Text Detection and Recognition from Natural Scene

Tajinder Kaur

Department of Computer Engineering
Punjabi University
Patiala, India

Nirvair Neeru

Department of Computer Engineering
Punjabi University
Patiala, India

Abstract— **Many techniques and algorithm have been developed to solve the problem of text extracted from natural scenes. Text extraction is emerging and challenging era in the computer vision. Text which is embedded into the image contains semantic information which is used in many other applications such as information retrieval of complex images, robot navigation, useful for visually impaired persons, street signs, automatic read the sign board and use in so many other applications. Most of the research work in this area has been done only on printed text, a very few research is addressing the LED scene text. Scene text is difficult to extract due to blur image, variations in color, noise problem, complex background, discontinuity, poor lighting conditions, and variation in illumination. LED is which is widely used in displaying the information in boards. Now days LED display that is natural scene is being widely used for displaying announcements, sign boards, banners for displaying information. To extract the text from the LED display is not an easy task, it is very complex due to its discontinuity. So, The aim of this paper to propose a technique to extract the two type of LED text from natural scene image. The first step of the algorithm is preprocessing of the image where the image is converted from RGB to grayscale, noise is removed and the image is converted to binary image, etc. Then the text is localized. After that connected component approach is used for text detection and finally the text has been recognized using template matching with correlation. The experimental results of the proposed method show the detection and recognition rate is 82.87 and 57.6 .**

Keywords— ***Connected Component method, Corelation, Dilation, FFT method, Morphological operations, Template matching, Text detection and recognition.***

I. INTRODUCTION

Text extraction method is a challenging task due to rapidly increase the digitization of all the material. It is complex because we need to find out where the text actually located in the image and how much part is not relevant to the recognition process. A text extraction in natural scene contains useful and valuable information and makes it easy which can be understood by human and computer. This research topic is very active and challenging task in computer vision applications. Text extraction process involves text detection, localization, extraction, segmentation and recognition of text [1]. Text extraction in natural scene image use in many applications such that mobile text recognition, scene understanding, automatic recognition sign board, supports for visually impaired persons, license plate detection, robot navigation, extract traffic sign board text that uses for intelligent transport system, navigational support for tourist guide, information retrieval etc..These are the variety of application which is develop with the use of mobile phone because mobile phone captured the scene text and directly convert into the recognition process. So, when we perform any text recognition, it is very important to extract the text region accurately. Text extraction in natural scene image use for so many applications, but still it is challenging task due complexity of its complex background, color variation, noise problem, image illumination changes, image distortion, blurring problem and lighting condition [2]. To extract the text from the LED display is not an easy task, it is very complex due to its discontinuity.The text which is superimposed into an image contains a useful text which represents the whole image information. Text is mainly classified into two categories:

a. Scene Text
b. Artificial Text.

The scene text is also known as ‘graphics text’. Scene text means the text which is shown in natural image .The artificial text is also known as ‘caption text’ which is superimposed on the image[3]. Following show the two kinds of images:

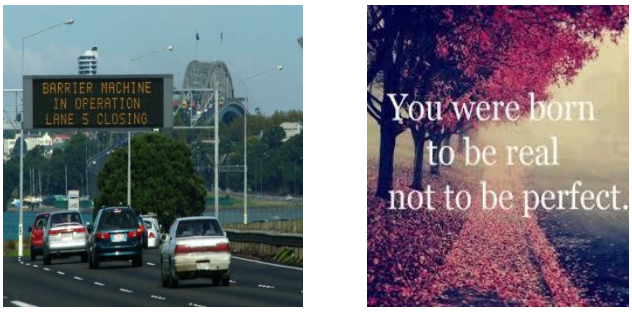

Figure 1 a) . Scene text b). Caption Text

Scene text (Figure. 1 (a)) is a natural type of text which is accidently happening when we capture the image. It contains the useful information which helps to understand the whole

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

idea about the image. There are some examples of scene text is like a vehicle number plate, street signboards, banners, traffic sign board and so on. This text is difficult to extract due to their various styles, font, color, contrast, complex background, low and high resolution, orientation, alignment, blurring and shadowing effects.

Artificial text (Figure 1 (b)) is also called the 'caption text' which is inserted in the image or video. This text could be segmented, detected and extracted using various techniques. The caption text is added into news channels, movies and videos where the subtitle is superimposed. Caption texts are rotating text, subtitle text, moving text. Artificial text may or not a fixed in position and shape and low resolution problem.

Various existing methods of text detection and extraction for natural scene can be roughly classified into two categories: region based technique, texture based techniques, hybrid technique and morphological based method [4-5].

A. Region based method

Region based method is also known as sliding window based method that uses a bounding box or sliding window to detect a text from a natural scene and use some heuristic technique to recognize text. This method uses a bottom up approach in which small component is successively combined into large until all the area are identified in the image. A morphological operation is used to merge the component and filter out the false region and mark the boundary around each text. In this approach a text region is identified from a complex background and removes the false or non-text region. This approach is based upon color, edge, shape, contour and geometry features [3] [4]. On the basis of these features separates text or non-text region. The speed of region based method is slow as compared to other techniques. Edge based and Connected Component is a further classification of the region based approach.

a. Edge based Method: Edge based method are an efficient method for text extraction.Its aim to find out the high contrast between text and the background. Edges are the most important feature of text character rather than orientation, color, layout, etc [5].For detecting the text in images we can use these operators i.e The Canny edge detector, Robert edge detector , Prewitt edge detector and Sobel edge detector .There are the main three properties of the text which is density, edge strength and orientation variance that is superimposed into the image[6].For detecting the boundary of text reliably these properties are mainly used.After that merge the boundary of text and separate the text or non text from the image.

b. Connected Component based method: Connected Component method directly extract the Candidate character from natural scene image by using color clusering and edge detection[6]-[7].Segment the connected component and then mege it after that the false positive components are removed using some classifier or heurstic method.CC method computation cost is low as compared to other techniques and the extracted candidate component directly use in recognition process.Without the prior information about the text position or scale the connected component mehod can't extract the candidate character effeciently.

B. Texture based method

The texture based method uses different texture properties to extract a text or to decide wheather or not the pixels belongs to the text of the image[8]. This technique uses the textual properties that separate them from complex background or non text region. Various methods are used in this approach to extract textual information like Wavelets, Fast Fourier Transform and Gabor filters, DCT Transform Wavelet and variance is used to find the textual properties of the text region in the image. A train classifier is used to extract the features of the target image region [9]-[10]. The main aim of train classifier is to distinguish the text or non-text region for a scene.

C. Hybrid technique

The hybrid technique uses a combination of both techniques, i.e. region based and texture based approach. In this, the first step region based approach is used to detect a text or character candidate using the CC method. The features are extracted from text region and use a classifier to decide which region contains a text or non-text on the basis of texture based method[11]-[12]. The main disadvantage of these approaches that the single method is not suitable for all the natural scene images due to size, color, font variation varies from one image to another image.

D. Morphological based Method

Morphological method is based upon geometrical and mathematics approach for character recognition and image analysis[13].It is used to extract the constrast feature of text in the input image.The feature of the geometrical image is never changed when a specfic transformation is applied on it the transformation like translation, scaling and rotation.The feature of the image is still maintain when the text color or even lighting conditions is changed of the image.

As mention above methods, we use a connected component method and PCA for recognition purpose in our proposed method.Our proposed method we work on LED text rather than printed text. LED is a Light Emitting Diode which is widely used in displaying the information in LED boards[14]-[15]. Now days LED display that is

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

natural scene is being widely used for displaying announcements, sign boards, banners for displaying information.LED text is difficult to extract because of its discontinuous nature. A matrix of segments is used to display the character of LED, which is combined together to generate an LED text.The character of the LED text is displayed in matrix form which is coming in rectangular or circle shape.The size of the matrix of the LED character 3× 5, 4×6, 5×7, 5×8 and 6×7 dots [15].

II. PROPOSED METHODOLOGY

The objective of this proposed method to develop a system that will able to detect and recognize the two types of LED display board text in natural scenes. The first step of the algorithm is preprocessing of the image where the image is converted from RGB to grayscale, noise is removed and the image is converted to binary image, etc. Then the text is localized. After that connected component approach is used for text detection and finally the text has been recognized using template matching with correlation. The flow chart of the proposed method describe below:

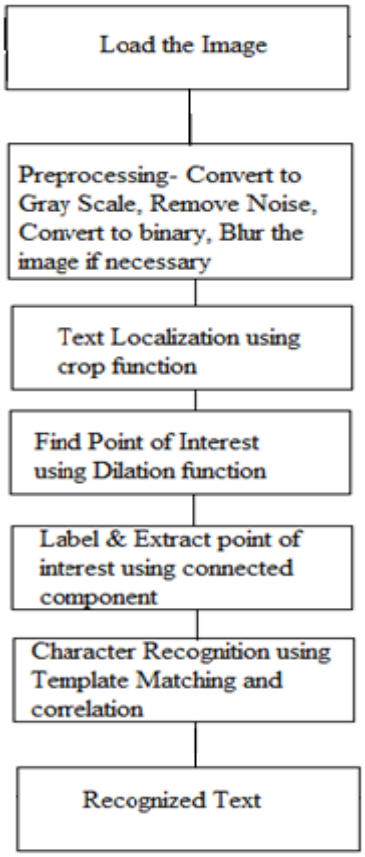

Figure 2. Flow Chart

A. *Load the Image:* Load the input image.

B. *Preprocessing*: The preprocessing is a very important step in text extraction method. Preprocessing steps are necessary to improve the performance and make the process efficient to the time. This includes gray-scaling, filtering to remove noise, binarization of image and blurring the image.

C. *Text Localization:*
Text localization means where the text in the image is accurately located. In this step, the text region is crop the blur text in the image. Then text region is continuously filled by morphological processes (dilation and erosion) and then the text region on the input image can be cropped as an output image result by using region labeling and regional properties.

D. *Find Points of Interest and Label and Extract the point of interest*
Point of interest refers to each text character in the image. Each character is itself a connected component. A set S of pixels is a connected component if there is at least one path in S that joins every pair {p, q} of pixels in S, The path must contain only pixels in S.
For finding the connected components the following iterative formula is used.

$$X_k = (X_{k-1} \oplus B) \prod A \qquad (1)$$

Where k=1,2,3,….

- Step by step method has been explained below for finding, labeling and extracting the connected components.
- Define the structuring element.
- Initialization the label matrix with zeros.
- Find the non-zeros element position in the input image matrix A.
- Initialize the matrix X with zeros and place 1 in the non-zeros element position found in the previous step.
- Perform dilation using the structuring element B on matrix X(imdilate(X,B)).
- Perform intersection with the matrix A(Y = A&imdilate(X,B)).
- Check whether Y==X. If no, then perform the previous two steps again else stop the iteration.
- Find the non-zeros element's position in the Y.N is for labeling the connected component.
- Similarly, place the zeros in those positions in input matrix A.
- Again, find the non-zeros element's position in the matrix A.
- Using the label the connected components can be extracted.

E. *Represent the character with different colors*
After the text is label and extracted from the natural scene image, then represent the extracted character

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

using RGB colors. In this for each label, unique color is used, i.e. every label is represented by different component.

F. Character Recognition

Recognition used to determine similarities between two entities (points, cures, or shapes) of the same type. In the proposed system for recognition template matching has been used. In template matching, a template or prototype of the pattern to be recognized is available and is matched against the stored template. The best match found in the stored template is chosen as the recognized character. The best match is chosen on the basis of correlation between the character to be recognized and the character in the stored template.The correlation method computes the correlation coefficient between two character images-trained and test image and the trained images which have maximum correlation with the character image to be tested is chosen as recognized character.

Correlation method:

When two variables co-vary, there exists a relationship between them. Correlation Analysis is one of the most widely used and reported statistical methods in summarizing medical and scientific research, emphasize on the Interpretation of Correlation Coefficient.The sample correlation coefficient (r) measures the degree of linearity in the relationship between X and Y or strength and direction of linear association between two variables.

$$r = \frac{\sum x.y}{\sqrt{\sum x^2 \ \sum y^2}} \qquad (2)$$

where,

$$x = X - \overline{X}$$

and

$$y = Y - \overline{Y}$$

r = 0 indicates no linear relationship exists.

Direct Method for Calculating "r":

$$r = \frac{N.\sum XY - \sum X.\sum Y}{\sqrt{N\sum X^2 - (\sum X)^2}.\sqrt{N\sum Y^2 - (\sum Y)^2}} \qquad (3)$$

The r is the correlation coefficient between train character component and tested character component. The character in the stored template which has the maximum correlation coefficient with the character to be recognized is selected as the recognized character from the stored template.

III. EXPERIMENTAL RESULTS

For obtaining the results first of all databases containing 15 natural scene images has been collected and both the systems have been tested on these 15 images.For detection precision, recall and f-measure has been calculated and for recognition purpose Recognition true rate (RTR) and Recognition false rate (RFR) has been calculated for each image using the following formula:

- Detection Recall (DR) = TDT/ATC
- Detection Precision (DP) = TDT/(TDT+FDT)
- Detection F-measure (DF) = (2*DR*DP)/(DR+DP)
- Recognition True Rate (RTR) = TRT/ATC
- Recognition False Rate (RFR) = FRT/ATC

The result of the number of detected and recognized characters has been shown in following table.

Image No.	Total Characters	(Proposed system)	
		Detected	Recognized
Image 1	10	10	7
Image 2	13	13	8
Image 3	10	8	3
Image 4	18	12	6
Image 5	4	4	3
Image 6	8	8	3
Image 7	6	6	5
Image 8	7	7	3
Image 9	5	5	3
Image10	5	5	1
Image11	5	5	4
Image12	3	3	3
Image13	8	7	5
Image14	6	5	5
Image15	3	1	1

Table 1. Number of detected and recognized characters

Based on the above table results the parameters- recall, precision and F-measure has been calculated for both the systems and result has been shown in the table 2.

Image No.	Proposed Method		
	Recall	Precision	F-measure
Image 1	100	100	100
Image 2	100	100	100
Image 3	80	100	88.8
Image 4	66.66	60	62.52
Image 5	100	57.14	72.72
Image 6	100	88.88	94.11
Image 7	100	100	100
Image 8	100	77.77	87.49

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

Image 9	100	100	100
Image 10	100	83.33	90.9
Image 11	100	62.5	76.92
Image 12	100	100	100
Image 13	87.5	58.33	69.99
Image 14	83.33	55.55	66.66
Image 15	25	14.2	18.11

Table 2. Precision, Recall and F-measure results

For recognition results, recognition true rate (RTR) and recognition false rate (RFR) has been calculated and results have been shown in following table.

Image No.	Proposed Method	
	RTR	RFR
.Image 1	70	30
Image 2	61.5	38.46
Image 3	30	50
Image 4	33.3	33.3
Image 5	75	25
Image 6	37.5	62.5
Image 7	83.33	16.66
Image 8	42.8	57.14
Image 9	60	40
Image 10	20	80
Image 11	80	20
Image 12	100	0
Image 13	62.5	25
Image 14	83.35	0
Image 15	25	0

Table 3. Table for Recognition true rate and Recognition false rate

On the basis of all the above results overall detection rate and recognition rate of the proposed has been calculated.

	Overall	
	Detection Rate	Recognize Rate
Proposed method	82.87	57.6

Table 4. Overall detection and recognition rate

IV. CONCLUSIONS

In this paper, an efficient method of text extraction from a natural scene image has been presented. Text detection and extraction from natural scene images has applications in many fields such as license plate detection, automatic street sign translation, image retrieval and help for visually impaired people. Text extraction in natural scene images has become a challenging problem due to complex background, variations in scene, scene text distorted by the perspective projection and camera settings. A database of 15 different images has been collected for testing which contains LED images and simple natural scene images. The basic steps for text extraction and recognition are – preprocessing, text localization, finding and labeling connected components and then extract the connected components and finally recognize the characters.
For detection precision, recall and F-measure has been calculated and for recognition truly recognized text (TRT) and falsely recognized text(FRT) has been calculated. By calculating these parameters, it has been found that proposed system gives better results than existing system with detection rate 82.87% and recognition rate of 57.6%. The proposed has worked on the images which contain single color text. In future, the work can be extended to images having different color of text.This technique can't detect text which is continuous. Sometimes the proposed system is not able to detect and extract text properly because of some factors like the image may be tilted, there may be some shadow area or the background may be complex background. So in future these aspects should be considered for better accuracy.

REFERENCES

[1] Yi-Feng Pan, Xinwen Hou, and Cheng-Lin Liu, "A Hybrid Approach to Detect and Localize Texts in Natural Scene Images", IEEE Transactions on Image Processing, Vol. 20, No. 3, March 2011.

[2] Anhar Risnumawan, Palaiahankote Shivakumara, Chee Seng Chan and Chew Lim Tan, "A Robust Arbitrary Text Detection System For Natural Scene Images", Expert System with Application 41(2014) 8027-8048.

[3] Prachi R. Dussawar and Prof. Parul Bhanarkar Jha, "A Survey on Comparision and Performance Analysis of Text Extraction Techniques", IJARSE, Vol. No.4, Issue No.01, January 2015.

[4] H. Raj, R. Ghosh, "Devanagari Text Extraction from Natural Scene Images", 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2014, pp. 513-517.

[5] Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao, "Robust Text Detection in Natural Scene Images", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 36, no. 5, May 2014.

[6] Shivananda V. Seeri, J. D. Pujari and P. S. Hiremath, "Multilingual Text Localization in Natural Scene Images using Wavelet based Edge Features and Fuzzy Classification", IJETTCS , Volume 4, Issue 1, January-February 2015.

[7] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in Proceedings of International Conference on Computer Vision and Pattern Recognition , 2012, pp. 3538–3545.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

[8] Cong Yao, Xiang Bai and Wenyu Liu, "A Unified Framework for Multi-oriented Text Detection and Recognition", IEEE Trans. On Image Processing, Vol. 23, No. 11, November 2014.

[9] Shraddha Naik and Sankhya Nayak, "Text Detection and Character Extraction in Natural Scene Images", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015.

[10] Chucai Yiand Yingli Tian, "Localizing Text in Scene Images by Boundary Clustering, Stroke Segmentation, and String Fragment Classification", IEEE Trans. on Image Processing, Vol. 21, No. 9, September 2012.

[11] Neha Gupta, V.K Banga, "Localization of Text in Complex Images Using Haar Wavelet Transform", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-1, Issue-6, November 2012.

[12] Xu-Cheng Yin, Wei-Yi Pei , Jun Zhang and Hong-Wei Hao, "Multi-Orientation Scene Text Detection With Adaptive Clustering", IEEE Trans. On Pattern Analysis and Machine Intelligence, 2013.

[13] R.Chandrasekaran and RM. Chandrasekaran, "Morphology based Text Extraction in Images", IJCST Vol. 2, Issue 4, Oct. - Dec.

[14] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs", IEEE Trans. on Intelligent Transportation Systems, 2014.

[15] Wahyono and Kanghyun Jo, "LED Dot matrix text recognition method in natural scene" Neurocomputing 151, pp 1033–1041, 2015.

[16] Cunzhao Shi, Chunheng Wang, Baihua Xiao, Yang Zhang and Song Gao, "Scene text detection using graph model built upon maximally stable extremal regions" Pattern Recognition Letters 34 107–116, 2013.

[17] P.Subbuthai, Azha Periasamy and S.Muruganand, "Identifying the Character by Applying PCA Method using Matlab", International Journal of Computer Applications (0975 – 8887) Volume 60– No.1, December 2012.

[18] J.Ashok and Dr.E.G.Rajan, "Principal Component Analysis Based Image Recognition" , International Journal of Computer Science and Information Technologies, Vol. 1 (2) , 2010, 44-50.

[19] Pramod Kumar Pandey, Yaduvir Singh and Sweta Tripathi, "Image Processing using Principle Component Analysis", International Journal of Computer Applications , Volume 15– No.4, February 2011.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET