IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 6, Dec-Jan, 2014
**ISSN: 2320 - 8791**
**www.ijreat.org**

# Assistive System For Product Label Detection With Voice Output For Blind Users

**T.Rubesh Kumar[1], C.Purnima[2]**

[1,2]Department of ECE,
Prathyusha Institute of Technology and Management, Tamil Nadu, India.

**ABSTRACT**

We propose a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. We first propose an efficient and effective motion based method to define a region of interest (ROI) in the video by asking the user to shake the object. 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 Ada boost model. Text characters in the localized text regions are then binarized and recognized by off-the-shelf optical character recognition (OCR) software. The recognized text codes are output to blind users in speech.

***Keywords***-*Assistive text, blind persons, distribution of edge pixels, hand-held objects, optical character recognition (OCR), stroke orientation, text localization.*

## 1. INTRODUCTION

Of the 314 million visually impaired people worldwide, 45 million are blind. Reading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages, instructions on medicine bottles, etc. Today, there are already a few systems that have some promise for portable use, but they cannot handle product labeling. For example, 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.

Our main contributions embodied in this prototype system are: 1) a novel motion-based algorithm to solve the aiming problem for blind users by their simply shaking the object of interest for a brief period; 2) a novel algorithm of automatic text localization to extract text regions from complex background and multiple text patterns; and 3) a portable camera-based assistive framework to aid blind persons reading text from hand-held objects.

## 2. SOFTWARE AND HARDWARE SPECIFICATIONS

### 2.1 Software specifications

| | |
|---|---|
| Operating System | :Ubuntu 12.04 |
| Language | :C,C++ |
| Platform | :OpenCV (linux-library) |

### 2.2 Hardware specifications

The Raspberry Pi is a credit-card-sized single-board computer developed in the UK by the Raspberry Pi Foundation. It has a Broadcom BCM2835 system on a chip (SoC), which includes an ARM1176JZF-S 700 MHz processor.And primarily uses Linux kernel-based operating systems. Memory (SDRAM) of 256 MB shared with GPU which can be upgraded to 512MB.

Fig.1. Raspberry pi is the simple hardware system which ensures portability of the assistive reading system

## 3.IMAGE CAPTURING AND PRE-PROCESSING

The live video is captured by using web cam and it can be done using OPENCV libraries.The image format from the webcam is in RGB24 format. The frames from the video is segregated and undergone to the pre processing. The capturing videos are projected in a window with a size of 320x240. Totally 10 frames per second can be captured by using the webcam. First, we get the frames continuously from the camera and send it to the process. Once the object of interest is extracted from the camera image cascade classifier is used for recognizing the character from the object , the system is ready to apply our automatic text extraction algorithm

## 4.TEXT REGION LOCALISATION

Text localization is then performed on the camera-based image. The Cascade Ada boost classifier confirms the existence of text information

### 4.1 Cascade classifier

The work with a cascade classifier includes two major stages: training and detection. Detection stage is described in a documentation of object detect module of general OpenCV documentation. Documentation gives some basic information about cascade classifier.

### 4.2 Training data preparation

For training we need a set of samples. There are two types of samples: negative and positive. Negative samples correspond to non-object images. Positive samples correspond to images with detected objects. Set of negative samples must be prepared manually, whereas set of positive samples is created using opencv_create samples utility.

## 5.CHARACTER DETECTION

We use OpenCV (open source computer vision) library to process the images so that features for each letter could be extracted.

### 5.1 Image capturing

First, we get the frames continuously from the camera and send it to the process. Once the object of interest is extracted from the camera image using cascade classifier, subsequent process can be done using following steps.

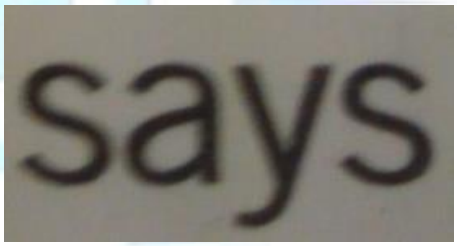


Fig.2. Original captured color image

### 5.2 Conversion to gray scale

Conversion to gray-scale can be done in OpenCV. The reason we had to convert our image to gray-scale was because thresholding could be applied to monochrome pictures only.In an 8 bit image each pixel is represented by one number from 0 to 255 where 0 is black and 255 is white. The simplest way to convert the image to black and white pixels would be to select one value, lets say 128 and consider all pixels that have higher value to be white and the others black. The biggest problem with this approach is that the brightness can vary from picture to picture and as a result some images might become totally black while others are entirely white. The function *MinMaxLoc* finds minimum and maximum element values and their positions. The extremums are searched over the whole array, selected *ROI* (in case of *IplImage*) or, if *mask* is not *NULL*, in the specified

array region. In case if multi-dimensional arrays *minLoc->x* and *maxLoc->x* will contain raw (linear) positionsofextremums. .

After, getting the min and max value convert Scaling function. The cvConvertScale has several different purposes and thus has several synonyms. It copies one array to another with optional scaling, which is performed first, and/or optional type conversion, performed and get the binary output.


Fig.3. Binary output image

## 6.TEXT RECOGNITION AND AUDIO OUTPUT

Text recognition is performed by OCR to output of informative words from the localized text regions. The recognized text codes are recorded in script files. Then, we employ the e-speak engine to load these files and display the audio output of text information. Blind users can hear the voice according to their preferences.

### 6.1 Tesseract OCR

Tesseract is a free optical character recognition engine, originally developed as proprietary software by Hewlett-Packard We used a very naïve approach to separate individual letters from the words. First we scanned the vertical columns of pixels from left to right and stored the numbers of the columns in which all pixels where white. Next we looked at the vector of column numbers and determined where they are interrupted. These points became the cutting points.


Fig.4. Cutting out the letters

After doing the vertical cuts we apply the same algorithm to pixel rows to remove the white areas from above and below of each letter separately. Now that we have each letter cut out we need to convert them to feature vectors that can be used as an input to classification algorithm. First thing that we noticed was that it is very easy to find the aspect ratio for each character and we decided to use it as the first feature. Another obvious thing to look at are the pixels. The simplest approach would be to use the values of all pixels as the features

The solution that we came up with was to resize the picture corresponding to each character to 16 x 8 pixels. These numbers are actually quite arbitrary and we chose to use them mainly because the same dimensions are used in a publicly available OCR data set. Using 18 and 9 or anything reasonably similar would probably not change the results too much. In OpenCV the resizing is done using cvResize(), which in its default behaviour uses bi-linear interpolation. It would also be possible to use nearest neighbour interpolation, which would result in only black or white pixels. To get the feature vectors we used the cvSave() function to save the picture as an XML file.

Fig.5. Letter a after resizing

## 7.CONCLUSION

In this paper, we have described a prototype system to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming problem for blind users, we have proposed a motion-based method to detect the object of interest, while the blind user simply shakes the object for a couple of seconds. This method can

effectively distinguish the object of interest from background or other objects in the camera view. To extract text regions from complex backgrounds, we have proposed a novel text localization algorithm based on models of stroke orientation and edge distributions. The corresponding feature maps estimate the global structural feature of text at every pixel. An Adaboost learning model is employed to localize text in camera-based images. Off-the-shelf OCR is used to perform word recognition on the localized text regions and transform into audio output for blind users.Further work would include bank-note identification for blind users.

## ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their constructive comments and insightful suggestions that improved the quality of this manuscript.

## REFERENCES

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

[2] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," *IEEE Trans. Image Process.*,vol. 13, no. 1, pp. 87–99, Jan. 2004..

[3] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the Nav-belt: A computerized travel for the blind," *IEEE Trans. Syst., Man, Cybern. C. Appl. Rev.*, vol. 28, no. 3, pp. 459–467, Aug. 1998.

[4] J. Zhang and R. Kasturi, "Extraction of Text Objects in Video Documents: Recent Progress," In *IAPR Workshop on Document Analysis Systems*, 2008.

[5] N. Nikolaou and N. Papamarkos, "Color Reduction for Complex Document Images," *International Journal of Imaging Systems and*
*Technology*, Vol.19, pp.14-26, 2009.

[6] N. Otsu, "A threshold selection method from gray-level histograms,"In *IEEE Tran.s on system, man and cybernetics*, pp. 62-66, 1979.

[7] L. Ma, C. Wang, B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification," In *the Int. Congress on*
*Image and Signal Processing (CISP)*, 2010.

[8]B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," In *CVPR*, pp. 2963-2970, 2010.

[9] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text Extraction and Document Image Segmentation Using Matched
Wavelets and MRF Model," *IEEE Trans on Image Processing*, Vol. 16, No. 8, pp. 2117-2128, 2007.

[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," presented at the IEEE Comput. Soc. Conf. Comput.Vision Pattern Recognit., Fort Collins, CO, USA, 1999.

[11] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," In *CVPR*, Vol. 2, pp. II-366 – II-373, 2004.

[12] X. Chen, J. Yang, J. Zhang and A. Waibel, "Automatic detection and recognition of signs from natural scenes," In *IEEE Transactions on*
*image processing*, Vol. 13, No. 1, pp. 87-99, 2004.

[13]D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," In *IEEE Transactions on*
*systems, man, and cybernetics*, Vol. 40, No. 1, pp. 25-35, 2010

[14]World Health Organization. (2009). 10 facts about blindness and visual impairment [Online]. Available: www.who.int/features/factfiles/
blindness/blindness_facts/en/index.html

[15]Advance Data Reports from the National Health Interview Survey (2008)[Online]. Available: http://www.cdc.gov/nchs/nhis/nhis_ad.htm

[16] Y. Tian, M. Lu, and A. Hampapur, "Robust and efficient foreground anal- ysis for real-time video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit*, 2005, pp. 1182–1187

[17] N. Otsu, "A threshold selection method from gray-level histograms,"In *IEEE Tran.s on system, man and cybernetics*, pp. 62-66, 1979.

[18] M. Shi, Y. Fujisawab, T. Wakabayashia and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image,"In *Pattern Recognition*, Vol. 35, Issue 10, pp. 2051-2059, 2002.