

# Estimation of Traffic Sign Visibility Considering Temporal Environmental Changes for 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, murase}@murase.m.is.nagoya-u.ac.jp

**Abstract**—We propose a visibility estimation method for traffic signs considering temporal environmental changes, as a part of work for the realization of nuisance-free driver assistance systems. Recently, the number of driver assistance systems in a vehicle is increasing. Accordingly, it is becoming important to sort out appropriate information provided from them, because providing too much information may cause driver distraction. To solve such a problem, we focus on a visibility estimation method for controlling the information according to the visibility of a traffic sign. The proposed method sequentially captures a traffic sign by an in-vehicle camera, and estimates its accumulative visibility by integrating a series of instantaneous visibility. By this way, even if the environmental conditions may change temporally and complicatedly, we can still accurately estimate the visibility that the driver perceives in an actual traffic scene. We also investigate the performance of the proposed method and show its effectiveness.

## I. INTRODUCTION

Recently, the demand for driver assistance systems is increasing. In particular, the development of a object detection and notification system with an in-vehicle camera is an important task. This system is mainly composed of two processing steps. One is the detection of target objects from an input image captured by an in-vehicle camera, and the other is the notification of the information about them to the driver. There are few researches that discuss the techniques for the latter in depth, whereas there are many of those focusing on the former. The latter technique is very important in practical applications, since providing too much information to the driver may cause driver distraction [1], and may increase the risk of a traffic accident. Therefore, we focus on the technique for sorting out an appropriate amount of information provided from the systems.

One approach is based on the driver's gaze estimated with an eye-gaze tracking system [2]. However, just because a driver gazes at an object does not mean he/she recognizes it. Therefore, it is dangerous to directly control the information only with the information from a driver's gaze. Another approach is based on the target's visibility estimated with an in-vehicle camera [3]–[5]. In particular, the visibility of a traffic sign changes largely depending on the environmental conditions despite its great importance in a traffic scene. For



(a) A scene with good visibility



(b) A scene with poor visibility

Fig. 1. Comparison of traffic scenes with different visibility of traffic signs.

example, in the scene shown in Fig. 1(a), the driver will be aware of the traffic signs because of their good visibility. On the other hand, in the scene shown in Fig. 1(b), he/she may not be able to do so because of their poor visibility. From this point of view, we consider that nuisance-free systems can be realized by providing appropriate information to the driver according to the visibility of the target.

So far, we have previously proposed a method to estimate the visibility of a traffic sign from an in-vehicle camera image [6]. However, our previous method estimated the

visibility from only one in-vehicle camera image, that is, it evaluated the visibility at a moment (hereafter called “instantaneous visibility”). In an actual traffic scene, it is considered that the driver judges the visibility of a target not from an instantaneous visibility but the visibility accumulated for a certain amount of time (hereafter called “accumulative visibility”). Moreover, even if the visibility may be good at a moment, it may be poor at the next moment, since the visibility could change temporally and largely by various factors [7]. Therefore, the driver assistance system should not control the information of the target based on only the instantaneous visibility. Thus, in this paper, we propose a visibility estimation method for traffic signs considering temporal environmental changes for smart driver assistance. The proposed method integrates a series of instantaneous visibility values calculated from an in-vehicle camera image sequence, and evaluates the accumulative visibility. By this way, we expect to accurately estimate the visibility that the driver perceives in an actual traffic scene.

This paper is organized as follows. First, Section II introduces related works. Next, Section III describes the proposed method in detail. Then, Section IV reports experimental results, and Section V provides some discussions. The paper concludes with a summary and future work in Section VI.

## II. RELATED WORKS

In general, our scene understanding is considered to be composed of two aspects: “vision at a glance” and “vision with scrutiny” [8]. Therefore, it is necessary to distinguish two types of visual attention: pop-out (involuntary attention) and visual search (voluntary attention). In this section, we introduce works related to each of the two types.

### A. Research on pop-out

There are many computational models of the pop-out to estimate salient regions which attract visual attention of human beings in an input image. Itti et al. have proposed a model for estimating such regions with a saliency map [9]. This model has been applied for various research areas, and its effectiveness has been shown [10], [11]. However, the pop-out is greatly influenced by human states such as psychology, interest, and anticipation. A driving task puts a heavy load to a driver since it always requires appropriate actions depending on the surrounding environments in real-time, which affects the pop-out of the driver. Therefore, it is considered that Itti’s model is not applicable to driving situations [12].

### B. Research on visual search

Many computational models of the visual search are summarized in [13]. Unfortunately, most of the existing models for the visual search are applicable only in a well-designed laboratory environment.

For practical use, some research groups including the authors themselves have proposed methods to estimate the visibility of a traffic sign with an in-vehicle camera [3], [4], [6]. Siegmann’s model [3] calculates the visibility level

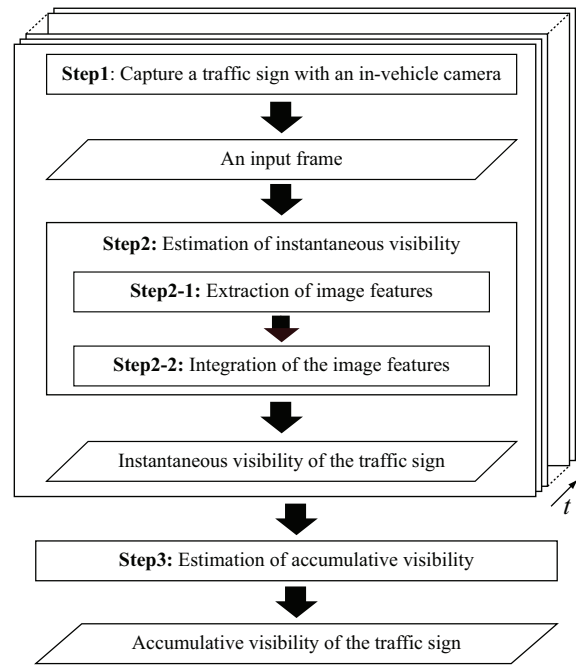


Fig. 2. Process flow of the proposed method.

based on only luminance, and visual properties of humans are not considered adequately. Simon’s model [4] learns a massive number of appearances of a target traffic sign in advance. Although this model directly calculates the saliency of traffic signs from the SVM discriminant function, the distance in the feature space may not correspond to the saliency perceived by humans. Moreover, the effect of the contrasts between a target and its surroundings are not well-considered in this method, since this model evaluates only the appearance of the target. To solve these problems, we have previously studied a visibility estimation method based on the integration of color, edge, and texture contrasts [6]. However, our previous method estimated the instantaneous visibility from only one in-vehicle camera image, and did not consider the temporal changes of the visibility. In an actual traffic scene, the visibility of a traffic sign could temporally change depending on the lighting conditions, the degree of occlusion, the visual size, etc.. Thus, it is important to consider such temporal changes for more accurate visibility estimation.

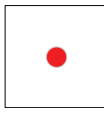
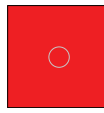

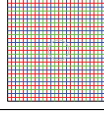


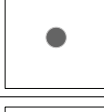
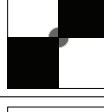
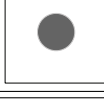
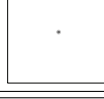
## III. PROPOSED METHOD

Fig. 2 shows the process flow of the proposed method. The first step captures a traffic sign by an in-vehicle camera. The second step estimates the instantaneous visibility of the traffic sign. The final step estimates the accumulative visibility by integrating a series of each instantaneous visibility. The details for each step are described below.

### Step 1) Capture a traffic sign with an in-vehicle camera

First of all, a target traffic sign is captured by an in-vehicle camera. The captured image is used in the following step for estimating the instantaneous visibility of the traffic sign.

TABLE I  
RELATION BETWEEN IMAGE FEATURES AND VISIBILITIES.

Factor	Visibility	
	High	Low
Color contrast		
Edge contrast		
Texture contrast		
Visual quality		
Visual size		

Here, we assume that the position, size, and category of the traffic sign in the input image can be obtained with an existing technique for traffic sign detection and recognition (e.g. [14], [15]).

### Step 2) Estimation of the instantaneous visibility

As shown in Table I, the visibility of a target is affected by several factors such as the contrasts between foreground and background, the visual quality (e.g. lighting conditions, the degree of occlusion), and the visual size. The proposed method extracts feature values  $f_i$  ( $i = 1, \dots, 5$ ) to evaluate the impacts of their factors, and then estimates the instantaneous visibility by integrating them. Note that  $f_i$  ( $i = 1, \dots, 3$ ) (the contrasts) have been previously proposed in [6], and  $f_4$  (the visual quality) and  $f_5$  (the visual size) are introduced in this paper. The extraction and integration of these feature values is performed as follows.

#### Step 2-1) Extraction of image features

The feature values  $f_i$  ( $i = 1, 2, 3$ ) based on the contrasts are calculated by the following steps. First, a sub image surrounding the sign region (hereafter called “surrounding region”) is cropped from the input image  $I$ . Second, the surrounding region is divided into a sign region  $s$  and several background sub-regions  $b_n \in \mathbb{B}$  as shown in Fig. 3. Third, the color, edge, and texture contrasts  $c_i^{(b_n)}$  ( $i = 1, 2, 3$ ) defined below are calculated [6].

- Color contrast  $c_1^{(b_n)}$ : The distance between the average color in the sign region  $s$  and that in a background sub-region  $b_n$

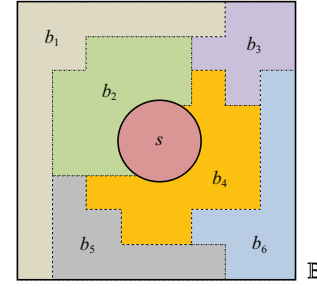


Fig. 3. Example of a surrounding region  $\mathbb{B}$  surrounding a sign region  $s$  and sub-regions  $b_n$  ( $n = 1, \dots, 6$ ).

- Edge contrast  $c_2^{(b_n)}$ : The difference between the average edge strength in the sign region  $s$  and that in a background sub-region  $b_n$
- Texture contrast  $c_3^{(b_n)}$ : The distance between the color histogram in the sign region  $s$  and that in a background sub-region  $b_n$

Fourth, the feature values  $f_i$  ( $i = 1, 2, 3$ ) based on the contrasts  $c_i^{(b_n)}$  are calculated by

$$f_i = \sum_{b_n \in \mathbb{B}} \frac{a^{(b_n)}}{a^{(\mathbb{B})}} c_i^{(b_n)}. \quad (1)$$

Here,  $a^{(*)}$  ( $* \in \{b_n, \mathbb{B}\}$ ) is defined by

$$a^{(*)} = \sum_{p^{(*)} \in *} \frac{1}{d(p^{(*)}, q^{(s)})}, \quad (2)$$

where  $d(p^{(*)}, q^{(s)})$  is the distance between each pixel  $p^{(*)}$  in a background sub-region  $*$  and the centroid  $q^{(s)}$  of the sign region  $s$ .

The feature value  $f_4$  based on the visual quality of the traffic sign is calculated by

$$f_4 = S(s, s_t), \quad (3)$$

where  $S$  is the similarity based on the SSD (Sum of Squared Difference) between the sign region  $s$  and a template  $s_t$ . Here,  $s_t$  is an ideal traffic sign image without any deterioration of visual quality.

The feature value  $f_5$  based on the visual size of the traffic sign is calculated by

$$f_5 = \frac{A^{(s)}}{A^{(I)}}, \quad (4)$$

where  $A^{(s)}$  and  $A^{(I)}$  are the areas of the sign region  $s$  and the input image  $I$ .

#### Step 2-2) Integration of the image feature

Based on the feature values  $\mathbf{f} = \{f_1, \dots, f_5\}$ , the instantaneous visibility value  $\hat{v}$  is calculated by

$$\hat{v} = \mathbf{w}^T \phi(\mathbf{f}) = \sum_{z=1}^Z w_z \phi_z(\mathbf{f}), \quad (5)$$

where  $\mathbf{w} = (w_1, \dots, w_Z)^T$  is the weight vector for the vector  $\phi(\mathbf{f}) = (\phi_1(\mathbf{f}), \dots, \phi_Z(\mathbf{f}))^T$  composed of basis

functions. Eq. (5) represents the linear combination of the vector  $\mathbf{f}$  on the  $Z$ -dimensional feature space.

*Step 3) Estimation of the accumulative visibility*

The proposed method calculates the accumulative visibility value  $\hat{V}$  by integrating a series of  $\hat{v}^{(t)}$  for each time  $t$  as

$$\begin{aligned} \hat{V} &= \frac{1}{T_p} \sum_{t=0}^{T_p-1} \hat{v}^{(\tau-t)} \\ &= \frac{1}{T_p} \sum_{t=0}^{T_p-1} \sum_{z=1}^Z w_z \phi_z(\mathbf{f}^{(\tau-t)}) \\ &= \sum_{z=1}^Z w_z \left[ \frac{1}{T_p} \sum_{t=0}^{T_p-1} \phi_z(\mathbf{f}^{(\tau-t)}) \right] \\ &= \mathbf{w}^T \Phi, \end{aligned} \quad (6)$$

where  $\tau$  is the current time,  $T_p$  is the number of input images, and

$$\Phi = \frac{1}{T_p} \left( \sum_{t=0}^{T_p-1} \phi_1(\mathbf{f}^{(\tau-t)}), \dots, \sum_{t=0}^{T_p-1} \phi_Z(\mathbf{f}^{(\tau-t)}) \right)^T. \quad (7)$$

We consider that the larger the  $\hat{V}$  is, the higher the visibility of the traffic sign is.

#### IV. EXPERIMENT

We investigated the effectiveness of the proposed method through experiments. For comparison, we evaluated the performances of the proposed method and a comparative method that estimates the visibility of a traffic sign from only an in-vehicle camera image [6]. We targeted several sign categories shown in Table II, considering the similarity in shape and color, and also the importance in traffic safety. The experimental preparations, the evaluation conditions, and the results and discussion are described below.

##### A. Experimental preparations

We prepared a test set and a training set for the parameter  $\mathbf{w}$  according to the following steps. First, with an in-vehicle camera (1,920 × 1,080 pixels, 15 fps), we captured various traffic signs under different weathers, locations around Nagoya in Japan. Then, from the captured videos, we extracted  $N = 100$  video clips (19–169 frames) as the test set, and  $M = 59$  still images as the training set. Here, in the test and training sets, each frame / image contained the wholes of the target traffic signs. Note that each test video clip did not include the training images.

We determined the ground-truth for the test and the training sets based on the following experiments with eight male and female subjects in their 20’s and 30’s. First, we showed a subject a test video clip on a computer screen only once. Next, he/she answered the visibility of the traffic sign in the range [0,1]. Here, when he/she could not find the target traffic sign, we regarded the visibility value as 0. Then, we used the average for the answers of all the

TABLE II  
TRAFFIC SIGN CATEGORIES TARGETED IN THE EXPERIMENT.

Category	Components
Warning sign	
Regulatory sign	
Indication sign	

subjects as the ground-truth for the video clip. The above procedure was performed for each subject and for each test video clip. As a result, we obtained a set of the ground-truth  $V_n$  ( $n = 1, \dots, N$ ) for the test set. Similarly, we obtained a set of the ground-truth  $U_m$  ( $m = 1, \dots, M$ ) for the training set.

##### B. Evaluation conditions

In the proposed method and the comparative method, we calculated the contrasts  $c_i^{(b_n)}$  in the RGB color space, and used the  $Z = 20$ -dimensional feature space defined by the second-order polynomial basis functions. These parameters were chosen based on results from preliminary experiments. Next, the parameter  $\mathbf{w}$  in each method was determined by the linear regression with  $U_m$  ( $m = 1, \dots, M$ ). Then, for each frame in the test set, we clipped the instantaneous visibility value  $\hat{v}$  into [0,1], since the value calculated with the obtained  $\mathbf{w}$  may be out of the range.

We evaluated the performances of each method on the test set with the MAE (Mean Absolute Error) defined by

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |V_n - \hat{V}_n|. \quad (8)$$

Note that the range of the MAE is [0,1] because the instantaneous visibility value is in the range [0,1], and the lower the MAE is, the more accurate the method is. In the proposed method, the MAEs were calculated while changing  $T_p$  from 1 to 169 (maximum number of frames in the test set). Here, when the length of the video clip was shorter than  $T_p$ , the accumulative visibility value (Eq. (6)) was calculated by averaging the instantaneous visibility values for all frames in the video clip. In the comparative method, the MAE was calculated based on the instantaneous visibility values for each last frame of the video clip. Note that the comparative method is equivalent to  $T_p = 1$  in the proposed method.

##### C. Results

Fig. 4 shows the relation between  $T_p$  and the MAEs for the proposed method ( $T_p > 1$ ) and the comparative method ( $T_p = 1$ ). The MAE of the comparative method was about 0.27. On the other hand, for all  $T_p$ , the MAE of the proposed method was lower than that of the comparative method.

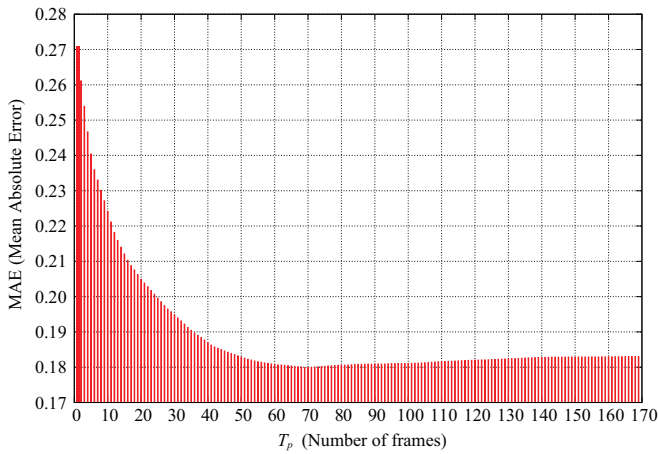


Fig. 4. Experimental results: the relation between  $T_p$  and the MAEs for the proposed method ( $T_p > 1$ ) and the comparative method ( $T_p = 1$ ).

Thus, we confirmed the effectiveness of the proposed method which evaluates the visibility of a traffic sign with an in-vehicle camera image sequence.

We aim at realizing the system that controls the information provided to the driver according to the visibility of a target. In addition, for practical use, we consider that the estimation error should be within plus / minus one level in a five-level warning system. To realize such a system, at least less than 0.20 error in the range [0,1] is required. In this experiment, the minimum MAE was 0.18 that is less than 0.20 when  $T_p = 70$ . Therefore, we conclude that the performance of the proposed method has enough feasibility considering our research goal.

## V. DISCUSSION

The effectiveness of the proposed method, and the importance of using a scene context are discussed below.

1) *Effectiveness of the proposed method:* Fig. 5 shows the transition of the instantaneous visibility values in the last 70 frames of a test video clip. For reference, Fig. 6 shows corresponding frames in the video clip. The comparative method estimates from only the last frame at time  $t = 0$ , where it output a visibility value 0.92. On the other hand, the proposed method estimates from all the last 70 frames in this case, where it output a visibility value 0.67 (almost equal to the ground-truth). Incidentally, the instantaneous visibility value estimated from the frame at around time  $t = -42$  was also nearly equal to the ground-truth. However, this was only for this video clip, not for the others. Therefore, it is difficult to estimate the visibility of a traffic sign from only the instantaneous visibility at a moment. We consider that this is why lower MAE values (Eq. (8)) were obtained by the proposed method which evaluates a series of instantaneous visibility values.

As for the image features proposed in this paper, the MAE calculated without the visual quality  $f_4$  and the visual size  $f_5$  was over 0.19 (higher than that calculated with these features). Therefore, we confirmed the effectiveness of

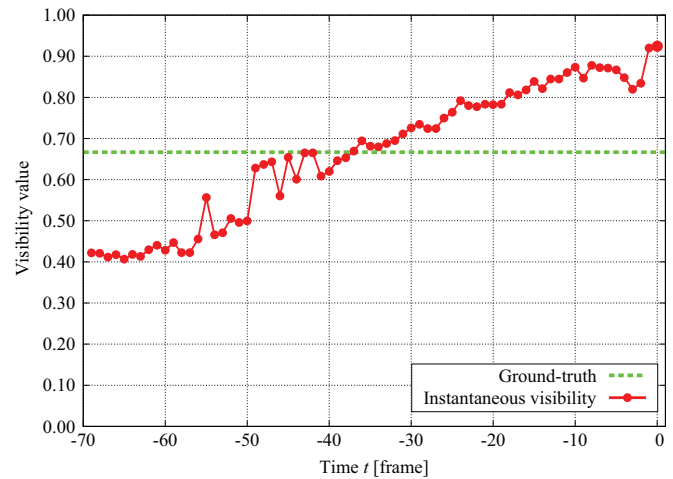


Fig. 5. Example of the transition of the instantaneous visibility.

considering the visual quality and the visual size for visibility estimation.

2) *Importance of using a scene context:* In an actual traffic scene, the visibility of a traffic sign would be affected by other objects such as sign boards, other traffic signs, traffic signals, pedestrians, and vehicles. In fact, the variance of the visibility values by subjects tended to become greater in a complex scene including many such objects. For example, in the scene shown in Fig. 7(a), there is no distractor which attracts a driver's attention. In such a scene with a simple context, the visibility of a target can be estimated by the proposed method based on the local image features. On the other hand, in the scene shown in Fig. 7(b), there are many objects such as several types of traffic signs, a bicycle and a frontal vehicle overtaking it. In such a scene with a complex context, it is considered that it is difficult to evaluate the visibility of a target only with local image features. Thus, to achieve more accurate visibility estimation, we will study on the combination of local image features and global image features considering the scene context in our future work.

## VI. CONCLUSION

In this paper, we proposed a visibility estimation method for traffic signs considering temporal environmental changes, as a part of work for the realization of nuisance-free driver assistance systems. By integrating the series of instantaneous visibility values, we expect to accurately estimate the visibility that the driver perceives in an actual traffic scene. Experimental results showed the effectiveness of our method and feasibility of our research goal. Future work includes a study on the use of global image features considering the scene context.

## VII. ACKNOWLEDGEMENT

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



(a) Time  $t = -70$  [frame]



(b) Time  $t = -42$  [frame]



(c) Time  $t = 0$  [frame] (last frame)

Fig. 6. The frames in the test video clip corresponding to Fig. 5.

## REFERENCES

- [1] M. Pettitt, G. Burnett, and A. Stevens, "Defining driver distraction," in *Proc. 12th World Cong. on Intelligent Transportation 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] 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.
- [4] 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.
- [5] F. Kimura, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and Y. Tamatsu, "Measurement of visibility conditions toward smart driver assistance



(a) A scene with a simple context



(b) A scene with a complex context

Fig. 7. Comparison of a simple scene and a complex scene.

- for traffic signals," in *Proc. 2007 IEEE Intelligent Vehicles Symposium*, pp. 636–641, June 2007.
- [6] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and Y. Tamatsu, "Estimation of traffic sign visibility toward smart driver assistance," in *Proc. 2010 IEEE Intelligent Vehicles Symposium*, pp. 45–50, June 2010.
- [7] Commission Internationale de l'Eclairage (CIE), "The conspicuity of traffic signs in complex backgrounds," CIE Tech. Rep. 137, 2000.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. 2009 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 1597–1604, June 2009.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, and H. Murase, "Generation of training data by degradation models for traffic sign symbol recognition," *IEICE Trans. on Information and Systems*, vol. E90-D, no. 8, pp. 1134–1141, August 2007.