ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, January 2014

International Conference on Engineering Technology and Science-(ICETS'14)
On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

Vision Based Assistive System for Label Detection with Voice Output

Vasanthi.G[1] and Ramesh Babu.Y[2]
Department of ECE, DMI College of Engineering, Chennai, India[1, 2]

Abstract--A camera based assistive text reading framework to help blind persons read text labels and product packaging from hand-held object in their daily resides is proposed. To isolate the object from cluttered backgrounds or other surroundings objects in the camera view, we 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. 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.

Keywords-Assistive devices, blindness, distribution of edge pixels, hand-held objects, optical character recognition (OCR), stroke orientation, text reading and text region localization.

I. INTRODUCTION

Of the 314 million visually impaired people worldwide, 45 million are blind. 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.

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 labelling. 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 scanner might be employed in these and similar situations.

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. 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.

II SOFTWARE SPECIFICATIONS AND FRAMEWORK

2.1 Software Specifications

Operating System : Ubuntu 12.04

Language : C and C++

Platform : OpenCV (linux-library)

2.2 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 data processing component is used for deploying our proposed algorithms, including 1) object- of- interest detection to

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, January 2014

International Conference on Engineering Technology and Science-(ICETS'14)
On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

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.

III. IMAGE CAPTURING AND PRE-PROCESSING

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 haar 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.

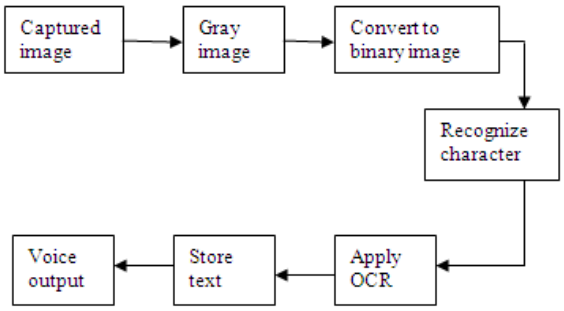

Fig 1: Block Diagram of Text Reading

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.

IV. 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 Adaboost based text classifier.

V. 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.

VI. 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 characters. However, this experiment show that OCR generates better performance text regions are first assigned proper margin areas and binarized 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.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, January 2014

International Conference on Engineering Technology and Science-(ICETS'14)
On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

VII. HARDWARE DESCRIPTION

7.1 Block Diagram

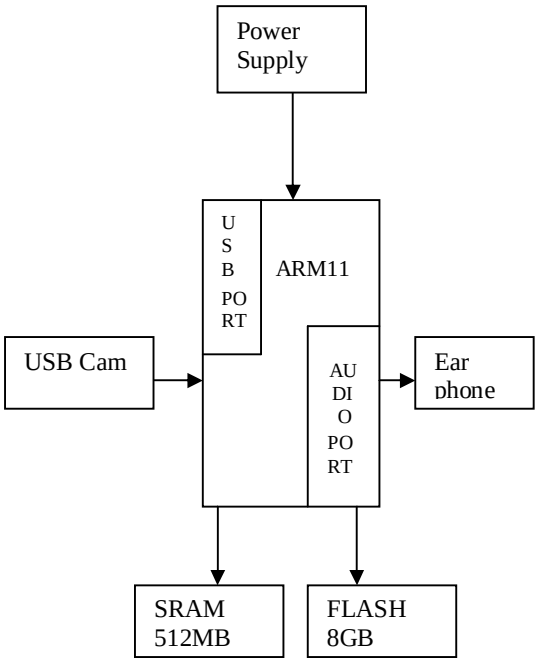

Fig 2: Block Diagram of Vision Based Assistive System

7.2 Block Diagram Description

The block diagram consists of ARM11 micro controller (BCM2835), USB cam, Power supply, Flash memory, SRAM and Earphone. The camera is connected to an ARM11 (BCM2835) by a USB connection and it can capture the hand-held object appears in the camera view. LAN9512 is interface with the system to view the monitor. SRAM is used for the temporary storage and flash memory is used for the permanent storage. The ARM11 (BCM2835) performs the processing and provides audio output through earphone.

7.3 Hardware Requirements

- Power Supply Unit
- USB cam
- SRAM
- ARM11 (BCM2835)
- Flash memory
- Ear phone

USB cameras are imaging cameras that use USB 2.0 or USB 3.0 technology to transfer image data. USB cameras are designed to easily interface with dedicated computer systems by using the same USB technology that is found on most computers. Static random-access memory (SRAM) is a type of a semiconductor memory that uses bi-stable latching circuitry to store each bit.

ARM11 features are, Supports 4-64k cache sizes, Powerful ARMV6 instruction set architecture, SIMD (Single Instruction Multiple Data) media processing extensions deliver up to 2x performance for video processing, and High-performance 64-bit memory system speeds data access for media processing and networking applications. LAN9512/LAN9512i contains an integrated USB 2.0 hub, two integrated downstream USB 2.0 PHYs, an integrated upstream USB 2.0 PHY, a 10/100 Ethernet PHY, a 10/100 Ethernet Controller, a TAP Controller and EEPROM Controller. Flash memory is an electronic non-volatile compute storage medium that can be an electrically erased and reprogrammed. Earphones either have wires for connection to a signal source such as an audio amplifier, radio, CD player, portable media player or have a wireless receiver, which is used to pick up signal without using a cable.

VIII. CONCLUSION AND FUTURE WORK

The proposed system ensures to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming problem for blind users, a motion-based method to detect the object of interest is projected, 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. 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.

The future development will be an obstacle detection process, using haar cascade classifier for recognizing the obstacle from the object. A novel camera based computer vision technology to automatically recognize currency to assist visually impaired people will be enhanced.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their constructive comments and insightful suggestions that improved the quality of this manuscript.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, January 2014

International Conference on Engineering Technology and Science-(ICETS'14)
On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

REFERENCES

[1] A.Shahab, F.Shafait, and A. Dengel, "Multi-script robust reading competition in ICDAR 2013" in *Proc. Int. Conf. Document Anal. Recognit*, 2013, pp. 1491-1496.

[2] C. Yi and Y. Tian, "Text string detection from natural scenes by structure based partition and grouping," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2594-2605, Sep. 2011.

[3] Jun Baba and Akihiro Yamamoto, "Text Localization in Scene Images by Using Character Energy and Maximally Stable Extremal Regions" *IEEE Trans Image Process*, vol. 9, pp. II-252-303, Sep. 2009.

[4] K. Kim, K. Jung, and J. Kim, "A Novel character Segmentation Method for text images Captured by Cameras" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631-1639, Dec. 2003.

[5] Chucai Yi and Ying Li Tian, "Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification", *IEEE Transactions on Image Processing*, vol. 13, No. 1, pp. 87-99, 2004.

[6] P. Viola and M.J. Jones, "Robust real-time face detection", *Int. J. Comput Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[7] 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.

[8] Wen Wu, Xilin Chen, and Jie Yang, "Detection of Text on Road Signs From Video", *IEEE Transaction on intelligent transportation systems*, vol. 6, no. 4, Dec 2005, vol. 6, no. 4, DEC 2005.

[9] 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.

[10] X. Yang, S.Yuan, and Y. Tian, "Recognizing clothes patterns for blind people by confidence margin based feature combination." in *Proc. ACM Multimedia, 211, pp. 1097-1100.*

[11] X. Yang, Y. Tian, C. Yi, and A. Arditi, "Context-based indoor object detection as an aid to blind persons accessing unfamiliar environments," in *Proc. ACM Multimedia*, 2010,pp. 1087-1090.

[12] Y. Tian, M. Lu, and A. Hampapur, "Robust and efficient foreground analysis for real-time video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit*, 2005, pp. 1182-1187.

[13] 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.

[14] 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.