Handy assistance tool for the visually impaired people using TTS technique with voice assistance

Bindhu B P Kavya M Spoorthy L Vinushree M

Guided by: Mrs. Shenbagavalli, *Associate professor*

Depatment of Electronics and Communication, AMC Engineerong College

Abstract: *Out of the total population of around 7.4 billion people a whooping 285 million people are visually impaired across the world according to the reports of the WHO.With such growing number of visually impaired population t is important that necessary steps should be taken to assist this population in their day to day activities with the recent emerging technologies to cope up with the fast moving world inspite of their inabilities. With this vision in mind we have proposed a system where the blind or the nearl yblind people can make use of this technology of text detection producing voice output. This helps the blind and the nearly blind people to read any text in the printed format thus producing a voice output which is read out to them through a headphone or speaker as per convinieth the vision of supporting the blind A camera which is inbuilt in the specs captures the image of the desired text to be read which is preferably in printed format and subjected to Tesseract-Optical character Recognition(OCR) analysis and converted into text. The text detected id further converted into voice output and read out to the user using eSpeak software.Here Raspberry Pi is targeted as the main part of the implementation because it can provide interface between camera, sensors, and results of image processing along with performing functions which manipulate peripheral units such as Keyboard, USB etc*

Keywords: eSpeak, Raspberry Pi, Tesseract OCR

I. INTRODUCTION

Taking a look at the share of blind population worldwide, its alarming to note that 215 million people are visually impaired of which 40 million people are blind. According to a survey conducted in the united states, National Health Interview Survey (NHIS), 25.2 million adult Americans were blind or visually impaired which accounts to 8 percent of the population.

It is feasible to assist these visually impaired and blind population with the recent developments in the fields of digital cameras, computer vision and portable computer systems.These products work with the combination of computer vision technology with the already existing products such as OCR. Getting access to printed text

ISBN : 978-93-87793-89-7

documents is difficult for the visually challenged population in various circumstances such as reading text immediately or text acess in less than or no ideal conditions.

The main purpose is to allow the visually challenged users to get access to printed text and receive real time voice output. This system requires the assistance of two central technologies namely OCR or Optical Character Recognition to carry out the process of Text Information Extraction (TIE) and Text to Speech conversion (TTS) which is used to convert the printed text into speech. TTS synthesizer is a software based operator that reads out any detected text

aloud when an oprator directly introduces into the computer system. The 'Grapheme to Phenome' transcription is the best way t describe this process.

The primary and important function of any assistive reading software or system is undoubtedly Text Information Extraction(TIE). This determines the itelligibilty of the voice output hence this forms the intergarl part of the Optical Character Reader. By the recent develpoments of camera based services in the field of computer vision and dogital cameras, The existing technologies like OCR easily merge with the computer vision tevhnologies. The OCR here is used to recognize block of words and sentences without any errors. OCR recognises electronically converted computer readable text at high rates.

II.LITERATURE SURVEY

1) The authors Billi et al. observed and stated that In the field if information and technology in the present society, mobile devices pose new oppurtunities such as ubiquitous access and portability. The unobtrusive nature of the mobile assisted technologies takes the fundamental advantage in delivering assistive technolo-gies in various platforms. Such technologies enable the visually challenged users not to be albelled and less stigmatized because these subtle applications can be incorporated into devices like mobile phones and can be individually operated by the users themselves. Also, such kind of assistive technologies are typically adapt to multiple platforms assisting individuals with multiple challenged disabilities.
2) The authors Brewster et al.19 proposed two novel technologies solving the problem of reading for the blind proposing an for eyes-free, mobile device use. The users are first presented the information of the items via the 3D radial menu in pie fashion. The users are required to nod their head in the direction of the appropriate sound to select an item. Each sound represents an item. For an example, in an application about the current affairs, different options such as weather, traffic news, sports,etc were represented using snippets of identifiable audio clips- traffic sounds, weather sounds, the theme song tune of a television show to a news channel respectively. The user is required to nod in the direction from which the sound originated from in order to listen to that particular type

ISBN : 978-93-87793-89-7

of information. There was another application where a music application with dfferent sounds assigned for different genres of music oprated in much the same way and allowed the user to pick the desired music genre, album, artists or tracks which were represented by clips of music, by nodding according to the direc-tion of appearance of the sound

3) Authors Neff et al worked on the issue of touchscrenen acessibility by splitting the screen into two where one half of the screen into icon presentation and other half into effective interaction with the icons. They have developed a design framework which is based on the use of non speech, spatialized sounds to represent the icons and make use of gestures to interact with the icons. But they have provided details of their framework yet no results of user studies of their work have yet been published.

4) The authors Jayant et al. introduced their work V-Braille 48 which allow the Braille literate users to interact with the touch screen using the vibration pro-duced on the touch screen. Its exclusively for the users who know the Braille and cab use the touchscreen or mobile phone interfaces via the vibrations generated. The mobile screen is divided into six different parts to convey the traditional Braille structure on the mobile scree. So when the user touches the screen within these six parts, vibrations of different strngths are produced which allows the user to relate each vibration to a cor-responding Brialle character.Preliminary observation and evaluation of this kind of mobile phone Braialle iterface among nine potential users shows the scope of expansion of V braille in intorducing Braille as an efficient alternative presentation paradigm.

5) MoBraille 9 is a appreciable framework for allowing access to many of the Android smart phone features by connecting the phone to a Braille structure display which acts as an input/output interface platform. Such produced Braille structured mobile displays electron-ically operate by increasing and decreasing different pin combinations to reproduce what appears on the portion of the smartphine screen in Brialle. MoBraille brings in the possibility of any Android application to interface with a Braille structured display over a Wi-Fi connection, thus allowing Braille display users to gain access to applications, including the GPS-based facilities and compass right on their smartphone.

III. PROPOSED SYSTEM

The proposed system has been developed to assist the visually challenged masses. It has been built on a Raspberry Pi 3 board which runs the Raspbian Operating System based on python and Open cv libraries. Raspbian OS is a de-facto standard operating system which is a linux based debian system. It comes presinstalled with libraries. The text caputed by the Pi camera in the for of jpeg,png,bmp,jpg formats are considered for analysis. The image captured is further processed using the Tesseract OCR. The Tesseract OCR

ISBN : 978-93-87793-89-7

converts the captured text into readable letters. This readable text is finally converted into voice ouput using the eSpeak software. All this happens in real time.

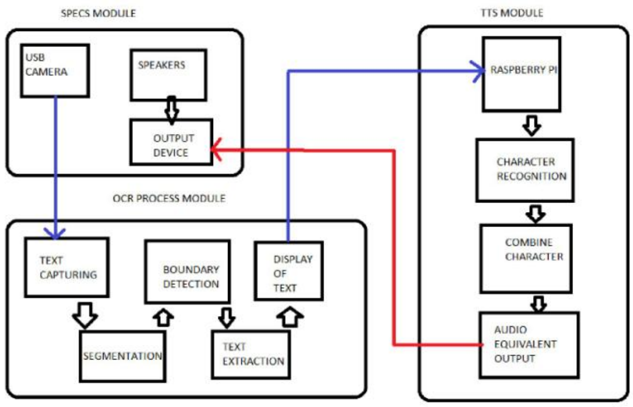

Fig. 1. Block diagram

IV. IMPLEMENTATION AND RESULTS

As represented in the block diagram above,the proposed system implemented on the Raspberry Pi 3 takes input via the Raspberry Pi camera of resolution 5MP and then the captured image is converted to text using the Tesseract optical character Reader. The readable text is then converted to voice output using the speech synthesizer eSpeak. The concept of proposed system is the idea of developing an assistive eye wear which reads text for the visually impaired people. The different modules used in the system are camera module, optical character recognition module and text-to-speech module.

With the help of Pi camera athe Raspberry Pi is enabled to capture a sequence of images. The letters in the image are extracted and converted into text format using the Tesseract Optical Character Recognition (OCR). The real time voice assisted output is provided by converting this text file into audio file using the eSpeak speech synthesizer. The above block diagram indicates the flow of the implementation of the prototype model proposed. The desired output at the end of this stage of implementation is thresholded and edge processed image. The input is taken by the Pi camera. The input is a raw image of the jpg format.The camera port of the Paspberry Pi is connected to the Pi camera of resolution 5 MP

The delay time for capturing the image is set to 3 seconds and is changeable according to the convinience of the user. The user is expected to capture the image using the camera in the specified time else the camera goes to sleep. Once the image is captured it is thresholded and then sent for edge detection. The image captured is in the RGB format. To enable easy thresholding this captured image in the RGB fromat is converted ito gray format. If the image is thresholded directly in the RGB format , it consists of many layers thus involves many levels of conversion and it becomes complicated. Thus

ISBN : 978-93-87793-89-7

the image isconverted into gray format which has only 2 primary layers thus easy to carry out thresholding.

Fig. 2. Working model

V.ISSUES AND IMPACTS:

According to the process of thresholding only the inten-sity of the image is considred and not the relationship between the pixels thus there is no specifies assurance that the pixels identified by the thresholding process are contiguous.

Extraneous pixels can easily be icluded in the required desired region and also isolated pixels especially near the boundaries of the image which may form a apart of the image can easily be missed out. As the noise in-creses these effects get worse just because the intensity of the pixel dont represent the normal intensity of the region.

Shodows also pose a problem in the processing of the image. Not just where the shadows fall but they are often included mistakenly as a part of dark object on a fairly lighter background. This may alter the desired results of image processing and noise filtering.

Sometimes in edge detection the differential masks act as high pass filters and results in noise in the output image.

VI. CONCLUSIONS

The main objective of our project is for building a spec-tacle prototype for the visually challenged which helps them to read through a miniature camera fitted to the frame which takes photos of text and using OCR converts it into text and reads back to the user through an ear piece using an open source speech synthesizer Espeak,also to recognise people faces and identify those that have been previously stored at an affordable cost(below 5000 rupees)for the general masses.

REFERENCES

[1] AlexandreTrilla and Francesc Alas. (2013), SentenceBased Sentiment Analysis for Expressive Text-to-Speech, IEEE Transactions onAudio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.

[2] Alas F. Sevillano X. Socor J. C Gonzalvo X. (2008), Towards high-quality next-generation text-to-speech synthesi, IEEE Trans. Audio, Speech, Language Process, Vol. 16, No. 7. pp. 1340-1354.

[3] Balakrishnan G. Sainarayanan G. Nagarajan R. and Yaacob S. (2007) Wearable real-time stereo vision for the visually impaired, Vol. 14, No. 2, pp. 614.

[4] 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.

[5] Deepa Jose V. and Sharan R. (2014), A Novel Model for Speech to Text Conversion, International Refereed Journal of Engineering and Science (IRJES) Vol.3, Issue.1, pp. 39-41.

[6] Goldreich D. and Kanics I. M. (2003), Tactile Acuity is Enhanced in Blindness, International Journal of Research And Science, Vol. 23, No. 8,pp. 34393445.

[7] Joao Guerreiro and Daniel Gonalves (2014), Text-toSpeech: Evaluat-ing the Perception of Concurrent Speech by Blind People, International journal of computer technology, Vol. 6, No. 8, pp. 1-8.

[8] 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. Enhances Tactile Acuity and Haptic 3-D Shape Discrimination, Proceedings of the 4thAugmented Human International Conference, Vol. 73, No. 7, pp. 2330.

[9] 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. 10991108.

[10] PriyankaBacche. ApurvaBakshi, KarishmaGhiya. PriyankaGujar. (2014), Tech NETRA (New Eyes to Read Artifact), International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 3, Issue. 3, pp. 482-485.

[11] Rissanen, M. J.Fernando S.Pang .N. (2013), Natural and Socially Ac-ceptable Interaction Techniques for Ringterfaces:Finger-ring Shaped User Interfaces, Springer - In Distributed Ambient and Pervasive Interactions, Vol. 19, No. 6, pp. 52-61.

[12] RohitRanchal .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.

ISBN : 978-93-87793-89-7

[13] Rubesh Kumar and Purnima (2014), Assistive System for Product Label Detection with Voice Output for Blind Users, International Journal of Research in Engineering and Advanced Technology, Vol. 1, Issue. 6, pp. 30-45.

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

[15] Shen, H. and Coughlan, J. M. (2012), Towards a realtime 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.

[16] 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 Pro-cessing, Vol. 8, Issue. 2, pp. 239-250 .

[17] 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.

[18] 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.