# Portable Camera-Based Product Label Reading For Blind People

Rajkumar N
Anand M.G
Barathiraja N

*Department Of Electrical And Electronics Engineering,*
*The Kavery Engineering College, Mecheri, 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 life. To isolate the object from untidy backgrounds or other surrounding objects in the camera vision, we initially propose an efficient and effective motion based method to define a region of interest (ROI) in the video by asking the user to tremble the object. This scheme extracts moving object region by a mixture-of-Gaussians-based background subtraction technique. In the extracted ROI, text localization and recognition are conducted to acquire text details. To automatically focus the text regions from the object ROI, we offer 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 identification software. The renowned text codes are converted into audio output to the blind users. Performance of the suggested text localization algorithm is quantitatively evaluated on ICDAR-2003 and ICDAR-2011 Robust Reading Datasets. Experimental results demonstrate that our algorithm achieves the highest level of developments at present time. The proof-of-concept example is also evaluated on a dataset collected using ten blind persons to evaluate the effectiveness of the scheme. We explore the user interface issues and robustness of the algorithm in extracting and reading text from different objects with complex backgrounds.**

*Index Terms*—**Assistive devices, blindness, distribution of edge pixels, hand-held objects, optical character recognition (OCR),stroke orientation, text reading, text region localization.**

## I. INTRODUCTION

In worldwide the visually impaired peoples are 314 million, in that 45 million are visual impairment which was released by "World Health Organization" in 10 facts regarding blindness. The estimation of The National Health Interview Survey 25.2 million adult Americans (over 8%) are blind or visually impaired. This number is increasing rapidly as the baby boomer generation ages.

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. And while optical aids, video magnifiers, and screen readers can help blind users and those with low vision to access documents, there are few devices that can provide good access to common hand-held objects such as product packages, and objects printed with text such as prescription medication bottles. 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 so here we are going to propose a system that it useful to blind people.

## II. EXSISTING SYSTEM

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. Some reading assistive systems such as pen scanners might be employed in these and similar situations. Such systems integrate OCR software to offer the function of scanning and recognition of text and some have integrated voice output.

However, these systems are generally designed for and perform best with document images with simple backgrounds, standard fonts, a small range of font sizes, and well-organized characters rather than commercial product boxes with multiple decorative patterns. Most state-of-the-art OCR software cannot directly handle scene images with complex backgrounds.

A number of portable reading assistants have been designed specifically for the visually impaired "K-Reader Mobile" runs on a cell phone and allows the user to read mail, receipts, fliers, and many other documents. However, the document to be read must be nearly flat, placed on a clear, dark surface (i.e., a non-cluttered background), and contain mostly text. In addition, "K-Reader Mobile" accurately reads black print on a white background, but has problems recognizing colored text or text on a colored background. It cannot read text with complex backgrounds. Furthermore, these systems require a blind user to manually localize areas of interest and text regions on the objects in most cases.

Even though a number of reading assistants have been designed specifically for the visually impaired, to our knowledge, no existing reading assistant can read text from the kinds of challenging patterns and backgrounds found on many everyday commercial products. Such as text information can appear in various scales, fonts, colors, and orientations.


Fig.1. Examples of printed text from hand-held objects with multiple colors, complex backgrounds, or non flat surfaces.

**III.PROPOSED SYTEMS**

To overcome the problems defined in problem definitions and also to assist blind persons to read text from those kinds of challenging patterns and backgrounds found on many everyday commercial products of Hand-held objects, then have to conceived of a camera-based assistive text reading framework to track the object of interest within the camera view and extract print text information from the object. Proposed algorithm used in this system can effectively handle complex background and multiple patterns, and extract text information from both hand-held objects and nearby signage.

To overcome the problem in assistive reading systems for blind persons, in existing system very challenging for users to position the object of interest within the center of the camera's view. As of now, there are still no acceptable solutions. This problem approached in stages.

The hand-held object should be appears in the camera view, this thesis use a camera with sufficiently wide angle to accommodate users with only approximate aim. This may often result in other text objects appearing in the camera's view (for example, while shopping at a supermarket). To extract the hand-held object from the camera image, this system going to develop a motion-based method to obtain a region of interest (ROI) of the object. Then, perform text recognition only that ROI.

It is a challenging problem to automatically localize objects and text ROIs from captured images with complex backgrounds, because text in captured images is most likely surrounded by various background outlier "noise," and text characters usually appear in multiple scales, fonts, and colors. For the text orientations, this thesis assumes that text strings in scene images keep approximately horizontal alignment. Many algorithms have been developed for localization of text regions in scene images. We can divide them into two categories: Rule-Based and Learning-Based.

In solving the task at hand, to extract text information from complex backgrounds with multiple and variable text patterns, here propose a text localization algorithm that combines rule-based layout analysis and learning-based text classifier training, which define novel feature maps based on stroke orientations and edge distributions. These, in turn, generate representative and discriminative text features to distinguish text characters from background outliers.

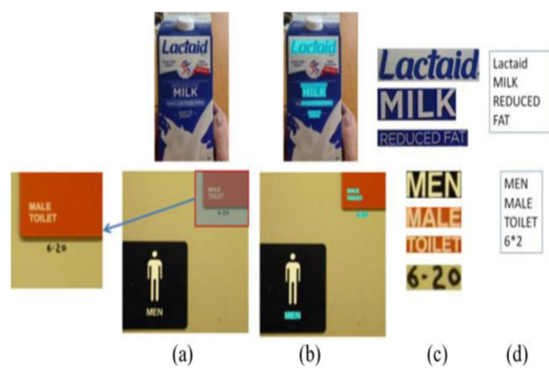


Fig.2. Two examples of text localization and recognition from camera captured Images. (Top) Milk box. (Bottom) Men bath room signage. (a) Camera captured Images. (b) Localized text regions (marked in blue). (c) Text regions cropped from image. (d) Text codes recognized by OCR.

**IV.FRAMEWORK AND ALGORITHM OVERVIEW**

The system framework consists of three functional components they are:

- Scene Capture
- Data Processing and
- Audio Output

The scene capture component collects scenes containing objects of interest in the form of images or video, it corresponds to a camera attached to a pair of sunglasses. 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 data processing component is used for deploying our proposed algorithms, including following processes

• Object-of-interest detection to carefully extract the image of the object held by the blind user from the cluttered background or other neutral objects in the camera view.
• Text localization to obtain text containing image regions then text recognition to transform image-based text information into readable codes.

The audio output component is to inform the blind user of recognized text codes in the form of speech or audio. A Bluetooth earpiece with mini microphone or earpiece is employed for speech output.

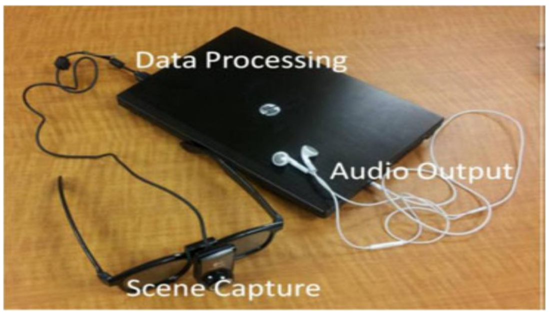


Fig.3. Snapshot demo system including three functional components for Scene capture, data processing, and audio output.

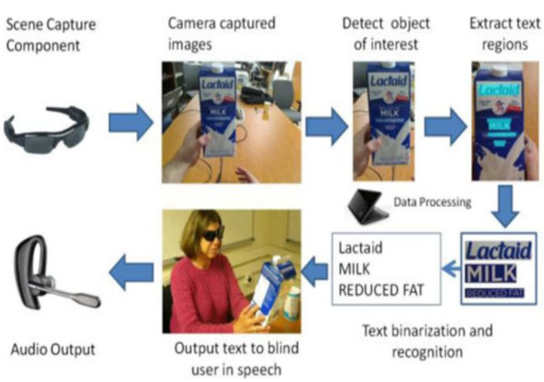


Fig.4. Flowchart of the proposed framework to read text from hand-held objects for blind users.

A frame sequence V is captured by a camera worn by blind users, containing their hand-held objects and cluttered background. To extract text information from the objects, motion based object detection is first applied to determine the user's object of interest S by shaking the object while recording video

$$S = \frac{1}{|V|}\sum_i R(Vi, B) \qquad (1)$$

Where Vi denotes the $i^{th}$ frame in the captured sequence, |V | denotes the number of frames, *B* denotes the estimated background from motion- based object detection, and *R* represents the calculated foreground object at each frame.

Next, the novel proposed text localization algorithm is applied to the object of interest to extract text regions. At first, candidate text regions are generated by layout analysis of color uniformity and horizontal alignment

$$X^C = \text{argmax}_{s \in S} L(s) \qquad (2)$$

Where L ( • ) denotes the suitability responses of text layout and $X^C$ denotes the candidate text regions from object of interest S.

Then, a text classifier is generated from a Cascade-Adaboost learning model, by using stroke orientations and edge distributions of text characters as features.

$$X = H[X^C] = H[\text{argmax}_{s \in S} L(s)] \qquad (3)$$

Where H denotes the Cascade-Adaboost classifier and X denotes the localized text regions. After text region localization, off-the-shelf OCR is employed to perform text recognition in the localized text regions. The recognized words are transformed into speech for blind users.

The main contributions include in this prototype system are:

- A motion-based algorithm to solve the aiming problem for blind users by their simply shaking the object of interest for a brief period.
- A novel algorithm of automatic text localization to extract text regions from complex background and multiple text patterns
- A portable camera-based assistive framework to aid blind persons reading text from hand held objects.

Algorithms of the proposed system are evaluated over images captured by blind users using the described techniques.

## V.TEXT RECOGNITION AND AUDIO OUTPUT

The audio output component is to inform the blind user of recognized text codes in the form of speech or audio. A Bluetooth earpiece with mini microphone or earpiece is employed for speech output. Text recognition is performed by off-the-shelf OCR prior to output of informative words from the localized text region. The accommodations of characters inside a text region are labels in the minimum rectangular area, so the border of the text region contacts the edge boundary of the text character. However, the experiments show that OCR generates better performance if text regions are first assigned proper margin areas and diarized to segment text characters from background. Thus, each and every localized text region is enlarged by enhancing the height and width by 10 pixels, respectively, and then, Here Otsu's method is used to perform binarization of text regions, where margin areas are always considered as background. We test both open- and closed-source solutions that allow the final stage of conversion to letter codes.

The recognized text codes are recorded in script files. Then, employ the Microsoft Speech Software Development Kit to load these files and display the audio output of text information. Blind users can adjust rate of speech, volume, and tone according to their preferences.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we have described a prototype system read printed text on hand- held objects for assisting blind persons. In order to solve the common problem for blind users, we proposed a motion-based technique to detect the object of interest, while the blind user simply shakes the object for few seconds. This method can effectively distinguish the object of interest from background or other objects in the camera vision. To extract text regions from complex backgrounds, we have proposed a "Haar Cascade Classifier Algorithm" text localization algorithm the corresponding feature maps estimate the global structural feature of text at every pixel. Block patterns project the proposed feature maps of an image patch into a feature vector. Adjacent character grouping is performed to calculate candidates of text patches prepared for text classification. An Adaboost learning model is applied 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.

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. Furthermore, we will address the significant human interface issues associated with reading text by blind users.

## REFERENCES

[1] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition:ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognition, 2011, pp. 1491–1496.

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

[3] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Comput. Vision Pattern Recognition, 2010, pp. 2963–2970.

[4] 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.

[5] 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.

[6] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011). [Online]. Available: http://www.m.cs.osakafuu.ac.jp/cbdar2011/

[7] J. Zhang and R. Kasturi, "Extraction of text objects in video documents: recent progress," in Proc. IAPR Workshop Document Anal. Syst., 2008, pp. 5–17.57

[8] K. Kim, K. Jung, and 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. Intel., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[9] 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.

[10]N.Nikolaou and N. Papamarkos, "Color reduction for complex document images," Int. J. Imaging Syst. Technol., vol. 19, pp. 14–26, 2009.

M.G.Anand was born on 1985 at Erode. He received B.E degree in Electronics Communication Engineering from Vellalar College of Engineering And Technology in the year of 2007.He got M.E degree in electronics and drives From K.S.R College of Engineering in the year of 2011. He is now working as an assistant professor at The Kavery Engineering College in the Department of Electrical and Electronics Engineering. His research interest includes power electronics, Renewable power Generation and distribution systems. He is a life time member of ISTE.

Rajkumar N pursued the B.Tech Degree in Information technology in the Department of Information Technology, on May 2010 from Krishnasamy college of Engineering and Technology, Cuddalore Affiliated to Anna University, Chennai, India. Since 2012, he has been working towards the M.E Degree in Embedded Systems in the department of Electrical and Electronics Engineering at The Kavery Engineering College, Salem Affiliated to Anna University, Chennai, India. His research interests includes in the field of Embedded Systems, Portable Camera-Based Assistive Text and Product Label Reading for Blind People.

Barathiraja N pursued the B.E Degree in Electronics and communication Engineering in the Department of Electronics and communication Engineering, on May 2010 from Annai Theresa Engineering College, Villupuram Affiliated to Anna University, Chennai , India. Since 2012, he has been working towards the M.E Degree in Embedded and Technologies in the department of Electrical and Electronics Engineering at Systems V.R.S.College of Engineering and Technology, Villupuram Affiliated to Anna University, Chennai, India. His research interests includes in the field of Embedded Systems, Portable Camera-Based Assistive Text and Product Label Reading for Blind People.