Perspectives in Communication, Embedded-Systems and Signal-Processing (PiCES) – An International Journal
ISSN: 2566-932X, Vol. 4, Issue 11, February 2021
Part of the Proceedings of the 1st All India Paper writing Competition on Emerging Research - PaCER 2020

Third Eye for Blind

Sandhya B R
Information Science & Engineering,
Sri Krishna Institute of Technology,
Bangalore, India

S Sahana
Information Science & Engineering,
Sri Krishna Institute of Technology,
Bangalore, India

S Sneha
Information Science & Engineering,
Sri Krishna Institute of Technology,
Bangalore, India

Abstract: Independent travel is a well known challenge for blind or visually impaired persons and also the increasing availability of portable digital image devices, high performance and cost efficiency has created a vast opportunity for supplementing traditional scanning for document image acquisition. A camera based visual assistance framework for object detection, face recognition, news and book reading is built. It converts into a voice output to help the blind people. Thus, we developed this system from a single camera-captured video streaming for the visually impaired persons, so they can lead their lives with some independence.

Keywords: Visually impaired; Object Detection; Face Recognition; KNN classifier; Independence for the visually impaired; Raspberry Pi

I. INTRODUCTION

Everyone has their own responsibilities and tasks that need to be carried out by themselves. Visually impaired people especially find it difficult to perform these tasks. There are many visually impaired people in this world who are always in need of helping hands. To help them perform these tasks with some independence a camera based visual assistance system which can be used to detect and recognize the face, obstacles, inform about the latest news a prototype is built. There are many devices built to help the visually impaired such as devices that detect objects which are intimated through vibrations as discussed in [2][4], devices which can detect text on objects as discussed in [3], to identify the signs on doors as discussed in [5] etc. In the existing system the devices which intimate the user about the obstacle have sensors attached to the device which senses the object in front of the user and produces a vibration. The sensors are mounted on a stick or a belt which the user has on him at all times whenever he or she needs it. By detecting texts on objects it will help the visually impaired person in shopping malls to buy required items without someone's help. The system has a camera that is attached to a cap, but the camera might take time to detect the text on objects. When the blind person needs to go through doors or enter a room as discussed in [5] the person can use that system to read signs on doors which will help in identification of exits, restrooms etc. These devices can help the user know that there is an obstacle ahead or read the texts on objects, but he/she would not know what kind of object it is or if the obstacle is a person or not.

In our system, whenever an obstacle is detected on the way, the camera will capture the object and it is sent to the processor to identify the type of the object as discussed in [1] and then it is intimated as voice command through earphones connected with Raspberry pi. In this way the blind people will be able to identify the object in-front of them. As a concern of the authenticity, the user can also identify the person in front of him by using Face Recognition which can differentiate between a known person and an unknown person. We use k-nearest neighbor algorithm for face recognition. An email is sent to the care taker when a person is recognized with the name and picture of the person. Similarly, if an unknown person is recognized an email is sent with the picture of the unknown person. Reading is very important and one of the most important aspects in our day-to-day lives. The text to speech gives access to putting in lesser efforts in everyday chores but for the visually impaired people it can be a great tool to lead a more normal life. Here, the text in books which are already stored in the system will be extracted and converted to speech using optical character recognition (OCR) and Google text to speech (gtts). There is always something new and informative going on in the world. Visually impaired person can know about this information through the latest news just by pressing a button. Controller with buttons is used to switch between different modes such as object detection, face recognition and book reading. Sometimes there might be situations where the blind person would want to contact a trusted person or a relative in case of any danger. An email can be sent to that person just by pressing a button in those kinds of situations. The whole system is built on a glove and a supporting band to hold the model. The blind person can wear the gloves and easily press the buttons on the controller which is also attached to the glove.

II. METHODOLOGY

A. Object Detection

Object Detection deals with identifying and tracking objects present in images and videos. Mobilenet SSD is used to detect the objects. Mobilenet is a neural network which is used for classification and recognition. SSD is a single shot multibox detector which detects objects within a frame. These two are used for object detection. The Mobilenet layers are used to extract features from an input image, where the layers convert the pixels in the image into features that describe the contents of that image. Therefore,

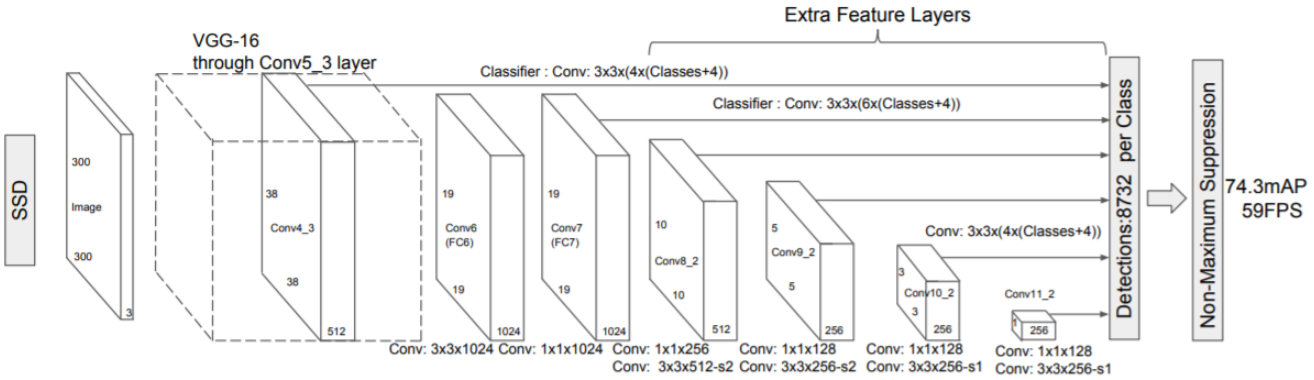

Fig 1. SSD Architecture

Mobilenet acts as a feature extractor. Single Shot Detector (SSD) detects the objects in a single shot using multibox. SSD adds many additional convolutional layers on top of the base network, so it's important to have a feature extractor. The additional convolutional layers are used for detecting the object. At the end of the base network convolutional layers decrease the feature map size and increase the depth. As the feature map size decreases, the objects can be detected. Reference [6] suggests that the additional convolutional layers are responsible for detecting object. Fig 1 shows the SSD architecture, where the extra convolutional layers are used for object detection. Instead of VGG-16 as base network, we use mobilenet as base network.

B. Face Recognition

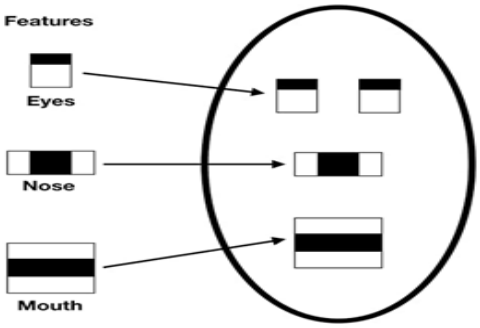

Fig 2. Haar Cascade Feature Extraction

Face recognition is finding points of interest on a human face and recognizing the person based on the features. Here, Haar Cascade and KNN are used for face recognition. Haar Cascade is used to detect the face region in the frame and KNN is used as a classifier to classify the face as known or unknown. Haar Cascade is an algorithm which is used for face detection that detects the face region in the frame. Initially lots of positive and negative images are used to train the cascade classifier. Then the features of the face are extracted which is done by using the Haar features as shown in Fig 2. Each feature is a single value obtained by subtracting the sum of pixels under white and sum of pixels under black. This process in this algorithm is called feature extraction. After the feature extraction process, the KNN algorithm is used for face recognition. The features that are extracted are compared with the trained images and the KNN algorithm maps the predicted face with the name of the person recognized. Based on the majority votes of its neighbours (training set) the face is recognized as known or unknown. When the user presses the button for Face Recognition as shown in Fig 3 the Pi Camera gets activated and the face region in the frame is detected using Haar Cascade. Then the feature extraction takes place and compared with the images of the dataset. After that the face is recognized using KNN algorithm and the name of the person is given as voice output. If the person recognized is unknown then "unknown" is given as voice output.

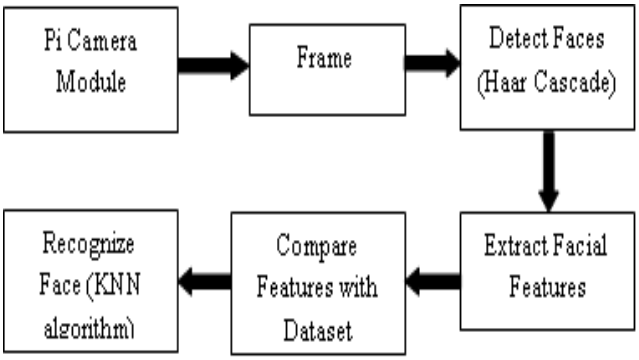

Fig 3. Block Diagram of Face Recognition

C. Optical Character Recognition (OCR)

Optical Character Recognition is a technology that is used to recognize texts in the image. It allows us to convert the texts in the image to editable documents; it extracts the text information that the image contains. We use OCR for book reading and news reading. The images of the books are already stored in the system. Using OCR the texts from those images are extracted and are given as voice output. The basic process of OCR involves examining the text and translating the characters into code that can be used for data processing. We use Python-Tesseract or Pytesseract for the conversion of image to text. Python Tesseract is a python

tool used for OCR. The image is split into zones identifying the areas of interest where the text is present and this helps in the extraction process. The areas containing text can now be broken down further into lines, words and characters and the software will be able to match the characters through comparison.

D. Text to Speech

Earphones / Speakers are attached to Raspberry Pi using which the output is obtained. For conversion of text to speech, Google text to speech is used. Based on the mode chosen the output is given; objects detected, known or unknown person recognized, the latest news or book chosen. For example, if the user chooses object detection mode, then the output obtained will be the detected object. Similarly, based on the mode chosen the other outputs are given.

E. Controller with Buttons

Controller is provided to the user so that the user can switch between different modes easily. Each button on the controller is responsible for a particular feature of the system. The user that is the visually impaired person can choose the feature which he/she would want to use with the help of a controller. It will be easier to handle the system with the help of a controller.

III. HARDWARE DESCRIPTION

A. Raspberry Pi

Raspberry Pi is a credit card sized microcontroller which acts as a minicomputer that connects all peripherals used by computer like keyboard, mouse, TV or monitor etc. It is a low cost microcontroller that helps people of all ages to explore computing. Raspberry pi has SD card slot for loading operating, Ethernet port for LAN cable connection, power supply, 4 USB ports for connecting I/O devices, memory, audio/video outputs and camera interface which can be connected to a Raspberry Pi camera. General purpose input output (GPIO) pins are used to connect the controller with Raspberry Pi. The GPIO pins can be used for both input and output. In our case, we use them for input; the buttons of the controller are connected to these pins through which the input is taken. The earphones and Raspberry Pi camera are connected to the board. The camera is connected to the camera slot and earphones to the jack slot. Fig 4 shows the Raspberry Pi 3 microcontroller in which the parts of the microcontroller are labelled. The operating system (OS) used with Raspberry Pi in our project is Raspbian OS which is a Debian-based free operating system. Raspberry Pi camera is used for taking the real time input for detecting obstacles and recognizing faces. The Raspberry Pi camera is a second generation module which is portable and lightweight that supports Raspberry Pi and has a fixed focus lens. The camera is normally used for image processing and machine learning projects. Fig 5 shows the Raspberry Pi camera which can be attached to Raspberry Pi through the camera interface on the microcontroller.

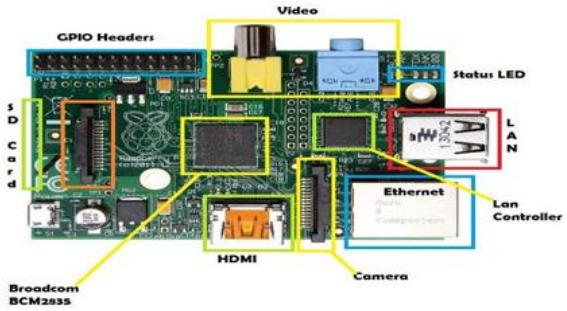

Fig 4. Raspberry Pi 3

Fig 5. Raspberry Pi Camera

B. Controller

Controller with buttons is provided to the user using which the mode can be chosen. The controller is also attached to glove along with the Raspberry Pi. There are four buttons on the controller and each of them is responsible for a specific mode. The first button is used for Face recognition, the second button for alert email, third button for object detection and fourth button for more features that is news reading and book reading. When the system is switched on the features allocated to the buttons are intimated through voice. After which the user can press the button which he or she would like.

IV. RESULT

A camera based visual assistance was built to help visually impaired people lead their lives independently. The system was attached to a glove. The system consists of a camera attached to Raspberry Pi, which is used for object detection using Mobilenet SSD and face recognition using Haar Cascade and KNN algorithm. Haar Cascade is used to detect the face region and KNN algorithm is used to classify the person as known or unknown. The system also fetches the latest news from the internet and can be used for book reading from the stored images in the system using optical character recognition. A controller with buttons was attached to the system using which the user could shift between different modes. An email could be sent to the caretaker or relative in case of emergency. The output of the system is in the form of voice output, where the obstacle detected, face recognized, book or news read, all give the output in the form of audio or voice output. Fig

6 shows how the system is built on a glove and a supporting band is attached to hold the model.

Fig 6. Prototype of Third Eye for Blind

V. CONCLUSION AND FUTURE ENHANCEMENT

Independence is a building methodology which is essential to achieve dreams and goals in life. Visually impaired people find it difficult to do their daily work, they need someone to guide them and help them. Therefore, our system which is a camera based visual assistance with Raspberry Pi which can detect objects, differentiate between known and unknown person, give the latest news, read books aloud; can help the visually impaired person lead their life with some independence.

In the future, more features can be added to the system such as text recognition through video streaming, using of HD camera for better clarity of the images, ultrasonic sensors can be used to calculate the distance between the user and the object. Since there is only limited storage of Raspberry Pi we can select the most important features and add it to system so the system can help visually impaired people in a better way.

REFERENCES

[1] Ayush Jain and Mohana Mohana, "Real Time Object Detection and Tracking using Deep Learning and OpenCV", Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1305-1308.

[2] Piyush Patil and Akshay Sonawane, "Environment Sniffing Portable Assistive Device For Visually Impaired Individuals", International conference on Trends in Electronics and Informatics (ICEI), 2017, pp. 371-321.

[3] Samrudhi Deshpande and Revathi Sriram, "Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People", International Conference on Automatic Control and Dynamic optimization Techniques, 2016, pp. 1020-1024.

[4] Radhika R, Payal G Pai, Rakshitha S and Rampur Srinath, "Implementation of Smart Stick for Obstacle Detection and Navigation", International Journal of Latest Research in Engineering and Technology (IJLRET), vol. 02, pp. 45-50, May, 2016.

[5] Yingli Tian and Aries Arditi, "Context-based Indoor Object Detection as an Aid to Blind Persons Accessing Unfamiliar Environments", October, 2010.

[6] Johnathan Hui, 2018 March 14, SSD object detection: Single Shot MultiBox Detector for real-time processing [Online]. Available:https://medium.com/@jonathan_hui/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06#:~:text=SSD%20uses%20lower%20resolution%20layers,be%20added%20for%20object%20detection.