# Smart Blind Stick using Artificial Intelligence

**Pruthvi S, Pushyap Suraj Nihal, Ravin R Menon, S Samith Kumar, Shalini Tiwari**

***Abstract: Smart Blind Stick is a device designed to help guide the visually impaired by detecting objects and portray the information to them in the form of speech. This reduces the human effort and gives better understanding of the surrounding. Furthermore it also provides an opportunity for visually impaired people to move from one place to another without being assisted by others. The device can also be used in old age homes where old age people have difficulty in their day to day activities due to decreased vision. With this paper, the aim to aid people in need to "see" the surroundings. Since the field of artificial intelligence is doing great progress now and features like object detection is getting easier and computationally feasible, these features are implemented in the paper. The paper focuses on object detection and classification on pictures which are captured by the device mounted on a stick whose information can then be relayed to the user in means of sound or speech.***

***Keywords- Object detection, YOLO, Tensorflow, eSpeak, Raspberry pi, Blind, Visually impaired.***

## I. PROBLEM STATEMENT

According to World Health Organization (WHO), there are over 1.3 billion people who are visually impaired across the globe [1], out of which more than 36 million people are blind. India being the second largest population in the world, contributes 30% of the overall blind population. Although there are enough campaigns being conducted to treat these people, it has been difficult to source all the requirements. It is the era of artificial intelligence and it has gained immense traction due to large amount of data and ease of computation [2]. Using artificial intelligence it is possible to make these people's life much easier. The goal is to provide a "secondary sight" until they have enough resources required to treat them. People with untreatable blindness can use this to make their everyday tasks much easier and simpler.

## II. INTRODUCTION

The paper focuses on making a device which is portable and provides a "secondary vision". The device consists of a Raspberry Pi, a Hi-Res Camera, an object detection algorithm (YOLO) and a text to speech unit (eSpeak).This unit can be mounted on a stick from where, it can capture images and process them. The device alerts the user if they come across any obstacles and give the description of what is in front of them. It can classify objects using directory of self-learned models.

**Manuscript published on 30 May 2019.**
***** Correspondence Author (s)
**Pruthvi S,** School of C&IT, REVA University, Bengaluru, India.
**Pushyap Suraj Nihal**, School of C&IT, REVA University, Bengaluru,
**Ravin R Menon**, School of C&IT, REVA University, Bengaluru, India.
**S Samith Kumar**, School of C&IT, REVA University, Bengaluru, India.
**Shalini Tiwari.** Asst. Professor, School of C&IT, REVA University, Bengaluru, India.



Since computation on the raspberry pi is pretty good, the images captured by the camera can be locally processed. The insight gained from these self-learned models can generate what is present in the images. This information is then converted to text. Then the text-to-speech module can inform the user about their surroundings. Smart Blind Stick is capable of detecting 80 classes in real time. The steps involved to do so are, repeatedly capture the image from the camera, pass these images through the classifier and the results obtained from this are then read from the text to speech engine. The time it takes for the object detection is 0.426 seconds on average. The text to speech may take time to relay the information to the user depending on the number of objects present in front of them. There's an ultrasonic sensor which captures the distance of the nearest object and says it to the user along with the results from the object detection.

To achieve this, the device has to have a smaller footprint yet high computation requirement. The device should be affordable as well. So, raspberry pi was chosen for the price to performance ratio and its small size. It is also open-sourced and has a huge community support since it works on GNU/Linux platform. Wide range of distributions can be found and there are many libraries and frameworks which ease the programming. There's also a camera which can be interfaced with it very easily.

Section 1 defines the problem the paper aims on solving and the motivation for the paper. Section 2 gives us an introduction of the paper and how to go forth to solve this problem with the paper briefly. Section 3 contains all the related work of devices which help the blind. Section 4 gives us the methodology of the paper i.e. how the device is made and the required frameworks and programs in detail. Section 5 contains the results of the paper and what the user is going to get out of this paper. Section 6 contains the conclusions that can be drawn from this paper and the steps that can be taken for future improvements to make it even better.

## III. RELATED WORK

"Seeing AI", by Microsoft aims at solving this problem by using an app which sends pictures to the cloud for processing. It could detect objects and had more features like text recognition which could read books, pages etc. Also users can't do anything while they are using the app on their phone. So there was a need to come up with an entirely new device to do this. "Orcam MyEye2" is a product which costs around 80,000 rupees with some additional features like facial recognition etc. The aim is to make it cheaper but loose less features. Initially thought on using Google's Cloud Vision

API but this would mean slower response times, high number of requests and considering the network coverage for internet is not that good in India, had to resort to local processing.

"Wearable Object Detection System for the Blind" [4] was published, where they used RFID to deliver data to user about the items which were tagged to help the blind. But this involved someone going through the objects, attaching RFID tags to everything and updating the database. There was a need for easier approach than this where more efforts should not be necessary. "Effective Fast Response Smart Stick for Blind People" [5] was published which used ultrasonic, infrared and water sensors to detect any objects within 4 meters very quickly. It focused on detecting staircases, distance of objects and water puddles. More information had to be conveyed than the distance of an obstacle by providing the description of the obstacle to the user.

"An Implementation of an Intelligent Assistance System for Visually Impaired/Blind People" [7] proposed an intelligent system which composed of smart wearable glasses and an intelligent stick which would detect objects and tell the user about the distance. If the user falls down, it would send the information (GPS, fallen, etc.) through a mobile application. "Ultrasonic Sensor Based Smart Blind Stick" [8] presented a system which used ultrasonic sensor HC-SR04 to detect obstacles within 5 to 35 cm distance and relay it to the user using a buzzer.

"SWSVIP–Smart Walking Stick for the Visually Impaired People using Low Latency Communication" [9] proposes a system which uses new algorithms for low latency communication of the information from the ultrasonic sensor, GPS, RFID etc. This information can be used to inform emergency contacts if there is any inconvenience.

## IV. METHODOLOGY

### A. Device

The PiCam v2 is attached to the raspberry pi with the help of CSI camera port. The CSI bus has a high data transfer rate and it is exclusively for transferring pixel data.

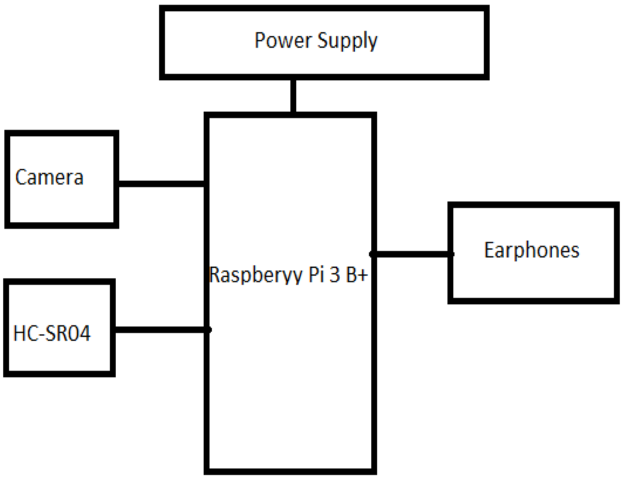


***Fig. 1: Block diagram of the device***

The ultrasonic sensor HC-SR04 emits a 40000 Hz ultrasonic wave which is used to calculate the distance. It is interfaced to the Raspberry pi using the GPIO pins. The board has a 3.5mm jack where earphones or headphones can be plugged in to listen to what is around the user. The battery pack is used to power the device.

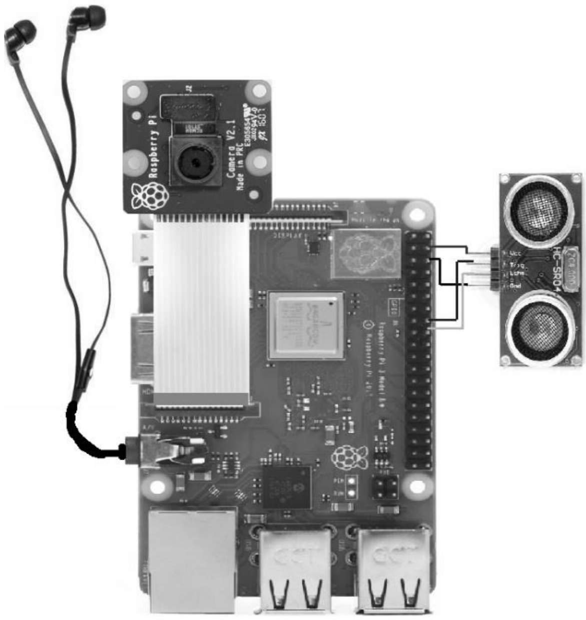

***Fig. 2: Schematics of the proposed design***

### B. YOLO and Darkflow

You Only Look Once [3] (YOLO) is an image classifier that takes parts of an image and process it. In classic object classifiers, they run the classifier at each step providing a small window across the image to get a prediction of what is in the current window. This approach is very slow since the classifier has to run many times to get the most certain result.But YOLO [3], divides the image into a grid of 13x13 cells. This means it just looks at the image just once and thus faster. Each grid box predicts bounding boxes and the confidence of these bounding boxes.

The confidence represents how accurate the model is that the box contains the object. Hence, if there is no object, then the confidence should be zero. Also an intersection over union (IOU) is taken between the predicted box and the ground truth to draw the bounding box.

As described in [3], each bounding box has 5 predictions:x,y,w, h and confidence. The(x,y)coordinates represent the centreof the box relative to the bounds of the grid cell. The widthand height are predicted relative to the whole image. Finallythe confidence prediction represents the IOU between thepredicted box and any ground truth box.

Retrieval Number: E10040585S19/2019©BEIESP
DOI:10.35940/ijeat.E1004.0585S19
Journal Website: www.ijeat.org



Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

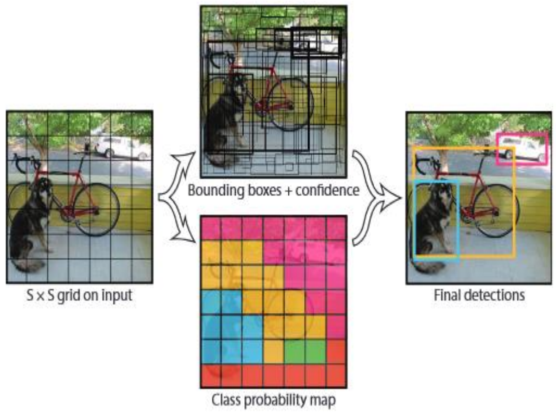


*Fig. 3: Model of YOLO*

For every cell, the classifier takes 5 of its surrounding boxes and predicts what is present in it. YOLO outputs a confidence score that lets us know how certain it is about its prediction. The prediction bounding box encloses the object that it has classified. The higher the confidence score, the thicker the box is drawn. Every bounding box represents a class or a label. Since there are 13×13 = 169 grid cells and each cell predicts 5 bounding boxes and end up with 845 bounding boxes in total. It turns out that most of these boxes will have very low confidence scores, so the boxes whose final score is 55% or more are retained. Based on the needs, the confidence score can be increased or decreased.

From the paper [3], the architecture of YOLO can be retrieved, which is a convolutional neural network. The initial convolutional layers of the network extract features from the image whereas the fully connected layerspredict the output probabilities and coordinates. YOLO [3] uses a 24 layer convolutional layers followed by 2 fully connected layers. The final output from this model is a 7X7X30 tensor of predictions.

The PASCAL VOC [6] dataset is used to train this model. There is an implementation of YOLO [3] in C/C++ called darknet. There are pre-trained weights and cfg which can be used to detect objects on. But to make the implementation more efficient on the raspberry pi, the tensorflow implementation of darknet called the darkflow is used. The images are passed to this image detection framework and get the output which contains the 5 predictions as discussed before. Darkflow outputs the file with bounding boxes or a json file. This json file is converted into text file and takes the count of objects and omits the rest. The objects along with their count are fed into the text to speech unit eSpeak.

### C. Distance

The ultrasonic sensor HC-SR04 is used to calculate the distance of the nearest object. The sensor has 4 pins: VCC, Ground, Echo and Trigger. VCC is connected to Pin 4, Ground to Pin 6, Echo to Pin 18 and Trigger to Pin 16. It emits a 40000 Hz ultrasound which bounces back if there is an object in its path. Considering the time between the emissions of the wave and receiving of the wave, the distance between the user and the nearest object can be calculated very accurately and it can be relayed to the user.The sensor can detect objects in range of 2cm–400 cm. The distance is calculated by the formula given below.

17150*Time = Distance.

### D. Workflow of Device

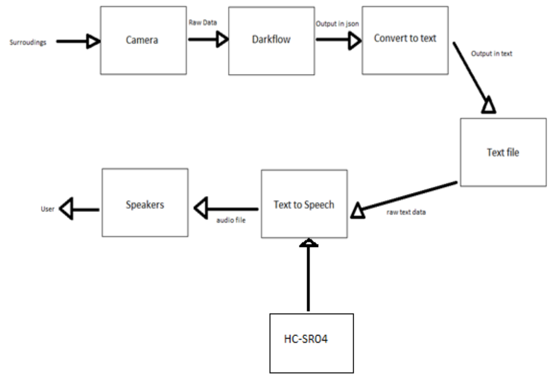


**Fig. 5: The working of the device.**

To change the output type of the object classifier from json to text, change a few lines in the predict.py in darkflow to output a text file with labels and count. To get count of the objects detected, dictionaries are used to store them as key-value pairs where name is the key and the count is the value. This dictionary is then written onto text files. The text files are read by the text to speech line by line where each line contains one key-value pair.

To capture images and predict continuously, the prediction is run through a loop which goes on till end i.e. when an escape sequence is invoked. Once it is invoked, it comes out of the loop and stops the program.

## V. RESULTS

The results from the object classifier are as follows. The text file which contains the labels and the count of objects are also shown. Distance from the ultrasonic sensor is received and the text to speech uses this to read it out to the user along with the text file which contains labels and count.

INPUT: OUTPUT:

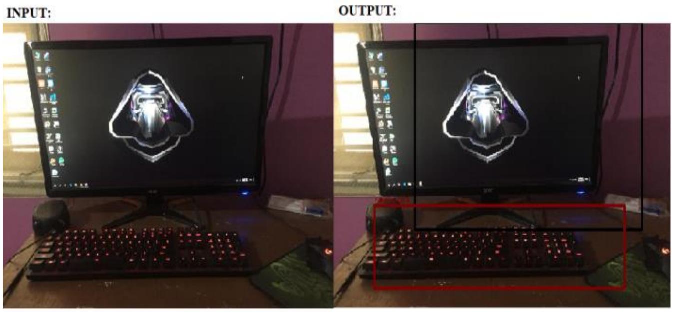

*Fig. 6: Result 1*

Text Output:
{“screen”: 1, “keyboard”: 1}
Distance: 43 centimetres

In this result as we can see, the object detection detects a screen and keyboard. This is output in the form of text file which has label (“screen”, “keyboard”) and the count associated to those labels. The distance from the ultrasonic sensor is retrieved.


Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

All these are passed to the text to speech engine which then reads it out to the user. Each cycle of taking picture, running object detection, retrieving information from ultrasonic sensor and relaying it to the user takes about 1.22 seconds on average depending on the number of objects in the image. If there are less objects, the lower the time and vice versa. There are 80 labels which can be identified.

## VI. CONCLUSIONS

Thepaper helps people who are visually impaired to get a better understanding of the surroundings. It speaks to the user what is present in front of them, hence helping them avoid obstacles and giving them the ability to "see" what's around them.

The paper aims to make the world a better living environment for people who are handicapped or have a tough time seeing. From the results seen above, it is clear that the user can greatly benefit in terms of knowing what is around them at a cheaper, efficient and easy way.

There are wide varieties of objects that the device can detect. Hence it can be used for everyday activities to enhance their experience and create a better place for them. In the future, face detection can be added, so that if there are any familiar faces, they can be recognised. Also text conversions would help the users read books or posters and signs to enhance their understanding even more. With increasing research in embedded systems, more computation in a smaller scale can be expected in the future. This can greatly impact the implementation of AI features like object detection, face recognition etc.

## REFERENCES


1. WHO," Universal eye health: a global action plan 2014-2019", ISBNNo: 978 92 4 150656 4.
2. Dominic Basulto, "Artificial intelligence is the next big tech trend.", Washington Post, March 25, 2014
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", Conference on Computer Vision and Pattern Recognition 2016, arXiv:1506.02640
4. Dionisi, A & Sardini, Emilio & Serpelloni, Mauro. (2012). "Wearable object detection system for the blind". 2012 IEEE I2MTC - International Instrumentation and Measurement Technology Conference, Proceedings. 1255-1258. 10.1109/I2MTC.2012.6229180.
5. Nada, Ayat & Mashali, Samia & Fakhr, Mahmoud & Seddik, Ahmed. (2015). "Effective Fast Response Smart Stick for Blind People". 10.15224/978-1-63248-043-9-29.
6. M.Everingham, S.M.A.Eslami, L.Van Gool, C.K.I.Williams, J.Winn, and A.Zisserman. The pascal visual ob-ject classes challenge: A retrospective. International Journal of Computer Vision, 111(1):98–136, Jan. 2015.
7. L. Chen, J. Su, M. Chen, W. Chang, C. Yang and C. Sie, "An Implementation of an Intelligent Assistance System for Visually Impaired/Blind People," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-2, 2019.
8. N. Dey, A. Paul, P. Ghosh, C. Mukherjee, R. De and S. Dey, "Ultrasonic Sensor Based Smart Blind Stick," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, pp. 1-4, 2018.
9. Shaha, S. Rewari and S. Gunasekharan, "SWSVIP-Smart Walking Stick for the Visually Impaired People using Low Latency Communication," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, pp. 1-5, 2018.


## AUTHORS PROFILE

**Pruthvi S** is currently pursuing B.Tech in Computer Science and Engineering at REVA University, Bengaluru, Karnataka, India.
Email: pruthvikanta@gmail.com

**Pushyap Suraj Nihal** is currently pursuing B.Tech in Computer Science and Engineering at REVA University, Bengaluru, Karnataka, India.
Email: surajnihal@gmail.com

**Ravin R Menon** is currently pursuing B.Tech in Computer Science and Engineering at REVA University, Bengaluru, Karnataka, India.
Email: ravinmenon9@gmail.com

**S Samith Kumar**is currently pursuing B.Tech in Computer Science and Engineering at REVA University, Bengaluru, Karnataka, India.
Email: samith6@gmail.com

**Shalini Tiwari** is an Assistant professor of School of C&IT in REVA University, Bengaluru, Karnataka, India.
Email: shalinitiwari@reva.edu.in

*Retrieval Number: E10040585S19/2019©BEIESP*
*DOI:10.35940/ijeat.E1004.0585S19*
*Journal Website: www.ijeat.org*



*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*