

A Smart Vision Based System for Assisting Visually Impaired People

Aysha Nafla P¹, SruthiRajan MK²

¹PG Scholar, Dept. Computer Science & Engineering, M.Dasan Institute Technology Ulliyeri, Koyilandi, Kozhikoode ,673620

²Assistant Professor, Dept. Computer Science & Engineering, M.Dasan Institute Technology Ulliyeri, Koyilandi, Kozhikoode ,673620

ABSTRACT

Eyes are most important sensory organ of a human body. A person is visually impaired if he cannot see objects as clearly as usual. So its difficult for a visually impaired one to live as normal people can do. They always required some assistive devices or assistants to guide them. To address the concerns, the suggested system will perform multiple tasks as quickly as feasible and accurately without requiring any specific talents. The proposed system bring together the deep learning and cloud APIs for smooth functioning of multiple tasks like face recognition, currency recognition, object labeling, text recognition, online news paper reading, current location ,weather and and date and time accessing. The user interact with the system by providing specific voice command like “who is in front of me” and system trigger corresponding module and return voice output. Face recognition is carried based on dlib’s face recognition. Image labeling and text recognition by Google cloud vision API, currency recognition for Indian currency notes is proposed to do on deep transfer learning model Resnet 101 ,online news paper reading is with the help of Google News API, Weather access is by Open Weather Map API , location can be accessed with the help of IP address and datetime module of python provide current date and time.

Keywords-Visually impaired, Deep learning, Google cloud vision API, Google News API, Open WeatherMap API, Resnet101.

1. INTRODUCTION

Among the sensory organs of human being eyes are the most important one. Vision impairment may be caused by a loss of visual acuity, where the eye does not see objects as clearly as usual. The World Health Organization (WHO) puts the total count of blind people in India at around 63 million, almost 20 per cent of the world blind population. It’s really hard for a visually impaired person to live their life without any assistance. They wish to recognize the barriers in front of them, recognise things and familiar faces, need to read texts, and so on. They may require a companion for assisting or any equipment for doing their needs. It’s quite burdensome to guide always these visually impaired by his friend or family. So its imperative to build assisting devices or system to guide these peoples. As the improvement in technologies there are several systems designed to support visually-impaired people and to improve the quality of their lives. Unfortunately, most of these systems are limited in their capabilities.

The simplest and the most affordable navigations and available tools are trained dogs and the white cane [1]. Although these tools are very popular, they cannot provide the blind with all information and features for safe mobility, which are available to people with sight [2,3]. The proposed smart vision based system will help the visually impaired people in many ways, such as by describing the surroundings, recognizing familiar faces, recognizing Indian currency notes, reading out texts, providing the latest information via an online newspaper and providing information about current location , weather condition and date and time.

The ability to recognize the faces of well-known people is one of the most difficult tasks for

visually impaired people. As technology advances, new methods for recognizing familiar face has emerged. Different technologies including early algorithms, artificial features and classifiers, deep learning and other stages are present[4]. The proposed system uses dlib's face recognition package for face recognition [5]. The dlib's HOG + Linear SVM based model [6] is used here. This model has a quick response time compared to others. Blind people do, in fact, have visual dreams. They want to know what amenities, things, or activities are in front of them. Google vision API will help them to do this endeavor[7]. The Google Vision API can also assist visually challenged people in reading text from an image. It promotes their confidence and autonomy by offering information that is beneficial in everyday life[8]. People who are visually impaired have a hard time distinguishing between different currency denominations. So, the proposed system e investigates the systems that can assist visually challenged or handicapped people in distinguishing between different types of Indian currencies deep learning technique. The framework utilizes the concept of transfer learning where a deep convolutional neural network already trained upon a huge dataset of natural images is re-utilized for the problem of classification of denomination from banknote images[9]. The pretrained model used is ResNet 101. This model recognizes currency notes of 10,20,50,100,200,500 and 2000 . The smart vision system uses the ip address for knowing the current location, region and the country, also uses the OpenWeatherMap API to get the current temperature and the weather description for the day [10]. The datetime module is used by the smart vision system to obtain the current date and time.

2. RELATED WORKS

Several assistive systems for the visually impaired have been developed. Visually impaired people face lot of difficulties in their daily life. Most of the times they depend on others for help. Several technologies for assistance of visually impaired people have been developed. Among the various technologies being utilized to assist the blind, Computer Vision based solutions are emerging as one of the most promising options due to their affordability and accessibility. The main objective of the proposed system is to create a wearable visual aid for visually impaired people in which speech commands are accepted from the user. Its functionality addresses identification of objects and sign boards. This will help the visually impaired people to manage day-to-day activities and to navigate through their surroundings. Raspberry Pi is used to implement artificial vision using python language on the Open CV platform. [11]. Google cloud vision's remote server API allows to integrate the software into any kind of android device and assist the visually impaired. This proposal attempts to explore the possibility of using the hearing sense to understand visual objects. Computer vision provides several algorithms and techniques to achieve this goal. This paper propose a real-time environment perception system, with the goal of informing the user about the presence around them and their spatial position using binaural sound[12].

With the advances in camera technologies, mobile platforms and light-based pointers, this propose a new and cost-effective solution for autonomous obstacle detection, classification and navigation. Without loss of generality, the proposed system point out and focus on the application of visual navigation and obstacle detection for the visually impaired [13]. Another work proposes image processing based identification of familiar places (Restroom, Pharmacies and Metro train station). This is done by point feature matching using template detection[14]. Another system helps the blind to navigate independently using real time object detection and identification. The proposed system consists of a Raspberry Pi-3 processor which is loaded with a pre-trained Convolutional Neural Network model (CNN) developed using TensorFlow. The pre trained object detection model is `ssd_mobilenet_v1_coco` [15]. One study, propose a deep learning framework for image detection, classification and person-currency recognition. Transfer learning is performed on the SSD-VGG16

model to predict outputs [16]. A smart glass system to recognize the family member using image processing is developed . Face recognition is performed by Haar features [17]

3. PROPOSED METHODOLOGY

The proposed system is a smart vision based system which assist visually impaired by doing multiple tasks for their better life style. The system is based on some deep learning models and uses some APIs such as Google vision API to carry out these tasks. The system will take the users audio as input and convert to text .These commands trigger corresponding module to function. The conversion of audio to text is based on using Google’s speech-to-text library. The corresponding text query for these audio input trigger parallel module. If the audio is like “Who is in front of me” then it will trigger face recognition module. If the voice command is like “which currency is this” then the system will start loading currency recognition module and so on. The system continue listening until the “stop” command is recognized.

3.1 Face Recognition Module

Face recognition is a two-step procedure that begins with face detection and ends with recognition. The dlibs face recognition library is used for face recognition. The first stage is to look for faces in the input image. For this, the Histogram of Oriented Gradients, or simply HOG, is used. HOG construct a reduced version of an image, and encode it. All we have to do to locate faces in this HOG image is seek for the area of our image that looks the most like a known HOG pattern generated from a bunch of other training faces. Fig 1 shows an example of how HOG of face look like.



Fig.1. Representation of HOG of an image

This is how a face is detected in an image. If a face is detected, the (x, y) coordinates of the 194 important spots on the face are detected using dlib's facial landmark detection model [18]. A 128-dimensional NumPy array or feature vector is generated which store the mapped facial features. The system already has a database of known individuals. Face recognition is accomplished by comparing the feature vectors of the input image and the feature vectors of the pre-stored image. The person's face is the name given to the feature vector that has the highest similarity value. Linear SVM machine learning classification algorithm can be used to do this.

Image Captioning Module

The Google Vision API can detect and retrieve data about items in a picture from a wide range of categories. This can be used to identify a variety of things, including generic objects, locations, activities, animal species, products, and more [19]. By delivering the contents of a local picture file as a base64 encoded string in the body of the request, the Vision API may perform feature detection on it.

3.2 Text Recognition Module

To read text, Google's Vision API is used, which provides accurate results without sacrificing latency. The API is capable of detecting text in both documents and natural scene photos. It provides

a structured hierarchical response of the identified text, divided into pages, blocks, paragraphs, words, and symbols, as well as their x and y coordinates [20]. It extracts machine-encoded text from any image and outputs it (e.g., photos of street views or sceneries). The returned JSON file contains the complete strings as well as individual words and their bounding boxes.

3.3 Currency Recognition Module

This module recognize Indian currency notes of 10,20,50,100,200,500,and 2000 using a deep learning approach. The system employs the notion of transfer learning, in which a deep convolutional neural network that has already been trained on a large dataset of natural images is re-used to solve the problem of denomination classification from banknote images..The pre-trained model used here is ResNet 101. Architecture of Resnet 101 is shown in figure 2.

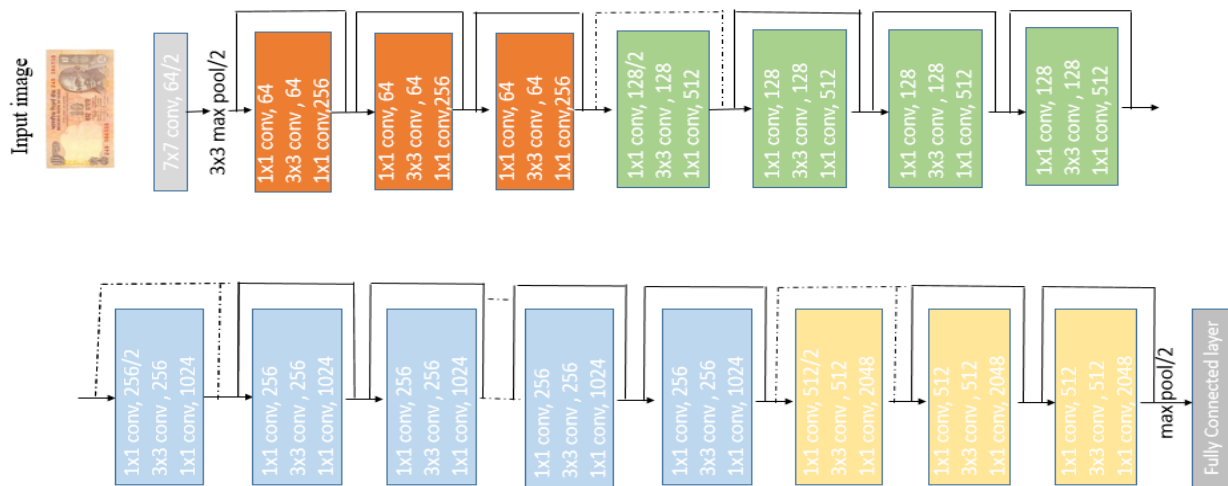


Fig .2.Resnet 101 Architecture

ResNet 101 is used because it not only solves the vanishing gradients problem, but it also has less learnable parameters, higher accuracy, and lower top-1 and top-5 error rates. It accomplishes this through its skip connection mechanism[21]. The skip connection bypasses a few stages of training and links directly to the output. ResNet 101 is made up of 99 convolution layers and two maximum pooling layers. ReLu is the activation function utilized here.

3.4 Online Newspaper Module

The free Google News API is used for extracting the online news article. It can extract top 10 headlines of the particular day of particular online newspaper. News API is a simple, easy-to-use REST API that returns JSON search results for live and historic news articles from all over the web. Using this, we can fetch the top stories running on a news website or can search top news on a specific topic (or keyword). The API key is needed to get started. API key will be available while sign in the API site. Get API results in JSON format via HTTP GET requests in any programming language and easily integrate them into the applications. To use the API, make a request, get the results, parse the JSON and the data is ready to be used.

3.5 Current location, weather condition, date and time providing Modules

Datetime module in python supplies classes to work with date and time. Open weather API allows to regularly download current weather for the corresponding location. The smart vision system uses the ip address for knowing the current location, region and the country.

3.6 Text to speech conversion

Every module generate output in text format. For visually impaired people the results should be available in audio. So the text to audio conversion is mandatory. This process of converting text to audio is done using python text to speech(pytsx3) library.

4. RESULTS AND DISCUSSION

If the user provide clear voice input the smart vision based system produces fast and accurate audio outputs for each module. The suggested system's performance is evaluated in terms of response time and accuracy. The performance analysis of the system is shown in the table 1. Speech-to-text and text-to-speech have average reaction times of 1-3 seconds and 5-7 seconds, respectively. All of the figures are based on a constant internet speed of 1 megabit per second. As a result, under ideal conditions, the system produces quick and precise results.

Table 1: System Performance Evaluation Table

Sl.No	MODULE NAME	RESPONSE TIME(s)
1	Describing surrounding	10-20
2	Recognizing familiar faces	20-30
3	Recognizing currency	10-20
4	Reading out text	10-20
5	Online Newspaper reading	0-2
6	Current location	3-8
7	Current weather situation	3-8
8	Date and time module	1-3

Face identification using Dlib is extremely accurate, with a maximum accuracy of 99.38% when faces are taken from numerous perspectives. Using the HOG+LSVM, the presence of a person in the input image is detected. People who are at least 14 feet away and facing the user can be appropriately recognized by the system. The figure 3 shows the result of face recognition module when an image of person in front of the user is captured through the webcam.



```
['anshad.jpg', 'aysha.jpg', 'mahira.jpg', 'mishah.JPG', 'naflu.jpg', 'shahruk khan.jpg']
[False, False, False, False, False, True]
[0.66350684 0.77586456 0.68552774 0.88591497 0.65563563 0.35864745]
5
shahruk khan
```

input image

Results showing the dataset labels, the comparison result, comparison distance, match index and output label

Fig.3. Result of face recognition module

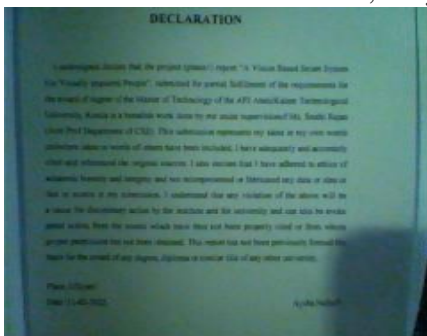
Google vision API can label an image in fast and detailed manner. When the user ask to describe his surrounding image is captured through the webcam and provided to the image captioning module. The system output the results in audio format speaking each entities in the image. The figure 4 shows the output labels of given input image.



Labels: Computer, Personal computer, Peripheral, Computer keyboard, Input device, Output device, Netbook, Space bar, Touchpad, Laptop, Activate Windows, For the PC software to

Fig.4. Output labels of given input image.

Like image labeling Google vision API is powerful enough to detect text in an image with different fonts and orientations, very accurately. Example is shown in the fig 5.



Labels:
DECLARATION
i undersigned declare that the project (phase) report "A Vision Based Smart System For Visually impaired People", submitted for partial fulfillment of the requirements for the mand of degree of the Master of Technology of the APJ AbdulKalam Technological University, Karala is a bonafide work done by me under supervision of Ms. Sruthi Rajan (Ass Prof Department of CSE). This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurly cited and referenced the original sources. I also declare that I have adhered to ethics of mis honesty and integrity and not misrepresented or fabricated any fact or source in my submission. I understand that any viciation of the above will be se for disciplinary action by the institute and the university and can also be evoke penal action from the source which have thus not been properly cited or any degree, diploma or similar tile of any other university.

Fig.5. Example of text detection

News API is used for online newspaper reading return top 10 headlines of the day. An example is shown in the fig.6. The average response time of the News API is between 100 and 200 ms.

- 1 Ukraine war latest: Civilians trapped as last bridge to key city destroyed
- 2 Russia-Ukraine war: Some of UK's top journalists barred from Russia
- 3 Taiwan: Are the US and China heading to war over the island?
- 4 January 6 hearing: Trump slams inquiry as 'Kangaroo Court'
- 5 Why is inflation in US higher than elsewhere?
- 6 Timed Teaser: Which team is Ed Sheeran backing again?
- 7 EU set to take legal action against UK over post-Brexit deal changes
- 8 Whiskey Wars: Denmark and Canada strike deal to end 50-year row over Arctic island
- 9 US aircrew cleared in review of deadly incident during flight from Kabul
- 10 Monkeypox to get a new name, says WHO

Fig .6.Ten top headlines of BBC news

Currency recognition done by training the dataset image using Resnet 101 model. The model obtained accuracy of 98.07% in 10 different epochs. The training and validation accuracy vs epochs graph is shown in the figure 6. The output prediction for a given 10 rupees note is also figured out in fig 7.

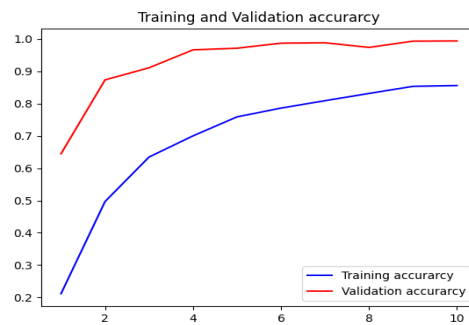


Fig.6.The training and validation accuracy vs epochs graph

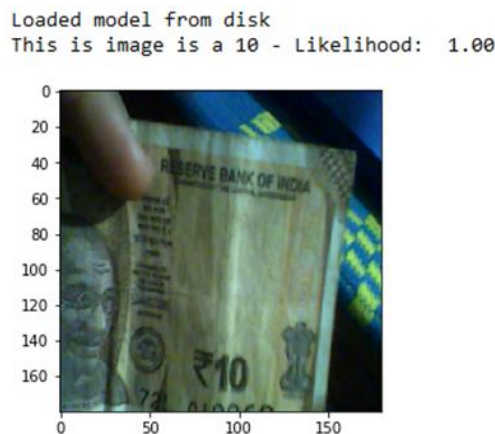


Fig.7.Prediction of 10 rupees note

The system voice out the current date and time, current weather, and current location within a few seconds.

CONCLUSION

A simple multi-purpose,fast responding, cheap and easily configured system assisting visually impaired so that they can live regular, independent lives like everyone else. This system proposes a single system that combines multiple separate aspects, such as face recognition, image labeling, text recognition, currency recognition, online newspaper reading, providing current location, weather condition and date and time.The proposed system is built using a combination of deep learning, machine learning, and numerous powerful APIs.The system respond promptly and accurately output each request as audio.

REFERENCES

1. Baldwin, Douglas. "Wayfinding technology: A road map to the future." *Journal of Visual Impairment & Blindness* 97.10 (2003): 612-620.
2. Wiener, William R., Richard L. Welsh, and Bruce B. Blasch. *Foundations of orientation and mobility*. Vol. 1. American Foundation for the Blind, 2010.
3. Shah, Chirayu, et al. "Evaluation of RU-netra-tactile feedback navigation system for the visually impaired." 2006 International Workshop on Virtual Rehabilitation. IEEE, 2006.
4. Li, Lixiang, et al. "A review of face recognition technology." *IEEE access* 8 (2020): 139110-139120.
5. Davis E. King. *Dlib-ml: A machine learning toolkit*. *Journal of Machine Learning Research*, 10:1755–1758, 2009.



6. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
7. Chen, Shih-Hsin, and Yi-Hui Chen. "A content-based image retrieval method based on the google cloud vision api and wordnet." Asian conference on intelligent information and database systems. Springer, Cham, 2017.
8. Pazio, Marcin, et al. "Text detection system for the blind." 2007 15th European Signal Processing Conference. IEEE, 2007.
9. Mittal, Shubham, and Shiva Mittal. "Indian banknote recognition using convolutional neural network." 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU). IEEE, 2018.
10. Hanen J abnoun I, F aouziBenzarti I, Hamid Amiri I, "Object Detection and Identification for Blind People in Video Scene' 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)
11. Dasila, Raghavendra Singh, et al. "Real time environment perception for visually impaired." 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). IEEE, 2017.
12. Cornacchia, Maria, et al. "Deep learning-based obstacle detection and classification with portable uncalibrated patterned light." IEEE Sensors Journal 18.20 (2018): 8416-8425.
13. Dahiya, Dhruv, et al. "Computer vision technique for scene captioning to provide assistance to visually impaired." 2018 41st International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2018.
14. Nishajith, A., et al. "Smart cap-wearable visual guidance system for blind." 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2018.
15. Bhole, Swapnil, and AniketDhok. "Deep Learning based Object Detection and Recognition Framework for the Visually-Impaired." 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2020.
16. Islam, MdTobibul, Mohiuddin Ahmad, and AkashShinghaBappy. "Real-Time Family Member Recognition Using Raspberry Pi for Visually Impaired People." 2020 IEEE Region 10 Symposium (TENSymp). IEEE, 2020.
17. Mangayarkarasi, P., et al. "Smart Vision System for Visually Impaired People." International Research Journal of Engineering and Technology 4.7 (2020): 943-946.
18. Z. Wei, Y. Sun, J. Wang, H. Lai and S. Liu, "Learning Adaptive Receptive Fields for Deep Image Parsing Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 3947-3955, doi: 10.1109/CVPR.2017.420.
19. "Document Text Tutorial," Google Inc. Accessed on: Feb. 1, 2022. [Online]. Available: <https://cloud.google.com/vision/docs/labels>.
20. "Document Text Tutorial," Google Inc. Accessed on: Feb. 1, 2022. [Online]. Available:<https://cloud.google.com/vision/docs/fulltextannotations>.
21. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.