An End-to-End Sequence Learning Approach for Text Extraction and Recognition from Scene Image

G. Lalitha[1†] and B. Lavanya[2††],

Department of Computer Science, University of Madras, Chennai-25, TamilNadu, India

Abstract
Image always carry useful information, detecting a text from scene images is imperative. The proposed work's purpose is to recognize scene text image, example boarding image kept on highways. Scene text detection on highways boarding's plays a vital role in road safety measures. At initial stage applying pre-processing techniques to the image is to sharpen and improve the features exist in the image. Likely, morphological operator were applied on images to remove the close gaps exists between objects. Here we proposed a two phase algorithm for extracting and recognizing text from scene images. In phase I text from scenery image is extracted by applying various image preprocessing techniques like blurring, erosion, tophat followed by applying thresholding, morphological gradient and by fixing kernel sizes, then canny edge detector is applied to detect the text contained in the scene images. In phase II text from scenery image recognized using MSER (Maximally Stable Extremal Region) and OCR; Proposed work aimed to detect the text contained in the scenery images from popular dataset repositories SVT, ICDAR 2003, MSRA-TD 500; these images were captured at various illumination and angles. Proposed algorithm produces higher accuracy in minimal execution time compared with state-of-the-art methodologies.

Keywords:
MSER – Maximally Stable Extremal Region, OCR – Optical Character Recognition, CC – Connected Components, SWV-Stroke Width Variation.

1. Introduction

Scene text detection plays vital role in road safety measures, by recognizing context present in scene images i.e., boarding kept on highways which serve traveller in various aspects. So detecting text from scene image is imperative. Text Extraction from scene images is challenging rather than extracting text from scanned images or document images, since scene images can also be captured by low-cost devices in uneven lightning, so as to detect text from such image is rigid. Scene text extraction plays majority of role in directing autonomous vehicles, directing vehicle is an imperative task since people always aware of the route for their destination place, so directing route eases tourist. Text extraction from images can be done in various ways i) Edge based detection ii) Connected Components-based detection iii) Texture based detection. Edge detection is used to threshold/segment the image. Edge detection is done by calculating the width, height and strokes of character candidate. The connected components methods, it assumes that pixels of each character candidates have similarity than non-character candidate present in the image, so based on such similarity measures character candidate regions segmented from background i.e., non-character region present in the image.

In connected components (CC) various features of character candidate such as aspect ratio, stroke width and size were used to identify candidate region from non-character region. In Texture based methods it assumes that text present in image has a special texture. Those features were used to classify text from non-text region. OCR (Optical Character Recognition) is another technique which is used to extract text from images, though it too suffers to detect text from image when text embedded with complex background. Here various measures have been done to extract text from scenery images; though it is complex to extract if texts of various fonts, size and illumination exist in the image. So an effective algorithm is mandatory to recognize text from scene images even from complex backgrounds since the aim of the proposed work is to guide the people who are driving on highways by recognizing such images which will guide them in their emergency situation. Here proposed work designed in two phase which extracts, detects and recognize text from scene images which have not performed in the state–of-art methodologies.

The proposed method organized as follows as related works, proposed method, dataset, results and discussion, pseudocode, performance evaluation, conclusion and future work.

2. Objective of the work

The proposed work's primary goal is to serve driving people who drives vehicle in highway, many types of boarding will be there in road which serves many purpose for people, direction boarding, nearby places boarding and hospital boarding will be kept, so recognizing text is imperative. Proposed work is to extract and recognize text from scenery images for that dataset from three popular dataset repositories were used namely SVT, ICDAR 2003, and MSRA-TD 500. Proposed work is done in two phases in first phase text

Manuscript received July 5, 2022
Manuscript revised July 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.7.27

extraction is done and in second phase text recognition is done. For extracting text from phase I initially image pre-processing techniques like blurring, erosion, tophat is applied. For segmenting text from scene images, simple thresholding is applied; later morphological gradient applied to enrich the edges exist in the image, followed by canny edge detector is applied to extract character region present in image. In phase II for recognizing text present in scene image, following steps have to be done.

- Pruning non-text regions by applying MSER based on text geometric properties
- Pruning non-text regions based on strokes of text regions
- Generating bounding box around recognized text from scene images

3. RELATED WORKS

` Parallel structures were used to create character candidate regions and in-depth segmentation have been applied to extract text from scene image, while extracting to filter non-character candidate, double-layer filtering were utilized to prune non-character region from character region and the edges of connected graph were pruned using SVM (Support Vector Machine) classifier[1]. At some cases wavelet transform have been applied on the original image, then image transmission have been done and K-means technique image is clustered into three specified clusters, such as text, simple and complicated backdrop, after sliding window is moved over the altered image[11]. The text and non-text regions must be separated using CC (Connected Component), which can be applied across the scene image by computing the geometric properties of the text. Four channels, including the grey, hue, saturation, and Cb channels, were used to extract the ER (Extremal Region) of the character region. After removing the superfluous character candidate using the RLS method [2], text line generation and verification were completed using the CNN algorithm (Convolutional Neural Network). Using MSER region detector to detect text region and by applying adaptive thresholding to segment character candidate region from non-character candidate from scene image. After segmentation the extracted character candidates were grouped into words using HOG (Histogram of Gradient) classifier [4]. By applying edge detection four edge maps were detected using those features of text candidate were extracted and by applying k-means algorithm extracted text candidates were combined based on edge information of those candidates[2]. Color and texture detection approaches were applied to detect text from color image and from video frames. Those techniques were applied at each pixel, texture feature of an image, using k-means algorithm, the extracted features were clustered into two clusters such as text cluster and Non-text cluster i.e. background content [13]. While applying high pass filter, background of the image will be suppressed to accurately to detect text, which is done at low computation. By using CC-based method text components will be detected accurately. In order to prune non-text regions morphological operations [6, 8, and 9] were applied to remove too large and small regions. Gradient based and color-based partitions were done to remove non-text region; grouping of text to form line were performed by taking measurements of characteristics like height, width, centroid area, and distance between centroids [3, 5]. CC based and Region-based approaches were used to detect and extract features of text, and classification is done by applying HOG [10]. Image decomposition is done by using wavelet transform and SVM [12], Since many state-of-art methodologies exists but it still lags the accuracy in recognizing scene text images, and it suffers to recognize text from challenging environment.

4. Proposed method

Proposed algorithm consists of two phases, phase I for text extraction and phase II for text recognition. Fig 1 demonstrates the phase I work flow. As an initial step, a grayscale version of the input image is created, and immediately Gaussian blur is applied on image. Blurring effect is applied on image to remove the noise and extra added details in the image, which is added to make the image look enhanced. So before heading in to the processing of image, it's mandatory to remove the extra added effects on the image and to extract the features. By applying blurring technique, image will look smoothened. By applying Gaussian blur image, will be blurred naturally, than using other blurring techniques and it preserves more of the edges in the image. In an image, erosion thins and weakens the foreground object. It alters pixels in an image close to an object's edge. A structuring element is first defined by erosion, which then moves the structuring element up and down the input image from left to right. If the structuring element's pixel is greater than 0, a foreground pixel won't erode. Erosion can be used to get rid of little blobs in an image or separate two related items. tophat also known as a whitehat, which is used to reveal brighter regions in an image from dark backgrounds. Thresholding is used to segment image; it segments foreground and background of the image. It's a binarization of an image. It creates a binary image with pixels that are either 0 or 255 from a grayscale image. Here simple thresholding is applied to segment image, it select a threshold value T, and it check a condition whether pixel intensity value greater than or less to T value, based on thresholding applied on image. Here, contour highlighting is done using morphological gradient. It is useful for determining the outline of a particular object from an image. Morphological gradient is applied here to recognise the text regions present in an image. In order to retrieve the structuring element within the image, kernel size has to be fixed. So that morphological gradient operator can be applied by looping over the kernels of an image, to highlight the contours present in the image.

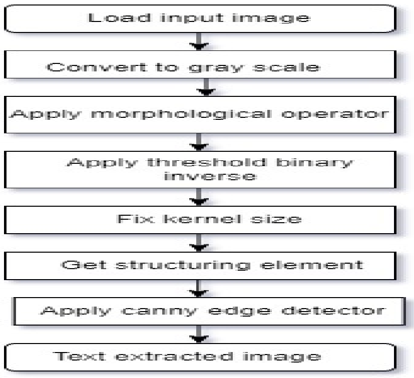


Fig.1. Phase I of proposed work

Fig 4 depicts the work flow of phase II for text recognition. Phase II also proposed to recognize text, so MSER is applied to detect text region. MSER performs well at recognizing text ROI (Region of Interest), variation of color region from text to non-text regions contributes more for identifying text ROIs. MSER is a key-point detector, it exhibit a key point detector, where it identifies areas of an image which exhibits connected components, near uniform i.e. almost identical pixel intensities, from contrasting background. If all three of these cases hold, then the image region can be marked as a ROI.

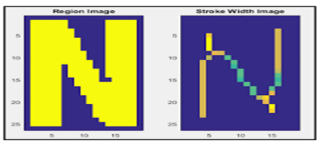


Fig 2. Image after applying SWT

By applying MSER it identifies maximum text region in image but some non-text region which appears like text region also recognized as False Positive (FP) text regions here. Text has many properties which can be applied on image to calculate its properties so that it non-text regions can be pruned in next level from MSER. To measure MSER properties some region properties such as bounding box, eccentricity, solidity, extent, and Euler were measured, aspect ratio is computed based on bounding box data, thresholding the pixel based upon metrics to determine non-text regions to get pruned. Threshold value varies for each image. Dataset images which has taken has un-structured images which have many co-relevant non ROIs in the images. Those regions has to be classified has non-ROIs while recognizing text regions. As a next consecutive step stroke width of the image is calculated in order to filter non-text region from text region. Text is formed with lines and curve, which defines texts, even non text region too has such strokes and curves but both have a major difference in terms of strokes. It’s an extent of the measure of width of the lines the image operator called Stroke Width Transform (SWT) calculates the width of each pixel's strokes. A continuous section of a picture that has a uniform width is called a stroke. [14].

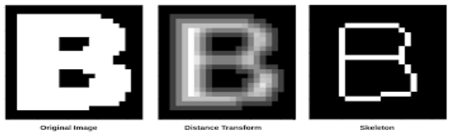


Fig 3. Image after applying distance transform and skeletonization

In proposed work to prune non text region from text region, Stroke Width Variation (SWV) of the image is calculated. To calculate SWV of an image, each region is manipulated in a binary form and it is padded to avoid boundary effects while computing stroke width computation. Strokes play a major role in segregating non text regions; by evaluating the stroke width of regions detected after applying MSER region detector, by applying SWT image will look like Fig 4. Stroke width of the ROI region and non ROI region can be evaluated using distance transform and binary thinning operation [15]. Distance transform [16] transforms a binary image with feature and non-feature pixels into a new image where each non-feature pixel is assigned a value that represents the distance from the feature pixel that is closest to it. Thinning [17] is the process of obtaining the skeleton of the region using skeletonization. Skeletonization is the process of extracting skeletons of an object in a given image. It is done by applying morphological operator which deletes dark i.e., zero intensity foreground pixels in an iterative manner which is done in layer by layer until completely obtaining skeleton of the object. After applying distance transform and skeletonization image will look like Fig 3.

Difference between region image and stroke width image has a very minute variation has shown in Fig2 which depicts that a region has much similar characteristics of a text because the lines and curves which form a Similar widths are present throughout the region, which is a frequent textual feature. Using a threshold value T and considering the fluctuation throughout the entire region, SWV prunes non-text parts which is measured as shown in eq (1) and stroke width threshold value is fixed to stroke width measure is measured in terms of eq (1), based upon stroke width measure value, Then image will be applied threshold to prune non-text regions. Threshold value T may vary for images with various font styles. Detected regions composed of distinct text characters. Next step in identifying characters in a scene image is to recognise the text that is contained in the scene image.

$$\text{stroke width measure} = \frac{\sigma(\text{stroke width values})}{\text{mean}(\text{stroke width values})} \quad (1)$$

OCR is used to recognise the actual words that are present in an image; text present in scene image carries a meaningful data than just the individual characters. Recognizing correct characters alone doesn't yield the purpose of recognition, if order of character changed which may yield to wrong information. So it's mandatory to recognize the characters in correct order. So for recognizing in correct order, a method for integrating distinct finding neighbouring text regions and creating a bounding box around each independent letter turns text regions into words or lines. When a bounding box is stretched to identify neighbouring text regions, the bounding boxes of those regions overlap, forming a chain of overlapping bounding boxes for text regions that are a part of the same word or text line.

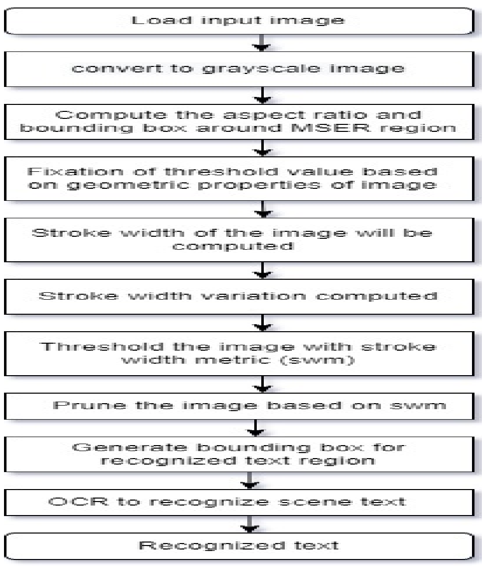


Fig 4. Phase II of proposed work

A single bounding box is created around each word and each text line by combining overlapping bounding boxes. Since all bounding box pairs overlap, the distance between any pair of text regions can be calculated, making it possible to identify clusters of nearby text regions by looking for non-zero overlap ratios. After computing the pair-wise overlap ratios, a graph is utilised to identify all the text sections in the image that are linked by an overlap ratio. The pair-wise overlap ratios for each of the extended bounding boxes are calculated using a bounding box overlap function, and all of the related regions are then identified using a graph. The result of connected text regions were used to combine all nearby bounding boxes into one bounding box, which is computed by finding minimum and maximum sizes of each bounding boxes which regions share text properties. Before computing final outcome falsely detected text region i.e., non-text region will be reduced by grouping all independent bounding box regions into one single bounding box, this process removes the separately recognized false positive region i.e., non-text region from bounding box which has ROI region.

As a final step, recognizing original text present in scene image that too in exact order; OCR function is used to recognize text; reason for finding text region before recognizing ROI is final outcome will be with more of false positive and true negative so before applying OCR it's mandatory to detect ROI first.

5. Dataset

Proposed work is to extract and recognize text from scene images, so three popular repositories images were used they are SVT, ICDAR 2003, MSRA-TD 500. Among the dataset's images were captured at varying angles, illumination and at varied lightning condition. Many images were skewed tilted, some images were blurred. MSRA-TD has the other national languages scene image too. Though dataset chosen in proposed work has many challenging aspects; proposed algorithm works best in extracting and recognising text from those scene images. ICDAR2003 consists 234 scene images, SVT consists 350 images, MSRA-TD 500 466 images, totally 1050 images were used in proposed work to check the efficiency of the algorithm.

6. Results and discussion

Proposed work aimed to recognize scenery images kept in road side, algorithm is tested evaluated using three popular dataset namely SVT, ICDAR 2003, MSRA-TD 500. Proposed algorithm is done in two phase: phase I text extraction, phase II text recognition. Phase I follows following sequences pre-processing, segmentation followed by edge/gradient detection of text region from scene images. Phase II follows the following workflow applying MSER, geometric properties, SWV, OCR. Experiment was done on integrated laptop with an Intel Core i5, 270 GHz, 8 GB of RAM, and Windows 10 64-bit OS in Anaconda Python. Comparing the proposed algorithm to existing approaches, it generates higher accuracy. as discussed in Table 11. Fig 5, Fig 6 depicts the various stages of correctly predicted results of proposed work. Fig 7 depicts the various challenging images which contains curved text image, other language images proposed algorithm recognizes challenging scene text images from dataset images.

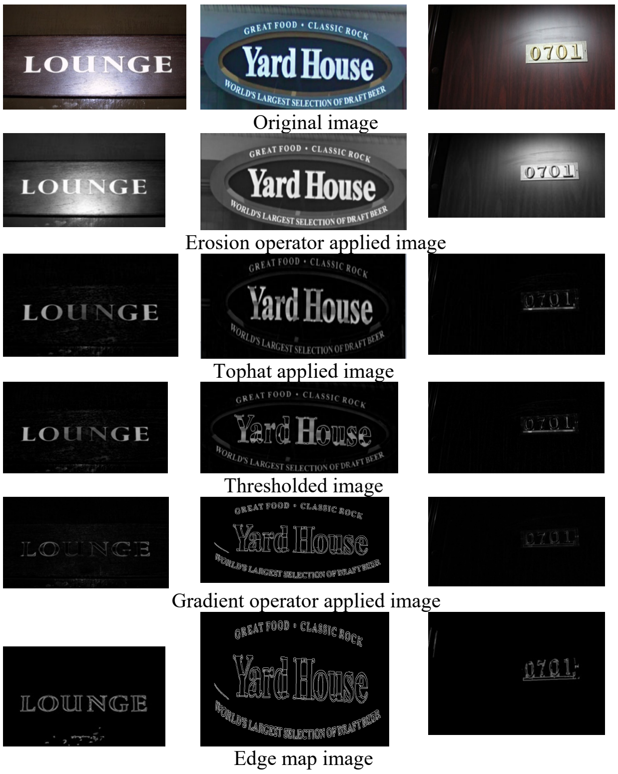


Fig 5. Various stages of correctly extracted results for phase I text extraction

7. PSEUDOCODE

Phase I: Text Extraction

Input: Scene Image

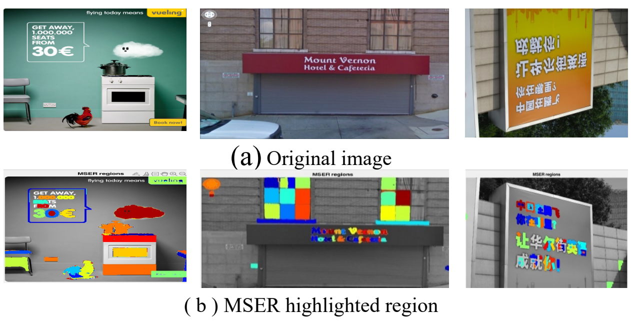
Load image and convert to grayscale
Apply Gaussian blur, erosion and tophat operation on image
Apply thresholding binary inverse to segment image
Fix kernel and apply morphological gradient over the image
For Kernel size in Kernel sizes
Kernel = Get Structuring Element ()
Gradient = Morphology Ex (image, Kernel)
End For
Apply canny edge detector
Output: Text extracted image

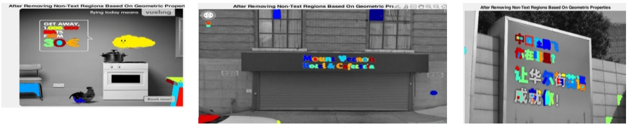
Phase II: Text Recognition

Input: Scene Image

Load image convert to gray_scale
Detect MSER regions = detect MSER_Features
Applying geometric region properties (‘bounding box’, ‘eccentricity’, ‘solidity’, ‘extent’, ‘euler’)
Compute aspect ratio using bounding box value measures
Aspect ratio=width/height
Apply thresholding to prune non-text regions
Pruning index=aspect ratio>3
ecentricity>1000
solidity<0.5
extent <0.3 || extent>1
Prune non text region display text region
Convert the image to binary region and apply padding to prune boundary effects during computing stroke of image
region_roi=pad array(region_roi,[1,1])
Compute stroke of image
compute stroke width variation measure eq(1)
segment image based on SWV value
SWV_threshold_index=0.5
SWV_filter_index=stroke_width_value>SWV_threshold_index
Prune non-text region based on SWV
Get bounding boxes for all the text regions in image
Bounding_boxes=verticalcat(mser_stats.bounding box)
Expand bounding box and bound image to it
Expansion ratio=0.03
Compute the overlap values
Overlap_value=bounding_box_overlap_value(expanded_boxes)
Assign the overlap value between bounding box to simply graph format
size (overlapvalue,1)
Generate graph, find all text regions
Graph generate(overlap_values)
Merge all the bounding boxed based on dimension of boxes
Recognize the text present in scene image by applying OCR.
OCR recognize=OCR (image,BBoxes)
Output: Text extracted image

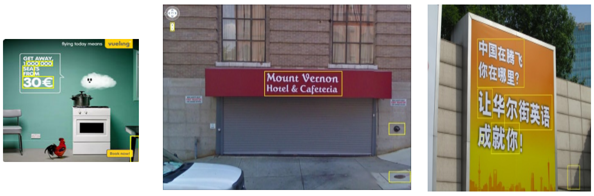
Fig 7. Proposed algorithm recognizes challenging scene text images from dataset images





(c) Non-text region pruned by applying geometric properties

(d) SWT applied image



(e) Recognized text region

Fig 6. Various stages of correctly extracted results for phase II text recognition

8. Performance Evaluation

In this proposed work, text extraction and recognition from scene image were evaluated using dataset images taken from SVT, ICDAR 2003, MSRA-TD 500, for extracting and recognizing boarding images kept on highways. The standard evaluation protocol has been followed to evaluate the proposed algorithm with other methods. The performance is calculated using the f-measure, recall rate, and precision rate. [18]. Recognition algorithms were evaluated using retrieval systems. More clearly, precision, recall measures were used here to determine the recognition rate of proposed algorithm.

$$\text{Precision rate (Pr)} = \frac{\text{No.of correctly retrived images}}{\text{Total no.of retrived images}} \quad (2)$$

$$\text{Recall rate (Rr)} = \frac{\text{No.of correctly retrived images}}{\text{No.of relevant images in the database}} \quad (3)$$

$$\text{F-measure score (Fs)} = 2 * \frac{Pr * Rr}{Pr + Rr} \quad (4)$$

Table 1 Different dataset images

Dataset names	*No. of. Images*
SVT	234
ICDAR 2013	350
MSRA-TD 500	466

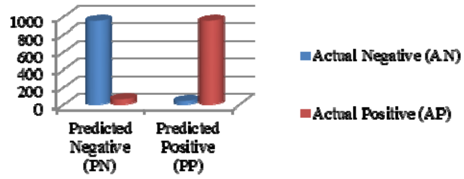


Fig 8. Pictorial representation of phase I of proposed work

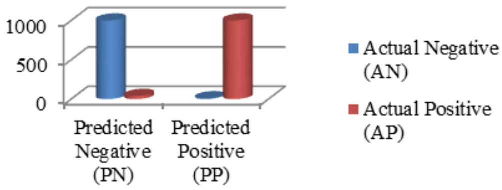


Fig 9. Pictorial representation of phase II of proposed work

Table 2 Confusion matrix for text extraction using SVT dataset

	Predicted Negative (PN)	*Predicted Positive (PP)*
Actual Negative (AN)	214	10
Actual Positive (AP)	10	214

Table 4 Confusion matrix for text extraction using MSRA-TD 500 dataset

	Predicted Negative (PN)	*Predicted Positive (PP)*
Actual Negative (AN)	421	15
Actual Positive (AP)	30	421

Table 5 Confusion matrix for text recognition using SVT dataset

	Predicted Negative (PN)	*Predicted Positive (PP)*
Actual Negative (AN)	222	2
Actual Positive (AP)	10	222

Table 6 Confusion matrix for text recognition using ICDAR 2013 dataset

	Predicted Negative (PN)	*Predicted Positive (PP)*
Actual Negative (AN)	**332**	**3**
Actual Positive (AP)	**15**	**332**

Table 7 Confusion matrix for text recognition using MSRA-TD 500 dataset

	Predicted Negative (PN)	*Predicted Positive (PP)*
Actual Negative (AN)	446	5
Actual Positive (AP)	15	446

In recognition phase if both the actual text and recognized text were same it represents True Positive (TP), If the actual text recognized wrongly means it represents False Negative (FN), if the non-text region recognized has text region means False Positive (FP) and non-text region recognized correctly has non-text region which represents True Negative (TN). In proposed algorithm if actual text region i.e Actual Positive (AP) region recognized correctly at recognition phase which is called as Predicted Positive (PP), if actual non-text region i.e Actual Negative (AN) region recognized correctly has non- text region which is also called has Predicted Positive (PP). If actual text region recognized has non-text region called as Predicted Negative (PN), if actual non-text region wrongly recognized has text region that is also called has Predicted Negative (PN). Performance evaluation of proposed algorithm is calculated in terms of precision, recall, f-measure score has mentioned in eq (2), (3), (4). Precision rate will be high if predicted negative (PN) is less. Recall metrics is used to enhance the precision rate when it does not produce high precision rate. F1 score is used to search for balance for accuracy between precision rate and recall rate, when the Uneven class distribution that is more of actual negatives.

Table 1 depicts the different dataset images used in proposed method along with the no. of images in dataset. ICDAR 2013, SVT, MSRA-TD 500 these are popular dataset repositories which has enormous scene text images which can be found on road side. Those dataset images have skewed blurred, out of focused images. It's really challenging to recognize text from such complex images. Proposed algorithm is designed efficiently to recognize text from complex images. Fig 8 and Fig 9 depicts the pictorial representation of confusion matrix values for phase I and phase II of proposed algorithm. Table 2 depicts the confusion matrix values for phase I for SVT dataset. By splitting and evaluating different dataset images for both two phases can identify the efficiency of proposed algorithm narrowly. Table 3 depicts the confusion matrix values for phase I for ICDAR 2013. Table 4 depicts the confusion matrix values for phase I for MSRA-TD 500 dataset. Likewise, for phase II text recognition, confusion matrix computed for all three dataset separately those values mentioned in Table 5, 6, 7 separately. Fig 10 depicts the pictorial representation of precision, recall; f-measure score for both phase I and phase II and it represents the precision, recall, f-measure values for proposed algorithm. Table 8, 9, 10 depicts precision, recall, f-measure score for different dataset images SVT, ICDAR 2013 and MSRA-TD 500. Table 11 depicts the proposed work comparison with existing methodologies.

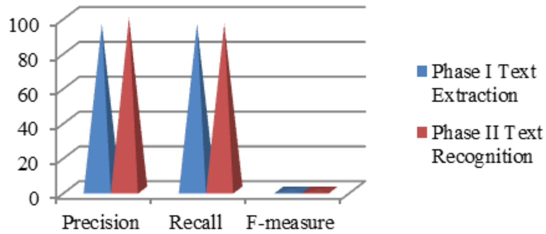


Fig 10. Pictorial representation of Precision, Recall, F-measure values of proposed algorithms

Table 8 Precision, Recall, F-measure values for proposed algorithm for SVT dataset

	Precision	*Recall*	*F-measure*
Phase I Text Extraction	95.5	95.5	95.5%
Phase II Text Recognition	99	95.7	97.3%

Table 9 Precision, Recall, F-measure values for proposed algorithm for ICDAR 2013

	Precision	*Recall*	*F-measure*
Phase I Text Extraction	95.4	94	94.7%
Phase II Text Recognition	99.1	95.6	97.3%

Table 10 Precision, Recall, F-measure values for proposed algorithm for MSRA-TD 500 dataset

	Precision	**Recall**	**F-measure**
Phase I Text Extraction	96.5	93.3	94.8%
Phase II Text Recognition	98.8	96.7	97.7%

Table 11 Comparison with existing algorithms

S. no	*Author's*	*SVT*	*ICDAR2013*	*MSRA-TD*
1	**Proposed algorithm phase I phase II**	**96% 99%**	**94% 96.15%**	**94.8% 97.7%**
2	Cheng et al.[19]	97.1%	93.3%	-
3	Shi et al.,[20]	91.8%	97.4%	-
4	Roy et al.,[21]	-	-	58.41%
5	Jadberg et al.,[22]	90.8%	80.7%	-

9. Time complexity

Phase I: Text Extraction

TC=TC (Pre-processing) +TC (morphological operation) +TC (Image segmentation) +TC (morphological gradient) +TC (text extraction)

TC=TC (1) +TC (1) +TC (1) +TC (1) +TC (1)

TC=ϴ (1)

Phase II: Text recognition

TC = TC (Detect region applying MSER) + TC (Based on geometric properties, remove the non-text area.) + TC (SWV applied image) + TC (Text Recognition)

TC = TC (1) + TC (1) +TC (1) +TC (1)

TC=ϴ (1)

10. Conclusion

In this proposed work text extraction and recognition from scene image were proposed to extract character candidate from scene image. Main purpose of recognizing scene images is to guide person driving via highway, so as to serve emergency and need of driving person. Understanding content present in boarding images plays a vital role in road safety purpose. Extracting text from scene image is intricate; here scene images were taken from standardized familiar standardized repositories such as SVT, ICDAR 2003, MSRA-TD 500. Dataset images composed of various types of images such as blurred, low resolution, affine rotated and Chinese scene images. Our proposed algorithm works better than many existing methodologies in extracting text from scene images.

Proposed work is done in two phase in phase I text is extracted from scene images, in phase II text is recognized from scene images. Scene images have same intensity range for both text and background details, in many images background details dominate text present in scene image in the aspect of gradient intensity. So it's really very hard to isolate text areas from scene images using pruning non-text regions. Phase I work was proposed using following techniques such as incorporating image preprocessing techniques along with morphological operations, and thresholding followed by applying canny edge detector.

In phase II, text from scene images is recognized by applying MSER to detect text region present in the image, based upon geometric properties, non - text region pruned, second filtering approach is done using SWV to prune non-text region, finally using OCR scene image texts were recognized. As a future work images several from multi-lingual benchmarks will be tried to recognize.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

References

[1] Wei, Yuanwang, et al. "Text detection in scene images based on exhaustive segmentation." *Signal Processing: Image Communication* 50 (2017): 1-8.

[2] Zheng, Yang, et al. "A cascaded method for text detection in natural scene images." *Neurocomputing* 238 (2017): 307-315.

[3] Ye, Qixiang, et al. "Fast and robust text detection in images and video frames." *Image and Vision Computing* 23.6 (2005): 565-576.

[4] GonzáLez, ÁLvaro, and Luis Miguel Bergasa. "A text reading algorithm for natural images." *Image and Vision Computing*31.3 (2013): 255-274.

[5] Yi, Chucai, and YingLi Tian. "Text string detection from natural scenes by structure-based partition and grouping." *IEEE Transactions on Image Processing* 20.9 (2011): 2594-2605.

[6] Kumuda, T., and L. Basavaraj. "Hybrid Approach to Extract Text in Natural Scene Images." *International Journal of Computer Applications* 142.10 (2016).

[7] Sahare, Parul, and Sanjay B. Dhok. "Review of text extraction algorithms for scene-text and document images." IETE Technical Review 34.2 (2017): 144-164.

[8] Bhattacharya, Ujjwal, Swapan Kumar Parui, and Srikanta Mondal. "Devanagari and bangla text extraction from

natural scene images." *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*. IEEE, 2009.(4)

[9] Ezaki, Nobuo, Marius Bulacu, and Lambert Schomaker. "Text detection from natural scene images: towards a system for visually impaired persons." *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 2. IEEE, 2004.(5)

[10]Pan, Yi-Feng, Cheng-Lin Liu, and Xinwen Hou. "Fast scene text localization by learning-based filtering and verification." *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010.(7)

[11]Gllavata, Julinda, Ralph Ewerth, and Bernd Freisleben. "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients." *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 1. IEEE, 2004.(8)

[12]Yi, Chucai, and Yingli Tian. "Text detection in natural scene images by stroke gabor words." *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011.(9)

[13]Liu, Chunmei, Chunheng Wang, and Ruwei Dai. "Text detection in images based on unsupervised classification of edge-based features." *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*. IEEE, 2005.

[14] Epshtein, Boris, Eyal Ofek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010.

[15] Li, Yao, and Huchuan Lu. "Scene text detection via stroke width." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.

[16] Borgefors, Gunilla. "Distance transformations in digital images." *Computer vision, graphics, and image processing* 34.3 (1986): 344-371.

[17] Abhishek, L. K. "Thinning approach in digital image processing." *Special Issue-SACAIM* (2017): 326-330.

[18] Wolf, Christian, and Jean-Michel Jolion. "Object count/area graphs for the evaluation of object detection and segmentation algorithms." International Journal of Document Analysis and Recognition (IJDAR) 8.4 (2006): 280-296.

[19] Zhanzhan Cheng, Yangliu Xu, Fan Bai, Yi Niu, Shiliang Pu, and Shuigeng Zhou. Aon: Towards arbitrarily-oriented text recognition. In CVPR, 2018.

[20] Baoguang Shi, Xinggang Wang, Pengyuan Lyu, Cong Yao, and Xiang Bai. Robust scene text recognition with automatic rectification. In CVPR, pages 4168–4176, 2016.

[21] Mishra, Anand, Karteek Alahari, and C. V. Jawahar. "Top-down and bottom-up cues for scene text recognition." *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012.

[22] Max Jaderberg, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Reading text in the wild with convolutional neural networks. IJCV, 2016.

Dr. B. Lavanya works as an Associate professor at the Department of Computer Science, University of Madras, Guindy, Chennai. She received an M.C.A degree from the University of Madras in 2000 and a Ph.D. degree in Computer Science from Bharathiar University in 2015. Her current research interests include Data Mining, Big Data Analytics, Bioinformatics, and Natural Language Processing.

G. Lalitha received M.C.A degree from Ethiraj College for the women, Affiliated to University of Madras, Chennai in 2013 and M.Phil. in Computer Science from University of Madras in 2016. Currently, she is pursuing a Ph.D. from the University of Madras, Chennai. Her current research interests include Machine and Deep Learning , Image Processing , Data Mining and Big Data Analytics.