SENSORY SUPERVISION SPECTACLES: A Navigation Assistance for the Blind

Shamnaz Mehar, M. Tech Scholar, Department of Communications & Engineering, Rajagiri School of Engineering & Technology Kerala, India, shamnazmehar.mec@gmail.com

Naveen N, Associate Professor, Department of Communications & Engineering, Rajagiri School of Engineering & Technology Kerala, India, naveenrset@gmail.com

Abstract: In today's fast moving world taking each steps without obstacles is a real challenge for the visually impaired. A smart assistant like 3S (Sensory Supervision Spectacles) can assist the visually impaired to read images, avoid obstacles and track routes. So, this paper presents a thought of building up a shrewd framework which can help the visually impaired individuals in their day-to-day life. Some of the main challenges include trouble in moving without the help of someone, difficulty in understanding text or pictures, detecting hindrances, etc. The proposed 3S (Sensory Supervision Spectacles) is capable of text recognition using Optical Character Recognition (OCR), speech synthesis using Text-To-Speech (TTS), and obstacle detection using ultrasonic sensors (HC-SR04) and a GPS tracker. The segments are associated with Raspberry Pi for being processed. The gadget is a voice empowered framework that would coordinate the visually impaired individual in their everyday works.

Keywords* — *Arduino, Camera, GPS, Obstacle detection, OCR, Raspberry pi, Text recognition, TTS, Ultrasonic sensors (HC-SR04).

I. INTRODUCTION

The estimated number of people visually impaired in the world is 285 million. This number is huge to the point that it drastically affects the economy of the nation. If we look in this busy world, average person doesn't have the opportunity to try and spend some time to assist these distinctively capable ones. In this manner, the visually impaired individuals continually need the help of somebody in their daily life, particularly while on streets and this dependency sometimes leave them to think about themselves in a helplessness or loneliness state. Some other challenges include difficulty in reading signs, symbols, texts or images, detecting obstacles, tracking geographical position, etc.

This paper proposes a compact model which can be used everywhere without much difficulty. It incorporates an automated virtual text framework. This is completely processed on a Raspberry. The whole framework is integrated with a camera module, OCR (Optical Character Recognition) mechanical conversion, a TTS (Text-To-Speech) engine, ultrasonic sensors (HC-SR04) and a GPS (Global Positioning System) satellite-based radio navigation system. The camera module snaps a photo when a user presses the button which is linked with the Raspberry Pi, the saved image is then taken care by the OCR for electronic conversion. OCR accomplishes character recognition via performing some picture handling techniques, (for example, eliminating negative and positive spots). The extracted character will then be saved in a document format (e.g. .txt). This document will then be processed by the TTS software which reads out the text/characters to the user via an output speaker.

This model also integrates an object detection module which utilizes ultrasonic sensors which is capable of alerting the user with the exact distance information of the object and the speaker is utilized to produce the alarm sound. Totally four ultrasonic sensors are utilized for covering four directions and alarm sound is produced via output speaker giving four different signals (example: 'left', 'right', 'front', 'back') along with the distance of the object (similar to automobile sensors). Furthermore, the device also gives a vibration signal which aids the user in detecting obstacles in a noisy environment. Hence, the proposed framework is designed in cautioning the user through both voice and vibration. This framework utilizes an Arduino microcontroller, which gathers all the data from sensor unit and appropriately send signals to response unit to alert the user.

The model is additionally integrated with a GPS system, which utilizes internet service for navigation assistance and enables live position tracking by the user's family members.

II. LITERATURE REVIEW

With the main intention of assisting the visually impaired, proposing a smart spectacle which can perform text detection and voice output [1]. This can aid the visually

 DOI : 10.35291/2454-9150.2021.0230

impaired to read any form of printed text and give output in vocal form. The inbuilt camera in the specs is used to capture the image with texts and the captured image is analyzed using Tesseract-Optical Character Recognition (OCR). The recognized text is then converted into speech using an open source software speech synthesizer, eSpeak.

Shubham Melvin Felix [6], proposed a framework to help those individuals who are visually impaired utilizing Artificial Intelligence, Machine Learning, Image and Text Recognition. This thought is actualized through Android mobile application that primarily centers around voice assistant and image recognition. The application features include voice assistant by giving voice commands to detect objects in the surrounding.

Ananth Noorithaya, Kishore Kumar M, Dr. Sreedevi A [2] showcases a navigation device prototype that aids the visually impaired people. This system was created with the aim of being as simple as possible. Fully self-sustaining, with as little reliance on virtual mapping methods as possible. It has an on-board power source and helps the user to detect objects in their surroundings. The ultrasonic sensor detects an obstacle, and depending on the distance travelled by the ultrasound, an audio is played. The module can detect an obstacle as near as 4cm away. The blind user is given an appropriate audio instruction. The navigator successfully plays a suitable audio file into the headphones that corresponds to the distance between the user and the obstacle. GPS can be used to improve the project. Since a GPS module can actuate the location using GPS co-ordinates, this lets the user navigate more precisely and efficiently.

One main limitation in this method is OCR's handling of color, as OCR algorithm to separate color words from black and white letters is not available. Due to this reason, whenever a color paper is processed, we always get negative result. The Tesseract OCR also has limitation with detecting Underlined words, it detects underlined words as alphabets with noise. Here too, when the image is processed it recognizes as negative results [7].

Another smart device which aid the visually impaired people is a complete read out system. The device is made up of a webcam that is connected to a Raspberry Pi and accepts a page of printed text [8]. The Raspberry Pi's OCR (Optical Character Recognition) package scans it and converts it to a digital text. A text to speech conversion device reads out the text after it has been classified (TTS engine). Before being read out, the signal is fed into an audio amplifier. The simulation is simply the start of image processing, which includes image to text and text to speech conversions using the OCR programme built on the Raspberry Pi. The system has potential uses in libraries, auditoriums, and offices where directions and notices must be read, as well as in assisting with the completion of application forms.

Attacks on speech-to-text devices are an example of audio adversarial examples. Keiichi Tamura, Akitada Omagari, Shuichi Hashida proposes a new protection strategy for speech-to-text transcription neural networks against audio adversarial examples [10]. It's difficult to tell whether an audio adversarial example is represented by waveform data reflecting voice tone. As a result, the proposed protection method's main structure is focused on the sandbox approach. They try to remove perturbations and analyse the feedback to see whether it's an audio adversarial example in the proposed protection process. They verified that their protection method for speech-to-text systems can detect audio adversarial instances.

Arunima B Krishna, Meghana Hari, Dr. Sudheer A.P proposes a paper that describes a wearable assistive device for blind people that transforms text into acoustic output, allowing the user to read any type of text [11]. This paper focuses on a standalone Raspberry Pi-based device with a camera mounted on the finger that can assist visually impaired people in word-based reading of textual data pointed to by the finger. The images are captured using a webcam. The proposed wearable assistive device aids the visually impaired in reading any text accessible to them using an efficient algorithm, making it easy and accurate. When compared to other existing technologies, the model has many benefits, including portability, computational effectiveness, and affordability.

Another smart device which assist blind people is a crutch with a bracelet [3]. When using it, the blind guy may easily activate the system by pushing the buttons on the walking stick and the bracelet, both of which are square, big, and easy to locate. During walking, two ultrasonic detectors in the stick begin to emit ultrasonic waves and use the echo to measure the distance to the barrier. When the distance between the stick and the bracelet is less than three meters, the stick will send signals through Bluetooth to the bracelet, which will vibrate to alert the blind. When the distance between the blind and the obstacle is less than one meter, the vibration increases dramatically, allowing the blind to modify their walking direction to avoid the obstruction.

Crutch and bracelet is replaced by shoe and a portable audio device by Chaitali Kishor Lakde and Dr. Prakash S. Prasad [4]. The designed shoe is having multiple depth obstacle detection and RGB sensor, a control board to detect multiple level of obstacle and the ground object, and a sound recording and playing module for voice assistance. Once an obstacle is detected these sensors get activated and the user get an alert in the form of vibration and voice assistant.

 DOI : 10.35291/2454-9150.2021.0230

RADAR are used to find the angle, range and velocity of the object using its radio waves. But it gives only short range values and also it requires a long time. So instead of RADAR a low cost ultrasound sensor HC-SR04 is used to determine the distance between object and user [5]. For rotation, the HC-SR04 is coupled to a Servo Motor (SG90). Object detection is also notified by message/SMS using the SIM808 module. These components are wired to an Arduino Uno and a Raspberry Pi 3 to be processed in order to detect and alert the object. The advantage of this sensor over RADAR is that it is unaffected by object colour or transparency, may be used in low-light conditions, and is not impacted by dust, filth, or high-moisture settings.

Another smart device proposed for visually impaired people is a smart eye system [9]. The device is a voice-activated technology that guides visually impaired people through their daily tasks. The device merges numerous available technologies into a single multipurpose gadget that may be utilized by visually challenged people. The paper examines the architecture of such a system as well as the issues that come with it.

III. PROPOSED SYSTEM

A. Image to Speech Conversion

Visually impaired individuals experience various challenges in converting into text utilizing existing technologies, which faces issues with arrangement, centering, precision, portability, and productivity. We present a smart gadget that helps the visually challenged to successfully and effectively reads the printed or written text. The proposed project utilizes the technique of a camera based assistive device that can be utilized by individuals to read printed or written text. The structure is for executing image scanning method in an integrated framework on Raspberry Pi board. In this project, we have proposed a text reading framework for the visually impaired. The proposed completely coordinated framework has a camera as an input device to take care of the printed or written text for digitization, then the document is scanned and processed by a software module called the OCR (Optical Character Recognition engine. Majority of the accessibility tools developed for individuals with visual deficiency and impaired vision are programmed using OCR programming and Text-to-Speech (TTS) engine. OCR makes it conceivable to apply methods, for example, machine interpretation, text-to-speech and text mining to the captured image. And finally the recognized text document is fed to the output device as per user's preference, either a headset or a speaker. This output device reads out the text document loud. The below Fig. 1 shows the basic blocks of the image to Speech Processing Technique.

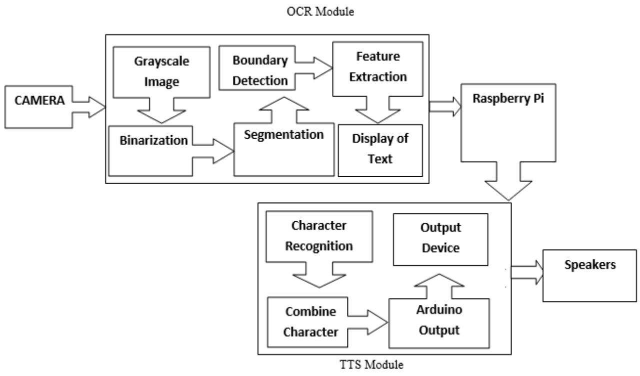

Fig. 1. Building blocks

The following are the steps to be followed to image to speech conversion.

- First step of the process is to read the image. This is done with the help of inbuilt camera.
- The next step is preprocessing. Here the image is converted into gray scale image. Then it is converted into black and white image with some threshold value.
- Next the image is cropped into its edges by finding its boundary.
- Once it is finished, the image is converted into word matrix. So that we can match the letters with templates when template is loaded.
- Each time the extracted letter is compared with the templates. The output is written on a text file.
- Next step is to concatenate. Then it is given to espeak. Thus text is converted into speech.

1) *Optical Character Recognition*: OCR is an electronic conversion tool which recognizes the text characters in the image which can be in printed text as well as in written form. The OCR layouts of each character are used to recognize the character. Next, the character image is converted into ASCII code, to be used in data processing. ASCII code limits with image recognition, where each character from the image is being picked and sent for recognizing. There are some pre-measures required to make the picture noise free which includes process like binarization, normalization and skew correction. here we can see the image going through some enhancements, for example, contrast correction and filtering out noise. After these pre-processing, the real role of OCR begins with the segmentation process, in this process each and every characters are separated. This level is assisted by some associated segment examination and projection analysis to again assist with the text segmentation process. Further the sequence is followed by feature extraction. All the real time processing in the OCR lies in the feature extraction.

2) *Text to Speech:* We realize that the artificial creation of human sound is generally 'speech'. A system is utilized to generate speech known as a speech synthesizer. In that,

 DOI : 10.35291/2454-9150.2021.0230

we can give the sound of human and robot. The text-to-speech framework is divided into two: namely a back-end and front-end. Here, the process begins with the conversion of the raw text like symbols, numbers, abbreviations, etc. This process includes text preprocessing & normalization. Next by allotting the phonetic transcription of the front ends of every word and divides the text into clauses into conditions, prosodic units, expressions, phrases and sentences. The phonetic transcription is assigned into the word. The output of the front end is nothing but the prosody information and phonetic transcriptions of the symbolic semantic portrayal. Though back-end is transformed into symbolic linguistic portrayal into Audio structure, which is also referred as the synthesizer.

B. Object Detection

Using ultrasonic sensors which is capable of detecting the object present in a specific range and once the object is identified; warning signal is given as voice and vibration. Four ultrasonic sensors are utilized for four directions for example right, left, front & back. The sensor gives different signals for different directions. The signals received from the ultrasound sensor are processed by a microcontroller so as to detect an object or obstacle in front and appropriately send signals to response unit to caution the user. Here the Arduino microcontroller controls the incoming and outgoing signals to different components. The benefits of ultrasonic sensing are its exceptional ability to probe into object or obstacles non-destructively on the grounds that ultrasound can propagate through any sort of medium including solid, liquid and gas, an exception is vacuum. The flow chart given in the Fig. 2 portrays the working.

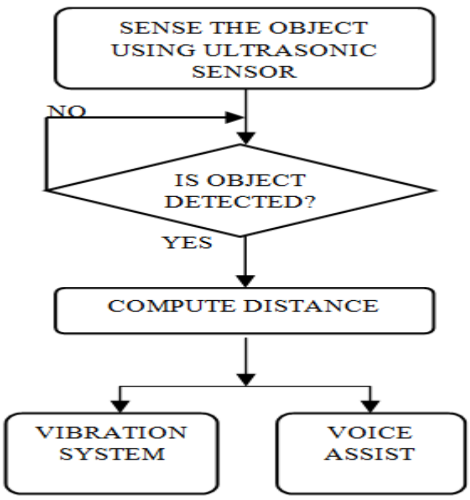

Fig. 2. Working of Object Detection System

Whenever an object is caught and recognized by the camera, the distance is determined by the ultrasonic sensors. Consequently, text to speech transformation is performed. Sound signs would thus be able to be heard from the headphones associated with Raspberry Pi. Along with the sound signal user can experience vibration. So in a noisy environment vibration system can help the user to detect the obstacle even if he could not hear the voice assistant. This is how object detection system works.

The ultrasonic sensor utilized in the task is HC-SR04 as appeared in Fig. 3. It has 4 pins-VCC, Trigger, Echo and ground.

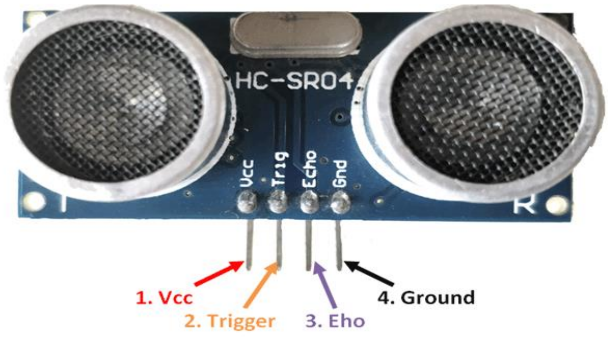

Fig. 3. HR-SO4 Sensor

Ultrasonic sensors have the following advantages: they are inexpensive, simple to set up and use, can be interfaced with any IoT module, and have a high resolution. Frequency, sensitivity, penetrating ability, and pinpoint accuracy are all features of this device. They can quickly determine the object's nature, form, and orientation, and are unaffected by sun, smoke, colour, dust, or material

The sensor sends out an ultrasonic wave and gets the wave reflected back from the target. The HC-SR04 utilizes sonar to evaluate the distance to an object. When the signal finds an entity, the transmitter (trig pin) sends a signal: a high frequency sound. The transmitter (echo pin) receives the reflected signal.

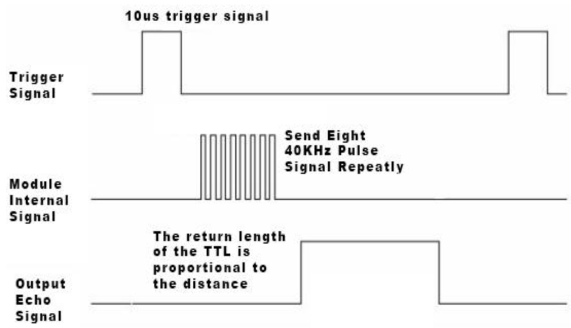

Fig. 4. Timing diagram of HR-SO4

At the point when the ultrasonic is triggered by a pulse sent to the trigger pin (min. 10 milli second pulse width), it sends an 8-burst of a directional 40 KHz ultrasonic waves through the transmitter and consequently makes the echo pin to go high as appeared in Fig. 4. The ultrasonic waves reflected from an object are picked by the receiver and the echo pin goes low. The span for which the echo pin was high is corresponding to the distance of the object. The status of the echo pin is continually checked. The exact time

 DOI : 10.35291/2454-9150.2021.0230

ISSN : 2454-9150 Vol-07, Issue-02, MAY 2021

of echo pin being high is calculated and relating count value is enlisted in a microcontroller.

The distance (D) travelled by any object is given by product of speed (S) and time (T) as in Eq. (1).

$$D=S*T \tag{1}$$

In the case of ultrasonic sensor, duration of travel of the ultrasound is the time taken by the signal to travel from sensor to the object and reach back to sensor, i.e. the total time corresponds to twice the distance. Hence actual time of travel, t = (T/2),

c = speed of sound (cm/microseconds)

D = distance measured (cm)

Consider an example of T= 3.28ms

Substituting values in the equation, we obtain

$$D=S*T \tag{2}$$

$$D = 340*(3.28/2)$$

$$D = 0.557m$$

Therefore, an obstacle as close as 4cm can be detected using this module. Hence from the Eq. (2) it can be understood that the state or the width of the echo pin being high corresponds to the distance at which obstacle is located. By knowing this basic function of ultrasonic sensor and the response of echo pin, the microcontroller is designed to sense the echo pin status and also to initiate vibration and voice assist.

C. Positioning using GPS

The GPS catches the current area of the user and also checks whether the location is valid or not. If the location is a valid one, they the way to arrive at that destination is given in form of voice commands. The GPS system consistently monitors latitude and longitude values at intervals, the geographical position of visually impaired will be sent to the family via 4G terminal, with the goal that the family can know his/her exact position. In case if the blind individual feels tired and need his families to pick him home, he could press the button on the gadget. When this button is pressed, the particular location is send to his family through 4G terminal.

IV. RESULTS AND DISCUSSIONS

A. Results obtained from image to speech conversion

The proposed system is tested with different input sets. In this work, the OCR system is implemented for the recognition of English capital alphabets (A to Z) and numbers 0 to 9. The first step is to capture the image. This is the input image which needs to undergo further process. Each character is recognized at one time. The recognized character is saved as text with notepad file, as shown in Fig.9. Next, it converts that text into the speech.

Fig. 5. Input image with noise

The Fig.5. shows the input image given to the system. The image with noise is further preprocessed to obtain speech output.

Fig. 6. Gray scale image output

Once an image is captured, first step is to convert the input image into gray scale image as shown in Fig.6. The gray scaled image is further converted into binary image as in Fig.7.

Fig. 7. Output after binarization

Threshold value is calculated from gray scale image and binary image is formed according to the threshold value. Fig.7. shows the binary image formed from the gray scale image.

Fig. 8. Boundary of characters

Edge detection decreases the amount of data in an image and filters out redundant information while maintaining the image's essential structural properties. Fig.8. shows the edges of each character. Canny filter is used to detect the boundaries of each letters. Canny edge detection is a technique for extracting useful structural information from a variety of vision artefacts while reducing the amount of data to be processed drastically. It's been used in a variety of computer vision systems. Canny discovered that the criteria for using edge detection on a variety of vision systems are remarkably similar.

DOI : 10.35291/2454-9150.2021.0230

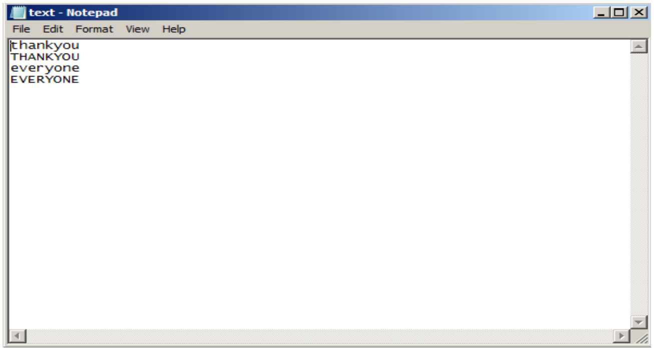

Fig. 9. Output text image

The black and white image is later converted into word matrix. Each character in the matrix is compared with the loaded template. For easy comparison the size of each letter is resized into template size. After comparison output is written in a text file as shown in Fig.9.

B. Results associated with Ultrasonic Sensors

Sl.No.	Object Detection		
	Calculated Distance (cm)	Actual Distance (cm)	Percentage Error
1	19.34	19.56	1,124
2	3.23	3.2	1.22
3	22.86	22.92	1.144

Table I: Object detection using sensors

The above Table I shows the output of object detection system. The table consists of actual distance between the user and obstacle and the distance calculated by the sensor. There occurs a small percentage error in actual distance and calculated distance maximum of 2%.

DIRECTION OF OBSTACLE	ANNOUNCEMENT
Front	Obstacle on your front
Left	Obstacle on your left
Right	Obstacle on your right

Table II: Voice result at various situations

Once an obstacle is detected, voice message is given to the user. The above table shows the output obtained by voice assistant once an obstacle is detected. The system is able to detect the object on the right left and front of the user.

V. CONCLUSION

Machine learning and artificial intelligence are two of the most rapidly developing technologies. These technologies are critical to the growth of the IT industry. We already have Attempts were made to use these devices for visually impaired people so that they could live independent and normal lives.

Blind people who have been denied the gift of sight have the right to live normal lives. In our research, we looked at developing a smart device for blind people to help them live more independently by using digital technology and artificial intelligence.

The proposed system integrates the working of a HD camera, OCR tool, TTS engine, sensors, speakers and a GPS. All these are integrated using Raspberry pi. The device has been designed in such a way that it is portable and easy for a blind person to use. The real time GPS tracker will assist the blind person to navigate his way to the predefined destination [for e.g. destination set from the home] with the help of the voice assistance feature in the GPS the person can walk with turn-by-turn voice guided navigation feature, on the way if there are any obstacles in the way, the sensors will give signal through voice both direction and the distance of obstacle from the person. This will help the blind person to avoid obstacles in his ways. The blind person can easily capture images with a push of a single button and the text in the images will be read out loud using the proposed system in a short time.

As a result, the ultimate goal was to build and design a compact, convenient, and low-cost system that will assist visually impaired people in moving around in unfamiliar environments. The proposed system is built with all ages in mind, is user-friendly, and does not require any prior experience or knowledge of advanced technologies. Additionally, in future development, we can train the system to detect the person present in front of the user. Along with that the system's functionality can be upgraded by training the system with more text font. So that the user can easily detect any text as like a normal person.

ACKNOWLEDGMENT

The author would like to thank Mr. Naveen N for his constructive comments and insightful suggestions that helped in improving the practicality and quality of this paper. I also would like to thank my institution for their support to carry out this project and providing with valuable materials related to this topic.

REFERENCES

[1] Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V, and Dr. M. A. Raja (2017), "Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method," IEEE International Conference on Innovations in Green Energy and HealthcareTechnologies(ICIGEHT'17).

[2] Ananth Noorithaya, Kishore Kumar M, and Dr. Sreedevi A. "Voice Assisted Navigation System for the Blind," Proceedings of International Conference on Circuits, Communication, Control and Computing (I4C 2014).

[3] Yi-Qing Liu, Zhi-Kai Gao, Zhang-Jian Shao and Gu-Yang Liu, "Intelligent ultrasonic detection of walking sticks for the blind," ECAI 2017 - International Conference – 9th Edition.

 DOI : 10.35291/2454-9150.2021.0230

[4] Chaitali Kishor Lakde, and Dr. Prakash S. Prasad, "Navigation System for Visually Impaired People," International Conference On Computation of Power, Energy, Information and Communication, 2015.

[5] Akshaya U Kulkarni, Amit M Potdar, Suresh Hegde and Vishwanath P Baligar, "RADAR based Object Detector using Ultrasonic Sensor," 1st International Conference on Advances in Information Technology, 2019.

[6] Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People" Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).

[7] Hasan U. Zaman, Saif Mahmood, Sadat Hossain and Iftekharul Islam Shovon, "Python Based Portable Virtual Text Reader", IEEE 2018.

[8] Dr. B. Muthusenthil and Joshuva J, "Smart Assistance for Blind People using Raspberry Pi," International Journal of Advance Research, Ideas and Innovations in Technology (Volume 4, Issue 2), 2018.

[9] Joe Louis Paul I, Sasirekha S, Mohanavalli S, Jayashree C, Moohana Priya P, Monika K, "Smart Eye for Visually Impaired-An aid to help the blind people", Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019).

[10] Keiichi Tamura, Akitada Omagari, Shuichi Hashida, "Novel Defense Method against Audio Adversarial Example for Speech-to-Text Transcription Neural Networks", 2019 IEEE 11th International Workshop on Computational Intelligence and Applications November 9-10, 2019, Hiroshima, Japan

[11] Arunima B Krishna, Meghana Hari, Dr. Sudheer A.P, "Word Based Text Extraction Algorithm Implementation in Wearable Assistive Device for the Blind", Unpublished.

 DOI : 10.35291/2454-9150.2021.0230