Estimation of Traffic Sign Visibility Toward Smart Driver Assistance

Keisuke Doman*, Daisuke Deguchi*, Tomokazu Takahashi** *,
Yoshito Mekada***, Ichiro Ide*, Hiroshi Murase* and Yukimasa Tamatsu****

* Graduate School of Information Science, Nagoya University, Japan
** Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan
*** School of Information Science & Technology, Chukyo University, Japan
**** DENSO CORPORATION, Japan
EMail: {kdoman, ddeguchi, ttakahashi, mekada, ide, murase}@murase.m.is.nagoya-u.ac.jp

***Abstract*— We propose a visibility estimation method for traffic signs as part of work for realization of nuisance-free driving safety support systems. Recently, the number of driving safety support systems in a car has been increasing. As a result, it is becoming important to select appropriate information from them for safe and comfortable driving because too much information may cause driver distraction and may increase the risk of a traffic accident. One of the approaches to avoid such a problem is to alert the driver only with information which could easily be missed. Therefore, to realize such a system, we focus on estimating the visibility of traffic signs. The proposed method is a model-based method that estimates the visibility of traffic signs focusing on the difference of image features between a traffic sign and its surrounding region. In this paper, we investigate the performance of the proposed method and show its effectiveness.**

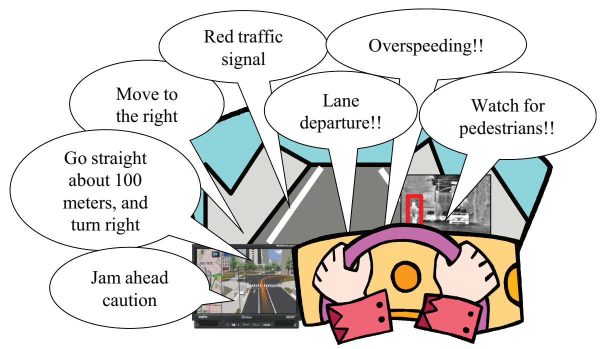


Fig. 1. Example of providing too much alerts to drivers.

I. INTRODUCTION

In recent years, the number of DSSS (Driving Safety Support System) in a car has been increasing. For example, a car navigation system that informs a driver the vehicle position in real-time improves the ease of driving dramatically. Furthermore, it is expected that a night vision system that detects pedestrians ahead at night provides safer driving. Many other systems such as lane keep assist system and adaptive cruise control have been put to practical use. However, when a lot of information from many DSSSs is provided to a driver as shown in Fig. 1, it may be an overload to the driver, which causes a "driver distraction" problem [1]. Thus, techniques to control the amount of information provided from DSSSs according to circumstances and drivers are required.

There is an approach which uses an eye-gaze tracking system for controlling information from DSSSs [2]. However, it is dangerous to directly control the information from DSSSs only with the information from a driver's gaze since a driver's gaze towards an object does not always indicate recognition of the object. On the other hand, there are several researches on the visibility estimation of various targets (e.g. [3]–[6]) with an in-vehicle camera image which is similar to human vision. For example, concerning the visibility of traffic signs, Fig. 2 shows a comparison between traffic signs captured by an in-vehicle camera in different scenes. Each image contains two traffic signs, but there are large differences in their visibility because of the positions of the traffic signs and the lighting conditions. If we can provide only easily-missed (i.e. low visibility) information to a driver, nuisance-free DSSS can be realized by means of preventing too much information, which can reduce the risk of traffic accidents. Additionally, such a visibility-based approach can be combined with an approach based on an eye-gaze.

This paper particularly focuses on traffic signs which provide important information for traffic safety. We propose a method for the estimation of traffic sign visibility using an in-vehicle camera. First, Section II introduces related works. Next, Section III describes the proposed method in detail, and results of an experiment are reported in Section IV. The paper ends with a summary and discussion of future work in Section V.

II. RELATED WORKS

To estimate the visibility of an object, it is necessary to distinguish two kinds of visual attention: pop-out (involuntary attention) and visual search (voluntary attention). Because human scene understanding is considered to be separated into two aspects: "rapid perception of scene gist (vision at a glance)" and "layout and explicit scene understanding (vision with scrutiny)" [7]. In this section, we introduce related works focusing on each of the pop-out and visual search, respectively.

(a) Traffic signs with high visibility.

(b) Traffic signs with low visibility.

Fig. 2. Comparison of traffic sign visibility.

A. Research on pop-out

As computational models of pop-out, there are many methods to estimate conspicuous regions which are easy to attract human visual attention in an input image. Itti et al. have proposed a computational model which calculates conspicuous regions with a saliency map [8]. The saliency map has been applied for various research areas and its effectiveness has been shown [9], [10]. However, especially in visual attention, involuntary pop-out is greatly influenced by factors such as psychological state, interests and anticipation. In addition, a driving task always puts a heavy load to a driver since it requires appropriate actions in real-time based on the surrounding environments. Therefore, the computational model proposed by Itti et al. is not applicable to such a situation. In fact, Simon et al. reported that representations calculated by the computational model do not correspond with visual attention in subject experiments in driving simulations [11].

B. Research on visual search

Many computational models of visual search are summarized in [12]. Unfortunately, most of the existing computational models for visual search are available only in a well-designed laboratory environment. As for practical use, a model-based method to estimate visibility of traffic signs with an in-vehicle camera has been proposed by Siegmann et al. [4]. However, visual features of human are not considered adequately since this method calculates the level of visibility only with luminance information. On the other hand, an appearance-based method for visibility estimation has been also proposed by Simon et al. [5]. This is a SVM-based method which learns appearances of a traffic sign (“No entry” sign) in advance, and then calculates the saliency of traffic signs from the SVM discriminant function directly. However, the distance in the feature space does not necessarily correspond to the saliency perceived by a human as well as it will take thousands of man-hours to collect various appearances of every target traffic sign exhaustively. Moreover, since the difference of image features between the visual target and its surrounding (background) area greatly affects the visibility, the effect of the background is not well-considered in this method. As for visibility estimation for other targets, Kimura et al. have proposed a model-based method to estimate the visibility of traffic signals with an in-vehicle camera [6]. Focusing on the effect of the background of traffic signals in this method, they use the difference of each image feature in the traffic signals and the background as a measure for evaluating the visibility. In addition, they divide the background into several sub-regions and then crop image features from each sub-region. By this method, they also consider the effects of the partial background regions. Note that there are a few but significant differences between traffic signals and traffic signs, especially in the variation of color and shape.

Considering these problems, in this paper, we propose a model-based approach focused on the effect of the background of traffic signs.

III. VISIBILITY ESTIMATION METHOD FOR TRAFFIC SIGNS

The proposed method focuses on the relation between traffic signs and its background. As shown in Fig. 3, the proposed method is composed of three steps: 1) detection of traffic signs, 2) segmentation of surrounding region and 3) estimation of the traffic sign visibility. Firstly, traffic signs in an input image are detected. Secondly, for each detected traffic sign, a sub image that is the background region surrounding each traffic sign (hereafter called “surrounding region”) is cropped from the input image, and then each surrounding region is divided into several sub-regions. Examples of surrounding regions used in the later experiments are shown in Fig. 4. Finally, visibility of each detected traffic

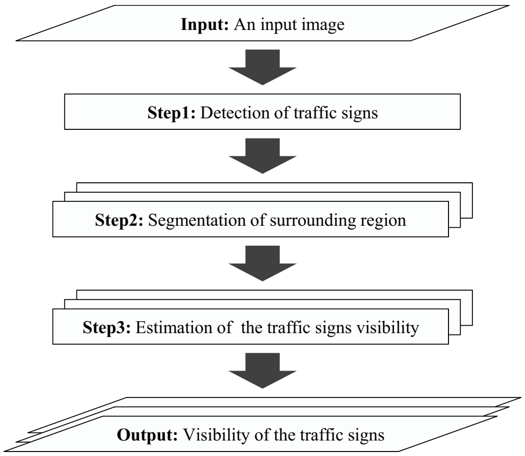


Fig. 3. Flow of the proposed method.

(a) Crossroads (b) Speed limit

Fig. 4. Examples of surrounding regions.

sign is estimated based on the feature value calculated by an adaptive integration of several kinds of image features. The detail of the process flow is described below.

Step 1: Detection of traffic signs

It is necessary to obtain the positions and scales of traffic signs in an input image to estimate the visibility of traffic signs. The category information of the traffic signs can also be useful for the visibility estimation if it is available, since a method for calculating the visibility measure can be adaptively controlled depending on the category. There are mainly two approaches for the detection and recognition. One is to prepare traffic sign detectors as many as the number of target categories. The other is to prepare one or more category recognizers in addition to a traffic sign detector. As for the former approach, We have previously proposed a method that detects a set of traffic signs whose appearances are similar in shape and color [13]. As for the latter approach, Ishida et al. proposed a method that constructs a category recognizer for the detected traffic signs [14]. In actual use, traffic signs can be detected from an input image captured by an in-vehicle camera with either of these methods or their likes.

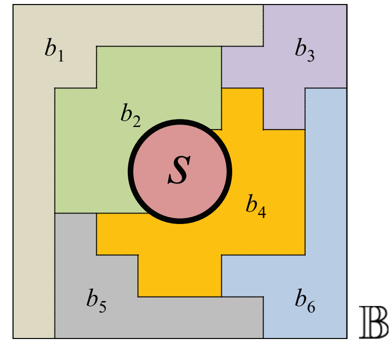
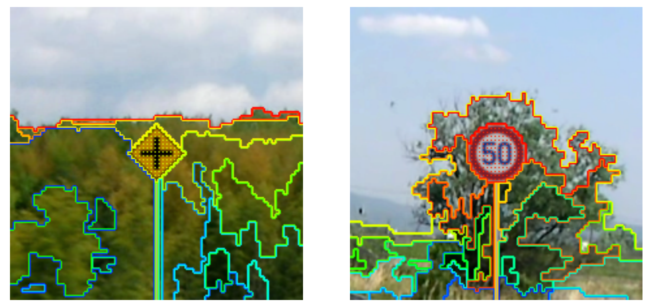


Fig. 5. Example of a surrounding region $\mathbb{B}$ surrounding a sign region s and sub-regions b_n ($n = 1, \ldots, 6$).



(a) Crossroads (b) Speed limit

Fig. 6. Segmentation results of Fig. 4.

Step 2: Segmentation of surrounding region

In an actual traffic scene, some regions in the background may have different image features among each other. In such case, visibility measurement should not be calculated from the whole background region at the same time. Thus, for each detected traffic sign region s, the surrounding region $\mathbb{B}$ is cropped and then divided into several sub-regions b_n as shown in Fig. 5.

The segmentation is done based on the CIE 1976 L*a*b* color representation where the distance is designed so that it matches the difference of human perception. The detailed algorithm is described below.

1) Divide the surrounding region into several blocks and set an initial label to each block.
2) Set the labels of each block in the region of the traffic sign to s, and exclude them from the following process.
3) Merge the adjacent blocks if the Euclidean distance of the average L*a*b* value between the blocks is smaller than a threshold.
4) Iterate the above process until convergence is reached.

Examples of the segmentation results of the surrounding regions in Fig. 4 are illustrated in Fig. 6, where the region of the sign class is represented by halftone dots and each boundary of the segmented sub-regions is shown in different colors.

Step 3: Estimation of traffic sign visibility

We estimate the visibility of a traffic sign based on the level of the feature value calculated from each sub-region $b_n \in \mathbb{B}$. That is, our method regards a traffic sign in an input image as being in low/high visibility if the feature value is low/high. The calculation of the feature value is described below.

1) Calculation of feature values: We calculate the feature values based on the following three image features.

- Color feature: average color
- Edge feature: complexity
- Texture feature: color distribution

Average color: The difference between average colors in the sign region and each sub-region is calculated as the color feature. The feature value X_1 based on average color is calculated as follows:

$$X_1 = \sqrt{(R_s - R_{b_n})^2 + (G_s - G_{b_n})^2 + (B_s - B_{b_n})^2} \quad (1)$$

where (R_s, G_s, B_s) and $(R_{b_n}, G_{b_n}, B_{b_n})$ are the average RGB values in the sign region s and each sub-region b_n, respectively.

Complexity: The difference between complexities in the sign region and the each sub-region is calculated as the edge feature. The feature value X_2 based on complexity is calculated as follows:

$$X_2 = |E_s - E_{b_n}| \quad (2)$$

where E_s and E_{b_n} are the average edge strengths calculated with Sobel filter in the sign region s and each sub-region b_n, respectively.

Color distribution: The difference between the color distribution in the sign region and each sub-region is calculated as the texture feature. The feature value X_3 based on color distribution is calculated as follows:

$$X_3 = \sqrt{\{D^{(R)}\}^2 + \{D^{(G)}\}^2 + \{D^{(B)}\}^2} \quad (3)$$

where $D^{(C)}$ ($C \in \{R, G, B\}$) is the Bhattacharyya distance defined by Eq. (4).

$$\{D^{(C)}\}^2 = \sum_j \left(\sqrt{H_s^{(C)}(j)} - \sqrt{H_{b_n}^{(C)}(j)} \right)^2 \quad (4)$$

where $H_s^{(C)}(j)$ and $H_{b_n}^{(C)}(j)$ are the j-th bin in the normalized histograms of channel C in the sign region s and each sub-region b_n, respectively.

2) Weighting the feature values: Concerning the integration of the three image features obtained by the previously described methods, we focus on the following facts about the background of traffic signs. 1) a background is composed of a set of sub-regions which have various image features, and 2) the larger and the closer the area of the sub-region is, the more significant impact for the visibility of the visual target becomes. Thus, we use the weighted sum of feature values Y_i calculated by Eq. (5), for each $X_i^{(b_n)}$ calculated in a sub-region $b_n \in \mathbb{B}$.

$$Y_i = \sum_{b_n \in \mathbb{B}} w_{b_n} X_i^{(b_n)} \quad (5)$$

where $w_{b_n} = A_{b_n}/A_{\mathbb{B}}$, A_{b_n} is the distance-weighted area based on the inverse of the distance to a traffic sign in b_n, and $A_{\mathbb{B}}$ is the sum of A_{b_n} in $\mathbb{B}$.

3) Integration of the feature values: There are various appearances of traffic signs in shape or color and each contains different image characteristics. Thus, we calculate the final image feature Y obtained by an adaptive integration of each image feature Y_i as follows:

$$Y = \sum_{i=1}^{3} \alpha_i Y_i \ , \quad (6)$$

$$\sum_{i=1}^{3} \alpha_i = 1 \quad (7)$$

where α_i is a positive weighted coefficient for each Y_i. We can control each α_i depending on the category of traffic signs. As a measure for the final decision of the visibility level, our method regards a traffic sign in an input image as being in low/high visibility if the final feature value Y is low/high.

IV. Experiments

In this section, we describe the results of subject experiments to investigate the performance of the proposed method. As described above, the proposed method is a model-based method that estimates the visibility of traffic signs focusing on the difference of image features between a traffic sign and its surrounding region. A conventional method that could be directly compared does not exist since no model-based method like our method has been proposed yet. From this reason, for a comparative method, we applied the original Kimura's method [6] to visibility estimation for traffic signs. There are some differences between the proposed method and the comparative method:

- calculation method of feature values
- way of setting weights to each sub-region
- availability of the category information of traffic signs

The last difference is the most significant. In the comparative method, α_i in Eq. (7) cannot be controlled depending on category information of traffic signs since only one image feature is used for visibility estimation. In this experiment, we evaluated each method above and then discuss the results.

A. Methods

The methods used for subject experiments are described below.

1) Preparations for surrounding regions: To eliminate the influence of quality, posture, lighting conditions, we cropped the surrounding regions as follows. Firstly, we prepared artificial sign images whose scale, shade of color, brightness, degree of blurring, etc. are the same and standardized [1]. Here, in view of similarity of color and shape, we set 6 categories of traffic signs shown in Fig. 7 as the visibility estimation

[1]Artificial sign images were purchased from `http://www.riguru.com/`.

(a) Crossroads

(b) No entry

(c) Follow the directions

(d) Speed limit

(e) Stop

(f) Crosswalk

Fig. 7. Japanese traffic signs used in this experiment as targets.

targets. Then, we synthesized artificial sign images by embedding the sign images to the center of 20 kinds of scenery images (225×225 pixels) captured by an in-vehicle camera.

2) Procedures for subject experiments: We conducted subject experiments as follows: Firstly, a pair of different surrounding regions whose categories of traffic signs are the same but with different backgrounds was selected randomly. Then we showed the subjects the selected pair, and then let them answer the traffic sign with a higher visibility from three choices: “left”, “right” or “hard to say”. We iterated the above procedure 100 times for each subject while adjusting the number of answers to each category to be almost the same. We obtained a total of $2,000$ answers from 20 subjects. Here, the experiment was performed by using the interface shown in Fig. 8.

3) Evaluation conditions: We evaluate the proposed method and the comparative method with the Degree of Agreement DoA calculated by Eq. (8).

$$DoA = \frac{N_a}{N_t} \tag{8}$$

where N_t is the number of pairs whose “left” or “right” answers are more than 80% of the total answers excluding ”hard to say” answers. N_a is the number of the pairs in N_t where the evaluations by subjects and the visibility estimation method agree with each other.

As for the proposed method, we searched all the combinations of α_i in Eq. (7) for each category of the traffic signs to find the best combination that obtains the highest DoA. As for the image feature used in the comparative method, we searched and chose the best one among average color, complexity, and color distribution.

B. Results

Experimental results are shown in Table I. The DoA with the proposed method was 0.80. This result shows that the proposed method is effective for estimating the visibility of traffic signs. Meanwhile, the DoA with the comparative method was 0.77. This is the result when average color was used as the best image feature for visibility estimation. For reference, the DoA with the comparative method based on complexity was 0.68, and the DoA with the comparative method based on color distribution was 0.70.

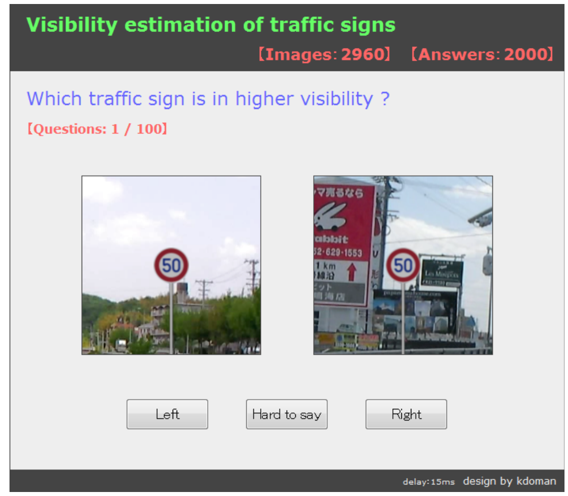


Fig. 8. The interface used for pair comparisons in this experiment.

TABLE I
DEGREE OF AGREEMENT (DoA) BETWEEN THE SUBJECT’S ANSWERS AND THE OUTPUTS OF EACH VISIBILITY ESTIMATION METHOD.

Visibility estimation method	$\boldsymbol{DoA\ (N_a/N_t)}$
Proposed method	$0.80\ (574/720)$
Comparative method (average color)	$0.77\ (551/720)$

C. Discussion

Here, we will discuss the effectiveness of 1) the integration of image features and 2) the use of category information.

1) Integration of image features: A pair of surrounding regions whose answers from all subjects agree with each other is illustrated in Fig. 9. All subjects who compared this pair judged that the traffic sign in (b) is in higher visibility than that in (a). To the contrary, the estimation results with the comparative method output the opposite. We consider that the disagreement was caused since not only average color but also texture complexity and color distribution influences the visibility. In fact, we confirmed that estimation results based on complexity or color distribution agreed with the subject’s answers. In such a case, the integration of various image features should be effective for visibility estimation. Actually, the estimation results with the proposed method agreed with the subject’s answers. Therefore, we consider that this is one reason that a higher DoA was obtained with the proposed method.

2) Use of category information: In this experiment, we set the best combination of α_i. However, if the category information is not available, we can only use the same combination of α_i for all categories of traffic signs. In this respect, a comparison of DoA with and without the use of category information is shown in Table II. From this result, we confirmed that we can obtain higher DoA when the category information is available. However, the choice of the best combination of α_i is a remaining issue of our future work, even if the category information of traffic signs is available.

TABLE II

EFFECTIVENESS OF USING CATEGORY INFORMATION OF TRAFFIC SIGNS IN THE PROPOSED METHOD.

	Category of traffic signs						Total
							
DoA with category information	0.80	0.75	0.81	0.79	0.83	0.80	0.80
(N_a/N_t)	$(103/128)$	$(85/114)$	$(93/115)$	$(85/108)$	$(109/131)$	$(99/124)$	$(574/720)$
$(\alpha_1:\alpha_2:\alpha_3)$	$(10:0:0)$	$(10:0:0)$	$(6:0:4)$	$(8:1:1)$	$(4:6:0)$	$(8:1:1)$	$(-:-:-)$
DoA without category information	0.80	0.74	0.78	0.79	0.79	0.77	0.78
(N_a/N_t)	$(103/128)$	$(84/114)$	$(90/115)$	$(85/108)$	$(104/131)$	$(96/124)$	$(562/720)$
$(\alpha_1:\alpha_2:\alpha_3)$	$(9:1:0)$	$(9:1:0)$	$(9:1:0)$	$(9:1:0)$	$(9:1:0)$	$(9:1:0)$	$(9:1:0)$

(a) Lower / Higher visibility (b) Higher / Lower visibility

Fig. 9. Subject's answers / Estimation results to a pair.

V. CONCLUSION

In this paper, we proposed a visibility estimation method for traffic signs as part of the work for realization of nuisance-free driving safety support systems by preventing to provide too much information to a driver. The proposed method estimates the visibility of traffic signs based on the difference of image features between a traffic sign and its surrounding region. In the proposed method, higher performance can be obtained with an adaptive integration of multiple image features if the category information is available. We investigated the performance of our method with pair comparisons of surrounding regions in subject experiments. As a result, we obtained 80% DoA with the proposed method, which shows that the proposed method is effective for estimating the visibility of traffic signs. Moreover, we confirmed that higher performance is obtained with the adaptive integration of multiple image features depending on the categories of traffic signs.

Future work includes considering the use of image features which is not focused in this paper, the choice of the best combination of α_i in Eq. (7) and evaluation with actual traffic sign images captured by an in-vehicle camera.

ACKNOWLEDGEMENT

Parts of this research were supported by JST CREST and MEXT, Grant-in-Aid for Scientific Research. This work was developed based on the MIST library (`http://mist.murase.m.is.nagoya-u.ac.jp/`).

REFERENCES

[1] M. Pettitt, G. Burnett, and A. Stevens, "Defining driver distraction," in *Proc. 12th World Cong. on Intelligent Transport Systems*, pp. 1–12, November 2005.

[2] L. Fletcher, G. Loy, N. Barnes, and A. Zelinsky, "Correlating driver gaze with the road scene for driver assistance systems," *Robotics and Autonomous System*, vol. 52, no. 1, pp. 71–84, July 2005.

[3] K. Mori, T. Takahashi, I. Ide, H. Murase, T. Miyahara, and Y. Tamatsu, "Recognition of foggy conditions by in-vehicle camera and millimeter wave radar," in *Proc. 2007 IEEE Intelligent Vehicles Symposium*, pp. 87–92, June 2007.

[4] P. Siegmann, S. Lafuente-Arroyo, S. Maldonado-Bascón, P. Gil-Jiménez, and H. Gómez-Moreno, "Automatic evaluation of traffic sign visibility using SVM recognition methods," in *Proc. 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision*, pp. 170–175, September 2005.

[5] L. Simon, J.-P. Tarel, and R. Brémond, "Alerting the drivers about road signs with poor visual saliency," in *Proc. 2009 IEEE Intelligent Vehicles Symposium*, pp. 48–53, June 2009.

[6] F. Kimura, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and Y. Tamatsu, "Measurement of visibility conditions toward smart driver assistance for traffic signals," in *Proc. 2007 IEEE Intelligent Vehicles Symposium*, pp. 636–641, June 2007.

[7] J. Saiki, "Representation of objects and scenes in visual working memory in human brain," in *Representation and Brain*, S. Funahashi, Ed. Springer-Verlag, 2007, pp. 103–119.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, November 1998.

[9] K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, "Real-time estimation of human visual attention with dynamic Bayesian network and MCMC-based particle filter," in *Proc. 2009 IEEE Int. Conf. on Multimedia and Expo*, pp. 250–257, June 2009.

[10] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, "Saliency-based video segmentation with graph cuts and sequentially updated priors," in *Proc. 2009 IEEE Int. Conf. on Multimedia and Expo*, pp. 638–641, July 2009.

[11] L. Simon, J.-P. Tarel, and R. Brémond, "A new paradigm for the computation of conspicuity of traffic signs in road images," in *Proc. Int. Conf. of the 26th session of the CIE*, vol. 2, pp. 161–164, July 2007.

[12] J. M. Wolfe, "Guided Search 4.0: Current progress with a model of visual search," in *Integrated Models of Cognitive Systems*, W. D. Gray, Ed. Oxford University Press, 2007, pp. 99–119.

[13] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "Construction of cascaded traffic sign detector using generative learning," in *Proc. 4th Int. Conf. on Innovative Computing, Information and Control*, ICICIC-2009-1362, December 2009.

[14] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, and H. Murase, "Identification of degraded traffic sign symbols by a generative learning method," in *Proc. 18th Int. Conf. on Pattern Recognition*, vol. 1, pp. 531–534, August 2006.