International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 4919–4926

In-house Object Detection System for Visually Impaired

Nilakshi Mule
Government College Of Engineering Karad, India

Dipti D. Patil
Mksss's Cummins College Of Engineering For Women, Pune, India

Yashpalsing D. Chavhan
Government College Of Engineering Karad, India

Abstract

The worldwide population of visually impaired is 2.2 billion. The visually impaired person has to face many challenges while performing their everyday activities. In this paper, the proposed system mainly focused on providing in-house object detection. The various household objects like TV, Chair, Remote, and Bottle are used as objects for object detection using Raspberry pi 3 kits, Tensorflow, OpenCV, and SSDlite MobileNet V2. The system generates an audio output based on hand gestures and the detected object labeled in English or Hindi languages. The system also calculates the distance between the users and objects. The system has shown precision 0.85 and recalls 0.8 with a 2-second delay in generating audio output.

Key Words: Object detection, hand gesture detection, Visual impairment, Raspberry pi 3 kit, Tensorflow, OpenCV, and SSDlite MobileNet V2.

1. INTRODUCTION

According to the World Health Organization report [16], worldwide, 2.2 billion people are suffering from visual impairment. The visual impairment can be caused due to various reasons such as age-related muscular degradation, cataract, trachoma, refractive error, corneal opacity, glaucoma, diabetic retinopathy and so on. The visually impaired person has to face many challenges while performing normal day to day activities. The visually impaired person has to rely on their friends and family. Visual impairment can also affect the social and economic condition of a person. The literacy rate of visually impaired people is also less as compared to other disabilities.

Self-navigation is the most important issue for the visually impaired person. The navigation can be categories into outdoor and indoor navigation [14]. Object detection and object identification are the main tasks in assisting with navigation. Various products are available for a visually impaired person to provide navigation assistance, but affordability is the main hurdle as per the economic condition. It is a laborious task for them to navigate from one place to another, and the conventional blind walking stick with its limitations.

The existing systems have provided solutions such as smart glasses and pair of shoes [4] to integrate with different sensors and raspberry pi. Using Raspberry Pi 3 for the elderly, a wearable device [10] is built to help see articles and tell the sort and area of items through voice. The YOLO neuron network is picked based on speed and asset use. The informational indexes of photos of familiar objects were arranged and cared for as preparing contributions to the YOLO organization. A tale multi-sensor combination based deterrent evasion calculation is proposed, not typical for existing work [8], which utilizes both the profundity sensor and the ultrasonic sensor to take care of the issues of little snag location and straightforward impediments. Article location algorithms [3], such as face recognition, skin discovery, shading identification, shape identification, target discovery, are reproduced and executed utilizing

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

MATLAB 2017 to distinguish various kinds of items with upgraded exactness video observation applications. The Smart Glass application [5] is equipped to record and perceive public signs in urban areas and send discourse signs to visual impairment. The proposed model design depends on Intel Edison and, with OpenCV libraries, the program in the engineering is written in C++. The impediment recognition module [12] comprises an ultrasonic sensor, a control module comprises a preparing unit, and a signal comprises a yield unit. The control unit screens the ultrasonic sensors get data on the deterrent before the person and examinations the information and automatically sends the yield through the signal. For identifying obstacle VIZIYON [15] IoT based device designed for real time object detection. In this identification is carried on the distance vector and object detection. Object recognition is done using CNN. The review of [1] deep learning systems used with navigational for visually impaired. It also compares the CNN and FCNN for the development of multifunctional technology. The real-time objects are detected [7] using RFID and IR technologies. In real-time detection, it is difficult to identify the smaller objects or object which cover a smaller area. The object is detected by creating a bounding box [11] around the object which is done using background subtraction technique. The fast R-CNN algorithm [2] used for object detection but cannot recognize multiple objects at the same time detection time is also more. Mobile camera [13] used for obstacle detection in an indoor environment. In an indoor environment, floor images are used and stored in a unique frame—segmentation and threshold algorithm [9] used for the frames. CNN is used for the 2.5D object detection according to the location and orientation of the object.

In this paper, we proposed IOT based system to address visually impaired people's indoor object detection problem using image processing and machine learning algorithms. The system can generate audio output in English and Hindi language. The language selection is based on hand gesture identification. . The scope is kept limited to household appliances for the purpose of experimentation.

2. PROPOSED SYSTEM

After reviewing the related literature survey, a custom dataset was designed for household objects, for audio feedback used TTS Engine (Espeak). SSDlite provides more efficiency than YOLO on devices; it lets computation power such as Raspberry Pi. Instead of using the Ultrasonic sensor, which would increase power consumption, it would be challenging to detect the user's particular object. The hardware and software used in the development of the system are Raspberry Pi with the camera module, ultrasonic sensor, SSDlite MobileNet V2.

2.1 Raspberry Pi Camera Module

The Raspberry Pi Camera Board is an add-on module for Raspberry Pi hardware that is custom-built. For mobile ventures, the lightweight and compact camera module makes it the perfect choice. It links the hardware through a custom CSI interface to the Raspberry Pi. The sensor has a local 5-megapixel goal in still-catch mode. Video mode allows goals of up to 1080p at 30 casings for each second for shooting. Utilize the camera module to catch a picture that can be utilized to distinguish the item.

2.2 Gesture Recognition

The hand gesture is used to select the language, either Hindi or English, a closed hand for Hindi and an open hand for English. The hand gesture used to give input to the Pi Camera further informs Raspberry Pi, and accordingly, the language was selected. If no input would have been given, then English as a default language get selected.

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

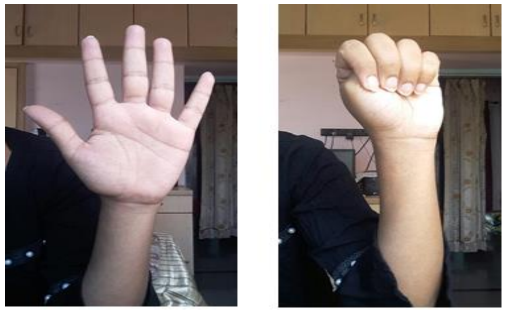

Fig. 1 Hand Gesture Recognition

2.3 Object Detection

The object detection is performed by using SSDlite MobileNet V2 [2]. The MobileNet V2 provides the advantage of requiring less computational power to work with raspberry pi. It also provides an output at higher frames per second, an essential part of the system. This decreases the latency in instructions provided to the user in real-time. MobileNetV2 comprises two types of blocks. One is a block residual with a period of 1. Another one for downsizing is a 2-step block.

- For both block shapes, three layers exist.
- This first layer is a convolution of 1 *1.
- The second layer is wisdom's deep convolution.
- A further 1 * 1 convolution is the third sheet but without any non-linearity.

Profound organizations are accepted to have a direct classifier on the yield area's non-zero volume segment of the just if the ReLU is utilized once more.

2.4 Distance Estimation

$$\textit{Image Distance} = \frac{\textit{focal length (mm)*average height*image height (pixels)}}{\textit{Height of object (pixels)*height of sensor (mm)}}$$

Height of object = ymax-ymin.

Focal length for pi-camera.

This simple formula helps to illustrate object distance from the camera. Image height can be calculated from the bounding boxes generated by the model.

2.5 Algorithm

Algorithm for Object Detection, distance estimation with voice feedback

1. Start
// Raspberry pi powered by power bank.
2. Check for User hand gestures
2.1. If open hand then
2.1.1. Hindi language.
2.2. Else if close hand then
2.2.1. English language.
2.3. Else
// No gesture recognized then

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

2.3.1. Wait for time=10 S && choose default English language

3. *End*
4. *Start object detection*
   4.1. Capture live video frames && feed to Model // SSDlite Mobile Net V2.
   4.2. Preprocessing on each individual frames.
   *4.3. 1*1 convolution with ReLu6.*
   4.4. Depth-wise Convolution with ReLu6.
   *4.5. 1*1 convolution.*
   *4.6. 1*1 convolution with ReLu6.*
   4.7. Depth-wise Convolution with ReLu6 [Stride=2].
   *4.8. 1*1 convolution.*
5. *Predict object classes*
6. *Extract attributes of bounding boxes.*
7. *Estimate object distance.*
8. *Print object name and estimated distance // Text to speech engine.*
9. *Voice Feedback*
10. *Stop*

3. SYSTEM ARCHITECTURE

Most of the existing systems used for object detection depend on sensors like ultrasonic, proximity sensors, etc., which affects the system performance as it increases the risk of failure due to maintenance required. Also, some systems are used to translate signboards, but none can provide live choose based audio feedback as an output [6]. Visually impaired people are most vulnerable to casualties in an unknown environment. The assistance in the form of such systems decreases their dependencies on others.

The general working of the system is described in fig 2. The user wearing Pi-Camera enabled glasses would activate raspberry pi, which will trigger Pi-Camera. At each instance, the user receives instructions from the system. The language of these instructions can choose by the user based on hand gestures. The system provides two language choices to the user, either English or Hindi. After the selection of language using gestures, the system automatically starts detecting objects in the frame. Each frame captured from the pi camera is fed to a neural network (SSDlite MobileNet V2). As this model name itself suggests, it goes easy on computation resulting in less power consumption. The model's output consists of an image with a bounding box for each object detected in the frame. The system also counts the number of objects which belong to the same class. Providing a user with such minor details help them get a rough idea of their environment. The attributes obtained from bounding boxes are used to calculate an object's distance from the camera lens. At last, the voice feedback is provided through earphones. The Espeak text to speech engine is used to provide voice assistance.

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

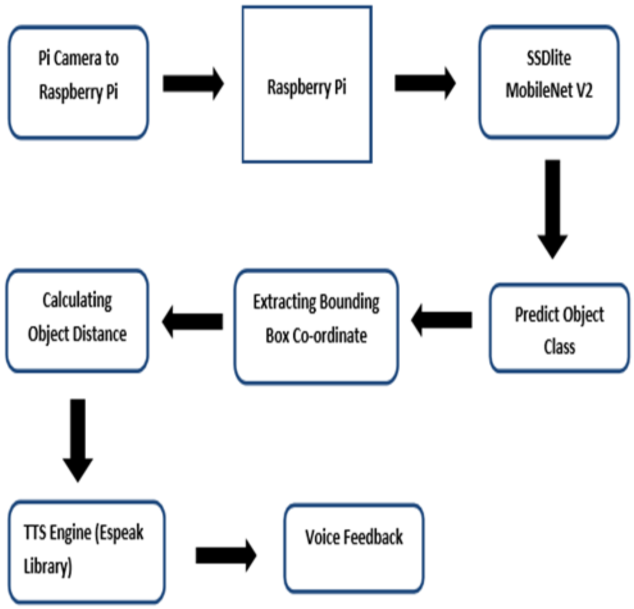

Fig. 2 Block diagram of system

4. SYSTEM IMPLEMENTATION

The Raspberry pi 3 and the camera module with 5-megapixel resolutions are used for image capture and computational purposes. The Tensor flow and OpenCV are used for image processing. The system asks at the beginning for the selection of a language. The selection of language based on hand gestures, either English or Hindi. The system setup is shown in figure 3.

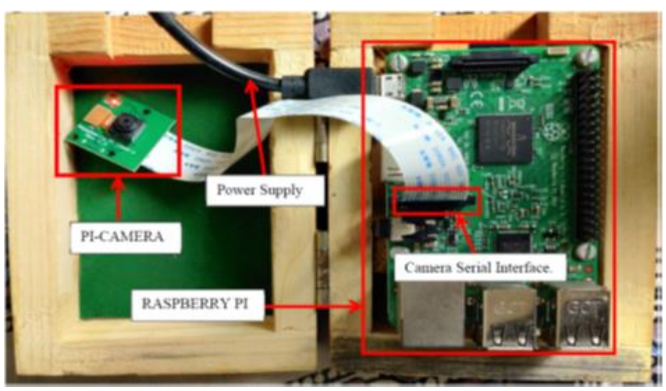

Fig. 3 System Hardware Setup

5. RESULT AND DISCUSSION

The experimentation is carried on 10 household objects. The selected objects are water bottles, chairs, table, sofa, Television set, TV remote, cupboards, doors, windows, and shoe stand of different shapes and sizes. The dataset contains a total of 200 images. The 20 images of different sizes and shapes for each object were captured from different houses. The system is trained on 70% of each object's images, and testing is carried on the remaining 30% of images. The system gives precision 0.85 and

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 4919–4926

recalls 0.8. The output image label is provided as voice output to the user in either English or Hindi languages. The time latency for generating voice output for 2 seconds.

Table I. Performance comparison

Method	Precision	Recall
Our	0.85	0.8
YOLO[3]	0.83	---
VIZIYON[8]	0.94	0.92

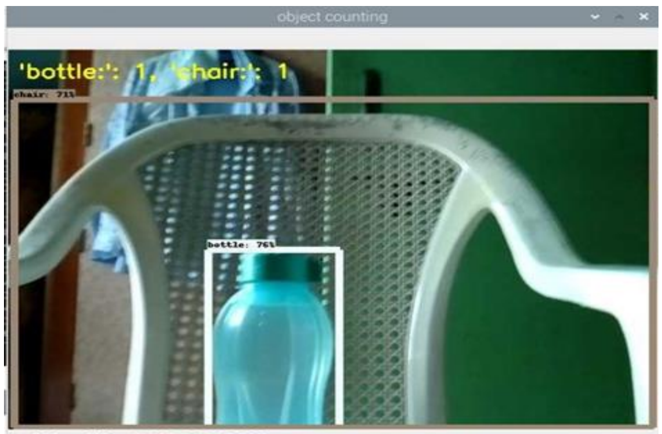

Fig. 4 Object detected Chair and Bottle

Fig. 5 Object detected Bottle and TV Remote

The above figure 4 and figure 5 shows the model's final output, where the object is detected, and the voice output is given to the user through the earphone.

6. CONCLUSION

The visually impaired person has to face many difficulties to perform day to day work. The results show that the proposed system improves house experiences by detecting objects and providing necessary information. The proposed system results give precision 0.85 and recall 0.8 and the audio output in English or Hindi languages at low cost.

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

The implementation of a system with better hardware and a high-resolution camera may increase the cost. However, it may also help improve the results and minimize the time delay for generating audio output. More household and outdoor objects included in this System with GPS and navigation facilities provide much-needed ease for visually impaired people.

REFERENCES

1. ABINASH BHANDARI,P.W.C. PARSAD, ABEER ALSADOON, ANGELIKA MAAG, 2019, *Object detection and recognition: using deep learning to assist the visually impaired, Disability and Rehabilitation: Assistive Technology, 2019*
2. ANITHA J, SUBALAXMI A, VIJAYALAKSHMI G, 2019, *Real Time Object Detection for Visually Challenged Persons, IJITEE, Vol. 8, Issue-8, June 2019*
3. APOORVA RAGHUNANDAN, MOHANA, PAKALA RAGHAV, H. V. RAVISH ARADHYA, 2018, *Object Detection Algorithms for Video Surveillance Applications, In Proceedings of the IEEE International Conference on Communication and Signal Processing, Chennai, India, 2018.*
4. ARJUN PARDESHI, PRITHVIRAJ N INDI, SASHWATA BANERJEE, ADITYA KAMAL, 2019. *Smart Assistive Navigation Devices for Visually Impaired People, In Proceedings of the 4th International Conference on Compute and Communication Systems. Singapore February 2019.*
5. FENG LAN, GUANGTAO ZHAI, WEI LIN, 2105, *Lightweight smart glass system with audio aid for visually impaired people, In Proceedings of the TENCON 2015 - 2015 IEEE Region 10 Conference, Macao, China, November 2015.*
6. GULBAKSHEE J DHARMALE, DIPTI D PATIL, *Evaluation of Phonetic System for Speech Recognition on Smartphone", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-10,*
7. HEETIKA GADA, VEDANT GOKANI, ABHINAV KASHYAP, MAIT DESHMUKH,2019, *Object Recognition for The Visually Impaired, In Proceedings of the IEEE International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, January 2019.*
8. JINQIANG BAI, SHIGUO LIAN, ZHAOXIANG LIU, KAI WANG, DIJUN LIU, 2017, *Smart Guiding Glasses for Visually Impaired People in Indoor Environment, IEEE Transactions on Consumer Electronics, Vol. 63, No. 3, 258-266.*
9. JINQIANG BAI, ZHAOXIANG LIU, YIMIN LIN, YE LI, SHIGUO LIAN AND DIJUN LIU, 2019, *Wearable Travel Aid for Environment Perception and Navigation of Visually Impaired People, MDPI, Electronics 2019, 8,697.*
10. PRUTTAPON MAOLANON, KANJANAPAN SUKVICHAI,2108, *Development of a Wearable Household Objects finder and Localizer Device using CNNs on Raspberry pi 3, In Proceedings of the IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Chonburi, Thailand, August 2019.*
11. REBEIRO SHARLENE SARA CARLTON, HUDA NOORDEAN, 2018, *A Review on Object Recognition for Blind People Based on Deep Learning", IJSR Vol. 7 Issue 8, August 2018.*
12. ROHIT AGARWAL, NIKHIL LADHA, MOHIT AGARWAL, KUNTAL KR. MAJEE, ABHIJIT DAS SUBHAM KUMAR SUBHAM KR. RAI ANAND KR. SINGH, SOMEN NAYAK SHOPAN DEY RATUL DEY HIMADRI NATH SAHA,2017, *Low cost ultrasonic smart glasses for blind, In Proceedings of the 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) IEEE Conference, Vancouver, BC, Canada, November 2017.*
13. SAMI UR RAHMAN, SANA ULLAH, SEHAT ULLAH, 2018, *Obstacle Detection in Indoor Environment for Visually Impaired Using Mobile Camera, Journal of Physics: Conference Series 2018.*

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC

14. SHWETA A. TIWASKAR, DR. DIPTI D. PATIL,2017, *Development of Framework for Moving Object Detection and Trackingin Video Sequences, International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS), Volume VI, Issue VII*
15. SREERAJ M, JESTIN JOY, ALPHONSA KURIAKOSE, BHAMEESH M B, ANOOP K BABU, MERIN KUNJUMON, 2020 VIZIYON:*Assistive handled device for visually challenged", In Proceedings of the 3rd International Conference on Computing and Network Comunications (CoCoNet'19) Trivendrum, India, December 2019.*
16. "World report on vision", report published by World Health Organization 2019 https://www.who.int/publications-detail/world-report-on-vision.

ISSN: 2233-7857 IJFGCN
Copyright ⓒ2020 SERSC