International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-6, Issue-6)

Research Article

June 2017

Techniques of Text Detection and Recognition: A Survey

Shivani, Dipti Bansal
Department of Elec. & Comm. Engineering, Punjabi University, Patiala,
Punjab, India

Abstract-

The pattern recognition is the technique which is applied on the image to detect similar type of patterns from the image. The text detection and recognition are the techniques of patterns detection. To detect text area in the image techniques of image segmentation is required which will segment the area in which text is present. To mark the text from the image technique of neural networks is required which will learn from the previous values and drive new values on the basis of current network situations. In this paper, various techniques of image segmentation and neural networks has been reviewed and discussed in terms of their outcomes.

Keywords-Segmentation, Neural Networks, Edge Detection

I. INTRODUCTION

In the Text recognition scenario from the detected lines, text remains a challenging trouble due to the variety of shades, fonts, colors as well asthe presence of complex backgrounds and the short length of the text strings. Accordingly, text detection and recognition in natural images have received increasing attention in computer vision and image understanding, due to its numerous applications in image retrieval, scene understanding, visual assistance etc. Text detection and recognition in images and video frames, which aims at integrating advanced optical character recognition (OCR) and text-based searching technologies, is now recognized as a key component in the development of advanced image and video annotation and retrieval systems. Unfortunately, text characters contained in images and videos can be any gray-scale value (not always white), low-resolution, variable size and embedded in complex backgrounds.

A. Steps in image segmentation:

The general steps that are involved in Image segmentation systems are:

1. Image acquisition
2. Pre-processing
3. Segmentation

1) Image Acquisition: This is the stage where the image under consideration is taken. In the case of online recognition system, a specialized hardware is implemented as explained earlier whereas for offline systems, the images are obtained either through a scanner or a camera. On any occasion an image is acquired, there will be some variations in the intensity levels along the image. Also noise gets added to the image. Hence preprocessing is required for adjusting the intensity levels and to denoise the image.

2) Preprocessing: The Pre-processing is most important part of a better performing recognition system. In this scenario, the acquired image is processed to remove any noise that may have incurred into the image during the time of acquisition or during the time of transmission. A colored image then it will be transformed to a gray image before proceeding with the noise removal procedure. The de-noised image is then converted to a binary image with suitable threshold.

3) Segmentation: This segmentation refers to a process of segregation an image into groups of pixels which are homogeneous with consideration to some benchmark. This distribution algorithms are area oriented instead of pixel oriented. At the output of this process is the splitting up of the image into connected areas. Thus, segmentation is concerned with dividing an image into meaningful regions. Image segmentation can be broadly classified into two types.

1. Local Segmentation: It deals with the segmenting sub images which are small windows on a whole image.

Global segmentation: It deals with the images subsist of comparably large number of pixels and makes estimated parameter values for global segments more robust.

For character segmentation, first the image must be segmented row-wise (line segmentation), then each rowmust be segmented column-wise (word segmentation). Certainly,aspect can be extracted using advisable algorithms such as edge detection technique; histogram based methods or connected component analysis

B. Text extraction Techniques

1) Compression Based Algorithm: This algorithm pre-suppose that the optimal segmentation is the one that minimizes the overall conceivable segmentation, coding length of the data. The connection between these two concepts is that segmentation tries to find patterns in an image and any consistency in the image can be used to compares it. The algorithm explain each segment by its texture and boundary shape. This algorithm was implemented by W.J Teahan, Yingying Wen, Rodger Mcnab and Lan H.

2) Corner Response Based Method: A novel text detection and localization method based on corner response consist of 3 stages: (1) Computing corner response in multi-scale space and thresholding it to get the candidate region of text; (2) Verifying the candidate region by combining color and size range features;(3) Locating the text line using bounding box. Corner is a special two-dimensional feature point which has high curvature in the region boundary. It can be located by finding the local maximum in corner response (CR). In corner points in video frame are used to generate connected component. But they use just the number of corner points, not CR, to classify text and non-text region.

3) Edge Detection Algorithm: This algorithm is well developed field on its own within image processing. The region boundaries and edge are closely related, since there is often a shape adjustment in intensity at the region boundaries. This detection techniques have therefore been used as the base of another segmentation technique. The edge identified by edge detection is often disconnected. To portion an object from an image however, one needs closed region boundaries. Salem Saleh Al-amril, Dr N.V kalyankar implemented image segmentation by using Edge detection .They did a comparative study using seven technique of the edge detection segment. They arerobert, canny, laplacian, and edge maximum technique on the Saturn original image and found that EMT and Perwitt techniques respectively are the best techniques for edge detection.

4) Nearest Neighbor Clustering Based Method (NNC): In this process, a novel approach for line and character segmentation in an epigraphically script based on closest neighbor clustering process is presented. The expected algorithm scans the given input image from the left corner. When it confrontation the first black pixel, it describe the complete character through connected component. This character is segmented and placed at different location. The centered of the character is computed. Similarly the second character is identified and the centered is computed. The Euclidean distance between the centroids is computed to know whether the character belongs to the same line or next line. This is determined based on the threshold which is based on the assumption that the space between the text lines is greater than that between the characters. In this way, the text lines and characters are segmented which could be used for the classification process.

II. LITERATURE SURVEY

In 2011 K.Wang, B. Babenko and S.Belongie"End-to-End scene text detection" proposed a novel scene text recognition method using part based tree structured character detection, different from conventional multi-scale sliding window character detection strategy, which does not make use of the character specific structure information. They have used part based tree structure to model each type of character to detect and recognize the characters at the same time. However, since text in natural images differs from text in traditional scanned document in terms of resolution, illumination condition, size and font style, the binarization result is usually unsatisfactory. Moreover, the loss of information during the binarization process is almost unrecoverable, which means if the binarization result is poor, the chance of correctly recognizing the text is quite small [1].

In 2012 Cong Yao, Xiang Bai, Wenyu Liu , Yi Ma, Zhuowen Tu "Detecting Texts of Arbitrary Orientations in Natural Images'' proposed a system which detects texts of arbitrary orientations in natural images. The proposed algorithm is equipped with a two-level classification scheme and two sets of features specially designed for capturing both the intrinsic characteristics of texts. To better assess this algorithm and analyze it with other clash algorithms, a new dataset has been generated, which includes various texts in diverse real-world scenarios. Experiments on benchmark datasets and the proposed dataset demonstrate that proposed algorithm compares favorably with the state-of-the-art algorithms when handling horizontal texts and achieves significant performance on texts of arbitrary orientations in complex natural scenes [2].

In 2013 Tao Wang, David J. Wu, Adam Coates, Andrew Y. Ng," End-to-End Text Recognition with Convolutional Neural Networks".In this paper, an alternate route is brought and consolidated with the representational power of large, multilayer neural networks together with recent developments in unsupervised feature learning, which allows us to utilize a typical framework to train highly-accurate text detector and character recognizer modules. Then, utilizing just simple off-the-shelf methods, these two modules are integrated into a full end-to-end, lexicon-driven, scene text recognition system that accomplishes state-of-the-art performance on standard benchmarks, to be specific Street View Text and ICDAR 2003 [4].

In 2013 Alessandro Bissacco, Mark Cummins, Yuval Netzer, Hartmut Neven,," Photo OCR: Reading Text in Uncontrolled Conditions", proposed Photo OCR, a system for text extraction from images. This particular concentration is dependable text extraction from smartphone imagery, with the goal of text recognition as a client input modality like discourse recognition. Financially accessible OCR performs ineffectively on this assignment. New development in machine learning has considerably improved isolated character classification; we expand on this progress by demonstrating an entire OCR system utilizing these methods. They likewise incorporate present day datacenter-scale distributed language demonstrating. This approach is capable of perceiving text in a variety of challenging imaging conditions where traditional OCR systems fail, notably in the presence of substantial obscure, low resolution, low contrast, high image noise and other distortions. It likewise operates with low latency; mean processing time is 600 ms per image. This system is evaluated on public benchmark datasets for text extraction and outperforms all beforehand reported results, more than halving the error rate on multiple benchmarks [6].

In 2014," Robust Text Detection in Natural Scene Images", proposed an accurate and powerful method for detecting texts in natural scene images. A quick and active pruning algorithm is designed to extract Maximally Stable Extremal Regions (MSERs) as character candidates utilizing the strategy of limiting regularized variations. Aspect candidates are grouped into text candidates by the single-link clustering algorithm, where distance density and clustering

edge are found out naturally by a novel self-training distance metric learning algorithm. The back contingency of text candidates comparing to non-text are estimated with a aspect classifier; text candidates with high non-text probabilities are eliminated and texts are identified with a text classifier. The proposed system is evaluated on the ICDAR 2011 Robust Reading Competition database; the f-measure is more than 76%, much superior to the state-of-the-art performance of 71%. Experiments on multilingual, street view, multi-orientation and even born-digital databases additionally demonstrate the effectiveness of the proposed method [5].

In 2014 Weilin Huang, Yu Qiao, and Xiaoou Tang,"Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees", a novel framework to handle this problem by utilizing the high ability of convolutional neural system (CNN). In contrast to the recent methods utilizing an arrangement of low-level heuristic features, the CNN system is capable of learning high-level features to heartily recognize text components from text-like outliers (e.g. bicycles, windows, or clears out). This approach takes favorable circumstances of both MSERs and sliding-window based methods. The MSERs operator dramatically reduces the quantity of windows scanned and enhances location of the low-quality texts. While the sliding-window with CNN is connected to effectively separate the associations of multiple characters in components. The proposed system accomplished strong heartiness against various extreme text variations and genuine real-world problems. It was evaluated on the ICDAR 2011 benchmark dataset, and accomplished more than 78% in F-measure, which is altogether higher than past methods [8].

In 2015 Xu-Cheng Yin, Wei-Yi Pei, Jun Zhang and Hong-Wei Hao,"Multi-Orientation Scene Text Detection with Adaptive Clustering", proposed in this paper a unified distance metric learning framework for adaptive progressive clustering, which can at the same time learn similarity weights and the clustering edge.Then, an effective multi-orientation scene text location system is proposed which constructs text candidates by grouping characters in view of this adaptive clustering. This text candidates construction method comprises of a few sequential coarse to-fine grouping steps: morphology-based grouping through single-link clustering, orientation-based grouping by means of divisive various leveled clustering, and projection-based grouping additionally by means of divisive clustering. The effectiveness of the proposed system is evaluated on a few public scene text databases, e.g., ICDAR Robust Reading Competition datasets (2011 and 2013), MSRA-TD500 and NEOCR. In particular, on the multi-orientation text dataset MSRA-TD500, the f measure of our system is 71%, much superior to the state-of-the-art performance [9].

In 2015 Zheng Zhang, Wei Shen, Cong Yao, Xiang Bai," Symmetry-Based Text Line Detection in Natural Scenes",proposed in this paper that recently, a variety of real-world applications have triggered tremendous demand for procedures that can extract textualinformation from natural scenes.

Therefore, scene text location and recognition have turned out to be dynamic research themes in PC vision. In this work, they research the problem of scene text recognition from an alternative perspective and propose a novel algorithm for it. Not the same as traditional methods, which principally make utilization of the properties of single characters or strokes, the proposed algorithm exploits the symmetry property of character groups and allows for direct extraction of text lines from natural images. The experiments on the most recent ICDAR benchmarks demonstrate that the proposed algorithm accomplishes state-of-the-art performance. In addition, compared to routine approaches, the proposed algorithm indicates stronger adaptability to texts in challenging situations [10].

In 2016 Tong He, Weilin Huang, Yu Qiao, and Jian Yao," Text-Attentional Convolutional Neural Network for Scene Text Detection", proposed a novel Text-Attentional Convolutional Neural Network (Text-CNN) that particularly concentrates on extracting text-related regions and features from the image components. One more learning structure is developed to train the Text-CNN with multi-level and rich managed information, including text region veil, character label, and binary text/non-text information. Also, a powerful low-level detector called Contrast-Enhancement Maximally Stable Extremal Regions (CE-MSERs) is developed, which extends the broadly utilized MSERs by enhancing intensity contrast between text patterns and background. This allows it

to identify highly challenging text patterns, resulting in a higher recall. This approach accomplished promising results on the ICDAR 2013 dataset, with a F-measure of 0.82, improving the state-of-the-art results substantially [3].

In 2016Max Jaderberg, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman," Reading Text in theWild with Convolutional Neural Networks",proposed an end-to-end system for text spotting—localizing and perceiving text in natural scene images—and text based image retrieval is introduced. This system depends on a region proposal mechanism for identification and deep convolutional neural networks for recognition. Pipeline employ a novel consolidation of reciprocal proposal generation systems to guarantee high recall, and a fast resulting filtering stage for improving precision. Rigorous experiments are performed over various standard end-to-end text spotting benchmarks and text-based image retrieval datasets, demonstrating a large improvement over every past method. At last, a real-world use of this text spotting system is demonstrated to allow thousands of hours of news footage to be in a split second searchable by means of a text query [7].

Table I of Comparison

AUTHOR	YEAR	DESCRIPTION	RESULT
Wang, B. Babenko &S.Belongie	2011	This paper focuses on the problem of word detection and recognition in natural images.The results establish a baseline for using generic computer vision methods on end-to-end word recognition in the wild.	Achieved superior performance and results demonstrate the suitability of applying generic computer vision methods.

Cong Yao, Xiang Bai, Wenyu Liu , Yi Ma, Zhuowen Tu	2012	Author proposed an algorithm which is equipped with a two-level classification scheme and two sets of features specially designed for capturing both the intrinsic characteristics of texts. To better evaluate proposed algorithm and compare it with other competing algorithms, they generate a new dataset, which includes various texts in diverse real-world scenarios; we also propose a protocol for performance evaluation	This method outperforms on the Oriented Scene Text Database (OSTD), with an improvement of 0.19 in F-measure.
Tao Wang, David J. Wu, Adam Coates, Andrew Y. Ng	2013	In this paper, they combine the representational power of large, multilayer neural networks together with recent developments in unsupervised feature learning.Then, using only simple off-the-shelf methods, they integrate these two modules into a full end-to-end, lexicon-driven, scene text recognition system.	Achieved accuracies of 90% on I-WD-50,84% on I-WD,70% on SVT-WD.
Alessandro Bissacco, Mark Cummins, Yuval Netzer, Hartmut Neven	2013	PhotoOCR, a system for text extraction from images is presented .This approach is capable of perceiving text in a variety of challenging imaging conditions where traditional OCR systems fail, notably in the presence of substantial obscure, low resolution, low contrast, high image noise and other distortions.	Accuracy rises to 82.83%.
Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao	2014	A quick and active pruning algorithm is designed. Aspect candidates are grouped into text candidates by the single-link clustering algorithm, where distance density and clustering edge are found out naturally by a novel self-training distance metric learning algorithm. The back contingency of text candidates comparing to non-text are estimated with a aspect classifier.	The proposed system is evaluated on the ICDAR 2011 Robust Reading Competition dataset; the f measure is over 76% and is significantly better than the state-of-the-art performance of 71%.
Weilin Huang, Yu Qiao, and Xiaoou Tang	2014	The CNN system is introduced.This approach takes favorable circumstances of both MSERs and sliding-window based methods. The proposed system accomplished strong heartiness against various extreme text variations and genuine real-world problems.	Achieved over 78% in F-measure, which is significantly higher than previous methods.
Xu-Cheng Yin, Wei-Yi Pei, Jun Zhang and Hong-Wei Hao	2015	An effective multi-orientation scene text location system is proposed which constructs text candidates by grouping characters in view of Adaptive clustering.	Achieved F-measure of 71% on the multi-orientation text dataset MSRA-TD500.
Zheng Zhang, Wei Shen, Cong Yao, Xiang Bai	2015	A novel algorithm is proposed which mainly make use of the properties of single characters or strokes, the proposed algorithm exploits the symmetry property of character groups and allows for direct extraction of text lines from natural images.	The symmetry feature works better than the appearance feature. These two types of features are indeed complementary. Their combination leads to a significant boost in F-measure (from 0.72 to 0.80).
Tong He, Weilin Huang, Yu Qiao&Jian Yao	2016	A novel Text-Attentional Convolutional Neural Network (Text-CNN) is proposed that particularly concentrates on extracting text-related regions and features from the image components. One more learning structure is developed to train the Text-CNN with multi-level and rich managed information, including text region veil, character label, and binary text/non-text information.	Achieved a F-measure of 0.82, improving the state-of-the-art results substantially.
Jaderberg, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman	2016	Proposed an end-to-end system for text spotting—localizing and perceiving text in natural scene images—and text based image retrieval is introduced.	Achieved high recall, and a fast subsequent filtering stage for improving precision.

III. CONCLUSION

In this paper, it is been concluded that text detection and text recognition are the techniques of patterns recognition. The technique of segmentations are required which will segment the text portion from the input image. The neural networks technique will recognize the text from theinput image. In this paper , various techniques of image segmentation, edge detection and neural networks are reviewed and discussed.

ACKNOWLEDGMENT

We thanks the reviewers for comments and suggesting and additional references. We also thanks to ECE department faculties to giving some useful suggestions. Finally we thanks all those authors who made their technical reports and publications readily available on the internet and the world wide web.

REFERENCES

[1] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," 2011, in IEEE International Conference on Computer Vision (ICCV).

[2] C.Yao, X.Bai, W.Liu, Yi Ma, Z.Tu "Detecting Texts of Arbitrary Orientations in Natural Images'',2012,IEEE Part IV, LNCS 8692, pp. 497–511

[3] T.He, W.Huang, Yu Qiao, and Jian Yao," Text-Attentional Convolutional Neural Network for Scene Text Detection", 2016, IEEE Transactions on Image Processing.

[4] T.Wang, David J. Wu, A.Coates, Andrew Y. Ng," End-to-End Text Recognition with Convolutional Neural Networks", 2013, ICPR

[5] X.Yin, X.in, K.Huang, and H.Wei Hao," Robust Text Detection in Natural Scene Images", 2014, IEEE Transcations on Pattern Analysis and Machine Intelligence, Vol. 36, No. 5

[6] A.Bissacco, M.Cummins, Y.Netzer, H.Neven," Photoocr: Reading Text in Uncontrolled Conditions", 2013, IEEE International Conference on Computer Vision (ICCV).

[7] M.Jaderberg, K.Simonyan, A.Vedaldi, A.Zisserman," Reading Text in theWild with Convolutional Neural Networks", 2016, IEEE, Int J Comput Vis (2016)

[8] W.Huang, Y.Qiao, and X.Tang," Robust Scene Text Detection with Convolution Neural Network Induced MSER Trees", 2014, IEEE Part IV, LNCS 8692, pp. 497–511

[9] X.Cheng Yin, W.Yi Pei, J.Zhang and H.Wei Hao," Multi-Orientation Scene Text Detection with Adaptive Clusteri.ng", 2015, IEEE Transcations on Psttern Anslysis and Machine Intelligence.

[10] Z.Zhang, W.Shen, C.Yao, X.Bai," Symmetry-Based Text Line Detection in Natural Scenes", 2015, IEEE Computer Vision and Pattern Recognition (CVPR).

[11] B.Singh, and R.Maini," Skew Detection and Correction of Gurmukhi Words from Natural Scene Images", International Journal of Signal Processing, Image Processing and Pattern Recognition., Vol.9, 2016.

[12] C. Yao, X. Bai, and W. Liu, "A unified framework for multi oriented text detection and recognition," Image Processing, IEEE Transactions on, vol. 23, no. 11, pp. 4737–4749, 2014.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, in International Conference on Learning Representation (ICLR).

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding,'' 2014, in Proceedings of the ACM International Conference.

[15] 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.

