

VISION BASED ASSISTIVE SYSTEM FOR LABEL DETECTION WITH VOICE OUTPUT

M. BEERENDAR¹, M. VIJAYA LAKSHMI²

¹ Student of M.Tech, Dept Of ECE, St. Martin's Engineering College, Hyderabad, Telangana, India.

² M.Tech, Associate professor, Dept Of ECE, St. Martin's Engineering College, Hyderabad, Telangana, India.

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 Ad boost model. Text characters in the localized text regions are then binaries 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.

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

localization, ARM9(S3C2440), Bluetooth, android mobile and MATLAB.

I. Introduction

OF the 314 million visually impaired people worldwide, 45 million are blind. Even in a developed country like the U.S., the 2008 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 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 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. Mobile 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), and contain mostly text. 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). Further more, 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.



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

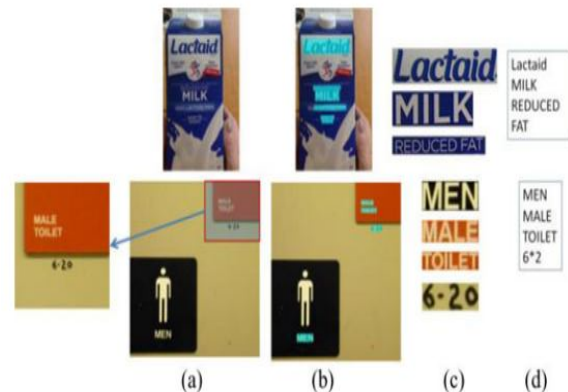


Fig: 2. Two examples of text localization and recognition from camera captured images. (Top) Milk box. (Bottom) Men bathroom signage. (a) camera captured images. (b) Localized text regions (marked in blue). (c) Text regions cropped from image. (d) Text codes recognized by OCR. Text at the top-right corner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

In assistive reading systems for blind persons, it is very challenging for users to position the object of interest within the center of the 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. analyzed edge pixel density with the Laplacian operator and employed maximum gradient differences to identify text regions are used gradient difference maps and performed global binarization to obtain text regions.

These are designed stroke width transforms to localize text characters, applied color reduction to extract text in uniform colors. In 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. There are five types of Haar based block patterns to train text classifiers in an Ad boost learning model. Considered text as a specific texture and analyzed the textural features of characters by a Support vector machine (SVM) model. They are used

Globally matched wavelet filter responses of text structure as features.

It performed classification of text edges by using histograms of oriented gradients and local binary patterns as local features on the SVM model. 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 .A survey paper about computer-vision-based assistive technologies to help people with visual impairments can be found in. 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.

II. The Hardware System

Micro controller:

This section forms the control unit of the whole project. This section basically consists of a Microcontroller with its associated circuitry like Crystal with capacitors, Reset circuitry, Pull up resistors (if needed) and so on. The Microcontroller forms the heart of the project because it controls the devices being interfaced and communicates with the devices according to the program being written.

S3C2440A:

S3C2440A is a 16/32-bit RISC microprocessor. SAMSUNG's S3C2440A is designed to provide hand-held devices and general applications with low-power, and high-performance microcontroller solution in small die size. To reduce total system cost, the S3C2440A includes the following components. The S3C2440A is developed with ARM920T core, 0.13um CMOS standard cells

and a memory complier. Its low power, simple, elegant and fully static design is particularly suitable for cost- and power-sensitive applications. It adopts a new bus architecture known as Advanced Micro controller Bus Architecture (AMBA). The S3C2440A offers outstanding features with its CPU core, a 16/32-bit ARM920T RISC processor designed by Advanced RISC Machines, Ltd.

Liquid-crystal display (LCD) is a flat panel display, electronic visual display that uses the light modulation properties of liquid crystals. Liquid crystals do not emit light directly. LCDs are available to display arbitrary images or fixed images which can be displayed or hidden, such as preset words, digits, and 7-segment displays as in a digital clock. They use the same basic technology, except that arbitrary images are made up of a large number of small pixels, while other displays have larger elements.

II. Design of Proposed Hardware System

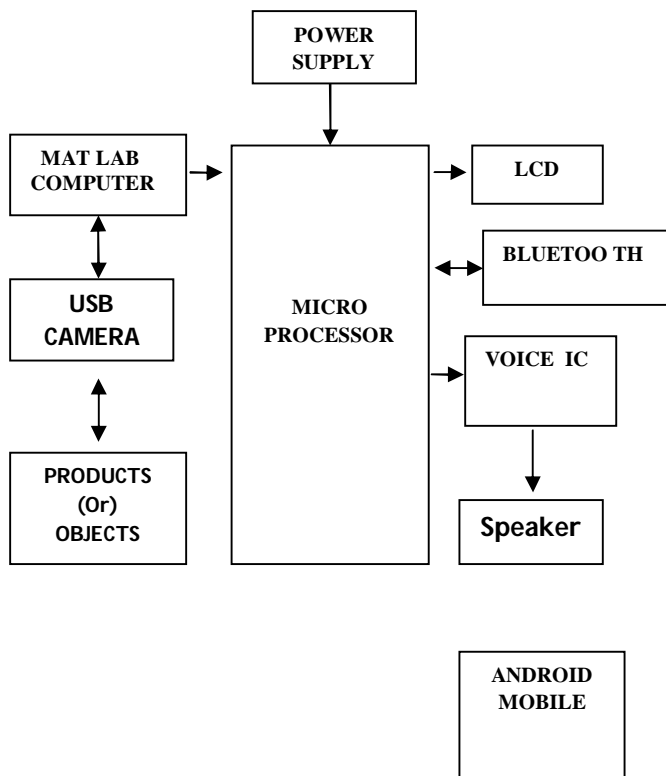


Fig.3.Block diagram

III. Board Hardware Resources Features

Bluetooth:

Bluetooth was selected as our way of communicating PDA/Mobile with a central system. The reason Bluetooth was selected over Wi-Fi for various reasons. First of all, Bluetooth security is less complex and more stable than that of Wi-Fi. Bluetooth manages a security measure of only permitting certain selected devices to interact with them; Wi-Fi in the other hand establishes a WEP key that has been known to be cracked. Another reason that Bluetooth was selected over Wi-Fi is that Bluetooth has a shorter range of signal emission than Wi-Fi. This is a pro because the shorter the range the less the amount intruders that will try to infiltrate your home system



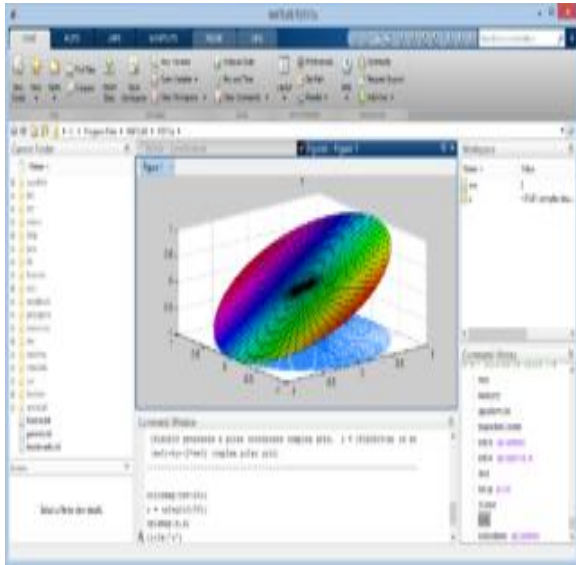
AUBTM-22 is a Bluetooth v1.2 module with SPP profiles. The module is intended to be integrated into another HOST system which requires Bluetooth functions. The HOST system could send commands to AUBTM-22 through a UART. AUBTM-22 will parse the commands and execute proper functions, e.g. set the maximum transmit power, change the name of the module. And next the module can transmit the data receive from the uart with SPP profiles.

MATLAB COMPUTER:

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and fourth generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, Fortran and Python. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain

simulation and Model-Based Design for dynamic and embedded systems.

In 2004, MATLAB had around one million users across industry and academia. MATLAB users come from various backgrounds of engineering, science, and economics. MATLAB is widely used in academic and research institutions as well as industrial enterprises.



USB CAMERA:



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. The accessibility of USB technology in computer systems as well as the 480 Mb/s transfer rate of USB 2.0 makes USB Cameras ideal for many imaging applications. An increasing selection of USB 3.0

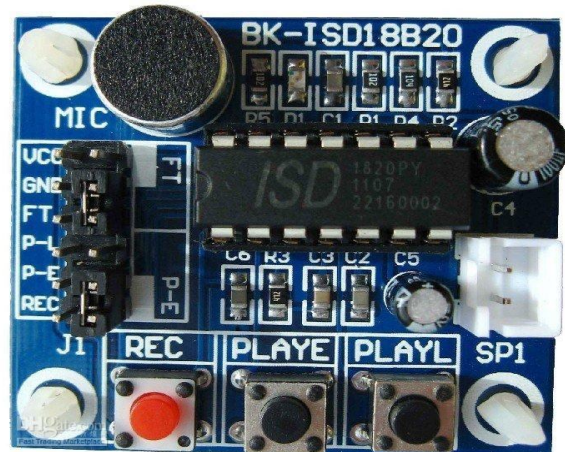
Cameras is also available with data transfer rates of up to 5 Gb/s. Edmund Optics offers a variety of USB Cameras suited to meet many imaging needs. EO USB Cameras are available in both CMOS as well as CCD sensor types making them suitable across a larger range of applications. USB Cameras contain out-of-the-box functionality for quick setup. USB Cameras using low power USB ports, such as on a laptop, may require a separate power supply for operation.

PRODUCTS (OR) OBJECTS:



VOICE IC:

A **voice IC** (or) **sound chip** is an integrated circuit (i.e. "chip") designed to produce sound. It might be doing this through digital, analog or mixed-mode electronics. Sound chips normally contain things like oscillators, envelope controllers, samplers, filters and amplifiers. During the late 20th century, sound chips were widely used in arcade game system boards, video game consoles, home computers, and PC sound cards.



SPEAKER:

A speaker driver is an individual transducer that converts electrical energy to sound waves, typically as part of a loudspeaker, television, or other electronics device. Sometimes the transducer is itself referred to as a speaker, particularly when a single one is mounted in an enclosure or as surface-mounted device (as in a wall-mounted speaker, car audio speaker, and so on). There are many different types of speaker drivers. The most common ones are the woofer, mid-range and tweeter, as well as subwoofers which are becoming very common. Less common types of speaker drivers are super tweeters and rotary woofers.

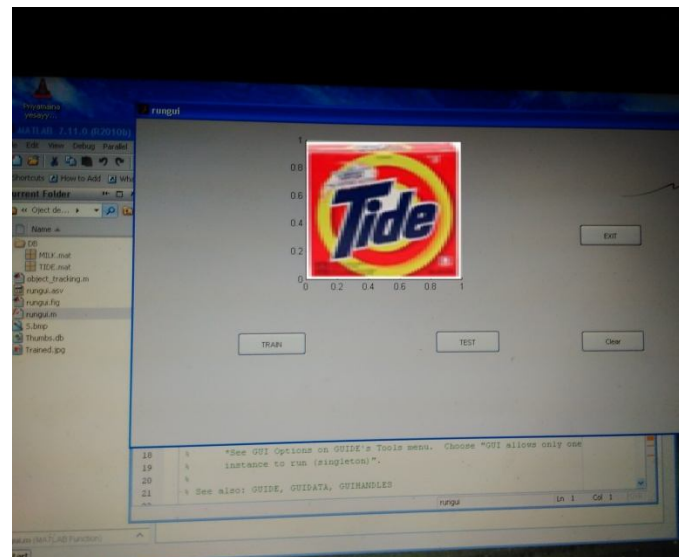


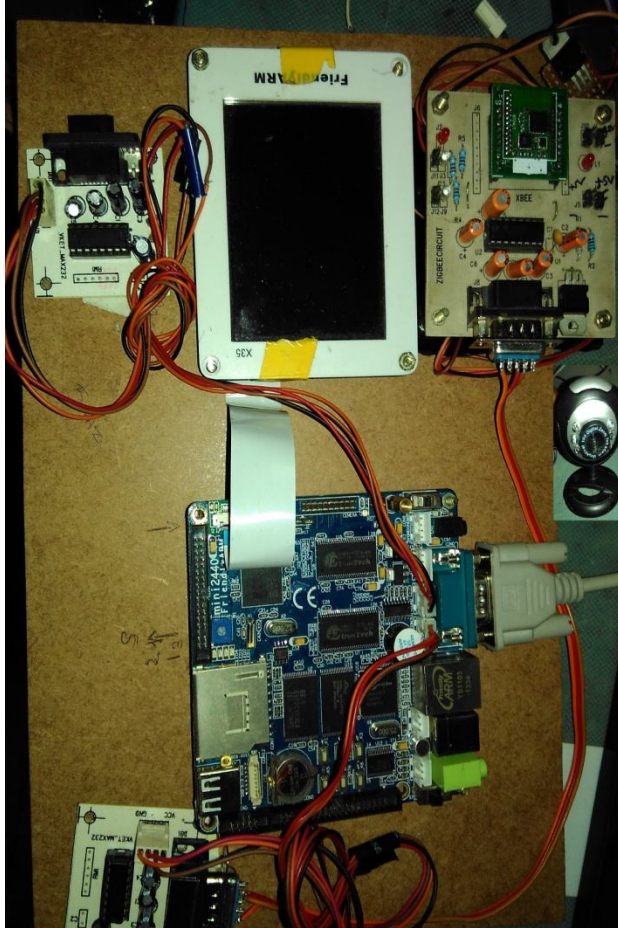
ANDROID MOBILE:

A smart phone (or Android) is a mobile phone with an operating system. Smart phones typically include the features of a phone with those of another popular consumer device, such as a personal digital assistant, a media player, a digital camera, and/or a GPS navigation unit. Later smart phones include all of those plus the features of a touch screen computer, including web browsing, Wi-Fi, 3rd-party apps, motion sensor, mobile payment and 3G.



RESULTS





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.

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.

CONCLUSION



[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.

[8] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Machine Learning*, 1996, pp. 148–156.

[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, Eds. Hoboken, NJ, USA: Wiley, 2008.

[10] A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition: ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in *Proc. Int. Conf. Document Anal. Recognit.*, 2011, pp. 1491–1496.

[11] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[12] KReader Mobile User Guide, knfb Reading Technology Inc. (2008). [Online]. Available: <http://www.knfbReading.com>

[13] 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

[14] S. M. Lucas, "ICDAR 2005 text locating competition results," in *Proc. Int. Conf. Document Anal. Recognit.*, 2005, vol. 1, pp. 80–84 department Jawaharlal Nehru Technological University, Hyderabad.

AUTHORS



M. Vijaya Lakshmi received her B.Tech degree in Electronics & Communication Engineering from JNTU, Hyderabad in 2009. Presently, she is working as Associate Professor in ECE department in St. Martin's Engineering College, Hyderabad.



M. Beerendar was born in Telangana, India, in 1991. He received B.Tech degree in Electronics and Communication Engineering from Jawaharlal Nehru Technological University, Hyderabad, Telangana, India. He is currently working towards the M.Tech degree in the department of Electronics and Communication Engineering from Jawaharlal Nehru Technological University, Hyderabad.