ISSN 2395-1621

Text & sign interpretation for visually impaired person using raspberry pi

#1Amol Funde, #2Sushil Jagdale, #3Abhishek Nikam,
#4Prof. D.S. Gaikwad

1amolfundef@gamail.com
2jagdale.sushil24@gmail.com
3abhisheknikam10@gmail.com
4deepak.gaikwad@sinhgad.edu

#1234Department of Electronics and Telecommunication,

NBN Sinhgad School of Engineering, Ambegaon (BK), Pune,
Pune University, Maharashtra.

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 method is a camera based assistive text reading to help blind person in reading the text present on the text labels, printed notes and products. The proposed project involves Text Extraction from image and converting the Text to Speech converter, a process which makes blind persons to read the text. This is the first step in developing a prototype for blind people for recognizing the products in real world, where the text on product 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.

ARTICLE INFO

Article History

Received: 20th March 2017

Received in revised form : 20th March 2017

Accepted: 23rd March 2017

Published online : 24th March 2017

Keywords: Raspberry pi, Text extraction, text labels.

I. INTRODUCTION

Machine replication of human functions like reading is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Today, there are already a few systems that have some promise for portable use, like portable bar code readers designed to help blind people identify different products in an extensive product database can enable users who are blind to access information about these products through speech and Braille. But a big limitation is that it is very hard for blind users to find the position of the bar code and to correctly point the bar code reader at the bar code. Speech is probably the most efficient medium for communication between humans. To extract the text from image we use optical character recognition technique (OCR). Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Character recognition or optical character recognition (OCR) is the process of converting images of machine printed or handwritten text (numerals, letters, and symbols) into a computer format text. Speech synthesis is the artificial synthesis of human speech. A Text-To Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system.

II. OBJECTIVE

Our aim in this project is to build a prototype system for recognition of text present in the image using raspberry pi. As illustrated in the block diagram (Figure 1) the system framework consists of five functional components:

- Image acquisition
- Image pre-processing
- Text extraction
- Text to speech conversion and
- Speech output.

Existing systems for text recognition are typically limited either by explicitly relying on specific shapes or colour

masks or by requiring user assistance or may be of high cost. Therefore, we need a low-cost system that will be able to automatically locate and read the text aloud to visually impaired persons. The main idea of this project is to recognize the text character and convert it into speech signal.

III. LITERATURE SURVEY

Nagaraja L, Nagarjun R S , Nishanth M Anand have worked on "Human communication which is mainly via speech and text. The proposed method is a camera based assistive text reading to help blind person in reading the text present on the text labels, printed notes and products. The proposed project involves Text Extraction from image and converting the Text to Speech converter, a process which makes blind persons to read the text. This is the first step in developing a prototype for blind people for recognizing the products in real world, where the text on product 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."

Anusha Bhargava, Karthik V. Nath, Pritish Sachdeva and Monil Samel have worked on "A Majority of the visually impaired which use Braille for reading documents and books which are difficult to make and less readily available. This gives rise to the need for the development of devices that could bring relief to the agonizing tasks that the visually impaired has to go through. This project aims to study the image recognition technology with speech synthesis and to develop a cost effective, user friendly image to speech conversion system with help of Raspberry Pi. The project has a small inbuilt camera that scans the text printed on a paper, converts it to audio format using a synthesized voice for reading out the scanned text quickly translating books, documents and other materials for daily living, especially away from home or office. Not only does this save time and energy, but also makes life better for the visually impaired as it increases their independency."

K Nirmala Kumari, Meghana Reddy have worked on "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 Synthesiser (TTS) in Raspberry pi. This kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction from colour 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.

POONAM. S. SHETAKE, S.A. PATIL, P. M JADHAV have worked on"A text to speech converter which convert's normal language text into speech. Text to speech converter is useful in different applications. Customer support dialog systems Interactive voice response (IVR) systems etc and are also useful in an applied research. This application is more helpful in banking, toys and many other applications like checking marks, railways, aid to the physically challenged persons, language education and fundamental and applied research. etc. But text to speech conversion is not that much easy for machine as it is for human. Basic steps that machine has to follow for text to speech analysis are database creation, character recognition and text to speech conversion. This paper surveys methods related to character recognition as well as approaches used for text to speech conversion for machine.

IV. PROPOSED SYSTEM

The Raspberry pi is a single computer board with credit card size, that can be used for many tasks that your computer does, like games, word processing, spreadsheets and also to play HD video. The raspberry pi comes in two models, they are model A and model B. The main difference between model A and model B is USB port. The Raspberry Pi board comes equipped with an SD card. This slot permits us to insert an SD card and that can use it as our devices. Python is the Pi's recommended programming language, but Linux is its recommended operating system. Nearly every flavour of OS that works on Raspberry Pi—Raspbian. Optical character recognition, or OCR, is a method of converting a scanned image into text. When a page is scanned, it is typically stored as a bit-mapped file in TIF format. When the image is displayed on the screen, we can read it. But to the computer, it is just a series of black and white dots. The computer does not recognize any "words" on the image. This is what OCR does. OCR looks at each line of the image and attempts to determine if the black and white dots represent a particular letter or number. OCR was actually developed originally to assist sight-impaired individuals gain access to printed information. That same technology has been updated and improved and is now used to "read" computer files.

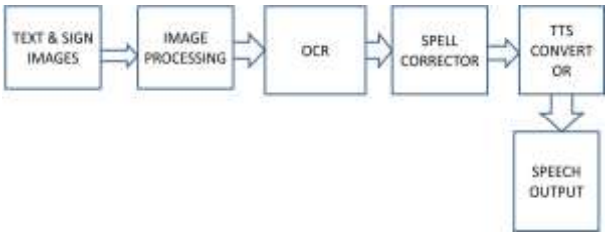

Figure 1. Basic Block Diagram

Image Acquisition

In this step the image of the text is captured using raspberry pi camera or an HD webcam with high

resolution. The acquired image is then applied to the image pre-processing step for reduction of unwanted noise.

Image Pre-processing

In image pre-processing the unwanted noise in the image is removed by applying appropriate threshold (OTSU), morphological transformations like dilation and black hat transformation, discrete cosine transformations generating the required contours and drawing the bounding boxes around the required text content in the image. Initially the captured image is rescaled to appropriate size and converted into grey scale image such that it will be more useful for further processing. Then the discrete cosine transformation is applied to the grey image to compress the image which helps to improve processing rate. Then by setting the vertical and horizontal ratio unwanted high frequency components present in the image are eliminated. Then the inverse discrete cosine transform is applied for decompression. Then image undergoes morphological operations like black top-hat transformation and dilations. The black top-hat transformation is applied to the image by generating appropriate structuring elements and extracts the objects or elements which are smaller than the defined structuring elements and darker than their surroundings. Then dilation operation is performed, which adds the pixels to the boundaries of the objects present in the image. The number of pixels added to the objects depends on the size and shape of the structuring element defined to process the image. After the morphological operations, thresholding is applied to the morphologically transformed image. Here the OTSU's thresholding algorithm is applied to the image, which is an adaptive thresholding algorithm. After thresholding, the contours for the image are generated using special functions in OpenCV. These contours are used to draw the bounding boxes for the objects or elements present in the image. Using these drawn bounding boxes each and every character present in the image is extracted which is then applied to the OCR engine to recognize the entire text present in the image.

Text Extraction

In this step the recognized text present in the image are extracted using OCR engines. In this project, we use tesser act OCR engine which helps to extract the recognized text.

Spell Corrector

The output of OCR is not 100% accurate, hence a spell correction method is proposed. In this a user, defined database is defined which is used for comparing the output of OCR so that the misspelled words are corrected. Comparison is done by using edit distances where either we insert or delete or substitute based on the requirement.

Text to Speech (TTS) System

In this step the extracted text is first converted into speech using the speech synthesizer called TTS engine which is capable of converting text to speech using predefined libraries. In this project the festival TTS engine is used for conversion of Text to speech.

V. HARDWARE AND SOFTWARE IMPLEMENTATION

Hardware implementation:

The main hardware components used in this project are Raspberry Pi 2, HD webcam, monitor, keyboard, mouse and ear phones. The interfacing diagram of these hardware components is shown below in Figure 2. Raspberry pi is a SoC (System on Chip), that integrates several functional components into a single chip or chipset. The SoC used in Raspberry Pi 2 is the Broadcom BCM2836 SoC Multimedia processor. The CPU of the Raspberry Pi contains an ARM Cortex-A7 900MHz processor which makes use of the RISC Architecture and low power draw. It is not compatible with traditional PC software. Hence it has to be connected to a monitor separately. Hence it is often called as a mini computer. Raspberry pi has a on-chip DSP processor which is used to perform the floating-point operations. The raspberry pi uses AMBA (Advanced Microcontroller Bus Architecture) which is an on-chip interconnect specification for the connection and management of functional blocks in system-on-chip (SoC) designs. It facilitates development of multi-processor designs with large numbers of controllers and peripherals. The GPIO pins of the Pi differ by the model. In model B there are 40 pins, out of which there are 4 power pins and 8 ground pins. Rest of the pins is used as GPIO's. The networking capabilities of the Pi can be used as a wired Ethernet (IEEE 802.3) or the wireless IEEE 802.11 Wi-Fi. Raspberry pi has an internal memory of 1GB RAM and external memory is extendable up to 64GB. HD webcam or raspberry pi camera which has a 5MP HD camera with a resolution of 1920x1200 can be used to capture the images. The speech output is given through the earphones connected to the raspberry pi's 3.5mm audio port.

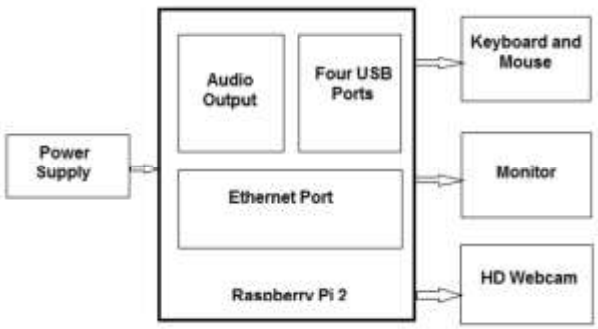

Figure 2

Software implementation:

Operating system: Raspbian (Debian)

Language: Python2.7

Platform: OpenCV (Linux-library)

Library: OCR engine, TTS engine

The operating system under which the proposed project is executed is Raspbian which is derived from the Debian operating system. The algorithms are written using the python language which is a script language. The functions in algorithm are called from the OpenCV library. OpenCV [4] is an open source computer vision library, which is written under C and C++ and runs under Linux, Windows and Mac OS X. OpenCV was designed for computational efficiency and with a strong focus on real-time applications. OpenCV is written in optimized C and can take advantage of multi-core processors.

VI. CONCLUSION

For the first stage project presentation, the required research work has been completed and the validation of project has been proved. The proposed system ensures to read text present in the image for assisting blind persons. Pre-processing part ensures efficient foreground extraction, which possess the required text region to be analysed. But the system fails to extract the foreground when they possess a complex background. An improved algorithm for background subtraction can reduce the effects of complex backgrounds. The extracted text is then given to a spell corrector as OCR output is not perfect. After getting the corrected output we send it to the TTS engine which provides a speech output. By providing a battery backup to the raspberry pi we can achieve the main aim of the proposed project of portability. The future work will be concentrated on developing an efficient portable product that can extract text from any image enabling the blind people to read text present on the products, banners, books etc.

REFRENCES

[1] Portable Camera Based Assistive Text and Product Label Reading From Hand-Held Objects For Blind Persons. Chucai Yi, Student Member IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi .

[2] Digital Image Processing, 3rd edition, Rafael C Gonzalez, Richard E Woods.

[3] “Text Pre-processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology”, ARCHANA A. SHINDE, D. pp. 810-812, 2012.

[4] ”Optical Character Recognition" International Journal of Recent Technology and Engineering (IJRTE)”, R. Mithe, S. Indalkar and N. Divekar. ISSN: 2277-3878,Volume- 2, Issue-1, March 2013.

[5] "High quality text-to-speech synthesis: a comparison of four candidate algorithms," Acoustics, Speech, and Signal Processing, T. Dutoit, 1994. ICASSP-94., 1994 IEEE International Conference on, vol.i, no., pp.I/565-I/568 vol.1, 19-22 Apr 1994.

[6] “OCR for printed Kannada text to machine editable format using database approach” WSEAS Transactions on Computers Volume 7, Pages 766-769, B.M. Sagar, Shobha G, R. P. Kumar, 6 June 2008.

[7] Implementing Optical Character Recognition on the Android Operating System for Business Cards Sonia Bhaskar, Nicholas Lavassar, Scott Green EE 368 Digital Image Processing.

[8] “Geometric Rectification of Camera-captured Document Images,” IEEE Transactions on Pattern Analysis and Machine Intelligence, J. Liang, et. al. pp. 591-605, July 2006.

