Obstacle Detection and Text Recognition for Assisting Visually Impaired People

Aishwarya. M
Department of Computer Science and Engineering SRM Institute of Science and Technology, Vadapalani
Chennai, India
ma5515@srmist.edu.in

Shivani. S
Department of Computer Science and Engineering SRM Institute of Science and Technology, Vadapalani
Chennai, India
ss7497@srmist.edu.in

Deepthi. K
Department of Computer Science and Engineering SRM Institute of Science and Technology, Vadapalani
Chennai, India
dk9974@srmist.edu.in

Dr. Golda Dilip
Department of Computer Science and Engineering SRM Institute of Science and Technology, Vadapalani
Chennai, India
goldad@srmist.edu.in

***Abstract*— As a potential medium for informing blind people, the project remodels the visual world into the audio world. In computer vision, Obstacle detection is the most widely used field with incredible achievement. The obstacles and texts can be enlightened to the visually impaired people where obstacles detected from the frame are addressed by their names and converted to speech. The image-to-text framework is an advanced innovation that is utilized to get the message in the image which will be extremely helpful to read any content. The voice result of the input text is produced by evaluating the adjoining and the recurrence of events of the words by the system. It is used to find the distance to the object that the user wants. This application is compact and helpful to economical society as well as an efficient assistant for visually impaired people. Thus, ultimately this could increase the confidence level of the user and make him/her feel secure.**

Keywords: Visually Impaired, Object Detection, Text Recognition, Text-to-Speech, Distance Estimation, Find Objects, Convolutional Neural Network, Computer Vision.

I. INTRODUCTION

Dealing with sight loss, already, could be a challenge in itself. But the biggest challenge for a visually impaired person is to navigate around places. To understand the environment, visually impaired people rely on their other senses such as touch and auditory signals. It is very difficult for visually impaired people to know what object is in front of them without touching it. This project aims at assisting visually impaired people to be aware of surrounding objects and learn about them through speech generated from the textual content. It is always challenging to cope with the external environment for a blind person. The idea is to identify and track the dynamic obstacles by recognizing the data in the current frame. "Vision" application will help the person to detect the objects and texts around them. Visually impaired people will be guided in a better way by estimating their distance from the object. For every action, the user will be assisted via voice. Vision helps blind people to explore the world around them and helps them to visualize the world with a new scenario. It is easily accessible and gives a few highlights that can be utilized by the visually impaired for doing everyday activities without the requirement for others' help.

II. MOTIVATION

For utilization of any facilities like food or travel, visually impaired people need support, so our project will be useful for them by interacting and helping them recognize the surroundings and environment by conveying about it. This project focuses on assisting users in detecting objects with the utilization of innovation and technology we have which our engineering career inspires us to carry out.

III. RELATED WORKS

In 1986, simple object detections on images are performed with algorithms like the Histogram of Oriented Gradients, and Support Vector Machine which resulted in a decent accuracy. [2] Now, there are many modern architectures such as YOLO, RetinaNet, Faster R-CNN, and Mask R-CNN. The Existing system uses slower algorithms for object detection (like R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, and Libra R-CNN) which typically consists of a two-stage approach. The conventional object detection models are primarily divided into 3 stages: informative region selection, feature extraction, and classification. All of these regions are passed to classification which is a tedious activity. Hence, when compared to two-stage detection one-stage detection performs faster.

Before the existence of smart blind sticks, visually impaired people used the normal cane, but normal cane could get stuck in cracks and on uneven surfaces. It also makes navigation difficult. The Smart blind stick cannot differentiate between a person or object, it will simply sense it is an obstacle. A few disadvantages of the wearable sensors are specialized hardships, unfortunate plans, or the unfashionable design of the device. There is often a problem with waterproofing designs. Sweat and bad weather, like heat and precipitation, causes damage to the technology. An assistive device such as a smart blind stick, Kinect goggles system, and wearable technologies have hardware components like sensors and wired circuits which causes difficulties and discomfort for the person. [1] Generally, wearable technology consists of Raspberry Pi, sensor-based obstacle avoidance, a camera, and advanced algorithms for obstacle detection like MobileNet-SSD. [1] To determine the distance between the objects, traditionally ultrasonic sensors or other high-recurrence gadgets that produce sound waves were used. [5] Previously, the Flutter framework was used to build applications, which utilize Dart SDK and the flutter_tts library to convert the text into speech.

IV. System Design and Architecture

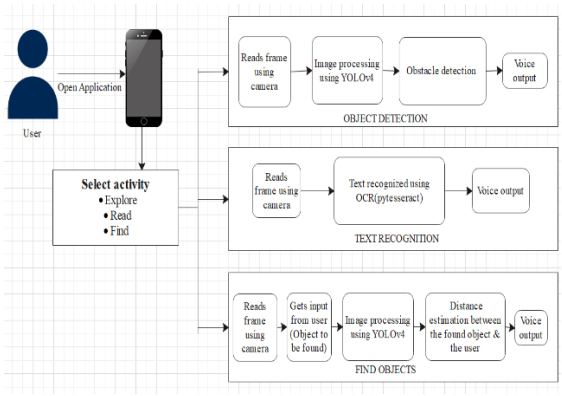

Fig. 1. System Architecture

In Fig. 1, the Vision Application first analyses the current frame according to the user's preference. For Obstacle Detection, the objects in the frame are identified with the help of the YOLOv4 algorithm. The name of the detected obstacles is passed to voice output. To read the text from a document or an image, the input from the frame is processed using Optical Character Recognition- Pytesseract. The recognized text from the frame is passed to the voice output. To find a specific object, vision gets input from the user. Based on the input, if the object is present in the frame, the distance of the obstacle from the user is calculated and is passed to the voice output. The respective output from each module is converted into speech using a Text-to-Speech synthesizer (pyttsx3) and conveyed to the user.

V. Methodology

A. *Proposed Work*

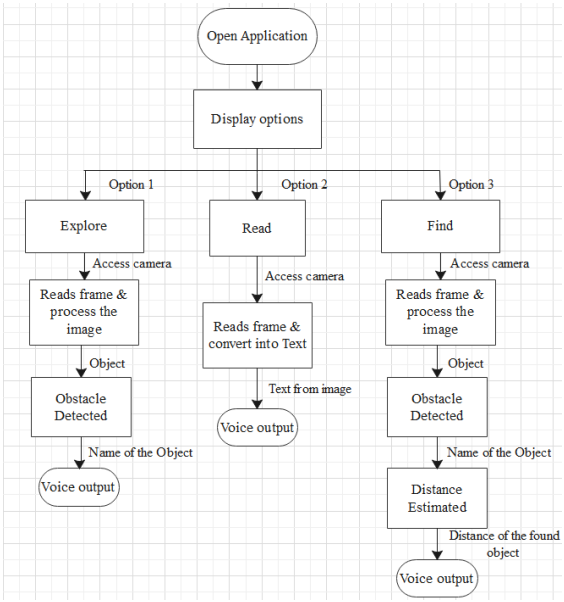

Fig. 2. Work Flow

This project is implemented using YOLOv4 and OCR- Pytesseract. In terms of both speed and accuracy, YOLOv4 is superior to the quickest and most accurate detectors compared to others. This version has established its high performance in a broad range of detection tasks. In YOLOv4, Transfer learning is a strategy to reuse the weights in at least one layer from a pre-trained network model in another model by keeping the weights, fine-tuning the weights, and adjusting the weights when training another model. To make the visually impaired person able to listen to messages, OCR is the most effective way for a machine to peruse any text. We prefer not to build an embedded device heavier by including unnecessary hardware modules while changing it into a compact design that has functionalities like obstacle detection. By making use of depth information that the camera uses to detect obstacles, we can calculate the distance between the specific obstacle and the camera that we have already incorporated for obstacle detection. The other applications that have been created before, require the person to catch a picture unlike ours, which doesn't require the person to catch a picture instead it just scans the object and informs them. By doing so, the visually impaired person will not have to periodically erase pictures from their phone's storage. Since it runs in real-time, it will be very useful for visually impaired people.

B. *Algorithms Used*

a) YOLOv4 (You Only Look Once- Version 4)

YOLOv4 is an incredibly quick multi-object detection algorithm that utilizes a convolutional neural network (CNN) to distinguish and recognize objects. It is a one-stage object detection model that enhances YOLOv3 with several bags of tricks and modules. The process is done by partitioning the frame into a grid and predicting bounding boxes and class probabilities for every cell in a grid. YOLOv4 recommends using IOU loss for bounding box regression like Distance IoU or Complete IoU loss function, leading to faster convergence and better performance. This model comprehends a generalized object representation which means the real-world scene prediction and artwork are fairly precise.

b) OCR (Optical Character Recognition)

Optical character recognition is the conversion of text in an image, transcribed, or printed text into the machine-encoded text from a photo of a document or a scene. Our goal is to determine the bounding boxes for each word of a sentence in a frame. Once we have got those regions, we can then OCR them. The best part of pytesseract is that it supports a wide variety of languages. It is through wrappers that Tesseract can be made compatible with totally different programming languages and frameworks.

C. *Obstacle Detection*

Obstacle detection is an important computer vision task used to detect instances of visual objects of certain classes (for example, humans, animals, cars, or buildings) in digital images such as photos or video frames. The object detection techniques are dealing with multiple object classification and its localization. In this module, we'll be using the YOLOv4 algorithm. The weights and configuration files for YOLOv4 are downloaded and they are fine-tuned according to the specification of the project. The coco name file contains 80 objects names that the pre-trained model can classify. Using the cv2 function, the weights and configuration files are loaded to the network. After loading the network as well as the objects' names, the input is taken from the frame using the camera which is pre-processed and communicates with the trained model. We get the output layers' names and then pass them into the forward function which will do a forward pass through the trained model. We loop through the layers to get the bounding boxes and confidence of the detected objects and will eventually make the prediction. For detected objects, a voice output is generated and it is communicated to the user.

D. *Text Recognition*

Text recognition is the process of localizing where an image text is. One can think of text recognition as a specialized form of obstacle detection. In this module, we will be using the python wrapper named pytesseract. It takes input from the frame using the camera which will be used to recognize text from a large document, or it can also be used to recognize text from an image of a single text line which is then converted into an array. We loop through the array to get the bounding boxes and confidences of the recognized text. Voice output is generated for recognized text and it is communicated to the user.

E. *Find Objects*

Speech Recognition incorporates technology and linguistics which permits computers to grasp human language. Using speech recognition, in python, the spoken words are converted into text, make a query, or, give a reply. With the assistance of a microphone, speech recognition starts by taking the sound energy produced by the person speaking. It then converts the electrical energy from analog to digital, breaks the audio information down into sounds, and analyses the sounds using algorithms to search out the most probable word that matches that audio.

To recognize the voice, we first initialize the recognizer using speech recognition. Speech to text takes input from the microphone as a source. Using adjust_for_ ambient_noise, we adjust the energy threshold based on the surrounding noise level. We get inputs from the user using listen function. Then the inputs are converted into speech using recognize_google. The converted speech is compared with the obstacles detected in the frame which is used to find the particular object.

$$\text{distance} = (2 \times 3.14 \times 180) \div (w + h \times 360) \times 1000 + 3$$

where-:

w - bounding box width,

h- bounding box height (1)

The width and height are used for estimating the obstacle and depicting the detail of the detected obstacle which is retrieved from the bounding box coordinates. The distance of the obstacle from the camera will vary depending on both variables.

F. Voice Assistant

Text-to-speech is the process of converting text into synthesized speech. It is used to communicate with users when reading on a screen is either impracticable or inconvenient. In this project, we are using the pyttsx3 package which has built-in say() function that takes a string value and talks it out. This function monitors when the engine starts converting text to speech and waits for that much time, and doesn't permit the engine to close. If it is not initialized, the engine probably won't work as expected as the processes won't be synchronized. The engine is shut down by calling the stop() function when all the processes are over. Voice output is generated for the detected objects, recognized text, and the estimated distance to an object.

VI. RESULT

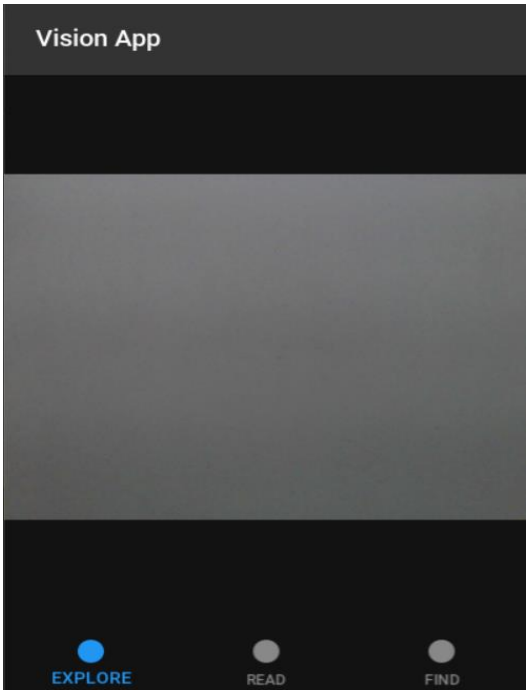

Fig. 3. Home Screen

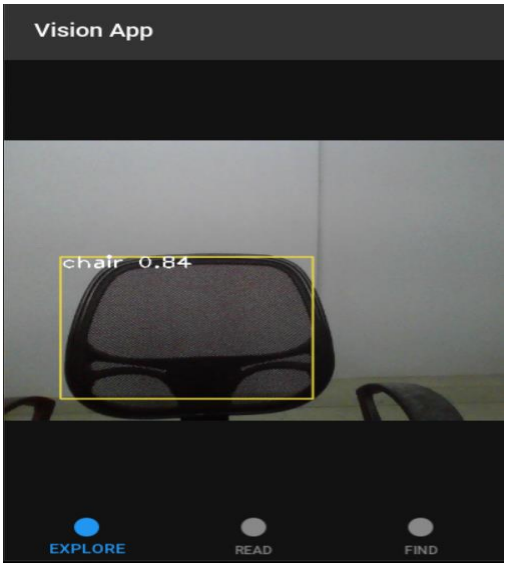

Fig. 4. Obstacle Detection

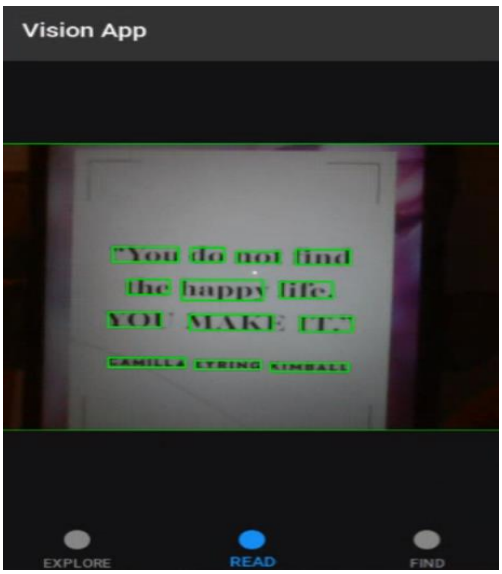

Fig. 5. Text Recognition

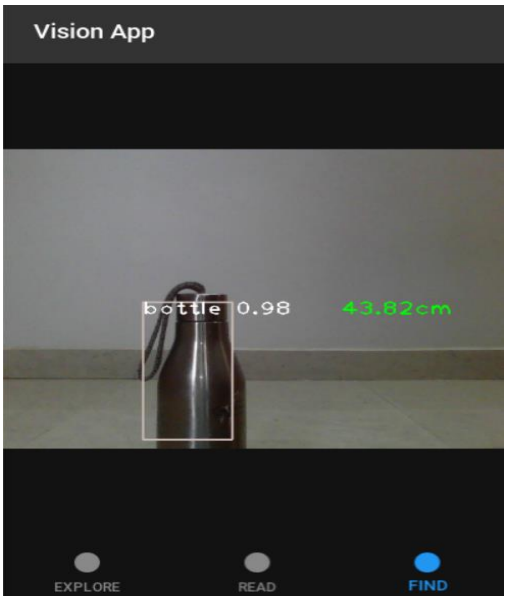

Fig. 6. Distance Estimation

VII. Conclusion

For all individuals, vision is the most important approach to receiving and decoding information from the world. However, individuals with vision impairment expertise the world through entirely different mechanisms. This paper deals with an application called "VISION" which is created mainly for visually impaired people. We have created this project with various options, which might facilitate them to build their life easier. The user gets assisted by knowing what's around them by exploiting our application. It provides them with a list of information like the objects near them and the textual content. The results are recited so the person can hear them through a voice-driven interface that is custom-built. By leveraging these technologies, this app was created that benefits visually impaired people with their everyday tasks.

Acknowledgment

We deeply express our sincere thanks to SRM Institute of Science and Technology, Vadapalani, for encouraging and allowing us to present the project successfully.

References

[1] Muiz Ahmed Khan, Pias Paul, Mahmudur Rashid, Mainul Hossain, and Md Atiqur Rahman Ahad, "An AI-Based Visual Aid with Integrated Reading Assistant for the Completely Blind", IEEE Transactions on Human-machine Systems, Vol. 50, No. 6, December 2020.

[2] Zhong-Qiu Zhao, Peng Zheng, Shou-Tao Xu, and Xindong Wu, "Object Detection with Deep Learning: A Review", IEEE Transactions on Neural Networks and Learning Systems, Vol. 30, No. 11, November 2019.

[3] V. Padmapriya,R. Archna,V. Lavanya,CH. VeeralakshmiKrishnaveni Sri, "A Study on text recognition and obstacle detection techniques", IEEE 2020 International Conference on System, Computation, Automation and Networking (ICSCAN).

[4] Md. Amanat Khan Shishir, Shahariar Rashid Fahim, Fairuz Maesha Habib, Tanjila Farah, "Eye Assistant using mobile application to help the visually impaired", IEEE 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019 (ICASERT 2019).

[5] Rajesh Kannan Megalingam; Palnati Teja Krishna Sai, Aditya Ashvin, Pochareddy Nishith Reddy, Bhargav Ram Gamini, "Trinetra App: A Companion for the Blind", IEEE 2021 Bombay Section Signature Conference (IBSSC).

[6] G Chandan, Ayush Jain, Harsh Jain, Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV", IEEE 2018 International Conference on Inventive Research in Computing Applications (ICIRCA).

[7] Bojan Strbac, Marko Gostovic, Zeljko Lukac, Dragan Samardzija, "YOLO Multi-Camera Object Detection and Distance Estimation", IEEE 2020 Zooming Innovation in Consumer Technologies Conference (ZINC).

[8] Nouran Khaled, Shehab Mohsen, Kareem Emad El-Din, Sherif Akram, Haytham Metawie, Ammar Mohamed, "In-Door Assistant Mobile Application Using CNN and TensorFlow", IEEE 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE).

[9] Roshan Rajwani, Dinesh Purswani, Paresh Kalinani, Deesha Ramchandani, Indu Dokare, "Proposed System on Object Detection for Visually Impaired People", International Journal of Information Technology (IJIT) – Volume 4 Issue 1, Mar - Apr 2018.

[10] Miss Rajeshvaree Ravindra Karmarkar, Prof.V.N. Honmane, "Object Detection System for the Blind with Voice guidance", International Journal of Engineering Applied Sciences and Technology, 2021, Vol. 6, Issue 2.

[11] Rakesh Chandra Joshi, Saumya Yadav, Malay Kishore Dutta, Carlos M. Travieso-Gonzalez, "Efficient Multi-Object Detection and Smart Navigation Using Artificial Intelligence for Visually Impaired People", *Entropy* 2020, *22*, 941.

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", IEEE 2016 Conference on Computer Vision and Pattern Recognition (CVPR).

[13] J. Nasreen, W. Arif, A. A. Shaikh, Y. Muhammad and M. Abdullah, "Object Detection and Narrator for Visually Impaired People", *IEEE 2019 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*.

[14] Siva Ganga, Hema, "A Review on Text Recognition for Visually Blind People", International Research Journal of Engineering and Technology (IRJET), 2020, Volume 7, Issue 2.

[15] Hao Jiang, Thomas Gonnot, Won-Jae Yi, Jafar Saniie, "Computer Vision and Text Recognition for Assisting Visually Impaired People using Android Smartphone", *IEEE 2017 International Conference on Electro Information Technology (EIT)*.

[16] Samruddhi Deshpande, Ms. Revati Shriram, "Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People", International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) 2016.

[17] M. Murali, Shreya Sharma and Neel Nagansure, "Reader and Object Detector for Blind", International Conference on Communication and Signal Processing (ICCSP) 2020.

[18] S. A. Jakhete, P. Bagmar, A. Dorle, A. Rajurkar, P. Pimplikar, "Object Recognition App for Visually Impaired," IEEE 2019 Pune Section International Conference (PuneCon).

[19] A. Badave, R. Jagtap, R. Kaovasia, S. Rahatwad, S. Kulkarni, "Android Based Object Detection System for Visually Impaired," 2020 International Conference on Industry 4.0 Technology (I4Tech).

[20] Venkata Naresh Mandhala, Debnath Bhattacharyya, Vamsi B, Thirupathi Rao N, "Object Detection Using Machine Learning for Visually Impaired People", 2020 International Journal of Current Research and Review (IJCRR).

[21] Sanghyeon Lee, Moonsik Kang, "Object Detection System for the Blind with Voice Command and Guidance", IEIE Transactions on Smart Processing and Computing, vol. 8, issue 5, 2019.

[22] Trupti Shah, Sangeeta Parshionikar, “Efficient Portable Camera Based Text to Speech Converter for Blind Person”, International Conference on Intelligent Sustainable Systems (ICISS 2019).

[23] Eliganti Ramalakshmi, Dixitha Kasturi, Gouthami V, “Object Detector for Visually Impaired with Distance Calculation for Humans”, International Journal of Engineering and Advanced Technology (IJEAT), Vol 9 Issue 4, April, 2020.

[24] Fatima ZahraeAitHamou Aadi, Abdelalim Sadiq, “Proposed real-time obstacle detection system for visually impaired assistance based on deep learning”, International Journal of Advanced Trends in Computer Science and Engineering, Vol 9, Issue 4, 2020.

[25] P. Zhihao and C. Ying, "Object Detection Algorithm based on Dense Connection," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2019