Asian Journal of Applied Science and Technology (AJAST)
Volume 1, Issue 3, Pages 66-69, April 2017

Portable Camera Based Text, Product Label and Currency Note Reading from the Hand Held Objects for Blind Person

E.Pavithra[1] and M.Anantha Kumar[2]

[1]*UG Scholar, Department of Electronics and Communication Engineering, IFET College of Engineering, Villupuram, India.*
[2]*Assistant Professor, Department of Electronics and Communication Engineering, IFET College of Engineering, Villupuram, India.*

Article Received: 22 March 2017 *Article Accepted: 31 March 2017* *Article Published: 03 April 2017*

ABSTRACT

A portable camera based text, product label and currency note reading from the hand held objects for blind person using Raspberry Pi. Camera is used to capture the image of the product. Open CV is used for processing the captured image. The captured image is converted into the gray scale image by using RGB to grayscale conversion and canny edge detector is used to finds the edges of the image. Then four point transform is used to detect the contours in the edged image. OCR software is used to recognize the characters and Pytesseract library is used to extract the recognized text from the image. Finally by using espeak library, the obtained texts are converted as an audio output.

Keywords: Raspberry Pi, Open CV, OCR, Pytesseract and espeak.

1. INTRODUCTION

In worldwide there are 285 million people are estimated to be visually impaired. Nearly 39 million are blind and 246 have low vision. About 90% of the world's visually impaired live in low-income settings. 80% of people living with blindness are in the age of 50 and above. The census 2001 has revealed that over 21 million people in India as suffering from one or other kind of disability. The total disabled in the country, 12.5 million are males and 9.1 million are females. Although the number of disabled is more in rural and urban areas. There are five types of disabilities on which data has collected that disability in seeing at 48.7% emerges as the top category.

Reading is necessary in today's society. Printed text appears everywhere in the form of product label, restaurant menus, instructions on the medicine bottles, receipts, reports, bank statements, classroom notes etc. Blind people face a number of challenges when interacting with the environment because so much information is encoded visually in daily life. This project intended to assist the blind person to easily identify the currency note.

Fig. Examples of printed text from hand held objects with

2. EXISTING SYSTEM

Today there are number of devices for the blind person to read the text such as portable bar code readers, pen scanners, Laser scanners, CCD readers, Camera based readers, Video camera readers, K-Reader Mobile and camera based assistive devices. A barcode reader (or barcode scanner) is an electronic device that can read and output printed barcodes to a computer. It consist of a light source, a lens and a light sensor translating optical impulses into electrical ones. Though the barcode readers are helpful for the blind person to identify the products but it is very difficult for the blind person to correctly point the barcode reader at the place of bar code. Pen scanners is one of the assistive technology which is the most basic pens work just like a highlighter, and help with reading of text in books, newspaper and magazines. Some devices can download the scanned text into a computer or mobile device. OCR software is also integrated in the system. It is difficult for the software to work with the complex background. K reader mobile runs on a multifunction cell phone and allows users to snap a picture of virtually any document including mail, receipts, handouts, memos and many other documents. Our proprietary document analysis technology determines the words and reads them aloud to the users. But it reads only the black print on a white background and there is a problem to recognizing the color text or text with color background. A camera based assistive technology is used to read the text. For processing the captured image from the camera, a processing device (Laptop) must be required.

3. PROPOSED SYSTEM

In this system a portable camera based text, product label and currency note reading from the hand held object for the blind person using Raspberry Pi 3 model B. A switch (key) is connected to the raspberry pi board. When the key is pressed, the web camera which is connected to the Raspberry Pi capture the image of the product. Because when the blind person need to capture the image of the product, at the time the camera is in ON condition. Otherwise it will be in OFF condition. So that key input is used. Open CV is used for processing the captured image. The captured image is converted into the gray scale image by using RGB to grayscale conversion and canny edge detector is used to finds the edges of the image. Then four point transform is used to detect the contours in the edged image. Pytesseract library is used to extract the text from the image. OCR software is used

to recognize the characters and finally by using espeak library, the obtained texts are converted as an audio output. Head phone is used to receive the audio. Python software is used for coding purpose.

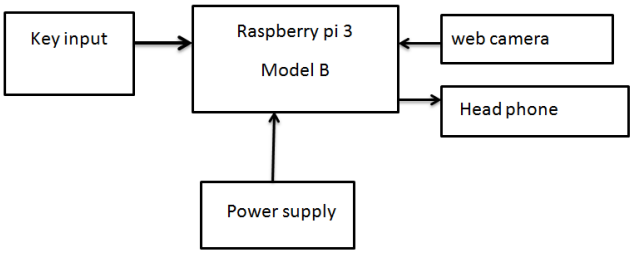


Fig.2. System block

4. BLOCK DIAGRAM DESCRIPTION

Raspberry pi is an ARM based credit card sized SBC (Single Board Computer). Raspberry pi runs Debian based GNU/Linux operating system Raspbian and ports of many other OSes exist for this SBC. It has onboard Wi-Fi/Bluetooth support and a 64 bit improved processor.

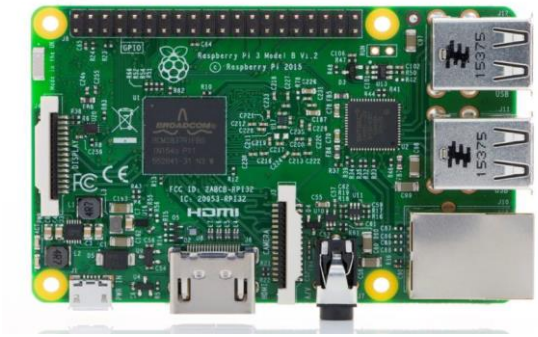


Fig.3. Raspberry Pi 3 model B

Ethernet port is a way to connect to the internet is via wire called Ethernet cable. Raspberry Pi 3 has four USB ports, allowing it to keyboards, mice, Wi-Fi dongles and USB sticks containing all our files. GPIO is to allow us to connect to the Raspberry Pi with an electronic circuit. Audio out looks like a headphone socket allow us to connect the Pi to computer's speakers. A little SD card is used as the Raspberry Pi's hard drive for memory purpose. Power is a small charging port. Use a standard HDMI cable to connect the Raspberry Pi to our chosen screen.

4.1 Basic Framework of the System

This paper presents a prototype framework for the blind person to read the text and product label from the hand held objects. The system consists of 3 functional components:

- Scene capture
- Raspberry Pi
- Audio output

Here web camera is used to capture the image of the product. The captured image has been resized and converts it into the gray scale image by using RGB to gray scale conversion; gray scale image is a range of shades of gray without apparent color. The darkest possible shade is black, the lightest possible shade is white, and from the gray scale image the edges are detected by using canny edge detector. Edge detection is used for finding the boundaries of objects within image.

Then find the contours of the edge detected image. Contours are the curve joining all the continuous points (along the boundary), having same color or intensity. Perspective transform is applied. Transformation is used to transfer the object from one state to another. Perspective transformation deals with the conversion of 3d world into 2d image. The text which is present in the image can be obtained by using pytesseract library function. For conversion of text into speech espeak library is used here. Finally the blind person hears the audio as an output.

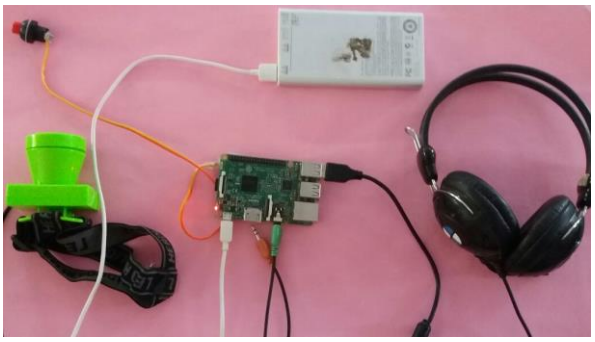


Fig.4. Snapshot demo system including three functional components for camera, raspberry pi and headphone

4.2 Process of Execution

Step 1: Edge detection
Step 2: Find contours of paper
Step 3: Apply perspective transform
Step 4: Image to text conversion
Step 5: Text to speech conversion

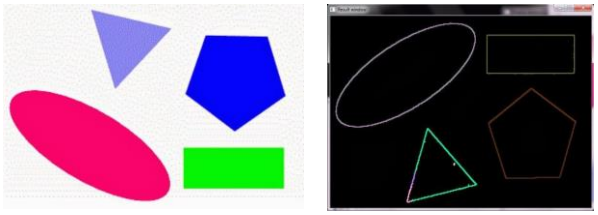
4.3 Edge Detection

To extract the text from the complex background, the edges of an image must be identified. Edge detection is used for finding the boundaries of the object within the image. It works by detecting discontinuities in brightness. Edge detection is used for image segmentation and data extraction in the areas of image processing.

Canny edge detector is used to detect the edges. The algorithm involves 5 steps. 1. Apply Gaussian filter to smooth the image to remove the noise 2. Find the intensity gradients of the image 3. Apply non maximum suppression to get rid 4. Apply double threshold to determine the potential edges 5. Track edge by hysteresis

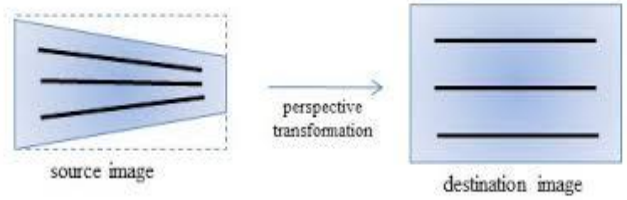
4.4 Find Contours

Contours are the curve joining all the continuous points (along the boundary), having same color or intensity. In case of shape analysis, detection and recognition of objects by using contours. In order to extract the text from the captured image, Contours must be detected. For getting better accuracy, we can use binary images. So before finding contours, we have to apply threshold or canny edge detection.



4.5 Perspective Transform

Perspective transform is used to transfer the object from one state to another. When the human eye sees near things they look bigger as compare to those who are far away. This is generally referred to as perspective. Perspective transformation deals with the conversion of 3d world into 2d image. Because the blind person cannot be properly fixes the product in front of the camera. The direction which the person holds the product can be changed. So in order to get the 2 dimensional images perspective transform is applied.



4.6 Image to Text Conversion

To extract the text from the image, Pytesseract (Python-tesseract) is an optical character recognition (OCR) tool for python. It will recognize and read the text embedded in image. Recognizing text from the image plays a major role here. Optical character recognition (also optical character reader, OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast).

4.7 Text to Speech Conversion

For converting the text into speech, espeak library function is used in open CV. Text to speech abbreviated as TTS is a form of speech synthesis that converts text into spoken voice output. Text to speech systems is useful for the blind person. It will provide a computer generated voice as an output. The output voice is used to recognize the product.

5. RESULT AND DISCUSSION

This project will help the blind person for self-reading purpose. The accuracy of text depends on the extraction of text. It is difficult for the OCR to separate the product from the complex background. Therefore by using four point transform it focus the image which has four point and extract the text from that image. It automatically clears the background even though which has some information. So the blind person can hear only the audio of the desired product.

6. PERFORMANCE MEASURES

The performance of the system can be measured by using precision, recall rate, F-score. True positives (TP) are those correctly identified characters, false positives (FP) are the incorrectly identified characters, true negatives (TN) are the correctly rejected characters and finally false negatives (FN) are the incorrectly rejected characters.

Precision is referred to as positive predictive value (PPV). Precision rate is defined as the ratio of correctly identified characters to the sum of correctly identified characters plus incorrectly identified characters as represented in equation below

$$P = [TP/(TP+FP)]*100$$

Recall rate (R) is defined as the ratio of the correctly identified characters to sum of correctly identified characters plus incorrectly rejected characters as represented in equation below

$$R = [TP/(TP+FN)]*100$$

F-Score is the harmonic mean of recall and precision rates. It is represented in equation below

$$F\text{-}Score = (2*P*R)/(P+R)$$

Accuracy (A) is defined as the ratio of the sum of correctly and incorrectly identified characters to the sum of true positive, true negative, false positive, false negative. It is represented as below

$$Accuracy = [(TP+TN)/(TP+TN+FP+FN)]*100$$

7. PERFORMANCE MEASURES OF THE PROPOSED SYSTEM

Images	Precision Rate (%)	Recall Rate (%)	F-Score (%)	Accuracy (%)
1	100	100	100	100
2	41.78	100	57.14	45.45

3	53.17	100	80	90.91
4	59.13	100	58.33	47.37
5	100	100	100	100

8. CONCLUSION

In this paper, a prototype system framework to read printed text on hand held object for blind person is described. This application will helpful for the blind person to read the text of their own and independent of others. Here open CV library functions such as (numpy, CV2, pytesseract, four-point transform, and imutils) are used for processing the captured images. OCR is used to perform text recognition on text localized regions and then recognized text codes are transformed to speech for blind person.

9. FUTURE WORK

In the future work, we can also use the smaller size camera for capturing the image of the product. It is so easy to handle for the blind person when it is fixed in the center of the glass. It will not produce any harmful effect to the blind person when it is in use.

.

REFERENCES

[1] KReader Mobile User Guide, knfb Reading Technology Inc. (2008). [Online]. *Available: http://www.knfbReading. Com*

[2] X. Chen and A. L. Yuille. "Detecting and reading text in natural scenes." Computer Vision and Pattern Recognition, *IEEE Computer Society Conference*, 2:366–373, 2004.

[3] C. Yi and Y. Tian, "Text string detection from natural scenes by structure based partition and grouping," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp.2594–2605, Sep. 2011.

[4] L. Ma, C. Wang, and B. Xiao, "Text detection in natural images based onmulti-scale edge detection and classification," *in Proc. Int. Congr. Image Signal Process.*, 2010, vol. 4, pp. 1961–1965.

[5] Nanayakkara, S., Shilkrot, R., Yeo, K. P., and Maes, P., "EyeRing: a finger-worn input device for seamless interactions with our surroundings". *In Augmented Human*, 2013.

[6] Carlos Merino-Gracia,, Karel Lenc, Majid Mirmehdi, "A Head- Mounted Device for Recognizing Text in Natural Scenes", *Lecture Notes in Computer Science*, Volume 7139, pp 29-41, 2012.

[7] Chengjun Zhu, yuanxin Quyang, Lei Gao, Zhenyong Chen, Zhang Xiong, "An Automatic Video Text Detection, Localization and Extraction Approach". *School of computer Science and Technology, Beihang University, SITIS* - Pages 166-175-2006.

[8] Christin Wolf, Jean-Michel Jolion, and Francoise Chassaing. "Text localization, enhancement and binarization in multimedia documents". *In Int. Conf. on Pattern recognition, pages 1037–1040, Quebec City, Canada.*, August 2002.

[9] Rajkumar N, Anand M.G, Barathiraja N, "Portable Camera-Based Product Label Reading For Blind People", *International Journal of Engineering Trends and Technology (IJETT)* – Volume 10, Issue No: 1 - Apr 2014.

[10] Ashwini V.Mhaske, Priti V.Chaudhari, Panchashila G.Sabale, Amit B.Sanghai, Pravin B.Patil, Rashmi S.Dshmukh, "Portable Camera Based Assistive Product Label Reading For Blind And Visually Impaired Individuals", *International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering*- Volume 4, Issue No:6, June 2016.