

Development of a Vision-Based Smart Cane with RGB-Only Perception for Collision Avoidance and Scene Recognition

Sama Nikanfar, Aref Hebri, Harish Ram Nambiappan, Advitya Garg

Supervisors: Farnaz Farahanipad, and Fillia Makedon

Department of Computer Science and Engineering

Introduction

- Vision-impaired users need support for both **safe navigation** and **environment understanding**.
- Many smart canes rely on **multiple sensors** such as LiDAR, laser scanners, ultrasonic sensors, or depth cameras.
- These designs often increase **cost, weight, power use, and system complexity**.
- We propose a **lightweight smart cane** that uses only a **single RGB webcam** for:
 - obstacle detection,
 - distance estimation,
 - scene recognition.

Background

- Prior smart canes often combine **multiple sensing modalities** to support navigation.
- Multi-sensor systems can require **sensor fusion, calibration, and extra hardware integration**.
- RGB-only methods have shown promise, but they are less commonly implemented in **practical smart cane prototypes**.
- This work addresses that gap with a **single-camera, low-complexity design**.

System Design & Methodology

- Built on a **traditional white cane** platform
- Core Components:** Raspberry Pi 5, 10,000 mAh battery, forward-facing Logitech Brio 4K RGB camera, vibration motor, and push-button mounted near the grip.

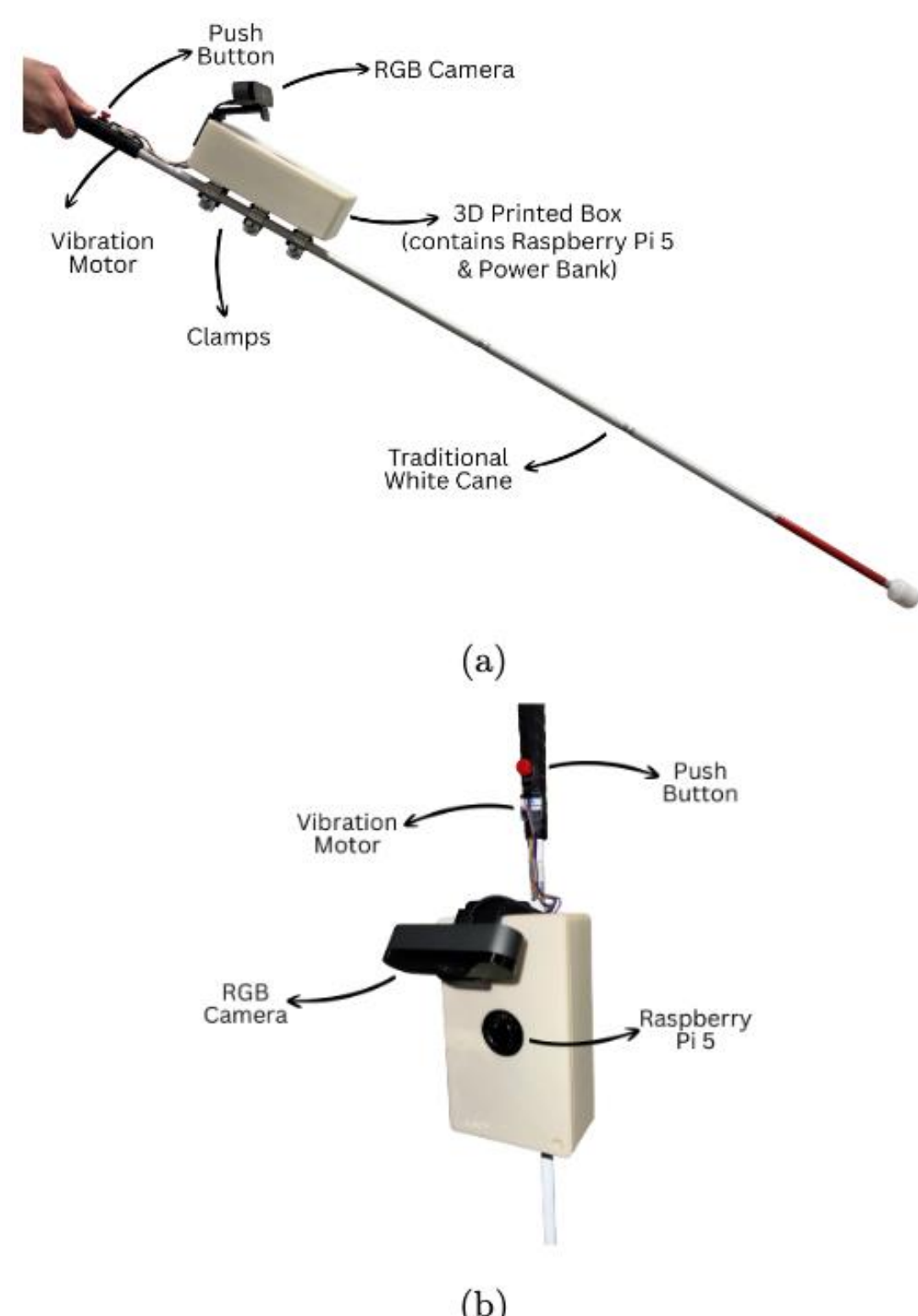


Fig. 1: Hardware configuration of the proposed smart cane

The system operates in two modes.

- Passive Alert Mode:** Uses RGB frames for Object detection & Distance estimation and provides real-time obstacle alerts based on that
- Active Query Mode:** Triggered by button press and generates audio description of surroundings

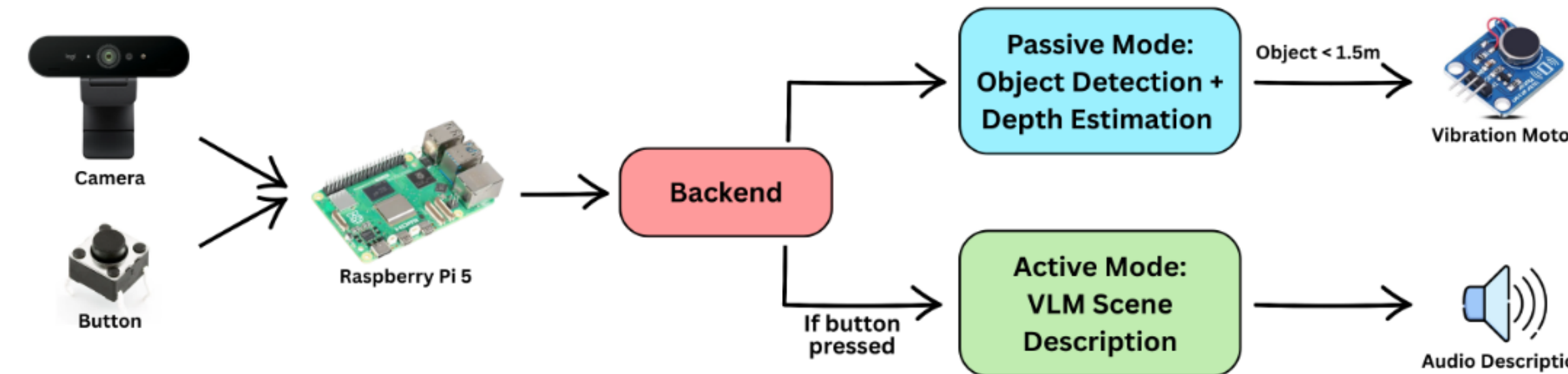


Fig. 2: System architecture of the smart cane

Passive Sensing Pipeline:

- RGB images are sent to a backend processing unit
- YOLOv11 performs object detection and Depth Anything V2 estimates depth
- Object midpoints are used to compute distances and identify the closest object
- If the closest object is within a 1.5 m threshold, vibration alerts and/or auditory feedback are triggered

Active Scene recognition:

- Captures an RGB frame when triggered by the user
- The frame is sent to a Vision-Language Model (VLM)
- Generates a navigation-focused audio description of the scene

User Interaction:

- Single button press provides a one-time scene description
- Double press enables continuous monitoring mode
- Audio updates are given only when meaningful scene changes occur
- Reduces repetitive feedback and cognitive load

Technical Evaluation

- Evaluation focused on:
 - end-to-end latency** of obstacle alerts,
 - distance estimation accuracy**.
- Testing was conducted indoors under **controlled conditions**.
- Cane was held at a **45° configuration** to approximate normal use.
- Obstacles were placed at known distances of **1 m, 2 m, and 3 m**.
- Latency testing used a detection threshold of **1.5 m**.

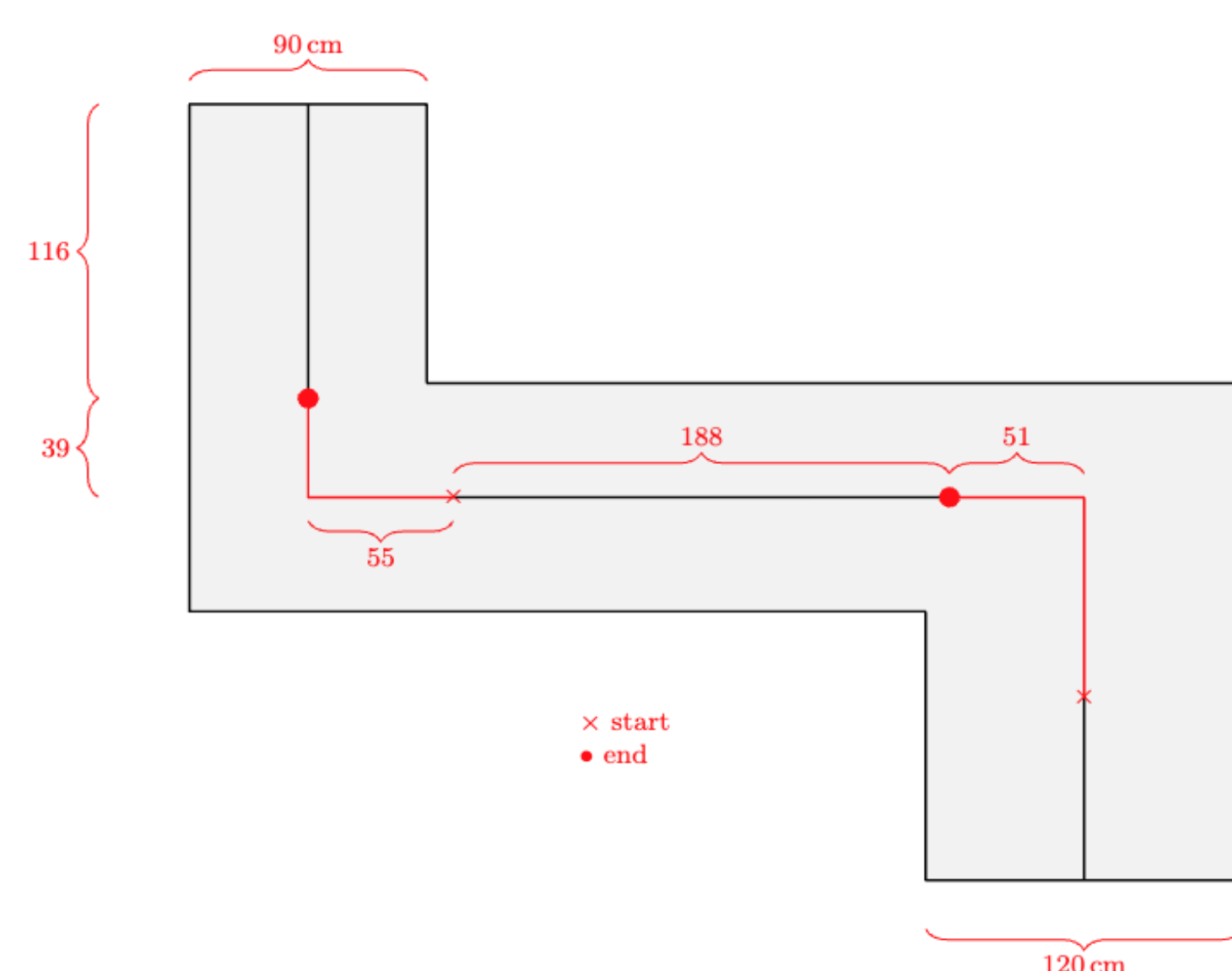


Fig. 3: Top-down view of the navigation path in the corridor.

The red segments indicate intervals when the smart cane vibration feedback is active (obstacle detected), while the black segments indicate intervals when vibration feedback is inactive (no obstacle detected).

Results & Evaluation

- The smart cane delivered haptic warnings with an overall average latency of **0.695 sec**, with mean latency values of **0.652 sec** for chair obstacles and **0.738 sec** for wall obstacles.
- These findings suggest that the system can react quickly enough to provide timely obstacle warnings under controlled conditions.

Table 1. Latency of Obstacle Detection

Obstacle Type	Mean (sec)	Standard Deviation (sec)	Median (sec)	Min-Max (sec)
Chair	0.652	0.118	0.68	-0.26
Wall	0.738	0.173	0.66	-0.39
Overall	0.695	0.147	0.67	-0.45

- For distance estimation, the system achieved an overall **MAE of 0.20 m** and an overall **percentage error of 12.52%** across tested distances from 1-3 m.
- Chair estimates showed a consistent absolute error of about **0.26 m**, while wall estimates were more accurate overall, with errors ranging from **0.12 m to 0.16 m**.
- In addition, corridor-path visualization showed that vibration feedback was spatially localized rather than continuously active, indicating that alerts were triggered mainly in regions where obstacles were encountered.

Table 2. Accuracy of Distance Measured

Obstacle Type	Distance (Ground Truth) (m)	Distance (Estimated) (m)	Mean Error (m)	MAE (m)	% Error (m)
Chair	1.0	1.26 ± 0.055	-0.26	0.26	26.0%
	2.0	2.26 ± 0.055	-0.26	0.26	13.0%
	3.0	3.28 ± 0.084	-0.28	0.28	9.3 %
Wall	1.0	1.15 ± 0.103	-0.148	0.148	14.8%
	2.0	1.84 ± 0.114	0.16	0.16	8.0%
	3.0	3.12 ± 0.110	-0.12	0.12	4.0%
Overall			-0.15	0.20	12.52%

Key Contributions & Conclusions

- Developed a **low-cost, lightweight smart cane** using only a **single RGB camera**.
- Achieved both **collision avoidance** and **scene recognition** without LiDAR, stereo vision, or depth cameras.
- Introduced a **user-controlled active query mode** for on-demand scene understanding.
- Demonstrated that **RGB-only perception is feasible** for real-time indoor assistive navigation.

Poster Based on

- Harish Ram Nambiappan, Aref Hebri, Sama Nikanfar, Advitya Garg, Farnaz Farahanipad, and Fillia Makedon. "Development of a Vision-Based Smart Cane with RGB-Only Perception for Collision Avoidance and Scene Recognition", To be presented at the 28th International Conference on human-Computer Interaction, 2026