

A Survey on Text Detection in Natural Images

Amritha S Nadarajan¹, Thamizharasi A²

¹Computer Science Engineering

²Asst.Prof. Computer Science Engineering,

Mohandas College of Engineering and Technology, Anad, India

Abstract— In natural images, some documents are embedded. Text detection is identifies text from natural images. Basic digital image processing techniques are used to detect text from the images. This includes preprocessing, extraction or text localization, classification and character detection. The different classification methods used are SVM, AdaBoost, CNN; Text-CNN etc. This paper provides a detailed study of evolution of text detection in natural images. It compares, analyzes and also discusses the different methods to overcome existing challenges in text detection. This paper presents the different types of datasets which are used to identify text from natural images and comparative study of different text detection methods. The comparative study proves that CNN is a better technique to detect text in natural images.

Index Terms- Text detection, survey, natural image, SVM, CNN, AdaBoost.

Introduction

The text has an important role in our life. Text become help to some vision based application. Text detection is to identify text in a given input images. In previous years, researchers have proposed different techniques to detect text from natural images. Text means either scene text or caption text.

The text detection in the image is carried out by constructing the bounded rectangle box around the text in the natural image. There are many techniques exists for the text detection [1]. Retrieving text from both in indoor and outdoor is difficult due to variety of backgrounds. Text detection can be easily done by the help of image retrieval algorithm. Fig.1 shows the basic text detection method in image processing technique.

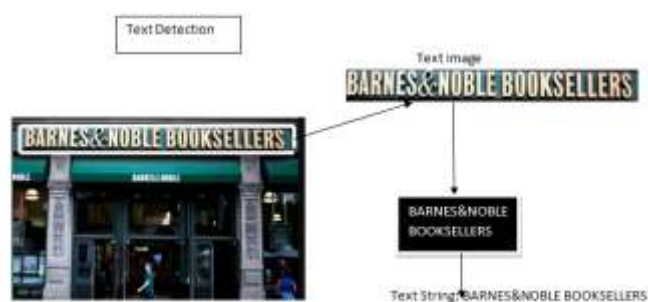


Fig.1 Sample screenshot of Automated Text detection System

In early days, text detection is carried out from images captured through camera. Optical Character Recognition (OCR) was the technology to detect text from scanned images and printed images. But it faces difficulties like Complexity backgrounds, diversity of scene text and interference factors etc. So to overcome the above difficulties researchers have developed techniques like Adaboost, SVM, Filters, and Convolutional Neural Networks (CNN) etc. [19]. Now high performing capability text detection techniques has been available. The recent text detection techniques are based on machine learning techniques. It helps to automatically detect the text from complex back grounds. Text detection can have variety of vision based application such as robot navigation [14], geolocation [15], industrial automation [16] etc.

This paper gives a detailed survey of existing text detection techniques. In this paper the difficulties, methods, applications and benchmark datasets are discussed.

The Challenges of Text Detection

However, reading texts from images is a difficult task. The main text detection challenges are categorized into three groups [17]:

- i) Diversity of natural images: The characters in natural images are in different font style, different size, unique colour, unique font alignment etc.
- ii) Complexity in backgrounds: The backgrounds in natural images can be more complex. The background includes grasses bricks, rocks, sign boards are there, and it leads to more complexity for identifying text.
- iii) Inference factors: The main inference factors are blurring, noise and low resolution of input images [17].

The remainder of the paper is arranged as follows: The section 2 describes the methodology of text detection. Discusses the disadvantages of recent technologies, in section discussing about what are the new methodologies and their problems and also discussing the present issues.

METHODOLOGY

The text detection task is to draw a bounding box or rectangle box for each line of text in the natural images. Many methods have been proposed for text detection. Some of the commonly used text detection methods are given below [7]. Mainly there are three different types of text detection techniques. They are as follows:

Texture based method describes, text as a type of texture and it has texture properties. This method handles horizontal texts and scale change. Tables 1 describe different texture based methods used to detect text from natural images. Mainly different types of classifiers are used to increase the speed and accuracy of detection process.

TABLE 1. TEXTURE BASED METHOD FOR TEXT DETECTION.

Method	classifier	dataset
Detection in color images	Algorithms	Color images
Text detection in images and videos	SVM classifier	Complex natural scene images.
Weak classifiers are trained	Cascade Adaboost classifier	Common scene images

In component based method, first extract the candidate components from the image and then filter out the by manually designed rules or classifiers [8], [11], [12]. However, these methods are more efficient for text detection, because the processing speed becomes high. This method is adaptable for changes in scale and font variations.

The component based method include several colour clustering, grouped components into text line, removed non-text components for text detection[13],[14]. The table 2 shows different component based text detection methods.

TABLE 2. COMPONENT BASED METHOD FOR TEXT DETECTION.

Method	Classifier	Dataset
Component by colour clustering [13]	Clustering	Simple scene images
Based on character Stroke from edge [14].	SWT (Stroke Width Transform)	Only horizontal text
Extract MSER regions from original images.[14]	MSER (Maximally Stable External regions)	Horizontal text, larger inclination angle.
Based on SWT.[13]	SFT (Stroke Feature Transform)	Standard dataset
MSER+CNN[14]	CNN (Convolutional Neural Network)	Scene text

In hybrid method, Combination of component method and texture method [3].This method is more efficient method for detecting text. In this method, researchers used gradient and geometrical properties for verifying texture portion from a natural image. The Table 3 Shows the detailed study of hybrid method.

TABLE 3. HYBRID METHOD FOR TEXT DETECTION

Method	Classifier	dataset
CRF(Conditional Random Field)[3]	Probability maps	Horizontal text

DATASETS USED FOR TEXT DETECTION

Some Specific types of public dataset are considered as the standard benchmark dataset for text detection technique. The widely used datasets are:

ICDAR 2003 and 2005: This is a basic dataset. It provides sample input images. This data set is provided in the form of zip file. The ICDAR 2003 Robust reading Competition held in the year 20013. It contains 509 text images.258 images are used for training and rest of the others are for testing. The ICDAR 2005 was also used for text locating competition.

ICDAR 2011 and 2013: It is a subset of QUWI. There are totally 475 images, 4 hand writing images. This is used for recent text detection technique. This dataset are inherited the previous dataset QUWI [18]. Fig.2 shows the sample ICDAR 2011 dataset images.



Fig.2. Sample ICDAR 2011 dataset images

OSTD: The Oriented Scene Text Database dataset contains 89 images. This can be used for multi oriented text in natural scene images. Fig.3 shows the sample OSTD dataset images.



Fig.3. Sample OSTD dataset images

MSRA-TD500: It is a benchmark dataset. It used for evaluating text detection algorithm for multi oriented text in natural images. It consists of almost 500 databases. It provides horizontal, slant and skewed text in natural scenes. Fig.4 shows the sample MSRA-TD500 dataset.



Fig.4 Sample MSRA-TD500 dataset

SVHN: The Street View House Numbers Dataset is a large dataset. It has 600,000 natural images. This is a real world images of dataset .These are numbers collecting from house numbers from go ogle street view images. The Table 4 describes the comparative study of different datasets.



Fig.5 Sample SVHN dataset images.

TABLE 4 . COMPARITIVE STUDY OF BENCH-MARK DATASETS USED FOR TEXT DETECTION

<i>Dataset</i>	<i>Features</i>	<i>Advantages</i>	<i>Disadvantages</i>
DAR 2003,2005	For scene detection and recognition	1.Robust read and text detection	Normal images
ICDAR 2011,2013	Word recognition	Much faster	Scene text images
OSTD	Detect text from logos,indoor scenes etc.	Multi oriented texts in natural images	Limited speed
MSRA -TD500	500 images in horizontal	Detect text in arbitrary orientation	Diversity of text and complexity in baccground
SVHN	Real world images of dataset	Detect arbitrary multi digit numbers	Dataset are obtained from house numbers from street view images.

COMPARITIVE STUDY

Earlier numerous method have been developed for text detection.Text detection was a natural extension of document analysis moving from scanned pages images into camera captured image, preprocessing, detection and OCR technology. Scene text has been regarded as presenting a more difficult challenge yet very little work had been done with it. Recently, researchers have explored approaches that prove effective for text captured in various configurations. The different approaches are Unsupervised feature learning, Convolutional neural network etc.

The Table 5 describes a detailed study about the text detection in past and current development techniques. Based on the literature survey the latest technique convolutional neural network provides 93% of accuracy result. This is a better method to apply for producing good result in future.Table.6 describes about the advantages and disadvantages of techniques used in following papers.

TABLE 5: COMPARITIVE STUDY OF DIFFERENT TEXT DETECTION TECHNIQUES.

<i>Title</i>	<i>Author's</i>	<i>Year of Publication</i>	<i>Method</i>	<i>Classifier</i>	<i>Accuracy</i>	<i>Dataset</i>
Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm [5].	Kwang In Kim, Keechul Jung, and Jin Hyung Kim	2003	A novel texture based method for detecting text in images. A SVM is used to analyze the external properties of text. No external texture feature extraction module is used.	SVM	71.5%	400 images from web images video from scanned documents
Detecting and reading text in natural scene [6].	Xiangrong Chen,Alan L. Yuille	2004	An AdaBoost machine learning algorithm to train strong classifier. OCR software is used to read the text or reject	AdaBoost	90%	SKERI

			it as a non text region.			
Detecting Text in Natural Scenes with Stroke Width Transform [7].	Boris Epshtein Eyal Ofek Yonatan Wexler	2010	By using text detection algorithm, defied the notion of a stroke and then explain SWT	filters	90.39	ICDAR2005
Text String Detection From Natural Scenes by Structure-Based Partition and Grouping [8].	Chucui Yi , YingLi Tian	2011	Text string detection method is used. It consist two step: -image partition to find text character. -Character candidate grouping to detect text strings. Two algorithms are used: -adjacent character grouping. -text line grouping method.	SVM learning classifier	71%	ICDAR 2003
Robust text detection in natural scene image [17].	Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang	2014	Proposed a novel MSER based scene text detection method. Linear reduction and tree accumulation algorithm are used in different situation	Character classifier	71%	ICDAR 2011
Symmetry-based text line detection in natural scenes[10]	Zheng Zhang, Wei Shen, Cong Yao, Xiang Bai	2015	Sliding window and connected component extraction are method for scene text detection	CNN classifier	80%	ICDAR 2011,ICDAR 2013
Image Net Classification with Deep Convolutional Neural Networks [11].	Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton	2017	A deep CNN to classify the 1.2 million high resolution images in image Net LSVRC 2010.CNN consist 650000 neurons and 5 convolutional layers.	Several classifiers trained on Fishers Vectors.	93%	ILSVRC 2010
The Text	Tong He,	2015	Recent deep	Text-	93%	MSRA-TD500 dataset

Attentional Convolutional Neural Network for Scene Text Detection [20]	Weilin Huang, Yu Qiao, and Jian Yao	learning models are used in this paper for text detection. It has capable of high level feature from whole image.	CNN classifier		
---	-------------------------------------	---	----------------	--	--

TABLE 6: ADVANTAGES AND DISADVANTAGES OF DIFFERENT TEXT DETECTION METHODS.

Title	Advantages	Disadvantages
Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm [5].	-Fast text detection. -It extracting texts from complex and texture back ground. -Robust method	Problems in classifying very small text or text with low contrast
Detecting and reading text in natural scene [6].	Combination of weak classifiers to Strong Classifiers	Choice of feature set is critical to success and transparency of the algorithm
Detecting Text in Natural Scenes with Stroke Width Transform [7].	Fast, robust	Need time
Text String Detection From Natural Scenes by Structure-Based Partition and Grouping [8].	Highly perform the state of the art result. Detect text strings the oriented scene Text dataset	Less performance in variation of scale, orientation clutter background
Robust text detection in scene image [17].	High speed	Fail to detect due to very complex background, on uniform illumination, highly blurred text
Symmetry-based text line detection in natural scenes [10]	Deal different sizes, discovering symmetric structures in natural images	Processing speed low
Image Net Classification with Deep Convolutional Neural Networks [11].	Training faster Very efficient GPU Reduce over fitting	Expensive for big NN
The Text Attentional Convolutional Neural Network for Scene Text Detection [20].	Efficient and accurate result	Problem in multiple text variations and significantly cluttered background.

CONCLUSION

This survey is based on text detection using image processing technique. Different types of difficulties have been faced during the time of text detection in from natural images. To overcome the difficulties, different techniques has been introduced: OCR, SVM, Adaboost, CNN etc. These techniques are different types of classifiers to detect text from natural images.

REFERENCES

- [1] Y. Li, W. Jia, C. Shen, and A. van den Hengel, "Characterness: An indicator of text in the wild," IEEE Trans. Image Processing (TIP), vol. 23, pp. 1666–1677, 2014.
- [2] L. Neumann and K. Matas, "Real-time scene text localization and recognition," 2012, in IEEE Computer Vision and Pattern Recognition (CVPR).
- [3] Y.-F. Pan, X. Hou, and C.-L. Liu, "Hybrid approach to detect and localize texts in natural scene images," IEEE Trans. Image Processing (TIP), vol. 20, pp. 800–813, 2011. Comput.Sci.DOI,2011.
- [4] P. He, W. Huang, Y. Qiao, C. C. Loy, and X. Tang, "Reading scene text in deep convolutional sequences," 2016, in 16).
- [5] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans Pattern Analysis and Machine Intelligence (TPAMI), vol. 25, pp. 1631–1639, 2003.
- [6] X. Chen and A. Yuille, "Detecting and reading text in natural scenes," 2004, in IEEE Computer Vision and Pattern Recognition (CVPR).
- [7] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scene with stroke width transform," 2010, in IEEE Computer Vision and Pattern Recognition (CVPR).
- [8] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Processing (TIP), vol. 20, pp. 2594–2605, 2011.

- [9] L. Sun, Q. Huo, W. Jia, and K. Chen, "A robust approach for text detection from natural scene images," *Pattern Recognition*, vol. 48, pp.2906–2920, 2015.
- [10] Z. Zhang, W. Shen, C. Yao, and X. Bai, "Symmetry-based text line detection in natural scenes," 2015, in *IEEE Computer Vision and Pattern Recognition (CVPR)*.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," 2012, in *Neural Information Processing Systems (NIPS)*.
- [12]. "Text Attentional Convolutional Neural Network for Scene Text Detection" Tong He, Weilin Huang, Yu Qiao, and Jian Yao.
- [13] K. Kim, K. Jung, and J. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 25, pp. 1631–1639, 2003.
- [14] S. Hanif and L. Prevost, "Text detection and localization in complex scene images using constrained Adaboost algorithm," 2009, in *International Conference on Document Analysis and Recognition (ICDAR)*.
- [15] D. B. Barber, J. D. Redding, T. W. McLain, R. W. Beard, and C. N. Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. *Journal of Intelligent and Robotic Systems*, 47(4):361–382, 2006.
- [16] Y. K. Ham, M. S. Kang, H. K. Chung, R. H. Park, and G. T. Park. Recognition of raised characters for automatic classification of rubber tires. *Optical Engineering*, 34(1):102–109, 2005.
- [17] X. C. Yin, X. Yin, K. Huang, and H. W. Hao, "Robust text detection in natural scene images," *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, pp. 970–983, 2014.
- [18]. <https://www.google.co.in/search?q=images+of+dataset+ICDAR+2011>
- [19] W. Huang, Y. Qiao, and X. Tang, "Robust scene text detection with convolutional neural networks induced msr trees," 2014, in *European Conference on Computer Vision (ECCV)*.
- [20]. Tong He, Weilin Huang, Yu Qiao and Jian Yao, "Text Attentional Convolutional Neural Network for Scene Text Detection", *IEEE TRANS. ON IMAGE PROCESSING.*, DOI 10.1109/TIP.2016.2547588 .

