2014 IEEE Intelligent Vehicles Symposium (IV)
June 8-11, 2014. Dearborn, Michigan, USA

Estimation of Traffic Sign Visibility Considering Local and Global Features in a Driving Environment

Keisuke Doman[1,2], Daisuke Deguchi[3], Tomokazu Takahashi[4,2], Yoshito Mekada[1,2], Ichiro Ide[2], Hiroshi Murase[2], and Utsushi Sakai[5]

***Abstract*— This paper proposes a camera-based visibility estimation method for a traffic sign. The visibility here indicates how a visual target is easy to be detected and recognized by a human driver (not a machine). This research aims at realizing a nuisance-free driver assistance system which sorts out information depending on the visibility of a visual target, in order to prevent driver distraction. Our previous study on estimating the visibility of a traffic sign considered only the effect of the local region around a target, assuming the situation that a driver's gaze is around it. The proposed method integrates both the local features and global features in a driving environment without such an assumption. The global features evaluate the positional relationships between traffic signs and the appearance around the fixation point of a driver's gaze, which considers the effect of the driver's entire field of view. Experimental results showed the effectiveness of incorporating the global features for estimating the visibility of a traffic sign.**

I. INTRODUCTION

The development of an object detection and notification system with an in-vehicle camera is important for a driver assistance system. Such a 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. Few researches focus on the latter step, whereas many researches focus on the former. The latter step is, however, very important in practical applications, because driver distraction may be caused by providing too much information to the driver [1], and may increase the risk of a traffic accident. Thus, we have been focusing on a technique for providing an appropriate amount of information to the driver.

One of the approaches to realize such techniques is the use of the visibility of a visual target. The visibility here indicates how a visual target is easy to be detected and recognized by a human driver (not a machine). For example, the visibility of a traffic sign changes largely depending on the environmental conditions despite its great importance in a traffic scene. As shown in Fig. 1(a), a driver will be aware of a traffic sign because of its good visibility. In contrast, as shown in Fig. 1(b), he/she may not be able to do so because of its poor visibility. We consider that nuisance-free systems can be realized by providing appropriate information to the driver depending on the visibility of the target. In this paper, we introduce a technique for estimating the visibility of a traffic sign with an in-vehicle camera.

(a) A scene with good visibility of a traffic sign

(b) A scene with poor visibility of a traffic sign

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

So far, there are some researches on camera-based visibility estimation for a traffic sign. Siegmann et al. and Simon et al. have proposed a luminance-based method [2] and a color-based method [3], respectively. In these methods, the effect of the contrasts between a target traffic sign and its surroundings are not well-considered. We have proposed a visibility estimation method integrating local image features based on several kinds of contrasts around a target and its

[1]School of Engineering, Chukyo University, Japan ({kdoman,y-mekada}@sist.chukyo-u.ac.jp)
[2]Graduate School of Information Science, Nagoya University, Japan ({ide,murase}@is.nagoya-u.ac.jp)
[3]Information and Communications Headquarters, Nagoya University, Japan (ddeguchi@nagoya-u.jp)
[4]Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan (ttakahashi@gifu.shotoku.ac.jp)
[5]DENSO CORPORATION, Japan

surrounding region [4], assuming the situation that a driver's gaze is around the target. As an extension to this study, we have also proposed a method that evaluates a series of instantaneous visibility values calculated from an in-vehicle camera image sequence [5]. This method still used only local image features, although it improved the accuracy of visibility estimation. If a driver's gaze is only around a target, the use of the local features may be sufficient for visibility estimation, since their effect on the visibility of the target is dominant. This assumption is, however, not always true. Accordingly, the visibility of a target would be strongly affected by not only local features but also global features in the driver's entire field of view.

Global features themselves have been introduced in literature as candidates for feature selection for estimating the detectability of a pedestrian [6]–[8], although they have not yet been introduced in visibility estimation for a traffic sign. We consider that the definition of detectability in these works is almost the same as that of the visibility on which we focus in this paper. However, there are a few significant differences between traffic signs and pedestrians. One is the number of categories of traffic signs (e.g. regulatory signs, warning signs, or indicator signs). We should not only detect traffic signs but also recognize their categories based on their appearances in order to visibly understand them. As a result, the optimal feature set for visibility estimation may be different between traffic signs and pedestrians. Another difference is that traffic signs are placed at various positions in a three dimensional traffic environment, whereas pedestrians usually appear on a walkway or a roadway. Due to these differences, it would be more important and more effective to use the global features based on the positions of traffic signs in order to estimate the visibility of a traffic sign.

This paper presents a method for estimating the visibility of a traffic sign based on both local and global features. The local features are the appearance and the size of a target, and also the contrast between the target and its surroundings. The global features are the positional relationships between traffic signs in addition to the appearance around the fixation point of a driver's gaze. The main contribution of this paper is to develop a more practical visibility estimation framework for a traffic sign using the global features, and also to investigate its effectiveness quantitatively.

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

II. PROPOSED METHOD

Figure 2 shows the process flow of the proposed system. The proposed method estimates the visibility of a traffic sign based on the framework [5] from an in-vehicle camera image sequence. First, the instantaneous visibility value of a target traffic sign is calculated by integrating both local and global features extracted from an input image for each time t. Next, a series of the instantaneous visibility values over T is averaged. The proposed method regards the average value as the visibility value of the target. Here, we assume that the position, the size and the category of the traffic sign in each input image can be obtained with an existing technique for traffic sign detection and recognition (e.g. [9], [10]).

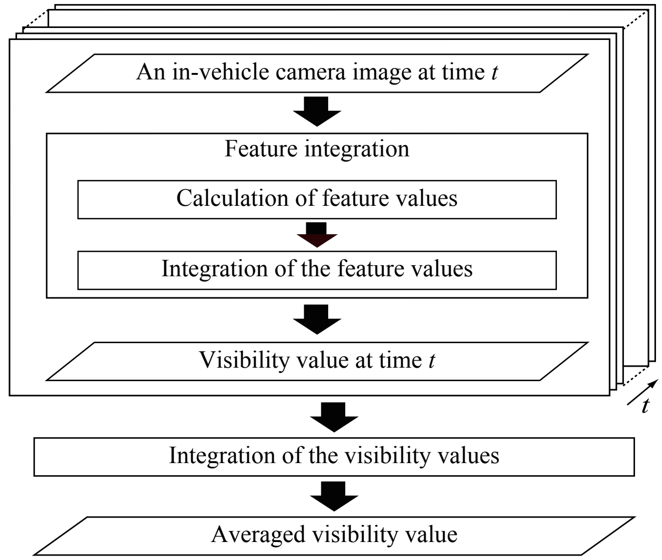


Fig. 2. Process flow of the proposed visibility estimation system.

The features for visibility estimation, and the feature integration model are described below.

A. Feature values for visibility estimation

The proposed method uses a feature vector $\boldsymbol{f} = (s_1, s_2, s_3, g, c_1, c_2, c_3, a_1, a_2, a_3)$ composed of the following features:

- Global features
  - s_1, s_2, s_3: the positional relationships between traffic signs
  - g: the appearance around the fixation point of a driver's gaze
- Local features
  - c_1, c_2, c_3: the contrast between a traffic sign and its surroundings
  - a_1, a_2, a_3: the appearance and the size of a traffic sign

Note that the global features are mainly focused in this paper. Each feature value is described below in detail.

1) Global features:

a) Positional relationships between traffic signs: In general, traffic signs are placed at various positions in a three dimensional traffic environment, and their positions in the driver's field of view are widely distributed while driving. According to the human visual characteristics, peripheral vision has lower spatial resolution than central vision, so it is difficult to capture the details of objects exactly by peripheral vision. Accordingly, in an in-vehicle camera image which corresponds to the driver's field of view, increasing distance from the image center to a target leads to the degradation of

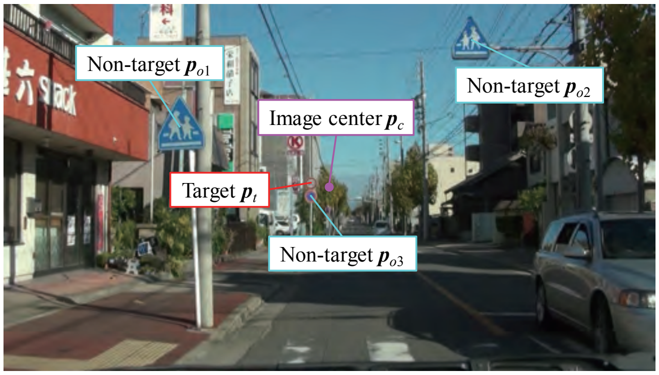


Fig. 3. Example of a target traffic sign whose centroid is $\boldsymbol{p}_t$ and non-target traffic signs whose centroids are $\boldsymbol{p}_{oj}$ $(j = 1, \ldots, 3)$.

the visibility of the target. In view of this fact, the proposed method uses the distance from the center of image to a target traffic sign as feature value s_1:

$$s_1 = \|\boldsymbol{p}_t - \boldsymbol{p}_c\|, \tag{1}$$

where $\boldsymbol{p}_t$ and $\boldsymbol{p}_c$ are the centroids of a target and the image center respectively, as shown in Fig. 3.

Also, multiple traffic signs often appear in the driver's field of view since they may be placed at short intervals, which also should affect their visibility. The visibility of more than one traffic sign is usually better than that of a traffic sign alone. In view of this fact, the proposed method uses the distance from a target to its closest traffic sign as feature value s_2, and also uses the average distance from a target to the others (non-target traffic signs) as feature value s_3:

$$s_2 = \min_{j=1,\ldots,N_o} \|\boldsymbol{p}_{oj} - \boldsymbol{p}_t\|, \tag{2}$$

$$s_3 = \frac{1}{N_o} \sum_{j=1}^{N_o} \|\boldsymbol{p}_{oj} - \boldsymbol{p}_t\|, \tag{3}$$

where $\boldsymbol{p}_{oj}$ is the centroid of the j-th non-target traffic sign, and N_o is the total number of non-target traffic signs in an input image, as shown in Fig. 3. If $N_o = 0$, the proposed method uses sufficiently larger values for both s_2 and s_3.

b) Appearance around the fixation point of a driver's gaze: The human eye has the function that controls luminance input by adjusting pupillary opening. Thus, the appearance of a traffic sign may change according to the intensity at the fixation point of a driver's gaze. In view of this fact, the proposed method uses the average intensity around the fixation point as feature value g:

$$g = \frac{1}{|G|} \sum_{\boldsymbol{p} \in G} I(\boldsymbol{p}), \tag{4}$$

where G is the $w_g \times h_g$ rectangular region centered at the fixation point, and $I(\boldsymbol{p})$ is the intensity at coordinates $\boldsymbol{p}$.

2) Local features:

a) Contrast between a traffic sign and its surroundings: Intensity and color contrast are the basic elements that affect the visibility of a visual target. In addition, evaluating edge contrast improves the accuracy of visibility estimation for a driving scene with complex background texture [11]. Considering these knowledge, the proposed method first obtains a region containing a target traffic sign and its surroundings. Then, it calculates the intensity, color and edge contrasts in the region [5], and uses them as feature values c_1, c_2 and c_3.

b) Appearance and size of a traffic sign: The appearance of a traffic sign affects its visibility, so considering its intensity and color would be effective for visibility estimation [2], [3]. In addition, a licensed driver has learned and memorized the representative appearances of traffic signs as templates, although there are several kinds of colors and shapes of traffic signs. Therefore, a traffic sign with an appearance similar to the memorized template will have good visibility for him/her. Considering these knowledge, the proposed method calculates the similarities in intensity and color between a target traffic sign and its corresponding template [5], and uses them as feature values a_1 and a_2.

As for the size of a traffic sign, the larger the size is, the better its visibility is. In view of this fact, the proposed method calculates the area of a target traffic sign region, and uses it as feature value s.

B. Feature integration model

After calculating the feature vector $\boldsymbol{f}$, the instantaneous visibility value $\hat{v}^{(t)}$ is calculated by

$$\hat{v}^{(t)} = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\phi}(\boldsymbol{f}^{(t)}) = \sum_{z=1}^{Z} w_z \phi_z(\boldsymbol{f}^{(t)}), \tag{5}$$

where $\boldsymbol{w} = (w_1, \ldots, w_Z)^{\mathrm{T}}$ is a weight vector for the basis function vector $\boldsymbol{\phi}(\boldsymbol{f}) = (\phi_1(\boldsymbol{f}), \ldots, \phi_Z(\boldsymbol{f}))^{\mathrm{T}}$. The proposed method uses the $Z = 20$-dimensional feature space defined by the second-order polynomial basis functions, considering the previous study [5].

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

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

where τ is the current time, T_p is the number of input images, and

$$\boldsymbol{\Phi} = \frac{1}{T_p} \left(\sum_{t=0}^{T_p-1} \phi_1(\boldsymbol{f}^{(\tau-t)}), \ldots, \sum_{t=0}^{T_p-1} \phi_Z(\boldsymbol{f}^{(\tau-t)}) \right)^{\mathrm{T}}. \tag{7}$$

Here, the weight vector $\boldsymbol{w}$ is set by regression with training data in advance. We consider that the larger the $\hat{V}$ is, the higher the visibility of the traffic sign is.

(a) Fix the eye position at the center of an image (1,500 ms) (b) Display an in-vehicle camera image sequence (200 ms) (c) Display a noise image (500 ms) (d) Input the position of traffic signs (No time limitation) (e) Display the positions of the traffic signs (3,000 ms)

Fig. 4. One cycle in a subject experiment ((a) and (d) are actions by a subject).

III. EXPERIMENT

We evaluated the effectiveness of the proposed method through experiments. We investigated the visibility estimation accuracy with the following methods:

- Proposed method: $\boldsymbol{f} = (s_1, s_2, s_3, g, c_1, c_2, c_3, a_1, a_2, a_3)$
- Conventional method: $\boldsymbol{f}' = (c_1, c_2, c_3, a_1, a_2, a_3)$

Here, the feature values used in the conventional method were based on our previous study [5]. The effectiveness of the global feature values s_1, s_2, s_3 and g was evaluated by comparing the accuracies of the above two methods. Note that we targeted warning signs, regulatory signs, and indicator signs in Japan, considering the similarity in shape and color, and also the importance in traffic safety.

The experimental preparations, the evaluation conditions, and the results are described below.

A. Preparations

1) Experimental Data: First, we captured various driving scenes under different luminance conditions, locations around Nagoya in Japan with an in-vehicle camera ($1,920 \times 1,080$ pixels, 30 fps). Next, we extracted 238 image sequences from the captured video, and used them as experimental data. The data contained 334 traffic signs in total, and each sequence was composed of six images (about 200 ms). Here, the traffic signs ranged in size from 18×18 pixels to 200×214 pixels. Table I shows the composition of the experimental data according to the number of traffic signs in an image.

2) Subject experiments: We determined the target visibility value for each traffic sign through the following subject experiments. We presented a subject one of the image sequences in random order on a computer screen, and asked him/her to indicate the position of a target traffic sign by using a computer mouse. This procedure is shown in Figure 4, and based on the subject experiment described in [6], [7]. The eye position of each subject was fixed to the image center as shown in Fig. 4(a). By this way, we simulated the situation that a driver's gaze is not always around a target traffic sign. Each subject practiced the procedure in advance with several image sequences apart from the experimental data.

The above experiment was performed for each of the twenty two male and female subjects in their 20's and 50's. Finally, for each traffic sign, we calculated the percentage of subjects who could correctly recognize its position. Here, we regarded a subject's input point inside the following ellipse as a correct recognition.

TABLE I
NUMBER OF TRAFFIC SIGNS IN THE EXPERIMENTAL DATA

Number of traffic signs	Number of image sequences
0	37
1	113
2	53
3	27
4	6
5	2
Total	238

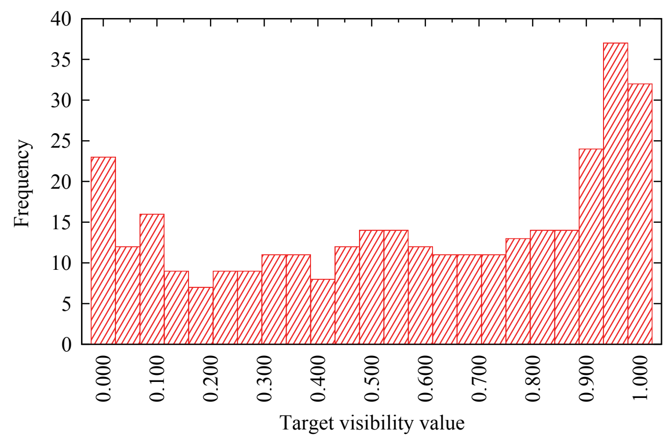


Fig. 5. Distribution of the target visibility values.

$$\frac{x^2}{w_s^2} + \frac{y^2}{h_s^2} \leq 1.5^2, \tag{8}$$

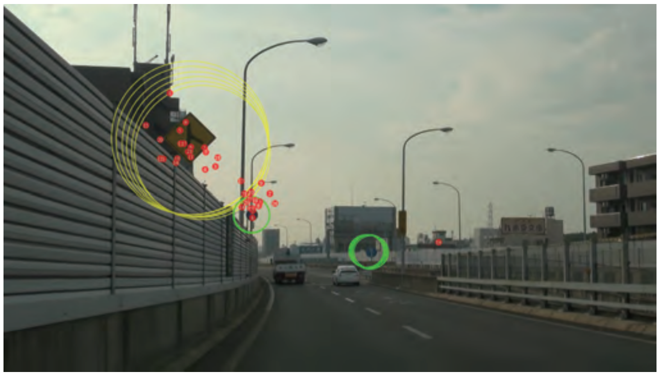
where (x, y) is the point input by a subject, and w_s and h_s are the width and the height of a target, respectively. The distribution of the obtained percentages is shown in Fig. 5. We referred to these percentages as the target values when evaluating the accuracy of each visibility estimation method.

B. Experimental conditions

1) Parameter settings: We set $T_p = 6$ in Eq. (6) for both the conventional method and the proposed method. That is, the visibility of a traffic sign was calculated from all of the images in an input image sequence. We defined the 101×101 pixels region centered at the image center as region G in

TABLE II
EXPERIMENTAL RESULT: THE ESTIMATION ERRORS OF THE CONVENTIONAL METHOD [5] AND THE PROPOSED METHOD.

Method	MAE	SD
Conventional [5]	0.237	0.178
Proposed	0.184	0.163



(a) From left to right: 0.909 / 0.918, 0.500 / 0.320, 0.000 / 0.011

(b) 0.909 / 0.811

Fig. 6. Examples of subjects' input and target value.

Eq. (4). This was because the eye position of each subject was fixed to the image center as shown in Fig. 4(a).

2) Performance evaluation: Test data for evaluating the accuracy of visibility estimation was separated from training data for the weight vector $\boldsymbol{w}$ in Eq. (5) and Eq. (6), for a five-fold cross validation. We calculated the mean absolute error (MAE) and the standard deviation (SD) between the estimated visibility values ($\hat{V}$) and their corresponding target values (Fig. 5). Here, the estimated value $\hat{V}$ was clipped into [0,1], since the value calculated with the trained $\boldsymbol{w}$ may be out of the range. Note that the MAE was in the range [0,1] as a result of clipping. The lower the MAE is, the more accurate the method is.

C. Results

The experimental result is shown in Table II. Some examples of subjects' input and the target values are also shown in Fig. 6. The MAE of the conventional method was 0.237, whereas that of the proposed method was 0.184. A significant difference between the two MAEs was observed through t-test ($p < 0.01$). We confirmed the effectiveness of the proposed method using the features s_1, s_2, s_3 and g.

TABLE III
EFFECTIVENESS OF THE PROPOSED FEATURES.

Method	Feature			MAE	SD
	s_1	s_2, s_3	g		
Comparative 1		✓	✓	0.211	0.175
Comparative 2	✓		✓	0.201	0.161
Comparative 3	✓	✓		0.200	0.171
Proposed	✓	✓	✓	0.184	0.163

IV. DISCUSSION

As explained in Section I, this research was motivated by an attempt to realize an accurate method for estimating the visibility of a traffic sign even in a situation that a driver's gaze is not around a target. This is why we introduced the additional features s_1, s_2, s_3 and g into our visibility estimation framework. This section focuses on the discussion about its effectiveness in addition to an improvement for realizing more accurate visibility estimation.

A. Effectiveness of the newly-introduced features

The difference between the conventional method and the proposed method was whether the features s_1, s_2, s_3 and g were used or not. Here, s_1 evaluates the position of a target traffic sign, whereas s_2 and s_3 evaluate the positions of non-target traffic signs. For detailed discussion, we investigated the MAEs of the methods using the full set of the features except 1) s_1, 2) s_2 and s_3, or 3) g, respectively. As a result, the proposed method outperformed all of the comparative methods, as shown in Table III. The results showed the effectiveness and the necessity of the newly-introduced features for the traffic sign visibility estimation.

B. Effectiveness under the situation in which a driver does not gaze at a target traffic sign

In the experiment, each subject's eye gaze was directed to the image center (cf. Fig. 4(a)). Accordingly, it would be more difficult to recognize a target located around the rim of an image, which should affect the visibility values. Figure 7 shows the MAE according to the distance from the image center to a target. The proposed method actually outperformed the conventional method [5] irrespective of the position of traffic signs. Thus, we consider that the proposed method could achieve accurate visibility estimation using the global features even in a situation that a driver's gaze is around a target.

C. Further improvement of the proposed method

We investigated the peak performance of the proposed method through a re-substitution method with the experimental data. As a result, the MAE and the SD were 0.144 and 0.118, respectively. The relationships between the target and the estimated values are shown in Fig. 8. We can see that there is still a large estimation error, although the estimated values were positively correlated with the target values.

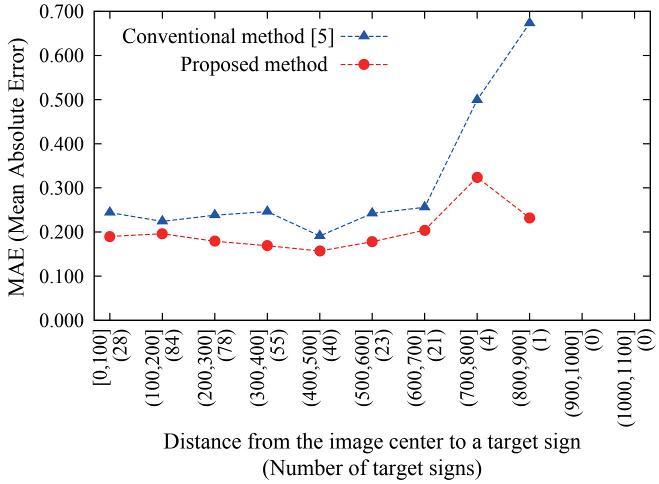


Fig. 7. Visibility estimation error according to the distance from the image center to a target.

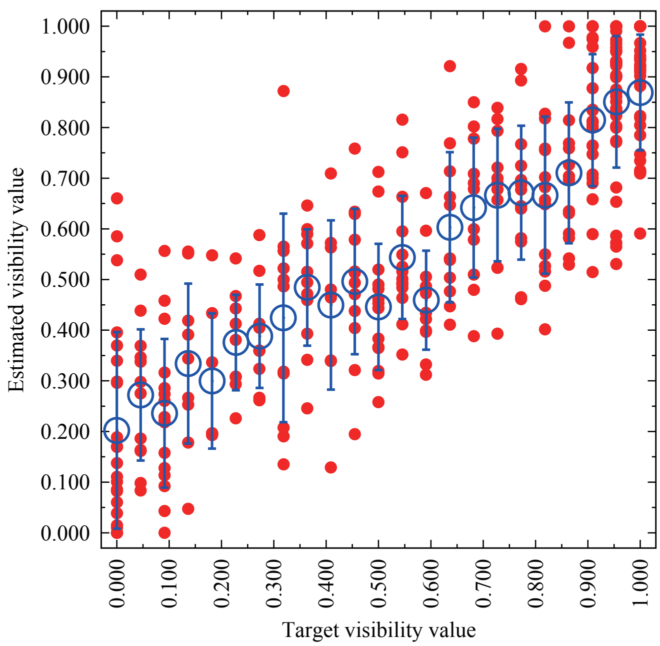


Fig. 8. Result of a re-substitution method: relationships between the target values and the estimated values ("◯" indicates the mean estimated value for each target value, and the error bar indicates ±1 standard deviation).

We consider that there are many factors yet to be introduced to the framework. One of the factors would be the effect of distractors in visual search. Many false detections by subjects were actually observed around visually-salient objects, or objects similar in appearance to traffic signs. Such objects may distract a driver to search a target, which leads to the poor visibility of the target. Therefore, we will study on improving the visibility estimation accuracy by introducing features that should evaluate the effect of distractors in visual search, such as the saliency map proposed by Itti et al. [12].

V. CONCLUSION

This paper proposed a visibility estimation method for a traffic sign with an in-vehicle camera towards a nuisance-free driver assistance system, that is more practical than the previously proposed method. The proposed method integrates both local and global features extracted from an input image sequence without the assumption that a driver's gaze is only around a target. The local features are the appearance and the size of a target traffic sign as well as the contrast between the target and its surroundings. The global features are the positional relationships between traffic signs in addition to the appearance around the fixation point of a driver's gaze. Experimental results showed the effectiveness and the necessity of incorporating the newly-introduced global features. Future work includes the study on considering the effect of distractors on visual search to achieve more accurate visibility estimation.

ACKNOWLEDGEMENT

Parts of this research were supported by JST CREST, JST COI, JSPS Grant-in-Aid for Scientific Research and JSPS Grant-in-Aid for Young Scientists.

REFERENCES

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

[2] 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*, Sept. 2005, pp. 170–175.

[3] 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*, June 2009, pp. 48–53.

[4] 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*, June 2010, pp. 45–50.

[5] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and Y. Tamatsu, "Estimation of traffic sign visibility considering temporal environmental changes for smart driver assistance," in *Proc. 2011 IEEE Intelligent Vehicles Symposium*, June 2011, pp. 667–672.

[6] M. Wakayama, D. Deguchi, K. Doman, I. Ide, H. Murase, and Y. Tamatsu, "Estimation of the human performance for pedestrian detectability based on visual search and motion features," in *Proc. 21st IAPR Int. Conf. on Pattern Recognition*, Nov. 2012, pp. 1940–1943.

[7] D. Engel and C. Curio, "Pedestrian detectability: Predicting human perception performance with machine vision," in *Proc. 2011 IEEE Intelligent Vehicles Symposium*, June 2007, pp. 429–435.

[8] D. Engel and C. Curio, "Detectability prediction for increased scene awareness," *IEEE Intelligent Transportation Systems Mag.*, vol. 5, no. 4, pp. 146–157, Nov. 2013.

[9] 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*, Dec. 2009, ICICIC-2009-1362.

[10] 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, Aug. 2007.

[11] 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*, June 2007, pp. 636–641.

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