International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-4 Issue-7 December-2012

Character localization and recognition application for Smartphone

Snehal Charjan[1], R. V. Mante[2], P. N. Chatur[3]
M.Tech 2nd year, Shegaon, Amravati (Maharashtra) India[1]
Assistant Professor, CSE, Department, Government College of Engineering, Amravati (Maharashtra) India[2]
HOD, Science & Engineering Department, Government College of Engineering, Amravati, Maharashtra,India[3]

Abstract

Smart Phones have Internet access anywhere. The automatic text localization and recognition of text within a natural image is very useful for many problems. Once identified, the text can be used for many purposes. User can get current information about the product, place or boards. More exciting applications can be developed over the text extraction method with a high performance while also being computationally inexpensive. There are various methods proposed for Text Localization, text area segmentation, sign recognition and translation, Optical Character Recognition. In this paper we have described these methods. We have also compared all methods on the basis of performance and accuracy. Finally we concluded some good methods for Smartphone OCR application.

Keywords

Text localization, text extraction, segmentation, OCR

1. Introduction

Optical Character Recognition (OCR) is branch of Pattern Recognition domain where text and characters in images are recognized and separated and converted into editable text. Until a few decades ago, research in the field of Optical Character Recognition (OCR) was limited to document images acquired with flatbed desktop scanners. The usability of such systems is limited as they are not portable because of large size of the scanners and the need of a computing system. Recently, with the advancement of processing speed and internal memory of hand-held mobile devices such as high-end cell-phones, Personal Digital Assistants (PDA), smart phones, iPhones, iPods, etc. having built-in digital cameras, a new trend of research has emerged into picture. Researchers have dared to think of running OCR applications on such devices for having real time results[6].

The processing speed and memory size of handheld devices are not yet sufficient enough so as to run desktop based OCR algorithms that are computationally expensive and require high amount of memory [6]. The processing speeds of mobile devices with built-in camera start with as low as few MHz to as high as 624 MHz. The handset 'Nokia 6600' with an in-built VGA camera contains an ARM9 32-bit RISC CPU having a processing speed of 104 MHz [1]. The PDA 'HP iPAQ 210' has Marvell PXA310 type processor that can compute up to 624 MHz [2]. Some mobile devices have dual processors too. For instance, 'Nokia N95 8GB' has a 'Dual ARM-11 332 M Hz processors' [3]. Processing speed of other mobile phones and PDAs are usually in between them.

Besides the lower computing speed, these devices provide limited caching. Random Access Memory (RAM) which is frequently referred to as internal memory in case of mobile devices is usually 2-128 MB. Among the mobile devices with high amount of RAM, are the PDA 'HP iPAQ 210' has a 128 MB SDRAM [2]. The cell-phone 'Nokia N95 8GB' has a 128 MB internal memory [4]. Compared to desktop computers, this much of memory is very less. Therefore, need is immensely felt to design computationally efficient and lightweight OCR algorithms for handheld mobile devices.

Many objects in natural images, such as tree branches or electrical wires, are easily confused for text by existing optical character recognition (OCR) algorithms. For this reason, applying OCR on an unprocessed natural image is computationally expensive and may produce erroneous results. Hence, robust and efficient methods are needed to identify the text containing regions within natural images before performing OCR [5].

Recognition is often followed by a post-processing stage. Applying efficient post-processing techniques, the accuracy will be higher and then it could be directly implemented on mobile devices [6].

2. Text Localization

Robust and efficient methods are needed to identify the text containing regions within natural images before performing OCR. Several approaches for automatic detection and localization of text in images and videos have been proposed [7]–[9]. These algorithms mainly focus on the features of the text itself, such as edges, corners, strokes, color and texture distribution.

For low-power mobile device, the detection and Localization algorithm must be implementable with a small number of relatively simple operations. In 2012, Katherine L. Bouman et al. [5] proposed an approach that uses a multiscale search technique to quickly rule out large regions of the image that are not homogenous is very useful, thereby dramatically reducing computation. The algorithm relies on a fundamental feature of text: text is usually surrounded by a contrasting, uniform back-ground. The proposed method of text segmentation searches for the text's background rather than the actual text. This allows for a large variation in the distribution of text features while requiring little computation. The algorithm has proved to have a high performance while also being computationally inexpensive.

3. Optical Character Recognition

A number of research works on mobile OCR systems have been found. In 2006, Laine et al. [20] developed a system for only English capital letters. At first, the captured image is skew corrected by looking for a line having the highest number of consecutive white pixels and by maximizing the given alignment criterion. Then, the image is segmented based on X-Y Tree decomposition and recognized by measuring Manhattan distance based similarity for a set of centroid to boundary features. However, this work addresses only the English capital letters and the accuracy obtained is not satisfactory for real life applications.

Luo et al. of Motorola China Research Center have presented camera based mobile OCR systems for camera phones in [21] –[22]. In [21], a business card image is first down sampled to estimate the skew angle. Then the text regions are skew corrected by that angle and binarized thereafter. Such text regions are segmented into lines and characters, and subsequently passed to an OCR engine for recognition. The OCR engine is designed as a two layer template based classifier. A similar system is presented in [22] for Chinese-English mixed script business card images. In 2005, Koga et al. [23] has presented an outline of a prototype Kanji OCR for recognizing machine printed Japanese texts and translating them into English. Moreover, research in developing OCR systems for mobile devices is not limited to document images only. In 2006, Shen at el. [24], worked on reading LCD/LED displays with a camera phone.

These studies reflect the feasibility and make a strong indication that OCR systems can be designed for handheld devices. But, of course, the algorithms deployed in these systems must be computation friendly. They should be computationally efficient and low memory consuming [6].

A lot of OCR software has been developed to accomplish text extraction. Tesseract, originally developed as proprietary software at Hwelett-Packard between 1985 and 1995, now sponsored by Google, is considered to be one of the most accurate open source OCR engine currently available. It is capable of recognizing text in variety of languages in a binary image format [10].

A complete OCR system has been presented by A.F.Mollah et al. [6] compared to Tesseract, acquired recognition accuracy (92.74%) is good enough. Experiments shows that the recognition system presented in this paper is computationally efficient which makes it applicable for low computing architectures such as mobile phones, personal digital assistants (PDA) etc.

4. Text Correction

In 2012, Wolf Garbe [19] gave Text Correction algorithms based on Edit distance. They try to find the dictionary entries with smallest edit distance from the query term. Three ways to search for minimum edit distance in a dictionary. First is Naive approach. In this approach, edit distance from the query term to each dictionary term is computed, before selecting the string(s) of minimum edit distance as spelling suggestion. So this is exhaustive search and is inordinately expensive. The performance can be significantly improved by terminating the edit distance calculation as soon as a threshold of 2 or 3 has been reached. Second is Peter Norvig Approach. It generate all possible terms with an edit distance <=2 (deletes + transposes + replaces + inserts) from the query term and search them in the dictionary. For a word of length n, an alphabet size a, an edit distance d=1, there will be n deletions, n-1 transpositions, a*n alterations, and a*(n+1) insertions, for a total of

2n+2an+a-1 terms at search time. This is much better than the naive approach, but still expensive at search time. Another one is Symmetric Delete Spelling Correction Approach that generate terms with an edit distance <=2 (deletes only) from each dictionary term and add them together with the original term to the dictionary. This has to be done only once during a pre-calculation step. The cost of this approach is the pre-calculation time and storage space of n deletes for every original dictionary entry, which is acceptable in most cases.

5. Experimental Analysis

As above, there are many methods proposed to localize text in image, segment text, recognize characters and text correction. But for OCR application on Smartphone has to be computationally efficient and lightweight OCR algorithms for handheld mobile devices. Also accuracy should be high.

5.1 Result

Table 1 lists the complete results of each method for accuracy measure. In this table percentage of accuracy of each recognition method is mentioned in different column along with researchers name as column name. Here, LCSD&TD - Low Complexity Sign Detection and Text Localization IS approach by Katherine L.Bouman, AATD - adaptive algorithm for text detection is approach by Gao and J.Yang.

5.2. Discussion

On the basis of above Table 1 we have compared the accuracy between tesseract [17] and other approaches. This shows that tesseract having highest accuracy and suitable for Smartphone application as most of the work is done on the server, instead of on the phone, because of the high complexity cost. Adaptive algorithm for text detection by Gao and J. Yang also has better accuracy than others and can be used for Smartphone application.

Table 1: Accuracy Results on Dataset based on recognition methods

Methods	Accuracy
LCSD&TD	92.74
AATD	93.3
Tesseract	93.51
Kim	81.5
Chen	87.1
X. Huang	90.2
Laplacian	85
X. Chen and A. L. Yuille	90

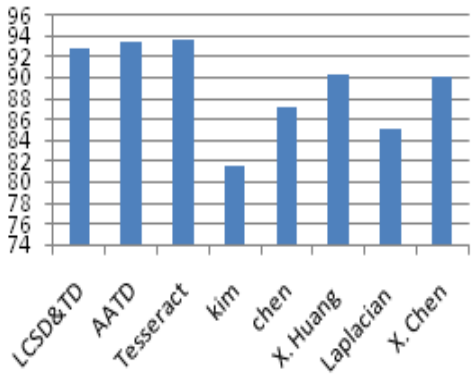

Figure 1: Accuracy of methods

6. Conclusions

This paper presented the methods for various methods proposed for Text Localization, text area segmentation, sign recognition and translation, Optical Character Recognition. Comparing all methods shows tesseract has highest accuracy. A lot of OCR software have been developed. Tesseract, sponsored by Google, is one of the most accurate open source OCR engine currently available. It is capable of recognizing text in variety of languages in a binary image format. Also Adaptive algorithm for text detection by Gao and J. Yang also has better accuracy and can be used for Smartphone application implementation. Future work includes implementing Smartphone application for localizing text in image using LCSD&TD and recognizing localized text using Tesseract which can be used for various functions.

References

[1] http://www.gsmarena.com/nokia_6600-454.php.
[2] http://h10010.www1.hp.com/wwpc/us/en/sm/WF06a/215348-215348-64929-314903215384-3544499.html.
[3] http://www.gsmarena.com/nokia_n95_8gb-2088.php.
[4] http://www.nokiaasia.com/findproducts/product/nokia-n95 8gb/technicalspecifications.
[5] Katherine L. Bouman, Golnaz Abdollahian,"A Low Complexity Sign Detection and Text Localization Method for Mobile Applications", IEEE Transactions on multimedia, VOL.13, NO. 5, October 2011.
[6] Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basuand Mita Nasipuri "Design of an Optical Character Recognition System for Camera-based Handheld Devices" IJCSI

International Journal of Computer Science Issues, Vol. 8, Issue 4, No1, July 2011.
[7] S. A. R. Jafri, M. Boutin, and E. J. Delp, "Automatic text area segmen-tation in natural images," inProc. ICIP, pp. 3196–3199, 2008.
[8] J. Yang, J. Gao, Y. Zhang, X. Chen, and A. Waibel, "An automatic sign recognition and translation system," inProc. 2001 Workshop Perceptive User Interfaces (PUI '01), pp. 1–8, 2001.
[9] Gao and J. Yang, "An adaptive algorithm for text detection from natural scenes," inProc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, vol. 2, 2001.
[10] Derek Ma, Qiuhau Lin, Tong Zhang "Mobile Camera Based Text Detection and Translation" Stanford University, Nov 2000.
[11] 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.
[12] K. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, "IEEE Trans. Pattern Anal. Mach. In-tell., vol. 25, no. 12, pp. 1631–1638, Dec. 2003.
[13] X. Huang and H. Ma, "Automatic detection and localization of natural scene text in video," inProc. 2010 20th Int. Conf. Pattern Recognition (ICPR), pp. 3216–3219, 2010.
[14] P. Shivakumara, T. Q. Phan, and C. L. Tan, "A Laplacian approach to multi-oriented text detection in video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412–419, Feb. 2011.
[15] E. Haneda and C. Bouman, "Multiscale segmentation for MRC docu-ment compression using a Markov random field model," inProc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP), pp. 1042–1045, Mar.2010.
[16] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," Proc. Computer Vision and Pattern Recognition (CVPR) Conf., vol. 2, pp. 366–373, 2004.
[17] Ray Smith,"An Overview of the Tesseract OCR Engine," Google Inc. IEEE 0-7695-2822-8/07, 2007.
[18] http://developer.android.com/about/index.html.
[19] http://blog.faroo.com/2012/06/07/improved-edit-distance-based-spelling-correction/.
[20] Mikael Laine and Olli S. Nevalainen, "A standalone OCR system for mobile camera-phones", Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium, pp.1-5, Sept. 2006.
[21] Xi-Ping Luo, Jun Li and Li-Xin Zhen, "Design and implementation of a card reader based on build-in camera", International Conference on Pattern Recognition, pp. 417-420, 2004.
[22] Xi-Ping Luo, Li-Xin Zhen, Gang Peng, Jun Li and Bai-Hua Xiao, "Camera based mixed-lingual card reader for mobile device", International Conference on Document Analysis and Recognition, pp. 665-669, 2005.
[23] Masashi Koga, Ryuji Mine, Tatsuya Kameyama, Toshikazu Takahashi, Masahiro Yamazaki and Teruyuki Yamaguchi, "Camera-based Kanji OCR for Mobile-phones: Practical Issues", Proceedings of the Eighth International Conference on Document Analysis and Recognition, pp. 635-639, 2005.
[24] Shen, H. and Coughlan, J., "Reading LCD/LED Displays with a Camera Cell Phone", Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, 2006.

Snehal Charjan received her B.E. degree in Computer science and Engineering, from G.H.Raisoni College of Engineering, Nagpur, Maharashtra, India, in 2011, pursuing M.tech Degree in Computer Science and Engineering from Government college of Engineering, Amravati, Maharashtra, India. Her research interests include Pattern recognition, Optical Character recognition. At present, she is engaged in Character localization and recognition application for Smartphone.

Ravi V. Mante received his B.E. degree in Computer science and Engineering, from Government college of Engineering, Amravati, Maharashtra, India in 2006, the M.Tech Degree in Computer science and Engineering, from Government college of Engineering, Amravati, Maharashtra, India, in 2011. He is Assistant professor in Government college of Engineering, Amravati, Maharashtra, India since 2007. His research interests include ECG signal analysis, soft computing technique, cloud computing. At present, He is working with Artificial Neural Network.

P. N. Chatur has received his M.E. degree in Electronics Engineering from Govt. College of Engineering, Amravati, India and PhD degree from Amravati University. He has published twenty papers in national and ten papers in international journal. His area of research includes Neural Network, data mining. Currently he is head of Computer Science and Engineering department at Govt. College of Engineering, Amravati.