

A Portable Prototype Label Reading System for Blind

Prachee H. Shah¹, G.P.Jain²P.G. Student, Dept. of Electronics Engineering, WIT, Solapur, Maharashtra, India¹Assistant Professor, Dept. of Electronics Engineering, WIT, Solapur, Maharashtra, India²

ABSTRACT: Self-Dependency for disabled persons is of prime importance. The ability of people who are blind or have significant visual impairments to read printed labels and product packages will enhance independent living and foster economic and social self-sufficiency. This paper presents a cost effective prototype system to help blind persons to shop independently. A camera-based assistive text reading framework is proposed to help blind persons read text labels and product packaging from hand-held objects in their daily lives. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, a region of interest (ROI) in the image is defined. In the extracted ROI, text localization and text recognition will be done to acquire text information. Text characters in the localized text regions are then converted into binary format and recognized by trained optical character recognition software. The recognized text codes are output to blind users in speech. The proposed framework is implemented on Raspberry pi board.

Performance of the text localization algorithm is quantitatively evaluated on ICDAR Robust Reading Datasets. Performance of the text recognition algorithm is evaluated on ICDAR-2003 Robust Reading Datasets and SVT dataset. The prototype framework is also evaluated on a dataset of captured images of commercial products with multiple decorative patterns to evaluate the effectiveness of the proposed framework.

KEYWORDS: Portable, Label Reading, Text detection, Text Recognition, Open CV, Raspberry Pi.

I. INTRODUCTION

The National Census of India has estimated around 21.9 Million disabled people in the country [17]. Out of which more than 15 million people in India are blind. This is considered to be the highest among all other disabilities. Blind people are an integral part of the society. However, their disabilities have made them to have less access to computers, Internet, and high quality educational software than the people with clear vision. Consequently, they have not been able to improve on their own knowledge, and have significant influence and impact on the economic, commercial, and educational ventures in the society. One way to narrow this widening gap and see a reversal of this trend is to develop a system, within their economic reach, and which will empower them to communicate freely and widely using the information infrastructure.

Many systems [1], [12], have been designed that have some promise for portable use. Portable assistive readers make use of Optical character recognition software. These systems can be used to recognize text information within the image. But these systems are designed for and perform better with document images which contain text snippets with simple background, well organized characters and standard fonts. These systems, thus, fail to recognize the product labels with complex and decorative backgrounds, different multiple numbers of fonts, different font styles and multiple colours. Portable bar code reader [1] helps a blind user to access information about the different commercial products. The limitation of this reader is, for a blind person it is very hard to exactly locate the barcode and to correctly point the barcode reader at the barcode.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

Figure 1.1 a) shows the different examples of commercial products. Texts printed on the commercial products vary in fonts, colours, font style and the backgrounds. Labels of such products is also designed in multiple decorative patterns. Figure 1.1b) gives the structure of the proposed framework.

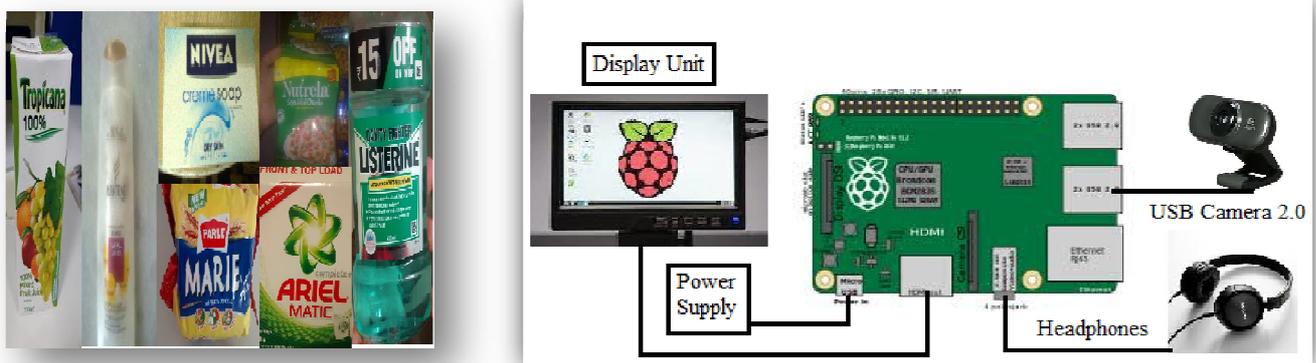


Figure 1.1 a) Examples of printed text from hand-held objects with multiple colours, Complex backgrounds or Non-flat surfaces. b) Structure of the proposed Framework.

Some applications (K-Reader Mobile Applications) which are designed specifically for blind users perform good when detecting text information from receipts, fliers, and many other documents [12]. But these systems/ device fail to give an economic solution of the problem and are available on specific platforms. No smartphones have designed for blind person until now. Thus accessibility of the Mobile application is a different question. Also, these systems cannot read text from commercial products with complex background, non-flat surfaces (text printed on cylindrical surface for e.g. medicinal bottles) and other packaged products. Although a number of assistive reader systems have been designed specifically for blinds, the proposed system presents a **portable and economic solution**.

The proposed framework is one of the techniques which work out as a portable assistive product label reader. The proposed framework is implemented on single board computer –Raspberry pi, in order to provide an economic and portable solution. The framework can be briefly described with the help of –

1. Defining the region of Interest (ROI) - ROI is that part of the captured image which contains the text regions (e.g. product label, sign boards) within the image. It is a challenging problem to automatically localize objects and text ROIs from captured images with complex backgrounds, as text in captured images is most likely surrounded by various background outlier “noise,” and text characters usually appear in multiple scales, fonts, and colours.
2. Locating the text snippets.
3. Recognition of the located text words.
4. Conversion of the text data into audio format.(.wav file)

The rest of the paper is organized as follows- Section II gives an overview of the previously published methods of robust text detection algorithms. Section III describes Text detection and Text recognition algorithm. Hardware and Software parts of the proposed system are described in Section IV. Section V discusses the Evaluation methods and Results. The paper is concluded in section VI.

II. RELATED WORK

Text extraction methods are widely studied as Text localization and recognition in images of real-world scenes is still an open problem, which has been receiving significant attention in the last decade. Textual Information System (TIE) mainly comprises three steps – the first step is Localization of Textual Information within the Image, the second step is recognition of the localized text information, which can be further subdivided as, recognition of the single

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

character and then grouping the text characters to form text snippets. The last but important step can be correcting the recognized words manually to make the TIE more robust.

Many Text localization algorithms have proposed which can be subdivided into categories depending on the approaches they use -sliding window method and methods based on stable regions analysis or connected component analysis. The methods based on sliding windows [2], [3] are more robust to noise, but they have high computational complexity as in sliding window method the input whole image is scanned with windows of multiple sizes. The other drawback of this method is it cannot detect slanted or perspectively distorted text. Methods published in [4], [5], [6] fall in the second category. The methods based on region based approach differ in their approach to detect individual character, which could be based on edge detection, character energy calculation or extremal region detection. Grouping of individual characters into words is performed based on heuristics or graph optimization methods and only unary and pairwise constraints are used. C. Yi and Y. Tian [7] proposed a method of adjacent character grouping to calculate the image patches that contain fragments of text strings.

Rule based and learning based methods are also proposed for text extraction. Learning based methods model text structure and extract representative text features to build text classifiers. Many different classifiers are proposed such as Haar classifier in an Adaboost learning model [8], support vector machine (SVM) model by using and analysing [9] the textural features of characters by. Ma et al. [10] performed classification of text edges by using histograms of oriented gradients and local binary patterns as local features on the SVM model. Chucai Yi and Ying Li Tian propose a camera-base assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives [11]. The system is built using Web camera, Mini Laptop and Microphone. The proposed framework implements the similar concept but on a portable platform Raspberry Pi board, which makes the proposed framework more handy, portable and within economic reach of any visually impaired person.

III. TEXT INFORMATION EXTRACTION SYSTEM (TIE)

The TIE is implemented on a portable platform –a stand-alone Computer Raspberry Pi board. For implementation of end to end recognitions of the product label (extraction of text information from a captured image), Open CV library is used.

Algorithm Overview

The text regions (ROI) within an image can be found out by two methods as specified in Section II. This approach is based on extracting the text regions based on the stroke feature of the character. The captured image is converted into grey-scale. Figure 1.2 describes the stages of the proposed algorithm. 2- Stage classifiers find the extremal regions from the input image. In the first stage classifier, text regions containing strokes are segmented, irrespective of that whether they correspond to a single character, a group of character or a non-text region. For the first stage classifier incrementally computable features such as Area, Perimeter, Height Ratio and Centroid are obtained and then the inferred values are used to limit the search space during next steps of the text formation.

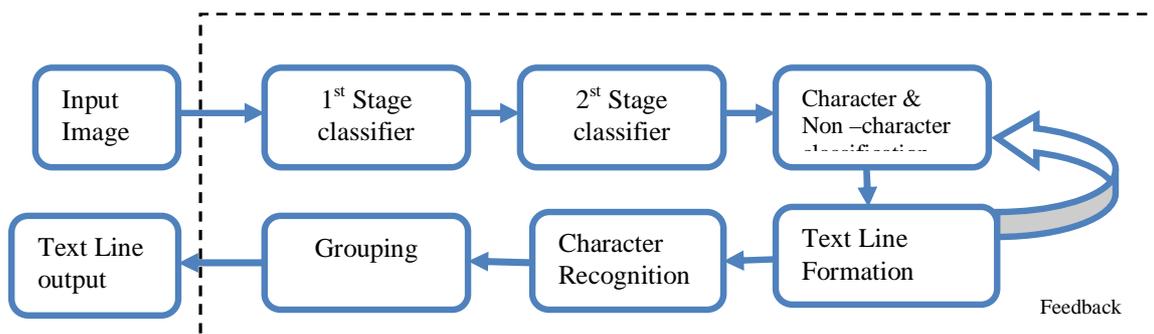


Figure 1.2 Stages of Text Information Extraction System .

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

In this stage, a threshold is increased step by step from 0 to 255, incrementally computable descriptors are computed for each ER (Extremal Region) and the descriptors are used as features for a classifier which estimates the class-conditional probability. The probability value is tracked using the inclusion relation of ER across all thresholds and only the ERs which correspond to local maximum of the probability are selected. Individual characters are recognized by obtaining Maximally Stable Extremal Regions (MSERs) and these regions are processed further for classification.

To classify these regions into non-text and text regions, an Adaboost trained classifier is built. In the second stage, the ERs that passed the first stage are classified into character and non-character classes using more informative but also more computationally expensive features such hole-area ratio and convex hull ratio. In the next step only classified regions, (classified in 2nd Stage) are taken into account and a region sequences can formed based on 1st stage classifier features. Grouping of such sequences into Text blocks is done using Exhaustive Search algorithm proposed in Neumann[16] for grouping horizontally aligned text. This method models a verification function for all possible ERs. The verification function consists of threshold based rules which compares 1st stage features of two regions. The verification function for ER triplets (a group of three characters recognized in 2nd stage) creates a word text line estimate using Least Median-Squares fitting for a given triplet and then verifies that the estimate is valid (based on thresholds created during training). Verification functions for sequences larger than 3 are approximated by verifying that the text line parameters of all (sub) sequences of length 3 are consistent.

For training sequential classifier, characters manually extracted from training samples from [13] are used and for training OCR stage, synthetically generated fonts are used. The extracted data contains Positive Samples and Negative Samples. Positive samples are the cropped images of characters. Negative Images are the non-text regions within the image and can be used as background images. The features used for training classifier are scale invariant but are not rotation variant. Character recognition is performed on a representation derived from the boundaries of extremal regions using a trained OCR. It has been observed that in real-world images a font rarely changes inside a word. This implies that certain character measurements such as character height, aspect ratio, spacing between characters, stroke width, etc. are either constant or constrained to a limited interval. Using these observations the detected characters are grouped together using the text-line hypothesis.

IV. HARDWARE AND SOFTWARE PARTS

A. Hardware

The electrical subsystems making up the assistive product label reader system includes interfacing camera, a power supply to Raspberry Pi board and connecting headphones or speakers to the 3.5mm audio jack. The Raspberry Pi [18] board will be powered from DC power supply through USB cable (Battery can be used). The Raspberry Pi is a series of credit card-sized single-board computers developed in the UK by the Raspberry Pi Foundation. The original Raspberry Pi is based on the Broadcom BCM2835 system on a chip (SoC), which includes an ARM1176JZF-S 700 MHz processor, Video Core IV GPU, and 1GB RAM (model B+). The system has Secure Digital Micro SD sockets for boot media and persistent storage. Raspbian Operating System is installed on Micro SD card. Raspbian is essentially linux based OS and it is based on Debian 'Wheezy' architecture port.

The packages such as *fswebcam*, *vlc*, *omxplayer*, *mpg321*, *imagemagick*, *alsa-utils* are installed for pre-processing of the images and audio files and for interfacing USB camera with board. For Character recognition, *tesseract-ocr* along with *Leptonica 1.70* is installed. *tesseract-ocr* is the optical character recognition software compatible with the debian-wheezy OS. This package is bundled with the other recognition data such as training data for English language –US and EN. Other training data files can be obtained from the source page of TesseractOCR [15]

B. Software

OpenCV [19] is an open source (see <http://opensource.org>) computer vision library available from <http://SourceForge.net/projects/opencvlibrary>. OpenCV was designed for computational efficiency and with a strong focus on real time applications. 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

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

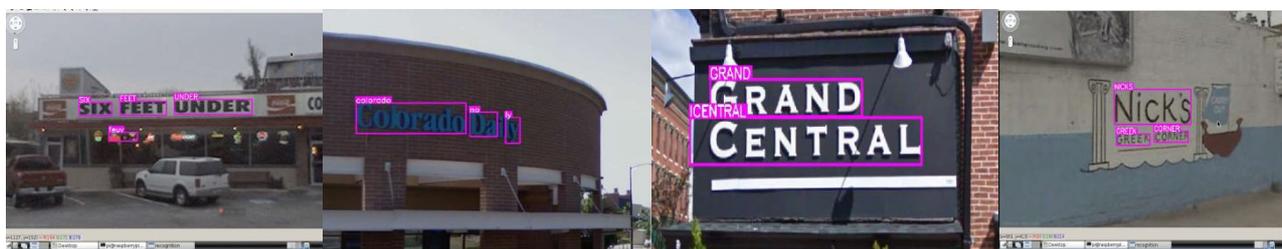
Vol. 4, Issue 9, September 2015

robotics. Because computer vision and machine learning of often go hand-in hand, OpenCV also contains a full, general-purpose Machine Learning Library (MLL). This sub-library is focused on statistical pattern recognition and clustering. The MLL is highly useful for the vision tasks that are at the core of Open CV’s mission, but it is general enough to be used for any machine learning problem. The ml (Machine Learning) module can be used to train and build the AdaBoostsequential classifier.

V. EXPERIMENTAL RESULTS

A. Datasets

The performance of Text Information Extraction System is evaluated on two datasets. First, the ICDAR Robust Reading Datasets[13], [14] are used to evaluate the proposed text localization algorithm. The ICDAR-2003 dataset contains 509 natural scene images in total. Most images contain indoor or outdoor text signage. The image resolutions range from 640 × 480 to 1600 × 1200. Since layout analysis based on adjacent character grouping can only handle text strings with three or more character members, images containing only ground truth text regions of less than three text characters are omitted. The Street View Text (SVT) dataset [13] contains 647 words and 3796 letters in 249 images harvested from Google Street View. The dataset is more challenging because text is present in different orientations, the variety of fonts is bigger and the images are noisy. From these datasets, a new dataset of in all 100 images is created and the proposed framework is tested on 100 images.



SIX FEET UNDER	Colorado nd ly	GRAND CENTRAL	NICKS GREEN CORNER
----------------------	----------------------	------------------	--------------------------

Figure 1.3 Text localization and recognition examples on the SVT dataset. Notice the robustness against noisy and blur. Incorrectly recognized letters marked red. Characters, not in image, but are recognized marked in Blue.



SUMMER WHATEVER THE WEATHER	CARLING	PROPFR FOOD PRONTO	Proceedwith caution Height when raised 115mm C41
-----------------------------------	---------	--------------------------	---

Figure 1.4 Text localization and recognition examples on the ICDAR dataset. Notice the robustness against reflections. Incorrectly recognized letters marked red. Characters, not in image, but are recognized marked in Blue.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

B. Prototype System Evaluation

The hardware of the prototype system includes a Logitech web camera, Raspberry Pi board which act as a Stand-alone computer. The camera is connected to USB port of Raspberry Pi. Open CV performs the processing on the image of the product captured by Web Camera and provides the recognized text words. The output text snippets are stored in .txt file. Pico Text To speech engine converts the contents of text file into .wav file. In order to avoid serious blocking or aural distraction, a wireless “open” style Bluetooth earpiece for presenting detection results as speech outputs to the blind travellers can be used in a full implementation.

The camera is fixed and it captures the image of the objects in held by user in his/her hand. The resolution of the captured image is 960×720 . There were 10 testing objects, including grocery boxes, medicine bottles, books, etc. The user is required keep his head (where the camera is fixed) stationary for a few seconds. Each object was then rotated by the user several times to ensure that surfaces with text captions are exposed and captured.

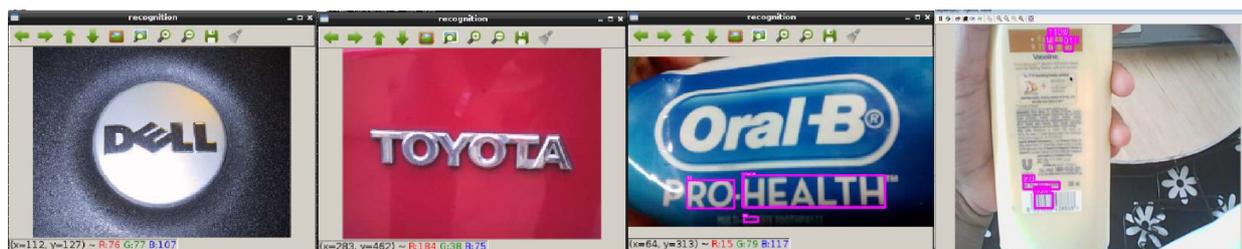
For the proposed framework, the precision is lower than that on the Robust Reading Dataset. The images in the user-captured dataset have lower resolutions and more compact distribution of text information, so they generate low-quality edge maps and text boundaries, which result in improper spatial layouts and text structure features.

Figure 1.5 shows the results of the proposed framework.



Figure 1.5 Results of text localization and recognition on the user-captured dataset. The characters incorrectly recognized marked in Red.

Figure 1.6 a) shows the samples on which the proposed algorithm fails to locate the text snippets. The problems of the proposed algorithm are- It fails to recognise the characters a) if multiple characters are joined together, b) the text embedded in logo c) Characters with no contrast. In 1.6 b) Proposed algorithm fails to recognise the product label –the goal of the proposed framework. For the images shown in 1.6c) , the audio output is not expected due to absence of product label in input image.



a) ICDAR Samples

b)

c)

Figure 1.6 a) ICDAR Samples on which proposed algorithm fails. b) The proposed framework fails to recognize the product label c) No audio output due to absence of Product Label

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2015

The results of the proposed algorithms can be summarized as given in Table No.1. *Matched* are the cases in which the text characters are localized and recognized correctly. *Mismatched* are the characters which are located correctly but are falsely recognized. The characters which are neither located nor recognized are the *Not found* cases.

DATASET	Matched	Mismatched	Not found
SVT and ICDAR 2003	81%	2%	11%
Captured Dataset	69%	12%	19%

Table No.1 Summary of the Results

VI. CONCLUSION AND FUTURE SCOPE

In this paper, a camera-based assistive text reading framework is proposed to help blind persons read text labels and product packaging from hand-held objects in their daily lives. The framework is implemented on a Portable platform Raspberry Pi board. To extract text regions from complex backgrounds, algorithm based on Extremal regions is used. The proposed framework can handle noisy, blur images and reflections. On the standard ICDAR 2003 dataset and SVT (Street View Text, the method achieves results as matched cases 81%. On user captured dataset the framework achieves matches cases as 69%.The precision is lower than that on the Dataset as the images in the user-captured dataset have lower resolutions.

Our future work will extend our localization algorithm to process text strings with characters fewer than three and to design more robust block patterns for text feature extraction. We will also extend our algorithm to handle non-horizontal text strings.

ACKNOWLEDGMENT

Authors would like to seize the opportunity to express sincere gratitude towards Head of Department Prof. Mr S.R.Gengaje and other electronics department faculties for their invaluable co-operation and guidance.

REFERENCES

1. ScanTalker, Bar code scanning application to help Blind Identify over one million products. (2006). [Online]. Available: http://www.freedomscientific.com/fs_news/PressRoom/en/2006/ScanTalker2-Announcement_3-30-2006.asp
2. X. Chen and A. L. Yuille. "Detecting and reading text in natural scenes." Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, 2:366–373, 2004.
3. R. Lienhart and A. Wernicke, "Localizing and segmenting text in images and videos," Circuits and Systems for Video Technology, vol. 12, no. 4, pp. 256–268, 2002.
4. Y.-F. Pan, X. Hou, and C.-L. Liu, "Text localization in natural scene images based on conditional random field," in ICDAR2009. IEEE Computer Society, 2009, pp. 6–10.
5. B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in CVPR 2010, pp. 2963–2970
6. L. Neumann and J. Matas, "A method for text localization and recognition in real-world images," in ACCV 2010, ser. LNCS 6495, vol. IV, November 2010, pp. 2067–2078.
7. 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.
8. X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. Comput. Vision Pattern Recognit., 2004, vol. 2, pp. II-366–II-373.
9. K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive meanshift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.
10. L. Ma, C. Wang, and B. Xiao, "Text detection in natural images based on multi-scale edge detection and classification," in Proc. Int. Congr. Image Signal Process., 2010, vol. 4, pp. 1961–1965.
11. Chucui Yi, Student Member, IEEE, YingLi Tian, Senior Member, IEEE, Aries Arditi, "Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons" IEEE/ASME Transactions on Mechatronics.
12. KReader Mobile User Guide, knfb Reading Technology Inc. (2008). [Online]. Available: <http://www.knfbReading.com>
13. ICDAR 2003 dataset -<http://algoval.essex.ac.uk/icdar/Datasets.html>
14. K. Wang, B. Babenko, and S. Belongie. "End-to-end scene text recognition" In ICCV 2011, 2011.
15. Tesseract Source :https://drive.google.com/folderview?id=0B7110Bj_LprhQnpSRkpGMGV2eE0&usp=sharing
16. L. Neumann and J. Matas. Text localization in real-world images using efficiently pruned exhaustive search. In ICDAR2011, pages 687–691, 2011
17. National Census of India- http://censusindia.gov.in/Census_And_You/disabled_population.aspx
18. www.raspberrypi.org
19. opencv.org