Design an Intelligent Driver Assistance System Based On Traffic Sign Detection with Persian Context

Alireza Khodayari*1, Mehrdad Yousefi3

1Assistant Professor, Mechanical Engineering Department, Pardis Brunch, Islamic Azad University, Tehran, Iran
3M.Sc. Student, Mechanical Engineering Department, Pardis Branch, Islamic Azad University, Tehran, Iran

mojallali@guilan.ac.ir

Abstract

In recent years due to improvements of technology within automobile industry, design process of advanced driver assistance systems for collision avoidance and traffic management has been investigated in both academics and industrial levels. Detection of traffic signs is an effective method to reach the mentioned aims. In this paper a new intelligent driver assistance system based on traffic sign detection with Persian context is designed. The main goal of this system is to assist drivers to choose their path based on traffic signs more precisely. To reach this purpose, a new framework by using of fuzzy logic was used for detection of traffic signs in videos which have has been captured from a vehicle path in highways. Fuzzy logic increases inference and intelligent capabilities in smart systems to make correct decision making in online conditions. Then, the combination of Maximally Stable Extermal Regions (MSER) and Canny Edge Detector Algorithms are used to detect road sign's texts detection. MSER algorithm is aimed at assists to detect regions in an image that differ in properties, for example in brightness or color, compared to surrounding regions. Also, canny edge detector uses a multi-stage algorithm to detect a wide range of edges in the images. Thereafter, morphological mask operator is used to join individual characters for final detection of texts in the traffic signs. Finally, MATLAB Optical Character Recognition (OCR) is employed to recognize the detected texts. This new framework gives an overall text detection and recognition rate of 90.6%.

***Keywords**: road sign detection, text detection, object detection from video, fuzzy logic, MSER..*

Introduction

Automatic detection and recognition of road sign's texts from videos is an effective task for advanced driver assistance systems. There have been extensive investigation efforts in the detection and recognition of road sign's texts until now [1], [2], [3]. Nevertheless, in this paper, the concentration is on the function of automatically detecting and recognizing of road signs from videos which are captured from vehicles path. The results of this research can be applied in the design of driver assistance systems.

Texts on road signs contain important information for the vehicle's path in highways. Detection of these texts can help drivers to a great extent to be aware of traffic situations. Furthermore, it can be added.to navigational systems such as global positioning systems (GPS) to help drivers with more efficient driving performances. Also, this system can be equipped by sound massages to have a better interaction with drivers [4], [5]. Road sign recognition systems usually are developed into two specific phases [6], [7]. The first phase is normally related to the detection of traffic signs in a video sequence or image using image processing. The second one is related to recognition of these detected signs.

More general techniques for detecting scene text from still images have been developed in pattern recognition and computer vision fields. Probably the most common approach is using some form of Hough transform. Approaches based on corner detection followed by reasoning or approaches based on simple template matching are also popular. Samples of these approaches are proposed in [8-10]. Also, some approaches use their prior knowledge of the problem (like the expected color and shape of a traffic sign) as a machine learning method for traffic sign recognition as used in [11].

Recently, several researchers were able to detect scenes [12], [13] and they reported that edge features can better handle lighting and scale variations rather than texture features in scene images [14], which are often used for detecting texts in news videos [15]. Inspired by their work, the edge-based feature is similarly used for text detection in this study. Myers et al. described a full perspective transformation model to detect three-dimensional (3-D) deformed text from still images [16].

To perform this action, the first step is to capture videos from the path which is in front of the vehicle. The designed system should detect texts on road signs to help drivers for managing the road ahead. Fig. (1-a to d) illustrates four examples of road signs.

Detecting texts on road signs raises many challenges. First of all, video images have relatively low resolution, and in addition, their environmental noises are effective on the results. Also, weather and lighting conditions are uncontrollable, and these all are added up to the problem of including the darkness of the path and sunlight reflectivity. In this paper, a new framework is designed to resolve the mentioned issues.

This framework consists of capture preprocessing, denoising, road sign detection from the background, text detection from the road signs and recognition of them. This new framework is also applicable for text detection of every other international languages.

The rest of this paper is organized as follows; section (2) describes the design of an intelligent driver assistance system and its framework. Afterward, in section (3), the experimental results of the designed system in usual conditions are presented. At the end, this paper is concluded in section (4).

Fig1. Example of road signs in different situations including different lighting conditions, weather and highlights.

Design an Intelligent Driver Assistance system

Principles of the presented system can be summarized in the following steps: Design of the road sign detector, Detection of signs on images, Text detection and then Text recognition. The flowchart of the system is shown in Fig. (2).

Preprocessing

To have a more accurate process, image space is mapped from RGB to HSV. Some specifications of HSV in respect to RGB make it more applicable to be used in text detection algorithm [17]. In HSV and HIS spaces, lighting variations only affects the intensity and have no effect on the hue that must be processed. Moreover, since color segmentation on HSV space is done on 1D of color component, it has much lower computational complexity regarding to RGB space (in which requires the same process being done on a 3D space). [18], [19].

Construction of Fuzzy Images

In fuzzy image, features of an each pixel such as color and lighting were presented by linguistic values. White, pale green, dark red, yellow-green, dark blue, are samples of verbal values that were allocated to each pixel. To transfer an image to fuzzy space, fuzzy logic was used. Fig. (3), shows a color image and its identical fuzzy image, where DB, B, PB, PR and R represent dark blue, blue, pale blue, pale red and red, respectively.

In this research, red, blue, green, yellow and black were used and other colors were deleted, and instead of others white color was used. For each pixel, three specifications of color, saturation and value were applied as inputs to fuzzy sets. Membership functions of these specifications are shown in Fig. (4).

The constructed fuzzy image needed the denoising process. Denoising In this process, the color of each pixel was compared by the color of its eight neighbors, then the decision was made about the necessary change for its color. Using this algorithm, diffused colored and undesirable pixels were corrected. However, the denoising process may destroy useful information of an image and as a result, smaller road signs are created which makes the road sign detection harder from further distances.

Edge Detection and Labeling

In this step of process, pixels were classified in to different sets. Useful sets consisted of homochromatic pixels and their neighbors. To present the pixels of a specific set and to separate it from other sets, a specific code showing the set number was dedicated to each pixel. This algorithm is called labeling.

For this algorithm, the fuzzy image shown in Fig. (5) was used. Fuzzy image indicates that the values related to each pixel are verbal values consist of red, green, yellow, blue and black. After the labeling process, the edges of the initial image were extracted by Sobel edge detection algorithm [20]. Then, the edge pixels were eliminated from the fuzzy image.

In the labeling process, all pixels were investigated from left to right and top to bottom. If the color of the pixel is white or either labeled before, new label wasn't used for it. Otherwise, a new label was declared to it and was considered as the start point of the labeling. Then, labeled pixel and its four neighbors were applied as the inputs of the fuzzy logic system. Each of the neighbors of the labeled pixels which had the same color, was labeled as a member of the set and was then saved. Fuzzy rules for the labeling process are as below:

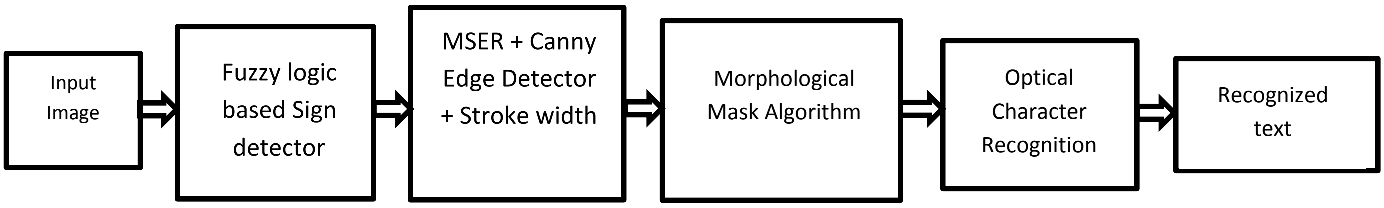
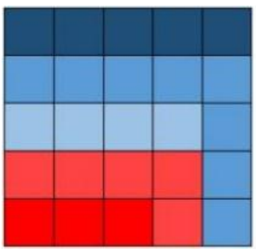


Fig2. Flowchart of the presented system



→

DB	DB	DB	DB	DB
B	B	B	B	B
PB	PB	PB	PB	B
PR	PR	PR	PR	B
R	R	R	PR	B

Fig3. Fuzzification of an image.

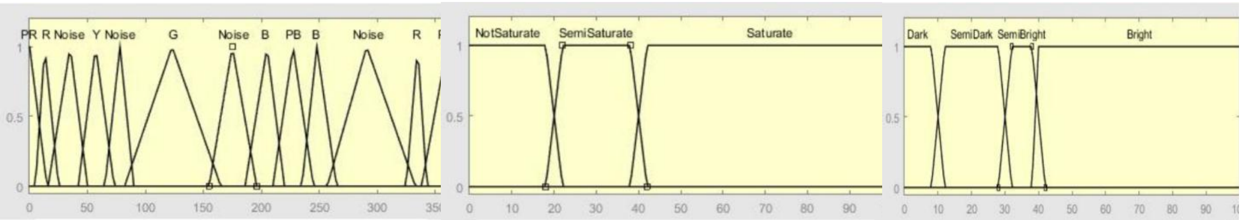


Fig4. Membership functions for inputs.

If	P(x+1,y) == P(x,y)	Then	save P(x+1,y) as label pixel
If	P(x,y+1) == P(x,y)	Then	save P(x,y+1) as label pixel
If	P(x-1,y) == P(x,y)	Then	save P(x-1,y) as label pixel
If	P(x,y-1) == P(x,y)	Then	save P(x,y-1) as label pixel

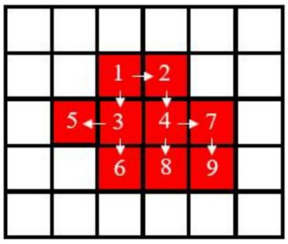


Fig5.. Labeling process.

Fig6. Example of denoising and edge detecting processes.

Edge Detection and Labeling

In this step of process, pixels were classified in to different sets. Useful sets consisted of homochromatic pixels and their neighbors. To present the pixels of a specific set and to separate it from other sets, a specific code showing the set number was dedicated to each pixel. This algorithm is called labeling.

For this algorithm, the fuzzy image shown in Fig. (5) was used. Fuzzy image indicates that the values related to each pixel are verbal values consist of red, green, yellow, blue and black. After the labeling process, the edges of the initial image were extracted by Sobel edge detection algorithm [20]. Then, the edge pixels were eliminated from the fuzzy image.

In the labeling process, all pixels were investigated from left to right and top to bottom. If the color of the pixel is white or either labeled before, new label wasn't used for it. Otherwise, a new label was declared to it and was considered as the start point of the labeling. Then, labeled pixel and its four neighbors were applied as the inputs of the fuzzy logic system. Each of the neighbors of the labeled pixels which had the same color, was labeled as a member of the set and was then saved. Fuzzy rules for the labeling process are as below:

Afterwards, the mentioned steps were repeated for the next pixel. This process continues for all the pixels of the set. At the end, the process gets back to the labeling start point, and at this point, all the pixels have been analyzed one by one.

In the labeling process, due to the definition of thresholds, sets of pixels which have very low or very high number of members, will be deleted. The reason is that in these sets the possibility of existence of any information about road signs is very low. As the number of labels decreases, the next necessary processes, will occur with more computational speed.

So far, the operations which should be implemented on images, are based on the color specification of road signs. In addition, each road sign has its own special colors and geometrical shapes. These two specifications rarely come together in urban environments. At this point, by finding out the same color and interconnected parts of the image, that are usually representative of a special shape, the location of the road signs in the image can be detected. Fig. (6) shows an example of denoising and edge detecting.

2-4. Invariant features

One of the methods of detecting objects in images, is to exploit its features [21]. Features are parts and unique properties that are representative of the intended object. By finding these features in an image, the existence of the object can be concluded. Features of image objects such as color, size, angle and intensity are always changing. Therefore, features should be chosen in a manner that they can be detected in case any variations occurred. Generally, features were classified in two types: local and global [22].

Local features refer to special parts of the object. Each pixel of the image, and its near neighbors, introduce special specification. Usually, those parts of the image which show different specifications are chosen as our desired features. Common specifications consist of intensity, color and structure. Points, edges or small parts of the image can be considered as local features. Then, measurement of a main part, considered as a local feature, was performed and the related modifiers were created. These modifiers can then be used in different applications.

Global feature is a model in which the whole parts of an image can be seen in a moment. For example, the histogram of the image can be a global feature. Global feature applications are image retrieval and object detection and have acceptable performance within low color variation images. Because of this property, users paid more attention to all the combinations of the considered images. However, by using global features, the foreground region cannot be detected from the background region, and besides, it cannot combine their information. Therefore, images with too many details and the images that are parts of the object were blocked and they are the main challenges.

To eliminate the limitation of global features, images are divided into several parts in a way that each one of these parts represents an object or a part of it. Generally partitioning of an image needs high perception of that image and it is a big challenge for this method.

A method named “Sampled Features” can remove the disadvantages of the global features and image partitioning. In this method, comprehensive examples from different parts of the image in different positions and sizes can be prepared. Then, for each part of the image, global features are extracted. Also, in this method, sliding window was used. Sampled features have various applications, special objects or specific sets of objects such as pedestrians or vehicles can be called.

Samples features focuses on sampled parts and searches for their likeness to model samples. But, on the other hand, for items which are parts of the objects which have been blocked, this method cannot be used for detection. Fig. (7). shows the result of detection of one of the signs in the capture.

Generally, three algorithms were used in text detection process which consists of: 1- MSER algorithm, 2- Edge Detection, and 3- Stroke Width Transportation

Fig7. Detecting one of the signs in the capture.

2-5. MSER Algorithm

This algorithm detects text specifications such as distinguished letters. Several algorithms were presented to detect this specification. Distinguished letters refers to the points in an image that have penumbra feature and specific geometrical shape. Also, these points have the most continuity in an image. This algorithm finds a set of distinguished regions in a gray space. MSER algorithm is used to extract external regions of the gray-scale image. Maximally, stable external regions are used as a method of text detection in images. MSER extracts a number of covariant regions in image. The relationship between these two objects always remains the same. External word means that the pixels which are inside the MSER have different intensity relative to the other pixels [23].

Image I is a mapping$I : D \subset Z^2 \rightarrow S$. Extremal regions are well defined on images if:

1. S is tottaly ordered i.e. reflexive, anitisymmetric and transitive binary relation ≤ exists. In this paper only $S = \{0,1, \dots ,255\}$ was considered, but extremal regions can be defined on e.g. real-valued images ($S = R$).

2. An adjacency (neighborhood) relation $A \subset D \times D$ is defined. In this paper 4-neighbourhoods were used, i.e. $p, q \in D$ are adjacent (pAq) iff $\sum_{i=1}^{d} |p_i - q_i| \leq 1$.

Region Q is a contiguous subset of$D$, i.e. for each $p, q \in Q$ there is a sequence $, a_1, a_2, \dots , a_n, q$ and $pAa_1, a_iAa_{i+1}, a_nAq$.

Region Boundary $\partial Q = \{q \in D \backslash Q : \exists p \in Q : qAp\}$, i.e the boundary ∂Q of Q is the set of pixels being adjacent to at least one pixel of Q but not belonging to Q.

Extremal Region $Q \subset D$ is a region such that for all $p \in Q , q \in \partial Q : I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region).

Maximally Stable Extremal Region (MSER). Let $Q_1 \dots , Q_{i-1}, Q_i, \dots$, Q1 … be a sequence of nested extremal regions i.e. $Q_i \subset Q_{i+1}$.Extremal region Q_{i*} is maximally stable iff $q(i) = |Q_{i+\Delta} \backslash Q_{i-\Delta}| / |Q_i|$ has a local minimum at $i *$(|. | denotes cardinality). $\Delta \in S$ is a parameter of the method.

The effect of MSER Algorithm is shown in Fig. (8).

2-6. Canny Edge Detector Algorithm

Canny edge detector Algorithm helps to detect a wide range of edges in images. Canny edge detection is a technique to obtain important information from all objects in an image and reduces the computational complexity by the removal of background data. It is used as a method to gain useful information from traffic signs in an image and to reduce computational complexity of texts.

The effect of Canny Edge Detector Algorithm is shown in Fig. (9). the intersection between MSER and Canny Edge Detector algorithms is used to detect the text. Fig. (10) Shows the result.

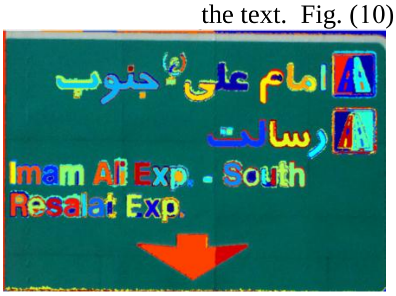


Fig8. Effect of MSER.

Fig9. Effect of Canny Edge Detector.

Fig10. Intersection between MSER and Canny Edge Detector.

Fig11. Stroke Width Transform of the image.

Fig12. Morphological Mask.

(a) (b) (c)

Fig13. Fig. (13- a, b and c). Text Box.

2-7. Stroke Width Transform

The Stroke Width Transform (SWT) is a local image operator which computes the width of the most likely stroke containing the pixel in each pixel. The output of the SWT is an image with the size equal to the size of the input image where each element contains the width of the stroke as a contiguous part of an image which forms a band of nearly constant width, as depicted in Fig. (11). after detecting of different characters in the image, these characters are joint together by the morphological Mask to display the text in the road traffic sign. The result is shown in Fig. (12).

To display the text in traffic signs, a box is made for each continuous region. Fig. (13-a, b and c) shows the results. Finally, MATLAB's Optical Character Recognition (OCR), was used for the recognition of detected texts.

3. Discussion and Results

To accomplish practical experiments, the required systems were mounted and applied on a vehicle as shown in the Fig. (14). This system was provided to study algorithm performance in urban circumstances and it was examined in real-life condition. The processing results of the captured image on Fig. (14) is shown in Fig. (15). This study was implemented at the Advance Vehicle Control Systems Laboratory

The results show that the designed algorithm has a suitable accuracy in real situations and the accuracy percent is about 90.6 percent. Table (1) shows the result of the automatic detection.

Finally the overall percentage of accuracy is

Fig14. Equipped vehicle for testing the provided algorithm.

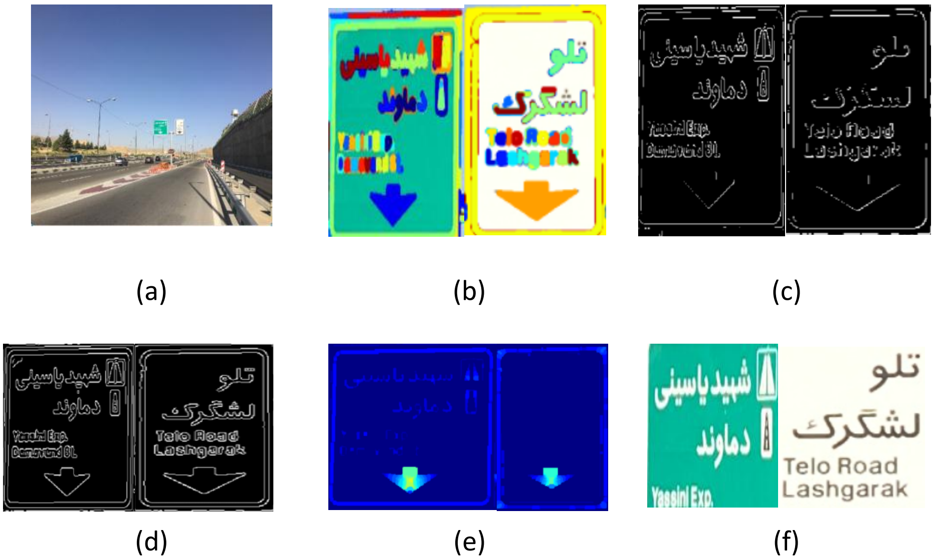


Fig15. Fig. (15-a to f). Results showing the live process of the captured image.

Table 1. Results of the automatic detection on 60 Iranian road signs.

Percentage of Accuracy: below of 50%	Percentage of Accuracy: between 50%-100%	Percentage of Accuracy :100%
4	12	44

4. Conclusion

The main goal of this paper was to design an intelligent driver assistant system for text detection of traffic signs. The designed system can help drivers to attain more information from traffic signs in order to have more time-efficient trips. To satisfy this purpose, a new framework was designed. As the first step, traffic signs were detected using fuzzy logic, then, by the combination of MSER and canny edge detector algorithms, the texts in traffic signs were detected. Afterward, a box was designed for each continuous objects using the morphological mask algorithm. Finally, the detected texts were recognized by the help of MATLAB’s OCR framework. The results show that this new framework has a suitable performance in real situations and can enhance traffic management significantly and also reduces waste of time to a great distance.

References

[1]. D. Chen, J. M. Odobez, "Text detection and recognition in images and video frames," Pattern Recogn., vol. 37, no. 3, pp. 595-608, Mar. 2004.

[2]. C. W. Lee, K. jung, and H. J. Kim, "Automatic text detection and removal in video sequences," Pattern Recogn. Lett., vol. 24, no. 15, pp. 2607-2623, nov. 2003.

[3]. G. Myers, R. Bolles, Q-T. Luong, and J. Herson, "Recognition of text in 3-D scenes, "in Proc. 4th symp Document Image Understanding Technology, Columbia, MD, pp. 23-25, Apr. 2001.

[4]. S.-L. Chang L.-s. Chen, Y.-C. Chung, and S.-W. Chen, "Automatic license plate recognition," IEEE Trans Intell Transp Syst., Vol 5, no. 1pp. 42-53 Mar. 2004.

[5]. H. Veeraraghavan O. Masoud and N. p. Papanikolopoulos, "Computer vision algorithms for intersection monitoring, " IEEE Trans Intell, Transp Syst., Vol. 4, no. 2, pp. 78-89, Jun. 2003.

[6]. R. Vicen-Bueno, R. Gil-Pita, M.P. Jarabo-Amores and F. L´opez-Ferreras, "Complexity Reduction in Neural Networks Applied to Traffic Sign Recognition", Proceedings of the 13th European Signal Processing Conference, Antalya, Turkey, September 4-8, 2005.

[7]. R. Vicen-Bueno, R. Gil-Pita, M. Rosa-Zurera, M. Utrilla-Manso, and F. Lopez-Ferreras, "Multilayer Perceptrons Applied to Traffic Sign Recognition Tasks", LNCS 3512, IWANN 2005, J. Cabestany, A. Prieto, and D.F. Sandoval (Eds.), Springer-Verlag Berlin Heidelberg 2005,

[8]. C. Paulo and P. Correia, "Automatic detection and classification of traffic signs," in Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop on, June 2007.

[9]. D. Gavrila, "Traffic sign recognition revisited," in DAGM-Symposium, 1999, pp. 86–93.

[10]. K. Brki ´ c, A. Pinz, and S. ˇ Segvi ´ c, " Traffic sign detection as a component of an automated traffic infrastructure inventory system," Stainz, Austria, May 2009.

[11]. X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection of signs with affine transformation," in Proc. Workshop Application Computer Vision (WACV), Orlando, FL, pp. 32–36, 2002.

[12]. P. Clark and M. Mirmehdi, "Estimating the orientation and recovery of text planes in a single image," in Proc. 12th British Machine Vision Conf., Manchester, U.K., Guildford, U.K.: BMVA, Sep, pp. 421–430, 2001.

[13]. "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.

[14]. A. K. Jain and B. Yu, "Automatic text location in images and video frames," Pattern Recogn., vol. 31, no. 12, pp. 2055–2076, Dec. 1998.

[15]. Young, I.T., Gerbrands J.J., and Van Vlient, L.J., Fundamentals of Image Processing. Printed in The Delft University of Technology, Netherlands, 1998.

[16]. Oruklu, E., Pesty D., Neveux J., and Guebey J.E., "Real-Time Traffic Sign Detection and Recognition for In-car Driver Assistance Systems" IEEE 55th International Midwest Aymposium on Circuits and Systems (MWSCAS), pp. 976-979., 2012.

[17]. Fleyeh H., "Traffic Signs Recognition by Fuzzy Sets". IEEE Intelligent Vehicles Symposium. Eindhoven University of Technology, Eindhoven, Netherlands, pp. 422-427., 2008.

[18]. Blackledge, J, "Digital Image Processing Mathematical and Computational Methods Horwood Publishing", ISBN: 1-898563-49-7. 2005.

[19]. Solanki, D.S, and Dixit, G. "Traffic Sign Detection Using Feature Based Method". International Jornal of Advanced Research in Comuter Science and Software Engineering, 5(2), Feb. pp.340-346, 2015.

[20]. Tuytelaars T., and Mikolajczyk, K., Foundation and Trends in Computer Graphics and Vision, Vol. 3, No. 3, pp. 177-280. 2007.

[21]. Chen, H, Tsai, S., Schroth, G., Chen, D., Grzeszczuk, R., Girod B., "Robust text detection in netural image with edge-enhanced maximally stable external regions". In: Proc. Of Internat. Conf. on Image Processing (ICIP), pp. 2609-2612., 2011.

[22]. Epshtein, B., Ofek, E., Wexler, Y., Detecting text in natural scenes with stroke width transform. In: IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2963-2970, 2010.