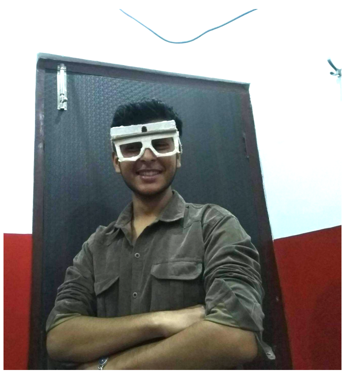
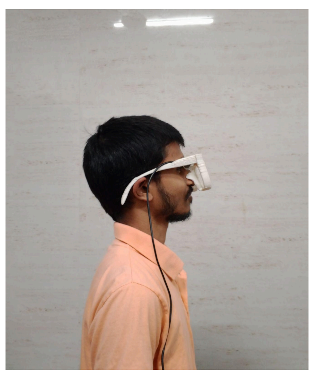
Samriddha Drishti: AI Powered Glasses for the visually impaired

Introduction

In the modern landscape the advancement of technologies and tools like AI are at an all-time high, even with these advancements the visually impaired people are facing challenges in their day to day living. Traditional smart blind sticks do not work effectively and solves the core challenges faced by visually impaired people. According to the WHO, as of late 2.2 billion of the world population is partially or fully visually challenged. The visually impaired generally face challenges like reading, navigation, and experiencing real time surroundings. To overcome those challenges the concept of Samriddha Dristhi was introduced. Samriddha Drishti is an AI Powered Glasses that helps visually impaired people to navigate, understand and explore the world independently.

As mentioned by the author of [1] the AI powered glasses prioritize the use of Deep Learning techniques such as RNN and CNN. Authors also mention how AI can assist the visually impaired . Samriddha Dristhi aims to implement features such as facial recognition, scene description, reading texts, cash recognition QR and bar code scanning to solve day to day problems faced by visually impaired person. Facial recognition feature helps visually impaired people to recognize there beloved ones independently . Scene description feature helps visually challenged people to explore and understand their surrounding effortlessly . Reading feature help visually impaired person to read and summarize handwritten and printed documents , books and newspaper etc. Cash recognition helps visually impaired people to recognize currency notes . QR and bar codes scanning features help visually challenged person to get the product information and read QR code information.


Related Works

The authors of [2] mention how the implementation of facial recognition was done by using an attachable camera to get faces as input and implement facial recognition in real time and provide the information to the user via TTS (Text to Speech) through an audio output (preferably earphones). Authors mention how the lighting affected the accuracy of the output.

Article [3] emphasizes the object detection methods were accomplished by the usage of models like YOLO v5 and providing the user with audio output.

There was a survey carried out by the authors of [4] where it was found that the mean age of the participants was around 23.5 years from a survey of people aged 6-56 with around 59% male, 51% from rural area and 64.4% students. In the survey around 90% of the volunteers were visually impaired. From the survey it was observed that people prioritized the use of reading with 72.9% positive experience, 44.7% for scene description, 36.5% for face recognition and 22.4% for walking. It was also observed that most ($2/3^{rd}$) participants utilized the device for at least one hour.

The use of computer vision, NLP (Natural Language Processing), and a wearable device was found to be the basics for the AI smart glasses as per [5]. The article also stated that the AI Powered technologies like smart glasses required high cost and lack of quality resources in the current state of the market which implied limited areas for experimentation and implementation for large scaling of the glasses.

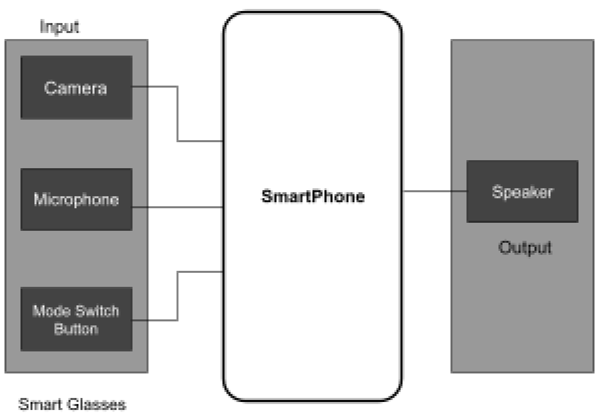
For the implementation of text reading for the visually impaired the authors of [7] had highlighted the use of OCR Optical Character Recognition which extracted the texts from documents which were then narrated to the user with the help of TTS.

Article [8] highlighted the use of cash recognition with the help of smartphone camera where the cash notes were identified. For the implementation of facial recognition, object detection and QR code/ Barcode scanning the authors of [9] had implemented the python library pyzbar which helps the user identify the products and its pricing. The authors of [6] had implemented the AI powered scene description models on the participants of their survey and Blip style models were found to be helpful to the visually impaired.

III SYSTEM ARCHITECTURE AND DESIGN

A. Overall System Architecture

Samriddha Drishti works on multi-layered systems. It has three primary parts: input, processing and output. Together they capture the visual & audio data , processes and provides real time audio feedback.Smart phone is used as the main computing device for our glasses that help us to manage the cost and make affordable assistive technology that is accessible to everyone.



- Input layer : captures the live visual , audio data through high resolution camera and microphone on smart glasses.
- AI Processing Layer : uses Pydroid3 and tensorflow lite to run models like yolov5s on device locally . A smart phone acts as a primary computing hub . It processes images, live visual and audio from camera and microphone and performs tasks like identifying objects and their location , facial recognition , scene captioning , cash recognition , visual question answering, QR bar code scanning , reading and summarizing text.
- Output layer : processed text information generated by the AI processing layer is converted to speech output using text-to-speech (TTS) module locally on device in smartphone and provides real time feedback to users by a small speaker situated on smart glasses or earphones connected to smart phone.

B. Hardware Components

Camera Module : High resolution camera Captures live video feed for identifying object , face recognition , cash recognition and images are captured for reading , summarizing text and scene captioning .

Microphone : Captures the audio data from the user and sends it to the smart phone for processing user requests .

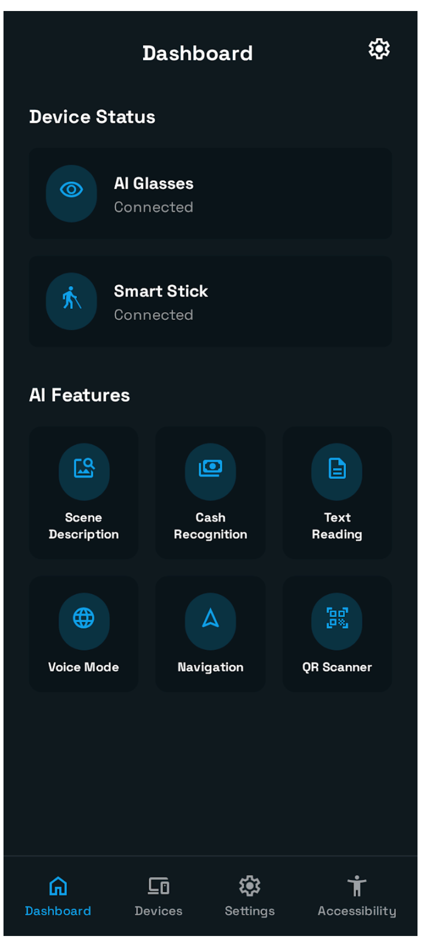
Smart phone : Smart glasses are connected through wired to smart phones that serve as the primary computing device to run deep learning models , neural networks and AI models.

Audio output : provides the audio feedback to the user through a small speaker or earphone to guide a visually impaired person.

Ergonomic frame : light weight frame designed to provide a comfortable experience for everyday long uses.

C. Software Modules

Samriddha Drishti AI glasses uses multiple AI software modules that provide realtime assistance , scene description and contextual understanding to visually impaired persons in real time . It uses the latest techstack like tensorflow light, OpenCV for deep learning and computer vision tasks. The major functional module includes :



- Identifying objects and their location : Uses convolutional neural networks(CNNs) to identify obstacles and multi layered decision making systems to provide realtime detailed audio feedback to enhance the safety of visually impaired persons and know the approximate location of obstacles.
- Scene Description and Visual Question Answering : uses mobile human vision language model to provide image captioning .
- cash recognition : uses image segmentation on pre trained currency dataset
- Reading and summarizing text : uses Deepseek OCR that can run offline on mobile devices and capable of text summarization, smart question answering , understands graphs, formulas , pie charts and figures etc.
- facial recognition : uses machine learning algorithms to detect the faces of known ones.
- QR bar code Scanning : Integrated Modules Use Computer Vision Library to detect and decode product or information tags.
- Text-to-Speech (TTS) Feedback System: Samriddha Drishti AI glasses uses native Text-to-Speech to convert the output text generated from an AI model into natural sounding audio feedback that the visually impaired person can easily understand in real time.

D. Data Flow Diagram

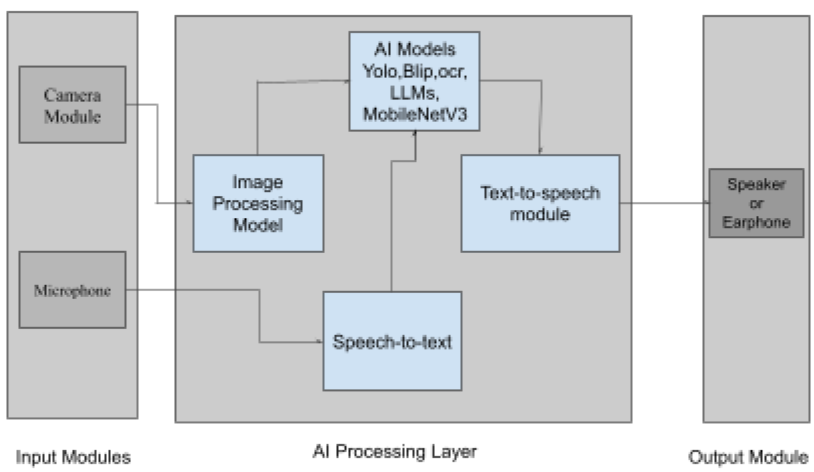


Fig 1 : Block Diagram

IV. Data Acquisition and Preprocessing

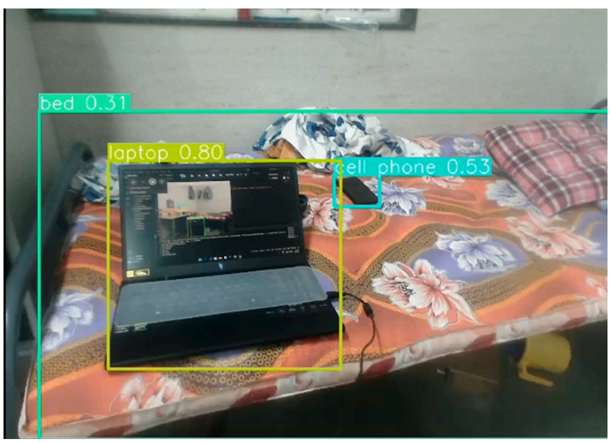
A high-definition camera on the glasses captures real-time frames. Each frame is processed using OpenCV that adjusts the contrast and brightness and enhances for low and high-light environments that ensure clear vision with advanced OpenCV algorithms so it can generate more accurate and reliable responses. This Document [10] shows the effective implementation of automatic adjusting brightness of an image with OpenCV . The same concept can be implemented for realtime video processing for more reliable and accurate prediction . Audio data is captured through microphone on glasses and transferred it to a smart phone and then its converted into text.

Methodology

Samriddha Drishti uses a multilayer deep learning system, convolutional neural networks (CNNs), computer vision , light weight human vision language models and LLMs , in a single wearable assistive device to help visually impaired persons to do day to day tasks. It uses flask as a backend for handling APIs for scene description and summarizing documents. AI models run using pydroid3 and tensorflow lite on smart phones. The goal was to design comfortable, lightweight wearable assistive glasses that can provide seamless real time guidance and help visually impaired people to do day to day tasks independently .

A. Indoor & outdoor Navigation

The navigation module integrates computer vision and GPS based guidance to help visually impaired persons to navigate safely and independently in an outdoor environment . Yolovs was used for recognizing and classifying objects due to its high speed and accuracy on low end devices . As mentioned in an article by the author. A multi layer decision making system is used to guide visually impaired people through speech output . It detects the object and position like left , right or center ,and approximate distance of object and count of the object for effective navigation in both indoor and outdoor environments . Yolovs was fine tuned on a custom dataset for common real world object detection and navigation . The bounding box and class labels are converted to audio feedback . for example : “Two persons detected far to the right side “ . The model was optimized into tensorflow lite for faster and efficient on-device mobile execution .



For outdoor navigation samriddha drishti uses phone GPS to estimate real-world positioning and estimates the distance between two locations. It detects obstacles and guides users with audio feedback for safer and precise navigation in outdoor environments. For example : “You’re on the right path. Turn left after 20 meters.”

B. Scene Description and Visual Question Answering

Initially For getting context of the environment, the BLIP (Bootstrap Language Image Pretraining) model was used to provide offline image captioning and visual question answering. Now we have adopted a lightweight Mobile Vision-Language Model (MobileVLM) which is highly optimized for mobile hardware using tensorflow lite.This significantly improves the speed and power consumption while maintaining reasonable accuracy. And when a smartphone is connected to the internet, the Gemini vision API is used to enhance response with detailed multilingual and semantic understanding.

Example :

User : “What is in front of me ? “

BLIP : “ Two people are standing near a car “

Gemini Vision : “Two individuals are standing near a white car parked in front of black gate. ”

This dual setup helps in both offline accessibility through MobileVLM and enhanced, high-quality responses via Gemini Vision when connected to the internet .

C. Reading and Document Summarization

samriddha drishti AI glasses can read printed text and handwritten notes, signboards. It uses DeepSeek-OCR, the latest next-generation optical-language model that combines **optical context compression** with **Mixture-of-Experts (MoE)** architecture. Combines image processing methods like binarization, contrast enhancement, and noise reduction to increase accuracy in different lighting environments. Once the text is processed, the system can read it instantly using native text-to-speech module for offline mode. It uses an open-source Deepseek OCR to generate a summary of the document trained on more than 30 million pdf documents. User questions about the text, and it simplifies complex information and provides processed text. This module helps visually-impaired people to understand, and interact with textual information from various real-world sources like documents, books, menus, signboards, charts , graphs formula etc.

Example :

User : “Summarize this document in one line .“

Output : “This document emphasizes the core features of samriddha drishti AI glasses that can help visually impaired persons to navigate and explore their surroundings independently. ”

D. Cash Recognition

This module helps visually impaired people to identify and handle currency notes independently in day to day cash payment. The module captures an image of currency notes using a camera module and classifies it using a MobileNetV3-based image classification model that is trained on Indian Currency Dataset. The model identifies the denomination of the currency note and recognizes its value.

The MobileNetV3 model was used because of its lightweight architecture, high accuracy and low power consumption in mobile devices. It effectively distinguishes between different currency notes such as Rs.10, Rs.20, Rs.50, Rs.100, Rs.200 and Rs.500.

The Identified currency notes are converted into speech through a text-to-speech system. For example, when a Rs.100 note is detected, the system says "Hundred Rupees Note Detected ." This module provides financial independence and confidence among visually impaired people by ensuring fast and accurate currency recognition in the real-world environment.

.

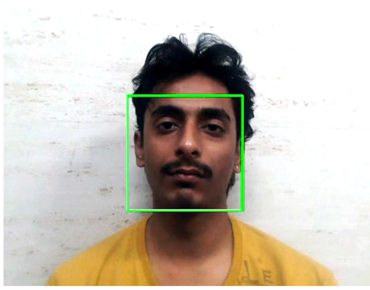
E. Face Recognition

Face recognition module helps visually impaired persons to recognize the faces of known people that are saved locally on device. Face recognition is implemented usingDlib's facial encodings and face_recognition python library.

During the operation, the camera captures facial features and compares them with locally saved encodings on the device to detect people. This feature completely works offline, ensuring privacy and fast response.user can toggled between different modules via touch command.

Example:

The photo of Abhinav was taken beforehand and saved in the device,when the camera was directed towards Aadil the following result was obtained.



Output → "I see Abhinav."

This module helps visually impaired people to recognize the faces of known people and increase social interaction and confidence.

F. QR code and bar code scanning

Samriddha Drishti includes a QR and Barcode scanning module that reads the information stored in QR and bar codes. It can be used to identify products , reads the QR codes on medical boxes and retrieves embedded digital information. . it is done using the pyzbar python library that provides the interface to

Zbar barcode reader library.it helps to detect and decode various types of one-dimension barcodes and QR codes from both live video and images in real-time.

The retrieved text information is converted to speech through the text-to-speech system that helps visually impaired people to identify products and its details without relying on others.

G. Multilingual Voice Interaction

Samridddha Drishti glasses are designed to help people from diverse communities to understand and experience the world in their own language . This system ensures that people from any background can use this device in their own local language .

This is achieved by using a multilingual speech interface :

- speech-to text : Implemented using Vosk for offline recognition.
- Text-to-speech : uses android native TTS engine , supporting Hindi , English and Nepali languages.

By adding Multilingual interaction , Samriddha Drishti ensures that language is never a barrier to use technology.

IV. System Integration

Samriddha Drishti glasses uses both hardware and software into a unified assistive device to provide real-time guidance to visually impaired people. The combined system ensures visual and audio data captured , processed , and converted into real-time audio guidance for seamless user interaction.

A. Overall Integration Architecture

This system is designed using a three-layered architecture:

- Input Layer: capture real world data.
- AI Processing Layer: perform all the computation and decision-making.
- Output Layer: converts processed information into speech output.

Samriddha Drishti glasses are connected to a smart phone via a wired interface (USB/OTG) which serves as the primary computing device. This system allows to run AI models on smart phones and make this assistive device cost-effective and lightweight.

B. Input Layer

The input layer consist of :

- Camera Module : High resolution camera Captures live video feed for identifying object , face recognition , cash recognition and images are captured for reading , summarizing text and scene captioning .
- Microphone : Captures the audio data from the user and sends it to the smart phone for processing users requesting information.
- Connection to Smartphone: The Camera and Microphone sends data to smartphone in real-time via a Wired connection to ensure minimum latency .

The smartphone acts as a central hub that runs AI models and executes the user requests. Manages module activation based on users requests .

C. AI Processing Layer

The AI processing layer works mainly on the smartphone using Pydroid3 and TensorFlow Lite (TFLite) to run models on mobile devices and low - latency execution. it integrates multiple AI models:

1.Indoor & Outdoor Navigation (Object Detection & Positioning):

- YOLOv5s detects and classifies objects (like vehicles, furniture etc.), obstacles (like potholes) and people in real-time.
- For outdoor navigation , the smartphone google maps and GPS module provides route guidance.
- Audio feedback guide and alerts visually impaired persons about the object type,position and approximate distance e.g " “Two persons detected far to the right side.”

2. Scene Description & Visual Question Answering:

- MobileVLM provides offline image captioning.
- When connected to the internet, Google Gemini Vision API enhances semantic understanding and multilingual support.

3.Reading & Document Summarization:

- Tesseract OCR extracts text from handwritten and printed documents.
- preprocessing includes noise reduction contrast and brightness adjustment.
- LLaMA 3 (1B parameters) generates summaries and answers questions about the text.
- Text is converted into audio feedback for real time support.

4. Currency Recognition:

- MobileNetV3-based image classification model classifies currency notes captured via the camera on glasses.
- Recognized currency notes are converted into speech e.g : “ Hundred Rupees Note Detected.”

5. Face Recognition:

- Uses dlib face encodings and face_recognition python library.
- captured faces are matched with a local database. it works completely offline for privacy.
- Example: “I see Aadil.”

6. QR & Barcode Scanning:

- Uses the pyzbar python library that provides the interface to Zbar barcode reader library
- Converts extracted information from QR or barcode to speech to inform the user about products or medications information.

7.Multilingual Voice Interaction:

- Vosk engine provides offline speech-to-text recognition.
- Android native Text-to-Speech (TTS) engine provides audio output in English, Hindi, and Nepali.

D. Output Layer

The processed text data from AI models are converted into real time audio feedback delivered through small speakers or earphones connected to the smartphone. This system ensures that user can receive immediate guidance.

E. Data Flow and Module Coordination

The system follows an event-driven workflow:

- The user activates the module they want to use via a button on smart glasses.
- Live input is captured through the camera and microphone.
- Input data is transferred to the smartphone for preprocessing.
- Relevant AI module is activated based on user command
- The AI model processes the data locally (or online for advanced queries).
- Text output is converted into audio feedback and delivered via speaker to the user.
- Temporary data (images, frames) are deleted automatically to respect users' privacy .
-

Advantages of Integration

- Cost-Effective:
- Offline Functionality:
- Real-Time Performance:
- Scalable:
- Privacy & Security:

Hardware-Software Co-Design

- Smart Glasses: Lightweight frame, camera, microphone, speaker/earphones.
- Smartphone: Runs AI models, TTS engine, and acts as a central computing hub.
- Connection: Wired USB/OTG connection ensures reliable and fast data transfer.
- Software Modules: Python, OpenCV, TensorFlow Lite, Pydroid3, Vosk, pyzbar, LLaMA, MobileVLM , Tesseract OCR.

Together , this integration ensures reliable , seamless and real time guidance for visually impaired person.

Performance Evaluation

Ethical Considerations

All captured videos and speech input are processed locally on device to maintain data privacy. No visual data or audio is transmitted or stored without users concern. The system is designed to increase user safety and inclusivity. None of the videos are stored in an online database. Temporary image storage is only used for real-time processing and automatically deleted after use. No any personal data is saved in cloud database. this system ensures the safety and privacy of users data .

Future Integrations

References

1. *Design and Implementation of Cash Recognition System for Visually Impaired People*. IJSRSET. Retrieved from https://ijsrset.com/index.php/home/article/view/IJSRSET25122194/IJSRSET25122194
2. Sharma, P., & Tiwari, A. (2020). Smart Currency Recognition System for Visually Impaired Persons. Saudi J Eng Technol. Retrieved from https://saudijournals.com/media/articles/SJEAT_106_270-276.pdf
3. Kumari, P. A., Devi, T. M., Aathira, B., & Devika, V. (2024). A Smart Currency Recognition System for Visually Impaired People. IJIRT. Retrieved from https://ijirt.org/publishedpaper/IJIRT176457_PAPER.pdf
4. Praneeth, P. K., Bhargav, B. R., Lokesh, K., & Sastry, A. V. S. (2024). Currency Recognition System for Visually Impaired using Deep Learning. PMC. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC12178407/
5. Sahu, H., & Rathore, P. (2025). Smart Currency Detection System for Blind People. IJCRT. Retrieved from https://ijcrt.org/papers/IJCRT2503499.pdf
6. Anu, A., Devi, G., Geetha, K., Prasanth, P. K. (2023). AI-Powered Smart Currency Recognition System for Visually Impaired. In Proceedings of the International Conference on Applied Intelligence and Computing (AIC 2023) (pp. 1–9). ACM. doi:10.1145/3613904.3642211
7. Patil, P. M., & Rane, S. A. (2023). Smart Currency Recognition System for Visually Impaired People. IJARSCT. Retrieved from https://ijarsct.co.in/Paper11233.pdf
8. Intel. ($\text{n.d.} $). Using AI to Help Visually Impaired People Identify Cash. Retrieved from https://www.intel.com/content/dam/develop/external/us/en/documents/using-ai-to-help-visually-impaired-people-identify-cash.pdf

9. Shinde, S. D., Patil, S. S., Patil, P. P., & Shinde, S. P. (2024). Smart Currency Recognition System for Visually Impaired People. IJIRT. Retrieved from https://ijirt.org/publishedpaper/IJIRT182948_PAPER.pdf
10. https://stackoverflow.com/questions/57030125/automatically-adjusting-brightness-of-image-with-opencv