MSER BASED TEXT EXTRACTION FOR VEHICLE LICENSE PLATES: A TEN-FOLD MODEL

Kannav Dhawan, Akshay Goyal, Ishank Sharma, Akriti Tyagi

ABSTRACT

In our paper, we have developed a novel and efficient TEN-FOLD model to detect, extract and transform the extracted noiseless text from Vehicle License Plates into speech by eliminating the distorted objects appearing like text. We have used "Maximally Stable Extremal Regions" (MSER) feature detector along with geometric properties of image such as Eccentricity, Extent (area region) for text detection and removal of non-text regions instead of commonly used edge or salient based region detector along with stroke width transform, which altogether resulted in increased accuracy, reduced time and space complexity with minimal false positives in the output. Existing algorithms face a problem of confusing certain characters with a letter and vice-versa which was also faced by our model.

Index Terms— Text Extraction, Detection, MSER, Connected Component, OCR.

1. INTRODUCTION

Text Extraction has a wide range of applications ranging from document processing to web crawling. The topic of Text Extraction has always been a major concern for the researchers to dive in due to its application in large scale automated recognition of text and features in Information retrieval systems and geographical systems. Although there has been a substantial research done in the area and despite the inclusion of the deep learning methodology which needs a lot of training data and requires complex data models with expensive GPUs, still there exists a room for more extensive research in the traditional algorithmic approach to improve the accuracy, time and spatial complexity. The traditional Edge and Connected Component based text extraction approaches still lacks the optimal extraction, resulting in noisy outputs. For an instance, even the AdaBoost algorithm, if not trained for the vertical text with less optimal adaptive binarization leaves the hazy text undetected. In our review paper, we went deeper into the various techniques viz. Edge, Connected Component, Texture and Morphological based along with the comparison of algorithms based on precision rate, recall rate, Efficiency and few other factors. This piece of information has helped us in building our TEN-FOLD Text Extraction Model.

The basic purpose of our paper is to develop a novel model for text extraction by using most efficient algorithms for each step involved in extraction process, which gives the most accurate results with minimum number of false positives as compared to the existing algorithms present in the market. In this paper, we have developed a TEN-FOLD novel approach which illustrates the extraction of text from the license plate images using linearly applicable multiple filtering and thresholding techniques followed by the application of MSER feature detector. The removal of the false text candidate regions is done by using geometric properties of the image viz. eccentricity, solidity and area extent which gave the desired output with least false positives when tested on the customized image dataset. Furthermore, we have expanded our model features to develop a more commercially usable and user-friendly model by exporting the extracted text into an audio file (i.e. Text to Speech Output) and including an alert message popping up on the screen indicating the license plate number.

2. RELATED WORK

Many algorithms have been proposed for text detection, text recognition and text extraction in the area of image processing. Previously, most of them have used canny edge detection, connected component analysis and morphological detection. Later the researchers shifted to more efficient technique of Maximally Stable Extremal Regions (MSERs). Peter Tarabek in [1] presented an algorithm for license plate text detection that implements two stage text detection technique using sliding window to label the windows which meets the edge density criteria followed by removing the false positives using the image properties (geometrical and textural). Venkateswarlu and Velaga in [2] designed an algorithm for text detection and extraction from natural images by generating Maximally Stable Extremal Regions (MSERs) then the output is provided to the canny edge detector which maps the text with edges. Further they applied region filtering in order to eliminate the non-text region from the image. Tabassum and Dhondse [3] implemented a two-fold text detection technique by applying MSER detector to detect the regions with variable properties followed by the Canny edge detector for edge improvement and then the Stroke width transform (SWT) was applied to remove the non-text regions but this algorithm was not able to handle the blurred images. In [4] Ullah and Lee had proposed an effective license plate text extraction algorithm which is based on geometrical features and multi thresholding levels. They have used a dynamic image dataset ranging in contrast, sharpness, skew and other properties and successfully attained 75% average accuracy. In [5] Wang et al. designed a novel method with k-shortest paths optimization algorithm to extract text line, where they first detected the MSER from an input image, then took a directed graph approach where different cost functions have been used to identify the cost of all the edges. After this they used k-shortest paths optimization method for extracting text line. They also suggested that further work must be done on robust k-shortest paths optimization with multilingual multi-oriented text extraction. In [6] Islam et al. developed an innovative method for detecting and recognizing text, his method uses enhanced (MSERs) thereby finding regions having identical intensities to get likely text regions followed by unification of canny edges and the MSER regions to find the more accurate text regions. Then, the text candidate regions having large variation in thickness of text characters are removed and the detection of text is done. In [7] Mohamed and

Sanaa had developed a method for 'detecting' text from License Plates where they have implemented a bilateral filter and adaptive thresholding technique to obtain MSER and then deployed a character classifier to eliminate false text candidate regions. Our proposed approach looks like this approach, however in [7] they performed union operation between MSER and thresholding output image to obtain a contour on which filtering and grouping operations was performed whereas in our work we have implemented a linear approach i.e. performed MSER algorithm on image obtained after thresholding and also extended the whole approach for text 'extraction'.

3. PROPOSED WORK

We have developed a novel ten-fold model for text extraction from the license plate images and further rendered the extracted text to form a speech output. The approach uses the following cycle of steps: License plate image as input, color to gray image conversion, gray to negative image conversion, Bilateral filtering of negative image, Adaptive Thresholding of smooth image, object removal whose pixel is less than 100, MSER feature detector for detecting the characters, Non-text region removal by using geometric properties of image, bounding box formation around characters, image resizing and finally using OCR for text extraction. Furthermore, pop up alerts and text to speech conversion is done as additional enhancement in our model for our research. Each processing block and their unification is explained below:

3.1 Image Preprocessing and Segmentation Process:

Being a two-step process, it first converts the colored image to a gray scale or black and white image and then to a negative image to enhance the properties viz. contrast, brightness and transform the image resolution [1].The noise reduction and smoothening of the image is done using bilateral filtering which also helps in preserving the edges of the image from getting distorted by substituting the pixel-level intensity with the average intensity of the surrounding pixels. Adaptive Thresholding is performed to separate and partition the foreground and background into segments which in turn is followed by setting the 'foreground polarity' to categorize the foreground as darker or lighter than the background. We segment the pixels by cleaning the binarized image i.e. by removing the connected components of pixels less than "n" where "n=100" in our case. This process is called AREA OPENING.

3.2 Text Detection Algorithm Implementation Process:

For the purpose of Text Detection, we have used MSER feature detector which is a robust blob detection method to detect the regions with variable properties viz. contrast, sharpness and others. It is an optimal feature detection method to sort the text and the background. After this step, resulted image has certain non-text candidate components which are eliminated with the help of geometric characteristics of image (Eccentricity, Solidity, Extent, Aspect Ratio etc.). We also carried out our research using Stroke Width Transform (SWT) approach, but it resulted in less accurate results with more false positives due to less accurate selection and separation of the text-background regions.

3.3 Text Extraction Process:

In order to recognize the characters in a more optimal way and eliminate the problem of letter-character confusion for elements like 8 and B, we have resized the image by the factor of 2 (which is a generic approach for the whole image dataset). We have constructed the bounding boxes for each detected character which is to be extracted. Finally, extracted the characters using OCR functionality and prompted it on the MATLAB command window as well as on a separate pop up. To add a commercially usable feature, our model converts the popped-up extracted text into speech to make it more user friendly.

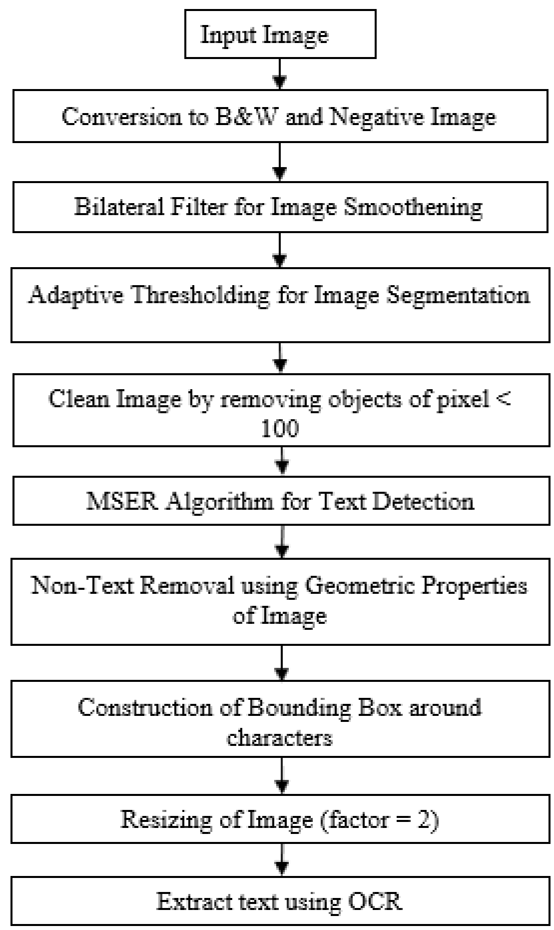


Fig 1: TEN-FOLD Text Extraction Model

4. PERFORMANCE EVALUATION

In order to get the insights of our model and verify the outperformance of our approach, we have tested our model with two different algorithms of text detection. First is MSER algorithm along with geometric properties (Eccentricity, Solidity, Extent, Euler) of the image and then existing models with certain algorithmic variations like models using SWT, for various parameters like time complexity, accuracy and so on. We have also tested our model by implementing different Thresholding techniques for the purpose of Image Segmentation (partition the foreground and background) such as Global or Multi-level Thresholding and Adaptive Thresholding. We have provided the experimental results of various comparisons of different algorithms and techniques that we have performed on our model. We have used MATLAB to code the program and the experiment is conducted on an ASUS Machine having AMD A8 Processor and 6GB RAM.

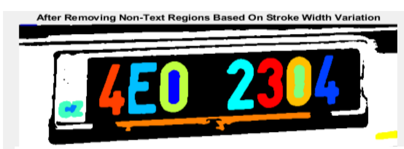
4.1 MSER using Geometric properties vs Stroke Width Transform:

Time Complexity: Our method of using the geometric properties of the image outperforms the SWT approach on time complexity parameters by an average difference of 0.92 seconds which is shown in the results below. Our approach takes average of 0.4032 seconds to process the image whereas the SWT takes 1.33 seconds for the same process.

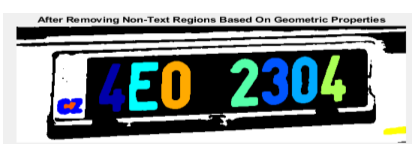
Accuracy: The criteria that we have used to calculate accuracy is:

$$\text{Accuracy} = \frac{\text{No. of correctly detected characters on VLP}}{\text{Total no. of characters on VLP}}$$

Original Image



Using SWT (1.33 sec)



Using MSER with Geometric Properties (0.40 sec)

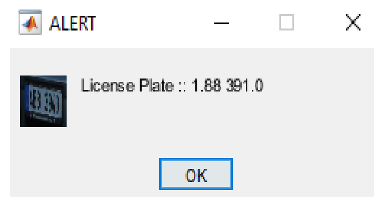
Our approach gives far better accuracy than the one we got using SWT when tested on a big dataset of images.

Parameter	MSER with Geometric properties	SWT
Average Time (in seconds.)	0.40	1.33
Accuracy	92.3%	61.45%

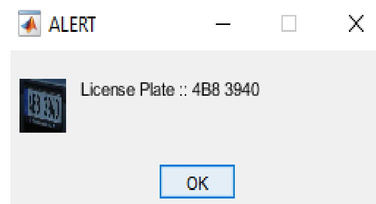
4.2 Adaptive Thresholding vs Global Thresholding:

Time Complexity and Accuracy: While doing the segmentation using adaptive thresholding, we found out that it outperformed the global thresholding in terms of accuracy in segmentation and text detection but lags the global thresholding approach on time complexity parameter by a very small average difference of 0.54 seconds which is shown in the results. Our approach takes average of 0.81 seconds to process the image whereas the global thresholding takes 0.27 seconds for the same image.

Original Image



Using Global/Multi-level Thresholding (0.27 sec)



Using Adaptive Thresholding (0.81 sec)

Parameter	Adaptive Thresholding	Global/Multi-level Thresholding
Average Time (in seconds.)	0.81	0.27
Accuracy	97.5%	76.41%

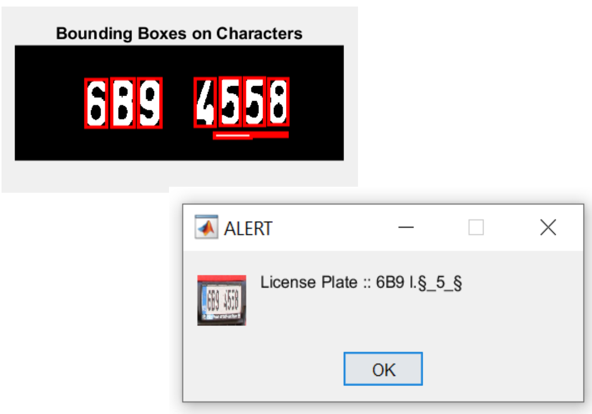
4.3 Bilateral Filter vs Gaussian Filter:

Time Complexity: On using Bilateral filtering for our model, it outperforms the Gaussian filtering approach on time complexity parameter by an average difference of 0.10 seconds which is shown in the results below. Our approach takes an average of 0.67 seconds to process the image whereas the Gaussian filtering takes 0.77 seconds for the same process.

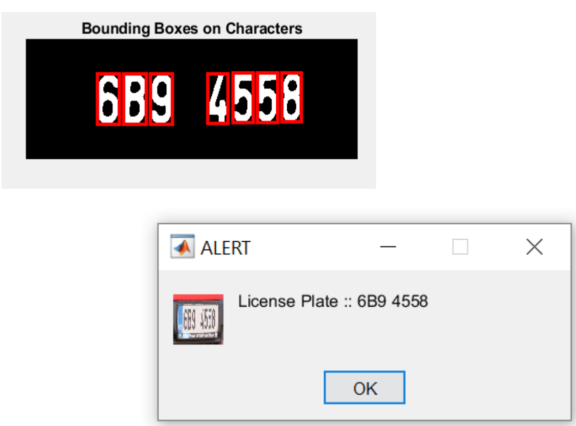
Accuracy: We have used the criteria of accurate bounding boxes for measuring the accuracy of our bilateral filter and the Gaussian filter. Gaussian filtering gives us inferior results on image samples with a constant detection of false bounding boxes in the images which results in the wrong output containing special characters. Whereas, our approach of using bilateral filtering gives us the correct results with the minimum false bounding boxes.

$$\text{Accuracy} = 1 - \frac{\text{Number of falsely identified bounding boxes}}{\text{Total number of correct bounding boxes}}$$

Original Image



Using Gaussian Filter (0.77 sec)



Using Bilateral Filter (0.67 sec)

Parameter	Bilateral Filtering	Gaussian Filtering
Average Time (in seconds)	0.67	0.77

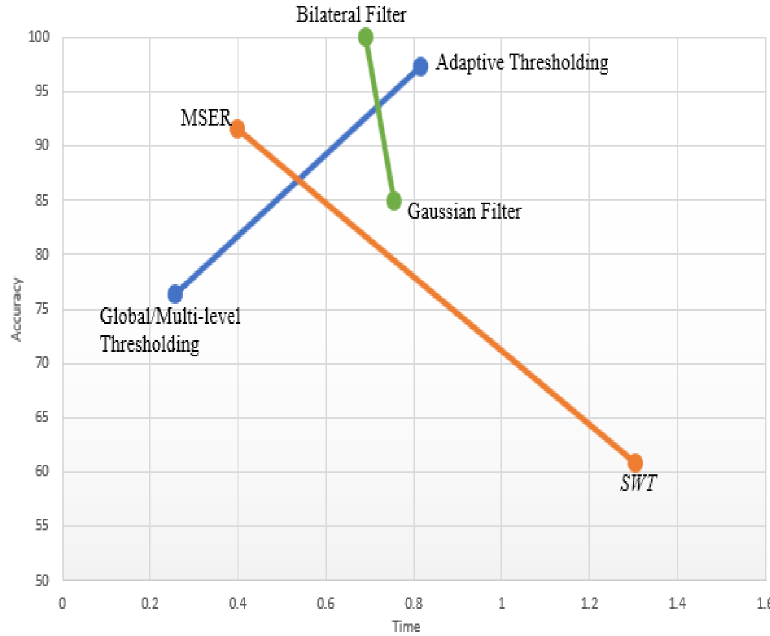


Fig 2: Performance Comparison of Algorithms

5. CONCLUSION AND FUTURE ADVANCEMENTS

We have developed a novel, efficient and fast text extraction algorithm using MSER feature detector for Vehicle License Plates which is tested on a large customized dataset of images with varying contrasts, skew and other properties. We have achieved the optimal results with respect to the time complexity and the accuracy in detection of the text over the other techniques. Our method of using the geometric properties of the image takes an average of 0.4032 seconds to process the image which is far better than the existing techniques resulting in the equally accurate detection. When we consider accuracy as the efficiency parameter, both MSER and adaptive thresholding gave far better results than SWT and Global or Multi-level Thresholding.

As a part of extension to our research we will be extending our approach and will implement the TEN-FOLD model with required advancements for natural (Indoor and Outdoor) images as well as medical images. We are also in queue to add an additional feature of detecting the curved text from images.

6. REFERENCES

[1] P. Tarabek: "Fast license plate detection based on edge density and integral edge image.", Applied Machine

Intelligence and Informatics Conference (SAMI), IEEE 10th International Symposium, 2012.

[2] K Venkateswarlu, SM Velaga: “Text detection on scene images using MSER”, International Journal of Research in Computer and Communication Technology (IJRCCT), Volume 4, Issue 7, pages 45-456, July-2015.

[3] A. Tabassum, S.A. Dhondse: “Text detection using MSER and stroke width transform.”, Fifth International Conference on Communication Systems and Network Technologies, April-2015.

[4] I. Ullah, H.J. Lee: “License plate detection based on rectangular features and multilevel thresholding.”, International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), 2016.

[5] L. Wang, W. Fan, J. Sun, S. Uchida: “Globally optimal text line extraction based on K Shortest paths algorithm.”, 12th IAPR Workshop on Document Analysis Systems, Santorini, Greece, April- 2016.

[6] Md. Islam: “Text Detection and Recognition Using Enhanced MSER Detection and a Novel OCR Technique”, 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2019.

[7] A. Mohamed, S. Fkihi, & R. Faizi: “An Enhanced MSER Based Method for Detecting Text in License Plates”, International Conference on Big Data, Cloud and Applications (BDCA), Kenitra, Morocco, April- 2018.