Volume 1, Issue 9, December 2016 ISSN: 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

Portable Camera for Blind Person and Web Extraction

Idris Shaikh
Computer Department
Sir Visvesvaraya Institute Of Technology
Nashik, India

Priyanka Varandal
Computer Department
Sir Visvesvaraya Institute Of Technology
Nashik, India

Taranjum Inamdar
Computer Department
Sir Visvesvaraya Institute Of Technology
Nashik, India.

Sneha Sandhan
Computer Department
Sir Visvesvaraya Institute Of Technology
Nashik, India.

Abstract: ***We offer 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 identify the object from cluttered backgrounds or other neighboring objects in the camera view, we first offer an efficient and effective motion-based method to state a region of interest (ROI) in the video by asking the user to shake the object. This method remove moving object region by a mixture-of-Gaussians based background subtraction method. In the extracted ROI, text localization and recognition are conducted to obtain text information. To automatically localize 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 Ad boost model. Text characters in the restricted text regions are then binaries and identify by off-the-shelf optical character recognition (OCR) software. The identified text codes are output to blind users in speech. Performance of the proposed text localization algorithm is quantitatively explained on ICDAR-2003 and ICDAR-2011 Robust Reading Datasets. Experimental results clearly show that our algorithm reach the state-of-the-arts. The proof-of-concept prototype is also evaluated on a dataset collected using 10 blind persons, to describe the effectiveness of the system's hardware. We discus user interface issues, and assess robustness of the algorithm in extracting and reading text from different objects with complex backgrounds. Internet has brought about an incredible being improved in human access to knowledge and information. However, blind people face problems in retrieving these text materials. Web browsers for the visually handicapped people in the past have been limited to changing documents to Braille or speech, or removing text and filtering. However, the human aspects of web surfing for blind people have not been adequately addressed. This paper offers architecture of an open source, light weight web browser that forms it easy for the visually handicapped people to surf the web.***

Keywords: Web Browser for blind, content reorganization, speech, blindness, assistive devices, text reading, hand-held object, OCR.

I. INTRODUCTION

The National Census of India has estimated around 21.9 million disabled people in the country [3]. Out of which more than 15 million people in India are blind. This is considered to be the highest among all other disabilities . Due to their inability in accessing information from written text documents, blind people face tremendous problems in accessing information through web. Thus, in order to provide proper information access and to bridge the communication gap between the visually impaired and the sighted community, the need to build some advance technologically supported systems are utterly essential. Several attempts have been taken to build web browsers for the blinds [1]. Most of the web browsers do not support any automatic speech recognizer that will enable blind users to navigate a web page through speech commands. Most of the web browsers do not support any speech output to any navigation related operations performed by a blind user. Most of the web browsers for blind are integrated with text to speech engines, however, very few browsers support text to Braille representation that may be used by blind users to archive the web document in printed form. Although many web browsers are integrated with Text-to Speech technology. However, most of these browsers read out the information on a web page to the user in a sequential pattern. This creates problems during navigation to a blind person. There are two

aspects of web surfing - a) surfing the content of a page, b) navigation through these pages. When a sighted user views a page, he/she can get a brief idea about that page by a mere glance. The user can then look for his/her area of interest and then read that particular content or may decide to leave that page. He/she may choose to skip any amount of information in between to reach the desired area of interest. However, visually handicapped users may face a great disadvantage here. They will not have any idea about the current active page. A screen reader merely reads out the whole content of the page as is. But that is very inconvenient for the users as in that case they may not have any control over the information in the page and will have to listen to the screen reader all through as it reads out the whole page, to get an idea about what is present. The user should be given the flexibility to explore the page in his own desired fashion. However, without prior knowledge about the page, it is difficult for a blind user to do so. If the page contains forms or any place which requires user input, a sighted user can just click on that area and enter the information. But a visually handicapped person does not have this advantage. So it is also required to inform him about what is desired and where to input the information.

A. 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 texted 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 move the hand-held objects containing the text they wish to identify and then employ a motion-based method to localize the objects from fill background. Background subtraction (BGS) is a conventional and effective approach to detect moving objects for video surveillance systems with stationary cameras. To discover moving objects in a dynamic scene, many adaptive background subtraction techniques have been developed.

Fig. 1. Localizing the image region of the hand-held object of interest. (a) Capturing images by a camera mounted on a pair of sunglasses; (b) An 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.

B. Automatic Text Extraction

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

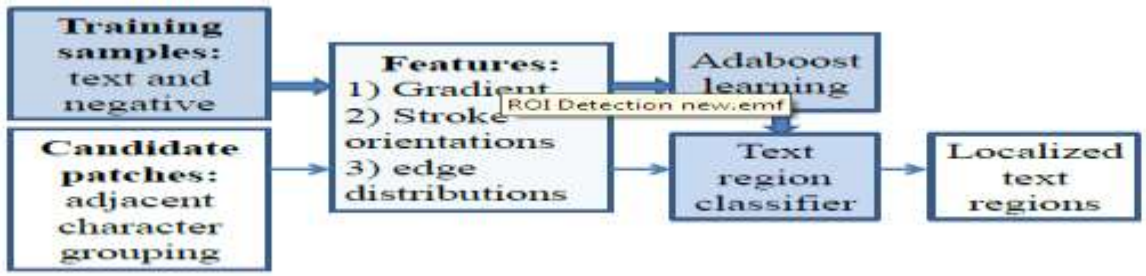

Fig. 2. Diagram of the proposed Adaboost learning based text region localization algorithm by using stroke orientations and edge distribut

II. RELATED WORK

A number of attempts were made worldwide to build web browsers and enhance usability of web for the visually handicapped people. In this paper we have thoroughly studied some of the most widely used web browsers for the blinds. Further, we have compared each of the browsers based on their key GUI features. Based on our study we consider the following GUI features to be essential for a web browser for the blinds.

- Open source architecture
- Text to speech output
- Navigations through voice feedbacks
- Speech recognition
- Text extraction and representation

• Text to Braille conversion
• Voice feedback for keyboard operations
• Mouse based gesture recognition

Features	WebbIE	eGuide Dog	Shruti Drishti	IBM Home Page Reader
Open Source	NO	YES	YES	NO
Speech input method	NO	NO	NO	NO
Mouse gesture recognition	NO	NO	NO	NO
Voice feedback for keyboard operations	NO	NO	NO	NO
Text extraction and filtering	YES	YES	YES	YES
Hierarchical representation	YES	YES	NO	NO
Text to Braille	NO	NO	YES	NO
text to speech	YES	YES	YES	YES

Table 1. Comparison Between Different Web Browsers For Visually Handicapped people

III. LITURATURE SURVEY

Depending on the purpose and content of a web page, these can be classified in various ways. It is important to understand and capture these because the presentation of a page to a blind person would depend on its type. We give below a broad classification with this in mind. This classification is based on the various formats followed by web page designers to design the pages based on what they want to present.
• Single article page
• Multiple article page
• E-mail
• Search Engine
• Portals
• Blogs
• Forms
• Social Networking
• Forums
• Online shops and auction web site
It may be noted that all these pages may contain multiple frames, tables, JavaScript, servlets, etc. These also have to be Identified and handled when the content of the web page is presented.

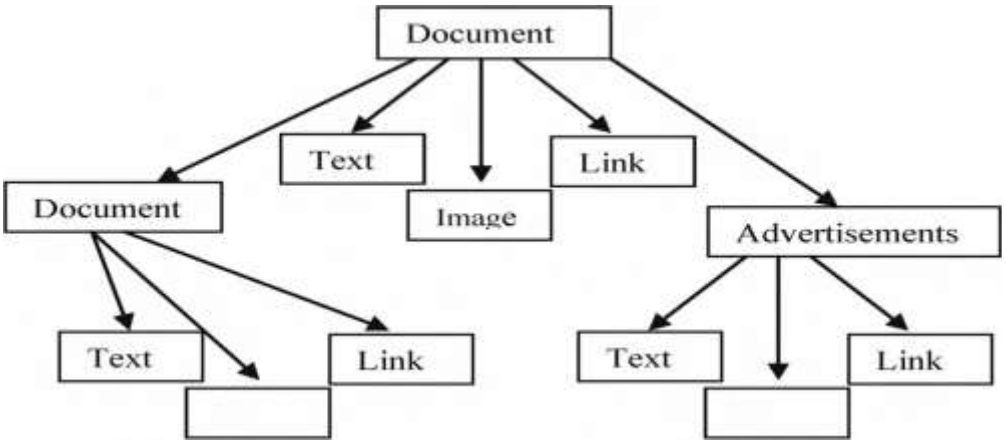

Figure 3: Hierarchical Representation of a Web Page

IV. PROPOSED SYSTEM

Our proposed system essentially consists of three different modules:
(1) User Input Module,
(2) Text Extraction and re-organization, and
(3) Output Representation.

A. User Input Module
The user input module deals with the different input methods. Operations on the browser, which are usually performed by mouse clicks, can be performed by speech commands. A speech recognizer is integrated with the browser, which is trained

to recognize all the commands. The system is speaker independent. Apart from browser operations, different navigational operations can also be performed by voice commands. For ease in data input using keyboard, an optional voice feedback for every key operation is provided by the system. Figure 4 shows the input module

B. Text Extraction and Re-organization

The extraction of text and presenting it to a visually handicapped person has many difficult aspects to it. With innumerous web pages present in the web, there is a varied diversity in the type of the pages. A prototype hierarchical structure of a web page is illustrated in figure 3. From the figure we can observe that, a web page may contain more than one kind of contents like, links, images, advertisements, and animations. These contents may not provide valuable information to a visually impaired person. Further, the document structure of an email page is also different from other page

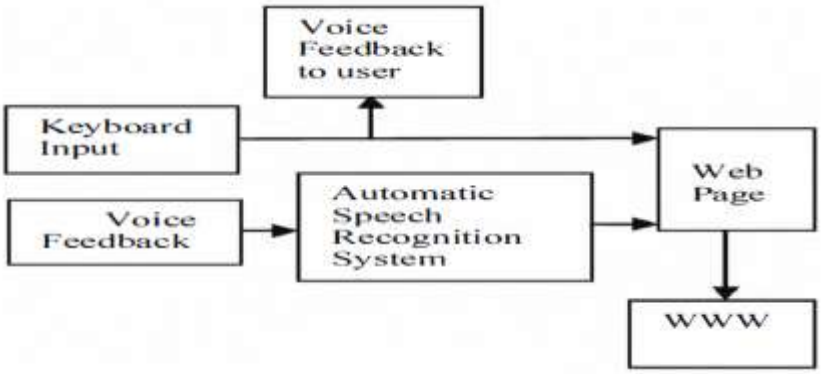

Figure 4: Block Diagram of the Input Module

In order to handle such kind of diversity in web pages, it is required to do a set of pre-processing on the corresponding pages before presenting it to the user. For that, the whole structure of the page along with its contents has to be extracted. This can be performed using an HTML parser. Using the parser, the syntactic information of a page is extracted from the html tags. For example, the header tags give information about the content following it and divider tags divide the pages into section which can be used to isolate different portions of the page as convenient. After getting an idea about the format of the page, the whole information in the page can be divided into sections and then presented to the user in a format convenient to them.

CONCLUSION AND FUTURE WORK

In this paper, we have explained a prototype system to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming difficulties for blind users, we have proposed a motion-based method to discover the object of interest while the blind user simply shakes the object for a couple of seconds. This method can effectively divide the object of interest from background or other objects in the camera view. To remove 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 evaluate 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.

REFERENCES

1. _http://www.synapseadaptive.com/wynn/IBM_Home_Page_Reader.htm I Indian Standard Code for Information Exchange.
2. Dasgupta T., Basu A., "A Speech Enabled Indian Language Text to Braille Transliteration System", International Conference on Information and Communication Technology for Development (ICTD),2009,
3. Chu, c.: Two Dimension Interactive Voice Browser for the VisuallyImpaired. In: K. Miesenberger eds. ICCHP 2004, LNCS 3118. Berlin: Springer,(2004),721-724.Qatar.
4. Chu, c.: Two Dimension Interactive Voice Browser for the Visually Impaired. In: K. Miesenberger eds. ICCHP 2004, LNCS 3118. Berlin: Springer, (2004), 721-724.
5. Liang, Z., Song, X., Zhu, Z., Liu, R.: Design and Implementation of a New Browser for Blind Person. Computer Engineering and Applications. (2004), 14, 106-108. [8] Mercedes Macias, Fernando Sanchez, Improving Web Accessibility for Visually Handicapped People Using KAI, Proceedings of the 3rd International Workshop on Web Site Evolution (WSE.OI)
6. A. Huete, J. Victores, S. Martinez, A. Gimenez, and C. Balaguer. Personal Autonomy Rehabilitation in Home Environment by a Protable Assistive Robot. In *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 42, No. 4, pp. 561-570, 2011.
7. ICDAR 2011 Robust Reading Competition: A. Shahab, F. Shafait, and A. Dengel. ICDAR Robust Reading Competition Challenge 2: Reading Text in Scene Images. In Proceedings of ICDAR, pp. 1491-1496, 2011.
8. K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using SVM and continuously adaptive mean shift algorithm," In *IEEE Trans. on PAMI*, Vol.25, No.12, pp.1631-1639, 2003.
9. "KReader Mobile User Guide", knfb Reading Technology Inc, http://www.knfbReading.com
10. S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model," In *IEEE Trans on Image Processing*, Vol. 16, No. 8, pp. 2117-2128, 2007.
11. J. Liang, D. Doermann, and H. Li, Camera-based analysis of text and documents: a survey. In *International Journal of Document Analysis and Recognition (IJDAR)*, No. 2-3, pp. 84-104, 2005.
12 R. Liu and Y. Wang, "Auditory Feedback and Sensory Substitution During Teleoperated Navigation," In *IEEE/ASME Transactions on Mechatronics*, Vol. 17, No. 4, pp. 680-686, 2011.