Traffic Sign Detection and Classification based on Combination of MSER Features and Multi-language OCR

Ali Retha Hasoon Khayeat*
Department of Computer Science, College of Science, University of Kerbala, Karbala, Iraq.
E-mail: ali.r@uokerbala.edu.iq, aliretha@gmail.com

Ashwan A. Abdulmunem
Department of Computer Science, College of Science, University of Kerbala, Karbala, Iraq.
E-mail: ashwan.a@uokerbala.edu.iq

Rafeef Fauzi Najim Al-Shammari
Department of Computer Science, College of Science, University of Kerbala, Karbala, Iraq.
E-mail: rafeef.fauzi@uokerbala.edu.iq

Xianfang Sun
School of Computer Science and Informatics, Cardiff University, Cardiff, United Kingdom.
E-mail: sunx2@cardiff.ac.uk

Received July 03, 2020; Accepted September 05, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V17I2/WEB17040

Abstract

Road signs are so important because they help preserve safe driving conditions; they also influence the safety of drivers and pedestrians. Without these signs, no one would know the driving speed limit, on which direction to drive down a road, any upcoming hazard, or whether they are approaching a merge. It would be chaotic to drive in such situations. Moreover, these signs help new drivers to find their way in the absence of navigators. Therefore, traffic sign recognition takes a critical place in computer vision applications to develop an effective algorithm. In order to tackle this challenge, we proposed the use of Multi-language Traffic Sign Detection and Classification. One of our contributions in this work is that, instead of using the standard grayscale image, we used the RGB colored image. This image is converted into the 2D highest-level grayscale image using the largest values of each pixel in the RGB channels. The novel generated image has the strongest features of the RGB image that make the features distinct and more informative in the classification step. Consider that, in general, the traffic sign has two colors only, the foreground (text location) and background (non-text location). The Maximally Stable Extremal Regions (MSER) used to extract features from the 2D image where the locations of interest are well-identified exclusively by an extremal property of the intensity function in the location and on its outer boundary. The geometrical properties and thinning operations were used to remove the non-text locations. A multi-language OCR was used to understand multi-language. This proposed method has been tested using 240 images which were collected from the Internet and two datasets. The experimental results demonstrated the performance of the proposed method where the traffic sign detected in 92% of the tested images with a very high percentage of localization.

Keywords

Traffic Sign Detection and Classification, Maximally Stable Extremal Regions (MSER), Multi-language OCR.

Introduction

The importance of traffic signs resides its daily use for safety and to reduce on-road accidents. Moreover, the car technology development which creates an urgent need for employing fast and effective algorithms for information retrieval from multimedia content to extract vital information can be used to safeguard people's lives. One of the most effective applications in this scope is text detection in traffic signs, and many studies have been conducted in this area. Localization and detection of text in images can be achieved with different scenarios based on the nature of the sense make these scenarios as a challenge in text detection in the image. These challenges can be divided based on the variability of the text, background complexity and texture when the text has been written.

Variability of the text refers to the font, the number of the color which the text is written in and the size, while the background complexity refers to the blur, noise and distortion in the image. Finally, the texture points to the material when the text has been written such as bricks, grasses and fences. All of these aspects create attractive avenues for new research.

Generally, the paper structure is introduced as follows: section 2 depicts the related work; section 3 explains the proposed method which includes dataset, the channel selection, features extraction, eliminates non-text locations by using the geometric characteristics and the variation of the stroke width and recognizes detected text using OCR; and sections 4 and 5 present the results of the proposed method and discuss the conclusions, respectively.

Related Work

General methods of this work can be categorized into two classes: methods based on text-locations and those based on the texture of the text. In the former class, text detection is based on the difference between the text color and the background; this can be achieved by using an approach of connected component (CC) and edge detection [1]. On the other hand, texture-based methods depend on extracting informative texture features from the content of the image and applying a classification step to detect text region [2].

Various related works have addressed the traffic-sign recognition and detection problem [3] [4]. In 2011, C. Yi and Y. Tian proposed a mechanism for extracting text phrases with random orientations. It is based on text image partitioning and the connected components approach [5]. Yao Li proposed MSER clustering to detect text in sense. Geometric restricts were applied to remove non-text areas. After this, integrating stroke width generated from skeletons of those candidates resulted in the rejection of false positives. In the end, MSERs were clustered into text locations [6].

In the paper [7], the authors used image descriptors to enhance the performance of traffic sign recognition. These features are Histogram of Gradients (HOG) and Soft HoG (SHOG). They reported that SHOG works perfectly when compared to the HOG descriptor.

In recent years, there has been work aimed towards the use of deep learning in text detection [8] because of the efficiency and reliability of these algorithms in classification and recognition. One such work about applying deep learning in text detection is introduced by Tabernik and Skocaj [9]. They adopted CNN to detect sign traffic. The results show that deep learning is an effective method for traffic sign and text detection. Another work using deep learning in traffic signs recognition is produced by Cireşan et al. [10]. The authors used a simple, completely parameterizable GPU implementation of a Deep learning Network which would not require immediate post-wired feature extractors to be precisely constructed and trained rather in a supervised manner.

In this work, we propose the use of MSER and multi-language OCR to detect and recognize traffic signs in a colored image.

Proposed Method

1) Dataset

Figure 1 The 200 categories of the DFG Dataset [9]

We used 20 images from the Internet and 200 images from two datasets (the DFG traffic-sign dataset [9] and the Tsinghua-Tencent 100K dataset [11]) to evaluate our proposed work. The DFG traffic-sign dataset [9] has 200 traffic-sign classes cover 13000 traffic-sign cases in 7000 high-resolution images. This is a novel benchmark dataset for complex traffic signs designed to use along with deep learning; it has a significant number of classes containing sufficient examples to guarantee the learning of deep features. Figure 1 shows the 200 annotations of traffic-sign categories.

The Tsinghua-Tencent 100K dataset [11] has 45 categories with 90000 background images and 10000 images containing at least one traffic sign (see Figure 2).

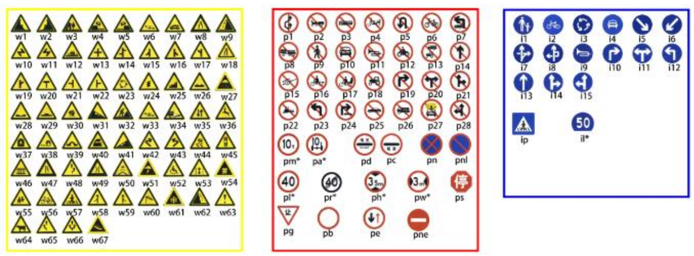

Figure 2 Three classes of the Chinese traffic-sign: the yellow signs are warning, the red signs are prohibitory and the blue boxes are mandatory [11].

2) Channel Selection

Firstly, we found the maximum value of the RGB channels in the colored image and neglected the rest. In other words, for each pixel in the RGB images, the most significant value of the RGB color has been chosen. The resulted image has the highest intensity values from the RGB channels. Experimentally, we found that the use of the MSER with the highest intensity channel can produce a better result than the use of a grayscale image where the MSER can detect the traffic signs more accurately (see Figure 3).

Figure 3 (left) an example of the MSER-detected areas in the grayscale image; (right) an example of the MSER-detected areas in the highest channel.

3) Features Selection

Matas et al. introduced the Maximally Stable Extremal Regions (MSER) [12] which is commonly used to detect the region types due to their high repeatability and their ability to match with many other commonly used detectors [13] [14] [15]. Consider that the standard MSER algorithm uses the flooding simulation algorithm to generate the watershed segmentation [16].

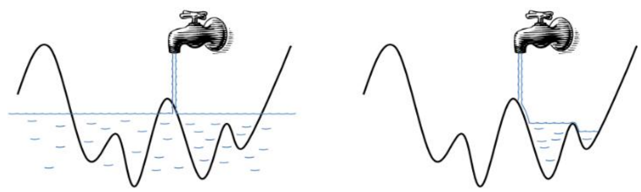

Figure 4 (left) immersion in the standard MSER; (right) immersion in the new version of the MSER algorithm [17]

We extracted the Maximally Stable Extremal Regions (MSER) [17] from the selected image channel. Instead of the standard connected-component algorithm, the immersion analogy was used in the MSER to create various computational scheduling of the pixels. This new computational ordering, which is like a flooding that adjusts to the parts of the grey-level where the pixels visited at any point, has one connected component for the surrounding pixels. Therefore, this version of MSER algorithm [17] needs less memory and shorter time to execute than the standard MSER [12] in which, rather than flooding the image with a similar water level all over the place, the water first tops off the sink where water is poured on and afterwards, overflows to different parts (see Figure 4).

We utilized the MSER detector to locate the traffic sign (text locations) in the image due to the uniform color and the high contrast of text. The non-text areas could be detected together with the text, and we plotted these primary results. Although the MSER algorithm picks out most of the text, it also discovers many other stables non-text locations in the image. We used the text geometrical properties (the aspect ratio, eccentricity, Euler number, extent and solidity) to eliminate non-text locations by employing simple thresholds. Alternatively, we could utilize a machine learning technique to classify text and non-text locations which is time-consuming.

4) Remove Non-Text Locations based on Thinning Technique

Yao Li and Huchuan Lu [6] Stroke width can be applied to differentiate between characters (text) and non-text locations where it measures the width of the lines and the curves that create each character.

Typically, text regions are subject to have little variation in stroke width, however, non-text locations incline toward bigger variations.

5) Understanding Detected Text based on OCR

In order to use the OCR for recognition tasks, the single text characters must be merged to form text lines or meaningful words. We expanded the bounding boxes computed earlier with connected-component analysis to find the neighboring text regions. This obligates the bounding boxes around neighboring text locations to overlap and form a single bounding box around each word.

We used the OCR MATLAB built-in tool [4] to understand the English characters inside each bounding box. Consider that without detecting the text locations, the OCR would generate a considerably noisier result. A Multilingual OCR [18] was used to recognize Multilingual text. Thus, the OCR language data file package (Pre-trained Language Data) was used to customize the standard OCR and added the Arabic and Chinese languages. Moreover, 61 more languages can be recognized using this package.

The proposed algorithm of Detection and Classification Traffic Sign is based on combination of MSER Features and Multi-language OCR.

```
Input: RGBimage
Output: SignImage, OCRtxt
% traffic Sign Image, traffic Sign Image Text
% Create the Max RGB channel image
M,N = size(RGBimage)
for i=1 to M
 for j=1 to N
 maxImage (i,j)= max(RGBimage (i,j,:));
 end
end
% detect MSER regions in maxImage
[MSERreg,MSERcc]= MSER(maxImage,area)
% measure MSER properties
stats=RegionGeometricalProperties(MSERcc)
% Threshold regions
MSERreg(index) = MSERreg(:)>= threshold
% remove small regions
stats(index) = [];
MSERreg(index) = [];
% compute the stroke width image
for k = 1:numel(stats)
 RegionImage = stats(k).Image;
```

SignImage=**morphological**(*RegionImage*,**thin**)
End
% generate B/W mask for the detected image
BWmask = ***binary*** *(SignImage)*
% find the connected text regions
cInd= ***connected_competent*** (*BWmask*);
% merge the detected areas based on the minimum and maximum dimensions
[*xmin,ymin,xmax,ymax*]= *BWmask* (*cInd*)
% create the merged bounding boxes
textBox=[*xmin ymin xmax-xmin+1 ymax- ymin+1*]
OCRtxt = ***OCR*** (*SignImage*, *textBox*);

Experimental Results

As explained in the section under Dataset, we used 40 images from the Internet and 200 images from two datasets (the DFG traffic-sign dataset [9] and the Tsinghua-Tencent 100K dataset [11]) to benchmark the performance of our proposed algorithm.
All RGB colored images are converted into a 2D image using the highest-level channel. The MSER was used to detect the traffic signs locations in the image. The geometric properties of the characters and the width of lines and the curves were used to refine the result and to filter out non-text areas. Finally, the Multilanguage OCR was used to understand the text inside the traffic sign(s) in the image. We implemented all our proposed algorithms using MATLAB 2017b on a 2.8 GHz Windows PC with 8 GB RAM.

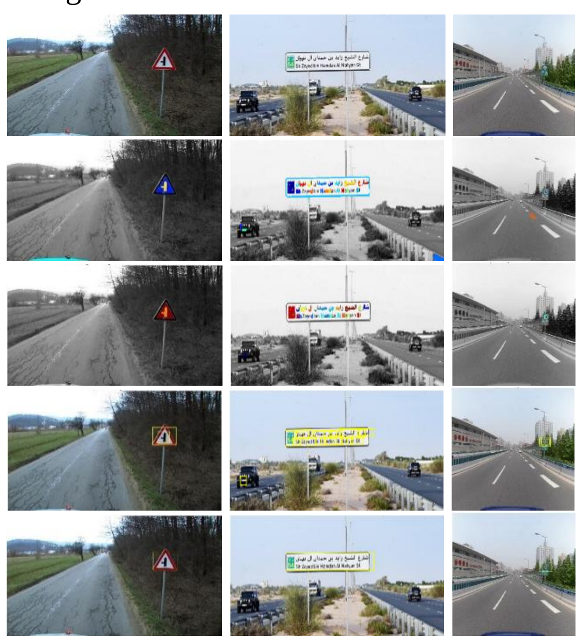

Figure 5 (Top to bottom) 1st-6th rows: input image, MSER regions, after eliminating non-text areas based on geometric properties and thinning operation, expanded the text of bounding boxes and final result

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{2}$$

$$\text{F} - \text{Measure(F1)} = 2 \cdot \frac{(\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \tag{3}$$

In order to calculate the accuracy of the proposed algorithm, we used the F-measure [19] at the image level (see (1), (2) and (3)). The outcomes show excellent performance with different types of scenes and can detect traffic signs with high F-measure (92%) of the tested images with high accuracy of localization (see Figure 5).

According to our information, no one has used the Multilanguage OCR to understand the text inside the detected sign image, (see Figure 6). Therefore, it's unfair to compare our work with that of other people.

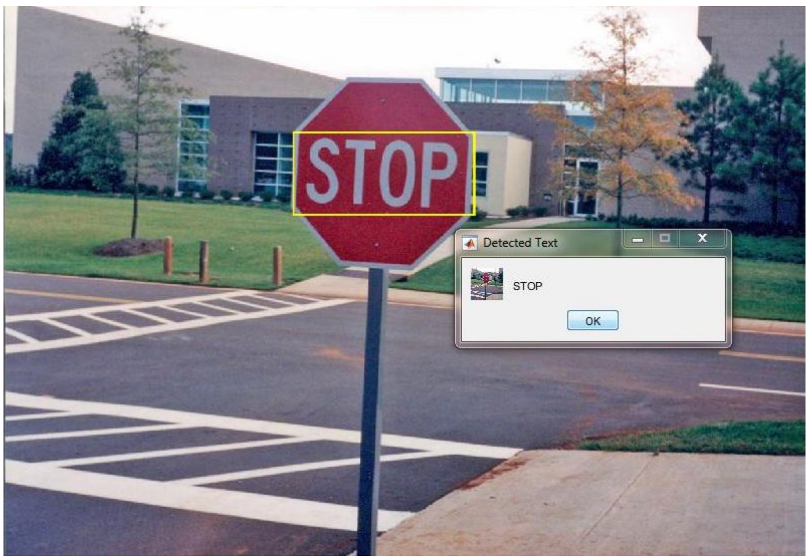

Figure 6 Examples of detect text in the detected traffic sign image

Conclusion and Future Works

In this paper, a novel algorithm has been proposed for traffic sign detection and recognition based on MSER and multi-language OCR. The highest-level grayscale 2D image has been generated which contains the strongest features of the input image. The experimental work illustrates that using the MSER with the highest intensity image can produce better results than using the standard grayscale image. Moreover, the required computation time has been reduced and enhancement has been achieved in the findings by using the following two steps:

Firstly, the use of MSER instead of a machine learning approach to train a classifier to detect and understand the traffic sign(s) which is time-consuming. Typically, a combination of the two methodologies delivers better results, and we will consider this idea in our future work.

Secondly, the use of multi geometrical properties of text to remove non-text areas within simple thresholds. This method removes the non-text (non-sign) region(s) and retains the interest area(s). Followed by thinning operation to detect the character and merge all characters in a text-line(s). Consider that using a multi-language OCR can help the driver to understand the traffic sign in foreign countries.

Acknowledgment

This research was supported by the Computer Science Department, Science College, University of Kerbala, Karbala, IRAQ.

References

Zhu, Y., Yao, C., & Bai, X. (2016). *Scene text detection and recognition: recent advances and future trends.* Frontiers of Computer Science, Higher Education Press, *10*(1), 19–36.

Aher, N. (2015). Text Localization and Extraction in Natural Scene Images. *International Journal on Recent and Innovation Trends in Computing and Communication*, *3*(2), 512–516.

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2012). Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, *32*, 323–332.

Smith, R. (2007). An overview of the Tesseract OCR engine. *In Ninth International conference on document analysis and recognition (ICDAR 2007)*, *2*, 629-633.

Chavre, P.B., & Ghotkar, A. (2015). A Survey on Text Localization Method in Natural Scene Image. *International Journal of Computer Applications*, *112*(13), 15–19.

Li, Y., & Lu, H. (2012). Scene Text Detection via Stroke Width. *21st International Conference on Pattern Recognition (ICPR 2012)*, 681–684.

Kassani, P.H., & Teoh, A.B.J. (2017). A new sparse model for traffic sign classification using soft histogram of oriented gradients. *Applied Soft Computing*, *52*, 231–246.

Arcos-García, Á., Álvarez-García, J.A., & Soria-Morillo, L.M. (2018). Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods. *Neural Networks*, *99*, 158–165.

Tabernik, D., & Skočaj, D. (2016). Deep Learning for Large-Scale Traffic-Sign Detection and Recognition. *IEEE conference on computer vision and pattern Recognition (CVPR)*, 2110–2118.

Cireşan, D., Meier, U., Masci, J., & Schmidhuber, J. (2012). Multi-column deep neural network for traffic sign classification. *Neural Networks*, *32*, 333–338.

Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-Sign Detection and Classification in the Wild. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *48*(10), 2110–2118.

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, *22*(10), 761–767.

Lowe, D.G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Mikolajczyk, K., & Schmid C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, *60*(1), 63–86.

Harris, C., & Stephens, M. (1988). A Combined Corner and Edge Detector. *In Procedings of the Alvey Vision Conference*, 1-23.

Vincent, L., Vincent, L., & Soille, P. (1991). Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(6), 583–598.

Nistér, D., & Stewénius, H. (2008). Linear time maximally stable extremal regions. *In European Conference on Computer Vision, Springer, Berlin, Heidelberg*, 183-196.

Smith, R., Antonova, D., & Lee, D.S. (2009). Adapting the Tesseract open source OCR engine for multilingual OCR. *In Proceedings of the International Workshop on Multilingual OCR*, 1-8.

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *Journal of Machine Learning Technology*, *2*(1), 37–63.