A Survey Paper on Text Detection and Recognition on Traffic Signs

Mr. Raushan Kashypa[1]
II Sem, M.Tech, Computer Science & Engineering
City Engineering College
Bengaluru, India

Mrs. Sowmya Naik[2]
Asst. Professor, Dept. of CSE
City Engineering College
Bengaluru, India

***Abstract*— Traffic sign detection and recognition has been thoroughly studied for a long time. In this survey paper we are talking about a system that will automatically detect and recognize text based traffic signs. It presents a method to detect traffic panels in street level images and to recognize the information contained on them. Google image search is also useful to translate a text based query into images. These automatic detection of text on road sign can help to keep a driver aware of the traffic situation and surrounding environments by seeing and highlighting signs that are ahead and have been passed. Scene structure will be used to define search region within the image, in that image the traffic sign candidates are then found. For locating a larger no of candidates initially saturation value color threshold are used. The larger number of candidates are again reduced by applying some constraints that is based on temporal and structural information. These regions are called Maximally Stable External Region (MSER).There will be two stages involved in decoding a text based board: first stage is recognition stage and second one is interpretation stage. First a candidate region will be selected , a recognition state will find text and these will be grouped into lines before interpreting by optical character recognition (OCR).Recognition accuracy is improved by using temporal fusion of text results across consecutive frames. This method gives an overall $F_{measure}$ of 0.87. Other method that include detecting text based information from a video currently running on a display board ,is also used by the drivers.**

Index Terms*— Sign detection, search region, maximally stable region, recognition, temporal fusion.

I. INTRODUCTION

Text on road signs carries much useful information for driving; it describes the current traffic situation, defines right of-way, provides warnings about potential risks, and permits or prohibits roadway access. Automatic detection of text on road signs,[1] can help to keep a driver aware of the traffic situation and surrounding environments by seeing and highlighting signs that are ahead and or have been passed. The automatic detection and recognition of traffic signs is a challenging problem, with a number of important application areas, including advanced driver assistance systems, road surveying, and autonomous vehicles. While much research exists on both the automatic detection and recognition of symbol-based traffic signs. Without the use of additional temporal or contextual information, there is few information to determine traffic signs from on the fly, while driving, other than basic features, such as shape or color.

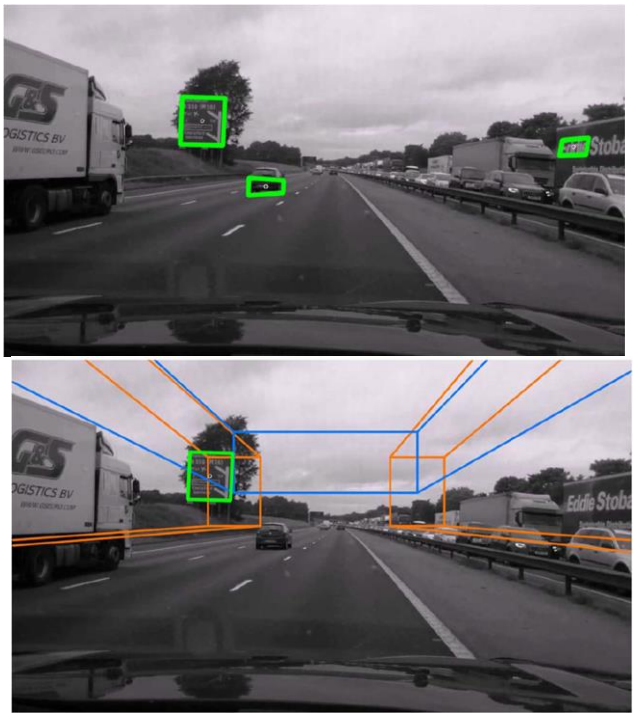

Fig. 1. System output showing detection of traffic signs (top) without and (bottom) with the use of structural and temporal information.

This is demonstrated in the example in Fig. 1, where although the traffic sign present in both images is successfully detected, more FPs[2] are detected by the system (in the top scene) when additional structural and temporal information is not deployed. The proposed system comprises two main stages: detection and recognition. The detection stage exploits knowledge of the structure of the scene, i.e., the size and location of the road in the frame, to determine the regions in the scene that it should search for traffic text signs. These regions are defined once the vanishing point (VP) of the scene and, hence, the ground plane are determined. On this basis, the number of false positives (FPs) likely to occur in a cluttered image, such as a road scene, is high. Potential candidate regions for traffic signs are then located only within these scene search regions, using a combination of MSERs and hue,

saturation, and value (HSV) color thresholding. By matching these regions through consecutive frames, temporal information is used to further eliminate FP detected regions, based on the motion of regions with respect to the camera and the structure of the scene.

A traffic sign recognition system monitors a complex and ever changing environment and must do so accurately and continuously. Here in lies the challenges for this type of system: complex environment, expectation of accuracy, and short response time. Essentially it must identify the road signs that are in view in real time. These efforts to determine the presence of a road sign in real time are complicated by the fact that the environment is continually changing. Road signs will appear significantly different to an artificial (*i.e.*, computer) vision system depending on the amount, direction, and type of light as well as weather conditions. Road signs may also be damaged or tilted confusing an automated system.

II. . RELATED WORKS

Traffic sign detection and recognition using computer vision techniques has been an active area of research over the past decade. A good survey about the main vision-based proposals of the state-of-the-art for Intelligent Driver Assistance Systems can be found in [8], where a discussion about the future perspectives of this research line is there included. Additionally, the work in [10] presents a recent contribution about an intelligent road sign inventory based on image recognition, which is related to the application we propose in this paper but for traffic signs instead of traffic panels and using images taken from a vehicle instead of images served by Google Street View. Lai *et al.* [5] present a sign recognition scheme aimed for intelligent vehicles and smart phones. Color detection is used and is performed in HSV color space. Template based shape recognition is done by using a similarity calculation. OCR is used on the pixels within the shape boarder to determine provide a match to actual sign. The description is purely algorithmic and implemented in software. Andrey *et al.* [2] use a very similar approach involving color segmentation and shape analysis. Histograms, however, are used as the shape classification method after connected regions are labeled. Actual sign recognition is done via template matching by using a weighted direct comparison of the interior portion of each shape to templates.

On the other hand, [1] proposes a method to detect text on traffic panels from video. Firstly, regions of the same color are extracted using a k-means algorithm and traffic panels candidates are detected by searching for flat regions perpendicular to the camera axis. The orientation of the candidate planes are estimated using three or more points in two successive frames, so this method needs an accurate tracking method to detect corresponding points in successive frames. Further, a multiscale text detection algorithm is performed on each candidate traffic panel area. The text detection method integrates edge detection, adaptive searching, color analysis using Gaussian Mixture Models (GMM) and geometry alignment analysis. A minimum bounding rectangle is fitted to cover every detected text line. A feature-based tracking algorithm is then used to track all detected areas over the timeline as they are merged with other newly detected texts in the sequence. Finally, all detected text lines are extracted for recognition, but the authors do not comment how the recognition is carried out. In terms of text detection, this method provides good results under different lighting conditions and it is not affected by rotations and projective distortions. It achieves an overall text detection rate of 89% in their own dataset, which is not publicly available.

In addition to being a topic of active academic research, Traffic Sign Recognition is also a technology that is being researched and implemented in the industry. This technology is developed by many car manufacturers who are partnering with traditional automotive suppliers such as Continental Automotive, TRW, Bosch [6] and newer image recognition software product developers like Ayonix. Continental Automotive developed several products for traffic sign recognition. Its Multi-Function Camera specification details its abilities for use in TSR [4]. This traffic sign recognition system [9] began production in 2010 on the BMW 5-series. In addition to BMW, many other carmakers have rolled out some version of this technology. Volkswagen has done so on Phaeton and the Audi A8. Mercedes-Benz E and S class both have an implementation of TSR. As well as the Saab 9-5, Opel, Insignia, and the European 2011 Ford Focus. Additionally, Google has developed technology that allows a vehicle to drive itself. Using a combination of data stored in its map database and data that it collects from its environment in real-time, the Google Car is able to safely navigate complex urban environments.

III. SYSTEM MODELS

A. Hue Calculation and Detection

Traffic signs consist of solid color text, symbols, or shapes on a solid color background as seen in **Figure 2**. Scanning an image looking for this color signature will allow for the quick identification of possible traffic signs and the rejection of the remaining parts of the image.

Stop, *yield*, *do not enter*, *wrong way*, and prohibition signs such as *no left turn*, all contain red backgrounds with white text or white backgrounds with red text. Main distinguishing color for these signs is red. Similar groupings can be done for signs that are primarily green, yellow, blue, or black and white. The algorithm described here and in the sections following must be performed for each of these color groupings.

Figure 2. Example traffic signs (red).

For each group of similarly colored signs, the algorithm begins by scanning the image to calculate the *hue* of each pixel. There are a variety of ways in which to express the color of a pixel. Perhaps the most common is by using the color's primary color components or the RGB value. Although this is very useful when displaying that color, it is not as helpful when trying to extract all the pixels of a specific hue. If, as in this case, the desire is to identify all the pixels that would be considered red, there are colors that contain a significant about of red as a primary color contributor, but are themselves not red. The color yellow is one such example. **Figure 3(a)** depicts the color representations using RGB. To determine the hue of a pixel, or its color regardless of shade, a conversion must be made. Each RGB pixel is converted to a different triplet called HSV: Hue, Saturation, and Value. HSV represents the color spectrum by having a value for the color (hue), the amount of that color (saturation), and the brightness of that color (value). The hue parameter represents the angle where the pixel's color lies on the cylinder depicted in **Figure 3(b)**. Thresholds can be chosen to categorize any hue value found. Once this conversion is made, the hue values for each pixel can be scanned. Detection is the process of identifying the pixels whose hue value falls between the thresholds for the relevant color. This will split the image into two categories: pixels that have the hue of interest and those that do not. At this point the full color image being processed can be simplified into a binary image. Active pixels had the desired hue while inactive pixels do not. This step is called *Hue Detection*.

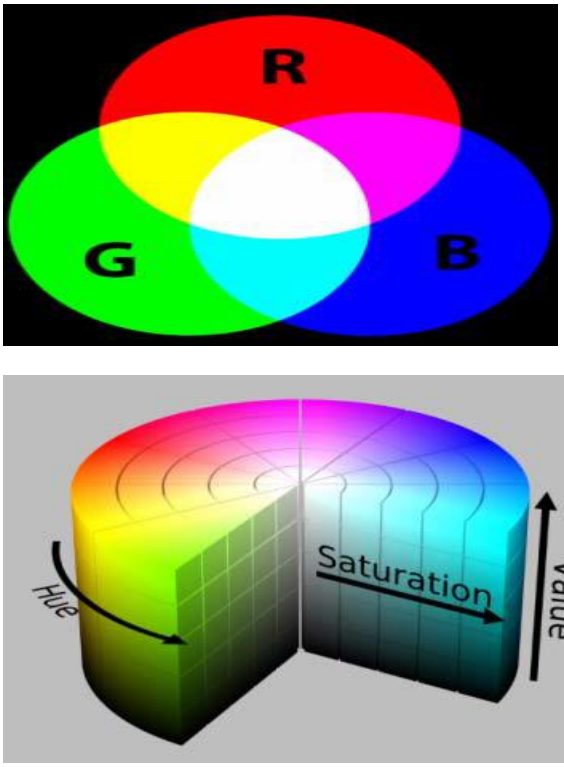

Figure 3.a (top) RGB color space, 3.b (bottom) HSV color space.

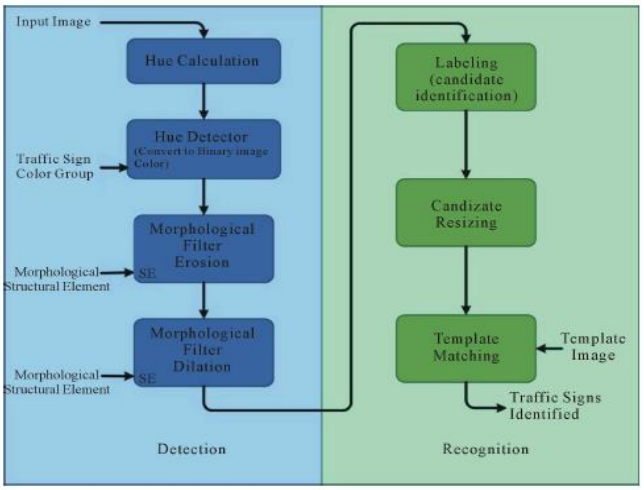

Figure 4. Algorithm for detection and recognition

B. Detection and Recognition

The first stage of the proposed system detects candidates for text-based traffic signs. This consists of three phases: determination of search regions (regions of interest where the text sign is expected to be found), detection of all possible candidates within these regions, and reduction of candidates using contextual constraints. Search regions of interest for traffic signs are found within the image, by first locating the sides of the road in the image and then defining 3-D search boxes, which are projected back onto the original 2-D frame. These search regions are shown in Fig.5, where the orange region is for traffic signs on either side of the road, and the blue box is for overhead gantries.

Figure 5. Camera position and captured frame.

The second stage of the system recognizes text contained within the detected candidate regions. To increase the chances of OCR in recognizing our noisy text regions, we first apply an approximate perspective transform to the rectangular candidate regions to vertically align them and their text characters. Individual text characters are then segmented, formed into words, and then sent to OCR. Results from several instances of each traffic sign are then combined, in order to further improve recognition. The steps involved in this stage are given as follows:-

i. Correction of Detected Candidate Region.
ii. Detection of Text Lines.
iii. OCR for Individual Candidate.
iv. Temporal Fusing of OCR Results.

Before text is read from the detected region, an approximate perspective transform is applied to vertically align the text characters and reduce perspective distortion. 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 noncharacter CCs and hence improving the chances for higher OCR accuracy. The set of detected text lines (in grayscale) are passed on to the open-source OCR engine for recognition. To improve the accuracy of OCR, results are combined across several frames.

C. Incremental Spatio Temporal Text Detection

Some previous research work has paid particular attention to detecting and recognizing symbols on road signs, particularly warning signs such as "STOP," "YIELD," and "DO NOT ENTER." Since only a finite number of shapes and colors can be applied on these warning signs, color and edge-based shape features are normally used to train the detector [4]. In this work, however, we are interested in detecting not only symbols, but also text, on road signs. Text appearing on road signs can have a variety of appearances. Color and shape features are not enough to train a robust detector. Without knowing text on the signs, drivers cannot obtain correct information about current traffic situation and appropriate driving instructions.

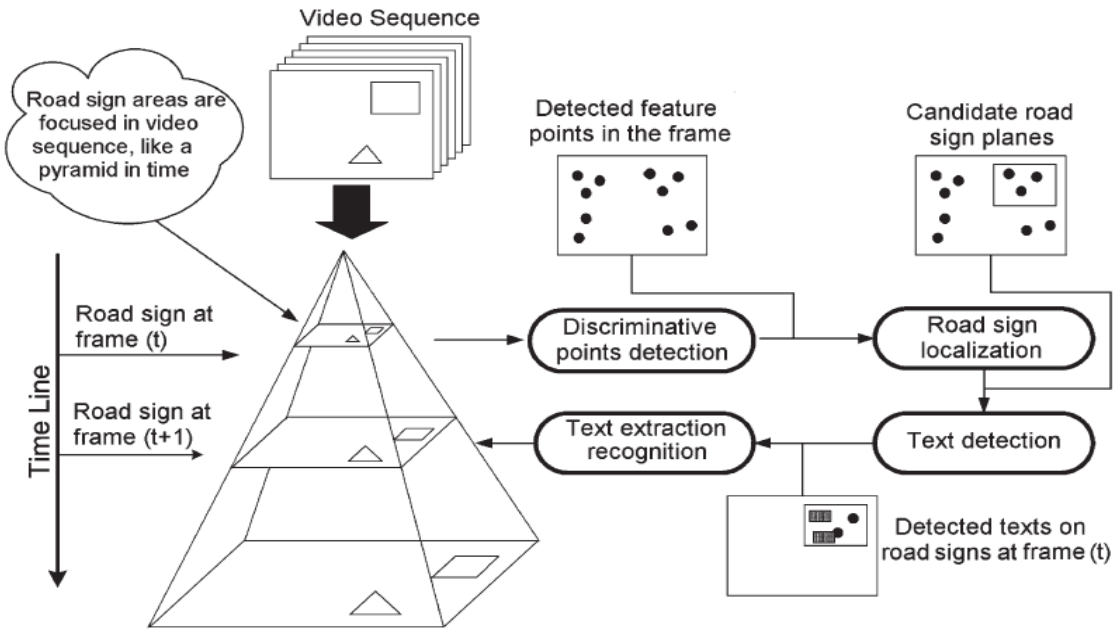

Figure 6. Architecture of Proposed Framework

Accurate real-time sign detection with few false positives is an essential requirement for the proposed framework to improve the safety and efficiency. The proposed framework considers the whole period of appearance of a road sign in a video as a pyramid of sign image patches along the time line. Fig. 6 shows the architecture of the framework from which four main steps are summarized as follows:-

1) Discriminative point detection and clustering—detect discriminative feature points in every video frame using the algorithm and partition them into clusters.
2) Road sign localization—select candidate road sign regions corresponding to clusters of feature points using a vertical plane criterion.
3) Text detection—detect text on candidate road sign areas and track them.
4) Text extraction and recognition—extract text in candidate sign plane for recognition given a satisfactory size.

In step 1), a number of discriminative points are selected in the current frame and are clustered using local region analysis. Then, the road sign localization step detects candidate road sign areas from the point clusters.

Finally, all detected text lines are extracted for recognition, given a satisfactory resolution in the text extraction and recognition step.

IV. PERFORMANCE

In this paper three purposed systems have been comparatively studied and some of the others are referred. The performance of the first proposed system is basically related with the timing. The key requirement for a traffic sign recognition system is that it should operate in real application, real-time may be defined very differently. The term real time refers to a system that is able to take data and process it sufficiently rapidly to able to take the action required of the system. For a system whose purpose is to detect traffic sign while travelling on the road, a threshold for real time could be calculated as the time required for detecting a sign in 50 feet of travel at 65 mph. This time is 525 millisecond. Therefore a system of this nature can be considered as real time if it can reliably detect a traffic sign in approximately half a second.

The second purposed system was based on combining the detection and recognition stages. So the performance can be also be discussed around the two stages. To evaluate the performance of the detection stage of our system, comparative analysis was performed against two these were the methods proposed by Reina *et al.* [3] and González *et al.* [7]. Since both methods were designed to recognize Spanish road signs, which are blue and white, it was necessary to adapt the algorithms to detect U.K. road signs, which also feature green and brown backgrounds. The method of González *et al.* [7] detected blue road signs as MSERs in the blue channel of a normalized red, green, and blue (RGB) image. Therefore, to extend this to green road signs, MSERs were also detected in the green channel. Brown road signs were detected as dark-on-light MSERs in a grayscale frame. The method of Reina *et al.* [3] uses hue, saturation, and intensity (HSI) thresholding to find candidate blobs for blue road signs. Therefore, additional thresholds were added to their method for the detection of brown and green road signs. These algorithms were optimized using the same validation data set used to develop our proposed method. To evaluate the performance of the recognition stage, ***Precision, Recall***, and $F_{measure}$ were computed based on the number of individual words correctly classified. For a word to be considered a TP, all characters must be correctly recognized in the correct letter case. If a single character is recognized incorrectly, then the entire word is considered to be an FP. Symbols such as "airport" were included in the training set merely to avoid their misclassification as characters, and are therefore classified as true negatives (TNs) when recognized, and have no effect on the result. There are 15 of these symbols in total, examples of which include directional arrows and the airport symbol.

The third proposed system used the concept of the incremental spatio temporal text detection. The proposed framework has been evaluated through experiments with a large and diverse set of road sign video sequences. From a video database of 3-h natural scene videos captured by a digital video (DV) camera mounted in a moving minivan, we selected 22 video sequences with different driving situations including different road conditions (straight, curve), vehicle speed (low, high), weather conditions (sunny, cloudy), and daylight variations. The objective of the selection was to be as diverse as possible and cover the range of difficulty as well as the generality of the task. Thus, we did not include the extreme cases, such as crooked lateral signs. Each video sequence is about 30 s, contains an average of 92 road signs, and 359 words (including numbers such as a speed limit), and has a frame size of 640 × 480.

V. CONCLUSION AND FUTURE WORKS

This paper presented a comparative analysis on detecting the text based information from a text board or from a video display and interpret it accordingly. A novel system for the automatic detection and recognition of text in traffic signs based on MSERs and HSV thresholding has been proposed. The search area for traffic signs was reduced using structural information from the scene, which aided in reducing the total number of FPs. Perspective rectification and temporal fusion of candidate regions of text were used to improve OCR results. Both the detection and recognition stages of the system were validated through comparative analysis, achieving the $F_{measure}$ of 0.93 for detection, 0.89 for recognition, and 0.87 for the entire system. This paper presented a real application of the text detection and recognition algorithm including some adaptations and new functionalities. It consists in reading the information depicted in traffic panels using panoramic images downloaded from the Google Street View service. The main use of this application is to automatically create up-to-date inventories of traffic panels of whole regions or countries. This information is very useful for supporting road maintenance and for developing future driver assistance systems. The proposed framework efficiently embeds road sign plane localization and text detection mechanisms with feature-based tracking into an incremental detection framework using a divide-and-conquer strategy. This strategy can significantly improve the robustness and efficiency of text detection. The new framework has also provided a novel way to detect road sign text from video by integrating image features and the vertical plane assumptions of road signs.

Interesting future work may include detecting variable message signs from video and exploring other robust image and video features for text detection. Finally, the recognition of the information depicted in the traffic panels was done frame by frame. Typically, a panel appeared in several consecutive frames. As future work, we intend to do a multi-frame integration of the recognized information at each single frame. In addition, the use of the a priori knowledge that we know about the design of traffic panels would improve the recognition rates, because certain objects, especially symbols and numbers, are located only at certain parts of the panels.

REFERENCES

[1] W. Wu, X. Chen, and J. Yang, "Detection of text on road signs from video," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 4, pp. 378–390 Dec. 2005.

[2] Andrey, "Automatic Detection and Recognition of Traffic Signs Using Geometric Structure Analysis," *Pro- ceedings of SICE-ICASE International Joint Conference*, Busan, 18-21 October 2006, pp. 1451-1456.

[3] A. Reina, R. Sastre, S. Arroyo, and P. Jiménez, "Adaptive traffic road sign panels text extraction," in *Proc. WSEAS ICSPRA*, 2006, pp. 295–300.

[4] Continental AG, "MFC 2 Multi-Function Camera," Data- sheet, 2009. http://www.conti-online.com/generator/www/de/en/continental/industrial_sensors/themes/mfc_2/mfc_2_en.html

[5] C. Lai, "An Efficient Real-Time Traffic Sign Recognition System for Intelligent Vehicles with Smart Phones," *Pro- ceedings of* 2010 *International Conference on Technolo- gies and Applications of Artificial Intelligence*, Hsinchu, 18-20 November 2010, pp. 195-202

[6] Frost & Sullivan, "Development of Low-cost DAS Tech- nologies to Help Reach European Union's Target to In- crease Road and Driver Safety," 2011. http://www.frost.com/prod/servlet/press-release.pag?docid=251082001

[7] A. Gonzalez, L. M. Bergasa, J. J. Yebes, and J. Almazan, "Text recognition on traffic panels from street-level imagery," in *Proc. IVS*, Jun. 2012, pp. 340–345

[8] A. Mogelmose, M. Trivedi, and T. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, 2012