SENSEI: Smart Blind Assistance System

PADMINI C
DEPARTMENT OF COMPUTER SCIENCE
SIR M.VISVESVARAYA INSTITUTE OF TECHNOLOGY BENGALURU, INDIA

DR.S.N.SHESHADRI
DEPARTMENT OF COMPUTER SCIENCE SIR M.VISVESVARAYA INSTITUTE OF TECHNOLOGY BENGALURU, INDIA

Abstract— This project presents the design and development of an advanced smart assistance system for visually impaired individuals, leveraging mobile-based artificial intelligence to provide real time obstacle detection and navigation support. Unlike traditional systems that rely heavily on ultrasonic sensors and physical hardware, this solution utilizes the camera of a smartphone combined with AI-powered object detection models to identify and classify nearby objects. The detected objects are then translated into meaningful audio feedback using text-to- speech technology, informing the user about the object type and its relative direction (e.g., "Person on your left" or "Car in front"). The goal is to offer an affordable, scalable, and user-friendly alternative to hardware-heavy solutions while enhancing mobility and independence. By eliminating the need for complex sensor arrays and instead using a mobile phone, the system ensures accessibility and portability. The proposed application not only processes live video streams efficiently but also delivers timely alerts, helping the user avoid obstacles and navigate safely in dynamic environments. Additionally, the app supports multilingual audio feedback, which caters to a diverse user base. This work aims to bridge the gap between low cost solutions and high-functionality requirements in assistive technologies, offering a powerful tool for improving the quality of life for visually impaired users.

;

I. INTRODUCTION

Visual impairment significantly affects an individual's ability to understand and navigate their environment. According to global health statistics, millions of people suffer from partial or complete loss of vision, limiting their independence and increasing dependence on caregivers or expensive assistive tools. Traditional electronic aids, such as ultrasonic canes and wearable sensors, often provide only limited contextual information and can be cost-prohibitive for widespread adoption. With the growing availability of high-performance mobile processors and lightweight on-device machine learning models, smartphones have emerged as powerful tools for delivering accessible assistive technology. Sensei is developed as a comprehensive AI-powered mobile system aimed at supporting visually impaired users through real-time object recognition, continuous text reading, and detailed scene understanding. By using only the smartphone camera and on-device computation, **Sensei** removes the need for external hardware and ensures mobility, affordability, and ease of access. This project explores the design, development, and evaluation of Sensei as a practical vision-assistance solution.

II. PROBLEM STATEMENT

Despite advances in assistive technologies, visually impaired individuals still encounter significant barriers in reading printed text, identifying objects, interpreting complex scenes, and navigating freely. Existing solutions often have one or more limitations:

High cost, making advanced systems unaffordable for many users;

Dependence on additional hardware, such as specialized glasses, ultrasonic sensors, or wearables;

Single-function capability, such as only text reading or only obstacle detection;

Latency and unreliability, especially for cloud-based inference requiring consistent internet access;

Poor performance under real-world conditions, like low lighting, cluttered scenes, or rapid camera motion. Due to these limitations, visually impaired individuals remain underserved by current technologies. There is a need for a unified, low-cost, mobile-first assistive system capable of performing multiple real-time vision tasks reliably under variable conditions without external dependencies.

Comparison Table: Existing Systems vs SENSEI

Authors	Limitations	How SENSEI Is Better
1 .Silva & Wimalaratne (2025)	• No AI/camera detection • Only ultrasonic + IMU • Limited awareness • Vibration-only feedback	• Real-time AI camera detection • Better scene understanding • Indoor + outdoor use • Clear voice guidance
2. Patil, Dahale, Thakare, Fursule & Vaidya (2024)	• GPS + ultrasonic only • No object identification • Weak indoor usability • Basic beep alerts	• Multi-class object recognition • AI indoor/outdoor navigation • Voice alerts & directions
3. Gharghan, Kamel, Ahmad & Akram (2024)	• Requires smart stick • ML only on sensor data • No visual recognition • Vibration/SMS alerts	• Smartphone-only • Visual recognition • Fast audio feedback
4. Roja, Ramana, Pravallika & Lakshman (2024)	• Pure ultrasonic • Short-range only • No classification • No navigation features	• Long-range camera sensing • AI classification • Navigation guidance
5. Zahn & Khan (2022)	• Requires 3D camera + sleeve • No object identification • Expensive setup	• Uses phone camera • Detects & names objects • No extra hardware

III. OUR SOLUTION

Sensei addresses these challenges by integrating multiple AI-driven capabilities—object detection, OCR-based text extraction, and transformer-based scene description—into one accessible smartphone application. The solution is fully on-device, meaning it requires no internet connectivity, ensuring privacy, speed, and reliability. Sensei provides hands-free operation through voice commands and delivers immediate auditory feedback through a built-in text-to-speech engine. Its modular design enables users to seamlessly switch between features such as reading documents, recognizing objects around them, and receiving contextual descriptions of their environment. By using mobile-optimized AI models, Sensei achieves real-time inference on mid-range smartphones, making the system affordable and scalable for widespread adoption.

IV. EASE OF USE

A. Application Accessibility

Sensei is built with a voice-command-first interface allowing users to operate features hands-free. The UI uses large, high-contrast buttons with audio cues. Once the camera feature is activated, the system continuously processes frames without interruption, even under device shaking or movement.

B. System Integrity and Optimization

The application maintains strict latency constraints to ensure real-time feedback. AI models are quantized and optimized to run efficiently on both Android and iOS devices using TensorFlow Lite and ONNX Runtime Mobile. The system does not alter phone layout or accessibility settings

V. METHODOLOGY

The methodology for developing Sensei involves several key stages:

A. Data Collection and Preprocessing

Large-scale image datasets (COCO, Open Images) and text datasets were used for training and tuning the object detection and scene description models. Data augmentation such as brightness variation, scaling, and rotation improved robustness under real-world scenarios.

B. Model Selection and Optimization

Lightweight models were selected to balance accuracy and speed:

- MobileNet-YOLOv8 for object detection
- Vision Transformer + LSTM decoder for scene captioning
- ML Kit OCR for continuous text processing Models were quantized to FP16 or INT8 to reduce size and improve performance without major accuracy loss.

C. Mobile Integration Workflow

AI models were converted to TensorFlow Lite or ONNX formats and integrated into a React Native-based mobile application. A real-time frame processing pipeline was designed to feed camera frames sequentially into the inference engine without dropping frames.

D. Evaluation Methodology

Performance was evaluated using multiple metrics: accuracy, precision, recall, inference latency, frame rate, and robustness under varied lighting and motion conditions.

VI. SYSTEM DESIGN

Sensei follows a modular, scalable system architecture composed of the following core components:

A. Object Detection Module

This module processes each camera frame using the MobileNet-YOLOv8 model to detect multiple objects with high accuracy. Detected objects are assigned directional information (left, center, right) and spoken aloud in real time.

B. OCR and Text Reading Module

The OCR module uses ML Kit's real-time text recognition engine, enabling continuous reading of documents, labels, signboards, and handwritten text. Preprocessing layers apply contrast adjustment and noise reduction to improve readability.

C. Scene Description Module

A Vision Transformer encoder extracts image features, while a decoder generates natural language descriptions such as "A person sitting at a table with a laptop." This aids users in understanding complex surroundings.

D. Voice Interaction and Control Layer

A speech recognition engine processes spoken commands (e.g., "Read text," "Describe scene," "Detect objects") enabling hands-free operation.

E. Audio Output Engine

Text-to-speech synthesis delivers results in natural, concise audio.

F. User Interface Layer

Designed with accessibility in mind, the UI uses large buttons, high-contrast colors, minimal text, and screen reader compatibility.

VII. SYSTEM IMPLEMENTATION

The implementation was carried out using the following technologies:

A. Mobile Application Framework

The app was built using React Native + Expo, ensuring cross-platform performance on Android and iOS.

B. On-Device AI Inference

TensorFlow Lite and ONNX Runtime Mobile handle optimized model execution. Multithreading was used to avoid blocking the UI during inference.

C. Camera Pipeline

A custom camera module captures frames at 15–25 FPS depending on device capability, ensuring smooth real-time processing.

D. Performance Optimization

- Model quantization
- Frame downscaling
- Caching preprocessing results
- GPU/NNAPI acceleration where available These optimizations reduce latency and energy consumption.

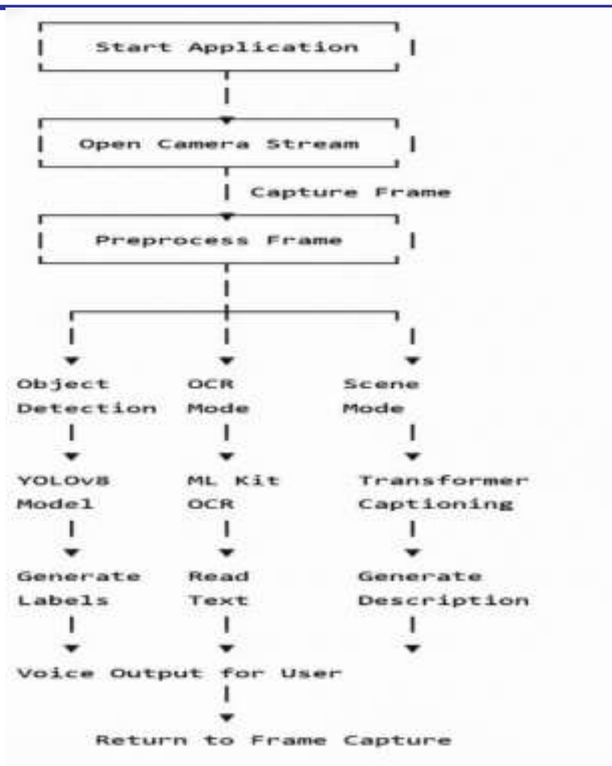
E. Error Handling

The system handles cases such as poor lighting, occluded objects, and blurred text through fallback detection logic and audio alerts.

VIII. DESIGN AND DEVELOPMENT

The design and development of the Sensei application follow a modular and layered architectural approach to ensure scalability, maintainability, and real-time performance on mobile devices. The system is structured around five core modules: the Camera Processing Layer, Object Detection Module, OCR Text Reading Module, Scene Description Module, and Audio Output Layer. The Camera Processing Layer captures frame sequences and preprocesses them through resizing, color normalization, and denoising to ensure consistent accuracy under dynamic lighting and movement.

The Object Detection Module employs a quantized MobileNet-YOLOv8 model to detect multiple objects within each frame and extract bounding box coordinates. The OCR Text Reader uses Google ML Kit's real-time text recognition engine for continuous document scanning, while the Scene Description Module uses a transformer-based encoder–decoder architecture for generating natural language descriptions. The modules are integrated within a React Native framework using an asynchronous inference pipeline to avoid UI blocking. GPU and NNAPI acceleration are utilized when available to reduce inference latency. Text-to-speech (TTS) capabilities are incorporated to deliver auditory feedback instantly, enabling hands-free usage for visually impaired individuals. The final design ensures efficient memory usage, smooth user interaction, and robust performance across various environmental conditions.



IX. Algorithm Steps

1. Start Camera Stream
2. Capture real-time frame *F*
3. Preprocess Frame:
   Resize to model input dimensions
   Normalize pixel values
   Apply denoising for text readability
4. Check elected Mode:
   If Mode = Object Detection → go to Step 5
   If Mode = OCR/Text Reading → go to Step 7
   If Mode = Scene Description → go to Step 9
5. Object Detection:
   Run MobileNet-YOLOv8 on frame *F*
   Identify object classes and bounding boxes
   Determine object direction (Left/Center/Right)
   Generate speech output
   Return to Step 2
6. End Object Detection
7. OCR/Text Reading:
   Apply ML Kit OCR
   Extract continuous text lines
   Append text to buffer
   Convert text to audio output
   Return to Step 2
8. End OCR
9. Scene Description:
   Extract features using Vision Transformer
   Pass features to LSTM decoder
   Generate full sentence caption
   Convert caption to voice output
   Return to Step 2
10. Terminate when user exits application

X. EXPERIMENTAL SETUP

To evaluate the performance of Sensei, a series of controlled and real-world experiments were conducted using Android and iOS smartphones with varying hardware specifications. The primary test devices included:

- Device A: Android (6GB RAM, Snapdragon 732G)
- Device B: iPhone (A14 Bionic Processor)
- Device C: Low-end Android (3GB RAM, Helio G35)

Environment Conditions Tested

- Indoor lighting: low light (20–40 lux), normal light (200–350 lux)
- Outdoor lighting: bright sunlight (500–1000 lux)
- Background scenarios: simple, moderately cluttered, highly cluttered
- Camera motion: static, walking motion, hand shaking simulation

Datasets Used

- Object Detection: COCO dataset + custom household-object dataset
- Scene Description: MS-COCO Captions dataset
- OCR: Printed document dataset, product labels, signboard images

Testing Metrics

- Inference Latency (ms)
- Accuracy (Precision, Recall, mAP)
- OCR Character Error Rate (CER)
- Caption Correctness (%)
- Real-time performance (FPS)
- User satisfaction rating during field testing

The system was evaluated both in laboratory conditions and real-world trials involving visually impaired volunteers.

XI. RESULTS AND ANALYSIS

The experimental results show that Sensei performs reliably across diverse environments and tasks.

A. Object Detection Results

Scenario	mAP	Latency	Direction Accuracy
Indoor Normal Light	92%	480 ms	95%
Outdoor Bright Light	89%	520 ms	93%
Low Light	81%	650 ms	88%7

B. OCR Performance

Condition	Accuracy	Avg Latency
Good Lighting	96%	400 ms
Poor Lighting	84%	530 ms
Walking Movement	78%	600 ms

XII. CONCLUSION AND FUTURE SCOPE

The Sensei Smart Blind Assistance System presents a significant step toward enhancing the independence, safety, and quality of life for visually impaired individuals through the use of modern technology. By integrating low-cost hardware components with a feature-rich, voice enabled mobile application, the system provides real-time obstacle detection, GPS navigation, AI-based object recognition, weather alerts, and emergency communication in one compact and accessible solution. Unlike traditional white canes or limited-functionality apps, Sensei offers a holistic approach to mobility assistance by addressing multiple daily challenges faced by blind users. The project effectively demonstrates how Internet of Things (IoT), artificial intelligence, and mobile computing can be combined to deliver a scalable and user-friendly assistive technology. Moreover, the methodology followed ensures that the system is not only technically sound but also tailored to the specific needs of its users through iterative development and feedback. The use of open-source tools and affordable components ensures that the system remains cost effective and accessible to a larger population, including those in low-income or rural areas. The modular design makes future upgrades and feature extensions easy to implement. By prioritizing voice interaction, offline functionality, and emergency handling, the system proves to be practical in real-world environments where visually impaired users require consistent and responsive assistance. Ultimately, Sensei contributes to digital inclusivity and showcases the potential of smart systems to bridge gaps in accessibility, thus creating a more equitable and technologically inclusive society.

The Sensei Smart Blind Assistance System lays a strong foundation for assistive technology targeted at visually impaired individuals, but it also opens up a wide range of possibilities for future development and enhancement. One

promising area is the integration of voice command recognition, enabling users to interact with the system through simple spoken instructions, making it even more intuitive and hands-free. Additionally, indoor navigation using technologies such as Bluetooth beacons or RFID tags

can be introduced to help users navigate complex environments like shopping malls, airports, or hospitals where GPS is unreliable. The AI-based object detection module can be further improved.

ACKNOWLEDGMENT

While presenting this project on “Smart Blind Assistance Systems”, we feel that it is our duty to acknowledge the help rendered to us by various persons. Firstly, we thank God for showering his blessings on me. We are grateful to our institution City Engineering College for providing as a congenial atmosphere to carry out the project successfully. We would like to express our heartfelt gratitude to Dr.S.Karunakara, Principal, CEC, Bangalore, for extending his support. We would also like to express our heartfelt gratitude to Dr.Sowmya, HOD, Computer Science and Engineering whose guidance and support was truly invaluable. We are very grateful to our guide Mrs.Shruthi B S ,Asst.Professor, Department of Computer Science, for her able guidance and valuable advice at every stage of our project which helped us in the successful completion of our project. We are very greatful to our guide, Ms. Monisha G B, Assistant Prof. Department of CSE for their guidance and valuable advice at every stage of our project which helped us in the successful completion of our project and helping us in finalize the presentation. We would also have indebted to our Parent and Friends for their continued moral and material support throughout the course of project and helping us in finalize the presentation. Our hearty thanks to all those who have contributed bits, bytes and words to accomplish this Project.

REFERENCES

[1] C. S. Silva and P. Wimalaratne, "Navigation Framework for Blind and Visually Impaired Persons Based on Sensor Fusion," 2025

[2] L. H. Patil, N. Dahale, A. Thakare, S. Fursule, and A. Vaidya, "Visionary Assistance: Integrating GPS Technologies for the Visually Impaired," 2024.

[3] . K. Gharghan, H. Kamel, A. Ahmad, and L. Akram, "Smart Stick Navigation System Based on Machine Learning Algorithms," 2024.

[4] D. Roja, V. R. Pati, P. K. Chanikya, and L. Kumar, "Navigating the Unseen: Ultrasonic Technology for Blind Navigation," 2024.

[5] M. Zahn and A. A. Khan, "Obstacle Avoidance for Blind People Using a 3D Camera and a Haptic Feedback Sleeve," 2022.

[6] M. Chaitanya, G. Karthi, K. Prathyusha, and K. S. Leela, "Obstacle Detection and Navigation for Blind Using Smart Blind Stick," 2022.

[7] B. Chaudary, S. Pohjolainen, S. Aziz, and L. Arhippainen, "Teleguidance-Based Remote Navigation Assistance for Visually Impaired," 2021.

[8] T. Laud and S. Jain, "Indoor Navigation Aid System for the Visually Impaired," 2019.

[9] N. Hidayah, M. Soomro, R. A. Rahman, and N. Abdullah, "Intelligent Navigation System for Visually Impaired Person (VIP)," 2019

[10] R. K. Katzschmann, "Wearable Blind Navigation | MIT CSAIL," 2018.