Visual Product Identification

For Blind Peoples

Prajakata Bhunje[1], Anuja Deshmukh[2], Dhanshree Phalake[3], Amruta Bandal[4]

[1] Prajakta Nitin Bhunje, Computer Engg., DGOI FOE,Maharashtra, India
[2] Anuja Ramesh Deshmukh, Computer Engg., DGOI FOE,Maharashtra, India
[3] Dhanshree Viajysinh Phalake, Computer Engg., DGOI FOE,Maharashtra, India
[4] Amruta Avinash Bandal, Computer Engg., DGOI FOE,Maharashtra, India

---***---

Abstract - We propose a camera-based assistive text reading framework to help blind persons ad 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 rest propose an ancient and active motion-based method to dine 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 Ad boost model. Text characters in the localized text regions are then binaries and recognized by o -the-shelf optical character recognition (OCR) 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 10 blind persons, to evaluate ether activeness of the systems hardware. We explore user interface issues, and assess robustness of the algorithm in extracting and reading text from di erent objects with complex backgrounds.

Keywords: *blindness; assistive devices; text reading; hand-held objects, text region localization; stroke orientation; distribution of edge pixels; OCR; Key Words*

I. INTRODUCTION

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 mangier 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. 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 optical character recognition (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. KReader Mobile runs on a cell phone and allows the user to read mail, receipts, fliers and many other documents [13]. 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. Furthermore, 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, 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.

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

Fig. 1. Example of printed text.

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 back- grounds 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 cameras 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 se a camera with sufficiently wide angle to accommodate users with only approximate aim. This may often result in other text objects appearing in the cameras 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.

II. MOTIVATION

Today, there are already a few systems that have some promise for portable use, but they cannot handle product labeling. For example, portable bar

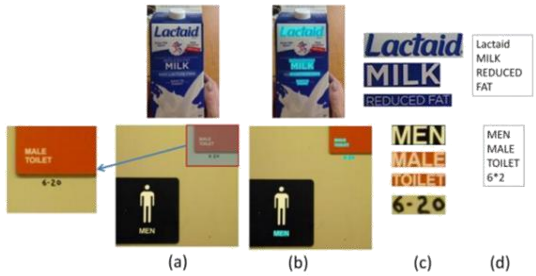

Fig. 2. Example of Text localization.

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. Two of the biggest challenges to independence for blind individuals are difficulties in accessing printed material [1] and the stressors associated with safe and efficient navigation. Access to printed documents has been greatly improved by the development and proliferation of adaptive technologies such as screen-reading programs, optical character recognition software, text-to-speech engines, and electronic Braille displays. By contrast, difficulty accessing room numbers, street signs, store names, bus numbers, maps, and other printed information related to navigation remains a major challenge for blind travel. Imagine trying toad room n257 in a large university building without being able to read the room numbers or access the you are here map at the buildings entrance. Braille signage certainly helps in identifying a room, but it is difcult for blind people to nd Braille signs. In addition, only a modest fraction of the more than 3 million visually impaired people in the United States read Braille. Estimates put the number of Braille readers between 15,000 and 85,000.

III. OBJECTIVES

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. The scene 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 a mini laptop as the processing device in our current prototype system. The audio output component is to inform the blind user of recognized text codes.
In solving the task at hand, to extract text information from complex backgrounds with multiple and variable text patterns, we 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.

IV. APPLICATIONS

1. Image capturing and pre-processing
2. Automatic text extraction
3. Text recognition and audio output
4. Prototype System Evaluation

V. SYSTEM DESIGN

This paper presents a prototype system of Assistive text reading. The system framework consists of three functional components: scene

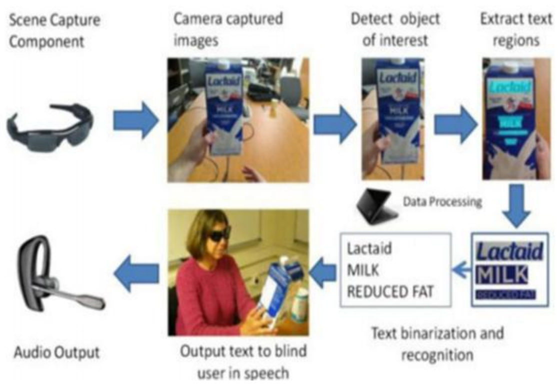

Fig. 3. Block Diagram of Text Reading.

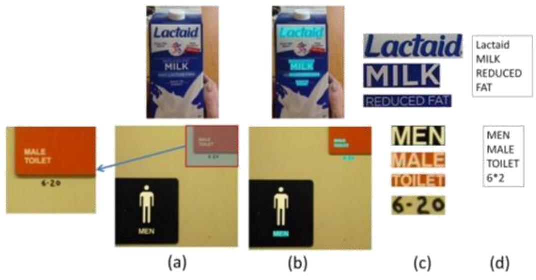

Fig. 4. Flowchart of the proposed framework to read text from hand-held objects for blind users.

Processing, and audio output. The scene 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

2. Text localization to obtain image regions containing text, and text recognition to transform image based text information into readable codes. We use a mini laptop as the processing device in our current prototype system. The audio output component is to inform the blind user of recognized text codes.

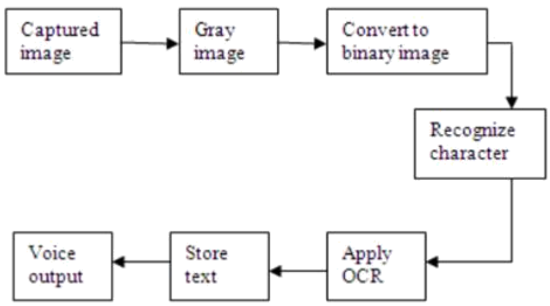

Fig. 5. Two examples of text localization and recognition from camera captured Images.

VI. MODULES

A. IMAGE CAPTURING AND PREPROCESSING

The video is captured by using web-cam and the frames from the video is segregated and undergone to the pre-processing. First, get the objects continuously from the camera and adapted to process. Once the object of interest is extracted from the camera image and it Converted into gray image. Use hears cascade classifier for recognizing the character from the object. The work with a cascade classifier includes two major stages: training and detection. For training need a set of samples. There are two types of samples: positive and negative. To extract the hand-held object of interest from other objects in the camera view, ask users to shake the hand-held objects containing the text they wish to identify and then employ a motion-based method to localize objects from cluttered background.

B. AUTOMATIC TEXT EXTRACTION

In order to handle complex backgrounds, 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 Ad boost based text classifier.

C. 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 it cannot the whole images, 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.

D. 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 commodation of characters inside it, so the border of the text region contacts the edge boundary of the text characters. However, this experiment shows that OCR generates better performance text regions are first assigned proper margin areas and binaries to segments text characters from background. 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 speech rate, volume and tone according to their preferences.

VII. CONCLUSIONS 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 are defined to 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 captured 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 3 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.

VIII. REFERENCES

[1] 10 facts about blindness and visual impairment, World Health Organization: Blindness and visual Impairment, 2009.

[2] Advance Data Repor ts f rom the Nat ional Heal th Interview Survey, 2008.

[3] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005,2007, 2009, 2011) .

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

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

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

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

[8] Y. Freund and R. Schapire, Experiments with a new boosting algorithm, In Int. Conf. on Machine Learning, pp.148156, 1996.

[9] N. Giudice and G. Legge, Blind navigation and the role of technology, in The engineering handbook of smart technology for aging, disability.