MJCIS Vol .13 No.1 Dec2017

Mansoura Journal of Computers and Information Sciences

Enhancement Of Text Recognition In Scene Images

Moayed Hamad
Faculty of computers and information systems , C.S dep.
Mansoura University, Egypt
Moayed_hamad@ yahoo.com

Osama Abu-Elnasr
Faculty of computers and information systems , C.S dep.
Mansoura University, Egypt
mr_abuelnasr@ yahoo.com

Sherif Barakat
Faculty of computers and information systems , I.S dep.
Mansoura University, Egypt
sherifiib@yahoo.com

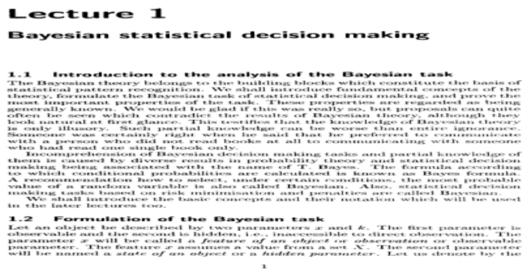
ABSTRACT
Text detection and recognition in natural scene images has received significant attention in last years. However, it is still an unsolved problem, due to some difficulties such as some images may have complex background, low contrast, noise, and /or various orientation styles. Also, the texts in those images can be of different font types and sizes. These difficulties make the automatic text extraction and recognizing it very difficult. This paper proposes the implementation of an intelligent system for automatic detection of text from images and explains the system which extracts and recognizes text in natural scene images by using some text detection algorithms to enhance text recognition. The proposed system implements various algorithms, such as Maximally Stable Extremal Regions (MSER) algorithm to detect the regions in the image, Canny edges algorithm to enhance edge detection and Bounding Box algorithm to detect and segment area of interest. Once the text is extracted from the image, the recognition process is done using Optical Character Recognition (OCR). The proposed system has been evaluated using public datasets (ICDAR2003 and the experimental results have proved the robust performance of the proposed system.

General Terms
Text Detection, Text Recognition.

Keywords
Scene Text Detection, Maximally Stable Extremal Regions, Bounding Box, OCR, OCR Spell Checker.

1. INTRODUCTION
In last years, detecting and recognizing texts from natural scene images has become one of the important research topics in the fields of machine learning, computer vision, and pattern recognition. Although the advancements in machine learning and computer vision have improved the accuracy of text detection and recognition, text detection and recognition in scene image and reading posters is still an open problem [1].

Generally, the images can be classified into two types of images: (a) Document image or (b) Scene image, as shown below in Figure 1. Document images may be looks like scanned book covers, while text in natural images are called scene text and caption text [2].



(a)

(b)

Fig. 1: The difference between printed document recognition (OCR) and scene text localization and recognition. (a) a scanned book page. (b) a sample from the ICDAR 2003 Robust Reading dataset [3].

Text localization and recognition in natural scenes image usually consists of a set of stages including text detection, text segmentation and OCR based text recognition for printed documents [4]. As everyone knows, detecting and recognizing text in natural scenes has proved to be more challenging than its counterpart in documents. Some text may not be determined or recognized correctly because of some difficulties [5]. These difficulties are due to representing the text in the natural scenes images using different font styles, different font sizes and different colors and languages. Also, the natural scene images may have complicated backgrounds and uneven lighting which make the text detection and recognition in natural scenes image more difficult [6].

Optical character recognition (OCR) is an important technology which is used for recognizing the text in the image. Using OCR, we can convert different kinds of documents, like PDF files, scanned documents, or images captured by camera into editable texts [7].

This paper proposes the implementation of an intelligent system for automatic detection of text from images and explains the details of a system that is able to extract and recognize text from a natural scenes image using some text detection algorithms for enhancing text recognition. This paper implements various algorithms including Maximally Stable Extremal Regions (MSER) algorithm to detect the regions in image, Canny edges to enhance edge detection and Bounding Box algorithm to detect and segment area of interest. Once the text is extracted from the image, the recognition process is done using Optical Character Recognition (OCR). In spite of fast processing of OCR, the output usually contains many errors. These introduced errors negatively affect the linguistic accuracy and retrieval accuracy of the OCR documents. In order to deal with this problem, many tools for spell checking and correction have been developed as support tools for OCR. The role of the spelling correction system is to determine the candidate words based on the information gathered from multiple resources and to automatically replace the incorrect word with the correct one.

This paper is organized as follows: Section 2 provides a brief review on the related works of text detection and recognition algorithms. Section 3 introduces a description of the proposed text detection and recognition system. Section 4 evaluates the proposed system and discusses the experimental results. Finally, section 5 presents the conclusion and future work.

2. RELATED WORKS

In the recent years, many researchers have suggested many approaches for determining the texts in the natural scene images and recognizing it. Various related works in text detection and text recognition are introduced below.

2.1 Text detection

Text detection methods could be divided into two main types:

A- Texture based methods.

B- Region based methods.

Texture based methods consider text as a special texture. It usually uses Fast Fourier transform (FFT), Discrete Cosine Transform (DCT), wavelet and Gabor filters, to extract texture features. Then, it uses a sliding window to search for possible text blocks throughout the whole image, and a classifier is finally used to verify it as a text region or not [4].

Wang et al [8] proposed a method for highlighting words detection. The research revolves around the difficulties found in the text. It also highlights upon Optical Character Recognition (OCR) as an application of scanned – pages printed texts which are considered as successful step.

Palaiahnakote et al [9] proposed a method for video text detection based on the laplacian frequency. The proposed method is able to treat with texts of arbitrary orientations. The process of analysis is done through some steps. Firstly, the input image is to be filtered with Fourier – Laplacian. Then, K– means clustering is used to indicate text region concerned with maximum differences. At last, there is a mixture of text straightness and edge density as an evidence of false positive elimination. The study conclude that laplacian method has the ability to deal with both graphics and scene text horizontally and non – horizontally.

Lee et al [10] proposed a method for detecting text regions. The process is done through choosing a novel text which extracts about six various classes features of texts. The text uses the technique of Modest AdaBoost with a search composed of multi – scale sequence. The study concluded that the algorithms can detect text regions including various fonts with different sizes, colors, and scripts.

On other side, Region-based methods is one of the first methods used to extract connect components, which usually use Maximally Stable Extremal Regions, color, edge, stroke width, and uses geometric constraint to exclude non-text region [4].

Epshtein et al [11] presented a new method through which the value of stroke width is found separately for each image pixel. The process is called stroke with transformation (SWT). The method is done through dependent data which makes the method very fast and sufficient to stamp out the need of scanning windows. The study concluded that extensive testing show a simplicity allowing the algorithms to detect texts through various fonts and language.

Shivakumara et al [12] proposed new Fourier – statistical features (FSF) for detecting text in video frames that have different fonts, scripts, and sizes. The process consists mainly of two parts. The first part is the text frame which represents novel features based upon three axes; sharpness, straightness of edge, and proximity of the edge with the aim of identifying true text frame. The second part is concerned with text blocks through determination of projection profiles. Both parts are tested by some experiments on a variety of frames. The study concludes that the proposed method has the superiority to any existing approaches according to detection rate, false positive rate, and midsection.

Neumann et al [13] proposed a method of tri-dimensional stage. The first stage displays a strict feedforward pipelines replaced by a verification framework. The second stage uses a synthetic font with the aim of training the algorithms that gets rid of the need of time consuming. The third stage exploits Maximally Stable Extremely Regions (MSERs) which generate geometric conditions. Then an evaluation is introduced on two standard datasets. The study concludes that there is an adaptation of the method in use through the recognition of other scripts.

2.2. Scene Text Recognition

Natural scenes image have different properties from document image. Applying traditional text recognition method for recognizing scene text would result in more difficulties. Various related work of text recognition is introduced below.

Mona Saudagar et al [14] stated that the extraction of a text in an image is a great obstacle in computer vision. The researchers admit the popularity of computer manual use as a main problem of the reassert. The text data displayed in images contain useful information for automatic annotation. The main problem lies in the natural scene images containing complex background which makes it difficult to extract information. The study concluded that the probability of solving errors, unsolved problems may have improvements to get over those issues.

Yao et al [15] stated that the main research problem is concerned with localizing and reading text which is the akin challenging of the research due to many factors. The research displayed a set of detectable primitives, strokelets (essential substructures character and some advantages of strokelets). The reprocess of strokelets has four various advantages namely usability, robustness, generality, and expressivity. The study concluded that the effectiveness of strokelets on proposed algorithms has some advantages.

Neumann et al [16] presented a free end-to-end text localization and recognition method. The method combined the merits of sliding-window and connected methods for detection and recognition of characters. The process of detection and recognition is done through orientations of specific strokes via a system of image gradient field with a set of oriented bar filters. The introduction of a novel character is done through values detection. The process has a great effectiveness through classification of real word characters. The study concluded that characters are selected from obtained set of target regions.

Bissacco et al [17] proposed an approach for text detection and recognition from smart phone image. The proposed approach has the ability to recognize text in a variety of challenging images that have some conditions such as substantial blur, low resolution, low contrast, and high image. The study concluded that there is a great importance of training even in designs where the classifier is bounded by some constraints.

3. PROPOSED METHOD

This paper proposes the implementation of an intelligent system for automatic detection of text from images and explains the system that extracts and recognizes text in natural scene images by using some text detection algorithms to enhancing text recognition. This paper implements various algorithms; Maximally Stable Extremal Regions (MSER) algorithm to detect the regions in image. Canny edges to enhance edge detection. Bounding Box algorithm to detect and segment area of interest. Once the text is extracted from the image, the recognition process is done using Optical Character Recognition (OCR). Then, a spell checker tool is applied on the output of the OCR. The spelling correction system is specially designed for OCR. Figure 2 shows the proposed scene text detection and recognition system.

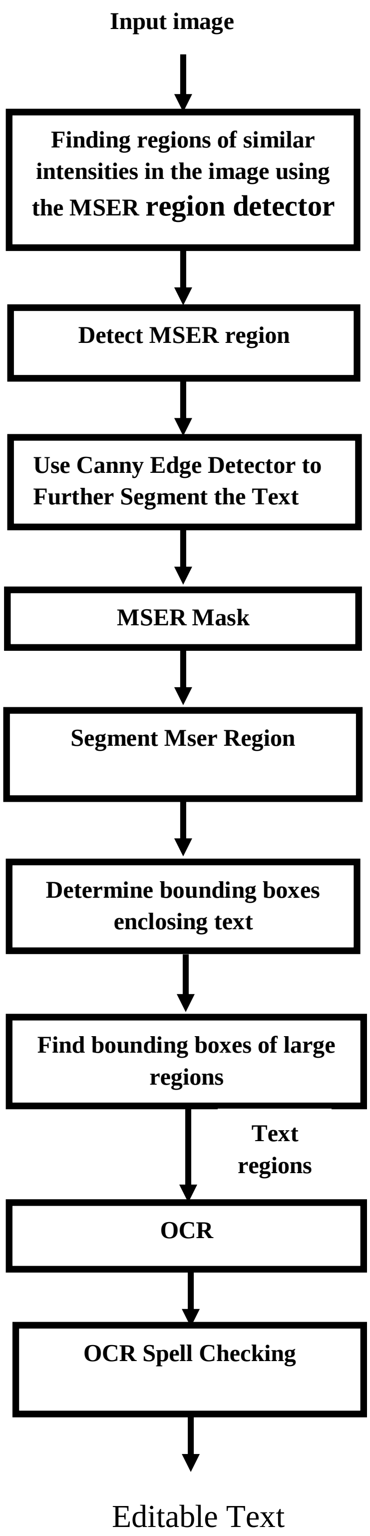


Fig 2: An overview of the proposed system

3.1 Maximally Stable Extremal Regions (MSER)

MSER is an algorithm for detecting the text in an image. It has been selected as one of the best region detectors, due to its high capability for detecting low quality texts [18]. Normally, text characters usually have consistent color. Therefore, in order to find the text, we start, with the analysis of the entire image to determine the regions of similar intensities to find a set of distinguished regions in the image, which differ in brightness and color compared with other regions in the image. However, natural image contains many non-text regions. So, geometric filtering, masking and preprocessing are used to determine the text part from the image as shown below in Figure 3.

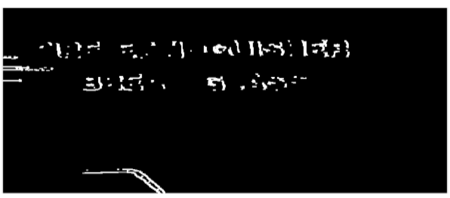
(a) Original image (b) detected region MSER

Fig 3: The performance of the MSER region detector

3.2 Canny Edge Detector

Canny edge operator is a robust edge detector. It is able to localize the correct edges and to ignore the false ones. The process of edge detection using canny algorithm occurs at four phases: (a) A Gaussian blur is applied for noise removal, (b) A gradient operator is used for obtaining the gradients' intensity and direction, (c) Non-maximum suppression that determines if the pixel is a better candidate for an edge than its neighbor's, and (d) Hysteresis thresholding that finds where edges begin and end. In the previous stage when applying the MSER, the small letters can't be detected because of blurring, noise and/or low resolution. In order to overcome these problems, we suggested combining the properties of canny edges and MSER [19]. The combination of MSER and canny edge detection algorithm produces the region that is likely to be a text as shown in Figure 4.

(a) Mser mask



(b) Edge gradient mask

(c) **Segment Mser Region**

Fig 4: Canny edge with MSER region detector

3.3 Determine Bounding Boxes

Bounding Box algorithm is used in the proposed system to detect and segment area of interest. This algorithm extract words and text lines by the rectangular box. Bounding boxes, as simply rectangles, are enclosing the group of connected components. Each and every bounding box has its unique width and height. To generate the bounding box, there are some steps:

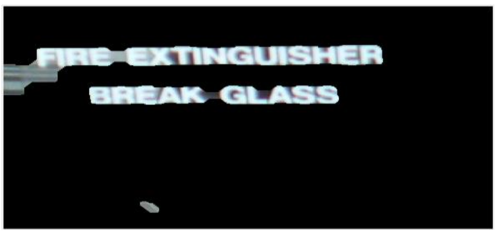
- Find starting and ending points of a connected component
- Height of bounding box={y coordinate of last pixel of connected component-y coordinate of first pixel of connected component}
-Width of bounding box {x coordinate of last pixel of connected component-x coordinate of first pixel of connected component} [20].

In this method, we will merge the individual characters into a single connected component, to compute a bounding box of the text region. This can be accomplished using morphological closing, followed by morphological opening to clean up any outliers and to connect nearby large regions as shown below in Figure 5.

Closing **Opening**

Fig 5: Morphology Mask

As a final step, text lines are split into individual words by classifying the inter letter distances into two classes: character spacing and word spacing as shown below in Figure 6.



(a) Image region under mask created by joining individual Characters

(b) Text region

Fig 6: Bounding Box generation, examples from ICDAR2003 dataset

3.4 OCR and OCR Spell Checking Method

Optical Character Recognition (OCR) is an important technology or software used to recognize the text in image and video and have the ability to convert different types of documents, such as PDF files or scanned paper documents, or images captured by camera into editable data [21]. A lot of errors may occur during performing OCR. Therefore, correction tools are needed to fix these errors. In the proposed system, an OCR spell checking method is used to correct the spelling mistakes that might appear after OCR is performed. It helps in attaining a high percentage of accuracy in recognition. OCR spell checking is done to recognize the unrecognized characters. The OCR spell checking software tries to match the words with unrecognized letters, with similarly spelled words in its dictionary. For example, assume that the character recognition software has recognized "tne". The OCR spell checking software will find and replace the incorrect letter "n" with the right one `h', after consulting its dictionary. In this way, the errors caused by the OCR programs can be minimized [22] as shown below in Figure 7.

(a) Input image

FIRE EXTINGUISHERBREAK GLASS

(b) Result of applying OCR

Fire Extinguisher Break Glass

Suggestions OCR:

1: break 3: fire

2: extinguisher 4: glass

(c) OCR after correction

Fig 7: Text recognition step using OCR and OCR Spell checker

4. EXPERIMENTAL RESULTS

In this section, we evaluated the performance of the proposed system using a public dataset (ICDAR2003) [3].The proposed system has been tested and its performance was compared other related system. The system was implemented using MATLAB 2015 platform, and a computer of Windows7 operating system and Intel core i5-2400@1.60GHz CPU. The experimental results are shown in the following subsections.

4.1 Word recognition datasets.

In our experiments, we have used the public datasets ICDAR2003 [3], a robust word recognition datasets, to evaluate the performance of the proposed approach. The dataset contains images taken from Google View Street, covers book, etc. Figure 8 shows a good set of sample images contained in this dataset while Figure 9 shows another set of sample images that suffer from various problems.

Fig 8: Image examples from ICDAR2003 dataset [3]

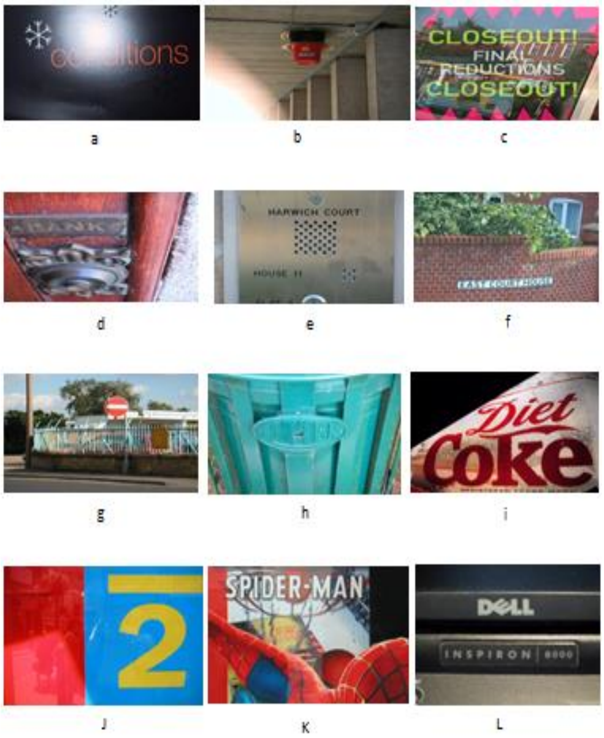


Fig 9: Image examples of failure cases in proposed system

The images shown in Figure 9 are suffering from different problems including strong high lights (a), text that is too small (b), transparent text (c), the text strokes are too subtle and do not have strong edges and low contrast (d), the text region is partially occluded by other structures (e), and text like outliers (f), and low quality texts (g), The text components in the image do not include strong text information and are easily confused with its background (h), Multiple characters joined together (i), Not enough letters to form a text line (J), and letters are connected to the surrounding area which has the same color (k), and low resolution (L).

4.2 Evaluation parameters

The proposed system has been evaluated using precision, recall and f-measure performance metrics. The Precision rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false positives. False positives are those regions in the image which are actually not characters of a text, but have been detected by the algorithm as text regions [23].

$$\textbf{Precision rate} = \frac{\textbf{correctly detected word}}{\textbf{correctly detected word+false positive}} * \mathbf{100}\ (\mathbf{1})$$

The Recall rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false negatives. False Negatives are those regions in the image which are actually text characters, but have not been detected by the algorithm.

$$\textbf{Recall rate} = \frac{\textbf{correctly detected words}}{\textbf{correctly detected words+false nagative}} * \mathbf{100}\ \ (\mathbf{2})$$

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score [23].

$$\mathbf{f} = \mathbf{2}.\frac{\textbf{precision .recall}}{\textbf{precision+ recall}} \qquad \mathbf{(3)}$$

4.3 Experimental Results and Discussions

Experiments on ICDAR 2003 dataset: Two different experiments have been conducted to evaluate the performance of the proposed system. The first one used OCR While second one used OCR with spelling correction tool (correct OCR). The OCR software used to recognize the text in image. OCR Spell Checking is used to correct the spelling mistakes that might appear after OCR is performed. It helps in attaining a high percentage of accuracy in recognition. The results are shown in Table 1. From table 1, we conclude that the OCR with spell checker gets better result than the OCR. In ICDAR 2003 dataset, the proposed method achieved a precision, recall and F-measure of 0.84, 0.85 and 0.84, respectively. The three values are higher than the closest reported results by large margins (0.1, 0.8 and 0.4), which demonstrates a significant improvement over existing methods. After carefully examining the results on each test images, we conclude that our significant performance improvement is mainly due to two factors: (1) the excellent performance of the proposed MSER and canny edge, which reliably detects letter candidates in most examples, leading to high recall, and (2) the effectiveness of the bounding box algorithm, OCR and OCR spell chick which lead to high precision. As we can see from the experimental results, the proposed method is robust to blur (Figure 8 (a), (d) and (j)), low resolution (Figure 8 (c)), low contrast (Figure 8 (b), (e) and (L)), strong high lights (Figure 8 (h) and (i)) and small letters and different fonts (Figure 8 (g) and (k)).

Table 1: The results of proposed system on some images ICDAR2003 dataset

		OCR	**Correct OCR**
	Precision	1	1
	Recall	0.85714	1
	f-measure	0.9230	1
		OCR	**Correct OCR**
	Precision	1	1
	Recall	0.7	0.777
	f-measure	0.8235	0.875

<table>
<tr><th></th><th></th><th>OCR</th><th>Correct OCR</th></tr>
<tr><td rowspan="3"></td><td>Precision</td><td>0.7777</td><td>1</td></tr>
<tr><td>Recall</td><td>0.4375</td><td>0.8181</td></tr>
<tr><td>f-measure</td><td>0.56</td><td>0.9</td></tr>
<tr><td></td><td></td><td>OCR</td><td>Correct OCR</td></tr>
<tr><td rowspan="3"></td><td>Precision</td><td>0.6</td><td>0.8</td></tr>
<tr><td>Recall</td><td>0.6</td><td>1</td></tr>
<tr><td>f-measure</td><td>0.6</td><td>0.88889</td></tr>
</table>

Finally, Table 2 shows a comparison between the proposed system and the other methods. From the results presented in Table 2, we can conclude that the proposed system has a superior performance regarding the different performance metrics.

Table 2: Text detection and recognition results on the ICDAR2003 dataset, compared with other methods

Algorithms	precision	recall	f_measure
Epshtein [10]	0.73	0.60	0.66
A Laplacian [11]	0.76	0.86	0.81
Neumann [12]	0.59	0.55	0.57
Luka´s, Jiri Mata's [25]	0.65	0.64	0.63
SFT-TCD[24]	0.81	0.74	0.72
Lee, Jung [9]	0.66	0.75	0.70
Cunzhao Shi [27]	0.833	0.631	0.7128
Neumann [28]	0.82	0.71	0.76
Yuanwang [29]	0.835	0.772	0.802
Proposed system	0.84	0.85	0.84

5. CONCLUSION

The detection and recognition of text in natural scene images are still unsolved problems, because this type of images may suffer from some problems, such as low contrast and complex backgrounds. Also, the text presented in these images may have various orientation, different font sizes, different font types and different font styles. This paper proposed an enhanced Maximally Stable Extremal Regions (MSER) based scene text detection method. It is able to determine the text regions in the natural scenes images and can recognize the text from the selected text regions. When applying the MSER, the small letters can't be detected because of blur or noise and low resolution. To overcome these problems, we suggested combining the properties of canny edges and MSER. Lastly, optical character recognition (OCR) is applied on the selected text regions to recognize the text. The proposed system showed improvements in performance in terms of the precision, recall and f-measure when compared to previous methods. Although the proposed system can remove most of the non-text regions and can successfully determine most of the text regions, some text regions are mistakenly removed. In the future, we aim to overcome these problems and to enhance the performance of the text recognition methods in natural scene images.

6. REFERENCES

[1] Huang, W., Qiao, Y., & Tang, X. 2014. Robust scene text detection with convolution neural network induced mser trees. In European Conference on Computer Vision (pp. 497-511). Springer International Publishing .

[2] Niti Syal and Naresh Kumar Garg. 2014. A Study of Text Localization Algorithms For Complex Images. International Journal of Innovative Research in Computer and Communication Engineering, pp. 35- 42.

[3] ICDAR 2003 Robust word recognition dataset, http://algoval.essex.ac.uk/icdar/RobustWord.html, [Accessed at 12\1\2016].

[4] Liu, X., Lu, Z., Li, J., & Jiang, W. 2014. Detection and Segmentation Text from Natural Scene Images Based on Graph Model. WSEAS Transactions On Signal Processing, 10.

[5] Gaikwad, B. P., Manza, R. R., & Manza, G. R. 2013, Video scene segmentation to separate script. In Advance Computing Conference (IACC), 2013 IEEE 3rd International, pp. 1269-1274.

[6] Elagouni, K., Garcia, C., & Sébillot, P. 2011. A comprehensive neural-based approach for text recognition in videos using natural language processing. In Proceedings of the 1st ACM International Conference on Multimedia Retrieval (p. 23).

[7] Sitafale, R. G., & Kulkarni, P. P. 2015. Character Recognition in Natural Scene Images using Novel Technology.

[8] Wang, K., & Belongie, S. 2010. Word spotting in the wild. In European Conference on Computer Vision , pp. 591-604. Springer Berlin Heidelberg.

[9] Shivakumara, P., Phan, T. Q., & Tan, C. L. 2010. New Fourier-statistical features in RGB space for video text detection. IEEE Transactions on Circuits and Systems for Video Technology, 20(11), pp. 1520-1532.

[10] Lee, J. J., Lee, P. H., Lee, S. W., Yuille, A., & Koch, C. (2011, September). Adaboost for text detection in natural scene. In Document Analysis and Recognition (ICDAR), 2011 International Conference on (pp. 429-434). IEEE.

[11] Epshtein, B., Ofek, E., & Wexler, Y. (2010, June). Detecting text in natural scenes with stroke width transform. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (pp. 2963-2970). IEEE.

[12] Shivakumara, P., Phan, T. Q., & Tan, C. L. 2011. A laplacian approach to multi-oriented text detection in video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(2), 412-419.

[13] Neumann, L., & Matas, J. 2011. A method for text localization and recognition in real-world images. Computer Vision–ACCV, pp. 770-783.

[14] Saudagar, M., & Jain, S. V.2015. Text Localization and Recognition in Natural Scene Images , pp. 54-58.

[15] Yao, C., Bai, X., Shi, B., & Liu, W. 2014. Strokelets: A learned multi-scale representation for scene text recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4042-4049.

[16] Neumann, L., & Matas , J. 2013. Scene text localization and recognition with oriented stroke detection. In Proceedings of the IEEE International Conference on Computer Vision , pp. 97-104.

[17] Bissacco, A., Cummins, M., Netzer, Y., & Neven, H. 2013. Photoocr: Reading text in uncontrolled conditions. In Proceedings of the IEEE International Conference on Computer Vision, pp. 785-792.

[18] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., & Van Gool, L. 2005. A comparison of affine region detectors. International journal of computer vision, 65(1-2),pp. 43-72.

[19] Chen, H., Tsai, S. S., Schroth, G., Chen, D. M., Grzeszczuk, R., & Girod, B. 2011. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In Image Processing (ICIP), 2011 18th IEEE International Conference on , pp. 2609-2612.

[20] Saluja, S., Patnaik, T., & Jain, T.2012. Text Extraction and Non-text Removal from Colored Images. *International Journal of Computer Applications*, 0975-8887.

[21] Mohapatra, Y., Mishra, A. K., & Mishra, A. K. 2013. Spell checker for ocr . International Journal of Computer Science and Information Technologies, 4(1),pp. 91-97.

[22] Sharma, S. (2007). Extraction of text regions in natural images, pp.5-8,

[23] https://en.wikipedia.org/wiki/Precision_and_recall, [Accessed at 10\8\2016].

[24] Huang, W., Lin, Z., Yang, J., & Wang, J. 2013. Text localization in natural images using stroke feature transform and text covariance descriptors. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1241-1248.

[25] Neumann, L., & Matas, J. 2011. Text localization in real-world images using efficiently pruned exhaustive search. In Document Analysis and Recognition (ICDAR), 2011 International Conference on , pp. 687-691.

[26] Yao, C., Bai, X., Liu, W., Ma, Y., & Tu, Z. 2012. Detecting texts of arbitrary orientations in natural images. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on , pp. 1083-1090.

[27] Shi, C., Wang, C., Xiao, B., Zhang, Y., & Gao, S. 2013. Scene text detection using graph model built upon maximally stable extremal regions. Pattern recognition letters, 34(2),pp. 107-116.

[28] Neumann, L., & Matas, J. 2016. Real-time lexicon-free scene text localization and recognition. IEEE transactions on pattern analysis and machine intelligence, 38(9),pp. 1872-1885.

[29] Wei, Y., Zhang, Z., Shen, W., Zeng, D., Fang, M., & Zhou, S. 2017. Text detection in scene images based on exhaustive segmentation. Signal Processing: Image Communication, 50,pp. 1-8.