PROPOSED SMART SPECS: VOICE ASSISTED TEXT READING SYSTEM FOR VISUALLY IMPAIRED PERSONS USING TTS TECHNIQUE

M.Thillai Rani

Department of Electronics and Communication Engineering
Sri Krishna College of Technology, Coimbatore

K.P.Sai Pradeep

Department of Electronics and Communication Engineering
Dr.N.G.P. Institute of Technology, Coimbatore

R.Sivakumar

Department of Electronics and Communication Engineering
Muthayammal College of Engineering, Namakkal

Abstract- World Health Organization has given an estimation of visually impaired amid worldwide people in which out 285 million people are affected out of 7.4 billion populations. Hence there is a great necessity for taking adequate measure by utilizing emerging technologies for leading day today life successfully without any difficulty which is considered as a highly challenging. Smart specs are suggested primarily for supporting them: Text to Speech (TTS) System is employed for Visually Impaired Persons for reading the passage without visually via voice technology which is useful for blind persons for text detection thus yielding a voice output. The printed text can be read by visually impaired persons in vocal form. Raspberry Pi is considered as crucial part offering an interface amid camera besides image processing outcomes. Text image is captured from printed text besides captured image by specs inbuilt camera and Optical Character recognition (OCR) is greatly exploited for analysis. A compact open source software speech synthesizer, eSpeak is employed for detected text conversion. As a final point, synthesized speech is obtained by headphone by Text-To-Speech (TTS) technique.

Keywords – OCR, Voice Technology, Text Information Extraction, TTS, e-Speak, Reading System, Impaired Persons

I. INTRODUCTION

Screen Readers is been considered as a crucial tool for blind users assistance in reading digital information for many years, report has been given National Health Interview Survey (NHIS) in 2008 Visual impairment estimates has been ensued at worldwide and in six WHO Regions. The visual deficiency and blindness pervasiveness are obtained for 6 WHO regions for three economies' groups: Low Income (Rural), Upper Middle Income/High Income (Urban) along with total population for diverse age groups. The existing data sources are considered to be reliable with age groups [1].A smart spec reader based text reading system for producing voice output for visually impaired persons is the key notion of suggested system. The system comprises of an assistive text reading prototype scheme. The self-dependency of visually impaired users is considered as chief importance of visually impaired system which is greatly stressed [2]. Visually impaired people are facing various challenges for surviving in education and employment field. For eliminating this issue in society [3], OCR software integration is done for offering various functions such as scanning, text recognition and integrated voice output.

TTS technology is mainly meant for computer to speak which utilizes natural language processing. Various techniques are available in text reading applications however text to speech can be performed through creating datasets. Sound into text conversion was done by the TTS [4]. Textual info yet still stays typical info exchange function, irrespective of technology progression. Practically individuals may well reinstate regular perspective with eye cups or maybe communication lenses are approximately twenty percentage from ABF survey run into their normal lives of theirs. Besides them ninety percent of the earth's visually damaged individuals who are now living in the lower, center also maybe even in many identified nations, cataract rests the primary loss of sight lead to [5]. The believed amount of individuals visually damaged around the earth is 285 billion and 39 billion people are

possessing very low vision; sixty-five percentage of folks visually impaired as well as eighty-two percentage of the oblivious are forty years and above are noted in Table 1.

Table 1 Blindness Prevalence of WHO (2016)

Types of Blindness	Urban	Rural	Total
Total Blindness	4.43	5.99	5.40
Economic Blindness	11.14	15.44	13.83
One eye Blindness	8.23	7.00	7.46

Visually impaired people suffer from text documents accessing in various circumstances like reading text lively and accessing text in lesser than perfect environments. The study goal is prompting oblivious owners for analyzing created and printed written text by coming in contact with as well as getting speech result within living fashion. Non-Verbal and haptic cd cues are generally deployed for leading user's finger along with every series. 2 important methods especially Optical persona recognition for Text Information Extraction (TIE) as well as TTS for written text transformation to speech are significantly necessitated for improving a method for visually damaged individuals [6]. Textual content Information Extraction may be the major feature for just about any assistive reading through structure and it is a significant OCR process because this treatment describes paper speech intelligibility [7].

Text-to-speech quality along with capability extension helps in producing expressive synthetic speech [8] [9]. Partitioning video blocks directly into the copy as well as nontext areas is done through video clip textual content recognition in addition to removing wearing automated fashion. The breakthroughs attained within laptop eyesight, digital camera models, as well as pcs aids within creating digital camera dependent items merging laptop perception engineering with various prevailing important products such as for instance optical persona recognition methods. OCR has a tendency to include excellent recognition fee mostly intended for typewritten photographed photos or even is printed textual content electric change in computer-readable textual content.

Visually impaired people are facing various challenges for surviving in education and employment field. For eliminating this issue in society [10]. OCR software integration is done for offering various functions such as scanning,text recognition and integrated voice output and this might be used through normal person for reading massive document in a short time span. Good accuracy rate as well as performance device has not been established so far as far as concerning related research work.

Additionally, existing research work prolongs work towards expansion of collecting information at ease and self-dependent. For sticking to the desired final result, modality replacing framework pools dissimilar modules established, like spectacle reading through Tesseract-OCR component. Numerous modules within this method: Camera component, Tesseract Optical Character Recognition Module as well as TTS Module that elucidates textual content reading through method for visually damaged owners for the self-in-dependency of theirs.

II. RELATED WORK

In [11], Alías et al (2008) outlined the latest innovations in high-quality next generation text-to-speech (TTS) synthesis systems development. The main key aspects of this research arena are speech expressiveness besides synthesis enhancement flexibility. A novel stratagem named multidomain TTS (MD-TTS) for synthesizing amongst diverse domains is introduced in this research. While multi domain philosophy has been extensively utilized in spoken language systems, limited research determinations have been shown for TTS field extension. Initially, text classifier (TC) is exploited for traditional TTS architecture for inevitably directing most suitable domain selection for input text synthesizing. In contradiction to typical topic text classification tasks, text contents are not considered along with text content by MD-TTS TC. A new text modeling system on basis of an associative relational network helps in representing texts as a directional weighted word-based graph. Thereby it is validated through experimental outcomes pertaining to objective (TC efficacy) along with subjective (perceived synthetic speech quality) assessment conditions.

In [12], Yi et al (2013) recommended an interesting textual content localization algorithm via mastering gradient facets of stroke orientations aside from advantage pixels distributions within an Adaboost design. Textual content figures binarization is carried out in localized textual content areas and also recognized via off-the-shelf optical persona recognition software program and then recognized textual content codes are paper to oblivious owners in deep speech. The suggested textual content localization algorithm efficiency numerical evaluation is completed on ICDAR-2011 and ICDAR-2003 Robust Reading Datasets. The algorithm is validated via experimental results that outshine conventional techniques. It's additionally evaluated on a dataset gathered up by 10 oblivious individuals for system's hardware strength evaluation. The operator user interface challenges are checked out in addition to algorithm robustness evaluation within taking away as well as reading through written text out of several items with multifaceted circumstances.

In [13], Joao Guerreiro et al (2007) elucidates that people ability in focusing distinct speech source amid numerous conversations, nonetheless yet relevant content in background has to be identified. Consequently, in contrast to sequential speech channel, it is hypothesized that simultaneous speech channels leveraging are done for rapidly obtaining digital information gist. The experimentation is performed with 23 participants which greatly supports in understanding blind people's capability in searching pertinent. The outcomes obtained infers that pertinent source identification with two and three concurrent talkers can be easily performed. Furthermore, both two and three sources might be utilized for understanding pertinent source content on basis of task intelligibility demands besides user characteristics but lower accuracy is only attained.

In [14], Balakrishnan et al (2014) suggested reading printed text on hand-held objects for supporting blind persons. For solving general issue for blind users, a motion-based technique has been suggested for object of interest detection while blind user merely wobbles object for two seconds. In camera view, object of interest from background or other objects can be easily differentiated by this approach. Text regions extraction from composite backgrounds can be attained by text localization algorithm on the basis of suggesting stroke orientation and edge distributions models. The text global structural feature at every pixel is estimated by conforming feature maps and assistive text reading prototype system is suggested in which handling non-horizontal text strings is considered to be great challenge.

III. PROPOSED METHODOLOGY

3.1 Text Reading System -

Text Reading System block comprises of three parts specifically Specs in-built camera module, the hybrid scheme technology for impaired persons is exposed in Fig 1. An in-built camera is used to design spectacle reader device along with a Raspberry Pi affording an edge among camera, antennas, and Processing of Image's outcomes.
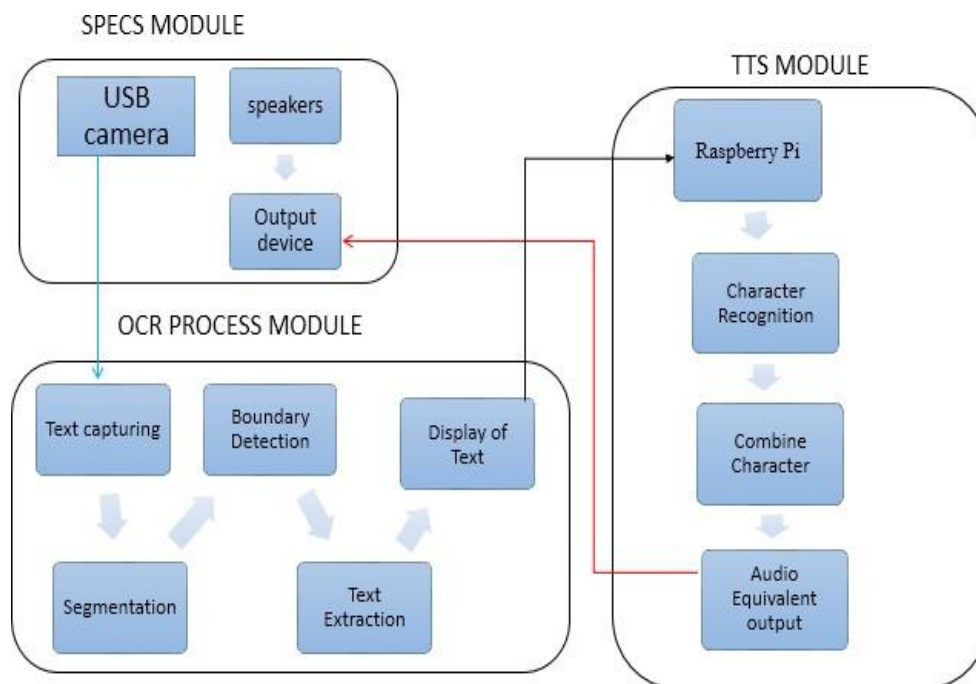

Figure 1. Proposed Hybrid Reading Arrangement

SPECS-MODULE: CAPTURING AND PROCESSING IMAGE

A Raspberry Pi has the ability in capturing an images sequence promptly through its video imprisonment utilization. Nonetheless, numerous facts must be considered before image capturing:

- Preferred image resolution/size may not be obtained when video recording is accomplished by video-capture port at all time for instance (distorted, rotated, blurred etc.).

- There is no adequate information such as coordinates, time about bagged images and those are not interchangeable.

- Bagged Images are deliberated based on their usage as well as capabilities.

- Capture sequence technique was deployed due to its increased speed.

- Raspberry Pi camera has image capturing capability in20fpsata640×480 resolution rate by capture sequence technique. The bandwidth is the main issue with Raspberry Pi when capturing images promptly. Raspberry Pi input output bandwidth is always and format makes process even less efficient which it is accessing pictures

- Raspberry Pi input output bandwidth is very restricted, besides format for accessing pictures creates process even less proficient. Furthermore, SD card size ought to be huge ample.

3.2 Image Processing -

Multithreading structuring is preprocessing algorithm phase for pictures as a result of limits interested in Raspberry Pi Input/output Bandwidth. A picture from logi tech digital camera needs to be shot largely consequently processing it. Picture queues are taken care of by Raspberry Pi as well as refined as shot pictures. Raspberry Pi's impression processing algorithm should work faster compared to recording videos frame fee, for not stalling encoder. In addition, good synchronization needs to be made certain via unique focus. At this moment GIL (Global Interpreter Lock) is tough to learn in Python in deep contradiction with low-level languages' multithreading. Python language can perform more processing on image to filter the exact area and it can perform the conversion on time [15]. A multi-threaded program is significantly necessitated in addition boosts a lot more intricacy within improving as well as debugging when as opposed with the single-threaded counterpart of its. A psychological item is ideal for sequential programming is learned in facilities for sequential programming and that mismatches with parallel delivery version. Worldwide Interpreter Lock is in fact supporting for guaranteeing consistency amid method of contemplating as well as amid threads.

OCR MODULE ENGINE

OCR is an optical recognition component that is primarily intended for typed, handwritten or even created and printed textual content pictures electric change in the machine-encoded textual content. It's the normal method of digitizing created and printed written text to ensure that it can certainly be electronically edited, browsed, saved much more heavily, shown online as well as utilized in printer procedures like textual content to speech. OCR is generally an offline procedure for examining a fixed paper. Handwriting campaign evaluation could be used as a type-in to hand composing recognition. For a substitute just utilizing glyphs as well as text styles, this particular method is adept to grabbed movements, as an example the purchase within what sections are pulled, pattern and direction of placing the dog pen lower and also picking up it. This particular additional info affords a much more exact end-to-end procedure. This particular concept can also be called online persona recognition, true period persona recognition as well as smart persona recognition [16].Fig 2 presents detection text flow by tesseract-OCR.

Figure 2. Flow of Detection of text using Tesseract-OCR

OCR is tremendously employed in this specific study for term reading through inside a part effectively. Tesseract OCR is a receptive supply motor started as a Ph.D. analysis task at HP Labs, Bristol. Tesseract OCR was demonstrated to outclass the majority of industrial OCR engines looking for a joint task with HP Bristol Labs as well as the HP scanner division. A customary step-by-action pipeline is necessitated for processing within Tesseract OCR. This particular stage is predominantly computational challenging, nevertheless, it requires a variety of advantages including becoming smart within reading through turned around textual content, determining smoothly on the text that is black on white-colored history [17]. Next point, outlines & areas are analyzed as blobs. The content collections are fragmented directly into persona cells with algorithm nested found OCR for spacing.

After, the recognition stage includes 2 components. Each term qualifies to an adaptive classifier, along with an adaptive classifier does textual content recognition much more exactly. An adaptive classifier was suggested for OCR engines use within figuring out the font that is non-character or character. In contrast to most OCR engines, Tesseract doesn't use a template (static) classifier. The main dissimilarity amid a template classifier as well as an adaptive classifier is, which adaptive classifiers deploy baseline x-height normalization while fixed classifiers acquire figures roles on the time frame of sizing normalization.

The baseline x-height recognition supports top situation, reduced situation figures as well as digits accurate recognition and detection, nevertheless it necessitates additional computational energy of the reaction. By default, Tesseract OCR makes use of Otsu's thresholding way to every picture, nevertheless, as it has a customized pre-processing algorithm, this particular phase was bypassed for quickness enhancement. For disabling Tesseract OCR inner thresholding; tesseract delegate choice is established as "self".

TTS MODULE

eSpeak is a small receptive supply software program speech synthesizer for Other and English languages, for Linux as well as Windows using "formant synthesis" strategy exactly where a number of languages may be within situation within a tiny measurement. The primary good thing about utilizing eSpeak is distinct speech, as well as could be exploited at higher speeds. eSpeak is provided as being a command type plan to talk written text coming from stdin or file plus shared library edition for using by some other applications. Raspbian OS uses Advanced Linux Sound Architecture (ALSA) for sound systems management.

The succeeding can be installed using the packages mentioned below:

```
apt-get install alsa-utils
```

```
apt-get install mpg321
```

```
apt-get install lame
```

```
modprobe snd-bcm2835
```

Proposed modules are installed using subsequent command:

```
sudo apt-get install espeak
```

```
sudo apt-get install espeak python-espeak
```

The module performs transformed text conversion into audible form. The on-board audio jack plays a vital role in Raspberry Pi and generated using a PWM output along with negligible filtering. There is no existing product as per the knowledge in the form of a compact specs module that helps blind people for text reading and understanding in the form of voiceoutput.

FLOW CHART OVERVIEW

Initial step is mainly importing image besides RPi GPIO and set GPIO pin 2 and 3 as input. The main program offers functions for retrieving and processing input image, thereby text into a speech signal conversion is done. Picture will be obtained or accuracy improvement. Fig 3 illustrates the entire program flow describing text detection workflow and conversion of text to speech with dissimilar input sets as black and white image, a colored image, text with image, text and image merged and text with dissimilar font styles.
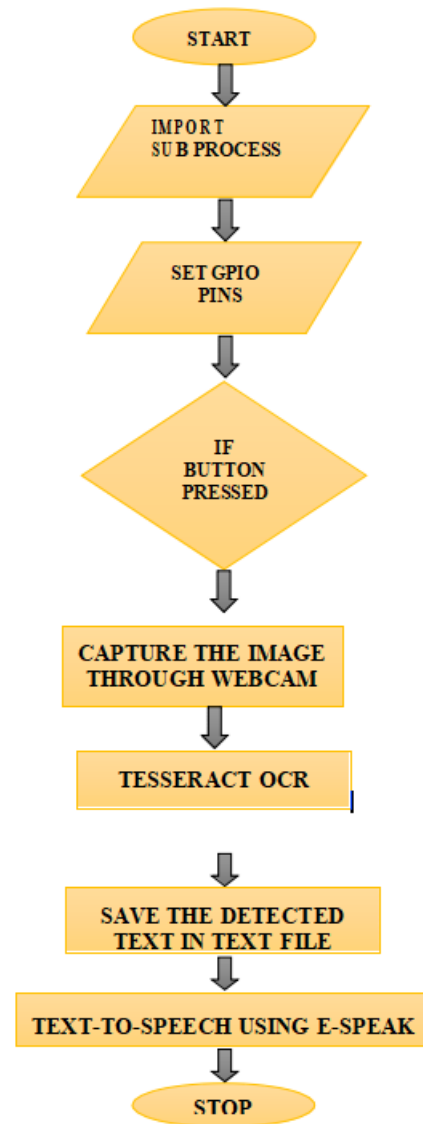
Figure 3. Flow Chart

IV. EXPERIMENTAL RESULT

Hardware Implementation

Facilitating the Visually Impaired persons with an appropriate reading is the objective behind the development of Smart Specs. During this development, the speakers are designed to deliver the audio as an output. Even though each variant of Linux compatible with Raspberry Pi (e.g. Mandrake, Debian, Red Hat and Gentoo) can be utilized, this study prefers Raspbian operating system since GPIO pins on Raspberry Pi have been predominantly utilized for camera module as well as USB WiFi-dongle in this work. Besides, it includes a menu based tool called raspi-config that eases the management of Raspberry Pi configurations through setting up SSH, facilitating Raspberry Pi camera module, etc., hence it is better than other OS.

Based on the specifications given below, the testing is performed on Raspberry Pi platform
- SBU Raspberry Pi 2 900 MHz Quad Code ARM Cortex-A7
- Logitech 3MP Camera Module
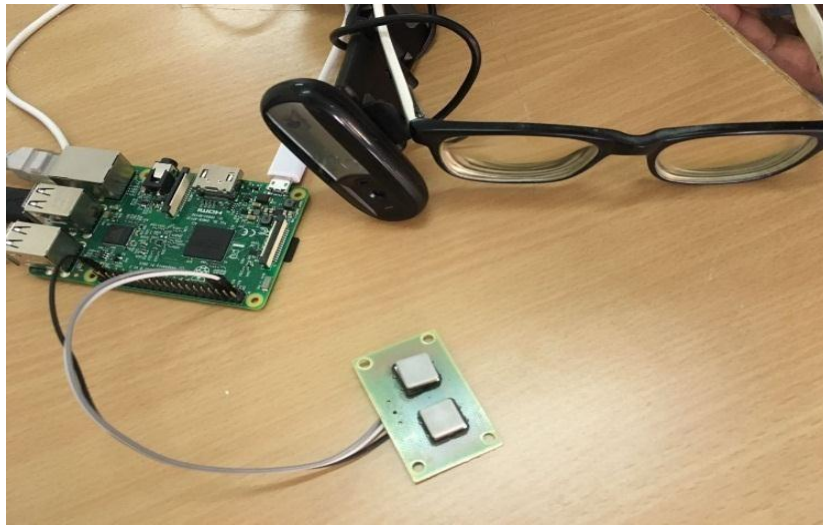- Bootable SanDisk Ultra 8GB microSD Card

Figure 4. Hardware implementation

In the proposed system's hardware, Wi Fi dongle is included that is linked to Pi via LAN cable. Besides, one of the USB ports is connected with a 5mp camera. From the system, 5V supply is given to Raspberry pi via power bank. At first, image is captured by the specs in-built web cam, once push button switch is pressed. Prior to feeding the captured image into OCR, it is thresholded for maximizing the accuracy. Further involved a compact speech synthesizer called eSpeak, which is an open-source software that converts the detected text into voice. It supports English and other languages, besides compatible with Windows and Linux operating systems. In this model, formant synthesis method belongs to SAP15 version is utilized, which is capable of making the speech clearer, even at higher speed. The text is converted to phoneme codes (MBROLA diphone voices) by Espeak, and the speech output is generated as WAV format file. Post-conversion, the output is transferred through headphone.

Software Implementation

For recognizing and "reading" text embedded in images, python includes an optical character recognition (OCR) tool, namely Python-tesseract that is a wrapper for Google's Tesseract-OCR. Because, Tesseract-OCR can only support bmp, and tiff formats of image, but Python-tesseract is capable of reading each of the image formats (e.g. tiff, png, bmp, gif, jpeg, etc.), whichever is supported by the Python Imaging Library. In context of using this as a script, the recognized text will be printed rather than writing it to a file.

In Tesseract 4.00, the greater accuracy rates are predominantly derived through a new neural network-based recognition engine that makes Tesseract 4.00 to outperform other existing methods, but with increasingly power consumption. Yet it is proved to be efficient and faster on complicated languages than base Tesseract. In this work, In the function of OCR engines, adaptive classifier is highly recommended for determining that0 font is either a character or non-character. The template (static) classifier usage is not preferred by Tesseract, similar to many OCR engines. Preferably, the images with clear black text and white background are fed into the OCR module to enhance the quality of outcomes. Each image is applied with Otsu's thresholding technique in Tesseract OCR. Nevertheless, this phase is avoided to accelerate the process due to its possession of custom pre-processing algorithm. Accordingly, it is necessitated to set tesseract delegate option as "self" (tesseract.delegate = self) for disabling Tesseract OCR's internal thresholding.

Text to Speech using eSpeak:

As an open source software speech synthesizer, eSpeak supports for English and further languages and compatible for Windows and Linux, in which "formant synthesis" approach is significantly utilized. Consequently, numerous languages are allowed to be given in a small size. A clear speech even at high speeds is considered as the major benefit of eSpeak.

Text speaking is available as a command line program (Windows/Linux). (This is a DLL on Windows). With regard to handle audio devices, Advanced Linux Sound Architecture (ALSA) is adopted by Raspbian OS.

Post-completion of image processing in model image set, it is enabled to compare the obtained outcomes with the original textual images. Throughout the process of data collection and scrutiny, various observations are obtained potentially. But, sometimes it is unavoidable that the misinterpretation done by Tesseract OCR at the corner of the captured text, though it is an efficient software for detecting text. The reason is that the frames cornermay be converted into letters by Tesseract OCR or the speed signs are rotated.

Working

At first, Logitech c270 webcam that is in-built in the spectacles captures the textual image. Then, it is stored in a text file and fed into OpenCV Python Simulator to be processed, since it is capable of supporting wide-ranging programming languages, such as python, C++. In this simulator, Tesseract OCR package plays a vital role in recognizing and reading the text embedded in images, during which the image is imported from the text file by employing Python Imaging Library (PIL). From each image in a directory, text is grabbed by Python script, whereas PyTesser module saves asubdirectory (/converted- text) as text files.

```
python main.py <directory_path>
```

For Python, it is considered being an Optical Character Recognition module, in which an image/image file is fed as an input and a string is delivered as an output. To convert the images into to an accepted format as well as to call Tesseract executable as an external script, the Tesseract OCR engine (an Open Source project at Google) is greatly utilized by PyTesser. Alongside the Python scripts, a Windows executable is provided, which also supported by Linux

Subsequently, eSpeak Package converts the detected text into speech. Besides, a normal language text is converted into speech by text-to-speech (TTS) method.

As a compact open source software speech synthesizer, eSpeak can be installed as given below,

```
sudo apt-get install espeak
```

Create code speech2.py:

```python
import os
os.system("espeak 'The quick brown fox'")
```

It is more robotic like pyttsx, yet simple to be used that involves formant synthesis technique called SAP15 version. Consequently, clear speech is enabled even at high speed. The speech output is generated as WAV format file, where the text is converted to phoneme codes (MBROLA diphone voices) by Espeak. Post-conversion, the output is transferred through headphone.

Simulation result:

This work tends to deliver the fast-processed outcomes for demonstrating. With regard to accuracy, Tesseract OCR is proved to be efficient among optimal character recognition engines, since it is capable of recognizing letters in several kinds of images, besides the open source C library Leptonica library is utilized by it. In this study, the images are successfully fed into Tesseract OCR and read. Prior to passing the images to Tesseract OCR engine, they would have to be gone through pre-processing. In this segment, the acquired results are simulated.

In all assistive reading system, the vital and significant function is Text Information Extraction, which is an integral part of OCR, since the intelligibility of the output speech is determined during this process. In espeak, speech engine belongs to libespeak shared library is utilized after installing libespeak library. Since the PortAudio sound library (version 18) is used in espeak, libportaudio0 library package needs to be installed first. since it might be used by other software like OpenOffice.org and the Audacity sound editor.

INPUT IMAGE SETS

The suggested approach analysis is carried out through numerous input sets, such as jpeg, png, jpg, bmp, etc. Fig 5 depicts that the text as black & white, in which a mixture of contextual through pure white & text through black color are involved. Besides, it shows text in colored imageform that includes text with black accompanied by colored background.
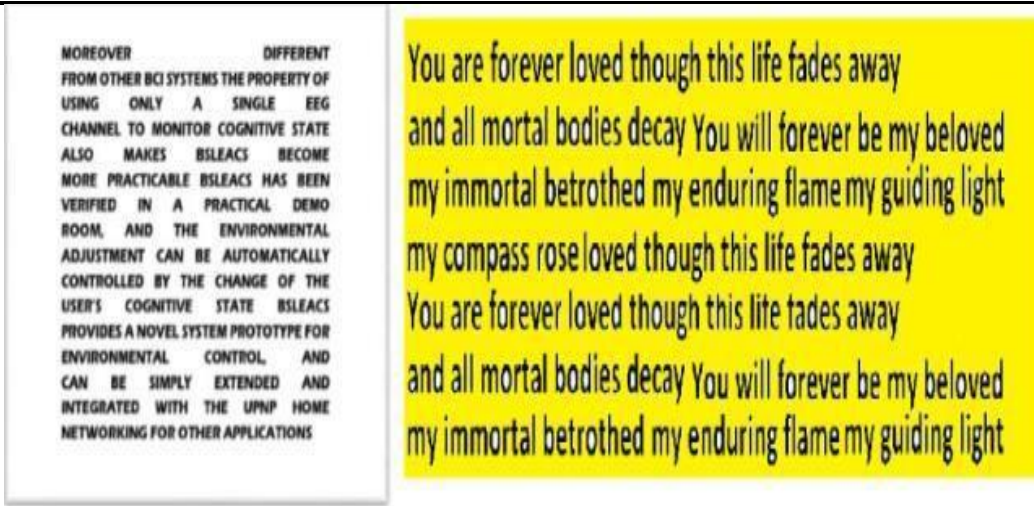
Figure 5. Black & White image and Colored image

In Fig 6, input consists of a text with image are depicted.

Figure 6. Merged Image and Text

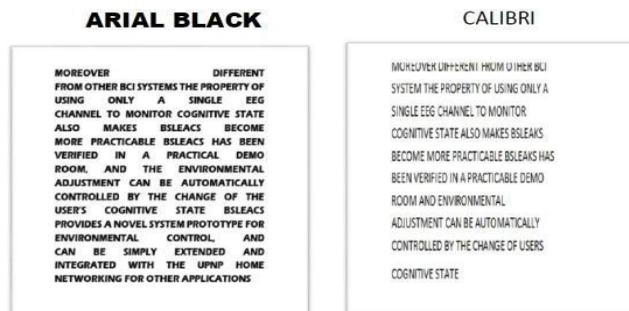In Fig 7, Text through various text styles are exhibited.

Figure 7. Text Style Scheme

Post-completion is enabled to compare obtained outcomes with the original textual images. Throughout the process of data collection and scrutiny, various observations are obtained potentially. But, sometimes it is unavoidable that the misinterpretation done by Tesseract OCR at the corner of the captured text, though it is an efficient software for detecting text. The reason is that the frames corner may be converted into letters by Tesseract OCR or the speed signs are rotated.

It represents that applied algorithm is proved to be accurate, where the total average time is ~1.5 frames per second (fps) on a 700 MHz Broadcom chip, for both the process of detection and recognition.
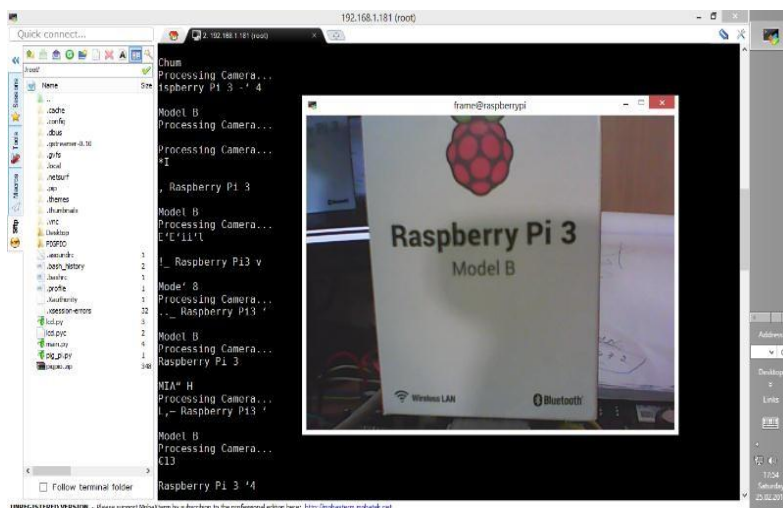
Figure 8. Capturing of image

Figure 9. Text Detection

INPUT IMAGE: BLACK AND WHITE IMAGE
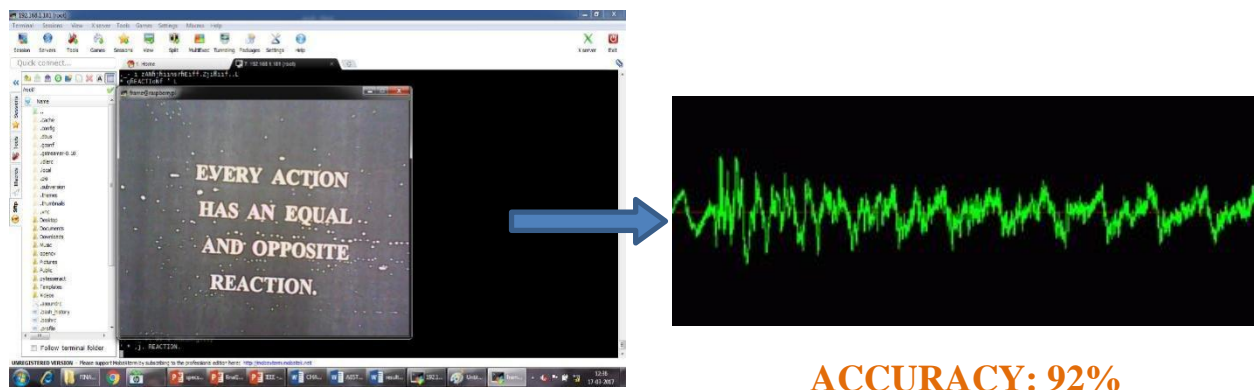

ACCURACY: 92%

Figure 10. Speech Signal for Input image: Black and White Image

In the comparison table, the various input images are compared, in terms of their corresponding accuracy, word error rate and time, due to the variation set by input sample image.
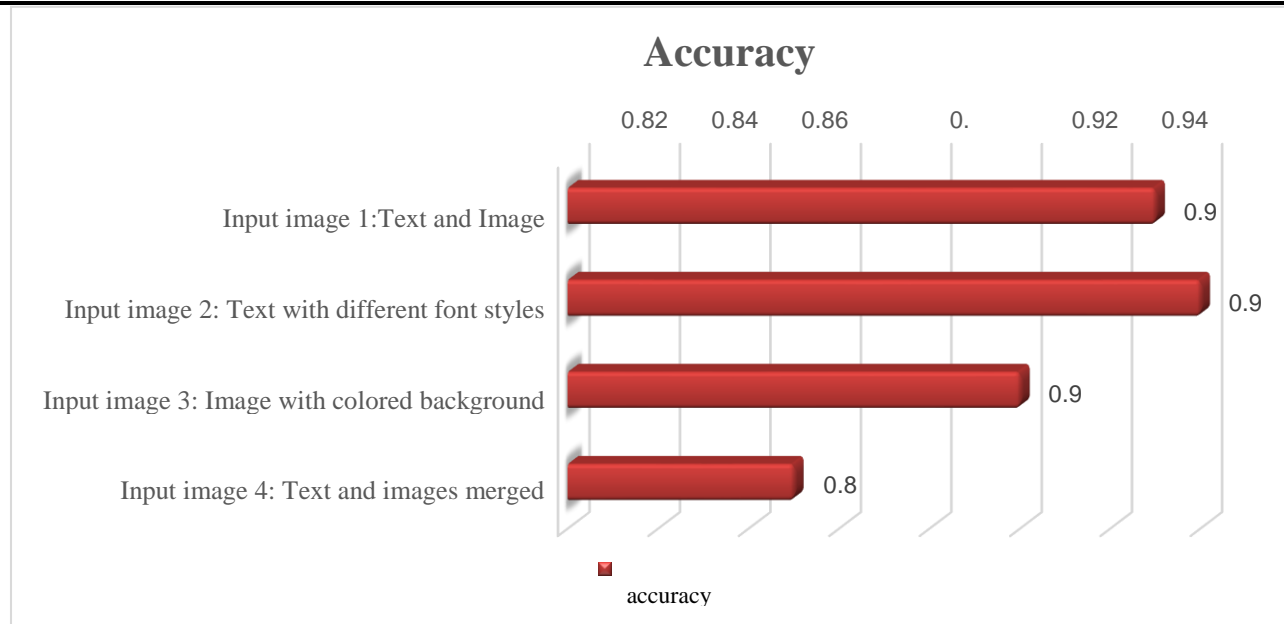
Accuracy

Figure 11. Comparison of accuracy for various input image sets

V.CONCLUSION

In this paper, the proposed model is summarized, besides the overall operations of the each block involved in the architecture of proposed framework are explained. Moreover, it confers the basic steps and specifications of proposed method. A few other classifiers, such as Euclidian distance and k nearest neighbor are further involved for maintaining the cohesive run of Raspberry Pi that is used to perform text-to-speech conversion method. By using various samples, verification of simulation results is carried out, and testing of hardware output is accomplished in an efficient manner. The algorithm is proved to be effective for processing the image and capable of reading it out perfectly. Thus, it is considered being an efficient, yet economical tool and appropriate for the people who are visually impaired. Over numerous images, this algorithm is applied, where it proves its efficiency to accomplish the conversion task. Hence, this device can be considered as handy, and valuable one to the society.

REFERENCES

[Deepa Jose V. and Sharan R. (2014), "A Novel Model for Speech to Text conversion",InternationalRefereedJournalofEngineeringandScience(IRJES) Vol.3, Issue.1, pp.39-41

[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

[??] Rissanen, M. J.Fernando S.Pang .N. (2013), "Natural and Socially Acceptable Interaction Techniques for Ringterfaces: Finger-ring Shaped User Interfaces", Springer - In Distributed Ambient and Pervasive Interactions, Vol. 19, No. 6, pp

[??] Goldreich D. and Kanics I. M. (2003), "Tactile Acuity is Enhanced in Blindness", International Journal of Research And Science, Vol. 23, No. 8,pp. 3439–3445

[??] Marion A. Hersh Michael A. Johnson (2013), "Assistive Technology for Visually Impaired and Blind",Springer-International Journal of Engineering and Technology, Vol. 4, No. 6, pp.50-69

[??] Rupali and Dharmale (2015), "Text Detection and Recognition with Speech Output for Visually Challenged Person", Research Gate- International Journal of Engineering Research and Applications, Vol. 5, Issue. 3, pp.84-87

[??] S. Mascaro. And H. H. Asada. (2001) "Finger posture and shear force measurement:Initial experimentation,"inProc.IEEE Int.conf.robot.Autom,Vol. 2,pp.1857-1862

[??] PitrelliJ. and BakisR.(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

[??] 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

[☐   Shen, H. and Coughlan, J. M. (2012), "Towards a real-time system for Finding and Reading Signs for Visually Impaired Users", Springer-International Journal in Computers Helping People with Special Need, Vol.2 , No. 1, pp. 41- 47.

[☐   Alías, Francesc, et al. "Towards high-quality next-generation text-to-speech synthesis: A multidomain approach by automatic domain classification." IEEE transactions on audio, speech, and language processing 16.7 (2008): 1340-1354.

[☐   Yi, Chucai, Yingli Tian, and Aries Arditi. "Portable camera-based assistive text and product label reading from hand-held objects for blind persons." IEEE/ASME Transactions On Mechatronics 19.3 (2013): 808-817.

[☐   Joao Guerreiro and Daniel Gonçalves (2014), „Text-to-Speech: Evaluating the Perception of Concurrent Speech by Blind People", International journal of computer technology, Vol. 6, No. 8, pp. 1-8.

[☐   Balakrishnan, G., et al. "Wearable real-time stereo vision for the visually impaired." Engineering Letters 14.2 (2007).

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

[16] 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

[17] Xiang Peng. Fang Liu.Tianjiang Wang. Songfeng Lu (2008), "A Density- base Approach for Text Extraction in Images", Research Gate- International Journal of Communication and Engineering Vol. 8, No. 4, pp. 239-248