International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Voice Assisted Text Reading and Google Home Smart Socket Control System for Visually Impaired Persons using Raspberry Pi

Chaitali S. Hande[1], Vijay R.Wadhankar[2]

[1]*M.Tech. Student of Electronics Engineering, Agnihotri College of Engineering Maharashtra, India*
[2]*Professor, Department of Electronics Engineering, Agnihotri College of Engineering, Maharashtra, India*

Abstract- *Speech and text is main medium for human communication. A person needs eyesight access the information in a text. However those who have very poor eyesight can collect information from voice. This paper offers a camera based assistive text reading to help blind person in reading the text present on the captured image. The proposed idea involves text extraction from scanned image using Tesseract OCR (Optical Character Recognition) and converting the text to speech by e-Speak tool, which is a process that makes visually impaired persons to read the text. This is a prototype for blind people to recognize the products in real world by extracting the text on image and converting it into speech. For blind people it is very difficult to switch on /off electrical devices and some time they get shock. We made use of Google Home's voice recognition with the conception of machine-learning to prove the feasibility analysis about fulfilling the users' needs by a smart home. So blind people van control devices by using voice .Proposed method is carried out by using Raspberry pi and portability is achieved by using a battery backup. Thus the user can carry the device anywhere and able to use at any time. This technology helps millions of people in the world who experience a significant loss of vision.*

Key Words: Tesseract OCR, e-Speak, HT 12 D Decoder, Raspberry Pi.

1. INTRODUCTION

Due to eye diseases, uncontrolled diabetes, accidents and other reasons the number of visually impaired persons increased every year. The most significant difficulty for a visually impaired person is to read. In this proposed system text recognition is done by Open Computer Vision, a library of functions used for implementing image processing techniques. An image or a set of parameters related to image is the output of image processing. Text extraction from an image is carried out by OCR. OCR is the method of conversion of images of printed books, sign boards etc. to text. The binary image is converted to text by Tesseract library in OCR engine that detects the outline, slope, pitches, white spaces and joint letters. It also checks the quality of the recognized text. In this system the conversion of text to voice output is by e-Speak algorithm. The e-Speak is a Text- To-Speech (TTS) system which converts text into speech. The artificial production of human speech is known as speech synthesis. The speech synthesizer can be implemented in a software or a hardware product. The platform used for this purpose is known as a speech synthesizer. The storage of entire words or sentences allows for high-quality output in specific usage domains. A synthesizer can incorporate the model of a vocal tract and other human voice characteristics. This project aims to build an efficient camera based assistive text reading device. The idea involves text extraction from image taken by a camera installed on a spectacle. The extracted text is then converted to audio signals and to voice output. It is also used to detect a person's face in the frame. This is carried out by using Raspberry pi where the portability is the main aim, which is achieved by providing a battery backup.

2. LITERATURE REVIEW

In [1] A smart spec for the blind persons can perform text detection thereby produce a voice output. This can help the visually impaired persons to read any printed text in vocal form. A specs inbuilt camera is used to capture the text image from the printed text and the captured image is analyzed using Tesseract-Optical Character recognition (OCR). The detected text is then converted into speech using a compact open source software speech synthesizer, eSpeak. Finally, the synthesized speech is produced by the headphone by TTS method. In this project Raspberry Pi is the main target for the implementation, as it provides an interface between camera, sensors, and image processing results, while also performing functions to manipulate peripheral units (Keyboard, USB etc.,).

In [2]the design involves human face, object and textual recognition which make vision for visually challenged. The smart kit contains an eye glass provided with camera, an earphone, a microphone and the system where the processing is carried out. The camera present at the nose head of eye glass captures the intended image of the user as snapshots and transfer to the system where it get processed and produces the specified audio description as output. The system's database holds a corresponding text for each image which is then converted to audio stream when sounded. In case of any mismatch or no entry of the image in the datasets then the new image is stored as new dataset with the name specified through microphone. This avoids mismatch of that image in the future search.

In [3] This paper proposes a camera based assistive text reading to help visually impaired person in reading the text present on the captured image. The faces can also be detected when a person enter into the frame by the mode

control. The proposed idea involves text extraction from scanned image using Tesseract Optical Character Recognition (OCR) and converting the text to speech by e-Speak tool, a process which makes visually impaired persons to read the text. This is a prototype for blind people to recognize the products in real world by extracting the text on image and converting it into speech. Proposed method is carried out by using Raspberry pi and portability is achieved by using a battery backup. Thus the user can carry the device anywhere and able to use at any time. Upon entering the camera view previously stored faces are identified and informed which can be implemented as a future technology. This technology helps millions of people in the world who experience a significant loss of vision**.**

In [4] the basic framework is an embedded system that captures an image, extracts only the region of interest (i.e. region of the image that contains text) and converts that text to speech. It is implemented using a Raspberry Pi and a Raspberry Pi camera. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. Two tools are used convert the new image (which contains only the text) to speech. They are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines. The audio output is heard through the raspberry pi's audio jack using speakers or earphones.

In[5] Voice Based Home Automation System using Raspberry Pi is the project which will be very useful for old age people and disabled people, basically for one's who cannot perform basic activities efficiently. It is the idea which corresponds to the new era of automation and technology. The main aim of the home automation system is to make life easier. Mobile devices are very common among everyone due to its user friendly interface and portability features. In this project we aim to control electrical home appliances by android voice commands using Wi-Fi as communication protocol between Raspberry Pi and Android device. Raspberry Pi 3 becomes a better option for home automation via internet due to its feature of inbuilt Wi-Fi and Bluetooth.

In[6] The aim of this project to develop a system that will voice control the home appliances and also provide security against intrusion when the home owner is not in home. This paper mainly concerned with the automatic voice control of light or any other home appliances. It is used to save the electric power and human energy. This project is made with help of Raspberry Pi 3 and Relay driver circuit. The various appliances are connected to the relay circuit and the microphone connected to Raspberry Pi 3. After successful recognition of voice command the Raspberry Pi 3 drives the corresponding appliances. Voice recognition is developed by using Google API's.

3. METHODOLOGY

Raspberry pi

Raspberry pi is a small computer which could be programmed.It works like a Linux based computer which could do all the normal operations in PC. Raspberry Pi works in open source platform. Raspberry Pi 3 Model B 1GB is used in this system. This model comes with 40 GPIO pins and 4 USB ports which makes it more useful. Also it has camera interface and 3.5mm audio jack. USB ports available on this board are used to connect the camera with raspberry pi. Three GPIO pins are used, for capturing image, for mode control and for shutting down the system respectively. The board is operated in such a way that the code starts executing when it is powered ON. The audio output is available through the audio jack.

E-speak

E-Speak is a compact open source software speech synthesizer for English and other languages, for Linux and Windows. The program 'E-speak' is a simple speech synthesizer which converts written text into spoken voice. The espeak program does sound a bit robotic, but its simple enough to build a basic program. E-Speak uses a "formant synthesis" method. This allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds, but is not as natural or smooth as larger synthesizers which are based on human speech recordings.

4. BLOCK DIAGRAM

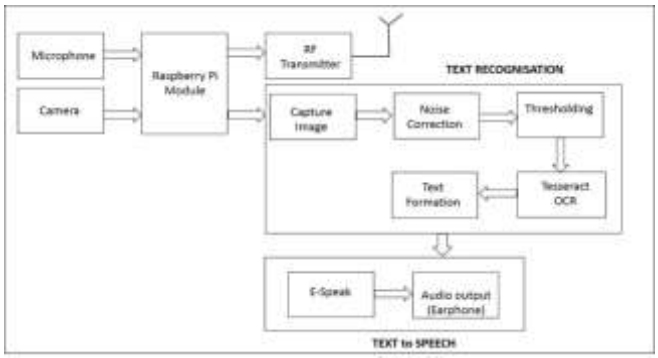

Fig.1(a): Block Diagram of Text Mode

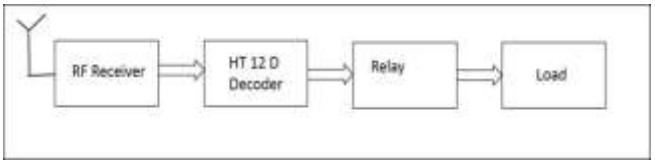

Fig.1(b): Block Diagram of Automation Mode

The proposed system has two different modes as shown in fig.1(a) the text mode and fig.2(b) automation mode.

Text Mode:-

The system captures the frame and checks the presence of text in the frame. If a character is found by the camera the user will be informed that image with some text was detected. Thus if the user wanted to hear or to know about the content in the image he can use a switch to capture the image. The captured image is first converted to grayscale and then filtered using a Gaussian filter to reduce the noise in the image. Here adaptive Gaussian thresholding is used to reduce the noise in the image. The filtered image is then converted to binary. The binarized image is cropped so that the portions of the image with no characters are removed. The cropped frame is loaded to the Tesseract OCR so as to perform text recognition. The output of the Tesseract OCR will be text file which will be the input of the e-Speak. The e-Speak creates an analog signal corresponding to the text file given as the input. The analog signal produced by the e-Speak is then given to a headphone to get the audio output signal.

Automation Mode:-

In the automation mode using Microphone the blind person give command to the system in the form of audio. This auto input process in to raspberry pi and signal send to the receiver side by using RF Transmitter. At receiver side signal received by using RF Receiver and signal is decode by using HT 12 D decoder ic and these decoded signal transfer relay. Using relay particular devices on and off by user voice command.

5. DESIGN AND IMPLEMENTATION

Conversion of image to text using OCR tool

Tesseract is an open source-OCR engine. It assumes that its input is a binary image with optional polygonal text region defined. The first step is a connected component analysis in which outline of the components is stored. By the inspection of the nesting of outlines, it is easy to detect inverse text and recognize it as early as black on white text. At this stage, outlines are gathered together, purely by nesting, into blobs. Blobs are organized into text lines, and the lines and regions are analysed for fixed pitch or proportional text. Slope across the line is used to find text lines. Quality of recognized text is verified. If clarity is not enough the text is passed to associator.

Conversion of text to voice using E-Speak

Normal text to speech conversion is done using E-Speak which is a TTS system. The artificial production of human speech is known as speech synthesis. Speech computer or speech synthesizer is used for this purpose and can be implemented in software or hardware products. To create a completely "synthetic" voice output a synthesizer can be used to incorporate a model of the vocal tract and other human voice characteristics.

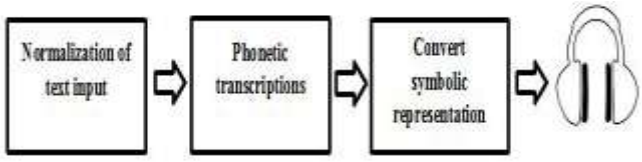

Fig. 2. Schematic diagram for E-Speak

Fig. 2 explains the e-Speak algorithm. A TTS (or"engine") is composed of two parts a front-end and a back-end. The front-end has two major tasks, the normalization and phonetic transcription of text. Normalization, pre-processing, or tokenization of text is the conversion of text containing symbols like abbreviations and numbers into equivalent written-out words. The front-end then assigns phonetic transcription to each word. The prosodic units like clauses, sentences, and phrases are marked and divided. Text-to phoneme conversion is the process of assigning phonetic transcriptions to words. The output from the front-end is a symbolic linguistic representation from the Phonetic transcriptions and the prosody information. The back-end performs the function of a synthesizer. The symbolic linguistic representation to sound conversion is achieved using this back end. The output sounds like human speech which describes the naturalness, the output is ease with the intelligibility of understanding. Speech synthesis systems usually try to maximize both natural and intelligibility which are the characteristics of an ideal speech synthesizer.

Voice Control Home Automation

The software used Google Voice and Speech API's. The voice command from user is captured by the microphone. This is then converted to text by using Google Voice API. The text is then compared with the defined commands inside the command configuration file. If it matches with any of them, then the bash command associate with it will be executed. This is achieved by using the Google speech API, which converts the text into speech. Below block diagram showing basic working for voice control home automation.

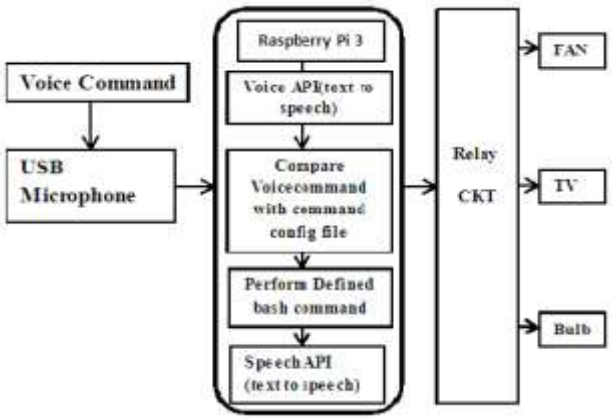

Fig.3: Block Diagram of Voice Control Home Automation using Rpi

6. RESULT

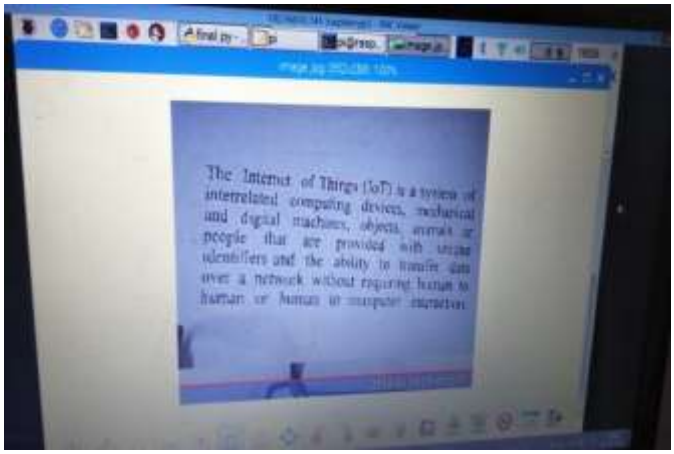

Fig. 4: Actual (Input) Image taken by camera

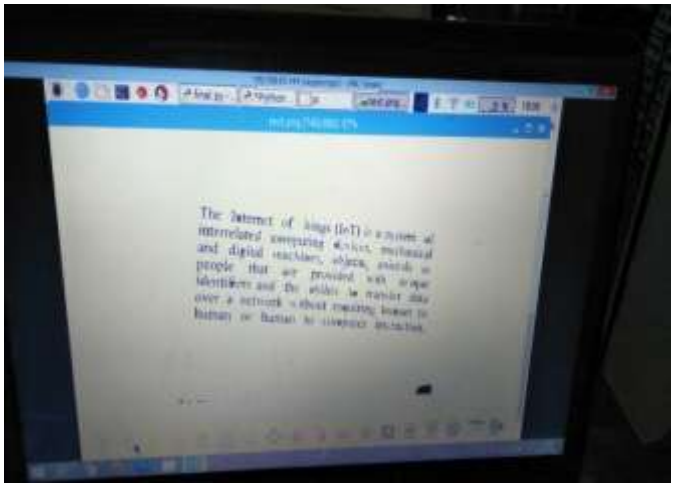

Fig. 5 : Converted (Black & White) image by OCR Technology

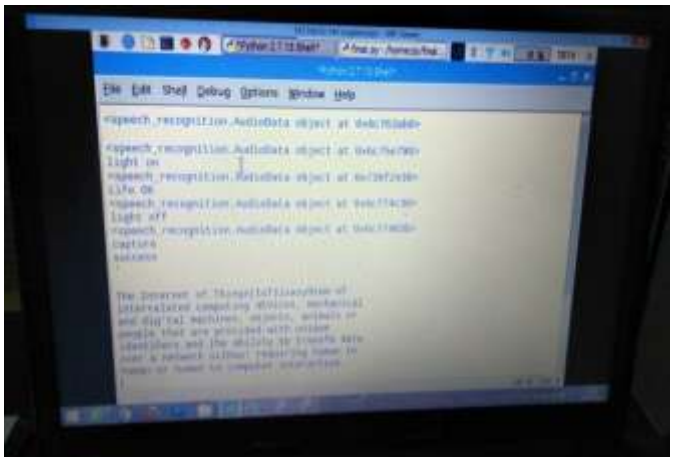

Fig. 6 : Output image when converted into speech

Fig.7: Automation Setup when light ON

Fig.8: Automation Setup when light OFF

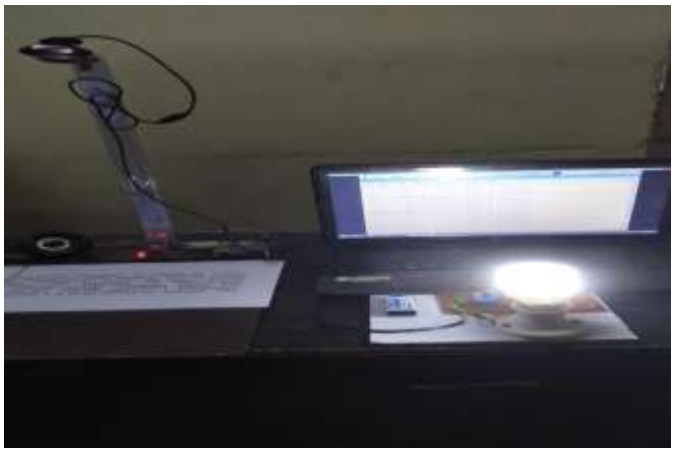

Fig. 9: Hardware implementation

7. CONCLUSION

We have implemented text to speech conversion technique and voice controlled home automation for the benefit of easy use and control of devices by visually impaired people using raspberry pi. The simulation results have been successfully verified. Our algorithm successfully processes the image and reads it out clearly. We have applied our algorithm on many images and found that it successfully does its conversion.

This system allow user to regulate the home appliances. This is compact, economical as well as efficient device for the visually impaired people and helpful to the society.

8. REFERENCES

[1] "Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method" by Ani R, Effy Maria, J Jameema Joyce , Sakkaravarthy V , Dr.M.A.Raja , In IEEE International Conference on Innovations in Green Energy and Healthcare Technologies(ICIGEHT'17), 978-1-5090-5778-8/17/$31.00©2017 IEEE.

[2] "Hot glass –human face, object and textual recognition for visually challenged" by diwakarsrinath, Praveen Ram in 978-1-5090-6221-8/17/$31.00_c 2017 IEEE.

[3]"Text recognition and face detection aid for visually impaired person using raspberry pi" by Mr.Rajesh M. in 2017 International Conference on circuits Power and Computing Technologies [ICCPCT]

[4] "Image to Speech Conversion for Visually Impaired by Asha G. Hagargund (International Journal of Latest Research in Engineering and Technology-2017)

[5] "Voice Based Home Automation System Using Raspberry Pi" by Harshada Rajput in International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 | Apr-2018.

[6] "voice controlled home automation using raspberry pi 3" by shubham olulkar in International Journal of Recent Innovation *in Engineering and Research Volume 02 Issue: 01 January – 2017.*

[7] "Voice Recognition by Google Home and Raspberry Pi for Smart Socket Control," by Chen-Yen Peng and Rung-Chin Chen, in 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI) March 29–31, 2018, Xiamen, China, 978-1-5386-4362-4/18/$31.00 © 2018 IEEE.

[8] "Text Detection and Recognition with Speech Output for Visually Challenged Person", by Ms.Rupali, D Dharmale, Dr. P.V. Ingole, in vol. 5, Issue 1, January 2016.

[9] "Vision based text recognition using raspberry PI," by Nagaraja, L., et al in National Conference on Power Systems, Industrial Automation (NCPSIA 2015).

[10]"Portable Camera Based Product Label Reading For Blind People.",by Rajkumar N, Anand M.G, Barathiraja N, in IJETT, Vol. 10 Number 11 - Apr 2014.

[11] "Portable Camera-Based Assistive Text and Product Label Reading from Hand-Held Objects for Visually impaired persons," by Chucai Yi, Yingli Tian and Aries Arditi, in IEEE/ASME Transactions on Mechatronics, Vol. 19, No. 3, pp.808, June 2014.

[12] "Improved text-detection methods for a camera-based text reading system for
visually impaired persons," by Ezaki, Nobuo, et al. in Eighth International Conference on Document Analysis and Recognition (ICDAR'05). IEEE, 2005.