Development of Self Assisted Voice Module for Visually Impaired

Sharath S[1], Ramu Muthangi[2], Premananda B. S.[3]
[1]M. Tech Student, Department of Telecommunication, RVCE, India, snsharath111@gmail.com
[2]CEO, Smart Health Global, India, ramumuthangi@gmail.com
[3]Associate Professor, Department of Telecommunication, RVCE,India, premanandabs@gmail.com

ABSTRACT

The visually impaired persons face challenge in recognizing the people around them and also reading the text content present in the documents. In order to overcome these issues, there is a need of assisted devices. This paper proposes a portable system which is able to recognize known faces and apprehends text. The proposed system has both Facial Recognition and Optical Character Recognition (OCR). Facial Recognition is carried out with the use of Local Binary Pattern Histogram (LBPH) face recognizer which makes use of HaarCascader to stumble on the faces. Initially the device identifies presence of person within its range and then applies the LBPH operator to recognize the face. In order, to access the statistics of textual content visually impaired person needs vision. OCR is used where text extraction is performed from scanned images. Tesseract OCR engine is used here. Extracted text is converted into speech using Google Text to Speech (gTTS) method which helps visually understand the text. This portable system is implemented using Raspberry pi. Upon getting into the digital camera view formerly stored faces are identified and acknowledges to the visually impaired through voice. The system can recognize face and characters with the highest accuracy.

Key words: Face Recognition, gTTS, LBPH, OCR, Raspberry pi, Tesseract.

1. INTRODUCTION

According to WHO there are 285 million humans estimated to be visually impaired out of which 37 million people are completely blind which reduces human's capability to perform day by day tasks without assistance [1],[2],[7]. Due to eye diseases, age associated reasons, out of control diabetes, injuries visually impaired people elevated every year. These include difficulties of moving in complete autonomy and the ability to seek and recognize objects and reading the texts. With all of the various issues confronted through blind, the hassle of analyzing is spotting the sector of interest.

The visually impaired rely largely on their other senses consisting of listening to, touch, and scent a good way to apprehend their environment. Navigation systems also helps visually impaired persons to navigate around the surrounding environment through voice [4]. Though there is advancement in technology the techniques used by the blind are old.Most of the studying substances for the blind are within the form of Braille [5]. A character has to analyse the usage of Braille only for the purpose of reading, and if such people is not able to learn Braille then one may be not able to understand the text.

Current technological tendencies in transportable computer systems, digital cameras and pc imaginative and prescient make it viable to assist those people via products that integrate laptop imaginative and prescient technology with different existing business products such as OCR structures [8]. Recent tendencies such as mobile phones, computers, and use of virtual cameras make it viable to help the blind by growing digital camera [6].

This paper aims to build a voice assisted module for visually impaired person in order to triumph over the issues that a visually challenged character is facing. The proposed system is able to identify the person around them and recognize the characters in the documents. This method is used to discover the faces from the database that is stored when the person comes in the vicinity of the camera.

The Organization of the paper is as follows: Section II gives the methodology carried out for the proposed system. Section III gives the algorithmic description used for Face and Optical Character Recognition. Section IV includes results obtained from the experiment. Finally, section V concludes the work with future scope.

2. METHODOLOGY

The proposed system helps visually impaired in recognizing the faces in front of them and character recognition present in the documents. The block diagram of voice assisted module is shown in the Figure 1. Raspberry Pi 3 is used as the main processor where all other components are connected to it. VL53L0X distance sensor is interfaced with the processor

using I2C pins for the purpose of recognition as well as text extraction such that if the image is captured only within the certain range, only such images will be processed. Raspberry pi camerais used for capturing the images for both Facial Recognition and OCR. Charging circuit provides constant supply output of 5V, which is used to power up the entire system. Thus, the system can work for 3 to 4 hours when completely charged.
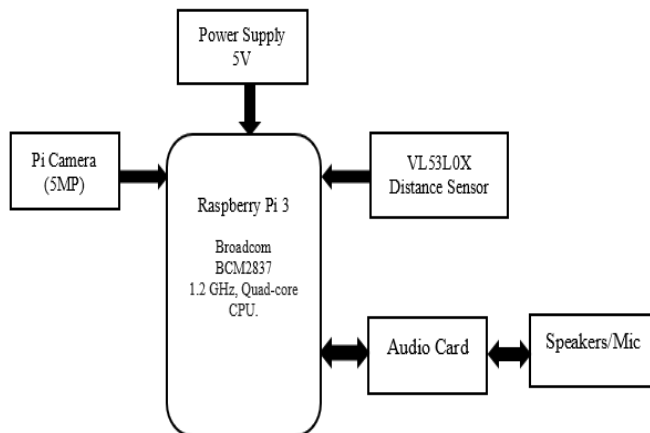

Figure 1: Voice Assisted Module

Microphone is used to capture the input voice for storing the name tag for the faces and to provide voice output for Face Recognition and OCR. Both Facial Recognition mode and OCR are activated by switching between the modes using the touch sensors used in the system. Initially, the system will be in standby mode. Touch sensor is used for mode control. When the sensor is touched the system goes to optical character recognition mode. It enters OCR mode only when there is no touch of the sensor for the duration of 3 sec. During this mode the image will be captured and is sent to processor and will be processed by Tesseract OCR engine. The system will read out the text from the image through speakers.

When the sensor is touched again the system goes to facial recognition mode where face will be detected to a certain range. This completely depends on lighting and image quality. When the system enters facial recognition mode image is captured, device asks to whether to record the name, if touched then the device saves the name and image. The name that is recorded will be appended to the image and is stored in the database. Touch and hold of sensor return the device to standby mode. Raspberry pi is configured using Raspbian OS. Python programming language has been used with PyCharm as an ideal environment. PyCharm is an integrated development Environment employed for creating computer programs, specifically for the Python language. OpenCV and machine learning library for developing both the algorithms and database is used for storing the images as well as information of the person.

3. IMPLEMENTATION

3.1 Facial Recognition Algorithm

Facial Recognition system is an emerging technology which is able to identify or verify someone with the image captured from the source. There are several techniques in which facial recognition system is used.They compare selected facial features from given photograph with faces that are present in database. Face recognition algorithms understand facial functions with the useful resource of landmark extraction or capabilities from a picture of the human face. Algorithms analyze parameters of face such as size of eyes, nostril, cheekbones, and jaw [12]. These parameters are then used to analyze for different photographs with matching features within the database.

There are several tough and ranging parameters in face detection and identity such as illumination, exclusive poses, change in expressions, low image qualityetc. For detection and recognition of faces, there are various perspectives such as some projects focuses on illuminations of the image, some of them focuses on resolution of the images such as low and high. Recently some researchers follow frontal view of the face from different angles with varying light illuminations which has been used in this project. Facial recognition system follows primary steps to detect the face and to recognize those faces which are listed as follows:1. Face Detection, 2. Preprocessing. 3. Feature Extraction,4. Feature Matching.

Local Binary Patterns Histogram algorithm has been used for Face Recognition which was introduced in the year 2006 [12]. It is primarily based on local binary operator and adapted in facial recognition because of its computational simplicity and discriminative strength [3], [13]. The steps involved in LBPH Face recognition are as follows: 1. Creation of Dataset, 2. Face Detection, 3. Feature Extraction, 4. Classification.

Initially, the face datasets are captured via image acquisition from the camera where important features of face are extracted from the feature extraction phase. Based on these facial features face landmarks are considered based on that human identity is recorded. The classifier used here is skilled to identify the face. In the final step the device can recognize the face and gives the information about the person in front from the database. The flowchart of face recognition system is shown in the Figure 2.

Creation of Dataset. Datasets is created which consist of 40 individual images. It is created on the basis of face detected from the pi camera. These images are stored in same folder to generate the face dataset. During this stage, these images are converted into gray scale in order to extract the facial features of the image and then they are usually normalized for better results. In order to detect the faces HaarCascaders has been used which detects the features of the given input image.

HaarCascade classifiers are powerful way for object detection. It was proposed by Paul Viola and Michael [13]. HaarCascade Classifier is a machine learning approach where both positive images and negative images are used to train the classifier. Positive images contain the object that has to be detected by the classifier (images with faces) where negative images do not contain the object that our classifier is required (images which doesn't consist of faces). On the basis of positive as well as negative images faces are detected.
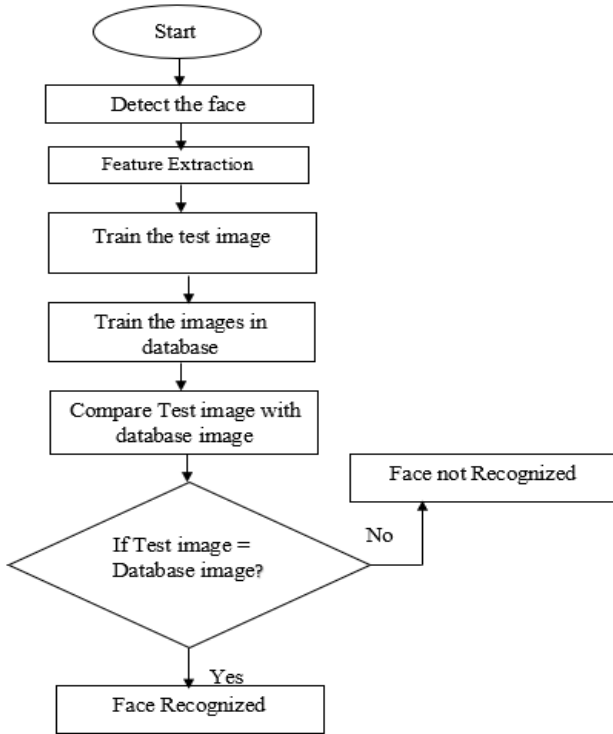
Figure 2: Flowchart for Face Recognition

Face Detection. When the device enters the facial recognition mode raspi camera captures the image in front of the visually impaired. When the face is detected a rectangular box is drawn around it. Flowchart for Face Detection is shown in the Figure3. The rectangular box is used to select the face of the person by cropping the required area. Faces detected are mainly concentrated on frontal face as the identification is only based on the person present in front of the visually impaired. The image captured is converted into gray scale so that grayscale simplifies the algorithm and decreases computational necessities. Face is detected with the help of HaarCascade classifier.

HaarCascade classifier uses Adaboost Algorithm which selects a small range of essential features from a large set and offers relatively efficient classifier in order to locate faces whether it is present in the frame or not. If at all the face is not detected then again image will be captured. More complicated classifiers are combined to shape a cascade which discards any non-face regions in a photo, thereby spending extra computation on promising item-like regions.

Feature Extraction. LBPH is the algorithm used for face recognition. It is primarily based on neighborhood binary operator and is one of the first-rate appearing texture descriptors. The LBP is the binary ratio of pixels intensities inside the middle pixel and it's round eight pixels. In order to extract facial features of the captured image Local Binary pattern operator is used that compares the depth fee (intensity) of every component with the 8 nearest neighbor pixels values. An example of LBPH operation is shown in Figure 4.
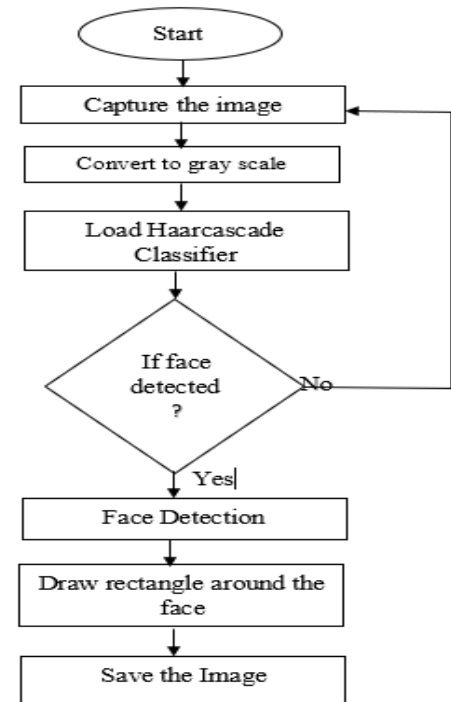
Figure 3: Flowchart for Face Detection

Decimal numbers are marked for the pixels of image with LBP operators which are called Local binary patterns (LBP) codes, which encode the local structure round each pixel. The photograph having dimensions M x N is split it into regions of same top, width which is represented in mxm dimension for every vicinity. LBP operator is defined in window of 3x3 which is shown in Figure 4. Each pixel in 3x3 matrix is subtracted from the value of central pixel value. For the example central pixel value considered is 90. The mathematical description for LBP operation is represented as follows

$$LBP\ (x_c, y_c) = \sum_{p=0}^{p-1} 2^p (i_p - i_c) \tag{1}$$

where intensity of central pixel is shown as i_c and (x_c, y_c) represents the eight surrounding pixels[13]. If the resulting values from the above operation are negative than those numbers are replaced with 0 and others with 1.
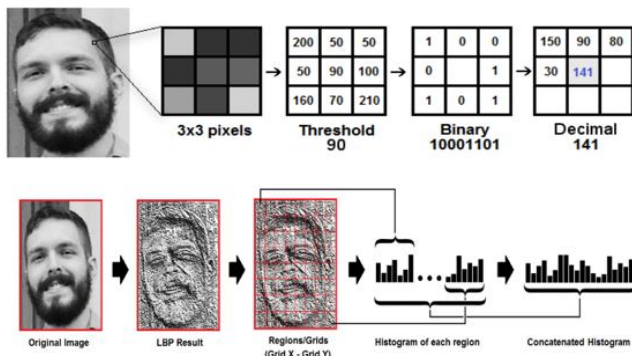
Figure 4: Example of LBPH operation

The mathematical representation for converting into binary numbers is as follows

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x \leq 0 \end{cases} \qquad (2)$$

The resulting matrix will contain simplest binary values. These binary values have to concatenate from every role from the matrix line by using line into a new binary number (10001101). The binary codes are converted into decimal number by concatenating binary numbers in clockwise direction [12]. Thus, resulting value will be the pixel value for the input image. Based on LBPH parameters the histograms are generated for the image. Histograms are generated for both input image as well as image present in the database.

Classification. The histograms are generated for both input image and image in database if these two matches accordingly the face is recognized which is represented by the mathematical equation.

$$D = \sqrt{\sum_{i=1}^{n}(hist1 - hist2)^2} \quad (3)$$

The recognition of person face is done by detecting the face of the person. When face recognition code is executed, the system looks at the image in internal database for the match using test image. If the match is greater than 60 percent then the face will be recognized by voice as the output.

3.2 Optical Character Recognition

OCR is defined as, the electronic conversion of photos of typed, handwritten or published textual content into device-encoded text. OCR systems provide individuals who're visually impaired with the capacity to scan revealed textual content and then have to be conveyed in artificial speech or stored to a computer report. In Text reading technology there consists of three key elements. They are capturing the image (scanning), recognizing and reading text from the image [11]. Initially a document is scanned by a digital camera. Images are then converted into recognized characters using OCR engine. Recognized texts are then converted to voice as the output. The document is represented in the form of pixels which is computed by OCR engine [10]. OCR makes it possible to use strategies which include

machine translation, textual content-to-speech for pdf and scanned documents. Finally diagnosed textual content report is fed to the output devices. The output is the headset related to the raspberry pi board or a speaker that can have voice as the output for text report. The steps followed for OCR are: Image acquisition, Preprocessing, Segmentation, Text Extraction, Text to speech conversion. The flowchart of OCR is shown in the Figure5.

OCR Engine: OCR engine provides high speed processing of image and also high recognition results. Advanced OCR engines provide better results irrespective of different image formats, varying font sizes and languages. For this project Tesseract OCR has been used. Tesseract is able to recognize more than 80 languages [6].
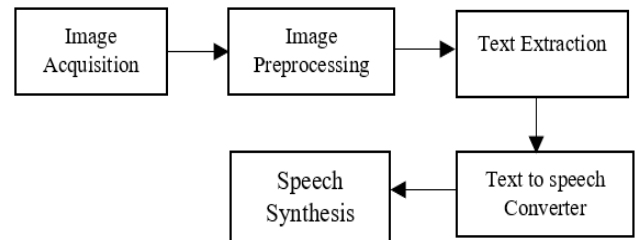

Figure 5: Flowchart for OCR

This is powered by google and is recognized to be one of the most accurate open source engines. Captured photos have to be carried out with scaling, edge detection, contour detection, and noise removal in order get the output.

TTS module: This module is used for speech synthesis. TTS module is used to convert recognized text in to voice as the output. As visually impaired people have disability to read the documents, when OCR engine recognizes the text, the TTS is used to translate the text document into speech [9]. Text is converted in to symbol by the front end where each symbol has phonetic assigned to it, whereas back end converts generated phonetic to sound. Google-text-to-speech has been used which is an open source software which is having clear voice output. The steps followed for OCR are described as follows:

Image Acquisition. Initially system is moved to the page in which one has to know the text present in the book, paper etc. and then pi camera captures the photos of the textual content. The captured image will be having good quality as camera is having high resolution.

Pre-processing. Pre-processing in particular consists of following steps Binarization, Edge Detection, Contour Detection and Noise elimination. Initially image is binarized. The input image is converted into grayscale image (0 up to 256 grey ranges) in to black and white picture (0 or 1) with binarization method. The Binarization converts image to grey scale, where result of OCR mainly depends upon the binarization. The excessive quality binarized photo can provide more accuracy as compared unique photograph because noise is present within the original photo. After binarizing the image is scaling is performed to obtain better

output results. Edge detection is used for locating the boundaries on the picture. Adaptive thresholding is completed with a view to separate foreground picture from that of background. This additionally helps in elimination of noise inside the image.

Segmentation: After pre-processing, image freed from noise is surpassed to segmentation phase. In this phase image having characters that are present in series are subdivided in to images of individual characters [5]. The image which is binarized is checked for inter line areas. If inter line spaces are detected then image segmented into units of the paragraphs across interline hole. The traces within the paragraphs are scanned for horizontal area intersection with respect to background. In order to hit upon the width of the horizontal line, histogram of the photograph is used. The lines are scanned vertically for vertical area intersection. Width of the phrases is detected with the aid of histogram. Words are decomposed to characters.

Text Extraction. The characters present in the image are transformed to corresponding ASCII values using raspberry pi processor [11]. Characters are matched with its corresponding template, will be stored as text transcription. This text transcription is fed to the voice output.

Text to Speech converter. This module converts the recognized text into speech. TTS is performed by gTTS. Sound is generated with the aid of audio card and is filtered. Audio card can significantly enhance the quality of sound.

4. RESULTS AND DISCUSSION

The proposed system is implemented using raspberry pi. Facial recognition and Optical character recognition algorithms are implemented using Python language. Face recognition system is implemented by preparing the set of Training images. The dataset used for training the classifier is shown in the Figure 6. In order to develop the set of training images, initially Face detection method is used which captures the image using pi camera. The captured images are initially converted in to grayscale which is shown in the Figure 7. Face is detected with the help of HaarCascade method. The detected face is represented using rectangular box which is shown in the Figure 8.
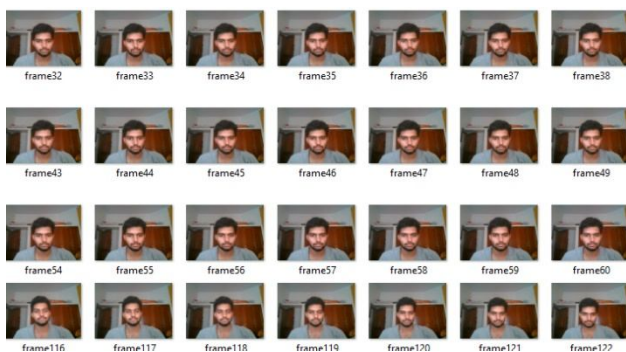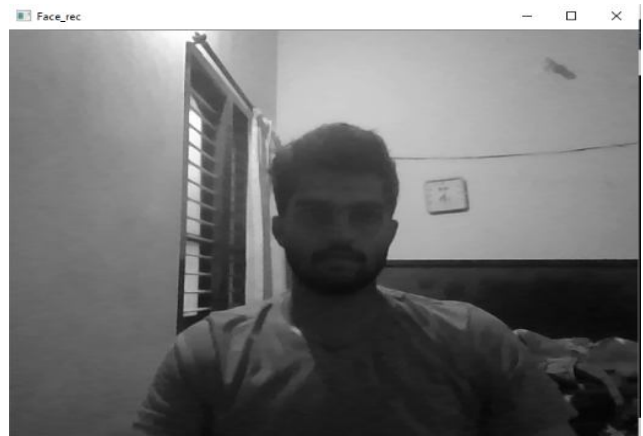

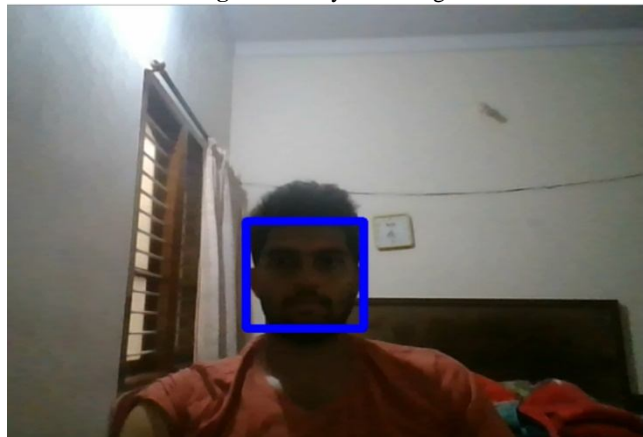**Figure 6:** Set of training images

Figure 7: Grayscale image

Figure 8: Face Detected

The captured images are saved into dataset folder for extraction of features and training processes. Feature extraction is performed using LBPH method. The vectors generated by the captured image are compared with the vector of Training set images using LBPH operator. If the confidence (matching) value is greater than 60 % face will be recognized. Information (name) of the recognized face is fetched from the Database with the voice as the output. The recognized image which is represented by text is shown in the Figure. 9. The output of OCR is acquired after all the preprocessing stages. The image captured using Pi camera is shown in Figure 10. Test image is converted to gray scale for further processing which is shown in Figure 11.

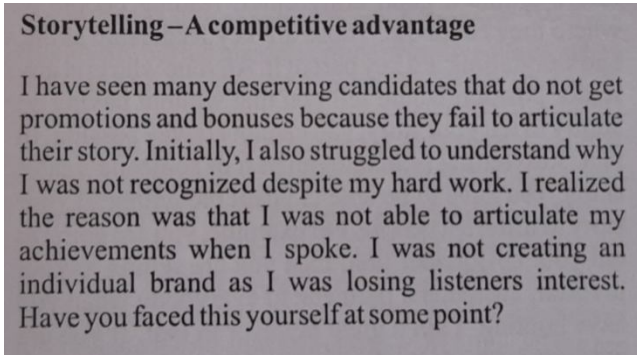**Figure 9:** Recognized face represented by text
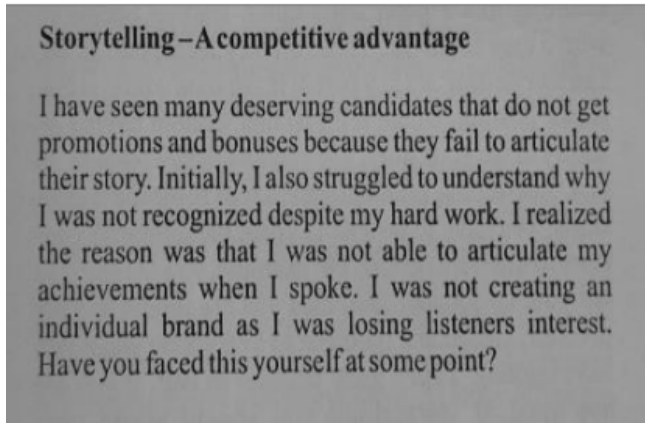
Figure 10: Test image

Figure 11: Gray scale image

In order to detect the edges in the text canny edge detection algorithm is used. The result of canny edge detection is shown in the Figure 12. Adaptive thresholding is performed to suppress the background image by highlighting only the foreground image which is shown in the Figure 13. Figure 14 shows the final Tesseract output and this text is converted in to speech using gTTS method. Preprocessing stages are performed to enhance the quality of the acquired image by eliminating the noise to obtain the accurate output results for OCR. Thus, the overall system will help the visually impaired people to recognize the face in their view and recognize the characters from the image through voice as the output. As battery is used portability is achieved which makes visually impaired to carry the system anywhere.
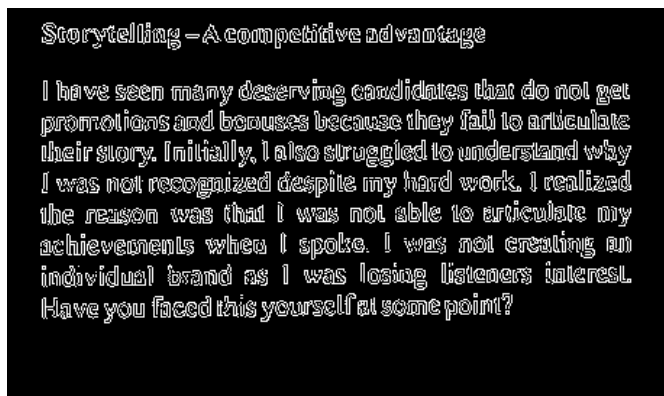

Figure 12: Cranny edge detection

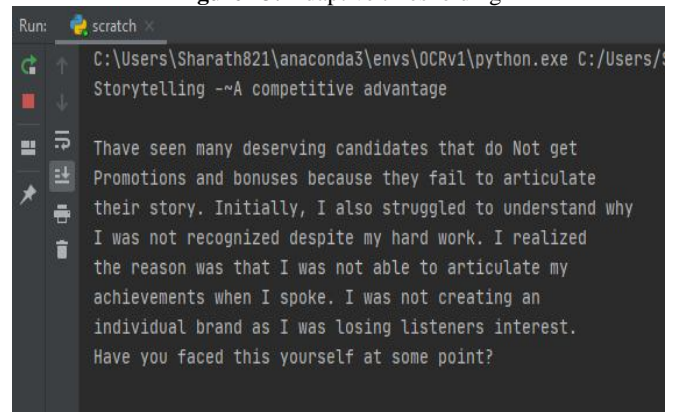**Figure 13:** Adaptive thresholding

Figure 14: Tesseract Output

5. CONCLUSION

As visually impaired people find most difficulty to recognize the persons in front of them as well as to read the text. The proposed system allows visually impaired people to recognize the faces and read the text with the highest accuracy. For Facial recognition LBPH has been used as it is having computational simplicity and discriminative strength. Haarcascade classifiers have been used to detect the face which consists of set of positive as well as negative images. In order to extract facial features, based on facial landmarks feature vector of the image will be generated. On the basis of parameters of LBPH, histograms will be computed for each of the faces on the basis of vectors. If histogram matches face will be recognized. OCR is performed with help of Tesseract OCR engine. This is one of the most effective OCR engine with high speed and recognition results. The output of the OCR engine is fed to the TTS module in order to convert the text into speech as the output. Advantage of this system is that portability and compact size. As visually impaired people move to different places, they can carry this system with them.

The proposed system has only face recognition and OCR. In future this system can be implemented with navigation system which can assist the blind. Currently English language is present in OCR, In the future, system may be implemented with other languages.

REFERENCES

1. S. Murali, R. Shrivatsan, V. Sreenivas, S. Vijjappu, S. J. Gladwin and R. Rajavel. **Smart walking cane for the visually challenged,** in *Proc,IEEE Region 10 Humanitarian Technology Conference,*pp. 01-04, 2018.

2. S. M. Felix, S. Kumar and A. Veeramuthu. **Smart Personal AI Assistant for Visually Impaired People**, in *Proc, International Conference on Trends in Electronics and Informatics,* pp. 1245-1250, 2018.

3. Phichaya Jaturawat and Manop Phankokkruad. **An Evaluation of Face Recognition Algorithms and Accuracy based on Video in Unconstrained Factors,** in *Proc. IEEE International Conference on Control System,* pp. 25–27, 2016.

4. M. Gharat, R. Patanwala and A. Ganaparthi. **Audio guidance system for blind.** in *Proc. International Conference of Electronics, Communication and Aerospace Technology,* pp. 381-384, 2015.

5. Sahu, Narendra Sonkusar and Manoj. **A Study on Optical Character Recognition Techniques.***International Journal of Computational Science, Information Technology and Control Engineering,* 2017.

6. G. Li, X. Li and B. Xu. **Numerical Simulation Technology Study on Automatic Translation of Foreign Language Images Based on Tesseract-ORC.** *International Conference on Robots & Intelligent System.* pp. 86-89, 2019.

7. Anitha P, Gowrishankar S, P., Gowtham Pandiyan, Guhan. S, Hari Vignesh. **Audio Navigation System for Visually Impaired People.** *International Journal of Emerging Technologies in Engineering Research (IJETER).*Vol. 8, pp. 16-19,2020.

8. R. Kasthuri, B. Nivetha, S. Shabana, M. Veluchamy and S. Sivakumar**. Smart device for visually impaired people.** In *Proc,International Conference on Science Technology Engineering & Management,* pp. 54-59, 2017.

9. S. Jadav.**Voice-Based Gender Identification Using Machine Learning.**in *ProcInternational Conference on Computing Communication and Automation,* pp. 1-4, 2018.

10. T. C. Wei, U. U. Sheikh and A. A. A. Rahman. **Improved optical character recognition with deep neural network.***IEEE International Colloquium on Signal Processing and its Applications (CSPA),* pp. 245-249, 2018.

11. Hamad, Karez and Kaya Mehmet. **A Detailed Analysis of Optical Character Recognition Technology.** *International Journal of Applied Mathematics, Electronics and Computers,* pp. 244-244, 2016.

12. V. Aza, Indrabayu and I. S. Areni. **Face Recognition Using Local Binary Pattern Histogram for Visually Impaired People.** *International Seminar on Application for Technology of Information and Communication (iSemantic),* pp. 241-245, 2019.

13. Farah Deeba, Hira Memon and Fayaz Ali Dharejo. (2019). **LBPH-based Enhanced Real-Time Face Recognition.** *International Journal of Advanced Computer Science and Applications,* pp. 1-4, 2019.