INSIGHT: In-device Navigation and Scene Interpretation Glasses for Human-centered Travel of Low-vision Users

EECS 473: Advanced Embedded System | Guanyu Xu[1], Haobo Fang[1], Ruopu Dong[1], Yizhe Shen[1], Zhuoyang Chen[1], Jinlin Li[1]
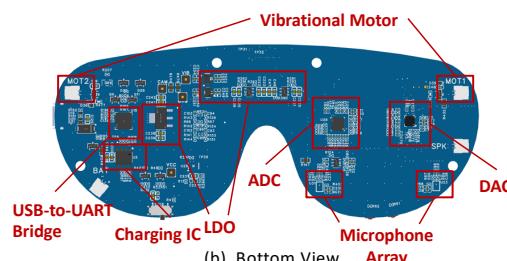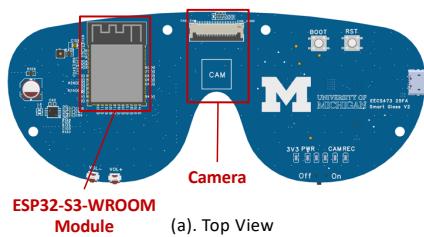
{xuguanyu, fanghb, druopu, wshenyz, janchen, jinlinli} @umich.edu

Introduction

- A system built to assist **visually impaired people** in **navigation** and **scene description.**
- **Low-cost**: Uses commodity hardware and compact models so the system costs less than $300, making it accessible to everyday users.
- **Real-time**: Processes sensor and camera data on-device to deliver instant audio and haptic feedback for safe navigation and scene description.
- **Privacy**: All perception and language models run locally and offline, so no video or audio ever leaves the device.

Smart Glass System

Customized PCB

ESP32-S3-WROOM Module
Camera
(a). Top View

Vibrational Motor
USB-to-UART Bridge
Charging IC **LDO**
ADC
Microphone Array
DAC
(b). Bottom View
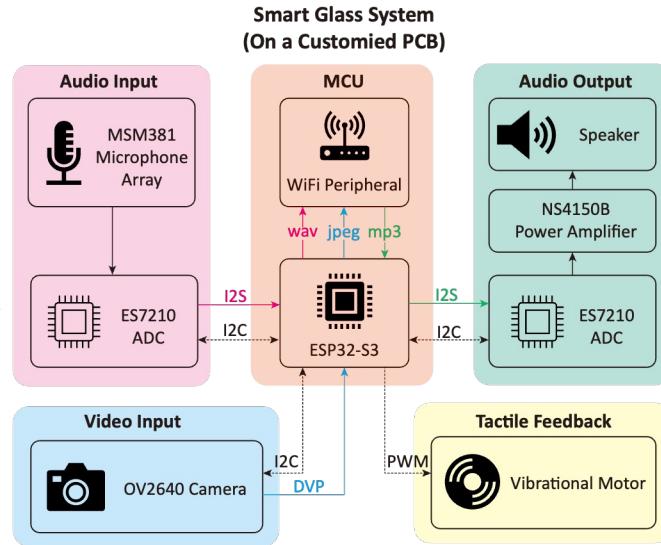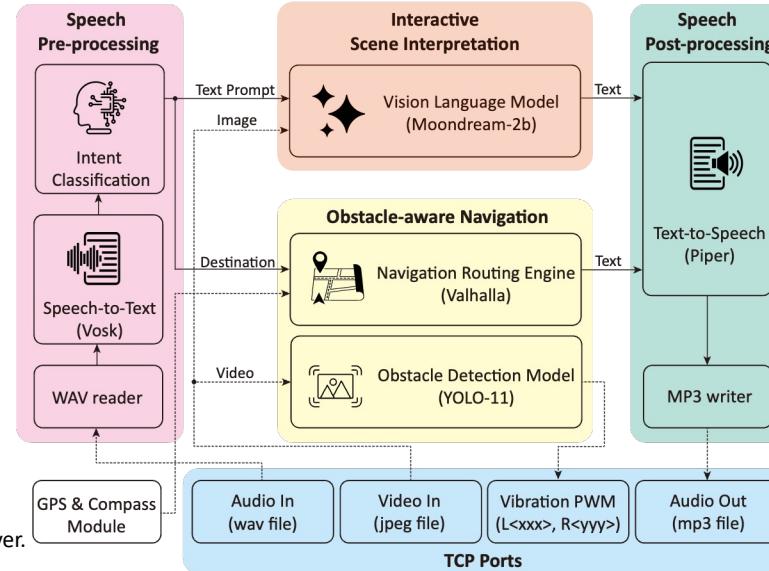
Software Overview

- Task 1: Wake word ("Hi ESP") detection.
- Task 2: Audio recording and streaming.
- Task 3: Take video from camera and stream out.
- Task 4: Receive audio and play on the speaker.
- Task 5: Vibrational motor PWM receiver and driver.

Function Block Diagram

Smart Glass System (On a Customied PCB)

Audio Input
MSM381 Microphone Array
ES7210 ADC

MCU
WiFi Peripheral
wav jpeg mp3
ESP32-S3

Audio Output
Speaker
NS4150B Power Amplifier
ES7210 ADC

I2S — I2C — I2S — I2C

Video Input
OV2640 Camera
I2C — DVP

Tactile Feedback
PWM
Vibrational Motor

Base Station System (On a Jetson Orin Nano)

Speech Pre-processing
Intent Classification
Speech-to-Text (Vosk)
WAV reader

Interactive Scene Interpretation
Vision Language Model (Moondream-2b)

Obstacle-aware Navigation
Navigation Routing Engine (Valhalla)
Obstacle Detection Model (YOLO-11)

Speech Post-processing
Text-to-Speech (Piper)
MP3 writer

Text Prompt — Image — Destination — Video — Text — Text

GPS & Compass Module

TCP Ports
Audio In (wav file) | Video In (jpeg file) | Vibration PWM (L<xxx>, R<yyy>) | Audio Out (mp3 file)

Results

Scene Interpretation

Two young men are sitting at a table in a modern office space. One man is using a laptop while the other is looking at the screen. The room has a minimalist design with white walls, a curved ceiling, and recessed lighting. There are several small tables and chairs in the background and a long bench or seating area. The overall atmosphere appears calm and focused.

What's in front of me?

Obstacle-aware Navigation

Walk west on the walkway

YOLO obstacle detection
person 0.87 0.83

Vibrational Motor Feedback
L = 8
R = 30

Planned GPS Coordinates
Actual GPS Coordinates

Walk west on the walkway
Turn left onto the walkway
Keep right to take the walkway
Your destination is on the left

Acknowledgements

[1]Electrical and Computer Engineering Department, University of Michigan, Ann Arbor, MI, 48109, USA.