International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 12995-12999
© Research India Publications. http://www.ripublication.com

Portable Camera Based Text Reading of Objects for Blind Persons

Sonal I Shirke [1], Swati V Patil [1]

[1] *PCET's PCCOE, Pune, India.*

Abstract

This paper proposes a portable camera-based assistive text reading framework to help blind persons to read text labels and product packaging from hand-held objects in their daily lives. The system framework consists of three functional components: First, scene capture-using a mini camera, the text which the user needs to read gets captured as an image and has to be sent to the image or data processing platform., second, data processing -where text will be filtered from the surrounding and will be recognized by optical character recognition (OCR) software, and finally, Speech output - A filtered text will be passed into this system to get an audio output. In addition to text recognition using OCR, the project also includes template matching technique as a separate method for recognizing certain objects like currency notes. Since the project should be portable to assist the blind persons, the entire application is based on Raspberry Pi.

Keywords: Portable, image processing platform, optical character recognition (OCR), template matching, Raspberry Pi

INTRODUCTION

There are over billions of visually challenged people worldwide. A major part of this population is still blind even in developed countries like United States. The national health interview survey conducted in 2008 reported that over 85 of the adult Americans lack the ability to see. In recent times development in computer vision, digital cameras and portable computers help to aid these individuals by developing camera based products that integrate computer vision technology with already existing products such as optical character recognition (OCR).

There are few devices that can provide better access to common hand-held objects much as product packages and objects printed with text. Formulating devices which are even more portable and sophisticated can promote independent living and foster economic and social self-dependency.

The most challenging part in assistive reading system for blind people is- positioning of object of interest within the camera view. In order to focus the object within the camera_view, a camera with wide angle is used as an approximate solution. Often text from the surrounding areas is also included. Thus to extract the hand-held objects from the image we proposed motion based method to isolate the region of interest and text recognition is done only on the area of interest.

RELATED WORK

The paper- Text detection in indoor/outdoor scene images- proposes a novel methodology for text detection in indoor/outdoor scene images. It is based on an efficient binarization and enhancement technique followed by a suitable connected component analysis procedure. Connected component analysis is used to define the final binary images that mainly consist of text regions.

Portable Camera-Based Assistive Text and Product Label Reading from hand-held objects for blind persons- propose an efficient and effective motion based method to define a region of interest in the video by asking the user to shake the object for isolation of objects from the cluttered background. This method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. In the extracted ROI, text localization and recognition are conducted to acquire text information. To automatically localize the text regions from the object ROI, we propose a novel text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model. Text characters in the localized text regions are then binarized and recognized by off-the-shelf optical character recognition software. The recognized text codes are output to blind users in speech.

Compact Camera Based Assistive Text Product Label Reading and Image Identification for Hand-Held Objects for Visually Challenged People- propose camera-based assistive text reading framework and to process the captured object and obtain its details to help blind persons read text labels and product packaging from hand-held objects in their daily lives. It is based on Raspberry Pi.

PRELIMINARIES

(i) Optical Character Recognition (OCR)

Optical character recognition (OCR) is the recognition of printed or written text characters by a computer. This involves photo scanning of the text character-by-character, analysis of the scanned-in image, and then translation of the character image into character codes, such as ASCII, commonly used in data processing.

When we read from the computer screen our eyes and brain are carrying out optical character recognition without us noticing. Our eyes recognize the patterns of light and dark that make up the characters (letters, numbers, and things like punctuation marks) printed on the screen and our brain uses those to figure out what the opposite person is trying to say. Computers can do this too, but it's really hard work for them. The first problem is that a computer has no eyes, so if we want it to read something like the page of an old book, we have to present it with an image of that page, generated with an optical scanner or a digital camera. The page we create this way is a graphic file (often in the form of a JPG) and, as far as a computer is concerned, there is no difference between it and a photograph of any monument or any other graphic: it's a completely meaningless pattern of pixels (the coloured dots or squares that make up any computer graphic image. In other words, the computer has a *picture* of the page rather than the text itself—it can't read the words on the page like we can, just like that. OCR is the process of turning a picture of text into text itself—in other words, producing something like a TXT or DOC file from a scanned JPG of a printed or handwritten page.

(ii) Raspberry Pi

The Raspberry Pi is a low cost, credit-card sized computer that plugs into a computer monitor or TV, and uses a standard keyboard and mouse. The Raspberry Pi is slower than a modern laptop or desktop but is still a complete Linux computer and can provide all the expected abilities that implies, at a low-power consumption level. System on a Chip(SoC), is a method of placing all necessary electronics for running a computer on a single chip. Instead of having an individual chip for the CPU, GPU, USB controller, RAM, everything is compressed down into one tiny and tidy package. The Raspberry Pi was designed for the Linux operating system, and many Linux distributions now have a version optimized for the Raspberry Pi. As far as the specifications, the Raspberry Pi is a credit card-sized computer powered by the Broadcom BCM2835 system-on-a-chip (SoC). This SoC includes a 32-bit ARM1176JZFS processor, clocked at 700MHz, and a Videocore IV GPU. It also has 256MB of RAM in a POP package above the SoC. The Raspberry Pi is powered by a 5V micro USB AC charger. A Raspberry Pi features HDMI and composite video outputs, two USB 2.0 ports, a 10/100 Ethernet port, SD card slot, GPIO (General Purpose I/O Expansion Board) connector, and analog audio output (3.5mm headphone jack). For cost reduction, the Raspberry Pi omits any on-board non-volatile memory used to store the boot loaders, Linux Kernels and file systems as seen in more traditional embedded systems. Rather, a SD/MMC card slot is provided for this purpose. After boot load, as per the application program Raspberry Pi will get executed.

METHODOLOGY

Figure 1 gives the block diagram of the steps to be followed for text reading using OCR and further give audio output.

Image Acquisition:

The Image acquisition component collects scenes containing objects of interest in the form of images. Here, generally available and low cost webcam is used for image acquisition.

Pre-processing and Gray Scale Conversion:

To make the system more robust i.e. work for noisy conditions, image pre-processing methods like noise filtering are applied. For the purpose of reducing the processing time of the overall process, the input is converted into Gray Scale.

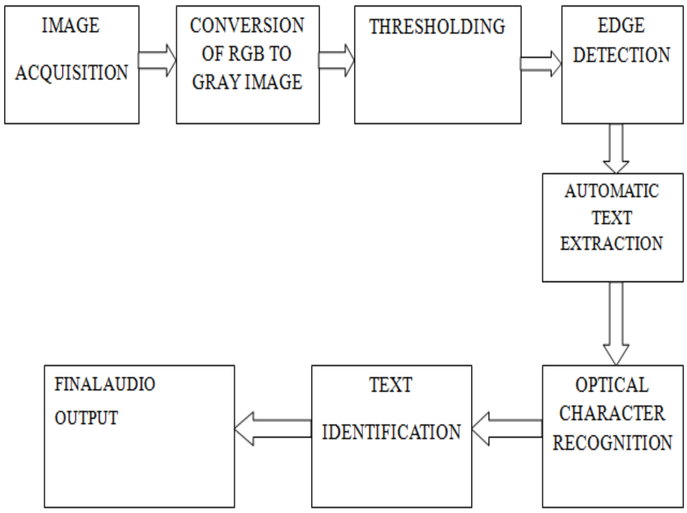

Figure 1.Block diagram of project using OCR

Pre-processing of document images is the way of using mature image processing techniques to improve the quality of images. Its purpose is to enhance and extract useful information of images for later processing purposes. Two pre-processing tasks, thresholding and noise removal, are performed here.

Edge detection:

Edge detection is a set of mathematical method which aims at identifying point in an image at which image brightness changes sharply or has discontinuities. Such points are typically organized into a set of curved line segments called as edges. The Canny Edge Detection Algorithm is used in edge detection.

The algorithm runs in various separate steps:

- Smoothing
- Finding gradients
- Edge tracking by hysteresis (Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge).

Thresholding:

It is a simplest method of image segmentation from a gray scale image. Thresholding can be used to create binary image. The method used in thresholding is Otsu's method. Otsu's method selects the threshold by minimizing the within-class.

Automatic Text Extraction:

Then, an automatic text extraction algorithm is implemented to detect the region containing the label text. In order to handle complex backgrounds, two novel feature maps to extracts text features based on stroke orientations and edge distributions, respectively are used. Maximally stable external region is used in automatic text extraction. MSER can be used to define image regions with their outer boundary according to the intensity of a scene street image.

Optical Character Recognition:

Text recognition is performed by off-the-shelf OCR prior to output of informative words from the localized text regions. A text region labels the minimum rectangular area for the accommodation of characters inside it, so the border of the text region contacts the edge boundary of the text character. However, OCR generates better performance if text regions are first assigned proper margin areas and binarized to segment text characters from background. We propose to use Template matching algorithm for OCR. The output of the OCR is nothing but a text file containing the product label (its name) in textual form. Audio output component is to inform the blind user of recognize text code in the form of speech or Audio.

Text Identification and Audio Output:

After extracting the text from the background, the text of the product is recognized and is converted into the audio form.

Here we are going to capture the image of a product or any kind of image which has a text on it through a mini camera mounted on spectacles. After acquisition of image various operations are needed to be performed in order to extract text from the image.

Firstly pre-processing steps are performed in which filtering of the image, removal of noise is done. After filtering we convert the image into gray scale image and thresholding operations are performed. The purpose of these operations is to enhance and extract useful information of images for later processing purposes. Next task is to separate the text from the background in which we are going compare the binary and inverted binary image Now the text frame is separated (i.e characters are separated) by using segmentation. By using OCR (object character recognition) characters are analyzed and these characters are compared with the database created. Now when these characters are recognized, these characters are finally converted into the audio output.

RESULTS

(i) Results using OCR

In order to be easy to perform morphological operations on the image, it is better to convert the original coloured image into gray image so that complexity reduces. Figure 2 shows the conversion results.

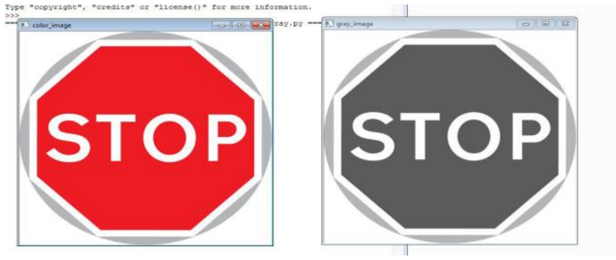

Figure 2. RGB to gray conversion

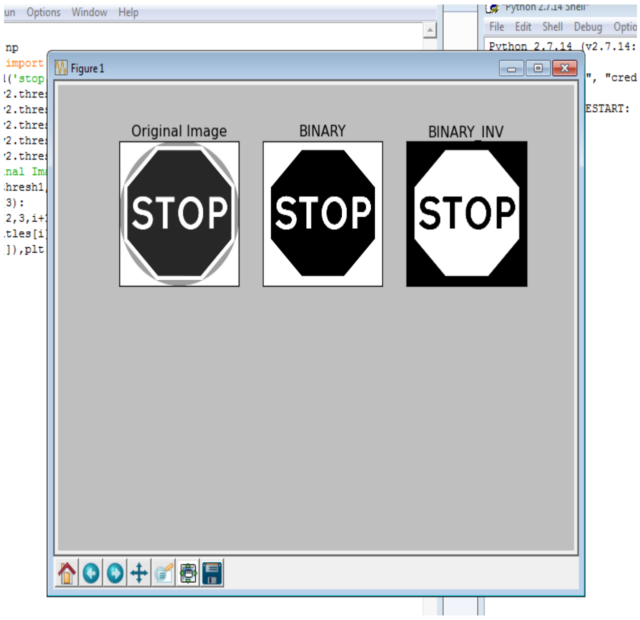

Figure 3. Conversion from binary to inverted binary

As seen in figure 3, conversion into inverted binary is done.

Due to this, the background of the image is suppressed and we get a clear text for further operations.

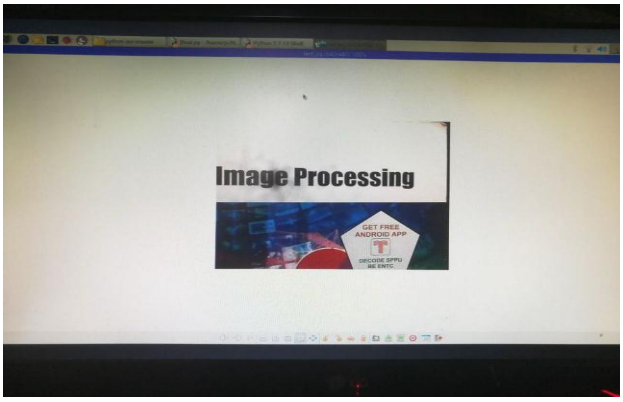

Figure 4. Test image for Optical character recognition

Figure 4 shows the test image of the OCR method. This is the image captured by the camera. Figure 5 shows the final text recognised output which is highlighted on the screen.

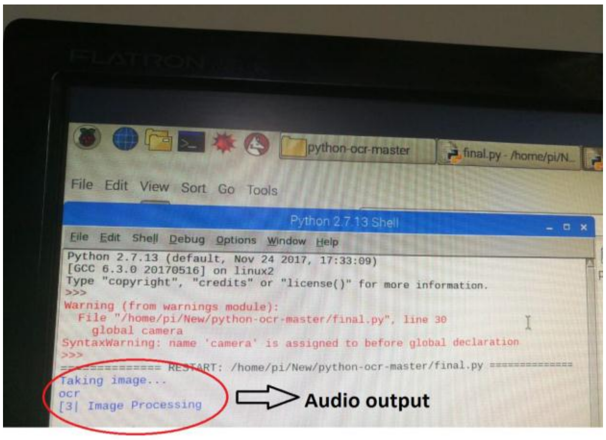

Figure 5. OCR output

Hereafter, processing of text is done and it is converted into audio output. If earphones are connected, we can hear it.

(ii) Results using template matching method

Figure 6 shows the template images stored in the database for template matching.

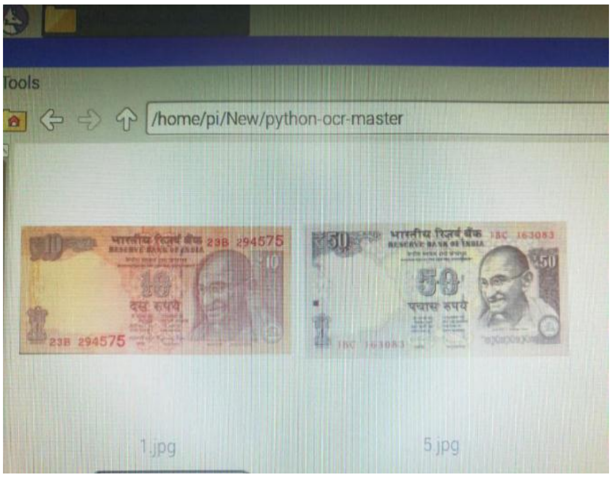

Figure 6. Template images in database

Figure 7 shows the test image captured by the camera.

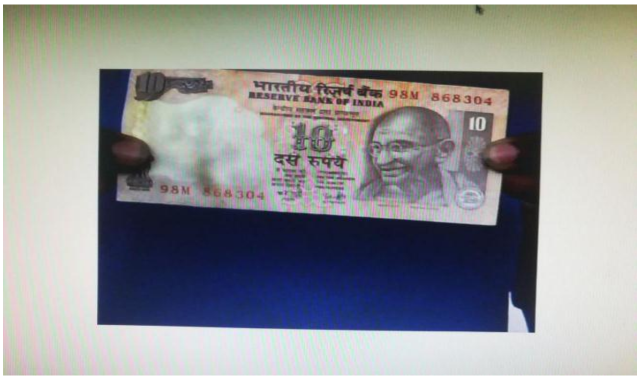

Figure 7 Test image for template matching method

Here captured test image of the product is compared with template images in the database and at the end we get the output as the product name as shown highlighted on the screen.

Figure 8 Output of template matching method

Hereafter the processing of text is done and the output is converted into audio output which can be heard through earphones.

CONCLUSION

Thus, this paper proposes a camera-based assistive text reading framework to help blind persons to read text labels and product packaging from hand-held objects in their daily lives. Results are successfully obtained for all texts having a font size of around 1 inch and more on non-complex backgrounds using OCR method. Successful results are also obtained using template matching. This method is more suitable for currency notes. Currency notes are frequently used by everybody including blind persons and its recognition accuracy ought to be 100%. In future, we will use more effective algorithms so that our system can identify text which are small in size and also the texts which have various complex backgrounds.

REFERENCES

[1] Ms Komal Mohan Kalbhor, ‚Kale S.D.," A Survey on Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 03| Mar -2017

[2] R. Infant Abinaya, E.Esakkiammal, Mrs.Pushpalatha," Compact Camera Based Assistive Text Product Label Reading and Image Identification for Hand-Held Objects for Visually Challenged People", International Journal of Computer Science and Information Technology Research ,Vol. 3, Issue 1, pp: (87-92), Month: January - March 2015

[3] Chucai Yi, , Yingli Tian, Aries Arditi," Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons", IEEE/ASME Transactions on Mechatronics,2013

[4] C. Yi, Y. Tian, "Assistive text reading from complex background for blind persons", Proc. Int. Workshop Camera-Based Document Anal. Recognit., vol. LNCS-7139, pp. 15-28, 2011.

[5] B. Epshtein, E. Ofek, Y. Wexler, "Detecting text in natural scenes with stroke width transform", Proceedings of Computer. Vision Pattern Recognition., pp. 2963-2970, 2010

[6] K. Kim, K. Jung, J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm", IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631-1639, Dec. 2008.

[7] L. Ma, C. Wang, B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification", Proc. Int. Congr. Image Signal Process., vol. 4, pp. 1961-1965, 2006.

[8] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proceedings of Computer. Vision Pattern Recognition., 2004, vol. 2, pp. II-366–II-373.

[9] M. Shi, Y. Fujisawab, T. Wakabayashia, F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image", Pattern Recognit., vol. 35, no. 10, pp. 2051-2059, 2002.

[10] J. Zhang, R. Kasturi, "Extraction of text objects in video documents: recent progress", Proc. IAPR Workshop Document Anal. Syst., pp. 5-17, 2008.

[11] B.Gatos, I.Pratikakis, K.Kepene, S.J.Perantonis,"Text detection in indoor/outdoor scene images"