GateGuide

E4: Opalina Khanna, Krrish Jain, Daniel Kim

18-500 Capstone Design, Spring 2025

Electrical and Computer Engineering Department

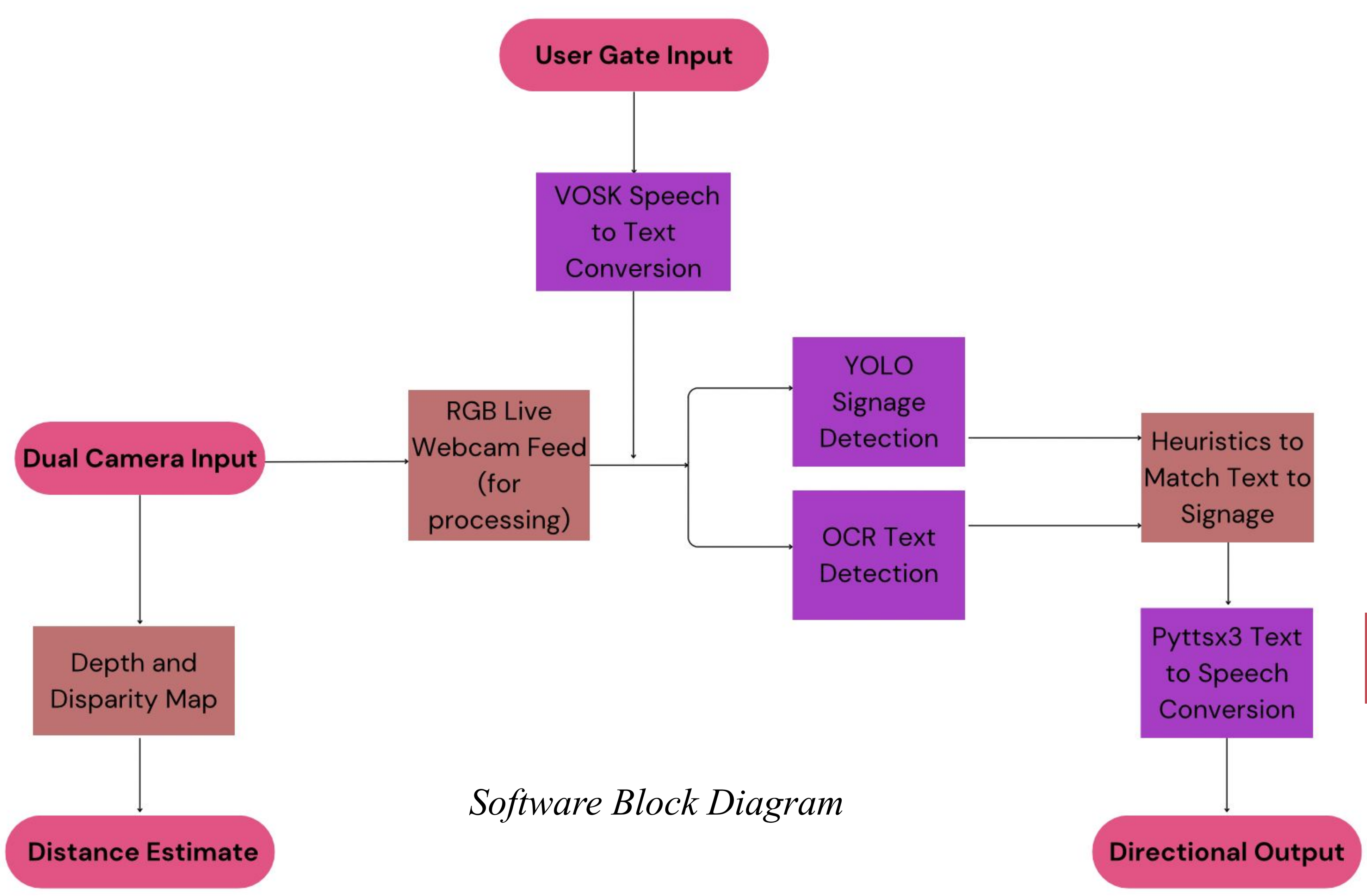
Carnegie Mellon University

Product Pitch

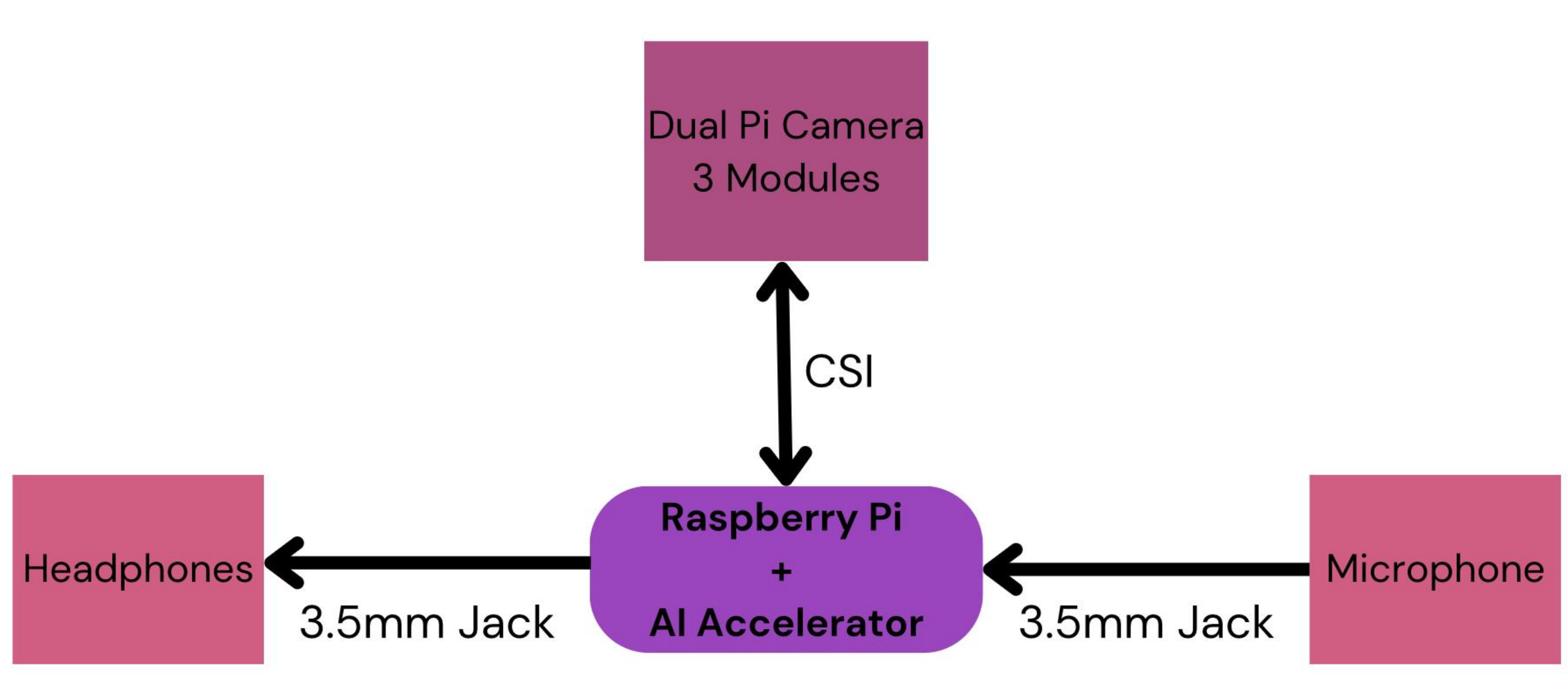
GateGuide is an indoor navigation system, tailored specifically to help visually impaired people find their way through airports. GateGuide attempts to make unfamiliar environments more accessible through real-time sign detection, interpretation, and auditory feedback. The device aims to provide a seamless experience from security to the user's gate, and ensure complete independence in an otherwise daunting setting.

Through this project, we aimed to create an **accurate, fast, portable, and power efficient system** to maximize its impact on our intended demographic. Test results indicated that most of these requirements that we initially outlines were fulfilled, with the exception of certain compromises and tradeoffs (as outlined in detail below) made to preserve latency and prioritize usability of the product.

System Architecture



Software Block Diagram



Hardware Block Diagram

Conclusions & Additional Information

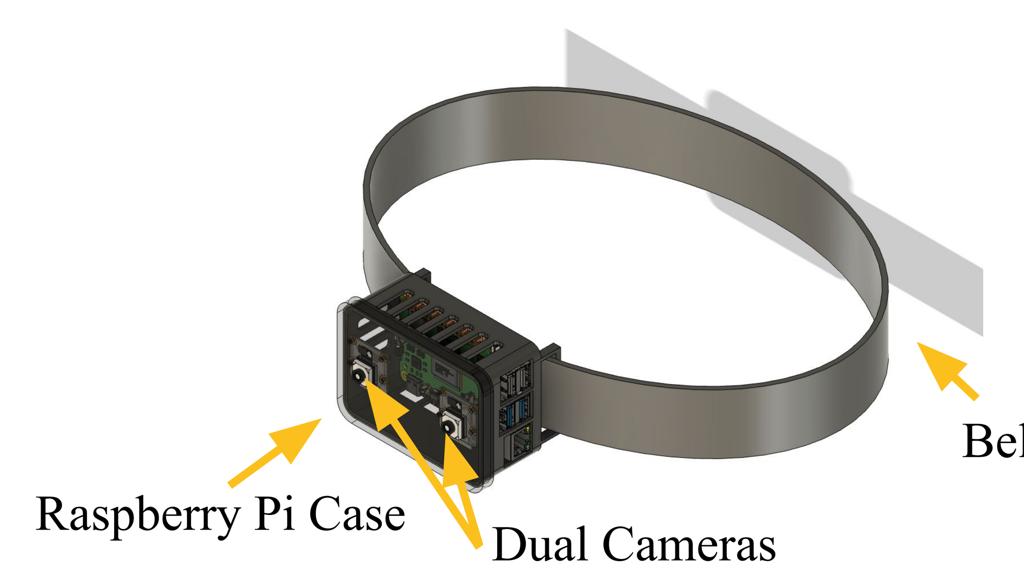
Check out our website!

This product employs technology that is widely applicable to purposes beyond our use case, including but not limited to language aid for non-native speakers. Furthermore, if we were to continue this project, we would expand the navigation capabilities to other modes of transport, including train stations, subways, etc. Overall, this project formed a foundation for us to dive deeper into the tools we used, and understand how individual software and hardware components integrate into a larger, more complex system.

System Description

- **Software Components**
  - **VOSK speech-to-text** to interpret the user's destination gate
  - A **custom trained YOLOv8** model for capturing and interpreting common airport signage. Specifically trained to extract arrows, bathrooms, gate / airplane signs, and handicapped symbols
  - An open-source rapid **Optical Character Recognition** (OCR) model, to extract gate ranges associated with signage and guide the user accordingly
  - **Pyttx3 text-to-speech** to output auditory feedback to guide the user once the directions have been processed
- **Hardware Components**
  - **Raspberry Pi 5 + AI HAT+** – 2.4 GHz quad-core processor with Hailo-8L accelerator for <2s latency for inference
  - **2× Raspberry Pi Camera Module 3** – Sony IMX708 sensors on a 60 mm stereo baseline. Images from Left and Right cameras are rectified. Semi-global block matching (SGBM) is performed on the rectified image to get a disparity map. Depth is calculated using disparity and the baseline.
  - **USB headset** – combined microphone and headphones for STT/TTS I/O
  - **10 000 mAh power bank** – 5V output, ~5 h runtime at ≈10 W draw (≈250 g)
  - **3D-printed PLA enclosure** + electronics – housing, PCBs, mounts & cabling all in ≤150 g package

3D Printed Case



CAD Model for Wearable Device

System Evaluation

Testing Approach

- **Signage**: The ML models used to interpret signage were tested both manually and using existing datasets of images found online. The reported accuracy is a combination of the results obtained (about 200 instances).
- **User Interface**: The UI was tested completely manually, by each of our team members individually, and were also user tested on a group of 5 people (about 100 total trials).
- **Hardware**: The power testing was done by using the device until the battery was exhausted (3 trials).

Design Trade-Offs

Chosen Solution	Alternative (Not Used)
Raspberry Pi 5 • 10–12 W draw vs • ~100g	**Jetson Nano** • Dedicated GPU but 25W draw • ~300g
2× Pi Camera Module 3 • Worse Accuracy • Native CSI encoding → 60 FPS capture	**eYs3D Depth Camera** • True IR-TOF depth, better accuracy • USB interface limits to 10–12 FPS, higher latency
YOLOv8n (Nano) • 80% accuracy @ <1 s latency → meets our ≤2 s target	**YOLOv8 (Full)** • 90% accuracy but >2s latency → fails latency requirement
RapidOCR (Every 10 Frames) • 90% accuracy - 10fps	**EasyOCR (Every Frame)** • 95% accuracy - 3fps

Use-Case Requirements

Metric	Target	Actual
Signage Accuracy	90%	80%
Gate Text Accuracy	95%	90%
End-to-End Latency	<2s	1s
Battery Life	5hr	4.5hr
Weight	<2kg	400g

Technical Challenges

Problem	Outcome / Solution
Running CNN-based networks on the Pi was very slow (No dedicated GPU)	• Frame interval for OCR • Reduced video resolution • Multi-threading
Handling real-time audio and video input while offline (no Wi-Fi dependence)	• Quantized models • Offline TTS and STT
Compatibility issues while compiling custom models on Hailo-8 NPU	• Compromised accuracy and latency

Electrical & Computer ENGINEERING

Carnegie Mellon