International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Hat for vision

Ankit Agarwal[1], Beulah Alexander[2], Pratik Chaurasia[3], Sushant Patil[4]
[1, 2, 3] *Department of Computer Engineering,*
[4]*Assistant. Professor, Department of Computer Engineering, St. John College of Engineering and Management, Village Vevoor, Manor Road, Palghar (East)-401404.*

Abstract*: These days, huge measure of data is accessible on the web or present as e-books and printed media. Outwardly tested individuals think that it's hard to perusfrom books in the event that it isn't in Braille dialect. However the individuals who are denied of vision can assemble data utilizing their listening ability capacity. The proposed framework plans to produce a voice portrayal of printed reports utilizing PC vision and sound age procedures. The entire equipment will be stuffed inside an ordinary Hat which can be conveyed by client on his head. Outwardly tested individuals are extremely intense on the grounds that they sense, smell and hear to carry on with their life. Shockingly, they are reliant on Braille dialect. To make them ready to peruse, this framework enhances the innovation which gives sound portrayal to literary contents. The Hat comprises of a self-ruling framework chipping away at Raspberry pi which incorporates a pi-camera. For a client to peruse printed reports, the framework utilizes Text Recognition calculation in view of machine realizing which forms content pictures into content record which would then be able to be at that point changed over to sound streams. Hence, 1-D flag change of the above picture is delivered so as to channel each content line. At last, every expression of every content line is recognized through an OCR (Optical Character Recognition) strategy and the clients hear it by means of TTS (Text To Speech) methodology. The yield will be given to the client through headphones or headset.

Keywords*: Computer Vision (CV), Optical Character Recognition (OCR), Text-to-Speech (TTS), Raspbian OS, Pi Camera.

I. INTRODUCTION

The project depends on the space of Computer Vision. Objective of computer vision is to compose computer programs that can translate pictures. Computer vision tries to robotize undertakings that the human visual frameworks can do. Computer vision in our project helps in helping people to distinguish printed archives. Manmade brainpower and Computer vision share different subjects, for example, design acknowledgment and learning systems. Thus, PC vision is now and then observed as a piece of the manmade brainpower (Artificial Intelligence) field or the software engineering field all in all.

The image processing module catches picture utilizing camera, changing over the picture into content. Voice handling module changes the content into sound and procedures it with particular physical attributes so the sound can be comprehended.

Text-to-Speech (TTS) can change the content picture contribution to sound with an execution that is sufficiently high and a decipherability resilience of under 2% [1], with the normal time preparing under three minutes. This convenient gadget, does not require web association, and can be utilized freely by individuals.

II. LITERATURE SURVEY

Our framework is motivated by the Reading assistant for visually impaired [2] which goes for removing letters from books and change over them into advanced shape and after that present it as needs be. Once the picture is being stacked, we can change over it into gray scale picture which at that point can be changed over to a content record. This content record is changed over to sound document utilizing TTS (Text To Speech) Algorithm.

Another inspiration is by the Real time text tracking for TTS translation camera for blind [3], this arrangements with the improvement of a structure of perusing associate gadget with a scene content locator that demonstrates the content area by sound flags, and assembled a model gadget by joining the scene content locator, Optical Character Recognition (OCR) motor, and content to-discourse (TTS) motor.

Regarding the yield, Autonomous OCR dictating system [4] manages the fundamental thought of improvement of an independent framework for directing content report by means of picture preparing calculation for daze individuals.

The framework is constituted by the Raspberry Pi 2B- the portable preparing unit and a couple of uniquely planned glasses with a HD camera and Bluetooth Headset.

III. ABBREVIATIONS AND ACRONYMS

A. *OCR* - Optical Character Recognition is the mechanical or electronic change of written pictures, transcribed or printed content into machine-encoded content.

B. *TTS* - Text-to-Speech is an artificial production of human speech. TTS system converts normal language text into speech.

IV. PROPOSED SYSTEM

The proposed innovation presents the idea of a HAT containing the camera for taking the depictions of the visually impaired environment. The taken depictions are prepared with the comparing calculation and an applicable content report of that picture is recovered. At that point the content is changed over to a sound record and sounded in the headphone. The components required are:
Raspberry pi 3 is the main processing unit of the system. It is placed in the HAT. Mostly processing is done by Raspbian operating system. Raspberry pi 3 comes up with 1 GB RAM. It also has a Wi-Fi and a Bluetooth 4.1 module.
The Raspberry Pi Camera Module v2 is a high quality 5 megapixel Sony IMX219 image sensor. The high resolution camera is placed on the HAT captures the snapshot of the surroundings and the captured image is transmitted to the system for further processing of details in the image.
A high audio clarity earphone outputs the transmitted audio file which is processed by the system. The system converts the text into (.wav) audio file format.
An LED flash light used to take clearer and non-blurred image of the surroundings in low light for a better image.

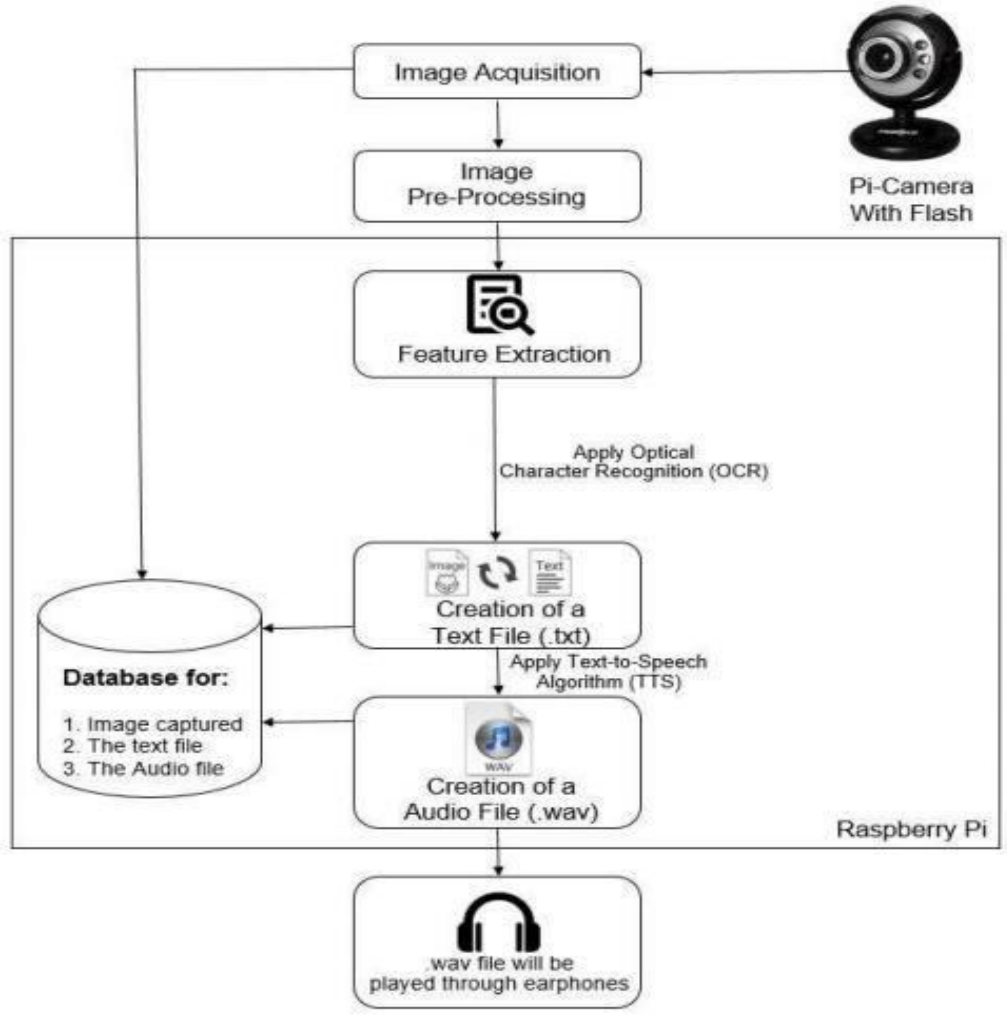

Fig. 1 Architecture diagram

A. Image Acquisition

The initial phase in which the individual is holding a book and the camera catches the pictures of the content. The nature of the picture caught will be high in order to have quick and clear acknowledgment because of the high determination camera.

B. Image Pre-processing

Image processing is essentially an arrangement of functions that is utilized upon a picture configuration to reason some data from it. The information is a picture while the yield can be a picture or set of parameters acquired from the picture. Once the picture is being stacked, we can change over it into gray scale picture. The picture which we get is presently as pixels inside a particular range. This range is utilized to decide the letters. In gray scale, the picture has either white or gray substance; the white will be the dividing between words or clear space.

C. Feature Extraction

In this stage we assemble the basic highlights of the picture called feature maps. One such technique is to distinguish the edges in the picture, as they will contain the required content. For this we can utilize different axis detecting methods like: Sobel, Kirsch, Canny, Prewitt and so forth. The most exact in finding the four directional axes: flat, vertical, right corner to corner and left inclining is the Kirsch locator. This strategy utilizes the eight point neighbourhood of every pixel.

V. METHODOLOGIES

A. OCR

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of printed records. It is crucial to the computerization of printed texts so that they can be electronically searched, stored more compactly, displayed on-line and used in machine processes such as machine translation, text-to- speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

B. Tesseract

Tesseract is a free software optical character recognition engine for various operating systems. Tesseract is considered as one of the most accurate free software OCR engines currently available. It is available for Linux, Windows and Mac OS. An image with the text is given as input to the Tesseract engine that is command based tool. Then it is processed by Tesseract command. Tesseract command takes two arguments: First argument is image file name that contains text and second argument is output text file in which, extracted text is stored. The output file extension is given as .txt by Tesseract, so no need to specify the file extension while specifying the output file name as a second argument in Tesseract command. After processing is completed, the content of the output is present in .txt file. In simple images with or without colour (gray scale), Tesseract provides results with 100% accuracy. But in the case of some complex images Tesseract provides better accuracy results if the images are in the gray scale mode as compared to colour images. Although Tesseract is command-based tool but as it is open source and it is available in the form of Dynamic Link Library, it can be easily made available in graphics mode.

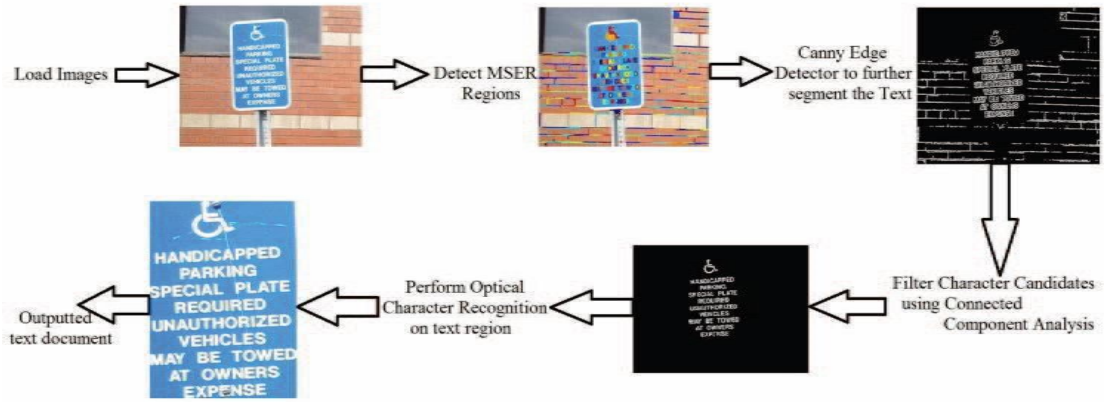

Fig. 2 Detection and conversion using OCR

1) *Step 1:* Initially the image is loaded for processing and the textual area is identified.
2) *Step 2:* This algorithm detects MSER regions and uses Canny Edge Detector to further segment the text.
3) *Step 3:* Some of the connected components are removed by the region properties.
4) *Step 4:* OCR then filters character candidates using Connected Component Analysis.
5) *Step 5:* Characters in most languages have similar stroke width and thickness throughout. Now in filters the characters candidates using the stroke width image.
6) *Step 6:* The bounding boxes that enclose the text region are determined and the Optical Character Recognition Algorithm is applied in the text regions.
7) *Step 7:* The image is now transformed to a text document which contains the same words.

C. Text-to-Speech

Text-to-Speech (TTS) synthesizer is a computer based framework that can read message out loud naturally, paying little heed to whether the content is presented by a PC input stream or an examined input submitted to an Optical character Recognition (OCR) unit. A discourse synthesizer can be actualized by both equipment and programming. Discourse is frequently in view of link of regular discourse i.e. units that are taken from regular discourse set up together to frame a word or sentence. There are two or three options for discourse acknowledgment device, in any case I thought the best response for this instructional exercise was to use Google's Speech recognition Tool. This API grants us to exchange the voice we basically recorded and transforms it to content substance.

D. Microsoft Translation API

Since we can record our voice and afterward change over it into content utilizing the Google's Speech Recognition, we need to make an interpretation of that content to our coveted outside dialect. I would have wanted to use Google's Translate API for this, since deplorably there is a 20$ join cost for the use of this API.

E. Google Text-to-Speech API

Google Text-to-Speech API again changes over back the yield content substance into voice which can be heard on headset.
The TTS system comprises of these 5 fundamental components:

1) Text Analysis and Detection
2) Text Normalization and Linearization
3) Phonetic Analysis
4) Prosodic Modelling and Intonation
5) Acoustic Processing

The input text is passed through these phases to obtain the speech.

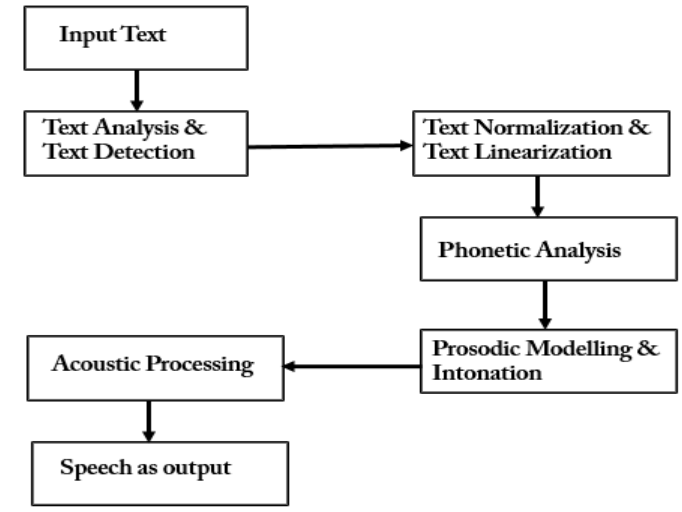

Fig. 3 Phases of Text-to-Speech Conversion

The Text Analysis part is pre-processing part which examines the information message and sorts out it into reasonable rundown of words. It comprises of numbers, abbreviations, acronyms and colloquial and changes them into full content when required.

Content Normalization is the change of content to pronounceable shape. The primary target of this procedure is to recognize accentuation checks and delays between words. Typically the content standardization process is improved the situation changing over all letters of lowercase or capitalized, to evacuate accentuations, highlight marks, stop words or excessively basic words and different diacritics from letters.

Phonetic Analysis changes over the orthographical images into phonological ones utilizing a phonetic letter set, fundamentally known as grapheme-to-phoneme transformation.

The idea of prosody is the blend of pressure example, musicality and pitch in a discourse. The displaying depicts the speaker's feeling. Late examinations recommend the distinguishing proof of the vocal highlights which flag passionate substance may make an exceptionally regular incorporated discourse.

The discourse will be talked by the voice attributes of a man, there are three sort of Acoustic blending accessible

1) Concatenative Synthesi
2) Formant Synthesi
3) Articulatory Synthesis

VI. CONCLUSIONS

The HAT for Vision isn't only a project that enables the oblivious in regards to end up free, but on the other hand is an asset saver. It chops down the cost of printing Braille books alongside the time and vitality spent into doing as such. This is a less expensive answer for one of the numerous difficulties that the outwardly debilitated face.

In future there is an extension for it can likewise stretch out for the long separation catching, and it can likewise execute for vertical perusing of the picture. Usage of an enhanced OCR strategy that can recognize scientific conditions, hand composing and different dialects (multilingual help).

REFERENCES

[1] S. Venkateswarlu, D.B.K. Kamesh, J.K.R. Sastry and Radhika Rani, Text to Speech Conversion, Indian Journal of Science and Technology, 2016.
[2] Anusha Bhargava, Karthik V. Nath, Pritish Sachdeva and Monil Samel, Reading Assistant for the Visually Impaired, International Journal of Current Engineering and Technology, 2015.
[3] Hideaki Goto and Takuma Hoda, Real-Time Text Tracking for Text-to-Speech Translation Camera for the Blind, Springer International Publishing, 2014.
[4] Christos Liambas and Miltidis Saratzidis, Autonomous OCR dictating system for blind people, Global Humanitarian Technology Conference (GHTC), 2016.
[5] K. Lakshmi and T. Chandra Shekhar Rao, Design And Implementation Of Text To Speech Conversion Using Raspberry PI, International Journal of Innovative Technology and Research (IJITR), 2016.
[6] Mallapa D. Gaurav, Shruti S. Salimath, Shruti B. Hatti, Vijayalaxmi I. Byakod and Shivleela Kanade, B-LIGHT: A Reading aid for the Blind People using OCR and OpenCV, International Journal of Scientific Research Engineering & Technology (IJSRET), 2017.
[7] Diwakar Srinath A., Praveen Ram A.R, Siva R., Kalaiselvi V.K.G., and Ajitha G., HOT GLASS – Human Face, Object and Textual Recognition for Visually Challenged, IEEE, 2017.
[8] Rajkumar N., Anand M.G., and Barathiraja N., Portable Camera-Based Product Label Reading For Blind People, International Journal of Engineering Trends and Technology (IJETT), 2014.
[9] Shagufta Md.Rafique Bagwan1 and L.J.Sankpal, VisualPal: A Mobile App for Object Recognition for the Visually Impaired, IEEE International Conference on Computer, Communication and Control (IC4-2015).
[10] Suraj.A.Khandare, Machine Learning with Text Recognition, International Journal of Engineering Reasearch &Technology (IJERT), Vol 3, Issue 3, March-2014.
[11] Roberto Neto, Nuno Fonseca, Camera Reading for Blind People, ELSEVIER, 2014.
[12] R. Shilkrot, J. Huber, C. Liu, P. Maes and N. S. Chandima, FingerReader: A Wearable Device to Support Text Reading on the go, CHI ’14 Ext. Abstr. Hum. Factors Comput. Syst., no. VI, pp. 2359-2364, 2014.