

Object Detection and Recognition for Blind Assistance

Rohini Bharti¹, Ketan Bhadane², Pranav Bhadane³, Ajjay Gadhe⁴

^{1,2,3}Students, Dept. of Electronics and Telecommunication Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India

⁴Assistant Professor, Dept. of Electronics and Telecommunication Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India

Abstract – Vision is one of the essential human senses and it plays the most important role in human perception about surrounding environment. The blind people face difficulties in safe and independent mobility, issues of communication and access to information. We propose an object recognition algorithm, and an assistive system which is very useful for their safety, quality life and freedom from other person all the time. The purpose of this system is to make the visually challenged able to take decisions independently based on the audio output from this model which can be trained according to their needs. Convolutional Neural Network is designed to achieve better accuracy. It is implemented using multimedia processor equipped embedded board and OpenCV.

Key Words: Object Recognition Algorithm, Convolutional Neural Network, Training, Accuracy, OpenCV

1. INTRODUCTION

The visually impaired people face a lot of problems in their day to day life. Unlike a normal sighted person they are unable to view their surroundings. Hence, they have limitation in almost every aspect of their lives like mobility, decision making etc. They have to face difficulties in accessing information and communicating the same. Thus, their personal, social as well as professional life is affected. Related works show that visual substitution devices accept input from the user's surroundings, decipher it to extract information about entities in the user's environment, and then transmit that information to the subject via auditory or tactile means or some combination of these two. They can only be used to recognize simple patterns and cannot be used as tools of substitution in natural environments. Also, they don't identify objects (e.g. whether it is a table or chair) and they have in some cases a late detection of small objects. Among the problems in object identification, we note the redundancy of objects under different conditions: the change of viewpoint, the change of illumination and the change of size. We have the concept of intra-class variability (e.g. there are many types of chairs) and the inter-class similarity (e.g. television and computer).

For this reason, we are interested in a fast and robust computer vision application to recognize and locate objects. Thus, it is important to design a system based on the recognition and detection of objects to meet the major

challenges of the blind in three main categories of needs: displacement, orientation and object identification.

Thus we propose a system with following feature:

Object Detection:

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class such as humans, buildings or cars in digital images and videos. Well-researched domains of object detection include face detection and pedestrian detection. Object detection has applications in many areas of computer vision, including image retrieval and video surveillance.

Object Recognition:

Human beings are highly dependent on vision sensor for daily tasks such as walking, eating, finding food, searching, driving vehicle, reading book etc., object recognition is the core algorithm in most of vision related task. The approach to solve object recognition problem is using Deep learning based methods where a deep neural network is designed and then train it. Deep neural network architecture can be trained in unsupervised learning where network is trained with unlabeled data and it takes very less training time. In our object recognition method, we are using convolutional neural network (CNN) and feature extraction followed by pooling stage and then the combined features vector is used to train the classifier.

1.1 Objective

- To develop a system using object and colour recognition which could enhance the capability of visually impaired people.
- To implement an object recognition system which adapts and continuously learns from user inputs.

1.2 Problem Statement

To design automated smart glasses capable of performing environment adaptive object detection and accurate real time recognition henceforth enhancing the capability of visually impaired people.

2. LITERATURE SURVEY

Real-Time Visual Recognition with Results Converted to 3D Audio

Real-time image recognition with YOLO and 3D location of the objects is estimated from the location and the size of the bounding boxes from the detection algorithm. Although YOLO provides real time, portable solution but it proves to be resource costly as the method exploits a platform that utilizes portable cameras, fast HD video link and powerful server to generate 3D sounds.

Integrating Computer Vision Object Recognition with Location Based Services for the Blind

In this paper it is proved that it's extremely time consuming and also known to fail the detection when there is a mismatch between the template's size and pose when compared to the actual size and of the object in the image. Another method of K-means clustering proves to be a simple approach on CIFAR-10 and NORB datasets but requires large number of features. Similarly, the Geographic Information System (GIS), template matching technique although is a novel approach for location based assistance but suffers from faulty detection and time lag.

3. PROPOSED METHOD

3.1 Block diagram

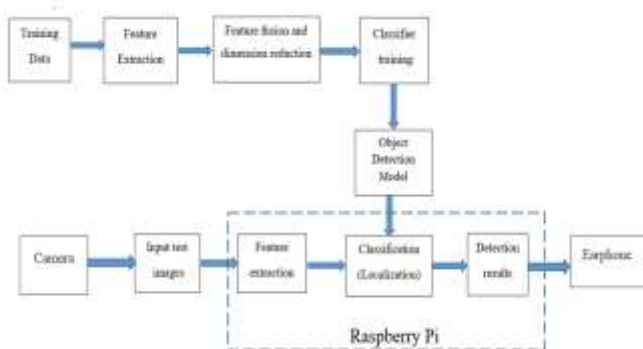


Fig - 1: Block diagram

Here we are going to capture the image of commonly used products through a mini camera mounted on spectacles. After acquisition of image it has to be pre-processed and compressed.

Various daily use objects' images are used to train the model. It is trained by performing feature extraction on the image to obtain the required pattern in the image. Followed by feature fusion and dimension reduction to compress the image for reliable and real time performance. Then this dataset is used to train the classifier. Comparing the

performance of various classifiers we select the optimum one, and thus the object recognition model is achieved. Now any test image may be given to this model which will be classified into one of the classes the model has been trained into.

3.2 Hardware Implementation

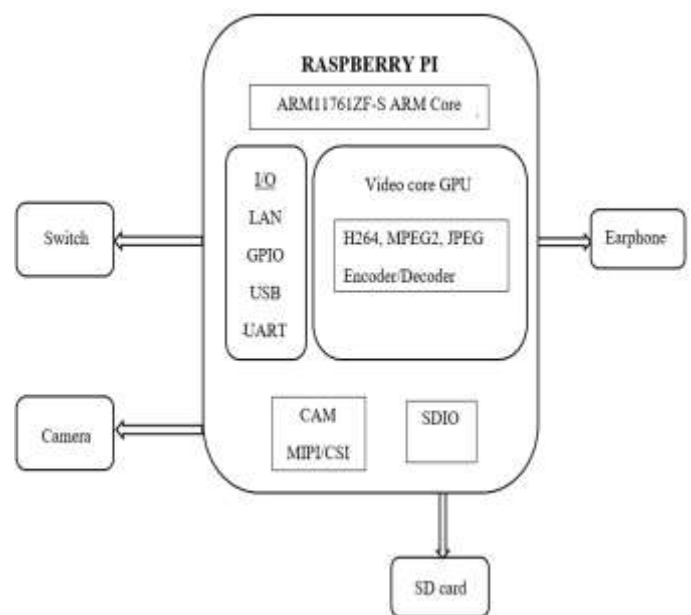


Fig - 2: Interfacing diagram

Raspberry Pi is single board computer which offers many useful functionalities in a tiny form factor. Thus it can be used for portable applications such as ours. The boards have one to four USB ports. For video output, HDMI and composite video are supported, with a standard 3.5 mm tip-ring-sleeve jack for audio output. Lower-level output is provided by a number of GPIO pins, which support common protocols like I²C. The B-models have an 8P8C Ethernet port and the Pi 3 and Pi Zero W have on-board Wi-Fi 802.11n and Bluetooth. It also provides a wide array of open source software development tools. And finally Tensorflow, an open source software library for high performance numerical computation supports Raspbian, a Raspberry Pi operating system. This makes Raspberry Pi a perfect fit for the system.

3.3 Software Implementation

1. Data Preparation is done by running histogram equalization on all training images for adjusting the contrast of images.
2. Model definition :
 - i) Input data layer - This is the first layer of the network where input image is specified. Image is a 3D matrix of pixels divided in 3 channels (red, green, blue).
 - ii) Convolutional layer - In this layer we define the parameters for the first part of convolutional layer

which is filter. The need of this layer is to extract primitive features such as edge detection, sharpening, Gaussian blurring, normalization etc.
 iii) Pooling layer - This layer is essential for dimensionality reduction. Here we are using MAXPooling.
 iv) Classifier layer - In this layer we calculate the probability values of the object belonging to one of the trained classes. Then the class with maximum probability is selected and that string is converted into audio.

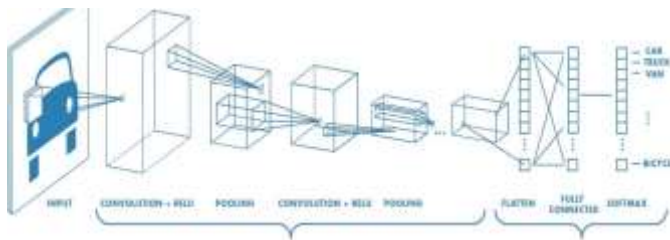


Fig - 3: Layers of Convolutional Neural Network

4. CONCLUSIONS

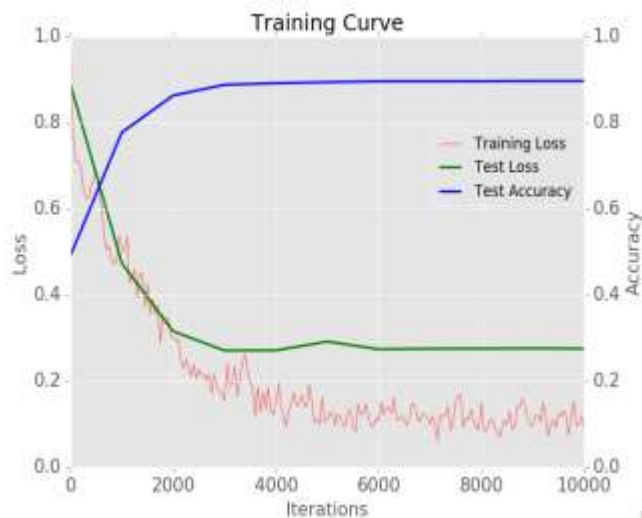


Fig - 4 : Loss vs Accuracy plotted for number of iteration

The model achieved a validation accuracy of 90%, and it stopped improving after 3000 iterations. We have trained our model for the following classes Aeroplane, Bicycle, Bird, Bottle, Bus, Car, Cat, Chair, Dining table, Dog, Motorbike, Person, Plant, Sofa, Train, Monitor.



Fig - 5 : Object detected - bottle

REFERENCES

1. Rui (Forest) Jiang and Qian Lin, "Let Blind People See: Real-Time Visual Recognition with Results Converted to 3D Audio," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
2. Shuihua Wang and Yingli Tian, "Camera-Based Signage Detection and Recognition for Blind Persons," International Conference on Computers for Handicapped Persons, 2012.
3. Vítor M Filipe, Paulo Costa and Hugo Fernandes, "Integrating Computer Vision Object Recognition with Location Based Services for the Blind," 8th International Conference, UAHCI 2014.