Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018)
IEEE Xplore Compliant Part Number:CFP18N67-ART; ISBN:978-1-5386-2456-2

SMART CAP – WEARABLE VISUAL GUIDANCE SYSTEM FOR BLIND

Nishajith.A
Department of ECE
LBS Institute of Technology for Women
Thiruvananthapuram,India
nishajith1996@gmail.com

Nivedha.J
Department of ECE
LBS Institute of Technology for Women
Thiruvananthapuram, India
nivedhavinayak@gmail.com

Shilpa.S.Nair
Department of ECE
LBS Institute of Technology for Women
Thiruvananthapuram,India
shilpasnair96@gmail.com

Prof.Mohammed Shaffi.J
Assistant Professor
Department of ECE
LBS Institute of Technology for Women
Thiruvananthapuram,India
shaffitvm@gmail.com

Abstract—Science and technology always try to make human life easier. The people who are having complete blindness or low vision faces many difficulties during their navigation. Blindness can occur due to many reasons including disease, injury or other conditions that limit vision. The main purpose of this paper is to develop a navigation aid for the blind and the visually impaired people. In this paper, we design and implement a smart cap which helps the blind and the visually impaired people to navigate freely by experiencing their surroundings. The scene around the person will be captured by using a NoIR camera and the objects in the scene will be detected. The earphones will give a voice output describing the detected objects. The architecture of the system includes the processor Raspberry Pi 3, NoIR camera, earphones and a power source. The processor collects the frames of the surroundings and convert it to voice output. The device uses TensorFlow API, open-source machine learning library developed by the Google Brain Team for the object detection and classification. TensorFlow helps in creating machine learning models capable of identifying and classifying multiple objects in a single image. Thus, details corresponding to various objects present within a single frame are obtained using TensorFlow API. A Text to Speech Synthesiser (TTS) software called eSpeak is used for converting the details of the detected object (in text format) to speech output. So the video captured by using the NoIR camera is finally converted to speech signals and thus narration of the scene describing various objects is done. Objects which come under 90 different classes like cell phone, vase, person, couch etc are detected.

Keywords: Raspberry Pi 3, TensorFlow API, TTS, eSpeak, NoIR camera

I. INTRODUCTION

Vision is one of the very essential human senses and it plays the most important role in human perception about surrounding environment [1]. The blind people face a lot of problem in their daily life. Some face this problem from the time they are born while others due to accidents or of different causes they encountered in this life. For blind persons it will be really difficult to recognize a product of their daily use. This smart cap aims to help the blind in object detection and to provide an audio information about the object detected.

The system helps the blind to navigate independently using real time object detection and identification. The proposed system consists of a Raspberry Pi-3 processor which is loaded with a pre-trained Convolutional Neural Network model (CNN) developed using TensorFlow. The processor is connected to a NoIR camera. The processor is coded in python. The NoIR camera captures the image in real time and will be provided to the Raspberry Pi-3 processor for processing it. The python code uses the COCO model to detect and classify the objects. It will draw boundary boxes around the detected and will also show the category index of the object. The category index of the detected objects will be stored in a text file. The category index consist of the class name and class id of the detected object. The contents of the text file is converted to voice using the Text to Speech Synthesiser (TTS) software eSpeak. This system is portable and the user can easily carry it.

II. RELATED WORKS

Several systems had been proposed for helping the blind people. Here, we will just mention the most related ones to the theme of our system. One of these systems is called intelligent electronic eye for visually impaired people [9]. The visual data from the surroundings is collected by using image and obstacle sensors. It will be processed by an AVR microcontroller and necessary voice information is provided. The electric power for this unit is availed through solar photo voltaic module, piezoelectric source and also from electricity generated from body temperature. An ultrasonic sensor based system is proposed in [10]. In this paper, an ATmega2560 based Arduino Mega 2560 is used for the object detection and distance measurement. Objects/staircase detection and distance measurement is performed by this system. Also the status of traffic signals can be

identified in this system. In the paper Computer vision guidance system for indoor navigation of visually impaired people [11], it is mentioned that indoor navigation help can be provided to blind by using mobile application and the remote processing computer.

III. THE PROPOSED SYSTEM

A. Block Diagram

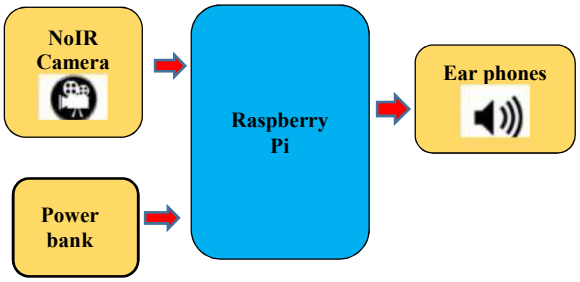

Fig.1 Block diagram of the proposed system

Smart Cap is based on TensorFlow and text to speech synthesizer software. With a single object detection model, it is possible to classify multiple classes present within an image and also it can specify the exact position of the image (if monitor provided) with a bounding box framing the object. The project is able to detect objects which come under 90 various classes.

The working of the system starts by suitably powering the raspberry pi processor. Thus, the web camera interfaced through one of the USB ports of pi is initialized. Real time video is captured using the NoIR camera which in turn is converted to a set of frames using python command. Here, we are using a simplest and fastest pre trained object detection model 'ssd_mobilenet_v1_coco' offered by TensorFlow to detect various objects present within our image. Identification of various objects present in the image is done using detection graphs and weights.

The output contains a box representing a part of the image where a particular object was detected, score representing level of confidence for each of the objects and class label. This can be displayed when the raspberry pi is interfaced with a display system.

By using the text to speech converter software, eSpeak, the text documents like class label, scores etc are converted to voice output. The earphones connected to audio jack of raspberry pi provides voice description corresponding to the objects present in the image.

B. Flowchart

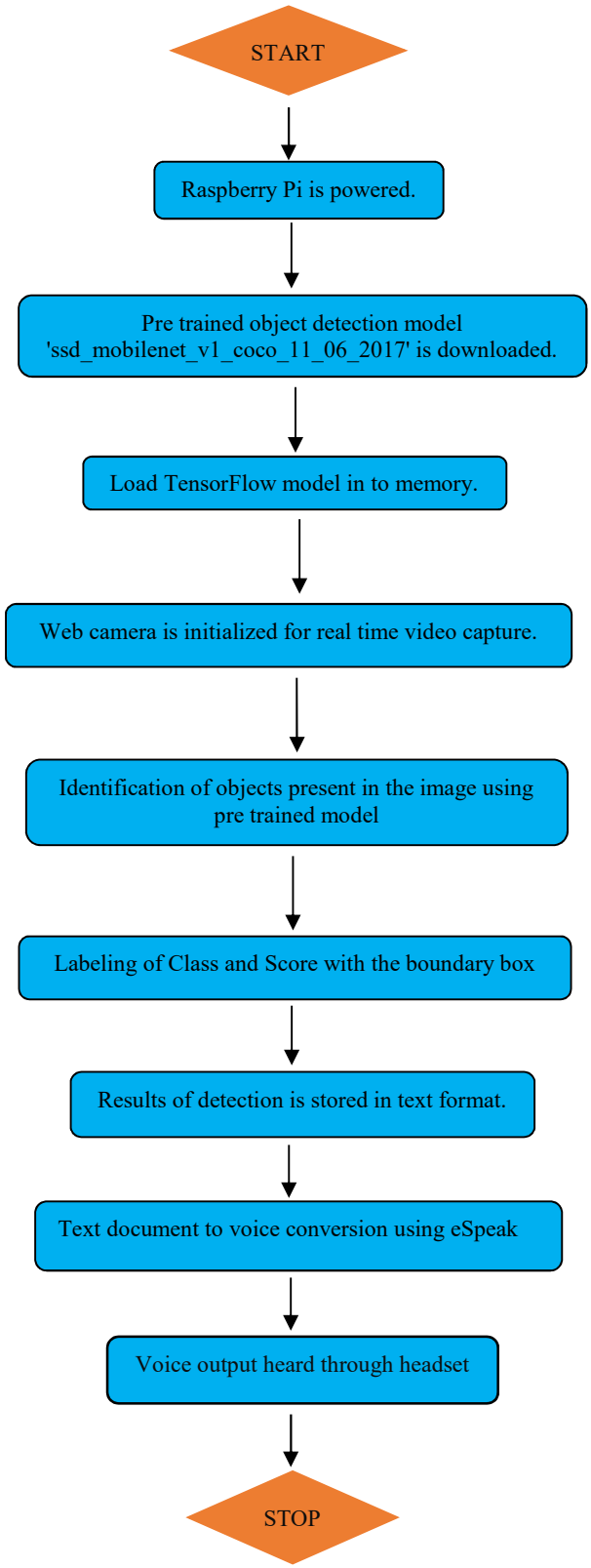

Fig.2. Flowchart showing working of the system

IV. SYSTEM DESCRIPTION

The different software used in the proposed system are explained below. The system uses TensorFlow API for the object detection and classification. OpenCV helps in the image processing operations and eSpeak performs the text to voice conversion.

A.TensorFlow Object Detection API

The TensorFlow API is widely used in the field of object detection. It is an open source software library for numerical computation using data flow graphs [8]. Second generation of Google artificial intelligence learning system got much attention and affirmation in the field of machine learning all over the world [2]. The API is trained by using the COCO dataset (Common Objects in Context). This is a dataset which contains 300k images of 90 most commonly found objects. Examples of objects includes:

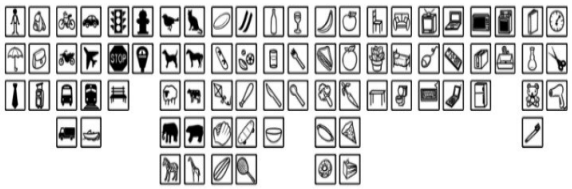

Fig.3. Objects in COCO model

The COCO model is achieved by gathering images of complex everyday scenes containing common objects in their natural context. The API provides 5 different models for object detection [4].

Model name	Speed	COCO mAP	Outputs
ssd_mobilenet_v1_coco	fast	21	Boxes
ssd_inception_v2_coco	fast	24	Boxes
rfcn_resnet101_coco	medium	30	Boxes
faster_rcnn_resnet101_coco	medium	32	Boxes
faster_rcnn_inception_resnet_v2_atrous_coco	slow	37	Boxes

Fig.4. Different models for object detection

In this paper, ssd_mobilenet_v1_coco model is used for the object detection. Single Shot Multibox Detector (SSD) [5] differs from the R-CNN based approaches by not requiring a second stage per-proposal classification operation [6]. This makes it fast enough for real-time detection applications. However, this comes with a price of reduced precision [6]. The "SSD with MobileNet" refers to a model where model meta architecture is SSD and the feature extractor type is MobileNet.

B. OpenCV

OpenCV uses c/c++ library functions which directly provides the computer with the machine language code and hence helps in faster execution. Using OpenCV results in more utilization of time and resources in image processing and less in interpreting [3].

C. eSpeak

eSpeak is a compact, open source, software speech synthesizer for Linux, Windows, and other platforms [7]. It can provide many languages since it use formant synthesis method. It supports Speech Synthesis Markup Language (SSML). eSpeak helps in changing the characteristics features of voice such as pitch range, frequency and add effects such as echo, whisper etc [7].

V. RESULTS AND ANALYSIS

The prototype of the proposed system is shown below.

Fig.5. Prototype of the system proposed

Fig.6. Output screen

The above figure shows the image of the output screen. It shows that in a single frame three objects like television, person and chair is detected. Anyway the

disadvantage of the system is that it has left many objects in the frame undetected. Also sometimes there will be error in the object classification that the system will misinterpret one object as another.

VI. CONCLUSION

The system has a simple architecture that transforms the visual information captured using a camera to voice information using Raspberry Pi. Unlike other systems available in the market, the subject needs only to wear the cap and doesn't require any particular skills to operate it. The proposed system is cheap and configurable. The person who uses it does not require any particular skill to operate it. Any blind or visually impaired person can use it simply since he/she has to only power up the device. The system helps in clear path indication and environment recognition. The device is a real-time system that monitors the environment and provides audio information about the environment making his/her navigation more safe and secure.

The smart cap will be really helpful for the blind people in their navigation. The object detection can be developed to count the number of objects in a scene. In this paper, the COCO model is used to train the SSD mobilenet which can detection only 90 classes of objects. The number of objects can be increased by training the model by ourselves. Face detection can be also incorporated so that the blind person can easily identify his/her family members and friends.

REFERENCES

[1] Hanen J abnoun I, F aouzi Benzarti I, Hamid Amiri I, "Object Detection and Identification for Blind People in Video Scene' 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)

[2] Nitin R. Gavai, Yashashree A. Jakhade, Seema A. Tribhuvan, Rashmi Bhattad "MobileNets for Flower Classification using TensorFlow", 2017 International Conference on Big Data, IoT and Data Science (BID),Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017

[3] Kruti Goyal , Kartikey Agarwal, Rishi Kumar "Face Detection and Tracking Using OpenCV", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[4] Tsung-Yi Lin, Michael Maire, Serge, Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Doll ´ ar, "Microsoft COCO: Common Objects in Context", arXiv:1405.0312v3 [cs.CV] 21 Feb 2015

[5] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C & Berg AC (2016) "Ssd: Single shot multibox detector. European conference on computer vision" , Springer: 21-37.

[6] Huang J, Rathod V, Sun C, Zhu M, Korattikara A, Fathi A, Fischer I, Wojna Z, Song Y & Guadarrama S (2016) Speed/accuracy trade-offs for modern convolutional object detectors. arXiv preprint arXiv:1611.10012 .

[7] http://espeak.sourceforge.net/

[8] https://www.tensorflow.org/.

[9] M.Maiti,P.Mallick,M.Bagchi,A.Nayek,"Intelligent electronic eye for visually impaired people",978-1-5386-2215-5/17/$31.00 ©2017 IEEE

[10] Namita Agarwal, Anosh Iyer, Sonalakshi Naidu, Snedden Rodrigues, "Electronic Guidance System For The Visually Impaired -A Framework", 2015 International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04 – 06, 2015, Mumbai, India

[11] Kabalan Chaccour, Georges Badr, "Computer vision guidance system for indoor navigation of visually impaired people", 2016 IEEE 8th International Conference on Intelligent Systems