Driver Adaptive Prediction for Pedestrian Detectability using In-Vehicle Camera Image

Ryunosuke Tanishige[1], Daisuke Deguchi[2], Keisuke Doman[3], Yoshito Mekada[3], Ichiro Ide[1], Hiroshi Murase[1], and Naoki Nitanda[4]

[1]: Graduate School of Information Science, Nagoya University, Japan
[2]: Information and Communications Headquarters, Nagoya University, Japan
[3]: School of Engineering, Chukyo University, Japan
[4]: DENSO CORPORATION, Japan
E-mail: tanishiger@murase.m.is.nagoya-u.ac.jp, ddeguchi@nagoya-u.jp, kdoman@sist.chukyo-u.ac.jp, y-mekada@sist.chukyo-u.ac.jp, ide@is.nagoya-u.ac.jp, murase@is.nagoya-u.ac.jp

Abstract In recent years, advances in pedestrian detection technology have resulted in the development of driving assistance systems that notify the drivers of the presence of pedestrians. However, warning of the presence of all pedestrians would confuse the driver. Therefore, the driver should only be notified of the less detectable pedestrians to avoid confusion. To achieve this, it is necessary to develop a method to predict the driver's perception performance of pedestrian detectability. This paper proposes a method that predicts the pedestrian detectability considering the difference between individual drivers. The proposed method constructs a predictor specific to each driver, in order to predict the pedestrian detectability precisely. To obtain the ground truth of the pedestrian detectability, we conducted an experiment by human subjects using images from an in-vehicle camera including pedestrians. From the comparison between the output of the proposed method and the actual detectability, we confirmed that the proposed method significantly reduces the prediction error in comparison with the existing methods.

Keywords ITS, driver assistance system, pedestrian, detectability, personalization

1 Introduction

In recent years, advances in pedestrian detection technology using in-vehicle cameras and sensors have resulted in the development of driving assistance systems that 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 driver's perception performance of pedestrian detectability. Figure 1 shows an example of the detectability of pedestrians in different conditions.

Several research groups have proposed methods for predicting the pedestrian detectability. Engel et al. [1] proposed a method for predicting the pedestrian detectability using image features and information on the structure of the road. Wakayama et al. [2] proposed a method considering Visual Search [3] and pedestrian motion. They used a saliency map [4] and motion features. The aim of these methods is to estimate the average pedestrian detectability for all drivers in general. However, in practice, the visual performance of individual drivers affects the pedestrian detectability.

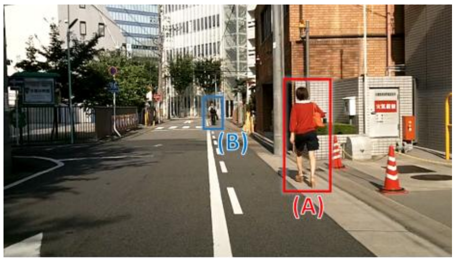

Figure 1. Example of the difference of pedestrian detectability. Pedestrian (A) is near the camera, and is easier to detect. Pedestrian (B) is far from the camera, and is more difficult to detect.

In this paper, we focus on the difference of visual performance between drivers, and propose a method for personalized prediction of the pedestrian detectability. To achieve this, we construct predictors optimized for individual drivers and predict the pedestrian detectability incorporating these predictors.

In the following, section 2 describes the details of the proposed method. Then, dataset construction by human subjects using in-vehicle camera images is reported in section 3. Next, evaluation of the proposed method is reported in section 4. Finally, we conclude this paper in section 5.

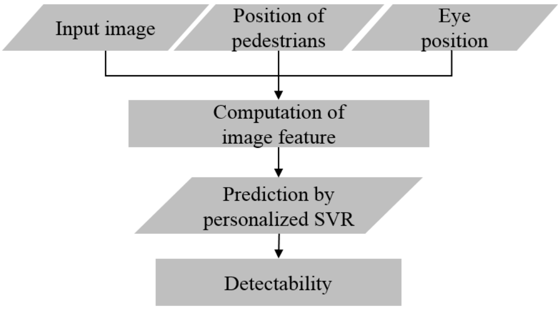

Figure 2. Process flow of the proposed method.

2 Personalized detectability prediction

Figure 2 shows the process flow of the proposed method. The input is an in-vehicle camera image, positions of pedestrians and the driver's eye position. 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) [5] trained using these features.

2.1 Features

The features used in the proposed method are categorized into:

1. Target pedestrian features
2. Contrast features
3. Global features

Table 1 lists the features used for predicting the pedestrian detectability. The following sections describe the details of these features.

2.1.1 Target pedestrian features

The proposed method calculates features on the appearance of a target pedestrian. They are extracted from within a pedestrian region as shown in Figure 3. First, the size of the pedestrian region P_{area}, P_{width}, P_{height} are extracted. Next, since the luminance of the pedestrian region may also affect the detectability, $P_{\mu(\mathrm{lum})}$ and $P_{\sigma(\mathrm{lum})}$ are calculated. Here, the proposed method assumes that the position of the pedestrian is obtained by a pedestrian detection method [6].

2.1.2 Contrast features

Contrast features are extracted by calculating the contrast between the pedestrian region and its surrounding region. As shown in Figure 3, the surrounding region is determined in proportion to the size of the pedestrian. The proposed method calculates several types of contrasts against luminance, color, edge, texture, frequency characteristics, and color histogram.

Table 1. List of features used for predicting the pedestrian detectability.

Category	Abbreviation	Description
Target pedestrian features	P_{area}	Area, width, and height of a pedestrian.
	P_{width}	
	P_{height}	
	$P_{\mu(\mathrm{lum})}$	Average, and standard deviation of luminance within a pedestrian region.
	$P_{\sigma(\mathrm{lum})}$	
Contrast features	$C_{\mu(\mathrm{lum})}$	Difference of luminance, color, edge, texture and frequency, between a pedestrian region and its surrounding region.
	$C_{\sigma(\mathrm{lum})}$	
	$C_{\mu(\mathrm{RGB})}$	
	$C_{\mu(\mathrm{Lab})}$	
	$C_{E(\mathrm{gray})}$	
	$C_{E(\mathrm{RGB})}$	
	C_{TEX}	
	C_{FFT}	
	$H_{R,G,B}$	Difference between color histograms of R,G,B and L,a,b.
	$H_{L,a,b}$	
Global features	N	The number of pedestrians in an image.
	$D_{(p,c)}$	Distance from a pedestrian to eye position, and the nearest pedestrian.
	$D_{(p,p')}$	

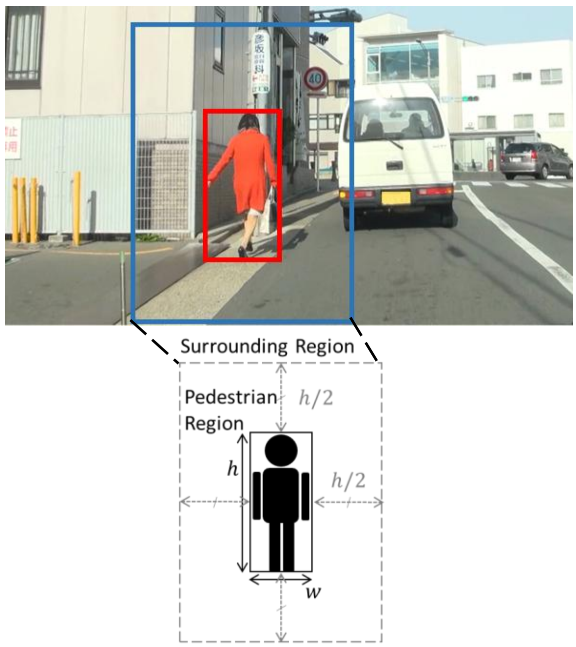

Figure 3. Definition of the surrounding region.

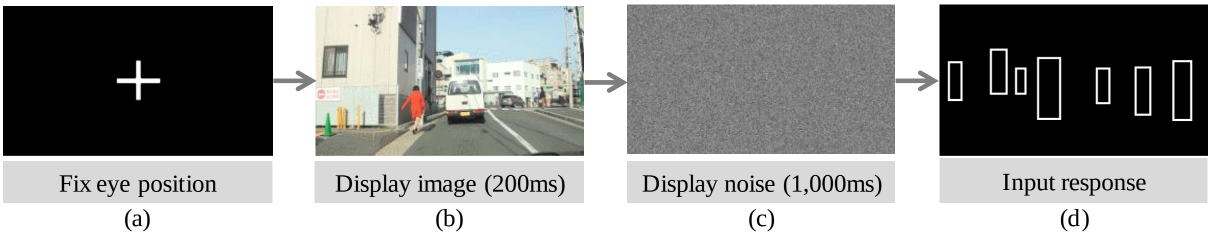

Figure 4. Experiment steps. (a) A subject fixes his/her eye direction at the center of the screen. (b) An in-vehicle camera image is displayed for 200 msec. (c) A noise image is displayed for 1,000 msec. (d) The subject inputs his/her response by selecting rectangles among multiple choices.

The luminance contrast features are calculated as the average luminance $C_{\mu(\text{lum})}$ and its standard deviation $C_{\sigma(\text{lum})}$. $C_{\mu(\text{lum})}$ is calculated as

$$C_{\mu(\text{lum})} = \left| P_{\mu(\text{lum})} - S_{\mu(\text{lum})} \right|, \quad (1)$$

where $P_{\mu(\text{lum})}$ and $S_{\mu(\text{lum})}$ are the average of luminance values within the pedestrian region and the surrounding region, respectively.
The color contrast feature is calculated as

$$C_{\mu(\text{color})} = \sqrt{\left| \vec{P}_{\mu(\text{color})} - \vec{S}_{\mu(\text{color})} \right|^2}, \quad (2)$$

where $\vec{P}_{\mu(\text{color})}$ and $\vec{S}_{\mu(\text{color})}$ are the average vectors of the color value inside the pedestrian region and the surrounding region, respectively. Here, RGB and L*a*b color space are used.
The edge contrast feature is extracted by calculating the edge strength of the pedestrian region and the surrounding region.
The texture contrast feature is extracted using gray level co-occurrence matrix. This feature is calculated as

$$C_{\text{TEX}} = \sum_{a=0}^{k} \sum_{b=0}^{k} (M_P(a,b) - M_S(a,b))^2, \quad (3)$$

where k is the size of the co-occurrence matrices, and M_P and M_S are the co-occurence matrices of the pedestrian region and the surrounding region, respectively.
The frequency contrast feature is extracted as the difference of the power spectrum strength between the pedestrian region and the surrounding region. To calculate the spectrum strength, the proposed method applies Fourier transform to the input image. This feature is calculated as

$$C_{\text{FFT}} = \sum_{u=0}^{U} \sum_{v=0}^{V} \left| F_P(u,v) - F_S(u,v) \right|, \quad (4)$$

where U and V are the size of Fourier transformed image $F_P(u,v)$ and $F_S(u,v)$ respectively.

The color histogram contrast is evaluated by calculating the distance of color histograms between the pedestrian region and the surrounding region. Here, the proposed method uses Earth Mover's Distance (EMD). This feature is calculated as

$$H = d_{\text{EMD}}(H_P, H_S), \quad (5)$$

where d_{EMD} represents the Earth Mover's Distance. H_p and H_s are color histograms of the pedestrian region and the surrounding region, respectively.

2.1.3 Global features

As global features, the proposed method evaluates the locations of the target pedestrian and other pedestrians. In a driving environment, the more number of pedestrians exist on the road, the more difficult it is to recognize all of them correctly. Therefore, two features are considered: the number of pedestrians, and the distance from the target pedestrian to his/her closest pedestrian. In addition to these features, the distance from the target pedestrian to the driver's eye position (the center of the image) is calculated. This feature was selected since human vision has a high resolution around the center of the field of vision compared to that of its periphery.

2.2 Prediction of the detectability

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

2.2.1 Construction phase

The predictor is trained by using pairs of feature values and a ground truth of the pedestrian detectability. In addition, the proposed method aims to adapt a predictor to individual drivers. To achieve this, the proposed method selects effective features for each driver and constructs predictors specific to the driver. RBF (Radial Basis Function) kernel is used in the SVR, and LIBSVM [7] is used for training the SVR.

2.2.2 Prediction phase

In the prediction phase, features are extracted from images captured by an in-vehicle camera. Then pedestrian detectability is calculated by using the predictor specific for each driver.

3 Dataset construction by human subjects

To predict the pedestrian detectability, we need its actual value. Therefore, we performed an experiment to obtain the ground truth of the detectability of

Table 2. Comparison of features effective for individual human subjects. The results show that the effective features are different between subjects.

Feature	Subject					
	A	**B**	**C**	**D**	**E**	**F**
P_{width}	✓	✓	—	✓	✓	✓
$C_{\mu(\mathrm{lum})}$	✓	—	✓	—	—	—
N	—	—	✓	✓	✓	—

Table 3. The result of the MAE of predicted pedestrian detectability. This table compares the proposed method with personal adaptation and the comparative method [1,2] without personal adaptation.

Method	Subject					
	A	**B**	**C**	**D**	**E**	**F**
Comparative	0.190	0.194	0.203	0.185	0.222	0.206
Proposed	0.172	0.184	0.196	0.175	0.204	0.195

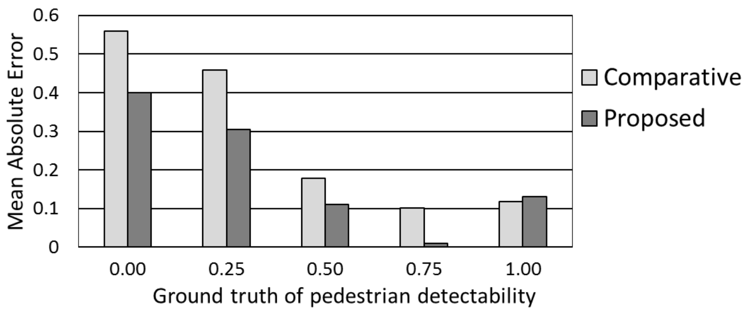

Figure 5. Comparison of the prediction accuracy by MAE for subject **E** between the proposed method and the comparative method.

pedestrians. Engel et al. [1] and Wakayama et al. [2] conducted experiment with several human subjects, then decided the ground truth of the pedestrian detectability by taking the average of the correct answer rate among subjects. However, in the proposed method, we need the ground truth for individual subjects. Therefore, we extended their experimental framework as follows.

Figure 4 shows the flow of the proposed experimental framework. At first, a subject was instructed to fix his/her point of view at the center of the screen. Then the subject was shown an image captured from an in-vehicle camera for 200 msec. After that, to reduce the influence of afterimage, the subject was shown noise images for 1,000 msec. Finally, the subject was asked to respond the locations of pedestrians by selecting rectangles containing pedestrians in the image.

We performed this experiment with six male subjects in their 20s. Every subject took the experiment for four times. Finally, the ground truth of the pedestrian detectability was calculated as the ratio of correct answers by each subject. In this experiment, we prepared 200 images whose sizes were 1,280 × 720 pixels. The number of pedestrians in each image was between 0 and 4, and 271 pedestrians in total were observed in the images without occlusions.

4 Experiments and Discussion

To evaluate the proposed method, we compared between the output of the proposed method and the actual detectability. We constructed predictors for individual subjects by their own pedestrian detectability and effective features selected for them from 18 features shown in Table 1. Using a personalized predictor, we evaluated the performance of the proposed method by 10-fold cross validation. To evaluate the effectiveness of personalization, we compared the prediction accuracy between the proposed method and a comparative method that uses a non-personalized predictor trained by the average of all subjects' results [1,2].

Table 2 shows the comparison of effective features for each subject. From this result, we confirmed that effective features were different between drivers; While some features (e.g. P_{width}) were effective for overall performance, some others were effective to evaluate the difference between drivers' visual performance.

Table 3 shows the accuracies of the predicted pedestrian detectability between the proposed method and the comparative methods. As can be seen in the table, the effectiveness of personalization was different between individual subjects. Meanwhile, Figure 5 shows a comparison of prediction accuracies between the proposed and the comparative method. This graph shows the Mean Absolute Error (MAE) between the ground truth of the pedestrian detectability and the predicted value for subject **E**. As can be seen in the graph, the proposed method showed more effect for low detectability pedestrians.

From these results, we confirmed that the proposed personalization for individual drivers significantly contributed to improve the prediction accuracy. However, the accuracy of the proposed method might be able to be improved by considering other human visual property.

5 Conclusion

This paper proposed a method for personalized pedestrian detectability prediction from in-vehicle camera image. To improve the accuracy, the proposed method considered differences between individual drivers. Evaluation results showed that the adaptation for a driver is effective for the prediction of the pedestrian detectability. Future works include: (1) investigation of features that can represent the difference of drivers, and (2) evaluation of the proposed method through larger experiment with many subjects.

Acknowledgement

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

Reference

[1] D. Engel and C. Curio, “Detectability Prediction in Dynamic Scenes for Enhanced Environment Perception,” *Proceedings of the IEEE Intelligent Vehicles Symposium 2012*, pp.178–183, Alcalá, Spain, June 2012.

[2] 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 International Conference on Pattern Recognition*, pp.1940–1943, Tsukuba, Japan, November 2012.

[3] J. M. Wolfe. “Visual Search,” In H. Pashler, editor, *Attention*, pp.13–73. University College London Press, 1998.

[4] 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.

[5] A. J. Smola and B. Schölkopf, “A Tutorial on Support Vector Regression.” *Statistics and Computing*, vol.14, no.3, pp.199–222, August 1998.

[6] 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, San Diego, USA, June 2005.

[7] 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.