21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Estimation of the human performance for pedestrian detectability based on visual search and motion features

Masashi Wakayama
Graduate School of Information Science, Nagoya University

Daisuke Deguchi
Strategy Office, Information and Communications Headquarters, Nagoya University

Keisuke Doman, Ichiro Ide, Hiroshi Murase
Graduate School of Information Science, Nagoya University

Yukimasa Tamatsu
DENSO CORPORATION

Abstract

This paper proposes a method for estimating the human performance of pedestrian detectability from in-vehicle camera images in order to warn a driver of the positions of pedestrians in an appropriate timing. By introducing features related to visual search and motion of the target, the proposed method estimates the detectability of pedestrians accurately. Support Vector Regression (SVR) is used to estimate the detectability. Here, SVR is trained using features calculated by the proposed method with the ground truth obtained through experiments with human subjects. From experiments using in-vehicle camera images, we confirmed that the proposed features were effective to estimate the detectability of pedestrians.

1. Introduction

In recent years, driving safety support systems are becoming important to prevent car accidents. One of the most important functions of such systems is to warn a driver of the positions of pedestrians by making a sound or indicating on a head-up display. Figure 1 shows an illustration of a driver's vision with two pedestrians that have different detectabilities. As can be seen in the image, since pedestrian A can be observed in a large size at the center of the image, it seems that it is easy to be detected by a driver. In contrast, it seems to be difficult to detect pedestrian B due to its small size and complex background. Since the driver may not be able to detect pedestrian B, he/she cannot cope with a sudden action of the pedestrian B. From these points of view, it can be considered that warning systems based on the detectability of the pedestrian will be important. However, over-warnings may decrease the concentration of the driver. Therefore, it is important to develop a method to select appropriate information for driving and to provide them to a driver. In this paper, to provide useful information that is important to prevent collisions with pedestrians, we focus on the detectability estimation of a pedestrian from in-vehicle camera images.

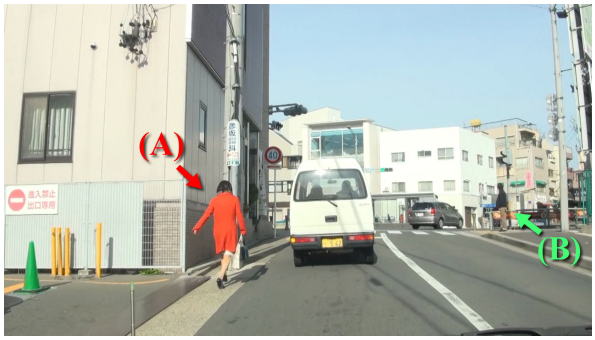

Figure 1. Examples of various appearances of pedestrians.

Several research groups have proposed methods for estimating the detectability (or the visibility) of pedestrians, traffic signals and traffic signs [3, 4, 6]. Kimura et al. proposed a method for estimating the visibility of traffic signals by evaluating the contrast of image features between a traffic signal and its surroundings [6]. Doman et al. extended this idea to the estimation of traffic signs [3]. Engel et al. proposed a method for estimating the detectability of pedestrians from a still image [4]. In this research, the detectability of pedestrians was estimated by Support Vector Regression (SVR) [1] using several types of image features extracted from a

978-4-9906441-1-6

still image. To obtain the ground truth of the detectability of pedestrians, they conducted experiments with human subjects. First, an in-vehicle camera image was displayed for an instant to each human subject, and they were asked the positions of pedestrians that can be seen in the image. Finally, the detectability of each pedestrian was computed as the percentage of pedestrian detected by human subjects. However, all of the above methods used still images for estimating the detectability (or the visibility) of targets, but the use of motion features have not been considered. In addition, although the detection of the targets by drivers is strongly related to the task of visual search, this was not considered either.

Therefore, this paper proposes a method for estimating the detectability of pedestrians by using an in-vehicle camera image sequence. The main contributions of this paper are as follows:

1. Estimation of the detectability by using a visual search feature that is strongly related to human perception.
2. Introduction of motion feature for estimating the detectability of pedestrians.

In the following, section 2 describes the details of the proposed method. Then, experiments using in-vehicle camera images are reported in section 3. Finally, we will conclude this paper in section 4.

2. Detectability estimation method

Figure 2 illustrates the processing flow of the proposed method. As seen in the figure, the proposed method computes several types of image features from an image sequence during a short period. Then, SVR based on these features is used for estimating the detectability of pedestrians. The following sections describe the details of this process.

2.1. Computation of features

In this paper, image features are categorized into five categories as

(a) Visual search feature: $\mathcal{D}$,
(b) Motion feature: $\mathcal{M}$,
(c) Contrast feature: $\mathcal{C}$,
(d) Others: $\mathcal{O}$.

To indicate each image feature in a compact style, this paper uses a combination of a category and a feature name (e.g. "$\mathcal{C}_{\text{hist (color)}}$" for the feature corresponding to the color histogram). The following sections introduce the overview of these features.

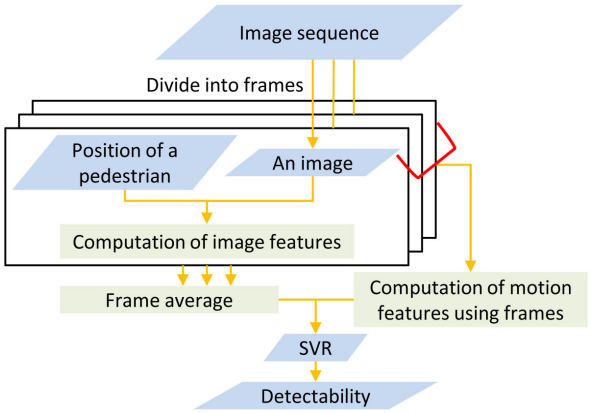

Figure 2. The processing flow of the proposed method.

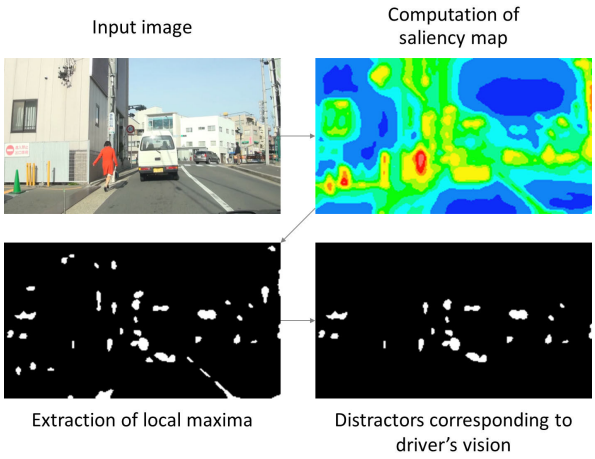

Figure 3. The procedure of distractor extraction.

2.1.1 Visual search features

Visual search is a task to find target objects from an image. It is known that this task is affected by the number of distractors observed in the visual field [7]. That is, if the number of distractors increases, the difficulty of finding a target object increases. Especially, its difficulty is strongly affected by the increase of distractors similar to the target. Based on this assumption, the proposed method evaluates the influence of distractors. Figure 3 shows the procedure of the distractor extraction. First, the proposed method extracts distractors by using the saliency map proposed by Itti et al. [5]. Then, the proposed method computes the difference $\mathcal{D}_{\text{dist}}$ of saliencies between the target and each distractor, and the number of distractors $\mathcal{D}_{\text{num}}$ from this map.

2.1.2 Motion feature

The motion of a target is one of the factors to enhance its detectability. Therefore, the proposed method extracts motion feature $\mathcal{M}_{\text{flow}}$ from an image sequence. First, optical flows are extracted from both pedestrian and its surrounding regions. Then, a histogram of optical flows inside each region is calculated. Finally, $\mathcal{M}_{\text{flow}}$ is calculated as a histogram intersection between the two histograms.

2.1.3 Contrast features

Kimura et al. reported that the contrast between a target and its surroundings is one of the key factor to affect the visibility of traffic signals [6]. Based on this idea, the proposed method extracts several types of contrast features from an in-vehicle camera image, and uses them for evaluating the detectability of pedestrians.

At first, the proposed method computes four image features from an in-vehicle camera image directly, such as the difference of luminances $\mathcal{C}_{\text{lum}}$, the difference of the standard deviation of luminances $\mathcal{C}_{\text{std(lum)}}$, the distance in L*a*b* color space $\mathcal{C}_{\text{color}}$, and the difference of edge strength $\mathcal{C}_{\text{edge}}$. Next, the histogram intersection is computed. Here, color histogram $\mathcal{C}_{\text{hist(color)}}$, and Histogram of Oriented Gradients (HOG) $\mathcal{C}_{\text{hist(HOG)}}$ are used. Finally, the contrast in frequency domain is computed. The difference of amplitude spectrum $\mathcal{C}_{\text{freq}}$ is used as the contrast feature.

2.1.4 Other features

In addition to the features described above, the proposed method computes the following six features. Three features related to a pedestrian are computed, which are the number of pedestrians $\mathcal{O}_{\text{num}}$, the area of a pedestrian $\mathcal{O}_{\text{area}}$, and the aspect ratio of a bounding box of a pedestrian $\mathcal{O}_{\text{asp}}$. Also, the proposed method computes the distance between the initial eye position $\boldsymbol{c}$ and the target $\boldsymbol{p}$ as $\mathcal{O}_{\text{d}(\boldsymbol{p},\,\boldsymbol{c})}$, and the distance between the target $\boldsymbol{p}$ and its closest pedestrian $\boldsymbol{p'}$ as $\mathcal{O}_{\text{d}(\boldsymbol{p},\,\boldsymbol{p'})}$. In addition, the difference of brightness between the target and the surroundings of the initial eye position is computed as $\mathcal{O}_{\text{lum}(\boldsymbol{p}\text{ and }\boldsymbol{c})}$.

2.2. Estimation of detectability

In the learning stage, parameters for the SVR are trained by using the features described in section 2.1. In the estimation stage, the same image features are calculated from an image sequence, and the detectability of a pedestrian is calculated by using the SVR constructed in the learning stage. Here, RBF kernel is used in the SVR, and LIBSVM is used to train the SVR [2].

Table 1. Acuracy of the constructed SVR and the importance of image feature.

# of features (After removal)	Removed feature	Accuracy	
		$\mathcal{R}^2$	Error
16	—	0.276	0.240
15	$\mathcal{C}_{\text{freq}}$	0.299	0.236
14	$\mathcal{C}_{\text{edge}}$	0.302	0.235
13	$\mathcal{O}_{\text{lum}(\boldsymbol{p}\text{ and }\boldsymbol{c})}$	0.298	0.235
12	$\mathcal{C}_{\text{lum}}$	0.299	0.234
11	$\mathcal{C}_{\text{color}}$	0.305	0.233
10	$\mathcal{C}_{\text{hist(HOG)}}$	0.316	0.231
9	$\mathcal{O}_{\text{asp}}$	0.318	0.228
8	$\mathcal{C}_{\text{std(lum)}}$	0.322	0.226
7	$\mathcal{D}_{\text{num}}$	0.315	0.226
6	$\mathcal{M}_{\text{flow}}$	0.319	0.227
5	$\mathcal{C}_{\text{hist(color)}}$	0.314	0.227
4	$\mathcal{D}_{\text{dist}}$	0.296	0.228
3	$\mathcal{O}_{\text{num}}$	0.290	0.228
2	$\mathcal{O}_{\text{d}(\boldsymbol{p},\boldsymbol{p'})}$	0.239	0.236

$\mathcal{O}_{\text{area}}$ and $\mathcal{O}_{\text{d(p,c)}}$ are remained in the end.

3. Experiments and discussions

3.1. Experimental setup

To obtain the ground truth of the detectability of pedestrians, we extend the framework proposed by Engel et al. [4]. Instead of a still image used in Engel's experiment, we use an image sequence capturing pedestrians during a short period. Figure 4 shows the procedures used in the experiment. At first, the human subjects fixed their eye direction at the center of the screen. After they watched an image sequence for 200 msec., we asked the positions of pedestrians that they could recognize. Finally, the detectability of each pedestrian was evaluated as the percentage of human subjects who could correctly recognize its position. In this experiment, we prepared 585 image sequences whose sizes were $1,280 \times 720$ pixels. The number of pedestrians in each image sequence was between 0 and 4, and 864 pedestrians in total were observed in the image sequences without occlusions. A bounding box for each pedestrian was marked manually.

3.2. Results

We evaluated the performance of the proposed method by changing the number of features used for es-

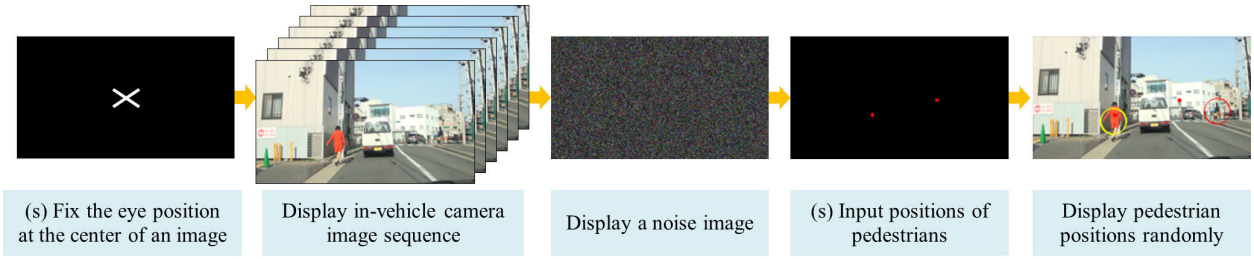

Figure 4. The procedure to evaluate the detectability by a human subject. (s) indicates the subject's action.

timating the detectability. First of all, we constructed a feature pool containing N ($N = 15$) features described in section 2.1. Next, SVR was trained by using $N - 1$ features (one feature removed from the pool). N types of SVR were trained by changing the feature removed from the pool. In this step, we obtained a SVR that has the maximum $\mathcal{R}^2$ evaluated by

$$\mathcal{R}^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (1)$$

where f_i is the output of the SVR, y_i is the detectability of a target, and $\bar{y}$ is the average of y_i. Then, we removed the feature not included in the construction of this SVR. By repeating this process, we evaluated the importance of each feature. That is, important features remain in the pool until the end of this procedure.

Table 1 shows the result of this experiment. Features $\mathcal{M}_{\text{flow}}$, $\mathcal{D}_{\text{num}}$, and $\mathcal{D}_{\text{dist}}$ proposed in this paper are used for constructing an SVR whose $\mathcal{R}^2$ is maximized. From this result, we concluded that visual search and motion features are effective to estimate the detectability of pedestrians. In our future work, we will investigae other features to improve the accuracy.

4. Conclusions

This paper proposed a method for estimating the human performance of pedestrian detectability from in-vehicle camera images in order to warn a driver of the positions of pedestrians appropriately. To improve the accuracy, the proposed method introduced features related to visual search and the motion of the target, and SVR was used to estimate the detectability of pedestrians. To evaluate the performance, the ground truth of the detectability was obtained through an experiment with human subjects. Then, we evaluated the effectiveness and the importance of features proposed in this paper. Experimental results showed that the proposed features were effective to estimate the detectability of pedestrians. Future works include: (i) investigation of features related to scene context, and (ii) evaluation by applying the method to many more cases.

ACKNOWLEDGMENTS

Parts of this research were supported by a Grant-in-Aid for Young Scientists from MEXT, a Grant-In-Aid for Scientific Research from MEXT, a Grant-in-Aid for JSPS Fellows, and a Core Research for Evolutional Science and Technology (CREST) project of JST, the Research Grant (K23-XVI-373) from Kayamori Foundation of Informational Science Advancement. MIST library (http://mist.murase.m.is.nagoya-u.ac.jp/) was used for developing the proposed method.

References

[1] D. Basak, S. Pal, and D. C. Patranabis. Support vector regression. *Neural Information Processing –Letters and Reviews*, 11(10):203–224, October 2007.

[2] C.-C. Chang and C.-J. Lir. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, April 2011.

[3] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, H. Murase, and Y. Tamatsu. Estimation of traffic sign visibility toward smart driver assistance. *Proceedings of the 2010 IEEE Intelligent Vehicles Symposium*, pages 45–50, June 2010.

[4] D. Engel and C. Curio. Pedestrian detectability: Predicting human perception performance with machine vision. *Proceedings of the 2011 IEEE Intelligent Vehicles Symposium*, pages 429–435, June 2011.

[5] 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*, 20(11):1254–1259, November 1998.

[6] 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*, pages 636–641, June 2007.

[7] J. M. Wolfe. Visual search. In H. Pashler, editor, *Attention*, pages 13–73. University College London Press, 1998.