

Visionary Eye for Visual Impairment

Ruchir Shah¹, Dhaval Tamboli¹, Ravindra Baria¹, Ajay Makwana¹, Kishori Shekokar², Dr. Sheshang Degadwala³

¹U.G, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

²Assistant Professor, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

³Associate Professor, Department of Computer Engineering, Sigma Institute of Engineering, Vadodara, Gujarat, India

ABSTRACT

Article Info

Volume 7, Issue 2

Page Number: 215-221

Publication Issue :

March-April-2021

Article History

Accepted : 05 April 2021

Published : 11 April 2021

In this Research paper, we have discussed a proposed system that can be a visionary eye for a blind person. A common goal in computer vision research is to build machines that can replicate the human vision system. For example, to recognize and describe objects/scenes. People who are blind to overcome their real daily visual challenges. To develop a machine that can work by the vocal and graphical assistive answer. A machine can work on voice assistant and take the image taken by a person and after an image processing and extract the result after neural networks. And it all happens in real-time.

Keywords : Computer Vision, Artificial Intelligent, Human-Computer Interaction, Visual Impairment, Text to Speech, Speech to Text, Real-Time.

I. INTRODUCTION

There are more than 50 million people live with visual impairment (blindness). They cannot do small daily tasks such as reading newspapers, cannot watch a movie, read the information on daily transportation, they face difficulty in the public area, looking in the mirror is among the daily challenges that they face on their routine. They have to ask questions about their surroundings to people. This makes them depend on other people.

Main goal for our research is to build an application that can replicate the human vision with more advanced attributes and quality. This can recognize objects and describe the scenes. Towards this aim, we introduce our project on developing application for assistive technologies with help of artificial

intelligence and computer vision. We hope this work will be helpful for those people who are blind and by using this they can eliminate their accessibility barriers.

II. SYSTEM

One of the biggest challenges for blind people is when they are moving outdoors where there are uneven terrains, obstacles such as cars, rocks, and holes and not knowing where to go without the ability to see. Traditionally, blind people use walking sticks or guide dogs to help them move around and they are limited to move only around familiar places.



Fig 1 – Our object detection application

We have created this application that can use the camera and detects objects in real time. User has to on the camera after opening the application. The camera will be turned on after giving the permission successfully. Then user has to start the object detection in application, so that the application will starts the object detection. it will detect more than one objects in frame. We have use SSD (Single Shot Detector) algorithm in this application. it has fast speed and 71.00% mAP (Mean Average Precision). YOLO is also very good algorithm when it comes to object detection, it has fast speed in output and 76.00% mAP.

These type of proposed systems follows on 2 aims either navigation or object detection. We have tried to make things in one with the help of our proposed systems a blind person not only can navigate around but also find useful information from objects they needs.

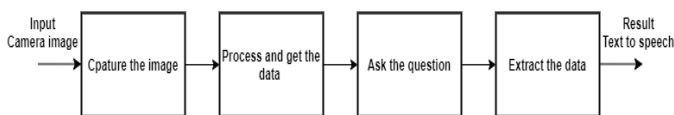


Figure 1.0

Fig 2 – Basic steps in object detection of application

A. Capture The Image

The user first opens a camera and takes a picture of an object to get the information. Then the process starts as follows.

1. The picture/image is taken.
2. The proposed system will check that if the image is taken properly or not i.e. blur, the object is not visible, etc.
3. If the image taken by the user is not in proper format then, the proposed system will tell the user to take a picture again.
4. Removes the noise in the image.
5. Process the picture in the proper frame and corrects the color.

B. User Ask Question

The user asks the question regarding the information that the user needs clearly and formally.

C. Processing Data

As per users question the process starts to defragmenting the words in speech given by user. Proposed system will focus on words where it can get more information. i.e. “what”, “where”, “find” “price”, “calories”, “amount”, “who”, “bill”, “is it”, “etc”.

D. Extracting The Data

The next step is to extract the data as per the user's needs. Which is done by Computer Vision.

E. Result

The result will be given after processing the image and question by the user. The result will be given to the user in an audio manner using text to speech approach.

III. CATEGORIZATION

The classification as well as object localization makes it one of the most challenging topics in the domain of computer vision. In simple words, the goal of this detection technique is to determine where objects are located in a given image called object localization and

which category each object belongs to, which is called object classification. Detecting instances of semantic objects of a certain class (such as humans, buildings, or vehicles) in digital videos and images. Object detection has proved to be a prominent module for numerous important applications like video surveillance, autonomous driving, face detection, etc. Feature detectors such as Scale Invariant Feature Transform and Speeded up Robust Feature are good methods that yield high-quality features but are too computationally intensive for use in real-time applications of any complexity. Based on the normalized corner information, support vector machine and back-propagation neural network training are performed for the efficient recognition of objects.

Here is a flowchart that shows the entire process in an abstract way.

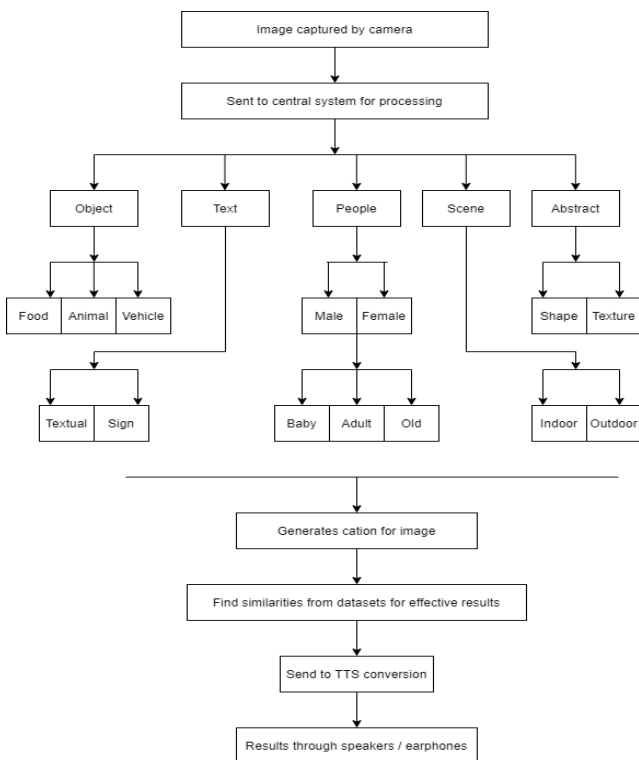


Fig – 3 Flowchart

It is a flowchart illustrating the method of operation of the application for the visually impaired. According to the present device, The camera

captures the images of the objects which is in real time. The images will be sent to the application and there using algorithms and mathematical functions the objects will be detected by application in collaboration with the different layers of processing. The output result is displays to user which can be heard and seen through the speakers and Monitor screen.

IV. ALGORITHMS AND TECHNIQUES

The modern world is enclosed with gigantic masses of digital visual information. To analyze and organize this devastating ocean of visual information image analysis techniques are major requisites. In particular, use would be methods that could automatically analyze the semantic contents of images or videos. The content of the image determines the significance in most of the potential uses. One important aspect of image content is the objects in the image. So there is a need for object recognition techniques.

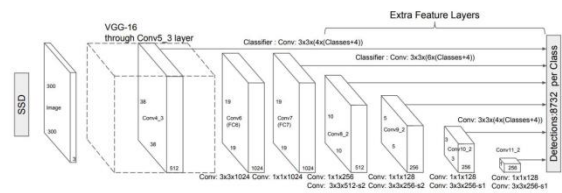


Fig 4 – Single Shot Detector (SSD) Architecture

Object recognition is an important task in image processing and computer vision. It is concerned with determining the identity of an object being observed in an image from a set of known tags. Humans can recognize any object in the real world easily without any efforts; on the contrary machines by itself cannot recognize objects. Algorithmic descriptions of recognition tasks are implemented on machines; which is an intricate task. Thus object recognition techniques need to be developed which are less complex and efficient.

Algorithms	Speed	mAP
R-CNN	Slow	66.00%
Fast R-CNN	Medium	66.90%
Faster R-CNN	Fast	69.00%
SSD	Fast	71.00%
YOLO	Fast	76.80%

Fig 5 – Object detection algorithm compared

The different algorithms are explained for suitable for different types of systems and scenarios.

Many successful approaches that address the problem of general object detection use a representation of the image objects by a collection of local descriptors of the image content. Global features provide better recognition. Color and shape features can also be used. Various object recognition techniques are presented in this paper. Difficulties may arise during the process of object recognition; a range of such difficulties are discussed in this paper.

1) SSD

The SSD Stands for “Single Shot Detector”. As per the name suggest Single Shot Detector (SSD) is a method for detecting objects in images using a single deep neural network. The SSD approach discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios. After discretizing, the method scales per feature map location. The Single Shot Detector network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. The advantage of SSD is that SSD eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. Easy to train and straightforward to integrate into proposed systems that require a detection component. SSD has competitive accuracy

to methods that utilize an additional object proposal step, and it is much faster while providing a unified framework for both training and inference.

2) YOLO

YOLO is a new and novel approach to object detection. It stands for "You Only Look Once". Prior work on object detection repurposes classifiers to perform detection. YOLO frames object detection as a regression problem to spatially separate bounding boxes and associated class probabilities.

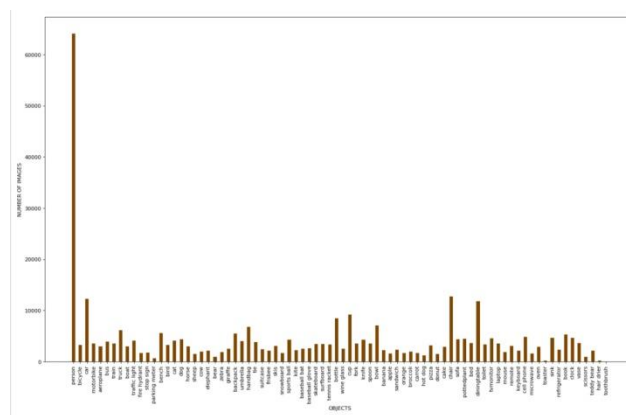


Fig 6 – Frequency distribution of dataset in YOLO

A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. Unlike sliding window and region proposal-based techniques, YOLO sees the entire image during training and test time so it implicitly encodes contextual information about classes as well as their appearance. Fast R-CNN, a top detection method, mistakes background patches in an image for objects because it cannot see the larger context. YOLO makes less than half the number of background errors compared to Fast R-CNN.

V. APPLICATION

Object detection is a computer vision technology that localizes and identifies objects in an image. Due to object detection's versatility in different uses, object detection has emerged in the last few years as the most commonly used computer vision technology. Object detection is useful in any scenario where computer vision is needed to localize and identify objects in an image. Object detection multiplies in settings where objects and scenes are more or less similar. The goal of feature extraction is to reduce a variable-sized image to a fixed set of visual features. Image Picture classification models are ordinarily developed utilizing solid visual component extraction strategies. Regardless of whether they depend on conventional PC vision draws near, such as, channel-based drew closer, histogram techniques, and so on, or profound learning strategies, they all have precisely the same objective.

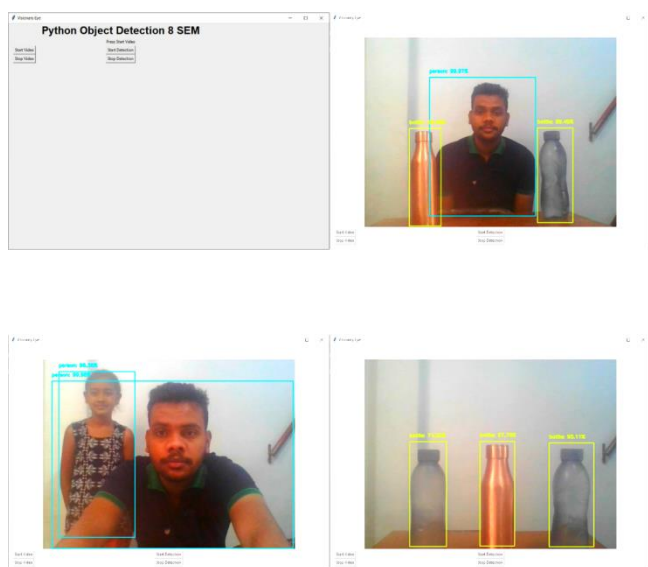


Fig 7 – Our proposed system application

In our application is a very simple. To use application for object detection the user will have to start the camera and then click on start video and then

detection on GUI interface. After that the application will start to detecting objects that comes in front of camera frame and all this happens in real time.

Navigation is a purposeful process which comprises the traveler updating their orientation and position of the travel task such as how to get from point A to point B. And can accomplish in three ways;

1. Speech to Text using Google API
2. Object Detection
3. Text to Speech

```

find computer
find
cotmtr
find
cotmtr
find computer
you are finding something
computer
[INFO) Loading model...
Model information obtained
[INFO] starting video stream.
[INFO] elapsed time: 3.50
[INFO] approx. FPS: 27.12
object found
computer found at 20 Degree left
    
```

Fig 8 – Running application finding result

The user gives input in the form of speech. Example: The user will say that, Find me a computer, Find bottle etc. The YOLO algorithm model detects the object in search through the camera in real-time. And then it gives the output. We modified it to tell us the degrees left or right at which object is present. Pytsx python text to speech library of python was used to give back the output in the form of speech. Example: Computer found at 90° degrees left, Bottle found ahead etc. 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 A text-to-speech (TTS) system converts normal Artificial or unified speech can be created by concatenating pieces of recorded speech that are stored in the database. Systems differentiate in the characteristics of the stored speech units; a

system that stores these data provide the largest output range but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Then again, an artificial speech can fuse a model of the vocal parcel and other human voice qualities to make a totally "engineered" voice output. The quality of a unified speech is judged by its similarity to the human voice and by its ability to be understood clearly. The nature of an AI voice is decided by its likeness to the human voice and by its capacity to be seen obviously. An intelligible text-to-speech program allows people with visual impairments or even reading disabilities to listen to written text easily.

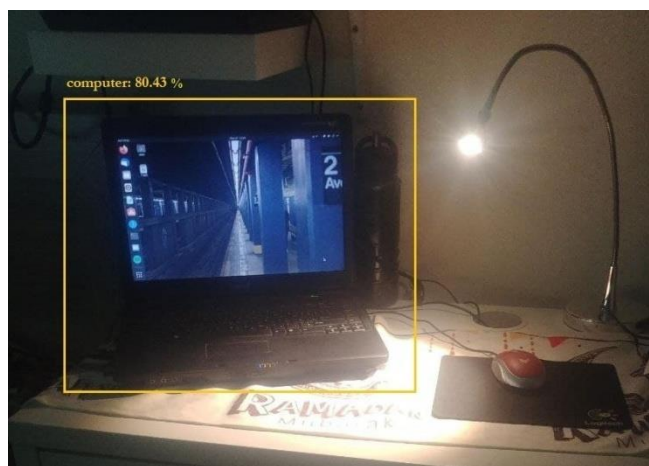


Fig 9 – Application detecting objects

The whole framework is at first on rest mode. By giving the "Wake up" order the framework gets initiated. The user can then give their voice query in the form of find me a bottle, find bed etc. A video stream starts using the imutils library of python for detecting. The object detection unit gets activated, where it matches the object against its set of predefined classes. Deep Neural Network (DNN) is used as the model. If object is found, it makes a box around the object. We have fetched the coordinates of the outlined box. For getting the user's viewpoint, we have considered the center of the bottom line of the box. If object is not present in its training dataset, the system returns object not found output. The

calculation of degrees of object has been gained by the mathematical formula: $\text{math.degrees}(\text{math.atan2}(\text{abs}(\text{centerX} - \text{user}_x) / \text{abs}(\text{centerY} - \text{user}_y)))$

The video stream stops. Finally, the text output is converted to speech by using the pyttsx library of python. Visually impaired people to navigate them indoors and outdoors environments. The most common sensor-based method for outdoor navigation relies on GPS, but it is not precise enough for the purposes of the present research, especially in indoor environments. This can identify and perceive different classifications of impediments that might be confronted while strolling, objects of the day by day use and machines, various kinds of vehicles, food, and so forth This is helpful for clients from various areas like Banking, Travel and Tourism, Food also, Beverages, Education, and so on Criticisms taken from individuals working in the financial area make a similar statement viable. A discovery application is expected to assist individuals with a visual impedance to discover all the more unequivocally where articles are situated through the proposed framework.

VI. CONCLUSION

We used the OpenCV library to detect objects with the laptop's front-camera. The detected objects were classified into different classes based on our trained model. This application has graphical user interface for easy to use. The possible future work for the application is to make improvements to the model. These include adding a distance measurer in the module or application such that the audio output can also tell the distance to the person. This can be in the format, "Objct_Name: 2m or 6.5 Ft".

VII. REFERENCES

- [1]. F. Puente-Mansilla, G. Boza-Quispe, G. Lapa-Velasquez, C. Matos-Avalos, and J. Rosales-Huamani, "A wearable UV sensor and accessible

- smartphone application for blind people," 2016 IEEE International Symposium on Consumer Electronics (ISCE), Sao Paulo, 2016, pp. 81-82, doi: 10.1109/ISCE.2016.7797382.
- [2]. R. A. Minhas and A. Javed, "X-EYE: A Bio-smart Secure Navigation Framework for Visually Impaired People," 2018 International Conference on Signal Processing and Information Security (ICSPIS), DUBAI, United Arab Emirates, 2018, pp. 1-4, doi: 10.1109/CSPIS.2018.8642718.
- [3]. A. Zeb, S. Ullah, and I. Rabbi, "Indoor vision-based auditory assistance for blind people in semi-controlled environments," 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), Paris, 2014, pp. 1-6, doi: 10.1109/IPTA.2014.7001996.
- [4]. B. Rajapandian, V. Harini, D. Raksha, and V. Sangeetha, "A novel approach as an AID for blind, deaf and dumb people," 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), Chennai, 2017, pp. 403-408, doi: 10.1109/SPPS.2017.8071628.
- [5]. M. Nassih, I. Cherradi, Y. Maghous, B. Ouriaghli and Y. Salih-Alj, "Obstacles Recognition System for the Blind People Using RFID," 2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies, Paris, 2012, pp. 60-63, doi: 10.1109/NGMAST.2012.28.
- [6]. A. A. Nada, M. A. Fakhr, and A. F. Seddik, "Assistive infrared sensor-based smart stick for blind people," 2015 Science and Information Conference (SAI), London, 2015, pp. 1149-1154, doi: 10.1109/SAI.2015.7237289.
- [7]. David A. Forsyth, Jean Ponce, "Computer Vision: A Modern Approach", Publisher: Prentice Hall Professional Technical Reference, ISBN:978-0-13-085198-7, March 2002
- [8]. S. Kumar KN, R. Sathish, S. Vinayak, and T. Parasad Pandit, "Braille Assistance System for Visually Impaired, Blind & Deaf-Mute people in Indoor & Outdoor Application," 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 2019, doi: 10.1109/RTEICT46194.2019.9016765.

Cite this article as :

Ruchir Shah, Dhaval Tamboli, Ravindra Baria, Ajay Makwana, Kishori Shekokar, Dr. Sheshang Degadwala, "Visionary Eye for Visual Impairment", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 2, pp. 215-221, March-April 2021. Available at doi : <https://doi.org/10.32628/CSEIT217254>
Journal URL : <https://ijsrcseit.com/CSEIT217254>