INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3450 - 3455
RECEIVED : 25.03.2018 PUBLISHED : 25.04.2018
April 25, 2018

OPTICAL CHARACTER RECOGNITION BASED SPEECH SYNTHESIS SYSTEM USING RASPBERRY PI

R.Venkatesh, K.Vijaya Kumar , Mr.M.Sudhakaran,
UG Student, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore
Assistant professor, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore
Associate professor, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore

Abstract

In this project an innovative, efficient and real- time cost beneficial technique that enables user to hear the contents of text images instead of reading through them as been introduced. 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 is a device 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 v3 with 1.2GHz processor speed.

Keywords : Raspberry Pi; OCR; Camera; Image Processing; Voice Processing

1. INTRODUCTION

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 converts it into voice. 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 processorspeed. Optical character Recognition (OCR) is a process that converts scanned or printed text image, handwritten text into editable text for further processing. This paper has presented a robust approach for text extraction and its conversion into speech. Testing of device was done on raspberry pi platform.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3450 - 3455
RECEIVED : 25.03.2018 PUBLISHED : 25.04.2018
April 25, 2018

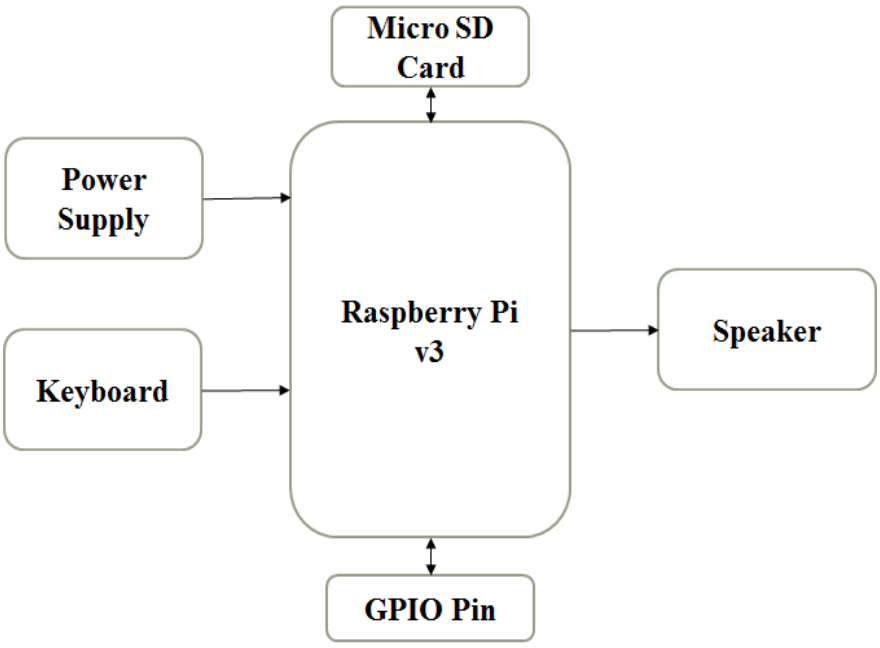

Fig.1.Block diagram

The following steps are implemented for character recognition. Firstly acquire the character image and the image was read. Second step is pre-processing step. In this step the color image is converted into gray scale, then this gray image is converted into binary image by performing the threshold operation. Character is extracted and resized in this step. Letters are resized according to templates size. Load templates that it can be matched the letters with the templates. Open the text.txt as file for write. Write in the text file and concatenate the letters. Feature extraction and classification are the heart of OCR. The character image is mapped to a higher level by extracting special characteristics and patterns of the image in the feature extraction phase. The classifier is then trained with the extracted features for classification task. The classification stage identifies each input character image by considering the detected features. As Classifiers, Template Matching and Neural Networks are used. The character image is converted into text and then text into speech. The algorithm is as follows. Firstly check the condition that if Win 32 SAPI (speech application programming interface) is available in the computer or not. If it is not available then error will be generated and Win 32 SAPI library should be loaded in the computer. Gets the voice object from Win 32SAPI. Compares the input string with Win 32 SAPI string. Extracts voice by firstly select the voice which are available in library. Choose the pace of voice. Initializes the wave player for convert the text into speech. Finally get the speech for given image. Automatic speech recognition (ASR) transforms speech into text. ASR is developed based on speech coding theory, at the same time as simulating certain spectral analysis performed by the ear. Speech synthesis is the automatic creation of a speech waveform, naturally from an input text.

2. SOFTWARE DETAILS

Open CV

Open CV is an open source computer vision library available from open CV library. The library is written in C and C++ and runs under Linux, Windows and Mac OS X. There is active development on

interfaces for Python, Ruby, Matlab, and other languages. Currently Open CV supports a wide variety of programming languages like C++, Python, Java etc and is available on different platforms including Windows, Linux, OS X, Android, IOS etc. Also, interfaces based on CUDA and OpenCL are also under active development for high-speed GPU operations. Open CV was designed for computational efficiency and with a strong focus on real- time applications. Open CV is written in optimized C and can take advantage of multicore processors. If you desire further automatic optimization on Intel architectures [Intel], you can buy Intel's Integrated Performance Primitives (IPP) libraries [IPP], which consist of low-level optimized routines in many different algorithmic areas. Open CV automatically uses the appropriate IPP library at runtime if that library is installed. One of Open CV's goals is to provide a simple-to-use computer vision infrastructure that helps people build fairly sophisticated vision applications quickly. The Open CV library contains over 500 functions that span many areas in vision, including factory product inspection, medical imaging, security, user interface, camera calibration, stereo vision, and robotics.

IDLE Python

IDLE (short for integrated development environment or integrated development and learning environment) is an integrated development environment for Python, which has been bundled with the default implementation of the language. It is packaged as an optional part of the Python packaging with many Linux distributions. It is completely written in Python and the Tkinter GUI toolkit (wrapper functions for Tcl/Tk). IDLE has two main window types, the Shell window and the Editor window. It is possible to have multiple editor windows simultaneously. Output windows, such as used for Edit / Find in Files, are a subtype of edit window. They currently have the same top menu as Editor windows but a different default title and context menu. IDLE's menus dynamically change based on which window is currently selected. Each menu documented below indicates which window type it is associated with. Check the syntax of the module currently open in the Editor window. If the module has not been saved IDLE will either prompt the user to save or autosave, as selected in the General tab of the Idle Settings dialog. If there is a syntax error, the approximate location is indicated in the Editor window.

3. HARDWARE DETAILS

- Ethernet Socket
- HDMI Socket
- USB 2.0 Socket
- RCA Video Socket
- SD Card Socket
- Powered from micro USB Socket
- 3.5mm Audio out Jack
- USB camera

4. REUSLTS AND DISCUSSIONS

A.SOFTWARE RESULTS

SIMULATION PROGRAM

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3450 - 3455
RECEIVED : 25.03.2018 PUBLISHED : 25.04.2018
April 25, 2018

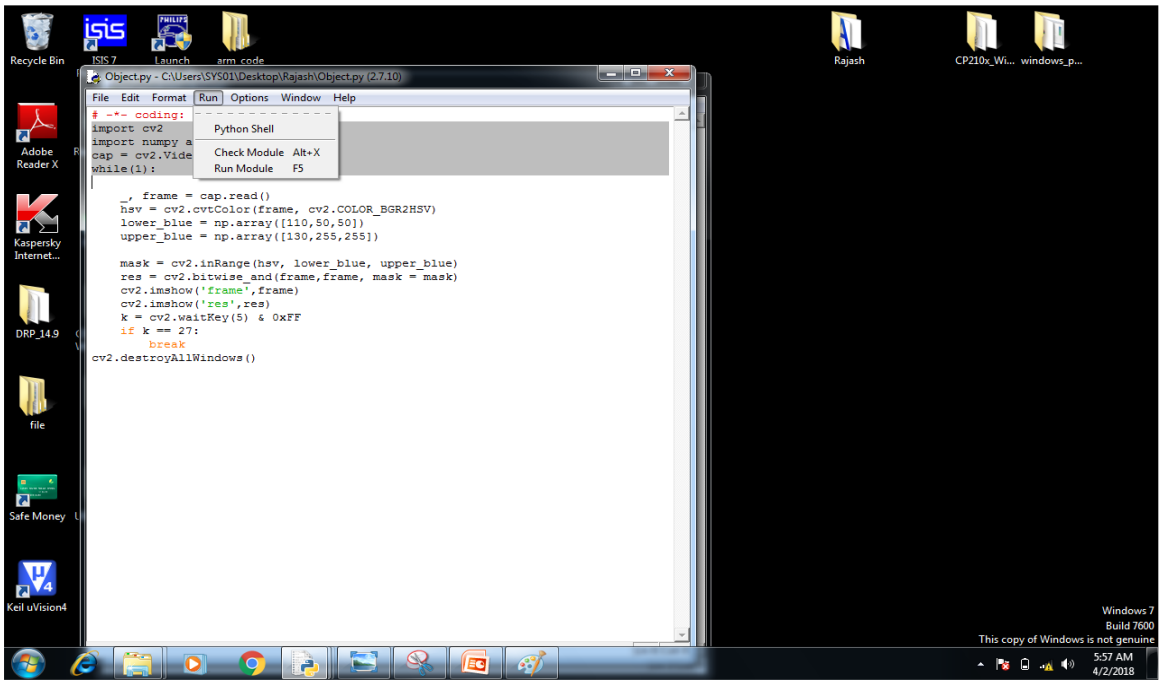

Fig.4.1. SIMULATION PROGRAM

SIMULATION OUTPUT

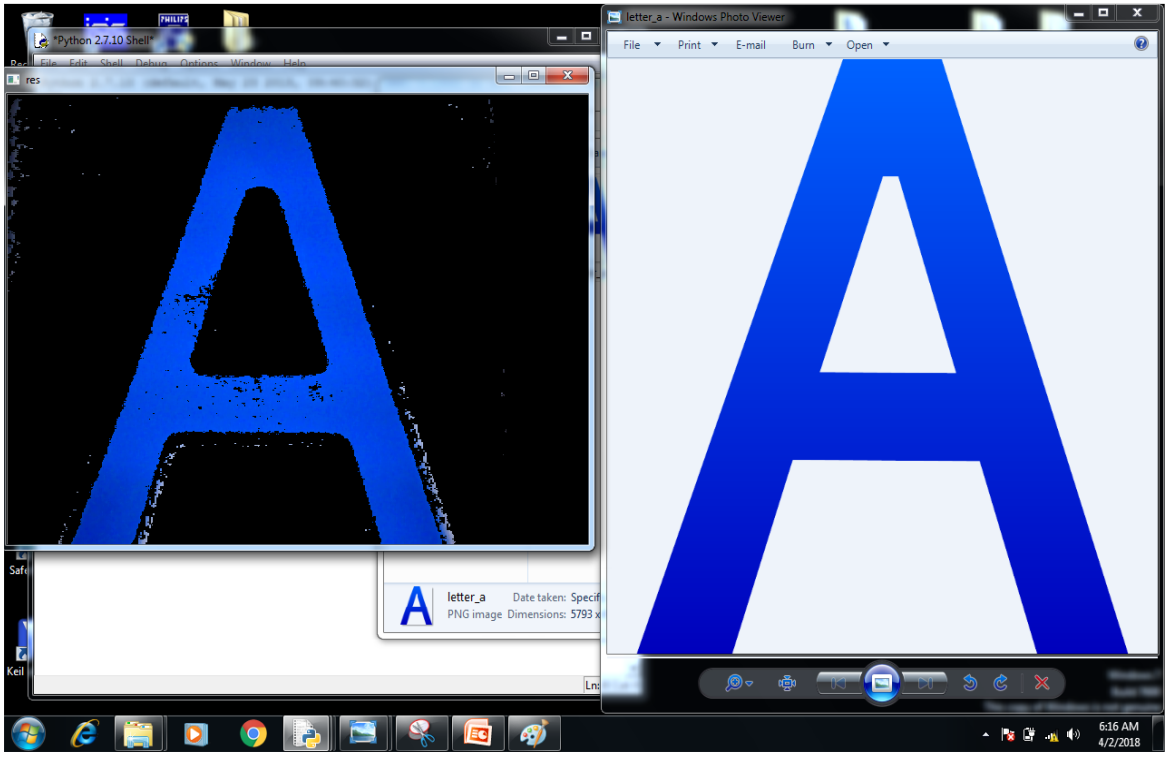

Fig.4.2. SIMULATION OUTPUT

 ©2018 R.Venkatesh | http://www.irjaet.com

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 4 ISSUE 2 (2018) PAGES 3450 - 3455
RECEIVED : 25.03.2018 PUBLISHED : 25.04.2018
April 25, 2018

B. HARDWARE RESULTS

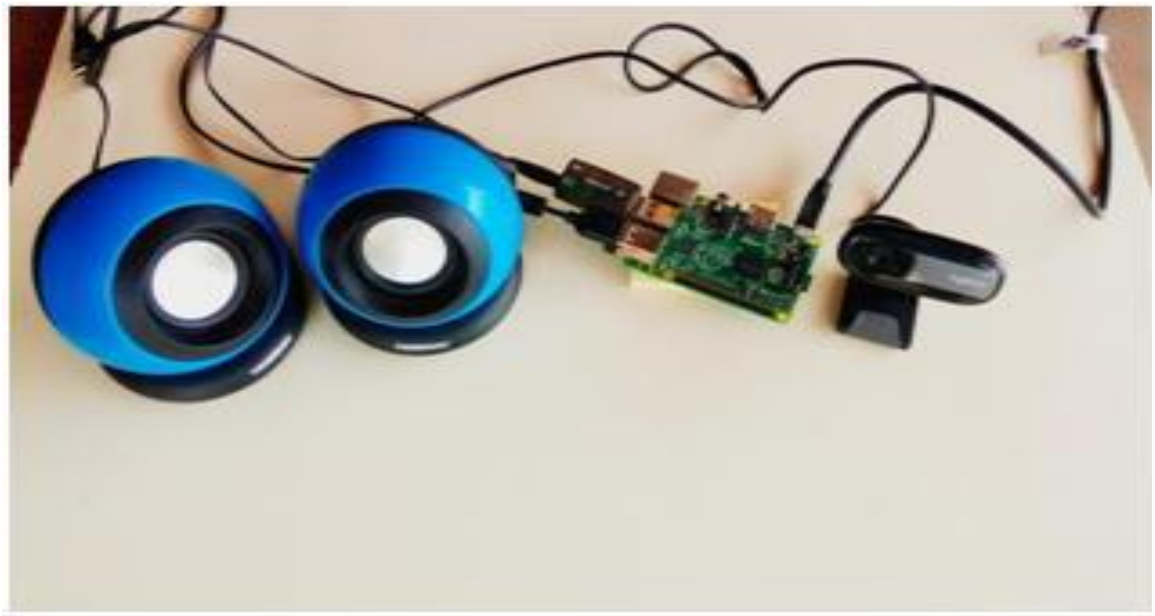

Fig.5.1. Hardware Kit

The above fig 5.1 Shows the experimental result of the project . Any number or character in English alphabet can be read by this pi. Here numbers are taken as input and analysed. A printed image of numbers is the input. The Logitech camera module is used to capture the image by a single command. The image is captured by the camera module and stored in a .jpg file format by a 15 pin ribbon cable. The image of text which was captured by the camera module. Text captured and conversion of text to flac file by espeak . The captured image is converted to .txt file. Image converted to text and stored in a file. The text file is then converted to .flac file which is given as input for translation. The proposed method captures 36 to 38 words and converts it into speech in duration of 10 to 20 seconds.

CONCLUSION

The text to speech system is implemented with the help of Raspberry Pi board. The results of simulation are verified successfully along with various Samples have been used to test the output. The input image is processed efficiently by algorithm used and clear output is generated. This is affordable and productive to blind people. The system is tested with different text and verified and it is useful for the society and compact. Future enhancements can be made in the image to text conversion for more accuracy and in the speech output to eliminate noise by using various algorithms and the application can be enhanced to capture images different languages and to output the voice in one or more languages. Thus conclude that this application is an attempt to help blind people in reading. With this application we can help the blind people.

REFERENCE

1. ARCHANA A.SHINDE, D. "Text Pre- processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology", pp. 810-812, 2012.
2. A.O.M.Asaduzzaman et al., "Printed bangla text recognition using artificial neural network with heuristic method," Proceedings of International Conference on Computer and Information Technology, 2002, pp. 27-28.

3. B.B.Chaudhuri and U.Pal, "A complete printed bangla OCR system," Pat. Rec., vol. 31, pp. 531-549, 1997.

4. B.M.Sagar et al., "OCR for printed Kannada text to machine editable format using database approach," WSEAS Trans. on Computers, vol. 7, pp. 766-769, 2008.

5. Endarpu Vanitha, Pradeep Kumar Kasarla, E Kuamarswamy, "Implementation of Text- to-Speech for Real Time Embedded System Using Raspberry Pi Processor", International Journal & Magazine of Engineering Technology Management and Research, IISN No:2348- 4845, Page 1995, July 2015.

51

6. H.Li et al., "Automatic Text Detection and Tracking in Digital Video," IEEE Trans. on Im. Proc., vol. 9(1), pp. 147-156, 2000.

. J.Yang et al., "An automatic sign recognition and translation system", Proceedings of the Workshop on Perceptive User Interfaces, 2001

8. K.Matsuo et al., "Extraction of character string from scene image by binarizing local target area," Trans. of The Ins. of Elec. Eng. of Japan, vol. 122-C(2), pp. 232– 241, 2002.

9. N.Ezaki et al., "Text detection from natural scene images: Towards a system for visually impaired persons", In Proceedings of the International Conference on Pattern Recognition, 2004, pp. 683-686

10. S.M.A.Haque et al., "Automatic detection and translation of bengali text on road sign for visually impaired," Daffodil Int. Univ. J. of Sc. and Tech,vol. 2, 2007.

11. Santosh kumar G, MNVLM Krishna, "Low cost speech recognition system running on Raspberry Pi to support Automation applications", International Journal of Engineering Trends and Technology, Volume 21 Number 5, March 2015.