Text to Speech Conversion

Article *in* Indian Journal of Science and Technology · October 2016

DOI: 10.17485/ijst/2016/v9i38/102967

4 authors:

S. Venkateswarlu
K L University
28 PUBLICATIONS **96** CITATIONS
SEE PROFILE

Duvvuri B K Kamesh Duvvuri
Malla Malla Reddy Engineering college for Women
34 PUBLICATIONS **178** CITATIONS
SEE PROFILE

Sastry Jammalamadaka
K L University
95 PUBLICATIONS **511** CITATIONS
SEE PROFILE

Chintala Radhika Rani
K L University
22 PUBLICATIONS **82** CITATIONS
SEE PROFILE

Indian Journal of Science and Technology, *Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Text to Speech Conversion

S. Venkateswarlu[1]*, D. B. K. Kamesh[1], J. K. R. Sastry[2] and Radhika Rani[2]

[1]Department of CSE, K L University, Vaddeswarm, Guntur – 522502, Andhra Pradesh, India; somu23@kluniversity.in, kameshdbk@kluniversity.in, radhikarani_cse@kluniversity.in
[2]Department of ECM, K L University, Vaddeswarm, Guntur – 522502, Andhra Pradesh, India; drsastry@kluniversity.in

Abstract

The present paper has introduced an innovative, efficient and real-time cost beneficial technique that enables user to hear the contents of text images instead of reading through them. It combines the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer (TTS) in Raspberry pi. This kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction from color images is a challenging task in computer vision. Text-to-Speech conversion is a method that scans and reads English alphabets and numbers that are in the image using OCR technique and changing it to voices. This paper describes the design, implementation and experimental results of the device. This device consists of two modules, image processing module and voice processing module. The device was developed based on Raspberry Pi v2 with 900 MHz processor speed.

Keywords: Image Processing, OCR, Text Extraction, Text-to-speech, Voice Processing

1. Introduction

Optical character Recognition (OCR) is a process that converts scanned or printed text images[1], handwritten text into editable text for further processing. This paper has presented a robust approach for text extraction and converting it to speech. Testing of device was done on raspberry pi platform. The Raspy is initially connected to the internet through VLAN. The software is installed using command lines. Following steps are to be followed:

1. The first setup is to download the installation script,
2. Second step is to convert it to executable form and
3. The last step starts the script which does the rest of the installation work.

Device set up is done as shown in Figure 1. The web-cam is manually focused towards the text. Then, it takes a picture; a delay of around 7 seconds is provided, which helps to focus the webcam, if it is accidently defocused. After delay, picture is taken and processed by Raspy to hear the spoken words of the text through the earphone or speaker plugged into Raspy through its audio jack.

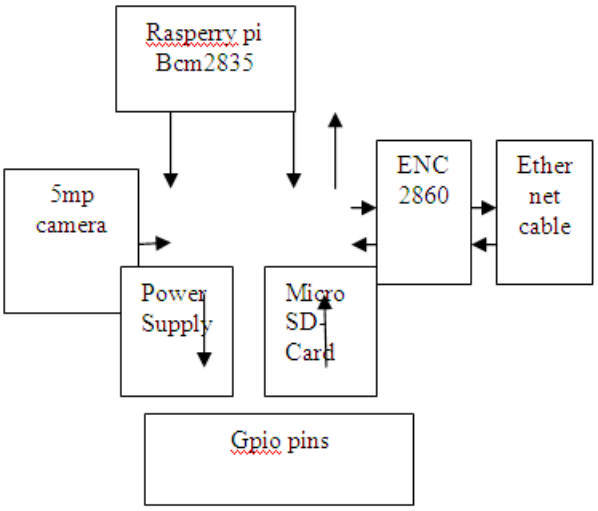

Figure 1. Block diagram of text to speech conversion.

2. Methodology

Text-to-speech device consists of two main modules, the image processing module and voice processing modules.

*Author for correspondence

Image processing module captures image using camera, converting the image into text. Voice processing module changes the text into sound and processes it with specific physical characteristics so that the sound can be understood. Figure 2 shows the block diagram of Text-To-Speech device, 1st block is image processing module, where OCR converts .jpg to .txt form. 2nd is voice processing module which converts .txt to speech

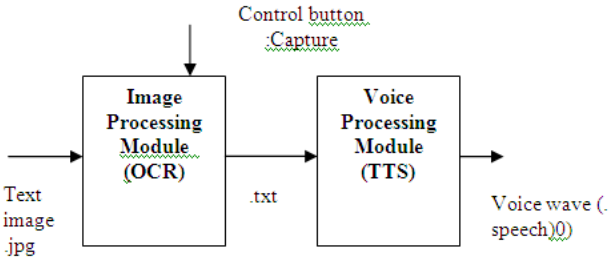

Figure 2. Block diagram of text-to-speech device.

Figure 2shows the block diagram of Text-To-Speech device, 1st block is image processing module, where OCR converts .jpg to .txt form. 2nd is voice processing module which converts .txt to speech. OCR is important element in this module. OCR or Optical Character Recognition is a technology that automatically recognize the character through the optical mechanism, this technology imitate the ability of the human senses of sight, where the camera becomes a replacement for eye and image processing is done in the computer engine as a substitute for the human brain[2]. Tesseract OCR is a type of OCR engine with matrix matching[3]. The selection of Tesseract engine is because of its flexibility and extensibility of machines and the fact that many communities are active researchers to develop this OCR engine and also because Tesseract OCR can support 149 languages. In this project we are identifying English alphabets. Before feeding the image to the OCR, it is converted to a binary image to increase the recognition accuracy. Image binary conversion is done by using Imagemagick software, which is another open source tool for image manipulation. The output of OCR is the text, which is stored in a file (speech.txt). Machines still have defects such as distortion at the edges and dim light effect, so it is still difficult for most OCR engines to get high accuracy text[4]. It needs some supporting and condition in order to get the minimal defect. *Tesseract OCR Implementation.*

2.1 Software Design

Software processes the input image and converted into text format. The software implementation is showed in Figure 3.

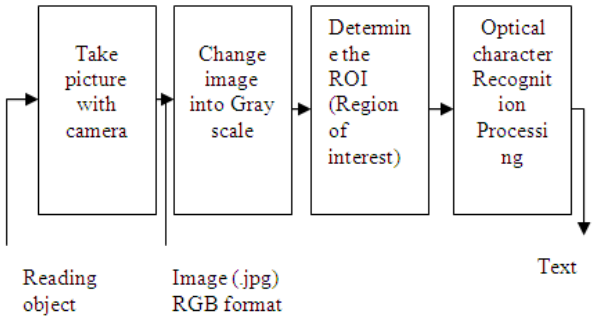

Figure 3. Software design of image processing module.

2.2 The Voice Processing Module

In this module text is converted to speech. The output of OCR is the text, which is stored in a file (speech. txt). Here, Festival software is used to convert the text to speech. Festival is an open source Text To Speech (TTS) [7,8] system, which is available in many languages. In this project, English TTS [9–11]system is used for reading the text.

3. Results

Observed outcome of project:

- Text is extracted from the image and converted to audio.
- It recognizes both capital as well as small letters.
- It recognizes numbers as well.
- Range of reading distance was 38-42cm.
- Character font size should be minimum 12pt.
- Maximum tilt of the text line is 4-5 degree from the vertical.

4. Conclusion

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. Through

this method, we can make editing process of books or web pages easier.

5. References

1. Archana A, Shinde D. Text pre-processing and text segmentation for OCR. International Journal of Computer Science Engineering and Technology. 2012:810–12.
2. Mithe R, Indalkar S, Divekar N. Optical character recognition. International Journal of Recent Technology and Engineering. 2013 Mar; 2(1).
3. Smith R. An overview of the Tesseract OCR engine, USA: Google Inc; 2007.
4. Shah H, Shah A. Optical character recognition of Gujarati numerical. International Conference on Signals, Systems and Automation. 2009; 49–53.
5. Monk S. Raspberry pi cook.
6. Text localization and extraction in images using mathematical morphology and OCR Techniques; 2013.
7. Vanitha E, Kasarla PK, Kuamarswamy E. Implementation of text- to-speech for real time embedded system using Raspberry Pi processor. International Journal and Magazine of Engineering Technology Management and Research. 2015 Jul:1995.
8. Kumar GS, Krishna MNVLM. Low cost speech recognition system running on Raspberry Pi to support Automation applications. International Journal of Engineering Trends and Technology. 2015; 21(5).
9. Bhargava A, Nath KV, Sachdeva P, Samel M. Reading assistant for visually Impaired. International Journal of current Engineering and Technology. 2015 Apr; 5(2).
10. Gomes LCT, Nagle EJ, Chiquito JG. Text-to-speech conversion system for Brazilian Portuguese using a formant-based synthesis technique. LPS-DECOM-FEEC-Unicamp.
11. Sim Liew Fong, Abdelrahman Osman Elfaki, Md Gapar bin Md Johar & Kevin Loo Tow Aik, Mobile Language Translator, 5th Malaysian Conference in Software Engineering (Misses); 2011.
12. Kamesh DBK, Nazma SK, Sastry JKR, Venkateswarlu S. Camera based text to speech conversion, obstacle and currency detection for blind persons. Indian Journal of Science and Technology. 2016 Aug; 9(30).