PHOTO OCR FOR BLIND

Meghana km[1],Ganavi.N[2],Chitra kumari[3],Khushbu kumari[4],Divya prabha kn Assistant Professor ,CSE.

SRI SAIRAM COLLEGE OF ENGINEERING

***Abstract*— The Photo OCR stands for 'PHOTO OPTICAL CHARACTER RECOGNITION'. With the growth of digital photography and more recently the growth of camera in our cell phones we now have tons of visual pictures that we take all over the place. And one of the things that have interested many developers is how to get our computers to understand the content of these pictures a little bit better. The photo OCR problem focuses on how to get computers to read the text to the purest in images that we take. It can not only help machines to understand our images better but also can help blind people in navigation. It can also help with car navigation system, with cars which can read street signs and help in reaching destination.**

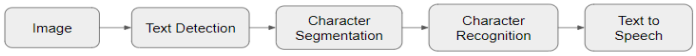
INTRODUCTION

Blind people have to face a lot of difficulty in performing day-to-day tasks. But the advanced technology era has tried to make their lives a bit easier. This paper talks about one such idea which can be a step towards achieving the same goal. The paper describes the process of developing a specs like device which can capture image in front of them and can read out any text, if present in the captured image. The technology uses own developed OCR with an advancement in precision and accuracy. The conversion of text to voice is done using the already existing TTS (The Technology of Speech) framework. The work described in this paper strictly focuses on detection of text and street signs from the image. This can help blind people in reading documents and navigating by hearing the street signs. Hence, technology if used in a correct way can definitely be a boon for the visually impaired people.

I. METHOD

The method comprises of three sections - capturing of image, extracting of text from image and finally converting text to voice.
An image can be taken with the help of a camera, here we use a raspberry pi camera. The image taken is sent for further processing where text is extracted from the image and predicted. From there, the predicted text is sent to TTS framework for converting text to speech.

The Data Flow Diagram



In Fig 1, text detection, character segmentation and character recognition comes under the text extraction phase.

Image will be taken in two modes - continuous and triggered. In continuous mode the images will be taken one after the other as the images appear in front of the camera. This mode is generally used when the user is walking. Example - Reading of street signs while navigating. In triggered mode, the image will be taken only when the user activates a trigger. Otherwise no image and no processing will take place. In text detection phase the algorithm will decide whether the image taken contains text or not. It does so with the help of tons of examples - Images which contain text, which is called dataset in machine learning terminology. The algorithm outputs 'yes' if the image contains text and 'no' otherwise. In character segmentation step, the extracted text portion is scanned to extract character from it.
The extracted character is checked to be one of the characters out of 26 alphabets. This process comes under the famous multi-class classification problem of machine learning. Again for this process a huge dataset is required. In this paper machine learning concepts are used and data is generated artificially. Hence tons of data can be generated and the accuracy of the algorithm increases. For converting the extracted text to speech TTS (The Technology of Speech) synthesis framework is used. The obtained voice is passed to the user with the help of earphones attached to the raspberry pi module.

II. WORKING PRINCIPLE

The proposed prototype works on the principles of image processing, machine learning and TTS (The Technology of Speech) framework. All these are arranged in a systematic way to obtain the desired prototype. Image processing is used to scan the captured image and extract the portion of the image

which contains text. Machine learning is used to process the extracted image portion (which contains text) to extract the text (character by character) and predict the text. The TTS framework takes input as the extracted text and converts that to speech so that a blind person who is not deaf can listen and can get an idea of the scene in front of him/her.

FIG2: TEXT DETECTION, CHARACTER SEGMENTATION AND CLASSIFICATION

Text detection – This is the first phase of processing taking place in the proposed prototype. In this phase the machine learning algorithm will decide whether the image taken contains text or not. It does so with the help of tons of examples which contains images having text, which is called dataset in machine learning terminology. The algorithm outputs 'yes' if the image contains text and 'no' otherwise. For this phase, Image processing is used. The image obtained from camera will be scanned top to bottom and left to right. If any text like appearance is found while scanning, that portion of the image will be distinguished from the rest of the image. Later those distinguished pixels will be extracted from the image and passed to the next phase described in the working pipeline.

Character Segmentation – This is the second phase of processing taking place in the proposed prototype. In this phase, the extracted image portion is scanned in one-dimension. For scanning sliding windows technology is used, which will scan the image from left to right(1-Dimension). The extracted image is extracted in 1-Dimension based on the property of the text that, generally the aspect ratio of the text have larger width than height. While scanning, individual character is extracted and queued in a queue like data structure, to be passed to next phase in the working pipeline. While scanning there will be two cases:

Case 1: the sliding window, which is used while scanning, contains partition between two text.

Case 2: the sliding windows, which is used while scanning, contain only a single text.

These cases and be represented as:

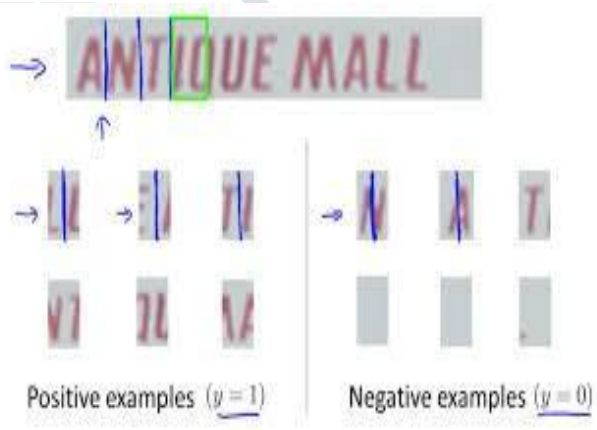


FIG3: CHARACTER SEGMENTATION

Case 1 will be treated as positive examples while case 2 will be treated as negative examples. On the basis of both the cases, words are classified from sentences. The characters queued up will be passed to the next phase.

Character Classification – This is the third phase of processing taking place in the proposed prototype. This phase is used to identify the characters which are extracted in the previous phase. The classification of the characters into respective alphabets comes under the famous machine learning multi-class classification algorithm. This phase makes use of the above stated algorithm and does the classification. The dataset for executing this phase is artificially generated by adding distortions in the pre-existing examples. Hence, with tons of examples the accuracy and precision of the algorithm increases largely.

III. RESULTS

The paper presents a description of making a prototype which can detect text from image and convert it to voice. This can make the life of blind people a bit easier in performing daily tasks like reading a document, moving from one place to another. The application makes use of raspberry pi module with a camera attached to it. Also it makes use of machine learning concepts in predicting the text present in the image and in increasing the accuracy of this prediction. There is also a usage of image processing in detecting text portion in the captured image. The proposed application makes use of TTS framework to convert text to speech. Combining all the technologies in a systematic way gives this application as a result which can help blind people. With the capability of producing tons of training examples the proposed prototype achieves high accuracy and precision in identifying the text correctly which makes the prototype interesting and reliable.

IV. CONCLUSION

The paper presents development of a prototype which is used to detect text from the captured image and converting the detected text to speech, thus helping blind people to visualize their surroundings. The process of developing the prototype consists of several phases which are arranged systematically to achieve the proposed goal. It consists of capturing the image, preprocessing the captured image. The output hence obtained will be sent to next phases of processing for text detection and identification. The text hence identified text will be sent to TTS (The Technology of Speech) framework to obtain speech as output corresponding to the identified text.The research, implementation and optimization developed allowed the design of the proposed prototype allowing the reading of texts; provided that the light conditions are the ideal for image recording and the equipment is properly directed to the text required to be identified so that recognition and reading are as satisfactory as possible. The proposed

V. REFERENCES

[1] https://www.researchgate.net/publication/275231020_Camera_Reading_for_Blind_People
[2] https://www.coursera.org/learn/machine-learning
[3] https://www.orcam.com/en/
[4] https://www.sciencedirect.com/science/article/pii/S2212017314003624
[5] http://www.afb.org/info/programs-and-services/professional-development/employment/accommodations-for-workers-with-vision-loss/optical-character-recognition-systems/12345
[6] https://ieeexplore.ieee.org/document/7098807/
[7] https://www.irjet.net/archives/V5/i4/IRJET-V5I4388.pdf
[8] https://arxiv.org/ftp/arxiv/papers/1710/1710.05703.pdf