ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

# Design of Embedded Linux For Label Reading By Using OCR Algorithm

**Vinusha**
M.Tech Student,
Malla Reddy Institute of Technology,
Hyderabad,TS India.

**Sowjanya Devi**
Assistant Professor,
Dept of ECE,
Malla Reddy Institute of Technology,
Hyderabad,TS 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. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, we first propose an efficient and effective motionbased method to define a region of interest (ROI) in the video byasking the user to shake the object. 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 Adaboostmodel. 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. Performance of the proposed text localization algorithm is quantitatively evaluated on ICDAR-2003 and ICDAR-2011 Robust Reading Datasets. Experimental results demonstrate that our algorithm achieves the state of the arts. The proof-of-concept prototype is also evaluated on a dataset collected using ten blind persons to evaluate the effectiveness of the system's hardware. We explore user interface issues and assess 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.

## INTRODUCTION:

OF the 314 million visually impaired people worldwide, 45 million are blind [1]. Even in a developed country like the U.S., the 2008 National Health Interview Survey reported that an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired [2]. This number is increasing rapidly as the baby boomer generation ages. Recent developments in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such optical character recognition (OCR) systems.

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. Andwhile 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.

## Existing System:

Today, there are already a fewsystems 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 [22] 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.

## Proposed System:

To assist blind persons to read text from these kinds of hand-held objects, we have 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. Our proposed algorithm can effectively handle complex background and multiple patterns, and extract text information from both hand-held objects and nearby signage. In assistive reading systems for blind persons, it is 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. We approach the problem in stages. To make sure the hand-held object appears in the camera view, we 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, we develop a motion-based method to obtain a region of interest (ROI) of the object. Then, we perform text recognition only in this ROI.

## FRAMEWORK AND ALGORITHM OVERVIEW:

This paper presents a prototype system of assistive text reading. The system framework consists of three functional components: scene capture, data processing, and audio output.

capture component collects scenes containing objects of interest in the form of images or video. In our prototype, it corresponds to a camera attached to a pair of sunglasses. The data processing component is used for deploying our proposed algorithms, including 1) object-of-interest detection to selectively extract the image of the object held by the blind user from the cluttered background or other neutral objects in the camera view; and 2) text localization to obtain image regions containing text, and text recognition to transform image-based text information into readable codes.We use aminlaptop as the processing device in our current prototype system. The audio output component is to inform the blind user of recognized text codes. A Bluetooth earpiece with minimicrophone is employed for speech output. This simple hardware configuration ensures the portability of the assistive text reading system.

## OBJECT REGION DETECTION:

To ensure that the hand-held object appears in the camera view, we employ a camera with a reasonably wide angle in our prototype system (since the blind user may not aim accurately). However, this may result in some other extraneous but perhaps text-like objects appearing in the camera viewfor example, when a user is shopping at a supermarket). To extract the hand-held object of interest from other objects in the camera view, we ask users to shake the hand-held objects containing the text they wish to identify and then employ a motion-based method to localize the objects from cluttered background. Background subtraction (BGS) is a conventional and effective approach to detect moving objects for video surveillance systems with stationary cameras. To detect moving objects in a dynamic scene, many adaptive BGS techniques have been developed.

Stauffer and Grimson [28] modeled each pixel as a mixture of Gaussians and used an approximation to update the model. A mixture of K Gaussians is applied for BGS, where K is from 3 to 5. In this process, the prior weights ofK Gaussians are online adjusted based on frame variations. Since background imagery is nearly constant in all frames, a Gaussian always compatible with its subsequent frame pixel distribution is more likely to be the background model. This Gaussian-mixture-model-based method is robust to slow lighting changes, but cannot handle complex foregrounds and quick lighting changes. Tian et al. [29] further improved the multiple Gaussian-mixturebased BGS method to better define foreground while remove background objects.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

First, texture information is employed to remove false positive foreground areas. These areas should be background but are often determined as foreground because of sudden lighting changes. A texture similarity measure is defined to evaluate whether the detected foreground motion is caused by lighting change or moving object. Second, in addition to quick lighting changes, BGS is also influenced by shadows.Many systems use color information to remove the shadow, but this does not work on grayscale videos. To solve this problem, the normalized cross correlation of the intensities is used for shadow removal. The grayscale distribution of a shadow region is very similar to that of the corresponding background region, except is a little darker. Thus, for a pixel in BGS-modeled foreground areas, we calculate theNCC between the current frame and the background frame to evaluate their similarity and remove the
influence of shadow.

## AUTOMATIC TEXT EXTRACTION:

In order to handle complex backgrounds, we propose two novel feature maps to extracts text features based on stroke orientations and edge distributions, respectively. Here, stroke is defined as a uniform region with bounded width and significant extent. These feature maps are combined to build an Adaboostbased text classifier.

### Text Stroke Orientation:

Text characters consist of strokes with constant or variable orientation as the basic structure. Here, we propose a new type of feature, stroke orientation, to describe the local structure of text characters. From the pixel-level analysis, stroke orientation is perpendicular to the gradient orientations at pixels of stroke boundaries. To model the text structure by stroke orientations, we propose a new operator to map a gradient feature of strokes to each pixel. It extends the local structure of a stroke boundary into its neighborhood by gradient of orientations. We use it to develop a feature map to analyze global structures of text characters.

### Distribution of Edge Pixels:

In an edge map, text characters appear in the form of stroke boundaries. The distribution of edge pixels in stroke boundaries also describes the characteristic structure of text.

Themost commonly used feature [34], [36] is edge density of text region. But the edge density measure does not give any spatial information of edge pixels. It is generally used for distinguishing text regions from relatively clean background regions. To model text structure by spatial distribution of edge pixels, we propose an operator to map each pixel of an image patch into the number of edge pixels in its cross neighborhood.

### Adaboost Learning of Text Features:

Based on the feature maps of gradient, stroke orientation, and edge distribution, a text classifier is trained from an Adaboost learning model. Image patches with fixed size (height 48 pixels, width 96 pixels) are collected and resized from images taken from the ICDAR-2011 robust reading competition [10] to generate a training set for learning features of text. We generate positive training samples by scaling or slicing the ground truth text regions, according to the aspect ratio of width w to height h. To train a robust text classifier, we ensure that most positive training samples contain two to four text characters.

We build a relationship between the width-to-height aspect ratio and the number of characters of ground truth text regions. It shows that the ground truth regions with two to four text characters have width-to-height ratios between 0.8 and 2.5, while the ones lower than 0.8 mostly have less than two characters and the ones higher than 2.5 mostly have more than four characters.Therefore, if the ratio is $w/h < 0.8$ with too few characters, the region is discarded. If the ratio $w/h \geq 2.5$ corresponding to more than four text characters, we slice this ground truth region into overlapped patches with width-to-height ratio 2:1. If the ratio w/h falls in [0.8, 2.5), we keep it unsliced and scale it to width-toheight ratio 2:1.

Then, the samples are normalized into width 96 and height 48 pixels for training. The negative training samples are generated by extracting the image regions containing edge boundaries of nontext objects. These regions also have widthto- height ratio 2:1, and we similarly scale them into width 96 and height 48. In this training set, there are a total of 15 301 positive samples, each containing several text characters, and 35 933 negative samples without containing any text informationfor learning features of background outliers.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

## Text Region Localization:

Text localization is then performed on the camera-based image. The Cascade-Adaboost classifier confirms the existence of text information in an image patch but cannot handle the whole image, so heuristic layout analysis is performed to extract candidate image patches prepared for text classification. Text information in the image usually appears in the form of horizontal text strings containing no less than three character members. Therefore, adjacent character grouping [33] is used to calculate the image patches that contain fragments of text strings.

These fragments consist of three or more neighboring edge boundaries that have approximately equal heights and stay in horizontal alignment, But not all the satisfied neighboring edge boundaries are text string fragments. Thus, the classifier is applied to the image patches to determine whether they contain text or not. Finally, overlapped text patches are merged into the text region, which is the minimum rectangle area circumscribing the text patches. The text string fragments inside those patches are assembled into informative words.

## TEXT RECOGNITION AND AUDIO OUTPUT:

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, our experiments show that OCR generates better performance if text regions are first assigned proper margin areas and binarized to segment text characters from background. Thus, each localized text region is enlarged by enhancing the height and width by 10 pixels, respectively, and then, we use Otsu's method [18] 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(e.g. OmniPage, Tesseract, ABBYReader). The recognized text codes are recorded in script files. Then, we employ the Microsoft Speech Software Development Kit to load these files and display the audio output of text information. Blind users can adjust speech rate, volume, and tone according to their preferences.

## CONCLUSIONS:

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

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

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

[4] 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.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
*A Peer Reviewed Open Access International Journal*

[5] 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.

[6] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronictravel aids for blind: A survey," IEEE Trans. Syst., Man, Cybern., vol. 40, no. 1, pp. 25–35, Jan. 2010.

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

[8] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in Proc. Int. Conf. Machine Learning, 1996, pp. 148–156.

[9] N. Giudice and G. Legge, "Blind navigation and the role of technology," in The Engineering Handbook of Smart Technology for Aging, Disability, and Independence, A. A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. Hoboken, NJ, USA: Wiley, 2008.

[10] 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. Recognit., 2011, pp. 1491–1496.

[11] 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. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003. nol., vol. 19, pp. 14–26, 2009.

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

[13] 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 Image Process., vol. 16, no. 8, pp. 2117–2128,Aug. 2007.

[14] S. M. Lucas, "ICDAR 2005 text locating competition results," in Proc.Int. Conf. Document Anal. Recognit., 2005, vol. 1, pp. 80–84.

[15] 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.

[16] R. Manduchi and J. Coughlan, "(Computer) vision without sight," Commun. ACM, vol. 55, no. 1, pp. 96–104, 2012.

[17] N. Nikolaou and N. Papamarkos, "Color reduction for complex document images," Int. J. Imaging Syst. Tech