Automatic Text Detection and Classification in Natural Images

C.P. Chaithanya, N. Manohar, Ajay Bazil Issac

Abstract--- Text detection is the method of locating areas in a picture wherever, text is present. Text detection and classification in natural pictures is very important for several computer vision applications like optical character recognition, distinguish between human and machine inputs and spam removal. Currently the challenge in text identifying is to detect the text in natural pictures due to many factors like, low-quality image, unclear words, typical font, image having a lot of color stroke than the background color, blurred pictures due to some natural problems like rain, sunny, snow, etc. The main aim of this work is to identify and classify the text in natural pictures. Here system detects the text and finds the connected regions, chainthem together in their relative position. Uses a text classification engine to filter chains with low classification confidence scores.

Keywords--- CNN, OCR, Extraction.

I. INTRODUCTION

Text detection in natural images has vital role in the field of artificial intelligence, augmented reality and other innovations. It helps to remove noise in images and identify text. Nonetheless, it is additionally a difficult issue because of the variability in imaging conditions, for example, lighting, specular reflections, commotion, obscure, and nearness of blocks over the content, and in the changeability of the content itself, for example, its scale, introduction, textual style, and style. Great text discovery calculations should accordingly be vigorous against such fluctuations. The text detection plays an important role in life as it is used for vision type applications. Currently it faces difficulties like background complexity, different direction of the text, complexity in backgrounds, diversity of scene text and interference factors etc. In order to overcome these difficulties, CNN (Convolutional neural networks) and related methods are used [21], and also in some cases [20] and valley and ridge techniques are used. Here the result is to detect the text and increasing the accuracy level. Here to implementing different methods for text detection and classification.

II. RELATED WORK

In this section, we present a brief literature about text detection and classification. Amritha S Nadarajan &; Thamizharasi A (2018), proposes an innovative algorithm to find value of stroke width in natural images. This algorithm helps to detect many font and languages. This includes pre-processing, 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, analyses 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. Chandio, A. A., Pickering, M., and Shafi, K. (2018), using the technique hybrid technique for enhancement of the image, robust technique for background subtraction, supervised machine learning algorithm for maps an input to an output based on example input output pair, applying Gaussian filter method for removing the noise, and use histogram-oriented gradient (HOG) for detection of the object. Tridib Chakraborty et al (2017), here check the noise and skew of the image. This is the first step. This is called pre-processing. The image can be converted in to grayscale. And then into binary. After the image can be segmented. That means text can be divided in to individual character. Feature extraction is the step that follow the segmentation. The classification of the text is done using the feature extraction. Text extraction process is the important method that is used in text detection and classification. There is some architecture used to extract the text. Like text detection, text localization, text tracking. Karaoglu, S., Tao, R., van Gemert, J. C., Gevers, T. (2017), they applying those methods like connected component approach, visual saliency, multi model fusion, fine grained classification and content for object categorization. Guan, L., Chu, J. (2017, character candidate extraction for finding the specific area in the image, preliminary filtering based on heuristic rules for solving the difficult problems, related text training, and text line aggregation methods used. Zhu, Q. H., Zhu, R., Li, N., Yang, Y. B. (2017, here using meiric learning method, using triplet selection strategy, and neural network structure equipped with triplet loss. Shi, B., Bai, X., Belongie, S. (2017), using the methodologies like convolutional neural network for deep learning, object detection for detecting the object, within link layer detection and cross layer link detection. Zhong, Z., Jin, L., Huang, S. (2017), to find text region proposal generation, text detection: ATC inter cooperation, and learning optimization for deep learning. Karaoglu, S., Tao, R., Gevers, T., Smeulders, A. W. (2017), Here to check word level textual cue encoding, multimodal classification for compared

Manuscript received February 01, 2019

C.P. Chaithanya, Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidhyapeetham, Mysuru Campus, Karnataka, India. (e-mail: chaithuraj97@gmail.com)

N. Manohar, Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidhyapeetham, Mysuru Campus, Karnataka, India. (e-mail: manohar.mallik@gmail.com)

Ajay Bazil Issac, Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa Vidhyapeetham, Mysuru Campus, Karnataka, India. (e-mail: ajaybazilissac93@gmail.com)

through the experiments, final ratio for. Pirker J, Wurzinger G (2016), the OCR (optical character recognition) system have introduce new techniques and algorithm. For this, applied many methods to improve the accuracy. Here mainly consider the old font text that means historical text. And here also we can use the different learning systems such as neural networks and export system. Jeong, M., Jo, K. H. (2015), here we can use edge detection for finding the objects in a specific boundary, edge component labelling for labelling the edge components, tree of edge component generating, component aggregation for finding relation between the components. Jacob, J., Thomas, A. (2015), here they using the methods image pre-processing, edge detection for finding the edges, stroke width transforms for finding stroke width, form connected component, filtering for removing noise, feature extraction to extract the feature components, and finally classification using random forest for learning m method Classification. Rani, N. S., & Vasudev, T. (2015), here they using OCR (optical character recognition) method for detect the text and extraction method applied for extract the text. Rani, N. S., & Vasudev, T. (2015), here using Optical Character Recognition method for recognize the text. Rong, L., Suyu, W., Shi, Z. (2014), connected component extracting, finding connected components group, using SVM classifier for learning the machine algorithms, classification of CCs, and CRF based post processing. Yao, C., Bai, X., Liu, W. (2014)., here they use the methods like classification scheme for text detection and character recognition, component linking and work partition, error correlation in character recognition, case disambiguation and training data. Pise, A.,Ruikar, S. D. (2014), text region detection method(HOG) is used to detect the text, image segmentation (Niklack's local binarization algorithm), image centroid and zone (ICZ) based distance metric feature extraction system for extracting the features, zone centroidand zone (ZCZ) based distance metric feature extraction system, and ICZ+ZCZ based distance metric feature extraction system. Meng, Q., Song, Y., Zhang, Y., Liu, Y. (2013), in this paper to implement the methodologies like, edge segmentation stroke width transforms for segmenting the edges using stroke width transform, color description for describing the colors, to merge edge for combining the edges, edge classification for edge and be classifying. A. Coates et al., (2011), computes the stroke an efficient algorithm is using a new feature. In a Photograph the text detection is the challenging problem. And the important key components are character recognition and text detection. Here mainly the learning algorithms are used for detecting the text. Example is unsupervised learning algorithm. The k-means clustering algorithm used to train the learning features. The scalable feature learning algorithm applied it to images of text in natural images. B. Epshtein, E. Ofek and Y. Wexler, (2010), the image retrieval algorithm is

used to detect the text. There are so many techniques are developing like SVM, filter, convolutional neural network (CNN). Shao Y., Wang C., Xiao B., Zhang Y., Zhang L., Ma L. (2010), the 'and valley image'(AVI) and 'and ridge image' (ARI) are the first extract input images. The components of the character are detected from AVI and ARI. And finally find the text in and each image. Text is the important way for convey the information. Some of the methods have been proposed to extract the text image. One of the major difficulties is that to retrieving the indoor and outdoor texts. Park, J., Lee, G. (2008), using the filtering algorithm for removing noise, using SVM for analyzing the learning algorithm, using feature extraction in candidate regions based on wavelet transform for representing a series of values.

III. METHODOLOGY

Our proposed methodology has 7 stages: pre-processing, segmentation, MSER, stroke width calculation, feature extraction, text detection. Figure 2 shows the architecture of the proposed system. Initially, filtering method is used for preprocessing to enhance the images. Later segmentation is done using thresholding to remove the background and extracting only region of interest. After that we find the MSER region, and removing the non-text region. Then calculate the stroke width using stroke width variation. Finally convolutional neural network is used to extract the features to identify and extract the text which will be given to OCR to get the text.

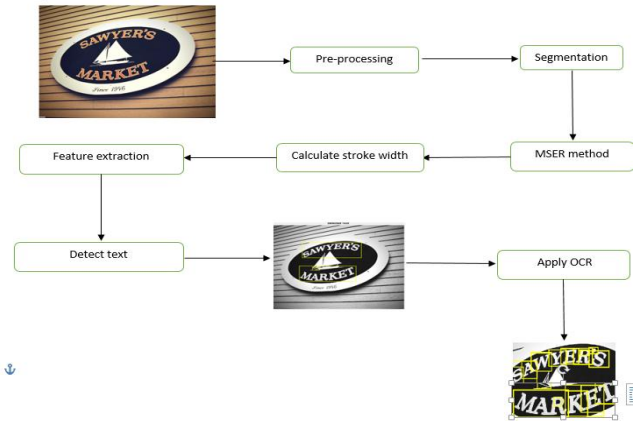

Figure 1: The architecture diagram of proposed system

i. Pre-processing

First, we consider text images which are of different forms, distorted horizontally and again vertically. Pre-processing may be a common name for operations with pictures at all-time low level of abstraction -- each input and output square measure intensity pictures. Pre-processing technique are used to improve the quality of the image. There are 4 ways of images preprocessing, they are: v element brightness, v preprocessing, v image restoration. Alternative classifications of image pre-processing ways exist.

Median filter pre-processing method is used in the proposed system which is a non-linear filtering method that is used to remove the noise. This type of noise reduction is a typical pre-processing step to improve the results of later processing. Median filter is used in digital image processing because under certain conditions, it preserve edge while removing noise also having applications in signal processing. Figure 3 shows after pre-processing images

Figure 3: sample pre-processing images

ii. Segmentation

Segmentation is used for partitioning the digital images in multiple regions or segments. These multiple segments are in the form of pixels. The segmentation result is a set of segments images. For segmentation I used thresholding method. The thresholding is the simplest segmentation method. The thresholding can be used to create the binary image from the grayscale image. Figure 3.1 shows the sample threshold images.

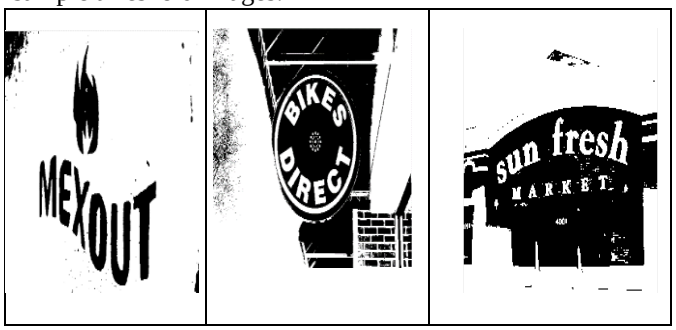

Figure 3.1: sample threshold images

iii. Finding Maximally Stable External Region

MSER is used to get the outline of the text in a proper form and to avoid the broken or non-uniform alignment of pixels. It is a method that used to extract number of co-variant regions in an image. The MSER not only represents the text region, it also represents the background region as well, which is helpful to classify the text into words or meaningful sentences. Figure 3.3 shows the text images of MSER region.

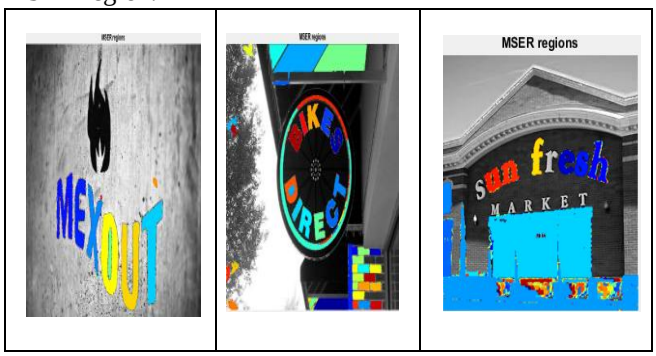

Figure: 3.3: MSER region images

a) Removing the non-text region based Geometric properties.

Using MSER, connected component regions are detected. Geometric properties are used to filter out non-text region using threshold. Some of the geometric properties that are used in this system include:

- Eccentricity: It is used to count the circular nature of the given regions.
- Solidity: It is the proportion of the pixels in the raised structure area that are likewise in a given region. It is calculated by,

  Area / raised area
- Extent: The location and size of the rectangle.
- Euler Number: It is a feature of the binary image.

Here figure 3.3.1 shows the removed non-text region based on geometric properties

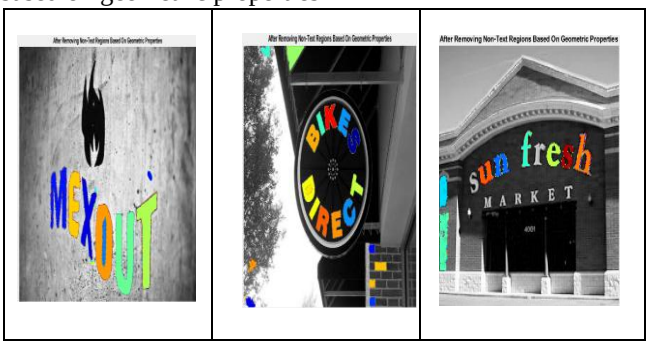

Figure 3.3.1: non-text region images based on geometric properties

iv. Region image and Stroke Width image

Stroke Width transform is an operator, which identify each pixel of a text in the image and groups the pixels in the letter candidate based on the stroke width. The figure 3.5 shows the Region image and Stroke Width images,

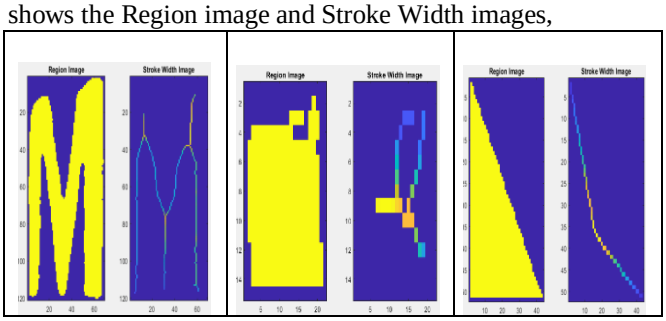

Figure 3.5: region images and stroke width images

b) Removing no-text region based on Stroke Width variation

As per the stroke width calculation, regions with presence of pixels of text are differentiated with background pixels and these background pixels are removed. Finally, the resulting text regions are displayed. Figure 3.4.1 shows only the text region,

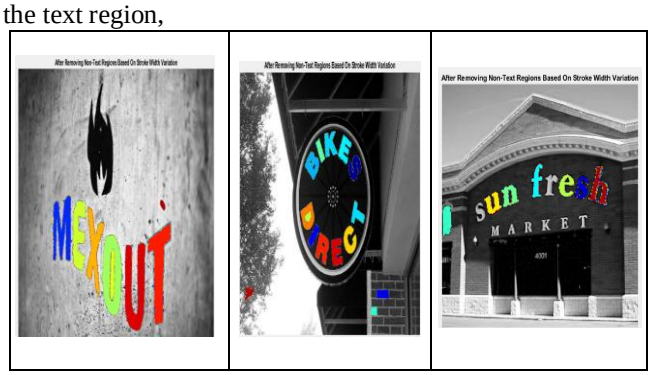

Figure 3.4.1: non-text region based on stroke width variation

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

v. Expanded bounding boxes

At this case all the detection result can be combined, and composed an indivichual text character. The meaning full character can be combined and form a new text region. Each bounding box can be represented each text. Figure 3.6 shows the bounding box texts.

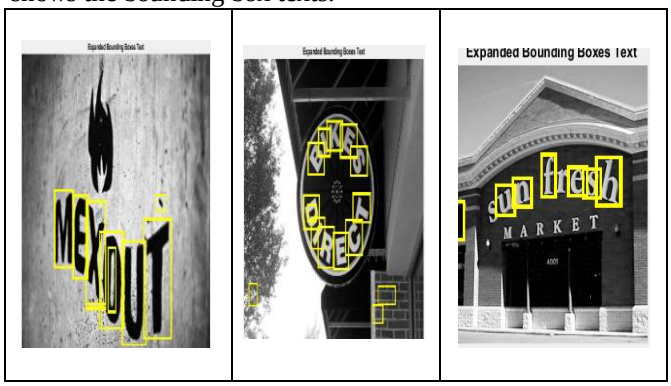

Figure 3.6: bounding box text images

IV. EXPERIMENTAL RESULTS

vi. Detected text

The detected text, which forms meaningful words or sentences are displayed. Figure 4 shows the detected text image,

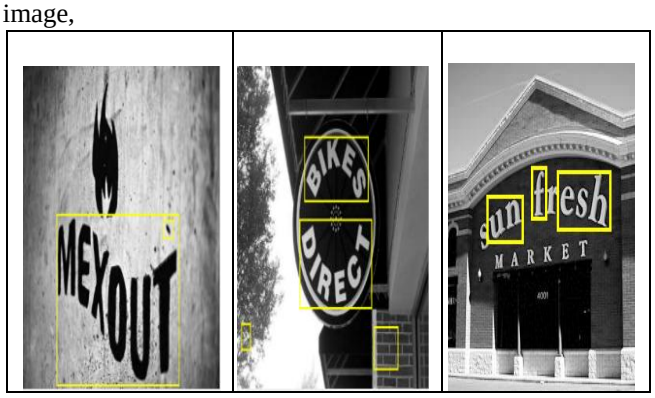

Figure 4: detected text images

vii. Data Set

Dataset used is from ICDAR (International Conference on Document Analysis) 2017. The ICDAR dataset consist of two types of images, namely the training image dataset and images which should be used for testing the proposed system. Figure 4.1(a) shows some of the train and, figure 4.1(b) shows

Figure 4.1(a): sample train images

Figure 4.1(b): sample test images

viii. Training Dataset

CNN is used to train dataset and classification. And the feature extraction is automatic. The Convolutional Neural Network is the one of the important classification method. It is used to find the accuracy of the training dataset. The accuracy of the training dataset is 92.31% percentage. Figure 4.2(a) shows the training dataset graph, and loss dataset is shows figure 4.2(b),

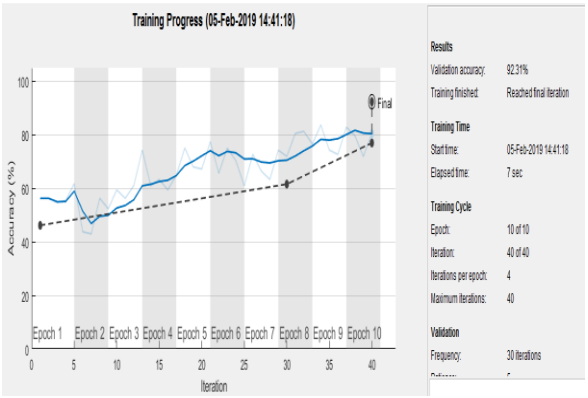

Figure 4.2(a) the training dataset graph

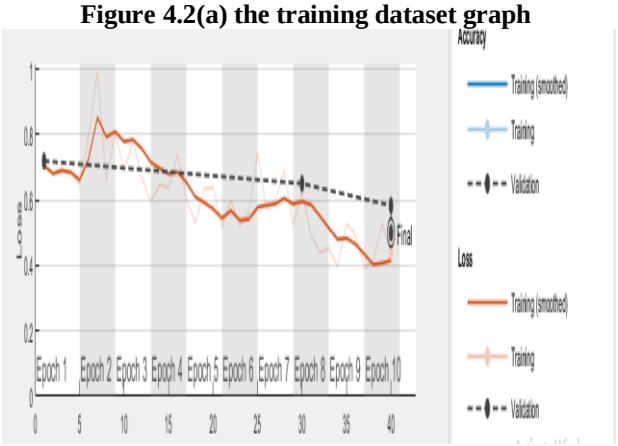

Figure 4.2(b): the loss in training dataset

V. CONCLUSION

Text detection is applicable in real world scenarios like optical character recognition, artificial intelligence, distinguish between human and machine inputs and spam removal. Text detection is the process of locating areas in an image where, a meaning full text is occurred. Variation in environment in which the image is captured makes it a difficult process.

The characters which are identified are classified in to meaning full word or sensible sentences. The identified words are chained together and checked weather they form the meaning full sentence. In this work, a system to detect text and classify of the same is presented. The accuracy of the proposed system is 92.31%.

REFERENCES

1. Amritha S Nadarajan, Thamizharasi A, "A Survey on Text Detection in Natural Images", International Journal of Engineering Development and Research (IJEDR), ISSN: 2321-9939, Volume.6, Issue 1, pp.60-66, January 2018.
2. Chandio, A. A., Pickering, M., & Shafi, K. (2018, March). Character classification and recognition for Urdu texts in natural scene images. In Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on (pp. 1-6). IEEE.
3. Tridib Chakraborty et al, (2017), Text recognition using image processing, International Journal of Advanced Research in Computer Science, 8 (5), May-June 2017, 765-768
4. Karaoglu, S., Tao, R., van Gemert, J. C., & Gevers, T. (2017). Con-Text: Text Detection for Fine-Grained Object Classification. IEEE Transactions on Image Processing, 26(8), 3965-3980.
5. Guan, L., & Chu, J. (2017, June). Natural scene text detection based on SWT, MSER and candidate classification. In Image, Vision and Computing (ICIVC), 2017 2nd International Conference on (pp. 26-30). IEEE.
6. Zhu, Q. H., Zhu, R., Li, N., & Yang, Y. B. (2017, October). Deep metric learning for scene text detection. In Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on (pp. 1025-1029). IEEE.
7. Shi, B., Bai, X., & Belongie, S. (2017). Detecting oriented text in natural images by linking segments. ArXiv preprint arXiv: 1703.06520.
8. Zhong, Z., Jin, L., & Huang, S. (2017, March). Deeptext: A new approach for text proposal generation and text detection in natural images. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 1208-1212). IEEE.
9. Karaoglu, S., Tao, R., Gevers, T., & Smeulders, A. W. (2017). Words matter: Scene text for image classification and retrieval. IEEE Transactions on Multimedia, 19(5), 1063-1076.
10. Pirker J, Wurzinger G. Optical Character Recognition of Old Fonts – A Case Study. The IPSI BgD Transactions on Advanced Research. 2016.
11. Jeong, M., & Jo, K. H. (2015, January). Multi language text detection using fast stroke width transform. In Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on (pp. 1-4). IEEE.
12. Jacob, J., & Thomas, A. (2015, December). Detection of multioriented texts in natural scene images. In Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2015 International Conference on (pp. 625-628). IEEE.
13. Rani, N. S., & Vasudev, T. (2015, December). Automatic detection of Telugu single and multi-character text blocks in handwritten words. In 2015 International Conference on Computing and Network Communications (CoCoNet) (pp. 234-240). IEEE.
14. Rani, N. S., & Vasudev, T. (2015, December). Post-processing methodology for word level Telugu character recognition systems using Unicode Approximation Models. In *2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)* (pp. 1-7). IEEE.
15. Rong, L., Suyu, W., & Shi, Z. (2014, April). A two level algorithm for text detection in natural scene images. In Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on (pp. 329-333). IEEE.
16. Yao, C., Bai, X., & Liu, W. (2014). A unified framework for multioriented text detection and recognition. IEEE Transactions on Image Processing, 23(11), 4737-4749.
17. Pise, A., & Ruikar, S. D. (2014, April). Text detection and recognition in natural scene images. In Communications and Signal Processing (ICCSP), 2014 International Conference on (pp. 1068-1072). IEEE.
18. Meng, Q., Song, Y., Zhang, Y., & Liu, Y. (2013, September). Text detection in natural scene with edge analysis. In Image Processing (ICIP), 2013 20th IEEE International Conference on (pp. 4151-4155). IEEE.
19. A. Coates et al., "Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning," 2011 International Conference on Document Analysis and Recognition, Beijing, 2011, pp. 440445.
20. B. Epshtein, E. Ofek and Y. Wexler, "Detecting text in natural scenes with stroke width transform," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 2963-2970.
21. Shao Y., Wang C., Xiao B., Zhang Y., Zhang L., Ma L. (2010) Text Detection in Natural Images Based on Character Classification. In: Qiu G., Lam K.M., Kiya H., Xue XY., Kuo CC.J., Lew M.S. (Eds) Advances in Multimedia Information Processing - PCM 2010. PCM 2010. Lecture Notes in Computer Science, vol 6298. Springer, Berlin, Heidelberg.
22. Park, J., & Lee, G. (2008). A robust algorithm for text region detection in natural scene images. Canadian Journal of Electrical and Computer Engineering, 33(3/4), 215-222.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication