AI Based App for Blind People

Tufel Ali Qureshi[1], Mahima Rajbhar[2], Yukta Pisat[3], Prof. Vijay Bhosale[4]

[1,2,3]*Student, Dept. of Computer Engineering, M.G.M. College of Engineering and Technology, Kamothe, Maharashtra, India*
[4]*HOD. Dept. of Computer Engineering, M.G.M College of Engineering and Technology, Kamothe, Maharashtra, India*

---***---

Abstract - *The World Health Organization (WHO) estimates that globally, approximately 1.3 billion people live with some form of visual impairment. Thanks to rapid advancements in technology, every now and then some form of innovation makes a positive impact in the lives of visually impaired people. The need for developing a low-cost assistive system for the visually impaired and blind people has increased with steady increase in their population worldwide. The system presented in the paper uses artificial intelligence in real time to help the visually disabled people to navigate their environment independently. Image recognition and obstacle detection are the tasks performed by the system. The image recognition task was performed using a smartphone application powered by artificial intelligence.*

Key Words: Deep learning, Image processing, Image recognition, Speech recognition, Voice output, YOLO Algorithm,

1. INTRODUCTION

In todays 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. Due to lack of necessary information in the surrounding environment visually impaired people face problems and are at disadvantage since visual information is what they lack the most. With the help of the advanced technology, the visually impaired can be supported. 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 may be the effective way blind people will interact with other people and may help blind people independent life.

Visually impaired are the ones who are completely or partially blind. According to an estimate made by the World Health Organization (WHO) 285 million of the population suffer from visual impairment and 39 people were blind and approximately 3% of all the ages in a nation are visually impaired [1].

The leading causes of blindness are cataract, trachoma, glaucoma and also deficiency like Vitamin A, onchocerciasis, and leprosy. People who are visually impaired suffer a lot and face great challenges in their day-to-day life for instance finding their way and directions and to places which they do not visit often.

The severe consequences visual impairment presents on certain capabilities related to visual function:

1.Day to day activities (requires average distance vision)
2.Communicating, reading, writing (requires a precise vision)
3.Estimating area and displacement (requires far vision)
4.Extended care of optical observation is needed for tracking activities.

2. PROBLEM STATEMENT

Blind people come across a number of challenges in everyday life from reading a book to walk on the street. Many tools are available to meet the challenges faced by them, but they are not sufficient. The most essentials thing a human can have is vision 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 to carry on work on daily basis. We have discussed the challenges faced by blind people and tried to provide a satisfactory solution to them for working everyday life.

3. EXISTING SYSTEM

Walking cane which detects obstacles on the way using IR sensors and issues a buzzer sound as warning. The white cane can only be used to detect obstacles up to knee-level within a limited range of 2-3 feet. The major drawback is that the buzzer sound can be unheard of in the traffic sounds on the roads. In addition to that Chaitrali et al proposed a navigation device uses voice output for. issuing warnings when obstacles are detected. Here, obstacles are detected with a help of IR sensors and Radio Frequency Identification (RFID) technology. This device is connected to an android phone through Bluetooth. An android application is designed which gives voice navigation based on RFID tags read and also updates person's location information on the server. Moreover, one more application is designed for family members to access the blind person's location through the server whenever needed [3].

Seeing AI is an iPhone app that uses Artificial intelligence to tell the visually impaired person what is around them. For example, if the phone is pointed at a park, the camera app, describing how the scene looks. Seeing AI is powered by Microsoft.
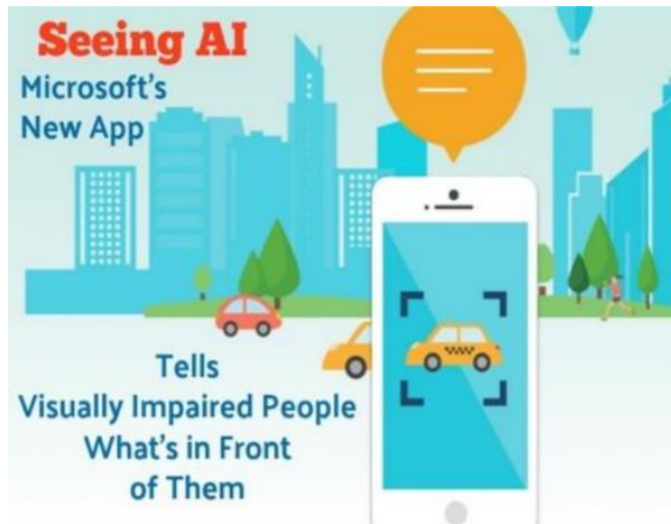

Fig -1: Seeing AI by Microsoft

AI Glass combines glasses with stereo sound sensors and GPS technology attached to the tablet, which can give spoken directions and recognize currency, read signs, identify colors, and other things. It also employs machine learning to recognize different places and objects. Because it uses ultrasound, it can also detect translucent obstacles, like glass doors.

4. LITERATURE SURVEY

The References surveys tell us that our society has a significant number of physically challenged people and large proportion of them are visually impaired. Vision impairment classifications:

(1) Mild impairment – presenting visual acuity worse than 6/12
(2) Moderate impairment– presenting visual acuity worse than 6/18
(3) Severe impairment– presenting visual acuity worse than 6/60
(4) Blindness – presenting visual acuity worse than 3/60

Among all the senses, ability to see through the eyes is one of the most important senses in human beings. And the loss of this ability severely affects all the possible movements an individual is likely to do in his/her life. Since such people are not expected to grow in their profession as much as an abled person, they often experience a violation of their rights and also discrimination on social platforms & at the workplace. Government and social groups are actively participating in making life of visually impaired more convenient and safer by organizing campaigns and providing education about the new tools and technology.

In [7], a method is proposed where the image is captured using the camera and the captured image is scanned from left to right for detection of the obstacle and then sound is generated. Sound is generated by analyzing the image where top image is altered into high frequency and the bottom portion into low frequency sound. And the loudness depends on the brightness of the image as well. The image is differentiated into foreground and background using image processing techniques. The foreground is assigned with high intensity value and background are assigned with low intensity values. Then this image is converted into stereo sound where the amplitude of the sound is directly proportional to intensity of image pixels, and the frequency of sound is inversely proportional to vertical orientation of pixels.

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

An Intelligent Assistant for Blind named Blind Reader 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 recognize Text through image Google's Vision API is used.

5. PROPOSED SYSTEM

Proposed methodology of object/obstacle detection works in a way that it involves several processes from extracting frames to the recognized output in the binary image. As the blind people are more vulnerable to the attackers than the sighted people, an assistant is required when they want to go outdoor, as we have discussed earlier it seems quite challenging to take an animal assistance as there is a high chance of the animal to be harm also. So, it is better to use technology as it can protect the user and does not get hurt. When people display a facial expression of emotion, we make judgments not only about their affective state, but also about their behavioral tendencies and traits. For example, when people display happy faces, we perceive them as having traits associated with high affiliation and high dominance. When they display angry faces, we perceive them as having traits associated with low affiliation and high dominance.

5.1 Image Pre-processing

The initial task of Image and Video pre-processing is to drop frames from the video data. A modern camera can record videos that have at least 20 – 30 frames per second. Since the person is not moving at a high pace, we can drop frames from the video to minimize the processing. To further reduce superfluous processing, we will convert the images/frames into a grayscale format. An image consists of pixels and every pixel holds its RGB color values. By reducing the number of values stored by our pixels, we can minimize our processing by a large margin.

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 [8].

5.2 Image Pre-processing

The initial task of Image and Video pre-processing is to drop frames from the video data. A modern camera can record videos that have at least 20 – 30 frames per second. Since the person is not moving at a high pace, we can drop frames from the video to minimize the processing. To further reduce superfluous processing, we will convert the images/frames into a grayscale format. An image consists of pixels and every pixel holds its RGB colour values. By reducing the number of values stored by our pixels, we can minimize our processing by a large margin. Grayscale conversion carried out by retrieving the RGB contents from the image taking average of it and assign to the new pixel. After converting it to grayscale, thresholding of the image is performed. The threshold depends upon the environment in which the system will be used. An optimal threshold is needed for proper edge detection. If a pixel exceeds certain thresholds, it is classified as an edge. We will get a barebones and skeletal representation of the environment through these methods. 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.

5.3 Object-based Classification

Object-based or object-oriented classification uses both spectral and spatial information for classification. The process involves categorization of pixels based on their spectral characteristics, shape, texture and spatial relationship with the surrounding pixels. Object-based

classification methods were developed relatively recently compared to traditional pixel-based classification techniques. While pixel-based classification is based solely on the spectral information in each pixel, object-based classification is based on information from a set of similar pixels called objects or image objects. Image objects or features are groups of pixels that are similar to one another based on the spectral properties (i.e., colour), size, shape, and texture, as well as context from a neighbourhood surrounding the pixels. Object-based classification is a two-step process, first the image is segmented or broken into discrete objects or features with and then each object is classified. This type of classification attempts to mimic the type of analysis done by humans during visual interpretation.

5.4 Information Generation

We will be generating info according to the classification of the objects in the frame. These instructions will be sent to the text to speech system and then communicated to the user using audio.
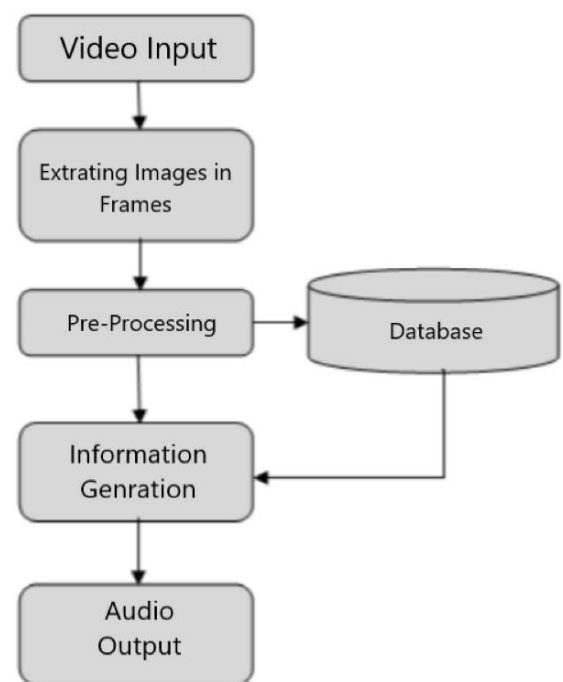

Fig -2: Architecture

6. IMPLEMENTATION

6.1 User Flow

User can request for surrounding information to which the AI system will respond by providing the surrounding information. It consists of two main actors which are the user and AI system. The AI system 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 -3: UR Diagram

6.2 Object Detection Pipeline

1. Camera will run in a loop and taking picture at rate of about 2Hz.
2. Picture will go through face detection, and any face bound will be drawn on screen as a preview.
3. System then decides whether it want to only detect the biggest face in the frame, and, the whole picture will be sent to CNN.
4. If user chose to only detect biggest face, then the head and shoulder area of the biggest face bound will be cropped out, and the new picture will be sent to CNN for classification.
5. CCN will output a N dimension vector of confidences of detected features.
6. Feature vector will be feed into the logistic regression prediction function, then give probability/score for each of the classes.
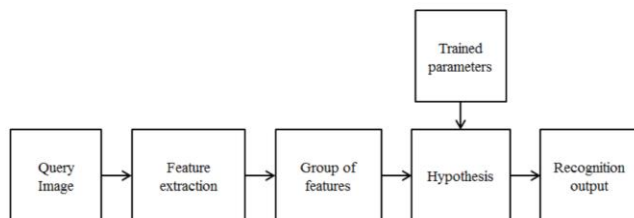

Fig -4: Object Detection Pipeline

6.3 Convolutional Neural Network

Convolutional neural network (CNN) is designed to recognize 2D shapes irrespective of object translation, rotation, scaling etc. CNN is neurobiologically motivated, it is first proposed by LeCun, the basic form of convolutional neural network consists of three main operations as follows:

- feature extraction.
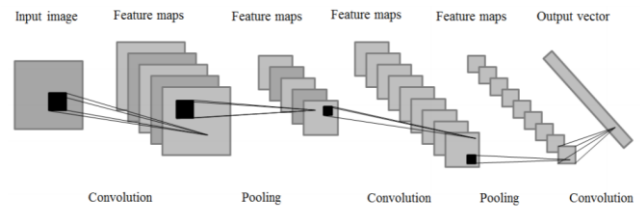- Feature mapping.
- Subsampling.

Fig -5: CNN Model

7. RESULT

As 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 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 2Hz which is very low compared to other algorithms.

Fig -6: Ouput Image

1) At first, we are capturing real time images from the rear camera of the mobile handset of blind people and a connection is established between mobile phone and system in laptop and then those images are sent from the mobile phone to laptop.
2) This connection is done by a app which is installed in the mobile phone of the person. All the real time images

which get captured by the rear camera of the mobile phone are first transferred to the app in the mobile phone and then those images are sent in laptop where they are processed for some further conclusions.

3) The system in laptop will test it using its APIs and SSD ALGORITHM and it detects the confidence accuracy of the image which it is testing. We reached 98% accuracy for certain classes like books, cups, remote.

4) After testing the images, we are generating an output on the laptop and its prediction is being translated into voice with voice modules and sent to the blind person with the help of wireless audio.

8. CONCLUSION

The technologies like artificial intelligence and machine learning plays a vital role in the development of the IT sector. We have made use of these technologies for the visually impaired people so that they too can lead a normal and independent life like other people. The friendly chat bot helps the visually challenged to recognize the objects and surroundings. Currency recognition helps in easy payment. Text recognition helps in reading and analyzing text. The development of the proposed system if is completed, it can serve the visually challenged people with a better assistant.

REFERENCES

[1] World Health Organization, "Visual Impairment and Blindness," WHO Factsheet no. FS282 , Dec. 2014.

[2] Mingmin Zhao, FadelAdib, Dina Katabi Emotion Recognition using wireless signals.

[3] N. Senthil kumar, A. Abinaya, E. Arthi, M. Atchaya, M. Elakkiya, "SMART EYE FOR VISUALLY IMPAIRED PEOPLE", International Research Journal of Engineering and Technology, Volume: 07 Issue: 06, June 2020.

[4] Liang – Bi Chen, Ming-Che Chen, "An implementation of an intelligent assistace system for visually impaired/blind people, "IEEE, 2018.

[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] Shreyash Patil, Oshin Gawande, Shivam Kumar, Pradip Shewale,"Assistant Systems for the Visually Impaired", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 01 | Jan 2020.

[8] Gagandeep Singh, Omkar Kandale, Kevin Takhtani, Nandini Dadhwal, "A Smart Personal AI Assistant for Visually Impaired People", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06 | June 2020.