NAYAN-DRISHTI: A Revolutionary Navigation/Visual Aid for the Visually Impaired.

Salil Fernandes [1], Jordan D'souza [1], Anthony Kattikaren [1], Mrs. Dipti Jadhav [1]

Don Bosco Institute of Technology, Mumbai-70.

salilfernandes72@gmail.com, jordandsouza99@hotmail.com, anthonyj3199@gmail.com, dipti.dbit@dbclmumbai.org

Abstract: The project/proposed product hinges on three domains of computational technology i.e., Machine learning, Convolutional Neural Networks and Internet of things. The aim of the project is to invent a product that is helpful to the disabled section of society as ideally as possible try to as well as to acquaint ourselves with the much talked about and ever growing domains of computer technology. The main functions that our proposed product will offer are detection of the obstructing object and alerting via a speaker (along with classification and distance of the object from the user) and a navigation system (which obtains live data of the current location of the user with the help of the UBLOX GPS module). The proposed product is designed in such a way so as to provide an all in one multitasking and hassle-free solution to our user and to ease the burden that come along with the disability of blindness. The proposed product is touted as a boon to our users since it not only will help them in identifying the obstructions ahead them but will also help them to navigate from their current location to their destination with freedom and no fear.

Keywords: Convolutional Neural Networks (CNN), Rectified Linear Unit (ReLU), Batch Normalization (BN).

1. Introduction:

We sometimes wonder how the disabled section our society manage to perform tasks that look impossible pertaining to their disability. We have to ponder as to how efficiently can they perform their day-to-day tasks and how well can they communicate with normal people. Keeping in mind the difficulties they face, their ability to perform such tasks and how to ease the fear related to it, technology has advanced in many fields.
Blindness can basically be classified into three types:
Complete Blindness, Night Blindness and Colour Blindness.

The proposed product will function accordingly as explained below using the following computing domains:
1. Convolutional Neural Networks: The product will take an input from the surrounding environment of the user and convolve it for a better prediction of the obstructing object which will be doneusing machine learning.
2. Machine Learning: The product with the output image can will predict the object that is in hinderance to the user's path with the help of pre-trained datasets.
3. Global Positioning System: The user will be able to communicate his/her destination to the software and the software will guide the user to the destination. The UBLOX GPS Module plays an important role in communicating the satellite in space to obtain the real time location of the user.
4. IOT (Internet of Things): Our project consists of raspberry pi, GPS module, camera module, etc. that will be effectively used in processing data, capturing images, displaying location of the person, etc.

2. Literature Survey:

A. Survey of Existing Systems:

"Electronic walking stick for the blind" published in the year 2018, the use of optical sensors hasbeen highlighted. This is a modern concept of a walkingstick which is completely digital. These sensors essentially convert light reflected from any surface and convert them into electric signal and acts as response to the stimuli and informs the blind person via a speaker onthe handle about the obstacle. The object aimed to give voice assistance to the user and was able to deliver in various cases.[1]
"Portable Blind aid Device" published in the year 2019, it highlights the use of a mobile-based project in which the user can switch his wireless device into blind assistance mode with thehelp of a button. With the help of the camera, GPS and a cloud-based architecture it will be able to give the real-time location of the user and also make himaware of his surroundings. Also plan to include an advanced image recognition algorithm to recognize the faces of stranger and to store them.[2]
"Intelligent glasses for the blind" published in the year 2016, the device is smart glass which uses a camera, an ultrasonic sensor

and an electronic touchpad to assist the blind person. Using a mobile device one can activate the glass and the camera will capture and convert the 3D image into a spatial matrix and give appropriate outputs. The touchpad giveslight electric shocks to make the user aware of the obstacles. Future scope was to add a walking cane with a button to give the user audio output of the real-time surroundings.[3]

"Object Identification for Visually Impaired." published in the year 2016, a simple image recognition system that makes use of camera to recognize images, an ultrasonic sensor to detect obstacles is explained. The camera captures the images and if the image can be recognized by the images in the dataset then the device gives and audio output using a speaker that will be attached to the users clothing. Future upgrades include to introduce face recognition and to use a wireless camera.[4]

"Real-Time Objects Recognition Approach for Assisting Blind People." Published in the year 2017, an object recognition project is depicted that with the help of SURF (Speeded Up Robust Features) and light machine learning is able to give accurate information about the objects captured. Using GPS and image recognition it gives 90% accurate results. It makes use of a database and uses machine learning algorithm it is able to identify the objects.[5]

"EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobile and Public Scenarios." published in the year 2019, a Blind aid device which uses ear gestures like long swipe, tap, slide, etc. for operating the various function on the smartphone. Making use of one-hand gestures and Ear- Touch feature it enables the user to use the device with ease. And it gives out audio output through the inbuilt speakers. The device was able to read the gestures accurately and activate the respective functions when theuser wanted it.[6]

"An Ultrasonic Navigation System for Blind People" published in the year 2017, introduces with us another device for object detection which not only uses an ultrasonic sensor but also includes an accelerometer, footswitch, microcontroller and vibration pad. User is made aware of the obstacle and closeness ofit by different degree of vibrations and also through an audio output.[7]

"Mobile Blind Navigation System UsingRFID". published in the year 2015, the paper introducesa mobile based communication device which makes useof the WIFI, GPS to determine the user's location. And also makes use of RFID tags in the surroundings to givemore accurate location to the user and guide them by giving appropriate instructions. The user makes use of asmart walking cane and RFID handled reader.[8]

"Voice assisted system for the blind" published in the year 2015, the paper introduces us witha simple object/obstacle detector which makes use of anultrasonic senor, microcontroller, mp3 module and SD card. The sensor sends the distance measured and the microcontroller depending the proximity limit it is programmed sends an output to the mp3 module. It makesuse of depth sensing for avoidance of pot holes etc. for the user to walk with smoothness and comfort with normalcy.[9]

"Virtual-Blind-Road Following Based Wearable Navigation Device for Blind People." published in the year 2019, the paper depicts a navigation device specifically for the indoor environment which makes use of SLAM (simultaneous localization and mapping). The device tries to give the person the best path to an indoor destination by keeping track of the positions of previously encountered obstacles. It makes use of PoI graph to store new points in the route to the destination and using A* algorithm onthe PoI graph it finds the shortest and optimum path to the destination. Obstacle distance is measured by an Ultrasonic range finder and provides audio output.[10]

B. Limitations of the systems on research gap:

All above proposed papers that we have researched haveshown 90% success rate in the field and in all test cases. Although with certain setbacks too. All devices are capable of giving the required results but the reactions time in real-world is less which the devices are not able to keep up with. For real-time outputs more complex andadvanced system and cloud-based architecture is required which is difficult on a limited budget. In image recognition few projects weren't able to identify objects in the dark or night. The components used in few projects are fragile and prone to damage. Hence when used by a user could get damaged. And using strong equipment makes the project expensive.

C. Problem Statement and Objective:

Our product introduces a three-way combined solution system that will serve as a navigation aid for the blind:

1. To ease the difficulty and uncertainty faced by the blind or any visually impaired person when they have to walk from one place to another. A simple task like walking is scary for blind people owing to the fact that they just do not know what obstacles, whether dangerous or not,may be in front of them.
2. Through our project we hope to create a product that makes navigation a safer and simpler task for blind people. Object detection and recognition is the core concept which our projectrevolves around.
3. Our project domain is machine learning and wehave decided to make use of the concept of a Convolutional Neural Network built using TensorFlow to process image data and performthe object recognition task after being trained with datasets.

D. Scope:

Our project aims to implement three mainfeatures:

1) Object recognition using a camera attached to the pi module and using our machine learning domain and implementing Convolutional Neural Network (CNN) break down the images into 2D frames andtrain the algorithm to identify various objects in the surrounding.
2) Object Detection using an Ultrasonic sensorwhich can detect obstacles at a distance of 4m and sends an alert to the user if the proximity of the object is within a certain given range.
3) Giving the real-time location of the user using a GPS module that is programmed tocommunicate with the satellite and on a press of a button give the user his/her location.

3. Models and Setups:

For object detection we have used a convolutional neuralnetwork to perform object recognition on a given image. With respect to sensing the distance of a user from a given obstacle, we have utilized the concept of ultrasonicdistance sensing.

A. Object Detection and Recognition:

The primary goal of our project is to perform image analysis to lead to object recognition which would enable a blind person to get a sense of his/her environment. By understanding what objects surround them, blind people can make better decisions about where they are and in what direction they should move. We have used the convolutional neural network (CNN) model to implement object detection and recognition. Convolutional Neural networks are a family of deep learning algorithms that take parts of images as inputs, apply certain weights and biases and differentiate one part of the image from another. Most images now a daysare RGB images meaning that there are three channels tothe image. CNNs make it easier to understanding the image by capturing the spatial and temporal dependencies by using filters are different levels or depths. It offers faster analysis by reducing the number of parameters and reusing weights.

Under CNNs there are several architectures that come into the picture like the Faster R-CNN model, Mask R- CNN Inception model and the SSD ResNet models to mention but a few. In our project we are using a Raspberry Pi 3 Model B+ to program and execute our programs. Due to limited memory of 1GB RAM and lower processing power available we have to make use of a neural network which is computationally less intensive but without making a compromise in accuracy.The architecture which suited this purpose is the SSD- MobileNet v2 architecture.

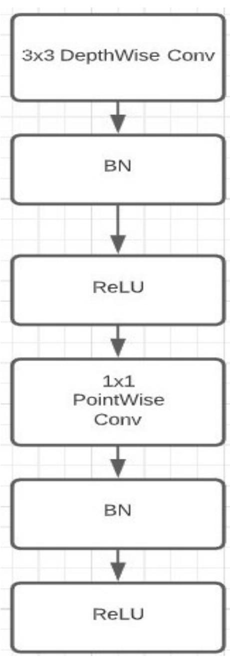


Fig 1. Single convolution layer in MobileNet.

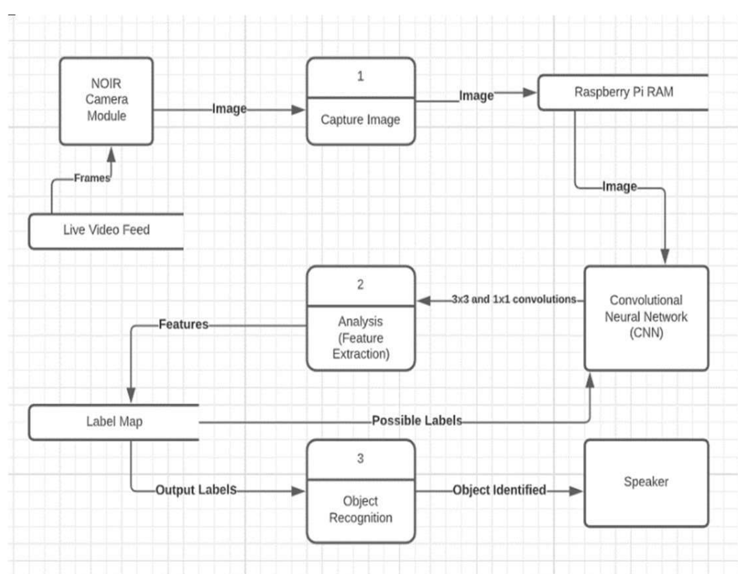


Fig 2. Data Flow diagram for Object Recognition.

The architecture uses the concept of depth-wise convolutions. Depth-wise Convolution is a type ofconvolution wherein a single convolutional filter isapplied to each input channel. In the regular
2D convolution performed over multiple input channels,the filter is as deep as the input and lets us freely mix channels to generate each element in the output. In contrast, depth-wise convolutions keep each channel separated from each other. In general, the steps used to perform a depth-wise convolution are:

1. Break up the input and filter into channels.
2. Convolve each input with the appropriate filter and combine the convolved outputs.

A depth-wise convolution consists of a depth separable convolution and a pointwise convolution. Spatial separable convolution works mainly with the spatial dimensions of an image and kernel which are the widthand height. (The third dimension is called depth which isthe number of channels of each image and is not taken into account by spatial separable convolutions). A spatial separable convolution breaks a kernel into twosmaller kernels. The most common case would be to divide a 3x3 kernel into a 3x1 and 1x3 kernel. In place ofperforming a single convolution by 9 multiplications, weperform two convolutions with 3 multiplications each to get the same result. With fewer multiplications, computational complexity goes down, and the network isable to run more efficiently. Unlike spatial separable convolutions, depth-wise separable convolutions use kernels that are not divided into two smaller kernels. The depth-wise separable convolution is so named because in addition to spatial dimensions, it deals with the depth dimension — the number of channels — as well. An input image can have3 channels: RGB. After a few convolutions, an image may have multiple channels. An image with 64 channelswould have 64 different versions of the same image. Analogous to spatial separable convolution, a depth-wiseseparable convolution divides the kernel into two kernelsthat do two convolutions: a depth-wise convolution and apointwise convolution.

B. Dataset Used:
The COCO Dataset has been used to train our MobileNetCNN for object detection and recognition. COCO in short is a dataset used for training object detection, segmentation and captioning networks. COCO stands forCommon Objects in Context which means that images are taken from everyday objects to prepare the dataset. It has approximately 330,000 images with more that 200,000 of them labelled.

C. Activation Function Used:
For each convolutional layer of the architecture, the ReLU activation function has been used along with BatchNormalization, the rectified linear activation function also called ReLU is the most commonly used activationfunctions in artificial neural networks. It returns a 0 as output for any negative input. If it receives a positive value x as input it returns x as the output. So in generalthis function is given as an equation $f(x) = \max(0, x)$.
Batch Normalization is a deep learning technique that normalizes the output of each sublayer. It allows for fastprocessing and deep network training by reducing internal covariate shift.

D. Obstacle Distance Sensing:
For the purpose of sensing the presence of an obstacleand sensing the distance between the obstacle and person, an ultrasonic sensor is being used. The input for this would in fact be an object placed in front of the sensor. The output is a distance reading if theobject is in the sensor's accurate range. By means of a wireless Bluetooth speaker, an audio warning will be sounded stating that the obstacle is nearby.

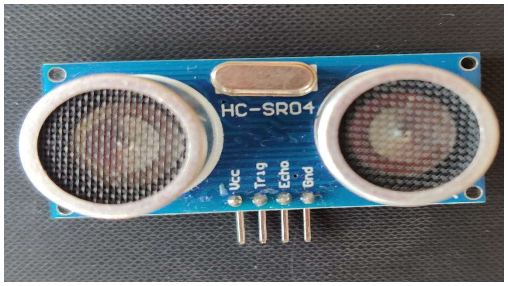


Fig 3. Ultrasonic Sensor.

The ultrasonic sensor has 4 pins:

a. VCC which is used to take the power supply.
b. TRIG which will on receiving the input signalfrom the Raspberry Pi will send out the ultrasonic pulse.

c. ECHO which will gather the echo pulse from theobstacle and send the signal that the obstacle hasbeen detected back to the Raspberry Pi.
d. GND which provides general grounding to the sensor.

For exchanging signals and for the program to receive readings from the sensor the GPIO pins of the RaspberryPi were connected with the pins of the sensor. We have used the HC-SR04 Ultrasonic Sensor to perform this functionality. The accurate range of this sensor is 2cm to 400cm which means that it can correctly measure distance between a user and obstacle when the obstacleis no less than 2cm and no more than 00cm away fromthe sensor. Extra caution had to be taken when implementing the circuit between the sensor and GPIO pins. The ultrasonic sensor operates at 5V and this voltage was supplied by connecting GPIO pin 1 to the VCC pin of the sensor. Since all the GPIO pins (except for pin 1) operate at 3.3V and the signals sent by the sensor at 5V, care has tobe taken to lower the voltage. To do this we made use ofa voltage divider circuit implemented on a breadboard. We used 560 ohm and 1000 ohm resistors to lower the 5V echo signal to a 3.3V signal.

E. Location Sensing:

In order to allow the blind person to know where exactly he/she is when navigating outdoors, a GPS sensor has been used. The Ublox Neo 6M GPS Module allows for easy interfacing with the Raspberry Pi.

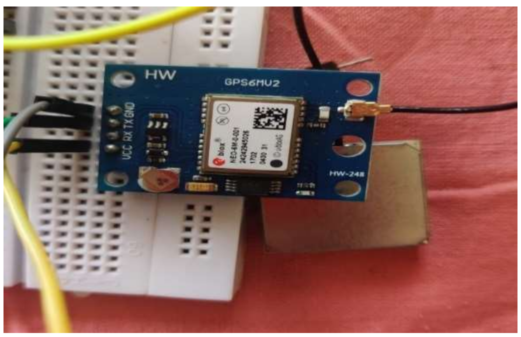


Fig 4. GPS Module.

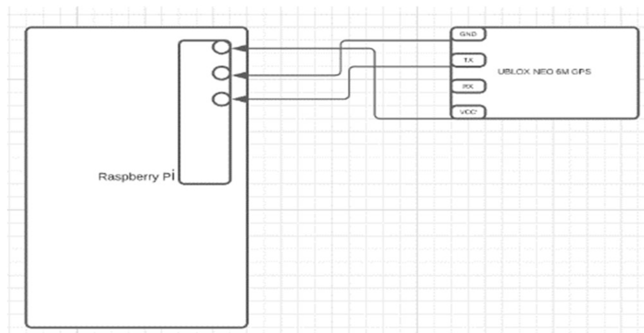


Fig 5. Diagrammatic location sensing setup.

The antenna once attached to the sensor allows for GPS data to be received. The antenna has to be left exposed to the sky in order for a satellite to lock onto it and start transmitting GPS data. There are 3 pins that are using to take the received data from the module:
a. GND: this pin is connected to pin 3 of the Pi for providing electrical grounding.
b. VCC: this pin is connected to pin 1 of the Pi which provides a 5V power supply.
c. TX: the transceiver pin of the module is connected to the RX pin of the Pi which is pin 5. RX is the receiver pin of the Pi which will transfer received GPS data to the serial port to be read by the Python program.

In python the GeoPy, PyNMEA2 and GeoPandas libraries have been used. The GPS Module receives location data in the form of NMEA sentences. In the Python code, the GPGLL (Geographic position, latitude / longitude) sentence has been extracted and used to obtain latitude and longitude coordinates of the module's current location. Reverse GeoCoding i.e. coordinates to address has been performed by the program to obtain the address. As usual this location address is sounded to the blind person by means of the Bluetooth speaker.

F. Wireless Sound:

The wireless Bluetooth chip that comes with the Raspberry Pi has been used to provide audio messages. A Bluetooth speaker is connected to the Pi with the help of this chip. For the purpose of making use of Bluetooth several packages had to be installed first namely: Bluez, Alsa, Bluetooth Manager and PulseAUDIO.
Since our device aims to help blind people navigate audio messages have to be delivered. This is the primary reason for using a wireless speaker. It is possible for blind people to get visual cues about their environment, hence audio hints have to be delivered to inform them.
In the object detection and recognition feature, once an object is successfully detected, the speaker delivers a message for eg. "Cup Detected".
In the obstacle sensing feature, whenever the ultrasonic sensor senses the obstacles distance to be less than 10 cm a message is

sounded, eg. “Careful Obstacle is 10 cm away”
In the location sensing feature, the speaker sounds the address of the current location obtained by reverse geocoding the coordinates.

4. Mathematical Model:

A. Object Detection and Recognition:

In each convolutional layer, the first layer is a pointwiseconvolution followed by ReLU. The second layer is a depth-wise convolution and the third is another pointwise convolution but without any non-linearity.

Input	Operator	Output
H x W x K	1x1 conv, ReLU	H x W x TK
H x W x TK	3x3 conv S=s,ReLU	$\frac{H}{s} \times \frac{W}{s} \times TK$
$\frac{H}{s} \times \frac{W}{s} \times TK$	1x1 conv	$\frac{H}{s} \times \frac{W}{s} \times K'$

T is called the expansion factor, s is called the stride, Hand W are the height and width of the image respectively. For our MobileNet, the depth-wise convolution applies asingle filter to each input channel. The pointwise convolution then applies a 1X1 convolution to produce alinear combination of the outputs the depth-wise convolution. All layers in MobileNet consist of a 3X3 depth-wise separable convolution except for the first layer which has a full convolution. A final average pooling reduces the spatial resolution to 1 before the fully connected layer. Counting depth-wise and pointwise convolutions as separate layers, MobileNet has 28 layers.

B. Ultrasonic Distance Sensing:

We have assumed the speed of sound to be approximately 340 metres per second or in other words 34,000 cm per second. To calculate distance between user and obstacle, only the distance one way has to be calculated since the time interval being measured is from the instant the ultrasonic pulse is sent to the instant theecho is received.

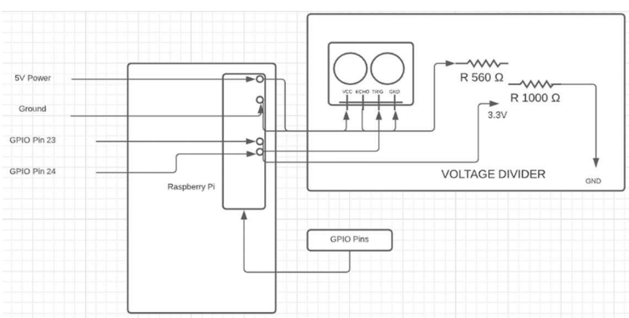


Fig 6. Diagrammatic ultrasonic sensing setup.

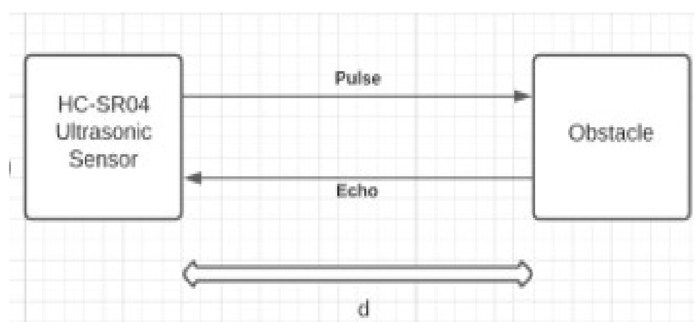


Fig 7. Distance Calculation.

Let t be this time interval. Therefore, the formula for distance is,

$$d = (t * 34000) / 2 \quad (1)$$

5. Results:

A. Object Detection and Recognition:

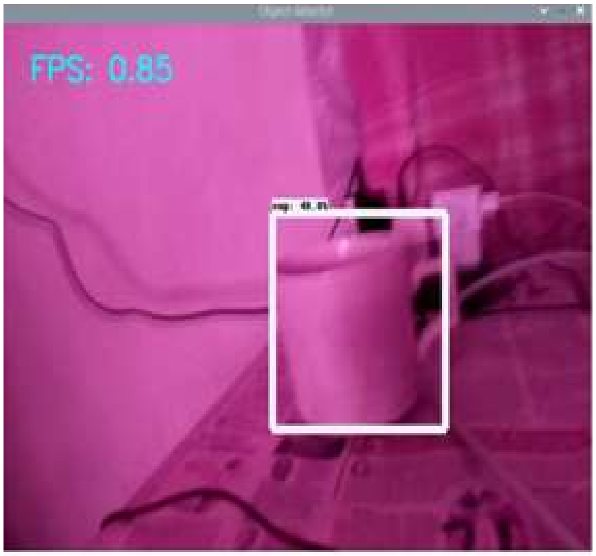
The frames per second (FPS) can be improved by manually changing the resolution of the detector. For objects that the model was trainedfor, the detections and recognitions were instantand highly accurate which is a very good result considering that the program was executed on the Raspberry Pi with only 1GB of memory.

B. Obstacle Distance Sensing:

In the program we preset the range for which analert message is to be given. Whenever the distance sensed was less than 10 cm an alert is generated. The range can be manually changed in the program.

C. Live Location Sensing:

This program enables the GPS module to acquire the signal from the orbital satellite and latch onto it. After the latching of the signal, it detects the location of the user and informs the user about the current location.



Figs 8, 9. Object Recognition output.

6. Future Scope:

To get more accurate obstacle/distance sensing results, a sonar sensor can replace the ultrasonic sensor as it has a larger sensing range. Whenever the Raspberry Pi foundation manages to integrate more RAM onto the computer board, the neural network can be trained a lot faster and give speedy and accurate predictions. If our prototype is manufactured and improved on by a professional company, the resulting product would change the lives of visually impaired people forever. The functionalities of the proposed system can also be increased from the proposed three to as much as we can in order to benefit the user even better. One such functionality can be a guided software to use the smartphone according to the user's comfort.

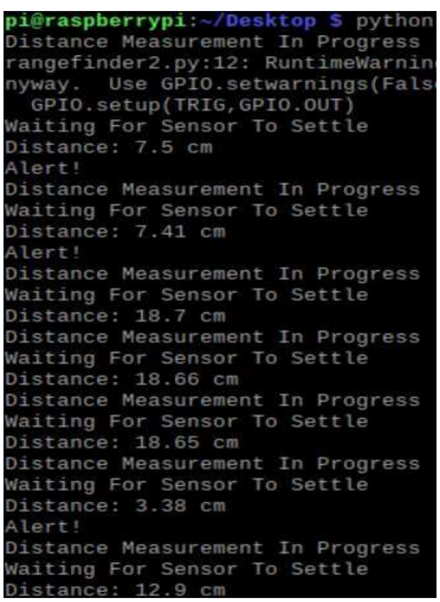


Fig 10. Ultrasonic Distance output depicting an alert for distances less than or equal to 10 cm.

Fig 11. Live Location Detection.

7. Conclusions:

Through the results depicted in the above section, it is clear that our object detection was overall a success in terms of efficiency and accuracy. The system could differentiate between basic objects easily with an efficiency rate ranging between 75-98% depending on the rate of frames per second. On testing the object distance measurement, we found out that the system was again efficient in displaying/sounding an alert as the object was nearing the user. An alert message will be sounded when the user is around 3-5m from the obstructing object. But here for test purposes we chose 10 cm as our minimal limit to sound an alert message. Thus,object detection and object recognition were successfully implemented.

8. References and Citations:

[1] Chi-Sheng, Hsieh. "Electronic walking stick for theblind." U.S. Patent No. 5,097,856, 24 Mar. 2010.

[2] Evanitsky, Eugene. "Portable blind aid device." U.S.Patent No. 8,606,316, 10 Dec. 2013.

[3] Cervantes, Humberto Orozco. "Intelligent glasses forthe visually impaired." U.S. Patent No. 9,488,833. 8 Nov. 2016.

[4] Jothimani, A., Shirly Edward, and G. K. Divyashree."Object Identification for Visually Impaired." Indian Journal of Science and Technology 9.S1, 2016.

[5] Zraqou, Jamal S., Wissam M. Alkhadour, and Mohammad Z. Siam. "Real-Time Objects RecognitionApproach for Assisting Blind People.", 2017.

[6] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, Yuanchun Shi. 2019. EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobileand Public Scenarios. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019),May 4–9, 2019, Glasgow, Scotland Uk. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300254

[7] Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 3247– 3256. https://doi.org/10.1145/1978942.1979424 Mounir Bousbia-Salah, Abdelghani Redjati, Mohamed Fezari, Maamar Bettayeb. "An Ultrasonic Navigation System For Blind People".2007. https://www.researchgate.net/publication/251851635

[8] Rachid Sammouda, Ahmad AlRjoub. "MobileBlind Navigation System Using RFID".2015. https://ieeeexplore.ieee.org/document/7353325

[9] Ananth Noorithaya, M. Kishore Kumar, A. Sridevi."Voice assisted system for the blind", 2015. https://doi.org/10.1109/CIMCA.2014.7057785

[10] Jinqiang Bai, Shiguo Lian, Zhaoxiang Liu, Kai Wang, Dijun Liu "Virtual-Blind-Road Following Based Wearable Navigation Device for Blind People", 2018. https://ieeexplore.ieee.org/document/8307352

[11] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[12] Abadi, Martn, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems."arXiv preprint arXiv:1603.04467 (2016).

[13] Ren, Shaoqing, et al. "Object detection networks onconvolutional feature maps." IEEE transactions on pattern analysis and machine intelligence 39.7 (2017): 1476-1481.

[14] Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Enhancing Independence and Safetyfor Blind and Deaf-blind Public Transit Riders. In Proceedings of the SIGCHI.

[15] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems, 2015.