ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website:www.ijircce.com

Vol. 5, Issue 2, February 2017

Survey on Image Text Detection and Recognition of Natural Scene Images

M. Rubani[1]Dr. J. L. Walter Jeyakumar[2]

M. Phil Student, Dept. of C.S., St. Xavier's College (Autonomous), Tirunelveli, Tamil Nadu, India[1]

Associate Professor, Dept. of C.S., St. Xavier's College (Autonomous), Tirunelveli, Tamil Nadu, India[2]

ABSTRACT- The detection and recognition of text in a scene image is a very challenging problem in image processing becauseof the high complexity of background in the image. Analyzing of the text that appears on image is required to understand the content of an image or to note the significance of the information. Various methods where used to address this problem in past years. None of these where found to be perfect, because even a single algorithm is not known to be common for all images. These methods have their own bench mark data sets. There aremany relevant applications available for automatic recognition and information retrieval. They further opened up the possibility for more improvement in research methods. This paper presents in a short survey on various scene text detection and recognition methods which were implemented recently.

KEYWORDS: Text detection, Text recognition, Scene image, Convolutional neural network, Edge detection algorithm.

I. INTRODUCTION

In recent year, the use of technology involves more and more databases. They are identified in multimedia. These databases usually contain images and videos in addition to textual information.Such images may contain several contents like face, human, object, scene, text etc. Among all the contents, textual information is veryusefulto understand the image fully. There aredifferent kinds of images that are enriched by the text such as document images, scene images, caption images, born-digital images, and heterogeneousimages. Document images may bein the form of scanned book covers, printed documents, historical documents, CD covers or video images.

Texts in images are classified into scene text and caption text. Scene text is also called graphics text. Natural images that are contained in a text are called scene images. The name of caption text is recognized as artificial text. Born-digital imagesare saved as digital images. Heterogeneous images consist of all kinds of images such as scene images, caption images, document images and born-digital images [11].

Whiledetecting text from an image, thefactors such as font, color, style, orientation, character arrangement are to be considered. An image environmentalfactorssuch as lighting, shadows, secularities, and occlusions can be found. Image acquisition factors areresolution, motion, focus, and blur.

II. LITERATURE SURVEY

Many text detection and text recognition methods have been initiated in the last decade. Some of these methods are briefly reviewed in this paper.

A. EDGE DETECTION ALGORITHM:

This method is proposed to magnitude the second derivative of intensity as a measurement of edge strength, as this allows better detection of intensity peaks that characterized text in an image. A feature map suppresses the false regions and enhances true candidate text regions. This method distinguishes the text regions from texture like regions and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website:www.ijircce.com

Vol. 5, Issue 2, February 2017

extracted accurate binary character from the localized text regions so that the existing OCR can directly recognize the characters. A thresholding algorithm segments the text regions into white characters in a pure black background[1].

B. SOBEL EDGE DETECTION:

This paper presented a simple approach for efficient morphological operations and sobel edge detection method. Images are converted into grayscale image. It segments all the letters and numbers that are used in the number plate by using bounding box method and detect image region by using morphology.it uses sobeloperator to calculate the threshold value[6].

C. DISCRETE WAVELET TRANSFORM:

This paper proposed a process for text extraction from scene image and also from the document image. The image can be either color or gray. Gray image cannot be directly used an input image. Color image is converted into RGB intensely image, which is processed with 2-D discrete wavelet transform. The 2-D DWT wavelet decomposesthe input image into four components or sub-bands (LL, LH, HL, HH) based on sobeledge detector produces three sub-bands detected text edges. The minimum and maximum value of the projection is considered as a threshold. In such case, the pixel segment text would be in a black background[2].

D. MAXIMALLY STABLE EXTREMAL REGIONS:

In this paper, an accurate and robust method for text in natural scene images was proposed MSER pruning algorithm is used to detect almost all characters even when the image is in low quality. It has lowered variations and has shaped borders that are more likely to be characters. This algorithm includes the following stages. Most of the non-character are reduced in the character candidateextraction stage.Single-link clustering algorithm is used to change the character candidates to text candidate. Characterclassifier cluster removes the non-text region. The lowest variation is eliminated first in this strategy which may be a parent or a child. Linear reduction algorithm is applied across the whole tree recursively which also eliminates the linear variation if the tree has only one child. Tree accumulation algorithm is processed if the tree has more than one child. Single link clustering is used to construct the text candidates. Thetermination threshold is specified a hierarchical specified cluster tree. Using feature space defines the similarity between the data points. The distance metriclearns the distance between the clusters andproduces two closed members in each cluster. The text character is eliminated to use the non- text candidates. The algorithm is exhibited for superior performance over state-of-art method [3].

E. LEARNING ALGORITHM:

This is the Text detection and character recognition using learning algorithm. Earlier, this algorithm was used for visual recognition and audio recognition. Using a variant of K-means clustering, the result was much simpler and faster. Feature extraction method converts each image into 9d- dimensional feature vector. The sliding window boxes are used to decide whether object is text or non-text. Large banks of feature are able to achieve increasing accuracy with top performance when compared to other area of computer vison and machine learning[5].

F. DEEP CONVOLUTIONAL NEURAL NETWORKS:

This paper proposed a unified approach. There are three steps in deep convolutional neural networks that are operated directly on the image pixel. This model is configured with multiple hidden layers with feedforward connections. Map is used to find the wrong locations so that it minimizes the incorrect transcriptions entering into the map. These convolutional neural networks are used to receive the input features that are extracted from softmax classifier. Each softmaxmodels has used exactly the same backprop learning rule. In, SVHN dataset the image resized into 64*64 pixel so that the image loses a half percentage point of accuracy. This system achieves97.84%-character level accuracy. This is better than the previous state-of-art individual character task 67.53%. This architecture contains maxout units with three filters per unit and other layers contain rectifier units. The dropout is applied to all hidden layers but not to the input layers. Internal street views dataset does not resize the image because the network must be robust to a wider variation of scales[4].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website:www.ijircce.com

Vol. 5, Issue 2, February 2017

G. CONVOLUTIONAL NEURAL NETWORK:

This paper proposed an end-to-end text spotting system and processed both text spotting and image retrieval taskto achieve the result. The convolutional neural network formulate is recognized as a multi-class classification problem with one class perword. This convolutional neural network algorithm used 5 convolutional layers and in that 3 layers are fully connected layers. These are connected across the dictionary words. Input image must be in a fixed size to localize and recognize the individual word enclosed by boundary boxes. Greedy non-maximum suppression has also used the same method. During image retrieval process the query word is optional[7].

In this paper, a convolutional neural network based scene text detection algorithm along with a new text region extractor waspproposed to improve the independency and completeness of the extract candidate text regions. Using filter, extracted region is merged to remove the obvious non-text regions and merge the regions describing the same character. This algorithm has two convolutional layers. They can be noted as 64 and 96 filters with 6*6 and 4*4 kernel size. The final output of the convolutional neural network is considered as the input to the simple softmax classifiers. The convolutional neural network model has advantage in identifying the test regions from the candidate's text region by achieving the precision close to the best 0.82 in comparison with the best precision 0.83. The evaluations of the proposed algorithm has achieved the best recall and the state-of -art result[9].

This paper presented a new system for scene text detection by proposing a novel text convolutional neural network that particularly focuses on extracting text related regions and features from the image components. The system mainly includes two parts: A text attentional convolutional neural network for text component filtering classification and a contrast enhanced MSERs detector for generating component candidates. The problem is formulatedas a MTL with deep convolutional neural network model. This algorithm has three convolutional neural network layertask with kernel size 9*9, 7*7, 5*5 respectively. The region is segmented as character label and binary mask indicating explicit pixel-level segmentation of text. This algorithm is able to robustly discriminate ambiguous text for complicated background and has alsoachieved the state-of-art performance on a number benchmarks[10].

H. SUPPORT VECTOR MACHINE(SVM) CLASSIFIER:

This method proposed the contourlet transform to extract potential text regions from image based on edge features and classifies whether the image is text or non-text by using GLCM feature and SVM classifier. Horizontalprojection to extract the lines separately from the potential text region. Each line is segmented to extract characters using vertical projection[8].

Table 1 - Performance Comparison Between Different Algorithms with their Merits

(Precision=P, Recall=R,Fmeasure=F, Accuracy=A)

S.No	Author	Method	Accuracy	Merits
1	Xiaoquingliu [1]	Multi-scale edge detector, Feature map generation, Text region localization, Character Extraction	P=91.8 R=96.6	Used in variety of application fields such as mobile robot navigation, vehicle license detection object identification, etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website:www.ijircce.com

Vol. 5, Issue 2, February 2017

2	Xu-cheng yin [3]	MSERspruning algorithm, Distance Metric learning, Linear Reduction	P=86.29 R=68.26 F=76.22	Detect the characters even when the image is in low quality
3	Adam Coates [5]	Features learning algorithm	A= 85.5%	High performance, scalable solution
4	Ian J. Goodfellow [4]	Deep CNN	A=99.8%	Simultaneous localization, increase the quality
5	Max Jaderberg [7]	Edge Boxes, ACF detector, RF filter, CNN regression, CNN recognition	P=90.3 F=90 A=98.7	Performance on all standard datasets
6	Xiaohang Ren [9]	Multi- layer CNN, Statistical Learning, Isolated MSER	P=0.83 R=0.71 F=0.77	Detection text region in addressing challenges such as noise, contrast etc.
7	Tong He [10]	Text CNN, contrast Enhanced MSER	P=0.91 R=0.74 F=0.82	To robustly discriminate ambiguous text from complicated background
8	Shivananda V. Seeri [8]	Contourlet Transform, GLCM, SVM Classifier, Morphology	P=98.85 R=90.85 F=94.68 A= 89.90	Robust to complex backgrounds, Multilingual, Robust edge feature

III. CONCLUSION

Text is born out of different types of illumination factors such as color, size, shape, effects, lighting etc. All these factors are seen in image background. The rich and exact information is embodied in text, which can assist a wide range of real-world applications. Therefore, the detecting and recognizing text in natural scenes have been recognized as important research areas in computer vision. Most of the existing systems are concentrated with text in English and some other languages like Chinese, Korean. Only the English language text is widely considered research field. It is prominent for developing the detection and the recognition systems which have capacity to handle texts of di□erent languages also. This literature review is aimed at tracing the recent advancements in scene text detection and recognition. Table 1 describes the comparisons made from the survey which is helpful to identify new methods that extracts the textual information from natural scenes accurately and robustly.

REFERENCES

[1]Xiaoqing Liu and JagathSamarabandu, "Multiscale Edge-Based Text Extraction From Complex Images", IEEE 17211-4244-0367-7/06/$20.00 ©2006.
[2] Neha Gupta, V .K. Banga, "Image Segmentation for Text Extraction", ICEECE Singapore April 28-29, 2012.
[3]Xu-Cheng Yin, Xuwang Yin, Kaizhu Huang, and Hong-Wei Hao,"Robust Text Detection in Natural Scene Images", Technical report, arXiv:1301.2628v3[cs.CV] 2 Jun2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website:www.ijircce.com

Vol. 5, Issue 2, February 2017

[4] IanJ.Goodfellow, YaroslavBulatov, JulianIbarz, SachaArnoud, VinayShet, "Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks", arXiv:1312.6082v4 [cs.CV] 14 Apr 2014.
[5] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, Andrew Y. Ng, "Text Detection and Character
Recognition in Scene Image with Unsupervised Feature Learning", ICDAR 2011.
[6] Ragini Bhat1, Bijender Mehandia2,"Recognition of Vehicle Number Plate Using Matlab", IJIREEICE Vol. 2, Issue 8, August 2014.
[7] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman, "Reading Text in the Wild with Convolutional Neural Network', Springer 2015.
[8] Shivananda V. Seeri1, J. D. Pujari2, P. S. Hiremath1, "Text Localization and Character Extraction in Natural Scene Images using Contourlet Transform and SVM Classifier",I.J. Image, Graphics and Signal Processing,2016, 5, 36-42Published Online May 2016 in MECS (http://www.mecs-press.org/)DOI: 10.5815/ijigsp.2016.05.02
[9]Xiaohang Ren, Kai Chen, Jun Sun**,"**A Novel Scene Text Detection Algorithm Based On Convolutional Neural Network"**,** arXiv:1604.01894v1[cs.CV] 7 Apr 2016.
[10]Tong He, Weilin Huang, Member, IEEE, Yu Qiao, Senior Member, , and Jian Yao, Senior Member,**"**Text-Attentional Convolutional Neural Network for Scene Text Detection", IEEE,arXiv:1510.03283v2 [cs.CV] 24 Mar 2016.
[11] Uma B. Karanje, Rahul Dagade, "Survey on Text Detection, Segmentation and Recognition from a Natural Scene Images", International Journal of Computer Applications (0975 – 8887) Volume 108 – No. 13, December 2014.
[12] Xiaohang Ren, Kai Chen and Jun Sun,"A CNN Based Scene Chinese Text Recognition Algorithm With Synthetic Data Engine", arXiv:1604.01891v1 [cs.CV] 7 Apr 2016.