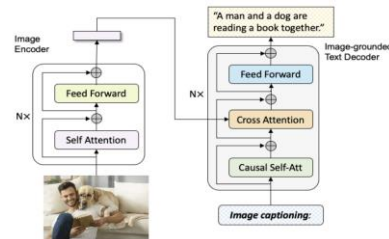
잠재적 위험 요소 감지로 사용자 요청 외로 잠재적인 위험 상황에 대해 탐지를 수행 하여 감지될 경우 사용자에게 음성 출력으로 안내해야 함

STT & TTS :

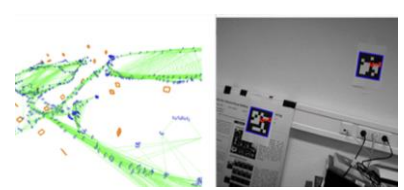
사용자는 시스템에게 음성을 통해 명령을 수행하고 음성을 텍스트로 변환하여 각각의 기능 컴포넌트에서 처리할 수 있도록 하고, 다시 음성으로 변환하여 출력 할 수 있음

- VQA : Visual Question Answer
- STT : Speech To Text
- TTS : Text To speech

Feature 1 : 주변 상황 설명



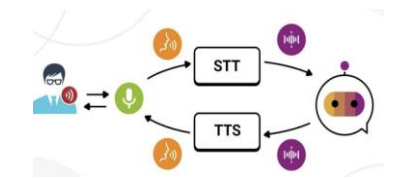
Feature 2 : 현재 위치 추정 및 목적지 경로 탐색



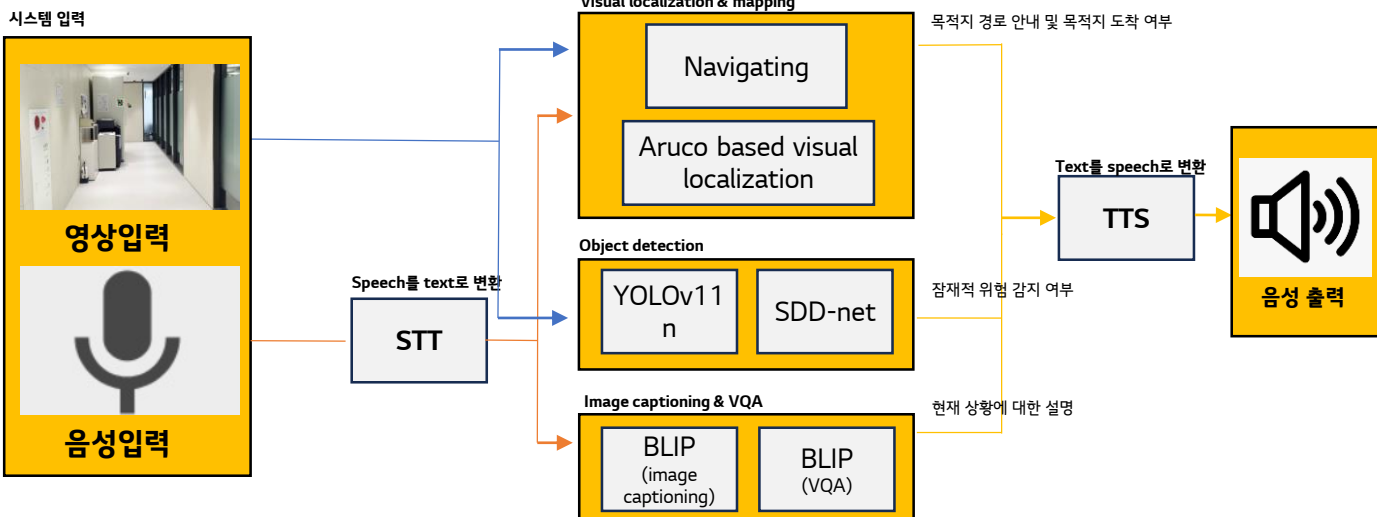
Feature 3 : 잠재적 위험 요소 감지



Feature 4 : STT & TTS



시스템 구현



구현 결과

Image captioning & VQA

STT로부터 질문(text)를 받으면, 이에 대한 답변 text로 생성

상황 변경 인지도 해당 상황에 대한 묘사 text로 생성 On board 구동 위해 모델 양자화 적용



Image captioning :

A) a hallway with a black door and a white wall

VQA :

Q) What kind of place does it look like?

A) Office

Object detection

YOLO v11n, SSD net을 병렬로 사용하여 정확한 성능을 요구하는 경우 전자를 빠른 성능을 요구하는 경우 후자를 inference할 수 있도록 병렬화 수행
잠재적 위험을 감지하면 해당 내용에 대한 text를 생성하여 TTS model에 전달

STT & TTS

Wakeword (SeeFar야)를 받으면 SST API가 동작하여 speech를 Text로 저장
목적지를 speech로 받으면, 해당 내용을 navigating에 text 전달
현재 상황을 알려달라는 요청을 받으면, VQA에 text 전달
각각의 컴포넌트로부터 응답 받은 text를 다시 speech로 변환하여 사용자에게 전달

Visual localization and navigating

STT로부터 목적지(text)를 받으면, 이에 대한 경로를 TTS에 전달하며, 현재 위치는 aruco marker based SLAM으로 추정

mapping 위해 사전에 정의된 marker를 특정 지점에 배치함.
Marker정보와 카메라 정보 기반으로 현재 위치를 추정할 수 있음



Navigating :

Q) Guide me to the classroom

A) Go straight



Object detection:

A) Warning : dining table is in front of you

결론

본 시스템은 영상 및 음성 입력 기반으로 동작하며, 시각 장애인을 위한 실시간 내비게이션 및 환경 인식 지원을 목표로 한다.

첫째, 이미지 캡셔닝(Image Captioning) 및 시각 질문 응답(VQA) 기능을 통해 주변 환경 변화를 감지하고, 이를 사용자에게 음성으로 안내하며, 사용자의 요청 시 현재 환경에 대한 설명을 제공한다.

둘째, 시각적 위치 추정 및 내비게이션 기능을 활용하여 현재 위치를 추정하고 목적지까지의 경로를 탐색하며, 사용자의 경로 안내 요청 시 음성 출력을 통해 목적지까지 안전하게 이동할 수 있도록 지원한다. 목적지에 도착하면 도착 여부를 안내한다.

셋째, 객체 탐지(Object Detection) 기능을 적용하여 잠재적 위험 요소를 실시간 감지하며, 사용자의 요청 여부와 관계없이 감지된 위험 요소를 음성으로 경고하여 사고를 예방한다.

넷째, 음성 인식(STT) 및 음성 합성(TTS) 기술을 기반으로 사용자가 음성을 통해 명령을 수행할 수 있도록 하며, 명령은 텍스트로 변환되어 시스템의 각 기능 모듈에서 처리된 후 다시 음성으로 변환하여 사용자에게 제공된다.

본 시스템을 통해 시각 장애 사용자는 실시간으로 주변 환경을 인식하고, 안전한 이동이 가능하며, 직관적인 음성 인터페이스를 활용하여 보다 원활한 일상생활을 영위할 수 있다.