Navigation Aid for the Blind and the Visually Impaired People using eSpeak and Tensor Flow

Nishkala H M, Anu S H, Ashwini B V, Kavya C M, Monika B S

Abstract: Applications of science and technology have made a human life much easier. Vision plays a very important role in one’s life. Disease, accidents or due some other reasons people may lose their vision. Navigation becomes a major problem for the people with complete blindness or partial blindness. This paper aims to provide navigation guidance for visually impaired. Here we have designed a model which provides the instruction for the visionless people to navigate freely. NoIR camera is used to capture the picture around the person and identifies the objects. Using earphones voice output is provided defining the objects. This model includes Raspberry Pi 3 processor which collects the objects in surroundings and converts them into voice message, NoIR camera is used detect the object, power bank provides the power and earphones are used here the output message. TensorFlow API an open source software library used for object detection and classification. Using TensorFlow API multiple objects are obtained in a single frame. eSpeak A Text to Speech synthesizer (TTS) software is used to convert text (detected objects) to speech format. Hence using NoIR camera video which is captured is converted into voice output which provides the guidance for detecting objects. Using COCO model 90 commonly used objects are identified like person, table, book etc. and can also identify the location of an image with a single frame. The classification reference be made of class name and class id of the identified things. Text to Speech Synthesizer (TTS) is a software called eSpeak, which changes the text file to audio. This Raspberry Pi system was easily transportable and it can be comfortably carried by the blind people.

Keywords: COCO model, eSpeak, NoIR camera, Raspberry Pi 3, TensorFlow API

I. INTRODUCTION

According to the human sense organs Vision plays a very vital role in the human’s life. Some people face this vision problem by birth or due to some road incidents. In the day to day life the visually impaired people face many difficulties in the vision senses. For the visually impaired people it was hard to confess the things in the everyday life. For this problem this model helps out the visually impaired people in the things identification and gives some hearing details about the things which are identified.

For real time object observation and recognition this smart cap helps the visually impaired people to guide alone themselves. The Raspberry Pi processor which is in the advanced system, the processor was filled accompanied by the instruction of Convolutional Neural Network model (CNN), which was advanced by make use of TensorFlow. The NoIR camera is connected by the processor and that processor was encrypted in Python. The image was captured by NoIR camera in live time and it provides the Raspberry Pi processor process the image. The COCO model is used to identify and differentiate the thing by the use of Python code. If the NoIR camera captures the image and that image was covered by squares, the camera detects the image with accuracy and name of the image. The text file which consists of the classification of the identical objects. The classification reference be made of class name and class id of the identified things. Text to Speech Synthesizer (TTS) is a software called eSpeak, which changes the text file to audio. This Raspberry Pi system was easily transportable and it can be comfortably carried by the blind people.

II. RELATED WORKS

Many models have been invited to guide the visually impaired people. Most related model of them is intelligent electronic eye for blind people [1]. Using image and obstacle sensors visual information about the surrounding is gathered. Required voice output is obtained by AVR microcontroller. By solar photo voltaic module, piezoelectric source and by body temperature electricity which is developed to provide electric power to this [2]. Here for object identification and distance calculation ATmega2560 based Arduino Mega 2560 is used. Traffic signals status can also be analyzed in this model. Computer vision guidance for indoor navigation of visionless people is explained in this paper [3]. Mobile apps and remote handling computer are used for indoor navigation to help the blind.

III. PROPOSED SYSTEM

This model works on the basis of TensorFlow and a software called eSpeak which convert text to speech format. We can arrange many classes which is occurred in an image and can also identify the location of an image with a single object detection model.

With this a bounding box will appear which indicates framing the object. This project has the capacity of detecting the things that appear in 90 various classes.

Initially the raspberry pi processor should be powered to start the working of the system.
Thus, the web camera is initialized which is gathered in USB ports of pi. With the help of NOIR camera real time video is captured which is further converted into a group of frames with the help of python rules. Here ‘ssd_mobilenet_v1_coco’ is used which is a fastest and simplest model that is used for the detecting an object. Here TensorFlow is allowed to detect the different things that occur in an image. Detection graphs and weights are used to identify the different objects which are present in an image.

In output the box can be seen which represents the image part in which a specific object is detected and the percentage represents the confidence level of every object and class name. It is displayed whenever the raspberry pi processor is subjected to interface with a display system.

eSpeak, a software is used to convert the text format like class tag, percentage to voice message. Objects in the picture are defined by voice message through earphones using audio jack of raspberry pi.

**FIG. 1: SCHEMATIC DIAGRAM OF THE PROPOSED SYSTEM**

**FIG. 2: STRUCTURAL OUTLINE OF THE SYSTEM**

**IV. SYSTEM DESCRIPTION**

This section focusses on the software that are used in the designed model. Multiple objects are identified and classified in a particular picture by machine learning model with the help of TensorFlow. Image processing is done by OpenCV and eSpeak that are used to convert text message to voice output.

**A. TensorFlow API**

TensorFlow API is a Python friendly open source library which is widely used for object identification [4]. By applying the COCO (Common Objects in Context) dataset API is trained. It consists of more than 300 thousand pictures of 90 frequently used things [5]. Some example objects are shown in Fig. 3.

COCO is a large-scale object identification, distribution, and captioning dataset. Five various models for object identification are implemented by API [6] as shown in Table-I.

For object identification ssd_mobilenet_v1_coco dataset is used. SSD (Single Shot multibox Detector) [7] is used to take one single shot to detect multiple objects in a picture. Where SSD is faster in object identification when compared with R-CNN [8]. Mobilenet architecture that are more suitable for mobile embedded based vision applications. It is used for classification and recognition of objects. By combining SSD and mobilenet object detection can be done.

**B. OpenCV**

OpenCV (Open source Computer Vision) is a library of c/c++ function usually aims on real-time computer vision which supports rapid execution. It results in extra usage of time and image processing resource and limited in interpreting [9].

**C. eSpeak**

eSpeak is a compact, open source, speech synthesizer software for Windows, Linux etc. It uses a formant synthesis method bringing different languages in a small size. Speech Synthesis MarkUp Language (SSML) is supported by eSpeak that are based on markup language for applications of speech synthesis. Some modifications in essential aspects of voice like frequency, pitch range, and effects such as echo, whisper etc [10] are supported by eSpeak.

**FIG. 3: COCO DATASET**
Table- I: Various dataset for Object identification

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Speed</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssd_mobilenet_v1_coco</td>
<td>Fast</td>
<td>Boxes</td>
</tr>
<tr>
<td>ssd_inception_v2_coco</td>
<td>Fast</td>
<td>Boxes</td>
</tr>
<tr>
<td>rfca_resnet101_coco</td>
<td>Medium</td>
<td>Boxes</td>
</tr>
<tr>
<td>faster_rcn_resnet101_coco</td>
<td>Medium</td>
<td>Boxes</td>
</tr>
<tr>
<td>faster_rcn_inception_resnet_v2_atrous_coco</td>
<td>Slow</td>
<td>Boxes</td>
</tr>
</tbody>
</table>

V. RESULTS AND ANALYSIS

The figure represents the detected objects with accuracy. In a single picture one or more objects can be identified like person, laptop, chair as shown in Fig. 4. It is not able to identify the objects which are not included in the COCO dataset and also there occurs fault in object identification where the system misleads the object name. These are the major drawbacks of the model.

VI. CONCLUSION

This is a simple model in which Raspberry Pi converts the visually captured picture using NoIR camera into voice output. Blind people have to simply wear it without any particular skills of operating it he/she has to only supply power to the device. Proposed model is reasonable and customizable. It is real-time model that guides the blind by providing voice output about the surrounding environment helping to navigate more safely.

This simple architecture is very useful for navigation for the blind. Objects in the picture can be recognised by object detector. SSD mobilenet which is trained by COCO model can identify 90 different classes of objects. Objects can be increased by training the model by showing the image in different angles. As face detections is also included family members and friends can be easily identified by the blind people.

REFERENCES


AUTHORS PROFILE

Nishkala H M
Assistant Professor in the department of CS&E, SJMIT, Chitradurga. I have pursued my M.Tech in CS&E from VTU in the year 2016. Area of interest includes Machine Learning Techniques and IoT.

Anu S H
Under graduated student B.E, completing graduation in the year 2020, in the department of computer science and engineering, SJM Institute of Technology, Chitradurga, India.
Navigation Aid for the Blind and the Visually Impaired People using eSpeak and Tensor Flow

Ashwini B V
Under graduated student B.E, completing graduation in the year 2020, in the department of computer science and engineering, SJM Institute of Technology, Chitradurga, India.

Kavya C M
Under graduated student B.E, completing graduation in the year 2020, in the department of computer science and engineering, SJM Institute of Technology, Chitradurga, India.

Monika B S
Under graduated student B.E, completing graduation in the year 2020, in the department of computer science and engineering, SJM Institute of Technology, Chitradurga, India.