Research Article

Visual Impairment Assistance System

Geethalakshmi.V [a], Surya Venkatraman[b] ,Mohammed Sinan Marikattay[c] , Poornima B.V[d]

[a] Senior Systems Engineer,Infosys Mangalore, India
[b]Programmer AnalystCognizant Technology Solutions Chennai, India
[c,d] Computer Science Department Bearys Institute of Technology Mangalore, India
Email:[a]geethu.rvr@gmail.com, [b]suryavenkatraman05@gmail.com, [c]meetsinan@gmail.com, [d]poornimabv.85@gmail.com

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: Visually challenged people have no knowledge of outdoor hazards and require visual aids to avoid threats. Furthermore, they face many difficulties compared to other people in doing their day to day work. Even though they have guide canes to help them it becomes difficult to recognize the faces of the people. They find it difficult to identify the familiar as well as unknown faces. Recognizing familiar faces confidently without anyone's help is a very difficult task for a blind person. This model uses deep learning techniques that can identify familiar and unknown faces for visually impaired people. Apart from recognizing faces, blind people find it difficult to read simple signboards and important instructions. A few of the examples are the menu cards of the restaurants, Direction boards, locating washrooms, etc. The proposed model will help the visually challenged people to read the images by converting the image into a text and then converting the same text which is read into audio. For converting the image to text, OCR technology was used. Recognizing face is a very important task for the visually challenged people as a failure of this may even lead to life-threatening problems and reading sign boards and other simple text is also very important in the daily life of visually challenged people. Hence the development of this visual impairment assistance system will help the visually challenged people to live a better life.

Keywords: Deep learning techniques, OCR technology.

1. Introduction

According to the world blind union, there are about 253 million people around the world with serious vision problems and about 47 million of them are blind [1]. Visually disabled people can only sense light and darkness, and cannot see things in front of them. Furthermore, they move around based on their senses and experiences with the aids of guidance cans to detect and avoid collision with moving and stationary obstacles. Sometimes, the guide canes don't offer their required safety levels because they don't provide perception of the obstacles or objects types and also, do not give information about the walking path. Guide canes don't help visually challenged people to identify the faces of the known and unknown people.

The main aim of this proposed system is to develop a system that helps blind people to survive in the community without the help of a second person. It will help the blind people to perform their daily activities normally and will help in problems such as communication with familiar and strange objects around them. This system involves three main modules.

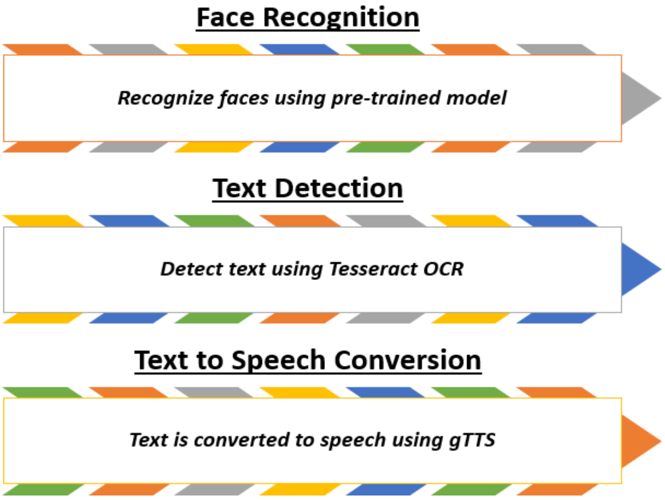
Face Recognition
Recognize faces using pre-trained model
Text Detection
Detect text using Tesseract OCR
Text to Speech Conversion
Text is converted to speech using gTTS

Fig 1.1: Modules in Visual Impairment Assistance System

2.Literature Survey

In Wael AbdAlmagee proposed deep learning face recognition [2], several pose-specific deep convolutional neural network (CNN) models process a face picture to

produce multiple pose-specific characteristics. They have used 3D rendering method for getting multiple images of the input image. The novel representation they present produces better outcomes in both the verification and recognition (i.e.earch) function than the state-of-the-art on IARPA's CS2 and NIST's IJB-A..

The face recognition system proposed by Sharma [3], explains about the process involved in face recognition like face alignment and feature extraction. They provide the information about face alignment. The research was conducted on the Face Recognition Grand Challenge (FRGC) dataset and with FAR of 0.1 providing 96 percent accuracy.

In the real time FR implementation system, which was proposed by the Neel Ramakant Borkar [4], the face of two similar person with different facial expression are compared. In his paper he has used Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). To implement this, he has used Raspberry Pi. In his system to identify the optimal features and to remove the noise he has used Principle Component Analysis.

In the algorithmic approach for detecting the text from the image, a paper presented by Neha Agarwal [5] explains how the Otsu’s algorithm and Hough transform algorithms which helps to read a set of information which is provided in the image as text. Her proposed was tested in Verdana font style with a specific font size also. It has given 93% accuracy. OCR was also used to identify the characters. This system was mainly used for converting the image which contained many lines or the documents which were in the image format. To enhance the quality of the text present in the document which is the key to improve the accuracy a system named QUARC was used. By using the QUARC system the document has which had an error rate of 38% was reduced to 24%.

Image Text to Speech conversion in the desired language by translating with raspberry Pi system proposed by Rithika[6] deals with the text images which is converted to audio in mainly in English language. The main aim of the system which was provided by the author was to help even the non- English speakers to under the text which is present in the image written in English language. Her system provides voice output and Raspberry Pi was used. Camera and speakers were used to read the text and give the output. Both the camera and the speakers were mounted on the Raspberry Pi hardware device. Tesseract OCR and Google Speech API was used in her system.

3.Proposed System

Module 1:

Face recognition is the process of detection and verification of the known person. The system will alert the user when an unknown person comes into picture. The four main steps involved in face recognition module:

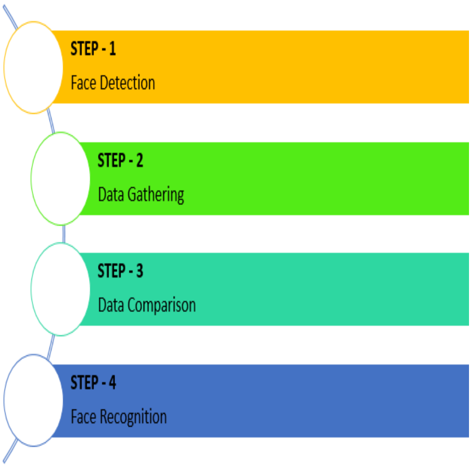

Fig 3.1: Steps Involved in the module 1

In order to perform face recognition, we can use OpenCV Python Libraries. In this proposed system we can use two external libraries named dlib and face_recognition libraries. OpenCV and Deep Learning both has to be used simultaneously for getting accurate results in this module. In face recognition module, known faces are stored in the repository and then the strangers are identified by comparing the image of them along with the images of the known people present in the repository. When an unknown face is detected then the user will be alerted stating that the person who is in front of them is a stranger via voice command. This module involves 4 common stages: Detection, Alignment, Representation and Verification. OpenCV and Dlib libraries provide face and eye detection packages. Deep face libraries offer both the face alignment and face detection functions.

Module 2:

Second module involves detecting the text from the image. The main advantage of this module is that it will detect the sentences and words in the images and convert the text in the image to normal text which can later be used for the audio conversion for the visually challenged people. Tesseract OCR is the tool which is used in this. Text detection can be done either by using machine learning or deep learning. When compared to machine learning methods, deep learning techniques provide more accuracy. OpenCV is the major library used in text detection. Reading the image is the main function performed by Open CV.

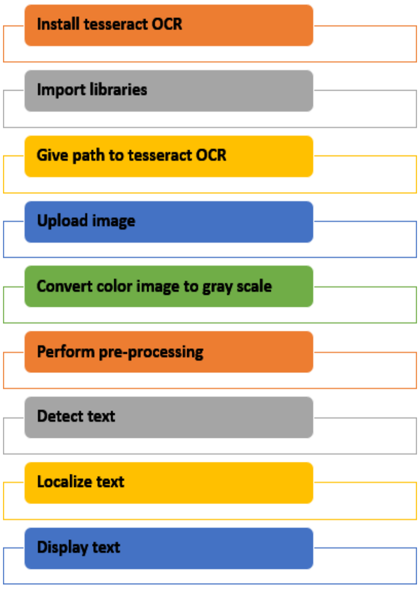

Fig 3.2: Steps involved in Text Detection

After uploading the image, it is necessary to convert the colored image to grey scale. There are various reasons for doing this, but one of the main reasons is that it reduces the complexity of the model. By using the image with the RGB we use three color medium channels but that is not the case in grey scale, and here the complexity is less other than dealing with the RGB images. OpenCV python code when an image is read it will be BGR format. There are some inbuilt functions which can be used to convert the BGR image to RGB. After converting from BGR to RGB, the images are converted to Greyscale.

Preprocessing is used to improve the quality or features of the image. The preprocessing method used in the text detection module is thresholding and blurring. This preprocessing is performed after the conversion of color image to gray scale. Thresholding can be applied to the image using cv2. Threshold function. There are three types of thresholding.

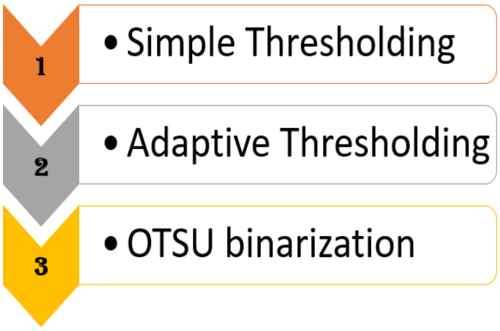

Fig 3.3: Types of Thresholding

In text detection module OTSU binarization can be used for preprocessing of the image. The main function of preprocessing is to remove the noise in the image, which will provide an image of better quality. By performing the preprocessing step, the accuracy can be improved. Three main steps involved in the image preprocessing is:

1. Reading the image

2. Resizing the image
3. Removing the noise.

After removing all the noise converting the image to greyscale, text from the image is detected and it is displayed as a text separately. This can be used by the visually impaired people while reading the boards and while reading the menu where there might be many fancy images in addition with the text. This model will delete all those additional details or pictures and read only the text from the image. This whole process is carried out using the Tesseract OCR.

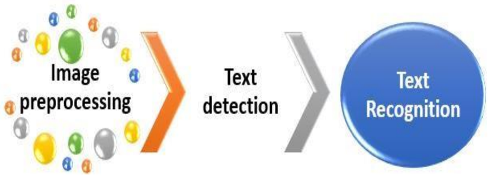

Fig 3.4: General Flow of OCR Module 3:

Third module is the final module where all the three modules are integrated into a single system. Text which is read in the second module is converted into audio. This module is designed mainly for the visually challenged people. There are many API's available to convert the Text to Speech. Google Text To Speech conversion in one among the several API's. gTTS is the most convenient and easy to use API. One of the major advantages of using this API is that it supports several languages. Pronunciation is also accurate in gTTS when compared to other text to audio conversion API's. The output of the text to audio is stored in

the mp3 format. Apart from visually challenged people, this text to audio conversion API can help people who does not have any knowledge about the language in which the text is written. As gTTS supports almost all the languages it can even act as a translator. The audio will be same as the human natural voice. This was one of the main reasons why gTTS API was chosen.

gTTS also has the function giving the audio in different speeds. Slow, medium and fast. But however, there is no functionality to change the voice of the audio. gTTS library is available in the python and in command line interface. gTTS can be installed using pip command in the Command Line Interface. The image with text content has to be loaded first. The image has to be preprocessed. This step is included in the text detection module. The output of the above module can be passed to produce the audio output.

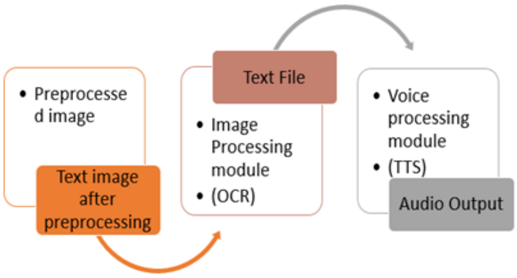
• Preprocessed image
Text image after preprocessing
Text File
• Image Processing module
• (OCR)
• Voice processing module
• (TTS)
Audio Output

Fig 3.5: General Flow of Text to Audio Conversion

4.Algorithm And Libraries Used:

CNN Algorithm:

Support Vector Machine and Convolutional Neural Network is used for face recognition. When it comes to image processing CNN outperforms the Support Vector Machine (SVM). [7] CNN perform the correlations in a large number of images and find the nonlinear correlations where as SVM slightly suffers in predicting the class labels when the size of the class labels is high and it provides a marginal classifier Also, it's difficult to parallelize SVM but the CNN architecture inherently supports parallelization. SVM provides an accuracy of 85%.[8] Thus, the accuracy of SVM algorithm was lesser than convolutional neural network algorithm.

Since the use of SVM did not provide with the accurate results we can use Convolution Neural Network (CNN) which uses deep learning for the face recognition. CNN was introduced in order to improve the training performance of the BP algorithm [4]. CNN reduces the input data preprocessing when compared to the Backpropagation neural networks. CNN which is used in the face recognition uses the Artificial Neural Network (ANN). The neural network contains many neurons in which the output of the previous neuron can be used as an input of the latter neuron.

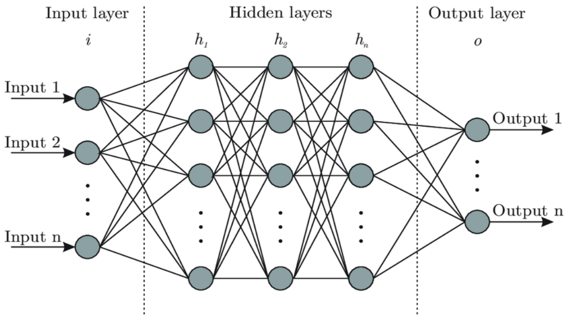

Fig 4.1: Basic Neural Network

When compared to the basic neural network the Convoluted Neural Network slightly different. However, it has the same neurons and different layers. A convolutional neural network is a special form of a feed forward neural network. The architecture integrates previous information about the input signal and its distortions. Convolutional neural networks are explicitly designed to deal with the obvious variation of the 2D forms. They combine local feature fields and mutual weights, and use spatial subsampling to ensure a certain degree of invariance of change, scale and deformation. Using the local receptive fields, the neurons will extract basic visual features such as corners, endpoints. The following layers then connect these simplistic features to detect more complicated characteristics.

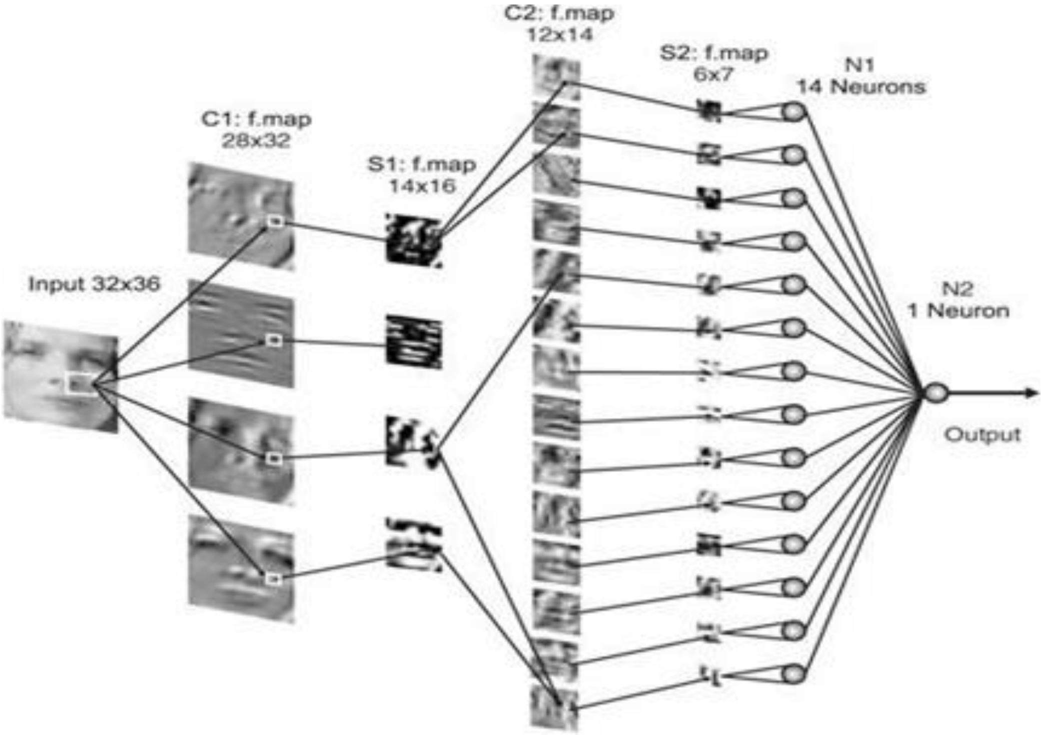

Fig 4.2: CNN in face recognition

In the convolution layer, the network tries to understand the pattern. In the first layer of the CNN the, the network tries to understand the edges in the pattern while in the second layer of the CNN the network tries to understand the color and shape of the image and in the final layer of the CNN which is the combination of other layers and it is also called as the fully connected layer tries to differentiate the image or find the similarity between the obtained image and the image which is present in the data set. CNN can have any number of layers starting from the input layer to the fully connected layer.

Face Recognition Library:

The face_recognition python library is used to recognize and manipulate faces. It works with the command line prompt and dlib library can be used in this process. face_recognition library can automatically recognize the features of the face and find the difference. Data set containing training set has to be provided, so that the system can be trained with the set of images provided in the training data set. While testing the system a set of images are given and compared with the images given in the training set.

Triplet loss function is a loss function used as an anchor set for the trained data set. It takes face encoding of three images, the anchor, positive and negative. Here the anchor and the positive are the same images. Whereas the negative is an unknown person's face. This is mainly used for face embedding in which the model analyses the given image and then returns a numerical vector value which can be used for filtering the image. By using CNN

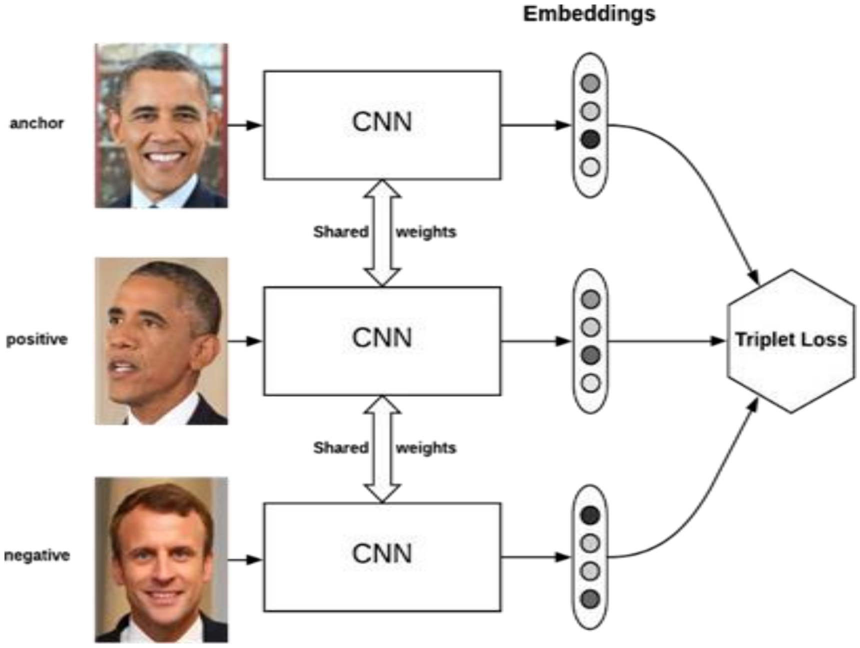

Fig 4.3: Triplet Loss Tesseract and gTTS:

For image to text conversion the system uses Tesseract. Pytesseract can be used in the proposed system. The system uses Python Tesseract Optical Character Recognition (OCR) tool [9] for recognizing the text from the image. Tesseract

OCR provides accurate results. Pytesseract can be used for recognizing text in different languages. OCR follows a set of steps before converting the image to text. After giving the image as the input, the image will be converted to RGB format and then to greyscale format. As OpenCV which uses Pytesseract will take the input in the form of BGR and its necessary to convert the image to gray scale to remove the unwanted noise in the picture which might affect the accuracy of the system. Pytesseract can also be used to find the confidence of the output which is obtained after using the Pytesseract. After this conversion using the Tesseract OCR tool the output text is converted to audio using the gTTS. Google Text To Speech conversion is used to convert the text to audio [10]. This system uses gTTS python library and using this we can explicitly provide the speed in which the text needs to be converted and the language which it has to be converted. Finally, the converted audio is stored as mp3 file.

5.Conclusion:

The proposed system was developed for the wellbeing of the visually impaired people. Three modules which is proposed here are face recognition, image to text conversion and the last module is text to audio conversion. All these three modules can be integrated as one single application using a framework. Flask framework can be used to combine all the modules into a single web application. The web application which is created can be deployed as a mobile application using Heruko. [11] Heruko is Platform as a Service (PaaS) that enables the users to build, run and operate the applications entirely in cloud. One major reason for using Heruko is data security. When a cloud platform is used for running the apps, user data is highly prone to security threats. Heruko uses many cryptography methods to protect the data. As the proposed system deals with some confidential data such as the images of people which needs to be stored. Apart from integrating the system and using as app we can also use Raspberry Pi [12] for integrating the modules as one single system.

References

1. World Blind Union Organization.
2. Wael AbdAlmageed; Yue Wu; Stephen Rawls "Face recognition using deep multi-pose representations". 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)
3. S.Sharma ; KarthikeyanShanmugasundaram ;

4. Sathees Kumar Ramasamy ”FARCE – CNN based efficient face recognition technique using Dlib” 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)
5. Neel Ramakant Borkar ;Sonia Kulwelkar “ Real-time implementation of face recognition system” Conference: 2017 International Conference on Computing Methodologies and Communication (ICCMC)
6. Neha Agrawal ; Arashdeep Kaur “An Algorithmic Approach for Text Recognition from Printed/Typed Text images” 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)
7. H. Rithika; B. Nithya Santhoshi “Image text to Speech conversion in desired language by translating with Raspberry Pi” 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)
8. Musab Coşkun ; Ayşegül Uçar ; Özal Yildirim ; Yakup Demir “Face Recognition based on Convolutional Neural Network” 2017 International Conference on Modern Electrical and Energy Systems (MEES)
9. Ivanna K. Timotius ; The Christiani Linasari ; Iwan Setyawan ; Andreas A. Febarinto “ Face Recognition using Support Vector Machines and generalized discriminant analysis” 2011 6th International Conference on Telecommunication Systems, Services, and Applications (TSSA)
10. M.H.O'Malley “Text to Speech Conversion Technology” Computer (Volume: 23 , Issue: 8 , Aug. 1990)
11. K Mona Teja ; S Mohan Sai ; H S S S Raviteja D ; P V Sai Kushagra “ Smart Summarizer for Blind People” 2018 3rd International Conference on Inventive Computation Technologies (ICICT).
12. Bih-Hwang Lee ; Ervin Kusuma Dewi ; Muhammad Farid Wajdi “Data Security in cloud computing using AES under Heroku Cloud”
13. Ishita Gupta ; Varsha Patil ; Chaitali Kadam ; Shreya Dumbre “ Face Detection and recognition using Raspberry Pi” 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)