INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 2)
Available online at: www.ijariit.com

Reading Aid for Visually Impaired People

S. Muralidharan
rajeeevmukesh@gnail.com
T. J. S Engineering College

D. Venkatesh
Venkat8697@gmail.Com
T. J. S Engineering College

J. P. Pritmen
Pritmen777@gmail.com
T. J. S Engineering College

Purushothaman R
purushothamanvishal14@gmail.com
T.J.S Engineering college

S. Jeya Anusuya
hodece@tjsec.in
T. J. S Engineering College

V. Saravanaperumal
sperumal181@gmail.com
T. J. S Engineering College

ABSTRACT

It is known that the technological advancements are increasing at a faster pace. But the utilization of technologies in various sectors is very low. Human communication today is mainly via speech and text. To access information in a text, a person needs to have vision. However, those who are deprived of vision can gather information using their hearing capability. It is known that most of the people find it difficult to detect the text from the paper and books. The proposed method is a camera-based assistive text reading to help a blind person in reading the text present on the text labels, printed notes from paper and books. The proposed project involves Text Extraction from the image and converting the Text to Speech converter, a process which makes blind persons to read the text. This is carried out by using Raspberry Pi, where The Pi processes the text images and reads out the content using the speaker. Optical character recognition (OCR) is the identification of printed characters using photoelectric devices and Computer software. To use OCR for pattern recognition to perform Document image analysis (DIA). In this portability is the main aim which is achieved by providing a battery backup and can be implemented as a future technology. The portability allows the user to carry the device anywhere and can use anytime. This enables the use of text to speech conversion.

Keywords: *Raspberry Pi, Camera-based Assistive Text Extraction, OCR (Optical Character Recognition), DIA(Document Image Analysis).*

I. INTRODUCTION

The beautiful gift that has given by the GOD to human beings is VISION. Vision allows people to perceive and understand the surrounding world. To access information in a text, a person needs to have vision. However, those who are deprived of vision can gather information using their hearing capability. A Majority of the visually impaired use Braille for reading documents and books which are difficult to make and less readily available and this type of reading methodology requires extra time for practice. This project aims to study the image recognition technology with speech synthesis and to develop a cost-effective, user-friendly image to speech conversion system with help of Raspberry Pi.

Many people suffer from serious visual impairments preventing them from exposing to the environment. In this research, we develop a protocol that assists visually impaired people which effectively and efficiently reads paper-printed Text. The proposed project uses the methodology of a Camera Based Assistive device that can be used by people to read Text document.

The main work of this prototype is to convert the text signal into speech signal. The idea in which the output is to be in the form of speech because Speech is probably the most efficient medium for communication between humans. This text to speech conversion is processed by both hardware and software components. The hardware components are raspberry orange pi 3, camera module and a headphone. The software components are OCR reader and there is other backend software which is responsible for this conversion.

The basic functioning of this prototype is to convert text to speech, the image is captured via camera, and the resultant will be in the form of a JPG or Jpeg format. The captured image is to be filtered and the edges of the image should be detected in which the pixels of the image is properly arranged. This image is given as input to OCR (Optical Character Recognition), the main function of OCR is to convert the image into machine code. Thus that machine code is converted into object code by processing the machine code using a raspberry pi and the speech output is taken from raspberry pi audio jack in their suitable language by using translators.

The text is converted into speech and it was taken from raspberry pi via earphone. And this makes the blind people know about the text.

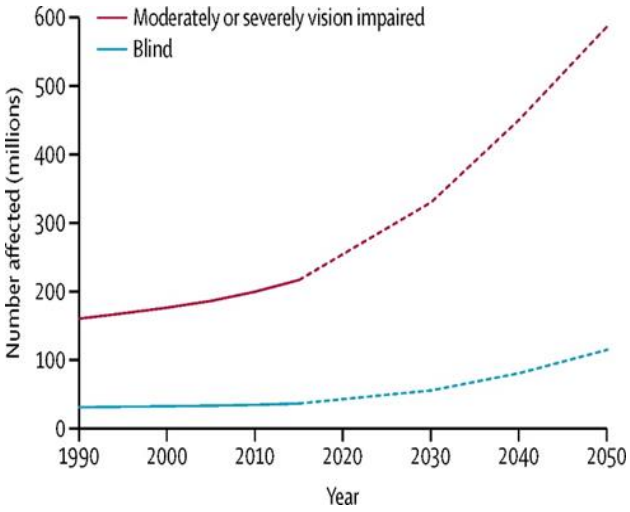

Chart-1 Survey of blind people

Age group (years) Number	Presenting visual acuity		Best-corrected visual acuity	
	Low vision from <6/18 to ≥ 3/60 Number (%)	Blindness <3/60 Number (%)	Low vision from <6/18 to ≥ 3/60 Number (%)	Blindness <3/60 Number (%)
40–49 - 1419	43 (3.0)	7 (0.5)	6 (0.4)	3 (0.2)
50–59 - 1120	87 (7.8)	8 (0.7)	20 (1.8)	3 (0.3)
60–69 - 906	133 (14.7)	38 (4.2)	46 (5.1)	9 (1.0)
70–79 - 366	78 (21.9)	24 (6.7)	30 (8.4)	10 (2.8)
≥80 - 49	12 (24.5)	9 (18.4)	10 (20.4)	6 (12.2)
Total 3850	353 (9.2)	86 (2.2)	112 (2.9)	31 (0.8)

Chart-2 Prevalence of blindness

2. LITERATURE SURVEY

The state of lacking the visual perception due to physiological or neurological factors is called blindness. The lack of integration in the growth of the optic nerve or visual center of the eye represents partial blindness. The total blindness is the full absence of the visual light perception. In this proposed work, a virtual eye which is simple, cheap, friendly user is designed and implemented, in which the mobility of both blind and visually impaired people is improved in specific areas. This proposed work is a wearable one and it consists of a head hat, mini hand stick and foot shoes are used, by which the blind person can navigate alone safely and easily and avoid any obstacles that may be encountered, whether fixed or mobile, prevent any possible accident. The ultrasonic sensor is the main component which is used for the blind people to scan predetermined area by emitting-reflecting waves. The input to Arduino microcontroller is the Signals received from the barrier objects. The commands which are issued are carried out by the microcontroller and the status of the given appliance is communicated or it is derived to the earphone using raspberry pi speech synthesizer The proposed system is fast, cheap and it is easy to use, it gives an innovative and affordable solution to the visually impaired and blind people in third world countries.

To help the blind persons to read text labels and product packaging from hand-held objects in their daily lives a camera-based assistive text reading framework is proposed and developed. An effective and efficient motion-based method is proposed first in which Region of Interest (ROI) is defined, a novel text localization algorithm is proposed by learning gradient features of stroke orientations and an Ada boost model is used for the distributions of edge pixels. Text characters in the localized text regions are then binarized and it is recognized by off-the-shelf optical character recognition (OCR) software. The output to the blind user in speech is the recognized text codes.

The braille system of reading is the method which is used by the majority of the blind people for reading books and documents, but this method found to be difficult to make and less readily available. This difficulty has to be minimized by developing a device that could bring relief to the agonizing tasks that the visually impaired has to go through. Due to the digitization of books, there are many excellent attempts at building a robust document analysis system in industries, academia and research labs, but this is only

for those who are able to see. This project aims to study the image recognition technology with speech synthesis and to develop a cost-effective, user-friendly image to speech conversion system with help of Raspberry.

To build texture representations, deep convolutional neural networks are used for recent approaches. Nevertheless, the representation of texture are unclear and it is invariances to categorical variations. The recent CNN-based texture descriptors for recognizing is the work of systematic evaluation and attempts to understand the nature of invariances captured by these representations. A general purpose texture descriptor which is excellent is bilinear CNN model[25], which is proposed recently and to other CNN-based descriptors, it is favorably comparable on various texture and scene recognition benchmarks. This model acquires translationally invariant and without requiring spatial jittering of data it obtains better accuracy on the image Net dataset of data compared to corresponding models trained with spatial jittering. The technique which we proposed is based on recent work in which the pre-images are visualized and by providing a means to understand categorical properties which are captured by these representations. Finally, we show preliminary results on how a unified parametric model of texture analysis and synthesis can be used for attribute-based image manipulation.

3. PROBLEM DETECTION

The braille system of reading document and books requires more practice and time. Nowadays Computers are designed to interact by reading the books or documents. Synthesized voice is used to read the content on the computers. We also have devices that scan the documents and use the interfaced screen to allow the blind to sense the scanned documents on the screen either in Braille or Anusha Bhargava et al, which is also a Reading Assistant for the blind people to read by using the shape of the letters itself with help of vibrating pegs. Various phone applications are also developed to help in reading or helping the blind in other ways.
But the system fails to perform the text to speech. And another problem is that developed application created by the computer source can read the document which has the higher font size, and it doesn't have the capability to read the smaller sized text.

4. PROJECT OVERVIEW

The main overview of our project is to convert the text signal to speech signal. The basic block diagram of our prototype is given below, in which the text is extracted from the camera and that text is to be processed and pre-processed in order to maintain the correct pixel density and to ensure that it has proper pixel quality. After the image processing technique, the image is converted into respective code, and the speech is finally obtained from the code generated and that speech signal is the output taken from the earphone.

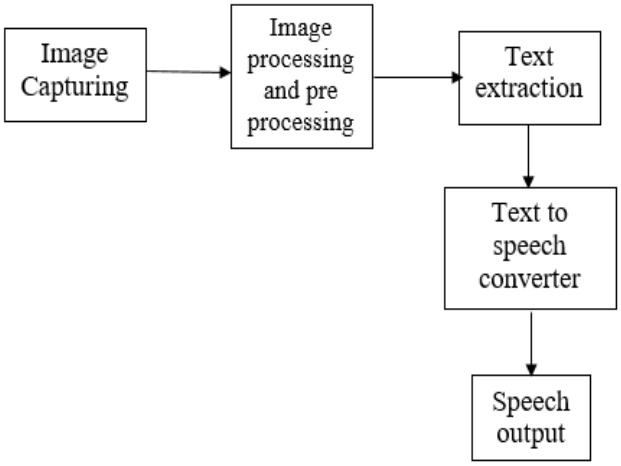

Chart.3 Basic Block Diagram

5. PROPOSED SYSTEM

The main block diagram of our prototype is given below. In this block diagram, each and every block has a unique and important functionality. The input for this prototype is an image, which is to be converted into speech signal i.e., the output of the prototype.

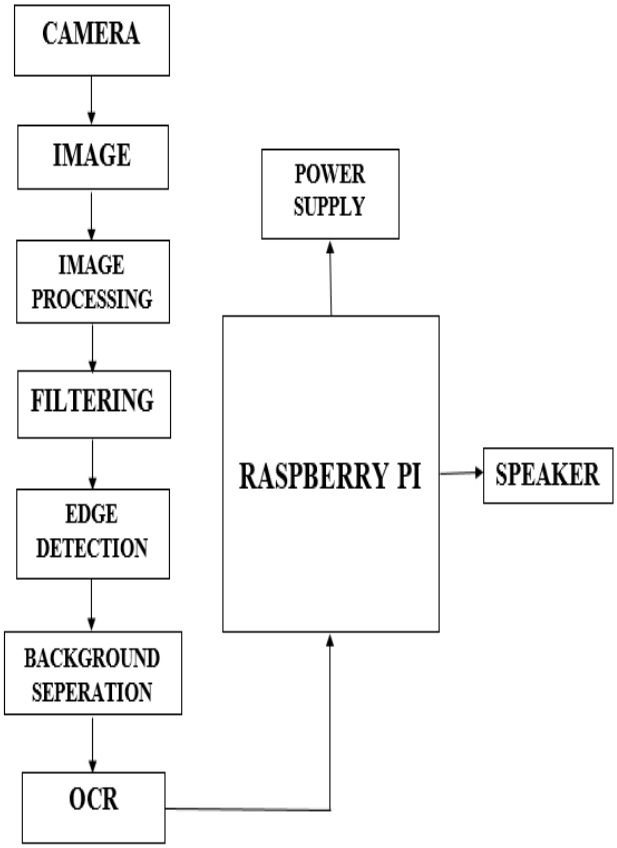

Chart 4: Architecture of the proposed system

6. METHODOLOGY

The proposed system overcomes the drawback experienced by the existing system. The system consists of a raspberry pi. Here we are using a camera for capturing the image of the text. The camera's image is provided to the Image processing section. The processed image output is given to the filter for the removal of noise signals. The filtered output is provided to the edge reduction unit. The output is provided to the background separation unit. At last the output is provided to the OCR. The output of the OCR is given to the raspberry Pi. The pi detects the image content and provides an output signal in the form of audio signal.

The entire process can be concluded to the following steps:

1) Capturing the Image
2) Processing the Image
3) Filtering
4) Edge Detection
5) OCR conversion
6) Raspberry pi Speech Output

A. Camera

An active Webcam can capture an image up to 30 frames per second. This capturing of the image can be accomplished from any video device including USB cameras, capture card which is connected with analog cameras, TV-boards, camcorders interfaced with FireWire(IEEE 1394) and from network cameras.

When a motion is detected in the motioned area by the program, then the alarm can sound, e-mail you the captured image, and it starts broadcasting or record a video. there is some additional feature in the program i.e., it can add text caption and image logos to the images, the date/time stamp should be placed each video frame, and the frame rate, picture size, and quality can be adjusted.

The video camera in which the feeding or streaming of image in real time is accomplished through a computer to a computer network is called Webcam. the "captured" of an image by the computer from a video stream, then the "captured" can be saved, viewed or sent on to other networks via systems such as the internet, and even it can send via email as an attachment.

When the "captured" is sent to a remote location, then the video stream may be saved, and it can be viewed or on sent there. The IP cameras are the one which requires a Wi-Fi or Ethernet connection, but this Webcam requires a USB cable for connection, or it can build into the computer hardware such as laptops, computers etc.

The features of a Webcam is that it has a Frame rate of 30 frames per second or above, it has good resolution another important feature is that has Continuous Autofocus.

Chart 5: Web Camera

B. Image Processing

The noise introduced during capturing or due to poor Quality page has to be cleared before further processing. This can be achieved by processing the image. The pixel density and quality is also adjusted and corrected by image processing. In image pre-processing, the unwanted noise in the image is removed by applying an appropriate threshold. The number of pixels added to the objects depends on the size and shape of the structuring element defined to process the image.
Image processing contains the following steps.

- Filtering
- Edge Detection
- Background Separation

a. Filtering

The technique which is used for modifying or enhancing an image filters. For example, to emphasize certain features or to remove other features of an image, the filter is the only solution to achieve these modifications. Filtering includes smoothing, sharpening, and edge enhancement of the image and it is implemented under image processing.

Filtering is a neighborhood operation, by applying some algorithm to the values of the pixels in the neighborhood then the pixels of the output image corresponding to the input is determined. The location relative to the pixel defines the pixel's neighborhood of some set of pixels. The filtering process in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood is called linear filtering.

b. Edge Detection

While performing image processing, one of the fundamental operations is edge detection. Edge detection is very useful to reduce the data (pixel) amount to process and the structural aspect of the image is maintained by this edge detection. The two edge detection schemes - edge detector which is based on the gradient (Sobel - first order derivatives) and another edge detector which is based on the Laplacian (2nd order derivative, so it is extremely sensitive to noise). Edge-detection is the same end goal for both the schemes and this is achieved by convolution of both the schemes.

In simple terms, we can say that the pixel density varies suddenly only on the edges which are very difficult for the computer to read the image without error. The pixel that lies on the boundary that is connected between the two regions with different pixel intensities is the pixel intensity in the edges.

Identifying an object with our naked eyes is easy, but how the edges of an image are detected by the computer?
Let’s find out.

In principle, the derivative of the image is calculated first. Then we look for the peak points, which has the values larger than the neighboring points. The set of connected ‘extrema’ points is an edge.

Every image is induced with at least one type of noise. These noises are to be filtered by using an appropriate filter in order to reduce the errors and to increase the efficiency. OpenCV offers a variety of filters for denoising.

Depending upon the nature of the noise the filter is chosen which would be appropriate to reduce all the noises. An example in our Edge detection with OpenCV example is Gaussian blur function.

c. Background Separation

The major pre-processing step in many vision base

the application is background separation or subtraction. Consider the cases like the information about the vehicle is extracted from the traffic camera or a static camera takes the number of visitors entering the visitor counter etc. In all the above cases it is necessary to extract the person or vehicle alone. Technically, the moving foreground from static background is to be extracted.

If there is a background image alone, like the image of the road
Without vehicles or image of a room without visitors etc. then it is very easy to get foreground image, just subtract the new image from the background, the resultant will be a foreground object alone. But in most of the cases, we are unable to see such an image, for this reason, it is necessary to extract the background from whatever images we have. The shadow of the vehicle becomes complicate while extracting the image. Because the shadow is also moving along with the object, simple subtraction will mark that also as foreground. It complicates things.

C. OCR(Object Character Recognition)

Our main aim is to build an application in which it can read the handwritten digits. Train_data and Test_data are required for this. Image digits.png(in the folder Opencv/samples/data/), from where the OpenCv comes from, and it has 5000 handwritten digits (500 for each digit). 20x20 image is the size of each digit. So it is required to split the image into 5000 different digits. For each digit, we flatten it into a single row with 400 pixels. It is mandatory to create the simplest feature set, i.e., intensity values of all pixels. From the sample, we use first 250 samples as train_data and next 250 samples as teat_data. Now we need to prepare them first.

There is a slight change in the data and feature set for English alphabets which is our next target. The data file, letter-recognition is the one in which the OpenCv comes from instead of images. Data in opencv/samples/cpp/ folder. You can see 20000 lines if you open it, which may on first sight look like garbage. The first column in each row is an alphabet which is our label. Next 16 numbers following it are its different features. UCI Machine Learning Repository is the one which is responsible to obtain these features. You can find the details of these features in this page.

The total available samples are 20000, in which first 10000 data as training samples and the remaining 10000 samples as test samples. We can't work with alphabets directly, so we have to change the alphabets to ascii character.

D. Raspberry Orange Pi

The third generation Raspberry Pi model is the raspberry pi 3 model. Raspberry pi 3 supersedes the original raspberry pi model B+ and Raspberry Pi 2 Model B and it is packed as powerful credit card sized single computer board. This processor brings you the most powerful processor i.e., 10x

times faster than the first generation raspberry pi 3 Model B.
Besides it provides additional features like wireless LAN and Bluetooth connectivity making a better solution for powerfully connected designs.

The third generation Raspberry Pi is the Raspberry 3, which replaced the Raspberry Pi 2 model B in February 2016. The raspberry pi 3 is also similar to previous Raspberry pi 2 (and Pi 1 Model B+) and has compatibility with previous versions of raspberry pi (pi 1 and 2).

The best part about all this is that the Pi 3 keeps the same shape, connectors, and mounting holes as the Pi 2.Dual Core Video Core IV® Multimedia Co-Processor. Provides an1080p30 H.264 high-profile decode and Open GL ES 2.0, hardware-accelerated OpenVG.

The features of the raspberry pi are, that it has a 1.2GHz 64-bit quad-core ARMv8 CPU, 802.11n Wireless LAN, Bluetooth 4.0, Bluetooth Low Energy (BLE), 1GB RAM, 40 GPIO pins and an Ethernet port.

Chart 6: Raspberry Pi

7. RESULT AND DISCUSSION

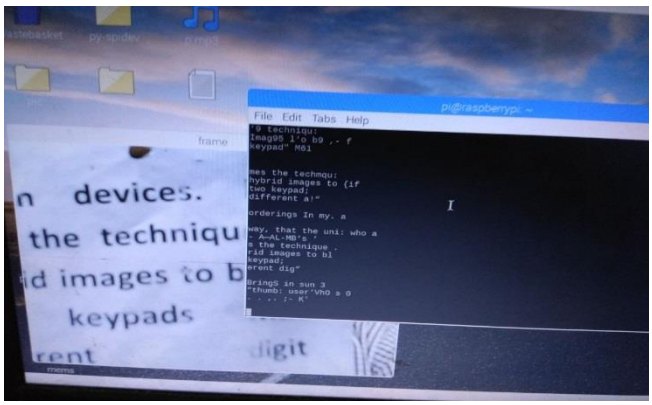

Chart7: Output Figure

The main objective of our project is to convert the text signal to speech signal. This is achieved by the components and methodology which is described above. The main problem that we faced while developing this prototype is that to get the machine operating code from the image and to convert that code into speech synthesizing code. But finally the prototype
is built and thus it converts the text image to the sound signal which can create a revolution for the blind people to read by listening to the speech, output from the speaker.

8. SUMMARY AND CONCLUSION

This paper presented an advanced way to convert the text signal into a speech signal. This paper gives a clear plot to design develop a protocol which can convert the text to the speech signal. This prototype can convert the text which has the font size of 18 and above to speech signal which is the output taken from the speaker or earphones.

Visually impaired people or blind people are those people who lack in their visual perception, and they are unable to see the object. But they have the hearing capability, and this capability makes them hear from the environment. Reading is an important aspect to each and everyone but the blind people use the braille method to detect and to read the text. This method is very difficult to perform and requires practice. To overcome this the prototype is developed in which the text is converted into a speech signal and that sound can be observed by the blind people and they can understand the text in the source of SOUND.

9. ACKNOWLEDGEMENT

We would like to thank all those who provide us the possibility to propose this project. A special gratitude I give to Final Year Project coordinator Mr. V. Saravanaperumal Assistant professor and our Head Of the Department Mrs. S. JeyaAnusuya Associate professor whose contribution in giving suggestions and encouragement helped us for this project to complete.

10. REFERENCE

[1] Pooja P. Gundewar and Hemant K. Abhyankar, “AReview on an Obstacle Detection in Navigation of VisuallyImpaired”, International Organization of ScientificResearch Journal of Engineering (IOSRJEN), Vol.3, No.1pp. 01-06, January 2013.
[2] BhuvaneshArasu and SenthilKumaran, “Blind Man‟sArtificial EYE An Innovative Idea to Help the Blind”, Conference Proceeding of the International Journal ofEngineering Development and Research(IJEDR), SRMUniversity, Kattankulathur, pp.205-207, 2014.
[3] Amjed S. Al-Fahoum, Heba B. Al-Hmoud, and Ausaila a. Al-Fraihat, “A Smart Infrared Microcontroller-BasedBlind Guidance System”, Hindawi Transactions on Activeand Passive Electronic Components,Vol.3, No.2, pp.1-7, June 2013.
[4] S.Bharathi, A.Ramesh, S.Vivek and J. Vinoth Kumar,“Effective Navigation for Visually Impaired byWearable Obstacle Avoidance System”, InternationalJournal of Power Control Signal and Computation(IJPCSC), Vol.3, No.1, pp. 51-53, January-March 2012.
[5] Shraga Shovel, Iwan Ulrich, and JohannBorenstien, “NavBelt and the Guide Cane”, IEEE Transactions on Robotics & Automation,Vol.10, No.1, pp. 9-20, March 2003.
[6] D. Yuan and R. Manduchi, “Dynamic EnvironmentExploration Using a Virtual White Cane”,Proceedings of the IEEE Computer Society Conference onComputer Vision and Pattern Recognition (CVPR),University of California, Santa Cruz, pp. 1-7, 2005.
[7] J. M, Benjamin, N. A. Ali, and A. F. Schepis, “A lasercane for the blind”, Proceedings of San Diego MedicalSymposium, pp. 443-450, 1973.
[8] S. Innet and N. Ritnoom, “An Application of InfraredSensors for Electronic White Stick”, Proceedings of theIEEE International Symposium on Intelligent SignalProcessing and Communication Systems Swissôtel LeConcorde,Bangkok, Thailand, pp. 1-4, 2008.
[9] E. Milios, B. Kapralos, A. Kopinska, and S.Stergiopoulos, “Sonification of range information for 3-Dspace perception,” IEEE Transactions on Neural Systemsand Rehabilitation Engineering, vol. 11, no. 4, pp. 416–421,2003.
[10] J. Borenstein and Y. Koren, “The vector fieldhistogram-Fast obstacle-avoidance for mobile robots”, IEEEJ. Robot. Autumat. Vol. 11, pp. 132-138, 1991.

