International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Index Copernicus Value (2015): 56.67 | Impact Factor (2017): 5.156

Automatic Text Detection in Scene Images

K. Vaidehi[1], S. Bhuvaneswari[2]

[1]Associate Professor, Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Hyderabad

[2]Assistant Professor, Department of Computer Science, Annai College of Arts and Science, Kovilacheri, Kumbakonam

Abstract: ***Automatically detecting and extracting text from the digital images present many challenging research issues. The proposed system detects text automatically using connected component labeling method. In this work, first, color scene images are taken as input image which is converted into gray scale image, then Otsu thresholding algorithm is used to binarize the gray scale image. To detect the connected regions, connected component labeling method is used and to retain the text regions alone, set of selection/rejection criteria is used.***

Keywords: Text Detection, Otsu's thresholding, connected component labeling, Morphological operations

1. Introduction

Text detection and recognition is an area of study which aims to develop a structure to read the text from images automatically. Due to the digitization, most of the ancient documents, land records, old books, periodicals, newspapers, etc are converted to images. An image may contain text embedded on to it, that give important information, and sometimes it needs to be removed for aesthetic and other reason, So text data extraction and manipulation of objects from digital media is essential, for understanding, editing, and retrieving information. Text extraction from images have many useful applications such as searching of images using keyword, license plate detection, document analysis, content based retrieval, object identification, street signs recognition, text based video indexing, video content analysis, etc. On the other hand, some applications need to eliminate the inscribed texts and to restore the image without blurring. So detection of the text is very important for removing unnecessary advertisements, and for artistic reasons also.

The different types of text images are: document text image, caption text image, and scene text image, and the text data which is present in an image is of different font styles, sizes, orientation, colors, and is mostly against a complex background.
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, color etc. In the case of color 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 color 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. 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. The work in this paper deals with detection of scene text only.

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

Fig.1. shows an text detection. Automatic methods for text extraction from images aim to detect the characters based on 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 proposed work is done in two stages: Detection of text in image and extraction of text from the detected region.

The first step in the proposed work is the RGB image converted into grey scale image which is binarized using Otsu thresholding algorithm for text detection. The second step uses morphological operation for extraction of text by eliminating non text regions by dilation and erosion.

2. Related Works

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 review 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. Text detection methods are broadly divided into three kinds: connected component based, texture based, and edge based. Though the connected component based method locates the text very easily, it fails in the complex background [2].

The computational complexity is more in the case of texture based method in the classification stage, and appearance of the text like regions may confuse the detection process [3]. The performance of edge based methods is not appreciable while handling large size texts [4]. The text detection methods range from using simple classifiers [5] to efficient multi-stage processes which employ many algorithms and layers [6] and [7]. Authors in [7] employed a method in which, after binarizing the input image, connected component analysis, using conditional random field (CRF) to detect the text lines. Authors in [8] introduced a new approach for segmenting 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.

In [9] 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 isolated candidate text edges, morphological dilation operator is used. Then, based on the edge map, the line feature graph is generated. Finally, according to the line features, the image is filtered and exact text regions are isolated.

Authors in [10] 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. Text region localization is the second stage that helps to detect the non-text regions. Two constraints are used to find and filter very small isolated blocks whose width is very small compared to the height of the blocks, and in the third stage, character extraction is done using the existing optical character recognition (OCR) engines.

Authors in [11] introduced a method for detecting and extracting text regions from natural scene images whose resolution is low. For removing the constant background, discrete cosine transform (DCT) based high pass filter is used. Then the text blocks are merged to obtain the new text regions. Then, post-processing is done to cover small portions of missed text in adjacent undetected blocks.

The proposed morphological technique is impervious to noise, skew and text orientation [12]. It is also free from artifacts that are usually introduced by both fixed/optimal global thresholding and fixed-size block-based local thresholding. In [13] compares the edge based, and connected component based approaches and analysed the advantages and disadvantages of these systems. In [14], by using multi-scale wavelet features, proposed a novel coarse-to-fine algorithm that is able to locate text lines even under complex background. The authorsin [15] propose an algorithm for selecting text regions and for region segmentation using horizontal projection and geometric properties.

3. Methodology

The work done in two steps .The first step detects text region in image .The second step extracts text from the image. The flow of proposed work is shown in the figure.

Techniques used in this work are:

RGB to Gray scale Conversion
Reading the RGB color image and converting into gray image.

$$I = 0.2989*r+0.5870*g+0.1140*b$$

Where, I is an intensity image with integer values ranging from a minimum of zero. r, g and b are the red, green and blue components respectively.

Otsu Thresholding Method:

- Redesign the two dimensional gray scale image to one dimensional.
- Discover the histogram of the image. (The bins are from 0 to 255)
- Calculate the weight, mean and the variance for the foreground and background
- Calculate weight of foreground* variance of foreground + weight of background* variance of background.
- Find the minimum threshold value.

Connected Component Labeling:
Connected components labeling scans an image and groups its pixels into components based on pixel connectivity, i.e., all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Once all groups have been determined, each pixel is labeled with a gray level or a color (color labeling) according to the component it was assigned to. Connected component labeling works by scanning an image, pixel-by-pixel (from top to bottom and left to right) in order to identify connected pixel regions, i.e., regions of adjacent

pixels which share the same set of intensity values. Connected regions are bounded using bounding boxes. Connected component labeling is applied to the dilated binary image using 8-pixel connectivity.

Morphology operations:
Morphology is a broad set of image processing operations that process **i**mages based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. Dilation adds pixels to the boundaries of objects in an image, while erosion eliminates pixels on object boundaries. The number of pixels added or detached from the objects in an image depends on the size and shape of the *structuring element* used to process the image.

The step by step procedure of proposed work is given below:
1) An RGB image is taken as input
2) The RGB image is converted into gray scale image.
3) The resultant gray scale image is binarized using Otsu thresholding method.
4) Connected component labeling is applied to obtain the various connected regions and the connected regions are bounded using bounding boxes.
5) After applying CCL, the first set of criteria are applied which eliminate non-text macro objects. Here, all objects whose area is greater than 3000 pixels and perimeter greater than 2000 pixels are removed. Still smaller objects of non-text regions are left behind.
6) To eliminate those micro non-text regions, CCL is applied to the resultant image once again and the second set of criteria is applied which eliminate non text micro objects. The objects area which is less than 2000 pixels and perimeter less than 1000 pixels are removed.
7) Then major axis length of each object is calculated. The major axis length lies between 20 and 95 are considered to be text. This length is almost suitable for all types of texts and which is found out empirically.
8) Then the resultant image is eroded, and then dilated to remove the undesired small objects and this image is the binarized text image.
9) Now the image contains only the text regions with the non-text regions fully removed.

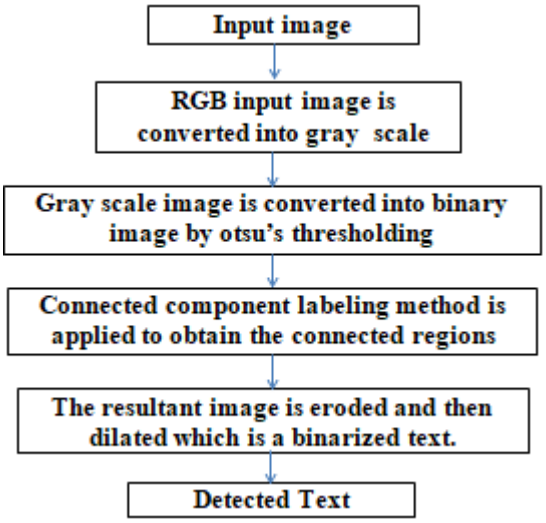

Figure 2: Methodology of the proposed work

4. Experimental Results

The images with simple and complex background are given as input to the proposed system and the outcomes are shown in Fig.3, Fig.4 and Fig.5.

5. Conclusion

In this paper, a method to detect text automatically in colour images is presented. Initially, Color input image is converted into gray scale image. Otsu thresholding method was applied for binarizing the gray scale image. Then connected component labeling method was carried out to detect the text. Finally morphological operations were used for removing the non-text regions. Experimental results have shown that the proposed method can be effectively used to detect the text automatically.

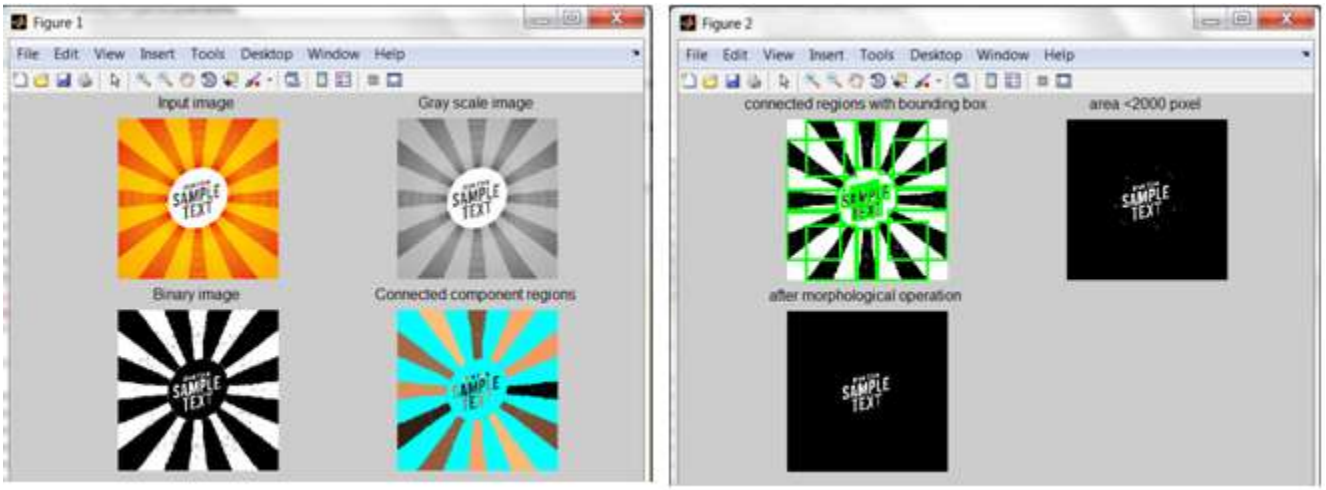

Fig.3. Input- Output of sample1

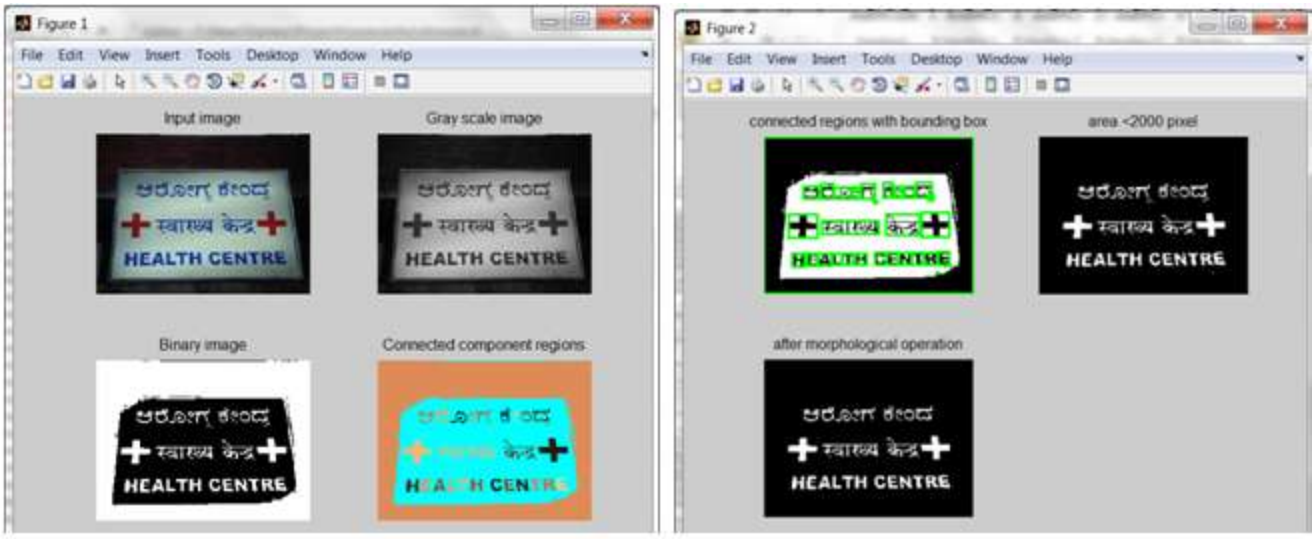

Fig.4. Input- Output of sample2

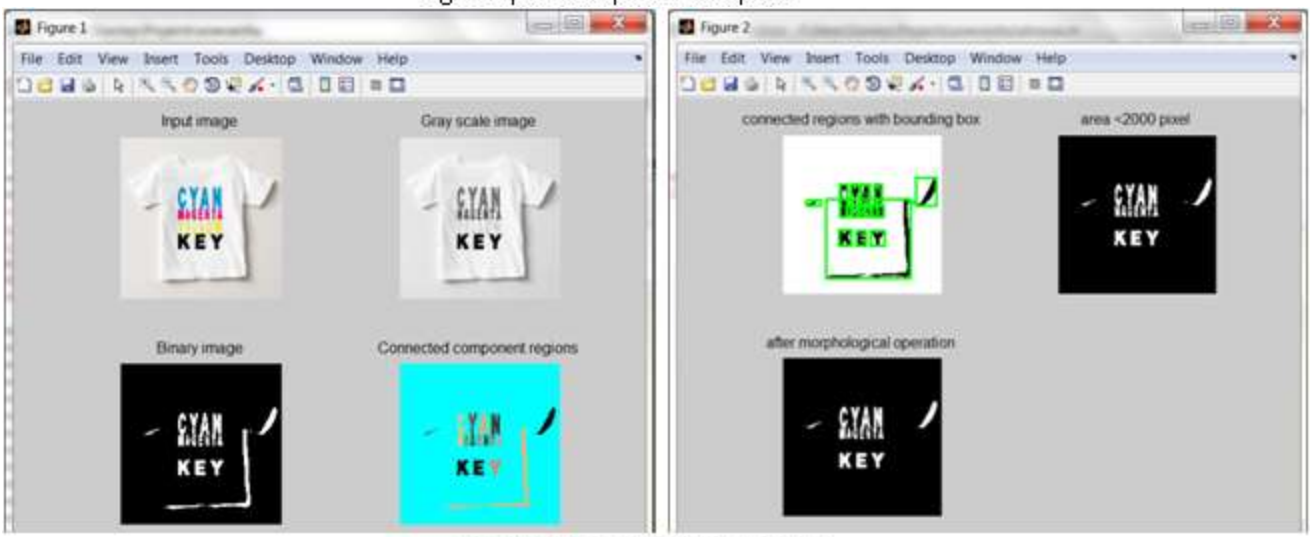

Fig.5. Input- Output of sample3

References

[1] J.Ohya,A.Shio,S.Akamatsu, "Recognizing Characters in Scene Images", IEEE transactions on PAMI, Vol.16, No. 2, pp. 214-224, 1994.

[2] Manjusha.K, SachinKumar.S, Jolly Rajendran, K.P.Soman, "Hindi Character Segmentation in Document Images using Level set methods and Non-linear Diffusion", International Journal of Computer Applications(IJCA), Vol.44-No.16, April 2012

[3] KhurramKhurshd, Imran Siddiqi, Claudie Faure, Nicole Vincent, "Comparison of Niblack inspired Binarization methods for ancient documents", SPIE-ISAT proceedings on the International conference on Document Recognition and Retrieval, 2009

[4] Graham Leedham, Chen Yan KalyanTakru, Joie HadiNata Tan and Li Mian, "Comparison of Some Thresholding algorithms for text/Background Segmentation in difficult document images", IEEE Proceedings on the Seventh international Conference on Document Analysis and Recognition (ICDAR)

[5] P. Dollár, S. Belongie, and P. Perona. The Fastest Pedestrian Detector in the West. In Proceedings of British Machine Vision Conference, volume 2, pages 1–11, 2010.

[6] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral Channel Features. pages 1–11, 2009.

[7] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: An Evaluation of the State of the Art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4):743–761, 2012.

[8] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification. Wiley, 2001

[9] B. Epshtein, E. Ofek, and Y.Wexler. Detecting Text in Natural Scenes with Stroke Width Transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2963–2970. IEEE, 2010.

[10] L. Fei-Fei, R. Fergus, and P. Perona. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories.Computer Vision and Image Understanding, 106(1):59–70, 2007
[11] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysisand Machine Intelligence, 32(9):1627–1645, 2010.
[12] de Campos, T. E., Babu, B. R., &Varma, M. (2009, February). Character Recognition in Natural Images. In *VISAPP (2)* (pp. 273-280).
[13] Sushma, J., &Padmaja, M. (2009, July). Text detection in color images. In *Intelligent Agent & Multi-Agent Systems, 2009. IAMA 2009. International Conference on* (pp. 1-6). IEEE.
[14] Ye, Q., Huang, Q., Gao, W., & Zhao, D. (2005). Fast and robust text detection in images and video frames. *Image and Vision Computing*, *23*(6), 565-576
[15] Choksi, A., Desai, N., Chauhan, A., Revdiwala, V., & Patel, K. (2013). Text Extraction from Natural Scene Images using Prewitt Edge Detection.