RESEARCH ARTICLE OPEN ACCESS

People and Object Identification for visually impaired Persons

Dr.R.Dharmaprakash*, J.Karthik**, S.Balaji***, K.Harirammohan****
Department of Electrical and Electronics Engineering, Panimalar Institute of Technology, Chennai-123
Corresponding Author : Dr.R.Dharmaprakash

ABSTRACT
This paper presents an application that allows blind persons to identify the people and objects. This system transforms the visual information into audio information. The audio information is given to the visually impaired persons through the audio device about the people and objects in front of them. In this proposed system a computer vision technology based deep neural network of Google API is utilized. Raspberry Pi is used to implement artificial vision python language on the Open CV, Dlib and YOLO platform. The prototype model has been developed and tested with various cases. The system exactly identifies the objects and the people in front of the camera and the results are presented here.
Keywords – Computer Vision, Object identification, Raspberry pi, Visual assistive devices, YOLO

Date of Submission: 30-03-2019 Date of acceptance: 13-04-2019

I. INTRODUCTION

The contents of each section may be provided to understand easily about the paper. Visual impairment is a global problem with important socio-economic consequences [1-2]. The key facts on Blindness and vision impairment given by World Health Organization (WHO) are, (i). Globally, it is estimated that approximately 1.3 billion people live with some form of vision impairment. (ii). With regards to distance vision, 188.5 million people have mild vision impairment, 217 million have moderate to severe vision impairment, and 36 million people are blind [3]. (iii). with regards to near vision, 826 million people live with a near vision impairment [4]. (iv). globally, the leading causes of vision impairment are uncorrected refractive errors and cataracts. (v). approximately 80% of all vision impairment globally is considered avoidable. (vi). the majority of people with vision impairment are over the age of 50 years [5]. The high importance of assistive technologies for overcoming difficulties and challenges has been addressed in the WHO Disability and Rehabilitation Action Plan 2006 – 2011, and the 2014 – 2021 draft [6-7].

So that the assistive technology for the visually impaired and blind people is a research field that is gaining more focus. The field has a very relevant social impact on increasing aging and blind populations. Alexy Bhowmick and Shyamanta M Hazarika research results reveal that there has been a sustained growth in this assistive technology field [8].

The scope of it extends from the physiological factors associated with vision loss, to the psychological and human factors influencing orientation, mobility and information access for an individual with visual impairments, to the technological aspects in the development of rehabilitation devices (for mobility, way finding, object recognition, information access, entertainment, interaction, education), to medical interventions and prostheses including both current and cutting edge research [8].The development of assistive technology starts in 1950s and continues to develop. Various prototypes are developed with similar operating principles and are mostly they are disappointing and the user acceptance was low. The computerized Electronic Travel Aids (ETAs) were developed and are wearable, computerized and sonar-equipped system that enabled blind users to safely walk through unknown, obstacle-cluttered environments [9-10]. Various useful ETA prototypes have resulted from efforts to leverage technology for blind mobility. An ETA that has been a commercial success and awarded primary mobility aid Ultracane, which combined the long cane with ultrasonic sensors [11]. Over the last decade many advanced wearable and embedded assistive devices have been developed. But a comprehensive assistive device is still remains an elusive goal. In recent years computer vision capabilities coupled with the freedom and flexibility of mobile computing has shown great promise and resulted in many interesting developments for providing assistance to the blind [12].

The existing approaches suffer from the drawbacks such as requirement of several sensors, system not being portable and fail to do real time processing. In the proposed system, uses trained convolution neural network model for object by darknet and person identification model by Dlib. The proposed system intends to assist the blind by taking voice commands to detect objects using image processing and provide audio output to navigate and reach the required object. The system also intends to recognize the necessary objects and the known persons (those who already stored in database). The paper is organized into six sections. Section I provides a brief introduction about the proposed system. Section II presents inferences drawn from carrying the literature survey. The proposed system is discussed in Section III. The implementation and algorithm of the proposed system is presented in Section IV. Section V gives the snapshots of the results obtained with the proposed system. The paper is concluded with Section VI.

II. PEOPLE AND OBJECT IDENTIFICATION SYSTEM

The system uses speech commands as the user interface. A microphone is used to capture speech input. The obtained input is recognized using Google API. Also, it uses image processing as its primary technique to identify objects and Person identification. Video is captured by the web camera, from which the frames are extracted. The frames are preprocessed for better results. Preprocessed frame is given into the YOLO object detection model process the frame and output the name of that object present in that frame with it count if person was detected on that frame. Frame is enters into the person identification process.

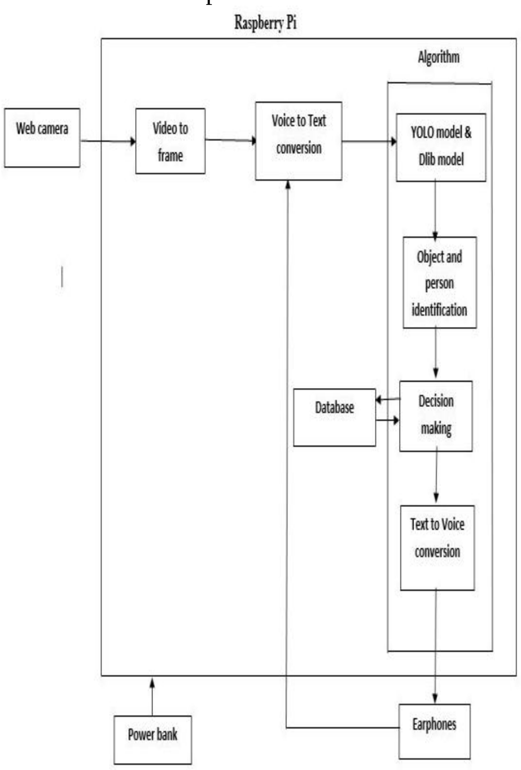

Fig.1. Block Diagram

Using Dlib model futures of the faces on that frame is extracted and that features are compare against the database and if matches name of that person is returned. Then the objects and name of the person identified in the frame is converted into audio message given to the User through earphone. Database of person identification is updated when the User request. The block diagram is shown in Fig. 1.

A. Raspberry Pi

The raspberry pi is a single-board computer based on Broadcom BCM2837 system on chip. It has a 1.2GHz CPU on board. It uses a 64-bit quad-core ARMv8 architecture based CPU. The raspberry pi version features 1GB of RAM. It uses an SD card to store the OS. It has USB port through which USB microphone is connected and speech input is given. It has a CSI port through which Pi camera is connected. The Raspberry PI features a 3.5 mm universal headphone jack for audio out. The Raspberry pi performs the task of taking video input, converting it to frames, does suitable image processing in Open CV and Dlib platform using Python language.

B. Web Camera

Logitech C310 web camera is used to take high resolution(720p) video, as well as still images. It has a resolution of 5 mega Pixel and 30 frames per second (fps). The output from the camera is fed to Raspberry Pi for further processing.

C. Open CV

It is a library of programming functions mainly aimed at real time computer vision. It is used for various applications such as augmented reality, gesture recognition, feature matching etc. With Open CV 3.3, we can utilize pre-trained neural networks with popular deep learning frameworks. It is imported by using the command “import cv2” in python.

D. Dlib

Dlib is a modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems. It is used in both industry and academia in a wide range of domains including robotics, embedded devices, mobile phones, and large high performance computing environments. Dlib's open source licensing allows you to use it in any application, free of charge. This library contains implementation of Deep metric learning model which is trained on a dataset of 3 million images and it is used to extract features of faces used in face recognition. It is imported by using the command “import dlib” in python.

E. YOLO Model ConvNet

The ConvNet architecture is shown in fig. 2. The network has 24 convolutional layers with 2 fully connected layers. The Convnet is to extract features from input images and the fully connected layers are to predict the probability of the boxes coordinates and confidence score. The accuracies of the predictions also depend on the architecture of the network. The loss function of the final output depends on the x, y, w, h, prediction of the classes and overall probabilities. The convolutional neural network of YOLO model is shown in Fig. 2.

F. Python

Python is a widely used high level programming language which has a dynamic type system and automatic memory management and supports multiple programming paradigms including object oriented, imperative, functional programming and procedural styles. Python is a light-weight programming tool that has many built-in functions and does not consume many resources while operating on the Raspberry pi.

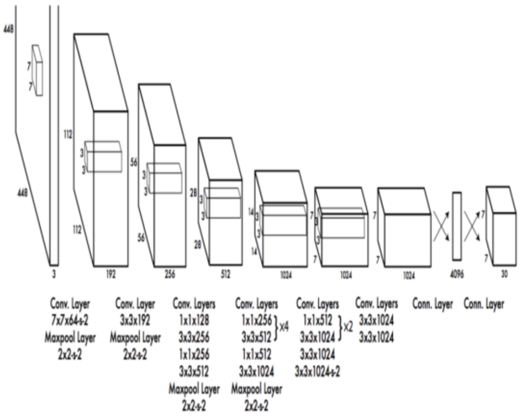

Fig. 2. Convolutional neural network of YOLO model

III. IMPLEMENTATION

Video captured by the web camera is converted into sequence of frames. Then its pass through the YOLO network model. Open CV is used to load the YOLO model. The model process the frame and output the name of the objects present in the frame with count. If person was detected on that frame. Frame is pass through the Deep metric model for features extraction and compared against database if matches found return the label (name of the person). Else ask the User for new face data entry to database with label (name of the person). At Finally the processed data (objects and person's name) converted text to voice given to the User through earphone.

A. You Only Look Once (YOLO) model:

You Only Look Once object detection model was trained by darknet. YOLO use a one-stage detector strategy. Its faster than R-CNN two-stage detectors. These algorithm treat object detection as a regression problem, taking a given input image and simultaneously learning the bounding box coordinates and corresponding class probabilities YOLO trained on the COCO dataset detect 80 different of objects. YOLO divides the input image into an **S×S** grid. Each grid cell predicts only **one** object. Each grid cell predicts a fixed number of boundary boxes. However, the one-object rule limits how close detected objects can be. For that, YOLO does have some limitations on how close objects can be. For each grid cell, it predicts B boundary boxes and each box has one box confidence score, it detects one object only regardless of the number of boxes B, it predicts C conditional class probabilities (one per class for the likeliness of the object class). Each boundary box contains 5 elements: (x, y, w, h) and a **box confidence score**. The confidence score reflects how likely the box contains an object (**objectness**) and how accurate is the boundary box. We normalize the bounding box width w and height h by the image width and height. x and y are offsets to the corresponding cell. Hence, x, y, w and h are all between 0 and 1. Each cell has 20 conditional class probabilities. The **conditional class probability** is the probability that the detected object belongs to a particular class (one probability per category for each cell). So, YOLO's prediction has a shape of (S, S, B×5 + C) = (7, 7, 2×5 + 20) = (7, 7, 30). YOLO is to build a CNN network to predict a (7, 7, 30) tensor. It uses a CNN network to reduce the spatial dimension to 7×7 with 1024 output channels at each location. YOLO performs a linear regression using two fully connected layers to make 7×7×2 boundary box predictions (the middle picture below). To make a final prediction, we keep those with high box confidence scores (greater than 0.25) as our final predictions.

Box confidence score $\equiv P_r(\text{object}) \cdot \text{IoU}$
Conditional probability $\equiv P_r(\text{class}_i|\text{Object})$
Class confidence score $\equiv P_r(\text{class}_i).\text{IoU}$
= box confidence score × conditional class probability

Where,

P_r(object) is the probability the box contains an object.

IoU is the IoU(intersection over union) between the predicted box and the ground truth.

$P_r(\text{class}_i|\text{object})$ is the probability the object belongs to class_i given an object is presence.

$P_r(\text{class}_i)$ is the probability the object belongs to class_i

The implementation flowchart of the proposed system is shown in fig. 3.

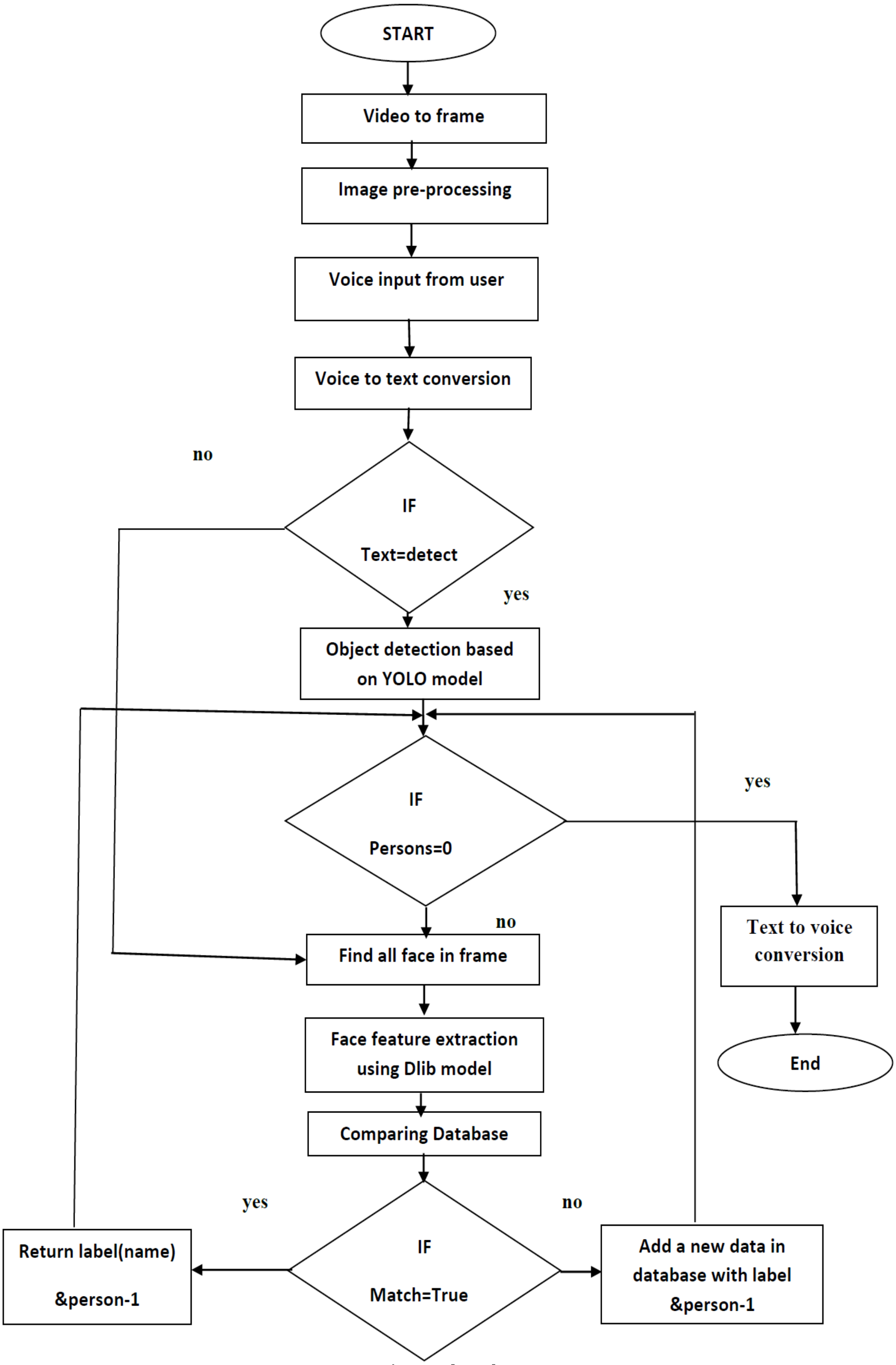

Fig. 3. Flowchart

IV. RESULTS

The proposed people and object identification system for visually impaired peoples has been developed and tested for various cases. The results are discussed below.

The remote is identified and the result is showed in fig.4. To identify the multiple objects a cell phone and a remote is kept in front of the system. It identifies the cell phone and remote and displays the same as shown in fig. 5. To check the other objects a laptop and a cup is also tested and verified as shown in fig. 6.

The people identification is done with existing database. The known person database is created. When the developed system seen the known person it identify the person with their name in the database. Else it will identify as unknown person. The fig. 7. Shows that the known person name Jothi is identified from the existing database and a laptop and a bottle in the computer vision is also identified. For justifying the results they are displayed. For the visually impaired persons the text are converted into voice. The information is send to their headphones.

Fig. 4. Identifying remote

Fig. 5. Identifying remote and cell phone

Fig. 6. Identifying laptop and cup

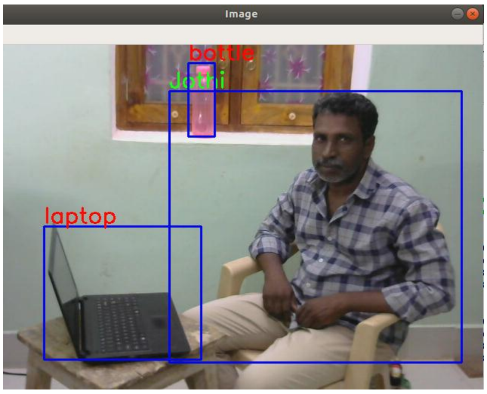

Fig. 7. Identifying laptop, person(Jothi) and bottle

V. CONCLUSION

This paper presents a technique for assisting visually impaired people. The proposed system has a simple architecture and makes it user friendly. This system helps the visually impaired person to identify the objects in their surroundings and also identify the known persons. The preliminary experiments shows promising results as the user can easily identify the people and the objects in their surroundings. This system converts the visual signal into voice signal which makes this more user friendly.

REFERENCES

[1]. N. Congdon, B. O'Colmain, C.C. Klaver, et al. "Causes and prevalence of visual impairment among adults in the United States**"**, Arch phthalmol, 122 (4) (2004), pp. 477-485

[2]. C.A. McCarty, M.B. Nanjan, H.R. Taylor, "Vision impairment predicts 5 year mortality**"**, Br J Ophthalmol, 85 (3) (2001), pp. 322-326

[3]. Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, et al.; "Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis". Lancet Glob Health. 2017 Sep;5(9):e888–97.

[4]. Fricke, TR, Tahhan N, Resnikoff S, Papas E, Burnett A, Suit MH, Naduvilath T, Naidoo K, Global Prevalence of Presbyopia and Vision

Impairment from Uncorrected Presbyopia: Systematic Review, Meta-analysis, and Modelling, Ophthalmology. 2018 May 9

[5]. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[6]. http://www.who.int/disabilities/publications/dar_action_plan_2006to2011.pdf

[7]. http://www.who.int/disabilities/policies/actionplan/disability_action_plan_en.pdf

[8]. Alexy Bhowmick and Shyamanta M Hazarika, “An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends”, J on Multimodal User Interfaces · January 2017, DOI: 10.1007/s12193-016-0235-6

[9]. Borenstein J (1990) The NavBelt—a computerized multi-sensor travel aid for active guidance of the blind. In: CSUN’s fifth annual conference on technology and persons with disabilities,Los Angeles, pp 107–116

[10]. Shoval S, Borenstein J, Koren Y (1998) Auditory guidance with the Navbelt-a computerized travel aid for the blind. IEEE Trans Syst Man Cybern Part C Appl Rev 28(3):459–467

[11]. Hoyle B,Waters D (2008) Mobility AT: the Batcane (UltraCane). In: Hersh MA, Johnson MA (eds) Assistive technology for visually impaired and blind people. Springer, London, pp 209–229

[12]. Coughlan JM, Shen H (2013) Crosswatch: a system for providing guidance to visually impaired travelers at traffic intersections. J Assist Technol 7(2).