I J C T A, 9(19) 2016, pp. 9063-9069

A Novel Approach for Text Extraction from Natural Scene Images

N. Jayanthi*, Abhilash Gaur* and S. Indu*

ABSTRACT

Digital images acquired by the camera have useful information of interest. Image retrieval systems provide a platform for retrieving the information for content based retrieval systems and image search engines. This paper addresses the problem of text region extraction for slanted images and presents a novel way to deal with concentrate content areas in a picture by application of Difference of Gaussian as a second order derivative edge detector combined with Niblack's Thresholding. Detailed comparison ofexisting algorithms using edge detectors is done with the stated method.The proposedalgorithm is testedon an image with various content size, textual style styles and content dialect.Execution is assessed in view ofprecision rate and recall rate for proposed strategy on sample data set and ICDAR 2015 data set.

Keywords: Difference of Gaussian; Niblack'sThresholding; Edge Detection; Text Extraction

I. INTRODUCTION

The recent technological boom and development of digital devices has made multimedia such as image and video, one of the most important carriers of information.More and more researchers are attracted to design an automated system for text extraction as the need to index, browse and retrieve the information present in an image or video is increasing. One, such as created application is the portable managing ofbank accounts using the mobile banking application provided by the banking institutions that allows the that permits the clients to play out the trades even on transferring the picture of the check to the server [1][3][4]. These sorts of uses rely on upon a Textual Information Extraction framework which can recognize, limit and concentrate the content data contained in an image[2]. An interpretation camera is another such applicationthat obtains the picture, recognize content from it and afterward decipher it in required dialect the local language [2] [7].

Numerous text detection algorithms have been proposed on text detection and localization. All these algorithms broadly belong to either edge-based techniques or connected-component based techniques or texture based techniques [6] [8]. Distinctive methods were expressed in the past for perceiving content information in a picture. Every technique suits for some predefined set of images like images with low light, images with sharp edges, and so forth. For such different types of images the output produced by algorithm may vary due to dimensions of image, colour, contrast, and orientation. In natural images, detection of text has some challenges because of variety in content size and alignment.This paper proposes a novel approach of text extraction which uses modified Difference of Gaussian as edge detector, Niblack's Thresholding for binarization [9] [10] [11] and morphological operations for post processing andalso compares the proposed method with the two other edge detectors namly, Gaussian Edge Detectorand Prewitt Edge Detector [10][8] by evaluating the performance parameters. Experiments on the dataset show that the promising resulthas been achieved by our proposed method.The rest of the paper is composed as takes

* Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India, *E-mails: njayanthi@dce.ac.in; itsmegaur@gmail.com; sindu@dce.ac.in*

after:The brief overview identified with the different techniques for text extraction is portrayed in segment 2, the proposed strategy is exhibited in area 3, Experimental consequences of execution investigation is outlined in segment 4, subdivision 5 finishes up the work and displays the future proposals.

II. RELATED WORK

Prewitt edge detection and Gaussian edge detection belong to the class of first order derivative. The Prewitt operator is a discrete differentiation operator which functions by calculating the gradient of the image intensity function and makes use of the maximum directional gradient. The Prewitt masks are simpler to implement but are very sensitive to noisealso Prewittoperator is more sensitive to **horizontal** and **vertical edges**. In general, the first order derivative operators are very sensitive to noise and produce thicker edges and are not efficient for localization of the edge[13][14]. In [12], U.Bhattacharya, S. K. Paruri & S. Mondal et al extricated associated segments (both highly contrasting) from the twofold picture. At that point, morphological opening operation alongside an arrangement of criteria is connected to concentrate features of Devanagari or Bangla text. Further, in [16], H. Raj & R. Ghosh et al used different geometrical properties for identifying the location of the whole-text part related to the detected headlines form morphological operations.

On the other hand,Laplacian of Gaussian belongs to a class of second order derivatives, i.e. the second derivative of Gaussian. The zero-crossings of the second-order derivatives are good for localization of the edge. The DoG function returns a nearby estimatethe scale-normalized LoG, as reviewed by Lindeberg (1994). Lindeberg demonstrated that the normalization of the Laplacian is required for genuine scale invariance. Also the use of LoG is computationally expensive. But processing time while using DoG is less and produce a strong response for poorly detected edges because they have fixed characteristics in all directions. [13][14] [15]

III. PROPOSED METHOD

Our aim is to extract the text region from the natural scene images which may include slanted and curved text with minimum computational cost using Difference of Gaussian method. A short description of the methods used in our proposed algorithm is given below.

2.1. Gaussian Filter and Difference of Gaussian

Gaussian filter is a filter having Gaussian function as impulse response. Gaussian filters have properties demonstrating no overshoot to a step function excitation at the same time minimizing the rise andfall time. By mathematical definition, a Gaussian filter modifies the input signal by convolving with a Gaussian function [3].As a gray image is a two dimensional object, two dimensional Gaussian function ischaracterized underneath.

$$g(u,v) = \frac{1}{2\pi\sigma^2} e^{-\frac{u^2+v^2}{2\sigma^2}} \quad (1)$$

Where v represents the distance from the origin in the vertical axis, u denotes the distance from the origin in the horizontal axis and σ is the standard deviation of the Gaussian distribution [3][8]. In image processing, difference of Gaussians (DoG) is an element improvement calculation that incorporates the subtraction of two blurred versions of the image under observation with different standard deviations thus subsequently sparing the edge information in a picture.

For a grayscale image, the blurred images are produced by convolution of the original grayscale image with a Gaussian function [3] with distinctive standard deviations. Obscuring an image using a Gaussian kernel smothers just high frequency spatial information. Subtraction of one image from other protects the spatial information that falls between the frequency sets that are preserved in the two blurred images. In this manner, the DoG is a band pass filter that disposes the modest bunch of spatial frequencies that are existent in the basic grayscale image.

2.1. Niblack's Adaptive Thresholding Algorithm:

The Niblack's algorithm estimates a threshold value to each pixel by scanning a rectangular window over the grayscale image [9][10]. The size of rectangular window can be adjusted. The threshold value is ascertained utilizing the local mean and standard deviation of the considerable number of pixels in the window and is computed by taking after conditions [9][10]:

$$T_{Niblack} = M + k * SD \tag{2}$$

$$T_{Niblack} = M + k * \sqrt{\frac{\Sigma(p_i - m)^2}{NP}} \tag{3}$$

Here NP is the total no of pixels in a grayscale image, T represents the threshold value, M is the mean of gray levels of pixels and k is fixed depending upon the noise in the background and is -0.2.

The algorithm is summarized below and each step is explained in detail in sub-sections.

i) Convert the image into a grayscale image.
ii) Filter the image for supressing noise using a Gaussian low pass filter.
iii) Use Niblack's Thresholding algorithm to binaries the image
iv) Apply modified Difference of Gaussian to detect the edges of the filtered image.
v) Apply morphological operations.
vi) Multiply the resultant image with binary image of the input image to obtain text in contrast with Plain background.

2.3. Pre-processing

Pre-processing encourages us to enhance the execution and make the procedure productive. The major steps are converting images to grayscale images and then binarization of images and noise elimination using some filtering.

a) Gray Scaling: The input image is an RGB image comprises of intensities R (Red), G (Green) and B (Blue). The direct application of an algorithm might not split up the text from the background, hence the input image is transformed to grayscale image which results intensities of each pixel in the range of [0,255].

b) Filtering: To enhance picture quality and for further handling, Gaussian filter which is rationally symmetric is used to eliminate noises such as blurred high frequency noise and white noise by smoothening the image.

c) Binarization: Binarization results in a black and white image consisting of pixel values either 0 or 1. In the proposed algorithm the binarization is performed using Niblack's thresholding algorithm. Binarization helps in separating the text from the background up to some extent, but alone binarization in not enough to get proper text information from the data.

2.4. Edge Detection using modified Difference of Gaussian

The difference of Gaussian calculation wipes out the high frequency elements that often include random noise removal. It reduces noise in the image by blurring and also preserves the edge information in the image. The DoG also eliminates the low frequency background blotches. Finally the resultant image is normalized to [0, 1]. In our method the modified equation for DoG is given by:

$$\Gamma\,(x, y) = I * (\mathcal{G}_{\alpha 1} - \alpha * \mathcal{G}_{\sigma 2})$$

Here $\Gamma(x, y)$ is the output image after application of DoG, I is the input gray scale image, $\mathcal{G}_{\alpha 1}$ and $\mathcal{G}_{\sigma 2}$ are Gaussian function provided $\sigma 2 > \sigma 1$ and α is a constant which lies between 1 and 2.

2.5. Morphological Operations

The previous described processes results in text regions. Morphological dilation is used to form a cluster of text regions. A reasonable estimated organizing component ought to be picked such that only least non-text area should be clustered within. Morphological opening is utilized to expel non-content objects.

2.6. Character Extraction

Extraction of the characters as is done in this step by multiplying the resultant image with black and white version of the original image.

IV. EXPERIMENTAL RESULT

The experiment is done on more than 100 images. The result for some test images is shown below. Results are compared with the output using the Prewitt Edge Detector and Gaussian Edge Detector. Also the proposed strategy is validated on distinct images of the training data set form ICDAR 2015, and demonstrated that the technique can enhance the text detection performance. The performance of the algorithm is validated based on Precision rate and Recall rate. The stated parameters are the average of 100 test images. Overall, 74.8% accuracy is obtained from this algorithm on precision value, 76.2% on recall value and an average processing time of 0.9 seconds. Performance comparison of the proposed method with some known method is given below in Table 1. However, if size of slanted or curved text is very small then the second order edge detection methods does not produce better text extraction which is also mentioned in[17].

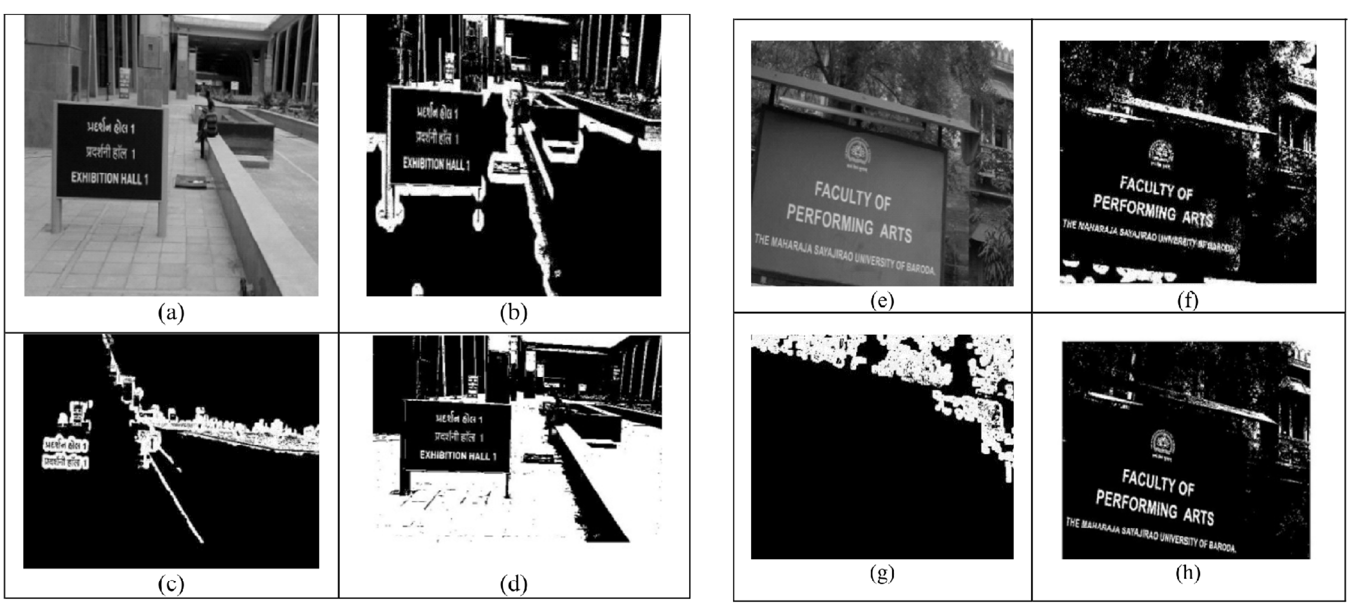

Figure 1: (a) (e) Input Image; (b) (f) Output of Prewitt Edge Detection Algorithm; (c) (g) Output of Gaussian Edge Detection Algorithm; (d) (h) Output of Proposed Method

Figure 2: (a) Input Image form ICDR2015 Training Dataset; (b) Output of Proposed Algorithm

Figure 3: (a) Input Image form [12] (b) Output of Proposed Algorithm

Figure 4: (a) Input Image form ICDR2015 Training Dataset (b) Output of Proposed Algorithm (c) Input image from [16] (d) Output of Proposed Algorithm

Table 1
Comparison of evaluated results

Method	*No. of Images*	*Precision rate (%)*	*Recall rate (%)*
U.Bhattacharya, S. K. Paruri & S. Mondal[12]	100	68.8	71.2
Neumann & Matas [18]	-	73.1	64.7
TH-TextLoc System [18]	-	66.97	57.68
H. Raj & R. Ghosh [16]	62	72.8	74.2
Proposed Method	100	74.8	76.2

II. CONCLUSION

A simple method for extraction of text regions is proposed. The algorithm is quite robust in locating text regions. The proposed strategy has created promising results for natural scene pictures consist of content of various size, text style, and arrangement with changing foundations. The methodology additionally distinguishes nonlinear text areas. The framework proposed in this study can be executed in content acknowledgment programs or significant software.

The proposed algorithm gives robust result for slanted text components as shown in Fig 3 and Fig 1(h). However, the small size of curved or slanted text limits the performance of the algorithm as bias in the localization of the text edges for second order derivative.

REFERENCES

[1] Sharma S., "Extraction of Text Natural Images", IJARCSSE, Vol.3, [2007].

[2] Agarwal P, Varma R "Text Extraction from Images", IJCSET, April 2012, Vol 2, Issue 4, 1083-1087

[3] Gonzalez R.C., Woods R.E., Eddins S.L., Digital Image Processing Using MATLAB, Pearson Education [2004].

[4] M. Swamy Das, B. HimaBindhu, A. Govardhan, "Evaluation of Text Detection and Localization Methods in Natural Images", IJETAE, ISSN 2250-2459, Volume 2, Issue 6, June 2012.

[5] S. Ferreira, V. Garin, and B. Gosselini, "A text detection technique applied in the framework of a mobile camera-based application," in Proc. Int. Workshop Camera-Based Doc. Anal. Recognit. 2005, pp. 133–139.

[6] S. A. Angadi and M. M. Kodabagi, "Text region extraction from low resolution natural scene images using texture features," Advance Computing Conference (IACC), 2010 IEEE 2nd International, Patiala, 2010, pp. 121-128

[7] Giri, P. S., "Text Information Extraction and analysis from Image Using Digital Image Processing Techniques". International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 11, November 2014 pg. 576-582.

[8] Yusuke Matsuda, Shinichiro Omachi and Hirotomo Aso, "String Detection from Scene Images by Binarization and Edge Detection", IEICE, D, Vol.J93-D, No.3, pp.336-344, 2010.

[9] P. K. Sahoo, S. Soltani, A. K. C. Wong, A survey of thresholding techniques, Computer Vision, Graphics Image Processing, 41 (2), 233&260, 1988.

[10] W. Niblack, "An Introduction to Image Processing", Prentice-Hall, Englewood Cliffs, NJ, pp. 115-116, 1986.

[11] B. Sankur and M. Sczgin. "Survey over image thresholding techniques and quantitative performance evaluation ", pp-148, Journal of Electronic Imaging, January 2004, Vol. 13(1).

[12] Bhattacharya, Ujjwal, Swapan Kumar Parui, and Srikanta Mondal. "Devanagari and bangla text extraction from natural scene images." 10th International Conference on Document Analysis and Recognition, pp. 171-175, IEEE, 2009.

[13] Albovik, "Handbook of Image and Video Processing", Academic Press, 2000.

[14] R.M.Haralick, L.G.Shapiro, "Computer and Robot Vision", Volume 1, Addition- Wesley Publishing Company Inc., 1992.

[15] Lindeberg, T. 1994. Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics, 21(2): 224-270.

[16] Raj, H. and Ghosh, R., 2014, September. "Devanagari text extraction from natural scene images". In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on (pp. 513-517). IEEE.

[17] P.R.S. Mendonca, D. Padfield, J. Miller, and M. Turek, "Bias in the Localization of Curved Edges," Proc. European Conf. Computer Vision, part 2, pp. 554-565, May 2004.

[18] Yin, Xu-Cheng, et al. "Robust text detection in natural scene images." IEEE transactions on pattern analysis and machine intelligence 36.5 (2014): 970-983.