SPEECH CONVERSION FROM THE EXTRACTED TEXT OF IMAGE FOR BLIND PEOPLE

[1]Anugna akkala, [2]S. Prem kumar

ABSTRACT--*This paper is about a device that is used to convert image text into voice. Language is one of the main problems in communication. The primary motivation of this project is to provide a user friendly interface for blind people. Input is text for books, papers, magazines, etc., output is in speech form. This system is for blind people. Whatever the input given to the device, the output is produced in speech. Optical Character Recognition (OCR) engine is mainly used for conversion process. Python coding is used for conversion process. The main components are raspberry pi, camera module, headphones or speaker. Worldwide many people are blind and unable to read the text present in papers; this is the system very useful for blind people for knowing information in papers.*

Key words*: optical character recognition, passive infrared sensor, gray values, accuracy, manipulation.*

I INTRODUCTION

The process of conversion of image text into speech involves two modules. The first module named as image processing module which uses many image processing approaches and the second module is voice processing module. The former module involves extraction of text by OCR engine. OCR is an android mobile phone application used to convert printed text of scanned images into editable text. It translates text from any printed document or images into machine encoded form.

OCR follows some steps for text recognition process. The steps are 1) scanning of input document or image, 2) segmentation of text region, 3) pre-processing of text region, 4) feature extraction, 5) text recognition. The input is the handwritten or printed text from newspapers, books etc The voice processing module is conversion of text into speech in desired language using Google speech API (application program interface). It consists of Microsoft translator. A raspberry pi camera module is used to capture the image. It is attached to raspberry pi using a cable. It also requires SD card for storage and internet connection for transferring data via internet or wi-fi. The images are converted into machine encoded text. For taking image input from book a dilution algorithm is used. Erosion algorithm is used for extracting text from the captured image.

II LITERATURE SURVEY

[1] *Professor, Department of CSE, RISE Krishna Sai Prakasham Group of Institutions, Ongole, AP, India, Email:ratnajoyal@gmail.com*

[2] *Professor, Department of CSE, RISE Krishna Sai Prakasham Group of Institutions, Ongole, AP, India,*

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

In R.Mithe, et., al., has proposed an OCR approach. It converts different type of documents like images of digital cameras, pdf files, scanned paper documents into editable data. The performance of OCR depends on quality of input documents. It involves adaptive binarization, component analysis phase, words detection, two steps recognition and finally editable document. In OCR scanning at first images is captured using camera module. In general the letters are printed in black on a white background. If the text is in color the through thresholding process the image is converted into binary image. The gray levels below to threshold level is considered as black and above the threshold level is considered as white. Segmentation is the process of locating printed regions. Sometimes it may get confused due to splits and joint characters due to scanning process. So, some noise may occur. Noise in scanned documents results in poor recognition of characters. So, pre-processing is done to overcome this drawback. It is smoothing and normalization i.e., thining, unifrom size, slant and rotation of characters. Defects like distortion at edges may occur. So, to overcome it tesseract is used for extensibility and flexibility. Feature extraction is removing unwanted symbols such as open spaces, lines, intersections, and open spaces etc., tesseract algorithm is used for recognizing characters. This characters recognition is done word by word.

2)Archana shinde et., al. proposed a text segmentation for OCR after a text pre-processing approach that defines the accuracy of OCR. Pre-processing involves skewed input image and input image spectrum. In segmentation process, vertical and horizontal projection methods are used to segment text into lines and words. Horizontal projection profile of text document is found to separate text lines. To get horizontal projection profile, the number of white pixels in each row is counted. After plotting peaks and valleys are formed for de-skewed text documents with horizontal projection. Line segmentation is done at valleys of zero heights between text lines. Vertical projection is used to segment words. To get vertical projection profile, the number of white pixels in each column is counted. In English spacing between words is higher than spacing between characters in a word. Zero valued valleys width between words is higher than zero valued valleys width between characters in a word. So, this is used for separating words and to count input text lines.

3)Mrunmayee patil, et., al., proposed object recognition in an image. Matrix of square pixels arranged in columns and rows is called an image. To understand the text from the image in speech form, the image to text and then into speech conversion system is used. This system is mainly for blind people. The image segmentation and edge detection are the two main techniques used in this system. The process of conversion involves a) image pre-processing, b) information or feature extraction, c) detection of objects d) detection of edges, e) image segmentation and f) conversion of text to speech. Edge is defined as set of connected pixels that forms a boundary between two disjoint regions. Edge detection is the process of segmenting image into regions of discontinuity. For extracting edges from images prewitt, robert and sobel edge detection techniques are used but this techniques are not efficient. Then canny algorithm is developed by john f. canny which provides high probability of edge detection and error rate. It mainly focuses on low error rate, minimum response; minimize distance between real edge and detected edge. Image segmentation involves division of image into regions and this helps for correctly identifying images. This paper mainly focuses on recognition of object in an image.

4)Anusha bhargava, et., al., proposed a reading assistant for system for blind people. Brallie is used by visually impaired people for reading books which is very difficult. In this paper they explained that this system consists of small inbuilt camera that is used for scanning the paper and then a synthesized voice is used to convert scanned text into a audio format. This process not only save time and energy but also makes life easy for blind people. The image will have either binary content, the white will be used to give spacing between words or letters. At first, proper techniques are used to extract words. After extracting words they are converted into alphabets. So image processing consists of two parts a) getting region of interest b) segmenting words into alphabets. Tesseract is free OCR software. It gives accurate output and this is available in Linux, windows and mac OS. A open source software e-speak is installed for synthesizing speech. A python code is given to raspberry pi for giving instructions to capture an image, then for processing it, calling the tesseract OCR, then saves the file and finally for reading the saved text for speech output. It is a resource saver and also a system for blind people for being independent.

5)sagar patil, et., al., has proposed a single desktop application for both speech to text and vice-versa.. By using this application different tasks like text to text extraction from images, speech synthesis, text translation and speech recognition are completed. Text to speech is conversion of any type of chosen text to speech. In this when we give text as input, speech will be output. This process is called speech synthesis. For this text normalization is major task. In normalization phonetic transcription is assigned for each word.this text normalization eliminates puntuations and changes uppercase and lowercase letters. Text to speech includes prosodic modeling and intonation i.e., pitch, rate of speech, time of speech and pause between words. Text extraction module extract text from images and displays it on screen and this is done by OCR engine. In speech recognition process input is the audio and output is text. It consists of database of all words. So, it tells how words to be pronounced. The final module is text translation in which input is text and output is same text in other language. English is the base language. For this conversion the words are splited and then searched in dictionary for matching word than that matching word is displayed.

6)Nirmala kumari, et., al., proposed a text to speech conversion process in two parts. The first part is image processing and second process is voice processing. Image processing is done using OCRthrough its optical mechanism , it recognizes the characters. Matrix matching uses tesseract OCR. The tesseract OCR is provides very flexible and extensible process. It can also support 149 languages. Before giving image to OCR for high recognition accuracy it is converted into binary image. That conversion is nothing but image manipulation. If the font size is 14pt the tesseract OCR accuracy may decrease. In voice processing text to speech conversion is done. The festival software is used for conversion of text output from OCR(OCR) into speech. It is open source system. For the process start button is pressed than with a delay of 7 seconds image is captured and it is sent to OCR for text extraction, then the text is saved with a file name and later sent to festival software for speech conversion. Both capital and small letters are recognised by OCR. 38 to 42 centimeters distance is the reading range. It also recognizes numbers. 4 to 5 degree is the maximum tilt from vertical. 12 pt is the font size of character.

7)sangramsing N. kayte, et., al., evaluated the performance of speech synthesis techniques for English language. In artificial speech synthesis quality is important. The speech synthesized quality can be evaluated by two methods. They are subjective measurement and objective measurement. The audible level and perceptible level of output speech is speech quality measurement. Subjective perception and judgement process are important for measuring quality of produced speech. For assigning grades to speech quality a 5-point scale is used. In that grades assigning grade1 is for least quality speech and grade 5 means very excellent quality speech. Objective quality measure is a real time automated quality measurement. The computational algorithm is used by perceptual listener. Objective quality measurement helps for real time speech quality monitoring. When we compare quality measures, the objective quality measure is accurate and reliable. This objective measure uses mean square error (MSE) and peak signal to noise ratio (PSNR). The average of squares of errors or difference between estimator value and estimated value is called as MSE. Logarithmic decibel scale is used for measuring PSNR. The reconstructed signal or image quality can be measured using PSNR. The original is the signal and the error during synthesis is noise. PSNR is the ratio of possible power of a signal and noise that decreases the speech quality.

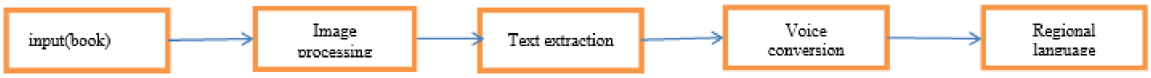
8)D. B .K .kamesh, et.,al.,explained about the system which is at low cost and helpful for blind people. According to some research in world 285 million people are visually impaired, blind people are nearly 39 million and people with low vision are nearly 246 million. This paper explains about combined technology of OCR and computer vision. Text to speech consists of three parts 1) image capturing 2) data managing 3) speech output. In conversion process noises occured in text is eliminated using some filters like gaussian filters etc., Blind people can face challenges in some places like banks, hospitals and other places where receipts are given in printed form. In this system a sensor named passive infrared sensor(PIR) is used for detecting obstacle. This sensor is connected to raspberry pi. If any person's hand is placed infront of this passive infrared sensor, than blind person is alerted by e-speak engine.Passive infrared sensor can detect the obstacle upto 7 meters.

9) nikisha jariwala, et., al.,developed a computer based system for text to speech conversion to hear the speech in Gujarati language . The ability to express the thoughts and emotions is possible only in speech form. The text to speech conversion is very difficult. Phone is the smallest sound unit that has definite shape. A group of phones with distintive units are called phonome. Diphone consists of pair of phonetic sounds that are kept adjacent to each other or two different sounds consisting of either vowels or consonants that are placed adjacent to each other. Text to speech conversion is artificially producing human from text. This text to speech conversion is mainly used for blind people and also have some other uses like public announcements at railway stations or airports, banks and all centers for providing services and to retrieve information.Education, Telecommunication, Games, multimedia, Aid to handicapped, Man machine communication etc., are fields that uses text to speech system.

III PROPOSED SYSTEM

There are already some systems that can convert text into speech, but this system is designed to for convert text into desired language speech output. This is not only for blind people but also for people who does not know English and want to translate text in their own native language. This device can give output in many languages.

BLOCK DIAGRAM



IV CONCLUSION

Received: 23 Dec 2019 | Revised: 05 Jan 2020 | Accepted: 27 Feb 2020

This paper is a review paper that explains about text to speech conversion system which is very useful in many fields. OCRplays a important role in conversion process. This system is used in many places like hospitals, education and other places where everything will be in text form and want it to convert into speech. If this is implemented with speech conversion in desired language then this device can also be used by the people who always travel from one place to another place. Dilusion algorithm and erosion algorithm is used for the conversion process.

REFERENCES

1. H. R. Mithe, S. Indalkar and N. Divekar. " Optical Character Recognition" International Journal of Recent Technology and Engineering (IJRTE), (2013) ISSN: 2277-3878,Volume-2.
2. ARCHANA A. SHINDE, D. "Text Pre-processing and Text Segmentation for OCR." International Journal of Computer Science Engineering and Technology (2012) pp. 810-812.
3. Runmayee patil, ramesh kagalkar "A review on conversion of image to text as well as speech using edge detection and image segmentation. International Journal of Science and Research (IJSR)" (2014) volume 3 issue.
4. Anusha bhargava, karthik V.nath, pritish sachdeva and monil samel."Reading assistant for the visually impaired". International journal of current engineering and technology" (2015) volume 5, No.2.
5. Sagar patil, Mayuri phonde, siddarth prajapati, saraga rane " Multilingual speech and text recognition and translation using image." International Journal of Engineering Research and Technology(IJERT), (2016) volume 5.
6. K Nirmala kumari , Meghana reddy J" Image text to speech conversion using OCR technique in raspberry pi." International Journal of Advanced Research in Electrical, electronics and instrumentation engineering (2016) volume 5.
7. Sangramsing, Monica mundada, santhosh gaikwad, bharti gawali N "Performance evaluation of speech synthesis techniques for english language." (2016) springer science +business media singapore.
8. D.B.K.kamesh, s. nazma, J.K.R. sastry and s.venkateswarlu " camera based text to speech conversion , obstacle and currency detection for blind people." International Journal of Science and Technology, volume 9(30).
9. Nikisha jariwala, bankin patel "A system for the conversion of digital gujarati text to speech for visually impaired people." 2018.
10. Maharaja, D., & Shaby, M. (2017). "Empirical Wavelet Transform and GLCM Features Based Glaucoma Classification from Fundus Image." International Journal of MC Square Scientific Research, 9(1), 78-85.
11. Saravanan, N. (2013). "Hand Geometry Recognition based on optimized K-means Clustering and Segmentation Algorithm." International Journal of MC Square Scientific Research, 5(1), 11-14.