ISSN: 2321-7782 (Online)
Volume 3, Issue 3, March 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Text Detection and Extraction from Natural Scene: A Survey

Tajinder Kaur[1]
Post-Graduation,
Department CE, Punjabi University,
Patiala, Punjab
India

Nirvair Neeru[2]
Assistant Professor,
Department CE, Punjabi University,
Patiala, Punjab
India

Abstract: ***Text detection and extraction is a popular and challenging research area in the image processing area because of complex background, variation of text size, font, orientation, scaling, alignment, and variety of color appearance, resolution of text or image and illumination changes. The text contains vital and useful information which is embedded in various types of documents and natural scene. The extraction of text from a natural image is a challenging task. The text is detected and is extracted in this way that it readable by another person without any difficulty. This extracted text contains valuable information that is usable in various applications like driving application so the driver can read the street sign easy, for a blind person it is a very effective technique, computer vision techniques, Automatic sign recognition technique, mobile's digital camera that shows good performance etc. Now days LED display that is natural scene is being widely used for announcements, sign boards, banners for displaying information. To detect and extract the text from the LED display is not an easy task, it is very complex due to its discontinuity. In this paper, we are discussing various proposed techniques of text detection and extraction from natural scene. The purpose of this review paper to classify various methods on the basis of performance parameters which are suggested by researchers.***

Keywords: Connected Component method, CRF method, Clustering technique, Gabor-based method, MSER, Text detection and extraction.

I. INTRODUCTION

Text detection and extraction method play an important role in many applications. It is a challenging task due to rapidly increase the digitization of all the material. 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 detection and extraction process involves text detection, localization, extraction, segmentation and recognition of text. Text extraction in natural scene image use in many applications such that mobile text recognition, automatic recognition sign board which supports for blind persons, license plate detection, text extracted from video, extract signboard text that uses for driving application, automatic text or form reading, mobile device text recognize, navigational support, document analysis, LED display text extracted that use of transport system etc.. Text extraction in natural scene and document in image use for so many applications, but still it is challenging task due complexity of its background, the appearance of different text due to text variation in size, color, orientation, alignment, font, shape, texture, geometry of text, image low and high resolution, image illumination changes, layout, image distortion, blurring problem and lighting condition [1].Figure 1 shows some examples of natural text or LED text to be extracted.

© 2015, IJARCSMS All Rights Reserved

Figure 1. ICDAR dataset sample image after text extraction and LED text recognition

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.

A. Scene text is a natural type of text which is accidently happening when we capture the image. The example of scene text is like a vehicle number plate, street signboards, banners, traffic sign board (simple or LED display) 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. Figure 3 shows a scene text.

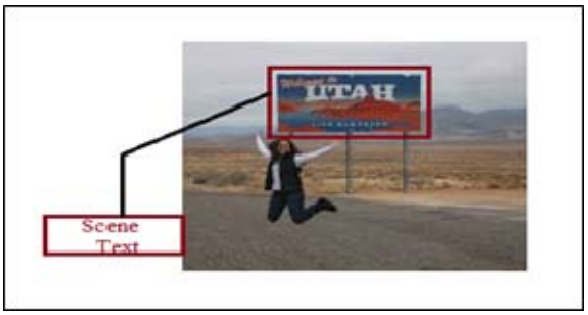

Figure 3. Scene Text

B. Artificial text 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 [2].Figure 4 show an example of artificial text.

Figure 4. Artificial Text

Various existing methods of text detection and extraction for natural scene can be roughly classified into two categories: region based technique, texture based techniques and hybrid technique [3].

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. 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 [1] [4]. On the basis of these features separates text or non-text region. The speed of region based method is

© 2015, IJARCSMS All Rights Reserved

slow as compared to other techniques. Edge based and Connected Component is a further classification of the region based approach.

The texture based method [5] uses different texture properties to extract a text from a complex image. Various methods are used in this approach to extract textual information like Wavelets, Fourier Transform and Gabor filters, DCT Transform Wavelet etc. A train classifier is used to extract the features of the target image region [1] [5]. The main aim of train classifier is to distinguish the text or non-text region for a scene.

The hybrid technique [3] 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. 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.

The rest of the paper is structured as follows. The literature survey is described in Section II. The performance parameters and comparison table are defined in Section III. In the Final step, we conclude the conclusion and future work are described in Section IV.

II. Related Work

Lots of research work has been done on text detection and extraction method in natural scene in the previous years. Various researchers have been proposed or studied many techniques of text extraction and give accurate results on the basis of performance metrics. In this section; we discuss a work done by various researchers on this active topic.

Anhar Risnumawan *et al.* [1] develop a robust method for text detection in natural scene use a properties or features, i.e. Mutual Magnitude Symmetry (MMS), Mutual Detection Symmetry (MMD) and Gradient Vector Symmetry (GVS) to detect the text candidate from natural scenes. Local descriptor SIFT exploring the pixel text , identify the text candidate or remove the non-text pixels(candidate).Then apply the ellipse growing method which is based upon the text orientation , extract the text , restore and eliminate the non-text character in it. In this, the proposed method work on three datasets and not depend upon contrast, orientation, fonts, resolution and text size.

Xu-Cheng Yin *et al.* [2] propose an accurate and robust text detection technique based upon the Maximally Stable Extremal Region (MSERs) the text character is extracted weather the condition of the image is bad. A self-trained distance metric method is used that learns weights and single link methods use the learned parameters. The classifier is used to identify the posterior probability of text character and remove the non-text candidate character. The proposed technique is very effective used the ICDAR 2013 dataset in “Text Localization in Born-Digital image” and “Text Localization in Real Scène”. In this paper, the ICDAR 2011 database is used and gives a 76% f-measure which is good.

Xu-Cheng Yin *et al.* [3] proposed a multi-orientation scene text detection method with adaptive hierarchical clustering. The main aim this method to accurately detect horizontal or non-horizontal natural scene text. They use a three step in their work. The unified distance metric learning framework with adaptive clustering because this method automatically detect and use parameters and constructs a text candidate .Then use a coarse-to-fine grouping of various techniques (morphology, orientation and projection clustering) and scene text is detected. The proposed method gives 71% of-measure on various public datasets which available on the internet.

Jack Greenhalgh and Majid Mirmehdi [6] present a novel approach automatic detect and recognize a traffic signs based upon the combination of MSERs and HSV thresholding techniques. In this, candidate characters detects first and afterward recognize the text detected region. It removes the false positive (FT) region using the temporal fusion information. It achieves an accurate and improved result off-measure of 87%.

© 2015, IJARCSMS All Rights Reserved

Wahyono and Kanghyun Jo [7] present a new method to recognition of LED Dot Matrix text from natural scene images because no method is capable to detect and extract the LED text accurately due to discontinuity of character. In this paper, they use a canny edge detector to detect a character pixel and calculate a center point of edges. Afterwards merge these point and obtain a character candidate k-nearest neighbors technique is used to classify the characters and combine the character into text line based upon their properties. This proposed method achieves a detection rate 68.8% and recognition rate 47%.

Cong Yao *et al*. [9] proposed a unified method of text detection and recognition of the multi-oriented natural scene using same features and classification method. They proposed a new dictionary search based method is used to detect and recognition errors and correct them. This technique work on different font, scales, orientation and color text in natural scenes .The proposed method mainly focus on multi-oriented text and achieves a performance parameter of f- measure 73 % work on four databases.

Jianqiang Yan and Xinbo Gao [10] discuss a new technique for text detection and recognition of text superimposed in complex image using a color clustering technique (SOM and FCM) with connected component based (CC) method. A trained cascade Adaboost classifier is used to distinguish whether the candidate text region is a real text or not in image layer. In the last step, OCR package is used to recognize the real text which is localized by the trained classifier. The accuracy or recognition rate of the text region improves on this layer method.

Chucai Yi and Yingli Tian [11] present a framework to extract and localization a text from the complex natural scene by using three steps, i.e. boundary clustering, stoke segmentation and string fragmentation classification method. The text is automatically extracted and localization from natural scene by using three phases: pixel, character and string on the basis of the features. They proposed a two method to combine a stoke text and filter the non-text region. After the Gabor-based method is used to string fragment classification on the basis of text features. The proposed method work upon a natural scene text, born-digital images, pictures captured by a blind person, broadcast videos, ICDAR 2003 and ICDAR 2011 dataset.

Yi-Feng Pan *et al.* [12] designed hybrid and robust technique to detection and localization of text in natural scene using scale-adaptive binarization to extract a candidate character. Then apply conditional random field (CRF) model is used to filter the non-text region. At last step, the energy minimization method is used to club the text line or text character region. The images come from ICDAR 2005 database and achieve a good result.

R.Chandrasekaran and RM. Chandrasekaran [13] proposed a method to extract and recognize a text for scene or caption type images using a morphological operator. Morphological dilation is used to text localization and then connected component method is applied to separate text or non-text features. These features are recognized by a support vector machine. For testing the images they use the ICDAR 2003 dataset and compare the result with existing techniques that is developed by past researchers. The proposed method achieves 93% of f-measure.

III. Performance Parameters

The accuracy of the text segmentation method is compared in terms of Precision measure (p), Recall measure(r) and F-measure (f). These performance parameters are described as follows:

Definition: 1

Precision measure (p) is defined as the ratio of the correctly detected text pixels to the sum of correctly detected text pixels and false positive means the number of pictorial pixels which are detected only texts.

$$p = \frac{TP}{TP+FP}$$

Definition: 2

Recall measure (r) is defined as the ratio of correctly detected text pixels to the sum of correctly detected pixels and false negative, which means the number of textual pixels which are detected pictorial ones.

© 2015, IJARCSMS All Rights Reserved

$$r = \frac{TP}{TP+FN}$$

Definition: 3

F-measure (f) is defined as the harmonic mean of the precision (p) and recall (r) measure.

$$f = \frac{Precision}{recall}$$

TABLE 1

COMPARISON OF VARIOUS TECHNIQUES DISCUSSED IN LITERATURE SURVEY

Sr. No.	Author	Year	Description / Techniques Used	Dataset/Images Type	F-measure	Advantages/ Disadvantages
1	Anhar Risnumawan *et al.*[1]	2014	For edge detection use canny and Sobel operators and work on symmetrical features. The extract candidate character uses local descriptors	ICDAR 2005	69%	Don't depend upon an orientation, script, font, font size, contrast and resolution. Accuracy is not good.
				ICDAR 2011	71%	
				MSRA-TD500	69%	
				Curved text	61%	
2.	Xu-Cheng Yin *et al.* [2]	2014	Connected component with MSER method, single-link clustering method and use character classifier.	ICDAR 2011	76%	It detects the text when image is in low quality. But it can't work when the text is highly blurred and low resolution image.
				Multilingual text	74.6%	
				Street view text	66%	
				Multi-oriented text	66%	
3.	Xu-Cheng Yin *et al.* [3]	2013	Adaptive hierarchical clustering method makes a single cluster. Morphological, orientation and projection clustering methods	USTB-SV1K	47.53%	It detects a multi-orientation text because earlier researcher works on horizontal and vertical scene text. Not detect text which is highly blurred in low resolution images
				MSRA-TD500	71%	
				ICDAR 2011	73.84%	
				ICDAR 2013	73.35%	
4.	Jack Greenhalgh and Majid Mirmehdi [6]	2014	Use a combination of MSER and HSV (Hue, Saturation and Value) color thresholding method for character component.	Traffic sign Board	87%	The search area reduces by using structural information. This method provides improve result than previous method.
5.	Wahyono *et al.* [7]	2014	Canny edge detector and k-nearest neighbor method.	LED Text	68%	It can extract the discontinuous LED text, but not continuous or titled text.
6.	Cong Yao *et al.* [8]	2014	Obtain a characters using SWT and clustering techniques. Dictionary based correction method to correct the errors. Random Forest classifier is work for recognition purpose.	ICDAR 2011	73%	This method improves the accuracy of character recognition method. It extracts the false positive character because they are very similar to true text.
				Chars74K (Matched)	75.9%	
				MSRA-TD500	61%	
7.	Yi-Feng Pan *et al.* [13]	2013	Hybrid approach i.e. combination of region	ICDAR 2005	92.5%	It combines the advantages of both region and CC based

© 2015, IJARCSMS All Rights Reserved

			based and connected component method, CRF work as a classifier.	Multilingual text	65.2%	method. This method fails to detect some complex background images and need to some improvements.
				MSRA-TD500	71%	
				ICDAR 2011	73.84%	
				ICDAR 2013	73.35%	

IV. Conclusion

In this paper we provide a review of different techniques of text detection and extraction from natural scene. There are so many techniques which are used in text extraction but no one single method is use for all applications due to variation in font, orientation, alignment, color, style, size, texture, font size and provide satisfactory results. Text detection and extraction in natural scene is very challenging task due to complexity of background. Many researchers can work on public database which is available online and give a better performance. On the basis of this review paper provide a light of different researchers which doing research on this area.

Acknowledgement

I am thankful to Nirvair Neeru mam Assistant Professor of Department of Computer Engineering, Punjabi University Patiala for her guidance and valuable support on this research topic and also in this paper.

References

1. 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.
2. 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.
3. 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.
4. Jing Zhang and Rangachar Kasturi, “A Novel Text Detection System Based on Character and Link Energies”, IEEE Trans. On Image Processing, Vol. 23, No. 9, September 2014.
5. Yao Li, Wenjing Jia, Chunhua Shen and Anton van den Hengel, “Characterness: An Indicator of Text in the Wild”,IEEE Trans. On Image Processing, Vol.23, No. 4, April 2014.
6. Jack Greenhalgh and Majid Mirmehdi, “Recognizing Text-Based Traffic Signs”, IEEE Trans. on Intelligent Transportation Systems, 2014.
7. Wahyono and Kanghyun Jo, “LED Dot matrix text recognition method in natural scene” Neurocomputing 151, pp 1033–1041, 2015.
8. 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.
9. 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.
10. Jianqiang Yan and Xinbo Gao, “Detection and Recognition of Text Superimposed in Images Based on Layered Method”, Neurocomputing, 134, pp 3-14, 2014.
11. 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.
12. 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.
13. R.Chandrasekaran and RM. Chandrasekaran, “Morphology based Text Extraction in Images”, IJCST Vol. 2, Issue 4, Oct. - Dec.

© 2015, IJARCSMS All Rights Reserved