As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 8(04):342-350, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0804059
Available online at: http://www.ijmtst.com/vol8issue04.html

Electronic Aid for Visually Impaired People Using Raspberry Pi

Gorli Nikhilanvitha | Perumalla Bogeswar Gopi Nadh | Vankayalapati Sowbhagya Lakshmi | Vinnakota Lokesh

Department of Electronics and Communication Engineering, R.V.R&J.C College of Engineering, Guntur, India.
Corresponding Author Email ID: gnikhilanvitha@gmail.com

To Cite this Article
Gorli Nikhilanvitha, Perumalla Bogeswar Gopi Nadh, Vankayalapati Sowbhagya Lakshmi and Vinnakota Lokesh. Electronic Aid for Visually Impaired People Using Raspberry Pi. International Journal for Modern Trends in Science and Technology 2022, 8(04), pp. 342-350. https://doi.org/10.46501/IJMTST0804059

Article Info
Received: 17 March 2022; Accepted: 13 April 2022; Published: 18 April 2022.

ABSTRACT

Among all the senses vision is one of the crucial human senses. Vision plays a key role in human perception of surroundings. But the perception is different for people with visual impairment. Blind people are facing many problems in their daily lives due to their impairment as they are unaware of the dangers in front of them. It is observed that they are finding it difficult to roll their day-to-day life without assistance. Our goal is to help them to live in the current world irrespective of their impairment. In the motive of supporting them, we came up with the idea of designing a device using raspberry pi which will be used as a real-time guiding system that will solve their navigation problem to travel without any difficulty with minimal assistance. This system will detect the objects and gives the necessary information about that object like its location, and the distance of the object from the visually impaired, and also performs optical character recognition. All this information will be conveyed to the person through audio commands. Object detection is done using the You Look Only Once (YOLO) algorithm. The process starts with capturing the real-time images and sending them for analysis to get the calculated results. The results obtained are conveyed through a hearing aid

KEYWORDS: Object Detection, Yolo, Optical Character Recognition.

1. INTRODUCTION

According to World Health Organization (WHO), there are 285 million visually impaired people worldwide. Among those 39 million are blind people. Of all the problems faced by blind people, the most important one is to find obstacles while traveling. In this current world where people were running to catch up with their dreams, it is not possible to normal people to stop running for a while to help the blind. Similarly, it is very difficult for the visually impaired to navigate and travel alone. Even the regular tasks such as walking, reading, traveling, and socializing become difficult for them. The traditional and oldest mobility aids such as the walking cane, guide dogs, and also the braille method where people used to read and understand through their fingers. However, there are pitfalls to these aids. With the rapid advancement in technology, it has become possible to design and develop technological solutions both in the hardware and the software field that can help a visually impaired person navigate freely. With the witness of a new era of deep learning our primary motivation for our team is how to build a real-world application that benefits the "blind" of a section of society.

In this context, we proposed a guidance system called Electronics Orientation Aid (EOA) which allows obstacle detection and intimation of obstacle location for the visually impaired person and helps them to reach the desired destination with greater awareness of the environment and also it performs the optical character recognition. The real-time images of objects in the surrounding are captured and then they are detected. Once the object has been identified, the user will be specified about the object type, whether it is a hindrance or not. Audio instructions will be given and will be heard by the visually impaired people, to ensure the objects in front of them. In this project, we will explore Object detection and localization using a deep learning algorithm and integrate it with Raspberry Pi and a camera. For the capturing of the images and processing, a raspberry pi board with a camera and speaker will be used to detect the obstacle and alert the user by reading out the distance and type of obstacle/object in the path of motion. The underlying algorithm is Yolo (You only see once) algorithm. The model will be trained to identify the type of object and calculate the approximate, if not exact, distance of the humans from the camera. Once the object is detected, the type of object for humans will be read out using a speaker. Thus, using EOA with a boot-start feature, the user can navigate with more ease and without assistance

2. STRUCTURE OF PAPER

Thepaperisorganizedasfollows: In Section 1, the introduction of the paper is provided along with thestructure, importantterms, objectives, and overalldescription. In Section 3 we discuss existing systems. In Section 4 wehave Literature Survey. Section 5represents the design implementation. Section 6 tellsusaboutthemethodology. Section 7 tells usaboutthe working mechanism. Section 8 concludes the paper with the future scope along with acknowledgment and references.

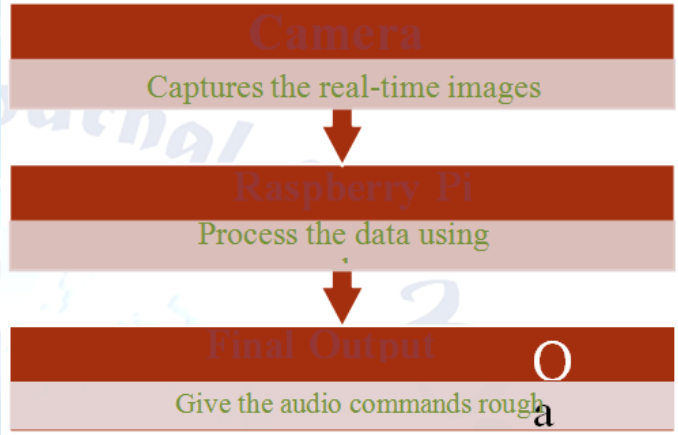
3.EXISTING SYSTEMS

Among the current frameworks, one framework incorporates an ultrasonic sensor distinguishing the article in three bearings for example in front, left, and right. The other framework incorporates object detection and distance computation through GPU [8].

4. LITERATURE SURVEY

Millions of visually impaired or blind people need helping hands. For many years a simple automated tool called “Walking cane” is preferred as it detects uneven surfaces and holes in the ground. Later it was improved with the addition of a remote sensor. Challenging the above one “An electronic walking stick” and also a “Smart stick for the blind” came into existence. Research papers such as “Electronic Aid ForVisually Impaired Through GPU” [8], “Wearable object detection system for the blind”, “A mobility aid for the blind with discrete distance indicator and hanging object detection”, "Object Detector for Visually Impaired with Distance Calculation for Humans" [7], etc were considered for our reference. Various algorithms were studied from different papers like “Object Detection Based on YOLO Network”, “Real-Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility” etc. encouraged us to select the correct algorithm and guided us to complete our project with better understanding.

5. DESIGN IMPLEMENTATION



6. METHODOLOGY

6.1 HARDWARE ORGANIZATION

6.1.1 Raspberry Pi:

The Raspberry pi is a series of single-board computers developed by a raspberry pi foundation in the United Kingdom which is a low-cost solution for many activities. All series and models feature a Broadcom system on a chip (SoC) with an integrated ARM, a compatible central processing unit (CPU), and an on-chip graphics processing unit (GPU)We have used the raspberry pi 3b+ model of Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64- bit SoC with processor speed ranging from 700 MHz

to 1.4 GHz for the Pi 3 Model B+; on-board memory ranges from 256 MB to 1 GB LPDDR2 SDRAM. First-generation of Raspberry Pi 1 Model B was released in February 2012, followed by the Model A and raspberry pi 1 model B+ with an improved design. Raspberry Pi 3 Model B+ was released in February 2018. It has several features, for example, remote LAN, Bluetooth 4.2, Gigabit Ethernet over USB 2.0, Extended 40-pin GPIO header, Full-size HDMI, CSI camera port for associating a Raspberry Pi camera, DSI show port Camera Module audio via Headset for interfacing a Raspberry Pi touchscreen show,4- port sound system and composite video port, Miniature SD port for stacking the working framework and putting away information. The Raspberry pi utilizes Raspbian, a Debian-based Linux working framework.

Fig 1: Raspberry Pi 3B+ Model

6.1.2 Camera:

The input from the camera is given to raspberry pi and we have used a 5MP Raspberry Pi 3 Model B. The 5MP camera module is perfect for small Raspberry Pi projects. The high-definition 5MP camera can also shoot video, ideal for drones or a CCTV project. The lightweight camera module can be utilized in more viable jobs, like a secret camera. The actual sensor has a goal of 5 megapixels and has a proper center focal point installed. The camera is fit for 2592 x 1944pixel static pictures, and furthermore upholds 1080p30, 720p60 and 640x480p60/90 video.

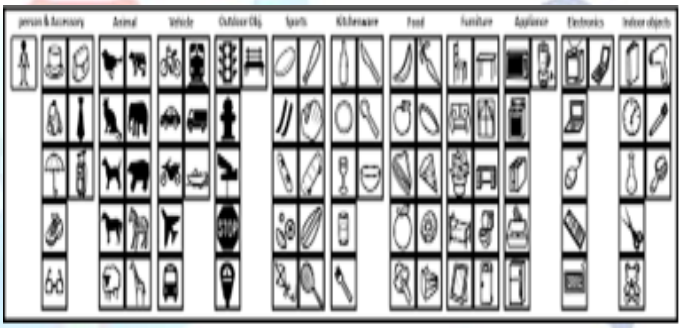
Fig 2: Raspberry Pi 3 Model B Camera

6.1.3 Headphones:

To convey the information through voice commands to blind people we need earphones or headphones. Any headphones that can be connected to the raspberry pi can be used.

6.2 Dataset:

We have utilized the "coco" dataset which has enormous scope for object detection, segmentation, and captioning datasets. COCO represents common objects in context. As implied by the name, pictures in the coco dataset are taken from ordinary scenes accordingly joining "context" to the objects captured in the scenes. Coco was an initiative to collect natural images that reflect everyday scene and provides contextual information. It gives availability to more than 200,000 pictures and 2,50,000man occasions named with central issues. It has 80 trained classes and the classes are as per the following:



6.3 Technologies Involved:

6.3.1 Computer vision:

Computer vision is a field of artificial intelligence that trains computers to decipher the visual world Through this, pictures can be caught, put away, examined, and comprehended. Here, the result is a distinguishing proof or numerical estimations. The rationale behind doing this is to pursue the vital choices given the profoundly powerful continuous information and its examination. computer vision is a field where comprehension of human vision can be copied. Frequently alluded to as image analysis. The thought is to involve Computer vision for object discovery, which could likely assist the visually impaired with envisioning the general climate better.

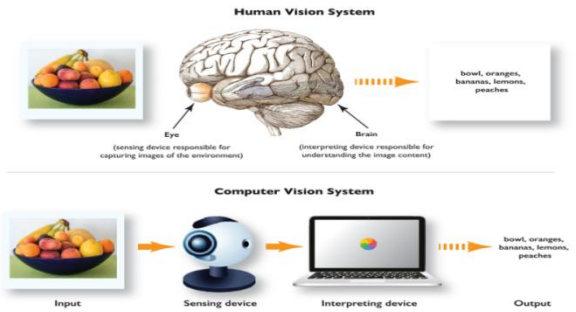


Fig 3: Representation of Human and Computer Vision Systems

6.4Object detection:

Object detection could be a computer vision operation that detects the images/videos and creates bounding boxes around them to locate that object. The annotated text on the bounding box is often translated into voice response and therefore the fundamental positions of the objects may be provided from the perspective of the person's location. By recognizing the front objects and making them aware of the danger, the article sensing module within the system will help the blind to also provide a secure route to succeed in the destination [6]. There are three main Object detection Algorithms and they are as follows:

6.4.1 Region-CNN (R-CNN):

To resolve the matter of choosing a sizable, amount of regions, Ross Girshick et al. proposed a technique employed to extract 2000 regions from a picture called region proposals. So, now rather than classifying huge no. of regions, RCNN works only with 2000 regions where region proposals are generated using a selective search algorithm [2].2000 district of recommendations is limited to a square and taken care of into a convolutional brain network that delivers a 4096 layered highlight vector yield.

6.4.2 Single Shot Detector (SSD):

Object detection and classification tasks are performed in a single forward network pass. Multi-Box is utilized in this to bound box relapse procedure. This strategy finds some kind of harmony between speed and accuracy. It runs a convolutional network on an input image only once and computes a feature map. It predicts bounding boxes after multiple convolutional layers to hold the scale [1].

6.4.3 You Only Look Once (YOLO):

Object detection is a task of computer vision involving the localization of one or more objects within an image and the classification of each object in the image. A troublesome Computer vision work incorporates both proficient place of objects to find and draw a bounding box around each item in an image, and classification of objects to determine the appropriate type of object that has been identified. The solution involves a single deep neural network (originally a Google-Net version, later revised and renamed Dark-Net based on VGG).

6.5 Deep Learning:

Deep learning is a subset of machine learning in AI that has networks capable of learning from data that is unstructured or unlabelled also known as deep learning. It uses algorithmic models that enable a computer to teach itself about the context of visual data. We have used the deep learning algorithm "Yolo" for object detection in our project.

6.5.1 You Only Look Once (YOLO):

YOLO is one every of the quickest algorithms to detect objects. Yolo trains whole pictures and increments discovery productivity straightforwardly. All the sooner algorithms were wont to classify the item within the image using regions. Parts of the image which contain the thing are highly likely. In YOLO the bounding boxes and sophistication probabilities for these cases are determined by one single convolutional network. YOLO's very quick. We simply run our CNN on an image to forecast detection [6]. You merely Look Once (YOLO) [5], the computer vision device capable of detecting a variety of objects in one image, with accuracy the same as Retina-Net, but with higher inferiority than with other existing systems, like Single Shot detector, R-FCN, and FPN FRCN. Its speed makes it one of the foremost suitable real-time object detection systems employed in systems like robotic operations. The algorithm processes the complete image, not only the region of its inferences, which reflects the general context of the image, thus making it less likely to detect background content in an object, which enables its inferences to contemplate the general significance of the image[14]. YOLO includes a unique, jointly trained pipeline. In contrast to other systems with different components, like Faster RCNN, which must be trained

separately. It is widely employed in real-time applications because of its tremendous speed. YOLO was one of the foremost popular algorithms for object detection. Some important things to grasp are:

YOLO takes a picture and divides it into an S x S grid, where S may be a number. Each pixel within the image causes a finite number (5 in our case) of boundary box estimates. The pixel is chargeable for predicting when the middle of the object is found. Of all the boxes found, it is liable for identifying just one object and therefore the other identifiers are rejected. It predicts C conditional class probabilities (one per class for the likeliness of the article class). Total detections to be done per image= SXS ((B ∗ 5) + C) Where SXS = total number of images Yolo divides the Input B into the Number of Bounding boxes detected within the image. For every bounding box, 5 elements will be detected which are Centre coordinates (x, y), Height and Width, Confidence percentage. C =Conditional probability for Number of Classes. For instance, if the image is split into a 2x2 grid, and 10 boxes are predicted with 3 classes (Bat, Ball, Gloves), we'll have 2 ∗ 2(10 ∗ 5 + 3) predictions= 212 predictions.

6.6 Different Versions of YOLO:YOLO requires a Neural Network framework for training and during this case, we've used Dark-Net.

YOLOv1: Version 1 contains a total of 26 layers, with 24 Convolution Layers followed by two Fully Connected layers. The most important problem with YOLOv1 is its inability to search out the tiniest things. Latter, 2 more versions of YOLO were released.

YOLOv2: After each Convolutional layer there are batch normalization layers. Compared to YOLO v1 it's 30 layers. Introducing the anchor boxes. Anchor boxes are predefined boxes that a user gives the network a concept of what number of things should be received on the Darknet. a collection of coaching materials should be accustomed to calculating it. there's no fully connected layer. Of coaching photos from 320-608, fixed sizes were taken. Many labels are often given to identical things, but still, the matter is with many symbols. YOLO v2 was still not efficient for little objects.

YOLOv3: In YOLOv3 there are 106 layers of neural networks. It uses a variant of Darknet, which originally has a 53layer network trained on Image-net and 53 more layers are stacked onto it for the sake of detection. The foremost important feature of YOLOv3 is that it makes detection at three different scales. It is fully conventional and specified that the output is obtained by applying a 1x1 kernel on the feature map. The detection kernel is of the form1 x 1 x (B x (5 + C)) where B is the bounding boxes a cell can predict on the feature map,'5' represents the full of bounding box attributes and confidence score and c is that the number of classes. YOLO v3 makes predictions at three scales, the primary detection is formed by the 82nd layer, the second detection is created by the 94th layer, and also the final detection is at the 106th layer. With three predictions it'll detect the objects in a picture and classify the objects with accuracy.

6.7 YOLO Algorithm for Object Detection:

YOLO algorithm passes the image through the complexities of a neural spec called Darknet [1]. Darknet could be a neural C and CUDA-based open-source spec that's simple to use and supports computation for CPU and GPU systems. In this proposed system YOLOv3 model which is more enhanced and sophisticated is employed. A pre-prepared model Common Objects. In Context (COCO) dataset is prepared on yolov3 [2]. Also, the python cv2 package includes a method to set up Darknet from yolov3.configuration file. OpenCV is the computer vision library/ framework that we are going to be using to support our YOLOv3 algorithm [3]. OpenCV has inbuilt support for Darknet Architecture. Darknet Architecture may be a pre-trained 80-class platform. Our goal is to spot objects using Python language with Darknet (YOLOv3) in OpenCV. The input from the camera is employed to feed images. Then the input is passed through resulting in the classification, localization, and calculation of assorted things like identification of the detected object from the 80 class of the trained model. The measurement of the object's position is set by its coordinates. This helps us to seek out the gap of the object from the camera.

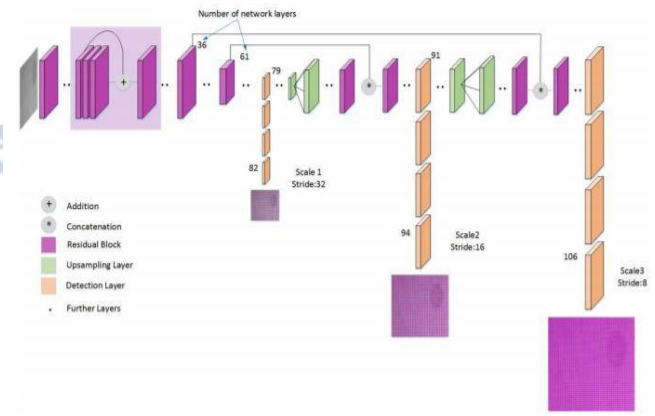


Fig-4: Yolo Network Architecture

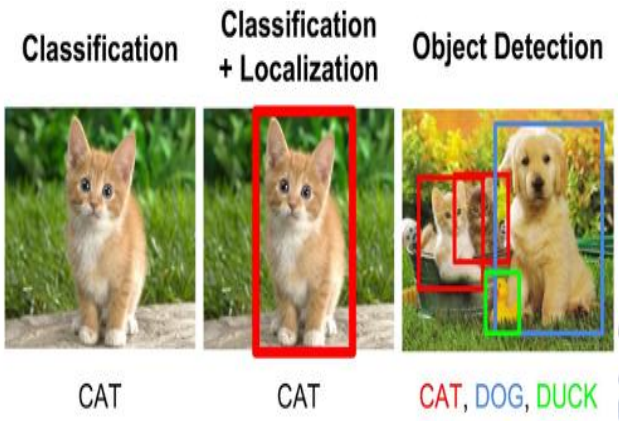


Fig-5: Different phases of Yolo

The following points explain the process of object detection and how bounding boxes are drawn. As shown in below fig,

1. An image is split into an S x S grid, and within each of the grids m bounding boxes are taken.
2. For each of the bounding boxes generated, a class probability and confidence scores values for the bounding box are given as output.
3. These confidence scores show how confident the model is that the box contains an object.
4. Each of the bounding boxes generated consists of 5 predictions: x, y, w, h, and confidence. The (x, y) coordinates represent the center of the box relative to the grid cell boundaries. Relative to the whole image, the width and height are predicted.
5. Only those bounding boxes that have the class probability greater than or equal to the set threshold value are selected. These bounding boxes are used to locate the object in the image/frame.

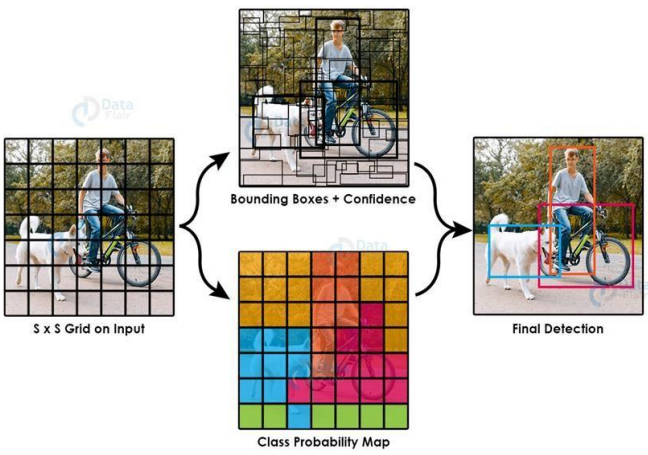


Fig-6: Working Mechanism of Yolo

6.8 Open CV:

Open CV is the huge open-source library for computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. Using it, one can process images and videos to identify objects, faces, or even the handwriting of humans [10]. When integrated with various libraries such as Numpy, python is capable of processing the OpenCV array structure for analysis. To identify image patterns and their various features we use vector space and perform mathematical operations on these features. We can see a lot of applications that are solved using Open cv, some of them are listed below 1. Face recognition 2. Automated inspection and surveillance 3. Vehicle counting on highways along with their speeds etc.

6.9 Text-to-speech-conversion:

Voice synthesis, characterized as TTS (an abbreviation for Text-To-Speech), may be a computing system that ought to be able to read aloud any text, irrespective of its origin. The use of TTS aims to supply human voices artificially. Voice synthesis is a complex process and complex algorithms are expected to deliver an understandable and regular outcome. TTS synthesis makes use of techniques of tongue Processing. Since the text to be synthesized is the first entry of the system, it must be the primary to be processed. Through open cv, we can convert images to text, and also the converted text is given as input to the TTS API which successively gives the voice commands of the recognized objects. we've used the speak3 engine for the conversion.

6.10 Optical Character Recognition (OCR):

Optical Character Recognition (OCR) is a technique of reading text from printed or scanned photos, and handwritten images and converting them into a digital format that can be editable and searchable. Pytesseract is an open cv tool for optical character recognition. We have installed the tesseract package in the open cv to read the text.

7. WORKING MECHANISMS

The process starts with capturing images through the camera in real-time. In other words, a live video or an image is captured and then converted into individual

frames are analyzed by a raspberry pi. For all this to happen we have to set up the raspberry pi. Initially, we formatted the sd card and then the os is installed on the card. The object detection program was written in a file and run in optimized Python IDE, Thonny. The "OCR" program was also integrated into it. For all this, we have imported several packages such as NumPy, cv2, os, etc. Pi can be accessed on the laptop using Putty, Xming Server, and Windows Remote Control. But instead of accessing in laptop and running the respective file which will be a tedious task for the blind, we came up with the idea of boot start which was enabled through a button. With a single press of a button, object detection code will be executed and it will detect the objects and give the voice commands that could be probably possible throughImage-to-Text and Text-to-Voice [9]. If the blind people want to recognize any text we can use OCR which can be executed with the long press of a button for more than five seconds.

For the task of object detection, the Yolo algorithm is used. The objects which are in a frame are captured and are passed through the layers of neural networks on which the Yolo algorithm works. No of the frames that were captured per second can be varied according to the model and the processing time of the frame to get results also depends on raspberry pi capacity and the load on it. As we have taken the Raspberry pi 3+ model the result can be obtained in 45 seconds approx. The captured images are transferred from one layer to another in which regions of interest from the frame are filtered, where the object is present.

For the task of object detection, the Yolo algorithm is used. The objects which are in a frame are captured and are passed through the layers of neural networks on which the Yolo algorithm works. No of the frames that were captured per second can be varied according to the model and the processing time of the frame to get results also depends on raspberry pi capacity and the load on it. As we have taken the Raspberry pi 3+ model the result can be obtained in 45 seconds approx. The captured images are transferred from one layer to another in which regions of interest from the frame are filtered, where the object is present. When the object is acquired, it is contrasted and the COCO informational collection. Along with the identification, we can obtain the distance of the object through which the blind can get alerted and take precautions in case of dangers[4]. We can even set an ideal distance limit, below which if any object is detected we can intimate the person about the obstacle. even the accuracy of the detection is also mentioned which reflects how much accuracy the object has been identified. The next important part of the system is the hearing mechanism without which the entire system would not be useful for the user. So the detections with their accuracy and distance can be informed through user headphones where the user will be more aware of the obstructions in front of them.

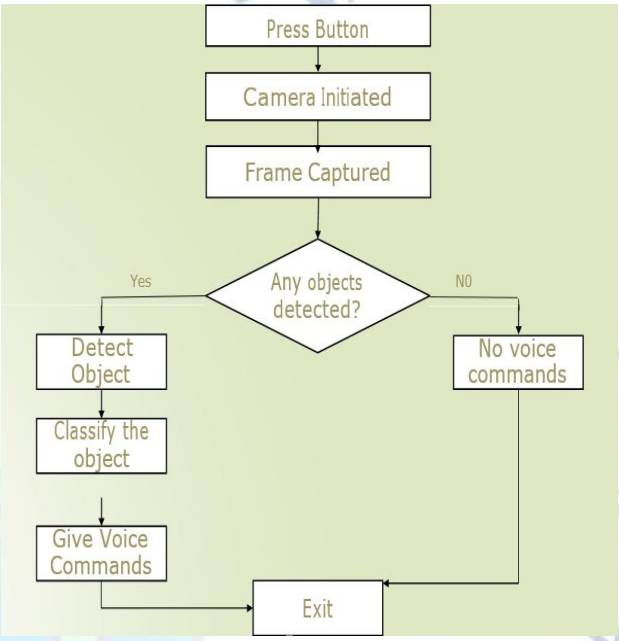


Fig-7 The working flowchart of the device is as follows:

Results:

We have taken many trails where we could capture and detect various objects along with their precision. through hardware as well as in the system by connecting the Raspberry pi to the pc and we have obtained the resultsbelow.

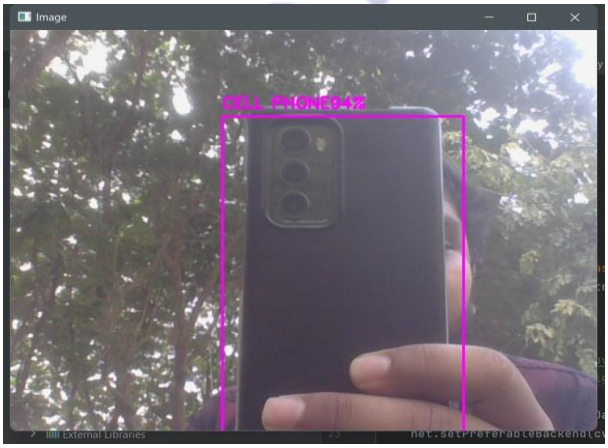


Fig-8The output of the YOLO algorithm for Frame-1.

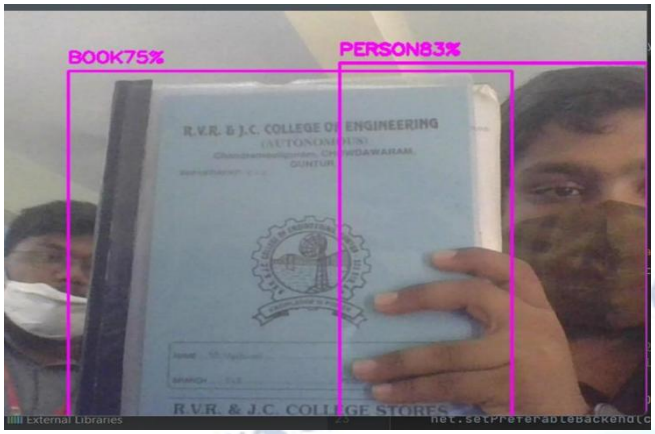


Fig-9The output of Frame-2

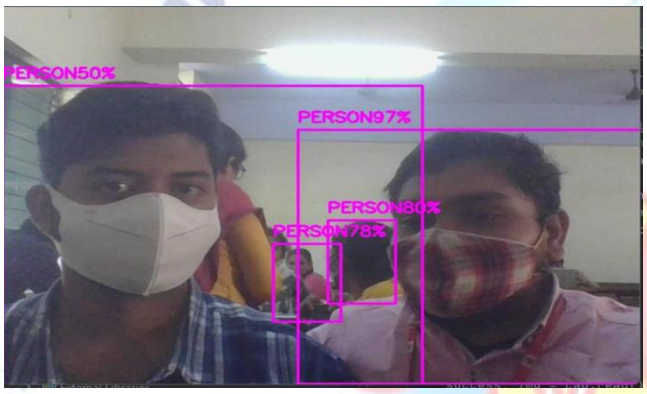


Fig-10 The output of Frame-3

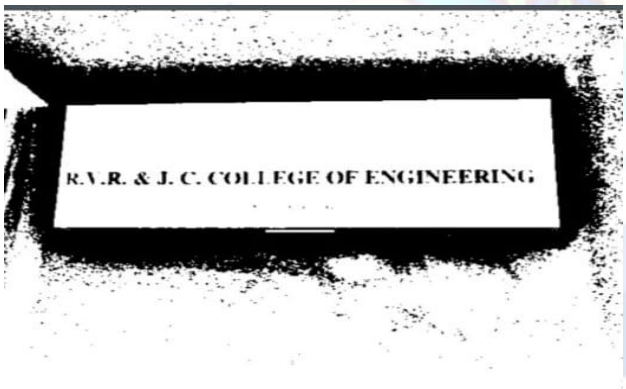


Fig-11 The result of OCR is as follows:

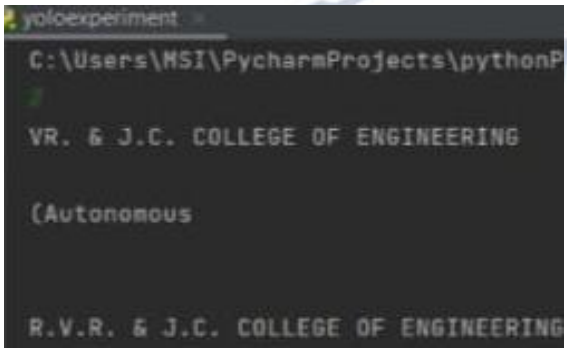


Fig -12 OCR results in the terminal:

8. Conclusion:

The purpose of this project is to guide the visually impaired about the obstacles. In this work, we have used OpenCV for capturing the images and Object detection respectively. Through YOLO V3 and Darknet the objects along with their position and precision are identified. The work likewise centers around the age of the outcomes in the least time conceivable which lessens the result dormancy. The Raspberry pi with its parallel processing capability is useful for this purpose. We aim to create a device that will provide minimal assistance through artificial eyes to visually impaired people. As a result, this is an aimed combination that collectively makes assistance a user and makes their commutation safe [22]. Adding further we can add various features like night vision or a safe path for overcoming the obstacles in front of them, Hand object detection. Even we can create custom datasets for the detection of more objects, other than the objects specified in the coco dataset. Also, we can use GPU or Jetson Nano for better efficiency. The designed system is a low-cost device that will reduce the overall cost of the system and make the system optimized and compact. This will make it easy and handy for visually impaired people to carry anywhere.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

References

[1] C. Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, "Object Detection Based on YOLO Network", IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China,2018.

[2] M. H. Putra, Z. M. Yussof, K. C. Lim, S. I. Salim "Convolutional Neural Network for Person and Car Detection using YOLO Framework", Journal of Telecommunication, Electronic and Computer Engineering,2014.

[3] S. Guennouni, A. Ahaitouf, and A. Mansouri, "Multiple object detection using OpenCV on an embedded platform", Third IEEE International Colloquium in Information Science and Technology (CIST), Tetouan, 2014.

[4] Z. Chen, R. Khemmar, B. Decoux, A. Atahouet and J. Ertaud, "Real-Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility", Eighth International Conference on Emerging Security Technologies (EST), Colchester, United Kingdom, 2019

[5] Zhihuan Wu, Xiangning Chen, Y Gao, Y Li, "Rapid target detection in high-resolution remote sensing images using Yolo model", International Archives of the Photogrammetry (ISPRS), Remote Sensing and Spatial Information Sciences, Volume XLII-3, 2018.

[6] Esteban Reyes, Cristopher Gomez, Esteban Norambuena, Javier Ruizdel-Solar, "Near Real-Time Object Recognition for Pepperbased on Deep Neural", Computer Science book series (LNCS, volume 11374) ,2018.

[7] Object Detector for Visually Impaired with Distance Calculation for Humans EligantiRamalakshmi, Dixitha Kasturi, Gouthami V International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020.

[8] Electronic Orientation Aid for Visually Impaired using Graphics Processing Unit (GPU) Akshata Parab1,1, Rashmi Nagare 2, Omkar Kolambekar3, and Parag Patil.

[9] Asha G. Hagargund, SharshaVanria Thota, MitadruBera, Eram Fatima Shaik, "Image to Speech Conversion for Visually Impaired", International Journal of Latest Research in Engineering and Technology (IJLRET) ISSN: 2454-5031.

[10] S. Guennouni, A. Ahaitouf, and A. Mansouri, "Multiple object detection using OpenCV on an embedded platform", Third IEEE International Colloquium in Information Science and Technology (CIST), Tetouan, 2014.

[11] P. Strumillo, "Electronic interfaces aiding the visually impaired in environmental access, mobility and navigation", 3rd International Conference on Human System Interaction, Rzeszow, 2010.

[12] . S. Hayat, S. Kun, Z. Tengtao, Y. Yu, T. Tu and Y. Du, "A Deep Learning Framework Using Convolutional Neural Network for Multiclass Object Recognition", IEEE 3rd International Conference on Image, Vision, and Computing (ICIVC), Chongqing, 2018.

[13] M Naveen Kumar, K Usha, "Voice-based guidance and location indication system for the blind using GSM, GPS, and optical device indicator", International Journal of Engineering Trends and Technology (IJETT), 2013

[14] "You Only Look Once: Unified, Real-Time Object Detection" published in CVPR 2016. This paper is "You Only Look Once: Unified, RealTime Object Detection" published in CVPR2016.