



# Real Time Indoor Object Detection Aid for Blind

Akhila Merin Binu<sup>1</sup>, Chinnu Kuruvilla<sup>2</sup>, Sony Thomas<sup>3</sup>, Preena Prasad<sup>4</sup>

<sup>1, 2, 3</sup>UG – Electronics and Communication Engineering, KTU University, Pathamuttom, Kerala

<sup>4</sup>Assistant Professor, Electronics and Communication Engineering, KTU University, Pathamuttom, Kerala

**Abstract :** Our proposal accentuates to help blind people to navigate independently using real-time indoor object detection and identification model. The project aims to help blind and visually impaired people to detect and recognize objects around them, through a small camera that is fixed on the glasses. The device setup gives an audio output that is sent to an earphone placed on the person's ear to help him/her find various items easily and independently. All the captured data will be fed to raspberry pi that will have an inbuilt code for object detection in python language and will contain Opencv and TensorFlow files in it. The unit will analyze the image based on the ssd\_mobilenet network and custom data-set and the processed data that is the name of the object will come as output. This output will be heard by the user in voice format by incorporating TTS in our model. This model will act as an aid for blind people and will help them to do the daily chores easily and independently.

**Keywords-** Tensorflow, Text to Speech converter, Raspberry pi, SSD\_ Mobilenet.

## 1. INTRODUCTION

Percentage of persons with disabilities in society, especially those with visual disabilities (blind) are estimated to be 1.5 % of the Indian population. As per the survey report, the maximum prevalence of blindness was seen in the age group of above 80 (11.6%), followed by 70-79 age group (4.1%), 60-69 age group (1.6%), and 50-59 age group (0.5%) [1]. The survey noted that most of the blindness (92.9%) and visual impairment (96.2%) cases were due to avoidable causes. These people face many difficulties in doing chores in day-to-day life. They are unable to find things and are also at a high risk of tripping over things and getting injured. The conventional methods for blind assistance like canes, electronic mobility aids[3], reading assistance[5], etc have limited features. From here the idea of our project began. The project of Blind assistance aims to promote a wide feature of computer vision that is recognition of objects of the surrounding area around the daily life of the blind. A customized home dataset with object detection of humans, cup, plate, chairs, doors, tables, bed, toothpaste, toothbrush, comb etc that are common in the daily scenes of a blind individual is trained and based on the locations of these objects in front of the blind, the camera captures the real-time image, and then processing the image takes place which detects the objects and then audio output is given via earphones.

## 2. LITERATURE SURVEY

Some of the existing and relevant works that were done in the field of aid for blind for visually impaired people are given in this section :

### **Real-Time Objects Recognition Approach for Assisting Blind People [2]**

In this paper, two cameras are placed on blind person's glasses along with it, GPS free service and ultra-sonic sensor are employed to provide the necessary information about the surrounding environment. A dataset of objects gathered from daily scenes is created to use the specified recognition. Objects detection is used to seek out objects within the real world from a picture of the world like faces, bicycles, chairs, doors, or tables that are common in the scenes of a blind. The two cameras are necessary to get the depth by creating the disparity map of the scene, GPS service is employed to make groups of objects based on their locations, and also the sensor is employed to detect any obstacle at a medium to long distance. The descriptor of the Speeded-Up Robust Features method is optimized to perform the recognition. The proposed method for the blind aims at expanding possibilities to people with vision loss to attain their full potential. The experimental results reveal the performance of the proposed work in a real-time system.

### **Voscal vision for visually impaired [3]**

This project is a vision substitute system designed to assist blind people with autonomous navigation. Its working concept is based on

'image to sound' conversion. The vision sensor captures the image in front of a blind user. This image is then fed to MATLAB for processing. Process intuit processes the captured image and enhances the significant vision data. This

processed image is then compared with the database kept in a microcontroller. The processed information is then presented as a structured form of the acoustic signal and it is conveyed to the blind user using a set of earphones. Color information from the interested objects evaluated to determine the color of the object. The color output is informed to the blind user through headphones.

### **Object Detection and Human Identification using Raspberry Pi [4]**

This paper aims to investigate the development of a navigation system for blind and visually impaired people along with human identification. It is based on a microprocessor with synthetic speech output. This solution is designed to give details about humans whomever they encounter. It is a self-contained portable electronic unit. The proposed system detects any obstacle in the path of the blind via a stereoscopic sonar system using ultrasonic waves and sends back vibrotactile feedback to inform the blind about its localization. And if there is an obstacle and then the second system kicks in to check whether the obstacle is human or not and if proven human, it matches with the existing system database to find out the personal details and provide them to the user. Face Detection is an important step in any face recognition system, to localize and extract the face region from the rest of the images. The proposed system has been developed and its main goal is to increase the capability of blind individuals. The technique used here is a well know name in imaging industries, where they reduce errors by using more advanced algorithms.

## **3. HARDWARE AND SOFTWARE COMPONENTS**

The proposed model is designed for object detection and its identification in real-time. The main aim behind the project is to able to help and guide the blind person about the objects in his surroundings and make him alert about his immediate surroundings. To achieve this goal we will be using both hardware and software components in our model.

### **3.1. HARDWARE COMPONENTS**

#### **3.1.1) Raspberry pi 4:**



**Fig.1: Raspberry Pi 4**

Raspberry pi is used as the interface that can be customized based on a user request whether that is vibrations, sounds, or the spoken word. The Raspberry Pi is a series of small single-board computers developed in the United Kingdom by the Raspberry Pi Foundation to push the teaching of basic computing in schools and colleges of developing countries. Raspberry pi 4 has 4GB SD RAM and an additional micro SD card slot thus providing more amount of storage for our custom dataset.



**Fig. 2: Raspberry pi camera module**



**Fig 2.1: Flex Cable**

The Raspberry Pi Camera Module is a custom-designed add-on for Raspberry Pi. It attaches to Raspberry Pi by way of one of the small sockets on the board's upper surface. This interface uses the dedicated CSI interface, designed especially for interfacing with cameras. The board itself is small, at around 25mm x 20mm x 9mm. It also weighs just over 3g, making it perfect for mobile or other applications where size and weight are important. It connects to Raspberry Pi by way of a short ribbon cable. The sensor itself features a native resolution of 5 megapixels and contains a fixed focus lens onboard. In terms of still images, the camera is capable of 2592 x 1944 pixel static images, and also supports 1080p30, 720p60 and 640x480p60/90 video. To interface Raspberry pi cam with Raspberry pi module, we will be using an additional long flex cable.

Board Features:



- Fully Compatible with Both the Model A, Model B and Model B+ Raspberry Pi
- 5MP Omni-vision 5647 Camera Module
- Still Picture Resolution: 2592 x 1944
- Video: Supports 1080p @ 30fps, 720p @ 60fps and 640x480p 60/90 Recording
- 15-pin MIPI Camera Serial Interface - Plugs Directly into the Raspberry Pi Board
- Size: 20 x 25 x 9mm
- Weight- 3g

### 3.1.3) Lithium ion battery



**Fig 3: 3.7V, 3000mAH (Lithium Polymer)**

Lipo Rechargeable Battery is also known as Lipo or Lipoly batteries are thin, light, and powerful. This battery has a capacity of 3000mAH. These Batteries are widely used in GPS, DVD, iPod, Tablet PC, MP4 Player, Power Bank, Mobile Backup Power Supply, Bluetooth Speaker, IoT, and other DIY and Industrial applications. The 3.7v 3000mAH provides power to the Raspberry Pi 4 module. Since the Raspberry Pi 4 used requires a 5.1v Supply here we use two 3.7v lithium-ion batteries. They are lightweight and compact. They don't suffer from the Memory Effect this makes a convenient option for our project. They have a lower self-discharge rate and able to recharge hundreds of times.

Specifications of 3.7V 3000mAH Lipo Battery Model KP-356070:- x 4mm

- Voltage:3.7V
- Capacity: 3000mAH
- Approx Size: 70mm x 60mm

## 3.2 SOFTWARE COMPONENTS

### 3.2.1) Tensor flow API

Tensor Flow is an open-source software library used in project software implementation for high-performance numerical computation.

-Object detection with Tensor Flow

Computations are done in two steps:

First: Build the model.

Second: Train and evaluate the model by doing object detection.

Both steps can be done in many languages (python, C++).In this project we use python.

- Object Recognition with Tensor Flow

A recognition algorithm (image classifier) takes an image as input and gives outputs as a label over the bounding box about the object name that the input image contains. In other words, the output is a class label which contains the information of the object ,that is to which class it belongs to.

Three Steps in Recognition: Preprocessing, Feature Extraction, Learning Algorithm for Classification.

### 3.2.2)Python

Python is the programming language used in Tensor Flow API. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. It is high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components. Python's simple, easy-to-learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

### 3.2.3) SSD-MobileNet

It is the pre-trained model we used in our project. It is a way for detecting objects in images employing a single deep neural network. Our approach uses:-

SSD[Single Shot Detector]:



The tasks of object localization and classification are done in a single forward pass of this network. MultiBox technique is used for bounding box regression. SSD network is an object detector that also classifies those detected objects. Detectors are convolutional filters, Each detector outputs a single value and discretizes the output space of bounding boxes into a group of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of varied sizes. Our SSD model is simply relative to methods that require object proposals because it eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to coach and easy to integrate into systems that need a detection component.

MobileNet:

It is the classification algorithm used. MobileNet uses depthwise separable convolutions. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets. This results in lightweight deep neural networks.

A depthwise separable convolution is formed from two operations.

1. Depthwise convolution
2. Pointwise convolution.

MobileNet is a class of CNN that was open-sourced by Google and thus, this provides us a superb start line for training our classifiers that are insanely small and insanely fast.

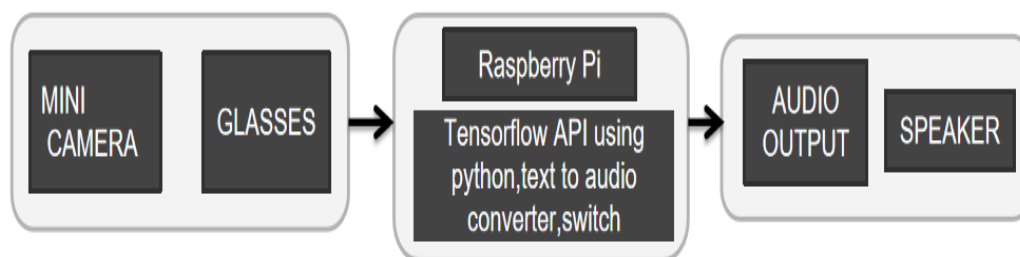
Feature extractor:

ssd\_mobilenet\_v1\_fpn\_keras is the feature extraction algorithm used in SSD MobileNet. Feature extraction refers to the portion of the training process by which a neural network learns to map input space to a latent space that can subsequently be used for classification via the final layer. Feature extraction includes several convolution layers followed by max-pooling and an activation function Relu\_6. The network consists of an input layer, followed by three convolutional and average pooling layers and followed by a softmax fully connected output layer to extract features.

Activation function:

In a neural network, the activation function is liable for transforming the summed weighted input from the node into the activation of the node or output for that input.

Rectified linear activation function or ReLU is the Activation Function used in our model. It is a piecewise linear function which will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for several varieties of neural networks because a model that uses it is easier to coach and often



achieves better performance.

The main advantage of using the ReLU function over other activation functions is that it doesn't activate all the neurons at a similar time. This means that the neurons will only be deactivated if the output of the linear transformation is a smaller amount than 0. For the negative input values, the result is zero, which means the neuron does not get activated. Since only a certain number of neurons are activated, the ReLU function is far more computationally efficient when compared to the sigmoid and tan activation functions.

#### 4. METHODOLOGY

**Fig 4: Block diagram of the Proposed Model**

Our project is based on tensor flow and text-to-speech synthesizer software. In our project we will be using a raspberry pi cam module that will be molded in the glasses of the user, this camera will capture the images of the user's indoor surroundings and send them to our microprocessor, raspberry pi in real-time. In raspberry pi, we will incorporate tensor flow object detection API along with datasets that will help him in customized indoor object detection. The image processing will take place inside the pi module and the identified objects will be given as output as voice via earphones. With a single object detection model, it is possible to classify multiple objects present within an image, and also it can

specify the exact position of the image (if monitor provided) with a bounding box framing the object. Real-time video is captured using the camera will be segmented into a set of frames using python command and digital image processing will be done on the image to identify the object. We will be using the simplest and fastest pre-trained object detection model [SSD\_Mobilenet] offered by Tensor Flow to detect various objects present within our image. In addition, we create a customized dataset model for better efficiency. The output is in form of a bounding box representing a part of the image where a particular object was detected with the object name as label. This will be displayed in form of text when the raspberry pi is interfaced with a display system but we will be using text to speech converter from where we extract text from the real image and this extracted text is given as the output in form of audio via earphones connected to raspberry pi's audio jack providing the voice description corresponding to the objects in the image.

### 5. SIMULATION AND RESULTS

We are using the simplest and fastest pre-trained object detection model SSD-MobilenetV2 FPNLite 320\*320, offered by Tensor Flow to detect various objects present within our image. To train our model we first labeled the objects in the captured image using the Labelling tool [An open platform tool for labeling]. We also incorporated transfer learning in our model. The activation function used in our model is RELU\_6, the classifier used is Mobilenet. The detector used is Single Shot Detector and the feature extractor is ssd\_mobilenet\_v2\_fpn\_keras. The labels assigned for our simulation model was human , cup, bed, cane, table, chair, comb ,door , toothpaste ,toothbrush ,thumbs up, thumbs down. The images collected using a camera were put inside a folder. The collected images were then divided into a train folder and a test folder. The images were resized to 320\*320 pixels for training. We then trained our model in 3 levels first of 1000 training steps, second 2000 steps, and third 3000 steps ,fourth 4000 and fifth 5000 training steps.

| Number of training steps | Per step time | Training loss |
|--------------------------|---------------|---------------|
| 1000                     | 9.246         | 0.986         |
| 2000                     | 2.113         | 0.845         |
| 3000                     | 2.159         | 0.911         |
| 4000                     | 2.115         | 0.623         |
| 5000                     | 2.139         | 0.956         |

Fig 5: Simulation result

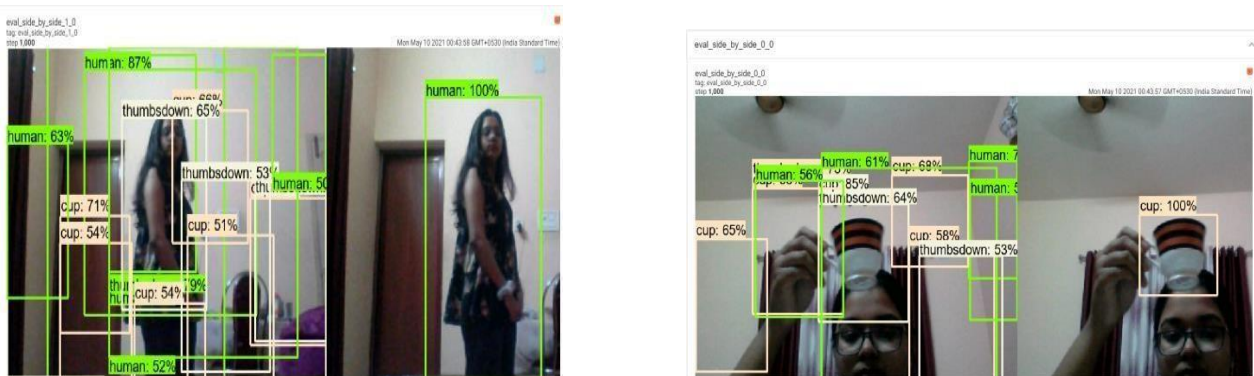


Fig 6: Object detected {here cup and human} from an image.

The results of the training and evaluation can be seen in figure5. In the table we could see the training loss for each 1000 steps (the step size was 100 steps). Apart from that we also got the evaluation loss that was the sum of localization loss, classification loss and regularization loss ,learning rate of the model and the average time required and the average recall rate and average precision rate for getting the bounding box at the correct aspect ratio. Both the training output and evaluation output were observed graphically using tensor flow graph tool- tensor board. From observing the tensor board graph and the above reading we got the result that for first simulation the learning rate was maximum and loss was minimum for steps around 3500 with 0.536 loss at 2.121 seconds and step 4900 with 0.568 loss in 2.118 seconds thus that was the ideal number of training steps. The model was then simulated to obtain output in real-time. The output displayed on the screen as an image containing a box, this box was around that part of the image where the labeled particular object was detected. This will be displayed in form of text when the raspberry pi is interface with a display system (as in fig 6).



But we will be using text to speech converter that will give us the output in form of audio via earphones. Headphones connected to raspberry pi's audio jack provide the voice description corresponding to the objects in the image. Thus this model aids in the commute of a blind person around his or her home and help them to do the basic chores and make them independent.

## 6. FUTURE SCOPE

For the future, the framework is ought to be extended to incorporate a bigger number of objects with a bigger dataset for the acknowledgment of outside and indoor objects. The framework can illuminate the visually impaired individuals about distinctive sorts of objects. Making it a 2 way communication system and adding extra features to make the system job efficient for blind. We can develop new feature types that could tell the user about the color of the object, distance of the object from the user and other features. The model can also be trained to recognize face of certain family members and friends thus decreasing the communication gap .We can also add features that will allow the family member to access the location of the user via using GPS and WIFI feature of raspberry module and find the location. We can also add features that could identify the amount of currency held in front of them by training our model to detect them thus preventing bad people with ulterior motive from robbing them. Overall our project could be modified to help and make the life of a blind person a little more easier.

## 7. CONCLUSION

Our projects aims at being an aid for a blind person and act as their eyes to the world where even though they cannot see but could identify the objects in their immediate environment. This project is a real-time object detection system that monitors the environment and provides audio information about the environment making the blind person's navigation indoors more safeand secure and helps them in their daily chores. The project has a simple architecture that transforms the visual information captured using a camera to voice information using Raspberry Pi in which object detection and recognition will take place in real time within a small fraction of time with minimum loss. This project incorporates computer vision, supervised learning, neural network and it's architecture object detection ,labeling ,recognition, bounding box regression, tensor flow API, open CV, TTS conversion and python language. With appropriate training and addition of more new complex features and data-sets according to the users need, we can turn this model to function as the artificial eyes of the blind user and make their life a little easier in this world, make them independent, increase their productivity and capacity and make them extremely fruitful.

## ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without acknowledging the people whose constant guidance and encouragement has crowned all the efforts with success. First of all, we thank God almighty, for lending us the talent to do this project and complete the work successfully. We take this opportunity to thank Dr. JOSEPHKUNJU PAUL.C, Principal, Saintgits College of Engineering, Kottayam, Kerala, for providing the best facilities and environment to prepare and present project. We are grateful for his support and co-operation. We also thank Dr. RIBOY CHERIYAN, Professor and Head of the Department, Electronics and Communication Engineering, for his advice and encouragement. We extend our special gratitude to our project guide Er. PREENA BASIL, Assistant Professor, Department of Electronics and Communication Engineering, for her wholehearted co-operation and support. We express our sincere gratitude to our project coordinator Er. ASWIN P V, Assistant Professor, Department of Electronics and Communication Engineering, for his helpful guidance, suggestions and encouragement in the completion of the project. Finally, we extend our special thanks to other staff members and to all our beloved friends for their timely help.

## REFERENCES

1. Zhigong Zhou ,Xiaosong Lan, Shuxioa Li, Chengfei Zhu, Hongxing Chang, "Feature Pyramid SSD: Outdoor Object Detection Algorithm for Blind People",in2019 IEEE 5th International Conference on Computer and Communications.
2. Jamal S. Zraqou Wissam M. Alkhadourand Mohammad Z.Siam,"Real-Time Objects Recognition Approach for Assisting Blind People ",Multimedia Systems Department, Electrical Engineering Department, Isra University, Amman-Jordan, Vol.7,No.1.
3. Shrilekha Banger , Preetam Narkhede , Rajashree Parajape, "Voscal vision for visually impaired", The International Journal Of Issn:2319– 1813 Isbn:2319–1805 Engineering And Science(Ijes)-01-07-2013
4. Dr.Sunitha M.R, Gowtham Ghatge R, Fathima Khan, Hemaya S Object Detection and Human Identification using Raspberry Pi [2019 1st International Conference on Advances in Information Technology]
5. Prakash T; Arun G; Arun Kumar B; Deepan S; KalyanaVenkataRamanan. "IoT Based Smart Irrigation for Multi Cropping System". International Research Journal on Advanced Science Hub, 2, 4, 2020, 7-12. doi: 10.47392/irjash.2020.20
6. Saraswathi S.; Aravindhan P.; Boovesh L.; SureshkumarV.. "IOT Based Future Technologies and Developments in Agriculture". International Research Journal on Advanced Science Hub, 2, 8, 2020, 131-135. doi: 10.47392/irjash.2020.107
7. Liming Wang, Jianbo Shi, Gang Song, and Ifan Shen ,"Object Detection Combining Recognition and Segmentation",Fudan University, Shan ghai, PRC,University of Pennsylvania.



8. Aniqua Nusrat Zereen and Sonia Corraya, "Detecting Real Time Object Along with the Moving Direction for Visually Impaired People", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE).
9. Tsung Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick "Microsoft COCO Common Objects in Context", 1 May 2014.
10. Prof. Radha Shankarmani, Sunit Vaidya, Niti Shah, Naisha Shah, "Real Time Object Detection for Visually Challenged People", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS2020) IEEE Xplore.
11. Ms. Revati Shriram, Samruddhi Deshpande, "Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People". 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) International Institute of Information Technology (IIT), Pune.
12. Rais Bastomi, Firza Puthra, Lucke Yuansyah Arif Tyras Putri, Septian Wahyu Saputra, Mohammed Rizki Maulana, Agus Khumaidi, Object Detection and Distance Estimation Tool for Blind People Using Convolutional Methods with Stereovision", 2019 International Symposium on Electronics and Smart Devices
13. Mounir Bousbia-Salah, Abdelghani Redjati, Mohamed Fezari, Maamar Bettayeb, "AN ULTRASONIC NAVIGATION SYSTEM FOR BLIND PEOPLE", 2018 IEEE International Conference on Signal Processing and Communications (ICSPC 2019), 24-27 November 2019, Dubai, United Arab Emirates.
14. Samkit Shah, Jayraj Bandariya, Garima Jain, Mayur Ghevariya, Sarosh Dastoor, "CNN based Auto-Assistance System as a Boon for Directing Visually Impaired Person", Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019).
15. Aniqua Nusrat Zereen and Sonia Corraya, "Detecting Real Time Object Along with the Moving Direction for Visually Impaired
16. People", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE).
17. Rais Bastomi, Firza Puthra, Lucke Yuansyah Arif Tyras Putri, Septian Wahyu Saputra, Mohammed Rizki Maulana, Agus Khumaidi,
18. "Object Detection and Distance Estimation Tool for Blind People Using Convolutional Methods with Stereovision", 2019 International Symposium on Electronics and Smart Devices.
19. Manjunath Jogin, Mohana, "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning" 2019 IEEE International Conference On Recent Trends In Electronics Information Communication Technology, (RTEICT) 2019, India