Text Reader for Visually Impaired Using Google Cloud Vision API

Paras Doshi[1], Yash Shirke[2], Tejas Hegde[3], Pranav Dhanvij[4]
[1,2,3,4] *B. Tech Electronics Final year, Department of Electrical Engineering, Veermata Jijabai Technological Institute, Matunga, Mumbai 400019, India*

***Abstract*- Visually impaired people confront a number of visual challenges every day – from reading the label on a frozen dinner to figuring out if they're at the right bus stop. Probable solutions include Braille wherein tactile information is converted into meaningful patterns. Other visual aids include liquid level indicators, coin sorters and large button telephones for daily living; electronic magnifiers, audio books, text to voice technology as a technological aid. Our aim through this paper is to propose a system that facilitates reading for a blind person. With the help of our system, we extract text from images using google cloud vision API. Our approach is capable of recognizing text in various challenging conditions where traditional OCR systems fail; in the presence of blur, low resolution, low contrast, high image noise, and distortions. The output text is converted into audio output in the form of synthetic speech. Thus, our proposed system will be very helpful to visually impaired person.**

***Index Terms*- OCR, Google Cloud Vision API, Text to Speech, Raspberry Pi.**

I. INTRODUCTION

Good vision is a precious gift. Not everyone is fortunate to enjoy this. Sometimes they turn out to have poor vision or sometimes even worse, blind. According to the World Health Organization, an estimated 253 million people live with vision impairment: 36 million are blind and 217 million have moderate to severe vision impairment. India has the highest population of blind people in the world. Approximately 1 out of every 4 individuals who are blind, live in India.

The World Health Organization (WHO) defines blindness as visual sharpness of less than 3/60, or a corresponding visual field loss to less than 10 degrees in the better eye, even with the best possible spectacle correction. The National Programme for Control of Blindness (NPCB) in India, on the other hand, defines blindness as vision of 6/60 or less and a visual field loss of 20 degrees or less in the better eye, after spectacle correction.

Vision is important not only for seeing objects but also for dark adaptations, contrast sensitivity, balance and colour perceptions. Though all these functions are lost in visually impaired people, yet they rely on other senses to carry out not only their activities of daily living but also participate in economically gainful employment. This is where Braille comes in handy.

Trouble reading is the most commonly reported problem of people with low vision, regardless of the underlying cause of their vision loss. In order to tackle the problem of difficulty in reading, we propose a system employing the OCR technology. This technology empowers the blind to scan images, extract the text from it and get the words spoken out as audio output. We propose a system wherein an image consisting of text is captured by the Raspberry Pi camera module. Earlier work[1] indicated that the noise present in the image distorted the final result. This image is then pre-processed to remove some noise in order to decrease the processing delay by Google Cloud. The filtered image is then sent to the Google Cloud which detects the text in the image and sends it to the Raspberry Pi. The Raspberry Pi then converts the text into speech.

II. LITTERATURE SURVEY

In paper [1] ,Google Cloud Vision API was used for image analysis. Their project detects individual objects and faces within images, also finds and reads printed words contained within images. Paper evaluates the robustness of Google Cloud Vision API to input noise. In particular, Set of images are taken and noise is added to them then API is unable to detect correct text or object were as if the noise is removed then the output is similar to that of the original image. cloud vision API can benefit from noise filtering

The paper [2] proposed a prototype that helps people to hear the text content of the image in their native language. The text is extracted from the image and then text is converted to translate speech of users native language. Camera captures the image and then OCR engine convert image to text. Then text is converted into speech using espeak TTS engine. The speech output is then stored in a flac file. This file is then converted into desired language by the Microsoft Translator using a python script.

Paper [3] proposed a system that reads text on a captured Image. It is performed as text is extracted from scanned image using Tesseract Optical Character Recognition (OCR) and then converting the text to speech by e-Speak tool. Fist captured image is converted to grayscale and then filtered using Gaussian filter to reduce noise adaptive Gaussian thresholding is used then it is converted to binary image and cropped and loaded to tesseract OCR for text recognition and output of tesseract is text file which is input for e-speak, which produce audio . Along with text, face detection is also possible

Figure 1: Raspberry pi 3

III. SYSTEM OVERVIEW

A computer vision system basically consists of a camera to capture the image and a computing device which analyses the image captured. The proposed system consists of Raspberry Pi 3 Model B (Fig 1),used for image analysis and a smartphone camera which works as an image sensor to capture the image. And lastly the Google Cloud Vision API is used which processes images on the Google Cloud Platform.

Raspberry Pi 3 Model B is an inexpensive credit card size Linux computer. It consists of Quad Core Broadcom BCM2837 64-bit ARMv8 processor, with a processor speed of 1.2 Ghz. Raspberry Pi runs Debian based GNU/Linux operating system Rasbian.

The operating system used on our embedded system is the latest Raspbian stretch which facilitates easy operation of the connected hardware to provide the best results.

The other major component of the proposed systemis the Google Cloud Vision API. It was released in the year 2015. The Google Vision API provides a RESTful interface that quickly analyses image content. This interface hides the complexity of continuously evolving machine learning models and image processing algorithms. It enables developers to analyze the content of images. It uses powerful machine learning tools to extract the necessary data from images. It can perform different functions such as label detection, face detection, Logo detection, Optical Character Recognition (OCR) etc. Google Cloud Vision API's performance in the presence of noise is not very dependable[1] and hence before sending the image to the Google Cloud .

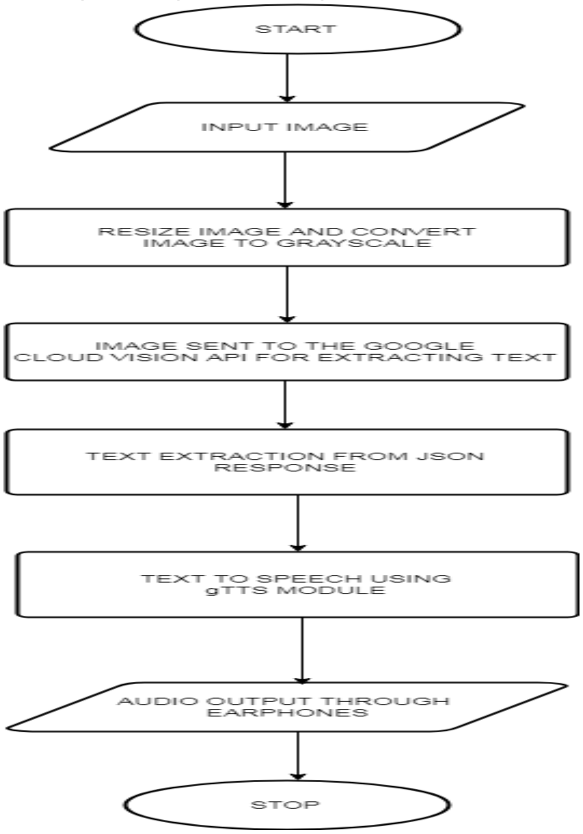

START
INPUT IMAGE
RESIZE IMAGE AND CONVERT IMAGE TO GRAYSCALE
IMAGE SENT TO THE GOOGLE CLOUD VISION API FOR EXTRACTING TEXT
TEXT EXTRACTION FROM JSON RESPONSE
TEXT TO SPEECH USING gTTS MODULE
AUDIO OUTPUT THROUGH EARPHONES
STOP

Figure 2: Flowchart

IV. WORKING METHDOLOGY AND IMPLEMENTATION

The android application developed has a user friendly interface for visually impaired people. It consists of speech recognition activity by which the camera is started the user can then point The camera at any point and capture the image. For sending the image, the application starts a Bluetooth activity which now displays the list of connected devices. From the list of devices raspberry –pi is selected and image is sent to the raspberry pi.

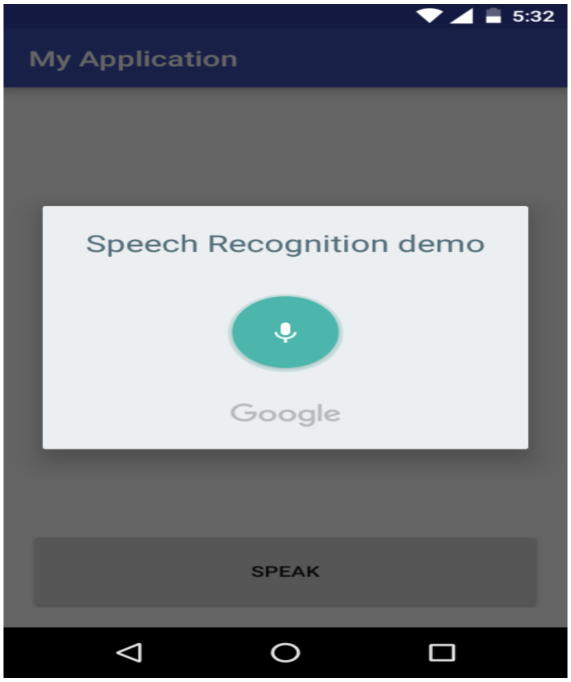


Figure 3: App Screenshot 1

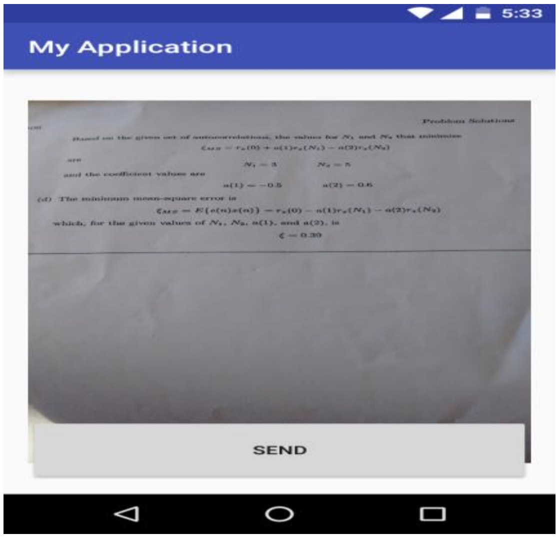


Figure 4: App Screenshot 2

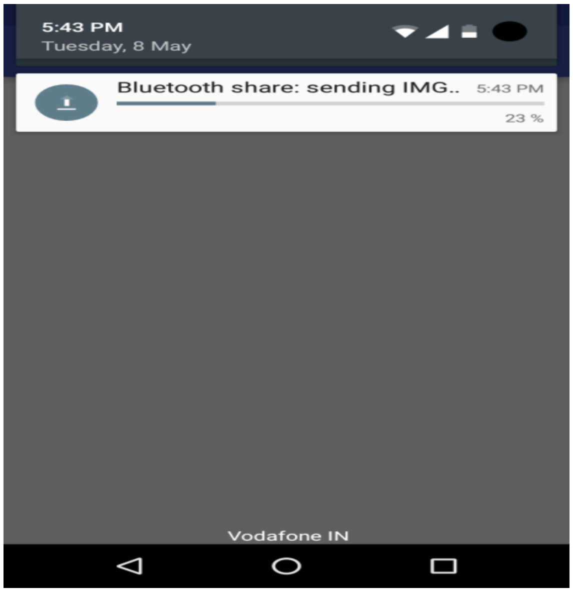


Figure 5: Image Transfer

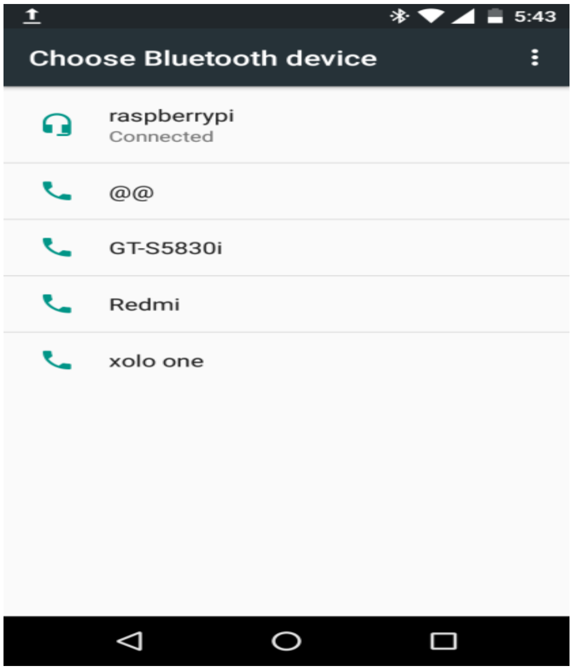


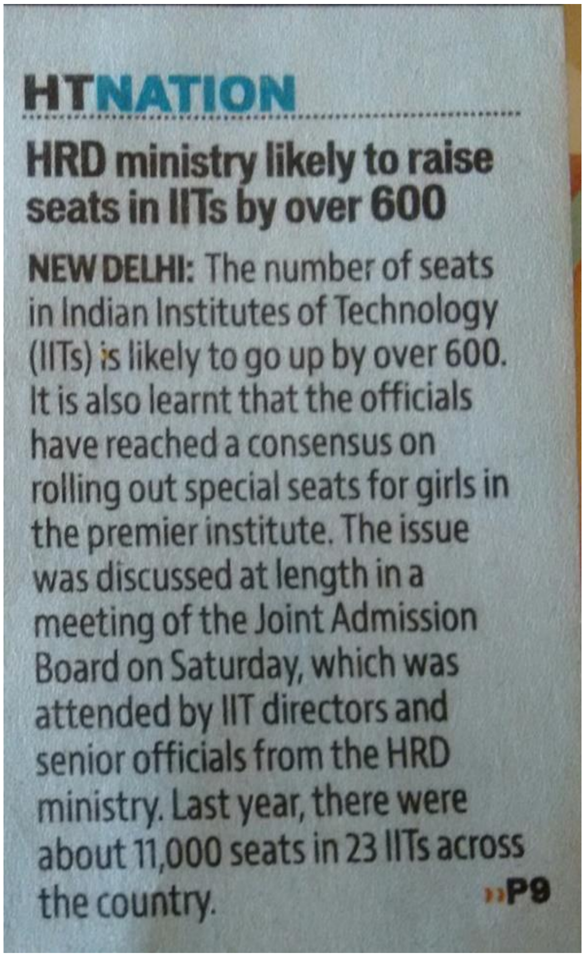
Figure 6: Connection Process

The image is then resized to 720 x 480 pixels since greater the size greater is the time taken for OCR, after resizing the image is converted into grayscale. All the preprocessing steps were performed to reduce the time required for image to text translation.

The Google Cloud then performs OCR(Optical Character Recognition) on the processed (resized and grayscale) image sent to it. In the Google Cloud everything is encapsulated in a RESTful API which returns a bounding box(information about the location of text in an image with the help of x and y co-ordinates).The Google Cloud can recognize different languages. The image resolution is so chosen that we get the correct output in minimum time. As the resolution of the image is increased the amount of time taken by the Google Cloud to perform OCR on it increases. The response obtained is in the form of JSON structure. The text is obtained and then converted into speech using the gTTS engine. The text is stored as an mp3 file and then played using an mp3 player.

V. RESULT

Text is extracted from image and converted to audio. It recognizes different fonts. Skewed text images are also identified and converted into speech. The model recognizes the text which is readable by human eyes. The initial system proposed the use of raspberry pi camera module but the image quality observed was poor and hence this system makes the use of a smartphone camera for better resolution pictures

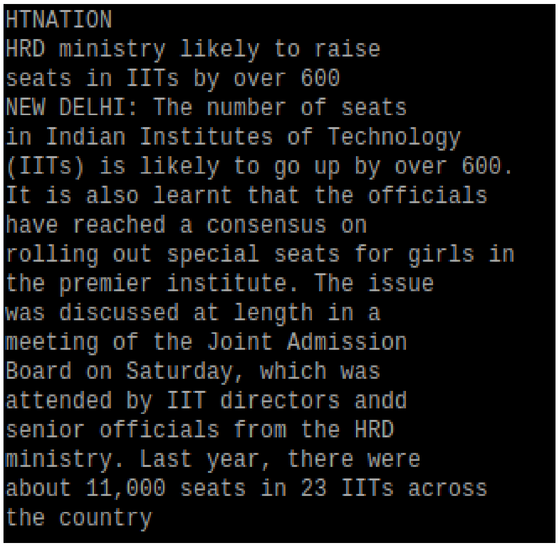


HTNATION

HRD ministry likely to raise seats in IITs by over 600

NEW DELHI: The number of seats in Indian Institutes of Technology (IITs) is likely to go up by over 600. It is also learnt that the officials have reached a consensus on rolling out special seats for girls in the premier institute. The issue was discussed at length in a meeting of the Joint Admission Board on Saturday, which was attended by IIT directors and senior officials from the HRD ministry. Last year, there were about 11,000 seats in 23 IITs across the country. »P9

Figure 7: Input Image 1



```
HTNATION
HRD ministry likely to raise
seats in IITs by over 600
NEW DELHI: The number of seats
in Indian Institutes of Technology
(IITs) is likely to go up by over 600.
It is also learnt that the officials
have reached a consensus on
rolling out special seats for girls in
the premier institute. The issue
was discussed at length in a
meeting of the Joint Admission
Board on Saturday, which was
attended by IIT directors andd
senior officials from the HRD
ministry. Last year, there were
about 11,000 seats in 23 IITs across
the country
```

Figure 8: Text Detected in Image 1

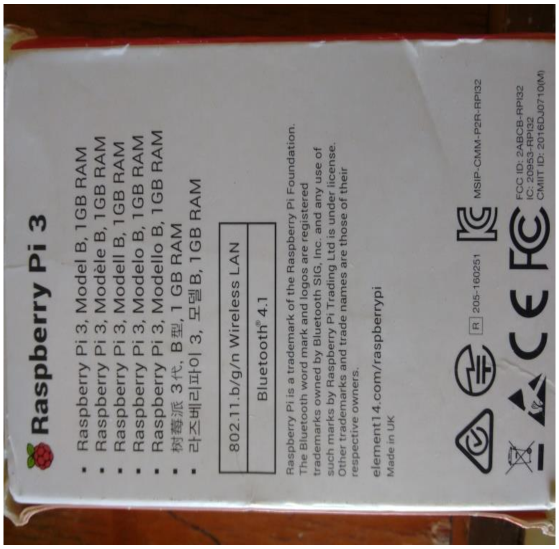
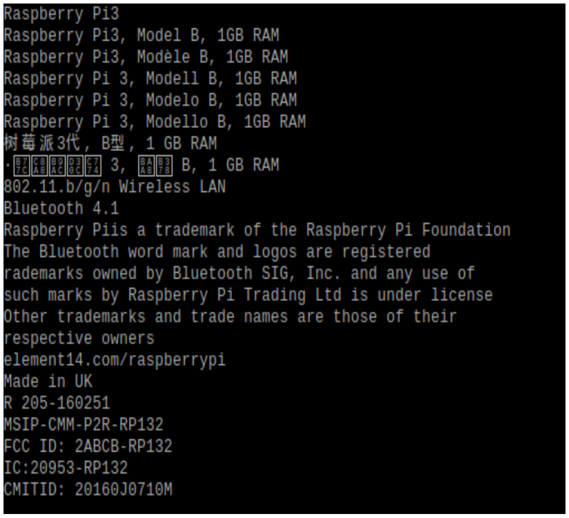
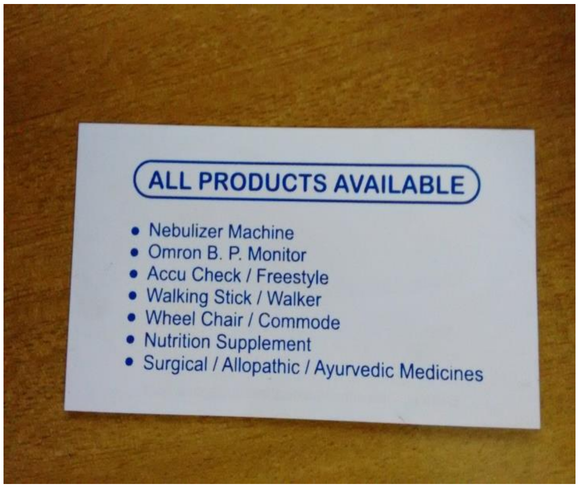

Raspberry Pi 3
Raspberry Pi 3, Model B, 1GB RAM
Raspberry Pi 3, Modèle B, 1GB RAM
Raspberry Pi 3, Modell B, 1GB RAM
Raspberry Pi 3, Modelo B, 1GB RAM
Raspberry Pi 3, Modello B, 1GB RAM
树莓派 3代, B型, 1 GB RAM
라즈베리파이 3, 모델 B, 1GB RAM
802.11.b/g/n Wireless LAN
Bluetooth® 4.1
Raspberry Pi is a trademark of the Raspberry Pi Foundation.
The Bluetooth word mark and logos are registered
trademarks owned by Bluetooth SIG, Inc. and any use of
such marks by Raspberry Pi Trading Ltd is under license.
Other trademarks and trade names are those of their
respective owners.
element14.com/raspberrypi
Made in UK
R 205-160251
MSIP-CMM-P2R-RPI32
FCC ID: 2ABCB-RPI32
IC: 20953-RPI32
CMIIT ID: 2016DJ0710(M)

Figure 9: Input Image 2



```
Raspberry Pi3
Raspberry Pi3, Model B, 1GB RAM
Raspberry Pi3, Modèle B, 1GB RAM
Raspberry Pi 3, Modell B, 1GB RAM
Raspberry Pi 3, Modelo B, 1GB RAM
Raspberry Pi 3, Modello B, 1GB RAM
树莓派3代, B型, 1 GB RAM
·라즈베리파이 3, 모델 B, 1 GB RAM
802.11.b/g/n Wireless LAN
Bluetooth 4.1
Raspberry Piis a trademark of the Raspberry Pi Foundation
The Bluetooth word mark and logos are registered
rademarks owned by Bluetooth SIG, Inc. and any use of
such marks by Raspberry Pi Trading Ltd is under license
Other trademarks and trade names are those of their
respective owners
element14.com/raspberrypi
Made in UK
R 205-160251
MSIP-CMM-P2R-RP132
FCC ID: 2ABCB-RP132
IC:20953-RP132
CMITID: 20160J0710M
```

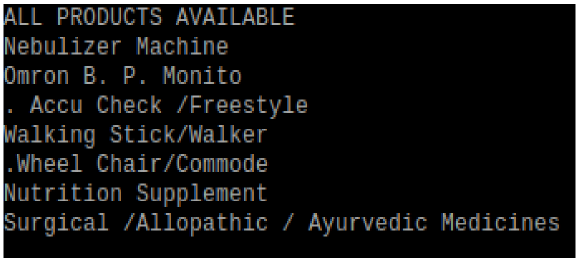
Figure 10: Text Detected in Image 2



ALL PRODUCTS AVAILABLE

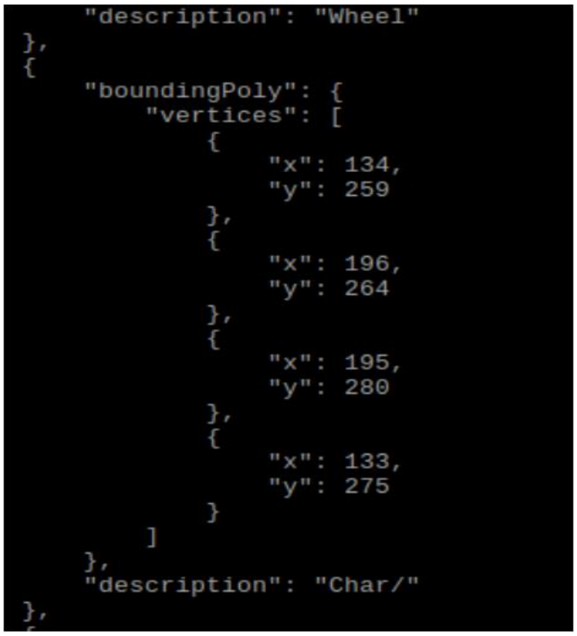
- Nebulizer Machine
- Omron B. P. Monitor
- Accu Check / Freestyle
- Walking Stick / Walker
- Wheel Chair / Commode
- Nutrition Supplement
- Surgical / Allopathic / Ayurvedic Medicines

Figure 11: Input Image 3



```
ALL PRODUCTS AVAILABLE
Nebulizer Machine
Omron B. P. Monito
. Accu Check /Freestyle
Walking Stick/Walker
.Wheel Chair/Commode
Nutrition Supplement
Surgical /Allopathic / Ayurvedic Medicines
```

Figure 12: Text Detected in Image 3



```
     "description": "Wheel"
  },
  {
     "boundingPoly": {
         "vertices": [
            {
                "x": 134,
                "y": 259
            },
            {
                "x": 196,
                "y": 264
            },
            {
                "x": 195,
                "y": 280
            },
            {
                "x": 133,
                "y": 275
            }
         ]
     },
     "description": "Char/"
  },
```

Figure 13: Output Obtained in JSON format for Image 3

The time taken for image to text translation is around 15 seconds. The gTTS engine then takes around 15 more seconds to convert text into audio.

VI. CONCLUSION

The text-to-speech can change the text in image into speech with high performance of the images in which text is readable by the naked eye accurately. As the resolution of image increases the time required for the processing also increases. Hence the images are first resized to 740 x 480 pixels and then the image is converted to grayscale in order to minimize the time required for image to text translation. If the pre-processing steps are skipped, then the time required increases to around 50 seconds and with the same result as above.

VII. FUTURE SCOPE

For future work in the model the time required for the processing can be reduced to a few seconds. Font independent model can be made. Text can be changed in non-English languages. Conversion of text image with multi-lingual script can be implemented. Cursive characters can be identified and converted to speech.

REFERENCES

[1] Hossein Hosseini, Baicen Xiao and Radha Poovendran “Google’s Cloud Vision API Is Not Robust To Noise” 16th IEEE International Conference on Machine Learning and Applications December 18-21, 2017

[2] Rithika.H , B. Nithya santhoshi “Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi” International Conference on Computational Intelligence and Computing Research 2016

[3] Mr.Rajesh M., Ms. Bindhu K. Rajan Ajay Roy, Almaria Thomas K, Ancy Thomas, Bincy Tharakan T, Dinesh C “Text recognition and face detection aid for visually impaired person using raspberry pi” International Conference on circuits Power and Computing Technologies [ICCPCT] July 2017

[4] Davide Mulfari, Antonio Celesti, Maria Fazio, Massimo Villari and Antonio Puliafito “Using Google Cloud Vision in Assistive Technology Scenarios” IEEE Workshop on ICT solutions for eHealth 2016