International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Development of Text to Speech Conversion System for Low Vision and Blind People

P. Vineetha, Dr. L. Padma Sree, M. Rama Devi

Abstract*: Around the world 285 million individuals are found to be visually challenged out of 7.4 billion populations found in a survey made by World Health Organization. These people face many problems but the major problem is reading. It is observed that they cannot read the text which is not written in braille. In the thought process of supporting them, here is a framework proposed for the visually challenged people which can perform content recognition and produce voice yield. This can assist the visually challenged people with reading any printed content and convey in speech output. A camera is utilized to capture the content from the printed content and the captured picture experiences progression of picture pre-preprocessing steps to get the content of the picture and expels the background. Characters are identified utilizing Tesseract-Optical Character recognition (OCR). The identified script is then changed into voice, utilizing open source speech synthesizer (TTS). Finally, the speech output is heard by the earphones.

Keywords*: Embedded Processor, Tesseract OCR, Speech Synthesis (TTS).*

I. INTRODUCTION

In a Survey made by the World Health Organization it is observed that 285 million individuals are assessed to be visually challenged around the world. Among the 285 million, 90 Percent live in advanced nations. Visually challenged and visual impaired individuals might be old or youthful.

There are many alternative solutions for the issue of helping people who are not able to read. The issue with the solutions is that they are not cheap and their execution is very complicate. Visually challenged individuals experience various troubles with reading the printed content utilizing existing innovation, it includes problems like arrangement, exactness, portability and performance. A gadget that helps the visually challenged which is effectively and efficiently reads the paper-printed content is designed.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
M. Rama Devi*, Assistant Professor, Department of Electronics and Communication Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.
Dr. L. PadmaSree, Professor, Department of Electronics and Communication Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.
Ms. P. Vineetha, Department of Electronics and Communication Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.

The Reader employs the philosophy of web cam-based gadget that can be utilized by individuals to recite script file. It uses combination of image processing strategy and embedded systems.

The system is designed with basic arrangements for visually challenged individuals. The arrangement has speaker read out for voice yield. The Reader has a camera as an input device to feed the printed text document. The obtained picture is digitized and the scanned record is dealt with a software module called the Optical Character Recognition.

The Open source platform libraries of Open CV are employed to capture the script and recognize characters. All the gadgets designed for visually challenged are based on the two essential strategies i.e. OCR and TTS.

Optical character recognition is the transformation of captured picture content into machine encoded content. OCR is a procedure that makes computer to perceive composed or printed characters, for example, numbers or letters and to transform them into a machine encoded form.

It is helpful for low vision people for getting the content of the text file. It is moreover used to digitize and impersonate writings that have been made with non-electronic system. Digitizing writings also diminishes the storage space. It is commonly used to change over books and archives into electronic records for capacity and document analysis.

OCR procedure is likewise applied in machine understanding, text to-speech and content mining to the catch or separated page. The perceived content archive is sent to the yield gadgets dependent on the choice of the customer. The yield contraption can be a headset or speaker related with the Raspberry PI board or a speaker which can illuminate the content report so anybody may hear.

OCR technique is also applied in machine interpretation, text-to-speech and text mining to the capture or filtered page. The recognized text document is sent to the output devices based on the decision of the user. The output gadget can be a headset or speaker associated with the Raspberry PI board or a speaker which can illuminate the text document so anyone might hear.

II. RELATED WORK

Selecting the usage of the assistance devices for the visually challenged is done by various strategies. It is discovered that the old style or customary strategies appeared to be tricky, expensive and inaccurate. The following are the various methodologies received in various text conversions, speech processing and text processing. Aaron James, designed an OCR Based automatic book reader for the visually impaired using Raspberry Pi.

It utilizes Energy Efficient Character Recognition, document image analysis and python programming for execution.

The reader is expensive in usage [1].D.Dakopoulos, presents wearable obstruction avoidance electronic travel aids for blind. Relative review among convenient or wearable obstruction detection and avoidance frameworks to educate about the advancement in assistive innovation for visually impaired individuals is taken [2].N. Giudice, adopted Blind navigation technology and clarified about the navigational innovations accessible to blind people to help autonomous travel on huge scale [3].Kumar S, extracted and documented the text image using Matched wavelets and MRF Models. The Method has a novel plan for the extraction of literary territories of a picture utilizing all around coordinated wavelet channels. A grouping based method is conceived for assessing comprehensive coordinated wavelet channels utilizing an arrangement of ground truth pictures [4].K Kim, proposed a Texture-based approach for text detection in images utilizing support vector machines and consistently adaptive mean shift algorithm for the textural properties of contents [5].Chen X, distinguished and read the texts in natural scenes. The author proposed an algorithm for detecting and reading the text from the city scenes as they stride through [6].

III. METHODOLOGY

The proposed framework incorporates both hardware and software. The hardware is composed of a Web Cam, Push button, Raspberry Pi, Monitor, and a speaker as shown in Figure1. The Raspberry Pi run the Raspbian Jessie OS 4.4, with OpenCV 2.3 and with Python 2 for compilation.

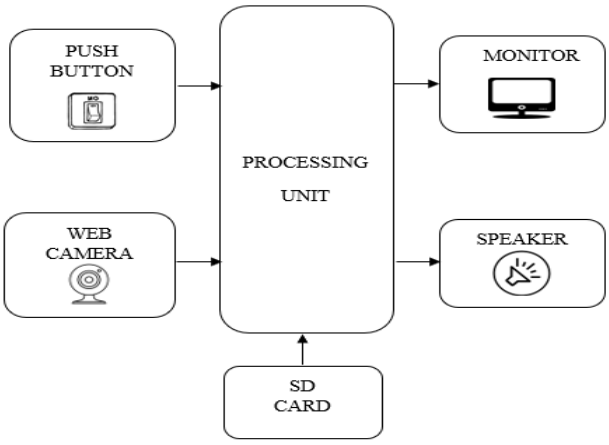


Figure1: Block Diagram

The software module consists of image acquisition, image pre-processing, segmentation, Feature Extraction, Classification, Linguistic Analysis, Speech Synthesis which are performed inside Raspberry Pi as shown in Figure2.

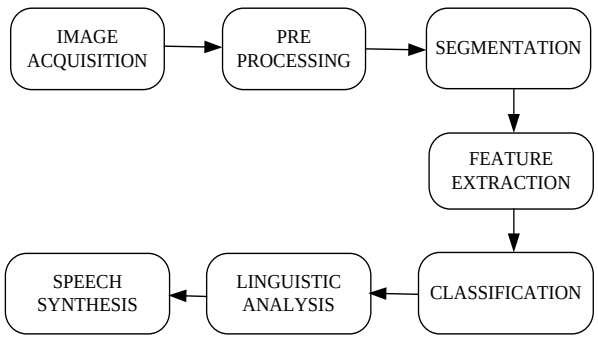


Figure2: Operation of Tesseract OCR

A. Raspberry Pi

The Pi board has a Broadcom BCM2836 (SoC), which includes an ARM1176JZF-S 700 MHz processor, Video Core IV GPU, and was initially worked with 512 megabytes of RAM.

B. Acquisition

Acquisition of sequence of images is done by camera with resolution of 1280*720 Pixels.

C. Pre-Processing

The acquired color image is transformed into grayscale image and then it is transformed to binary image. In pre-processing elimination of noise is done with median blur filtering, which increases the accuracy of the image.

D. Segmentation

Segmentation is nothing but splitting the entire image into sub-parts to process them further as shown in Figure3. The image is segmented in the following sequence: Line Segmentation, Word Segmentation and Character Segmentation.

Figure3: Segmentation

Histogram projection technique is used for all these stages of segmentation to decide where the segment should be done. In this technique after converting the colored image to binary representation, there will be only black and white pixels in the image, i.e. white pixels indicate the presence of foreground pixels and black indicates no foreground pixels.

There are two methods in histogram projection they are horizontal histogram projection, vertical histogram projection. Line segmentation is carried by horizontal projection. word segmentation and character segmentation are carried by vertical projection.

Text pixels have higher peaks, because of more no. of foreground pixels and the spacing between lines, words and characters has lower peaks because of more no. of background pixels by this image is segmented.

E. Feature extraction

Each character is defined as a feature vector in the extraction stage, which becomes its identity. Extraction of features can be described as seeking a set of parameters (features) that determine accurately and uniquely the shape of the underlying character.

It is based on two types of features: Structural approach and Statistical approach. For accurate output structural approach is used in the system.

Character strokes, character holes, end points, intersections between lines, loops,

and other character attributes such as concavities can be described as structural features.

F. Adaptive Classifier

The classification stage assigns labels to character images based on the extracted features and the relationships between the features. This section of the OCR eventually recognizes outputs of individual characters in a machine editable form.

The use of an adaptive classifier benefits OCR engines. Tesseract has input from adaptive classifier that is prepared utilizing static classifier yields that are broadly used to accomplish more prominent segregation in each record. Adaptive classifier is increasingly compelling and give the best outcome.

G. Linguistic Analysis.

Tesseract contains relatively little linguistic analysis. The linguistic module chooses the best available word string in each of the following classifications: Top recurrence word, Top lexicon word, Top numeric word, Top Upper-case word, Top lower-case word (with optional capitalized), Top classifier word.

H. Speech Synthesis (TTS)

A text-to-speech (TTS) framework changes over ordinary language script into speech whereas different frameworks render emblematic semantic portrayals like phonetic translations into speech.

Festival is an open source programming speech synthesizer for language like English and also many other languages. This allows various dialects to be given in tiny size. The standard preferred position of using festival is that the discourse is clear and can be used rapidly.

I. Flowchart

The flow of the system is as follows shown in Figure4.When the system is powered up the camera is ready to take the text image. If the person wants to read the text from the image, an interrupt is given by the button Raspberry Pi instructs the camera to take the picture.

Then camera takes few moments to record the picture and send it to the processor. The noise and background subtraction were eliminated by image processing. The picture undergoes through Optical Character Recognition technique to extract the text. The processor translate text into voice and it is heard through speaker.

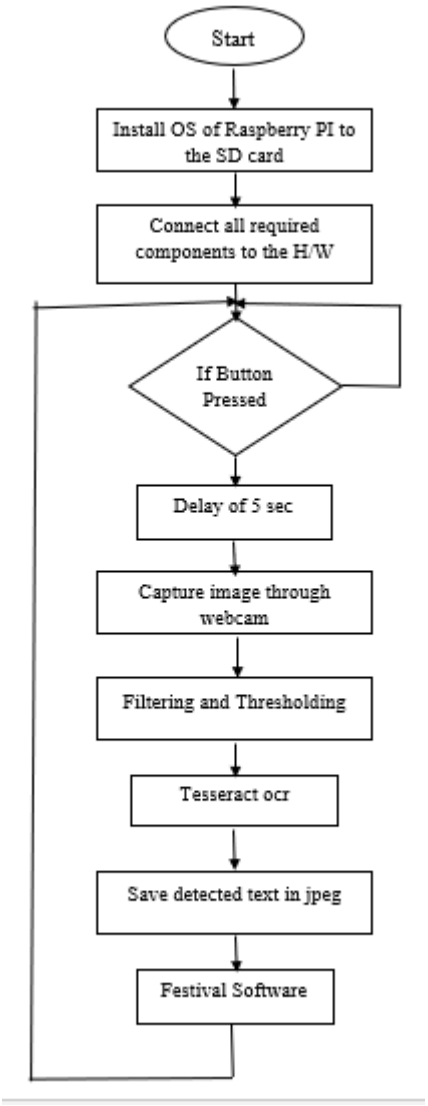
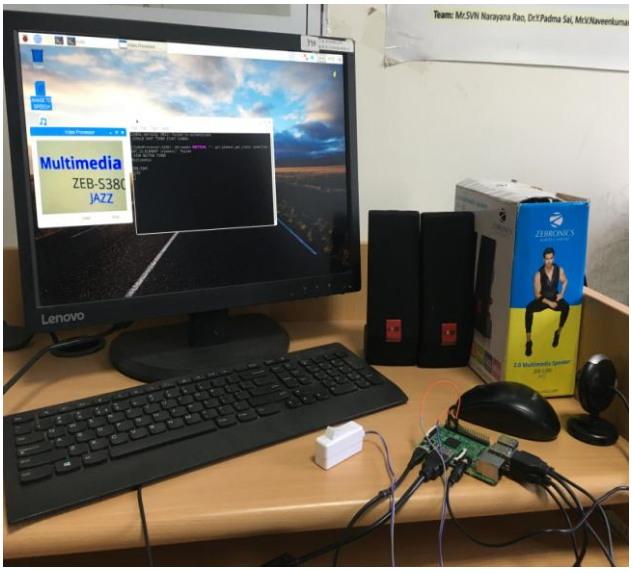
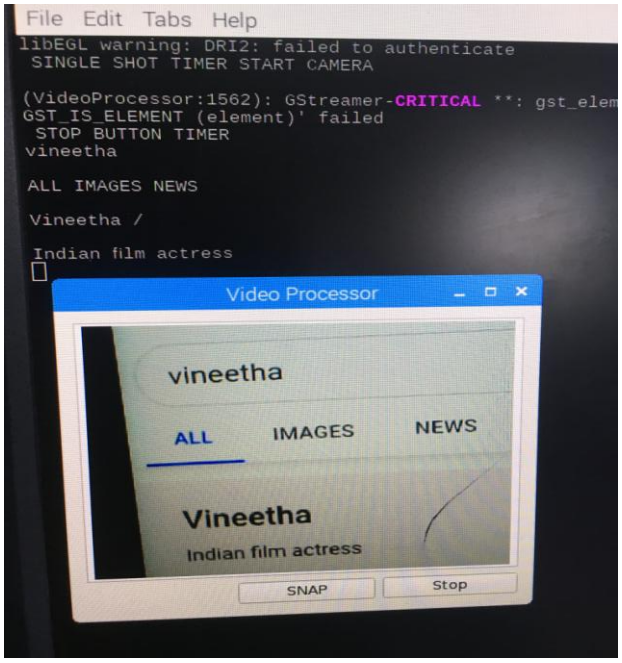


Figure4: Flowchart

IV. RESULTS



a. Hardware Setup



(b)

Figure4: Real time Text Detection

An evaluation exercise is conducted to verify effectiveness of the developed text reader. After the system is powered up it takes 5seconds for camera focusing. The font size of character should be minimal of 12. It can recognize alphabets and numbers.

A. Variation of Text

The initial assessment includes the readability check utilizing various textual styles with changing background colours. For this activity, the sample phrases is shown below.

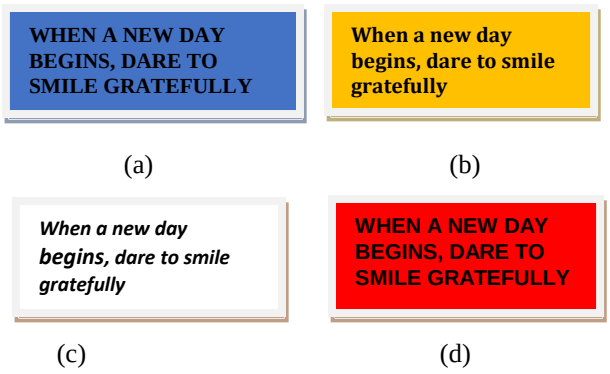


Figure5: Tests of few text styles with different background colors. (a)Times new roman, (b) Cambria. (c) Calibri (Italic), (d) Arial. The fonts used were Arial, Times New Roman, Cambria, Calibri (Italic). The background colors were white, yellow, red, and Blue Figure5. The percentage of error is calculated by using the formula, characters misspelled by total characters in the sentence as shown in below equation.

$$\%\text{Error}=\frac{\text{Number of Alphabets or Characters misspelled}}{\text{Total number of Characters in the sentence}}\times 100$$

The outcome of this evaluation exercise is provided in Table 1, indicates error percentage. As far as PC produced textual styles, Arial gives the best output.

Table1: Error in recognizing text with various textual styles utilizing different background colors.

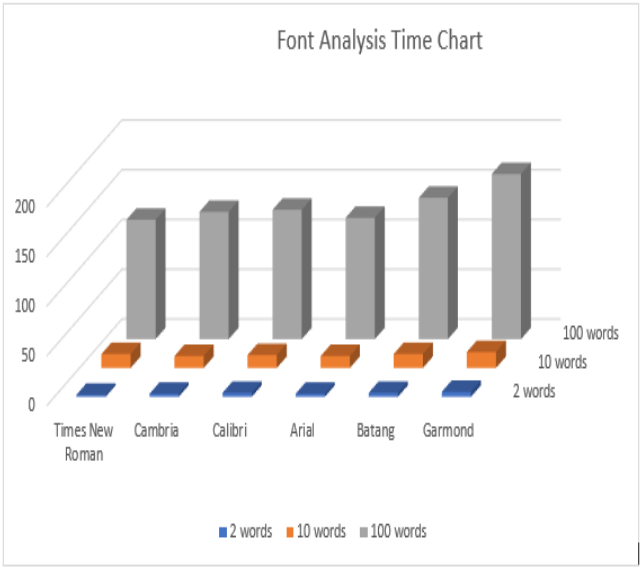
Font	Error %yellow	Error %White	Error %Blue	Error %Red
Times New Roman	0	0	8	0
Cambria	4	1.33	2.67	4
Calibri	1.33	2.67	4	0
Arial	0	0	0	0

B. Processing time for Reading Text.

One of the vital problems is the absolute processing time. Table1 show processing time with white background. With different background colour it takes more time because when the image is captured the background colour is reflected so the image will be blurred. To remove the noise, it takes more time for processing. As the characters are increased the processing time also increases.

Table2: Processing Time

TYPE OF FONT	DURATION TO READ(SEC)		
	2 Words	10 Words	100 Words
Times New Roman	2	14	120
Cambria	3	12	128
Calibri	4	13	130
Arial	3	12	122
Batang	4	14	142
Garmond	5	16	166



V. CONCLUSION

The implemented system read printed text for assisting the blind people. This translation does not require any internet connection. So, it is a simple device to utilise for blind. A small equipment is carried out by the blind person like a camera, an installed processor board, and an earphone or speaker. To make the device flexible a battery is needed to power up the system.

REFERENCES

1. Aaron James S, Sanjana S and Monisha M, OCR based automatic book reader for the visually impaired using Raspberry PI, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 7, January 2016, 1111-1118.
2. Dimitrios Dakopoulo Nikolaos G. Bourbakis, Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) ,Volume:40 , Issue: 1 , Jan. 2010 , 25-35.
3. Giudice, N. A., & Legge, G. E, Blind navigation and the role of technology. In A. Helal, M. Mokhtari & B. Abdulrazak (Eds.), engineering handbook of smart technology for aging, disability, and independence (pp. 479- 500): John Wiley & Sons.
4. Sunil Kumar, Rajat Gupta , Nitin Khanna, Santanu Chaudhury and Shiv Dutt Joshi, Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model, IEEE Transactions on Image Processing (Volume: 16 , Issue: 8 , Aug. 2007) 2117 – 2128.
5. Kwang In Kim, Keechul Jung, and Jin Hyung Kim, Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 12, DECEMBER 2003, 1631-1639.
6. Xiangrong Chen, Alan L. Yuille, Detecting and Reading Text in Natural Scenes, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.

AUTHORS PROFILE

Ms.P.Vineetha received her Bachelor of Engineering in Electronics and communication from Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, India in 2017. Her research of interest is in Embedded system. She is pursuing Master of Technology in Embedded system from VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.

Dr.L.PadmaSree completed Ph.D from the specialization of Computer Security in 2010 and M.E from Digital systems & computer electronics in 1997.She is working as a professor in the department of Electronics and communication Engineering at VNR Vignana jyothi institute if engineering and technology, Hyderabad, Telangana, India.

M.Rama Devi completed ME from Embedded Systems she is working as a assistant professor in the department of Electronics and communication Engineering at VNR Vignana jyothi institute if engineering and technology, Hyderabad, Telangana, India