

Object Detection and Recognition for Virtual Vision: Using Text-to-Speech Conversion Technique

Memoona Mushtaq¹, Muhammad Munwar Iqbal^{1,*}, Ayesha Mariam¹, Aatka Ali², and Muhammad Nabeel Asghar³

¹Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

²Department of Computer Science, Air University Islamabad, Multan Campus, Multan

³Department of Computer Science, Bahauddin Zakariya University, Multan

*Corresponding Author: Muhammad Munwar Iqbal. Email: munwariq@gmail.com.

Received: September 07, 2022 Accepted: November 11, 2022 Published: December 29, 2022.

Abstract: We see people with disabilities on a regular basis, and one of the most serious concerns is loss of sight or blindness. Blindness causes navigational issues for a person and leads to social challenges like readability. We are attempting to create an inexpensive project for common people who have lost sight. To assist them in their everyday lives by providing them with a virtual vision to have a better real-life experience. It will also assist people who are illiterate or unfamiliar with the area, i.e., tourists, who cannot interpret the texts placed in front of them. Along with them, it can aid auditory learners, such as pupils, in memorizing notes in a unique way. The goal of our project is to support people in a variety of everyday chores by utilizing the benefits of a design. As a proof of concept, this project provides several examples of uses, such as text recognition technology that can assist in reading from hard copy documents, sign boards, billboards etc., which will then be transformed to voice and audible to the user via earbuds. This research presents a developed prototype of a smart cane. The experimental results show an accuracy of 99.96%, an error rate of 0.0309%, and an average response time of 1.4 seconds.

Keywords: Text Recognition; Text to Speech Conversion; Object Detection & Recognition; Blindness and Visual impairment; Radio Frequency Identification.

1. Introduction

Along with navigational issues, blindness also provides considerable social obstacles because some activities are inaccessible to blind persons. Blindness often compromises a person's competence to carry out several tasks or duties, restricting their working options. Their finances and self-esteem can suffer as a result, which might adversely impact them. Visual impairment may also cause participation challenges in numerous activities outside of the job, such as recreation, shopping, internet research, and reading. For example, if a blind person cannot read the content on a web page immediately, it may be essential to rely on screen reading software to have the information read out. This can make web browsing sluggish and inconvenient. This can hinder a blind's ability to mingle in society and meet new people, which can impact their emotional health.

The below Figure 1, shows the GPS module, which indicates the blind person's directions to the family member via email. Another benefit that this proposed system will give them is reading capability. Our proposed system can read any text like signboards, captions on the shops, documents, etc. text. To speech conversion module will let them hear the text in front. The software-based solution comprises an app that will perform object detection and recognition, read the text, track the stick, and tell the blind about their

location. This proposed mechanism's primary goal is to help blind and visually impaired communities by providing a complete solution to their navigational, interactive, and reading problems. And this proposed model aims to achieve fast response time, efficiency, and cost-effectiveness to make blinds independent and fully rely upon this affordable solution.

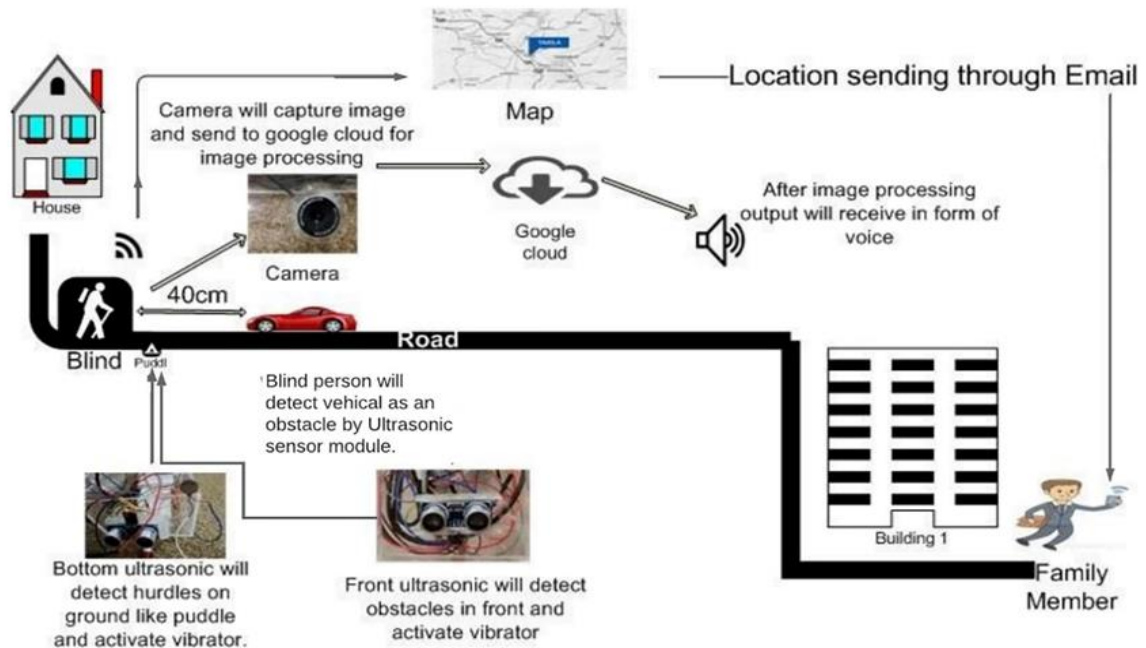


Figure 1. Previous working model similar to the proposed research

Our research also addresses the issues for other users, such as visually impaired people, audio learners, travelers, and illiterate masses like "it's difficult for auditory learners to absorb. The context from the book understanding foreign country voice communications for visitors. Understanding instructions printed in regional language for "tourists' difficulties understanding texts placed around them for uneducated or illiterate individuals. For these issues, we are attempting to propose a technical solution that is inexpensive to common people who have lost their sight in order to assist them in their everyday lives by providing them with a virtual vision to have a better living experience through utilizing the most recent technology. The proposed system works as the image is captured from pi-cam and processed by raspberry-pi. This synthesized image is then undergoing a text-to-speech conversion process, and the further procedure is performed on the converted text file to transform it into an audible mp3 file. The process begins with recording the scenario surrounding the individual with the pi-cam module. The raspberry-pi then recognizes the image and sends this image to the cloud for processing. Where object detection and recognition are accomplished by implementing different image processing algorithms, google vision API is created by the google brain team. This processed text file is sent back to the raspberry-pi, where the text file's contents are transformed into the audio alert message by making use of Text-to-Speech Synth (TTS) in the E-Speak program. An algorithm written in python is used to refine and control the speed of the audio file.

2. Literature Review

An intelligent electronic eye for those who are blind or have vision problems gathers visual information from the environment using obstacle sensors. It is common to employ an AVR microprocessor, solar photovoltaic module, and piezoelectric source [3]. The article "A Framework for an Electronic Guidance System for Visually Impaired Users" suggests an ultrasonic sensor-based system. This system uses an Arduino Mega 2560 with an ATmega2560 chip that can detect objects, stairs, and traffic lights [4]. The combination of mobile applications and remote processing computers allows blind persons to navigate easily indoors using computer vision-guided systems [5]. With the use of image-to-speech technology, ultrasonic sensors, and indoor object recognition, the Netra Smart Hand Gloves can identify impediments. To gauge the distance between the obstacles, a buzzer and trigger are linked to the glove [6]. Another

obstacle detection system for those with vision loss combines Arduino and the sonar concept, which perceives objects and obtains a vibro-tactile reaction to indicate directions [7]. The Dijkstra algorithm and smart phone are used to determine the shortest path and identify hurdles in path for blind and visually impaired ones. To help the blind person avoid obstacles, this gadget and the Bluetooth module also employ an ultrasonic sensor to detect obstacles in front and immediately respond to the user. When using this instrument, we speak about our desired destination. However, long-distance navigation is not possible with this gadget [8]. Another tool for blind people to use is an ultrasonic cane. This type uses ultrasonic signals to find ground obstructions like potholes. Speakers and LEDs are employed to present the user with a trigger [9]. Blind People's Obstacles Recognition System is another smart solution that tracks barriers and makes use of RFID technology [10]. A stereo vision-based gadget refers as Obstacle Detection of a Novel Travel Aid for Visually Impaired People combines picture morphological processing and an adaptive threshold method for detection [11]. Navbelt, is computerized travel assistance for people who are blind, is the further gadget that offers audio direction, assisting the blind in navigating new environments. The system functions similarly to mobile bot navigation, which are able to choose the most advantageous route on their own [12]. Using the GSM module, blind people can also send emergency messages to their guardians. the solar panel put over the charging circuit will provide power for the suggested system [13]. Another author of [14] contributed to an open research problem of automatic image classification techniques. In image processing, the classification of digital images plays a vital role in object detection and recognition techniques. A smart system introduced in [15] helps blind people detect obstacles in front of them and prevent them from any mishaps. This system also contains an android app with a GPS tracking system. Familiar or unfamiliar people can follow and track that person from home. They get information related to that person's current location. In this study [16] authors have presented an innovative and cost-effective smart stick that will assist blind or visually impaired people while walking in different outdoor and indoor environments. The stick developed can identify hurdles of any height in front or slightly to the side. This stick indicates the distance and location of obstacles by vibrating in hand and emitting audio into the person's ear. Changing frequencies of the generated vibration and distinct audio tracks alert and draw the individual's attention towards the distance from the obstacle. The modification technique to the existing mechanical stick form for the blind. This modification and sensor microcontrollers also sense and monitor the health parameters. Any abrupt changes in the reading will automatically trigger a message along with the current location to the specified member. The proposed mechanism is for visually impaired individuals and is intended to be inexpensive, easy to handle, and user-friendly. By offering extremely reliable navigational performance and a better user experience through various additional features. [17].

3. Method and Materials Used

3.1. Hardware Requirements

The proposed system is a combination of software and hardware technologies. As for hardware, the technologies used are Raspberrypi-3, HDMI-Cable, SD-Card, RPI-Camera Supports night vision, adjustable focus, Internet Connectivity Module, Toggle Switch, Ultrasonic Sensors, Mini breadboard, Stick, Jumper wires, and Power Bank. The properties and specifications are described in (Table 1). Raspberrypi is used as a whole processing unit. All the processing is done in raspberrypi. It best suits the proposed system as it contains high storage capacity, cost, and size effective and easy to handle. The following characteristics, such as atmospheric pressure, price/cost, environmental factors, the type of obstacle to be identified, the range of detecting objects, and the desired accuracy of measurement, frequency transmission, and collected information, all affect the choice of an appropriate sensor for object detection. Laser, radar, and infrared can also be used for this, but their performance is affected by air conditions, unlike that of ultrasonic, which is unaffected by changing atmospheric conditions. Because precision is crucial to our product, we have chosen an ultrasonic sensor for detecting purposes.

Table 1. Hardware Tools Properties and Specifications

No	Hardware Equipment	Specifications
----	--------------------	----------------

1	Raspberry-pi	The Raspberry-pi3 Model B is utilized in our offering. It is a little, on-board computer that is about the size of a credit card. The third version, Raspberry Pi, has wireless LAN and Bluetooth connections, four USB ports, and a 10 times faster CPU. The input for our suggested system comes from GPIO pins, which we have linked to modules like ultrasonic sensors, webcam, audio, power bank etc. It has a 32 G.B. memory card installed into it that serves as permanent memory. Python can be used to code the application.
2	Ultrasonic	The ultrasonic sensor has a 40 kHz transmission signal, a 5V input voltage, a 30-degree angle of sense, and a 15-degree angle of effect. It has the ability to produce 2.4644 thin beams. It may fit within the cane because of its appropriate size. For close obstructions, it works well within a range of 2 cm to 10. Laser, radar, and infrared can also be used for this, but their performance is affected by air conditions, unlike that of ultrasonic, which is unaffected by changing atmospheric conditions. Because precision is crucial to our product, we have chosen an ultrasonic sensor for detecting purposes.
3	Web-Cam	USB-Cam is used for streaming video and taking pictures. It is dependable and has extremely suitable characteristics with the suggested system, such as 2.0M pixel video streaming resolution, focus-adjustable images, Bayer RGB conversion to YUY2 color space, and input voltage of standard 3.3V to 5V. The following parts are included with it: a CMOS sensor lens, an image processing circuit and connection, a digital video device with 8 Megapixels and a maximum 640*480 resolution, and a USB type interface.
4	Evo-Wingle	Evo-Wingle 3.9Mbs provide a constant supply of internet in the device. Image processing is possible because the relationship between the Raspberry Pi and the Google Cloud may be formed using the internet. It has the following features: compatibility for the 1900MHz frequency band, rates of up to 9.3Mbs, and simultaneous connections for up to 5 Wi-Fi-enabled devices. Hi-Link technology for automatic setup and user-friendliness. The features that work best with our product are portable and require no software download.
5	Power Bank	Specifications for the power bank Model PB-5200 include an output voltage of 5.1 volts and input and output currents of 500/1000 mA and 2.1A, respectively. Raspberry Pi and the Evo-Wingle are powered by two USB ports for charging devices, allowing them to be processed and put to use as soon as Raspberry Pi needs electricity for processing.

3.2. Software Details

The software we will use for our proposed model is the Raspbian operating system to function on Raspberry Pi. Google vision API is used for object detection and recognition and then for text to speech, text to speech synthesizers. For Coding in python, PyCharm IDE is installed, which supports development in python.

3.2.1. Raspbian-OS

We have used Raspbian OS for our project. " 'It's a free operating system runs on the raspberry-pi. It was released back in July 2012, derived from Debian Linux, and used the LXDE desktop environment. Comes packed with all the software " 'you'll need for every basic task to perform within a computer. The benefit of Raspbian over its foundation operating system Debian is that some programs will experience a very noticeable performance improvement on Raspbian.

3.2.2. Google Vision API

For Image processing, we have used the Google Cloud Vision API encapsulates potent machine learning models in a user-friendly REST API to allow developers to comprehend the information of a picture. It instantly organizes photographs into thousands of categories, recognizes specific items and faces inside pictures, and locates and deciphers printed text inside pictures. The main purpose we used google vision API is that it processes images within milli seconds and gives response in real time and its accuracy is 99.9%, Which other image processing techniques 'can't achieve.

3.2.3. Text to Speech Synthesizer

With Google Text-to- Google created the screen reader program called Speech. It powers software that reads the text on the screen in a variety of languages. Speech synthesis is the creation of human speech using artificial means. A voice computer, sometimes known as a speech synthesizer, is a type of computer system that serves this purpose. It may be utilized in both software and hardware devices. Text in everyday language is translated into voice using text-to-speech (TTS) technology. The resemblance of a speech synthesizer's voice to the human voice and its clarity are used to assess its quality. People with visual impairments or reading difficulties can listen to printed words on a home computer using an understandable text-to-speech application.

3.2.4. PyCharm

Python specified IDE for the developers of python, PyCharm offers an extensive range of tools. For effective development in python, these tools are firmly integrated to produce a practical environment. These are the reason we choose PyCharm for the " 'task's implementation in python

4. System Design

The process begins with recording the scenario, surrounding the individual with the pi-cam module and ultrasonic sensors. The raspberry-pi collected the data from sensors in the form of images and distance to the hurdle. In (Figure 2) flowchart shows the flow of processes.

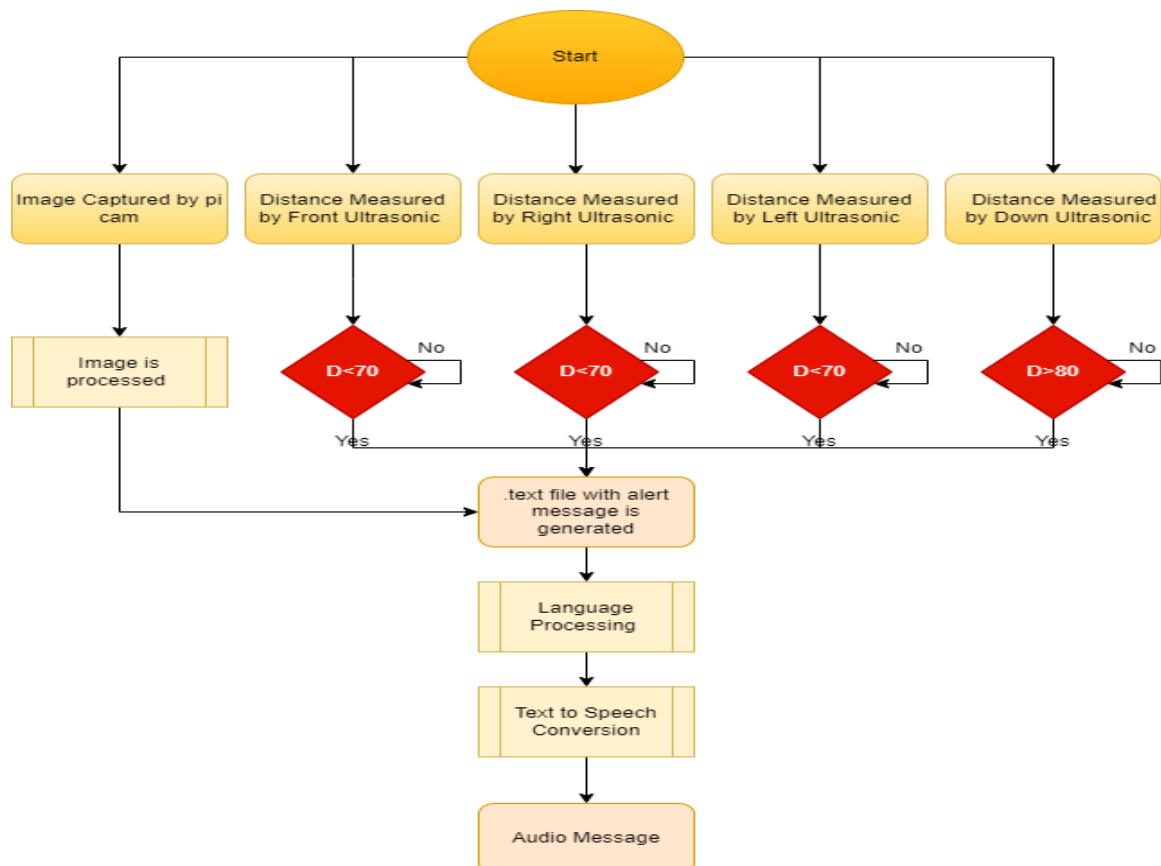


Figure 2. Proposed structure block diagram, show the working model

Image of the hurdle is captured by pi-cam and the distance from the hurdle is calculated by specified ultrasonic sensors for each direction. Raspberry-pi sends the captured image to the cloud for processing. Where object detection and recognition are accomplished by implementing different image processing algorithms, google vision API is created by the google brain team. This processed text file is sent back to the raspberry-pi where the text file's contents are transformed to audio by utilizing Text-to-Speech Synthesizer (TTS) in the E-Speak program. Along with an object detection and recognition for distance is also checked. If the distance to the hurdle in either direction is less than 70 cm then the alert message is generated, and audio output given for blind can be heard by using the earphone. The voice synthesizer could be applied either in software or hardware. Voice synthesizer program is employed to attain the desired purpose. An algorithm written in python is used to control the audio file's speed and translate the text in the desired language. This alert message will tell the blind which hurdle is present in which direction and at what distance so they may know in which direction they have to navigate to avoid hurdle.

This (Figure 3) shows the whole system's design and working schema. The scenario depicts if the blind person is navigating this proposed technology with him/her. Then if any hurdle comes in the path, the system tends to work as follows: a power bank is used to power up raspberry-pi. Raspberry pi is the processing unit where all processing is done. Model we are utilizing here is Pi 3-B+. Pi-cam tied-up to raspberry pi for capturing the real-time images. And four ultrasonic sensors for each direction i.e., left, right, front and down are connected to calculate distance from object in their respective paths. Then the captured image is sent to the cloud for image processing over internet. Evo is used for internet connection. After object detection and recognition data is sent back to the raspberry-pi. The processed image and distance from the hurdle are processed in raspberry pi to get audio output.

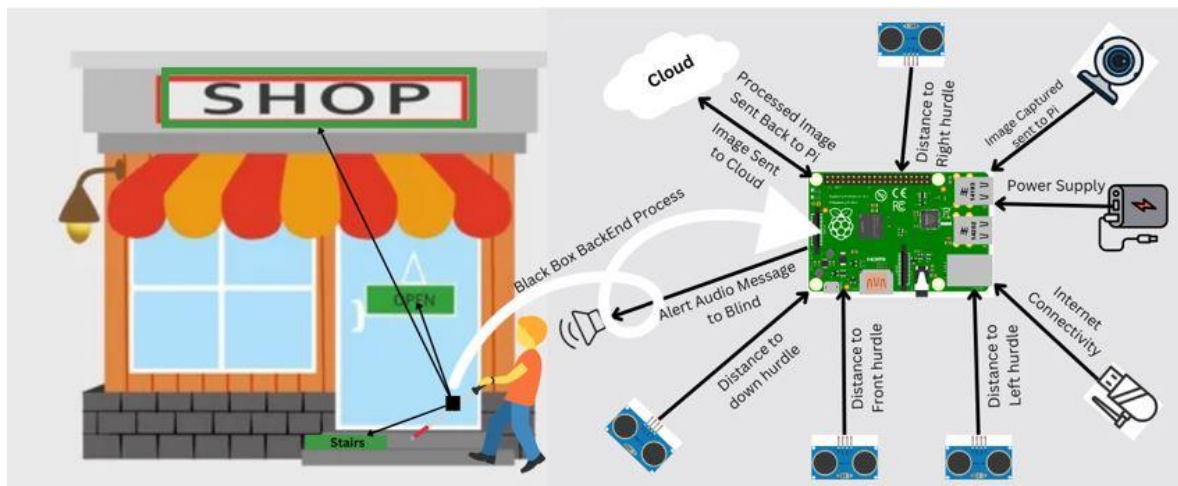


Figure 3. Image of Proposed System Whole Process

5. Result and Discussions

We have expected real-time, effective, cost-efficient, accurate, and competent results from the proposed structure. The testing of the model has been done with the help of different scenarios across the functioning of our modules. The results we get are much closed to our expected results. Our proposed system's major focus is to achieve real-time response and high accuracy as we can't afford negligence which leads to any mishap. Our system needs to be highly efficient so that the blind can safely rely upon this solution.

For this reason, we performed meticulous analysis and testing of our system. The parameters on which we have measured the efficiency of our model involve the time taken by our modules to give response and accuracy precision of our modules. Results depict that we obtained high accuracy and Real-time response as required.

Testing is a process of experimentation on an application or system to find errors and determine whether the system is fulfilling user needs. Different testing criteria are designed as mentioned in (Table 2) and tests are being performed. We have checked our proposed model in various scenarios to test its effectiveness and efficiency. Throughout the project's development, several tests are run to ensure

accuracy. Module tests are carried out during the creation of each module. The full unit is tested after integration. Real-time experimentation is done on a blind and found to make navigation simpler.

Table 2. Test cases for the Proposed Model

Test Cases	Objectives	CRITERIA	Test Status
1	Make sure that user can easily interact with the system.	For the proposed model, all the graphical user interface options display successfully.	Test successful
2	Make sure that user can hear correctly.	Audio files generated are audible	Test successful
3	Make sure that processing is done on run time.	Getting response on time	Test successful
4	Make sure that target Object detection & recognition work properly	Pi-Camera is working properly and Image processing is working on run time.	Test Successful
5	Make sure that all modules are working properly and guiding correct directions.	Ultrasonic Sensors get the distance properly Respected file for each direction works properly	Test successful

We have performed some real-time experimentation on the objects in the path at the specified distances and directions which can be seen through the (Figure 4). These processed images with labels were sent to the raspberry-pi where they were converted into specific text files and these text files are then converted to the audio alert message. This whole process takes the average time of 1.3 seconds and the accuracy of object detection and recognition attained is of 93% as shown in the (Table 3) For our proposed schema accuracy and real time response are the major concerned factors which will affect our efficiency and reliability of smart cane. From (Table 3) it can be seen that we have achieved the target and got the best results.



Figure 4. Processed and Labelled Images

Table 3. Average Accuracy and Response Time of Object Detection and Recognition

S. No	Objects	Accuracy of object detected and recognized	Response Time
1	Window	94%	1.3s
2	Refrigerator	91%	1.5s

3	Chair	98%	1.3s
4	Stairs	91%	1.3s
5	Bed	93%	1.4s
6	Couch	96%	1.5s
7	Television	93%	1.2s
8	Car	97%	1.3s
9	Door	90%	1.2s
10	Wardrobe	93%	1.6s
11	Mirror	93%	1.3s
12	Bench	92%	1.5s
Average Accuracy & Response time		93.42%	1.36s

The graph in () shows the difference between the readings of distance measured by ultrasonic sensors from the hurdles present in the path and the distance calculated manually. For the investigation purpose, we have placed the objects mentioned in the table at certain pre-calculated distances and then check the condition that whether the proposed model works according to the specified conditions of distance or not. So, from the two lines on the graph, it is concluded that there is no major difference in both calculations, and the system responds properly according to the given conditions. We placed the objects within or across the mentioned range of 70 cm and checked the criteria if the alert message was generated. The system has generated every time within the defined range alert message. Here in the graph at x-axis list of objects are presented in form of M-1 up to M10 and y-axis indicates the distances. Blue and Red lines denote distance measured by ultrasonic and manually measured distances from the hurdle. The steric signs on these lines show the hurdles at a particular distance.

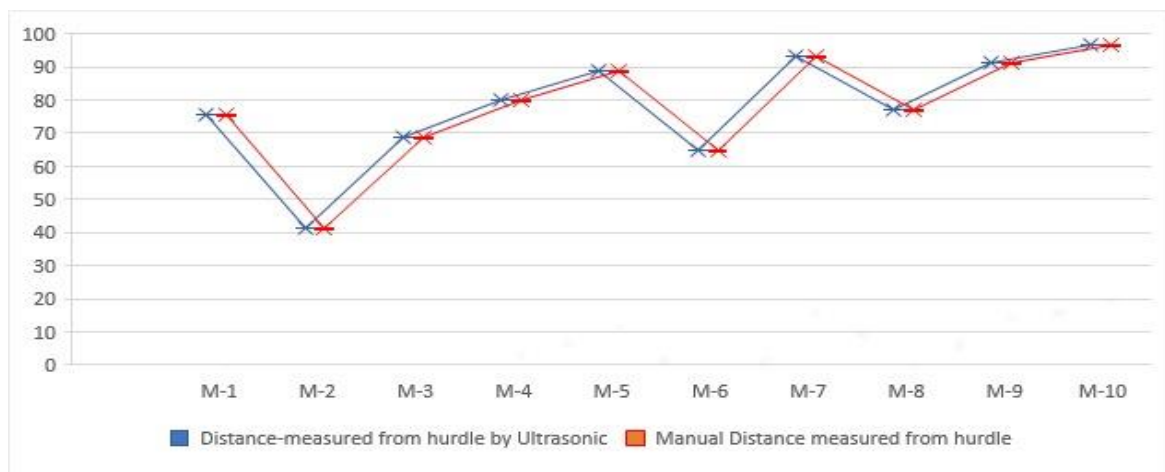


Figure 5. Graph of the actual distance to the hurdle detected through Ultrasonic

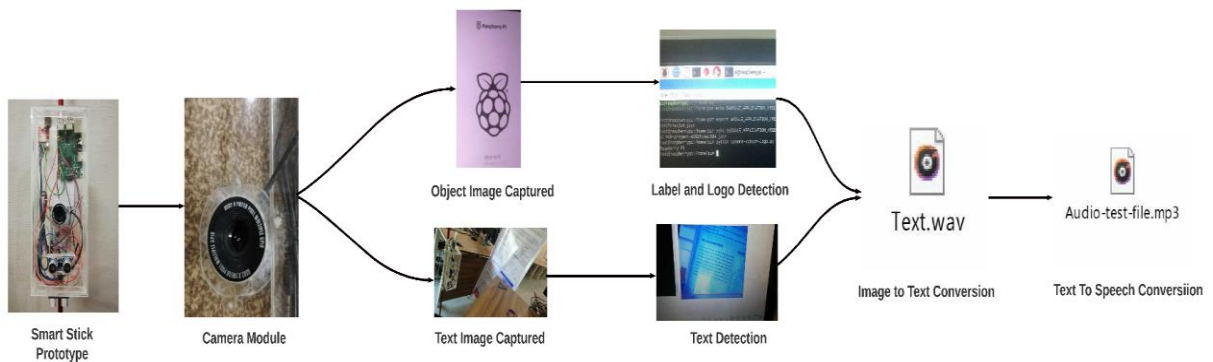


Figure 6. Experimentation on the Proposed Model

The model of the proposed system, the whole process through the real time experimentation performed, and the output obtained in the form of audio can be seen the (Figure 6). The figure shows how the sensors are mounted on stick, and the stick is assembled, their working as a camera is capturing real time scenario images. Obtaining results of object detection and recognition and then converting these to text files. This text file is then converted into audio which can be audible. This whole process is done within seconds as required. Different experimentation is performed for the text extraction from the image, and results are displayed in (Table 3). The column "caught picture" contains images taken by the Raspberry Pi cam module (USB-Pi Camera). The field under "Text Extracted" contains processed picture information and the image converted to text. The amount of time needed for a picture to process and produce the desired results in the form of speech is indicated in the "Time" box. The field of time span is used to calculate the smart cane's overall precision and accuracy. According to the data compiled after experimentation and displayed in (Table 3), the average real time response to extract text on images and convert them into speech is 1.4 seconds. This response rate is according to the expectation for an effective solution, as per required.

6. Conclusion

This research discusses the implementation of smart sticks for the assistance of virtually impaired ones. The white cane is a basic gadget that is used to assist blind people. There are several characteristics that distinguish our smart stick from other existing devices. The proposed device is a single-piece system which have been implemented, such as a simple cane with a box consisting of all the mechanism, making it portable and requiring no prior knowledge from the user. This system is extremely useful for navigation and object identification. This research clearly shows that the proposed system addresses all the problems and implements the features missing in the existing solutions. It gives a real time response, and makes the system accurate and efficient. By using this cane, blind can easily and independently navigate through hurdles as they can identify the object present in the path and can make a safer move through the guidance of an audio alert message. Also, they can be capable of reading the text, sign boards, which would increase their learning capability and will be beneficial for other activities like shopping, crossing roads etc. The average response rate of 1.6s is better than the techniques already existing techniques, which makes our proposed solution much more efficient, productive, and reliable.

7. Future Work

In the future, presented research work can opt for developing the whole working model of the proposed system. Also, this work can be extended by adding the guided path for the blind by using techniques defined in this research. Specific paths can be defined, and guidance throughout the whole path, from start to destination, can be provided to the blind. Along with blinds, we can also take advantage of this study in different healthcare-related scenarios. It can help in the detection of various types of disease. Provide solutions to cure some of the medical related issues like people suffering from Parkinson or partial or temporary paralysis. This work can be beneficial in the field of surveillance and the automotive industry.

References

1. Cang Ye, 3D Object Recognition of a Robotic Navigation Aid for the Visually Impaired, IEEE Transactions on Neural Systems and Rehabilitation Engineering.
2. Maiti, P. Mallick, M. Bagchi, A.Nayek, T.K. Rana, Shreya Pramanik, "Computer vision guidance system for indoor navigation of visually impaired people" 2016 IEEE 8th International Conference on Intelligent Systems .
3. M.Maiti, P.Mallick, M.Bagchi, A.Nayek, "Intelligent electronic eye for visually impaired people", 978-1- 5386- 2215-5/17/\$31.00 ©2017 IEEE.
4. Namita Agarwal, Anosh Iyer, Sonalakshi Naidu, Snedden Rodrigues, "Electronic Guidance System For The Visually Impaired- A Frame- work", 2015 International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04 – 06, 2015, Mumbai, India
5. Kabalan Chaccour, Georges Badr , "Computer vision guidance system for indoor navigation of visually impaired people""", 2016 IEEE 8th International Conference on Intelligent Systems.
6. N.K. Srivastava;Satyam Singh , "Netra: Smart Hand Gloves Comprises Obstacle Detection, Object Identification OCR Text to Speech Converter for Blinds", 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON).
7. D. M. Suraj ; Varun A. Prasad ; Shwetha Lokesh ; Sudhanva G. Hebbale ; Vimuktha Evangeleen Salis"" "" Obstacle Detection for the Visually Impaired Using Arduino" 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)
8. Md. Ashraf Uddin; Ashraf Huq Suny "Shortest path finding and obstacle detection for visually impaired people using smartphone",2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)
9. Krishna Kumar; Biswajeet Champerty; K. Uvanesh ; Ripunjay Chachan ; Kunal Pal ; Arfat Anis, ""Development of an ultrasonic cane as a navigation aid for the blind people".2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)
10. M. Nassih ; I. Cherradi ; Y. Maghous ; B. Ouriaghli ; Y. Salih-Alj; "Obstacles Recognition System for the Blind People Using RFID"2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies
11. Ying Jie ; Song Yanbin, "Obstacle Detection of a Novel Travel Aid for Visual Impaired People" 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics.
12. S. Shoal ; J. Borenstein ; Y. Koren:1998 "Auditory guidance with the Navbelt-a computerized travel aid for the blind",IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).
13. H. Nikhil, D. Narasimhanayaka, N. N. Gujjar, I. Vidya, and G. Nayana, "Solar Powered Smart Stick for Impaired People.", IJRESM, 2020, Vol 3, pp 308-309.
14. B. Zafar et al., "Intelligent image classification-based on spatial weighted histograms of concentric circles," Computer Science and Information Systems, 2018, vol. 15, no. 3, pp. 615-633.
15. S. Wankhade, M. Bichukale, S. Desai, S. Kamthe, and A. Borate, "Smart stick for blind people with live video feed," International Research Journal of Engineering and Technology,2017, vol. 4, no. 3, pp. 1774-1778.
16. S. Sharma, M. Gupta, A. Kumar, M. Tripathi, and M. S. Gaur, "Multiple distance sensors based smart stick for visually impaired people," in 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2017, pp. 1-5..
17. S. Ghosh, M. Bose, and A. Kudeshia, "GPS and GSM Enabled Smart Blind Stick," in Proceedings of International Conference on Communication, Circuits, and Systems, Springer, 2021, pp. 179-185.