Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI:10.21474/IJAR01/16228
DOI URL: http://dx.doi.org/10.21474/IJAR01/16228

RESEARCH ARTICLE

AUTOVISION - SPEECH BASED FEEDBACK SYSTEM FOR THE OPTICALLY DAMAGED PEOPLEUSING RASPBERRY PI

Remegius Praveen Sahayaraj L. and Mohamed Irfan A.
Loyola-ICAM College of Engineering and Technology, Chennai-600034.

Manuscript Info

Manuscript History
Received: 10 December 2022
Final Accepted: 14 January 2023
Published: February 2023

Key words:-
Object Detection, Optically Damaged, Text Recognition, Camera Module, Text-Speech Conversion, gTTS, Audio Jack, Raspberry Pi

Abstract

There are millions of people around the globe with various kinds of visual defects and blindness known as Optically Damaged People. As everything is dark for Optically Damaged People, they usually perform daily tasks based on calculations. This project aims to help the blind society to experience the world independently with the help of a speech-based feedback system. This project proposes (1) text recognition and text-to-speech, (2) identify and locate specific types of objects; **and walking navigation which can be incorporated into this project as a future scope.** Our project will help a visually-impaired person to walk easily by detecting 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.

Introduction:-

Visual impairment may be caused due to many reasons such as naturally, during birth or maybe due to terrible accidents leading to loss of sight.Though science has created a greater advancements in performing surgeries, in order to retrieve their eyes back. This method proves to be an expensive method.But in recent year's technology has shown its way of improving life.In 2021, WHO has an estimate of about 285 million people with severe as well as partial blindness despite the age.Primarily 43% are suffering from refractive disorders and 33% are suffering from cataract.

The fear of visual impairment among children is rapidly increasing. A survey conducted in Jordan stated that nearly one-hundred and eight were surveyed and that children both male and female with Optical damage had a fear for communicating as well as social communication. This also led to their general growth being affected in a massive scale. Also, these Children highly feared physical threat due to not being able to see things with their eyes.

In order to avoid this kind of fear and have less impact. Many arrived with solutions like mechanism which can sense the surroundings and give them mobility, suits which the optically damaged can wear and the system warning them in case of emergency, and also many other wearable devices which have been able to sense a flaw and beep the nervous system which will bring alert to the blind people. Also, the nearest yet arrived device was an application developed which will be connected to Bluetooth and internet continuously and when needed to locate where the person is standing, The application will guide them back to their home considering all vehicles as well as poth holes.

Corresponding Author:- Remegius Praveen Sahayaraj L.
Address:- Loyola-ICAM College of Engineering and Technology, Chennai-600034.

Also an additional feature such as when the Optically Damaged People when they get into any vehicle and step out in a new place will be, the current location will be delivered as speech feedback by the application.

The system proposed is primarily to make Optically Damaged People to get rid of the stick that they use and also to perform their basic functionalities such as walking, household activities, fetching water, a simple hold of the remote. The system is a complex yet unique combination consisting of Raspberry PI 3b+ and PI camera module with a camera clarity of 1080p. The PI has an audio jack for audio output. The system will recognise the objects in front of them and will convert them into speech based feedback which can be easily hear through the earphones connected to the audio jack

Literature Survey

Joao Jose in [1] proposed a system which is merely a prototype for blind aid and visually impaired in guidance for navigation to a certain destination considering all types of obstacles and avoidance from pot holes and sidewalks. It is detected based on how far the white cane they carry could reach and detect a small camera with a portable computer is worn with the headphones connected to navigate.

Roberto Manduchi [2] has suggested a worthy technology with mobile vision operated by blind users. It can be used to communicate with other blind person with the help of sign. A cell camera which captures the signs will make it audible through the earphones.

ShoroogKhenkar [3] also proposed a similar mechanism but using smartphones. This system will generate decisions to navigate the visually impaired along with a fusion of GPS technology for directions. "ENVISION", it uses a dynamic, accurate real time video streaming being recorded in a smartphone.

Jizhong [4] gives a system very close which will a wearable device near the waist with multiple sensors like RGB, IMU, vibro-tac. It is an IOT based system which will interact with the environment. It will identify the gestures and warn the blind person through this method using a speakers. Since these devices only send a feedback as output. These devices can be highly considered as only feedback devices and not wearable devices. This device has also made enhancements to note down numbers and alphabets in the wall and doors. It is primarily based on 3-D Map technology which will create different colour based on localization for the visually impaired. Also future requirements were made for embedding GPS technology into the belt when in situation of urban areas and or tall buildings.

Proposed System

The system we proposed is like a wearable device which can be used by the visually impaired. The device consists of raspberry PI and a PI camera module which will be attached to the PI and the PI will be connected to a power Bank. This will be tied to the spectacles and speech will be heard by the use of earphones. The process consists of custom-object-detection and a text-to-speech converted output.

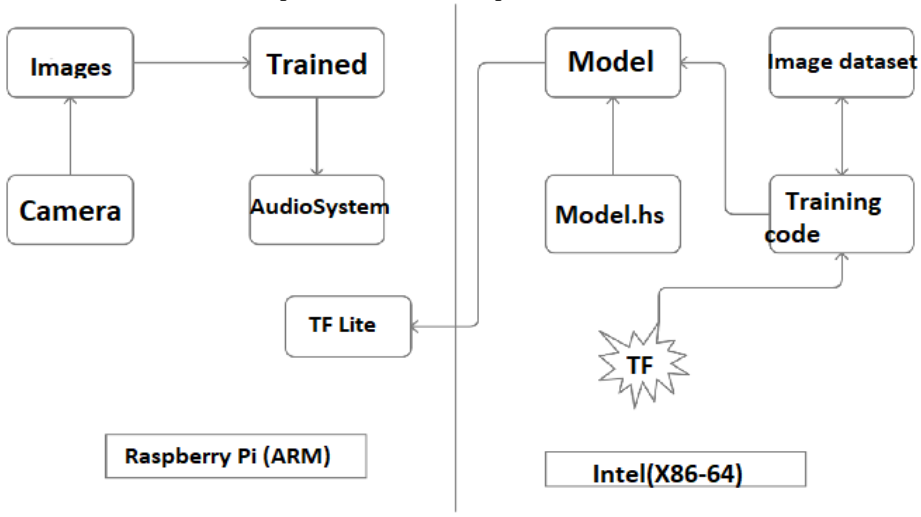

Fig 1:- Architecture Diagram.

Hardware Requirements and Operations

Fig 2:- Raspberry PI 3b+

Raspberry PI

Raspberry PI is a compact,small sized computer capable of running at a performance of workstations. Being low cost and its capability of modularity and open design it is mostly used in IOT devices as well robotics. Model B was a predecessor, whereas B+ is the latest one with faster processor of 1.4GHz 300 Mbits/s gigabit ethernet speed 2.4/5 GHz dual band wi-fi speed. It has a ram capacity of 1 Gb. Fig 2.Depicts the architecture of the Pi model

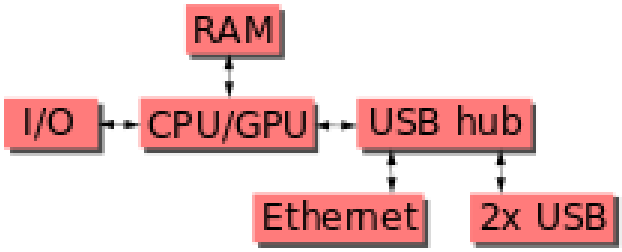

Fig 2:- PI architecture model.

This model lacked the power of handling 4K resolution images. Primarily we need to configure the raspberry Pi for programming and it is required to install certain packages, and further run the code. We can simply use a simulator to understand how the program runs or executes without the use of HDMI cable and a monitor and other necessary hardware kits.

Camera module

The method requires to capture live video, thus in order to do so, we require a PI camera Module which will be able to take high definition videos and images of 1080p quality. It can take pictures upto 12 megapixel and has no infrared. It has a maximum frame rate 50 to 90fps. There is no limitation on frame size. Once the camera module has been connected it is very easy to capture videos using openCV. Video is made to play and in the background OpenCV makes the video to run in an endless loop until the program gets aborted. Since the frame rate is only affected by the computation power of the model of raspberry Pi. So, It is recommended to have Raspberry Pi 4 or greater version to run videos in>40 fps.

Fig 3:- Camera Module.

Object Detection

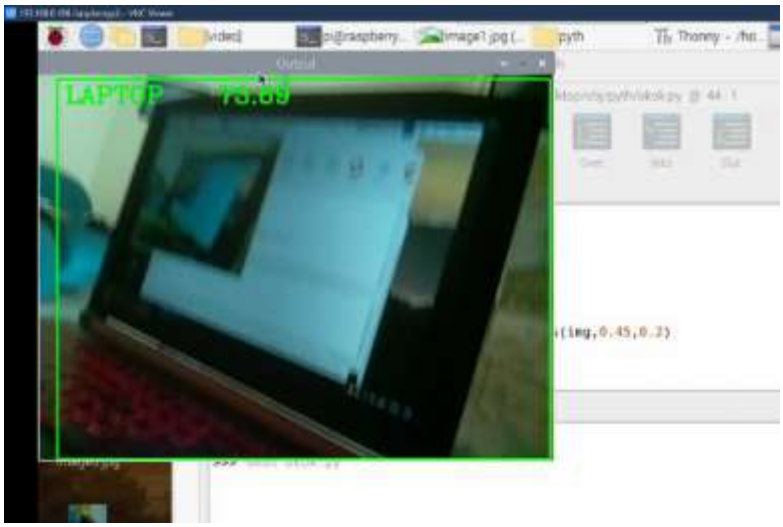

Fig 4:- Object Detection.

Object detection uses machine learning model which is a field related to computer vision and image processing. It can identify objects of any grade based on the training sample, irrespective of being static images or dynamic videos. This method is the primary one for identifying objects. It will help the blind people in identify the objects based on the video being captured. The objects will be trained and input will be processedbased on live objects and the objects which are regularly used by the Optically Damaged People.

Text-to-Speech

Since recognizing the text is already embedded inside the object detection for avoiding latency in running the program efficiently. Once the text has been identified it is passed through pre-built software module which is known as Google-text-to-speech translator which is capable of translating the identifies text into speech

Methodology:-

In order to complete the idea. The System must be integrated to real time object detection along with recognizing of text, finally text to speech conversion and all of this being done on an easily wearable device for real time speech feedback mechanism.

Object Detection& Text Recognition

In the proposed System we have used OpenCV and Deep neural networks for detecting the objects. We are using a pre-trained data set known as COCO (Common object Common Dataset). This reduces an overhead of training the data externally. Cv2 being a powerful package is capable of setting up MobileNet configuration dynamically for live capturing of video. In this model we are using MobileNet-SSD detection network. Since it has been efficiently trained on COCO data sets which has nearly 80 images being pre-trained. The pre-trained names are taken as input using array and is passed through the detection model containing weights. Also in order to avoid overlapping bounding boxes,we are using confidence threshold and nms threshold and we are flattening the image.

In order to avoid latency in running as a package we have also embedded the text recognition into the object detection module. It is not as an API but as boxes predicting the objects. Once the object is identified it will be appended with the coco names and stored as an array. Afterthat, it will return these array names. Finally if there is an object detected it will return the names as array,otherwise as a blank array.

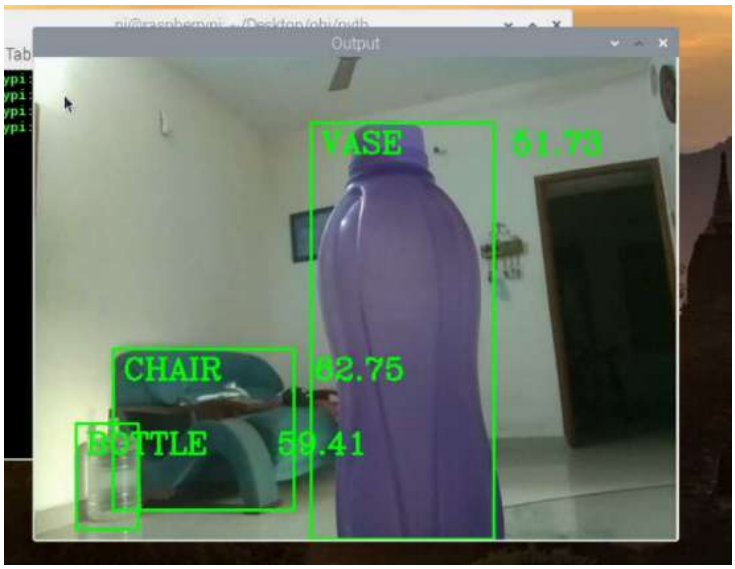

Fig 5:- Object Detection with multiple objects.

Text-to-speech conversion

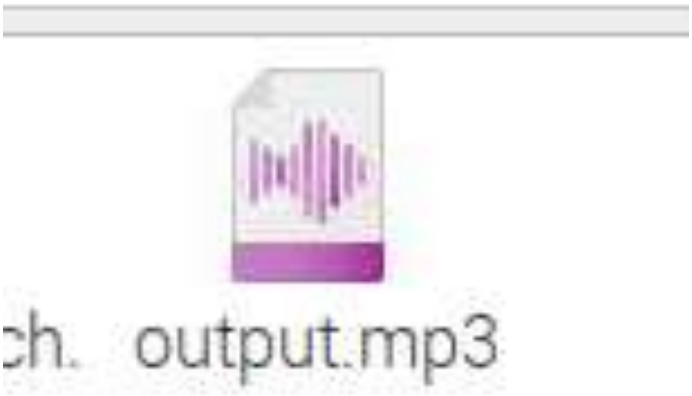

Fig 6:- Audio file after object detection.

Our system uses gTTS (Google-Text-to-speech).It is a python library and a CLI tool to interface Google Translate's Text-to-Speech API. It can detect files of .mp3 and .wav format. Since it is used as a tokenizer it can be used to read unlimited lengths of text, it keeps improper intonation in gaps, decimals. It supports multiple languages irrespective of the language of the text such supporting languages are German, French, Tamil, Hindi. It also has a feature to reduce as well as increase the speed of the text being read. In our system text to speech conversion is added as a package. At first, we are using a gTTS and playsound, if the object is not there it is made to hear "there is no object detected" and once each text is being heard, before the next object is appended "and" text is used in between the objects. Also, in order to know the recording of the objects I have stored it as an .mp3 file called output.mp3 and this output.mp3 is being read using playsound module.

Performance Metrics

MobileNet_SSD_v3 is a combination technique of COCO model and deep neural network as a caffe implementation framework. It uses deep convolutional neural network. Since we are using raspberry PI we cannot have a dedicated GPU built over the layer of raspberry PI. In order to increase the process we are using this type of network. The bounding boxes are completely eliminated but as a proof of concept for this system to be built in a real world scenario, we can train images and store it in Amazon S3 Bucket. But now it has certain objects which will be erroneous for certain pre-trained objects. Otherwise, the model is working at a high performance ratio. Future works can be extended if so, options like text recognition in the documents can be embedded along with object detection such as object character recognition.

Conclusion:-

This paper aims on allowing real time simulation of using the device and to use it for voice based feedback mechanism. The blind users can use real time objects as reference and can use it to hear the objects recognized by the model. The model will have a final output with an accuracy greater than 0.85. Apart from the fact that for a certain objects it will detect objects not live. Also in the future work adding such as navigation of direction for maps on real time analysis can be included.

References:-

[1] João José, M. Farrajota, Joao Rodrigues, J. M. H. du Buf , "The Smart Vision Local Navigation Aid for Blind and Visually Impaired Persons".

[2] Roberto Manduchi , " Mobile Vision as Assistive Technology for the Blind: An Experimental Study".

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

[6] Narasimha Saii Yamanoor;Srihari Yamanoor
"High quality, low cost education with the Raspberry PI"
2017 IEEE Global Humanitarian Technology Conference (GHTC)

[7] "Visual Impairment aid using haptic and sound feedback"
Syed Tahir Hussain Rizvi;M. Junaid Asif;Husnain Ashfaq
2017 International Conference on Communication, Computing and Digital Systems

[8] "Raspberry Pi Based Wearable Reader For Optically Damaged People with Haptic Feedback" S. Srija;P. Kawya;T. Akshara Reddy;M. Dhanalakshmi
2020 International Conference on Electronics and Sustainable Communication Systems

[9] "Bionic Kinect device to assist Optically Damaged People by haptic and voice feedback" Fawad Ahmad; Tanveerulhaq; Irfan Ishaq; Danish Ali; M. Faisal Riaz
Published in: 2016 International Conference on Bio-engineering for Smart Technologies.