ISSN No. 0976-5697

Volume 3, No. 2, March-April 2012

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A Survey on Image to Text Detection Methodology

Sandeep Sharma*, Jai Prakash
Department of Computer Science & Engineering,
Madan Mohan Malaviya Engineering College,
Gorakhpur-273010, U. P., INDIA
sndp.sharma01@gmail.com
jpr_1998@yahoo.co.in

Abstract- The automatic detection of text within a natural image is an important problem in many applications. Text detection in natural images has gained much attention in the last years as it is a primary step towards fully autonomous text recognition. It needs to be fast, efficient and robust in order to feed an OCR classifier with the correct input. In other words, segmented regions must correspond to the actual text. A lot of work has been done for detecting text in images and a lot has to be done. This paper gives detailed survey on image to text detection mechanism. This paper gives the description of work has been done for automatically detection of text from images, localize and extract horizontally aligned text in images (and digital videos) with complex backgrounds is presented.

Keywords – Text Detection, text segmentation, similarity measures.

I. INTRODUCTION

A text in any form or place has high purpose and contains more information related to the place and helps us to understand the objective more easily. The rapid growth in digital technologies and gadgets outfitted with megapixel cameras and invention of latest touch screen method in digital devices like mobile, PDA, etc., increase the demand for information retrieval and it leads to many new research challenges. Text detection and segmentation from natural scene images are useful in many applications. the primary property of scene text such as, high contrast against background, uniform colors are difficult to preserve in real application. When the system scans whole image for texts, text pixels with low contrast and non-uniform lighting could be confused as background due to similar colors.

Indexing images or videos requires information about their content. This content is often strongly related to the textual information appearing in them, which can be divided into two groups:

a. Text appearing accidentally in an image that usually does not represent anything important related to the content of the image. Such texts are referred to as scene text [1].
b. Text produced separately from the image is in general a very good key to understand the image it is called artificial text [1].

In contrast to scene text, artificial text is not only an important source of information but also a significant entity for indexing and retrieval purposes. Localization .of text and simplification of the background in images is the main objective of automatic text detection approaches. However, text localization in complex images is an intricate process due to the often bad quality of images, different backgrounds or different fonts, colors, sizes of texts appearing in them. In order to he successfully recognizable by an OCR system, an image having text must fulfill certain requirements, like a monochrome text and background where the background-to-text contrast should be high.

Recently, face detection and recognition is being applied in cameras and at large popular image sharing websites.

Several text detection methods have been proposed based on edge detection, binarization, spatial-frequency image analysis and mathematical morphology [2]. Generally text detection methods can be classified as either edge-based, connected-component based and texture-based methods [3].

According to [2] the best results were achieved using edge based text detection. It obtained top overall performance among 4 methods including mathematical morphology and color-based character extraction. Edge-based text detection has also been used in combination with edge profiles. Park et al. [2, 4] use them for automatic detection and recognition of Korean text in outdoor signboard images. However, they assume that a single text sign is located around the center line of the image. Edge profiles have also been used for detecting text in video data. Shivakumara et al. [5, 6] use edge profiles in combination with additional edge features to eliminate false positives selection.

The remaining part of our paper is organized as follows: In section II we will discuss the literature review done in field of image to text detection and in section III we will discuss the performance issues and research challenges. The comparative analysis of various images to text detection algorithms will be discussed in section IV and finally in section V we will conclude the paper and give the future scope of this paper.

II. LITERATURE REVIEW

Several approaches for text detection in images and videos have been proposed in the past. Several approaches for automatic detection and translation of text in images and videos have been proposed. Most of these methods aim to detect the characters based on general properties of character pixels. The distribution of edges, for example, is used in many text detection methods [7, 8, 9]. In these methods the edges are grouped together based on features such as size, color and aspect ratio. Texture is another commonly used feature for text segmentation [10, 11]. Many researchers working on text detection and thresholding algorithm with various approaches have achieved good performance based on some constraints. An early histogram based global thresholding Otsu's method is widely used in many applications [1]. Text detection and binarization method is proposed for Korean sign board images using k means clustering [2]. But finding a best value for 'k' to achieve a good binary image is difficult in images with complex background and uneven lighting. The linear Niblack method was proposed to extract connected components and texts were localized using a classifier algorithm [3]. Four different methods were suggested to extract text, depending on character size [4]. In the work of Wu et al. a method was proposed to clean up and extract text using a histogram based binarization algorithm [5]. The local threshold was picked at first valley on the smoothed intensity histogram and used to achieve good binarization result. A thresholding method was developed using intensity and saturation feature to separate text from background in color document images [6]. System using the gray-level values at high gradient regions as known data to interpolate the threshold surface of image document was proposed [7].

Layer based approach using morphological operation was proposed to detect text from complex natural scene images [8]. However, these method put few constrain and showed lots of missing and false positive detection on many natural scene images. This may confirm that the detection of text from natural scene is still a challenging issue. In our previous work we proposed a region based method using the color contrast of the text and their surrounding pixels. Due to limited number of color variation between text and its immediate background, finding a right threshold and detecting text pattern are key issues. Based on the methods being used to localize text regions, these approaches can be categorized into two main classes: connected component based methods and texture based methods.

Cai et a1.[2] have presented a text detection approach which is based on character features like edge strength, edge density and horizontal distribution. First, they apply a color edge detection algorithm in YUV color space and filter out non-text edges using a low threshold. Then, a local thresholding technique is employed in order to keep low-contrast text and simplify the background. Finally, projection profiles are analyzed to localize text regions. Lienhart and Effelsberg [SI have proposed an approach which operates directly on color images using the RGB color space. The character features like monochromacity and contrast within the local environment are used to qualify a pixel as a part of a connected component or not, segmenting each frame into suitable objects in this way.

Then, regions are merged using the criteria of having similar color. At the end, specific ranges of width, height, width – to - height ratio and compactness of characters are used to discard all non-character regions. Kim [6] has proposed an approach in which LCQ (Local Color Quantization) is performed for each color separately. Each color is assumed as a text color without knowing whether it is real text color or not. To reduce processing rime, an input image is converted to a 256-color image before color quantization takes place. To find candidate text lines, the connected components that are extracted for each color are merged when they show text region features. The drawback of this method is the high processing time since LCQ is executed for each color. Agnihotri and Dimitrova [11] have presented an algorithm which uses only the red part of the RGB color space, with the aim to obtain high contrast edges for the frequent text colors. By means of a convolution process with specific masks they first enhance the image and then detect edges. Non-text areas are removed using a preset fixed threshold. Finally, a connected component analysis (eight-pixel neighborhood) is performed on the edge image in order to group neighbouring edge pixels to single connected components structures. Then, the detected text candidates undergo another treatment in order to be ready for an OCR. Garcia and Apostolidis [4] perform an eight-connected component analysis on a binary image, which is obtained as the union of local edged maps that are produced by applying the band Deriche filter on each color. Jain and Yu [5] first perform a color reduction by bit dropping and color clustering quantization, and afterwards, a multi-value image decomposition algorithm is applied to decompose the input image into multiple foreground and background images. Then, connected component analysis combined with images performed on each of them to localize text candidates. This method can extract only horizontal texts of large sizes. The second class of approaches [7, 91 regards texts as regions with distinct textural properties, such as character components that contrast the background and at the same time exhibit a periodic horizontal intensity variation, due to the horizontal alignment of characters.

Methods of texture analysis like Gabor filtering and spatial variance are used to automatically locate text regions. Such approaches do not perform well with different character font sizes, and furthermore, they are computationally intensive. For example, Li and Doerman [7] typically use a small window of 16x16 pixels to scan the image and classify each of them as a text or non-text window using a three-layer neural network. For a successful detection of various text sizes, they use a three-level pyramid approach. Text regions are extracted at each level and then extrapolated at the original scale. The bounding box of the text area is generated by a connected component analysis of the text windows. Wu et al. [9] have proposed an automatic text extraction system, where second order derivatives of Gaussian filters followed by several non-linear transformations are used for a texture segmentation process. Then, features are computed to form a feature vector for each pixel from the filtered images in order to classify them into text or non-text pixels. In a second step, bottom-up methods are applied to extract connected components. A simple histogram-based algorithm is proposed to automatically find the threshold value for each text region, making the text cleaning process more efficient.

III. PERFORMANCE ISSUES AND RESEARCH CHALLENGES

Reading text in scene images challenge consisted of two tasks:

a) Text localization task: The target of text localization task was to identify text regions in scene images and mark their location with axis-aligned rectangular bounding boxes.
b) Word recognition task: The target of word recognition task was to recognize cropped word images of scene text. Cropping was done based on ground-truth word bounding boxes to evaluate recognition performance independently from text localization accuracy.

A. Dataset the dataset used in earlier Robust Reading Competitions organized in [12], [13], [14]. We carefully analyzed the dataset and observed following shortcomings:

a. Missing ground truth information for some of the files and text elements within some images.
b. Mixed interpretation of punctuation and special characters as part of words.
c. Bounding boxes around words are not tight. The ground truth is prepared in two phases.

In the first phase, we prepared text location ground truth using kolourpaint1. We converted all images to gray level and used colored bounding boxes to mark the word location and save them as 24 bit PNG image. We took special care of the following:

d. Space character is consistently used as word separation. All punctuation marks and special characters are considered as part of the word as long as there is no space character separating them.
e. The bounding boxes are tight so they touch most of the boundary pixels of a word.

In the second phase, we prepared word recognition ground truth. We prepared a simple ground truth GUI to annotate words in an image. The GUI allows users to draw rough bounding boxes around words and label them with ASCII string. We generated our ground truth automatically using the colored image files and labels generated using our GUI by evaluating bounding boxes overlap for a given image file. The ground truth consisted of bounding box co-ordinates which are stored in a separate text file for each of the image files. The same method is used to extract word images and its associated ground truth. Our word recognition dataset consisted of 1564 word images. These word images are actually cropped from images in the text localization dataset using word bounding box ground truth. Each word is stored in a separate file and the ground truth transcription for these words is provided in a line separated file.

B. Performance Evaluation:

a. Text Localization Task: The task of text localization can be evaluated using any standard methodology for evaluating page segmentation performance [15], [16] that takes into account different categories of segmentation errors (over-, under-, and missed-segmentation). The main question when choosing a method for scene text detection particularly is how to deal with under and over segmentation errors. In this competition, we employ the method by Wolf et al. [17] that is specifically designed to evaluate scene text detection approaches.

b. Word Recognition Task: To evaluate the word recognition accuracy, we simply use the edit distance with equal cost of deletions, substitutions, and insertions. We normalize the edit distance by the number of characters in ground truth word.

IV. COMPARATIVE ANALYSIS

The comparison between various algorithms used for detecting text from the natural scene is described in the following table.

Algorithms	**Parameters**	**Advantage**	**Disadvantages**
Yi's Method	Ad boost learning model	text regions are merged into rectangle boxes	Only localize the text reason
Kim's Method	MSER	localizing text region in a mobile phone	Only localize text region in a mobile phone
Text Hunter	Detection window Size varies from 32*16 pixels to 288*144 pixels in 9 steps.	maximize the F-measure on the given training database	case of scene text features based on gradient information of connected component are also taken into During validation.
KAIST AIPR System	Conditional Random Field	one or more neighboring super pixels together	localization task is not yet published
Neumann's Method	Maximally Stable External Regions	trained using synthetic Data.	compensat errors in text detection
LIP6-Retin	large character size variations	generate text hypotheses	Very complex

V. CONCLUSION

In This paper we have studied the define what is image to text detection paradigm the we have done the literature review in image to text detection and we have also analyze the various performance issues and research challenges as well as we have examines the various image to text detection algorithms and we have done the comparative analysis of the algorithms on the basis of various parameters and we have done the analysis on the basis of advantages and disadvantages. So we can say that image to text detection is a growing these days and become the important area of research and a number of work has been done in this field and various work has to be done in this field and can propose various fine algorithms in the field of image to text detection which can perform well.

VI. REFERENCES

[1] R. Lienhart and W. Effelsberg. Automatic Text Segmentation and Text Recognition for Video Indexing Multimedin Sysfem, Vol. 8, pp. 69-81,2000.

[2] N. Ezaki, M. Bulacu, L. Schomaker, “Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons”, *Int. Conf. on Pattern Recognition* (ICPR 2004), vol. II, pp. 683-686.

[3] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, S. Hwang, “Automatic detection and recognition of Korean text in outdoor signboard images”, *Pattern Recognition Letters*, 2010.

[4] T. N. Dinh, J. Park, G. Lee, “Korean Text Detection and Binarization in Color Signboards”, *Proc. of The Seventh Int. Conf. on Advanced Language Processing and Web Information Technology* (ALPIT 2008), pp. 235-240.

[5] P. Shivakumara, W. Huang, C. L. Tan, “Efficient Video Text Detection using Edge Features”, *Int. Conf. on Pattern Recognition* (ICPR 2008), pp. 1-4.

[6] P. Shivakumara, T. Q. Phan, C. L. Tan, “Video text detection based on filters and edge features”, *Int. Conf. on Multimedia & Expo* (ICME 2009), pp. 514-517.

[7] Q. Yuan and C. Tan, “Text extraction from gray scale document images using edge information,” *Sixth International Conference on Document Analysis and Recognition*, 2001, pp. 302–306.

[8] X. Chen, J. Yang, J. Zhang, and A. Waibel, “Automatic detection and recognition of signs from natural scenes,” *IEEE Transactions On Image Processing*, vol. 13, pp. 87–99, 2004.

[9] N. Ezaki, M. Bulacu, and L. Schomaker, “Text detection from natural scene images: towards a system for visually impaired persons,” *Proceeding of the 17th International Conference on Pattern Recognition*, vol. 2, 2004, pp. 683–686.

[10] 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 Transactions On Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1631–1638, 2003.

[11] K. Jung, J. Han, K. Kim, and S. Park, “Support vector machines for text location in news video images,” *TENCON*,vol. 2, 2000, pp. 176–180.

[12] S. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, “ICDAR 2003 robust reading competitions,” in Proc. Int. Conf. on Document Analysis and Recognition, Edinburgh, UK, Aug. 2003, pp. 682–687.

[13] S. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, K. Ashida, H. Nagai, M. Okamoto, H. Yamamoto, H. Miyao, J. Zhu, W. Ou, C. Wolf, J. Jolion, L. Todoran, M. Worring, and X. Lin, “ICDAR 2003 robust reading competitions: Entries, results, and future directions,” Int. Jour. on Document Analysis and Recognition, vol. 7, no. 2-3, pp. 105–122, Jul, 2005.

[14] S. Lucas, “ICDAR 2005 text locating competition results,”in Proc. Int. Conf. on Document Analysis and Recognition, Seoul, Korea, Aug. 2005, pp. 80–84.

[15] F. Shafait, D. Keysers, and T. M. Breuel, “Performance evaluation and benchmarking of six page segmentation algorithms,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 941–954, 2008.

[16] A. Antonacopoulos, S. Pletschacher, D. Bridson, and C. Papadopoulos, “ICDAR 2009 page segmentation competition,” in Proc. Int. Conf. on Document Analysis and Recognition, Barcelona, Spain, Jul. 2007, pp. 1370–1374.

[17] C. Wolf and J. Jolion, “Object count/area graphs for the evaluation of object detection and segmentation algorithms,” Int. Jour. on Document Analysis and Recognition, vol. 8, no. 4, pp. 280–296, Sep. 2006.