Vision based Text Recognition using Raspberry Pi

Nagaraja L
Dept. of ECE, BNMIT
12th Main, 27th Cross,
BSK 2nd Stage, Bangalore-70

Nagarjun R S
Dept. of ECE, BNMIT
12th Main, 27th Cross,
BSK 2nd Stage, Bangalore-70

Nishanth M Anand
Dept. of ECE, BNMIT
12th Main, 27th Cross,
BSK 2nd Stage, Bangalore-70

Nithin D
Dept. of ECE, BNMIT
12th Main, 27th Cross,
BSK 2nd Stage, Bangalore-70

Veena S Murthy
Associate Prof., Dept. of ECE,
BNMIT, 12th Main, 27th Cross,
BSK 2nd Stage, Bangalore-70

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 [1]. 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.

Keywords
Image Acquisition, Image Pre-processing, Black-hat transformation, Dilation operation, OTSU thresholding, Bounding boxes, Tesseract OCR engine, Spell Corrector, Festival TTS engine

1. 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 [1]. 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) [4] 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.

2. PROJECT OVERVIEW
This project presents 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 consist of five functional components: Image acquisition, Image pre-processing, Text extraction, Text to speech conversion and Speech output.

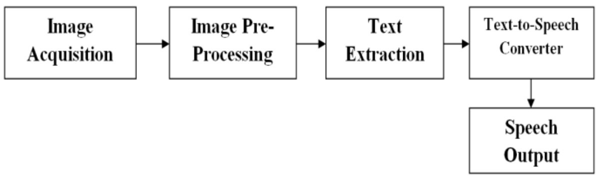

Figure 1 Basic block diagram

2.1 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

2.1.1 Image Pre-processing
In image pre-processing the unwanted noise in the image is removed by applying appropriate threshold (OTSU), morphological transformations [5] like dilation and black hat transformation, discrete cosine transformations [6], 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 gray scale image such that it will be more useful for further processing [3].Then the discrete cosine transformation is applied to the gray 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 [2] 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.

2.2 Text Extraction

In this step the recognized text present in the image are extracted using OCR engines. In this project we use tesseract OCR engine [7] which helps to extract the recognized text.

2.3 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. This works based on a probability model given by,

$\text{argmax}_c P(c|w)$ ……………………………..(2.1)

where P() is the probability c is the correct word and w is the word taken from the dictionary. Comparison is done by using edit distances where either we insert or delete or substitute based on the requirement.

2.4 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.

3. SOFTWARE SPECIFICATIONS

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. The OpenCV 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. Because computer vision and machine learning often go hand-in-hand, OpenCV also contains a full, general-purpose Machine Learning Library (MLL). To support OCR and TTS operations we need to install OCR [7] and TTS engines with predefined libraries.

4. HARDWARE SPECIFICATIONS

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.

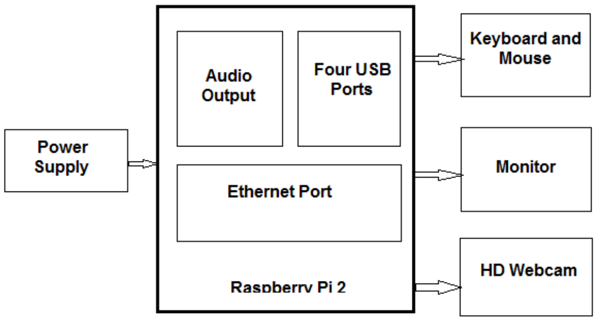

Figure 2 Interfacing of Raspberry Pi to hardware

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 upto 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.

5. RESULTS

The results obtained from the procedure described above are indicated in the figures below. Figure 3 indicates the image captured using the webcam, Figure 4 indicates the pre-processed image which is given to tesseract OCR engine to extract the text in the image. However due to the less resolution of the webcam, the output obtained is not 100% accurate. The accuracy can be improved by making use of a HD camera or mobile camera. Figure 5 shows the output of the tesseract OCR engine. Figure 6 indicates the output of the spell corrector to correct the misspelled words from the OCR engine.

Figure 3 Input image

Figure 4 Pre-processed image

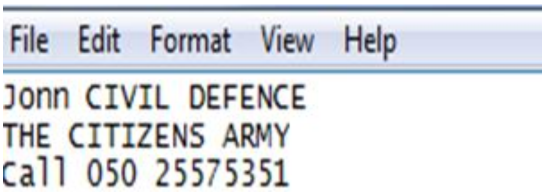

Figure 5 Extracted text from OCR engine

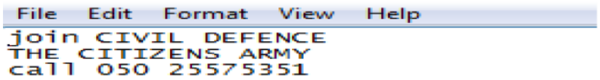

Figure 6 Spell Corrector Output

6. CONCLUSION AND FUTURE WORK

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 analyzed. 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

...R 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.

7. REFERENCES

[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] The watershed algorithm is a classic algorithm used for segmentation and is especially https://www.pyimagesearch.com/

[4] Vision Based Assistive System for Label Detection with Voice Output Vasanthi.G and Ramesh Babu.Y Department of ECE, DMI College of Engineering.

[5] Study And Implementation Of Various Morphology Based Image Contrast Enhancement Techniques. SumanThapar, Shevani Garg, CSE department, Ludhiana College of Engineering and Technology, Punjab.

[6] Image Compression Using the Discrete Cosine Transform. Andrew B. Watson, NASA Ames research Center.

[7] Short-term connection problems might affect your updatescode.google.com