RESEARCH ARTICLE OPEN ACCESS

Tech Assistance for Blind People

Jyoti Dange[1], Tejas Vaity[2], Gaurav Salvi[3], Pooja Sharma[4], Sachin Tyagi[5].

Department of Electronics and Telecommunications Engineering, Atharva College of Engineering
Malad-Marve Road, Charkop Naka, Malad (West), Mumbai-95
University of Mumbai, India
gauravsalvi@0104gmail.com,. s.poja1301@gmail.com, tejasvaity17@gmail.com, 14.sachintagi@gmail.com, jyotidange@atharvacoe.ac.in

Abstract:

Throughout the world, there are 39 million blind people. There is a lot of difficulty they encounter in their daily lives, whether it is accessing written scripts or even recognizing their knowns. The proposed system helps them to recognize people (using face recognition), detect any obstacle in the path (thus providing ease in their day-to-day life), and convert textual scripts into audio signals (text detection) that are provided to the blind person with the help of a microphone. An RPi camera is used to collect all the visuals and RASPBERRY PI 4B board. The conventionally used braille system of reading is not very efficient and all textual scripts are not available in braille language. Therefore, this device can be a great assistance to a blind person as it is low-weight, economical and efficient.

Keywords—RPi camera, Raspberry PI, face recognition, text detection, object detection.

Introduction:

The visual sense plays a primary role in guiding a sighted person through an unknown environment and assisting him or her to reach a destination safely. Unfortunately, people who are blind face difficulties in performing such tasks. Many people in our society are afflicted with various illnesses or disabilities. Globally, at least 2.2 billion people have near or distance vision impairments, according to the World Health Organization (WHO). To enable such individuals to live comfortably, certain amenities must be made available. For instance, if the atmosphere is suitable, students with vision impairments can study alongside their peers. We can deploy computer vision technology to create customized aids that will allow visually impaired persons to live as comfortably as possible in order to resolve this issue. The majority has a notion that those who are blind or have vision issues cannot live alone and constantly require assistance from others. They do not always require assistance; they have the ability to live independently most of the time. Using smart glasses for visually impaired people is one of the more well-liked solutions in this situation. These eyewear models employ software and hardware for computer vision (camera, image processing, image classification and speech processing) [1]. With such a method, visually impaired people can live comfortably among other people. This was the motivation behind designing and developing smart glasses to make learning easier for visually impaired students. These glasses are designed to use computer vision technology to capture images and extract English text. The main goals of the proposed system can be summarized as follows. Capture images, extract text from images, identify correct text, and convert text to speech. The rest of the article is organized as follows. Section II provides an overview of various solutions offered in the area of implementing smart glasses for the blind using deep learning-based computer vision techniques. Details of the design and implementation of the proposed system are given in Section III. Experimental results and their implications are discussed in Section IV. Section V describes the conclusions and future directions of the proposed solution.

Literature Survey:

Tesseract is an optical recognition engine wherein development is sponsored by Google Inc. The version 4 of Tesseract is based on a deep learning-based Artificial Recurrent Neural Network called as Long Short-Term Memory (LSTM) architecture. This OCR engine has high character recognition rates and supports 116 languages [13]. For accurate and robust solution of text detection from scene images MSER algorithm is used as it detects all the characters from any image. The datasets used for this system are ICDAR 2011 and Multilingual datasets. According to results it has 88.52% accuracy in character level recall [14]. An Efficient and Accurate

Scene Text (EAST) detector method is a simple and powerful pipeline that allows detecting a text in natural scenes. It has high accuracy and efficiency. In this study, three datasets have been used, that are ICDAR 2015, COCO- Text and MSRA-TD500. According to experimental results, this method has better result as compared to the previous method in terms of accuracy and efficiency [15].

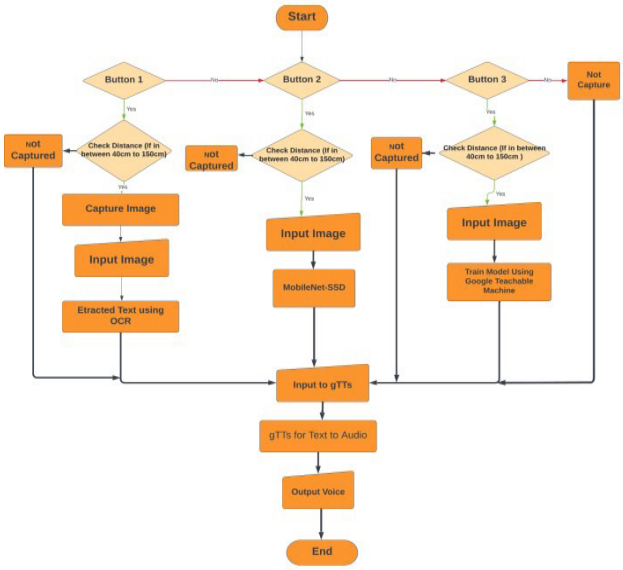
Methodology:

The proposed system has a fully functional model of smart glasses that are built to assist the blind person in several ways. The three assisting features in the device are face recognition of their family member, object detection, and text detection These smart glasses come with three buttons on it for executing the three different tasks.

1.Button 1- It allows the user to get the audio output of the detected texts in front of it.
It makes use of text detection techniques.

2.Button 2- It makes the user aware of the objects in front of it by using object detection techniques when the button is pressed.

3.Button3- When pressed, it allows the user to recognize the already trained dataset of family members and knowns. The overall flow of events can be understood easily with the help of the diagram below,



Block diagram fig. 1

System design and implementation:

•Image Capturing

First Image capturing is done so that error is reduced while performing any task. By using Rpi Camera we define the certain distance from which we get a perfect image and capture the image if the distance is not proper then send a message to a blind person that adjust your camera and then proceed.

•Image Processing

Image processing is the application of computer algorithms to manipulate and analyze digital images. It involves the use of various techniques to improve the visual quality of an image, extract useful information from an image, and transform an image into a more useful form for further analysis. Some common image processing techniques include:

1.Image filtering: This involves the application of various filters to an image to enhance its quality, remove noise, and extract useful features.

2.Image segmentation: This involves dividing an image into smaller, more manageable parts or segments. It is useful for identifying objects within an image.

3.Image restoration: This involves removing noise, blurring, or other types of degradation from an image to improve its visual quality.

4.Image compression: This involves reducing the size of an image by removing redundant or irrelevant information while preserving its essential features.

5.Object recognition: This involves identifying objects within an image and categorizing them based on certain criteria.

Image processing finds applications in a variety of fields such as medical imaging, robotics, surveillance, and satellite imagery analysis

•Text Extraction

Now by using OCR py-tesseract we extract the text which is written on the captured image or apply further operations to it.

•Object Detection

Object detection is a computer vision technique that involves identifying and localizing objects within an image or video. The goal of object detection is to automatically detect and locate all objects of interest within an image or video, and draw bounding boxes around them.

Object detection typically involves a combination of two main tasks: object classification and object localization. Object classification involves identifying the category or type of object within an image, while object localization involves determining the object's spatial location within

the image.

There are several techniques used for object detection, including:

•**Haar Cascades:** A feature-based object detection technique that involves training a classifier on a set of positive and negative samples.

•**Convolutional Neural Networks (CNNs):** A deep learning technique that involves training a neural network to learn features that are relevant for object detection

•**Region-based Convolutional Neural Networks (R-CNNs):** A variant of CNNs that involves generating region proposals within an image and classifying each proposal using a CNN.

•**You Only Look Once (YOLO):** A real-time object detection algorithm that involves dividing an image into a grid and predicting bounding boxes and class probabilities for each grid cell.

Object detection has many applications in areas such as surveillance, autonomous driving, and robotics. It is used to detect and track objects such as pedestrians, vehicles, and animals, and can be used to provide valuable information for decision-making in various scenarios.

•Face Recognition

Face recognition is the technology that enables machines to identify and verify the identity of a person using their facial features. It is a biometric technology that compares the features of a person's face captured by a camera with those in a database to find a match.The process of face recognition involves capturing an image of a person's face, extracting the facial features, such as the distance between the eyes, the shape of the nose, and the contours of the face, and then comparing these features with a database of known faces to identify a match.Face recognition technology has many practical applications, such as security and surveillance systems, mobile device authentication, and social media platforms. However, it also raises concerns about privacy, data security, and potential misuse.

Hardware design:

The various components that were used to build the smart glasses are-

a) Raspberry Pi

Raspberry Pi is a pocket-sized minicomputer that is available in various models with the latest model of Raspberry Pi 4 B. The model that we are using in this system is Raspberry pi 4 B. It requires a connected keyboard, mouse, display, power supply and a memory card and installed operating system. It includes GPIO (General Purpose Input/Output) pins to control various sensors. It is used for various educational purposes, coding and building software's. The operating system that we are going to use in this system is the Raspbian operating system.

b) Pi camera

The Raspberry Pi camera, also known as R-pi camera is also available in various variants such as V1, V2, etc. But the model used in this proposed system is R pi V1. With a field view of 160 degrees, the R-pi V1 stands out from other normal cameras which have merely 72 degrees of field view. It captures pictures in 1080p resolution and is compatible with major OS available in the market such as Linux, Windows, and Mac.

c)Earphones

Standard earphones are used in the given system. The Raspberry pi model 4B has a dedicated headphone jack which gives us an extra USB port for use for some miscellaneous. The headphones are given so that the person can hear the converted audio message of the text file and listen to the messages of obstacles in their path and the names of recognized family members. The headphones are chosen as lightweight and efficient ones as they have been connected to the glasses so the blind person does not lose them.

Software design

The various software components used to build the smart glasses are stated below

a) Pyttsx3

Pyttsx3 is used to convert text to speech in python language. It also works in offline mode and can be used with both python 2 and python 3. It is a very easy-to-use tool and converts all inserted text to speech. The pyttsx3 supports two voices,the first one is the female and the second one is the male provided by "Sapi5" for windows. It supports various types of engines [12].

b) Open CV library

Open CV stands for OPEN-SOURCE COMPUTER VISION which is used for machine learning projects [11]. Released under the BSD 3 Clause license, it is free to use. It is compatible on both desktop and mobile devices and supports all major operating systems. Here we are using it on the Raspbian operating system [4].

c) Mobile net SSD

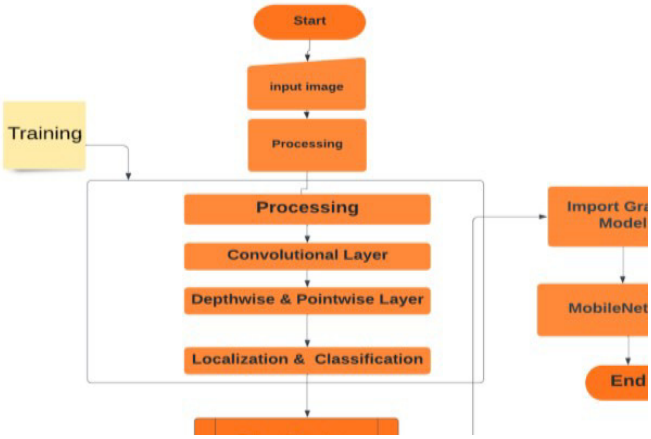
In the proposed system, object detection is carried out using the mobile net SSD algorithm. This Single Shot Detector model uses mobile net as a backbone for faster and more efficient object detection. The SSD architecture is based on CONVOLUTIONAL NEURAL

NETWORKS and follows the following two stages

1) Extracting features

2) Applying convolution filters for object detection.

It can be understood easily from the following figure [10]



d) Google teachable machine:

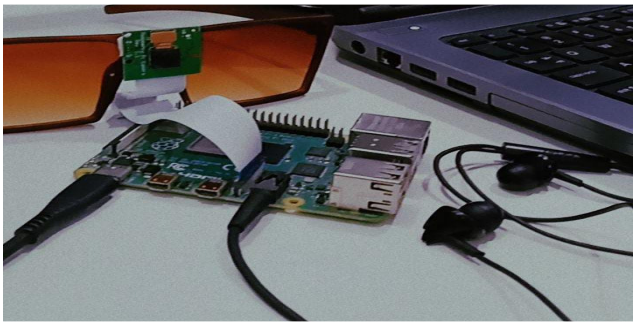
We are going to use Google's teachable machine for face recognition. It is designed by Google for training and testing machine learning models with various images and videos data sets. It is web-based tools that has user friendly interface 3 steps:

1. Data collection
2. Train of model (in which we can select batch size, Epochs, learning rate)
3. Testing the model

After training the model Google teachable machine provides the option of exporting the model which we can use in Raspberry pi 4[9]

Results and Implementations:

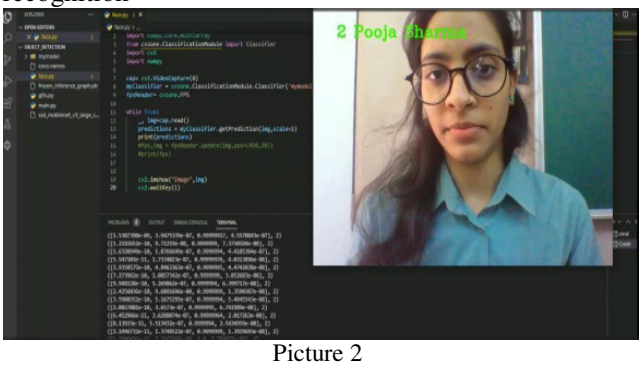
Hardware implementation:



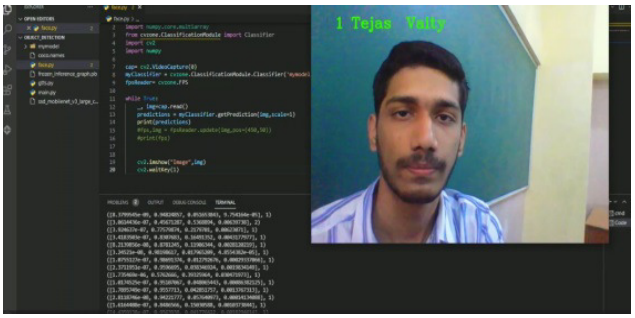
Picture 1

1. Face Recognition

The following images show the result of face recognition



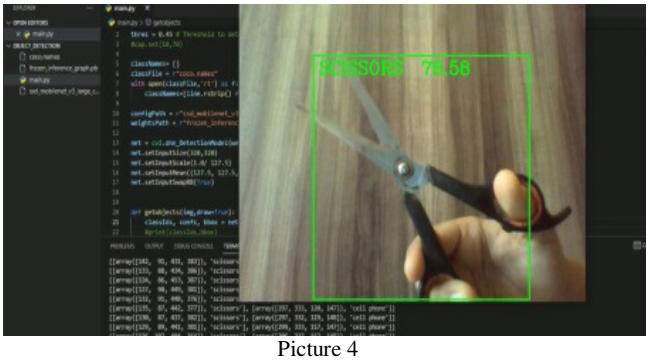
Picture 2



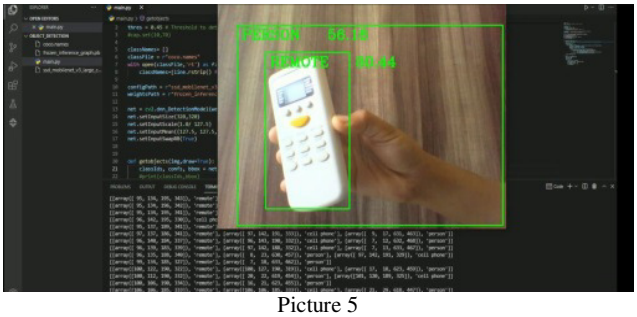
Picture 3

2. Object Detection

The following images display the result of object detection.



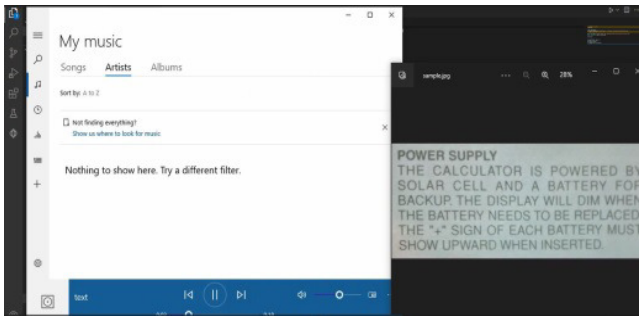
Picture 4



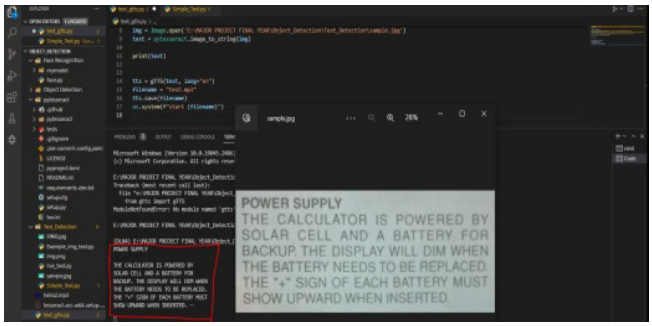
Picture 5

3. Text Detection

The following images show the results of text detection:



Picture 6(Google Text to speech)



Picture 7

	Input	Output
Text Detection	35 words in an image	All text is extracted successfully
Object Detection	We trained 90 plus objects	Detected remote, Tv, Phone, Toothbrush, Book, Bootle etc.
Face Recognition	Trained four team members faces	Successfully detected all trained faces

Conclusion:

This technical paper has proposed and explained implementation of smart glasses to assist the blind people. It recognizes faces, detects objects and recognizes texts and provides processed output in the form of audio output. Also, according to the market survey, the system here is much more dynamic, efficient, and low in cost. Approximate value of the manufacturing costs 15,000 INR. Although the system has certain limitations, those limitations can be covered in future scope.

Future scope

The paper has covered the basic requirement of a blind person. But certain additional upgrade can be added to it.

- A wide-angle camera can be added for more space visibility.
- Work on more accurate text detection algorithm try to train with YOLO V8 in future.
- For a better and more Real-time experience, video processing can be implemented.
- Captcha, Music, Skew text, Blurry text, and handwriting detection also added in future.
- Multi-lingual feature could be added for people with no knowledge of English.

Acknowledgement:

We sincerely thank Atharva College of Engineering and Madam JyotiDange for the cooperation and support throughout the project at every step.

References:

1. Improving mobility for blind persons using video sunglasses- https://ieeexplore.ieee.org/document/573972
2. Deep Learning Assisted Smart Glasses as Educational Aid for Visually Challenged Studenthttps://ieeexplore.ieee.org/document/8923044/
3. Raspberry Pi Based Wearable Reader for Visually Impaired People with Haptic Feedbackhttps://ieeexplore.ieee.org/document/9156005/
4. https://towardsdatascience.com/read-text-from-image-withone-line-of-python-code-c22ede074cac
5. https://towardsdatascience.com/read-text-from-image-withone-line-of-python-code-c22ede074cac
6. https://linuxhint.com/measure_distance_raspberry_pi/#:~:text=You%20can%20measure%20distance%20using,object%20and%20get%20reflected%20back.

7. Computer vision for the blind: A dataset for experiments on face detection and recognition- https://ieeexplore.ieee.org/document/7522323/
8. Deep Learning Assisted Smart Glasses as Educational Aid for Visually Challenged Students- https://ieeexplore.ieee.org/document/8923044/
9. Conferences>2020 IEEE Region 10 Symposium...RealTime Family Member Recognition Using Raspberry Pi for Visually ImpairedPeoplehttps://ieeexplore.ieee.org/document/9230937/
10. Raspberry Pi Based Wearable Reader for Visually Impaired People with Haptic Feedbackhttps://ieeexplore.ieee.org/document/9156005/
11. OpenCV. Retrieved from https://opencv.org/
12. Python Software Foundation, "Python". Retrieved from https://www.python.org/ (Last accessed on 30th May, 2019). 13. GitHub Inc., "Tesseract OCR". Retrieved from https://github.com/tesseract-ocr (Last accessed on 30th May, 2019).
13. X. C. Yin, X. Yin, K. Huang, H. W. Hao, "Robust text detection in natural scene images " IEEE transactions on Pattern Analysis and Machine Intelligence, 36(5), pp. 970980, 2014.
14. X. Zhou, C. Ylao, H. Wen, Y. Wang, S. Zhou, W. He, J. Liang, "EAST: an efficient and accurate scene text detector", Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 5551-5560, 2017

