Portable Camera Based Assistive Text and Product Label Reading from Handheld Objects for Blind People

M.Krushnaveni[1], Mrs.M.Sunitha Rani[2],
[1]PG Scholar, Vidya Jyothi Institute of Technology, Aziznagar, Hyderabad-500075,India
[2]Assistant Professor, Vidya Jyothi Institute of Technology, Aziznagar, Hyderabad-500075, India

Abstract: Assistive technologies are being developed for visually impaired people in order to live confidently. This project work proposes 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. The project work is framed into three stages. First, Image capturing – Using a mini camera ,the text which the user need to read get captured as an image and have to send to the image processing Platform. Secondly, Text recognition – Using text recognition algorithm, the text will get filtered from the image. Finally, Speech output - A filtered text will be passed into this system to get an audio output. This project work can be able to insist the blind people in their daily life. The entire application will run on Raspberry Pi.

The Raspberry Pi is a credit card sized single computer or SoC uses ARM1176JZF-S core. SoC, or System on a Chip, is a method of placing all necessary electronics for running a computer on a single chip. Instead of having an individual chip for the CPU, GPU, USB controller, RAM everything is compressed down into one tidy package. Raspberry Pi needs an Operating system to start up. In the aim of cost reduction, the Raspberry Pi omits any on-board non-volatile memory used to store the boot loaders, Linux Kernels and file systems as seen in more traditional embedded systems. Rather, a SD/MMC card slot is provided for this purpose. After boot load, as per the application program Raspberry Pi will get execute.

Keywords: Raspberry pi, OCR, espeak

1. Introduction:

Now a day's visually impaired peoples are increases, 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. Tesseract-ocr is an open-source OCR engine that was developed at HP between 1984 and 1994. Like a supernova, it appeared from nowhere for the 1995 UNLV Annual Test of OCR Accuracy, shone brightly with its results, and then vanished back under the same cloak of secrecy under which it had been developed. Now for the first time, details of the architecture and algorithms can be revealed. Tesseract began as a PhD research project in HP Labs, Bristol, and gained momentum as a possible software and/or hardware add-on for HP's line of flatbed scanners. Motivation was provided by the fact that the commercial OCR engines of the day were in their infancy, and failed miserably on anything but the best quality print. After a joint project between HP Labs Bristol, and HP's scanner division in Colorado, Tesseract had a significant lead in accuracy over the commercial engines, but did not become a product. The next stage of its development was back in HP Labs Bristol as an investigation of OCR for compression. Work concentrated more on improving rejection efficiency than on base-level accuracy. At the end of this project, at the end of 1994, development ceased entirely. The engine was sent to UNLV for the 1995 Annual Test of OCR Accuracy, where it proved its worth against the commercial engines of the time. In late 2005, HP released Tesseract-ocr for open source. It is now available at http://code.google.com/p/tesseract-ocr.

2. Literature survey:

ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Optical character recognition (also optical character reader, OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast). It is widely used as a form of information entry from printed paper data records, whether passport documents, invoices, bank statements, computerised receipts, business cards, mail, printouts of static-data, or any suitable documentation.

It is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

A number of portable reading assistants have been designed specifically for the visually impaired [2], [6], [10], [11], [12], [14], [17], [9], [13]. *KReader Mobile* runs on a cell phone and allows the user to read mail, receipts, fliers, and many other documents [12]. However, the document to be read must be nearly flat, placed on a clear, dark surface (i.e., a noncluttered background), and contain mostly 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 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,

System architecture:

Here, in this block diagram the whole system is controlled by Arm11 processor and this processor is implemented on Raspberry Pi Board. The system consists of Raspberry pi, Camera, SD card and personal computer. Those all components are connected by USB adaptors. Raspberry pi is the key element in processing module. First, Image capturing - Using camera image to be taken. Secondly Text recognition-Using text recognition algorithm text to be monitored. Finally Speech output the text content is changed into speech output.

Part of the recognition process for any character recognition engine is to identify how a word should be segmented into characters. The initial segmentation output from line finding is classified first. The rest of the word recognition step applies only to non-fixed pitch text. Processing follows a traditional step-by-step pipeline, but some of the stages were unusual in their day, and possibly remain so even now. The first step is a connected component analysis in which outlines of the components are stored. This was a computationally expensive design decision at the time, but had a significant advantage: by inspection of the nesting of outlines, and the number of child and grandchild outlines, it is simple to detect inverse text and recognize it as easily as black-on-white text. Tesseract-ocr was probably the first OCR engine able to handle white-on-black text so trivially. At this stage, outlines are gathered together, purely by nesting, into *Blobs*. Blobs are organized into text lines, and the lines and regions are analyzed for fixed pitch or proportional text. Text lines are broken into words differently according to the kind of character spacing. Fixed pitch text is chopped immediately by character cells. Proportional text is broken into words using definite spaces and fuzzy spaces.

Recognition then proceeds as a two-pass process. In the first pass, an attempt is made to recognize each word in turn. Each word that is satisfactory is passed to an adaptive classifier as training data. The adaptive classifier then gets a chance to more accurately recognize text lower down the page. Since the adaptive classifier may have learned something useful too late to make a contribution near the top of the page, a second pass is run over the page, in which words that were not recognized well enough are recognized again. A final phase resolves fuzzy spaces, and checks alternative hypotheses for the x-height to locate small cap text.

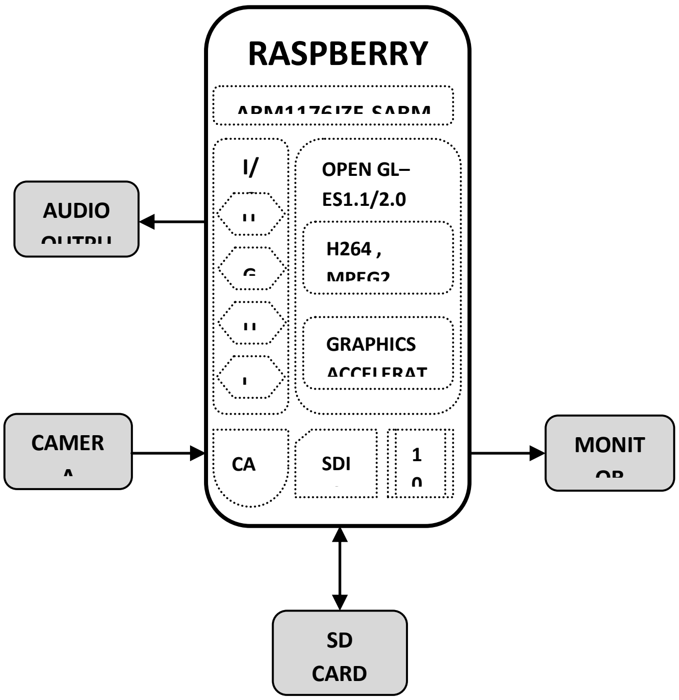

Figure: system architecture for product label reading

Raspberry pi:

A Raspberry Pi is a thirty five dollar, credit card sized computer board which when plugged into an LCD and attachment of a keyboard and a mouse, it is able to complete the functions of any regular PC can. Like a PC, it has RAM, Hard Drive (SD Card), Audio and Video ports, USB port, HDMI port, and Ethernet port. With the Pi, users can create spread sheets, word-processing, browse the internet, play high definition video and much more. It was designed to be a cost friendly computer for users who needed one. There are two models, Model A,B and 3. Model 3 is the faster containing 1GB of RAM as well as the ability to over clock.

Tesseract-ocr:

Tesseract is an optical character recognition engine for various operating systems. It is free software, released under the Apache License, Version 2.0 and development has been sponsored by Google since 2006. In 2006 Tesseract was considered one of the most accurate open-source OCR engines then available. Matrix Matching converts each character into a pattern within a matrix, and then compares the pattern with an index of known characters. Its recognition is strongest on monotype and uniform single column pages. This method defines each character by the presence or absence of key features, including height, width, density, loops, lines, stems and other character traits. Feature extraction is a perfect approach for OCR of magazines, laser print and high quality images.

Espeak

espeak is a compact open source software speech synthesizer for English and other languages, for Linux and Windows. espeak uses a "formant synthesis" method. This allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds, but is not as natural or smooth as larger synthesizers which are based on human speech recordings. The espeak speech synthesizer supports several languages, however in many cases these are initial drafts and need more work to improve them. Assistance from native speakers is welcome for these, or other new languages

Conclusion:

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. Our future work will extend our localization algorithm to process text strings with characters fewer than three and to design more robust block patterns for text feature extraction. We will also extend our algorithm to handle no horizontal text strings. Furthermore, we will address the significant human interface issues associated with reading text by blind users.

References

[1]. S.V. Rice, F.R. Jenkins, T.A. Nartker, The Fourth AnnualTest of OCR Accuracy, Technical Report 95-03, Information Science Research Institute, University of Nevada, Las Vegas, July 1995.

[2]. R.W. Smith, The Extraction and Recognition of Text from Multimedia Document Images, PhD Thesis, University of Bristol, November 1987.

[3]. R. Smith, "A Simple and Efficient Skew Detection Algorithm via Text Row Accumulation", Proc. of the 3rd Int. Conf. on Document Analysis and Recognition (Vol. 2), IEEE 1995, pp. 1145-1148.

[4]. World Health Organization. (2009). 10 facts about blindness and visual impairment [Online]. Available: www.who.int/features/factfiles/ blindness/blindness_facts/en/index.html

[5]. Advance Data Reports from the National Health Interview Survey (2008). [Online]. Available: http://www.cdc.gov/nchs/nhis/nhis_ad.htm

[6]. 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/

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

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

[12]. 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.

[13]. 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.

[14]. 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.

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

[16]. 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.

[17]. S. M. Lucas, "ICDAR 2005 text locating competition results," in Proc. Int. Conf. Document Anal. Recognit., 2005, vol. 1, pp. 80–84.

[18]. 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.