JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SURVEY ON PERSONAL ASSISTANT FOR VISUALLY IMPAIRED

[1]Prashanth H S, [2]R Dekshitha, [3]Swetha Bijanapalli,[4]Yogita Raikar,[5]T Hemanth Chowdary

[1]Associate Professor, [2,3,4,5]Undergraduates
[1]Department of Computer Science and Engineering,
[1]K S Institute of Technology, Bangalore, India

***Abstract**: Loss of vision that occurs due to some accident or disease from birth, is a huge problem faced by many people around the world. Access to information present on the internet or various other sources is essential in this fast paced world. It is quite hard for the non-sighted people to remain connected with the world and keep themselves informed and updated regularly. Hence, we propose a project which is a mobile application specifically designed to help blind or visually impaired people in their everyday lives, just like their human assistant. With this application, the visually impairs can get more autonomy as they can read with their ears and write with their voices. They can interact with the application through their voice as input to perform certain specific operations on their mobile such as make a call, write and send a message/email, capture the surroundings, recognize currency notes or people. This application sends voice output while reading new unread messages/emails, news highlights, current time and weather, or to produce warning of the obstacles. Objects and people present around them can be detected and identified when they instruct the application to capture the scene. The identified objects are given as output in the form of speech. Objects are detected using OpenCv and TensorFlow, and with the ratio of the box and the mainframe, the approximate distance of the object from the person is calculated and if it is too close a warning is produced. Object detection feature is extended to recognize currency notes and people.*

Index Terms **– Visually Impaired, Assistant, Speech-to-Text, Text-to-Speech, Object Detection, Object Recognition**

I. INTRODUCTION

At present, numerous virtual assistants such as Siri in iPhone, Google Assistant, Microsoft Cortana, and so on exist due to the rapid advancement in the field of artificial intelligence. Even after such progression, the visually impaired community face many challenges as very little has been done to implement these technologies specifically to assist them. Certain tasks such as recognizing a person or obstacles or distinguishing objects, are straightforward for common people but can be very difficult for people who 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 and also help them to perform essential tasks in their everyday lives.

We aim to develop a system/assistant that will serve to guide a visually impaired person. It listens to the commands provided by the person, performs the respective task and interacts back with the person by speaking through the microphone of their mobile phones. Speech recognition is used to understand the user's words and convert the speech to text. Modules such as Email/message writing are implemented by using speech recognition libraries.

This 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 continuously and feeds them as input 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 also 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.

The paper is organized as follows: Section I was the introduction. Section II shall be the overview of the literature review. Section III outlines the problem statement. Section IV describes the implementation methodology of our project. Lastly, Section V concludes the paper.

II. LITERATURE SURVEY

In this section, research papers related to personal assistants for the visually impaired are surveyed and highlighted. Numerous efforts have been made in the field of research to help the visually impaired people.

Kanchan Patil et. al paper [1], describes a voice-over Chat-bot that interacts with the user through voice commands and forms the backbone of the system. Five modules, namely, image captioning, object detection, face recognition, voice-over chat bot and text reading were implemented. They have built this system using several techniques. For object detection and image captioning, Tensorflow was used. Opencv is used for image processing. Pyaudio and speech recognition are used to obtain input from the user. gTTS and pyTTSX are used to output the dialogue through speech. One limitation in this was that the Chat-bot fails at times to recognize commands in noisy environments.

Paper [2] aims to help the visually impaired people with detection objects- value of currency notes, and color/pattern of clothes. TensorFlow along with deep-learning techniques were used for detection. Convolution layer, batch normalization, ReLU non-linear and linear layers are used for object classification. Recognition of currency notes (Egyptian) was implemented by CNN using MobileNet architecture. CNN resnet50 architecture is used to analyze color and pattern of clothing. It was concluded that, CNN VGG16 (Currency Recognition) achieves 92.7% test accuracy in ImageNet with a dataset of over 14 million images belonging to 1000 classes but low accuracy was observed when used with smaller dataset.

In-door mobile application as represented in paper [3] facilitates the process of guidance in indoor environments for visually impaired people. The system detects objects and calculates the distance for guiding the user toward the right direction and provides warning of obstacles. In this project, Coco dataset is trained using quantized Mobilenet SSD and Objects are detected using Tensorflow Lite object classification. Camera calibration with triangle similarity formula is used to calculate distance.

Abdul Majid Norkhalid, et. al paper [4] depicts the speech interface system. This mobile application was based on the Android platform with the combination of speech interface system and object detection. The mobile navigational application has four main components- automatic speech recognition to detect voice, natural language processing component to understand the context, Route processing middleware component to process the route commands and map API component to map the route.

Vinayak Iyer et. al paper [5] proposed a system to assist the visually impaired to access the internet and interact with any website with ease. Speech-to-text is used to take input or query from the user, the recognized query is searched by using Selenium and the website is scrapped by using Beautiful Soup module. For implementing the Gmail module, Python script is built to log in the user to gmail and takes relevant details and messages using Google speech-to-text library and then sends it with the user's confirmation. To obtain answers to questions from Wikipedia and summarize it, Stanford Question Answering dataset with BERT model is used.

Object Detection using SSD algorithm paper [6] describes the method to detect the objects in the surroundings and guide the visually impaired people about the objects. It gives voice instruction about the obstacles present on the way and helps visually impaired people to get direction. The algorithm also estimates the distance of the object. It uses SSD algorithm and Monodepth algorithm for estimation of object distance. R-CNN algorithm is used to reduce complexity of the problem. The selective search was created using Edge box algorithm and it passed directly into softmax (last) layer of the CNN.

Shrikesh Suresh et. al. paper [7], built an android application – 'Vision'. This android application is for blind people with a voice assistant that can recognize objects and speak out. It has several features such as, voice assisted navigation through Google Maps, news reader and alarm keeper. It uses two APIs - Tensorflow Object Detection API for detection and Google Maps navigation API to navigate to the saved locations. Latest news is read out from RSS feeds of popular news.

Artificial Intelligence-based Voice Assistant paper [8] performs tasks as given by the user such as play/download song or video from YouTube, search and read content from Google, find the exact location the user asks for, inform the exact weather of the location, take screenshot on voice command, give live news updates and also send user's location to alert police and relatives in case of any danger. The Automatic Speech Recognition (ASR) system is used for AI-based Voice Assistant. In this method the input audio from the microphone is taken and the recorded waveforms from speech are sent for analyzing the acoustics. After processing data the speech waveforms are transmitted to the decoder where they are transformed to text to use as command.

Deep learning based real-time obstacle detection system paper [9] is a low cost system which has the capability to detect and track the surrounding obstacle and estimate the object-camera distance in real-time. You Only Look Once (YOLO) approach was applied to detect objects in the frame and create a bounding box and the bounding boxes are tracked in subsequent frames. Kernelized Correlation tracker (KFC) is a kernel based tracking technique applied to the system for object tracking. DisNet - deep learning based method is enforced for estimation of distances.

Rais Bastomi et. al. paper [10], uses convolutional methods with Stereovision. This tool is built to utilize the camera to record the real-time video of the surroundings and measure the distance between object and camera with Stereo Vision. The input image is processed using OpenCv and set into 640*320 pixels. CNN consists of 2 layers - Feature extraction Layer (Convolution layer using Gaussian kernel and Max Pooling layer) and classification layer. Centroids are located in the bounding box through which distance is calculated.

Jawaid Nasreen et. al. paper [11], communicates to the users about the objects detected in the surroundings which are captured by the camera of the device. The proposed system uses TensorFlow python libraries for detecting the object and one of the CNN models called VGG16 algorithm trained in ImageNet dataset. The system sends the captured image from the phone camera to a web server where the image is detected using the YOLO model and the result is sent to the client where it is converted to voice and is narrated to blind person. The accuracy of the object detected depends on the opacity of the image.

Paper [12] builds a smart mobility application for visually impaired people which detects objects in real time, tracking and estimates distance of the object from the user. This smart mobility application also tracks movement and position of pedestrians and vehicles and gives an estimation of the direction of movement and speed of objects in motion. Two deep learning approaches were used: You Only Look Once (YOLO) V3 and Single Shot Detector (SSD) algorithm. Monodepth algorithm was used to estimate the object distance. The comparison between two methods for object detection showed that YOLO V3 is better than SSD as it detects more objects in an image with good accuracy.

III. PROBLEM STATEMENT

The problem statement is identified as:
"To develop an application for visually impaired people to assist them and to perform the tasks using voice commands."
Innovative and effective tools are needed to help them overcome some of the problems in their everyday life. A system which can help them perceive the objects/people in the surroundings and provide navigation guides or carry out tasks on their smartphone, can transform their lives by instilling a sense of independence. Hence, a personal assistant is developed to interact with the users through voice and perform the instructed tasks.

IV. RESEARCH METHODOLOGY

The software can be invoked by the user's voice command or by a predefined keyboard shortcut. When the user opens the application the user can login using face recognition**.** Once the user logs in, the guidance about the application usage is given to the user. It tells the different features in the application and the command to be used to perform a particular task. The application interface provides the available options which the user can perform with a voice command.

The user then gives the command as voice. If the command is invalid, the 'Repeat command' instruction will be given to the user. If the user gives the valid command, the particular task will be performed based on the given command. For efficient speech recognition, the voice command initiates the user to speak by giving a voice command. Once the user's command is received and recognized by the system, the user's command is played again to get the user's confirmation to reduce any errors. On getting confirmation from the user the commanded task is being performed by the system.
After performing the tasks the user can exit the application by giving 'Quit' command in the form of voice by the user.

Below are the methodology for various modules that we're implementing along with its supposed working. These modules are individually implemented across the app to ensure their working separately.

Call/Message Module- If the user gives the command 'Call', the application takes the number in the form of voice as input from the user then asks for the confirmation of the input from the user. Upon confirmation it places the call to the number. If the user gives the command 'Read messages/ emails', the application opens the unread messages/emails and reads out the messages/emails to the user in the form of voice. If the user gives the command 'Read messages/ emails', the application opens the unread messages/emails and reads out the messages/emails to the user in the form of voice.

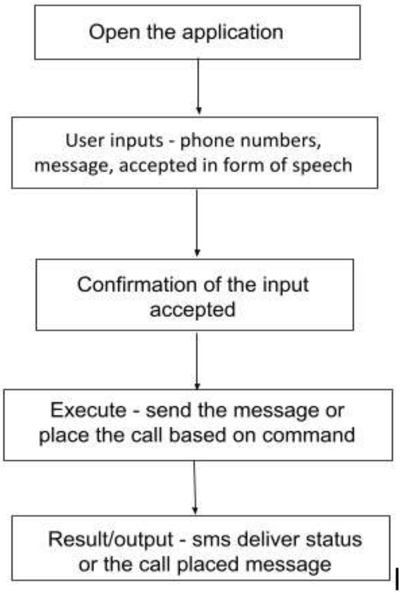
Open the application
User inputs - phone numbers, message, accepted in form of speech
Confirmation of the input accepted
Execute - send the message or place the call based on command
Result/output - sms deliver status or the call placed message

Fig.1. Call/Message Module

Time/ Date/ Location/ Battery Level/ Weather/ News Module- If the user gives the command, “Today’s News”, the application reads out current news to the user. If the user gives the command, “Today’s Date & Time”, the application reads out today's date and current time to the user. If the user gives the command, “Today’s Weather”, the application reads out current weather to the user. If the user gives the command, “Location”, the application reads out the location of the user.

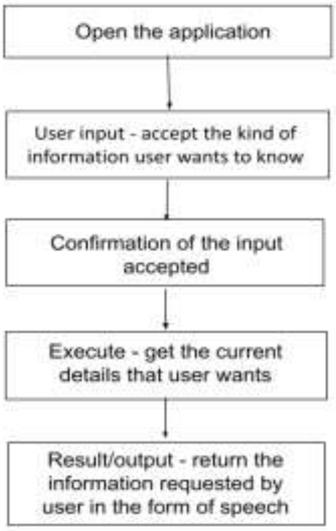
Open the application
User input - accept the kind of information user wants to know
Confirmation of the input accepted
Execute - get the current details that user wants
Result/output - return the information requested by user in the form of speech

Fig.2. Time/Date/Location/BatteryLevel/Weather/News Module

Face/ Currency/ Object Detection module- If the user gives the command ‘Detect the currency’, the application tells the value of the scanned image of the currency in the form of voice. If the user gives the command ‘Detect the objects’ or ‘Tell about the surroundings’, the application will capture the image of the surroundings as the user holds the camera and detects the objects in the captured image and tells the identified objects to the user in the form of voice and also generates a warning if any objects is in proximity.

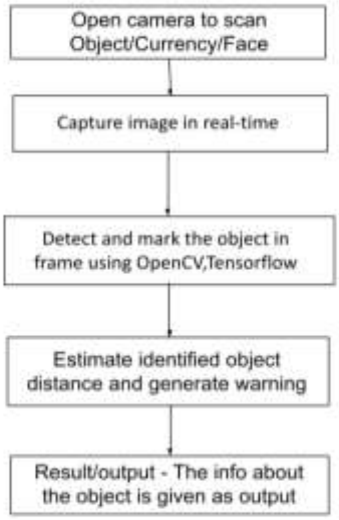
Open camera to scan Object/Currency/Face
Capture image in real-time
Detect and mark the object in frame using OpenCV,Tensorflow
Estimate identified object distance and generate warning
Result/output - The info about the object is given as output

Fig. 3. Face/Currency/Object Detection module

V. CONCLUSION

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 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-using TF, TTS, and speech recognition. 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. ACKNOWLEDGEMENT

We'd like to take this opportunity to thank our college, K.S. Institute of Technology, Bengaluru, for providing all resources, facilities and creating a learning atmosphere. We would like to express our gratitude to Dr. Rekha B. Venkatapur, Professor and Head, Department of Computer Science and Engineering, for all of the assistance we received to successfully complete this survey paper.

REFERENCES

[1] Kanchan Patil et. al., “Guidance System for Visually Impaired People”, Proc of IEEE, 2021

[2] Mayar Osama et. al., “Design and Implementation of Visually Impaired Assistant system", Proc of IEEE, 2021

[3] Nouran Khaled et. al., “In-door assistant Mobile application using CNN and Tensorflow”, Proc of IEEE 2020

[4] Abdul Majid Norkhalid, et. al., “Mobile Application: Mobile Assistance for Visually Impaired People-Speech Interface System”, Proc of IEEE, 2020

[5] Vinayak Iyer et. al., “Virtual Assistant for the visually impaired”, Proc of IEEE, 2020

[6] K.Vijayakumar et. al., “Object Detection for visually impaired people using SSD Algorithm”, Proc of IEEE, 2020

[7] Shrikesh Suresh et. al., “Vision: Android Application for the Visually Impaired”, Proc of IEEE, 2020

[8] Subhash S et. al., “Artificial Intelligence-based Voice Assistant”, Proc of IEEE, 2020

[9] Fatima ZahraeAitHamou Aadi et. al. , “Proposed real-time obstacle detection system for visually impaired assistance based on deep learning”, Proc of IJATCSE, 2020

[10] Rais Bastomi et. al., “Object Detection and Distance Estimation Tool for Blind People Using Convolutional Methods with Stereovision”, Proc of IEEE, 2019

[11] Jawaid Nasreen et. al., “Object Detection and Narrator for Visually Impaired People”, Proc of IEEE, 2019

[12] Zhihao Chen et. al., “Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility”, Proc of IEEE, 2019