AI Based Smart Wearable For Real Time Obstacle Detection With Voice Assistance

Mansi Ahire[1], Chetan Sonawane[2], Archana Bendale[3]
[1,2]*Student, M.Sc. Computer Science, K.K. Wagh Arts, Commerce, Science & Computer Science College, Saraswati Nagar, Nashik (Affiliated to SPPU, Pune) Maharashtra, India*
[3]*Assistant Professor, Department of Computer Science, K.K. Wagh Arts, Commerce, Science & Computer Science College, Saraswati Nagar, Nashik (Affiliated to SPPU, Pune) Maharashtra, India*
doi.org/10.64643/IJIRT211670

***Abstract*—Within busy areas the visually impaired can struggle to navigate safely due to unexpected hazards. To solve this problem, this project proposes a wearable device for visually impaired people to allow safe independent navigation. It uses computer vision, stereo vision for depth perception and sensor fusion. The wearable device contains stereo cameras, an edge computing device and optional time-of-flight/inertial measurement unit sensors. Objects are detected using YOLO family models, distances calculated using stereo vision and a risk assessment system to priorities alerts. Critical alerts will function offline. As this project is in the proposal stage, the system design and evaluation plan are outlined in this work.**

***Index Terms*—Visual Impairment, Smart Wearable, Obstacle Detection, Computer Vision, YOLO.**

I. INTRODUCTION

Blindness and poor eyesight make avoiding obstacles difficult. Although white canes are widely used, using electronic devices can provide additional information about far away, moving objects. Blindness and poor eyesight still have an ongoing effect on people worldwide.

Recent advances in embedded AI means image data can be efficiently processed within a small device. Objects such as people, animals and vehicles can be detected. By using stereo vision distances can also be estimated. However, without understanding what object has been detected, we cannot know whether it is safe to proceed. To produce a navigation aid, information on distance, direction, movement and uncertainty should be taken into consideration.

A wearable device is proposed in this paper to detect obstacles in real time in India. Obstacles being detected include motor bikes, auto rickshaws, potholes, speed humps, steps, open drains, poles and people. Warnings will be given in English, Marathi and Hindi.

The system will be an offline solution. Detection of objects, distance estimation and critical warnings will all take place at the device. As an optional feature, additional features may be given by using internet, but even if no internet connection is available, critical warnings will still be available.

II. LITERATURE REVIEW

There have been many pieces of research into wearable systems. Obstacle detection has been done by RGB-D, stereo vision, object detection and haptic/auditory feedback. For obstacle detection, a wearable solution has been created using RGB-D and haptic feedback. It was tested by blindfolding participants and asking them to complete an obstacle course. [1] This shows that depth sensing in wearables can be beneficial. Furthermore, users can be used in testing.

In another study, obstacle avoidance was performed in a wearable system using video and depth. [2] In addition, RGB-D sensing has been implemented in a wearable system using point cloud data, IMU and YOLO to perform detection. [3] This shows that depth sensing can be beneficial in wearable systems, and that multiple sensors can be incorporated.

Models have been made to improve accessibility. AI4Bharat have produced IndicConformer, which provides models that can recognise Indian languages. The Indic-TTS and Indic Parler-TTS allow speech. [4-6] The VL53L5CX provides a multi-zone TOF sensor.

[7] The OAK-D Lite is a small stereo camera with known depth abilities. [8, 9]

Overall, it is clear that depth sensing wearables, multi-modal sensing and assisting navigation are benefits. This proposal will incorporate these findings to assist with commonly found obstacles in Indian roads/pavements. Alerts in English, Marathi and Hindi will be provided.

III. PROBLEM STATEMENT

Insufficient information may be given by a camera. We would like to know how far away the object is, what direction it is coming from, if it is moving, and whether the ground looks clear. Transparent/reflective/thin/partly occluded objects may be problematic to detect using sensors. Reliance on cloud means the system may not be available.

Using a language may cause issues. This project will create a wearable multi-sensor system with no reliance on the cloud, which can detect objects, determine depth, prioritise risk and output speech in multiple languages. If the system cannot understand what is happening in the scene, uncertainty will be provided.

IV. OBJECTIVES

Evaluate existing systems to identify advantages and disadvantages
Wearable device with stereo cameras, edge processing unit and optional additional sensors.

Recognise obstacles and hazards common in Indian roads: people, vehicles, animals, poles, steps.
Estimate distance using stereo sensing. If short distances are required, time of flight can also be used.
If an inertial measurement unit is available, head orientation can be determined. Create risk module to identify priority hazards and add uncertainty.

Speech recognition in three languages:
English, Marathi and Hindi. The language will be selectable by the user. Basic voice commands will be available.
Evaluate prototype. Assess obstacle detection, distance estimation, time delay, number of false alarms, language recognition and usability.

V. METHODOLOGY

Design-based project. Includes depth perception and wearable parts. Testing will be done using video recordings and webcams.

5.1 Data Capture

There will be people, bicycles, motorbikes, cars, auto-rickshaws, animals, poles, steps, pothole-like depressions and speed bumps in the dataset. Videos will be taken in corridors, pavements, residential roads and simulated roads. Publically available datasets may be used to pre-train however local data must be captured to get data specific to India. People must have correct consent to be identified.

5.2 Object Detection

The YOLO family will be used to perform object detection. Bounding boxes and confidence measures will be given. Object detection will start from a pre-trained model. However, once data has been labelled within the local dataset, the model can be further tuned. Specific hazardous objects such as potholes, open drains and speed bumps may be needed in the training data to make sure they can be correctly identified by a general model.

5.3 Distance Calculation

Physical distance will be calculated using stereo vision. Features will be matched between images from the two cameras. The depth mapping needs to be calibrated and matched with the RGB image. This allows detection results to be associated with the correct distance. Relative distances can be calculated using monocular depth.

5.4 Optional: Time of Flight and Inertial Measurement Units

Time of flight data (ST VL53L5CX) can give more information. It can detect objects up to 4 meters away in 8x8 regions, and has a diagonal field of view of 65 degrees [7]. The inertial measurement unit can give extra information about the head depending on what version of OAK-D Lite is being used.

5.5 Sensor Fusion and Risk Assessment

Camera, depth, time of flight and inertial data will all be synchronised in time. Objects will be matched to their depth. By tracking objects, whether it is moving

towards, or across an expected walking path can be found. A rule-based approach will be used to prioritise hazards based on distance, direction, movement, object and confidence.

If there is conflicting information or it cannot be decided what path they are on, then a cautious alert will be output. For example, 'Stop, I cannot confirm a clear path’.

5.6 Multilingual Speech

Three languages will be supported (English, Marathi, Hindi). IndicConformer can be used for recognition, and Indic-TTS and Indic Parler-TTS for speech [4-6]. This will be trialed to see if low memory use, low latency, high clarity and usability in the local area can be achieved. Pre-written sentences will be used for urgent alerts.

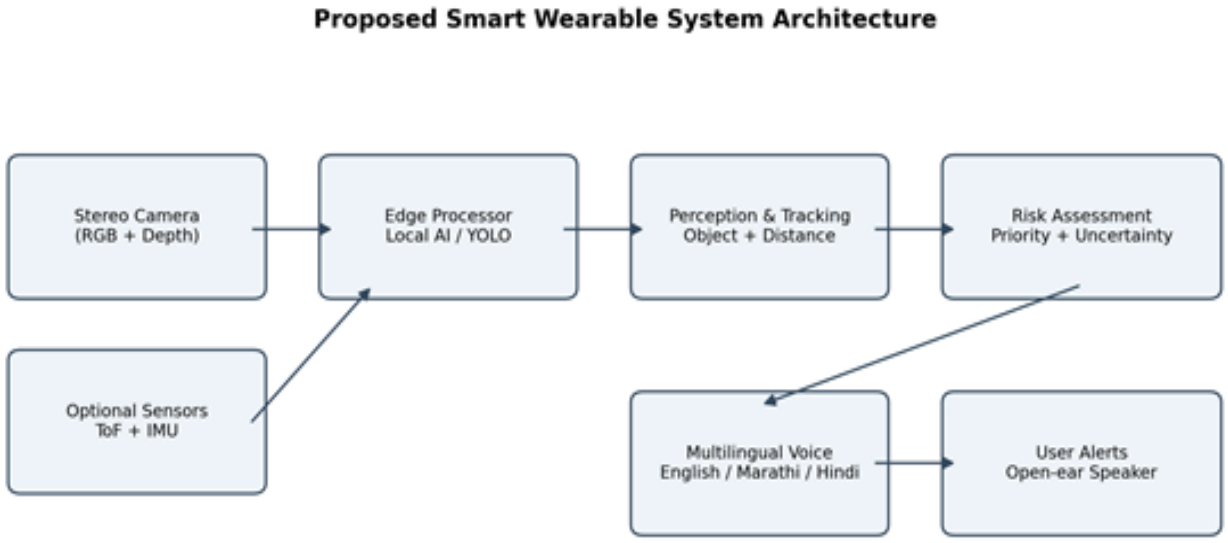
5.7 Alerts

The navigation module will give information, for example obstacle_front, pothole_front, vehicle_left or uncertain_path. These will be matched to pre-written sentences corresponding to the selected language. This is then sent to the text to speech module. Low priority alerts will be restricted but urgent alerts will be prioritised.

VI. SYSTEM DESIGN

The system will consist of sensor, perception, sensor fusion, risk assessment and user assistance modules. Main components will be in a waist/shoulder strap to leave the head lightweight. The main sensor will be stereo cameras facing forwards.

6.1 Overall Design



Continuously stream data from camera and sensors. Detect objects, estimate distance, detect motion and identify potential dangers. If an open area is detected, a short warning message will be shown in the selected language. Sensor failures and uncertainties will be reported.

6.2 Hardware

Table 2. Proposed hardware

Hardware Component	Proposed Choice	Purpose	Alternative
Stereo Camera	Luxonis OAK-D Lite (Check SKU)	RGB + Stereo Depth USB 3	Other calibrated stereo/RGB-D
Edge Processor	Jetson Orin Nano Super Developer Kit	Local AI	USB, I²C, UART Laptop for development Raspberry Pi 5 or phone for lighter models
Near Field Sensor	ST VL53L5CX (Optional)	8x8 ToF Zones I²C	VL53L1X, VL53L8CX or Small LiDAR

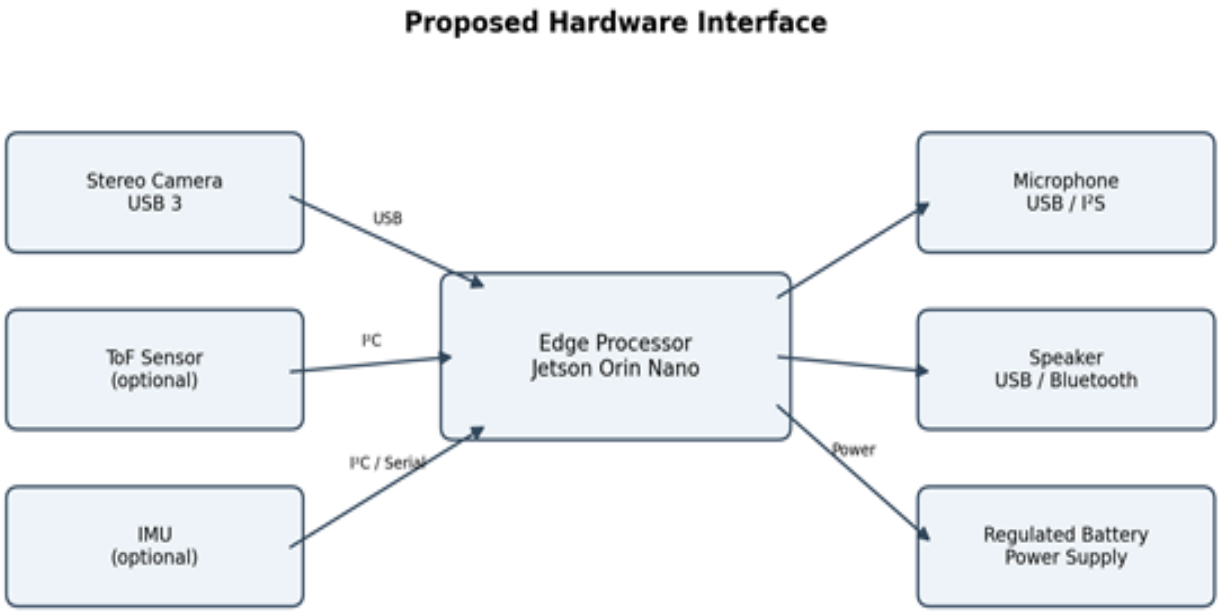
IMU	BNO085 (Optional if IMU available in camera)	Head Orientation I²C/UART	MPU6050 or ICM-20948
Microphone	USB or I²S MEMS	Speech input	Phone mic
Speaker	Open ear or Bone Conduction headset	Alerts	Open ear speaker Optional haptic motor
Power	Protected Battery with Regulated Power	Portable	Jet supply for development
GNSS	Phone GPS or Module (Optional)	Coarse Outdoor Location	Not for obstacle detection

This design contains two parts. There is a wearable on the head device and a computer and battery in the body. The camera is positioned close to the forehead. An open ear speaker is used so that the user can still hear sound around them, and get a notification.

6.3 Sensor Selection & Replacement

Minimum hardware: Camera, Processor, Depth/Range sensor, Speaker and Power. Optional sensors include: time of flight, GPS, additional cameras and IMU. If a sensor is replaced, it must be compatible with the voltage, interface, drivers and calibrated. Additional visibility can be achieved with wide angle camera. If stereo depth is required then it cannot be provided by wide angle camera unless valid distances are given. Alternative sensors should be considered for ranging, field of view, precision and latency.

6.4 Hardware Interface



Stereo Camera: Connect to Edge Processor using USB.

Time of Flight + IMU: Connect to Edge Processor using I²C or Compatible Serial.

If multiple Time of Flights are used, additional circuitry is required to either change address on the Time of Flights, or an I²C MUX.

Microphone: Connect to Edge Processor using USB or I²S.

Speaker: Connect to Edge Processor using USB or Bluetooth.

Power should be supplied according to Voltage and Current required by the processor. Lithium cells and vibrating motors should not be directly connected to the GPIO.

VII. ANALYSIS PLAN

As this is in the proposal phase, this section describes how the project will be analysed. There are three different configurations to understand how much each sub-system contributes (Table 3).

Table 3. Testing Configurations

Configuration	Modules	Purpose
A (Vision)	RGB Camera + Object Detection	Basic Detection
B (Depth)	RGB + Stereo Depth + Tracking	Benefits of Depth & Tracking
C (Multi-Sensor)	RGB-Depth + Optional ToF/IMU + Risk Engine + Multilingual Speech	Sensor Fusion & Structured Alerts

Each configuration will be analysed on the following parameters (Table 4).

Table 4. Parameters

Parameter	How it will be analysed
Precision, Recall & mAP	Using labelled test set. Results will be provided per hazard.
Distance Accuracy	Compare estimated and actual distance. MAE & RMSE
Hazard Recall	Percentage of hazards staged that were detected.
False Alarms	Number of false alerts per minute/route.
Latency (End to End)	Median and 95th percentile from sensor to audio.
FPS	Sustained FPS and percentage of frames lost.
Speech Recognition	Word error rate for English, Marathi, Hindi & Code Mixed.
Speech Understanding	Native speaker.
Runtime & Temperature	Runtime and temperature under normal load.
Structured Feedback	Observing structured feedback.

VIII. EXPECTED RESULTS & OBSERVATIONS

More relevant alerts can be provided than by using only the RGB camera. Motion can be tracked and time-of-flight improves short range detection in certain directions (must be verified experimentally). However, transparent/reflective/flat/obscured objects may not be detected well. In busy scenes there may be many false positives.

Alerts in multiple languages may make the device more accessible to users. However, it depends upon good object recognition, good translations, clear speech, and quick reactions. Operation offline means less dependence on internet connections; however, it may increase memory requirements, processing requirements, and power consumption.

More relevant information can be given by combining detections with measured distances.

Robustness can be improved by a combined sensor system, though this increases requirements for calibration, synchronisation and power.

Short priority alerts may be more effective, but this should be tested with users.

Useful only if translations and speech have been verified by native speakers.

Alerts given offline mean that key information can be supplied even if internet connection fails.

Uncertainty will be given. It cannot be guaranteed that a clear path will exist.

IX. CHALLENGES & FUTURE WORK

1. Incorporate information about hazards in pavements/roads (Potholes, Speed bumps, Open drains, Uneven terrain) within local data.
2. Multiple cameras and calibration, to increase awareness (possibly 360).
3. Better detection of edges/risers/dips using depth sensors below.
4. Optimise speech recognition and synthesis for low power offline operation. Include recognition for different regional accents.
5. Option to include haptic feedback for key alerts, but ensure that ambient audio is still possible.
6. Consideration of privacy. Local data storage. Limited video recording.
7. Supervised user testing with informed consent and Institutional Ethics approval.

X. CONCLUSION

The proposed wearable system includes stereo vision, object detection, distance measurement, (optional) time of flight/inertial measurement unit, sensor fusion

and voice assistance. Supporting multiple languages (English/Marathi/Hindi) and processing information locally means that important alerts can be given without internet. Since this project is not yet implemented, the method of design and analysis is detailed. In future work, this prototype will be developed. This will include validation of sensor and language modules, and supervised user testing. This device is intended as a support tool only. It must not be used as a substitute for a white stick/guide dog or professionally trained mobility skills.

REFERENCES

[1] A. D. P. Dos Santos et al., “NavWear: A wearable obstacle detection system for people with visual impairments and blindness,” *Disability and Rehabilitation: Assistive Technology*, 2025, doi: 10.1080/17483107.2025.2477681.

[2] Y. Gao et al., “Cross-modal learning based wearable obstacle avoidance device for visually impaired people,” *Nature Communications*, vol. 16, Art. no. 2857, 2025, doi: 10.1038/s41467-025-58085-x.

[3] Z. Li, F. Han, and K. Zheng, “A wearable device based on RGB-D cameras for visually impaired people: High navigation performance with less social discrimination,” *Electronics*, vol. 14, no. 11, Art. no. 2168, 2025, doi: 10.3390/electronics14112168.

[4] AI4Bharat, “IndicConformer ASR models.” [Online]. Available: https://ai4bharat.iitm.ac.in/areas/model/ASR/IndicConformer/

[5] AI4Bharat, “Indic-TTS models.” [Online]. Available: https://github.com/AI4Bharat/Indic-TTS

[6] AI4Bharat and Hugging Face, “Indic Parler-TTS model card.” [Online]. Available: https://huggingface.co/ai4bharat/indic-parler-tts

[7] STMicroelectronics, “VL53L5CX time-of-flight 8×8 multi-zone distance sensor.” [Online]. Available: https://www.st.com/en/imaging-and-photonics-solutions/vl53l5cx.html

[8] Luxonis, “OAK-D Lite hardware.” [Online]. Available: https://docs.luxonis.com/hardware/products/OAK-D%20Lite

[9] Luxonis, “OAK depth accuracy.” [Online]. Available: https://docs.luxonis.com/hardware/platform/depth/depth-accuracy

[10] World Health Organization, “Blindness and vision impairment,” 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment