International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

Advance Text and Product Label Reading from Hand-Held Objects for Blind Persons Using Portable Camera

[1]D.suresh [2]MP.Venkatesh

[1] M.Tech (DECS), Shree college, Tirupati,
degalasuresh8@gmail.com
[2]Assistant Professor, Shree College, Tirupati,
venkatesh3816@gmail.com

***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 motion based method to define a region of interest (ROI) in the video by asking 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 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. 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

I. INTRODUCTION

Of the 314 million visually impaired people worldwide, 45 million are blind. Even in a developed country like the National Health Interview Survey reported that , an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired. 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. And while optical aids, video magnifiers, and screen readers can help blind users and those with low

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

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. 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 [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. A number of portable reading assistants have been designed specifically for the visually impaired 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),

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

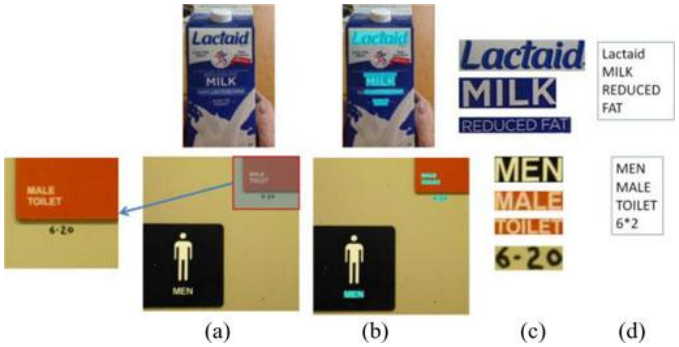

Fig. 2. Two examples of text localization and recognition from cameracaptured images. (Top) Milk box. (Bottom) Men bathroom signage. (a) camera captured images. (b) Localized text regions (marked in blue). (c) Text regions cropped from image. (d) Text codes recognized by OCR. Text at the top-right corner of bottom image is shown in a magnified callout.

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, text printed on cylinders with warped or incomplete images (such as soup cans or medicine bottles). Furthermore, these systems require a blind user to manually localize areas of interest and text regions on the objects in most cases.

Although 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. As shown in Fig. 1, such text information can appear in multiple scales, fonts, colors, and orientations. 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, as shown in Fig. 2. In assistive reading systems for blind persons, it is very challenging for users to position the object of interest within the center of the

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

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. 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 paper 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 divide them into two categories: rule-based and learning-based.

Rule-based algorithms apply pixel-level image processing to extract text information from predefined text layouts such as character size, aspect ratio, edge density, character structure, color uniformity of text string, etc. Phan *et al.* [19] analyzed edge pixel density with the Laplacian operator and employed maximum gradient differences to identify text regions. Shivakumara *et al.* [26] used gradient difference maps and performed global binarization to obtain text regions. Epshtein *et al.* [7] designed stroke width transforms to localize text characters. Nikolaou and Papamarkos [17] applied color reduction to extract text in uniform colors. In [5], color-based text segmentation is performed through a Gaussian mixture model for calculating a confidence value for text regions.

This type of algorithm tries to define a universal feature descriptor of text. Learning-based algorithms, on the other hand, model text structure and extract representative text features to build text classifiers. Chen and Yuille [4] presented five types of Haarbased block patterns to train text classifiers in an Adaboost learning model. Kim *et al.* [11] considered text as a specific texture and analyzed the textural features of characters by a support vector machine (SVM) model. Kumar *et al.* [13] used globally matched wavelet filter responses of text structure as features. Ma *et al.* [15] performed classification of text edges by using histograms of oriented gradients and local binary patterns as local features on the SVM model. Shi *et al.* [25] employed gradient and curvature features to model the grayscale curve for handwritten numeral recognition under a Bayesian discriminate function. In our research group, we have previously developed rule-based algorithms to extract text from scene images

II. FRAMEWORK AND ALGORITHM OVERVIEW

This paper presents a prototype system of assistive text reading. As illustrated in Fig. 3, the system framework consists of three functional components: scene capture, data processing, and audio output. The scene capture component collects scenes

Fig. 3. Snapshot of our demo system, including three functional componentsfor scene capture, data processing, and audio output.

Fig. 4. Flowchart of the proposed framework to read text from hand-held objects for blind users. 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 a min laptop 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 mini-microphone is employed for speech output. This simple hardware configuration ensures the portability of the assistive text reading system. Fig. 4 depicts a work flowchart of the prototype system. 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

$S = 1 |V|_{-} I R(Vi,B)$ (1)

where Vi denotes the ith 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. The object of interest is localized by the average of foreground masks . Next, our 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

$XC = \text{argmax}_{s \in SL}(s)$ (2)

where $L(\cdot)$ denotes the suitability responses of text layout and XC 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 .

$X = H\_XC\_ = H[\text{argmax}_{s \in SL}(s)]$ (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. Our main contributions embodied in this prototype system are: 1) a novelmotion-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. Algorithms of the proposed system are evaluated over images captured by blind users using the described techniques.

III. 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 view for 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 of$K$ 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.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

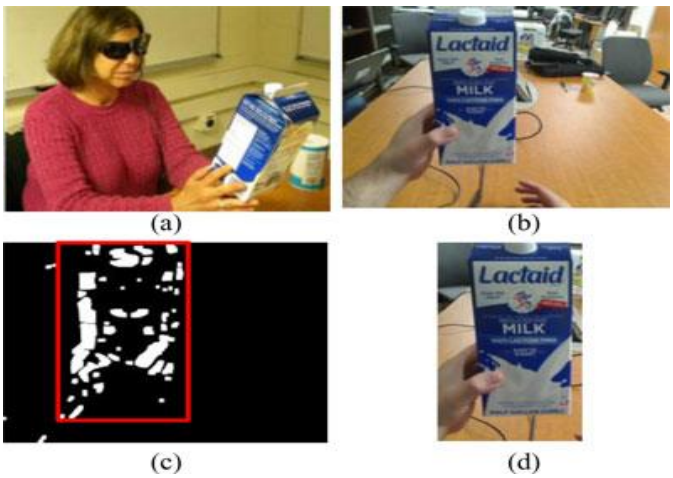

Fig. 5. Localizing the image region of the hand-held object of interest.
(a) Capturing images by a camera mounted on a pair of sunglasses.
(b) Example of a captured image. (c) Detected moving areas in the image
while the user shaking the object (region inside the bounding box). (d) Detected region of the hand-held object for further processing of text recognition. H

Handle complex foregrounds and quick lighting changes. further improved the multiple Gaussian-mixture based BGS method to better define foreground while remove background objects. 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 the*NCC* between the current frame and thebackground frame to evaluate their similarity and remove the influence of shadow. As shown in Fig. 5, while capturing images of the hand-held object, the blind user first holds the object still, and then lightly shakes the object for 1 or 2 s. Here, we apply the efficient multiple Gaussian-mixture-based BGS method to detect the object region while blind user shakes it. More details of the algorithm can be found in [29].Once the object of interest is extracted from the camera image, the system is ready to apply our automatic text extraction algorithm.

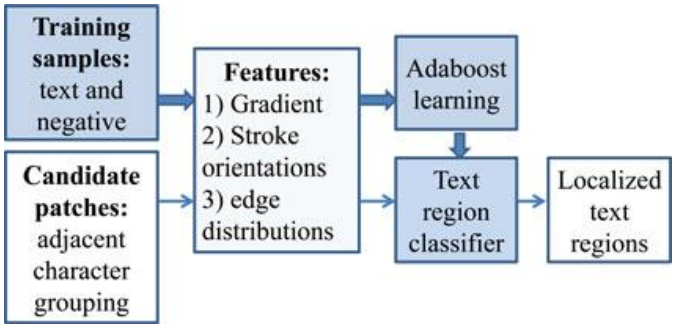

Fig. 6. Diagram of the proposed Adaboost-learning-based text region localization algorithm by using stroke orientations and edge distributions

IV. AUTOMATIC TEXT EXTRACTION

we design a learning-based algorithm for automatic localization of text regions in image. 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. Fig. 6. Diagram of the proposed Adaboost-learning-based text region localization algorithm by using stroke orientations and edge distributions. Fig. 7. Sample of text strokes showing relationships between stroke orientations and gradient orientations at pixels of stroke boundaries. Blue arrows denote the stroke orientations at the sections and red arrows denote the gradient orientations at pixels of stroke boundaries.

V. TEXT RECOGNITION AND AUDIO OUTPUT

` As shown in fig 6 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.

VI. EXPERIMENTS

A. Datasets

Two datasets are used to evaluate our algorithm. First, the ICDAR Robust Reading Dataset [10], [14] is 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, we omit the images containing only ground truth text regions of less than three text characters. Thus, 488 images are selected from this dataset as testing images to evaluate our localization algorithm.

To further understand the performance of the prototype system
and develop a user-friendly interface, following Human Subjects Institutional Review Board approval, we recruited ten
blind persons to collect a dataset of reading text on hand-held

Fig. 12. Examples of blind persons capturing images of the object in their hands.

objects. The hardware of the prototype system includes a Logitech web camera with autofocus, which is secured to the nose bridge of a pair of sunglasses. The camera is connected to an HP mini laptop by a USB connection. The laptop performs the processing and provides audio output. In order to avoid serious blocking or aural distraction, we would choose a wireless "open" style Bluetooth earpiece for presenting detection results as speech outputs to the blind travelers in a full prototype implementation.

B. Evaluations of Text Region Localization

Text classification based on the Cascade-Adaboost classifier plays an important role in text region localization. To evaluate
the effectiveness of the text classifier, we first performed a group of experiments on the dataset of sample patches, in which the patches containing text are positive samples and those without text are negative samples. These patches are cropped from natural scene images in ICDAR-2003 and ICDAR-2011 Robust Reading Datasets

Each patch was assigned a prediction score by the text classifier; a higher score indicates a higher probability of text information. We define the true positive rate as the ratio of correct positive predictions to the total number of positive samples. Similarly, the false positive rate is the ratio of correct positive predictions to the total number of positive predictions. Fig. 13 plots the variation of true positive against false positive rates.

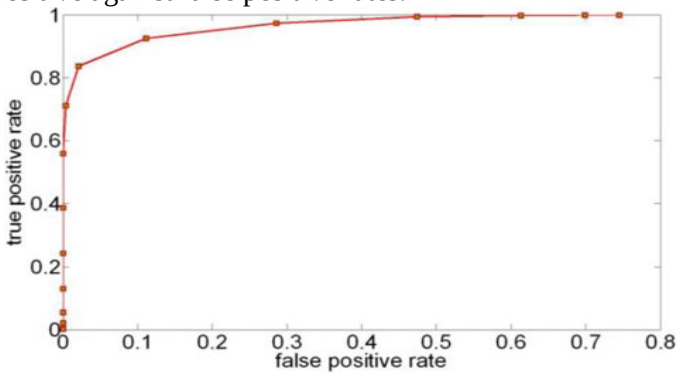

Fig. 13. Curve of classification performance, where horizontal axis denotes false positive rate and vertical axis denotes true positive rate.

C. Prototype System Evaluation

The automatic ROI detection and text localization algorithms were independently evaluated as unit tests to ensure effectiveness and robustness of the whole system. We subsequently evaluated this prototype system of assistive text reading using images of hand-held objects captured by ten blind users in person. Two calibrations were applied to prepare for the system test. First, we instructed blind users to place hand-held object within the camera view. Since it is difficult for blind users to aim their held objects, we employed a camera with a reasonably wide angle. In future systems, we will add finger point detection and tracking to adaptively instruct blind users to aim the object. Second, in an applicable blind-assistive system, a text localization algorithm might prefer higher recall by sacrificing some precision. We

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 02 Issue 11
November 2015

adjusted the parameters of our text localization algorithm and obtained another group of evaluation results, as precision 0.48, recall 0.72, *f*-measure 0.51. The higher recall ensures a lower miss (false negative) rate. To filter out false positive localizations, we could further employ some post processing algorithm based on scene text recognition or lexical analysis. This work will be carried out in future work.

Fig. 16. (a) Some results of text localization on the user-captured dataset, where localized text regions are marked in blue. (b) Two groups of enlarged text regions, binarized text regions, and word recognition results from top to down.

OCR is applied to the localized text regions for character and word recognition. Fig. 16 shows some examples of text localization and recognition of our proposed framework. We note that the recognition algorithm might not correctly and completely output the words inside localized regions. Additional spelling correction is likely required to output accurate text information. Our text reading system spends 1.87 s on average reading text from a camera-based image. The system efficiency can and will be improved by parallel processing of text extraction and device input/output, i.e., speech output of recognized text and localization of text regions in the next image are performed simultaneously.

VII. CONCLUSION AND FUTURE WORK

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. 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-u.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.

[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 electronic travel 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.