A Smart Personal AI Assistant for Visually Impaired People: A Survey

Gagandeep Singh[1], Kevin Takhtani[2], Omkar Kandale[3], Nandini Dadhwal[4]

1,2,3,4Dept. of Computer Engineering, SKN Sinhgad Institute of Technology and Science, Lonavala, Maharashtra, India

--***---

Abstract - *In today's world, Independent living is becoming a necessity where visually impaired people face a major problem of restrictiveness. Visually Impaired are at a disadvantage as they require a manual aid to get information about their surrounding environment. Visual information is the basis for most tasks and that is why Visually Impaired people faces restrictiveness in carrying out most of the basic tasks of their lives such as going for a walk, eating, communicating with a group of people etc. With the advancement in Technology, it is now possible to support the visually impaired. This paper is a survey on a solution to help the visually impaired using Artificial Intelligence, Voice Recognition and Image Recognition. The solution is implemented by developing an Android Application which can assist the user using voice command, recognizing and analyzing the surrounding environment and providing suitable response, chat bot. The application is also capable of providing text recognition to analyze the text. It provides an efficient way by which visually impaired people will be able to interact with the environment using their smartphone.*

Key Words: **Vision API, Artificial Intelligence, Natural Language Processing, Natural Language Conversion, Neural Network, Dialog Flow, Chatbot**

1. INTRODUCTION

One of the major disabilities faced by visually challenged people is Visual Impairment. The person is not able to recognize the surrounding nor able to feel the surrounding due to lack of eye vision. The main aim of this project is to provide the Visually impaired people a way to look at their surroundings. This aim is fulfilled using the technologies such as Artificial Intelligence, Machine Learning, Image Recognition and Text Recognition.

Visual Impairment leads to various difficulties in carrying out different tasks including

1. Basic day to day activities (Moving from one room to another, eating)

2. Travelling from one place to another for work, shopping etc.

3. Reading, Writing, Social Interaction at different places.

The most popular solution to provide accessibility to Visually Impaired people by helping them in travelling from one place to another is the Smart Stick that uses a GPS module to track the users location and various sensors and a microcontroller to alert user about the obstacles on the way[1]. The problem with this system is that it works on less crowded areas and does not provide details about the surrounding.

Another solution is the one that helps the visually impaired with reading using Speech Synthesis Technology. It is an application which recognizes the text from a pdf document and reads it for the user [2]. It uses a camera to take image and convert it into a document. However, it requires an Internet connection and cannot work offline as it also provides chatbot functionalities such as light conversations.

2. RELATED WORK

An Object Recognition in a mobile phone application for visually impaired users [3]. It uses camera to capture images using the available RGB camera which is then converted to HSI (Hue Saturation Intensity) and then the object is detected. It uses Camera as well as sensors to find the brightest source and other colours in the image. The tool was developed on android platform. In this method, the properties of the image are very important, the present text can be of different types depending on the source. The image is taken and is converted to greyscale. The major problem with this implementation is that as the complexity increases, the image recognition performance degrades as more processing is required to convert RGB to HSI.

Object Recognition for Blind People based on Features Extraction [4]. SIFT (Scale Invariant Features) algorithm was proposed to implement this solution. It doesn't require any form of image conversion. Pre-processing is done to clear the challenges created by noise and uneven lightning. The interest point is then found out with the help of local feature extraction method for which a feature or vector and descriptor is computed. The algorithm helped in image representation as an interest point collection of images which are invariant to image transformation and partial to illumination changes. It overcomes the drawbacks of previous implementation of RGB to HSI conversion since the complexity of image does not affect the performance. However, the algorithm used was closed source and is difficult to implement on different devices.

VisualPal was a mobile application for Object Recognition to aid the visually impaired [5]. It detects the direction of maximum brightness and major colours in the image. It made

use of Artificial Neural Network Technology along with Euclidean Distance measures together. It captured a video and categorized it into various frames. All frames are compared with previous frames and response will be given based on stored objects information

Blind Reader: An Intelligent Assistant for Blind is an Android Application [6]. It used Speech Synthesis and Text Recognition to recognize the text from a pdf file and synthesize it to the user. A text document or a .ppt file is converted into a .pdf by recognizing a collection of words. As the application is built on android, it uses pre-defined APIs for text-to-speech conversion which makes the process even more efficient. However, it doesn't recognise Text through image Google's Vision API is used.

A Smart Personal AI Assistant for Visually Impaired People [7] is an application developed for android platform and is the paper we have reviewed. It uses Google Vision API for Image recognition and also provides functionalities to recognize text from a .pdf file which is then read aloud using text-to-speech engine provided by google. It uses device camera to capture image and requires a network connection for the image to be processed by google vision API.

3. PROPOSED SYSTEM

Visual Impairment is one of the major problems of the blind people. Due to this, they face various problems in their daily activities such as reading, eating, walking etc. The main aim of this proposed system is to provide an efficient manual aid to the visually impaired. In this paper we will be reviewing the technology used to implement the proposed system and the features it will contain using Artificial Intelligence and Machine Learning [8]. Hence, a solution to the problem was provided. A simple diagram to provide an overview of functionality is given below.
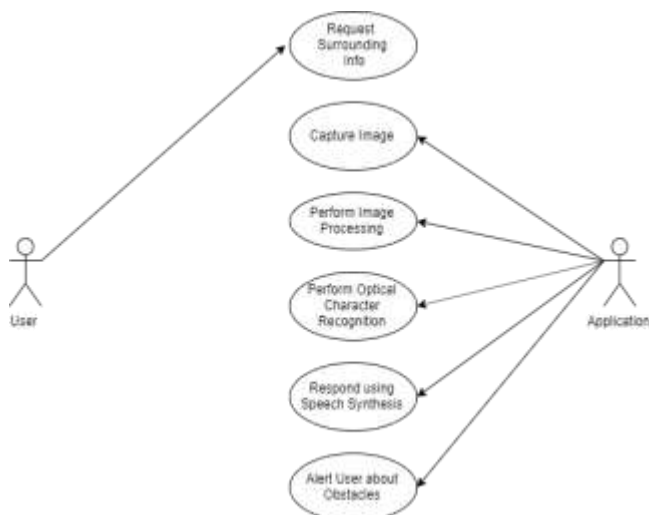

Fig -1: Application Use Case

3.1 System Architecture

Reviewed system is an application developed in android platform that uses device camera to capture the image which is used to detect objects in the image and calculate their distances which will enable alerting the user about the obstacle in the surrounding. Along with that a chatbot that will answer user's query is implemented. The Fig.1 shows the basic functionalities of the application.
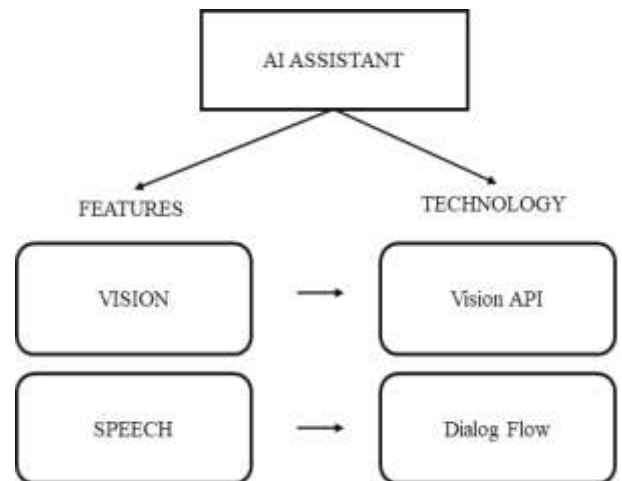

Fig -2: Application Functionality

3.2 Modules

A. Image Recognition

Image Recognition will be using Google Vision API. The basic data flow for image processing is represented in the Fig. 3. Here the image is captured with camera which is then uploaded to the cloud where Vision API runs on the image and object recognition takes place and result is sent to the user.
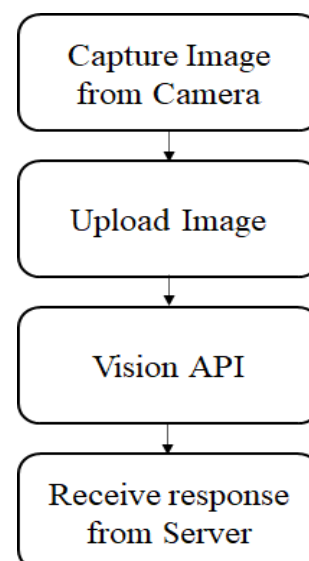

Fig -3: Image Recognition flowchart

Google Cloud's Vision API offers powerful pre-trained machine learning models through REST and RPC APIs. Assign labels to images and quickly classify them into millions of predefined categories. Detect objects and faces, read printed and handwritten text, and build valuable metadata into your image catalog.

B. Advantages of Google Vision API

By using specialized REST API, called Google Cloud Vision API, developers can implement Google Vision as per their requirement which is a huge advantage over other available solution. Detection and classification of multiple objects including the location of each object within the image is performed efficiently by the API. It uses AutoML Vision Edge to build and deploy fast, high-accuracy models to classify images or detect objects at the edge, and trigger real-time actions based on local data. Google AutoML Vision is a machine learning model builder for image recognition. Its machine learning code allows virtually anyone to provide the tagged images required to train a system that is learning computer vision, enabling it to perform categorization and other image recognition tasks.

C. Speech Recognition

Speech Recognition is used in the application to understand user's query and respond to it. The user can ask for surrounding information, to which the application will provide information such as objects in the image detected and their distances using Speech synthesis. If the user is walking, then the speech synthesis will be used to inform user that there is an obstacle of the way. The flow of speech recognition is explained by the Fig. 4 where the user's speech is given as input which is then converted to text and then sent to server for analysis. A suitable response is generated and sent to the application.
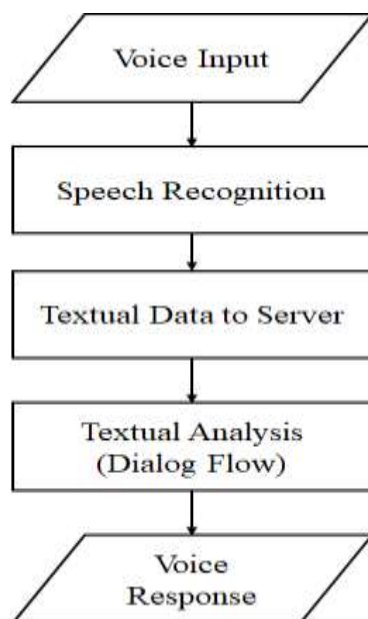

Fig -4: Voice Recognition flowchart

Application requires a bot to perform interaction with the end-user which is done using Dialog Flow API. Dialog-Flow is a platform for building natural and rich conversational experiences. Google uses its own implementation of Dialog Flow which can use Cloud Text-to-Speech powered by DeepMind WaveNet to generate speech responses. This conversion from intent text responses to audio is known as audio output, speech synthesis, text-to-speech or TTS.

D. Advantages of DialogFlow Interaction

It delivers deliver natural and rich conversational experiences using Natural Language Processing and carry forward conversation in a natural way. Machine learning makes Dialogflow intelligent enough to predict the hidden intention expressed in the natural input language. A Dialogflow chatbot can map the user's query with the database available with its backend server. The mechanism of mapping is called as Intent. Being an API, it is independent of the platform it is running on making it cross-platform.

E. Text Recognition

Text Recognition is another technology that will be used in the application to aid the user by reading the data. Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine text, whether from a scanned document, a photo of a document, a scene-photo or from subtitle text superimposed on an image. Widely used as a form of information entry from printed paper data records – whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation – it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

It will allow text recognition. Clicking an image of text will be used to recognize text using OCR (Optical character recognition) provided on Android phones. After that Android's Text-to-Speech engine will be used to speak out the recognised text. This will further increase the functionality of the application enabling the user to read.

4. CONCLUSION

In this paper, we have reviewed the technique to make the lives of Visually Impaired people easier by providing them a visual aid. A Chatbot application was implemented which enables the visually impaired to feel the surrounding up to certain extent using Artificial Intelligence and Machine Learning Technology. A chatbot provides friendly interaction and further makes it easy to interact with the application. It also completely eliminates the need to physically interact

with the device. This is because, we are directly interacting with the application using voice command. Therefore, this application enabled visually challenged people to improve their lifestyle.

REFERENCES

[1] Mukesh Prasad Agrawal, Atma Ram Gupta, "Smart Stick for the Blind and Visually Impaired People" in Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)

[2] Aatisha Cyrill, Shubham Melvin Felix, L. Mary Gladence, "Text Reader for Blind: Text-To-Speech", International Journal of Pure and Applied Mathematics Volume 117 No. 21, 119-125, 2017

[3] K. Matusiak, P.Skulimowski and P. Strumiááo," Object recognition in a mobile phone application for visually impaired users", Lodz University of Technology, Lodz, Poland.

[4] Hanen Jabnoun, Faouzi Benzarti, Hamid Amiri, "Object recognition for blind people based on features extraction", IEEE IPAS'14: International Image Processing Applications and Systems Conference 2014

[5] Shagufta Md.Rafique Bagwan, Prof. L.J. Sankpal," VisualPal: A Mobile App for Object Recognition for the Visually Impaired", IEEE International Conference on Computer, Communication and Control (IC4-2015).

[6] Shahed Anzarus Sabab, Md. Hamjajul Ashmafee, "Blind Reader: An Intelligent Assistant for Blind", 19th International Conference on Computer and Information Technology, December 18-20, 2016, North South University, Dhaka, Bangladesh.

[7] Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People" in Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018).

[8] N.G. Bourbakis, D. Kavraki, " An Intelligent Assistant for Navigation of Visually Impaired People", 2011.