ISME

Detection and Recognition of Multi-language Traffic Sign Context by Intelligent Driver Assistance System

A.R. Khodayari*
Associate Professor

M. Yousefi†
PhD Student

Design of a new intelligent driver assistance system based on traffic sign detection with Persian context is concerned in this paper. The primary aim of this system is to increase the precision of drivers in choosing their path with regard to traffic signs. To achieve this goal, a new framework that implements fuzzy logic was used to detect traffic signs in videos captured along a highway from a vehicle. Implementing fuzzy logic in smart systems increases its inference and intelligent capabilities that results in better decision making in real-time conditions. In order to detect road sign's texts, the combination of Canny Edge Detector Algorithms and Maximally Stable Extremal Regions (MSER) is used. Regions of an image that vary in properties, such as color or brightness, with respect to surrounding regions, are detected with the help of MSER algorithm. By using a multi-stage algorithm, Canny edge detector detects a wide range of edges in the acquired images. In order to join the individual characters for the final stage of detection of texts in traffic signs, a morphological mask operator is used. Finally, the recognition of the detected texts is carried out by employing MATLAB Optical Character Recognition (OCR). The overall accuracy of this new framework in detecting and recognizing texts is 90.6%.

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

1 Introduction

One of the most important tasks of an advanced driver assistance system is automatic detection and recognition of road sign's text from video taken from the road. There have been many works on the detection and recognition of road sign's texts in the past few years [1-3]. However, automatic detection and recognition of the road signs from videos that are captured from the road is concerned in this paper. The results of this research can be used in the design of driver assistance systems. The writings on the road signs have important information for a vehicle's path in highways. Acquiring the information on the road signs by means of detecting the text on it can raise the awareness of drivers about traffic situations to a great context. Moreover, by adding it to a navigational system such as GPS, individuals can have more efficient driving performance.

---

* Corresponding Author, Associate Professor, Mechanical Engineering Department, Pardis Brunch, Islamic Azad University, Tehran, Iran khodayari@pardisiau.ac.ir
† PhD Student, Mechanical Engineering Department, Shahid Rajaee Teacher Training University, Tehran, Iran 2015mehrdadyousefi@gmail.com
Receive : 2019/10/16 Accepted : 2019/11/10

Implementing voice assistants on this system can develop driver-system interaction [4], [5]. Road sign recognition systems usually consists of two characteristic phases [6], [7]. The first phase usually uses image-processing algorithms to detect traffic signs in an image or a video sequence. The second phase is the recognition of the information on these detected signs.
Common techniques have been developed in pattern recognition and computer vision to detect scene text from still images. Using Hough-transform based method is arguably the most common approach. Other popular approaches are based on simple template matching and corner detection followed by reasoning. Studies using these approaches are proposed in [8-10]. Using their prior knowledge of the problem (like the expected shape and color of a traffic sign) some approaches implement learning machine method for traffic sign recognition. An example of this approach is proposed in [11]. Several recent works are able to detect scenes [12], [13] and as they reported, edge features can handle lighting and scale variation better than scene images [14], which are often used in news videos for detecting texts [15]. In this study, the edge-based features used for text detection are inspired by the aforementioned works. A full perspective transformation model to detect three-dimensional (3-D) deformed text from still images has been described by Myers et al. [16].

First step is Capturing video from the path in front of the vehicle. In order to help driver to be able to manage the vehicle on the road ahead, the system should detect texts on traffic signs. In Fig. (1-a to d) four examples of traffic signs are pictured.

There exists many challenges for such a system to detect texts on road signs. Firstly, low resolution and environmental noise in video images can affect the results. Furthermore, uncontrollability of weather and lighting conditions, the darkness of the path and sunlight reflectivity all add up to the problem. With the introduction of a new framework in this study, all the aforementioned issues have been addressed.

(a)

(b)

(c)

(d)

Figure (1- a to d) Some examples of road signs under different environmental situations. Notice the effect of lighting conditions, weather and highlights on the signs.

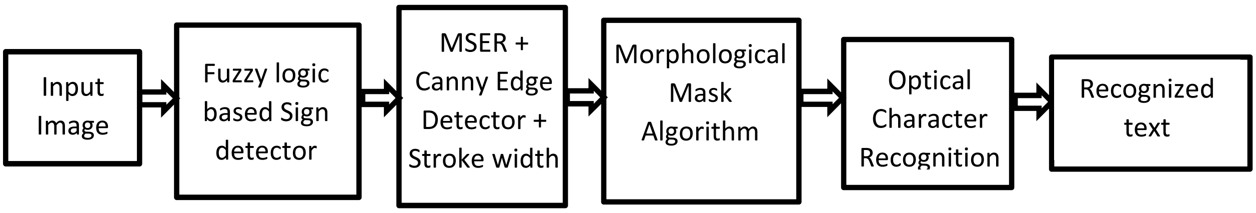


Figure 2 Design flowchart of the system

In the proposed framework, image preprocessing is consists of converting the RGB image to HSV. Image noise removal is implemented for the constructed fuzzy image and road sign is recognized and detected from the background by edge detecting algorithms. In the last phase, feature extraction is used to detect the texts from the road signs. Evaluation of this method on 60 road signs has a result of average accuracy of 90.6% in detecting the road signs with Persian context. On the plus side, text detection for other international languages is possible with this new framework.

Organization of the contents of this paper is as follows: design of an intelligent driver assistance system and a description of its framework is addressed in section (2), the test setup and experimental results obtained by using the system in real-life conditions are presented in section (3). Finally, section (4) concludes this paper.

2 Design of an Intelligent Driver Assistance system

Main steps in the design of this system can be summarized as follows: Design of the road sign detector, detection of signs on images, text detection followed by text recognition. These steps are pictured in the flowchart of Fig. (2).

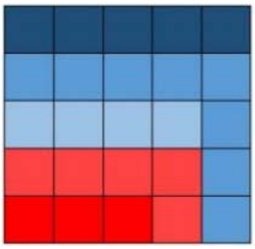
2.1 Preprocessing

In order to increase the accuracy of the process, RGB image is being converted to HSV. The difference in some features of HSV with respect to RGB makes it more appropriate to be used in text detection algorithm [17]. Since lighting conditions only affect intensity in HSV, lighting variations have no effect on the hue value to be processed. In addition, the segmentation process must be implemented on 1D of color component in contrast to RGB's 3D space, so the computational complexity and resources needed for color segmentation on HSV is less than RGB [18], [19].

2.2 Construction of Fuzzy Images

In fuzzy images, linguistic values present features of each picture such as color or lighting. These verbal values that are allocated to each pixel can be types of white, pale green, yellow-green, dark red and dark blue color. Fuzzy logic is used to transfer an image to fuzzy space. An example of a color image and its fuzzy image counterpart is shown in Fig. (3); where R, PR, PB, B, DB represent red, pale red, pale blue, blue, and dark blue respectively.

The colors that were used in this research were black, yellow, blue, green, and red. The white color was used instead of the other colors. Inputs to fuzzy sets for each pixel includes three characteristics of color, saturation, and value. The membership function of these specifications is depicted in Fig. (4).



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

Figure 3 The fuzzification of an image.

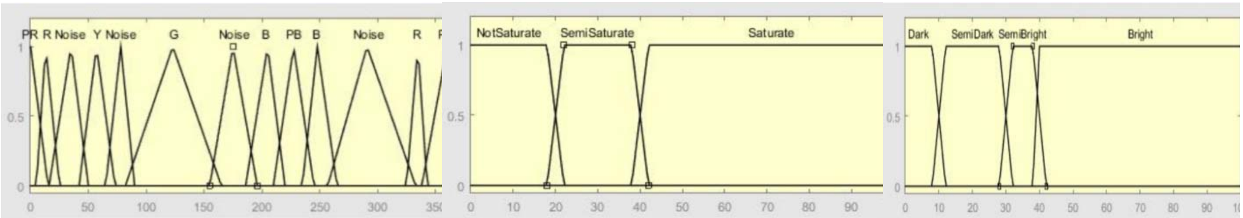


Figure 4 Membership functions for inputs.

Image noise removal is essential for the constructed fuzzy image. In noise removal a comparison between the color of each pixel and its eight neighbors id made to analyze the necessary changes to central pixel. By using this algorithm, undesirable and diffused colored pixels were corrected. As a result of this process, useful information of an image can be destroyed. Therefore, detecting road signs from further distances is harder.

2. 3 Edge Detection and Labeling

In this step, pixels are classified into useful sets consisting homochromatic pixels and their neighbors. To present the pixels of each set and to distinguish them, a specific code was dedicated to each pixel showing their set number. This process is called labeling.

The fuzzy image used for this algorithm is shown in Fig. (5). The fuzzy image indicates the verbal values related to each pixel that consists of red, blue, yellow, green, and black. After the labeling process, edges of the initial image were extracted by using Sobel edge detection algorithm [20]. Afterward, the edge pixels were removed from the fuzzy image.

The direction of investigating pixels in the labeling process was from top to bottom and from left to right. A new label was not used for a pixel if the color of the pixel is white or it was labeled before. Otherwise, it was considered as the starting point for labeling and a new label was declared for it. Afterward, the inputs of the fuzzy logic system, the labeled pixel, and its four neighbors were used. Neighbors of the labeled pixel that had the same color as the labeled pixel were labeled as a member of the set and were saved thereafter. Labeling process has the fuzzy rules as stated below:

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*

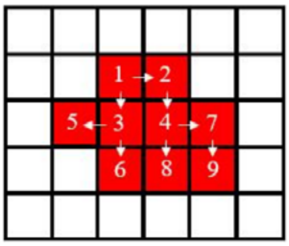


Figure 5 Labeling process.

Figure 6 Noise removal and edge detecting process of an image containing road signs.

These steps were repeated for all the pixels in the set. Finally, this process analyzes all pixels one by one and gets back to the starting point. There is a low probability that a very large or small set of pixels contains useful information about road signs, so it will be eliminated due to the definition of thresholds. This decrease in the number of labels results a higher computational speed in the following processes. Up until this point, all the processes were done on images were about the color specification of road signs. Moreover, road signs have different colors and geometrical shapes and in urban environments, these two characteristics rarely come together. At this point, the location of the road signs in the image was detected by finding the parts of the image that have the same color. Interconnected parts in an image are usually representing a special shape. An example of edge detection and noise removal is presented in Fig. (6).

2.4 Invariant features

Extracting object features in an image is a way of object detection [21]. Features are properties and parts that uniquely represent the target object. Since features of objects in an image such as size, color, intensity, and angle are ever-changing, feature selection should be in a way that they could be detected under various circumstances. In general, features are classified into two types: global and local [22]. Local features describe special parts on an object. Each pixel of an image and other pixels in that near vicinity introduce special specification. Usually, desired features are selected from parts of the image that have different specifications. Structure, color, and intensity are common specifications of an image. Small parts, edges, and points of the image are assumed as local features. Afterward, measurement of the main part that is assumed as a local feature was accomplished and the related modifiers were created and can be implemented in a variety of applications. The global feature is a model in which entire parts of an image can be seen in a moment. Histogram of an image is an example of a global feature. Global feature applications have acceptable performance within low color variation images in image retrieval and object detection. Users paid more attention to all combinations of the considered images because of this property. The downside of using global features is the inability to detect foreground from background regions.

Figure 7 Detection of a sign in the capture.

Also it cannot combine their information. Therefore, it is a challenge to detect highly detailed or partially blocked images. Limitations of global features can be eliminated by dividing images into several parts so that each of them represents a part or an object of it. Partitioning of an image can proved to be challenging since it needs high perception of the image.
These drawbacks of global features and image partitioning can be eliminated by using a method named "Sample Features". In this method, global features are extracted from comprehensive examples of different parts of an image in different sizes and positions. Sliding window was also used in this method. Various applications of sampled features consist of specific sets of objects, for example, vehicles or pedestrians, or special objects.

Sample features use sample parts and search for resemblances to model samples. However, this method cannot be used to detect items that are parts of a blocked object. The result of detecting a road sign in the capture can be seen in Fig. (7). The three algorithms used in the process of text detection consist of MSER algorithm, Edge detection, and Stroke Width Transportation.

2. 5 MSER Algorithm

MSER algorithm detects specifications of a text such as distinguished letters. In order to detect this specification, various algorithms were developed. Distinguished letters refer to the points in an image with specific geometrical shape and penumbra feature. In addition, these points have the highest continuity in an image. This algorithm works by finding a set of distinguished regions inside a gray space. Extremal regions are extracted by MSER algorithm in the gray-scale image. One of the methods used in text detection is Maximally Stable Extremal Regions. MSER extracts some covariant regions in the image. The word extremal in MSER refers to pixels that have different intensity in comparison to other pixels [23].

The image I is a mapping $I : D \subset Z^2 \to S$. Under these conditions extremal regions can be well-defined on images:

1- S is completely ordered i.e. antisymmetric, reflexive, and transitive binary relation $\leq$ exists. Extremal regions can be defined on S=R (real-valued images), but in this paper $S = \{0,1, \dots ,255\}$ was considered.

2- $A \subset D \times D$ is defined as an adjacency (neighborhood) relation. Four neighborhoods were used in this study, 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 exists 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 all pixels being in the immediate neighborhood of at least one pixel of Q but not belonging to Q.

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

Maximally Stable Extremal Region (MSER). Assuming $Q_1 \dots, Q_{i-1}, Q_i, \dots$,Q1, … to 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 in this method. In Fig. (8) the effect of MSER is evident.

2.6 Canny Edge Detector Algorithm

By using Canny edge detector, a wide range of edges can be detected in images. Canny edge detection is used to acquire key information from objects in an image and to reduce computational complexity by means of removing background data. This technique can also be used to get useful information from an image of a traffic sign and to reduce the computational complexity of texts. The effect of Canny edge detector can be seen in Fig. (9). The intersection between Canny edge detector and MSER algorithm for detecting texts is shown in Fig. (10).

2.7 Stroke Width Transform

The Stroke Width Transform or SWT is a local image operator that for each pixel calculates the width of the most probable stroke containing that pixel. As shown in Fig. (11), the output image of SWT is equal in size compared to the input image where each element has the width of the stroke as a contiguous part of the image that creates a band of approximately constant width. As can be seen in Fig. (12), after completing the detection of various characters in the image, these characters are joined together to display the text in the road sign by using morphological mask. As shown in Fig. (13 a to c), the text in the road signs is displayed by using a box for each continuous region.

Figure 8 Effect of MSER.

Figure 10 The intersection between Canny Edge

Figure 11 Stroke Width Transform of the image.

Figure 12 Morphological Mask.

Detector and MSER.

(a) (b) (c)

Figure 13 a to c Text Boxes.

3 Eigen decomposition of Images for detection of words

Firstly, pre-determined words that are investigated in this study were saved in the database. This system saves the maximum singular value of each textbox in order to use them for detecting texts during the processing of traffic signs. For this purpose, each detected word is transformed into gray-scale, then Canny algorithm was used to detect edges. Afterward, eigenvalues are obtained by Eigen decomposition of the matrix with tolerance limit of 0.1 for detecting the existing words. Eigen decomposition of the matrices is described as follows.
The matrix $A_{m\times n}$ with rank k can be decomposed as follows:

$$A = U\Sigma V^T$$

Where $U_{m\times m} = [u_1 \ \dots \ u_m]$ and $V_{n\times n} = [v_1 \ \dots \ v_n]$ are orthogonal matrices. Columns of $U_{m\times m}$ contain orthonormal eigenvectors of AA^T and columns of $V_{n\times n}$ contain orthonormal eigenvectors of A^TA and $\Sigma_{m\times n}$ is a diagonal matrix that its main diagonal entries are non-zero singular values of AA^T or A^TA .

$$\Sigma_{m\times n} = diag(\sigma_1, \dots, \sigma_p) \quad , \quad p = \min\{m, n\}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0 \quad , \quad \sigma_{k+1} = \cdots = \sigma_p = 0$$

Here σ_1 and σ_k are biggest and smallest non-zero singular value of A respectively.

4 Discussion and Results

To accomplish the required experiments, the system was mounted on a vehicle as shown in Fig. (14). This system was designed to study the real-life performance of the algorithm in urban environments. The results of the process on captured images is shown in Fig. (15).
This study was completed at the Advanced Vehicle Control Systems Laboratory.

The accuracy is defined as the percentage of correctly detected road signs. System performance is evaluated on 60 traffic signs with Persian context. All the mentioned steps are performed on each road image and the results of the automatic detection are presented in Table (1). As each image has a different percentage of accuracy, the average of all results is calculated. The results show that the average accuracy of the designed algorithm is the acceptable value of 90.6%.

Figure 14 The test vehicle used for examining the provided algorithm.

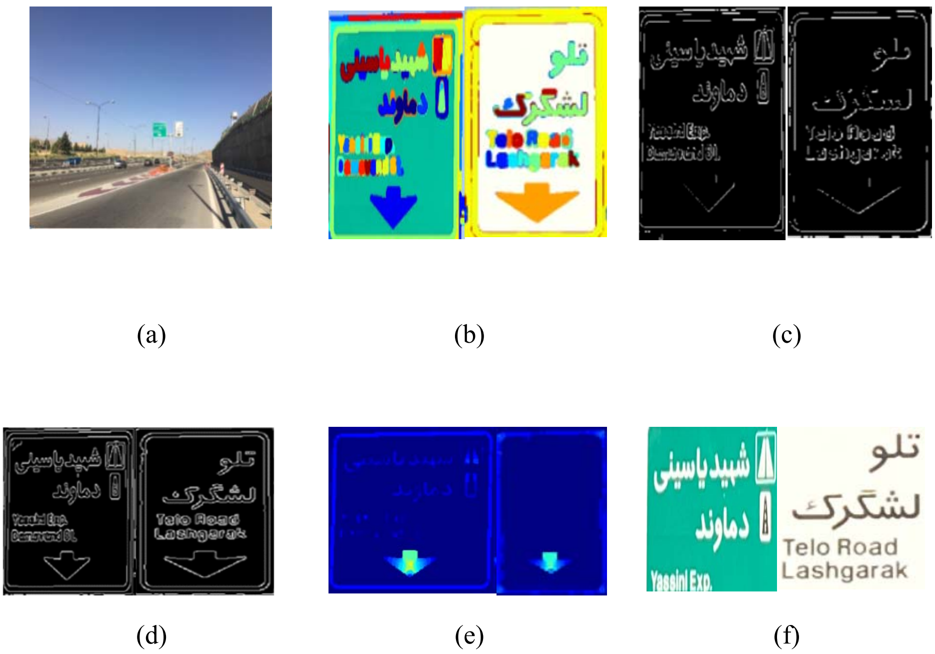


Figure 15 a to f Results of the real-time process of the captured images.

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

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

5 Conclusion

As stated before, the aim of this study was to design an intelligent driver assistant system for detecting road signs. This system can assist drivers to obtain more information from road signs so that they could have more time-efficient trips. This purpose was satisfied by designing a new framework. Firstly, road signs were detected by applying fuzzy logic to images, thereafter, by using a combination of Canny edge detector and MSER algorithms, the texts in the road signs were detected. Then, by using the morphological mask algorithm, each continuous object was assigned a box. Finally, by using MATLAB's Optical Character Recognition, the detected texts were recognized. The experimental results show that this new framework has a promising performance in real-life situations and can be used to enhance traffic management and also shorten the trip time by a great amount.
Our proposed method works on rectangular signs. In future, we will work on the all shaped traffic sign context detection and the comparison of proposed system and neural networks to find the best technique for recognition of traffic sign.

References

[1] Chen, D., and Odobez, J. M., "Text Detection and Recognition in Images and Video Frames", Pattern Recognition, Vol. 37, No. 3, pp. 595-608, (2004).

[2] Lee, C. W., Jung, K., and Kim, H. J., "Automatic Text Detection and Removal in Video Sequences", Pattern Recognition Letters, Vol. 24, No. 15, pp. 2607-2623, (2003).

[3] Myers, G., Bolles, R., Luong, Q.T., and Herson, J., "Recognition of Text in 3-D Scenes", Proceedings of the 4th Symposium on Document Image Understanding Technology, Columbia, MD, pp. 23-25, (2001).

[4] Chang , S. L., Chen, L. S., Chung, Y. C., and Chen, S. W., "Automatic License Plate Recognition", IEEE Transaction on Intelligent Transportation Systems, Vol. 5, No. 1, pp. 42-53, (2004).

[5] Veeraraghavan, H., Masoud, O., and Papanikolopoulos, N. P., "Computer Vision Algorithms for Intersection Monitoring", Transaction on Intelligent Transportation Systems, Vol. 4, No. 2, pp. 78-89, (2003).

[6] Vicen-Bueno, R., Gil-Pita, R., Jarabo-Amores, M.P., and L´opez-Ferreras, F., "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] Vicen-Bueno, R., Gil-Pita, R., Rosa-Zurera, M., Utrilla-Manso, M., and Lopez-Ferreras, F., “Multilayer Perceptrons Applied to Traffic Sign Recognition Tasks”, International Work-Conference on Artificial Neural Networks, IWANN 2005: Computational Intelligence and Bioinspired Systems, pp 865-872, (2005).

[8] Loy, G., “Fast Shape-based Road Sign Detection for a Driver Assistance System”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sendai, Japan, pp. 70–75, (2004).

[9] Paulo, C., and Correia, P., “Automatic Detection and Classification of Traffic Signs”, Eighth International Workshop on, Santorini, Greece, June (2007).

[10] Gavrila, D., “Traffic Sign Recognition Revisited”, in DAGM-Symposium, Germany, pp. 86–93, (1999).

[11] Brki´c, K., Pinz, A., and ˇSegvi´c, S., “Traffic Sign Detection as A Component of an Automated Traffic Infrastructure Inventory System”, Stainz, Austria, May (2009).

[12] Chen, X., Yang, J., Zhang, J., and Waibel, A., “Automatic Detection of Signs with Affine Transformation”, Proc. Workshop Application Computer Vision (WACV), Orlando, FL, pp. 32–36, (2002).

[13] Clark, P., and Mirmehdi, M., “Estimating the Orientation and Recovery of Text Planes in a Single Image”, Proc. 12th British Machine Vision Conf., U.K, Guildford, pp. 421–430, (2001).

[14] Chen, X., Yang, J., Zhang, J., and Waibel, A., “Automatic Detection and Recognition of Signs from Natural Scenes”, IEEE Trans. Image Process, Vol. 13, No. 1, pp. 87–99, (2004).

[15] Jain, A. K., and Yu, B., “Automatic Text Location in Images and Video Frames”, Pattern Recognition, Vol. 31, No. 12, pp. 2055–2076, (1998).

[16] Young, I.T., Gerbrands, J.J., and Van Vlient, L.J., “Fundamentals of Image Processing”, Printed in the Delft University of Technology, Netherlands, (1998).

[17] 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 Symposium on Circuits and Systems (MWSCAS), Boise, ID, USA, pp. 976-979, (2012).

[18] Fleyeh, H., “Traffic Signs Recognition by Fuzzy Sets”, IEEE Intelligent Vehicles Symposium, Eindhoven University of Technology, Netherlands, pp. 422-427, (2008).

[19] Blackledge, J., “*Digital Image Processing Mathematical and Computational Methods*”, Horwood Publishing, ISBN: 1-898563-49-7, (2005).

[20] Solanki, D.S., and Dixit, G., “Traffic Sign Detection using Feature Based Method”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, pp. 340-346, (2015).

[21] Tuytelaars, T., and Mikolajczyk, K., “Local Invariant Feature Detectors: A Survey”, Foundation and Trends in Computer Graphics and Vision, Vol. 3, No. 3, pp. 177-280, (2007).

[22] Chen, H., Tsai, S., Schroth, G., Chen, D., Grzeszczuk, R., and Girod, B., “Robust Text Detection in Netural Image with Edge-enhanced Maximally Stable Extremal Regions”, 18th IEEE International Conference on Image Processing, Brussels, pp. 2609-2612, (2011),

[23] Epshtein, B., Ofek, E., and Wexler, Y., "Detecting Text in Natural Scenes with Stroke width Transform”, IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, pp. 2963-2970, (2010).