

Portable Camera and Sensor Based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons

Priyadharshini.M¹, Rajarajeswari.M², Ramya.G³, Thejeswini.K.S⁴, Venkatesan.M⁵

UG Scholar, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India¹²³⁴

Assistant Professor, Department of ECE, Adhiyamaan College of Engineering, Hosur, Tamilnadu, India⁵

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 (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 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.

I. INTRODUCTION

Of the 314 million visually impaired people worldwide, 45 million are blind. Even in a developed country like the United States, the 2008 National Health 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 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 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 .



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

II. FRAMEWORK

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 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, 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.

A frame sequence 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 by shaking the object while recording video. Our main contributions embodied in this prototype system are: (1) a novel motion-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 using images captured by blind users using the described techniques.

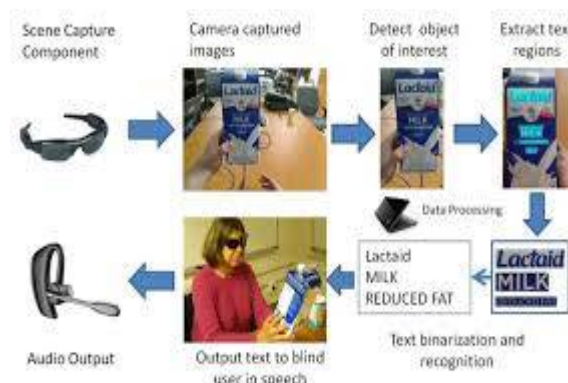


Fig. 2. Flowchart of the proposed framework to read text from hand-held objects for blind users.

III. OBJECT REGION DETECTION

To ensure 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 textured 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 background subtraction techniques have been developed. Stauffer and Grimson modeled each pixel as a mixture of Gaussians and used an approximation to update the model. A mixture of Gaussians is applied for BGS, where k is from 3 to 5. In this process, the prior weights of 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. further improved the multiple Gaussian mixtures 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 gray-scale videos. To solve this problem, the normalized cross-correlation (NCC) of the intensities is used for shadow removal. The gray-scale 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 the background 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 one or two seconds. Here, we apply the efficient multiple Gaussian Mixtures-based BGS method to detect the object region while blind user shakes it. Once the object of interest is extracted from the camera image, the system is ready to apply our automatic text extraction algorithm.

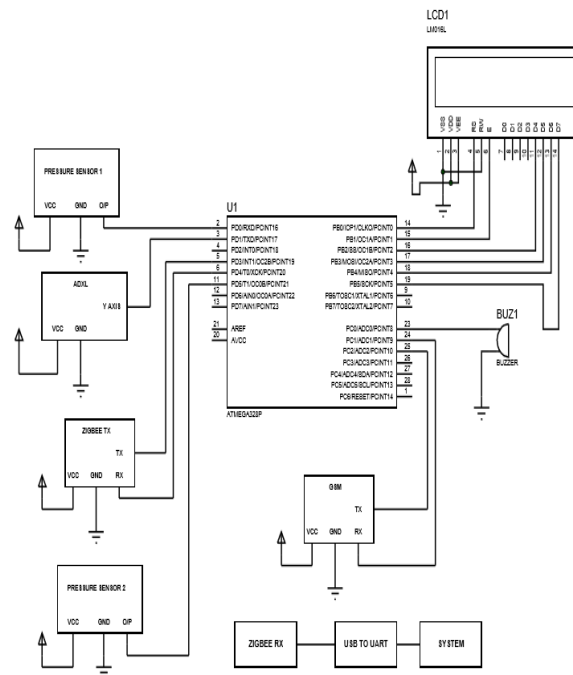


Fig.3. Block diagram

IV. 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 to perform binarization of text regions, where margin areas are always considered as background. We test both open- and closed-source solutions exist that have APIs 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 (SDK) 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.

V. 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 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.

REFERENCES

- [1] "10 facts about blindness and visual impairment", World Health Organization: Blindness and visual impairment, 2009 www.who.int/features/factfiles/blindness/blindness_facts/en/index.html
- [2] Advance Data Reports from the National Health Interview Survey, 2008. http://www.cdc.gov/nchs/nhis/nhis_ad.htm.
- [3] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011). <http://www.m.cs.osakafu-u.ac.jp/cbdar2011/>
- [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. 148-156, 1996.
- [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, Editors. 2008, Wiley: Hoboken, N.J.
- [10] A. Huete, J. Victores, S. Martinez, A. Gimenez, and C. Balaguer. Personal Autonomy Rehabilitation in Home Environment by a Portable Assistive Robot. In IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 42, No. 4, pp. 561-570, 2011.