Virtual Assistant for Visually Impaired

Vipul Sharma[1], Vishal Mahendra Singh[2], Sharan Thanneeru[3], Prof. Amol Kharat[4]
[1,2,3,4]Department of Information technology, Pillai College of Engineering, University of Mumbai

***Abstract*- The field of artificial intelligence has led to various virtual assistants such as Siri in iPhone, Google Allo, Microsoft Cortana, and so on. Even after such progression, very little has been done to implement these technologies to assist the visually impaired community. Recognizing a person or distinguishing an object, these tasks are straightforward for common people but can be very difficult for people that are partly or completely blind. Their lives can be made smoother by assisting them to detect what is present in front of them at that instant. We aim to develop a system/assistant that will serve to guide a visually impaired person and will indicate the person by speaking through the earpiece. The system will help the person recognize people, add new faces and detect objects that are in their vicinity. We will have a mobile application which will consist of numerous deep learning models that will help applications increase its administration. The primary working of the system will consist of the camera continuously feeding images for inputs, the core system processing this input information and the earpiece acting as the output device to provide this output to the user.**

***Keywords*- Face Recognition, Object Detection, Cognitive Services, Text-to-Speech, Deep Learning.**

I. INTRODUCTION

"Virtual Assistant for visually impaired", the said project applies the concept of Deep learning i.e. Neural networks. The models employed for our project are - Face Detection and Object Detection. The system comprises a camera that acquires images and sends them to the application, where a powerful processor derives information from them and explains them to the user through a distinct audible message. The device will continuously detect all the faces in front of the person and verify them against all the faces of the people who have been previously taught to the device.

II. PROPOSED SYSTEM

In the system level, we could say that the novelty lies in the real-time web application. The already existing system comprises modules such as Image processing, Speech processing, etc, therefore the problems faced by blind people are often reduced to a particular extent. But neither are these modules enough nor are they implemented purposefully such that they assist the visually impaired. Taking these limitations into consideration, the system we have developed overcomes these drawbacks and helps build a system that assists the needful in a better and more appropriate manner.

Modules focused upon by us:-

A. ***Text-to-Speech***

This module comprises text and speech processing. The main purpose of this module is to take into consideration all the text provided and convert these into the appropriate audio output using speech processing. We have implemented a dynamic system that makes use of Google API (Gttx) for the conversion of Text to Speech dynamically provided that good internet connectivity is present.

B. ***Object Recognition***

Object Recognition is a process in which Real-world objects are identified using Image processing. It is an important operation that will aid visually impaired to locate their frequently used day to day objects. The system that we have developed provides support in visual aid by assisting to dynamically locate and identify the objects in an image and providing the text output for the same.

C. ***Face recognition***

Some face recognition algorithms identify countenance by extracting landmarks, or features, from a picture of the subject's face that includes the features shape of the jaw, nose, cheek, facial hair and other such characteristics. The features of the image in consideration are then compared with other images having similar features. The algorithm normalizes a dataset of face embeddings then compresses these embeddings, only saving the information within the image that's useful for face recognition. Eventually what we will be obtaining is a bounding box surrounding the face in the live monitoring having the name of the person and the confidence attached to the bounding box.

III. IMPLEMENTATION

The system developed is deployed on the web as a website. The website is built on the backbone of flask, which serves the purpose of providing connectivity between the python code and the HTML.

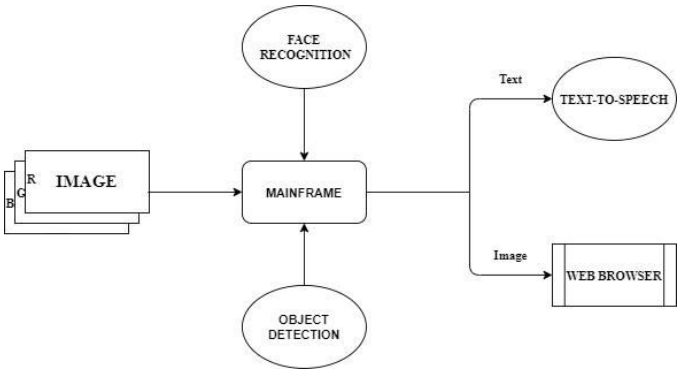

Fig 1. Implementation flow.

When the website is loaded, the object detection module starts its processing and the objects detected by this module are displayed on the page as well as delivered to the user via an earpiece/speaker. Along with this, we also have two buttons ('Switch to Face' and 'Stop') on the landing page that are well separated to be easily accessible. Clicking on 'Stop' results in pausing the Livestream until the 'Start' button is clicked. The 'Switch to Face' button on click will switch to the page where the Face Recognition processing begins. We have also included the buffer which can only contain a maximum of five entities (objects/people) at a time. Each entity will be converted to speech in every 20 seconds if it still exists in the frame.

The 'Face Detection' module is implemented similarly as the 'Object Detection' module using the same layout for the buttons. Here, the clicking of the 'Stop' button will have the same function as mentioned above whereas a click on the 'Add Face' button will capture the current frame and prompt the user to speak out the name of the person whose face is being added. The name is spoken into the microphone by the user and the speech-to-text model converts this audio into the text and stores the text with the captured frame into the database. All the processing is carried out in the python engine and is displayed using HTML to the user. Thus implementing all these, we obtain a system that is more relevant and more assistive to the user.

IV. RESULT

The system is deployed as a web application which, when opened on any mobile browser, gives us the landing page shown below. Along with the landing page, we have two additional pages that play an important role in our system and play a fundamental role in its deployment. When the user presses any of these buttons, the command will be addressed to the user via earpiece/speaker. All these buttons are large in size and are separated properly, so that it is convenient for the visually impaired user to distinguish between them.

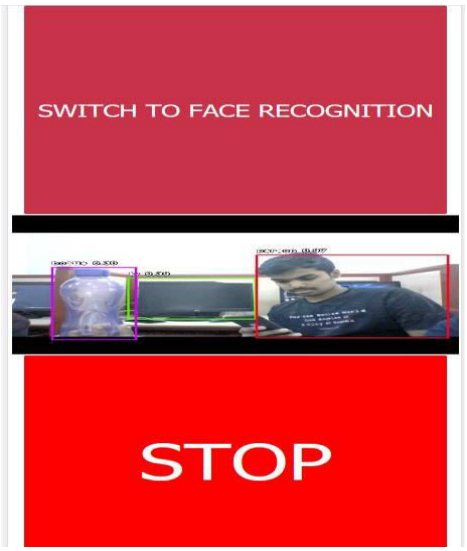

Fig 2. Landing Page - Object Detection.

This page consists of two buttons- one at the top and the other at the bottom of the page. We have a block in the center of the page which provides a continuous live stream that is displayed through the phone. Above the block is the "SWITCH TO FACE RECOGNITION" button which, when clicked, deploys the Face recognition model and directs the user to that page. The other button is named "STOP" and resides below the live stream block. When clicked, this button will stop the current processing model and redirect the user to the page having the "START" button.

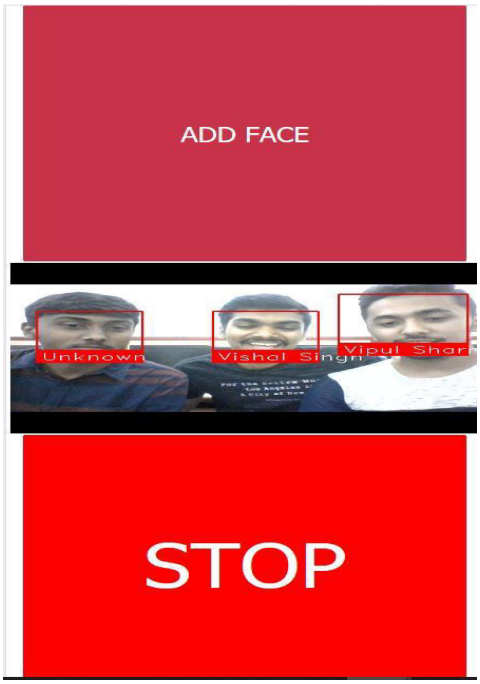

Fig 3. Face Recognition Page.

A click on the "SWITCH TO FACE RECOGNITION" button, the system is directed to a new page where the face recognition functionality begins its execution. Similar to the landing page, this page consists of two buttons- one at the top and the other at the bottom of the page. The button at the top is named "ADD FACE" whilst the button at the bottom of the page is named "STOP". If an unknown face is encountered, we can click on the "ADD FACE' button at the top of the page to add the unknown face into the Facial database. The "STOP" button executes the same functionality as before and will stop the current processing model and redirect the user to the page where the "START" button resides. The block in the center of the page separates the two buttons and continues to provide the live stream and displays it through the phone's browser window. All the faces recognized in the live stream are addressed to the user via earpiece/speaker.

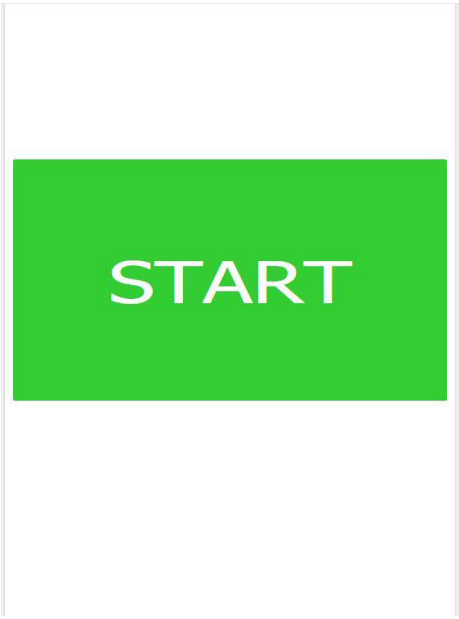

Fig 4. Page to START the system after it is stopped.

This page consists of a single large button named "START". When the "STOP" button on either of the Face Recognition page or the Object Detection page is clicked, the user is redirected to this page where the "START" button resides.This enables the user to start the system anew after it has been stopped. Hence, allowing the user to begin the system according to their convenience and usability. This project is available on https://github.com/Deimos-M/DL-Virtual-Assistant.

V. CONCLUSION

In this paper, various techniques to implement the aforementioned system are analyzed and summarized. Different systems have different ways of implementation along with some limitations and restrictions. These types of systems are very critical for multiple reasons and the occurrence of an error in such a system/device may cause catastrophic damage and loss. The system we are achieving overcomes the limitations of the already implemented systems. Our system consists of a basic UI on a web-based application and comprises several Deep learning models; some of them are object detection, face recognition, speech recognition and so on. These modules will work together and assist in vital activities like object detection as well as face detection and recognition for the visually impaired.

VI. FUTURE SCOPE

There are various applications of this domain system. The future scopes are listed below.

A. Alerting the visually impaired person about the Obstacle Position

We would implement the device in such a way that the sensors will be mounted on a spectacle and this would help the person wearing the spectacle detect the obstacle position in front of their vision in the walking path.

B. Voice Command and Emergency Voice Call Establishment

We would include the facility to save an emergency number in the application so that the visually impaired person can establish a voice call to the predefined number by using his/her voice command. When the visually impaired person wants to give a voice command, he/she need not touch the phone and just pressing the lock button thrice on the phone will lead to prompt command and by uttering “HELP” this voice command will connect through a voice call to a predefined number.

C. Text Reader

This system will help a visually impaired person to listen to the text which is written in any literature or any book. The system will take a pic and it will recognize the text written on it using image processing. This recognized text is then converted to speech using a text-to-speech model.

REFERENCES

1. Marcos Barata, Afan Galih Salman, Ikhtiar Faahakhododo, Bayu Kanigoro,"Android based voice assistant for blind people", Library Hi Tech News, Vol. 35 Issue: 6, pp.9-11 (2018).
2. Md. Siddiqur Rahman Tanveer, M.M.A. Hashem and Md. Kowsar Hossain , “ Android Assistant EyeMate for Blind and Blind Tracker” (2018).
3. Joseph Redmon, Ali Farhadi, “An Incremental

Improvement", Pjreddie (2018).

4. Vincent Gaudissart, Silvio Ferreira, Céline Thillou, Bernard Gosselin,"Mobile Reading Assistant for Blind People" (2018).
5. DR. Kavitha C, MR. Nithin V Gopal, MS. Nidhi Amarnath, MR. Prajwal G, MS. Supreetha R.R., "VIRTUAL ASSISTANT FOR BLIND PEOPLE" (2018).
6. Prof. Priya U. Thakare, Kote Shubham, Pawale Ankit, Rajguru Ajinkya, Shelke Om, "Smart Assistance System for the Visually Impaired" International Journal of Scientific and Research Publications, Volume 7, Issue 12, December 2017 378 ISSN 2250-3153 (2017).
7. Faizan Ahmad ,Aaima Najamand Zeeshan Ahmed, " Image-based Face Detection and Recognition" ,Arxiv (2015).
8. R. Velázquez, Wearable Assistive Devices for the Blind. Chapter 17 in A. Lay-Ekuakille & S.C. Mukhopadhyay (Eds.), Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization, LNEE 75, Springer, pp 331-349 (2010).