Image Text Conversion from Regional Language to Speech/Text in Local Language

[1]Grasha Khiyani, [2]Rashmi Mohansingh Solanki, [3]Sufiyan Momin
[1,2,3] *Student, Department of Electronics and Telecommunications, Pimpri Chinchwad College of Engineering, Pune, India*

***Abstract*- Human communication today is mainly via speech and text. To access information in a text, a person needs to have vision. However, those who are deprived of vision can gather information using their hearing capability. The proposed project involves Text Extraction from image and converting the Text to Speech converter, a process which makes a person unaware about written language to read the text. This is the first step in developing a prototype for such people for recognizing the products in the real world, where the text on the sign board is extracted and converted into speech. This is carried out by using Raspberry pi, where portability is the main aim which is achieved by providing a battery backup and can be implemented as a future technology. The portability allows the user to carry the device anywhere and can use any time. Recently according to the survey carried out in metro cities, it was observed that 78% of the accidents were due to the driver's fault. This project aims at effective and efficient interpretation of the signs or regional languages. Thus, it would not just help the driver to drive smoothly but also allow him to be attentive. As many times observed, in hurry one may often neglect sign boards but with this project the display on the screen will make us aware of the important message. It can even use audio to make the driver aware more effectively thereby make the travel reliable.**

***Index terms*- Google Speech API, Raspberry pi, Tesseract OCR engine, USB camera module, Speakers**

I.INTRODUCTION

The main problem in communication is language bias between the communicators. This device basically can be used by people who do not know English and want it to be translated to their native language. The novelty component of this research work is the speech output which is available in 53 different languages translated from English but since amongst them Indian languages are three those are: Hindi, Bengali and Tamil. This paper is based on a prototype which helps user to hear the contents of the text images in the desired language. It involves extraction of text from the image and converting the text to translated speech in the user desired language. This is done with Raspberry Pi and a camera module by using the concepts of Tesseract OCR engine, Google Text to Speech converter which is a python interface for Google's Text to speech API. This relieves the travellers as they can use this device to hear the English text in their own desired language. It can also be used by the visually impaired. This device helps users to hear the images being read in their desired language.

II. LITTERATURE SURVEY

In the paper "Road Traffic Accidents in India: Issues and Challenges" [1] the main aim of this paper is to analyze the road accidents in India at national, state, and metropolitan city level. Analysis shows that the distribution of road accidental deaths and injuries in India varies according to age, gender, month and time. Age group 30-59 years is the most vulnerable population group, though males face. Higher level of fatalities and injuries than their female counterparts. Moreover, road accidents are relatively higher in extreme weather and during working hours. Analysis of road accident scenario at state and city level shows that there is a huge variation in fatality risk across states and cities. Fatality risk in 16 out of 35 states and union territories is higher than the all India average. Although, burden of road accidents in India is marginally lower in its metropolitan cities, almost 50% of the cities face higher fatality risk than their mofussil counterparts. In general, while in many developed and developing countries including China, road safety situation is generally improving, India

faces a worsening situation. Without increased efforts and new initiatives, the total number of road traffic deaths in India is likely to cross the mark of 250,000 by the year 2025.

The paper "Image text to speech conversion in the desired language using Raspberry" [2] There are already many systems which read images and give voice output. But this system gives voice output in any language desired by the user. This is done by capturing the image which is to be read using a raspberry pi camera module. Raspberry pi is a credit card sized single board computer. The operating system used is Raspbian. A 15 cm ribbon cable is used to attach the camera module to the raspberry pi. The coding is done using python language. The Optical character recognition engine converts the images of text into machine encoded text and saves it in a text file. Tesseract is the OCR engine which is used for extracting the English text from the image and storing it in a text file. The text to speech engine converts text to speech output. eSpeak is a speech synthesizer which can easily be used in raspberry pi for speech output in English. For translating it to other languages Google text to speech engine and Microsoft translator is used [1][2]. Google text to speech is a screen reader which speaks the text on the screen. Microsoft translator is a multilingual statistical machine translation cloud service provided by Microsoft. It supports 53 different language systems. This translated speech output could be heard through speakers or headset. The platform being used for simulation of this model is putty in SSH (Secure Socket Shell). Generally, it is done using MATLAB, but it is different here because translation module is an added feature.

In the paper "Implementation of a reading device for Bengali speaking visually handicapped people.[3]" The motive of this paper is to implement the image processing techniques by extracting the texture features from the medical echocardiography images, combining intensity histogram features and Gray Level Co-occurrence Matrix (GLCM) features, then by applying neural network for automatic classification using back-propagation algorithm to classify heart valve diseases more accurately. The precision, recall and accuracy terms were evaluated for performance of the proposed system. The experimental results prove the efficiency of the proposed method of providing good classification efficiency.

In the paper "Detecting text-based image with optical character recognition for English translation and speech [4]". Smartphones have been known as most used electronic devices in daily life today. As hardware embedded in smartphones can perform much more task than traditional phones, the smartphones are no longer just a communication device but also considered as a powerful computing device which able to capture images, record videos, surf the internet and etc. With advancement of technology, it is possible to apply some techniques to perform text detection and translation. Therefore, an application that allows smartphones to capture an image and extract the text from it to translate into English and speech it out is no longer a dream. In this study, an Android application is developed by integrating Tesseract OCR engine, Bing translator and phones' built-in speech out technology. Final deliverable is tested by various type of target end user from a different language background and concluded that the application benefits many users. By using this app, travelers who visit a foreign country able to understand messages portrayed in different language. Visually impaired users are also able to access important message from a printed text through speech out feature.

Figure 1: Raspberry pi 2

III. SYSTEM OVERVIEW

In the block diagram as shown in Figure 2 the basic inclusions are power supply, USB camera, power supply speakers and push buttons. The USB camera will take the snaps it would be placed on the top of the vehicle these snaps will be processed by the Raspberry pi to get the conversion from image to text and text to speech be done satisfactorily. After conversion buttons are interfaced to choose the

language preferred by the driver. For this project the provision given of three languages i.e. Hindi, Bengali and Tamil. Driver can choose the language according to his/her preference. Once the conversion is done it could be heard by the audio speaker.

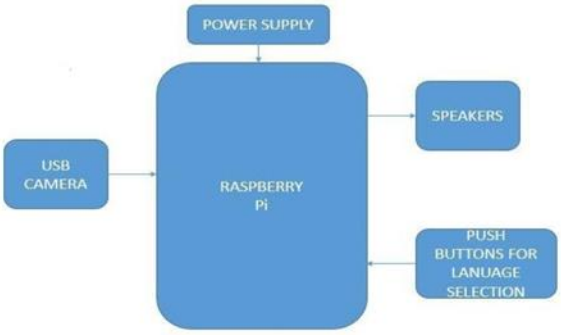

Figure 2: Block diagram

IV. WORKING METHDOLOGY AND IMPLEMENTATION

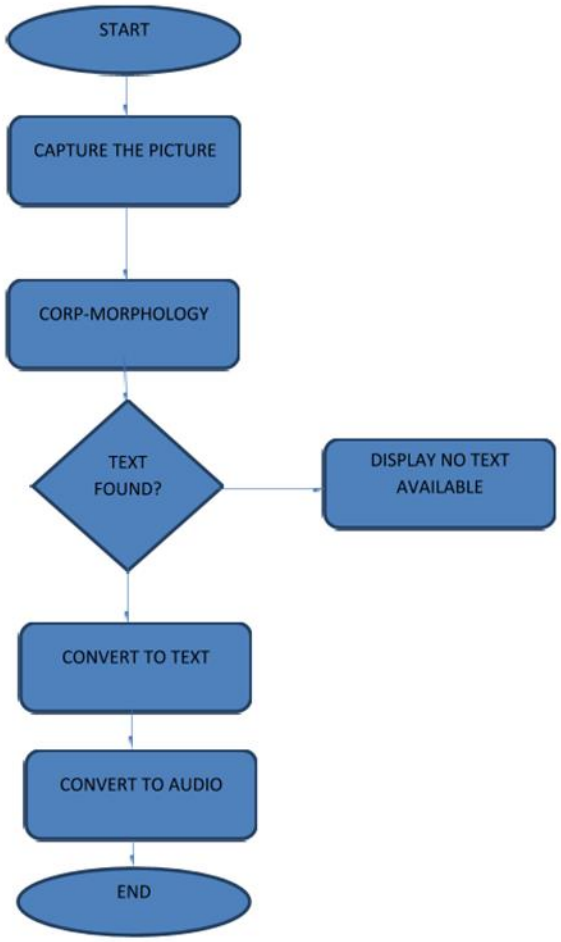

Figure 3: Flowchart

First captured image of the sign board using high resolution USB camera module 2.0 programmed it in such a way that it takes 20 frames of images and keeps the image which has best resolution and discard the other frames. Initially the resolution of camera is set by using cheese command.

Secondly by using crop morphology technique one can extract the image text and then convert it into desired audio. Morphology is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, one can construct a morphological operation that is sensitive to specific shapes in the input image.

The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. In the morphological dilation and erosion operations, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image.

Next activity is to convert the text into speech by using Google text to speech conversion and it translates the extracted text into the required language using the GPIO pins. In this project programmed the 3 GPIO pins according to the languages such as pin number 14 for Hindi and pin number 15 for Bengali and pin number 18 for Tamil. Thus, in this way image text can be converted into required regional language.

It consists of two main modules:

1. Image processing module
2. Voice processing module

Image processing module

Software processes the input image and converted into text format. This can be done using two techniques:

Optical Character Recognition

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.

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 processing is completed, the content of the output is present in .txt file. In simple images with or without color (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 grayscale mode as compared to color 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.

Voice processing module

A text to speech (TTS) synthesizer is a computer based system that can read text aloud automatically, regardless of whether the text is introduced by a computer input stream or a scanned input submitted to an Optical character recognition (OCR) engine. A speech synthesizer can be implemented by both hardware and software. Speech is often based on concatenation of natural speech i.e. units that are taken from natural speech put together to form a word or sentence.

Figure 4: Conversion module setup

The above figure 4 shows the experimental result in which user can capture the picture from the camera then it sends to the processor there are convertors for converting the image to the text.

Figure 5: USB camera module

Figure 5 shows the USB camera module which is used for capturing of the images.

Figure 6: Image processing on still images

Figure 6 shows the results of image processing on still images.

Figure 7: Image processing on captured images

Figure 7 shows the grey image of the captured image which is converted by using cv2.cvtColor (image, cv2.COLOR_BGR2GRAY) this statement. In this an output image is created which is used to store this gray image and which can be referred later for the reference purpose.

Detailed description of the commands and libraries used in our project are as follows:

Open CV

Open CV -Python is a library of Python bindings designed to solve computer vision problems. Python is a general-purpose programming language. Compared to languages like C/C++, Python is slower. That said, Python can be easily extended with C/C++, which allows us to write computationally intensive code in C/C++ and create Python wrappers that can be used as Python modules. This gives us two advantages such as first, the code is as fast as the original C/C++ code (since it is the actual C++ code working in background) and second, it easier to code in Python than C/C++. OpenCV-Python is a Python wrapper for the original OpenCV C++ implementation.

PIL

Python Imaging Library is used to access python for image processing.

NumPy

NumPy is the fundamental package for scientific computing with Python. It has powerful N-dimensional array object, sophisticated (broadcasting) functions for integrating. This makes it ideal for image processing applications.

Pytesseract

Python tesseract is an Optical Character Recognition (OCR) tool for python which can recognize and read the text embedded in images. It is useful as a stand-alone innovation script to tesseract, as it can read image type supported by PIL including .png, .tiff whereas tesseract OCR by default supports .tiff and .bmp formats. Additionally, it helps in printing the recognized text instead of writing it to the file.

gTTS

There are several APIs available to convert text to speech in python and one of such APIs is the Google Text to Speech API commonly known as the gTTS API. gTTS is a very easy to use tool which converts the text entered, into audio which can be saved as a mp3 file. The gTTS API supports several languages including English, Hindi, Tamil and many more. The speech can be delivered in any one of the two available audio speeds, fast or slow.

Counterpart removal

For efficient light intensity control the threshold adjustment was done. But apart from light the other problem faced was the counters present like trees, background texture and many more. To overcome this, here opted to crop morphology this method distributes the image into small parts and perform image to text conversion of each portion of the image. Thus, the converted text was more accurate and counterpart removal was successful.

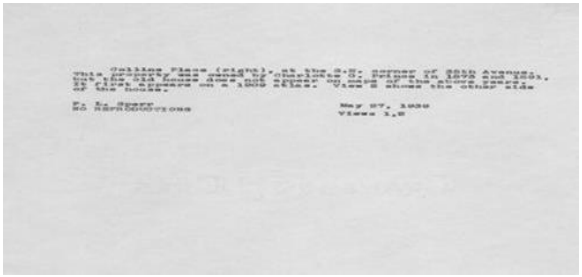

Figure 8: Sample image

Figure 9: Cropped sections

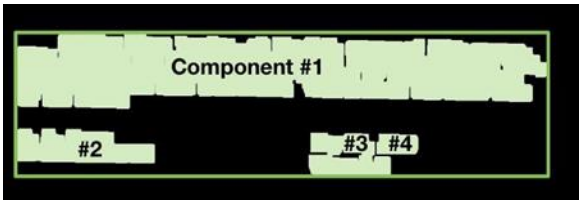

Figure 10: Sections encountered

The images above in figure 8, 9 and 10 show the counter removal using crop morphology.

V. CONCLUSION

The text-to-speech can change the text in image into speech with high performance of the images in which text is readable accurately. This system will help to reduce the number of accidents and will ensure safe

and reliable journey. By implementing this system language barrier problem can be eliminated. And with the help of the translation tools he/she can convert the text to the desired language and then again by using the Google speech recognition tool. Also, it is less costly as compared to other implementations. Text-to-Speech device can change the text image input into sound with a performance that is high enough and a readability tolerance of less than 2%, with the average time processing less than three minutes for A4 paper size. This portable device does not require internet connection and can be used independently by people if images mp3 files are downloaded and provided as a database using Google API.

VI. FUTURE SCOPE

For future scope, the delay in sound wave causes the speech to look very unnatural. Removing the delay manually every time may be a tedious job. To overcome this problem, further one can think of a method which recognizes the delay and automatically remove it. And the best method would be 'integration'. Through integration of the synthesized speech, one may not only avoid delay but also get a continuous flow of speech. In this project only on text documents are used. Further reading word files, scanned data, PDF files etc. will be a part of future scope.

REFERENCES

[1] Singh, S.K., 2017. Road traffic accidents in India: issues and challenges. Transportation research procedia, 25, pp.4708-4719.

[2] Rithika, H. and Santhoshi, B.N., 2016, December. Image text to speech conversion in the desired language by translating with Raspberry Pi. In Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on (pp. 1-4). IEEE.

[3] Sarkar, M.M., Datta, S. and Hassan, M.M., 2017, December. Implementation of a reading device for Bengali speaking visually handicapped people. In Humanitarian Technology Conference (R10-HTC), 2017 IEEE Region 10 (pp. 461-464). IEEE.

[4] Ramiah, S., Liong, T.Y. and Jayabalan, M., 2015, December. Detecting text-based image with optical character recognition for English translation and speech using Android. In Research and Development (SCOReD), 2015 IEEE Student Conference on (pp. 272-277). IEEE.

[5] Sumathi, C.P. and Priya, S.N., 2013, February. Text extraction from assorted images using morphological-region, texture and multiscale techniques-A comparative study. In Information Communication and Embedded Systems (ICICES), 2013 International Conference on (pp. 294-299). IEEE.