Proceedings of the Fourth International Conference on Trends in Electronics and Informatics (ICOEI 2020)
IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision

Leo Abraham
Computer Science and Engineering Department
Saintgits College of Engineering
leo.abraham98@gmail.com

Nikita Sara Mathew
Computer Science and Engineering Department
Saintgits College of Engineering
mathew.nikita@gmail.com

Liza George
Computer Science and Engineering Department
Saintgits College of Engineering
lizageorge2407@gmail.com

Shebin Sam Sajan
Computer Science and Engineering Department
Saintgits College of Engineering
shebinsam98@gmail.com

***Abstract*-- With the advent of new technologies, a whole lot realm of possibilities is open to mankind which otherwise would have been either impossible or a miracle. Computer vision is one such field which has millions of possibilities and this project itself being a primary example of it. This project aims to help the blind society to experience the world independently with the help of a speech-based feedback device.**

This project proposes (1) identifying walkable spaces, (2) text recognition and text-to-speech, (3) identify and locate specific types of objects; and walking navigation which can be incorporated into this project as a future scope. This is implemented with the YOLO algorithm for object detection, which uses the COCO dataset. Our project will help a blind person to walk easily by finding the path, detect obstacles in front of them and thus avoid it. It will help them read texts as well, which is done using OCR which uses python and an API for the text recognition. Thereafter, gTTS is used to convert text to speech, which is the final output for the users.

***Keywords*—Wearable device, computer vision, object detection, visually impaired, text recognition, camera module, audio jack, Raspberry Pi**

I. INTRODUCTION

Deficiencies in the visual system may contribute to visual impairment which can lead to blindness in the worst cases, which may prohibit individuals from performing many day today tasks, including learning, work, and even walking. According to the World Health Organization [2] around 38 million people worldwide suffer from blindness, while the other 110 million have other types of defects.

Recent statistics show that multiple degrees of blindness aff ect 7 in 1,000 people, with a total world population of 5.3 billion. Unfortunately, in developing countries there are over 90 percent of people suffering from blindness.

Under Vision 2020, India will lessen the predominance of visual deficiency to 0.3% of the all-out populace. India presently has around 12 million visually impaired individuals against 39 million all around, which makes India home to 33% of the world visually impaired populace.

Iraq is where the visual impedance is a huge issue because of them across the board fear based oppressor action and inherent variations from the norm in infants brought about by water and nourishment contamination. All things considered, innovative advances have permitted help to be given to the individuals who live in unfortunate conditions. Subsequently, visually impaired individuals are commonly ready to perform day-to-day tasks independently such as driving through the streets and traveling inside houses. In addition to increased autonomy, often individuals with visual difficulties require help to identify challenges and are generally supported by other trusted people. However previous research has suggested many strategies to overcome the issues of visually impaired people (VIPs) and offers higher mobility, but these strategies have not been able to fully address the safety measures when VIPs walk on their own. However, the proposed ideas haven't indicated a mechanism for blind people to be in constant interaction with their loved ones and are generally high in complexity and not cost effective. [2] In this paper, a system is proposed that aids visually impaired people in dealing with day-to-day activities like walking, working, doing house chores and reading texts. This system includes three modules, namely object detection, lane segmentation and finally text recognition. The system comprises of a wearable device which is worn by the visually impaired, whereby the device includes a camera module, a Raspberry Pi 4 and an audio jack. The camera will capture the object's image that is in front of the person, thereafter it gets processed using machine learning methods and in turn the output which is the name of the object will be converted into audio for the user through the audio jack.

II. LITERATURE SURVEY

Ali JasimRamadhan in [1] proposed a system to help visually impaired people guide themselves while walking, seek for assistance and navigate public places. The system consists of a microcontroller, set of sensors, a solar paneland a cellular communication. Buzzers on the wrist are used to alert the user in a more effective way in case of toomuch surrounding noise. These sensors are used to trace the path and inform the clients. If the clientcomes across an accident like stumble over anobject, the location will be sent to a registered number which is a family member, in the form of a text message.

Hseueh-Cheng Wang et al. in [2] proposed a system that provides situational awareness for the blind, with the help of a camera, a haptic device and an embedded computer. The haptic device alerts the user whether there is an obstacle in front of them and also alerts safe walkable areas for the blind. The system uses computer vision techniques to identify walkable spaces and alert safe areas. The user would as a result even navigate themselves through a maze without any obstacles, locate specific objects and walk through a crowded area without the collision with people while walking.

ShoroogKhenkar et al. in [3] proposed a system called "ENVISION" which involves navigation with the help of a smart phone for the visually impaired. It comprises of Global Positioning System technology navigations and innovative object detection methods. It avoids many risks like diversity of the background, poor quality of the video streams and weakprocessing capabilities. Hence it uses another method to attain dynamic and static obstacles by video streaming realtime videos on a smart phone with the help of average capacity hardware.

Jizhong Xiao et al. in [4] proposed a framework which provides context-awareness navigation services to the visually impaired people. This system requires conceptual features of the obstacles that are dealt with in the person's surroundings tointegrate advanced intelligence with navigation techniques. It consists of visual signs and distance calculation toobtain object locations in real time. It also uses data from online platforms like tweets as event-driven communication in order toachieve circumstantial occurrences.

S. Scheggi et al.in [5] stated that help dogs and helping walking canes contribute a descent level of independence to the visually impaired people, but it does not provide precise navigation and locating of specific objects, especially inunknown environments. The system comprises of a pair of camera glasses, two vibro-tactile arm bands and a walking cane in order to discard possible obstacles. For autonomous navigation, the system connects with a remote operator after the image of real time videos are captured, who in return alerts the user with vibro-tactile alerts through the bracelets.

Ja--e Sung Cha et al. in [6] states that second-class citizens +have not been getting supports as the society expands. Oneof the many supports that is important is mobility for the visually impaired people. This proposal utilizes text tospeech techniques for providing directions for the blind and the exact information of the location, using an Android-based smartphone. It uses Google MapAPI to apply mapping information. This system is comparatively cost effective and easyto install into smart phones.

II. PROPOSED SYSTEM

The proposed system includes a wearable device that helps visually impaired people to move around and get their day to day tasks done independently like every other person. The wearable device will be a pouch that contains a Raspberry Pi, camera module and audio jack, that will be worn by the visually impaired user. Fig 1 depicts the proposed system architecture. It consists of object detection, text recognition and text-to-speech conversion.

Basic Operations:

1. Object Detection:

Object Detection is a computer technology related to computer vision and image processing that is used in digital images and videos to recognize instances of symbolic objects of a certain type (such as humans, homes or cars).[7] So basically, this is incorporated in our project so that it can help the blind people identify commonly used objects in daily life as well as in identifying walkable space. Whenever there is an object or obstacle in front of the person, it alerts the person and helps them to avoid it. Objects are classified to help the person know of what objects are in front of them using image processing, therefore helping them to get a sense of their surroundings. [7]

2. Text Recognition:

It recognizes the textual data placed in front of the camera module with the help of text recognition and converts it into speech, thus helping visually impaired to identify the textual information in front of them.

3. Text-to-Speech:

gTTS is used to convert textual contents to audio feedback, which is the final output to the visually impaired user.

The hardware requirements are:

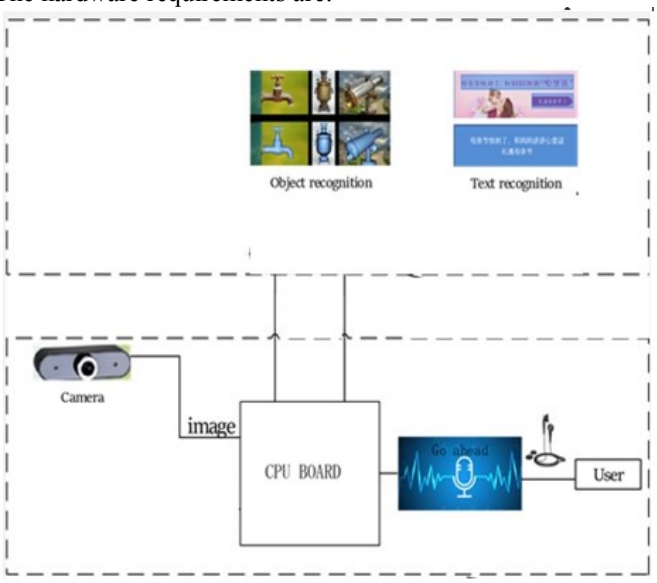

Figure 1. Proposed Model

- **Raspberry Pi 4 Model**

Raspberry Pi is a cost-effective, credit card-sized computer which connects to a computer monitor or television. It's quite beneficial for both personal and commercial use. Model B of the Raspberry Pi 4 features a 1.5GHz quad core 64bit ARMCortex- A72 Processor, 1 GB or 2 GB or 4 GB SDRAM, complete Gigabit Ethernet, Bluetooth 5.0 dualband 802.11ac,

two USB 3.0 and two USB 2.0 and supports up to 2 monitors of 4K resolution.[5] Fig 2 depicts a Raspberry Pi 4 Model B.
First needed to configure the Raspberry Pi from the programming side to communication mode and then upload the code which is in python programming language. The online simulation mode is after the compilation program. The online simulation mode is used to check how the program is running step by step.

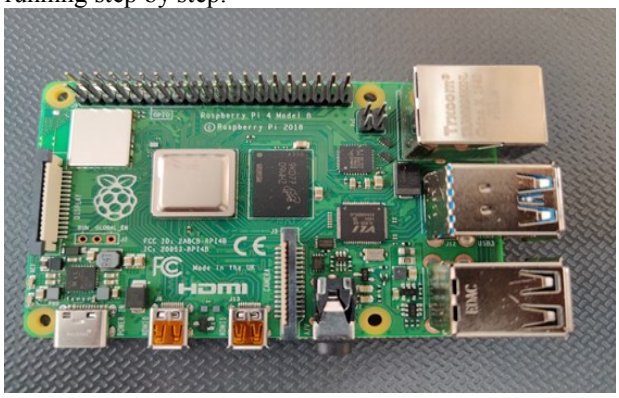

Figure2. Raspberry Pi 4 Model B

- **Camera Module**

The Raspberry Pi requires a camera module to take high definition videos and images which is later used as inputs for various training and analyzing purposes. It also supports 1080 p30, 720p60 and VGA90 video types and still captures. Capturing the video using the camera module is easy with OpenCV, as it does not require any additional softwares. Video capture is a function which is to be initialized for the camera to capture the video, and for that, the class Video Capture is called by assigning it a name. The class copy is used to capture frames from the camera at each loop, where the loop is an endless loop. The trained model is fed with images at the rate of thirty frames per second through the camera. Fig 3 shows the camera module used in this project.

Figure 3. Camera Module.

III. METHODOLOGY

To complete the proposed idea, the system needs to integrate real time object detection, text recognition, text to speech conversion into a smart wearable device for the visually impaired for a real time feedback.

- Text Recognition

OCR(Optical Character Recognition) Text recognition with Python and API is used in this system. OCR is the recognition of a text from an image with the use of a computer. Ocr.space is an OCR engine which offers free API. In short OCR does all the work of text detection by sending their API with the selected image with the required text that want to scan and the returned result is the text scanned. [10] Firstly to obtain an API key by registering to their website which is http://ocr.space/OCRAPI, following that step you have to import the required libraries (OpenCV, IO, numpy, requests and json), whereby IO and Json are already installed on python which makes it default, and then load the image. Thereafter can cut the image with the required area of text incase the image contains some sort of background disturbance. Then the image is to be sent to the ocr.space server to be further processed and need to compress the image into JPG format so that the image will be sent with a maximum size of 1MB using the free service, and this will shrink the size of the image. Later on convert the image into bytes, and then send the bytes to the server using the python library requests, by sending three parameters. First one is the url-api, second is the called “files” which has the name of the file and the bytes of the file that were generated after the compression of the image. And thirdly is the “Data” which includes the post parameters of the OCR engine. The API key then needs to be inserted where it is now written “YOURAPIKEYHERE”, where the language is the language of our text, which is English. Thereafter the function will send the image to the server and will get a response from the server. Finally the result from the server is a string which is converted into a dictionary.

- Text-To-Speech

This system uses gTTS (Google Text-To-Speech) which is a python module and a CL method to interact with the Google Translate Text-to-Speech API.GTTS is a very easy-to-use device that converts the text that the machine has recognized and then translates the text to audio that is saved in an mp3 format. The gTTS API supports many languages such as Hindi, English, Malayalam, French and many more. It is not possible to change the voice of the audio, however it is possible to change the audio speeds to fast or slow.
The first step of the code is to import the gTTS library and the “os” module which is required to hear the converted audio. Thereafter the text is sent to an object called “text”, and then an object called “speech” is created to pass the text and required language to the engine. “fast=false” indicates that the text should have slow speed when read out. The object “speech” that contains the converted file will be saved in a mp3 file named ‘text.mp3’. To play the converted file, the Windows command “start” is used following the mp3 file name.

- Object Detection

In this system, did with the help of OpenCV and the pre-trained deep learning model, ie, YOLOv3 to identify objects. The model “YOLO” algorithm is run through various high complexity convolutional neural network architectures which

is known as the Darknet. Common Objects in Context(COCO) dataset is used to train the model. COCO is pre-trained dataset, so there is no need for external training. [12] OpenCV, keras, and image modules are imported in our python program. The python cv2 package has a technique to set up Darknet from the configurations in the yolov3.cfg file and capture the live video from the Pi camera and then give it as input to a pre-trained YOLO model which has more than 80 objects classified. The prediction of the class of the objects identified in each frame is a string e.g. “dog”. This will also retain the coordinates of the objects in the image and include the position “top left” etc to the class “dog”. Then the text description is sent to the gTTS API using the gTTS package.

IV. RESULT AND PERFORMANCE ANALYSIS

YOLOv3 uses a combination of techniques like Darknet-9 and a residual network. It uses 3×3 and 1×1 convolution layers. It is called Darknet-53 as it consists of 53 convolution layers.[11] High floating point operations are obtained by Darknet-53, which in turn utilizes GPU much more efficiently, thus making it faster. This project uses YOLOv3 because it works extremely faster than the other detection systems with the same performance. YOLOv3 is also a very strong detector in which it outclasses others in obtaining boxes for objects.

The final output of this proposed system is a wearable device that is wearable by a visually impaired person. The device consists of a Raspberry Pi camera which captures the required image in a specific frame using a trigger by the user, thereafter the image is processed with the help of a Raspberry Pi 4, which detects the object, classifies the object, and also detects whether there is an object in front of them, and converts it to audio to the user. The coordinates of the bounding box of all object identified in the frames are also retained, which overlay the boxes on the objects identified, thereafter returning a video playback from the sequence of frames. An audio output is scheduled on the first frame of every second.

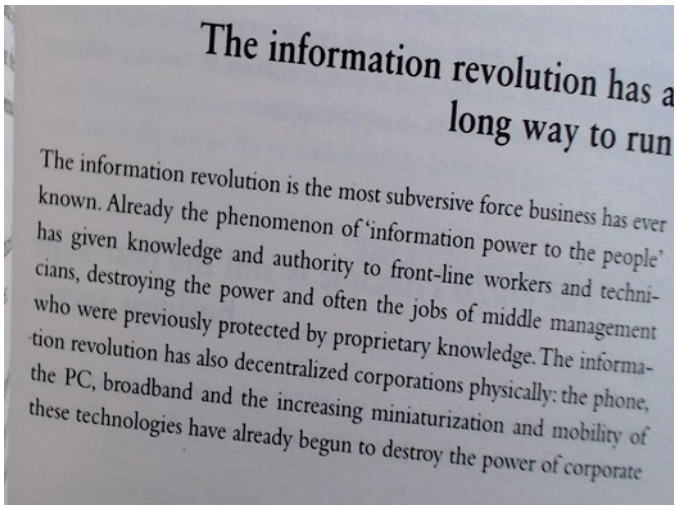

Figure 4. Text image

- Text Recognition

The required image is captured by the camera, processed (the required section is cut) and then recognized by the processor as per the code that is already programmed. The result includes the text that was read from the OCR engine including a few other values which depends on the post parameters that set.

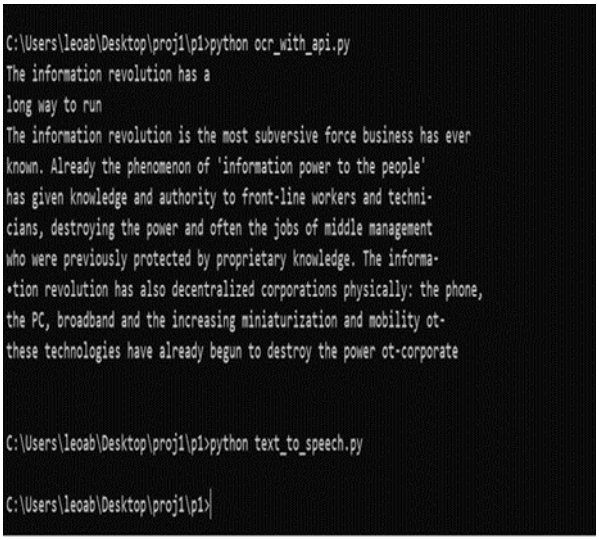

Figure 5. Text Recognized

- Text-to-Speech

An important requirement of gTTS is that is requires an internet connection to function as it solely depends on Google to receive the audio date. The output of the program which was saved as text.mp3 file will play the text which was shown in the above picture. The text will be read out according to the instructions that set: In English and in slow speed.

Figure 6. Output as mp3 file

- Object Detection

The image is captured by the camera and processed using the Raspberry Pi. The result of the image classification will be sent as an audio to the user.

Our system models detection as a regression problem. It divides the image into a S×S grid and for each grid cell it predicts B bounding boxes, confidences for those boxes and C class probabilities. The predictions are encoded as an S×S×(B*5+C) tensor. [9] The final predicted classes are shown as the boxes shown in Fig 7. This was done through a web cam initially. Later on tested it using the Raspberry Pi camera module.

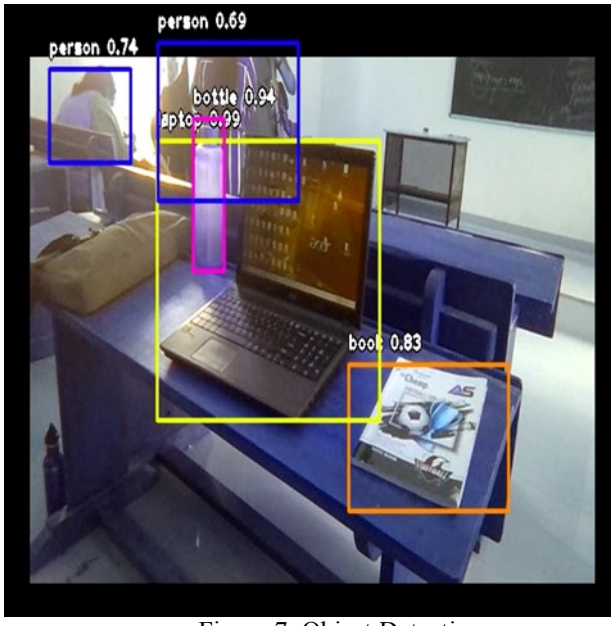

Figure 7. Object Detection

In overall studies, visually impaired aiding devices usually rely on sensors for object detection, where sonar sensors are mostly used. But the drawback of sonar sensors is that it is expensive. Most systems only include object detection and navigation features, whereas our proposed system includes object detection, text recognition and text-to-speech conversion. Our system does not involve the usage of sensors as sensors are not cost effective, and our primary aim was to create a device which is cost effective; which is practical for everyone to purchase. This device will help users to live more independently as make their lives easier. The aim for the future is to include additional features like location navigation functioning in dim lighting and maintain the cost efficiency throughout.

V. CONCLUSION

This paper presents a synopsis of enabling a real-world experience through a wearable speech-based feedback system. The idea of a wearable device that includes a Raspberry Pi 4 and camera module to provide feedback to signal obstacles to the users and also identify and read out texts is proposed. When a text is placed in front of the camera module, the text is first recognized and then read out to the user. Similarly, objects present in front of the user are identified and communicated to the person who is wearing the device. In a study conducted, it was found that visually impaired people had difficulty in identifying whether there are any hindrances in front of them or what textual content is present in front of them and hence our project. Our project solves these challenges and aids visually impaired people to get their tasks done in the same manner as that of a normal person. Our project therefore is aiming to make the living of visually impaired people easier as well as help them get through their daily activities without coming in contact with any dangerous obstacles and wish to incorporate several new features to the system like navigation, which helps the user with directions, like going left or right, and also wish to include object detection in dim light, which is a crucial feature as without it visually impaired people would be living a restricted life.

VI. REFERENCES

[1] Ali JasimRamadhan, University of Alkafeel, “Wearable Smart System for Visually Impaired People”, DOI: 10.3390/s18030843, Available : https://www.researchgate.net/publication/323743137_Wearable_Smart _System _for_ Visually _Impaired_People.

[2] Hsueh-Cheng Wang, Robert K. Katzschmann, SantaniTeng, Brandon Araki, Laura Giarré, Daniela Rus, “Enabling independent navigation for visually impaired people through a wearable vision-based feedback system”, 2017 IEEE International Conference on Robotics and Automation (ICRA), ISBN: 978-1-5090-4633-1, DOI : 10.1109/ICRA.2017.7989772

[3] ShoroogKhenkar, HananAlsulaiman, Shahad Ismail, AlaaFairaq, Salma KammounJarraya, Hanêne Ben-Abdallah, “ENVISION: Assisted Navigation of Visually Impaired Smartphone Users”, Conference on Enterprise Information Systems / International Conference on Project Management / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN /HCist 2016, October 5-7, 2016

[4] JizhongXiao,Samleo L. Joseph, Xiaochen Zhang, Bing Li, Xiaohai Li, Jianwei Zhang, “An Assistive Navigation Framework for the Visually Impaired”, IEEE Transactions on Human-Machine Systems (Volume: 45, Issue: 5, Oct. 2015), Electronic ISSN: 2168-2305, DOI: 10.1109/THMS.2014.2382570

[5] Sampa Jana, ShubhangiBorkar, “Autonomous Object detection and tracking using Raspberry Pi”, 2017 IJESC

[6] Jae Sung Cha, Dong Kyun Lim, Yong-Nyuo Shin, “Design and Implementation of a Voice Based Navigation for Visually Impaired Persons, available : https://www.semanticschol ar .org/paper/Design-and- Implementation-of-a-Voice-Based-for-Cha Lim/0f31cd177359a baa55871f4ebbac4af3c0db468d

[7] ApoorvaRaghunandan, Mohana, PakalaRaghav. “Object Detection algorithms for video surveillance applications.” 2018 International Conference on communication and signal processing.(ICCSP)

[8] ShashankShetty, Arun S Devadiga. “ Ote-OCR based text recognition and extraction from video frames.” 2014 IEEE 8th International Conference on Intelligent systems and control.

[9] DebalinaBarik, ManikMondal. “Object Identification for computer vision using image segmentation” 2010 2nd International conference on education technology and computer.

[10] RavinaMithe, SupriyaIndalkar, NilamDivekar. “ Optical Character Recognition” 2013 International Journal of recent technology and engineering.

[11] Joseph Redmon, Ali Farhadi. “YOLOv3: An incremental improvement” 2018

[12] Dae-Hwan Kim, “Evaluation of COCO Validation 2017 Dataset with YOLOv3”, 2019 Journal of Multidisciplinary Engineering Science and Technology