

Drishti: Voice Assisted Text Reading Smart Specs for Visually Impaired Persons

Chitranjali Tanwar¹, S. Dakshayini², R. Gagana³, Mythri M. Hegade⁴, Manjusha Mane⁵

^{1,2,3,4}Student, Department of Electronics and Communication Engineering, Sambhram Institute of Technology, Bangalore, India

⁵Assistant Professor, Department of Electronics and Communication Engineering, Sambhram Institute of Technology, Bangalore, India

Abstract: According to the World Health Organization, out of 7.4 billion population around 285 million people are estimated to be visually impaired worldwide. It is observed that they are still finding it difficult to roll their day today life and it is important to take necessary measure with the emerging technologies to help them to cope up the current world irrespective of the impairments. In the motive of supporting, an innovative, efficient and real-time, cost effective smart spec is proposed to read any printed text in vocal form. As accessing text documents is troublesome for impaired people in many situations, it provides self- independency. Pi camera is used to capture the text image from the printed text and captured image is analyzed using Tesseract-OCR. The text in the form of [jpeg, png, jpg, bmp etc.] are considered for the analysis. The image captured is then processed and with the help of OCR/Tesseract the captured text was converted into letters which was made readable with the help of eSpeak software. The OCR algorithm involves scanning, processing, classification and recognition. Finally, eSpeak command software is used to convert the obtained text from the OCR into speech command. This converted speech is read aloud by speaker connected to Raspberry Pi. A tilt sensor is as used to notify fall detection.

Keywords: Tesseract OCR, Optical Character Recognition, Text-to-Speech, eSpeak, Raspberry Pi

1. Introduction

Of the 215 million visually impaired people worldwide, 40 million are blind. Some of the developed countries like the united states, the 2008 national health interview survey (NHIS) reported that an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired. Recent development trends in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such OCR systems. Accessing text documents is troublesome for visually impaired people in many situations, such as reading text on the go and accessing text in less than ideal conditions. The objective is to allow blind users to touch printed text and receive speech output in real-time. The development of such systems requires the usage of two technologies that are central to these systems, namely optical character recognition for Text Information Extraction (TIE) and Text-To-Speech (TTS) to

convert this text to speech.

A Text-To-Speech (TTS) synthesizer is a computer based system that should be able to read any text aloud, when it is directly introduced in the computer by an operator. It is more suitable to define Text-To-Speech or Speech-To-Text synthesis as an automatic production of speech, by 'grapheme to phoneme' transcription. Text Information Extraction is the primary and important function of any assistive reading system and is an integral part of OCR because this process determines the intelligibility of the output speech. The output quality of text-to-speech as well as extending our capabilities to generate expressive emotional synthetic speech. Automatic video text detection and extraction was employed in order to partition video blocks into text and non-text regions. Recent developments in computer vision, digital cameras, and computers make it possible by developing camera based products that merges with computer vision technology with other existing beneficial products such as optical character recognition systems. OCR is used to recognize words. It can recognize characters, words and sentences without any mistakes. OCR has a high rate of recognition which is the electronic conversion of photographed images of typewritten or printed text into computer readable.

2. Literature review

- Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V (2017), proposed a smart spec for the blind persons which can perform text detection thereby produce a voice output. This can help the visually impaired persons to read any printed text in vocal form. 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.,).
- Joao Guerreiro and Daniel Goncalves (2014), the authors proposed that the portable digital imaging Devices has created enormous technique in many domains by describing typical imaging devices and imaging process.in this paper, author proposed that screen readers have been an essential

tool for assisting visually impaired users over many years in accessing digital information. Making this information accessible is crucial and has been the object of intensive research. They made use of concurrent speech which enables visually impaired people to listen to several irrelevant information items, get the gist of the information and identify the ones that deserve further attention.

- J. Liang D and Doremann (2005), the authors proposed that the advanced technology can play the tremendous role using mobile computer vision that facilitates environment exploration for visually impaired persons. In this paper, author present a survey of application domains, technical challenges and the solutions for the analysis of documents captured by digital devices and the imaging process. They provided document analysis from a single camera-captured image as well as multiple frames and highest some sample applications under development and feasible ideas for future
- Alexandre Trilla and Francesc Alias. (2013), as the author published about the improvement of TTS Method in the case of input text reading and conversion of speech. the article is about the research of analysis of input feature for speech synthesis.
- S. Mascaro. and H. H. Asada (2001), when the conference for robot tutor was held the author introduced finger posture and shear force in easuremed: initial experimentation. Pitrelli J. and Bakis R. (2006), the author proposed the IBM expressive text-to-speech synthesis system for American English which is based on the article published about the audio, speech, language process.
- Priyanka Bacche. Apurva Bakshi, Krishna Ghiya. Prianka Gujar. (2014), they proposed the idea as NETHRA - new eyes to read artifact which was included in the Journal of science, Engineering and technology research.
- Rissanean, M. J. Fernando S. Pang N (2013), they introduced it as “Natural and Socially acceptable interaction technique for Ringer faces. Finger-ring shaped user interfaces” which was published in distributed ambient and pervasive interactions. Norman J. F. and Bartholomew A. N. (2011), the author proposed the Blindness Enhance Tactile Acuity and Haptic 3-D Shape Discrimination. Published in the article based on human interaction conference.
- Rohit Ranchal. Yiren Guo. Keith Bain and Paul Robinson J (2013), they proposed as using speech recognition for real time captioning and lecture transcription in the class room and the article was learning technologies.

3. Description

The concept of proposed system is the idea of developing specs reader based text reading system for visually impaired persons. There are three different modules in this system: Camera module, Optical Character Recognition Module and Text-To-Speech Module This explains the text reading system for visually impaired users for their self-independent.

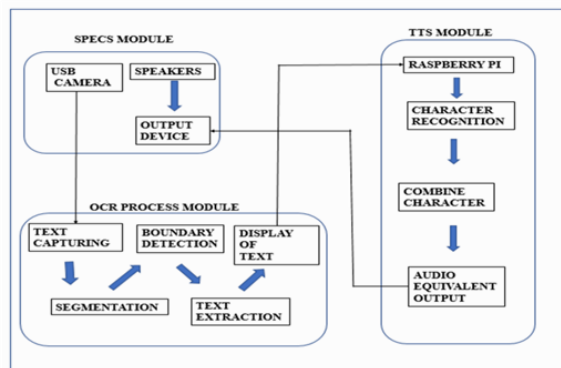


Fig. 1. Block diagram

4. Working principle

The system has been built on a Raspberry pi board running Raspbian OS and python/opencv libraries. Raspbian is a Debian based Linux distributed de-facto standard operating system, which comes with pre-installed peripheral units libraries. The texts in the form of (jpeg, png, jpg, bmp etc.) are considered for the analysis. The image which is captured from the pi camera is splitted in following conditions as described below for detecting corresponding text. The image captured is then processed and with the help of OCR/Tesseract the captured text was converted into letters which was made readable with the help of eSpeak software. We will also use ultrasonic sensor and tilt sensor to notify obstacles and fall detection. The fall detection is sent as a notification to concerned person using IoT.

The concept of proposed system is the idea of developing specs reader based text reading system for visually impaired persons. There are three different modules in this system: Camera module, Optical Character Recognition Module and Text-To-Speech Module.

A. Camera Module

All capture methods found in OpenCV (capture, capture continuous, capture sequence) have to be considered according their use and abilities. In this project, the capture sequence method was chosen, as it is the fastest method by far. Using the capture sequence method our Raspberry Pi camera is able to capture images in rate of 20fps at a 640x480 resolution.

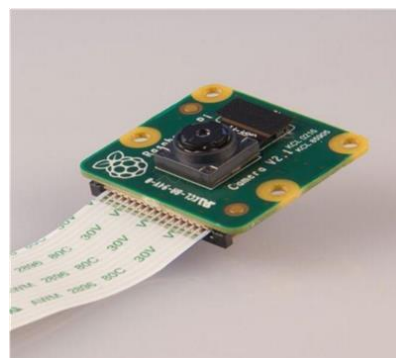


Fig. 2. Camera Module

B. Optical Character Recognition Module

In order to read the words in a paragraph accurately, Tesseract Optical Character Recognition (OCR) is chosen for the project.

Tesseract OCR was shown to outperform most commercial OCR engines. The processing within Tesseract OCR needs a traditional step-by-step pipeline. In the first step the Connected Component Analysis is applied, and the outlines of the components stored. This step is particularly computational intensive, however it brings number of advantages such as being able to read reversed text, recognizing easily on black text on white background. After this stage the outlines and the regions analysed as blobs. The text lines are broken into character cells with the algorithm nested in the OCR for spacing. Next, the recognition phase is set two parts: Adaptive classifier and Template classifier.

C. Text to Speech using eSpeak

eSpeak is a compact open source software speech synthesizer for English and other languages, for Linux and Windows which uses a "formant synthesis" method. This allows many languages to be provided in a small size. The main advantage of using eSpeak is that the speech is clear, and can be used at high speeds.

D. Hardware tools

- Raspberry pi
- Pi camera
- ESP 8266 Wifi Module
- Speakers
- Ultrasonic Sensor ADXL Tilt Sensor

E. Software Tools:

- Raspbian OS
- Python
- OpenCV
- eSpeak

- OCR Tesseract

5. Conclusion

A text detection and recognition with speech output system was with success incontestable on mechanical man platform. this technique is extremely handy and helpful for the purblind persons. Compared with a computer platform, the mobile platform is transportable and a lot of convenient to use. this technique is going to be useful for purblind persons to access info in written kind and within the encompassing. it's helpful to know the written communication messages, warnings, and traffic direction in voice kind by changing it from Text to voice. it's found that this technique is capable of changing the sign boards and alternative text into speech.

References

- [1] Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V (2017), "Smart Specs: Voice assisted Text Reading system for Visually Impaired Persons Using TTS method", IEEE International Conference on Innovations in Green Energy and Healthcare Technologies.
- [2] Alexandre Trilla and Francesc Alías. (2013). "Sentence Based Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.
- [3] S. Mascaro. And H. H. Asada. (2001) "Finger posture and shear force measurement: Initial experimentation," in Proc. IEEE Int. conf. robot. Autom, Vol. 2, pp. 1857-1862.
- [4] Pitrelli J. and Bakis R. (2006), "The IBM expressive text-to-speech synthesis system for American English", IEEE Trans. Audio, Speech, Lang. Process, Vol. 14, No. 4, pp. 1099-1108.
- [5] Priyanka Bacche. Apurva Bakshi, Karishma Ghiya. Priyanka Gujar. (2014), "Tech-NETRA (New Eyes to Read Artifact)," International Journal of Science, Engineering and Technology Research, Vol. 3, Issue. 3, pp. 482-485.
- [6] Rissanen M.J., Vu S., Fernando O.N.N., Pang N., Foo S. (2013) Subtle, Natural and Socially Acceptable Interaction Techniques for Ring Interfaces — Finger-Ring Shaped User Interfaces. In: Streitz N., Stephanidis C. (eds) Distributed, Ambient, and Pervasive Interactions. DAPI 2013. Lecture Notes in Computer Science, vol. 8028. Springer, Berlin, Heidelberg."
- [7] Norman J. F. and Bartholomew A. N. (2011), "Blindness Enhances Tactile Acuity and Haptic 3-D Shape Discrimination," Proceedings of the 4th Augmented Human International Conference, Vol. 73, No.7, pp. 23-30.