2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

Prediction of Driver's Pedestrian Detectability by Image Processing Adaptive to Visual Fields of View

Ryunosuke Tanishige[1], Daisuke Deguchi[2], Keisuke Doman[3], Yoshito Mekada[3], Ichiro Ide[1], and Hiroshi Murase[1]

***Abstract*— Recently, pedestrian detection technology using in-vehicle cameras or sensors are being developed, which supports safety driving by notifying the drivers of the existence of pedestrians. However, warning of all existing pedestrians would interfere with the driver's concentration. Therefore, the driver should only be alerted of pedestrians with low detectability to avoid distraction of his/her concentration. To achieve this, it is necessary to develop a method to predict the detectability of a pedestrian by the driver. This paper proposes a method for predicting the pedestrian detectability adaptive to the characteristics of the human visual field. We prepared image features effective for the different regions of the human visual field; central and peripheral, in order to predict the pedestrian detectability correctly. To obtain the ground truth of the pedestrian detectability, we conducted an experiment by human subjects using image sequences captured by an omnidirectional camera including pedestrians. From the comparison between the output of the proposed method and the ground truth of pedestrian detectability, we confirmed that the proposed method significantly reduces the prediction error in comparison with the existing methods.**

I. INTRODUCTION

In recent years, advances in pedestrian detection technology using in-vehicle cameras or sensors have led to the development of driving assistance systems that can notify the drivers of the presence of pedestrians. However, warning the driver of all visible pedestrians could be confusing and is thus prohibitive towards safe and comfortable driving. Therefore, it would be useful to develop a method to predict the detectability of pedestrians by the driver. Figure 1 shows an example of the difference of pedestrian detectability.

Recently, estimation methods of the human eye characteristics based on computer vision have been widely studied. Itti et al. [1] proposed the "saliency map" consisting of salient regions where humans may be attracted to look. It has also been used for the segmentation of foreground objects. Following Itti's research, other research groups improved the saliency map [2][3]. For example, Lee et al. [4] proposed a method for the estimation of human visual attention in a video sequence using a learning based saliency map.

As related to the saliency map, some researchers have proposed methods for estimating the visibility of road objects. The visibility value represents that how a visual object is easy to be recognized by drivers. Kimura et al. [5] proposed a method for estimating the visibility of traffic signals by evaluating the contrast of image features between a traffic signal and its background. Likewise, Doman et al. [6] proposed a visibility estimation method for traffic signs using several image features, such as textures and appearances of the target traffic sign.

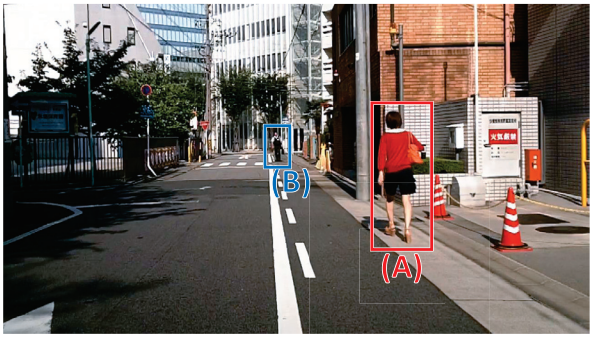


Fig. 1. Example of the difference of pedestrian detectability. Pedestrian (A) is near the camera, and is easier to detect. Meanwhile, pedestrian (B) is far from the camera, and is more difficult to detect.

Other research groups have proposed methods for predicting the pedestrian detectability. The pedestrian detectability is the probability of detecting a pedestrian by drivers. Engel et al. [7] proposed a method for predicting the pedestrian detectability using image features and the information of objects on the road . Wakayama et al. [8] proposed a method considering Visual Search [9] and pedestrian motion. They used the saliency map as a map of visual distractors and optical flow to represent the pedestrian's motion. However, these methods did not consider the influence of human visual characteristics such as fields of view, color vision, light adaptation, and so on. In practice, the human visual characteristics strongly affect the pedestrian detectability. Therefore, this paper proposes a method for predicting the pedestrian detectability adaptive to the characteristics of the human fields of view. The contributions of this paper are as follows:

1) Propose image features effective for the central and the peripheral fields of view.
2) Improve prediction accuracy by optimizing predictors for the central and the peripheral fields of view.

In the following, section II describes the basic idea and the details of the proposed method. Then, dataset construction by human subjects using omnidirectional camera images are reported in section III. Next, evaluation of the proposed method is reported in section IV. Finally, we conclude this paper in section V.

[1]Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601, Japan
[2]Information and Communications Headquarters, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601, Japan
[3]School of Engineering, Chukyo University, 101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393, Japan

II. BUILDING A PREDICTOR ADAPTIVE TO THE CHARACTERISTICS OF THE HUMAN VISUAL FIELD

A. Basic idea and overview of the proposed method

In this paper, we focus on the characteristics of the human visual field. In general, the human visual field is separated into two regions; the central and the peripheral fields of view. The central field of view mainly contributes to visual recognition. The cone cells which contribute to color vision are distributed densely in this region. On the other hand, in the peripheral region, humans cannot detect the color difference. However, the rod cells which are more sensitive to small difference of light than cone cells, are distributed in this region. These characteristics of human vision affect the pedestrian detectability. Therefore, we propose a method for predicting the pedestrian detectability according to the difference between the central and the peripheral fields of view.

The proposed method extracts features effective for each field of view and constructs predictors optimized for each of them for predicting the pedestrian detectability.

Figure 2 shows the process flow of the proposed method. The inputs are in-vehicle camera images, the positions of pedestrians, and the driver's eye gaze position. In this paper, the proposed method assumes that the position of the pedestrians are obtained by an external pedestrian detection method [10]. Then, the proposed method calculates several types of image features related to the pedestrian detectability. Finally, the pedestrian detectability is predicted by SVR (Support Vector Regression) [12] trained using these features. The following sections describe the details of the proposed method.

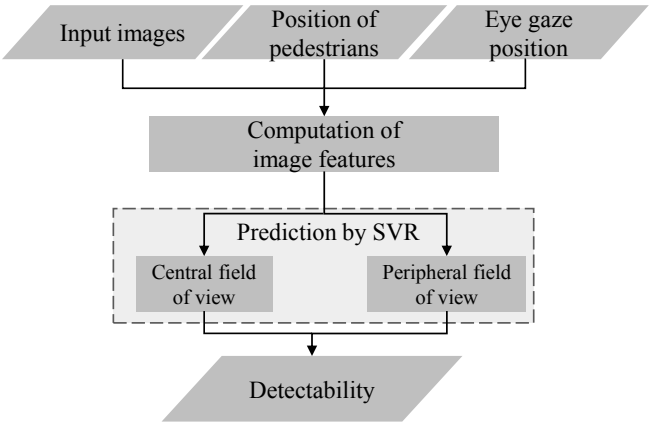


Fig. 2. Process flow of the proposed method.

TABLE I
LIST OF IMAGE FEATURES

Category	Abbreviation	Description
Features for the central field of view	$C_{\mu(\mathrm{lum})}$	Contrasts of luminance, color (RGB), and texture
	$C_{\mu(\mathrm{RGB})}$	
	C_{tex}	
Features for the peripheral field of view	P_{motion}	Movement of pedestrian region
	P_{lum}	Change of luminance of pedestrian region
	C_{flow}	Contrast of optical flow
Common features for both fields of view	P_{size}	Size of pedestrian region
	num	Number of pedestrians
	$D(p, c)$	Distance from a pedestrian region to the eye position, and to the nearest pedestrian
	$D(p, p')$	

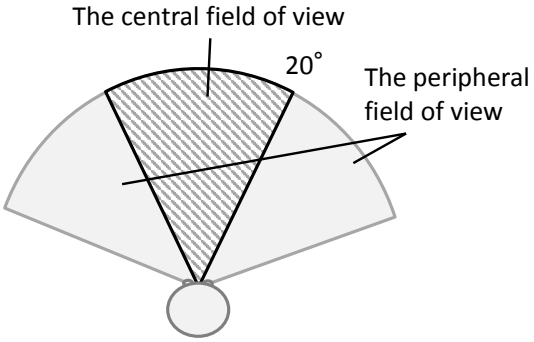


Fig. 3. Definition of the central and the peripheral fields of view.

B. Image feature extraction

The proposed method uses several image features to predict the pedestrian detectability. We assumed that the image features which are effective for the prediction might be different between the central and the peripheral fields of view. Therefore, we prepared three types of image features to represent the characteristics of the human visual field.

- Features for the central field of view
- Features for the peripheral field of view
- Common features for both fields of view

Table I shows the list of image features prepared for each field of view. The following sections describe the details of these features. Note that in this paper, the central and the peripheral fields of view are defined as shown in Fig. 3. Here, the range of the central field of view is defined as 20°.

1) Features for the central field of view: In the central field of view, a human can see objects in high resolution and can detect the color difference. Therefore, difference of the color and the texture of a pedestrian will strongly affect its detectability. We represented these characteristics by contrast features such as contrasts of luminance, color, and texture, between the pedestrian region and its surrounding region.

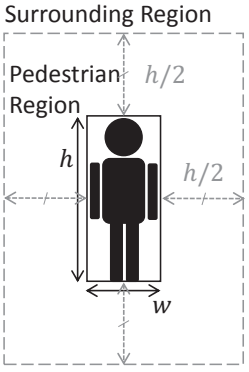


Fig. 4. Definition of the pedestrian region and its surrounding region.

The luminance contrast feature $C_{\mu(\mathrm{lum})}$ is calculated as

$$C_{\mu(\mathrm{lum})} = |P_{\mu(\mathrm{lum})} - S_{\mu(\mathrm{lum})}|, \tag{1}$$

where $P_{\mu(\mathrm{lum})}$ and $S_{\mu(\mathrm{lum})}$ are the average luminance values of the pedestrian region and its surrounding region, respectively. Here, the pedestrian region and its surrounding region are defined as shown in Fig. 4.

The color contrast feature $C_{\mu(\mathrm{color})}$ is calculated as

$$C_{\mu(\mathrm{color})} = ||P_{\mu(\mathrm{color})} - S_{\mu(\mathrm{color})}||, \tag{2}$$

where $P_{\mu(\mathrm{color})}$ and $S_{\mu(\mathrm{color})}$ are the average color values of the pedestrian region and its surrounding region, respectively. Here, $||\ ||$ represents the Euclidean norm. In the proposed method, the L*a*b* color space is used.

The texture contrast feature is extracted using a gray level co-occurrence matrix. This feature is calculated as

$$C_{\mathrm{tex}} = \sum_{a=0}^{k}\sum_{b=0}^{k}\left(M_p\left(a,b\right) - M_s\left(a,b\right)\right)^2, \tag{3}$$

where k is the size of the co-occurrence matrices, and M_p and M_s are the co-occurrence matrices of the pedestrian region and its surrounding region, respectively.

2) Features for the peripheral field of view: In the peripheral field of view, the rod cell can detect small light difference. That is, this region is sensitive to objects' motion and brightness change. Therefore, the pedestrian's motion is expected to strongly affect the pedestrian detectability in this region. We prepared the following three image features to represent the pedestrian's motion.

P_{motion} is the horizontal distance of the target pedestrian's motion, calculated as

$$P_{\mathrm{motion}} = |P_x(t+\Delta t) - P_x(t)|, \tag{4}$$

where $P_x(t)$ is the target pedestrian's horizontal position at the t-th frame. In the proposed method, the pedestrian's motion in a short period of time ($\Delta t = 5$ frames) is evaluated.

P_{lum} is the change of luminance in the pedestrian region, calculated as

$$P_{\mathrm{lum}} = |P_{\mathrm{lum}}(t+\Delta t) - P_{\mathrm{lum}}(t)|, \tag{5}$$

where $P_{\mathrm{lum}}(t)$ is the average luminance in the pedestrian region at the t-th frame. As shown in Fig. 5, we confirmed that the luminance change of the pedestrian region represents the pedestrian's motion.

C_{flow} is the contrast feature of optical flow calculated as

$$C_{\mathrm{flow}} = \frac{1}{N}\sum_{i=0}^{5}|P_{\mu(\mathrm{flow})}(t+i) - S_{\mu(\mathrm{flow})}(t+i)|, \tag{6}$$

where $P_{\mu(\mathrm{flow})}(t)$ and $S_{\mu(\mathrm{flow})}(t)$ are the optical flows in the pedestrian region and its surrounding region at the t-th frame, respectively.

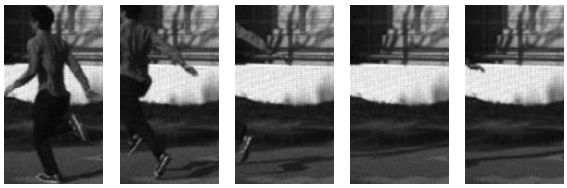


Fig. 5. Exmaple of luminance change of the pedestrian region.

3) Common features for both fields of view: We also prepared image features effective for both fields of view. In general, the closer a pedestrian exists, the bigger he/she appears to the driver. Therefore we focused on the area of the pedestrian region.

In addition to this feature, we focused on the locations of the target pedestrian and the other pedestrians. In driving situations, the more number of pedestrians exist, the more difficult it becomes to recognize all of them correctly. Besides, the human eye has a high resolution near the central field of view compared to that of the peripheral field of view. Therefore, we focused on the number of pedestrians, the distance from the target pedestrian to his/her closest pedestrian, and the distance from the target pedestrian to the driver's eye gaze.

C. Prediction of the pedestrian detectability

Detectability predictors are constructed by SVR. This section introduces an overview of the construction phase and the prediction phases.

1) Construction phase: The predictor is trained by using pairs of feature values and the ground truth of the pedestrian detectability. In addition, the proposed method aims to adapt the characteristics of the human visual field. To achieve this, the proposed method selects features effective for each field of view and constructs predictors specific to each of them. The RBF (Radial Basis Function) kernel is used in the SVR, and LIBSVM [13] is used for training the SVR.

2) Prediction phase: In the prediction phase, image features are extracted from images captured by an in-vehicle camera. Then pedestrian detectability is calculated by using the predictor specific for each field of view.

III. Dataset Construction

To construct a predictor for the pedestrian detectability, we need the ground truth. Therefore, we performed an experiment to obtain the ground truth of the pedestrian detectability. Engel et al. [7] and Wakayama et al. [8] used an in-vehicle camera to capture videos, and conducted experiments with a single display to present videos to subjects. Then they obtained the ground truth of the pedestrian detectability based on its result. However, in the proposed method, since we need to evaluate the characteristics of the human visual fields of view, the angle of a single in-vehicle camera is not sufficient. Therefore, we used an omnidirectional camera, (Ladybug5 from Point Grey Research, Inc.), to capture videos instead of an in-vehicle camera. The resolution of the

Fig. 6. Example of an image captured by an omnidirectional camera.

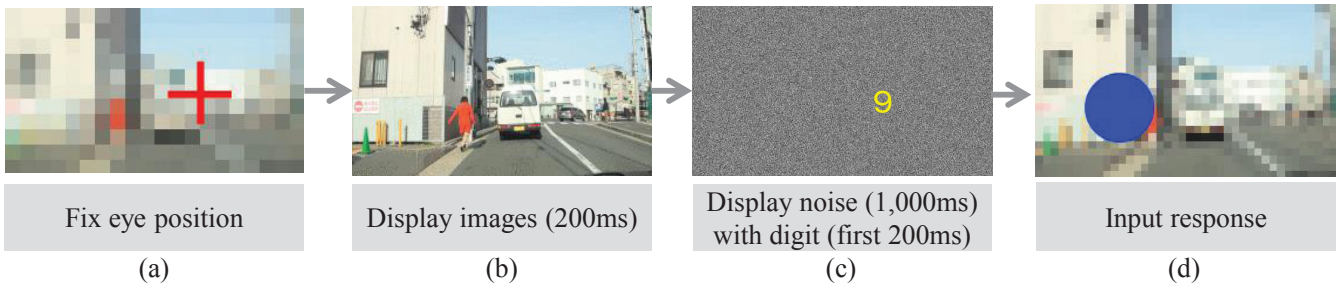


Fig. 7. Procedure to evaluate the detectability by a human subject.

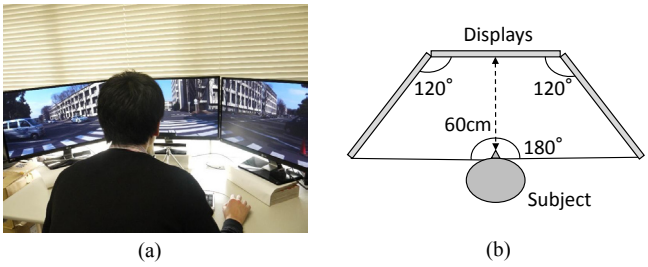


Fig. 8. Experimental setting. Three displays were used for the experiment.

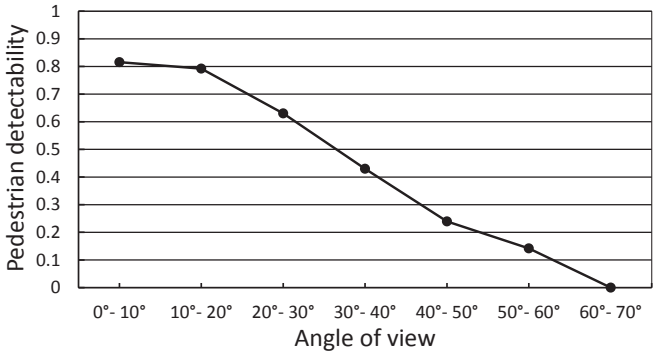


Fig. 9. Distribution of the pedestrian detectability.

videos was 5,760 × 1,200 pixels. Figure 6 shows an example of the image captured by the omnidirectional camera.

We extended Engel's experimental framework as follows. Figure 7 shows the procedure of the proposed experimental framework.

- Step (a): The subject is instructed to look at a certain position indicated by a red cross.
- Step (b): The subject is shown a sequence of images captured from an omnidirectional camera for 200 msec.
- Step (c): To reduce the influence of afterimage, the subject is shown noise images for 1,000 msec. To keep the subject's eye gaze to the initial position during Step (b), a random digit is displayed at the same position as the cross for the first 200 msec.
- Step (d): The subject is asked to report the locations of pedestrians which he/she had detected, and the displayed digit.

Prior to the experiment, the subjects were introduced to the experimental procedure, and took exercises for three times to get used to the task.

In the experiment, we assumed that the subject's eye gaze position was fixed to the initial position represented by the red cross and a random digit. However, the human's gaze position could be attracted to a salient region and may move to a different position from the initial position. Our research focus on the mechanism of human recognition without visual search. Therefore, the random digit was displayed in order to confirm that the subjects' eye gaze was fixed to the initial position during Step (b).

We performed this experiment with 13 subjects (11 males and 2 females) in their 20s. Finally, the ground truth of the pedestrian detectability was calculated as the ratio of correct answers by all subjects. In this experiment, we prepared 150 images. The number of pedestrians in each image was between 0 and 5, and 299 pedestrians in total were observed in them without occlusions.

Figure 8 shows the actual setting of the experiment. We showed the video to each subject on three 27 inch (wide) displays.

Figure 9 shows the distribution of the pedestrian detectability by the angle of view formed between the pedestrian position and the subject's gaze position. From this result, we can say that drivers can detect pedestrians in their central field of view better than their in peripheral field of view. In addition, we can see that the detectability gradually decreases according to the increase of the angle of view

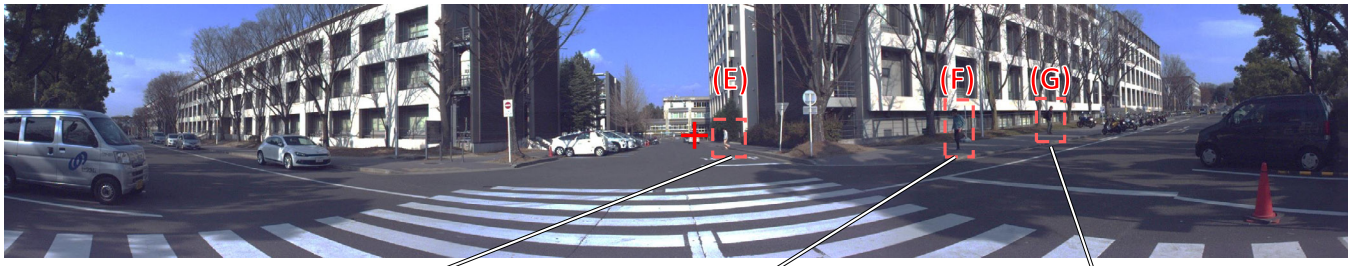
Pedestrian A	0.08
Comparative	0.33 (0.25)
Proposed	0.26 (0.18)

Pedestrian B	0.77
Comparative	0.57 (0.20)
Proposed	0.76 (0.01)

Pedestrian C	0.31
Comparative	0.50 (0.19)
Proposed	0.40 (0.09)

Pedestrian D	0.08
Comparative	0.21 (0.13)
Proposed	0.13 (0.05)

(a) Example with four pedestrians.



Pedestrian E	1.00
Comparative	0.89 (0.11)
Proposed	0.91 (0.09)

Pedestrian F	0.62
Comparative	0.25 (0.37)
Proposed	0.57 (0.05)

Pedestrian G	0.23
Comparative	0.15 (0.08)
Proposed	0.17 (0.06)

(b) Example with three pedestrians.

Fig. 10. Examples of prediction results. The cross indicates the fixation point. The first row in each table shows the ground truth of the pedestrian detectability. The second and third rows show the predicted value by the comparative method and the proposed method. The numbers in parentheses represent the prediction error (MAE).

formed between the pedestrian position and the driver's gaze position.

IV. EVALUATION AND DISCUSSIONS

To evaluate the proposed method, we compared the output of the proposed method with the ground truth. We constructed predictors for the central and the peripheral fields of view. Table II shows features effective for each field of view from the ten features introduced in section II, which were selected by a greedy algorithm. Using the predictors optimized for each field of view, we evaluated the performance of the proposed method by 10-fold cross validation. To evaluate the effectiveness of adaptation to the human visual field, we compared the prediction accuracy between the proposed method and a comparative method. The comparative method used a single predictor optimized for predicting general pedestrian detectability, in a similar way to [7] and [8].

Figure 11 and Table III show the comparison of the prediction error (Mean Absolute Error: MAE) between the proposed method and the comparative method. Figure 10 shows examples of the comparison of the prediction accuracy. Since pedestrian (A) in Fig. 10 (a) appears large, it may be easy to detect. However, since the position of pedestrian (A) is at the edge of the driver's visual field, it may be difficult to detect. Meanwhile, the position of pedestrian (F) in Fig. 10 (b) is almost the same with that of pedestrian (C). However, the detectability of pedestrian (F) is higher than that of pedestrian (C). This might be because pedestrian (F) is running, while pedestrian (C) is walking. Thus, the features of pedestrian's motion is effective for the prediction in the peripheral field of view.

TABLE II
COMPARISON OF FEATURES THAT WERE EFFECTIVE FOR THE PREDICTION

Features	Fields of view	
	Central	Peripheral
$C_{\mu(\mathrm{lum})}$	✓	✓
$C_{\mu(\mathrm{color})}$	✓	–
C_{tex}	–	–
C_{flow}	✓	–
P_{lum}	–	✓
P_{motion}	–	✓
P_{size}	–	–
num	–	✓
$D(p, c)$	✓	✓
$D(p, p')$	–	✓

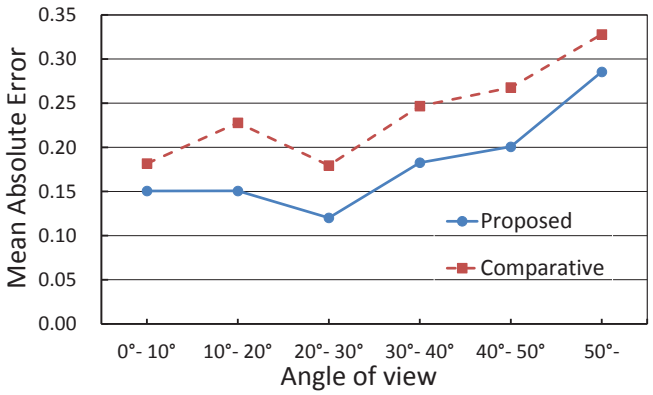


Fig. 11. Comparison of the prediction error for each angle of view.

TABLE III
COMPARISON OF THE MAE BETWEEN THE PROPOSED METHOD ADAPTIVE TO THE CHARACTERISTICS OF HUMAN VISION AND THE COMPARATIVE METHOD

Methods	Fields of view	
	Central	Peripheral
Comparative	0.200	0.269
Proposed	0.149	0.195

Table II shows the comparison of effective features for each field of view. For the prediction in the central field of view, the color feature $C_{\mu(\mathrm{color})}$ was effective while this feature was not effective for prediction in the peripheral field of view. For the prediction in the peripheral field of view, the features C_{flow} and P_{lum} which represent the target pedestrian's motion were effective. C_{flow} was also effective for prediction in the central field of view. From these results, as we expected, the effective features were different between the central and the peripheral fields of view.

From this evaluation, we confirmed that the proposed approach that considers the characteristics of the human visual field significantly contributed to improve the prediction accuracy. However, the accuracy of the proposed method in the peripheral field of view was slightly worse than that in the central field of view, as shown in Fig. 11. We consider that this is because there are many variations of pedestrians' appearance and position in the peripheral field of view. The number of training data for the peripheral field of view might have been insufficient in the experiment. To improve the prediction accuracy for pedestrians in the peripheral field of view, we need to prepare more training data.

V. CONCLUSION

This paper proposed a method for the prediction of the pedestrian detectability adaptive to human visual characteristics, especially for the central and the peripheral fields of view. The proposed method extracts image features related to the central and the peripheral field of views. Optimized predictors for each field of view were used for the prediction. Evaluation results showed that the approach that considers characteristics of the human visual field was effective for the prediction of the pedestrian detectability. In future works, we will evaluate the influence of ego-vehicle speed and the driver's age to the pedestrian detectability, since they were reported by Roge et al. that they influence the range of visual field [14]. We will also investigate features that can represent other human visual characteristics. In addition, the prediction method needs a larger pedestrian image dataset to adapt the variety of their appearance. Therefore we will conduct an experiment with more subjects and more captured images.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254–1259, November 1998.
[2] Q. Zhao and C. Koch, "Learning Saliency-based Visual Attention: A Review," Signal Processing, Vol. 93, pp. 1401–1407, June 2013.
[3] T. Judd, K.Ehinger, F. Durand, and A. Torralba, "Learning to Predict Where Humans Look," Proceedings of the 2009 IEEE International Conference on Computer Vision, pp. 2106–2113, September 2009.
[4] W. Lee, T. Huang, S. Yeh, and H. Chen, "Learning-based Prediction of Visual Attention for Video Signals," IEEE Transactions on Image Processing, Vol. 20, No. 11, pp. 3028–3038, November 2011.
[5] F. Kimura, T. Takahashi, Y. Mekada, I. Ide, H. Murase, T. Miyahara, and Y. Tamatsu, "Measurement of Visibility Conditions toward Smart Driver Assistance for Traffic Signals," Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, pp. 636–641, June 2007.
[6] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and U. Sakai, "Estimation of Traffic Sign Visibility Considering Local and Global Features in a Driving Environment," Proceedings of the 2014 IEEE Intelligent Vehicles Symposium, pp. 202–207, June 2014.
[7] D. Engel and C. Curio, "Detectability Prediction in Dynamic Scenes for Enhanced Environment Perception," Proceedings of the 2012 IEEE Intelligent Vehicles Symposium, pp. 178–183, June 2012.
[8] 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," Proceedings of the 21st IAPR International Conference on Pattern Recognition, pp. 1940–1943, November 2012.
[9] J.M. Wolfe. "Visual Search," In H. Pashler, editor, Attention, pp. 13–73, University College London Press, 1998.
[10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 886–893, June 2005.
[11] G. Paulmier, C. Brusque, V. Carta, and V. Nguyen, "The Influence of Visual Complexity on the Detection of Targets Investigated by Computer Generated Images," Lighting Research and Technology, Vol. 33, No. 3, pp. 197–205, September 2001.
[12] A.J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," Statistics and Computing, Vol. 14, No. 3, pp. 199–222, August 1998.
[13] C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," ACM Transactions on Intelligent Systems and Technology, Vol. 2, No. 27, pp. 1–27, April 2011.
[14] J. Rogé, T. Pébayle, E. Lambilliotte, F. Spitzenstetter, D. Giselbrecht, and A. Muzet, "Influence of Age, Speed and Duration of Monotonous Driving Task in Traffic on the Driver's Useful Visual Field," Vision Research, Vol. 44, pp. 2737–2744, October 2004.