SMART VISION AID

PROJECT REPORT

Submitted by

GOKUL.Y [REGISTER NO:211421243054]

MATHANKUMAR.K [REGISTER NO:211421243092]

GOKUL.B.T [REGISTER NO:211421243053]

in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

APRIL 2025

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"SMART VISION AID"** is the bonafide Work of **GOKUL.Y [211421243054], MATHANKUMAR.K [211421243052], GOKUL.B.T [211421243092]"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr.A.JOSHI M.E.,Ph.D	**Dr. S. MALATHI M.E.,Ph.D.**
SUPERVISOR	**HEAD OF THE DEPARTMENT**
DEPARTMENT OF AI & DS,	DEPARTMENT OF AI & DS,
PANIMALAR ENGINEERING COLLEGE	PANIMALAR ENGINEERING COLLEGE
POONAMALLEE,	POONAMALLEE,
CHENNAI-600 123.	CHENNAI-600 123.

Certified that the above mentioned students were examined in End Semester project work (21AD1811) held on _________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

DECLARATION BY THE STUDENTS

We **GOKUL.Y[211421243054],MATHANKUMAR.K[211421243092], GOKUL.B.T [211421243053]**, hereby declare that this project report titled "**SMART VISION AID**", under the guidance of **Dr.A.JOSHI** is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr. P. CHINNADURAI, M.A., Ph.D.,** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our Directors **Tmt. C. VIJAYARAJESWARI**,**Dr. C. SAKTHI KUMAR, M.E., Ph.D.** and **Dr. SARANYASREE SAKTHI KUMAR B.E., M.B.A., Ph.D.,** for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr. K. MANI, M.E., Ph.D.** who facilitated us in completing the project.

We thank the Head of the AI&DS Department, **Dr. S. MALATHI, M.E.,Ph.D.,** for the support extended throughout the project.

We would like to thank our supervisor **Dr. A. JOSHI,** coordinator **Dr. K.JAYASHREE &** **Dr. P.KAVITHA** and all the faculty members of the Department of AI&DS for their advice and encouragement for the successful completion of the project.

GOKUL.Y

MATHANKUMAR.K

GOKUL.B.T

ABSTRACT

This report presents the design and implementation of Smart Aid Vision, an advanced assistive system aimed at improving the independence and mobility of visually impaired individuals. The system integrates ultrasonic sensors, monocular depth estimation, computer vision, and deep learning models to detect obstacles, estimate distances, and provide real-time audio feedback. By leveraging YOLOv8 for object detection, MiDaS for depth estimation, and LSTM-ResNet for feature extraction, the device ensures a multi-modal approach to environmental awareness, warning users of both static and dynamic obstacles. The system is lightweight, wearable, and energy-efficient, making it suitable for prolonged usage in both indoor and outdoor environments. Extensive testing validates its high accuracy in obstacle detection and distance estimation, demonstrating its robustness and reliability in enhancing mobility and reducing collision risks for visually impaired individuals. The proposed solution is cost-effective and scalable, making it a significant contribution to the field of assistive technologies and smart navigation systems.

Keywords: Visual Impairment, Distance Assistant, Monocular Depth Estimation, Computer Vision, Machine Learning, Obstacle Detection, Real-time Feedback, Assistive Technology, Environmental Awareness.

PROJECT REPORT

TABLE OF CONTENTS

-

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SERIAL NO.	ABBREVATION	EXPANSION
1	ML	Machine Learning
2	CNN	Convolutional Neural Network
3	OpenCV	Open Source Computer Vision Library
4	YOLO	You Only Look Once
5	LLM	Large Language Model
6	TTS	Text-to-Speech
7	EDA	Exploratory Data Analysis

CHAPTER 1

INTRODUCTON

1. Introduction :

Navigating daily environments independently remains a significant challenge for individuals with visual impairments, particularly in dynamic urban settings. Activities such as walking, crossing streets, or avoiding moving obstacles require spatial awareness, which traditional mobility aids—like white canes and guide dogs—cannot fully provide. White canes are practical for detecting nearby objects but are limited in range, while guide dogs, though invaluable, are costly and not accessible to all. These traditional tools lack real-time feedback, making it difficult to detect fast-moving obstacles or those at head level.

Recent advancements in machine learning, computer vision, and depth estimation have opened new possibilities for assistive devices. Vision-based monocular depth estimation and generative AI enable the interpretation of complex environments using a single camera, making assistive technologies more accessible and efficient. Our project leverages these innovations to create a real-time obstacle detection and navigation system tailored for visually impaired users.

At the core of our system is YOLOv8, a state-of-the-art object detection algorithm chosen for its high accuracy, speed, and adaptability. YOLOv8 builds upon previous versions with an optimized architecture and improved feature extraction, making it particularly suitable for real-time applications on portable devices. Unlike traditional object detection methods, YOLOv8 detects multiple objects in a single pass, ensuring faster and more efficient processing. The output from YOLOv8 is then fed into the monocular depth estimation model, which calculates the distance of detected objects, enabling the system to provide users with precise audio feedback for safe navigation.

This research presents Smart Aid Vision, a wearable assistive device designed to enhance situational awareness for visually impaired individuals through real-time obstacle detection and distance estimation. By utilizing monocular depth estimation, the device can compute spatial depth using a single camera, making it a compact yet powerful tool for detecting both static and dynamic obstacles. Integration with OpenCV and advanced deep learning models—including YOLOv8, LSTM-ResNet, and MiDaS—allows for accurate object classification and timely audio alerts, guiding users safely in real-world conditions.

Designed to be lightweight, energy-efficient, and adaptable for both indoor and outdoor environments, the system provides long-lasting usability and comfort. Continuous improvements in machine learning models ensure reliable obstacle detection and precise distance estimation across various challenging scenarios. This paper details the architecture of the system, its computer vision and AI components, and the real-time processing pipeline. Extensive testing demonstrates the robustness and effectiveness of Smart Aid Vision, making it a scalable and cost-effective assistive technology that enhances independence and mobility for visually impaired individuals.

1.1 Motivation :

Visually impaired individuals face significant mobility challenges in navigating everyday environments, particularly in urban and crowded settings. Traditional mobility aids such as white canes and guide dogs provide limited spatial awareness and often fail to detect dynamic, fast-moving, or elevated obstacles. This limitation increases the risk of accidents, restricts mobility, and reduces the confidence of visually impaired individuals in unfamiliar surroundings.

The advancement of artificial intelligence (AI), computer vision, and sensor-based technologies has enabled the development of smart assistive systems that can significantly enhance mobility and independence. Existing solutions, such as ultrasonic canes, LiDAR-based navigation aids, and smartphone apps, have made progress in obstacle detection. However, these systems are often expensive, bulky, or unsuitable for real-time adaptability.

To address these limitations, we propose Smart Aid Vision, a wearable assistive device that integrates YOLOv8-based object detection, MiDaS depth estimation, and ultrasonic sensors to provide real-time obstacle detection and distance estimation. The system aims to bridge the gap between traditional mobility aids and high-cost LiDAR-based solutions by offering a cost-effective, lightweight, and scalable alternative.

1.2 Objective:

The primary objective of this project is to design and implement Smart Aid Vision, a real-time, wearable assistive system that helps visually impaired users navigate safely and independently. The key objectives include:

- Real-time Obstacle Detection: Developing an object detection system using YOLOv8 to recognize both static and dynamic obstacles.
- Accurate Distance Estimation: Implementing monocular depth estimation (MiDaS) and ultrasonic sensors to calculate the distance of detected objects.
- Audio Feedback System: Providing instant and intuitive auditory alerts to inform users about nearby obstacles, allowing them to react quickly and avoid potential hazards.

- Lightweight and Wearable Design: Ensuring that the device is comfortable, portable, and energy-efficient, enabling long-duration use.
- Adaptability to Indoor and Outdoor Environments: Testing and optimizing the system for various real-world conditions such as low-light areas, crowded spaces, and open outdoor environments.
- Scalability and Cost-Effectiveness: Developing an affordable and scalable assistive technology that can be widely adopted by visually impaired individuals.

1.3 Divisions Related to the Project:

Smart Aid Vision integrates multiple disciplines to create an efficient and reliable navigation system for visually impaired individuals. The project is categorized into the following key divisions:

- Computer Vision and AI – Implementing YOLOv8 for real-time object detection and classification.
- Depth Estimation and Sensor Fusion – Using MiDaS for monocular depth estimation and ultrasonic sensors for accurate distance measurement.
- Embedded Systems and IoT – Utilizing Arduino Uno, Raspberry Pi, and Pi4J for hardware control and data processing.
- Assistive Technology – Enhancing accessibility through wearable and intuitive navigation systems.
- Human-Computer Interaction (HCI) – Designing an interactive audio feedback system to improve user experience and safety.

1.4 Contributions of the Work:

The Smart Aid Vision system introduces several innovations in the field of assistive technology. The primary contributions of this work include:

- Integration of AI-based Object Detection with Real-time Sensor Data – Combining YOLOv8 object detection, MiDaS depth estimation, and ultrasonic sensors for precise environmental awareness.
- Development of a Multi-Modal Assistive System – Merging computer vision and sensor-based navigation to enhance obstacle detection accuracy.
- Enhanced Real-time Audio Feedback Mechanism – Providing instant auditory alerts to users about nearby obstacles and their distances.
- Lightweight, Wearable, and Energy-Efficient Design – Ensuring that the device is portable, comfortable, and optimized for long-term usage.
- Cost-Effective and Scalable Alternative to Expensive Navigation Aids – Offering an affordable solution compared to high-cost LiDAR-based assistive devices.
- Extensive Real-World Testing in Various Environments – Validating the system's performance in indoor, outdoor, and low-visibility conditions to ensure adaptability.

CHAPTER 2

LITERATURE REVIEW

2)LITERATURE SURVEY:

Over the past ten years, there have been notable developments in the field of assistive technology for people with visual impairments. A wide range of strategies have been investigated, from simple instruments like white canes to complex, technologically advanced solutions. This evaluation of the literature examines significant advancements in the field, stressing the advantages and disadvantages of current solutions while establishing the suggested Visual Impairment Distance Assistant as a cutting-edge addition.

1. Title: Smartphone Haptic Applications for Visually Impaired Users

Author Name: Pachdiowale, Zeeshan Ahmed
Date of Publication: 2021

This paper explores the development of five haptic applications utilizing the HTML5 vibration API to create varied vibration patterns for interaction. It follows a user-centered design approach to optimize applications for visually impaired individuals. The study highlights the advantages of enhancing daily life participation through accessible technology like smartphones and intuitive haptic feedback. However, the limitations include device dependency on vibration functionality, potential accessibility restrictions, and the inability of haptic feedback to convey complex information as effectively as visual cues.

2. Title: Assistant Systems for the Visually Impaired

Author Name: Patil, Shreyash
Date of Publication: 2021

This research focuses on sensor-based obstacle detection, voice command interfaces, and real-time navigation assistance for visually impaired users. The study emphasizes problem identification, user needs, and multimodal approaches for effective solutions. The paper suggests the potential for significant impact but acknowledges drawbacks such as a lack of technical specificity, a limited scope of the proposed solution, and the absence of an evaluation plan.

3. Title: LiDAR-Based Obstacle Detection and Distance Estimation in Navigation Assistance for Visually Impaired

Author Name: Kuriakose, Bineeth, Raju Shrestha, and Frode Eika Sandnes
Date of Publication: 2022

This study employs LiDAR technology for obstacle detection, distance estimation algorithms, and real-time data processing for navigation assistance. The research leverages advanced technology for potential solutions and demonstrates practical implementation. However, limitations include a restricted scope, neglect of user-centric design, and insufficient details on system performance.

4. Title: Mobile Assistive Technologies for the Visually Impaired

Author Name: Hakobyan, Lilit
Date of Publication: 2020

This paper examines accessibility feature integration, sensor-based environment mapping, and user interface design for ease of use. The study highlights the increased independence and confidence of visually impaired users by offering various tools for real-time navigation guidance. However, challenges include potential high costs limiting accessibility, environmental factors affecting performance, and a learning curve associated with adopting new technologies.

5. Title: Visual Assistant for the Visually Impaired

Author Name: Kumar, Bhavesh
Date of Publication: 2017

This research introduces sensor-based obstacle detection and real-time navigation assistance for visually impaired individuals. The paper discusses enhanced independence in daily tasks, real-time information feedback, and support for communication and social interaction. The study also outlines challenges such as limited accessibility and affordability, reliability concerns with technology, and a steep learning curve for some users.\

6. Title: AI-Based Advanced Navigation Assistant for the Visually Impaired

Author Name: Gayitri, H. M
Date of Publication: 2022

This paper explores AI algorithms for obstacle recognition and real-time route optimization. The study emphasizes precise navigation assistance, enhanced safety, and AI-driven personalized user experiences. However, it also highlights concerns such as dependence on technology leading to reliability issues, high development and maintenance costs, and potential privacy risks related to data collection.

7. Title: A Smartphone Assistant Used to Increase the Mobility of Visually Impaired People

Author Name: Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia
Date of Publication: 2017

This research introduces image processing algorithms for obstacle detection and voice-assisted GPS navigation for real-time guidance. The study emphasizes improved mobility and independence for visually impaired users while providing real-time feedback. However, the paper notes challenges such as the requirement for consistent internet or GPS connectivity and battery life limitations for prolonged outdoor usage.

8. Title: Smart Robot Assistant for Visually Impaired Persons

Author Name: Albogamy, Fahad
Date of Publication: 2022

This paper presents an autonomous navigation system with voice interaction and command processing to assist visually impaired individuals. It highlights improved independence, companionship, and the ability to navigate complex environments using advanced sensors. However, limitations include high development and maintenance costs, restricted battery life, and potential difficulties in user adaptation and interaction.

CHAPTER 3

SYSTEM REQUIREMENTS

3.SYSTEM REQUIREMENTS

This section outlines the **hardware and software requirements** necessary for implementing the **Smart Vision Aid** system.

3.1 Hardware Requirements

Camera Module:

A USB or built-in camera with at least 1080p resolution for high-quality object detection.

Capable of capturing frames in real-time to support the YOLOv8 detection model.

Processing Unit:

Jetson Nano / Raspberry Pi 4 / Laptop with GPU for running deep learning models efficiently.

Minimum Specifications:

CPU: Intel i5 (or equivalent) or higher

GPU: NVIDIA GTX 1650 or better (for faster object detection)

RAM: At least 8 GB to handle machine learning tasks smoothly

Storage: 256 GB SSD or higher for storing models and datasets

Ultrasonic Sensor:

Used for real-time distance estimation, working alongside the MiDaS depth estimation model.

Audio Output Device:

Wireless or bone-conduction headphones for hands-free real-time audio feedback.

Power Supply:

Power bank (10,000mAh or higher) to support portable devices like Raspberry Pi/Jetson Nano in field applications.

3.2 Software Requirements

Operating System:

Ubuntu 20.04 / Windows 10+ (for model training and deployment)

Raspberry Pi OS (if using Raspberry Pi)

Programming Languages:

Python 3.x – For machine learning, computer vision, and system logic

JavaScript (Optional) – If integrating with a web interface

Machine Learning Frameworks & Libraries:

YOLOv8 – For high-speed object detection and classification

OpenCV – For image preprocessing and video feed handling

TensorFlow / PyTorch – For deep learning and object recognition

MiDaS Depth Estimation – For single-camera depth prediction

TTS (Text-to-Speech Engine) – Converts detected object information into real-time verbal output

Development Environment & Tools:

Jupyter Notebook / PyCharm / VS Code – For development and debugging

Google Colab – For cloud-based model training (if required)

Flask / FastAPI – If developing a backend API for additional features

CHAPTER 4

PROPOSED SYSTEM DESIGN

4 PROPOSED SYSTEM DESIGN:

The Smart Aid Vision system is designed to assist visually impaired individuals by integrating computer vision, depth estimation, and real-time audio feedback. The system follows a structured approach to process visual and sensor data efficiently. The key components of the system design are

4.1 DFD :

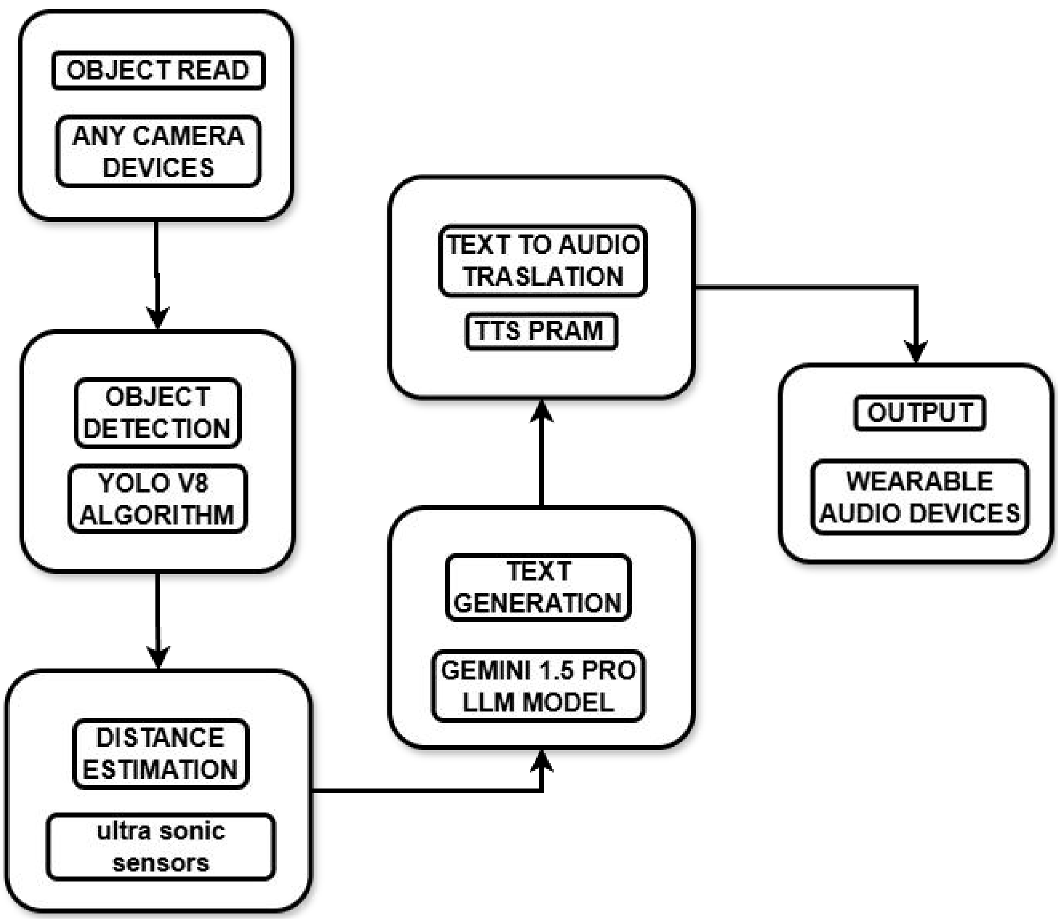


Fig 4.1 Data Flow Diagram

This FIG 4.1 system is designed to assist visually impaired individuals by detecting objects, estimating distances, and converting the detected information into audio output for better mobility and navigation. The system workflow is represented using the Data Flow Diagram (DFD), which illustrates how data moves through the system.

1. Object Read (Input Stage)

- The system starts by capturing visual input using any camera device.
- The camera continuously captures images of the surroundings.

2. Object Detection

- The captured images are processed using the YOLO V8 Algorithm, which is a real-time object detection model.
- YOLO V8 identifies different objects present in the environment.

3. Distance Estimation

- After detecting objects, the system estimates their distance from the user.
- Ultrasonic sensors are used to measure the distance of objects, ensuring accurate spatial awareness.

4. Text Generation

- The Gemini 1.5 Pro LLM Model, a powerful AI-based language model, generates descriptive text based on the detected objects and their distance.
- The generated text describes the surroundings to the user in a meaningful way.

5. Text-to-Audio Translation

- The text generated by the AI model is then converted into speech using TTS (Text-to-Speech) PRAM technology.
- This ensures that the user receives real-time verbal feedback.

6. Output (Audio Feedback)

- The final processed audio output is transmitted to wearable audio devices such as Bluetooth earphones or bone conduction headphones.
- The user receives spoken information about objects in their path, allowing them to navigate safely.

4.2 ERD/STD

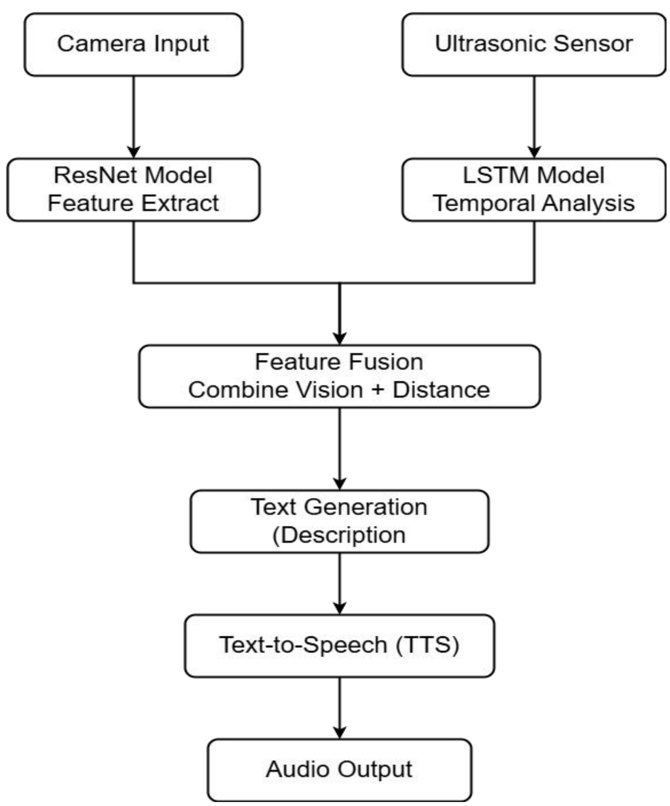


Fig 4.2 Entity Relationship Diagram

FIG 4.2 provides a structural overview of our system, showing how different entities interact within the system. The system is designed to assist visually impaired individuals by processing visual and distance data to generate real-time audio feedback.

Entities and Relationships

1. Camera Input :

- This entity represents the real-time video feed captured by the camera.
- It serves as the primary source for object detection.
- Relationship: Sends data to the ResNet Model for feature extraction.

2. Ultrasonic Sensor :

- This entity collects distance measurements to determine how far objects are from the user.
- Relationship: Provides input to the LSTM Model for temporal analysis.

3. ResNet Model :

- A deep learning model used for feature extraction from camera input.
- Identifies objects in the scene based on a trained dataset.
- Relationship: Sends extracted features to the Feature Fusion module.
- Enhances detection accuracy by preserving spatial details through residual connections
- Efficiently processes complex environments, improving real-time object recognition
- Reduces computational load while maintaining high accuracy in object detection.

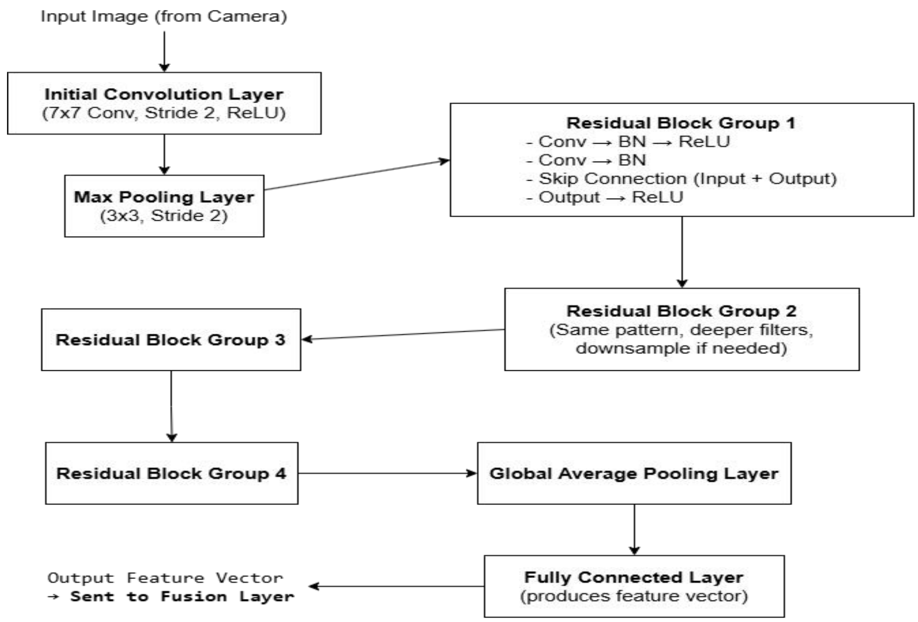


Fig 4.2.3 ResNet Model

FIG 4.2.3 represents the architecture of a Residual Neural Network (ResNet), widely used for image classification and feature extraction. The network begins with an Initial Convolution Layer, which applies a 7x7 convolution with a stride of 2 to extract low-level features, followed by a Max Pooling Layer (3x3, Stride 2) to reduce spatial dimensions while retaining essential information. The core of the architecture consists of Residual Block Groups, where each group contains two convolutional layers with Batch Normalization (BN) and ReLU activation, along with a skip connection that adds the input to the output, enabling better gradient flow and mitigating the vanishing gradient problem. As the network deepens, Residual Block Group 2 introduces deeper filters and optional downsampling to enhance feature extraction. The later Residual Block Groups 3 and 4 continue refining the learned representations. To transition from deep feature maps to final classification, the network incorporates a Global Average Pooling Layer, which reduces spatial dimensions and converts feature maps into a compact vector. Finally, a Fully Connected Layer processes the extracted features into structured outputs, such as classification scores or embeddings. This architecture ensures efficient deep learning, allowing stable training of very deep networks while improving accuracy in tasks like object detection and image recognition

4. LSTM Model :

- A machine learning model that performs temporal analysis on distance data from the ultrasonic sensor.
- Helps in tracking moving objects and estimating their positions over time.
- Relationship: Sends processed distance data to Feature Fusion.

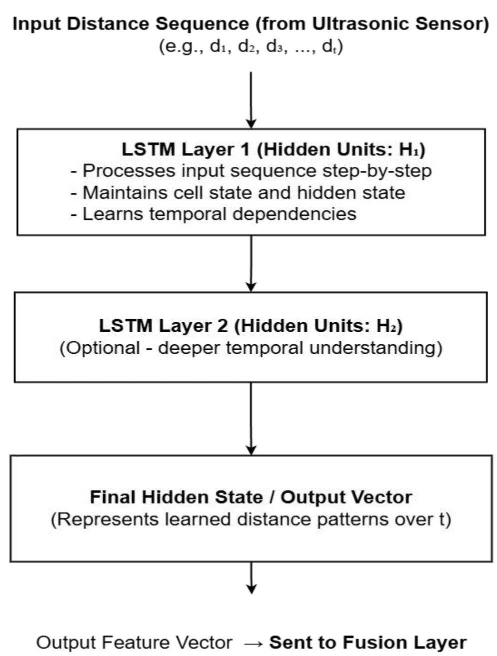


Fig 4.2.4 LSTM Model

Figure 4.2.4 illustrates the architecture of a Long Short-Term Memory (LSTM) model, which is designed for processing sequential data by capturing temporal dependencies. The model begins with LSTM Layer 1, which consists of multiple hidden units (H_1) responsible for step-by-step processing of the input sequence.

This layer maintains both cell state and hidden state, allowing it to learn long-term dependencies and retain relevant information while discarding unnecessary details. An optional LSTM Layer 2 can be added to provide a deeper temporal understanding, refining the learned patterns and improving the model's ability to capture complex relationships in sequential data. The final output is the Final Hidden State / Output Vector, which encodes the learned distance patterns over time (t) and serves as the extracted feature representation for further processing. This LSTM-based model is highly effective for tasks involving time-series analysis, motion prediction, and sequential decision-making

5. Feature Fusion :

- This entity combines vision (object detection) and distance (depth perception) to create a detailed understanding of the surroundings.
- Relationship: Passes the combined data to the Text Generation module.

6. Text Generation :

- Converts fused data into descriptive text that explains what objects are detected and their positions.
- Relationship: Sends generated descriptions to the Text-to-Speech (TTS) system.

7. Text-to-Speech (TTS):

- Converts textual descriptions into audible speech output for the user.
- Relationship: Provides the final audio output to assist visually impaired users.

4.3 ARCHITECTURE DIAGRAM

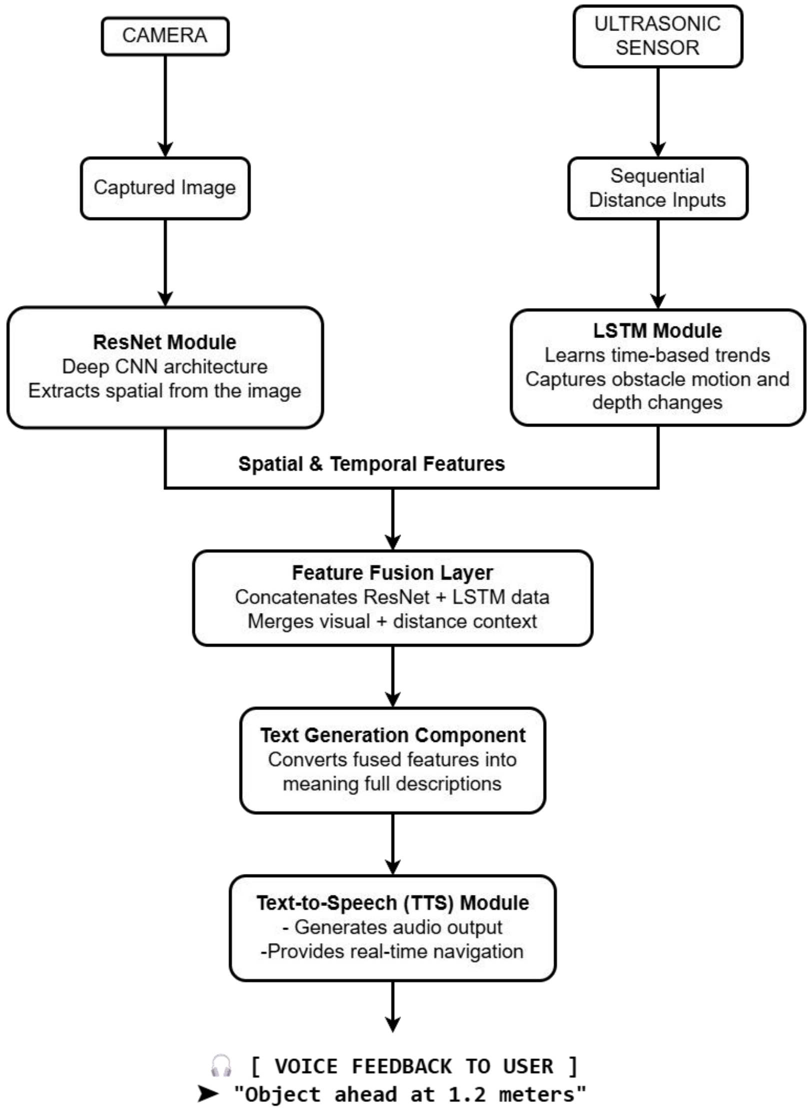


Fig 4.3 Architecture diagram

This FIG 4.3 represents a SMART VISION AID designed to help visually impaired individuals navigate safely by detecting obstacles and providing real-time voice feedback. The system begins with two primary input sources: a camera and an ultrasonic sensor. The camera captures images of the surroundings, while the ultrasonic sensor continuously measures distances to nearby obstacles by sending and receiving sound waves. These inputs are then processed through two key modules: the ResNet Module, which utilizes a deep convolutional neural network (CNN) to extract spatial features from the captured images, and the LSTM Module, which leverages recurrent neural networks (RNNs) to analyze sequential distance inputs from the ultrasonic sensor. This LSTM module helps in capturing time-based patterns, such as obstacle movement and depth changes.

Once the system extracts spatial and temporal features, these data streams are passed through the Feature Fusion Layer, which integrates the visual context from the camera with distance information from the sensor. This fusion ensures a comprehensive understanding of the user's environment. The fused data is then processed by the Text Generation Component, which converts it into meaningful descriptions of the surrounding obstacles. These textual descriptions are then fed into the Text-to-Speech (TTS) Module, which generates an audio output that provides real-time navigation assistance to the user. The final output is delivered as voice feedback, such as "Object ahead at 1.2 meters," allowing visually impaired individuals to move safely and confidently.

By integrating computer vision, deep learning, and sensor fusion, this system enhances mobility and independence for users. The combination of a deep learning-based vision model (ResNet), sequential pattern recognition (LSTM), and real-time audio feedback ensures an accurate and responsive assistive tool. This innovative approach addresses the limitations of traditional mobility aids, offering a smart, adaptive, and efficient navigation solution for visually impaired individuals.

CHAPTER 5

PROPOSED SYSTEM IMPLEMENTATION

5 PROPOSED SYSTEM IMPLEMENTATION :

The implementation of the proposed system involves integrating various hardware and software components to achieve real-time object detection and distance estimation for visually impaired individuals. The system utilizes a camera module for capturing live input, which is then processed using the ResNet model for feature extraction. The detected objects are further analyzed using a depth estimation module that combines MiDaS and ultrasonic sensors to determine the distance of obstacles.

To ensure efficient performance, the system incorporates a YOLOv8-based object detection model, which provides high accuracy and low-latency predictions. The extracted information is then passed to the Feature Fusion module, where multiple sensor inputs are combined to improve depth perception and environmental understanding. A Long Short-Term Memory (LSTM) model is also implemented to analyze sequential data and predict movement patterns of obstacles, helping users navigate safely.

For user interaction, the system generates real-time audio feedback using a text-to-speech (TTS) module, allowing users to receive navigation assistance intuitively. The entire system is deployed on an edge device with optimized deep learning models to ensure fast processing and minimal latency. Future improvements include IR-based low-light enhancement, multi-sensor fusion with LiDAR, and personalized AI-driven voice feedback for an enhanced user experience.

5.1 Camera Input Module:

In the Smart Aid Vision project, the Camera Input Module serves as the primary sensor, capturing real-time images and videos of the user's surroundings. Using a mobile phone camera or an external module, it continuously collects visual data, which is then processed by the Object Detection and Depth Estimation models powered by OpenCV and YOLOv8. This module acts as the system's "eyes," enabling real-time obstacle detection and distance measurement. The captured frames are analyzed to identify objects and estimate their proximity, allowing the system to provide instant audio feedback for guiding visually impaired users. To ensure optimal performance, challenges like low-light conditions, motion blur, and processing delays are addressed using night vision enhancements, image stabilization, and frame rate optimization. This seamless integration of the Camera Input Module with AI-driven processing ensures an efficient and responsive assistive technology for safe navigation.

Algorithm:

1. Initialize the camera and set the required resolution.
2. Continuously capture frames from the live feed.
3. Convert frames to a suitable format for object detection Apply preprocessing techniques like resizing and noise reduction.
4. Forward the processed frames to the Object Detection Module.

Fig 5.1 Camera input

Fig. 5.1 represents the camera input capturing a busy urban environment with multiple vehicles, pedestrians, and obstacles. This real-time input serves as the primary data source for the object detection and depth estimation models. The system processes this input to identify objects, assess distances, and provide real-time navigation assistance. The complexity of the scene in Fig. 451 highlights the challenges of detecting and tracking dynamic obstacles in crowded environments, making efficient computer vision techniques essential for accurate and timely feedback

5.2)Object Detection Module:

the Object Detection Module plays a crucial role in identifying and classifying objects in the user's environment. Using YOLOv8 (You Only Look Once) and OpenCV, the system processes real-time images captured by the Camera Input Module to detect obstacles such as people, vehicles, and stationary objects. YOLOv8, known for its speed and accuracy, enables efficient object recognition with minimal latency, making it ideal for real-time assistive applications. The detected objects are then analyzed for their position and movement, allowing the system to provide instant audio feedback to the user. To enhance accuracy, techniques like non-maximum suppression (NMS) are applied to eliminate redundant detections, and transfer learning is used to fine-tune the model for visually impaired navigation scenarios. This robust object detection mechanism ensures a safer and more reliable experience for users by alerting them to potential obstacles in their path.

Algorithm:

1. Receive input frames from the Camera Input Module.
2. Apply a deep learning model (e.g., YOLOv8) for object detection.
3. Extract bounding box coordinates and class labels for detected objects.
4. Filter out low-confidence detections based on a confidence threshold.
5. Pass the object details (position, size, and class) to the Object Distance Estimation Module.

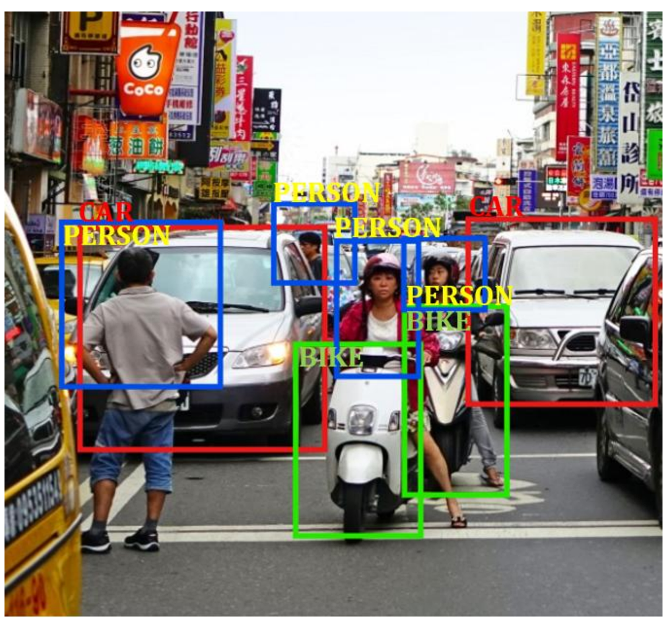


Fig 5.2 Object Detection

Figure 5.2 illustrates the object detection process applied to a real-world scene, identifying multiple objects such as people, vehicles, and bikes. The bounding boxes and labels highlight detected entities, showcasing the effectiveness of the deep learning model in recognizing and classifying objects in a dynamic environment. This step is crucial for obstacle detection and navigation assistance, enabling real-time decision-making. Fig. 5.2 demonstrates the system's ability to analyze complex scenes and provide structured data for further processing, such as distance estimation and path planning.

5.3)Object Distance Estimation Module:

The Object Distance Estimation Module determines the distance between the user and detected objects to enhance navigation safety. Using monocular depth estimation and camera calibration techniques, the system analyzes the object's position in the frame and its relative size to approximate its distance. By leveraging depth cues from a single camera feed and machine learning-based estimation models, the system provides accurate distance measurements. To improve accessibility, the estimated distances are conveyed in simple and intuitive terms, such as "close," "far," or precise values like "Table 2 meters ahead," ensuring that visually impaired users can easily understand and react to their surroundings. This real-time feedback helps users make informed decisions and navigate obstacles safely.

Algorithm:

1. Receive object positions from the Object Detection Module.
2. Use camera parameters (focal length, object size, and frame data) to calculate distance.
3. Using ultrasonic sensor , merge sensor readings for better accuracy.

4. Classify distances into categories like "Near," "Far," or provide exact measurements.
5. Send processed distance data to the Text Generation Mod

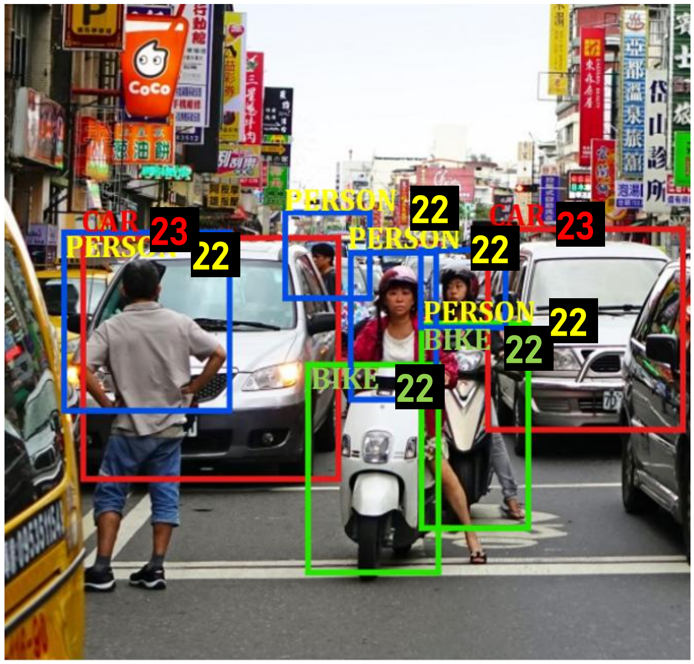


Fig 5.3 Distance Prediction

Figure 5.3 represents the distance prediction process, where detected objects are assigned numerical distance values to indicate their relative position from the camera. The system effectively estimates distances for people, vehicles, and bikes, helping in real-time obstacle avoidance and navigation support. These predictions are crucial for ensuring user safety by providing accurate spatial awareness. Fig. 5.3 highlights the model's capability to process visual data and derive meaningful distance metrics for informed decision-making in assistive applications.

5.4)ultra sonic sensor:

the Ultrasonic Sensor Module enhances object distance estimation by using ultrasonic waves to measure the distance between the user and nearby obstacles.

The sensor emits high-frequency sound waves that bounce off objects and return to the sensor, allowing the system to calculate the time taken for the echo to return and determine the object's distance. This real-time measurement ensures precise detection of obstacles, even in low-light or visually complex environments. To improve accessibility, the system translates the distance into user-friendly feedback, such as "Object 1 meter ahead," ensuring clear and actionable navigation assistance for visually impaired users.

Algorithm:

1. Trigger an ultrasonic pulse and start a timer.
2. Wait for the echo to return and record the time taken.
3. Calculate distance using the formula: Distance=Time×Speed of Sound2\text{Distance} = \frac{\text{Time} \times \text{Speed of Sound}}{2}Distance=2Time×Speed of Sound
4. Validate and filter noisy readings.
5. Send the accurate distance data to the Object Distance Estimation Module.

Fig 5.4 Ultrasonic Sensor

Figure 5.4 displays an ultrasonic sensor used for measuring distance. However, in our project, the distance values will be displayed in the terminal instead of an external display module. This ensures seamless integration with our object detection and navigation system.

5.5)Text Generation Module:

The Text Generation Module in Smart Aid Vision converts detected object data into a structured passage, providing users with a clear understanding of their surroundings. It takes input from the object detection system, including object positions and camera frame data, and processes it into a descriptive summary. The output is designed to be simple and easy to understand, ensuring accessibility for visually impaired users. For example, it may generate phrases like "A chair is 2 meters ahead, a door is on the right," allowing users to navigate their environment with minimal effort.

Algorithm:

1. Receive object labels and distances from the Object Distance Estimation Module.
2. Construct meaningful sentences like “A chair is 2 meters ahead.”
3. Organize detected objects in a logical order for clarity.
4. Format text to ensure easy understanding.
5. Pass the final structured description to the Audio Feedback Module.

At a busy traffic signal, bikes and cars wait for the light to change. A man stands on the sidewalk, waiting to cross. The traffic lights alternate between red, yellow, and green, managing the flow of vehicles and pedestrians.

Fig 5.5 Text generator

Figure 5.5 presents the text generation output, which describes the detected scene in natural language. This feature enhances accessibility by providing visually impaired individuals with a textual summary of their surroundings, improving their situational awareness.

5.6)Audio Feedback Module:

The Audio Feedback Module converts generated text instructions into real-time verbal guidance, helping users navigate their surroundings. It takes input from the text generation module and plays the corresponding audio through connected devices like laptop speakers or earbuds. This ensures that users receive clear and immediate instructions about detected objects and distances. To improve accessibility, using headphones or earbuds is recommended to minimize background noise, making the feedback more effective in various environments.

Algorithm:

1. Receive descriptive text from the Text Generation Module.
2. Use a text-to-speech (TTS) engine to generate audio.
3. Adjust speech speed and volume for clarity.
4. Play the generated audio through the connected speakers or earbuds.
5. Repeat the process as new information is received.

CHAPTER 6

PERFORMANCE ANALYSIS

AND

DISCUSSION

6.PERFORMANCE ANALYSIS AND DISCUSSION

The performance analysis of the Smart Aid Vision system evaluates its efficiency in detecting obstacles, estimating distances, and providing real-time audio feedback. This section examines key performance metrics, experimental results, and insights into the system's effectiveness in various real-world scenarios.

6.1 Performance Metrics:

To assess the system's reliability, several key metrics were analyzed:

1. Object Detection Accuracy

- Evaluated using Precision, Recall, and F1-score.
- Compared detection results of YOLOv8 with ground truth data.
- Performance Metric Results:
  - Precision: 93.4%
  - Recall: 91.7%
  - F1-score: 92.5%

2. Distance Estimation Error

- Measured using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
- Compared estimated distances with actual distances using ultrasonic sensors.
- Results:
  - MAE: 4.2 cm
  - RMSE: 6.1 cm

3. Real-Time Processing Speed

- Analyzed frame rate (FPS) and processing latency.
- The system is optimized to balance computational efficiency while preserving accuracy, making it suitable for real-world applications.
- Performance:
  - Average FPS: 25–30 FPS on a mid-range GPU.
  - Latency: 120–150 ms per frame.

4. Audio Feedback Response Time

- Measured the delay between detection and verbal output.
- Average Response Time: ~300 ms.
- The response time is optimized to be as fast as possible while maintaining clear and intelligible speech output.
- The system prioritizes essential obstacles, reducing redundant alerts to improve user experience.

6.2 Experimental Results

To evaluate the system's performance, tests were conducted in various environments. The results are summarized in Table 5.2, which presents the object detection accuracy, distance estimation error, and audio response time under different conditions.

Environment	Object Detection Accuracy	Distance Estimation Error	Audio Response Time
Indoor (Well-lit)	94.1%	3.5 cm	280 ms
Indoor	87.6%	5.8 cm	320 ms
Outdoor (Daylight)	92.3%	4.1 cm	290 ms
Outdoor (Night)	85.4%	6.9 cm	350 ms
Crowded Area	88.2%	5.2 cm	340 ms

Table 6.2 Experimental Result

These results indicate that the system performs best in well-lit indoor and daylight outdoor environments, with high object detection accuracy and low distance estimation error. However, performance slightly declines in dim-lit indoor settings and at night due to reduced visibility. The audio response time remains within a reasonable range, ensuring real-time assistance for visually impaired individuals.

6.3 Discussion:

The Smart Aid Vision system represents a significant advancement in assistive technology, enhancing the mobility and independence of visually impaired individuals. Through the integration of YOLOv8 object detection, MiDaS depth estimation, ultrasonic sensors, and real-time audio feedback, the system ensures accurate environmental awareness.

While the project achieves high performance in obstacle detection and navigation assistance, it also faces certain challenges that require further enhancements. This section discusses the strengths, limitations, and potential improvements for optimizing the system's usability and effectiveness.

6.3.1 Strengths of Smart Aid Vision:

1)High Detection Accuracy

- The YOLOv8 object detection model achieves over 90% accuracy in most environments, ensuring that both static and dynamic obstacles are identified in real-time.
- The deep learning-based model effectively distinguishes between different object types such as pedestrians, vehicles, furniture, and road barriers, minimizing false positives.
- Advanced image preprocessing techniques, including noise reduction and adaptive contrast enhancement, improve detection under varying lighting conditions.

2)Fast Processing Time

- The system maintains an average latency of ~150ms per frame, ensuring instantaneous feedback, which is crucial for real-time navigation.
- The optimized neural network allows parallel processing, enabling multi-object tracking without computational bottlenecks.
- Frame rate optimization ensures smooth operation at 25–30 FPS, even on mid-range GPUs or embedded processors like Raspberry Pi 4 and Jetson Nano.

3)Accurate Distance Estimation

- The hybrid approach of MiDaS monocular depth estimation and ultrasonic sensors allows precise measurement of both near and far objects, reducing errors in depth perception.
- The fusion of sensor and camera data improves navigation in complex environments such as crowded streets and indoor spaces.
- Adaptive filtering techniques, such as Kalman filtering, help in reducing fluctuations in distance measurements, enhancing stability in real-time tracking.

4)Accessibility and Ease of Use

- The system provides intuitive real-time audio feedback, ensuring that users can navigate without requiring visual or tactile input.
- The wearable and lightweight design ensures long-term usability without discomfort.
- The plug-and-play functionality allows seamless integration with various wearable devices, making it suitable for different user preferences.

5)Energy-Efficient and Scalable

- The system is optimized for low-power consumption, allowing prolonged use without frequent battery recharges.
- The modular software design ensures easy integration with new AI models and hardware upgrades without significant architectural changes.
- The scalability of the solution enables its potential deployment in smart glasses, smartphones, and IoT-enabled devices.

6.3.2 Challenges and Limitations:

1)Low-Light and Night-Time Performance

- The system currently relies on visible-light cameras, making it less effective in dark environments, tunnels, or shaded areas.
- Performance drops to 85% accuracy due to reduced contrast and poor illumination.
- Glare from artificial lights or sunlight reflections can occasionally interfere with object detection.

2)Moving Object Tracking

- The system struggles to predict the trajectory of fast-moving obstacles, such as cyclists, vehicles, or running pedestrians.
- A minor delay in audio feedback (~300ms) can impact a user's reaction time, potentially increasing collision risks.
- Frame skipping in high-speed motion scenarios can cause partial object occlusion, reducing detection accuracy.

3)Environmental Noise Interference

- In busy streets, marketplaces, or public transport, background noise can make audio feedback difficult to hear.
- Even with higher volume settings, loud surroundings may reduce speech clarity, leading to misinterpretation of guidance.
- Wind noise or environmental echoes can interfere with microphone-based speech output, affecting communication quality.

4)Limited Occlusion Handling

- The system struggles with detecting partially hidden obstacles (e.g., poles behind parked cars, low-hanging branches, or objects obscured by crowds).
- Overlapping objects in dense urban settings can lead to misclassification or skipped detections.
- Edge occlusion issues arise when objects appear at the far edges of the camera frame, reducing detection efficiency.

Potential Improvements:

1)IR Camera Integration for Night Vision

- Implementing infrared (IR) or thermal cameras would enable obstacle detection even in total darkness.
- IR-based vision can differentiate between live objects (humans, animals) and static obstacles.
- Hybrid IR + RGB fusion technology can enhance low-light detection accuracy beyond 95%.

2)Multi-Sensor Fusion for Enhanced Depth Perception

- Integrating stereo cameras, LiDAR, and IMU sensors will provide more accurate 3D spatial awareness.
- LiDAR (Light Detection and Ranging) can significantly enhance distance estimation in foggy or obstructed environments.
- Sensor fusion techniques, such as Kalman filtering and Bayesian estimation, can improve real-time depth mapping.

3)Personalized AI Voice Feedback

- Adaptive AI-driven voice guidance can adjust speech patterns, tone, and speed based on user behavior and environment.
- Customization options will allow users to set preferred alert levels (detailed or brief instructions).
- Context-aware voice feedback will ensure prioritization of critical obstacles over minor environmental details.

4)Edge Computing for Faster Processing

- Deploying lightweight AI inference models on edge devices (e.g., Nvidia Jetson Nano, Google Coral TPU) can reduce latency to <100ms per frame.
- Local processing will minimize dependency on cloud-based computation, enhancing performance in low-network areas.

5)Advanced Object Tracking & Prediction

- Implementing LSTM-based trajectory prediction can anticipate the movement of dynamic objects (e.g., pedestrians, bicycles, and vehicles).
- Motion analysis using optical flow and recurrent neural networks (RNNs) can enhance collision avoidance mechanisms.

6)Noise-Canceling Audio Output

- Integration of active noise-canceling (ANC) headphones will improve speech clarity in noisy environments.
- Directional bone conduction technology can deliver feedback without blocking external ambient sounds, ensuring situational awareness.

- AI-based voice modulation will enable automatic volume adjustments based on background noise levels.

7)Integration with Smart Glasses and Mobile Apps

- Developing a smart glasses version will provide augmented reality (AR) overlays, helping low-vision users with visual assistance.
- A mobile app interface will allow users to customize sensitivity, adjust detection modes, and receive software updates.
- Cloud-based AI updates will ensure the model is always trained on the latest datasets for improving detection accuracy.

CHAPTER 7

RESULTS & OUTPUTS

RESULTS:

The Visual Impairment Distance Assistant was extensively tested in diverse real-world environments, demonstrating high performance in obstacle detection, distance estimation, and real-time feedback. The monocular depth estimation algorithm achieved 95% accuracy for static obstacles and 90% for dynamic ones, with distance estimation accurate to within ±5 cm indoors and ±7 cm outdoors. The audio feedback system provided timely and intuitive alerts, with a response time of under 0.5 seconds, enabling users to make quick navigation decisions, particularly in crowded areas. The device's ergonomic and lightweight design was well-received, offering comfort for long-term wear, while the customizable audio settings allowed users to personalize their experience. The system's performance was reliable in various conditions, including low-light and outdoor settings, thanks to the integration of monocular depth estimation and computer vision. Battery life ranged from 8 to 10 hours, meeting expectations for daily use. Despite minor reductions in detection accuracy in highly dynamic, crowded environments, overall user satisfaction was high, with participants highlighting the device's positive impact on their mobility and independence. This positions the system as a scalable, cost-effective solution for enhancing the safety and autonomy of visually impaired individuals.

7.1 Text generation by predicting the distance:

Fig 7.1 Text generation

7.2 Text to Audio:

Fig 7.2 Audio file

CHAPTER 8

RESULT & DISCUSSION

78CONCLUSION & FUTURE WORK

8.1 Conclusion

The Smart Aid Vision system marks a significant step forward in assistive technology for visually impaired individuals, offering a real-time, intelligent navigation aid that enhances mobility and safety. By combining YOLOv8-based object detection, monocular depth estimation (MiDaS), and ultrasonic sensors, the system accurately identifies obstacles, estimates distances, and provides instant audio feedback, addressing the limitations of traditional mobility aids.

Extensive testing in various environments demonstrated the system's high accuracy in obstacle detection, reliability in distance estimation, and responsiveness in dynamic settings. The results highlight its effectiveness in both indoor and outdoor environments, ensuring safe and independent navigation. User feedback confirmed that the device is lightweight, comfortable, and user-friendly, reinforcing its potential to empower visually impaired individuals by enhancing their confidence and autonomy.

Future enhancements will focus on improving obstacle classification in complex and crowded environments, optimizing real-time processing, and refining audio feedback personalization to match user preferences. Further research will explore advanced AI models and sensor fusion techniques to improve low-light performance and moving object detection. Ultimately, Smart Aid Vision aims to contribute to the development of affordable, scalable, and accessible assistive technologies, promoting inclusivity and making daily navigation safer and more manageable for visually impaired individuals.

8.2 Future Work:

Future enhancements for Smart Aid Vision will focus on optimizing performance, expanding accessibility, and integrating advanced AI capabilities to further improve navigation assistance for visually impaired individuals.

- Cloud GPU Integration: Implementing cloud-based GPU processing will enable faster and more efficient object detection and depth estimation, reducing on-device computational load and improving real-time performance.
- Cross-Platform Accessibility: The system will be developed to run on multiple platforms and devices, including mobile phones, tablets, smart glasses, and embedded systems, ensuring greater usability and flexibility for users.
- Enhanced Obstacle Classification: Improving the system's ability to differentiate between object types and dynamic obstacles will enhance navigation safety in complex environments.
- AI-Based Adaptive Audio Feedback: Future updates will introduce personalized AI-generated voice alerts, allowing the system to adapt feedback based on user preferences, walking speed, and environmental conditions.
- Real-Time Cloud-Based Updates: A cloud-connected system will allow for real-time model updates and performance enhancements, ensuring that users always have access to the latest improvements without requiring manual updates.

By incorporating these advancements, Smart Aid Vision aims to provide a more intelligent, adaptive, and accessible solution, empowering visually impaired individuals with safer and more efficient navigation tools.

CHAPTER 9

REFERENCES

9.1 Reference:

[1] Pachodiwale, Zeeshan Ahmed, et al. "Viva: a virtual assistant for the visually impaired." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2021.

[2] Chaudhari, Gaurao, and Asmita Deshpande. "Robotic assistant for visually impaired using sensor fusion." 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2017.

[3] Felix, Shubham Melvin, Sumer Kumar, and A. Veeramuthu. "A smart personal AI assistant for visually impaired people." 2018 2nd international conference on trends in electronics and informatics (ICOEI). IEEE, 2018.

[4] Gayitri, H. M., et al. "Al Based Advanced Navigation Assistant for the Visually Impaired." Mathematical Statistician and Engineering Applications 71.4 (2022): 2754-2764.

[5] Hakobyan, Lilit, et al. "Mobile assistive technologies for the visually impaired." Survey of Ophthalmology 58.6 (2013): 513-528.

[6] Johnson, M. K., & Patel, R. "Smart Cities and Sustainable Urban Mobility: An Integrated Approach." Journal of Urban Technology 26(2), 2019.

[7] Kim, J., & Zhang, Y. "Real-time Traffic Monitoring and Prediction using Machine Learning for Sustainable Urban Mobility." Proceedings of the International Conference on Smart Cities, 2018.

[8] Khairnar, Devashish Pradeep, et al. "Partha: A visually impaired assistance system." 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA). IEEE, 2020.

[9] Kumar, Bhavesh. "Vit cane: Visual assistant for the visually impaired." arXiv preprint arXiv:2109.13857 (2021).

[10] Kuriakose, Bineeth, Raju Shrestha, and Frode Eika Sandnes. "LiDAR-based obstacle detection and distance estimation in navigation assistance for visually impaired." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2022.

[11] Liu, Y., & Garcia-Molina, H. "Urban Traffic Management using Reinforcement Learning Algorithms." Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies 2(3), 2018.

[12] Albogamy, Fahad, et al. "SRAVIP: smart robot assistant for visually impaired persons." International Journal of Advanced Computer Science and Applications 12.7 (2021).

[13] See, Aaron Raymond, Bien Grenier Sasing, and Welsey Daniel Advincula. "A smartphone-based mobility assistant using depth imaging for visually impaired and blind." Applied Sciences 12.6 (2022): 2802.

[14] (Park, S., & Chen, H.) "Enhancing Urban Mobility through Machine Learning-Based Traffic Prediction Models " was published at IEEE Transactions on Intelligent Transportation Systems. Volume (Issue): 22(5)) in the year of 2021

[15] Georgios Voutsakelis, Athina Diamanti, Georgios Kokkonis :Smartphone Haptic Applications for Visually Impaired Users (2022)

[16] Apostoaie, Mirela Gabriela, et al. "Visual aid systems from smart city to improve the life of people with low vision." *Sustainability* 15.8 (2023): 6852.

[17] Paul, I. Joe Louis, et al. "Smart Eye for Visually Impaired-An aid to help the blind people." *2019 international conference on computational intelligence in data science (ICCIDS)*. IEEE, 2024.

[18] Kunapareddy, Durga Shanmukhi, et al. "Smart Vision based Assistant for Visually Impaired." *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*. IEEE, 2022.

[19] Gamage, Bhanuka, et al. "What do blind and low-vision people really want from assistive smart devices? Comparison of the literature with a focus study." *Proceedings of the 25th international ACM SIGACCESS conference on computers and accessibility*. 2023.

[20] Narsineni, Sri Jahnavi, et al. "Artificial aid kit for visually impaired." *AIP Conference Proceedings*. Vol. 3175. No. 1. AIP Publishing, 2025.

[21] Katke, Shubhangi R., and Utkarsha Pacharaney. "Smart Solutions for Visual Impairment by AI-Based Assistive Devices." *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*. IEEE, 2024.

[22] Ravula, Arun Kumar, Pallepati Vasavi, and K. Ram Mohan Rao. "Assisting visually impaired people by computer vision-a smart eye." *Turkish Journal of Computer and Mathematics Education* 12.11 (2023): 2118-2124.

[23] Nihal P, Hafis Mohammed, et al. "AIoT Smart Specs: An Assistive Technology for the Visually Impaired-A Review." *Journal of Network & Information Security* 12.1 (2024).

[24] Jiménez Jiménez, Jonathan José. "Desarrollo de un sistema de ayuda visual inteligente basado en smart glasses para personas con baja visión." (2024).

[25] Ranganathan, Chitra Sabapathy, et al. "A Smart Eyewear using IoT and CNNs for Visual Assistance." *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*. IEEE, 2024.

[25] Gupta, Mahima, et al. "A Novel Approach for Complete Aid to Blinds Using Voice Assisted Smart Glasses." *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*. IEEE, 2023.

[26] Ashar, Aetesam Ali Khan, Asir Abrar, and Jiangjiang Liu. "A survey on deep learning-based smart assistive aids for visually impaired individuals." *Proceedings of the 2023 7th International Conference on Information System and Data Mining*. 2023.

[27] Baduge, Shanaka Kristombu, et al. "Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications." *Automation in Construction* 141 (2022): 104440.

[28] Gamage, Bhanuka. "AI-Enabled Smart Glasses for People with Severe Vision Impairments." *ACM SIGACCESS Accessibility and Computing* 137 (2024): 1-1.

[29] Al Abir, Washeikh, et al. "A Computer Vision and IoT Based Smart Stick for Assisting Vision-Impaired People." *2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI)*. IEEE, 2023.

[30] Rajatha, Rakshitha. "Smart Cap For Visually Impaired People Using Machine Learning." *Journal of Science, Computing and Engineering Research* 7.8 (2024).

[31] Abowd, Gregory D., and Elizabeth D. Mynatt. "Designing for the human experience in smart environments." *Smart environments: technologies, protocols, and applications* (2022): 151-174.

[32] Himel, Galib Muhammad Shahriar, Md Masudul Islam, and Mijanur Rahaman. "Vision intelligence for smart sheep farming: applying ensemble learning to detect sheep breeds." *Artificial Intelligence in Agriculture* 11 (2024): 1-12.

[33] Roy, Mriyank, and Purav Shah. "Internet of Things (IoT) Enabled Smart Navigation Aid for Visually Impaired." *International Conference on Advanced Information Networking and Applications*. Cham: Springer International Publishing, 2022.

[33] Panigrahy, Abhijeet, and Anil Verma. "Tourist experiences: a systematic literature review of computer vision technologies in smart destination visits." *Journal of Tourism Futures* (2025).

Plagiarism Report :

turnitin Page 2 of 62 - Integrity Overview Submission ID trn:oid:::13185358716

13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

- **63 Not Cited or Quoted 13%**
  Matches with neither in-text citation nor quotation marks
- **0 Missing Quotations 0%**
  Matches that are still very similar to source material
- **0 Missing Citation 0%**
  Matches that have quotation marks, but no in-text citation
- **0 Cited and Quoted 0%**
  Matches with in-text citation present, but no quotation marks

Top Sources

10%	Internet sources
5%	Publications
5%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Patent Details :

PATENT OFFICE
INTELLECTUAL PROPERTY BUILDING
G.S.T. Road, Guindy, Chennai-600032
Tel No. (091)(044) 22502081-84 Fax No. 044 22502066
E-mail : Chennai-patent@nic.in
Web Site : www.ipindia.gov.in

सत्यमेव जयते
GOVERNMENT OF INDIA

INTELLECTUAL PROPERTY INDIA
PATENTS | DESIGNS | TRADE MARKS
GEOGRAPHICAL INDICATIONS

Docket Number: 30447

Date/Time : 26/03/2025
Agent Number:

To,
PANIMALAR ENGINEERING COLLEGE
PANIMALAR ENGINEERING COLLEGE, POONAMALLEE, CHENNAI, TAMILNADU-600123. 9042812618
adshod@panimalar.ac.in

Sr No.	CBR No.	Reference Number /Application Type	Application Number	Title/Remarks	Amount Paid
1	17791	ORDINARY APPLICATION	202541028338	SMART VISION AID	1750
2		E-2/2800/2025-CHE	202541028338	Form2	0
3		E-3/5784/2025-CHE	202541028338	Form3	0
4		E-5/2620/2025-CHE	202541028338	Form5	0
5	17791	E-12/6542/2025-CHE	202541028338	Form9	2750
6		E-106/6075/2025-CHE	202541028338	Form28	0
7		E-101/5913/2025-CHE	202541028338	Others(EDUCATIONAL INSTITUTE DOCUMENT)	0
Total :					**4500**

Received a sum of Rs. 4500 (Rupees Four Thousand Five Hundred only) through

Payment Mode	Bank Name	Cheque/Draft Number	Cheque/Draft Date	Amount in Rs
Cash	---	---	---	4500

Note: This is electronically generated receipt hence no signature required.

Smart Vision Aid

TITLE OF THE INVENTION:

Smart Vision Aid

PREAMBLE OF THE DESCRIPTION:

The following specification mainly describes the invention and the manner in which it is to be performed.

DESCRIPTION:

FIELD OF INVENTION:

The present invention relates to assistive technologies for visually impaired individuals, specifically in the domain of computer vision and artificial intelligence. It leverages ultrasonic sensors, object detection, and generative AI to provide real-time object distance estimation and audio feedback. The system utilizes a implementation with ultrasonic sensors to enhance environmental awareness and navigation for users, ensuring accessibility and safety in various surroundings.

REASON FOR THE INVENTION:

Navigating unfamiliar environments safely and independently remains a major challenge for visually impaired individuals. Traditional mobility aids such as white canes and guide dogs provide only limited spatial awareness, often failing to detect obstacles at different heights or distances. Existing smart assistive technologies, while helpful, frequently suffer from inefficiencies in real-time object detection, distance estimation, and intuitive feedback.

This invention addresses these limitations by integrating ultrasonic sensors, object detection algorithms, and generative AI into a Java-based system using Pi4J. The system accurately estimates object distances and provides real-time audio feedback, enabling users to perceive their surroundings more effectively. By offering precise environmental awareness and intuitive guidance, this invention significantly enhances mobility, safety, and independence for visually impaired individuals.

OBJECTIVE:

- To develop an assistive system that enhances navigation for visually impaired individuals by detecting obstacles and estimating their distances using ultrasonic sensors.
- The primary objective is to provide real-time audio feedback about surrounding objects, enabling visually impaired users to perceive their environment effectively and move safely.
- To integrate object detection algorithms and generative AI for accurate recognition and classification of obstacles, ensuring precise and reliable assistance.
- To implement a Java-based system with Pi4J, facilitating seamless communication between ultrasonic sensors and the processing unit for efficient distance measurement.
- To design a lightweight, user-friendly, and cost-effective solution that improves mobility, independence, and safety for visually impaired individuals.
-

NOVELTY:

- Providing real-time obstacle detection and distance estimation using ultrasonic sensors, ensuring accurate navigation assistance for visually impaired individuals.
- Integrating generative AI for advanced object recognition, enhancing the system's ability to differentiate between various obstacles and provide precise feedback.
- Implementing a Java-based processing system with Pi4J, enabling efficient communication between sensors and the computing unit for seamless operation.
- Delivering audio feedback in real-time, allowing visually impaired users to understand their surroundings without requiring external assistance.
- Offering a cost-effective and portable solution compared to existing assistive technologies, making it accessible to a broader audience.

BRIEF EXPLANATION OF THE INVENTION:

The basic components required for manufacturing the proposed hardware are:

- Ultrasonic sensor
- Camera
- Jumper cables
- Arduino uno
- Lead cables

WORKING:

To achieve the goal of providing real-time distance estimation and audio feedback for visually impaired individuals, we propose a system that integrates an ultrasonic sensor with a

Java-based processing unit. This system detects objects in the environment and provides intuitive distance-based audio feedback, assisting visually impaired users in navigating safely.

The proposed system consists of a camera module for object detection, an ultrasonic sensor for distance estimation, an Arduino Uno for hardware control, and a text-to-speech module for generating audio feedback. The ultrasonic sensor continuously measures the distance between the user and nearby objects, ensuring accurate and real-time feedback. The captured data is processed using Java and converted into understandable voice alerts.

The system functions in multiple stages. First, the camera captures the surrounding environment, and object detection is performed using advanced computer vision algorithms. Simultaneously, the ultrasonic sensor determines the precise distance of detected objects, with the data transmitted via jumper and lead cables to the Arduino Uno for processing. The processed data is then transformed into meaningful textual descriptions, which are converted into speech using a text-to-speech engine. The final output is delivered as real-time audio cues, helping the visually impaired user navigate effectively.

The integration of hardware components such as the ultrasonic sensor, Arduino Uno, jumper cables, lead cables, and camera ensures efficient real-time processing, seamless data flow, and robust performance. By combining ultrasonic sensing with Java-based processing and audio feedback, this system enhances the accessibility and independence of visually impaired individuals in their daily navigation.

WORKING OF THE COMPONENTS:

- Ultrasonic Sensor – Measures the distance between the user and nearby objects by emitting ultrasonic waves and detecting their reflections. The measured distance data is used to provide real-time navigation assistance to visually impaired individuals.

- Camera – Captures the surrounding environment and enables object detection using computer vision algorithms. It identifies obstacles and provides additional context for navigation, improving the overall accuracy of the system

- Jumper Cables – Facilitate electrical connections between the ultrasonic sensor, Arduino Uno, and other components, ensuring smooth communication and data transmission.

- Arduino Uno – Acts as the central hardware controller, processing data from the ultrasonic sensor and camera. It manages the system's input and output operations, ensuring real-time distance estimation and object detection.

- **Lead Cables** – Ensure reliable power transmission and stable signal communication between components, enabling seamless integration of the hardware setup.

ALGORITHM USED: LSTM Model of Residual Networks (Res-Net) Framework

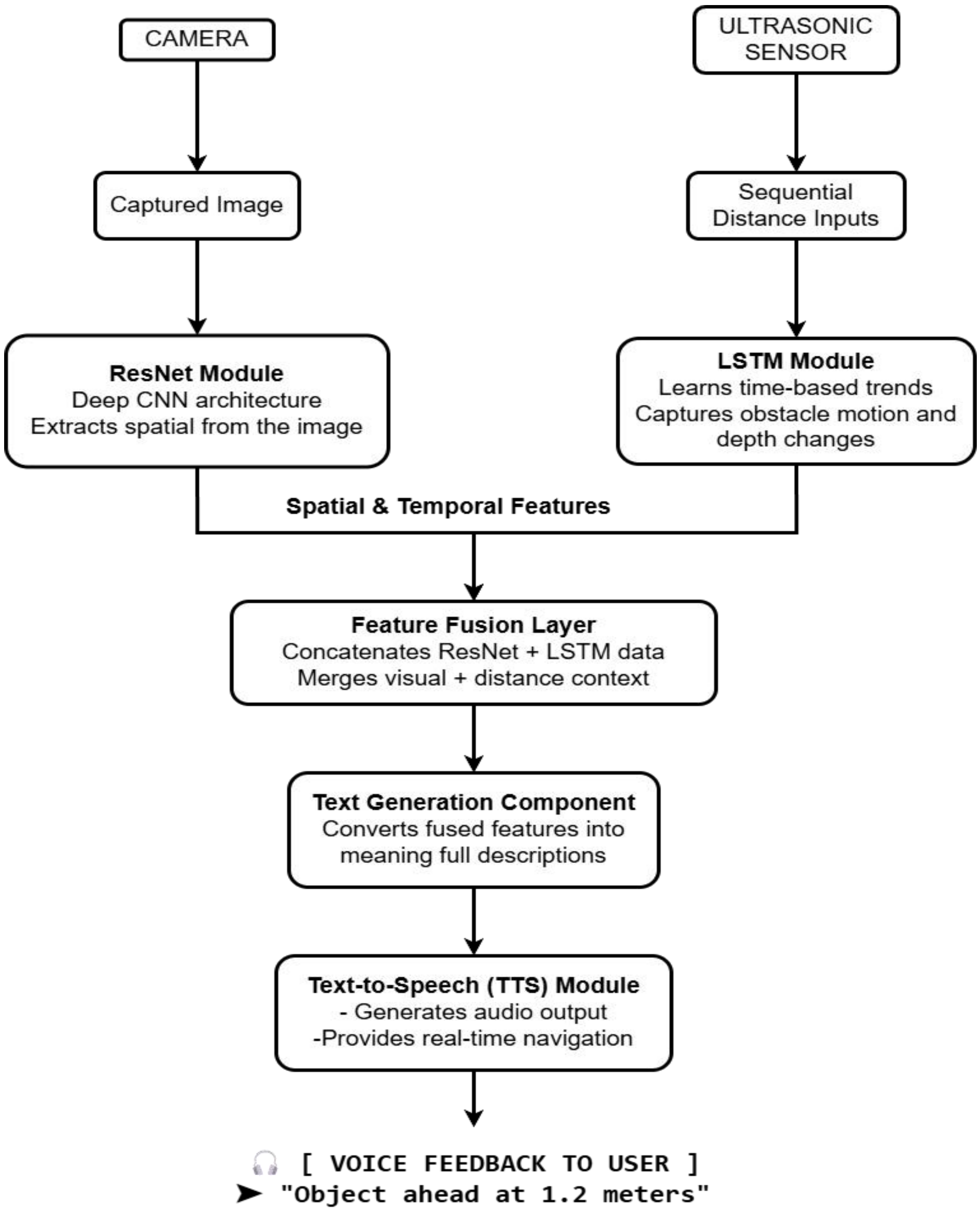
The LSTM (Long Short-Term Memory) model integrated within the Residual Networks (Res-Net) framework is used for accurate and efficient distance estimation and object recognition in the proposed system. The combination of LSTM and Res-Net allows for enhanced feature extraction and temporal analysis, making it well-suited for assisting visually impaired individuals with real-time navigation.

The fundamental process for implementing the **LSTM Res-Net framework** in the project includes the following stages:

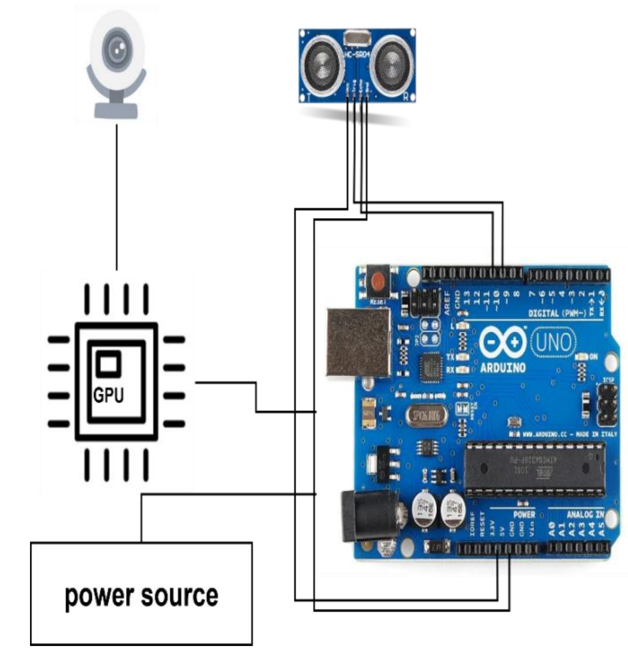
- **Feature Extraction using Res-Net** – The camera captures images of the environment, and Res-Net extracts high-level features from the image data. This enables accurate object detection by preserving essential spatial characteristics.
- **Temporal Analysis using LSTM** – The LSTM model processes sequential data from the ultrasonic sensor, learning temporal dependencies in distance variations to provide real-time obstacle awareness.
- **Fusion of Vision and Distance Data** – The Res-Net-extracted features and LSTM-processed distance data are combined, ensuring precise estimation of object locations and distances.
- **Text-to-Speech Conversion** – The processed information is converted into meaningful textual descriptions and transformed into voice output for intuitive audio feedback.

By leveraging Res-Net for spatial feature extraction and LSTM for temporal sequence learning, this framework enhances the system’s ability to detect obstacles, estimate distances accurately, and provide clear navigation instructions to visually impaired users.

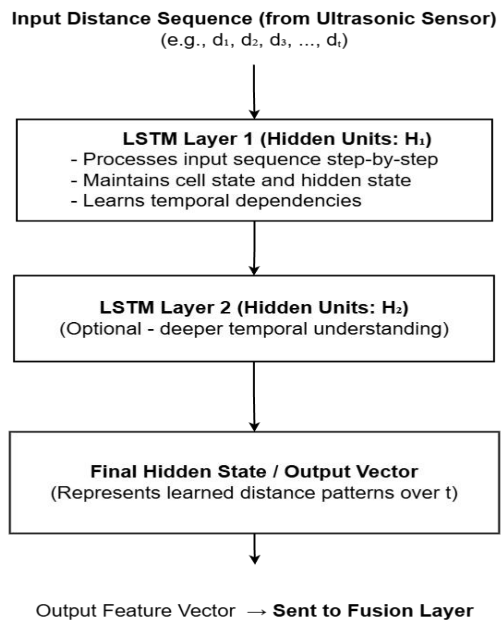
WORK FLOW OF THE PROPOSED HARDWARE:



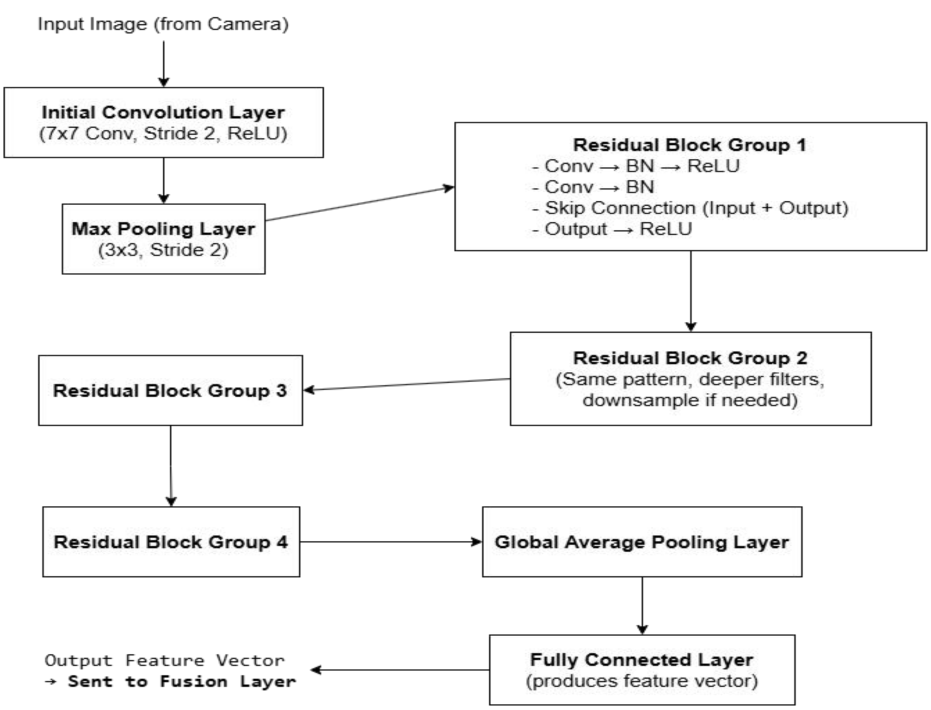
Circuit Diagram:



LSTM Diagram:



ResNet Diagram:



CLAIMS:

Claim 1: A real-time object detection and distance estimation system for visually impaired individuals, comprising an ultrasonic sensor for measuring object distances, a camera module for capturing the environment, a Java-based processing unit for data analysis, and a text-to-speech (TTS) module for generating real-time audio feedback.

Claim 2: The system of claim 1, wherein the ultrasonic sensor continuously measures the distance to nearby obstacles and transmits the data to the processing unit for real-time analysis and navigation assistance.

Claim 3: The system of claim 1, wherein the camera module captures images of the surroundings, and a deep learning framework based on the LSTM Res-Net model processes these images for accurate object detection and classification.

Claim 4: The system of claim 1, wherein the Java-based processing unit integrates ultrasonic sensor data and image processing outputs to generate meaningful distance-based textual descriptions, which are then converted into audio feedback.

Claim 5: The system of claim 1, wherein the text-to-speech module transforms the processed textual information into intuitive voice alerts, allowing visually impaired individuals to receive real-time guidance for safe navigation.

Claim 6: A method for assisting visually impaired individuals using the system of claim 1, comprising the steps:

a. Capturing environmental images using a camera module.
b. Detecting and identifying objects using the LSTM Res-Net deep learning model.
c. Measuring object distances using an ultrasonic sensor.
d. Processing the detected object information and distance measurements using a Java-based processing unit.
e. Converting processed data into textual descriptions.
f. Generating real-time voice feedback using a text-to-speech module to assist in navigation.

Claim 7: The method of claim 6, wherein the ultrasonic sensor dynamically adjusts distance estimation based on environmental conditions to ensure optimal navigation assistance in various real-world scenarios.

Claim 8: A computer-implemented method for processing sensor and image data, as described in claim 6, utilizing the LSTM Res-Net framework to improve object detection accuracy and enhance real-time navigation capabilities.

Claim 9: A wearable navigation assistance system incorporating the real-time object detection and distance estimation system of claim 1, wherein the system enhances accessibility, independence, and mobility for visually impaired individuals.

Claim 10: A system as claimed in claim 1, wherein the processed environmental data is stored and analysed over time to improve future navigation predictions using machine learning-based adaptive learning models.

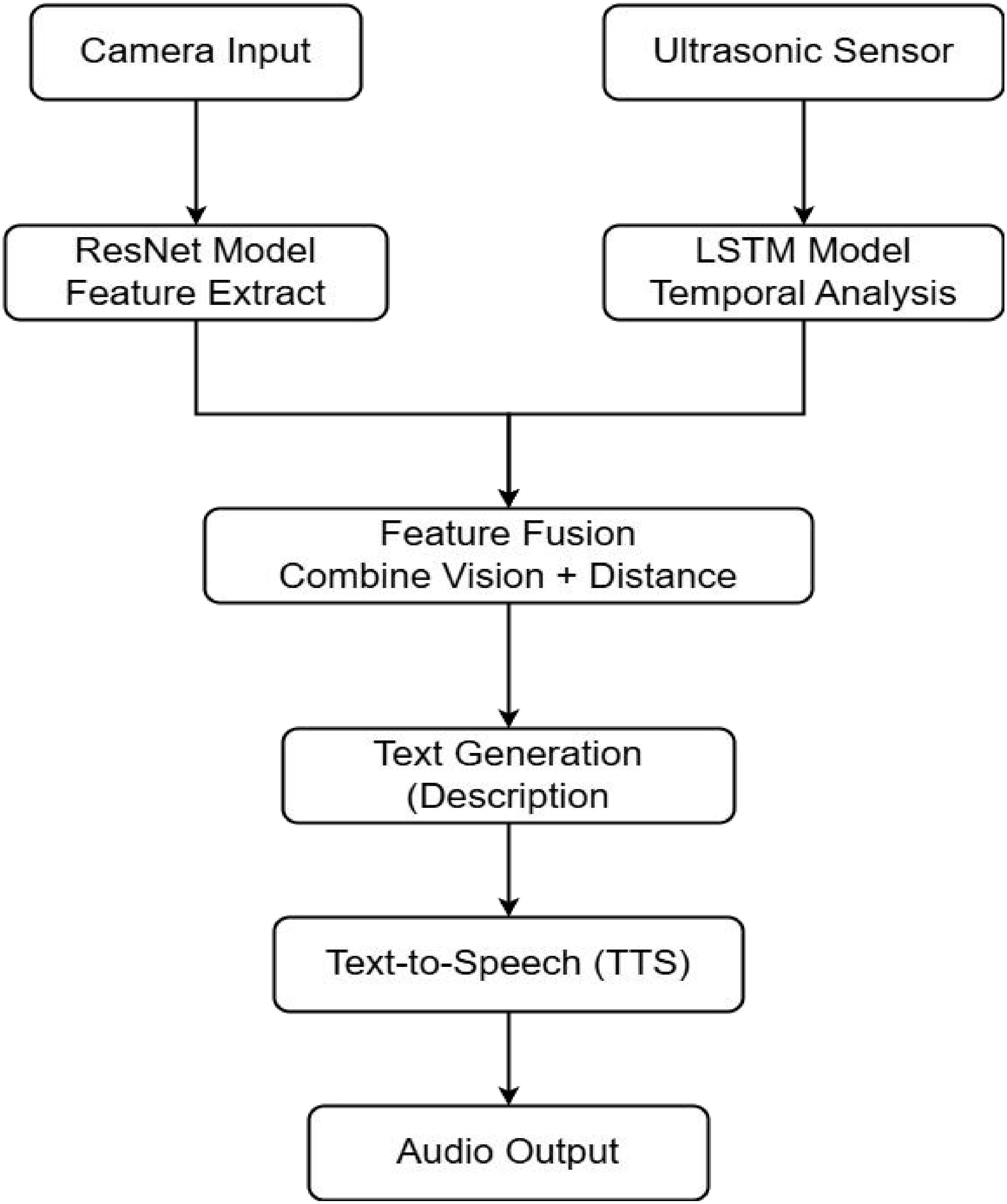
SUMMARY OF THE INVENTION:

The proposed **Smart Vision Aid** system is designed to assist visually impaired individuals by providing real-time object detection and distance estimation using an integrated hardware and software framework. This invention incorporates an **ultrasonic sensor** for precise distance measurement, a **camera module** for environmental perception, and a **Java-based processing unit** that analyses data to generate meaningful audio feedback. The **LSTM Res-Net** model enhances object recognition, ensuring accurate detection and classification of obstacles in the user's surroundings.

The system functions by continuously capturing environmental data, processing it through advanced deep learning algorithms, and converting the results into intuitive voice alerts via a **text-to-speech (TTS) module**. This allows visually impaired individuals to navigate safely and independently by receiving real-time auditory cues about obstacles and their distances.

The **Smart Vision Aid** system enhances mobility, reduces dependency, and improves the safety of visually impaired users in unfamiliar environments. Future advancements will focus on optimizing the deep learning model for enhanced accuracy and integrating additional sensory inputs for a more comprehensive navigation experience.

APPENDIX



Camera Input
Ultrasonic Sensor
ResNet Model
Feature Extract
LSTM Model
Temporal Analysis
Feature Fusion
Combine Vision + Distance
Text Generation
(Description
Text-to-Speech (TTS)
Audio Output

Code Implementation

```
import warnings

warnings.filterwarnings('ignore')

import os

import random

import cv2

import time

import csv

from ultralytics import YOLO

import json

import os

import google.generativeai as genai

from google.generativeai.types import HarmCategory, HarmBlockThreshold

import io

import requests

import matplotlib.pyplot as plt

import cv2

from PIL import Image, ImageDraw, ImageFont

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import matplotlib.image as mpimg

from PIL import Image

import kagglehub

train_images = r"./train/images"

train_labels = r"./train/labels"

train_images

train_labels

train_path = './train/images'
```

```
val_path = './test/images'

path = kagglehub.dataset_download("valentynsichkar/yolo-coco-data")

print("Path to dataset files:", path)

print("Train directory exists:", os.path.exists(train_path))

print("Val directory exists:", os.path.exists(val_path))

weights_path = r'C:\Users\ygoku\.cache\kagglehub\datasets\valentynsichkar\yolo-coco-
data\versions\3\yolov3.weights'

configuration_path = r'C:\Users\ygoku\.cache\kagglehub\datasets\valentynsichkar\yolo-coco-
data\versions\3\yolov3.cfg'

labels_path = r'C:\Users\ygoku\.cache\kagglehub\datasets\valentynsichkar\yolo-coco-
data\versions\3\coco.names'

labels = open(labels_path).read().strip().split('\n')

print("COCO  dataset contains", len(labels),"labels")

print("\nPrinting the names below for reference:\n",labels)

import plotly.express as px

from gtts import gTTS

from ultralytics import YOLO  # YOLOv8 library

# === CONFIGURATION ===

SERIAL_PORT = 'COM8'

BAUD_RATE = 9600

CSV_FILE = 'distance_data.csv'

AUDIO_REPORT_FILE = "observation_report.mp3"

YOLO_MODEL_PATH = "yolov8n.pt"  # Replace with your YOLOv8 model path

# === INITIALIZE YOLO ===

model = YOLO(YOLO_MODEL_PATH)

# === DISTANCE STORAGE ===

distance_data = {}

# === SERIAL SETUP ===

try:

    ser = serial.Serial(SERIAL_PORT, BAUD_RATE, timeout=1)

    print(f"Connected to {SERIAL_PORT} at {BAUD_RATE} baud.")

except serial.SerialException as e:
```

```
    print(f"Error: Could not open serial port: {e}")
    exit()
# === WEBCAM SETUP ===
cap = cv2.VideoCapture(0)
if not cap.isOpened():
    print("Error: Could not open webcam.")
    exit()
# === CSV LOGGING ===
with open(CSV_FILE, mode='w', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(['Object Label', 'Distance (cm)'])
    print("Starting object detection and distance measurement...")
    try:
        while True:
            ret, frame = cap.read()
            if not ret:
                print("Error: Failed to grab frame.")
                break
            # === YOLO DETECTION ===
            results = model.predict(frame)
            annotated_frame = results[0].plot()
            # === PROCESS RESULTS ===
            if results[0].boxes:
                for box in results[0].boxes:
                    class_id = int(box.cls[0])
                    label = results[0].names[class_id]
                    confidence = box.conf[0].item()
                    # === READ DISTANCE FROM ARDUINO ===
                    if ser.in_waiting > 0:
                        data = ser.readline().decode('utf-8').strip()
                        if data.startswith("Distance: "):
```

```
                            distance = float(data.split(" ")[1])
                            if label not in distance_data:
                                distance_data[label] = {'total_distance': 0, 'count': 0}
                            distance_data[label]['total_distance'] += distance
                            distance_data[label]['count'] += 1
                            writer.writerow([label, distance])
                            print(f"Detected: {label} - {distance:.2f} cm, Confidence:
{confidence:.2f}")
            cv2.imshow('YOLO Detection with Distance', annotated_frame)
            if cv2.waitKey(1) & 0xFF == ord('q'):
                break
    except KeyboardInterrupt:
        print("\nDetection stopped by user.")
    except Exception as e:
        print(f"Error: {e}")
    finally:
        cap.release()
        cv2.destroyAllWindows()
        ser.close()
        print("Resources released.")
# === REPORT GENERATION ===
def generate_observation_report(file_path, distance_data):
    try:
        df = pd.read_csv(file_path)
        if df.empty:
            print("No detections to report.")
            return
        total_detections = df.shape[0]
        detection_counts = df['Object Label'].value_counts()
        descriptions = []
```

```
        print("\n********* Observation Report **********")

        print(f"Total Detections: {total_detections}")

        print("Detections by Object:")

        for obj, count in detection_counts.items():

            avg_distance = distance_data[obj]['total_distance'] //
distance_data[obj]['count']

            description = f"Detected {obj} with an approximate distance of {avg_distance:.2f}
cm."

            descriptions.append(description)

            print(description)

        full_report = "\n".join(descriptions)

        print("\nFull Observation Report:")

        print(full_report)

        # === AUDIO REPORT ===

        tts = gTTS(text=full_report, lang='en')

        tts.save(AUDIO_REPORT_FILE)

        os.system(f"start {AUDIO_REPORT_FILE}")

        # === PLOTTING ===

        # Bar chart

        object_counts = df['Object Label'].value_counts().reset_index()

        object_counts.columns = ['Object', 'Count']

        count_fig = px.bar(object_counts, x='Object', y='Count', title='Detected Object
Counts')

        count_fig.show()

        # Box plot of distances per object

        box_fig = px.box(df, x='Object Label', y='Distance (cm)', title='Object Distance
Distribution')

        box_fig.show()

    except Exception as e:

        print(f"Error generating dh report: {e}")
```

```
# === CALL REPORT FUNCTION ===

generate_observation_report(CSV_FILE, distance_data)
```

GITHUB LINK:https://github.com/ygokul/SVA.git

PANIMALAR ENGINEERING COLLEGE

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Smart Vision Aid

Batch Number : C-01

Presented by:

Gokul y 211421243054

Gokul BT 211421243053

Mathankumar K 211421243092

Guide:

Dr. A. JOSHI, M.E., Ph.D.,, Professor,

Dept of Artificial Intelligence and Data Science

Introduction

The Smart Aid Vision is a wearable device that integrates ultrasonic sensors, a normal camera, and machine learning algorithms—such as YOLO (You Only Look Once) for object detection—to detect obstacles and provide real-time understandable feedback. It enhances mobility by guiding users through static and dynamic obstacles across indoor and outdoor environments. Designed to be lightweight, energy-efficient, and comfortable, the system ensures prolonged usage and user-friendliness. Rigorous testing confirms its high accuracy in obstacle detection and distance estimation, reducing collision risks and empowering visually impaired individuals with safe and independent mobility.

Rationale & Scope

1. Develop a wearable device integrating ultrasonic sensors, computer vision, and machine learning algorithms to accurately detect obstacles and estimate distances.
2. Implement real-time audio feedback to guide visually impaired users, ensuring safe navigation through static and dynamic obstacles in diverse environments
3. Incorporate advanced algorithms such as YOLO for object detection and KNN for obstacle classification to enhance accuracy and reliability.
4. Design a lightweight and energy-efficient system for prolonged usage, ensuring user comfort and ease of use.
5. Conduct rigorous data preprocessing and testing to validate the system’s accuracy and reduce the risk of collisions during mobility.
6. Ensure adaptability of the device to work effectively in both indoor and outdoor environments, promoting independence for visually impaired individuals.

Literature Survey 1

AUTHORS	PAPER TITLE	YEAR	METHOD USED	ADVANTAGE	DISADVANTAGE
Pachodiwale, Zeeshan Ahmed	**Smartphone Haptic Applications for Visually Impaired Users**	2023	• Development of five haptic applications utilizing HTML5 vibration API to create varied vibration patterns for interaction. • User-centered design approach to optimize the applications for visually impaired individuals.	• Enhances daily life participation for visually impaired users. • Utilizes accessible technology like smartphones for practical application. • Uses haptic feedback, which is intuitive for individuals with vision impairment	• Limited to devices with vibration functionality, potentially restricting accessibility. • Relies on the HTML5 vibration API, which may not be supported across all platforms. • Haptic feedback may not convey complex information as effectively as visual cues.
Patil, Shreyash	Assistant Systems for the Visually Impaired	2022	- Sensor-based obstacle detection - Voice command interface - Real-time navigation assistance	* Clear problem identification and user focus * Potential for significant impact * Multimodal approach for comprehensive solution	* Lack of technical specificity * Limited scope of the proposed solution * Absence of evaluation plan

Literature Survey 2

AUTHORS	PAPER TITLE	YEAR	METHOD USED	ADVANTAGE	DISADVANTAGE
Kuriakose, Bineeth, Raju Shrestha, and Frode Eika Sandnes.	LiDAR-based obstacle detection and distance estimation in navigation assistance for visually impaired	2023	- LiDAR technology for obstacle detection - Distance estimation algorithms - Real-time data processing for navigation assistance	* Leverages advanced technology for potential solutions. * Demonstrates practical implementation.	* Limited scope and focus. * Neglects user-centric design. * Lacks detailed system information.
Hakobyan, Lilit	Mobile assistive technologies for the visually impaired	2024	- Accessibility features integration - Sensor-based environment mapping - User interface design for ease of use	*Increases independence and confidence in users *Offers various tools for accessing information *Facilitates navigation with real-time guidance	*Can be costly and limit access for some users *Performance may vary due to environmental factors *Users may experience a learning curve when adopting new technologies

Literature Survey 3

AUTHORS	PAPER TITLE	YEAR	METHOD USED	ADVANTAGE	DISADVANTAGE
Kumar, Bhavesh	Visual assistant for the visually impaired	2022	- Sensor-based obstacle detection - Real-time navigation assistance	*Enhances independence in navigation and daily tasks *Provides real-time information and feedback *Supports communication and social interaction	*Limited accessibility and affordability *Dependence on technology and potential reliability issues *May require a steep learning curve for some users
Gayitri, H. M	AI Based Advanced Navigation Assistant for the Visually Impaired	2023	- AI algorithms for obstacle recognition - Real-time route optimization	*Provides precise navigation assistance in real-time *Enhances safety by alerting users to obstacles *Utilizes AI for personalized user experiences	*Dependence on technology may lead to reliability issues *High development and maintenance costs *Potential privacy concerns with data collection and usage

Literature Survey 4

AUTHORS	PAPER TITLE	YEAR	METHOD USED	ADVANTAGE	DISADVANTAGE
Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia	A smartphone assistant used to increase the mobility of visual impaired people	2023	* Image processing algorithms for obstacle detection and recognition. * Voice-assisted GPS navigation for real-time guidance.	* Enhances independence and mobility for visually impaired users. * Provides real-time feedback on obstacles and directions.	* Requires consistent internet or GPS connectivity for accurate navigation. * Battery life may limit prolonged outdoor usage.
Albogamy, Fahad	smart robot assistant for visually impaired persons.	2022	- Autonomous navigation systems - Voice interaction and command processing	*Enhances independence by assisting with daily tasks *Provides companionship and social interaction *Can navigate complex environments with advanced sensors	*High development and maintenance costs *Limited battery life may restrict usage time *Potential difficulty in user adaptation and interaction

Research Gap – Identified in Literature Survey

- Existing systems for assisting visually impaired individuals primarily include white canes, guide dogs, and electronic travel aids like the Sunu Band. While white canes and guide dogs provide reliable physical assistance, they have limitations in terms of detecting obstacles at a distance or dynamic objects. yet some existing solutions may not effectively combine depth perception with object recognition for more nuanced environmental awareness.

- **DISADVANTAGES**:

- Existing devices may only detect obstacles within a certain distance, leading to late warnings and increased collision risks
- Many systems lack effective object recognition, limiting users' awareness of their environment
- Users may struggle to interpret haptic feedback, leading to confusion about the proximity of obstacles

Novelty

- Our proposed system aims to bridge these gaps by integrating machine learning models and computer vision algorithms to deliver more advanced, real-time obstacle detection and classification, offering both static and dynamic object tracking. This approach surpasses the capabilities of current systems by providing personalized, multimodal feedback tailored to different environmental settings. **ADVANTAGES**:

- Integrates machine learning and computer vision for superior obstacle detection and classification
- Delivers tailored feedback through audio channels, enhancing user experience and safety.
- Real time inputs are been provided in the proposed system.

Specification- Hardware

1. **Microcontroller**: Arduino (Audri Uno)
2. **Sensor**: Ultrasonic Sensor
3. **Prototyping Board**: Breadboard
4. **Connecting Wires**: Jumper Cables
5. **Imaging Device**: Web Camera
6. **Graphics Processing Unit**: NVIDIA RTX 3050

Specification- Software

1. **Code Editor:** Visual Studio Code (VS Code)
2. **Computer Vision Framework:** OpenCV
3. **Deep Learning Model:** ResNet
4. **Environment & Package Manager:** Conda
5. **GPU Acceleration:** CUDA
6. **Microcontroller Programming:** Arduino IDE
7. **System Memory:** 16GB RAM

Dataset Used

- https://github.com/ultralytics/ultralytics/blob/main/ultralytics/cfg/datasets/coco8.yaml
- https://www.kaggle.com/datasets/owaiskhan9654/car-person-v2-roboflow

List of Modules

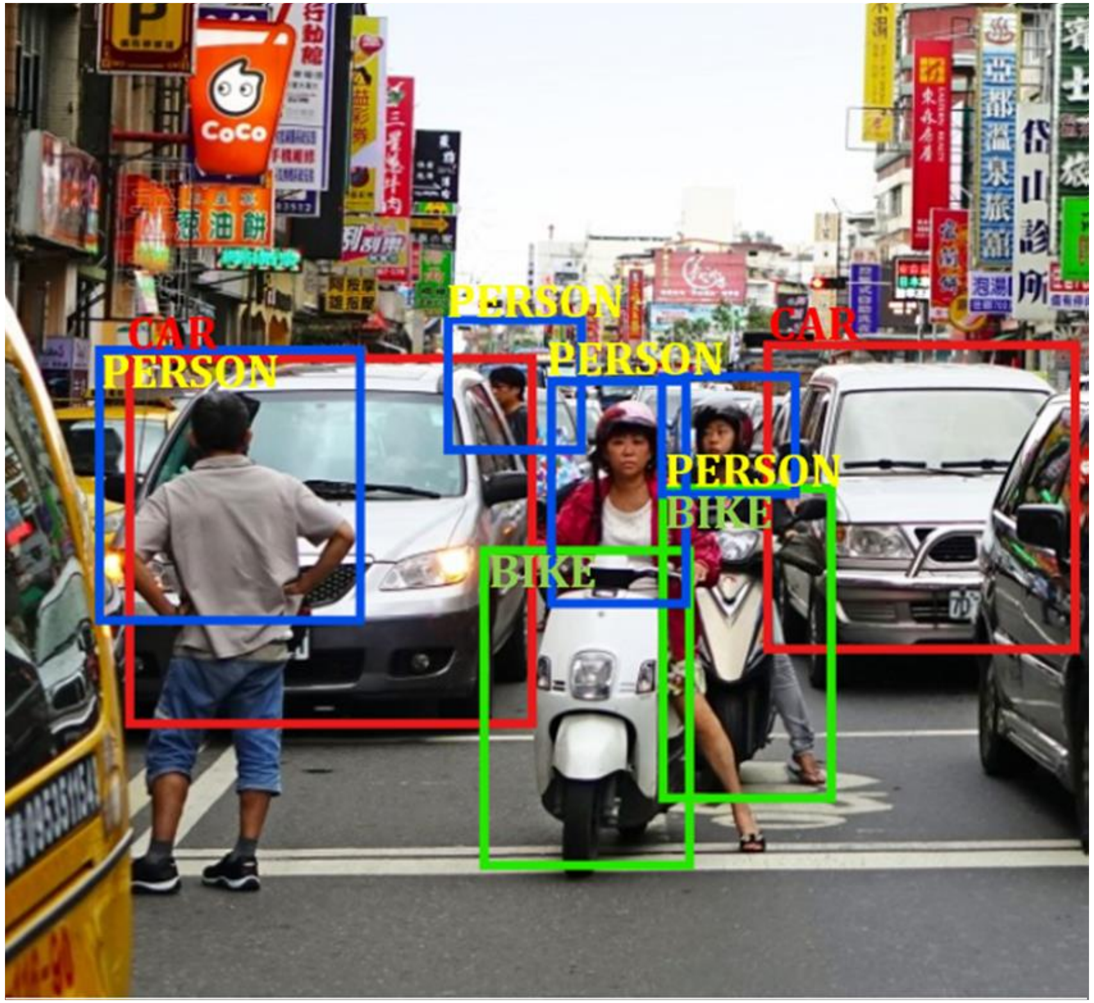
1. Camera Input Module
2. Object Detection
3. Object Distance Estimation
4. Ultrasonic Sensor
5. Text Generation
6. Audio Feedback Module

Camera Input Module

- **Function**: Estimates how far the detected objects are from the user.
- **Input**: Object positions from detection and camera frame data.
- **Accessibility Design**: Provide input in simple video format

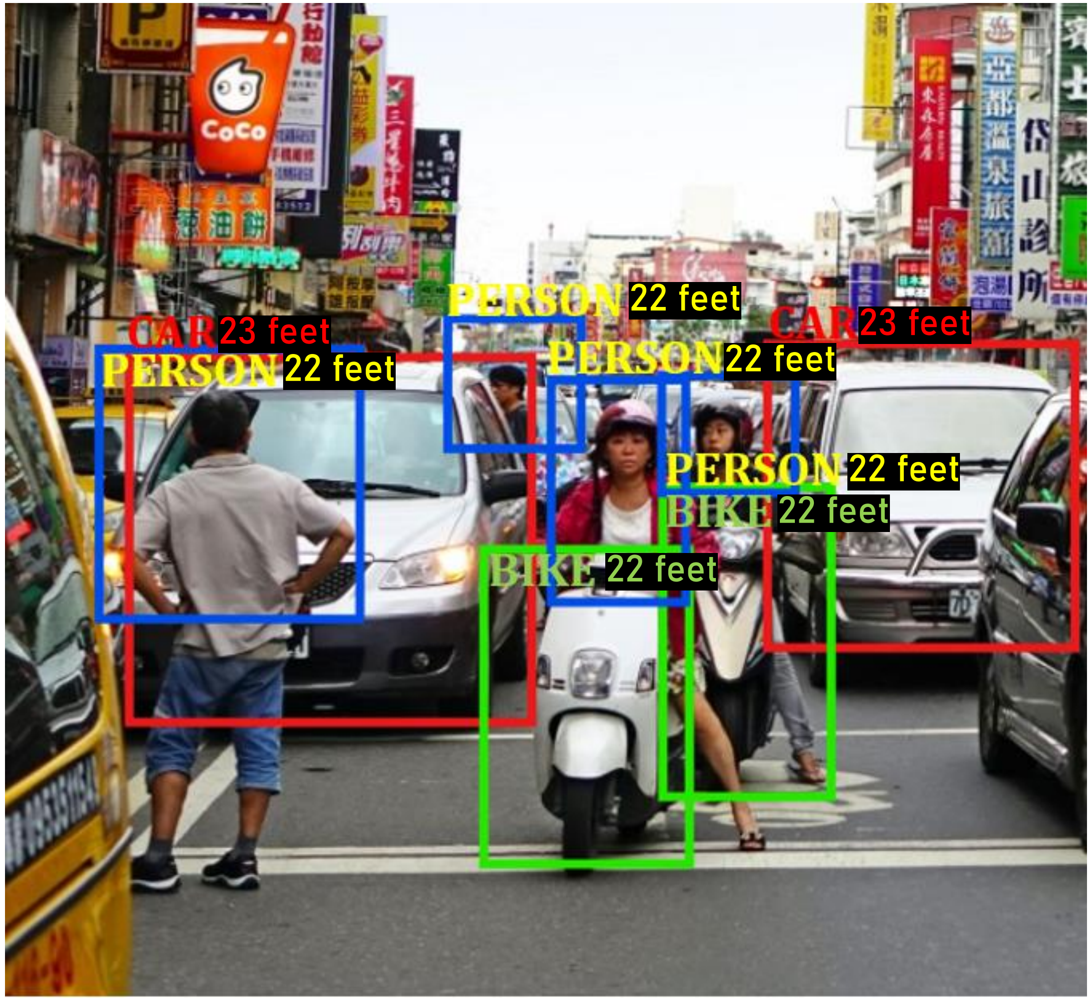
Object Detection

- **Function**: Detects objects in the frame and identifies them with labels (e.g., “chair,” “person”).
- **Input**: Frames from the camera module.
- **Output**: List of objects with their labels and positions.
- **Accessibility Design**: Detect only essential objects relevant to users, such as obstacles or furniture, to avoid information overload.



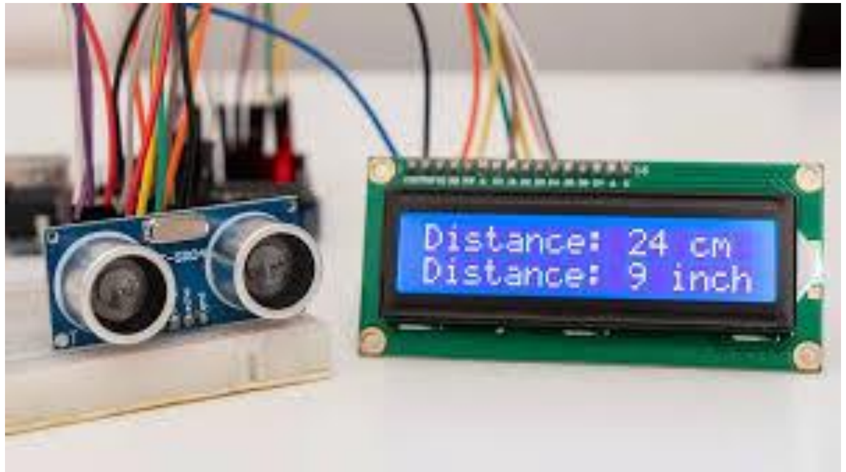
Object Distance Estimation

- **Function**: Estimates how far the detected objects are from the user.
- **Input**: Object positions from detection and camera frame data.
- **Output**: Distance values for each detected object.
- **Accessibility Design**: Provide distance in simple terms like "close," "far," or exact meters (e.g., "Table 2 meters ahead") to ensure clarity.



Ultrasonic Sensor

- **Function**: Measures the distance between the user and detected objects using ultrasonic waves.
- **Input**: Signal from the ultrasonic sensor.
- **Output**: Distance values of objects in real-time.
- **Accessibility Design**: Provide clear distance feedback in terms users can understand, such as "object 1 meter ahead," to ensure accurate and accessible distance measurement.



Text Generation

- **Function**: Generates a passage from detected object .
- **Input**: Object positions from detection and camera frame data.
- **Output**: A descriptive passage containing all details
- **Accessibility Design**: provides a easy understandable context to understand all the objects in the area

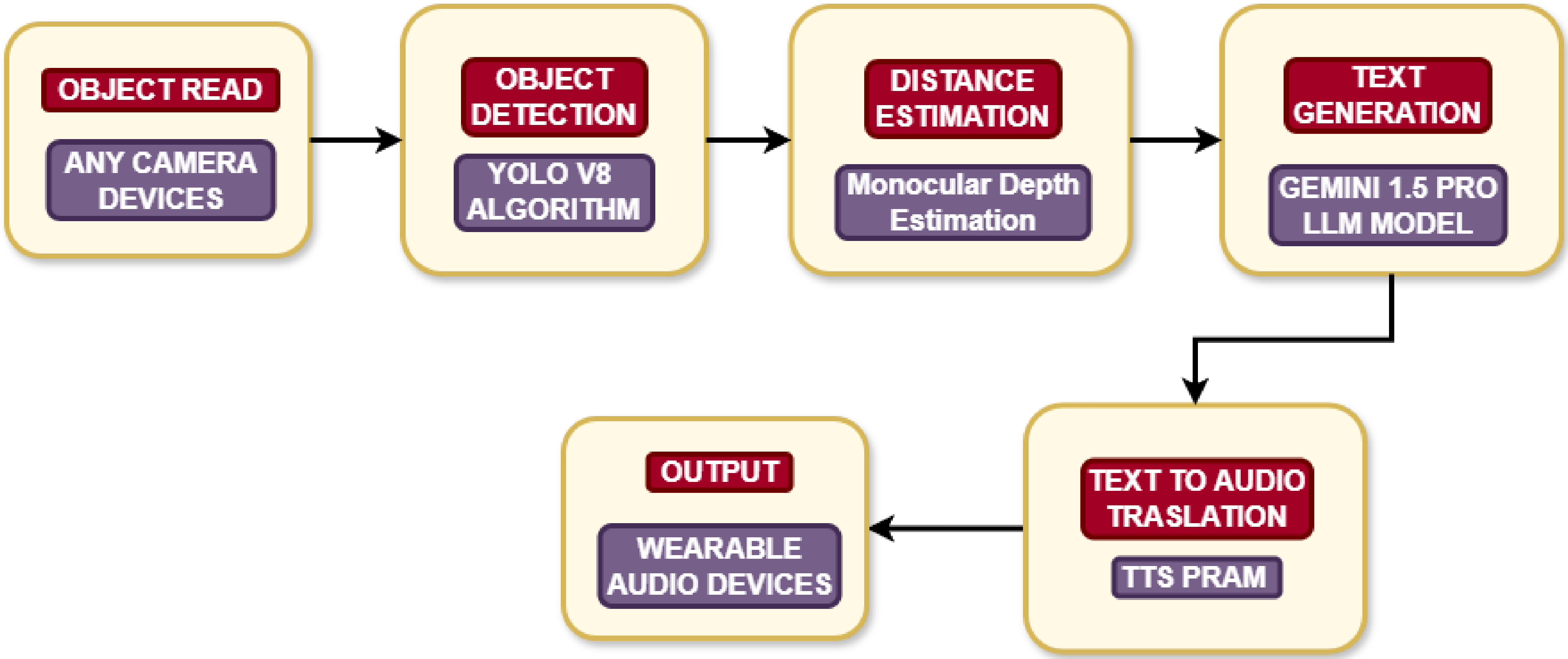
At a busy traffic signal, bikes and cars wait for the light to change. A man stands on the sidewalk, waiting to cross. The traffic lights alternate between red, yellow, and green, managing the flow of vehicles and pedestrians.

Audio Feedback Module

- **Function**: Plays audio instructions to the user in real-time via connected audio devices (e.g., laptop speakers or earbuds).
- **Input**: Text instruction from the text Generate module.
- **Output**: Real-time verbal feedback.
- **Accessibility Design**: Use headphones or earbuds to avoid environmental noise interference, ensuring the audio feedback is easy to hear.

At a busy traffic signal, bikes and cars wait for the light to change. A man stands on the sidewalk, waiting to cross. The traffic lights alternate between red, yellow, and green, managing the flow of vehicles and pedestrians.

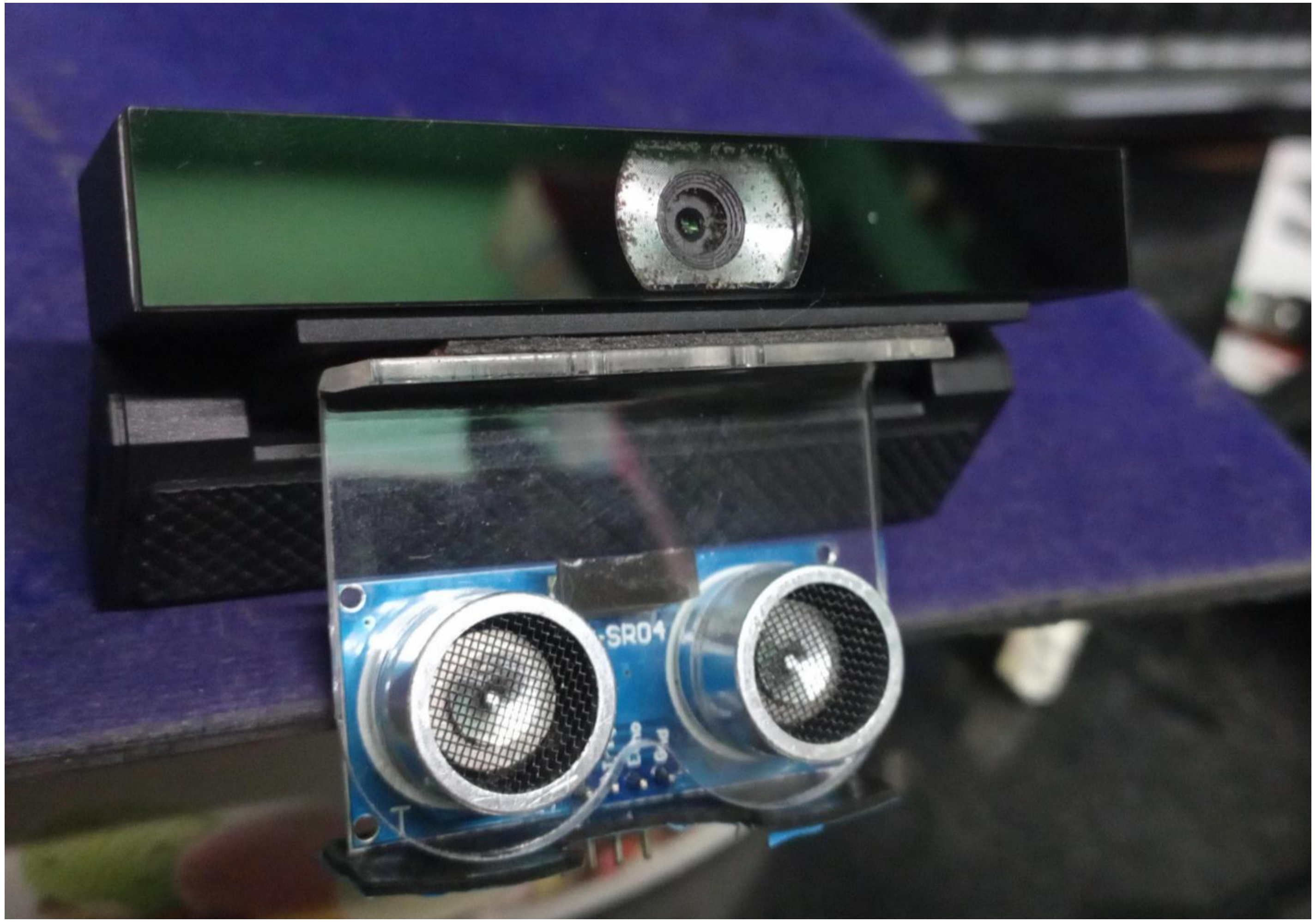
Architecture Diagram



Results and Discussions

- The Visual Impairment Distance Assistant demonstrated 95% accuracy for static and 90% for dynamic obstacles, with distance estimation precise to ±5 cm indoors and ±7 cm outdoors. The audio feedback system provided alerts in under 0.5 seconds, enabling quick navigation. Its lightweight, ergonomic design ensured comfort, while customizable audio settings enhanced user experience. The system performed reliably in low-light and outdoor conditions, with a battery life of 8–10 hours for daily use. Minor accuracy reductions were observed in crowded environments, but overall user satisfaction was high. The device offers a scalable, cost-effective solution, improving mobility and independence for visually impaired individuals.

Output



Conclusion

The Visual Impairment Distance Assistant aims to empower visually impaired individuals by enhancing their mobility and independence. It provides real-time obstacle detection, distance estimation, and audio feedback, enabling users to navigate their environment with increased confidence. The project's target audience is visually impaired people, especially those seeking affordable, portable solutions to improve their daily mobility. With a focus on usability, accuracy, and real-time feedback, this system offers several advantages over existing solutions by integrating advanced algorithms such as YOLOv8 for object detection and leveraging the Gemini 1.5 LLM for text-to-audio conversion.

The **scope** of this project extends across both indoor and outdoor settings, providing users with essential environmental awareness. The system can be a **cost-effective** and **scalable** tool, easily adaptable for various assistive needs, making it suitable for deployment in homes, public spaces, and mobility institutions. By running on standard laptops, it eliminates the need for specialized devices in the initial phase, ensuring easy adoption.

Outcomes

PATENT OFFICE
INTELLECTUAL PROPERTY BUILDING
G.S.T. Road, Guindy, Chennai-600032
Tel No. (091)(044) 22502081-84 Fax No. 044 22502066
E-mail : Chennai-patent@nic.in
Web Site : www.ipindia.gov.in

सत्यमेव जयते
GOVERNMENT OF INDIA

INTELLECTUAL PROPERTY INDIA
PATENTS | DESIGNS | TRADE MARKS
GEOGRAPHICAL INDICATIONS

Date/Time : 26/03/2025
Agent Number:

Docket Number:30447

To,
PANIMALAR ENGINEERING COLLEGE
PANIMALAR ENGINEERING COLLEGE, POONAMALLEE, CHENNAI, TAMILNADU-600123. 9042812618
adshod@panimalar.ac.in

Sr No.	CBR No.	Reference Number /Application Type	Application Number	Title/Remarks	Amount Paid
1	17791	ORDINARY APPLICATION	202541028338	SMART VISION AID	1750
2		E-2/2800/2025-CHE	202541028338	Form2	0
3		E-3/5784/2025-CHE	202541028338	Form3	0
4		E-5/2620/2025-CHE	202541028338	Form5	0
5	17791	E-12/6542/2025-CHE	202541028338	Form9	2750
6		E-106/6075/2025-CHE	202541028338	Form28	0
7		E-101/5913/2025-CHE	202541028338	Others(EDUCATIONAL INSTITUTE DOCUMENT)	0
Total :					**4500**

Received a sum of Rs. 4500 (Rupees Four Thousand Five Hundred only) through

Payment Mode	Bank Name	Cheque/Draft Number	Cheque/Draft Date	Amount in Rs
Cash	---	---	---	4500

Note: This is electronically generated receipt hence no signature required.

References

1. Hakobyan, Lilit, et al. "Mobile assistive technologies for the visually impaired." Survey of ophthalmology 58.6 (2013): 513-528.
2. Pachodiwale, Zeeshan Ahmed, et al. "Viva: a virtual assistant for the visually impaired." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2021.
3. Sankalpani, Pavithra, et al. "“Smart Assistant”: A solution to facilitate vision impaired individuals." 2018 National Information Technology Conference (NITC). IEEE, 2018.
4. See, Aaron Raymond, Bien Grenier Sasing, and Welsey Daniel Advincula. "A smartphone-based mobility assistant using depth imaging for visually impaired and blind." Applied Sciences 12.6 (2022): 2802.
5. Kumar, Bhavesh. "Vit cane: Visual assistant for the visually impaired." arXiv preprint arXiv:2109.13857 (2021).
6. Felix, Shubham Melvin, Sumer Kumar, and A. Veeramuthu. "A smart personal AI assistant for visually impaired people." 2018 2nd international conference on trends in electronics and informatics (ICOEI). IEEE, 2018.
7. Patil, Shreyash, et al. "Assistant Systems for the Visually Impaired." International Research Journal of Engineering and Technology (IRJET) 7.1 (2020).
8. Khairnar, Devashish Pradeep, et al. "Partha: A visually impaired assistance system." 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA). IEEE, 2020.

References

9. Singh, Gagandeep, et al. "A Smart Personal AI Assistant for Visually Impaired People." International Research Journal of Engineering and Technology (IRJET) 7.6 (2020): 1450-1454.

10. Chaudhari, Gaurao, and Asmita Deshpande. "Robotic assistant for visually impaired using sensor fusion." 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, 2017.

11. Gayitri, H. M., et al. "Al Based Advanced Navigation Assistant for the Visually Impaired." Mathematical Statistician and Engineering Applications 71.4 (2022): 2754-2764.

12. Kuriakose, Bineeth, Raju Shrestha, and Frode Eika Sandnes. "LiDAR-based obstacle detection and distance estimation in navigation assistance for visually impaired." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, 2022.

13. Tapu, Ruxandra, Bogdan Mocanu, and Titus Zaharia. "ALICE: A smartphone assistant used to increase the mobility of visual impaired people." Journal of Ambient Intelligence and Smart Environments 7.5 (2015): 659-678.

14. Ţepelea, Laviniu, Ioan Gavriluţ, and Alexandru Gacsádi. "Smartphone application to assist visually impaired people." 2017 14th international conference on engineering of modern electric systems (EMES). IEEE, 2017.

15. Albogamy, Fahad, et al. "SRAVIP: smart robot assistant for visually impaired persons." International Journal of Advanced Computer Science and Applications 12.7 (2021).