Text Region Detection And Extraction From Road Direction Sign Boards

Prashant Madavanavar [1], Vinaykumar Patil [2]

[1] *Department of CSE, Basaveshwar Engineering College, Bagalkot, India*
[2] *Department of CSE, Basaveshwar Engineering College, Bagalkot, India*

Abstract

Text detection and Extraction in natural scene images is important for automated systems for understanding display boards Which can be used for applications in providing location aware information and in a recognition system This problem is challenging due to the complex background, the non-uniform illumination, the variations of text font, size and line orientation. The objective of the proposed method is to design and develop a system to extract and segment the text part from a road direction sign images. Several software like OpenCV comes preloaded with OCR functionality by means of which the system can detect characters in images. However for the complex scene the detection of text part becomes complex as there are several other objects which presents a character like features. Moreover there are other issues like edge from hoardings and signboards which affects the performance of the system. Therefore in the proposed which is based on edge detection and morphological operation aim to extract the text part from the images using by appropriately removing the background such that the process The proposed method is robust and achieves a detection rate of 95.5% on a variety of 100 natural scene road direction sign boards images each of size 250x250.

Index Terms- Connected component, Natural Scene Images, Road Sign Board, and Text Extraction.

1. Introduction

With the increasing use of digital image capturing devices, such as digital cameras, mobile phones and PDA's [1], the image captured by these devices contain huge amount of useful data, that's why it is very important to detect and identify the text region as accurately as possible before performing any character recognition. The research field of scene text recognition receives a growing attention due to the proliferation of digital cameras and the great variety of potential applications, as well. Such applications include number plate recognition robotic vision, image retrieval, intelligent navigation systems and applications to provide assistance to visual impaired persons. Natural scene images usually suffer from low resolution and low quality, perspective distortion and complex background. In the last decade, many methods have been proposed to address image and video text detection and localization problems [7], and some of them have achieved impressive results for specific applications. However, fast and accurate text detection and localization in natural scene images is still a challenge due to the variations of text font, size, color and alignment orientation, and it are often affected by complex background, illumination changes, image distortion and degrading. Although the existing methods have reported promising localization performance, there still remain several problems to solve.

Generally text detection methods can be classified as both edge-based, connected-component based and region -based methods, Region-based methods attempt to detect and localize text regions by texture analysis [7]. Generally, a feature vector extracted from each local region is fed into a classifier for estimating the likelihood of text. Then neighboring text regions are merged to generate text blocks. Because text regions have distinct textural properties from non-text ones, these methods can detect and localize texts accurately even when images are noisy. On the other hand, CC-based methods directly segment candidate text components by edge detection or color clustering then the non-text components is then pruned with heuristic rules or classifiers [15]. The region based methods, the speed is relatively slow and the performance is sensitive to text alignment orientation. On the other hand, CC-based methods cannot segment text components accurately without prior knowledge of text position and scale, so here to overcome the above difficulties and to detect text regions in the digital image by making advantages of different methods.

In the proposed system, a edge detection approach along with morphological for text detection and extraction is carried out which is based on extracting the strong edges comparison to other edges present in the image and then the connected component analysis

is made in order label component and from the labeled components select the unique set of labels and mark those label in to the original image in order extract the text part appropriately removing the background. The proposed method is enough to detect text regions from road direction sign boards and achieves a detection rate of 95.5%. The system is developed in MATLAB and evaluated for 100 natural scene images by considering the challenges such as illumination and the different font size, this method is developed on Intel core i3 (2.2GHz) computer. It was observed that the processing time lies in the range of 1.02 to 2.5 seconds due to varying background.

2. Present System

There are several text extraction techniques being proposed by different authors. This technique relies on enhancing the edge description against background descriptor and assumes that text part's background is structure. Some of the commonly used techniques are DCT based techniques and wavelet based techniques. In DCT based techniques the frequency component of the images is obtained and it is threshold based on the type of image and text properties [7]. Inverse DCT eliminates the background and returns the foreground. Similarly in wavelet based techniques image is subsequently decomposed. As the text part is very high edge preserving in nature, text part is retained in decomposed image. Their position is tracked in the decomposed image and is applied as a mask to extract the actual text part from the main image. However it is important to know here is that several other parts of the image can be edge preserving like the hoarding edges and other objects like vehicles. Therefore both of the techniques fail to detect text against complex background. Further speed of both the techniques is high due to resource consuming operations and methods are implemented edge-based method and connected component labeling method known as blob extraction method [3], [13] and connected component labeling with CRF model [8].

One more category of Text extraction relies on classification problem. One of the most widely used classifier is Haar cascade. First a large database is build with positive sets and negative sets. haar features are extract ted from both the sets and are used as classification weight for a given image. However this technique relies on the strength of the database and that is ideally more than 2000 image large as par basic opencv haar documents.

3. Proposed System

Our work is based on our finding that Text part in an image not only depends upon the edge strength of the text part but also the dependence on type of edges and property of edge in comparison to other edges present in the image. Therefore we use morphological thresholding rather than thresholding the image in color domain. Finding out the connected components which are more prominent in comparison to other components gives better statistics of the text part. Hence we first apply median filter to smooth the images and reduce the edge affect of smaller edges and then binaries the image based on first a sobel descriptor and followed by morphological thresholding on number of connected components on each and every extracted text part. As the system does not need any complex resource consuming operation and does not need a classifier, it works extremely well in real time.
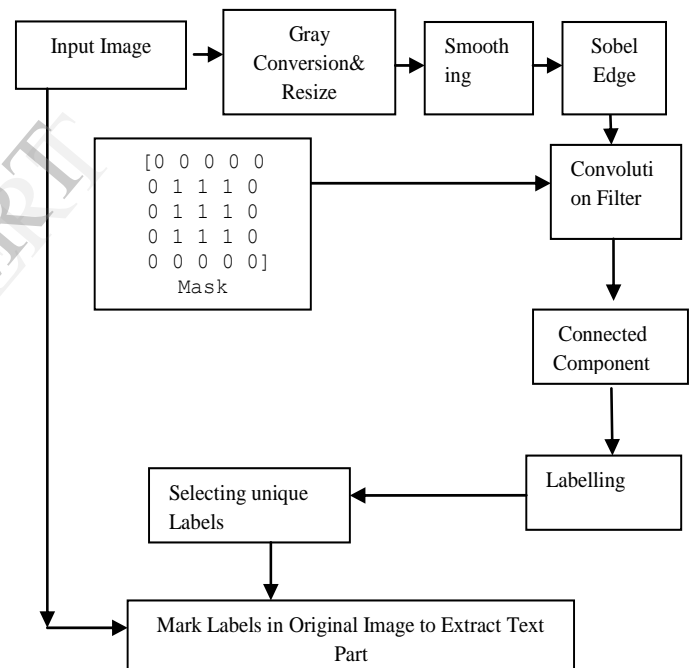

Fig 1: Block Diagram for Proposed Model

3.1 Methodology Steps

3.1.1. Median Filtering: A road side image is generally texture and structure images where the image may have homogeneous background or heterogeneous background. As the detection operator in our work is morphology based, we need to convert the image into a pure structure image by minimizing the texture effect. It is done through applying median filter. A Median filter presents a smooth image where numbers of colors

are reduced significantly in comparison to original image. Another role of median filter is to reduce number of edges. Median process smooth's the image around edges. As the text part presents strong edges, edge descriptor for text part after median filter are stronger than other parts.

3.1.2 Text Mask: A text part in the image is one where a character will be present against a background and there would exist a gap between two characters in both horizontal and vertical direction. Therefore we can say any character or text part in the image segment is one where center part is stronger in edges than their background. In order identify every part of the image that satisfies this principle we define a mask which clearly shows that the centre is stronger edges than the neighbors.

$$[0\ 0\ 0\ 0\ 0$$
$$0\ 1\ 1\ 1\ 0$$
$$0\ 1\ 1\ 1\ 0$$
$$0\ 1\ 1\ 1\ 0$$
$$0\ 0\ 0\ 0\ 0]$$

3.1.3 Edge Detection: From details presented in 3.1.2 it is clear that text part will present very strong edges, no matter the shape or font of the characters. Detection of text therefore must be considered as detection and selection of appropriate edges from the image. Characters could be either horizontal or vertical or diagonal depending upon the orientation of the camera. Sobel combines all three edge descriptor by defining two separate masks: one for horizontal and one for vertical gives the diagonal operator. Therefore Sobel is selected as edge detector rather than kris and prewitt which ideally detect edges in all directions with equal weight.

$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * A \quad \text{and} \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A$$

$$G = \sqrt{G_x{}^2 + G_y{}^2}$$

3.1.4 Convolution Filter: Convolution filter is applied locally on an image chunk of size of convolution kernel. Convolution process is defined as multiplication followed by addition. Edge image's components are multiplied with respective component of mask and the result of addition is put in the centre of the pixels. Hence this operation extracts the strongest edges that are present against a background. It includes text part

as well as edges in the board which are of course strong edge against the rest of background.

3.1.5 Text Extraction using Connected Component: As a result of step 3.1.4, a binary image with text part and some other parts reflecting strong edges against background, now need to remove edges that are isolated and get hold of only the strong edges. If it is known how many binary points are connected to each other then easily apply a threshold and eliminate those set of binary one pixels which are not spread over large number of pixels. This is done by first extracting the connected components. It returns number of pixels each group or set of binary one pixels are connected, Now need eliminate isolated groups by first labelling the groups and From the labeled image which contains text and non text regions, in order to eliminate non text regions the number of count of connected component lesser than 40% are considered this indicate that the region's which are smaller than three pixels are consider as non text part and clearing it from the resultant image this how the text objected are collated.

4. Experimental Results and Discussions

This method has been tested over a 100 road direction sign boards as well as other natural Scene text images which are selected from standard data set of type including handwritten text, scene text in which text has different font size, color, orientation, and alignments. These images are analyzed to demonstrate the performance of the proposed algorithm. Performance is verified with the oriented text in horizontal & vertical direction with different languages (English & Kannada). Various metrics have been evaluated from the tested results.

4.1 Performance Analysis

Metrics used to evaluate the performance of the system are Precision; Recall. Precision and Recall rates have been computed based on the number of Correctly Detected Characters (CDC) in an image, in order to evaluate the efficiency and robustness of the algorithm. The metrics are as follows:

Definition 1: False Positives (FP) / False alarms are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text.

Definition 2: False Negatives (FN)/ Misses are those regions in the image which are actually text Characters, but have not been detected by the algorithm.

Definition 3: Precision rate (P) is defined as the ratio of correctly detected characters to the sum of correctly detected characters plus false positives as represented in equation below.

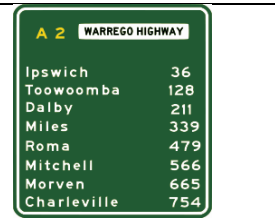**P= number of correctly detected text / [correctly detected text +FP]**

Definition 4: Recall rate (R) is defined as the ratio of the correctly detected characters to sum of correctly detected characters plus false negatives as represented in equation below.

R= number of correctly detected text / [correctly detected text +FN]

4.2 Result and discussion

In order to evaluate the performance of the proposed method. We used more than 100 test images which are natural Scene road direction sign boards and the main advantage of the work is that the system detects the texts without presence of any classifiers. Therefore detection is faster and can easily be adopted in the real time. The work is independent of tilt and background color. Hence detection speed is faster. Further the system is capable of extracting multiple text areas and lines. Therefore it can be used as a preprocessing step for character recognition in natural scenes. It can also be used in automated driver guidance system which relies on detecting road signs and messages to inform driver about various aspects of the road. The result obtained by the proposed algorithm is presented in the Table I and the number of true positive ,false positives and precision rate and the recall rate are and corresponding values are listed ,average of all measure is calculated and presented in the Table II.

Table I: Text Extraction results of processing natural scene images dealing with various issues.

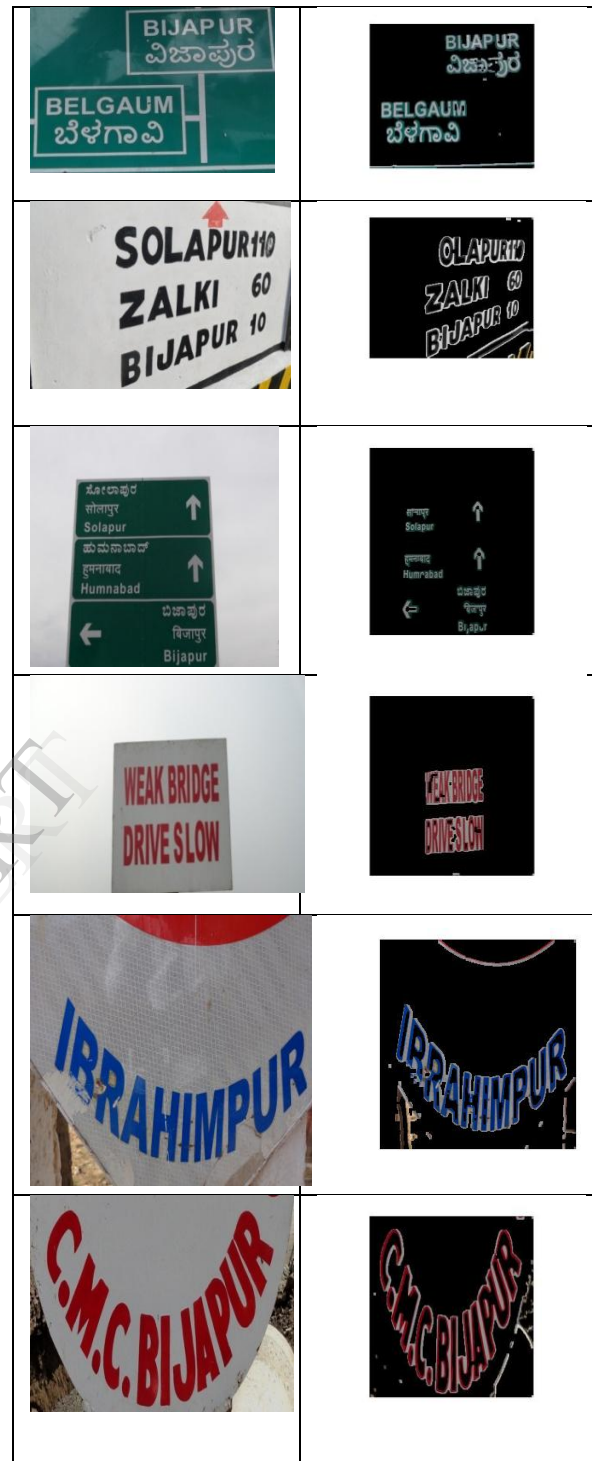| Natural Scene Input Images	Text Extracted output image

Table II: Presents the performance of the system

Operation	Precision rate (P)	Recall rate(R)	Genuine accept rate	False Accept ratio
Text Extraction	93.85%	95.01%	95.5%	5.34%

4.3 Comparison with other text extraction techniques

To give an average estimate of the performance of the text extraction the results have been compared against two existing algorithms [1] in which the four connected component method is used which is done using Sobel edge detection and connected component extraction and rule based connected component filtering and other method [22], have used the aspect ratio to identify the text and non text regions within an image and another method called adaboost for text extraction in natural scene image .The proposed method in order to extract the text from non text object not used any classifier or heuristic rules in the proposed method after applying the Convolution Filter it extracts the strongest edges that are present against a background. It includes text part as well as edges in the board which are of course strong edge against the rest of background. And then connected component is used In order to group the strong edges once we find strong edges against the background assumes as text part which are extract from the original image. This algorithm is in sensitive to skew and text orientation, the output of the text extraction algorithm is fed to an OCR system to recognize the contained information. The main objective of the text extraction algorithm is to reduce the number of false text candidate that may be fed to the OCR and the graph plotted by comparing other method is shown in figure2.
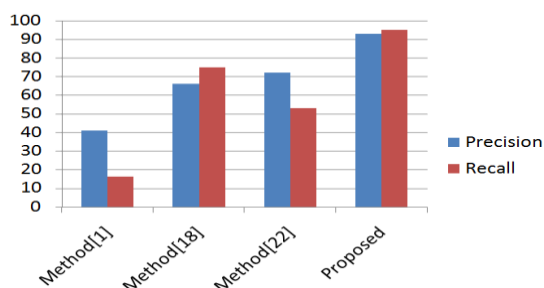
Figure 2 Comparison with other technique

5. Conclusion and future work

Text extraction is one of the most important parts of OCR. However it is not widely researched are. It is found out that the performances of good OCR systems are quite poor for road side images. It is due to the incapability of the techniques to separate the text part from the images. Haar based text detection is gaining popularity but is restricted due to large database requirement. It has also limitation as different text in different languages has different writing style and features. Training a classifier with several of thousands of languages written in the world is quite tough. Therefore in this work proposed an efficient system for text extraction without the need of any resource consuming transforms or classifiers. The technique works perfectly under most complicated backgrounds including several other objects and the experimental results show good recall rate and precision rate of the method is average of 95.01 % and 93.81 % respectively the only drawback observed was remaining of traces of edges for large edges. This can be overcome by introducing threshold for rejective large edges the way have removed smaller one. Future work will include the integration of an OCR module and extension of our test database. Using feedback information from this module, a further increase in robustness of text identification is expected with respect to natural scene images.

References

[1] Nobuo Ezaki 2004 "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons," Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04).

[2] G.Ramamohan babu "Text extraction from heterogeneous images using mathematical morphology," Journal of Theoretical and Applied Information Technology © 2005 - 2010 JATIT. All rights reserved.

[3] Xiaoqing Liu., "Multi-scale edge-based text extraction from complex images," The University of Western Ontario Department of Electrical & Computer Engineering London, Ontario, N6A 5B9, Canada,IEEE

[4] Krishna, "Character-Stroke Detection for Text-Localization and Extraction," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) IEEE.

[5] Palaiahnakote Shivakumara, "A New Method for Handwritten Scene Text Detection in Video".

[6] Apurva Srivastav, "Text Detection in Scene Images using Stroke Width and Nearest Neighbor Constraints''.

[7] S. A. Angadi , M. M. Kodabagi(2009) , "A Texture Based Methodology For Text Region Extraction From Low Resolution Natural Scene Images ", International Journal Of Image Processing (Ijip) Volume(3), Issue(5).

[8] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu(2009), "Text Localization In Natural Scene Images Based On Conditional Random Field," *ICDAR,*pp 6-10.

[9] .J. Fabrizio, M. Cord, And B. Marcotegui(2009), "Text Extraction From Street Level Images,", *CMRT*, Vol. Xxxviii, Part 3/W4 , pp. 199–204.

[10] Boussellaa , Aymen Bougacha, Abderrazak Zahour, Haikal El Abed, Adel Alimi(2009) ,"Enhanced Text Extraction From Arabic Degraded Document Images Using Em Algorithm", *10th International Conference On Document Analysis And Recognition*.

[11] Miriam Leon, Veronica Vilaplana, ''caption text extraction for indexing purposes using a hierarchical region-based image model", 2009 IEEE.

[12] Hinde ANOUAL, Driss," Features Extraction for Text Detection and Localization", 2010 IEEE.

[13] Kohei Arai1, Herman Tolle(2011)," Text Extraction From Tv Commercial Using Blob Extraction Method", *International Journal Of Research And Reviews In Computer Science* Vol. 2, No. 3.

[14] Wonder Alexandre Luz Alves And Ronaldo Fumio Hashimoto(2010),"Text Regions Extracted From Scene Images By Ultimate Attribute Opening And Decision Tree Classification", Proceedings of the 23rd Sibgrapi Conference On Graphics, Patterns And Images.

[15] Yi-Feng Pan, Xinwen Hou,"A Hybrid Approach to Detect and Localize Texts in Natural Scene Images". IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 3, MARCH 2011.

[16] Basilios Gatos, Ioannis Pratikakis and Stavros Perantonis," Towards Text Recognition in Natural Scene Images", Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, NCSR "DEMOKRITOS" Athens 153 10, Greece.

[17] Chucai Yi and YingLi Tian ," Text String Detection From Natural Scenes by Structure Based Partition and Grouping", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 9, SEPTEMBER 2011.
[18] Jung-Jin Lee," AdaBoost for Text Detection in Natural Scene", 2011 International Conference on Document Analysis and Recognition.

[19] Luk´aˇs Neumann,"Real-Time Scene Text Localization and Recognition",IEEE conference june 2012.

[20] Karin Sobottka," Identification of Text on Colored Book and Journal Covers," Institute of Informatics and Applied Mathematics University of Bern, Neubr¨uckstrasse 10, CH-3012 Bern, Switzerland.

[21] Shyama Prosad Chowdhury ," Robust Extraction of Text from Camera Images using Colour and Spatial Information Simultaneously," Journal of Universal Computer Science, vol. 15, no. 18 (2009), 3325-3342 submitted: 26/10/09, accepted: 20/11/09, appeared: 28/12/09 © J.UCS

[22] Andrej Ikica, Peter Peer," An improved edge profile based method for text detection in images of natural scenes".