Visual aid for the disabled

Jay Kamdar (60001160022) Samruddhi Pai (60001160036) Shaival Parikh (60001160038)
Guided by: Prof. Darshana Sankhe

Introduction

Our project is an application in the field of computer vision. The aim of the project is to provide assistance to visually impaired people. The product first receives input from a camera and processes the image and detects various objects in it. Then by the use of neural classifiers and trained data sets, identifies and classifies the objects detected. Then the distance between the most important objects out of all is calculated. After these processing steps the objects that are closest to the person or that might have the potential to harm them are detected and their positions are informed to the person using a speaker. This enables the person to become aware of their surroundings and thus act accordingly.

Materials and methods

The language used for the major development of this project is Python (ver. 3.7). We have employed the following software tools:

1. Open source computer vision tool OpenCV(ver. 3.4.1)
2. TensorFlow(ver. 1.13.1) for implementation of object classification algorithm
3. Keras(ver. 2.15.1)
4. Pyttsx3 – text to speech converter

Hardware used in the project are:

1. Raspberry Pi
2. Camera
3. Speaker

Image processing has been used extensively in this project for object detection, classification and depth perception of the obstacles from the impaired person. Classification algorithm YOLO (You Only Look Once) employing Convolutional Neural Networks (CNN) has been used in order to recognize the obstacles. Depth perception had been tried with multiple approaches, distance measurement using the triangulation approach, depth sensing using stereo vision, and storing the approximate heights of different objects and creating a machine learning model to calculate the distance of the object using focal length and height of pixel as parameters. The output is given in terms of feedback actions by a speaker.

Results

The steps of execution of our project has been summarized in the flowchart below:

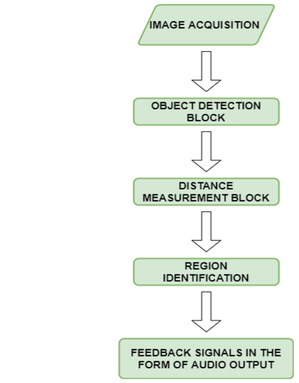

Image Acquisition is done by the web camera and the algorithm we use for object classification as well as depth perception uses RGB type of image.

Object recognition is done using YOLO algorithm which helps in classifying objects into 80 different object classes. Recognizing an object and identifying it is a very important step for our application since the feedback signal heavily depends on the type of obstacle in front of the human. Machine Learning algorithm of CNN is used to identify the object and map the features of the image to the corresponding features of the trained dataset.

Distance measurement has been one of the crucial steps of execution in the project. The approach used by us involved storing the average height of an object (in the COCO file) in a dataset and hence finding the distance by taking the focal distance and height of the pixel into consideration. The formula used for the same is:

$$\text{Distance} = \frac{f \times h}{h_p}$$

Where, f = focal length

h = height of object stored (real height)

h_p = height of the pixel (virtual height)

In accordance with the above formula, the relative distance and (x,y) coordinates of the object on the image are calculated in order to obtain the distance of the obstacle from the visually impaired person. The image below shows a photo taken from a bridge of a traffic scene below, the shell output on the left shows the objects recognized along with the distance and (x,y) coordinates.


```
>>>
 RESTART: C:\Users\Samruddhi Pai\Desktop\BE
Project\ObjectDetect\object_detection_img.py
bus, Distance = 47, x = 597, y = 337
car, Distance = 25, x = 139, y = 596
truck, Distance = 149, x = 104, y = 98
bus, Distance = 145, x = 284, y = 129
car, Distance = 92, x = 315, y = 205
truck, Distance = 192, x = 233, y = 37
car, Distance = 152, x = 495, y = 39
car, Distance = 152, x = 363, y = 43
car, Distance = 251, x = 181, y = 56
car, Distance = 148, x = 324, y = 72
car, Distance = 176, x = 408, y = 72
car, Distance = 176, x = 454, y = 71
motorbike, Distance = 108, x = 31, y = 89
car, Distance = 136, x = 578, y = 102
motorbike, Distance = 92, x = 233, y = 126
motorbike, Distance = 104, x = 556, y = 131
```

The above output is translated into feedback actions and given as voice output to the disabled person.

Conclusions

Since there is no technological and cost-effective product solution for the visually disabled, our project aims to fill the gap for this demand. We aim to serve a social cause in order to increase the efficiency and inclusion of the disabled people in the society. Through the use of various machine learning and computer vision algorithms, we have successfully implemented the vision aid project. By giving a voice output, we aim to provide a real-time feedback so that no additional help is needed in order to perceive the surroundings. By employing various algorithms and approaches for depth sensing, we get a deeper understanding in the field of depth perception and monocular vision. We further understand the drawbacks of a sensor- based approach and the possibilities of innovation in computer vision for perceiving surroundings.

Literature Cited

- Samartha Koharwal, Samer Bani Awwad, Aparna Vyakaranam “Navigation System for Blind - Third Eye” International Journal of Innovative Technology and Exploring Engineering (IJITEE)
- Joseph Redmon, Santosh Divvala, Ross Girshick , Ali Farhadi “You Only Look Once: Unified, Real-Time Object Detection” University of Washington, Allen Institute for AI , Facebook AI Research
- DRISTI: Dynamic Ranging by Image Segmentation and Terrain Imaging- Samanth S Mokshagundam, Suhas G, Suhas T Shanbhogue, Suraj S

Acknowledgments

We thank Prof. Darshana Sankhe for her valuable assistance throughout the duration of the project. We would also like to thank our HOD Prof. Prasad Joshi for extending his support towards the completion of the project and for the timely encouragement.

Dated: 28th March 2020