International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 21 (2018) pp. 14912-14917

Efficient Scene Text Recognition Using Energetic Edge Detectors, LDN Descriptor, Energy Enriched Self Organizing Map

Beula Bell. T[1], Dr. Jeyakumar M. K[2]

[1]*Research Scholar, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India.*

[2]*Professor, Dept. Of Computer Applications, Noorul Islam University, Kumaracoil, 629180, Tamil Nadu, India.*

Abstract

Onscene text recognition is a vital need for analysing scene images. The main challenges on this task are noise environment, intensity variation and font size variation. The existing methodologies developed for scene text recognition face the inaccuracy drawback in both in text localization and text character recognition. This scenario induces the real world to develop new scene text recognition systems to solve the demerits of the existing methods. This circumstance motivates this research to propose a novel method on scene text recognition namely *Efficient Scene Text Recognition Using Energetic Edge Detectors, LDN Descriptor, Energy Enriched Self Organizing Map*. The proposed approach is branched into two parts viz. scene text segmentation and scene text recognition. The scene text segmentation is constructed by Modified Decision based Unsymmetrical Trimmed Median Filter (MDBUTMF) Energetic *Edge* Detectors and *Local Directional number (LDN)* descriptor. The text character recognition which is influenced by the new method *Energy Enriched Self Organising Map* is constructed by *Histogram on Oriented Gradients (HOG) feature*, *Symlet transform* and *Self Organizing Map (SOM)*. The proposed method is employed on bench marked scene image databases and the experimental analysis is the evidence of significant improvement in accuracy of scene text recognition.

Keywords: Scene text segmentation, Text recognition, HOG, Symlet, EESOM

INTRODUCTION

Image processing is a wide technology for processing various data elements in an image. Now a day, images play a vital role in communicating a theme to the public in an effective manner. Due to the advent of image processing huge development occurs in the field of visual scene processing which enhances the science and technology.

Natural scene images are acquired via digital cameras from nature, which consists of huge capacity of text information that can be utilized to content matching in news reading, content examination by text matching and text removal in medical images. The dominant challenge in onscene text segmentation and recognition is the combination of graphics-and-text information located at the same document [1]. So the text region on onscene image should be extracted by the optimum techniques. The segmentation quality of onscene text induces the increment in text recognition results when employing optical character recognition (OCR) system. The onscene image may contain a sequence of words or even a single word which is in strong relationships with each other.

In this paper, text content from the complex on scene image is fully extracted in a wise manner. In order to extract the content of scene text images two steps are adopted. In the first step the text content area is marked by using the text detection algorithm which involves *Energetic Edge* detectors and *LDN* descriptor. This text area detection method is imparted from our previously published paper [10]. In the second step, the actual text character content in the onscene natural image is recognised by using the novel text recognition algorithm which is constructed based on HOG feature & Symlet transform induced Energy Enriched SOM. This new frame work efficiently segments and recognises the onscene text data.

Varieties of works for text analysis (detection, marking, extraction, recognition) in onscene digital images have been published in the past literature. According to the ways being used to identify text areas, most of the text detection techniques can be grouped as either connected component (or) texture based algorithms. In paper [2] edge based system is used for coarse detection accompanied by a multi resolution scheme for different sign size. It also combines the conventional 2-dimensional based recognition method with 3-dimensional pre-processing technology to enhance the texts in a 3-dimensional world. A hybrid approach to localize the Scene texts by integrating region information into a robust connected component (CC) based method [3] [4] [8]. In [5] the character specific part-based tree structure to model each category of characters so as to detect and recognize characters simultaneously. The final word recognition result is obtained by maximizing the character chain posterior probability via *Viterbi algorithm*. Stroke configuration map based on edge and frame [6] is projected for scene text recognition such as text recovery. Characterness representation reflects the probability of extracted region belonging to character which is constructed via fusion of novel character cues in the *Bayesian frame work*. In the character group model a criterion graph computes the characterness text, is capable to accomplish precise and strong results of scene text detection [7].

Convolution neural network is mainly focuses on extracting text related regions and features from the image components. It develops a new learning mechanism to train the Text-CNN with multi-level and rich supervised information, including text region mask, character label, and binary text/non-text information. The Low-level detector called *contrast-enhancement maximally stable external regions (MSERs*) is developed, which extends the widely used MSERs by enhancing intensity contrast between text patterns and background [9]. The Tracking based multi-orientation scene text recognition method uses multiple frames within a united framework via dynamic programming [11]. Text detection with multi-information fusion, text tracking with multiple tracking strategies, and integration of detection results with dynamic programming are highlighted in [12].

The papers [15] and [16], present an end-to-end fully convolution network for arbitrary-oriented text detection, which is highly stable and efficient to produce word proposals next to cluttered backgrounds. A fully convolution network to segment the candidate traffic sign areas, provides a fast neural system to detect texts on the extracted region on interest (ROI).

The section II describes proposed the new framework and the section III puts forth the analytic methodology to make a performance analysis of the proposed method. The Section IV confirms the peculiar advantages of the proposed method than the existing methods.

PROPOSED METODOLOGY

This research proposes an onscene text detection and recognition method for RGB colour images. This onscene text processing research is comprised of two main components. The first one is on scene text detection via *Energetic Edge Detector* and *LDN descriptor*. The second component is the onscene text recognition which separates each text characters and recognises it to built-up an *OCR* engine. The Fig.1 expresses the block diagram of proposed scene- text recognition method.

The combined process of this text detection can deeply referred with the paper [10].

B. Onscene Text Recognition

A novel method for text recognition is applied to character extraction and recognition. The query input image's text areas are marked by the previous module and the text characters are extracted one by one and that characters are recognized by the new *Energy Enriched SOM method*. This novel text recognition work is built by the following sub modules.

- Energy Enriched SOM based training
- Query text character extraction
- Query character's feature extraction
- Energy Enriched SOM testing
- Post Processing

a) Energy Enriched SOM based training

The training set of n fonts is chosen for each character. So for a single character there is n samples. The dimension of the specified character is assumed as [*H, W]*. Each character is altered into a standard block size in the dimension of *BH*BW* which are here maintained as *18*23*. The *BH* value 18 is derived by the highest character *Q* and the width *BW* is derived from the lengthiest character *W*. The *padding* technique is adapted to alter the block size of each character and now it is a *BH*BW* size matrix. The n font information of this dimensions are converted into vector format. Some of the characters are undergone the uncertainty issue and they are listed as

- i ,I and l
- 0,o and O
- s and S
- c and C
- v and V
- w and W
- x and X
- z and Z.

For example the characters i, I and l are confused among them in shape structure and there is a uncertainty in matching process, These type of characters are flagged with uncertainty marking.

The SOM networks are created based on the configuration like epochs=100, goal=0.01. The *Energy Enriched SOM* is trained using the n samples of each characters based on MATLAB'S built in function.

This SOM training yields the weight vectors for the trained network. The trained weight vectors contain more fractional values, so that they are optimized based on equation 1 and equation 2.

$$wv = WV\left(fix\left(\frac{n}{2}\right)\right) \quad (1)$$

$$wv' = \begin{cases} 1, if\ wv > 0.1 \\ 0, otherwise \end{cases} \quad (2)$$

Where

WV –SOM's weight vector contains n count with *l* length

wv – Center positioned weight vector

wv' –optimized weight vector

Normally SOM network optimizes the weight vectors by processing the neighbour vectors and the median vector contains much information or optimized values, so the median weight vector is selected as the optimized weight vector and this is referred by equation 1. The fractional

A. Onscene text detection In this module, the onscene text detection is carried out through the *Trimmed Median Filter, Energetic Edge Detector and LDN descriptor.* This research utilizes the paper [10]. The input onscene images may be affected by the salt and pepper noise problem. These noises can impact the output quality of text detection process. So Trimmed Median filter is employed to recover the image from noises by an efficient way. This manuscript uses the Trimmed Median Filter variant, namely, *Modified Decision based Unsymmetrical Trimmed Median Filter (MDBUTMF)* to accomplish the denoising work and this filter is deeply explained in the paper [19]. The onscene text detection process is mainly packed with the following subsequent functionalities.

- Modified Decision based Unsymmetrical Trimmed Median Filter
- Energetic Edge Detection
- Local directional number feature image generation
- Linked map generation
- Non scene text rejection
- Scene text marking

weight values are lifted to whole numbers through the equation 2 which will improve the matching process.

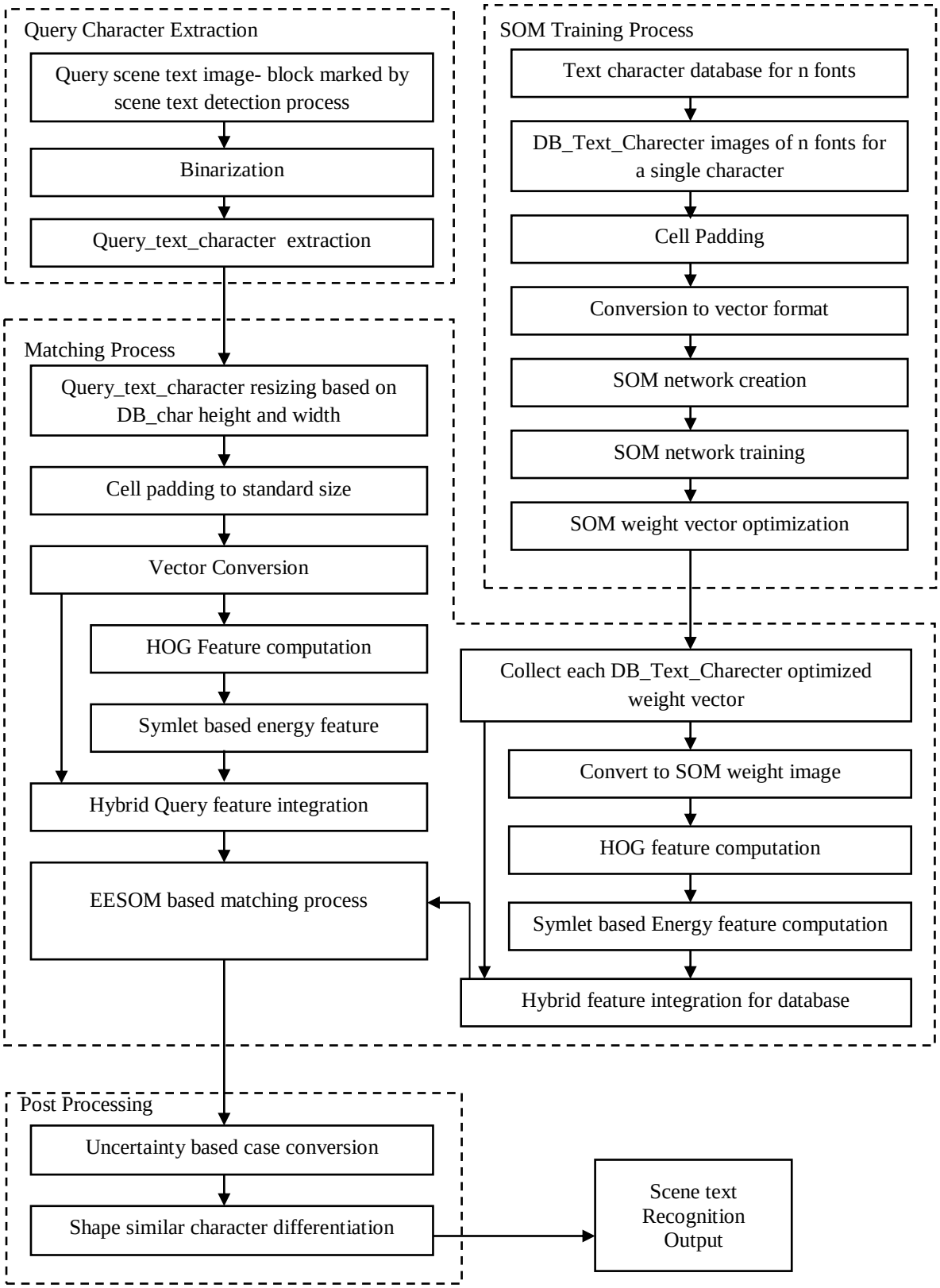

Fig. 1 Block diagram of the proposed method.

The HOG [23] transform is used to find features from Histogram On Gradients [20][21]. The HOG computation is done in five steps and they are:

- Gradient computation
- Create bins associated with orientation
- Block formation
- Block normalization
- HOG feature vector formation

The block normalization is obtained using equation 3.

$$b \leftarrow \frac{b}{\sqrt{||b||^2+e}} \quad (3)$$

Where

e- Small positive constant

b- Block information

The HOG features encode local shape information from regions with an image and they are converted in to energised features using symlet transform [22] which works based on wavelet concept. Symlets are better suitable for feature computation from *1D* signal model data. The hybrid feature set is constructed by the optimized weight vector and the HOG-Symlet feature corresponding with database training.

b) Query Text Character Extraction

The query characters are extracted from the scene text marked image one by one by processing lines and word-breaking-spaces.

c) Query Character Feature Extraction

The query characters are converted into standard size of BH and BW through the padding process. Then it is converted into *vector* type features. The query character images are undergone the HOG feature detector and Symlet based feature detection. The vector data and the HOG-symlet features are concatenated into the hybrid feature set.

d) Energy Enriched SOM Testing

The Energy Enriched SOM based Testing is achieved through the equation 4.

$$MS^k = \sum_{i=0}^{l-1}(DBF^i - QF^i)^2 \quad (4)$$

Where

MS –Matching score against k^{th} database character for the query image

l –Length of query image's combined feature

DBF- database features

QF- query feature

The less scoring database character is announced as the intermediately matching character.

e) Post Processing

The uncertainty flagged matching is further refined by the post-processing using the *font height* property. The small variation of shape –similar characters are considered and solved using the post-processing task through the extension differences in font.

In this way each character of the query is recognized and the text information are displayed.

EXPERIMENTAL RESULTS AND ANALYSIS

This research implements the proposed onscene text detection and recognition method and the experimental results are tabulated in tables and drawn as chart to get results with clarity. This paper analyzes the proposed method against the following three existing methods related with onscene text detection and recognition.

1. Tong He et al. method [9]
2. Youbao Tang et al. method [13]
3. Sezer Karaoglu et al. method [14]

This analysis part is analysed with the state-of-the-art analytic methods to construct a better report on the performance behaviour of the proposed method compared with the existing methods.

This research uses two state-of-the-art onscene image databases namely KAIST [17] and ICDAR [18].

(a) Input onscene image

(b) Scene text marked output.

Fig.2. Scene text detection by the proposed method

The Fig.2 illustrates the output of the onscene text segmentation of the proposed method. The onscene texts are separated and marked in the Fig.2.b.

The PSNR analysis is performed using the binarized version of segmented image and the ground-truth image using the equation 5.

$$PSNR = 10\log_{10}(\frac{255^2}{MSE}) \quad (5)$$

Where

MSE –mean square error

Table1: Peak Signal to Noise Ratio (PSNR) for Text segmentation

Data base Name	Image Name	PSNR(in db)			
		Tong. He method	Youbaou method	Sezer k method	Pro posed metho d
KAIST	KAIST Img 1	58.34	59.18	61.27	**64.23**
	KAIST Img 2	58.86	60.48	62.21	**65.14**
	KAIST Img 3	57.08	58.87	60.46	**63.17**
	KAIST Img 4	57.52	59.44	61.92	**64.57**
ICDAR	ICDAR Img1	56.82	58. 43	60.73	**63.51**
	ICDAR Img 2	58.91	60.27	62.03	**64.31**
	ICDAR Img 3	56.28	57.10	59.54	**62.84**
	ICDAR Img 4	57.13	58.12	59.94	**62.52**

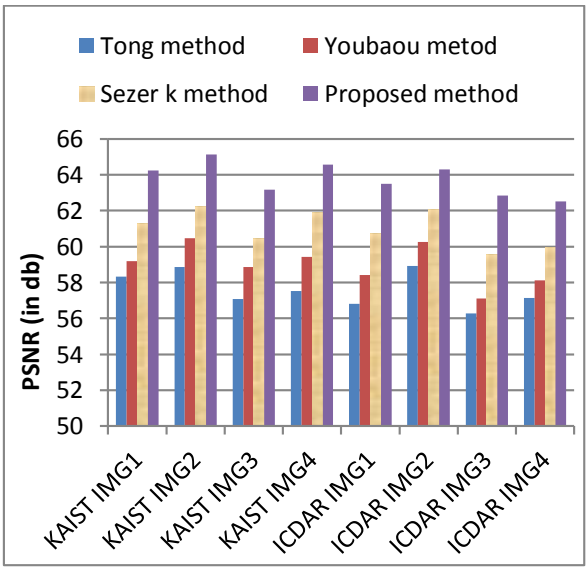

Fig.3. PSNR analysis chart for Scene text segmentation.

The proposed method obtains the higher PSNR than the existing methods and it can be proven by the Table 1 and Figure 3. The Figure 3 represents the PSNR values for the KAIST database and ICDAR database. This PSNR analysis clearly indicates the better segmentation of the proposed method than the existing methods because the proposed method reaches high PSNR than others for both databases. The highest value of PSNR related with proposed method for KAIST database is 65.14 and for ICDAR database it is 64.31. The *positive class* is referred by all the text in the recognized output file (whether it may be correctly or incorrectly recognized). The term *true positive* refers the correctly recognized items in the positive class. The *false negative* term indicates the items that are not labelled as belonging to the positive class, but should have been recognized. The highest recall percentage of KAIST database for the proposed method is 84 and the highest recall of ICDAR database is 83.

Table2: Recall analysis for Scene text recognition

Data base Name	Image Name	Recall			
		Tong. He method	Youbaou method	Sezer k method	Pro posed metho d
KAIST	KAIST Img 1	72	74	77	**82**
	KAIST Img 2	72	73	76	**81**
	KAIST Img 3	71	74	78	**84**
	KAIST Img 4	70	72	76	**82**
ICDAR	ICDAR Img 1	72	73	75	**81**
	ICDAR Img 2	71	74	78	**83**
	ICDAR Img 3	72	73	77	**82**
	ICDAR Img 4	70	73	76	**83**

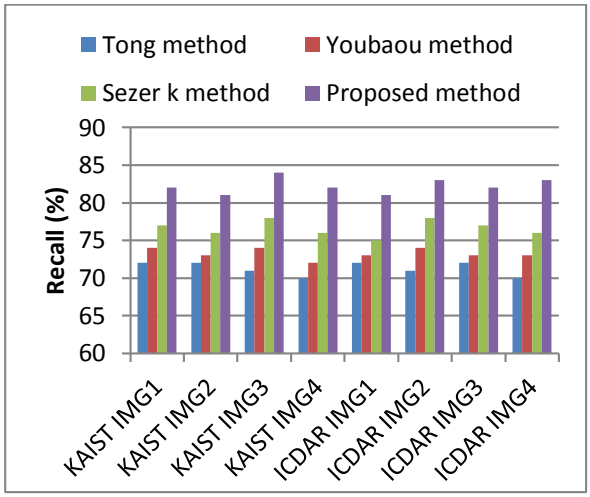

Fig.4. Recall analysis chart for text recognition.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Where

TP -True Positive
FN -False Negative

The proposed method achieves better recall percentage than the existing methodology which shows the extended character recognition power of the proposed method.

CONCLUSION

The proposed method efficiently detects the texts positioned in onscene images and recognized them with higher accuracy. The main contribution of this manuscript is set with the scene text character recognition. The proposed method can be calibrated with set of n fonts. The proposed method produces higher recall for text recognition than the existing methods. The average recall obtained by the proposed method by considering the two databases is 83.5 which are much higher than the existing methods. This novel research can compete

with the current onscene recognition schemes with better reliability and accuracy. In future this approach can be improved by optimized training with more fonts related with additional languages like Tamil and Hindi.

REFERENCES

[1]. Xilin Chen, Jie Yang, Jing Zhang, and Alex Waibel,"Automatic Detection and Recognition of Signs From Natural Scenes", IEEE transactions on image processing, vol. 13, no. 1, pp. 87-99, 2004.

[2]. Wonjun Kim and Changick Kim, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene", IEEE transactions on image processing, vol. 18, no. 2, pp. 401-411, 2009.

[3]. 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, pp. 800-813, 2011.

[4]. Hyung Il Koo, and Duck Hoon Kim, "Scene Text Detection via Connected Component Clustering and Nontext Filtering", IEEE transactions on image processing , vol. 22, no. 6, pp. 2296-2305, 2013.

[5]. Cun-Zhao Shi, Chun-Heng Wang, Bai-Hua Xiao, Song Gao, and Jin-Long Hu, "Scene Text Recognition Using Structure-Guided Character Detection and Linguistic Knowledge" IEEE transactions on circuits and systems for video technology, vol. 24, no. 7, pp. 1235-1250, 2014.

[6]. Chucai Yi and Yingli Tian,"Scene Text Recognition in Mobile Applications by Character Descriptor and Structure Configuration", IEEE transactions on image processing, vol. 23, no. 7, pp. 2972- 2982, 2014.

[7]. Yao Li, Wenjing Jia, Chunhua Shen, and Anton van den Hengel," Characterness: An Indicator of Text in the Wild", IEEE transactions on image processing, vol. 23, no. 4, pp. 1666-1677, 2014.

[8]. Huan Yang, Shiqian Wu, Chenwei Deng, and Weisi Lin,"Scale and Orientation Invariant Text Segmentation for Born-Digital Compound Images", IEEE transactions on cybernetics, vol. 45, no. 3, pp. 533-547, 2015.

[9]. Tong He, Weilin Huang, Yu Qiao, and Jian Yao,"Text-Attentional Convolutional Neural Network for Scene Text Detection", IEEE transactions on image processing, vol. 25, no. 6, pp. 2529- 2541, 2016.

[10]. Beula Bell T and M. K. Jeya Kumar, "Scene Text Segmentation and Recognition by Applying Trimmed Median Filter Using Energetic Edge Detection Schemes and OCR", Journal of Network Communications and Emerging Technologies, vol. 7, issue 12, pp. 57-66, 2017.

[11]. Sezer Karaoglu, Ran Tao, Jan C. van Gemert, and Theo Gevers,"Con-Text: Text Detection for Fine-Grained Object Classification", IEEE transactions on image processing, vol. 26, no. 8, pp. 3965 -3980, 2017.

[12]. Chun Yang, Xu-Cheng Yin, Wei-Yi Pei, Shu Tian, Ze-Yu Zuo, Chao Zhu, and Junchi Yan,"Tracking Based Multi-Orientation Scene Text Detection: A Unified Framework With Dynamic Programming", IEEE transactions on image processing, vol. 26, no. 7, pp. 3235-3248, 2017.

[13]. Youbao Tang and Xiangqian Wu,"Scene Text Detection and Segmentation Based on Cascaded Convolution Neural Networks", IEEE transactions on image processing, vol. 26, no. 3, pp. 1509- 1520, 2017.

[14]. Sezer Karaoglu, Ran Tao, Theo Gevers, and Arnold W. M. Smeulders," Words Matter: Scene Text for Image Classification and Retrieval", IEEE transactions on multimedia, vol. 19, no. 5, pp. 1063-1076, 2017.

[15]. Minghui Liao, Baoguang Shi, and Xiang Bai ,"TextBoxes++: A Single-Shot Oriented Scene Text Detector", IEEE transactions on image processing, vol. 27, no. 8, pp. 3676-3690, 2018.

[16]. Yingying Zhu, Minghui Liao, Mingkun Yang, and Wenyu Liu,"Cascaded Segmentation-Detection Networks for Text-Based Traffic Sign", IEEE transactions on intelligent transportation systems, vol. 19, no. 1, pp. 209-219, 2018.

[17]. KAIST Scene Text Database, Prof. Jin Hyung Kim, Email: Jkim @ kaist.ac.

[18]. ICDAR-TEXT-DATASET, https://github.com>Total-Text-Dataset

[19]. S. Esakkirajan, T. Veerakumar, Adabala N. Subramanyam, and C. H. PremChand,"Removal of High Density Salt and Pepper Noise Through Modified Decision Based Unsymmetric", IEEE Signal processing letters, vol. 18, no. 5, May 2011.

[20]. Boran Yu and Hongjie Wan,"Chinese Text Detection and Recognition in Natural Scene Using HOG and SVM", International Conference on Information Technology for Manufacturing Systems, pp.148-152, 2016.

[21]. P. Shiva Reddy and M.N.Giri Prasad,"Extracting text from natural scene images by HOG Character Descriptor", International Journal of Advanced Research in Electronics and Communication Engineering, vol. 5, pp. 2518-2523, 2016.

[22]. Ankush Gautam," Segmentation of Text from Image Document", International Journal of Computer Science and Information Technologies, vol. 4 (3), pp. 538-540, 2013.

[23]. Everton B. Lacerda and Carlos A.B Mello,"Segmentation of Touching Handwritten Digits Using Self-Organizing Maps", IEEE International Conference on Tools with Artificial Intelligence, 2011.