International Journal of Engineering Research and General Science Volume 4, Issue 2, March-April, 2016
ISSN 2091-2730

Comparative study of Image processing techniques used for Scene text detection and extraction

Akhilesh Panchal, Shrugal Varde, Dr.Prof.M.S.Panse

V.J.T.I. Mumbai, akhileshp31@gmail.com

Abstract— In recent years, wide variety of research has been done on Text detection and Extraction from Scene images. These techniques are used for large number of applications like aid for visually impaired people, Document analysis, Vehicle license plate recognition, etc. Text Extraction plays a major role in finding vital and valuable information from captured image. With rapid development in Multimedia Technology and growing requirement for information, identification, indexing and retrieval, several image processing techniques have been developed for extracting text. Each technique has its pros and cons depending on various conditions like Speed, Accuracy, Complexity, Processing time, etc. Hence, only single method is insufficient for overall text detection and extraction system. To achieve better performance, it is necessary to combine these techniques. So, we need to have adequate knowledge of various techniques proposed worldwide. On this background, this article discusses various schemes proposed earlier for extracting the text from an image. This paper also provides the performance comparison of several existing methods proposed by researchers in extracting the text from an image.

Keywords— Text detection, Image Enhancement, Image Preprocessing, Localization, Text extraction, Text Recognition.

I. INTRODUCTION

Text data is particularly interesting, because text describes the contents of an image. Text embedded in images is mainly classified as Caption/Artificial text and Scene/Natural text [9]. Caption text is laid over the image during editing e.g. score of match whereas Scene text is actual part of the scene e.g. street signs, name plates. The problem of Text detection in printed document has been focused for many years and has already reached high recognition rates made it the most successful applications of Compute vision and Machine learning techniques. However, characters recognition from scene images is still a challenging task due to complex background, non-uniform lighting condition, font size, styles, perspective distortion multilingual environment or blurring effects of natural images and active subject for many researchers nowadays [6] [9]. Hence, this paper focuses on extraction of text from Scene image. In order to overcome these problems in scene images, many preprocessing, image enhancement and extraction techniques are proposed and they are used in particular conditions. So, it is essential to study these techniques for employing simple, robust, high performance and cost effective system for Scene text recognition. To achieve this goal, current paper surveys most of the image processing techniques used for text detection and extraction in Scene images. The purpose of the survey is to compare text extraction techniques for selecting proper technique according to applications and conditions.

II. BACKGROUND

Typically, Text extraction consists of various steps like Preprocessing, Text detection, Localization, Binarization and Thresholding, Extraction, Enhancement and Recognition. Order may vary according to application and convenience. The methods cited in this paper are based on morphological operators, wavelet transform, Feature Learning algorithm, artificial neural network, edge detection algorithm, histogram technique etc.

Figure 1: Examples of Scene images

Earlier methods consider only 2-D image or B&W image but nowadays 3-D or Color images are also taken into consideration. They used mainly the image datasets such as ICDAR competitions and Chars 74k for experimentation which is shown in figure 1.Software used for simulation in most of the researches is MATLAB as it is simple to use and easily available image processing tool. It has various inbuilt commands for image processing. Also, Mathscript built on MATLAB can be used on different platforms. Lots of research work has been done to improve accuracy and performance of text extracting techniques. Recently, researchers have explored approaches that prove effective for text captured in various configurations, in particular, incidental text in complex backgrounds. Such approaches typically stem from advanced machine learning and

unsupervised feature learning ,convolutional neural networks (CNN), deformable part-based models (DPMs) , belief propagation and conditional random fields (CRF) [10] [14].

III. LITERATURE SURVEY

For convenience, we break the system into three stages: 1. Pre-processing stage 2. Processing stage 3. Post processing stage. Pre-processing stage use some enhancement algorithms to eliminate challenges created by noise, blurring effect and uneven lighting whereas Processing stage includes Text Detection, Extraction, Segmentation and Localization which uses sophisticated methods. Third stage is Text recognition stage which is applied after processing stage.

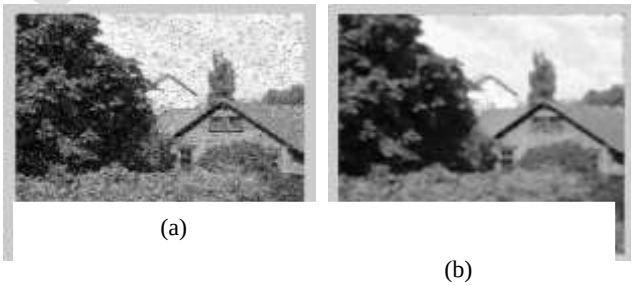
A) IMAGE ENHANCEMENT / PREPROCESSING

Before proceeding to text detection and extraction methods used, we have to first consider Scene image can be mixed with noise like Salt and pepper noise, Impulse noise etc. or it can be blurred due motion of camera. For that purpose, we should use some image preprocessing/enhancement Techniques. De-blurring techniques like Lucy Richardson algorithm, Blind de-convolution algorithm, Wiener de blurring techniques are generally used [17]. Out of them, Wiener filter is selected which is a natural extension of the inverse filter when noises are present. Figure 2 illustrates how de-blurring is achieved using Wiener filter on MATLAB. From figure, it is observed that binarization after wiener filtering produces better result which will be effective for further processing.

(a) (b) (c)

Figure 2: De-blurring of an image using Wiener filter. (a) Blurred image; (b) Binarized image without filtering; (c) Binarization after De-blurring

Salt and pepper noise is one type of impulse noise which can corrupt the image, where the noisy pixels can take only the maximum and minimum values in the dynamic range i.e. black dot on white background (pepper) and white dot on black background (salt) which degrades the text extraction performance of system [19]. Since, linear filtering techniques are not effective, standard median filter (SMF), which is a non-linear filter used to remove such noise due to its good de-noising power and computational efficiency. However, when noise level is more than 50%, edge details of the original image will not be preserved by the median filter
as shown in Figure 3. So, It is recommended that during the filtering (restoration) process the edge details have to be preserved without losing the high frequency components of the image edges.



(a) (b)

Figure 3: Salt and pepper noise removal using Median

Sometimes, image is captured in dark or uneven lighting for which Text extraction becomes difficult. So, application of contrast enhancement is necessary. Histogram Equalization method is mostly used for Contrast enhancement. Figure 4 shows how contrast enhancement done using Histogram Equalization. Hence, this leads to overcome Uneven lighting, Blurring and noise degradation problems which would adversely affect system performance.

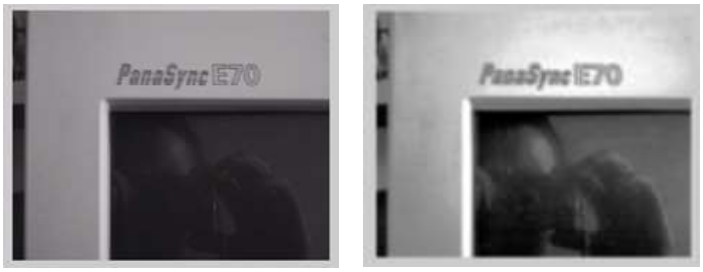


Figure 4: Contrast Enhancement using Histogram equalization.

B) PROCESSING STAGE:

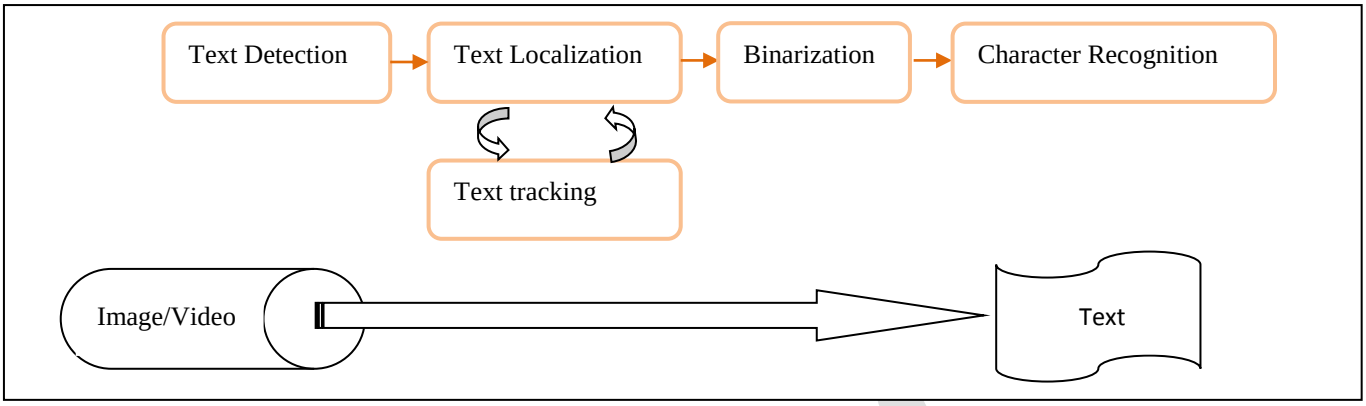


Figure 5: Architecture of Text Detection & recognition System

Text Detection phase takes enhanced image or video frame as input and decides it contains text or not. It also identifies the text regions in an image whereas *Text Localization* merges the text regions to formulate the text objects and define the tight bounds around the text objects. Figure 5 shows Architecture of Processing stage. Text detection, localization and tracking modules are closely related to each other and it is the most challenging and difficult part of extraction process as it feeds to character recognition system [11].

Text Tracking: This phase is applied to video data only. For the readability purpose, text embedded in the video appears in more than thirty consecutive frames. This phase exploits temporal occurrences of the same text object in multiple consecutive frames. It can be used to rectify the results of text detection and localization stage. It is also used to speed up the text extraction process by not applying the binarization and recognition step to every detected object [11].

Text Binarization: This step is a part of image segmentation, used to segment the text object from the background in the bounded text objects. The output of text binarization is the binary image, where text pixels and background pixels appear in two different binary levels like white text on dark background or vice versa. Many times Binarization can be applied before localization step.

For Text Detection, connected component analysis (CCA) and sliding window classification are two widely used methods, and color, edges, strokes, and texture are typically used as features [10]. CCA which is a graph algorithm, where subsets of connected components are uniquely labeled based on heuristics about features, i.e. color similarity and spatial layout. The use of statistical models in CCA significantly improves its adaptivity. In the sliding window classification method, multi-scale image windows that are classified into positives are further grouped into text regions with morphological operations, CRF [13] or graph methods. For text localization, color, edge and texture features were conventionally used, and stroke, point, region and character appearance features have recently been explored [10]. M. Swamy Das et.al [8] provides detail analysis of detection techniques such as Connected component based, edge based and Texture based method. From this article, it is observed that Texture based method is more efficient compared to that of the performance obtained with edge based method and connected component based method. But for better performance it is always advisable to combine this techniques.

C) POST PROCESSING STAGE:

Character Recognition: The last stage is the character recognition. This module converts the binary text object into the ASCII text. There are various sophisticated tools already developed which are used for recognition like OCR, Snooper text [22] etc. Figure 6 shows how 'hello' word wrapped in image gets recognized through Text Extraction process.

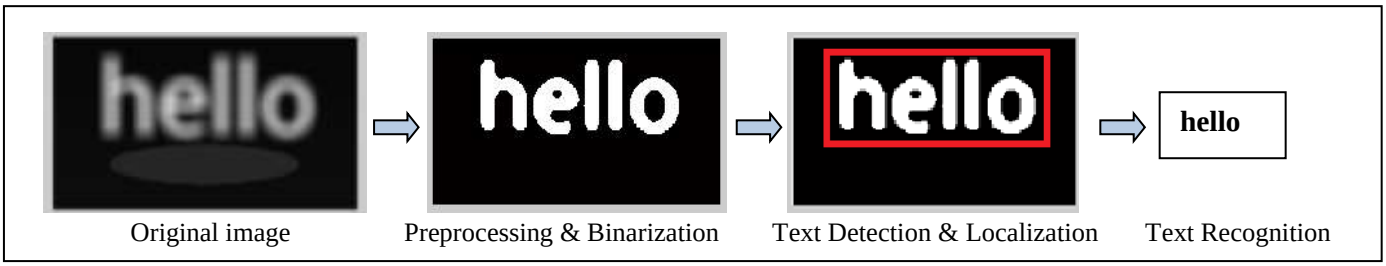


Figure 6: Text Extraction process on image

IV. ANALYSIS

The performance of each algorithm mostly evaluated based on parameters like precision rate, recall rate, average run time etc. The Precision and recall rates are calculated as

$$\text{Precision} = \frac{\text{Correctly detected}}{\text{Correctly detected} + \text{False positives}} X\ 100\% \qquad \text{Recall} = \frac{\text{Correctly detected}}{\text{Correctly detected} + \text{False negatives}} X\ 100\%$$

Where, False positives are the non-text regions in the image and have been detected by the algorithm as text regions and False negatives are the text regions in the image and have not been detected by the algorithm [8]. Both precision and recall rates are useful to determine the accuracy of each algorithm in eliminating the non-text regions and locating the correct text regions. Higher accuracy and Less run time is preferred for any application which requires text extraction from an image. Performance analysis based on such parameters on some of surveyed papers is given in Table 1 as shown below.

Table 1: Performance Analysis of Text Extraction methods

SR. NO.	AUTHOR	YEAR	METHOD/S USED	ACCURACY	ADVANTAGES	DISADVANTAGES
1.	Wahyono, et.al [27]	2015	Canny edge detector, Fast Stroke Width Transform (FSWT)	61% (Precision), 63% (Recall)	Fast (0.18 sec.) So, it can be used in real time. Also used for Multi Language text detection.	Complex in Design
2.	Hrishav raj, et.al [7]	2014	Binarization, Connected Components (CC), Morphological operations, Canny Edge detection	72.8% (Precision), 74.2% (Recall)	Independent of Font Size, Style and directions.	Trained only for extracting Devanagari Text from image
3.	C.P. Sumathi et.al [4]	2013	Wavelet transformation, Morphological operation, Feature extraction, Neural Network classifier	87.0%	Low fragmentation, low error rate, Tolerance to noise. Works on Video frames	Slow and Complex to design.
4.	Ho Vu, et.al [1]	2012	Feature learning method with Orthogonal matching pursuit for training & sparse coding as a mapping-function.	83.8%	Less affected by the categorization of images	Takes long time while extracting feature vectors.
5.	Huizhong Chen et.al [10]	2011	CC based Edge-enhanced Maximally Stable Extremal Regions (MSER), Stroke width Transform (SWT)	73% (Precision), 60% (Recall)	Simple & efficient, can be combined with visual search systems without further computational load	Detection fails due to excessive blur and out of focus as no preprocessing
6.	Andrej Ikica, et.al[20]	2011	Edge profile based detection with Canny edge map, Heuristic rules	70.9% (Precision), 55.2% (Recall)	Simple, fast and efficient	Sometimes Non Text areas get detected leads to low accuracy.
7.	Huang et.al [25]	2010	Stroke Map, Connected component analysis, Harris Corner Detection	90.2%	Robust to detect and locate video scene text with variation of text size, Good speed	Not suitable in low contrast background
8.	Pan et. al [13]	2009	Combination of CC & region based approach includes Conditional Random Field(CRF) model, Minimum classification error (MCE) learning, Graph cuts inference, Minimum spanning tree	67% (Precision), 71% (Recall)	Robust and accurately Localize texts	Takes More time and Complex

9.	Nobuo Ezaki et.al [18]	2004	Sobel edge detection, Otsu binarization, connected-component extraction, rule-based connected-component selection	48% (Precision), 76% (Recall)	Easy to design, Combination of these methods gives good overall performance	Low detection accuracy for small text in images
10.	Gllavata et.al [5]	2003	Color reduction technique, Edge detection, and localization of text regions using projection profile and geometrical properties	83.9% (Precision), 88.7% (Recall)	Works well in Grayscale as well as Color image.	Low quality images makes detection complex.

V. CONCLUSION

This paper covers detail analysis of the text detection, localization and tracking techniques. After comparison study of recent researches on Text extraction in scene images, it is observed that each proposed method has its own advantages depending on various conditions which are mentioned before. Some papers have modified the techniques while some invent new techniques. It is necessary to first preprocess the image before applying Text Extraction algorithms cause it can produce false detection and hence less accuracy. Accuracy and speed are important factors while considering performance and there is trade-off between two factors. So, keeping this in mind proper technique should be selected.

Recent methods developed using neural network, Fuzzy logic, DCT, Wavelet transforms are complex but produce good results rather than conventional methods. Connected component based, Edge detection based methods are comparatively easy to develop but are less accurate than modern techniques. Hence, to achieve good performance, System needs to be design by combining these techniques as per user's requirement.

REFERENCES:

[1] Ho Vu, Duongl and Quoc Ngoc, 'A Feature Learning Method for Scene Text Recognition', *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 176 - 180, 2012.*

[2] Jun Ohya, Akio Shio, and Shigeru Akamatsu, 'Recognizing Characters in Scene Images', *IEEE Transactions on Pattern Analysis And Machine Intelligence. Vol. 16, No. 2, February 1994.*

[3] Linlin Li, Chew Lim Tan, 'Character Recognition under Severe Perspective Distortion, *IEEE 19th International Conference on Pattern Recognition(ICPR), pp.1-4, 2008.*

[4] C.P. Sumathi, N. Priya, 'Analysis of an Automatic Text Content Extraction Approach in Noisy Video Images', *International Journal of Computer Applications (0975 – 8887) Volume 69– No.4, May 2013*

[5] Julinda Gllavata, Ralph Ewerth and Bernd Freisleben, 'A Robust Algorithm for Text Detection in Images', *IEEE Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (ISPA), Volume 2, pp.611 – 616, 2003.*

[6] Roberto Manduchi and James Coughlan**, '**Computer Vision Without Sight', *Communications of the ACM, Vol.55, no.1, January 2012.*

[7] Hrishav raj, Rajib Ghosh, 'Devanagari Text Extraction from Natural Scene Images', *IEEE International Conference on Advances in Computing,Communications and Informatics (ICACCI), pp.513-517, 2014.*

[8] M. Swamy Das, B. Hima Bindhu, A. Govardhan, 'Evaluation of Text Detection and Localization Methods in Natural Images', *International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 6, pp.277-282, 2012.*

[9] Qixiang Ye and David Doermann**, '**Text Detection and Recognition in Imagery: A Survey', *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.37, No.7, July 2015.*

[10] Huizhong Chen, Sam S. Tsai, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, 'Robust Text Detection In Natural Images with Edge-Enhanced Maximally Stable Extremal Region', *Image Processing (ICIP), 18th IEEE conference on Image Processing, 2011.*

[11] G. Gayathri Devi, T. Santhanam and C.P. Sumathi, 'A Survey On Various Approaches of Text Extraction in Images', *International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.4, August 2012.*

[12] Shyama Prosad Chowdhury, Soumyadeep Dhar, Karen Rafferty, Bhabatosh Chanda, 'Robust Extraction Of Text From Camera Images

Using Colour And Spatial Information Simultaneously', *Journal Of Universal Computer Science, Vol. 15, No.18, pp.3325- 3342, 2009.*

[13] Y. Pan, X. Hou, and C. L. Liu, **'**Text Localization in Natural Scene Images based on Conditional Random Field, *Proc. IEEE International Conf. Doc. Anal. Recognition, pp. 6-10, 2009.*

[14] Michael R. Lyu, Jiqiang Song and Min Cai, 'A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction', *IEEE Transactions on Circuits And Systems For Video Technology, Vol. 15, No. 2, pp.243-255, February 2005*.

[15] Ananta Singh, Dishant Khosla, 'Text Localization and Recognition in Real-Time Scene Images', *International Journal of Scientific Engineering and Research (IJSER) ,Volume 3 Issue 5, pp. 123-125, May 2015*.

[16] Adesh Kumar, Pankil Ahuja, Rohit Seth,'Text Extraction and Recognition from an Image Using Image Processing In Matlab', *Conference on Advances in Communication and Control Systems (CAC2S), pp. 429-433, 2013.*

[17] Sonia George, Noopa Jagdeesh , 'A Survey on Text Detection and Recognition from Blurred Images', *International Journal of Advanced Research Trends in Engineering and Technology(IJARTET), Vol. II, Special Issue X, pp. 1180-1184, March 2015.*

[18] Nobuo Ezaki, Marius Bulacu, and Lambert Schomaker, 'Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons', *Proc. of 17th Int. Conf. on Pattern Recognition (ICPR), IEEE Computer Society, pp. 683-686, vol. II, 23-26 August, Cambridge, UK, 2004.*

[19] Madhu S. Nair, K. Revathy, and Rao Tatavarti, 'Removal of Salt-and Pepper Noise in Images: A New Decision-Based Algorithm', *Proceedings of the International Multi-Conference of Engineers and Computer Scientists IMECS, Hong Kong, Volume I, March 2008.*

[20] Andrej Ikica, Peter Peer, 'An improved edge profile based method for text detection in images of natural scenes', *IEEE EUROCON - International Conference on Computer as a Tool (EUROCON), pp. 1-4, 2011.*

[21] Lukas Neumann, Jiri Matas, 'Real-Time Scene Text Localization and Recognition', *25th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Providence, RI, USA, June 16-21, 2012.*

[22] Rodrigo Minetto, Nicolas Thome, Matthieu Cord, Neucimar J. Leite, Jorge Stolfi, 'Snooper Text: A text detection system for automatic indexing of urban scene', *Computer Vision and Image Understanding , journal homepage: www.elsevier.com/locate/cviu, 2013.*

[23] Xiaoqian Liu, Weiqiang Wang ,"Extracting Captions From Videos Using Temporal Feature", *Proceedings Of The International Conference On Acm Multimedia, pp.843-846, 2010.*

[24] Liang Wu, Palaiahnakote Shivakumara, Tong Lu, and Chew Lim Tan 'A New Technique for Multi-Oriented Scene Text Line Detection and Tracking in Video', *IEEE Transactions on Multimedia, Vol. 17, No. 8, pp.1137-1152, August 2015*

[25] Xiaodong Huang, Huadong Ma**,** 'Automatic Detection and Localization of Natural Scene Text in Video', *International Conference on Pattern Recognition (ICPR 2010), IEEE Computer Society, pp.3216-3219, 2010.*

[26] N. Senthilkumaran and R. Rajesh, 'Edge Detection Techniques for Image Segmentation – A Survey of Soft Computing Approaches' *International Journal of Recent Trends in Engineering, Vol. 1, No. 2, pp.250-254, May 2009.*

[27] Wahyono, Munho Jeong and Kang-Hyun Jo, 'Multi Language Text Detection Using Fast Stroke Width Transform', *IEEE 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp.1-4, 2015.*

[28] Miriam Leon, Veronica Vilaplana, Antoni Gasull, Ferran Marques , 'Caption Text Extraction For Indexing Purposes Using A Hierarchical Region-Based Image Model", *Proceedings Of The 16th IEEE International Conference On Image Processing, pp.1869-1872, 2009.*

[29] Guowei Yang, Fengchang Xu, 'Research and analysis of Image edge detection algorithm Based on the MATLAB', *Elsevier Ltd., Procedia Engineering 15 ,pp.1313-1318, 2011.*