International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Electronic Eye for Visually Impaired People

[1]Disha Bhosale, [2]Aboli Chougule, [3]Shravani Kamble, [4]Prajkta Kore

[1,2,3,4] Department of Electronics and Telecommunication Engineering,
[1,2,3,4] Sharad Institute of Technology Polytechnic, Yadrav, India

ABSTRACT:

An essential aspect of human life is communication. One of the most crucial components of a fully communicated message is vision. To obtain information from a text or image, one must have vision. People with visual impairments solely use voice to get information. Because of their condition, those persons frequently have very difficult access to information. The notion for image to speech conversion is implemented in this suggested paper using a Raspberry Pi, a webcam, and either a speaker or an earphone. The suggested method generates speech from images that the camera captures by utilizing the Tesseract OCR (Optical Character Recognition) technology. This makes it easier for those who are blind or visually impaired to retrieve text information from printed sources like books and handwritten notes. Since a power bank is utilized for battery backup and produces speech output in different languages with the aid of Speech API and Microsoft translation, the implemented prototype will be a portable gadget. Millions of people who are blind or visually impaired could benefit greatly from this concept by being able to read, shop, navigate both inside and outdoor spaces, and acquire vision.

***Keywords* – Raspberry Pi, Tesseract OCR, Google Speech API, Microsoft translator, webcam.**

I. Introduction

Approximately 285 million people on our planet, which has a population of 7.4 billion, suffer from visual impairment. Of these, 39 million are completely blind, meaning they have no vision at all, and 246 million have visual impairment that is mild to severe: WHO, 2011. By 2020, 200 million individuals will have visual impairment and 75 million people would be blind, according to predictions [5]. People with visual impairments have many challenges because reading is an essential part of everyday life (text can be found in newspapers, commercial products, billboards, computer screens, etc.). Our gadget reads the text aloud to those who are blind or visually handicapped. Many developments in this field have made it easier for those who are visually impaired to read. The approaches used by the current technologies are comparable to those discussed in this study, however they have specific limitations. First off, the test inputs are printed on a plain white sheet; the input photos from earlier efforts don't have any complicated backgrounds. Without pre-processing, it is simple to convert such images to text; nevertheless, a real-time system will not benefit from this method [1][2][3]. In techniques that employ character segmentation for recognition, each character will be pronounced as a single letter rather than as a whole word. The user hears an unwanted auditory output as a result. We needed the gadget to be able to efficiently read text from any complicated background while still being capable of text detection for our project. Motivated by the approach taken by applications like "CamScanner," we reasoned that the text would probably be encased in a box on any complicated background, such as billboards, displays, etc. We assume that the region containing four points is the necessary region holding the text because we can detect it. Cropping and warping are used for this. After undergoing edge detection on the newly acquired image, a boundary is drawn across the letters. It gains additional definition as a result. After that, the image is processed using OCR and TTS to produce audio output.

II. Literature survey

Numerous studies have already been conducted in this area. A camera-based text reading system for the blind is described in [1]. Fisher's Discriminant Rate (FDR) is used to determine whether to use global or local thresholding in the proposed study in order to obtain a binary image. The method is mainly based on the binarization process from OTSU. It is a segmentation technique based on automatic threshold selection in regions. This approach results in a high value for the FDR when there are two peaks in the local histogram, indicating the presence of characters on a frame. The histogram only has one peak and the FDR is tiny for quasi-uniform frames. Although the FDR values are larger in complicated regions due to the dispersed histogram, they are still lower than in text areas. Image frame detection is done using the FDR in conjunction with a bimodal gray-level histogram. The local OTSU threshold is used to binarize images with high FDR values while maintaining the integrity of the specifications for frames with low FDR values.

A Raspberry Pi prototype for text extraction from photos is shown in the publication [2]. A web camera is used to take the photos, which are then processed using the OTSU algorithm and Open CV. The collected photos are first changed to a grayscale color mode. By adjusting the vertical and horizontal ratios, the photos are resized and cosine changes are applied. Images are subjected to OTSU's adaptive thresholding technique after undergoing various morphological modifications. Following thresholding, Open CV's unique functions are used to create the images' contours. Bounding boxes are drawn around the text and objects in the photos using these contours. Every character in the image is retrieved using these generated bounding boxes, and the OCR engine uses this extracted data to identify the text in the image.

To assist visually impaired people in reading text labels and product packaging from hand-held objects in daily life, a camera-based assistive text reading framework was developed in [3]. The suggested method separates things in the camera's field of view from dirty backgrounds or other nearby objects. The ROI Region of Interest is defined using motion-based techniques. A combination of background subtraction techniques based on Gaussians is used to extract the object region in motion.

To extract text details from the ROI and focus text regions from the ROI of an object concurrently, text localization and recognition are applied. The Novel Text Localization technique in an Adaboost model handles the ramp aspects of pixel edge distributions and stroke orientations. Optical character identification software that is readily available for purchase uses binarization to identify the characters that are present in text inside localized text regions.

With the use of SWT [4], text in natural settings is identified by combining pixels with comparable strike widths into connected components through bottom-up integration. This approach detects letters across a wide variety of scales in the same image. Because it doesn't employ a filter bank of distinct orientations, it can distinguish text lines and strokes in any direction. The information is handled properly to ensure correct text segmentation. Thus, a decent mask will be easily accessible for the recognized text. Filter orientations, the method's intrinsic attenuation to horizontal texts, and the requirement for integration across scales are its drawbacks. The medical imaging and remote sensing domains use linear features that are associated with the stroke definition. While the text contained within a picture can have a very wide range of scale variations, the road width range in a satellite or aerial photo is restricted and well-known in the road detection process. Text won't have the typical low-curvature, long linear structures found in roadways.

III. System architecture

The physical components utilized in the design of hardware include the Raspberry Pi 3, a webcam, a battery bank with a capacity of 10000 mAh, and Bluetooth earphones. The webcam collects the photos, which are then transferred to the Raspberry Pi 3 for processing. The generated speech output is Bluetooth-enabled and delivered to a wire-free headphone. A rechargeable power bank is used to provide the power source.

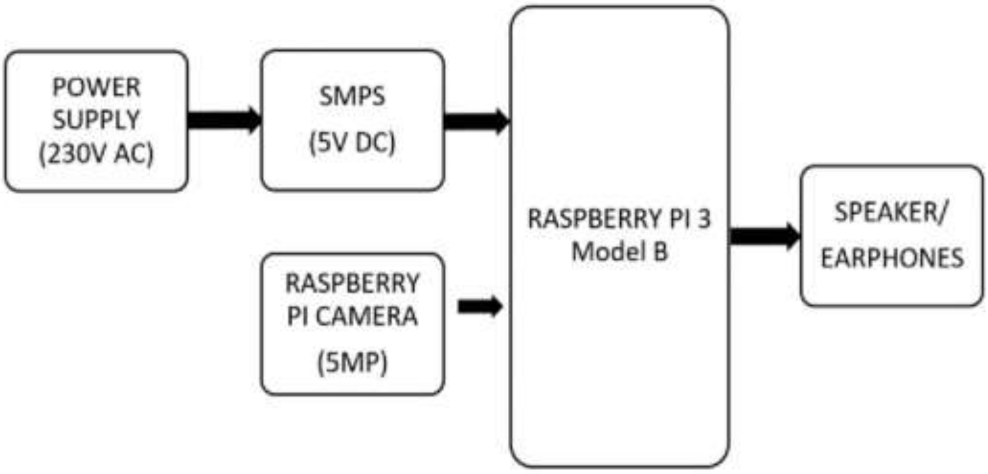
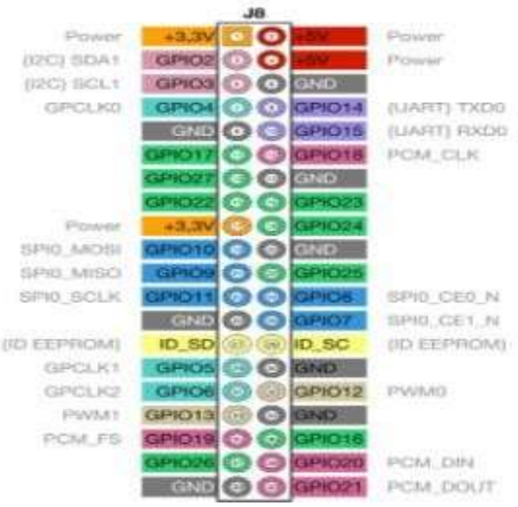


Fig. 1: Block Diagram of System implementation

A. Raspberry Pi 3 Model

The Raspberry Pi is a gadget that combines numerous critical functionalities on a single chip. It's a system on a chip (SoC). Broadcom BCM2837 SoC Multimedia processor is used in the Raspberry Pi 3. The Raspberry Pi's CPU is a quad-core ARM Cortex-A53 processor running at 1.2GHz. It includes a 1GB LPDDR RAM (900Mhz) internal memory and an external memory that can be expanded to 64 GB. The two primary new features of Raspberry Pi 3 are wireless internet connection 802.11n and Bluetooth 4.1 classic. It features 40 GPIO pins. The Raspberry Pi camera has a resolution of 25921944 and a resolution of 5MP. The Raspberry Pi features a 3.5mm audio connection, so it may simply be connected to earbuds or a speaker to hear sounds.

Fig (a): Raspberry Pi 3 Model B



Power	+3,3V	+5V	Power
(I2C) SDA1	GPIO2	+5V	Power
(I2C) SCL1	GPIO3	GND	
GPCLK0	GPIO4	GPIO14	(UART) TXD0
	GND	GPIO15	(UART) RXD0
	GPIO17	GPIO18	PCM_CLK
	GPIO27	GND	
	GPIO22	GPIO23	
Power	+3,3V	GPIO24	
SPI0_MOSI	GPIO10	GND	
SPI0_MISO	GPIO9	GPIO25	
SPI0_SCLK	GPIO11	GPIO8	SPI0_CE0_N
	GND	GPIO7	SPI0_CE1_N
(ID EEPROM)	ID_SD	ID_SC	(ID EEPROM)
GPCLK1	GPIO5	GND	
GPCLK2	GPIO6	GPIO12	PWM0
PWM1	GPIO13	GND	
PCM_FS	GPIO19	GPIO16	
	GPIO26	GPIO20	PCM_DIN
	GND	GPIO21	PCM_DOUT

Fig (b): pin description of Raspberry Pi 3
[Source: raspberrypi-spy.co.uk]

B. Camera Module

The Raspberry Pi camera module is 25mm square and has a 5MP sensor. It connects to the Raspberry Pi computer using a flat flex cable (FFC, 1mm pitch, 15 conductor, and type B).

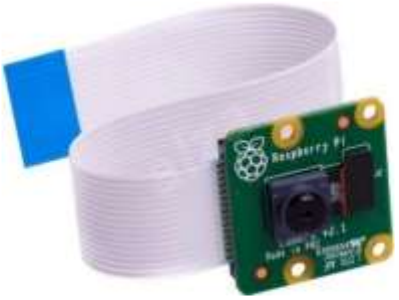


Fig: Camera Module

The Raspberry Pi camera module provides a unique new capability for optical instrumentation, including the following important capabilities:

SD flash memory cards are used to record 1080p video. Simultaneous 1080p live video broadcast via HDMI while recording. Omni Vision OV5647 Colour CMOS QSXGA (5megapixel) sensor, Sensor size: 3.67 x 2.74 mm, Pixel Count: 2592 x 1944, Pixel Size: 1.4 x 1.4 um, Lens: f=3.6 mm, f/2.9, Angle of View: 54 x 41 degrees, Field of View: 2.0 x 1.33 m @ 2 m Full-frame SLR lens, 35 mm equivalent, Fixed Focus: 1 m to infinity, Lens is detachable. Adapters for interchangeable M12, C-mount, Canon EF, and Nikon F Mount lenses. Image mirroring in-camera.

C. Powerbank

To make the model portable, a power bank with a capacity of 10,000 mAh is employed. This battery bank can be recharged and can keep the system running for a day with reduced power consumption by disconnecting the supply to the webcam and other components. It maintains a constant voltage of 5 V and 2 A for an extended period of time. It makes the system more compact and adaptable.

Fig: power bank

D. Bluetooth earphones

Any wired earphone or wireless earphone with Bluetooth connectivity can be used for the purpose. Here, wireless earphone has been used for the ease of convenience.

IV. System software design

Raspbian is a free operating system based on Debian that is designed for the Raspberry Pi device. Raspbian Jessie is used as the RPi's main operating system in our project. Our code is written in Python (version 2.7.13), and the functions are invoked by OpenCV. OpenCV, which stands for Open Source Computer Vision, is a library of functions used for real-time applications such as image processing and many others [14]. OpenCV now supports a wide range of programming languages, including C++, Python, and Java, and is available on a variety of platforms, including Windows, Linux, OS X, Android, and iOS. Our project makes use of opencv-3.0.0. OpenCV's application areas include facial recognition systems, gesture recognition, human-computer interaction (HCI), mobile robotics, motion understanding, and object recognition.

Segmentation and recognition, motion tracking, augmented reality, and many other applications are available. We use Tesseract OCR and Festival software to execute OCR and TTS operations. Tesseract is a free and open source optical character recognition engine distributed under the Apache 2.0 license. It can extract typed, handwritten, or printed text from photos either directly or through an API. It supports a large number of languages. The package is commonly referred to as 'tesseract' or 'tesseract-ocr'.Festival TTS was created by "The Centre for Speech Technology Research" in the United Kingdom. It is open source software with a framework for creating efficient voice synthesis systems. It supports three languages: British English, American English, and Spanish. Festival is simple to install because it is included in the Raspberry Pi package management.

Image Processing

Letters can be found in books and newspapers. Our goal is to extract these letters, transform them to digital form, and then repeat them. The letters are obtained using image processing. Image processing is essentially a set of procedures that are applied to an image file in order to extract information from it. The input is an image, and the output can be an image or a set of image-derived parameters. We can convert the image to grayscale after it has been loaded.

The image we get now takes the shape of pixels inside a certain range. The letters are determined by this range. In gray scale, the image has either white or black content; the white will mostly be the spacing between words or blank space.

Feature Extraction

At this stage, we create feature maps to collect the image's main elements. One such way is to recognize the image's edges, which will include the appropriate text. We can employ axis detection techniques like as Sobel, Kirsch, Canny, Prewitt, and others to do this. The Kirsch detector is the most precise in determining the four directional axes: horizontal, vertical, right diagonal, and left diagonal. This method makes use of each pixel's eight-point neighborhood.

Optical Character Recognition

The mechanical or electronic translation of scanned images of handwritten, typewritten, or printed text into machine encoded information is known as optical character recognition (OCR). It is frequently used as a method of data entry from a paper data source, such as documents, sales receipts, mail, or any number of printed records. It is critical to the computerization of printed texts so that they can be searched electronically, stored more compactly, shown online, and used in machine processes such as machine translation, text-to-speech, and text mining. OCR is a branch of pattern recognition, artificial intelligence, and computer vision research.

Tesseract

Tesseract is a free optical character recognition engine that runs on a variety of operating systems. Tesseract is widely regarded as one of the most accurate free software OCR engines available today. It is compatible with Linux, Windows, and Mac OS.

The Tesseract engine, a command-based tool, is given an image with text as input. The Tesseract command is then used to process it. The Tesseract command accepts two arguments: The first argument is the name of the image file that contains text, and the second argument is the name of the output text file in which the extracted text is saved. Tesseract provides the output file extension as.txt, therefore there is no need to specify the file extension when specifying the output file name as a second argument in the Tesseract command. When the processing is finished, the output is saved in a.txt file. Tesseract produces 100% accurate findings in basic images with or without color (gray scale). However, in the case of some complicated photos, Tesseract produces more accurate findings when the images are in grayscale mode rather than color mode. Tesseract is a command-based tool, but because it is open source and available in the form of a Dynamic Link Library, it may be readily converted to a graphical mode.

V. Translation and Speech Output

Output speech in the chosen language is acquired by passing a.flac file to a developed python program that includes a Google TTS engine, an MP layer, and different authorizations for translator services. The MPlayer supports a wide range of media formats, and any format supported by FFmpeg libraries can be played and saved to a local file. MEncoder is a program that takes an input file and translates it into various output formats using a variety of transformations. The Google text to speech engine will then convert the text to speech. Microsoft Translator, a Microsoft translation service[6], is used for speech translation in the user's preferred language. Typical language code has to be given in the command for execution for getting output in user required language.

To obtain Spanish as the destination language, try using the following command: sudo nano pitranslate.py –o en –d es "filename".The speech in Spanish can then be heard through headphones that are plugged into the Raspberry Pi's 3.5mm port or, if you're using wireless headphones, Bluetooth.

VI. Conclusion

This method makes it simple for visually impaired people to listen to anything they wish to. Additionally, he may translate the material into the desired language with the aid of translation tools, and he can turn that translated text into speech once more by utilizing Google's speech recognition engine. They can be independent in this way. Furthermore, it is less expensive than alternative implementations. With an average processing time of less than three minutes for A4 paper size, a text-to-speech device may convert text image input into sound with a performance that is high enough and a readability tolerance of less than 2%. People can utilize this portable device on their own and without an internet connection. This approach allows us to streamline the editing process.

References

[1] *Ms.Rupali, D Dharmale, Dr. P.V. Ingole, "Text Detection and Recognition with Speech Output Visually Challenged Person", vol. 5, Issue 1, January 2016.*

[2] *Rajkumar N, Anand M.G, Barathiraja N, "Portable Camera Based Product Label Reading For Blind People.",IJETT, Vol. 10 Number 11 - Apr 2014.*

[3] *Boris Epshtein, Eyal Ofek, Yonatan Wexler, "Detecting Text in Natural Scenes with Stroke Width Transform."IEEE, 2010,pp.2963-2970.*

[4] *N. Haering, P. L. Venetianer and A. Lipton, "The evolution of video surveillance: An overview," Machine Vision and Applications, vol. 19, no. 5–6, pp. 279–290, 2008.*

[5] *T. Winkler and B. Rinner. "Security and privacy protection in visual sensor networks: A survey," ACM Computing Surveys (CSUR), vol. 47, no. 1, article no. 2, 2014.*

[6] *Raspberry Pi 3 Model B,[Online].Available: https://www.raspberrypi.org*