INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

Artificial Vision for Visually Impaired Using Raspberry Pi

Darshan N[1], BhuvanG S[2], Lahari Shastri[3], Manoj G[4], Dr. Suma[5]

UG Students, Department of Electronics and Communication Engineering, Vidya Vikas Institute of Engineering and Technology, Mysore, India[1,2,3,4]

Associate Professor, Department of Electronics and Communication Engineering, Vidya Vikas Institute of Engineering and Technology, Mysore, India[5]

ABSTRACT: For those millions of us whose vision isn't perfect, there are glasses. But for those hundreds of thousands who are blind, devices that merely assist the eyes just aren't enough. What they need are alternative routes by which the sights of the world can enter the brain and be interpreted. "Artificial Vision for Visually Impaired" presents an electronic device which provides an audio output of the object detected for visually impaired and blind person. The system consists of Raspberry Pi, Camera, Push Button, and Speaker. The electronic device captures an image when a button is pressed. To inform the person about detected object, a voice output is passed through the speaker. The main objective is to provide a cost-effective system that allows blind individuals to understand theirsurroundings independently.

KEYWORDS: Artificial Vision, Object Detection, Raspberry Pi, Voice output.

I. INTRODUCTION

Humans can pick out objects in their line of vision in a matter of milliseconds. However, a visually challenged person is deprived of this. Visual impairment is a decreased ability to see to a degree that causes problems not fixable by usual means. The estimated number of people visually impaired in the world is 285 million, 39 million blind and 246 million having low vision as stated by World Health Organization (WHO). It's very difficult for them to every time perceive outside world.

This study is to enable these visually challenged people hear what they cannot see. Earlier work in object detection has had its application in areas like optical character recognition (OCR), self-driving cars, tracking objects, face detection, face recognition, image retrieval, security and surveillance. It has also been extended to help visually challenged people to recognise the objects. However, these detections have been carried out by developing extensive neural networks like CNN and numerous hidden layers. This decreases the speed of image retrieval and processing to a considerable amount. Other similar works are based on the local feature's extraction concept. The simulation results using SFIT algorithm and key points matching showed good accuracy for detecting objects.

The proposed work focuses on using The You Only Look Once (YOLO) algorithm for detecting the objects along with their position and uses the pytts API to give voice feedback. The algorithm runs through a variation of an extremely complex Convolutional Neural Network architecture called the Darknet. Previously, classification-based models were used and only high scoring regions of the image were considered as a detection and they could be very time-consuming. Instead,
YOLO is regression-based. Predicting classes and bounding boxes for the whole image is done quickly in one run of the algorithm, so that the predictions are informed by the global context in the image. Accuracy is achieved by setting the threshold and increasing the iterations.
The application uses technologies like machine learning, image processing.

Machine Learning: Machine learning is an application of artificial intelligence (AI) which provides a system which has the ability to automatically learn itself and improve from its experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and

make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Image Processing: Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing basically includes the following three steps:

- Importing the image to the system.
- Analysing and manipulating the image.
- Output in which result can be altered image or report that is based on image analysis.

II. PROPOSED SYSTEM

The system is implemented in a Raspberry Pi which detects various objects in real-time.
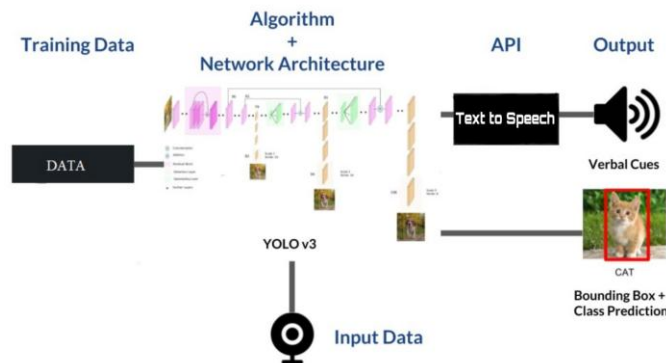

Fig.1 : System Overview

A. Hardware and Software Requirements:

i. Hardware requirements:
- Raspberry Pi Model 3B+
- Raspberry Pi Camera V2
- Speaker
- SD Card
- Push Button

ii. Software requirements:
- Raspberry Pi OS
- OpenCV
- Python
- Darknet
- PyTorch
- Python text-to-speech

B. Implementation

Input data is fetched from the webcam that captures the object. Feature Extraction of the image is done in the next step, where the splitting of frames into cells takes place along with Blob generation (stores images as bytes) and Bounding box construction.

The third step involves Object Classification where the confidence (X, Y, W, H) is updated and the relative position is found – Mid, Top, Bottom. Once the features have been extracted and the objects are classified, the corresponding text description are sent to the Pytts Library where the text is analysed and the letter to sound conversion takes place.

The output is predicted correctly from the results of the classification and the object is recognized with the audio.

The following gives a brief description about each module in the proposed model with Voice Feedback.

1. Image Capture: The first step in the working of the application is Image Capturing. Image Capturing is the process of obtaining images from a video which have to be converted into frames. Live stream is captured with a camera. OpenCV provides a very simple interface to this. A video from the camera is captured, converted it into grayscale video and displayed. To capture a video, VideoCapture() object is created. Its argument is either the device index or the name of a video file.

2. Feature Extraction: Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process the Splitting of Frames, Blob Generation, Feed to YOLO Network, Bounding Box Construction and Filtering Weak Prediction. YOLO uses Non-Maximal Suppression (NMS) to only keep the best bounding box. The first step in NMS is to remove all the predicted bounding boxes that have a detection probability that is less than a given NMS threshold.

3. Object Classification: Object Classification is a classification problem which tends to classify different objects which could flowers, faces, fruits or any object we could imagine. After removing all the predicted bounding boxes that have a low detection probability, the second step in NMS, is to select the bounding boxes with the highest detection probability and eliminate all the bounding boxes whose Intersection Over Union (IOU) value is higher than a given IOU threshold. The database is updated with confidence and class. The object classification process includes the following:

- Updating of Bounding Boxes, Confidence & Class.
- Finding Relative Position.
- Class Identification with Confidence.
- Text Description.

4. Speech Synthesis: Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. It is the process of generating spoken language by machine on the basis of written input. Speech synthesis involves the process of text analysis and letter to sound conversion. Text Analysis is about parsing texts in order to extract machine-readable facts from them. The purpose of Text Analysis is to create structured data out of free text content. The process can be thought of as slicing and dicing heaps of unstructured, heterogeneous documents into easy-to-manage and interpret data pieces. The annotated text is converted into voice response and gives the basic positions of the objects in the person/camera's view.

III.   RESULT AND ANALYSIS

Application can detect and recognize various categories of objects that may be faced while walking, objects of daily use and appliances, different types of vehicles, food, etc. application is beneficial for the users from different sectors like Banking, Travel and Tourism, Food and Beverages, Education, etc.

In this the percentage of accuracy differs as the object is placed at different angles and also when we vary distance of the object from the camera. When object is placed closer to the camera the percentage will be near to 99.99% whereas when we vary the distance slight far away from the camera the percentage decreases slightly. So, we can say placement of the object from the camera can change the accuracy of processing an image.

When multiple objects will be there in the scenery, every object will be detected and at the end of the result the total number of objects will be given in numerical form, but as a result of audio output only the object which has the highest probability among all the objects will be given as an audio output. Remaining other objects will be counted and will be given in the result.

Fig: Detection of Objects

IV.   ADVANTAGES

- Image to text generation.
- Object Detection.
- Can be used as a translator

- Text to speech.

V. CONCLUSION

The proposed system helps blind persons to understand the surroundings that appear in front of the camera through the earphone. It also helps them to detect the obstacle to prevent them from dangerous situations Therefore, it will help to prevent possible accidents. Application can detect and recognize various categories of obstacles that may be faced while walking, objects of daily use and appliances, different types of vehicles, food, etc.

Application is beneficial for the users from different sectors like Banking, Travel and Tourism, Food and Beverages, Education, etc. Feedbacks taken from people working in banking sector prove the same point under consideration.

REFERENCES

[1] Abhijeet Mohanta, Shah Yash Jitendra, Khandelwal Nikita Dinesh, Wable Saurabh Suhas, Aruna K. Gupta "Application for the Visually Impaired People with Voice Assistant", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-6, April 2020.

[2] Tom M. Mitchell "Machine Learning". McGraw Hill Education 2017.

[3] Rafael C. Gonzalez and Richard E. Woods "Digital Image Processing". Pearson 2018.

[4] Aditya Raj, Manish Kanauji, Ajeet Bharti, Rahul Prasad, Namrata Singh, Ishan Bhardwaj "Model for Object Detection using Computer Vision and Machine Learning for Decision Making" International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 43, March 2019.

[5] Jayshree R Pansare, Aditi Gaikwad, Vaishnavi Ankam, Priyanka Karne and Shikha Sharma "Real - Time Text Reader" International Journal on Computer Applications 182(34):42-45, December 2018.

[6] OpenCV, https://opencv.org/

[7] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", University of Washington, April 2018. Mr. Vaibhav V. Govekar, Prof. Meenakshi A, "A Smart Reader for Blind People", IJSTE - International Journal of Science Technology & Engineering | Volume 5 | Issue 1 | July 2018 ISSN (online): 2349-784X.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING