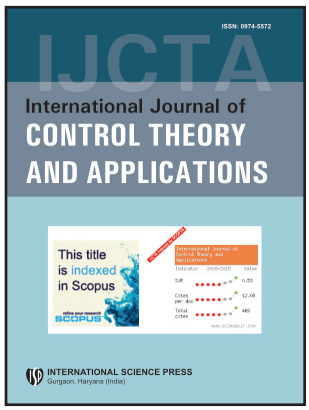


International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 16 • 2017

A New Approach for Automatic Scene Text Detection

S. Bhuvaneswari[a], T.S. Subashini[b], M. Saravanan[c] and S. Sumathi[d]

[a]*Assistant Professor, Department of Computer Science, Annai College of Arts and Science, Kovilacheri, Kumbakonam. Email: drbhuvanamagesh@gmail.com*
[b]*Associate Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Annamalainagar. Email: rtramsuba@gmail.com*
[c]*Professor & Head, Department of, Computer Science, Annai College of Arts and Science, Kovilacheri, Kumbakonam. Email: Saravanan.professor@gmail.com*
[d]*Assistant Professor, Department of Computer Science, Annai College of Arts and Science, Kovilacheri, Kumbakonam*

Abstract: Many applications need embedded text that will provide useful information. Text extraction is very challenging for its variety of text patterns and variant background interferences. These difficulties provide a path way for text recognition. In this paper, new extracted method in scene text detection is proposed. First preprocessing is done using gray image transformation, median filtering, and sobel edge operator. The process of text region extraction has different stages that include vertical edge detection, horizontal edge detection, background subtraction, removal of noise, and candidate text region extraction. Extracted candidate text regions are further segmented using background subtraction method, which helps to separate text as well as non-text regions without background. Experimental results proves the better performance of the proposed method compared with existing methods.

Index Terms: Pre-processing, background subtraction, Scene Text Detection, Scene Text Recognition.

1. INTRODUCTION

This Text detection and recognition is an area of research which aims to develop a system to read the text from images automatically. Due to the digitization, most of the resources such as historical manuscripts, land records, old books, journals, newspapers, etc., are converted to images.

Automatically detecting and extracting text from these digitized images gives the path way to new and challenging research issues. The text in image gives useful information about a scene. Extraction of this text data can be useful in navigation, image search and indexing, image understanding, and computer – human interaction. Further, text detection has its applications such as document analysis, forensic-analysis, content based image searching and retrieval, object identification, video content analysis, etc. Text characters are difficult to detect and recognize due to their variation in terms of size, font, style, orientation, alignment, contrast, background etc. The aim of automatic text extraction is to detect the characters using the general properties of text pixels,

namely: text contains a number of edges, text width is larger than height, and text is usually of uniform size. Text size is a major factor which is usually of uniform size and texture property of text is irregular and weak [1]. The different types of text images are: document text image, caption text image, and scene text image, and the text data in image has different font styles, size, colours, orientation and is mostly against a complex background. Figure 1 shows the different types of text present in digital images.

(a) (b)

Figure 1: Text detection: (a) Text image and (b) Detected text

Document text is a combination of text with few graphics components which usually appears in black on white background. In this type of image, the text differs in size, style, alignment, colour etc. In the case of colour images, the main problem is to separate the text region from the remaining parts of an image. If the background is white, then the separation becomes easy for extraction, but in the case of colour text images, the detection method becomes more complex.

Caption text is otherwise known as overlay text or cut-line text. The superimposed text can be easily detected, segmented, and recognized automatically. Caption text is the text inserted on the video/image during the time of editing, and it usually defines the meaning of the image or video content. This include text which moves, rotates, shrinks, and which is of arbitrary orientation and size. The Figure 2 illustrates the types of text.

(a) (b) (c)

Figure 2: (a) Document text, (b) Caption text and (c) Scene text

Scene text appears within the scene during the shot and occurs naturally as a part of the scene and contains important semantic information like street names, traffic signals, number plates, food containers etc. It can also be used as a cue in recognizing the content of the image. Even though, the detection of text is very interesting, there are certain challenges are also available while detecting the scene text images.

The major challenges in scene text detection and recognition can be roughly categorized into three types [2, 3]:

- **Scene text Diversity:** In contrast to characters in document images, which are usually with regular font, single colour, consistent size and uniform arrangement, texts in natural scenes may bear entirely different fonts, colours, scales and orientations, even in the same scene.

- **Background Challenges:** The backgrounds in natural scene images and videos can be very complex. Elements like bricks, fences, signs, and grasses are very difficult to separate from true text, and thus are needs techniques to separate.
- **Interference factors:** Various interference factors, for

Instance, noise, blur, distortion, low resolution, non-uniform illumination and partial occlusion, may give rise to failures in scene text detection and recognition.

To tackle these challenges, a rich body of approaches has been proposed and substantial progresses have been achieved in recent years.

This paper aims in detecting text from scene text images automatically without any manual intervention. In this paper a robust method for segmentation of text from the images is proposed. The proposed method uses RGB histogram for detection. For accurate detection of text preprocessing is done accurately. This method take cares of eliminating false negatives and recovering the full text even from the complex background.

The rest of the paper is organized as follows. Section 2 describes the related work in this area. Section 3 presents the key observations and methodology of this work. Section 4 shows the experimental results. Section 5 concludes the paper.

2. RELATED WORK

The text in an image gives detailed information about a scene, which is helpful in a wide range of applications, such as indexing, image understanding, monitoring/controlling the movement of an object, and human-computer related applications. This section discusses the work done so far related to the detection of text from images automatically both from simple and complex backgrounds. The aim of text detection is to identify candidate text regions in a given input image.

Scene text detection can be categorized under two categories. Detection of text is generally has preprocessing like binarization, angle correction, and removal of noise and then detected scene texts are parsed to the existing OCR engines. Before sending detected text to OCR for recognition authors in [4] performed Niblack's adaptive binarization algorithm [5] on the detected text region. Authors in [6] proposed an iterative binarization method on unique character using k-means for producing a set of potential binarized characters and then Support Vector Machines (SVM) is used to measure the degree of character likeness and the one with maximum character-likeness is selected as the optimal result. Markov Random Field (MRF) model is utilized for binarization in [7] and an auto-seeding technique is proposed for determining certain foreground and background pixel seeds and then MRF is used for segmenting text and non-text regions. The recent approach first extracts certain features from gray/color images and then trains classifiers for scene text recognition. Authors in [8] recognized the scene text and the performance is evaluated using feature descriptors like Shape Contexts, Scale Invariant Feature Transform (SIFT), Geometric Blur, Maximum Response of filters, patch descriptor, etc. in combination with bag-of-words model. But the results are not satisfactory to serve as the basis for word recognition. In [9], the authors address the problem by employing Gabor filters and then building a similarity model to measure the distance between characters in their text recognition framework. Authors in [10] introduced a new approach to segment the text from colour images. To locate the candidate text lines, the multi scale wavelet features and structural information are used. From the candidate text lines, the true text is identified using support vector machine (SVM) classifier. This approach has four stages. In the pre-processing step, input text blocks are rescaled with cubic interpolation, and Gaussian filter is utilized for smoothing and removing noise. Further processing is done using component filtering procedure. Further K-Means clustering algorithm is applied to cluster the remaining

components into multiple text layers after which, a set of proper constraints are applied to detect the real text layers.

An efficient document text extraction method which is computationally fast is proposed in [11]. Haar discrete wavelet transform is used to detect edges of the candidate text regions. Then thresholding technique is used to remove the remaining non-text edges from the image. To connect the discrete candidate text edges, morphological dilation operator is used. Then, based on the edge map, the line feature graph is generated. Further, improved Canny edge detector is utilized to detect text pixels and the spatial distribution of edge pixels helps to extract the stroke information. Finally, according to the line features, the image is filtered and exact text regions are isolated. Text embedded in complex coloured document image is detected by authors in [12]. They proposed edge based features for their work. The weighted sum of the R, G, and B components was used to convert the colour image to gray scale. Then Sobel masks are applied to detect the horizontal and vertical edges. This is followed by the elimination of weak edges. Then the edge image is split into non-overlapping blocks of m × m pixels, where m depends on the resolution of the image. Using predefined threshold, block classification is performed, and it differentiates the text from the image.

Authors in [13] introduced a new method for text extraction from complex colour document images. Edge detection is performed using Canny edge detector followed by the dilation morphological operator. This results in the creation of holes in most of the connected components that represent character strings, and the components without holes are removed, which corresponds to non-characters. The standard deviation of each connected component is computed for eliminating the remaining non-text components. Still, if noisy text region persists, it is replaced to improve the quality of the retrieved foreground.

The authors in [14] first performed colour space reduction after which segmentation and spatial regrouping was carried out for detecting text. The authors tried to tackle the problem of touching the text; however, the segmentation algorithm failed in the case of poor quality documents and specifically in case of video sequences. Authors in [15] proposed a method of text extraction which is done in three stages. Candidate text region is detected in the first stage by generating a feature map. Feature map is a binary image created using the edge characteristics like strength, density, orientation, and the pixel intensity of the feature map gives clues about the possibility of text regions.

In this paper automatic text detection is proposed with the focus on improvement in automatic detection of text without user intervention.

3. METHODOLOGY

The block diagram of the proposed method for text localization in a natural scene image is shown in the Figure 2. The input image may be a color or gray scale image. If the image is color image, then it is converted into grayscale image and preprocessing operation is applied on the resultant image. In this algorithm, a color image is input data to the system and the segmented text with black background is the output. The processing of scene text images is difficult due to the presence of noise. The proposed method uses 2D order-statistic filtering for preprocessing. The 2D order-statistic filtering is used to remove the noise and enhance the weak boundaries of images. The proposed method consists of three stages, namely, preprocessing, feature extraction and classification which are described in this section.

We propose text extraction method based on intensity information from natural scene images. The proposed method is done using binarization by an average intensity of the image, and then filtering is done using sobel operator and both vertically and horizontally using sobel and subsequently image background is removed using threshold and after and this method is appropriate to extract texts from complex backgrounds.

Methodology

The block diagram of the proposed method for text localization in a natural scene image is shown in the Figure 2. The input image may be a color or gray scale image. If the image is color image, then it is converted into grayscale image and preprocessing operation is applied on the resultant image. In this algorithm, a color image is input data to the system and the segmented text with black background is the output. The processing of scene text images is difficult due to the presence of noise. The proposed method uses 2D order-statistic filtering for preprocessing. The 2D order-statistic filtering is used to remove the noise and enhance the weak boundaries of images. The proposed method consists of four stages, namely, preprocessing, text region extraction, background subtraction and recognition which are described below.

Preprocessing

The preprocessing consists of gray image transformation, median filtering, and sobel edge operator. First, the input colour image is transformed to a gray scaled image [10]. Then median filter is applied to the gray image for removing noise followed by a 2D order-statistic filtering and an opening operation is done and last an edge image is extracted by sobel operator.

Text Region Extraction

The text region extraction consists of vertical edge detection, horizontal edge detection, background subtraction line and noise removal, and candidate text region extraction. In the first step, sobel edge operator is used to find the vertical and horizontal edges of the image. The Sobel edge detector uses a pair of *3 x 3* convolution masks, one estimating gradient in the x-direction and the other estimating gradient in *y*–direction. The Sobel detector is very sensitive to noise in pictures that effectively highlight as edges. Second step, clears the background using calculated threshold. The resultant image is filtered using 2-D order-statistical filtering, to enhance and connecting the edges. Threshold is found out experimentally and this value is compared with every pixel and if the compared value is lesser than threshold is set to one to extract the text regions. Connected components and their bounding boxes are extracted using their size, locations, and aspect ratio. Figure 2 shows the results of the proposed algorithm after Pre-processing and text region extraction. Figures 2a shows the input image. Figure 2b shows the image where long line and small noise are removed. Figure 2c shows the input image with bounding boxes of superimposed of the connected components.

Character segmentation:

Character regions located in images are required to be binarized before recognition. Generally, characters can be easily eminent with colour background, thus colour clustering is common for binarization. So that characters in natural scene images often suffer from uneven light, reflex and shadow, which likely to make stroke broken and bring about some isolate noise, hence the performance of color clustering is greatly affected. Hence we proposed a background subtraction method.

Background Subtraction

The newly formed image from the previous step is compared to the original image and the region lying within the bounding box is retained. This is done by subtracting original image from the newly formed image. The output of this step is the image comprising of only textual region.

4. EXPERIMENTAL RESULTS

The ICDAR dataset which contains ground truth data of the natural scene images with text from the ICDAR 2003 Robust Reading Competition tagged Dataset Collection Team [16] is used in this work to analyze the performance of the proposed text detection method. Fifty images having both simple and complex scene text images present in this database of varying size was taken for the study.

Precision

Precision is the ratio of number of correctly detected text (i.e. the number of text correctly labelled as belonging to the text class) to the total number of text and non-text.

$$\text{Precision} = \frac{\text{Correctly detected text}}{\text{Total number of objects in images}}$$

Recall

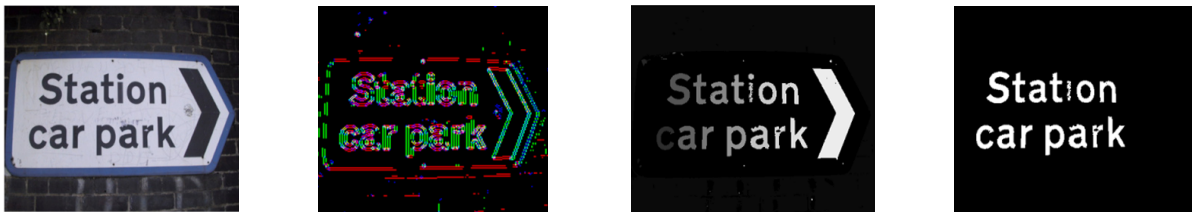
Recall is defined as the ratio of number of correctly detected text to the total number of elements that actually belong to the text class.

$$\text{Recall} = \frac{\text{Correctly detected text}}{\text{Total number of text in images}}$$

The comparison of the three proposed text detection method with existing methods in terms of precision and recall is given in Table 1. From the Table it could be seen that the proposed method was able to detect text better than the other three methods.

Table 1
Performance measure of proposed method

S.No.	*Methods*	*Precision %*	*Recall %*
1.	[17]	72.93	52.86
2.	[18]	78.03	63.35
3.	[19]	85.83	94.88
4.	Proposed method	87.54	96.28



(a) (b) (c) (d)

Figure 3: (a) Input image, (b)–(c) Images during processing and (d) Resultant image

Figure 3 shows the processing stages between input image and resultant image.

5. CONCLUSION AND FUTURE WORK

Due to the complicated background and unpredictable text appearances scene text detection is still a challenging problem. This papers a new improved scene text detection method is proposed that makes use of extracting scene

text alone. With the rapid growth of camera based application readily available on mobile phones, understanding scene text is very important. Comparison is made and different challenging tasks such as cluttered background and diverse text patterns. This paper has discussed a new scene text detection method which is done in four stages: Preprocessing, Text region extraction, character segmentation and background subtraction

In this application, maximum effort is made to identify the text from the image and to produce more accurate results than the existing systems. In future work, the accuracy rate of text detection and recognize text in videos also will be taken into an account.

REFERENCES

[1] Yen-Lin Chen, "Automatic text extraction, removal and inpainting of complex document images," in International Journal of Innovative Computing, Information and Control, 2011.

[2] C. Yao, X. Zhang, X. Bai, W. Liu, Y.Ma, and Z. Tu. Rotation-invariant features for multi-oriented text detection in natural images. *PLoS One*, 8(8), 2013.

[3] C. Yao, X. Bai, B. Shi, and W. Liu. Strokelets: A learned multi-scale representation for scene text recognition. In *Proc. of CVPR*, 2014

[4] X. Chen and A. Yuille, "Detecting and reading text in natural scenes," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 2. IEEE, 2004, pp. II–366.

[5] W. Niblack, An introduction to digital image processing. Strandberg Publishing Company, 1985.

[6] K. Kita and T. Wakahara, "Binarization of color characters in scene images using k-means clustering and support vector machines," in Proceedings of the 2010 20th International Conference on Pattern Recognition, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 3183–3186. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2010.779

[7] A. Mishra, K. Alahari, and C. Jawahar, "An mrf model for binarization of natural scene text," in Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011, pp. 11–16.

[8] T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images." in VISAPP (2), 2009, pp. 273–280.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1. IEEE, 2005, pp. 886–893

[10] Yaowen Zhan, Weiqiang Wang, and Wen Gao, "A robust split-and-merge text segmentation approach for images," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. IEEE, 2006, Vol. 2, pp. 1002–1005.

[11] S Audithan and RM Chandrasekaran, "Document text extraction from document images using haar discrete wavelet transform," European Journal of Scientific Research, Vol. 36, No. 4, pp. 502–512, 2009.

[12] Grover Sachin, Kushal Arora, and Suman K Mitra, "Text extraction from document images using edge information," in IEEE India Council Conference, 2009.

[13] P Nagabhushan, Shivananda Nirmala, et. al., "Text extraction in complex color document images for enhanced readability," Intelligent Information Management, Vol. 2, No. 02, pp. 120, 2010.

[14] Anil K Jain and Bin Yu, "Automatic text location in images and video frames,"Pattern recognition, Vol. 31, No. 12, pp. 2055–2076, 1998.

[15] Xiaoqing Liu and Jagath Samarabandu, "An edge-based text region extraction algorithm for indoor mobile robot navigation," in Mechatronics and Automation, 2005 IEEE International Conference. IEEE, 2005, Vol. 2, pp. 701–706.R. Crane, "A simplified approach to Image Processing", Prentice-Hall, 1997.

[16] Simon M Lucas, Alex Panaretos, Luis Sosa, Anthony Tang, Shirley Wong, and Robert Young, "Icdar 2003 robust reading competitions," in 2013 12th International Conference on Document Analysis and Recognition. IEEE Computer Society, 2003, Vol. 2, pp. 682–682.

[17] S. Bhuvaneswari, T.S. Subashini, "Detection and De-occlusion of Text in Images," in Centre for Promotion of Multidisciplinary Research – International Journal of Technology (CPMR-IJT) (ISSN:2250-1827), vol.2(2), pp. 6-10, 2012.

[18] S. Bhuvaneswari, T.S. Subashini, "Automatic Detection and Inpainting of Text Images," in International Journal of Computer Applications (ISSN:(09758887), vol.61(7), 2013.

[19] S. Bhuvaneswari, T.S. Subashini, "Automatic Text Detection using Wavelet Transform and Connected Component Labelling", in International journal of Applied Engineering Research, (ISSN: 0973-4562), vol.9(21), pp. 4871-4875, 2014.