Smart Reading Glasses: Conversion of Image Text into Speech

[1]Nayana G H, [2]Sowmya.N, [3]Yaduguri Sravani, [4]Beulah James
1 Snr. Asst. Professor, Department of Electronics and Communication Engineering, New Horizon College of Engineering
2,3,4 Student, Department of Electronics and Communication Engineering, New Horizon College of Engineering, Outer Ring Road, Marathahalli, Bangalore- 560 103

Abstract: Reading is essential in today's society. Printed text is everywhere in the form of bank statements, restaurant menus, classroom handouts etc. Learners with visual impairments are finding it very difficult to survive in both education and employment. With the aim of helping visually impaired people with the emerging technologies, we have proposed a smart spec for blind persons which performs text detection and the detected text is converted into voice output. A spectacle having an inbuilt camera is used to capture the text image from the printed text. Tesseract, an open source optical character recognition (OCR) engine is used to extract the text from the image. The detected text is converted into speech output using compact open source software called eSpeak and the output is fed to an audio amplifier before it is read out. Raspberry pi is used as an interface between the softwares, keyboard and the image processing results.

Index Terms **-** Raspberry Pi, Tesseract, OCR, eSpeak

I. INTRODUCTION

According to World Health Organization, around 285 million people live with visual impairments worldwide, of which 45 million are blind. To assist the people with visual impairments, we put forth a smart model that effectively reads out black and white printed text. The structure is based on implementing image processing in an embedded system based on Raspberry Pi board. As a part of software development the OpenCV (Open Source Computer Vision) libraries are utilised to capture images and extract text from the images. The technology tools used are OCR (Optical Character Recognition) and TTS (Text-to-speech) engines.

Optical Character Recognition is a software tool that converts printed text into machine readable text. It basically extracts text from an image. It also helps users to convert books and documents into electronic files for use in storage and document analysis. The OCR engine ensures high speed image processing and accurate recognition capability makes it useful for unique identification, business card recognition, forms processing, and more. OCR makes it likely to apply text-to-speech techniques. Text-to-Speech system converts the text into speech output. The output is fed to an output device depending on the user's choice. Output can be heard through headphones connected to audio jack of raspberry pi.

II. BLOCK DIAGRAM

The block diagram shows three modules: the input and output module, the OCR module and the TTS module.

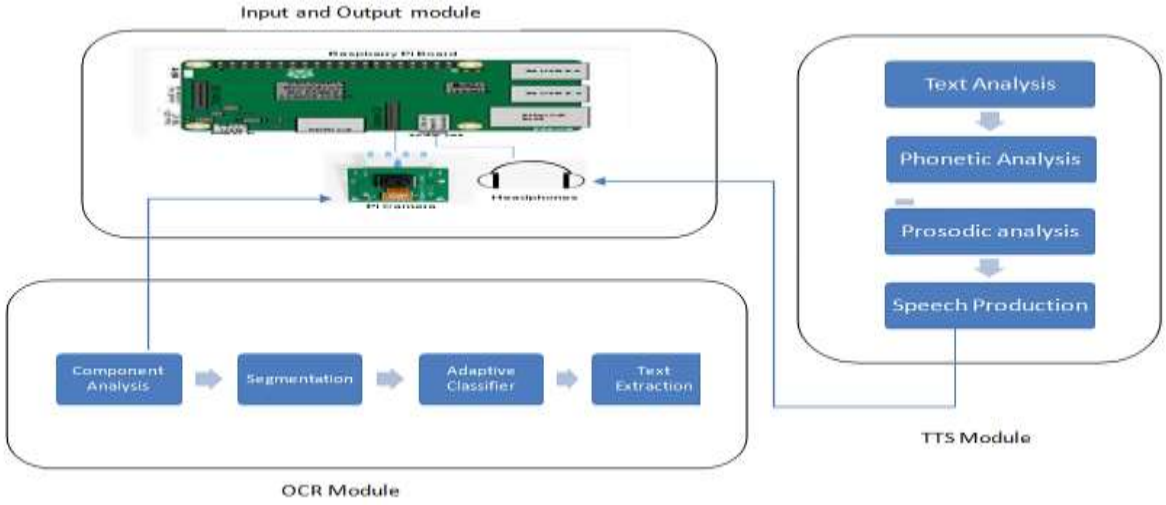

Fig.1 Block Diagram

The Raspberry Pi 3 Model B has a system on chip, a power supply, 4 USB ports, 1 Ethernet port, one audio jack, one CSI and one DSI and 40 GPIO pins. The input is given to Raspberry Pi using the CSI (Camera Serial Interface) port. The image is processed and the speech output is produced by the OCR module and the TTS module. The output is taken from the audio jack of Raspberry Pi board and is heard through headphones.

III. PROPOSED SYSTEM

The prototype consists of a Raspberry Pi and a Pi camera as the hardware and Tesseract OCR, OpenCV and eSpeak installed on the operating system on Raspberry Pi called Raspbian OS. Four major steps are involved in producing the voice output. They are given as:

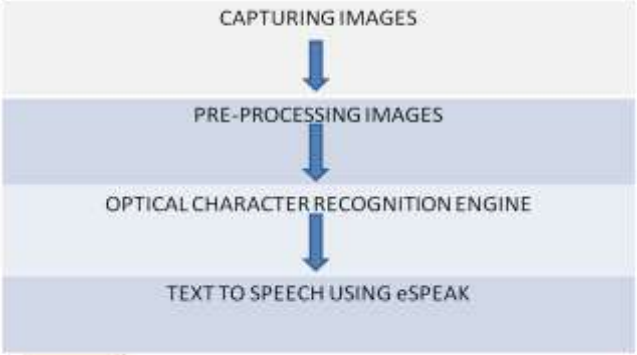

Fig.2 Process Flow

A. Capturing Images

The Pi camera connected to CSI port on Raspberry Pi board is moved on the printed text to be read to capture the image of the text. For a fast and clear recognition of text, we need a high quality image which can be achieved with the help of a higher resolution camera.

B. Pre-processing images

Pre-processing consists of two steps: Thresholding and Blurring.

Thresholding is the process of converting the gray scale image into a binary image by replacing each pixel with a black pixel if the image intensity is greater than a threshold and with a white pixel if the image intensity is less than a threshold value. Binary Thresholding algorithm uses Otsu's method which assumes that the image contains foreground and background pixels. It then calculates a threshold value such that the spread between both the classes (foreground and background pixels) is minimal.

Image filtering is done to remove noise from the image. Average filters, median filters or adaptive filters can be used. Median filtering is used because it is less sensitive to the extreme values. Median filtering, a nonlinear filtering technique is used for noise removal which determines the output pixel by calculating the mean of the neighbouring pixels.

C. Optical Character Recognition Engine

Tesseract is compact open source software used for optical character recognition. The first step in optical character recognition is component analysis which saves the outlines of the components. The advantage of this step is it is simple to detect inverse text and detect black and white text. Then the text lines are broken into words according to the character spacing. Text recognition is done in two phases. In the first phase, each word is passed through an adaptive classifier as training data. In the second phase, x-height normalization is performed to differentiate between capital and small text.

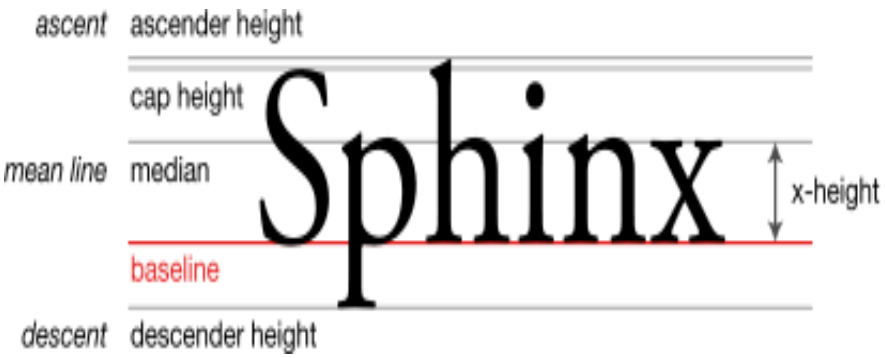

Fig.3 x-height normalization

D. Text to speech using eSpeak

Speech Synthesis converts printed text in an image into speech. An open source software eSpeak is used for this purpose. The steps involved in synthesis are: text analysis, phonetic analysis, prosodic analysis and speech production. The detected text is analysed by performing text tokenization and text normalization. Text tokenization is the process of fragmenting running text into words and sentences. Text normalization is the process of converting text into a form which it might not have before. The next step Phonetic analysis performs grapheme to phoneme conversion. Adding accent and melody to the phonemes is called prosodic analysis. Combining all these phonemes, the speech is generated.

IV. LITERATURE SURVEY

Previous work includes wearable stereo-vision devices [2] that can not only help avoid obstacles, but also stream real- time video feed via the 3G network. These help the blind to avoid obstacles and also share live video feed to a person who can guide them. Alternate methods for text to speech conversion have been used in models such as Multi-Domain TTS (MD-TTS) for synthesizing among different domains [1]. HMM-based TTS synthesis [5] allows higher flexibility to speech signal parameterization, but it is still not capable of achieving the typical high speech quality obtained by unit-selection concatenative approaches [7]. To create an effect while reading, domain-based expressiveness [4] was used in an advertising scenario, while affect-based expressiveness was used in news reading and storytelling.

Table 1: Literature Review

REFERENCE PAPER	OBJECTIVE	METHODOLOGY
[1]	Multi-domain Text to speech conversion	Represents text as directional weighted word based graph
[2]	TTS method	VoIP (Voice over Internet Protocol)
[3]	Text Extraction	Effective motion based algorithm Text localisation algorithm
[4]	Expressive TTS	Corpus driven approach Prosodic phonology approach
[5]	Speech Synthesis	HMM (Hidden Markov Model) Unit selection synthesis
[6]	Text Extraction	Histogram Segmentation Method Colour Clustering

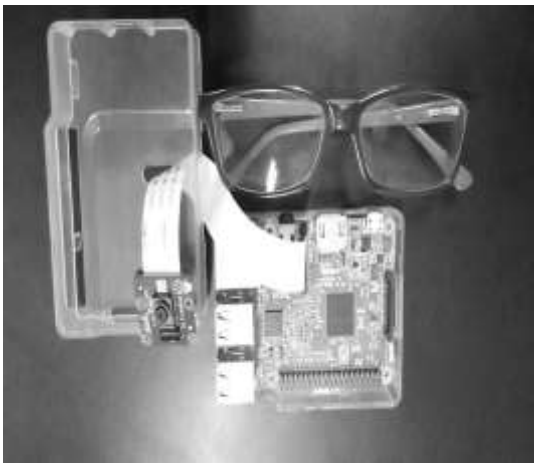

Fig.4 Hardware Implementation

V. IMPLEMENTATION AND EXPERIMENTAL RESULTS

The input to the prototype is an image taken by the pi camera attached to the CSI port of Raspberry pi. The images is pre-processed to remove noise and then converted into a binary image. Then the text is extracted and audio output is produced.

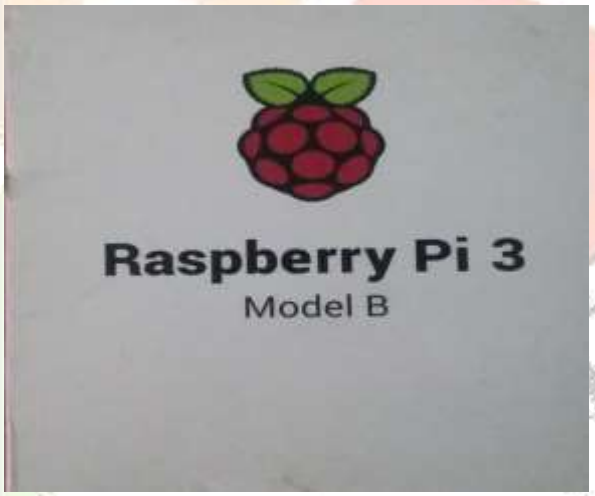

Fig.5 Original Image

Fig. 6 Gray Image

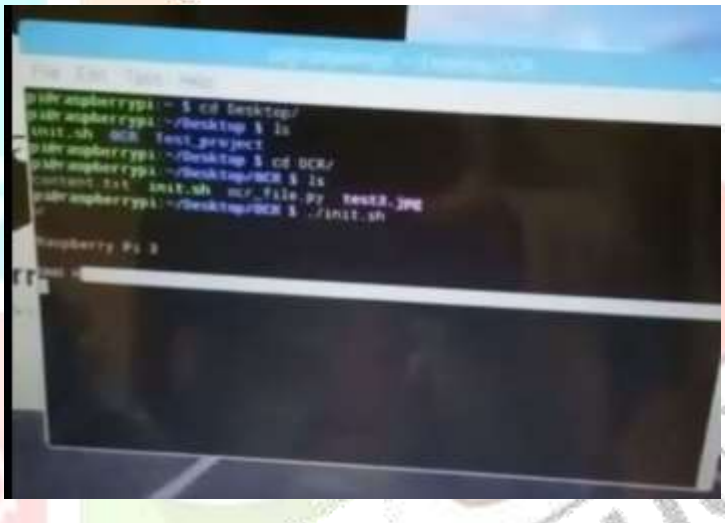

Fig.7 Recognized text displayed in the terminal

VI. Conclusion and Future Scope

This paper performs text to speech conversion using Raspberry Pi board. This can be implemented using MATLAB also. The audio output is clear and accurate. The main advantage of this model is that it is compact and portable which makes it an efficient device to support visually impaired people.

References

[1]. Alías F. Sevillano X. Socoró J. C Gonzalvo X. (2008), „Towards high-quality next-generation text-to-speech synthesis", IEEE Trans. Audio, Speech, Language Process, Vol. 16, No. 7. pp. 1340-1354.

[2]. Balakrishnan G. Sainarayanan G. Nagarajan R. and Yaacob S. (2007) „Wearable real-time stereo vision for the visually impaired", Vol. 14, No. 2, pp. 6–14.

[3]. Chucai Yi. YingLiTian.AriesArditi. (2014), „Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons", IEEE/ASME Transactions on Mechatronics, Vol. 3, No. 2, pp. 1-10.

[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]. Shinnosuke and Takamichi (2014), „Parameter Generation Methods With Rich Context Models for HighQuality and Flexible Text-To-Speech Synthesis" , IEEE Journal Of Selected Topics In Signal Processing, Vol. 8, Issue. 2, pp. 239-250 .

[6]. Tapas Kumar Patra and Biplab (2014) „„Text to Speech Conversion with Phonematic Concatenation", International Journal of Electronics Communication and Computer Technology (IJECCT) Vol. 2, Issue. 5. pp.223-226.

[7] A. Black, H. Zen, and K. Tokuda, "Statistical parametric speech synthesis," in Proc. ICASSP, Honolulu, HI, 2007, vol. IV, pp. 1229–1232.

[8] Rohit Ranchal .YirenGuo . Keith Bain and Paul Robinson J (2013), "Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom", IEEE Transactions On Learning Technologies, Vol. 6, No.4, pp. 12-17.

[9] Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method Ani R, Effy Maria, J Jameema Joyce, Sakkaravarthy V, Dr.M.A.Raja.