

AUTOMATIC NON-HORIZONTAL SCENE TEXT RECOGNITION FROM A DOCUMENT IMAGE

BASAVANNA M.¹, SHIVANAND GORNALE², SHIVAKUMARA P.³ AND SRIVATSA S.K. ⁴

¹School of Computing Science, VELS University, Chennai-Tamil Nadu-India

²Department of Computer Science, Government College (Autonomous), Mandya-Karnataka-India

³School of Computing, National University of Singapore-Singapore

⁴Senior Professor, St. Joseph College of Engineering, Chennai-Tamil Nadu-India

*Corresponding Author: Email- basavanna_m@yahoo.com, hudempsk@yahoo.com, profsks@rediffmail.com

Received: September 29, 2011; Accepted: November 03, 2011

Abstract- The detection and extraction of scene text from document images is one of the challenging research areas. Many researchers have detected and extracted the text from plain text background. But the multi-oriented scene text detection is one of the complex problems due to multi-oriented texts which have different orientation, font size, colors etc. In this work, we have proposed a new algorithm to detect and extract the multi-oriented scene text. Experiments have been carried out to find the robustness on the proposed method, by conducting various experiments on heterogeneous datasets. The proposed method achieves selectively high detection rate of 88.43% on the multi-oriented scene text.

Keywords - Multi-Oriented scene text, Sobel edge map, horizontal text image, Scene text detection, Scene text recognition

1 Introduction

Automatic identifying and extracting text lines of images has become an imperative need to help blind persons [1-4]. Efficient identifying and retrieving relevant information from multi-oriented text in an image becomes hard and challenging problem for the researchers [5]. Therefore, for the past decades, several ways based on annotation and content have been introduced to meet real challenges of the retrieval. But, to the best of our knowledge, none of the methods achieve good accuracy in filling the semantic gap between low level and high level features to understand the image [6-8]. This is because of unexpected and undesirable properties of image such as low resolution, complex background, different orientation, font, size and text [9-11]. Hence, an alternate way to fill the semantic gap is multi-oriented text detection, extraction and recognition to understand the image content. Text detection and recognition is quite familiar work for image analysis community but due to the above properties of image, image analysis based methods may fail to give satisfactory results [12-14].

Multi-Oriented text detection and extraction in image is usually addressed by four main approaches, namely, distance between two white pixels of horizontal text image, distance between two white pixels of oriented text image, connected components, and eliminating false positive using geometrical properties such as height, width, aspect ratio and PCA (Principal Component Analysis) of text block. These methods solve the problem to some extent, but still there is room for improvements especially for large image containing both graphics and scene text and more than 180 degree

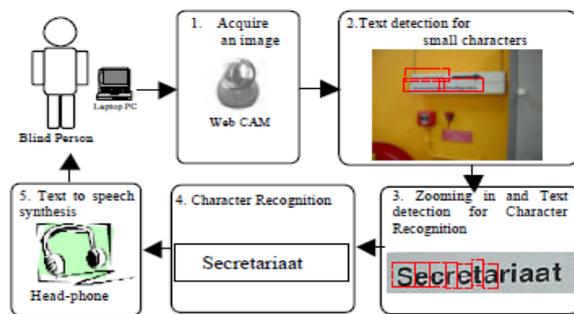
angle [5]. Many researchers have worked and noticed that there are several methods for oriented text detection in scene images. But none of the method to identify oriented text lines from a document image gives perfect solution to this problem in terms of simplicity and better accuracy. Hence, these factors motivated us to propose a new method for Multi-Oriented text detection in image. In this paper, we propose simple and effective method which considers regular spacing between characters and words as a basis for text detection. The proposed concept for text segmentation is to achieve better accuracy in text detection. The proposed method uses geometrical properties and PCA to find angle of all the blocks of image to eliminate false positives and retain true text blocks.

2 Methodology

The objective of image analysis is to recognize the text and graphics components in images of documents, and to extract the desired information as manually done. Two categories of document image analysis are defined viz. textual processing and graphical processing. Textual processing deals with the text components of a document image. In this, detection of skew (any tilt present in the document during scanning), finding columns, paragraphs, text lines, words, and finally recognizing the text (and possibly its attributes such as orientation, size, font etc.) through the Optical Character Recognition (OCR) system. Graphical processing deals with the non-textual line and symbol components that make up line diagrams, delimiting straight line between text sections, company logos etc. Pictures are the third

major component of documents, but except for recognizing their location on a page, further analysis of them is usually the task of other image processing and machine vision techniques. After application of these text and graphics analysis techniques, the several megabytes of initial data are called to yield a much more concise semantic description of the documents. Our system's aim is to guide blind person without anybody's help to walk around the city or busy area. Camera is fixed on head of the blind person, which captures images with text information; there is a system to recognize the text [15]. Once the text is recognized the system should translate it into speech to guide him.

2.1 System Design



The system works as follows a) Acquire the image. b) Oriented text detection from the acquired images. c) Store the detected text. d) Character Recognition e) Text to speech recognition. This research work is intended to carry out the second step of the proposed system i.e. Oriented text detection from the acquired image.

2.2 Proposed Algorithm

In this work, we propose new algorithm, for oriented text detection in scene images as we have observed that the space between characters and words in a text line of Sobel edge map of given gray input image is in regular pattern. This observation indicates strong clue for oriented text location in the scene image. The reason for using Sobel edge map of input image is that the Sobel operator detects edges when there is high contrast information in the images. It is true that text in the images usually have high contrast compared to its background. Since the aim of our work is to locate oriented text in images, few edges detected by Sobel edge operator are enough to identify the location of text in the images. The text detection is illustrated in Figure 1.

For the given horizontal gray image shown in Figure 1(a), we obtain Sobel edge map as shown in Figure 1(b) where we can see edge information corresponding to text information in the image. The horizontal run length is obtained by counting consecutive number of black pixel between the white pixels in horizontal direction for the Sobel edge map. The frequency of number of consecutives of black pixels

between white pixels is stored in the array say horizontal array (HA). Same procedures applied for oriented scene text input image Figure 1(c) and Figure 1(d). The frequency of number of consecutives of black pixels between white pixels of Figure 1(c) is stored in the array say oriented array (OA). We compare the frequency of number of consecutives of black pixels between white pixels of HA and OA, if the frequency of HA and OA is equal we fill corresponding oriented image Figure 1(e). To identify the frequency of number consecutives which represents text pixels, we introduce Max-Min clustering algorithm. The Max-Min clustering algorithm first selects maximum and minimum values in HA and then the values in HA are compared with Max and Min. As a result, we get two clusters. The cluster which belongs to Max values is considered as text cluster as shown in Figure 1(e) where text lines are separated with spaces compared to results in Figure 1(f). Then the output of Max-Min clustering algorithm is used for fixing bounding boxes for the text lines as shown in Figure 1(f) with false positive elimination. We use geometrical properties such as height, width, aspect ratio and using PCA angles of text block for eliminating false positives. The final text detection result after eliminating false positives can be seen with correct bounding boxes in Figure 1(f). One can notice from Figure 1(g) that false positives are not eliminated completely due to the problem of text and non-text separation.

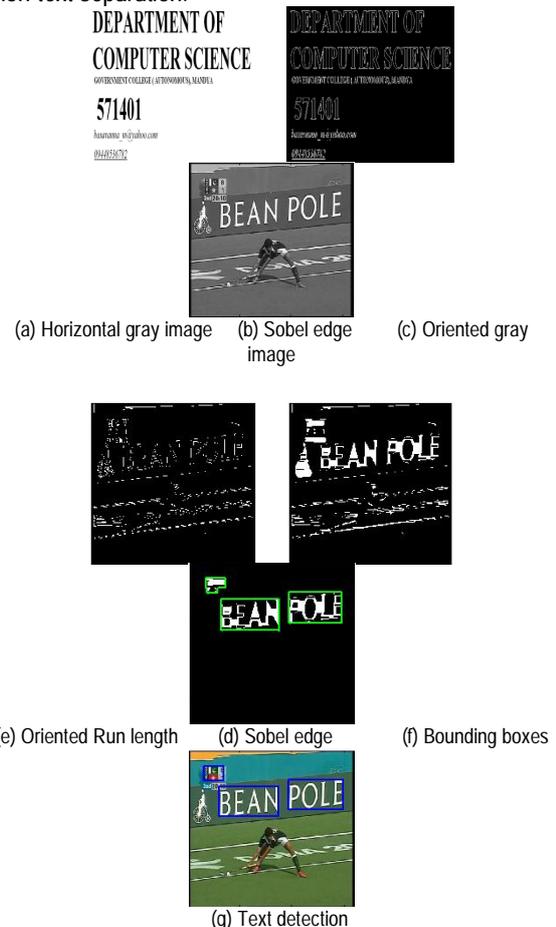


Fig.1- Steps for scene text detection

3. Experimental Results

In this experimental work we have proposed the new algorithm for oriented text detection from the acquired image, which is applied on 341 sample images. These images were used to test the proposed algorithm and compared with existing algorithms [16-17]. In all the three kinds of datasets, we have computed the number of true text block detection (TDB), Number of False Positives (NFP), number of Miss Detection Blocks (MDB) for the number of actual text blocks (ATB) of an image.

3.1 Performance Measures for Text Frames

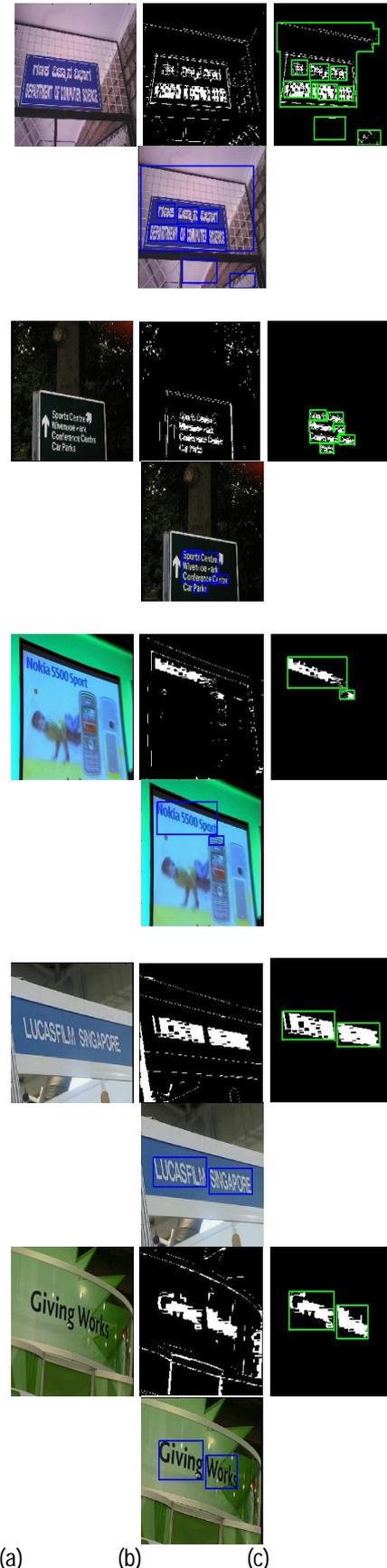
The performance of the proposed method is predicted as under by defining the following quality measures:

- Truly Detected Block (TDB): A detected block that contains a text string, partially or fully.
- Falsely Detected Block (FDB): A detected block that does not contain text.
- Text Block with Missing Data (MDB): A detected block that misses some characters of a text string (MDB is a subset of TDB).

For each image in the dataset, we manually count the number of Actual Text Blocks (ATB), i.e. the number of true text blocks in the frame.

- Recall (R) = TDB / ATB
- Precision (P) = $TDB / (TDB + FDB)$
- F-measure (F) = $2 \times P \times R / (P + R)$
- Misdetection Rate (MDR) = MDB / TDB

There are two other performance measures commonly used in the literature, Detection Rate and False Positive Rate; however, they can also be converted to Recall and Precision: Recall = Detection Rate and Precision = $1 - \text{False Positive Rate}$ [18]. Hence only the above four performance measures are used for evaluation. Table-1 and Figure-2 gives the objective subjective analysis of tested sample images for the performance of the two existing methods and the proposed method on the non-horizontal text dataset. The proposed method has the highest recall, the second highest precision (almost the same as that of Laplacian method) and the highest F-measure. This shows the advantage of the proposed method because it achieves good results while making fewer assumptions about text and MDR, which is good as those of the Skeleton-Based method. The proposed method achieves a relatively high detection rate (88.43%) on the non-horizontal text dataset, which is much more challenging than graphics text due to its arbitrary orientation and low contrast. However, the proposed method also has a high FPR.



Fig/ 2-Sample results (a) Input images (b) Oriented Run length (c) Bounding boxes (d) Text detection

Table 1- Performance of the proposed method and existing methods

Method	No. of Sample images	Recall	Precision	f-measure	MDR
Laplacian	220	0.85	0.77	0.81	0.14
Skeleton Based	220	0.72	0.63	0.67	0.16
Proposed	341	0.88	0.64	0.71	0.14

4. Conclusion and Future work

Recognizing the text and graphics component from images is one of challenging tasks in Digital image processing and machine vision. Due to unexpected and undesirable properties of images such as low resolution, Font, Size, Complex background and different orientation, it is very difficult to locate the exact text blocks in the image. In view of this we have proposed a new algorithm for Scene Text Recognition in document Image and the method works on the basis that the spacing between characters and words in an oriented text line with oriented regular pattern. The Experimental work has been carried out on 341 sample images and we have achieved a detection rate of 88.43%. It is also observed that in order to get more detection rate more false positives may be eliminated.

Further this work may be extended to strengthen the algorithm to get higher detection rate and for higher oriented text images with a complex background.

Acknowledgement

This research work is supported by University Grants Commission (UGC), New Delhi, India (F.No.UGC/MRP(s)/800/10-11/KAMY022).

References

- [1] Doermann D., Liang J. and Li H. (2003) *In Proc. ICDAR*, 606-616.
- [2] Jung K. (2001) *Pattern Recognition Letters*, 1503-1515.
- [3] Ye Q., Huang Q., Gao W. and Zhao. D. (2005) *Image and Vision Computing*, 565-576.
- [4] Chen D., Odobez J.M. and Bourlard H. (2004) *Pattern Recognition*, 595-608.
- [5] Clark P. and Mirmehdi M. (2000) *SPIE Conf. on Document Recognition and Retrieval VII*, 267-277.
- [6] Chucai Yi. and YingLi Tian. (2011) *IEEE Transactions*, Volume: 20, Issue: 9, 2594 – 2605.
- [7] Roy P.P., Pal U., Llados J. and Delalandre. M. (2009) *In Proc. ICDAR*, 11-15.
- [8] Jagath Samarabandu, Xiaoqing Liu. (2005) *IEEE International Conference*, 701 – 706.
- [9] Shivakumara P. Dutta A., Pal U. and Chew Lim Tan. (2010) *ICFHR*, 387 – 392.
- [10] Jung K., Kim K.I. and Jain A.K. (2004) *Pattern Recognition*, 977-997.
- [11] Wu V., Manmatha R. and Riseman E.M. (1999) *IEEE Transactions on PAMI*, 1224-1229.
- [12] Pan Y.F., Hou X. and Liu C.L. (2011) *IEEE Transactions on Image Processing*, 800-813.
- [13] Chen X., Yang J., Zhang J. and Waibel A. (2004) *IEEE Transactions on Image Processing*, 87-99.
- [14] Bhattacharya U., Parui S.K. and Mondal S. (2009) *In Proc. ICDAR*, 171-175.
- [15] Ezaki N., Bulacu M. and Schomaker L. (2004) *17th International Conference on Pattern Recognition*, 683-686.
- [16] Trung Quy Phan, Palaiahnakote Shivakumara, Chew Lim Tan. *DAS '10 Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 271-278.
- [17] Shivakumara P., Trung Quy Phan Chew Lim Tan. (2011) *Pattern Analysis and Machine Intelligence*, 412– 419.
- [18] Wong E.K. and Chen M. (2003) *Pattern Recognition*, 1397-1406.