An Intelligent Voice Assistant Engineered to Assist the Visually Impaired

Rishabh Chopda, Aayan Khan, Anuj Goenka, Dakshal Dhere, and Shiwani Gupta

Abstract Visually handicapped people's lives are subject to a multitude of unrelenting challenges because they've been made bereft of the gift of sight. The proposed solution is a wearable Smart Voice Assistant that is developed to accommodate the needs of the visually impaired to aid them in every aspect of their everyday lives. It takes advantage of recent breakthroughs in the fields of language processing and computer vision to provide a broad spectrum of applications, including emergency response functionality, object recognition, and optical character recognition. It comprises hardware components that provide feedback in the form of sound, haptics, and speech to help with obstacle avoidance. The voice assistant also interacts with a smartphone application to enhance the user's experience by enabling them to read the messages from their phone, send an SOS message to their closest connections in an emergency, customize the device settings through the mobile application, and find the device with the press of a button if it is misplaced. The proposed solution will enable the user to live a life in relative safety and comfort, which is essential for people suffering from varying levels of visual impairment.

Keywords Voice assistant · SOS · Object avoidance · Object recognition · Optical character recognition

1 Introduction

As we approach a stage in human civilization where the average age of the population is increasing at an unprecedented rate, human physical functions are failing, and that visual faculties are declining at a behooving rate. Globally, World Health Organization (WHO) [1] that 43 million people are visually disabled, with another 295 million suffering mild to severe vision impairment. It was essential to create a gadget that assists them in traversing their environment and empowering them to do tasks that would otherwise be difficult or simply impossible.

R. Chopda (✉) · A. Khan · A. Goenka · D. Dhere · S. Gupta
Thakur College of Engineering & Technology, Mumbai, Maharashtra 400101, India

V. E. Balas et al. (eds.), *Intelligent Computing and Networking*, Lecture Notes in Networks and Systems 632, https://doi.org/10.1007/978-981-99-0071-8_12

The condition of lacking vision is referred to as blindness, which is caused by a physiological or neurological imbalance. Despite tremendous advancements in technology, blindness remains a serious problem [2, 3]. Researchers have been concentrating on this topic to produce helpful tools or aides for those who are blind or visually impaired. For blind people, very few assistive tools and devices are already available, however, the efficacy of their applications is fairly limited by speed, scope, and above everything, the manner they have been implemented in.

The previous works in the domain of assisting the blind have yielded fruition in various domains of the field. Functionalities related to OCR, fruit ripeness estimation, object detection, and identification, and navigation assist for the blind have been developed, albeit not necessarily to the point of utility, but to generate a solution to the many problems faced by the blind. S. M. Felix and the team [1] have created a mobile application that works to provide functionalities similar to the proposed system like voice assistant, OCR, and even image recognition. Safe navigation for the visually impaired [4, 5] has been worked on extensively using stereo-cameras [6] and GPS. Similarly, P. Bose [2], has developed a mobile-based application to work in assisting the blind to perform functionalities like speech recognizing and speech synthesis as the means of interaction through voice input to recognize text on a real-life object and provide audio feedback [7]. The related works [3, 8] are a raspberry pi based smart device to assist the blind by providing object identification functionality and even obstacle detection and avoidance system. This is an idea involving the aforementioned hardware and machine learning based software. The premise for common object detection and identification is of a prime importance from a safety and utility standpoint and there is an abundance of related work in the object detection using Computer Vision [9–11]. Iyear [12] has used the current technology to increase the number of visually impaired users navigating the internet like reading articles or listening to music on Youtube, etc.

Additionally, there are a variety of technologies available to help the sight impaired navigate both indoors and outdoors. All of these devices rely solely on the Global Positioning System (GPS) to determine their location to navigate your way around. In reference [13], the paper offers a system that makes use of stereo vision a sonification approach and image processing methodology Support navigation for the blind. The system that has been created includes stereo cameras as vision sensors and stereo cameras as wearable computers. All of the earbuds are fashioned into a helmet.

To summarize, when considering the challenges of the visually handicapped, earlier attempts to solve this problem have focused on solving problems that have a limited range. Previous approaches have tended to solve only one key issue while leaving the others unaddressed. We hope to provide a one-stop solution to all the primary challenges that the visually impaired face with the proposed system. The proposed system not only assists the blind in walking by avoiding obstacles, but also allows them to read the newspaper, determine the maturity of fruits, and ask for assistance in an emergency. It is a significant improvement over prior art in that it addresses problems that were previously unresolved, provides a better user experience, and happens to be multifaceted.

2 The Proposed System

This paper describes a smart wearable voice assistant that leverages machine learning and deep learning to help blind individuals identify obstacles to help them in walking, also providing them with other recognition functionalities like facial and text recognition [14], which ultimately aims at decreasing the unfair challenges they encounter daily. When combined with Ultrasonic sensors, the device allows users to move freely without much caution as it alerts them of approaching impediments. Furthermore, the Voice Assistant will be used in combination with a companion app [15] that adds a slew of new functions to an already feature-rich device. The mobile app has a host of additional features that make the user's life easier. It uses the Voice Assistant to keep the user updated on recent activity on their phone, but it most importantly functions as an integral part of the SOS functionality, which involves sending a distress signal to the user's preferred emergency contacts.

Largely, the proposed idea can be divided into three major parts:

1. Wearable Wrist Device
2. Voice Assistant
3. Companion Mobile Phone Application.

2.1 Wearable Wrist Device

The proposed wearable wrist device is tailored to be as user-friendly as possible. The buttons are etched with symbols to make them distinguishable when felt by the user's fingertips. The device is engineered to handle all the rudimentary tasks of an individual's day-to-day life efficiently. Figure 1 shows the diagram of the proposed device that includes an embedded system module (Raspberry Pi 4 Model B) for handling all the computational tasks, an ultrasonic sensor that measures the distance to an object using ultrasonic sound waves wherein the user is alerted if an object lies within 40cms of the sensor's radius, a rechargeable LiPo battery to power the device, a Raspberry Pi camera for accommodating all the functionalities involving computer vision, five buttons with a distinct set of functionalities and a vibration motor to provide haptic feedback. The functions of the buttons are as followed:

- The Activate button is used to activate and deactivate the voice assistant.
- The Volume Up button is used to raise the volume of the voice assistant.
- The Power button is used to turn on/off the voice assistant
- The Volume Down button is used to lower the volume of the voice assistant.
- The SOS button is used to activate and deactivate the ultrasonic sensor when pressed once and activate the SOS functionality when pressed thrice.

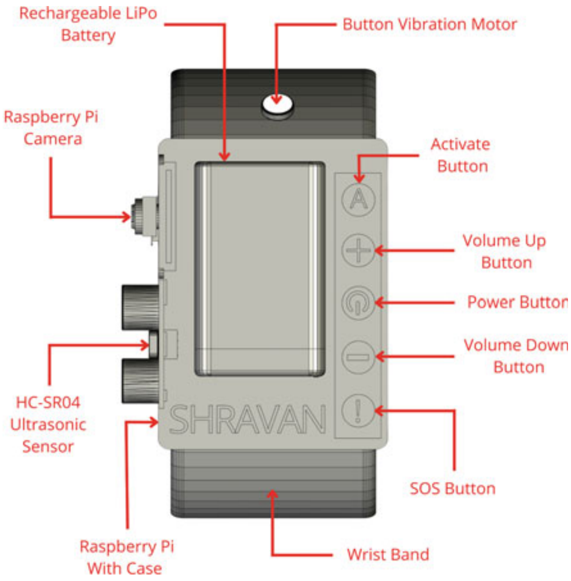


Fig. 1 Structure for the proposed device

2.2 *Voice Assistant*

The wearable wrist device includes Computer Vision-based functionalities like object recognition, fruit ripeness detection, and optical character recognition. After the user asks for the function through the voice assistant, the camera captures a snapshot of the desired subject, converts it to a vector, sends it to the central processing unit where it is identified, and the result is sent back to the user in audio format.

At the user's request, the voice assistant works in conjunction with its companion mobile application to read incoming messages from the paired phone. The voice assistant can perform the following key tasks in addition to the features listed above:

- Getting live cricket match scores by scraping data from the web using Beautiful Soup which is a Python package used for parsing HTML and XML documents.
- Describing weather conditions of the user's location by fetching the user's geolocation from the phone and getting the weather data for the desired location using the OpenWeatherMap API.
- Carrying out a quick Wikipedia search based on the spoken keyword and reading out the article summary of the keyword using the Wikipedia python library which makes it easy to access and parse data from the Wikipedia website.
- Updating the user with the latest news headlines by fetching data from the News API which is a straightforward and easy-to-use REST API that returns JSON search results for current and historic news articles gathered from a multitude of sources.

The results of the modules will be read out by the voice assistant upon the user's request.

2.3 Companion Mobile Application

In the age where staying connected 24/7 has become more of an unspoken societal norm than an option, Cell phones have become a necessity for many individuals all around the world. They are becoming increasingly instrumental for a large variety of reasons.

To give the Smart Voice Assistant a new dimension, we propose integrating a Mobile Application [16, 17] whose functionalities would communicate directly with the Voice Assistant Device's features. The mobile app also facilitates composing text messages without having to interact with the phone along with a feature for the user to locate the phone. Additionally, an on-the-fly settings configurator allow users to alter the Voice Assistant's settings directly from the app. The SOS functionality [18] intends to help the user if they find themselves in any of the following situations: Component failure or traumatic emergency. When these conditions are detected, a message is sent to 4 pre-determined contacts of user making them prompt in a possible emergency.

Message enable connection with the consumer that's less intrusive in nature. To overcome this stumbling block of no vision, we've integrated the connected phone's messaging system with the Voice Assistant, allowing the user to summon the Assistant and request a message readout or compose a text message. Unfortunately, some of the information that your brain may consider unimportant may be required for you to remember where you put your keys, phone, or wallet, and if it has been erased, you will have to spend time attempting to locate some of your daily things. The feature of FindMyDevice helps in locating the device while it is not strapped to the user. Single button press, triggers the device to make sound for the user to locate it (Fig. 2).

3 Architecture and Algorithms

3.1 Optical Character Recognition

Optical character Recognition is a procedure that includes multiple sub-processes that must be completed as precisely as possible. The first subprocess in the process includes pre-processing the image for which we have used the OpenCV library. This library contains a lot of tools to help us pre-process the image with simplicity.

- Firstly, we read the image in a grayscale format which is the only format supported by the next two algorithms.

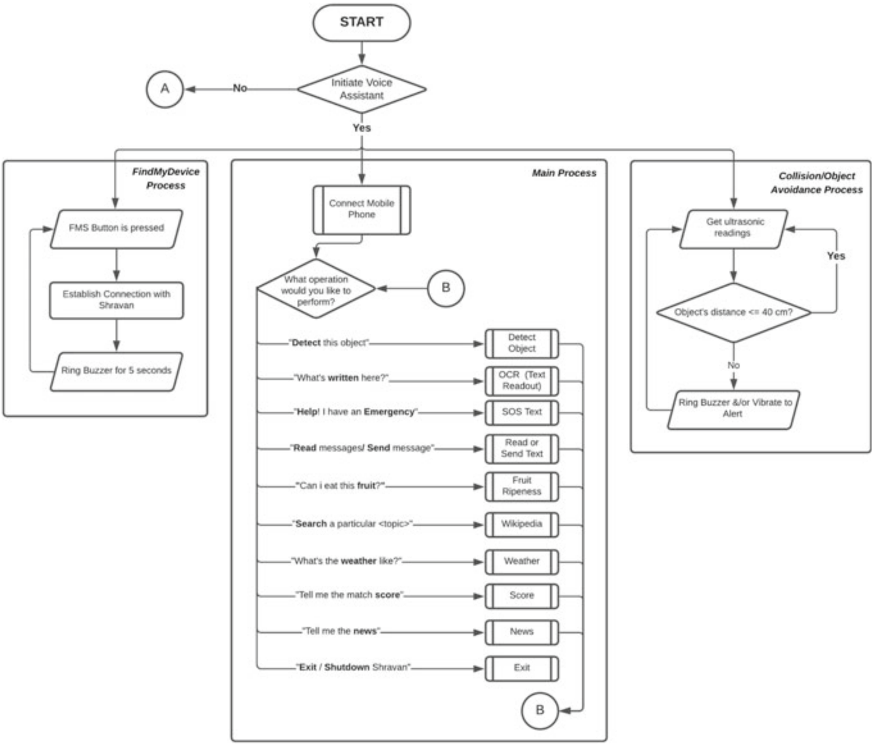

START
Initiate Voice Assistant
No
A
Yes
FindMyDevice Process
FMS Button is pressed
Establish Connection with Shravan
Ring Buzzer for 5 seconds
Main Process
Connect Mobile Phone
What operation would you like to perform?
B
"Detect this object"
Detect Object
"What's written here?"
OCR (Text Readout)
"Help! I have an Emergency"
SOS Text
"Read messages/ Send message"
Read or Send Text
"Can i eat this fruit?"
Fruit Ripeness
"Search a particular <topic>"
Wikipedia
"What's the weather like?"
Weather
"Tell me the match score"
Score
"Tell me the news"
News
"Exit / Shutdown Shravan"
Exit
B
Collision/Object Avoidance Process
Get ultrasonic readings
Object's distance <= 40 cm?
Yes
No
Ring Buzzer &/or Vibrate to Alert

Fig. 2 Flowchart depicting the voice assistant functionality

- Secondly, we use the bilateral filter algorithm [19] which is a picture smoothing filter that preserves edges while lowering noise. It uses a weighted average of intensity data from surrounding pixels to replace the intensity of each pixel. A Gaussian distribution can be used to calculate this weight. The weights are determined not just by the Euclidean distance between pixels, but also by the radiometric differences. Sharp edges are preserved as a result.
- Then we use a thresholding algorithm that converts blacks in the images to pitch black and white snow becoming white. We mainly chose a threshold value by calculating a mean of all the pixels in the image (pixels in a grayscale image range from 0 to 255), and any pixel above that threshold gets a value of 255 which is deepest black and anything below the mean gets a value of 0 which is lightest white.

The next subprocesses include text localization, character segmentation, character recognition and post processing. Tesseract OCR [20] was chosen to do the above mentioned subprocesses. This engine uses the Long Short Term Memory (LSTM) [21] network, which is a form of Recurrent Neural Network (RNN) [22]. To use the tesseract engine for our code we use pytesseract [23] which is a wrapper class for the

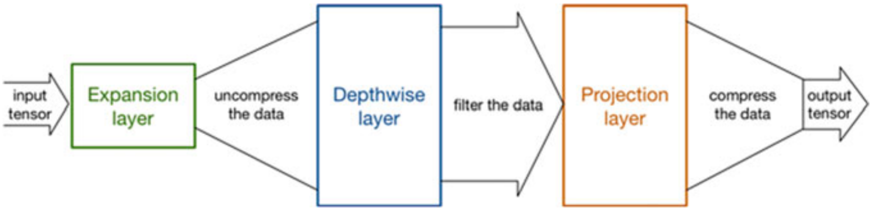


Fig. 3 MobileNetV2 architecture

Tesseract OCR Engine. Preprocessed image when passed through this library returns a text as an output using the text localization, character segmentation, character recognition and post processing subprocesses and the LSTM algorithm.

3.2 *Fruit Ripeness Detection*

The fruit ripeness detection model was trained on a dataset that was scraped from the internet using Selenium [24]. The scripts are executed by a browser-driver on a browser-instance on your device.

The first step is the data pre-processing. We will be going to use the ImageDataGenerator class in Keras. Images are converted to an input shape of 224, 224 since it is the acceptable input shape for our algorithm. The images are rescaled and divided by 255 which is mainly for normalization.

The next step is building the model for which we use transfer learning [25]. This improves the learning in the new task greatly. For this purpose, we will use the tensorflow hub to load a pre-trained MobileNetv2 [26] model (Fig. 3).

To this base layer of MobileNetV2, we add our global spatial average pooling layer, a fully connected layer and a logistic layer at the end. We use ReLu activation function in all the layers except the last one which is the logistic layer or the output layer. In that layer we use the sigmoid function which ret urns a probability for each class between 0 and 1. With 1 being the most probable class and 0 being the least (Figs. 4 and 5).

3.3 *Object Detection*

Object detection is a computer technology that deals with finding instances of semantic items of a specific class (such as individuals, buildings, or cars) in digital photos and videos. To obtain more accuracy, computer vision models are becoming deeper and more sophisticated. However, these advancements increase the size and latency of the system, making it incompatible with systems that are computationally challenged. MobileNet comes in handy in these situations. This is a model created

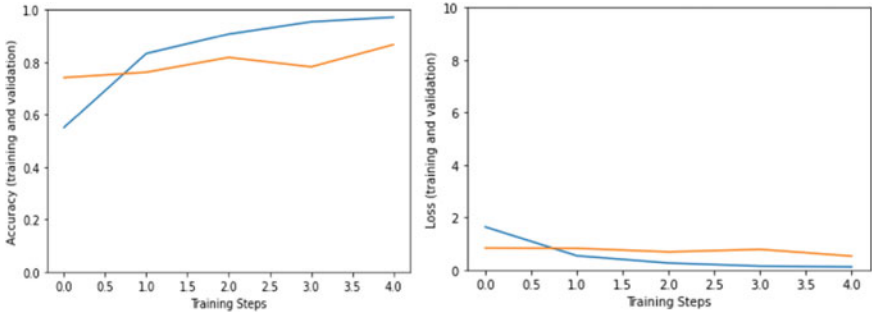


Fig. 4 Model training accuracy and loss chart

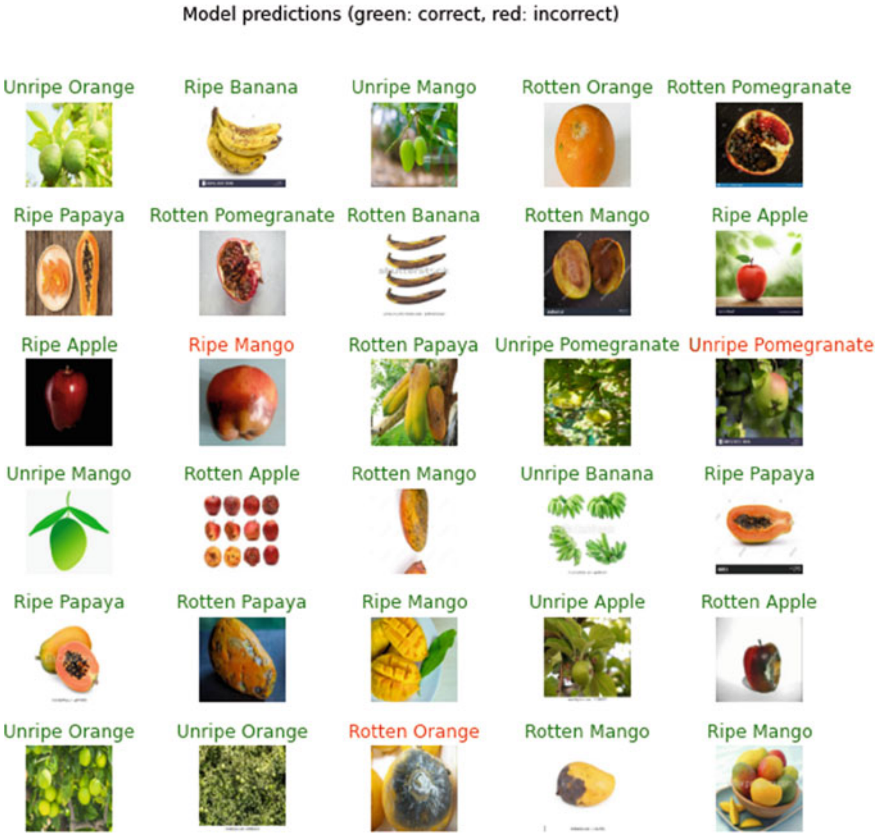


Fig. 5 Result of fruit ripeness detection model

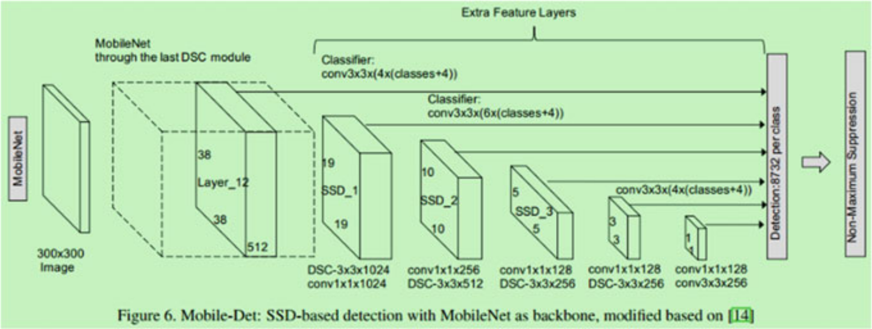


Fig. 6 MobileNetSSD architecture

primarily for high-speed mobile and embedded applications. Although it provided good frames per second on low computation, it lacked accuracy. To counteract this, we used the MobileNetSSD [27] model (Fig. 6).

MobilenetSSD is an object detection model that uses an input image to compute the bounding box and category of an object. Using Mobilenet as a backbone, this Single Shot Detector (SSD) object detection model may enable quick object detection optimised for mobile devices [28]. SSD requires only one shot to recognise many objects within an image, but RPN-based techniques such as the R-CNN [29] series require two shots, one for generating region suggestions and the other for detecting the item of each proposal. As a result, SSD is substantially faster than two-shot RPN-based techniques.

4 Result and Discussion

The proposed device will have modules ranging from object detection to voice assistant, from SOS functionality to fruit ripeness estimation and OCR. All of these would need to be implemented so that they work seamlessly with each other and be tailored to suit the needs of the blind. The voice assistant that is integrated into the device is a smart bot that is designed to answer user queries like the latest news, current weather, score of the current match etc.

The Fruit ripeness estimation works to determine whether a fruit is ripe enough or not and helps determine whether it's buyable. The machine learning module was made based on a supervised learning model. The OCR module has to be able to read printed text on hard copies in order to convert it into voice output for the blind user to listen to. Additionally, the sos functionality is an emergency alert feature that can be activated on the device. The voice assistant even enables the user to ask for weather related information from the voice assistant.

5 Conclusion and Future Scope

In this paper, we propose and smart wearable device wherein the system comprising: An Ultrasound sensor, used to detect the optical obstacles, objects and person during the walking of the person; A camera unit, used to read text using OCR, identify objects placed in front, identify faces of people; A feedback unit, used to alert the person on presence of the obstacles, articles and person in path of the person and also respond to the voice queries of the user including but not limited to; whether conditions, current time and location; A mobile application containing GPS to send information about location to the device, an SOS functionality that can be activated remotely via the Device to send current location of the user to saved contacts on the phone; and A processing unit used to process the information received from the ultrasound sensor and the mobile application, where the processing unit sends the processed information to the alert unit, and thereafter the alert unit sends the information of the obstacles, reads out written script taken from the camera, responds to user queries with computer generated voice, this forms the part of the OCR module of the proposed device. The proposed device on its own is sufficient and enough to enable the visually challenged to lead a comfortable and safe life and even provide an opportunity to go beyond hamstrung opportunities that blindness presents them with (Fig. 7).

One of the primary problems that the proposed device deals with concerns the safety [30] of the user and aims to direct help to the user in the event of an emergency. In view of that, events that involve the user stumbling or falling because an object or hurdle, for example, a raised platform, could not be detected by the proposed device, need to be considered. And should such events be fatal, the closest contacts should be immediately intimated of such occurrence right away so that their help can be directed. This paves way for a system that can detect if the user has stumbled or even fallen. This module can be called the Fall Detection module. Under the Fall Detection module, the device can identify an event and send an alert to predetermined contacts of the user that the user has fallen down and might be in need of help. To realize this module, hardware consisting of a fall detection circuit will have to be implemented. These circuits primarily consist of accelerometers that are a type of low-power radio wave technology sensor, to monitor the movements of the user. Some advanced systems may even consist of gyroscopes, infrared sensors, acoustic sensors, etc. to make the fall detection even more accurate. Once a fall is detected, the device will rely on the information to the companion mobile app of the proposed device. The mobile application can then alert the saved contacts of the user a location of the user along with a message informing them of the fall. Events like the user being involved in a car accident or any other road accident would also trigger the alert being sent. Such a functionality would add an extra level of security to the lives of the user in the case of post event damage control (Fig. 8).

Moreover, the other wider domains like navigation while avoiding obstacles, voice assistant, OCR can be further improved by working on their speed and accuracy. In

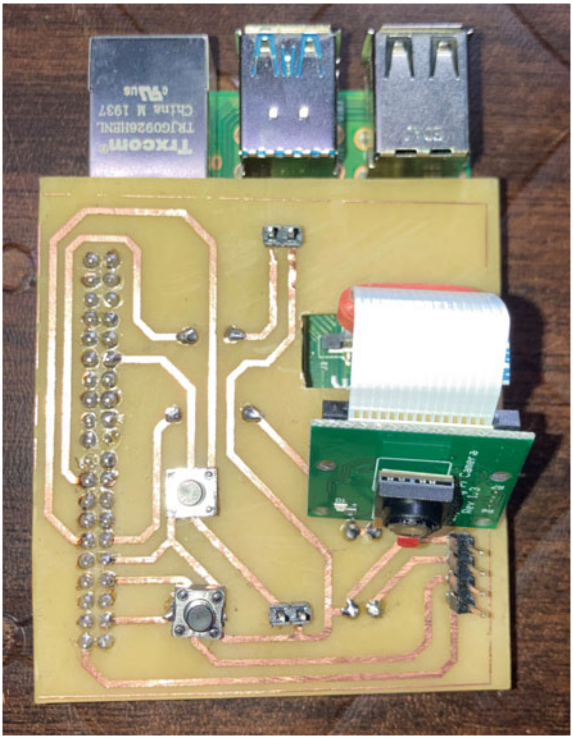
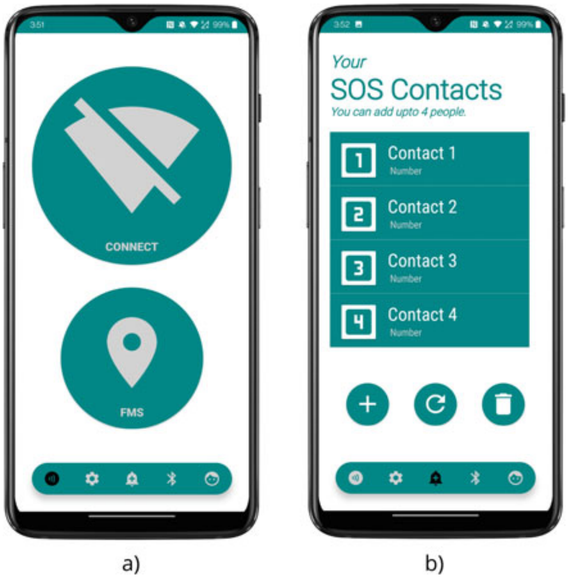


Fig. 7 Wearable top view



Your

SOS Contacts

You can add upto 4 people.

1 Contact 1
Number

2 Contact 2
Number

3 Contact 3
Number

4 Contact 4
Number

Fig. 8 **a** Home screen of the proposed mobile device application. **b** The SOS contacts screen in the application the voice assistant module, additional functionalities like setting an alarm, etc. can be added.

References

1. Felix SM, Kumar S, Veeramuthu A (2018) A smart personal AI assistant for visually impaired people. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI), pp 1245–1250. https://doi.org/10.1109/ICOEI.2018.8553750
2. Bose P, Malpthak A, Bansal U, Harsola A (2017) Digital assistant for the blind. In: 2017 2nd international conference for convergence in technology (I2CT), pp 1250–1253. https://doi.org/10.1109/I2CT.2017.8226327
3. Tahoun N, Awad A, Bonny T (2019) Smart assistant for blind and visually impaired people. In: Proceedings of the 2019 3rd international conference on advances in artificial intelligence (ICAAI 2019). Association for Computing Machinery, New York, NY, USA, pp 227–231. https://doi.org/10.1145/3369114.3369139
4. Bai J, Liu D, Su G, Fu Z (2017) A cloud and vision-based navigation system used for blind people. In: Proceedings of the 2017 international conference on artificial intelligence, automation and control technologies (AIACT '17). Association for Computing Machinery, New York, NY, USA, Article 22, pp 1–6. https://doi.org/10.1145/3080845.3080867
5. Khan A, Khan A, Waleed M (2018) Wearable navigation assistance system for the blind and visually impaired. In: 2018 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT), pp 1–6. https://doi.org/10.1109/3ICT.2018.8855778
6. Balakrishnan G, Sainarayanan G, Nagarajan R, Yaacob S (2008) A stereo image processing system for visually challenged impaired. World Academy of Science
7. Sharma V, Singh VM, Thanneeru S (2020) Virtual assistant for visually impaired. SSRN https://ssrn.com/abstract=3580035. https://doi.org/10.2139/ssrn.3580035
8. Saffoury R et al (2016) Blind path obstacle detector using smartphone camera and line laser emitter. In: 2016 1st international conference on technology and innovation in sports, health and wellbeing (TISHW), pp 1–7. https://doi.org/10.1109/TISHW.2016.7847770
9. Le V-H, Vu H, Nguyen TT (2018) A frame-work assisting the visually impaired people: common object detection and pose estimation in surrounding environment. In: 2018 5th NAFOSTED conference on information and computer science (NICS), pp 216–221. https://doi.org/10.1109/NICS.2018.8606899
10. Kim JU, Man Ro Y (2019) Attentive layer separation for object classification and object localization in object detection. In: 2019 IEEE international conference on image processing (ICIP), pp 3995–3999. https://doi.org/10.1109/ICIP.2019.8803439
11. Koskowich BJ, Rahnemoonfai M, Starek M (2018) Virtualot—a framework enabling real-time coordinate transformation & occlusion sensitive tracking using UAS products, deep learning object detection & traditional object tracking techniques. In: IGARSS 2018—2018 IEEE international geoscience and remote sensing symposium, pp 6416–6419. https://doi.org/10.1109/IGARSS.2018.8518124
12. Iyer V, Shah K, Sheth S, Devadkar K (2020) Virtual assistant for the visually impaired, pp 1057–1062. https://doi.org/10.1109/ICCES48766.2020.9137874
13. Young M (1989) The technical writer's handbook. University Science, Mill Valley, CA
14. Pise A, Ruikar SD (2014) Text detection and recognition in natural scene images. In: 2014 international conference on communication and signal processing, pp 1068–1072. https://doi.org/10.1109/ICCSP.2014.6950011
15. Bhowmick A, Prakash S, Bhagat R, Prasad V, Hazarika S (2014) IntelliNavi: navigation for blind based on kinect and machine learning. Multi-disciplinary trends in artificial intelligence (MIWAI '14), vol 8875, pp 172–183. https://doi.org/10.1007/978-3-319-13365-2_16

16. Awad M, Haddad JE, Khneisser E, Mahmoud T, Yaacoub E, Malli M (2018) Intelligent eye: a mobile application for assisting blind people. In: 2018 IEEE Middle East and North Africa communications conference (MENACOMM), pp 1–6. https://doi.org/10.1109/MENACOMM.2018.8371005
17. Mambu JY, Anderson E, Wahyudi A, Keyeh G, Dajoh B (2019) Blind reader: an object identification mobile-based application for the blind using augmented reality detection. In: 2019 1st international conference on cybernetics and intelligent system (ICORIS), pp 138–141. https://doi.org/10.1109/ICORIS.2019.8874906
18. Mohapatra S, Rout S, Tripathi V, Saxena T, Karuna Y (2018) Smart walking stick for blind integrated with SOS navigation system. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI), pp 441–447. https://doi.org/10.1109/ICOEI.2018.8553935
19. Kornprobst P, Tumblin J, Durand F (2009) Bilateral filtering: theory and applications. Found Trends Comput Graph Vis 4:1–74. https://doi.org/10.1561/0600000020
20. Patel C, Patel A, Patel D (2012) Optical character recognition by open source OCR tool tesseract: a case study. Int J Comput Appl 55:50–56. https://doi.org/10.5120/8794-2784
21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
22. Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S (2017) Recent advances in recurrent neural networks
23. Saoji S, Eqbal A, Vidyapeeth B (2021) Text recognition and detection from images using pytesseract. J Interdiscip Cycle Res XIII:1674–1679
24. Bressoud T, White D (2020) Web scraping. https://doi.org/10.1007/978-3-030-54371-6_22
25. Wang K, Gao X, Zhao Y, Li X, Dou D, Xu C (2020) Pay attention to features, transfer learn faster CNNs. ICLR
26. Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications
27. Chiu Y-C et al (2020) Mobilenet-SSDv2: an improved object detection model for embedded systems. In: 2020 international conference on system science and engineering (ICSSE). IEEE
28. Shuai Q, Wu X (2020) Object detection system based on SSD algorithm. In: 2020 international conference on culture-oriented science & technology (ICCST), pp 141–144. https://doi.org/10.1109/ICCST50977.2020.00033
29. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards realtime object detection with region proposal networks. In: Neural information processing systems (NIPS), pp 1–14
30. Gaikwad D, Baje C, Kapale V, Ladage T (2017) Blind assist system. IJARCCE 6:442–444. https://doi.org/10.17148/IJARCCE.2017.63101