IJO-Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

Comparative Study on Text Detection and Recognition from Traffic Image

Asit Kumar
M.Tech Scholar
Department of Electronic & Communication
OCT, Bhopal

Sumit Kumar
Assistant Professor
Department of Electronic & Communication
OCT, Bhopal

Abstract- **Text plays an significant role in day-to-day life because of its dissimilarities in text size, font, style, orientation and alignment as well as composite background and rich information, as a consequence automatic text detection in natural scenes has several attractive applications. Though, detecting and recognizing such text is all the time a challenging issue. Several text extraction techniques grounded on edge detection, connected component analysis, morphological operators, wavelet transform, texture features, neural network etc. have been established. This paper contributes comparative analysis of different technique which provides efficient performance.**
Keywords: **Scene structure, text detection, Maximally Stable Extremal Regions (MSER), traffic text sign recognition.**

I. INTRODUCTION

Text in images comprises of valuable statistics and is exploited in many applications that uses image and video applications, such as content-based web image search, video information retrieval, and mobile based text analysis and text recognition [1-5]. Due to composite background, and deviations of font, size, color and orientation, text in natural scene images has to be vigorously detected before being recognized and regained. Prevailing methods for scene text detection can coarsely be classified into three groups: sliding window based methods [6]–[8], connected component based methods [9]–[13], and hybrid methods [14]. Sliding window based approaches, also termed as region-based methods that uses a sliding window to find for probable texts in the image and further use machine learning techniques to recognize text. These approaches are slow as the image has to be handled in multiple scales. Connected component based methods derives candidates for character recognition from images by connected component analysis trailed by combination of character candidates into text lines. With a number of significant areas of application, i.e. advanced driver assistance systems, road surveying, and autonomous vehicles, text detection and recognition is a challenging task.

While most of the study exists on both the automatic detection and recognition of symbol-based traffic signs, and the recognition of text in real scenes, there is far less research motivated specifically on the recognition of text on traffic information signs. This could be partially due to the difficulty of the task affected by problems, such as illumination and shadows, blurring, occlusion, and sign deterioration.

To detect text candidates and extracts connected components as character candidates by Maximally Stable Extremal Regions (MSERs) grounded approaches have developed the attention of several recent works [15].

The reminder of this paper proceeds as follows. Section II presents the review of past work. Section III presents the methodology used for the detection of text-based traffic signs. Section IV explains comparative study of different methods to illustrate the performance of the system. Conclusion is illustrated in section V.

IJO-Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

II.RELATED WORKS

As described above, MSER-based methods have demonstrated very promising performance in many real projects. However, current MSER-based methods still have some key limitations, i.e., they may suffer from detecting of repeating components and also insufficient text candidates construction algorithms. The main advantage of MSER-based methods over traditional connected component based methods roots in the usage of the MSERs algorithm for character extraction – the MSERs algorithm is able to detect most characters even when the image is in low quality (low resolution, strong noises, low contrast, etc.)[16]. However, one severe but not so obvious pitfall of the MSERs algorithm is that most of the detected MSERs are in fact repeating with each other. Repeating MSERs are problematic for the latter character candidates grouping algorithm, thus most of the repeating MSERs, apart from the MSERs that most likely correspond to character, need to be removed before being fed to the character grouping algorithm. Carlos et al. [19] presented a MSERs pruning algorithm that contains two steps: (1) reduction of linear segments by maximizing the border energy function; and (2) hierarchical filtering with a cascade of filters. Neumann and Matas [21] proposed a MSER++ based text detection method, which exploits rather complicated features, e.g., higher-order properties of text and uses exhaustive search for pruning. Later, Neumann and Matas [20] presented a two-stage algorithm for Extremal Regions (ERs) pruning with the exhaustive search strategy. In the first stage, a classifier trained from incrementally computable descriptors (area, bounding box, perimeter, Euler number and horizontal crossing) is used to estimate the class-conditional probabilities of ERs; ERs corresponding to local maximum of probabilities in the ER inclusion relation are selected. In the second stage, ERs passed the first stage are classified as characters and non-characters using more complex features. The above methods all explore the hierarchical structure of MSERs, but have used different methods for estimating the probabilities of MSERs corresponding to characters. To deal with the large number of repeating MSERs, they have used relevant features (cascading filters and incrementally computable descriptors) in pruning. Another problem with MSER-based methods, or more generally, connected component based methods and hybrid methods, is the absence of an effective text candidates construction algorithm. The existing methods for text candidates construction fall into two general approaches: rule-based [18]–[20] and clustering-based methods [14]. Much research exists on the detection and recognition of text in natural scenes. Approaches to this problem can be broadly divided into two groups: region-based methods, e.g., [9], [12], and [18], and connected component (CC)-based methods. Region-based text detection methods use local features, such as texture, to locate text regions, whereas CC-based methods attempt to segment text characters individually by using information such as intensity, color distribution, and edges. They usually consist of three phases: a first stage to detect CCs within the image, a second stage to eliminate unlikely CCs based on their features, and a final stage that attempts to group the remaining CCs into words or lines. The existing state-of-the-art methods all consist of two stages: detection and recognition. In [16] segmented regions of interest based on color information, by applying a threshold to the chrominance and luminance channels in the Lab color space. Rectangular regions were found by

IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

comparing the fast Fourier transform (FFT) signature of each blob to the FFT signature of a rectangular shaped reference. The method presented by [22] made use of MSERs for the detection of both traffic signs and text characters. White and blue traffic panels were detected in each frame, using a combination of color segmentation and bag of visual words. These regions were then classified using both support vector machines and Naïve Bayes classifiers. The method was applied to single images, with no use of temporal information, and the emphasis placed on the geolocalization of traffic signs using Global Positioning System information. In [23] author proposed a novel system for the automatic detection and recognition of text in traffic signs. Scene structure is used to define search regions within the image, in which traffic sign candidates are then found. Maximally Stable Extremal Regions (MSERs) and hue, saturation, and value color thresholding are used to locate a large number of candidates, which are then reduced by applying constraints based on temporal and structural information. A recognition stage interprets the text contained within detected candidate regions. Individual text characters are detected as MSERs and are grouped into lines, before being interpreted using optical character recognition (OCR). Recognition accuracy is vastly improved through the temporal fusion of text results across consecutive frames.

III. DETECTION OF TEXT BASED TRAFFIC SIGN

The automatic detection and recognition of traffic signs is a challenging issue, with a huge area of application, comprising driver assisting devices, road surveying, etc. The detection stage exploits knowledge of the structure of the scene and next stage of the algorithm attempts to extract text within the detected search region [21]. While much research exists on both the automatic detection and recognition of symbol-based traffic signs and the recognition of text in real scenes. This could be partly due to the difficulty of the task caused by problems, such as illumination and shadows, blurring, occlusion, and sign deterioration. The text detection system comprises two main stages: recognition and extraction as shown in figure 1. Much research exists on detection of text in natural scenes. Approaches to this problem can be broadly divided into two groups: region-based methods and connected component (CC)-based methods [24]. Region-based text detection methods use local features, such as texture, to locate text regions, whereas CC-based methods attempt to segment text characters individually by using information such as intensity, color distribution, and edges. They usually consist of five phases: a first stage to filter image, a second stage to detect CCs within the image, a third stage to eliminate unlikely CCs based on their features, fourth stage to line formation and a final stage that attempts to group the remaining CCs into words or lines.

Text Recognition

In this step firstly traffic image is taken as input as shown in figure 1 in which we have to detect text. Due to imperfections in the imaging and capturing process, however, the recorded image invariably represents a degraded version of the original scene[18-20]. The degradation results in image blur, affecting identification and extraction of the useful information in the images. It can be caused by relative motion between the camera and the original scene, by an out of focus of optical system, atmospheric turbulences and aberrations in the optical system. Noise introduced by the medium through which the image is created can also cause degradation. The degradation phenomenon of the acquired images causes

IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

serious economic loss. Therefore, restoring the degraded images is an urgent task in order to expand uses of the images by using noise removal and de-blurring techniques. Further edge detection algorithm is used for image edge detection. Such algorithm runs in 5 separate steps i.e. smoothing, gradient findings, local maxima marking, finding threshold and finally edge tracking.

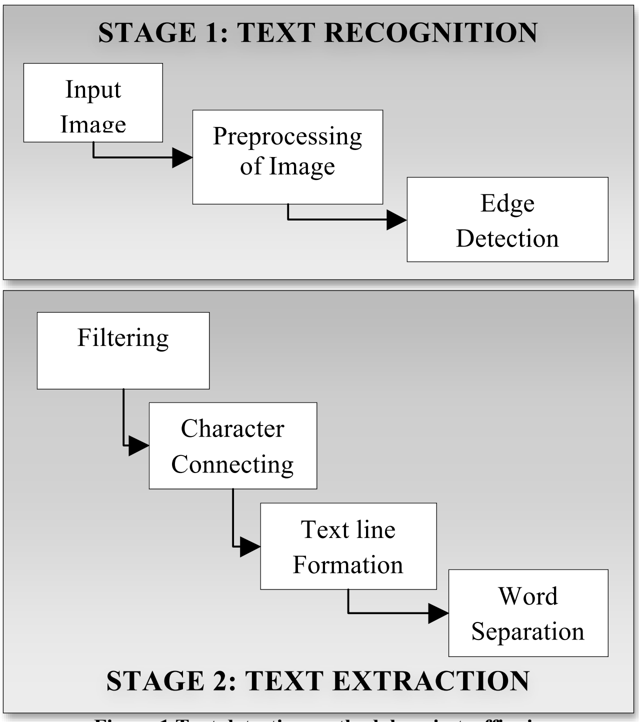

Figure 1 Text detection methodology in traffic sign

Text Extraction

With the extraction of enhanced edge, we obtain a binary image where the foreground CCs are considered as letter candidates. As in most state-of-the-art text detection systems, we perform a set of simple and flexible geometric checks on each CCs to filter out non-text objects.

Text lines are important cues for the existence of text, as text almost always appear in the form of straight lines or slight curves. To detect these lines, we first pairwise group the letter candidates using rules. The next stage of the algorithm locates lines of text within the detected candidate regions. This allows the total number of CCs to be reduced, removing non-character CCs and hence improving the chances for higher accuracy[23, 25]. As a final step, text lines are split into individual words by classifying the inter letter distances into two classes: the character spacing and the word spacing.

IV. COMPARATIVE ANALYSIS OF TEXT EXTRACTION

The several techniques of text extraction are as follow:

Sliding window based method: Region-based technique uses the assets of the color or gray scale in the text region or their alterations to the corresponding properties of the contextual. They are recognized on the condition that there is very small aberration of color restricted by text and this color is sufficiently separate from text's immediate background [30]. Text can be achieved by thresholding the image at intensity level amongst the text color and that of its instant background. This technique is not vigorous to complex background. This method is further sub-divided as connected component (CC) and edge based.

CC based approaches: CC-based approaches [23, 25] use a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is required to merge the text components using the spatial arrangement of those components so as to filter out non-text components and the boundaries of the text regions are marked. This method locate locates text quickly but fails for complex background.

Edge based approaches: Edges are a reliable feature of text regardless of color/intensity, layout, orientations, etc. Edge based method [31] is focused on high contrast between the

IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

text and the background. The three distinguishing characteristics of text embedded in images that can be used for detecting text are edge strength, density and the orientation variance. This method is not robust for handling large size text.

Texture based Method: This method uses the fact that text in images have discrete textural properties [25] that distinguish them from the background. This method is able to detect the text in the complex background. The only drawback of this method is large computational complexity in texture classification stage.

Morphological based Method: Mathematical morphology is a topological and geometrical based method for image analysis. Morphological feature extraction techniques have been efficiently applied to character recognition and document analysis. It is used to extract important text contrast features from the processed images [25]. These features are invariant against various geometrical image changes like translation, rotation, and scaling. Even after the lightning condition or text color is changed, the feature still can be maintained. This method works robustly under different image alterations.

Many researches have been done on various text extraction techniques such as region based (CC based and edge based) [23], texture based, morphological based or combination of these technique (i.e. hybrid approach). Researchers have used different type of images for their experimentation. The detailed analysis of text extraction techniques is shown in Table 1.

V. CONCLUSION

Detecting text regions in natural scene images and traffic images has developed an significant area due to its several applications. Text Extraction system includes finding text regions in a given image, localizing it, extracting the text part and recognizing text. In this paper, a number of methods such as region based, edge based, connected component (CC) based, texture based, morphological based etc. have been deliberated and a comprehensive evaluation of these procedures on the basis of numerous factors such as precision rate, recall rate, accuracy etc. has been done. Each method has its individual benefits and limitations. Even however there are various numbers of algorithms, there is no sole combined methodology that fits for all the applications due to difference in font, size, alignment, complex background of text etc. It is concluded that character connected method can detect and localize text accurately even when images are noisy or with complex background.

Author	Technique Used	Parameters	Remarks
Yao et al. [24]	CC and Support Vector Machine (SVM)	PR=64% RR=60%	Pixels of each character assumed to have similar color.
Zhang et al. [25]	Discrete Wavelet Transform (DWT), k-means clustering, morphology Operations	DR= 94.5%, FAR= 13.6%	Text character Color independent.
Fan et al. [26]	Stroke features and connected component	PR=95.2% RR= 94.5%	Color information is not fully used.

IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
ISSN : 2455-0108

Angadi et al. [27]	Discrete Cosine Transform and texture features extraction	DR=96.6%	Inefficient when background in the image is more complex like trees, vehicles.
Anoual et al. [28]	Edge detection, texture features, connected component analysis	PR=95% RR=89%	Robust and effective.
Kumar et al. [29]	CC Analysis	PR=90% RR=89%	Capable of Multilingual Text extraction.
Hassanzadeh et al. [30]	Morphological operator, Decision classifier	PR=95.6% Accuracy=86.9%	A novel and fast method for logo detection.
Seeri et al. [31]	Median filter, Sobel edge detector, connected component labeling, order static filter.	PR=84.21% RR=83.16% Accuracy =75.77%	Fails to extract very small characters
Azadboni et al. [32]	FFT Domain Filtering , SVM Classification, K-means clustering	DR= 98.10%	Text characters having uniform color.
Raj et al. [33]	CC based	PR= 72.8%, RR=74.2 %	Fails for small slanted/curved text.
Jack Greenhalgh and Majid Mirmehdi [23]	MSERs and HSV thresholding	F_measure= 87%	Not efficient for blurred images.

REFERENCES

[1] Y. Zhong, H. Zhang, and A. Jain, "Automatic caption localization in compressed video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 385–392, Apr. 2000.

[2] H. Li, D. Doermann, and O. Kia, "Automatic text detection and tracking in digital video," IEEE Trans. Image Process., vol. 9, no. 1, pp. 147–156, Jan. 2000.

[3] J. Weinman, E. Learned-Miller, and A. Hanson, "Scene text recognition using similarity and a lexicon with sparse belief propagation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1733–1746, Oct. 2009.

[4] X.-C. Yin, H.-W. Hao, J. Sun, and S. Naoi, "Robust vanishing point detection for mobile cam-based documents," in Proc. ICDAR, Beijing, China, 2011, pp. 136–140.

[5] P. Shivakumara, Q. P. Trung, and C. L. Tan, "A laplacian approach to multi-oriented text detection in video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412–419, Feb. 2011.

[6] X. Chen and A. Yuille, "Detecting and reading text in natural scenes," in Proc. IEEE Conf. CVPR, vol. 2. Washington, DC, USA, 2004, pp. 366–373.

[7] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, "Adaboost for text detection in natural scene," in Proc. ICDAR, Beijing, China, 2011, pp. 429–434.

[8] K. Kim, K. Jung, and J. Kim, "Texture-base approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1631–1639, Dec. 2003.

[9] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. IEEE Conf. CVPR, San Francisco, CA, USA, 2010, pp. 2963–2970.

[10] C. Yi and Y. Tian, "Text string detection from natural scenes by structure-based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sept. 2011.

[11] C. Mancas-Thillou and B. Gosselin, "Color text extraction with selective metric-based clustering," Comput. Vis. Image Und., vol. 107, no. 1–2, pp. 97–107, 2007.

[12] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4256–4268, Sept. 2012.

[13] C. Yi and Y. Tian, "Text extraction from scene images by character appearance and structure modeling," Comput. Vis. Image Und., vol. 117, no. 2, pp. 182–194, 2013.

[14] Y.-F. Pan, X. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. Image Process., vol. 20, no. 3, pp. 800–813, Mar. 2011.

[15] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proc. Int. Conf. Mach. Learn., San Francisco, CA, USA, 2001, pp. 282–289.

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. Brit. Mach. Vis. Conf., vol. 1. 2002, pp. 384–393.

[17] A. Shahab, F. Shafait, and A. Dengel, "ICDAR robust reading competition challenge 2: Reading text in scene images," in Proc. ICDAR, 2011, pp. 1491–1496.

[18] H. Chen et al., "Robust text detection in natural images with edge enhanced maximally stable extremal regions," in Proc. IEEE Int. Conf. Image Process., 2011, pp. 2609–2612.

[19] C. Merino-Gracia, K. Lenc, and M. Mirmehdi, "A head-mounted device for recognizing text in natural scenes," in Proc. Int. Workshop CBDAR, Beijing, China, 2011, pp. 29–41.

[20] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in Proc. IEEE Conf. CVPR, Providence, RI, USA, 2012, pp. 3538–3545.

[21] L. Neumann and J. Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in Proc. ICDAR, Beijing, China, 2011, pp. 687–691.
[22] A. González, L. Bergasa, and J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 228–238, Feb. 2014.
[23] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs", IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 3, pp 1360-1369, June 2015.
[24] J.L. Yao, Y.Q. Wang, L.B. Weng, Y.P. Yang, *Locating Text Based on Connected Component And SVM,* International Conference On Wavelet Analysis And Pattern Recognition, Vol. 3, 2007, pp. 1418 - 1423.
[25] X.W. Zhang, X.B. Zheng, Z.J. Weng, *Text Extraction Algorithm Under Background Image Using Wavelet Transforms,* IEEE Proceedings of International Conference On Wavelet Analysis And Pattern Recognition, Vol. 1, 2008, pp. 200-204.
[26] W. Fan, J. Sun, Y. Katsuyama, Y. Hotta, S. Naoi, *Text Detection in Images Based on Grayscale Decomposition and Stroke Extraction,* Chinese Conference on Pattern Recognition, IEEE, 2009, pp. 1-4.
[27] S. A. Angadi, M. M. Kodabagi, *Text Region Extraction from Low Resolution Natural Scene Images using Texture Features*, IEEE 2nd International Advance Computing Conference, 2010, pp. 121-128.
[28] H. Anoual, D. Aboutajdine, S.E. Ensias, A.J. Enset, *Features Extraction for Text Detection and Localization,*5th International Symposium on I/V Communication and Mobile Network, IEEE, 2010, pp. 1-4.
[29] M. Kumar, Y.C. Kim and G.S. Lee, *Text Detection using Multilayer Separation in Real Scene Images,* 10th IEEE International Conference on Computer and Information Technology, 2010, pp. 1413-1417.
[30] S. Hassanzadeh, H. Pourghassem, *Fast Logo Detection Based on Morphological Features in Document Image,* 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, 2011, pp. 283-286.
[31] S.V. Seeri, S. Giraddi and Prashant. B. M, *A Novel Approach for Kannada Text Extraction*, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, 2012, pp. 444-448.
[32] M.K. Azadboni, A. Behrad , *Text Detection and Character Extraction in Color Images using FFT Domain Filtering and SVM Classification,* 6th International Symposium on Telecommunications. IEEE, 2012, pp. 794-799.
[33] H. Raj, R. Ghosh, *Devanagari Text Extraction from Natural Scene Images, 2014 International* Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2014, pp. 513-517.