Sight Companion: App for People with Visual Impairment

Kolanu Varun
CSE, IIT (BHU)
Varanasi, Uttar Pradesh, India
kolanu.varun.cse22@itbhu.ac.in

Manvi Chauhan
CSE, IIT (BHU)
Varanasi, Uttar Pradesh, India
manvi.chauhan.cse22@itbhu.ac.in

ABSTRACT

People with visual impairments (PVIs) face a great challenge in navigating the world. They deserve to be included in the society like everyone. So, our app "Sight Companion" is a novel mobile application in Flutter, which helps users to know their environments smoothly. This paper describes the app's features like Object detection, Reading assistant, Prompt Navigation, Color Detection, QR Scanner, Text to Speech, their technologies involved and its unique contributions for enhancing the daily life experience of PVIs.

We have integrated many AI models in the app by leveraging flutter's capabilities and by providing the intuitive, easy UI for PVIs to increase their confidence in daily life. Experiments were conducted and our results support the accessibility and accuracy of the application furthermore.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**.

KEYWORDS

Visual Impairment, Object Detection, OCR, Color Recognition, QR Scanner, Document Reader, Blind, App Development, Flutter

ACM Reference Format:
Kolanu Varun and Manvi Chauhan. 2024. Sight Companion: App for People with Visual Impairment. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

People with visual impairment (PVIs) perceive the world around them in a whole different way than those who have working eyes. Eye is unarguably one of the most important organs in a human body. According to the World Health Organization (WHO), an estimated 191 million individuals globally face difficulty with visual impairment, with 32 million among them classified as blind [12]. Despite this huge number, the promise of social inclusion by the internet [15] is mostly overlooked when it comes to PVIs. They face difficulties in getting socialized, reading text, identifying objects, knowing colors in front of them and many more. Still many use white canes [20] and similar traditional methods, hoping for a better alternative approach to rely upon. These can't help them know what objects exactly they are interacting with.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

There are many electronic aids [13] made to help PVIs, but many are costly and some not so effective. Though they need help in some situations, many of the people show sympathy to show they're "helping" the blind, which seems to them condescending. As most of the people have smartphones with them, just providing them an app with AI embedded in it, which can act as their own eyes is cheap and effective, giving them confidence to not depend on anyone else. So, there are many AI mobile applications developed in the motivation of helping PVIs. These changed the life of PVIs and boosted confidence. But these lacked in some areas, which started us thinking of developing our own app.

In this paper, we propose a novel application developed in Flutter and discuss its unique features in detail. The features include detecting objects, reading out the text, determining the colors, scanning QR codes, text-to-speech conversion and all this with an efficient navigation system using speech recognition. Section 2 discusses the already existing research and solutions in this area. Section 3 describes our design and implementation of the application. Section 4 discusses the evaluation and experiments of solutions and Section 5 concludes the whole article while discussing the future works.

2 RELEVANT WORK

The research and experimentation to help the blind was an active topic among the people who understood the difficulties of them. Electronic gadgets as discussed in articles like [6, 9, 11, 14] are one of the solutions to help the PVIs, but many of them are either not accessible easily or costly to afford.

As smartphones have become an inseparable part of our life, developing an application using AI technologies is cost-effective and easy to access. According to the survey [8], more than 90% of PVIs find special mobile applications accessible and 95% people find them useful. This survey [22] analyzes different AI solutions that have been developed. Mobile applications using AI were developed. App BeMyEyes [4] was developed to provide confidence to the blind, telling that they're not alone and there is still humanity left in the form of volunteers. In this app, PVIs contact the volunteers through video calls to get assistance, and keeps people connected through their stories, blogs etc.

Seeing AI [7] was an AI application developed by Microsoft for the blind, which integrated AI technologies to help PVIs read the text in front of them and identify objects. This became pretty popular and many found it useful. The app [17] developed a solution for image recognition and obstacle detection with a smartphone. Numerous apps like Seeing Assistant Home, Dictate, Envision AI, One step Reader etc have been developing solutions to help the blind by incorporating one or two features like from Object detection, Text Scanning, Text to speech, Easy UI etc

These apps though had a keen motivation of providing the best solution, lacked in certain aspects. For example, BeMyEyes was

a successful attempt, but volunteers wouldn't be available every time someone needs assistance, and the results which PVIs get may be flawed, depends basically upon the sympathy and morality of volunteers. Hence we created a platform where users can rely on the app to detect the objects in front of them, reducing the human errors. The remaining features in other apps like Object Detection, Text reader etc were appreciable, but they all missed a point that, if a blind can't see what they are seeing, how will they even able to see what's on the screen to use the proposed features? Talk back and voice over can be helpful to an extent, but moving their fingers continuously to find the needed feature is tiring. So, we have introduced the idea of navigating through speech. The user asks app what they wish to get, app does it. Pretty simple right! And we have got the right mixture of what PVIs need in a single app which every app is trying to do individually, also adding unique features of ours like Color detection, speech navigation etc.

3 APPLICATION AND FEATURES

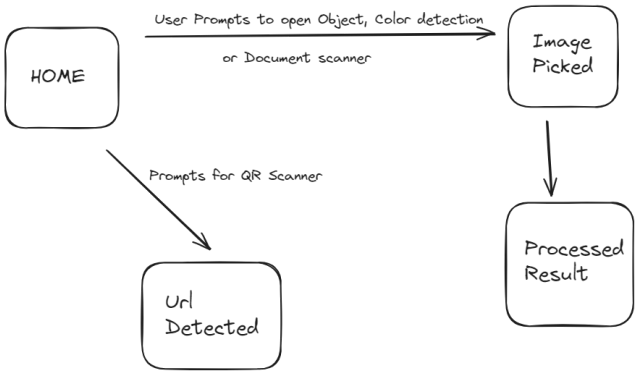

Figure 1: Workflow

The following features are available on the current SightCompanion app:

3.1 Feature Extraction

The app takes various user inputs to work. For the features Object Detection, Color Detection and Document Reader, image of the scene in front of user is taken through camera of the smartphone (Fig. 2). Before opening the camera, user is informed that the camera is being opened, so please capture the picture and tap on the tick mark on the bottom right. The tapping of buttons of camera can be done with help of Talk Back or Voice Over. This extraction of images is done with the help of flutter package **image_picker** and pre-processed before feeding to the model.

App takes speech from the user as input, for whole navigation of the app. This is discussed as a separate feature, Speech navigation.

3.2 Speech Navigation

This feature gives a unique taste to the app, making the UI/UX of the app to stand out among remaining applications in the eyes of PVIs. The user can navigate through the whole app by giving prompts. They can tap anywhere on the screen to start or stop speaking, and

Figure 2: Image Collection

trigger a feature to be used using certain keywords in their speech. For example, speaking "Open object detection" activates the object detection feature, "Repeat Instructions" repeats the instructions for using the app etc. The instructions to use app are shown on the home page of app (Fig. 3) .

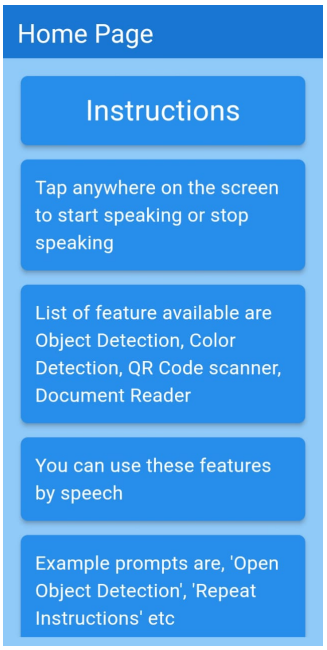

Figure 3: Home Screen

Keywords to trigger:

(1) **Object Detection**: *object*
(2) **Document Reader**: *document, ocr*
(3) **Color Detection**: *color, colour*
(4) **QR Code Scanner**: *qr, barcode, bar code*
(5) **Repeat Instructions**: *instruction*

The workflow of this feature is, speech is converted into text (using the package **flutter_stt** which gives the functionality of Speech To Text (STT) [18]), if keywords are detected in the speech, the correesponding feature is used.

3.3 Object Detection

Object detection is a major feature of Sight Companion that enables users to identify and recognize objects in their surroundings. By simply pointing their smartphone camera at a location and capturing the image, users can receive audio feedback about objects present in front of them. This feature is particularly useful for navigating unfamiliar environments and performing everyday tasks independently (Fig. 4). The objects in front are spoken out loud, informing the users about potential information they wouldn't want to miss out.

Figure 4: Object Detection

Sight Companion utilizes a state-of-the-art object detection model based on **Yolo** (You only look once) V2 tiny architecture [10] for this purpose. The Yolo model is trained on a diverse dataset of objects commonly encountered in daily life, allowing for accurate detection across various scenarios. Yolo (Fig. 5) is chosen as the model for our app because of its high accuracy, and speed.

The app uses the TensorFlow Lite framework for real-time inference on mobile devices, ensuring low latency and efficient utilization of hardware resources. The package **tflite_v2.** in flutter is used to integrate the YOLO model in the app.

3.4 Color Detection

Color detection is another important feature of Sight Companion that helps users identify and differentiate colors in their environment. Selecting clothing, matching accessories, or sorting items, color recognition provides the best assistance in daily life (Fig. 6 and 7).

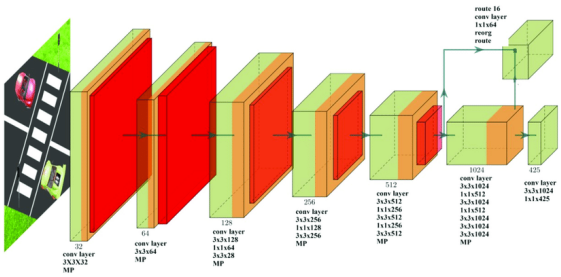

Figure 5: YOLO Architecture

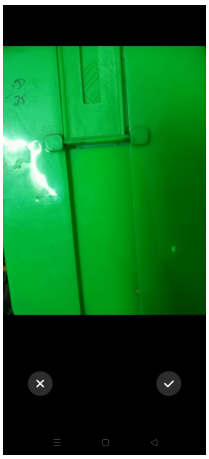

Figure 6: Image for Color Detection

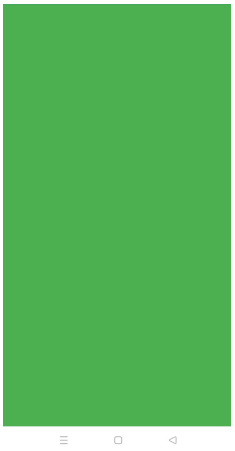

Figure 7: Color Detected

The **palette_generator** package in Flutter is utilized to obtain the RGB value of the dominant color. This package uses K-Means clustering and color quantization methods to get the dominant color in image as RGB. However, since this value cannot be directly mapped to a known color [3], we have selected specific known

colors such as Black, Brown, Yellow, Green, etc., and stored their RGB values in a map.

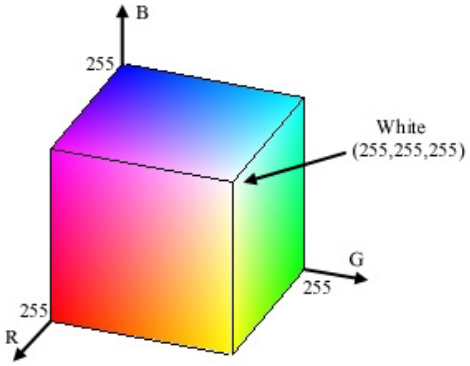

Figure 8: RGB Colors in 3D Space

To translate the obtained dominant color into a known color, we leverage the understanding that these RGB colors can be extrapolated into a 3D environment. Each RGB triplet corresponds to a unique point in this 3D space. We already possess static points for known colors in this 3D space.

To determine the closest known color to the obtained dominant color, we calculate the distance between their respective RGB values. The color with the least distance is considered the required color. This distance calculation is based on the Euclidean distance formula:

$$d = (r - r_i)^2 + (g - g_i)^2 + (b - b_i)^2 \quad (1)$$

Where r, g, and b represent the RGB values of the dominant color obtained from the package, and r_i, g_i, and b_i represent the RGB values of the known colors chosen.

3.5 Reading assistant

The Reading assistant feature of Sight Companion helps users to access printed text from documents, newspapers, books, and other sources. By simply capturing an image of the text (Fig. 9) , the app performs optical character recognition (OCR) and converts the text (Fig. 10) into audible speech, helping users to listen to the content within no time.

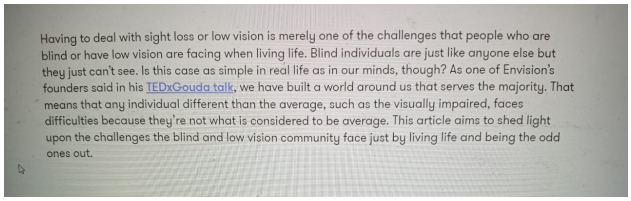

Figure 9: Document

Sight Companion uses the Google's ML Kit [19] for OCR, which provides accurate text recognition use case. The ML Kit's models are efficient for mobile devices, ensuring fast and accurate performance even in low lighting conditions. **google_mlkit_text_recognition** package in Flutter is used for this purpose.

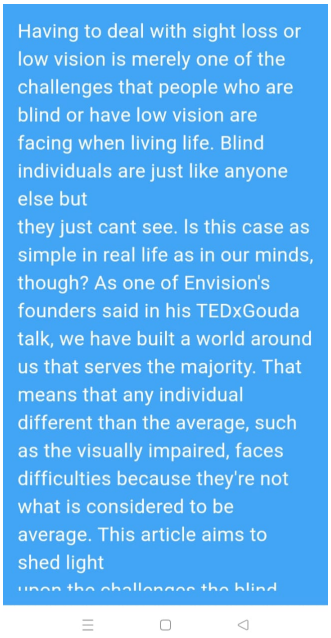

Figure 10: Text Extracted

3.6 QR Scanner

The QR scanner feature of Sight Companion allows users to scan QR codes [21] and access embedded information, such as website URLs, contact details, and product information. By pointing their smartphone camera at the QR code (Fig. 11), users can instantly fetch the embedded data and do useful actions. If the embedded data is a valid url, the app opens it directly, else, it reads out that text by showing it on the screen.

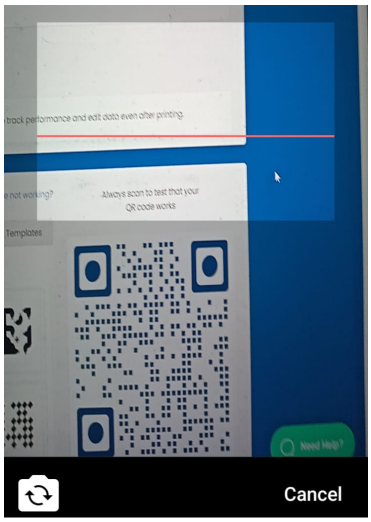

Figure 11: QR Scanner

Table 1: Some of Object Detection Results

S.No.	Real Object	Prediction
1.	Person	Person
2.	Bicycle	Bicycle
3.	Cat	Cat
4.	Chair	Chair
5.	Handbag	Vase
6.	Tv Monitor	Tv Monitor
7.	Cell Phone	Cell Phone
8.	Tv Monitor, Chair, Person	Tv Monitor, Chair
9.	Bicycle, Person	Bicycle, Person
10.	Potted Plant, Fork	Potted Plant, Spoon
11.	Cell Phone, Mouse	Remote, Mouse

The **simple_barcode_scanner** package from Flutter was used for QR code scanning, which offers fast and accurate detection of QR codes.

3.7 Text To Speech (TTS)

The Text To Speech (TTS) [5] functionality of Sight Companion converts text into speech, allowing users to listen to information such as object descriptions, color identifications, text extracted from images and the urls scanned from QR codes. This is the backbone of the app, making the users life easier by telling that "Hey! you are have low vision? Don't worry.. We will make use of your ears to make the life easier".

The app integrates the **flutter_tts** package for text-to-speech, providing speech output. This package acts as a bridge between Android's Text To Speech API and Flutter.

4 EXPERIMENTS AND EVALUATION

After performing extensive evaluations of the app, the following are the results:

4.1 Performance of Object Detection

Most of the objects which were scanned alone, were detected correctly, and gave some errors when were captured along with other objects (Table 1).

Though, most results were detected pretty accurately with 80% accuracy in daily lives and will help the PVIs significantly.

4.2 Performance of Reading assistant

Document Scanner scanned smaller to bigger texts accurately in no time and were spoken out loud clearly. This will help greatly in reading bill boards, newspapers, articles etc. Though, the correct pronunciation of punctuation was missed sometimes and feature of different language detection is to be implemented yet. Now it works for English accurately.

4.3 Performance of Color Detection

Lighter colors were more accurately detected with nearly 90% accuracy and darker colors with 70% accuracy. It depends on lighting conditions. If the lighting is darker, it was biased towards darker colors.

4.4 Performance of QR Scanner

This correctly scanned every QR and opened url, if valid, or else, read out the text embedded.

4.5 Performance of Text to Speech Conversion

This was perfect in most of the times. But some older versions (>4years back) of smartphones may face issue because, it was getting disconnected to API because of resource management race conditions.

5 CONCLUSION

This research paper introduced a novel mobile application using AI technologies incorporating features like Object detection, Reading assistant, Prompt Navigation, Color Detection, QR Scanner, Text to Speech. We are excited to continue enhancing the functionality of the app. Currently, one area we aim to improve is the navigation system for users by detecting obstacles. There are two potential methods for implementing obstacle detection: Stereo Vision and Monocular vision.

Stereo Vision requires two cameras positioned with a known baseline distance between them to calculate depth information. However, not all phones have dual cameras, so we plan to implement Monocular vision instead. In Monocular vision, we will utilize a depth estimation model such as MIDAS [2] or Zoe Depth [1] to generate a depth map from a single camera input. We will then use the depth map along with the SLIC approach to calculate the position of obstacles in a 3D environment, similar to the method discussed in [16] and navigate the users smoothly in their environment.

Additionally, we are planning to integrate navigation through maps in the real world by using Google Maps. This feature will provide users with guidance and directions to navigate their surroundings efficiently.

We are committed to continually improving the app's functionality to enhance the user experience and address the needs of individuals with visual impairments.

REFERENCES

[1] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. 2023. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288* (2023).

[2] Reiner Birkl, Diana Wofk, and Matthias Müller. 2023. Midas v3. 1–a model zoo for robust monocular relative depth estimation. *arXiv preprint arXiv:2307.14460* (2023).

[3] Alphonse Chapanis. 1965. Color names for color space. *American Scientist* 53, 3 (1965), 327–346.

[4] Sebai Mohamed Bachir Dassa Elhousseyne. [n. d.]. An Investigation into Visually Impaired Students and Teachers' Attitudes towards the Use of 'Be My Eyes' Application The Case of Pupils at Taha Houssine Middle School-Biskra. ([n. d.]).

[5] Thierry Dutoit. 1997. *An introduction to text-to-speech synthesis*. Vol. 3. Springer Science & Business Media.

[6] René Farcy, Roger Leroux, Alain Jucha, Roland Damaschini, Colette Grégoire, and Aziz Zogaghi. 2006. Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view. In *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion*, Vol. 12. Citeseer, 12.

[7] Christina Granquist, Susan Y Sun, Sandra R Montezuma, Tu M Tran, Rachel Gage, and Gordon E Legge. 2021. Evaluation and comparison of artificial intelligence

vision aids: Orcam myeye 1 and seeing ai. *Journal of Visual Impairment & Blindness* 115, 4 (2021), 277–285.

[8] Nora Griffin-Shirley, Devender R Banda, Paul M Ajuwon, Jongpil Cheon, Jaehoon Lee, Hye Ran Park, and Sanpalei N Lyngdoh. 2017. A survey on the use of mobile applications for people who are visually impaired. *Journal of Visual Impairment & Blindness* 111, 4 (2017), 307–323.

[9] Sameer Grover, Aeysha Hassan, Kumar Yashaswi, and Namita Kalyan Shinde. 2020. Smart blind stick. *International Journal of Electronics and Communication Engineering* 7, 5 (2020), 19–23.

[10] Peiyuan Jiang, Daji Ergu, Fangyao Liu, Ying Cai, and Bo Ma. 2022. A Review of Yolo algorithm developments. *Procedia computer science* 199 (2022), 1066–1073.

[11] L Kay. 1984. Electronic aids for blind persons: an interdisciplinary subject. *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)* 131, 7 (1984), 559–576.

[12] Moncef Khairallah, Rim Kahloun, Rupert Bourne, Hans Limburg, Seth R Flaxman, Jost B Jonas, Jill Keeffe, Janet Leasher, Kovin Naidoo, Konrad Pesudovs, et al. 2015. Number of people blind or visually impaired by cataract worldwide and in world regions, 1990 to 2010. *Investigative ophthalmology & visual science* 56, 11 (2015), 6762–6769.

[13] Jin Liu, Jingbo Liu, Luqiang Xu, and Weidong Jin. 2010. Electronic travel aids for the blind based on sensory substitution. In *2010 5th International Conference on Computer Science & Education*. IEEE, 1328–1331.

[14] Aslam Muhammad, Muhammad Umair Ahmad Khan, Haseeb Azhar, Ali Masood, and Maliha Saleem Bakhshi. 2010. Analytical study of intelligent assistants to help blind people in avoiding dangerous obstacles. *Journal of American Science* 7, 8 (2010), 480–485.

[15] Robert Pepper, Enrique J Rueda-Sabater, Brian C Boeggeman, and John Garrity. 2009. From Mobility to Ubiquity: Ensuring the Power and Promise of Internet Connectivity… for Anyone, Anywhere, Anytime. *The Global Information Technology Report 2008–2009* (2009), 1943.

[16] Andra Petrovai, Arthur Costea, Florin Oniga, and Sergiu Nedevschi. 2014. Obstacle detection using stereovision for Android-based mobile devices. In *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)*. 141–147. https://doi.org/10.1109/ICCP.2014.6936966

[17] Tufel Ali Qureshi, Mahima Rajbhar, Yukta Pisat, and Vijay Bhosale. 2021. AI based app for blind people. *Int Res J Eng Technol* 8, 03 (2021), 2883–7.

[18] Rustam Shadiev, Wu-Yuin Hwang, Nian-Shing Chen, and Yueh-Min Huang. 2014. Review of speech-to-text recognition technology for enhancing learning. *Journal of Educational Technology & Society* 17, 4 (2014), 65–84.

[19] Anubhav Singh and Rimjhim Bhadani. 2020. *Mobile Deep Learning with TensorFlow Lite, ML Kit and Flutter: Build scalable real-world projects to implement end-to-end neural networks on Android and iOS*. Packt Publishing Ltd.

[20] Philip Strong. 2009. The history of the white cane. *Tennessee Counc. Blind* (2009).

[21] Sumit Tiwari. 2016. An introduction to QR code technology. In *2016 international conference on information technology (ICIT)*. IEEE, 39–44.

[22] Hélène Walle, Cyril De Runz, Barthélemy Serres, and Gilles Venturini. 2022. A survey on recent advances in AI and vision-based methods for helping and guiding visually impaired people. *Applied Sciences* 12, 5 (2022), 2308.