Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)
IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-5386-3570-4

A Smart Personal AI Assistant for Visually Impaired People

Shubham Melvin Felix, Sumer Kumar, and A. Veeramuthu

***Abstract*— In today's advanced hi-tech world, the need of independent living is recognized in case of visually impaired people who are facing main problem of social restrictiveness. They suffer in strange surroundings without any manual aid. Visual information is the basis for most tasks, so visually impaired people are at disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology, it is possible to extend the support given to people with visual impairment. This project is proposed to help those people who are blind or visually impaired using Artificial Intelligence, Machine Learning, Image and Text Recognition. The idea is implemented through Android mobile app that focuses on voice assistant, image recognition, currency recognition, e-book, chat bot etc. The app is capable to assist using voice command to recognize objects in the surrounding, do text analysis to recognize the text in the hard copy document.**

It will be an efficient way in which blind people can also interact with the environment with the help of technology and utilize the facilities of the technology.

***Index Terms*— Impaired, AI, Machine Learning, Voice Assistant, Chat bot, Image Recognition.**

I. Introduction

VISUAL impairment is a major disability faced by the visually challenge people. A person who cannot see can never feel the emotion that a person feels who can see the world. This visibility problem is a black dot faced by billions of people around the entire world. Our aim is to remove this black dot with the help Artificial Intelligence and Machine Learning.

Visual impairment present severe consequences on certain capabilities related to visual function:

i. The daily activities (that requires a vision at an average distance)
ii. Conversation, reading, writing (which requires a precise vision and average distance)
iii. Estimation of area and the displacement (which require a far vision)
iv. The tracking of an activity involves an extended care of optical observation.

The currently existing system uses speech synthesis to read e-books for the visually impaired using mobile application and converts the document/soft copy of the books to the speech using natural language and Text-to-Speech [1].

Major problem with the existing system is that it works only for single language (English), and not compatible with other languages. It does not work offline and always requires internet connection for the feedback/response.

The proposed system uses Artificial Intelligence to assist the visually impaired people which is all based on voice command. It also does image recognition of the photographs clicked or uses camera to recognize the objects and describes them in audio and also a chat bot to have light and friendly conversations.

The remaining sections are organized in following ways: in section II discussed about various related work done in this area. The proposed work like problem statement, complete system architecture and algorithm description are given in this section III. In section IV discussed about the results and discussion of the work was done. Finally, in section V concluded the work and also given the future direction.

II. Related Work

An object recognition system for visually impaired people [2]. This system will be a boon to the visually impaired person as well as society. It also helps in the detecting the direction of maximum brightness and major colors. A hybrid algorithm is proposed for object recognition in which the Artificial Neural Networks and Euclidean Distance measures are used in combination.

The video captured with the help of camera will be categorized into several frames, all the frames will be compared with the previous frames [3] and the data will be stored and respond will be given on the basis of stored data about an object.

Edge detection is the first step and it is one of the preprocessing steps involved in the process. Edges may be defined as a boundary between two homogeneous regions. Important points found can be collected from the edges of an

Shubham Melvin Felix, Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai (e-mail: felix.melvin117@gmail.com).

Sumer Kumar, Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai (e-mail: sumer.sk57@gmail.com).

A.Veeramuthu, Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai (e-mail: aveeramuthu@gmail.com).

image which were detected. Canny operator is used in this project to figure out the edges. An edge is typically founded by computing the derivative of the image intensity function [4].

A file has to get selected in any format by the user it can be in either .ppt or .doc file format. When the file gets selected it is then converted into pdf file format and then the text present in the file format are reconstructed as a collection of words [5]. The information collected from the image is the filtered and shown on the output screen.

The technique proposed based on key points extraction and matching in video. A comparison is done between the numbers of frames and database objects is made to detect the object in each frames present [6]. When an object is found the audio containing the information about it is activated.

Three software tools developed for Android Smartphones image processing module, a module which detects colors, object as well as light source detector. The algorithm works on the images taken with a flash which is automatic with possibly very small resolutions. The HIS (Hue Saturation Intensity) [7] conversion takes place for the RGB color images.

An image reading tool which read the text present in image. The properties of image and text is a very important work to be checked in the image. The text present in the picture or image can be of many types, it can be of scene text or Graphics text which depends on the image source as like whether the image is a machine generated image or it is captured from a camera. The color of the image is taken and it get converted into Greyscale image [8]. Preprocessing of image is done to clear the challenges created by noise and uneven lighting in image.

The interest point is find out with the help of local feature extraction method for which a feature vector and descriptor is computed. SIFT the earliest schema [9] for feature extraction. It helps in the representation of that image is collection of interest points which are not variant to image transformation and partial to illumination changes. The method is based on Gabor filters [10] for texture characterization which is used for variety of texture segmentation and classification tasks.

III. Proposed System

A. Problem Description

Blind people come across a number of challenges in everyday life from reading a book to walk on the street. Although many tools are available to meet the challenges faced by them, but they are not sufficient. Vision is the most essentials thing a human can have and it plays a very essential role in the life of a person either a person can see or not. A visually challenged people need an assistant even for working a common daily routine work.

In this paper we have discussed the challenges faced by blind people and tried to provide a satisfactory solution to them for working everyday life.

B. System Architecture

Cloud APIs of Google are an integral of Google Cloud Platform, allowing user to easily add the power of everything from storage access to ML based image analysis to the user application. All the Cloud APIs works a simple JSON REST interface that are called directly or via the client libraries. Google Cloud API architecture mainly use chat-bot client used for the vision as well as Dialog-Flow which is shown in "Fig. 1". It is used for speech recognition as well as translation of the textual documents. A Web-Hook is a HTTP callback: a HTTP POST that happens when some request is made a direct notification is sent by means of HTTP POST. A web

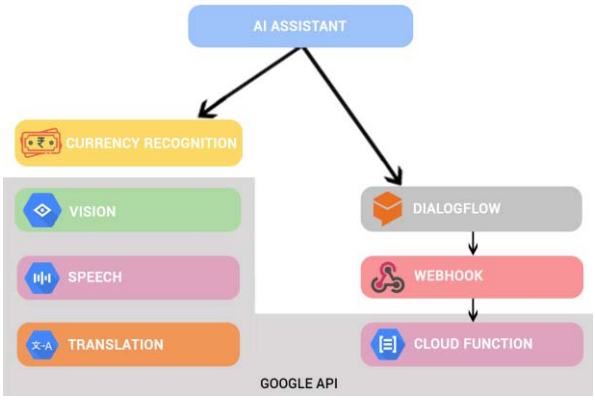

Fig. 1 System Architecture

application running WebHook will trigger a message to the link when request is made.

C. Modules

We all know the importance of vision. A person is incomplete if he/she cannot see. If a person cannot see, he/she has to face many challenges in the society to live.

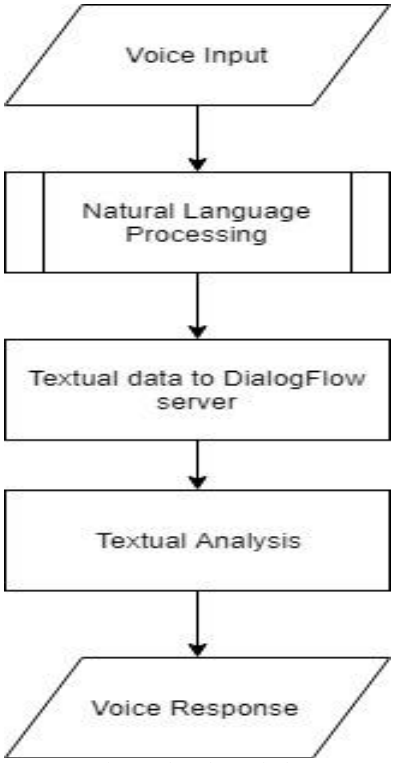

Fig. 2: Flowchart of Chat Bot

Our proposed system handles the difficulties faced by the visually challenged people. With the help of Artificial Intelligence [11] and Machine learning, we have tried to solve some of the problems faced by the visually challenged people. Dialog-Flow is a platform for building natural and rich conversational experiences. A chat-bot with an assistant as like google assistant, which will help them to talk and get the desired responses which is shown in "Fig. 2". A chat-bot is an interface, which takes the input from the user either in the term of voice or text and respond back with the satisfactory reply. A person can ask the chat-bot to guide the person with either voice or text to know about some place or if he has to visit any place the chat bot can guide with the map with voice assistant. Dialog-Flow platform has been used for training the chat-bot so that the person can interact with it. It uses keywords given by the user, the keywords are compared with the trained data, and the response is given back.

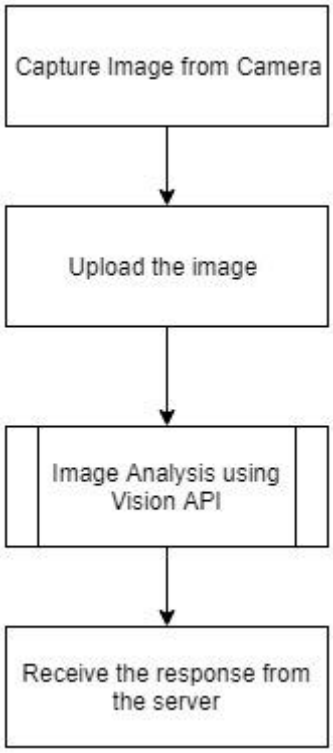

Fig. 3 Flowchart for working of Vision API

The REST API is used to analyze the image captured, which is provided by Cloud Vision API which encapsulate powerful machine learning. It rapidly groups pictures into many classifications (e.g., "Taj Mahal", "Deer", "Footwear"), identifies singular protests and face inside pictures and finds and peruses printed words, which contained inside pictures which is shown in "Fig. 3". This stage is utilized to distinguish the picture caught by the mobile camera. After the picture examination, the insights about the picture is react back. As like on the off chance that somebody click a picture of Taj Mahal, Google Vision API will process the picture and will react back with the coveted information in either voice or printed information. In Google Vision API the captured images can be categorized in following classes as like Landmark Detection, Logo Detection, Explicit Content Detection, Image Properties, Label Detection and Document Text Detection.

With the application of neural system and effective model in a very simple way to use an API. The Cloud Speech API helps the designer to change the voice input to text. The API provides 100 plus variations and dialects to help the customer all over the world. You can decipher the content of clients directing to an application's mic, it active summon and it is controlled by voice command, or interpret sound documents, among many other cases. Our proposed framework utilizes the Google Cloud Speech API to change over the voice contribution of the client into literary information and can send the printed information to informing android application as like Message application, WhatsApp and so on.

D. Algorithm

```
id3(exp, characteristics)
      Begin
    /*exp are the training examples and the characteristics is a list of attributes that will be tested by the trained decision tree.  It returns a tree that classifies the given exp. This algorithm uses to build the AI chat bot to give appropriate reply to the user based on the key words present in the question and those key words are searched in the trained dataset to produce the appropriate output in the sound format */
       n = dTNode(exp)
      // handle target characteristics with characteristics labels
     dict=sumUpExp(exp, tAttr)
      for key in dic:
      Begin
         if dict [key] == total no. of exp
           n.label = key
           return n
      End
      //test for no. of exp to avoid overfitting
    if characteristics are empty or no. of exp < min per branch:
         n.label = most common value in exp
         return n
      bestAtr = characteristics with the most infoGain
      n.decision = bestAtr
     for each possible value v of bestAtr:
        Begin
       subset = the subset of exp that have value V for bestAtr
        if subset is not empty:
           n.addBranch(id3(subset,tAttr,attrbestA))
      End
       return n
End
```
Algorithm 1: Decision Tree Algorithm

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm developed by Ross Quinlan used to produce a decision tree from a dataset which is given in Algorithm 1. ID3 is the antecedent to the C4.5 and is utilized as a part of the machine learning and natural language preparing spaces.

The accompanying algorithm is utilized as a part of Dialog-Flow to settle on a choice as indicated by the client voice summon and classify its reaction based on enter component in a sentence, then the key elements are searched using decision tree algorithm in the trained dataset.

Information gain IG(A) is the measure of the distinction in entropy from before to after the set S is part on an attribute A which is given in Algorithm 2.

```
infoGain(exp, characteristics, eOfSet)
    Begin
      g = eOfSet
      for value in characteristics Values(exp, characteristics):
      Begin
        sub = subset(exp, characteristics, value)
    gain-=(no. in sub)/(total no. of exp)* entropy(sub)
      End
      return g
  End
```

Algorithm 2: Information Gain

IV. Results and Discussion

In “Fig. 4” shows the desired output obtained from the android application. Output shown in the chat-bot are as per the training given to the machine with the help of Dialog-Flow and the responses according to the input given to the machine.

Fig. 5 Image Recognition GUI

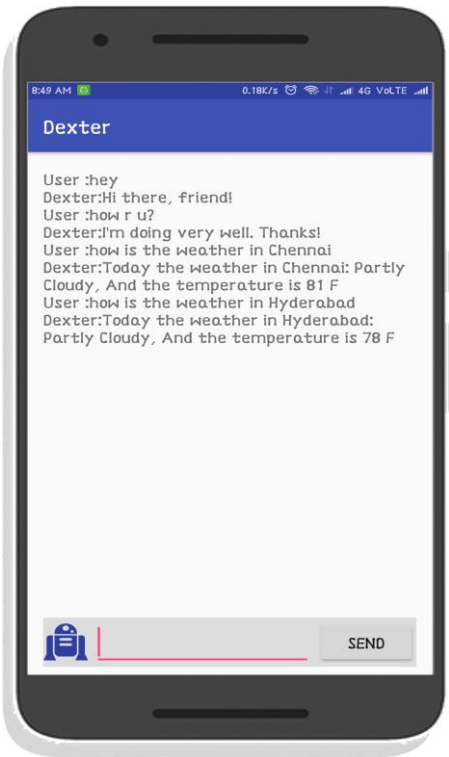

Fig. 4 Chat Bot GUI

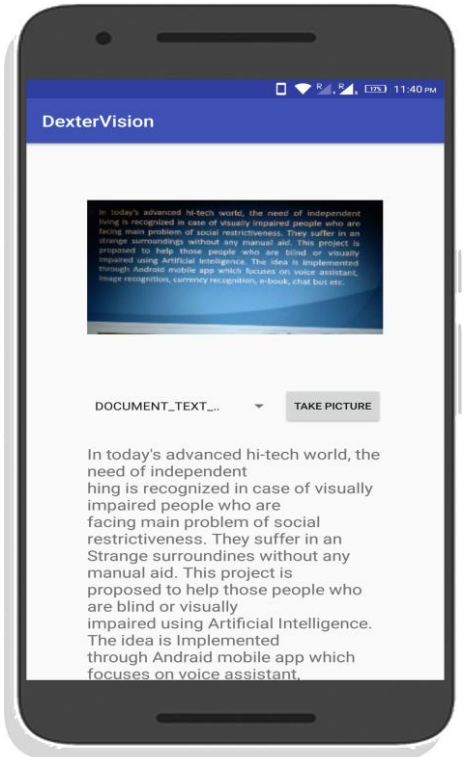

Fig. 6 Textual Recognition GUI

Google Cloud Vision API shows the data in the output screen is shown in "Fig. 5" after the analysis of the Image captured by the mobile camera. Thus, based on the image category the analysis of the image is done the JSON data responded back to the app with different parameters of the image. Thus, based on the retrieved JSON data the app tells the appropriate result with the help of confidence score.

With the application of neural system and effective model in a very simple way to use an API. The Cloud Speech API helps the designer to change the voice input to text. In "Fig. 6" demonstrates the working Chat-Bot and the discussion between the client and application.

All the results are shown in "Fig. 6" is as per the training given to the machine through Dialog-Flow platform. The machine needs the training so that it can capture the input keyword and respond with desired output.

V. Performance Analysis

The table compare features between different research papers according to the accuracy of features which is shown in "Table I". The comparison done between object recognition, landmark detection, textual analysis, interactive, voice input as well as response and multilingual properties. From the comparison table we can conclude that our proposed system acquires more features than others do.

Features	M 1	M 2	M 3	M 4
Object Recognition	✓	✓	✓	✗
Landmark Detection	✓	✗	✗	✗
Textual Analysis	✓	✗	✗	✓
Interactive	✓	✗	✗	✓
Voice Input & Response	✓	✗	✓	✗
Multilingual	✗	✗	✗	✗

Table 1: Features comparison; M 1: Personal AI Assistant for Visually Impaired (Proposed); M 2: Visual-Pal: A Mobile App for Object Recognition for the Visually Impaired; M 3: Image Recognition for Visually Impaired People by Sound; M 4: Character Detection and Recognition System for Visually Impaired

The following notations are used in graph; M1: Personal AI Assistant for Visually Impaired; M2: Visual-Pal: A Mobile App for Object Recognition for the Visually Impaired; M3: Image Recognition for Visually Impaired People by Sound; M4: Character Detection and Recognition System for Visually Impaired People.

The accuracy appears to have a higher percentile which is shown in "Fig. 7". The above bar graph shows the accuracy comparison between our paper and other papers. The chart shows that our proposed system has more accuracy with respect to other proposed system. Our proposed system M1 has accuracy in between 80 to 90, which is the maximum accuracy among all.

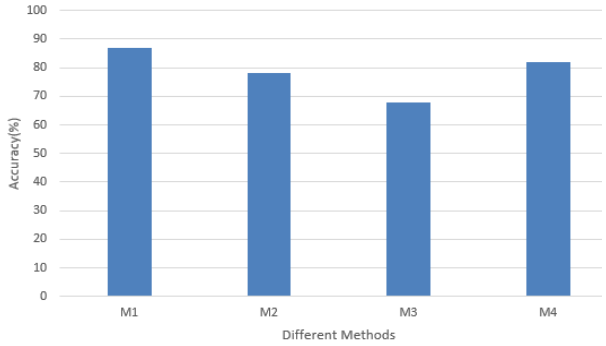

Fig. 7 Accuracy

The sensitivity of M2 appears to have a higher percentile above all which is shown in "Fig. 8". The sensitivity comparison between our work and other works. The chart shows that our proposed system has little less sensitivity with respect to M2 and higher with respect to other proposed system. Our proposed system M1 has sensitivity above 90, which is the second maximum sensitivity among all.

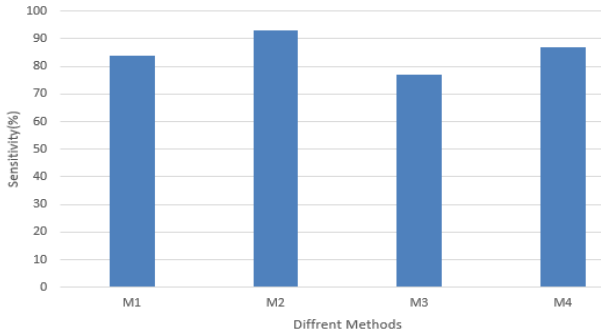

Fig. 8 Sensitivity

The feasibility of our proposed system appears to have a higher percentile above all which is shown in "Fig. 9". The chart shows that our proposed system has more optimal solution with respect to other proposed system. Our proposed system M1 has feasible above 90, which is the maximum among all.

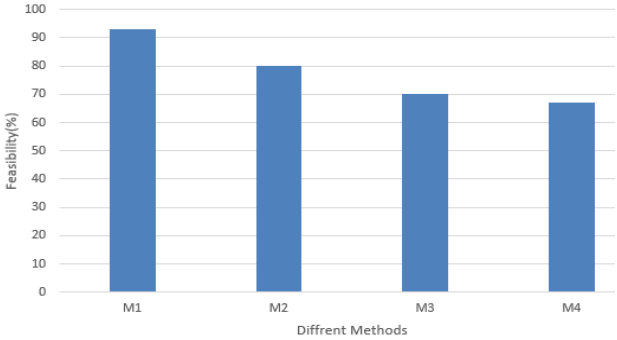

Fig. 9 Feasibility

VI. CONCLUSION

Artificial Intelligence and machine learning is one of the most growing technology. These technologies are the playing vital role in the development of the IT sector. Here we have tried to use these technologies for the visually challenged people so that they can also live an independent and normal life. The friendly chat with the bot Image recognition of the objects and surroundings. Currency recognition to help in the easy payment. Text recognition and reading of the analyzed text. The development of the proposed system if is completed, it can serve the visually challenged people with a better assistant.

In coming days our proposed system can be applied in multilingual application so that a person can use the application in their own language without any trouble. In addition, our proposed system can be deployed with the IoT. In future our proposed system will be able interpret the textual description in a much better way. The Image recognition can be enhanced with much more details about the image captured through the camera. Enhancement to this system can be done by adding the features of currency recognition [12]. The existing methodology for image and currency recognition can be done with more accuracy.

REFERENCES

[1] 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.

[2] 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).

[3] 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.

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

[5] 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.

[6] Hanen Jabnoun, F aouzi Benzarti, Hamid Amiri , "Object Detection and Identification for Blind People in Video Scene", Universite de Tunis EI Manar, Ecole Nationale d'Ingenieur de Tunis 1002, Tunis Le Belvedere, Tunisie.

[7] K.Gopala Krishnan, C.M.Porkodi, K.Kanimozhi, "Image Recognition For Visuall Impaired People By Sound", International conference on Communication and Signal Processing, April 3-5, 2013, India.

[8] Akhilesh A. Panchal, Shrugal Varde, M.S. Panse," Character Detection and Recognition System for Visually Impaired People", IEEE International Conference on Recent Trends in Electronics Information Communication Technology, May 20-21, 2016, India.

[9] Nada N. Saeed, Mohammed A.-M. Salem, Alaa Khamis," Android-Based Obect Recognition for the Visually Impaired", German University in Cairo, Ain Shams University.

[10] Vincent Gaudissart, Silvio Ferreira, Celine Thillou, Bernard Gosselin, "Mobile Reading Assistant for Blind People", SPECOM'2004: 9th Conference Speech and Computer St. Petersburg, Russia September 20-22, 2004.

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

[12] Noura A. Semary, Sondos M. Fadl, Magda S. Essa, Ahmed F. Gad," Currency Recognition System for Visually Impaired: Egyptian Banknote as a Study Case", Menoufia, Egypt, 2015.