International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

A Smart Personal AI Assistant for Visually Impaired People

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

[1, 2, 3, 4]*Dept. 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 big problem of restrictiveness. Visually Impaired are at a drawback as they require a manual aid to urge information about their surrounding environment. Visual information is that the premise for many tasks which is why Visually Impaired people faces restrictiveness in closing most of the elemental tasks of their lives like going for a walk, eating, communicating with a group of people etc. With the advancement in Technology, it's now possible to support the visually impaired. This paper may well be a survey on a solution to help the visually impaired using computer science, Voice Recognition and Image Recognition. the solution is implemented by developing an Android Application which could assist the user using voice command, recognizing and analyzing the encircling environment and providing suitable response, chat bot. It provides an efficient way by which visually impaired people are able to interact with the environment using their smartphone*

Key Words: **YOLO Algorithm, 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 and Image 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 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

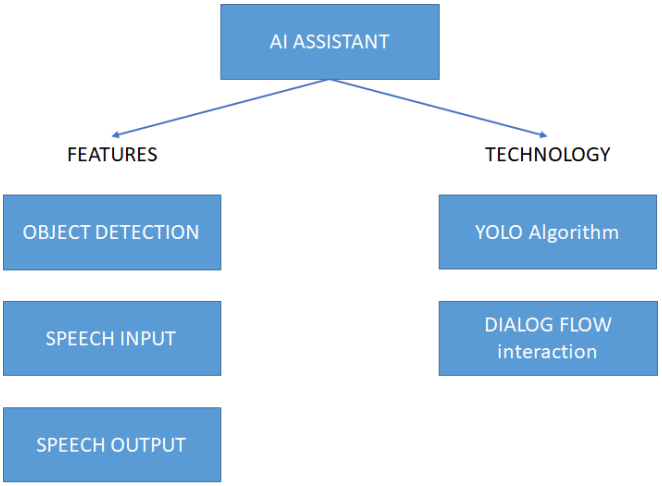

Fig -1: Application Architecture

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 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.

3.1 System Architecture

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.

3.2 Modules

A. Image Recognition

YOLO (You Only Look Once) real-time object detection algorithm, it is one of the most effective object detection algorithms that consists of many of many ideas coming out of the computer vision research community. Object detection is a critical capability of autonomous vehicle technology. It's an area of computer vision that's very popular and working so much better than just a few years ago.

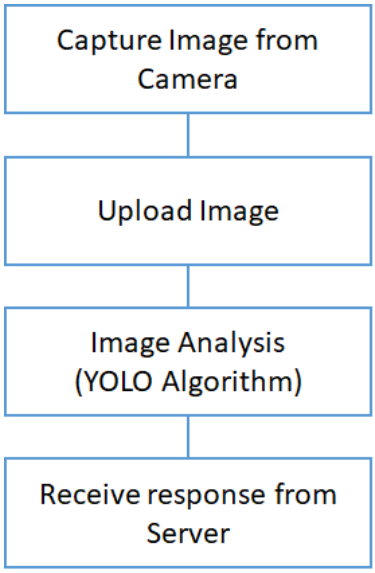

Fig -2: Image Recognition flowchart

B. Advantages of YOLO Algorithm

YOLO has a very high computation speed of (45 frames per second better than real time). Network understands generalized object representation allowing allowed to train the network on real world images and predictions on artwork was still fairly accurate. The network uses features from the entire image to predict each bounding box. It also predicts all bounding boxes across all classes for an image simultaneously. YOLO design enables end-to-end training and realtime speeds while maintaining high average precision. The system divides the input image into an S × S grid. If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object.

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 binary data and then sent to server for analysis. A suitable response is generated and sent to the application.

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.

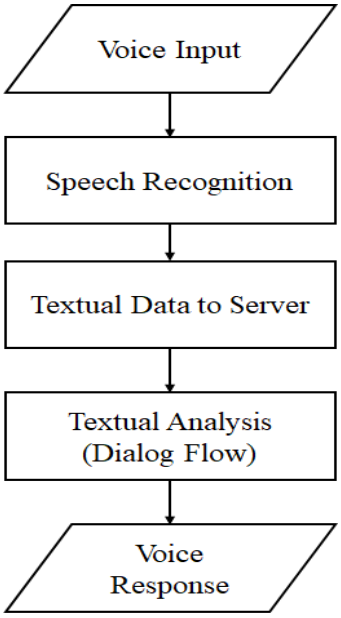

Fig -3: Voice Recognition flowchart

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.

4. IMPLEMENTATION

A. Functionalities

User can explicitly request for surrounding info to which the AI will respond providing the surrounding info. It consists of two main actor that is the user and AI. AI will respond to all user related queries. Application will capture the image and the image will be processed by AI and provide suitable response such as the object detected and their type.

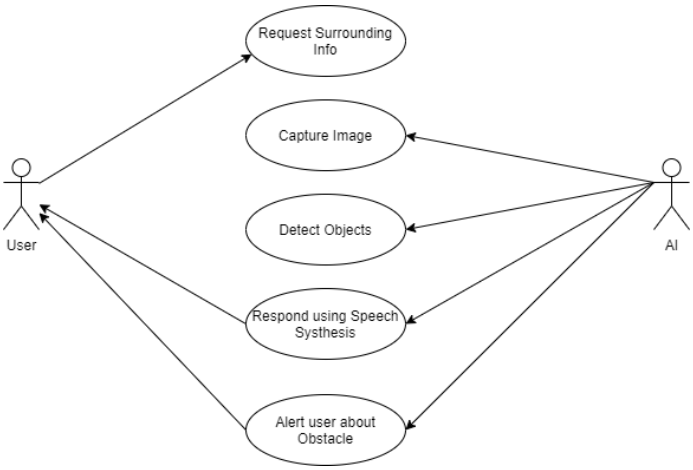

Fig -4: Use Case Diagram

Suitable human understandable response is generated using speech synthesis and provided to the user. This further enables user interaction to be limited to speech and makes the application friendly to visually impaired.

B. Application Flow

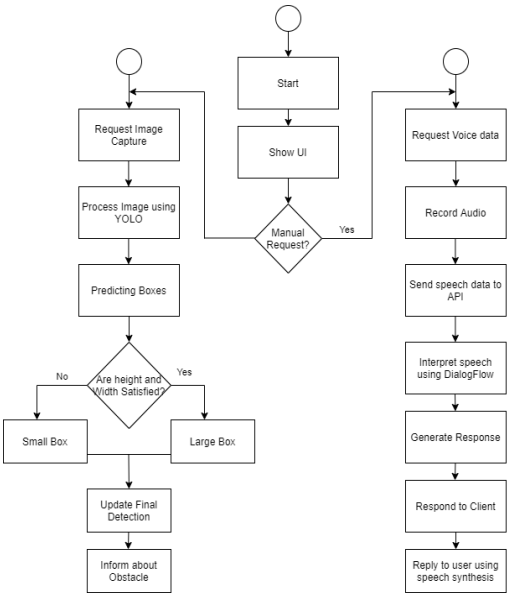

Fig -5: Activity Diagram

When application is started a UI showing surrounding with the help of a camera is displayed along with a microphone button. Clicking on the microphone will start a dialog flow communication enabling user to interact using speech directly. Basically when the user says something the data will be recorded and sent to server for processing and suitable response will be generated and sent back which will be provided to the user. An image processing service keeps running in the background enabling camera to capture surrounding data which is processed by YOLO algorithm and objects are detected. The user is then informed about the obstacles nearby

B. Working

To interact with the application, the user needs to launch app. After that the application will continuously request for video as shown in fig 6. The captured image data will be sent to service layer where all the processing in performed and the object is detected. A response will be sent from the service layer and the output will be displayed in the view or the Application screen. User can also request for manual object detection if required. Along with Image processing, Speech synthesis is also used to provide more accessibility to the user. The response from the application is in the form of speech, all the detected objects related information will be informed to the user using speech to assist the visually impaired

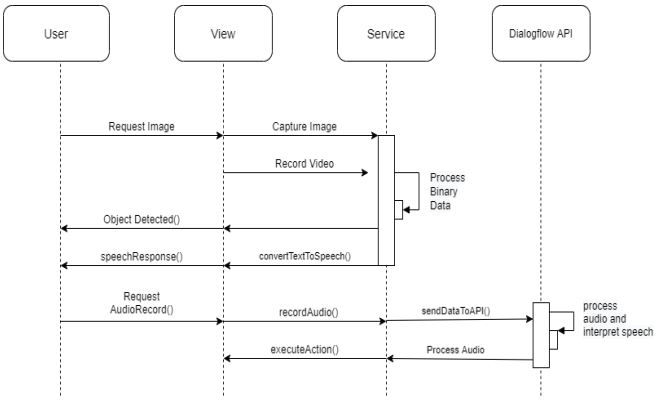

Fig -6: Sequence Diagram

5. OUTPUT

In the fig 7 we can see how the application is able to detect the object with the help of the camera. When the application is used, the camera used as a background service to assist the user by continuously processing surrounding data and providing information to the user. There is also a button for manual interaction, where the data will be processed by DialogFlow algorithm, a response will be generated and the User will be response in the form of a speech. The response time of the application is very less making it useful in a real-world scenario. In the fig 7 we can see that the teddy bear detected has an accuracy of 78% and the inference time is in the range of 200ms-300ms which is very low compared to other algorithms

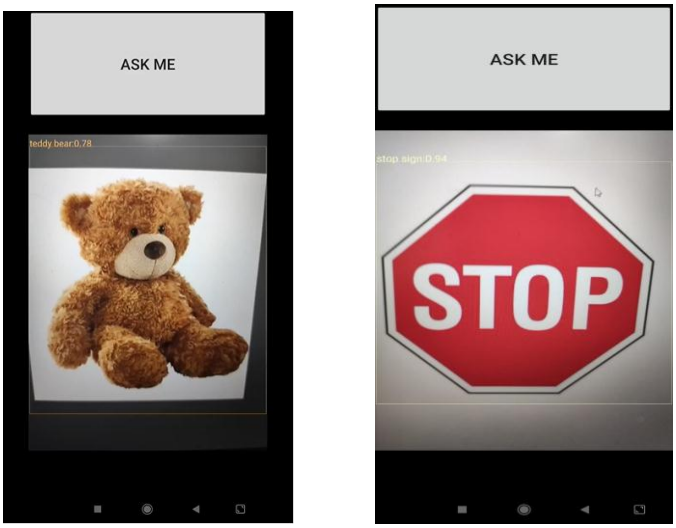

Fig -7: Output

6. CONCLUSIONS

In this paper, we have implemented the technology that attempts 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 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 enables visually challenged people to improve their lifestyle.

7. FUTURE SCOPE

We can provide more accessibility to the user by further adding features such as text recognition and live translation This will enable the user to interpret text detected in the image in front of them. Instead of using an Android Application, an IOT device can be implemented for the sole purpose of assisting the visually impaired. This will further reduce the cost and open more possibilities such as adding of other sensors to provide additional data such as distance and provide more precision and accuracy. It will also make the device more energy efficient as it won't be required to run heavy services in the background that are used in android to overcome restriction such as running of background service, energy usage etc.

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 forOject 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 IntelligentAssistant 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.