2015 IEEE Intelligent Vehicles Symposium (IV)
June 28 - July 1, 2015. COEX, Seoul, Korea

Pedestrian Orientation Classification Utilizing Single-Chip Coaxial RGB-ToF Camera

Fumito Shinmura[1], Yasutomo Kawanishi[2], Daisuke Deguchi[3], Ichiro Ide[2], Hiroshi Murase[2], and Hironobu Fujiyoshi[4]

Abstract— **This paper proposes a method for pedestrian orientation classification. In image recognition, the accuracy is often degraded by the influence of background. In addition, it is also difficult to remove the background and extract only the human body from an image. To overcome these problems, we utilize a single-chip RGB-ToF camera. This camera can acquire RGB and depth images along the same optical axis at the same moment, and thus segmentation of the RGB image becomes easier by using the coaxial depth image. Our proposed method segmented a human body from its background accurately, which lead to the improvement of the accuracy of pedestrian orientation classification.**

I. INTRODUCTION

In order to reduce traffic accidents and assist safety-driving, pedestrian recognition methods from in-vehicle camera images have been studied widely. For example, they are applied to avoid collision with pedestrians [1]. In order to improve the collision avoidance with a pedestrian, it is important to predict his/her orientation [2]; If a system could predict the location of a pedestrian in the next few seconds, we can prevent a collision earlier. In order to predict the location, his/her orientation is important, since it is related to his/her walking direction (Fig. 1).

Various methods to classify a pedestrian's orientation from an image have been proposed. They can be divided into two types; Methods using motion features and those using static features. To classify a pedestrian's orientation whenever he/she is not walking, say, when he/she is waiting to cross a street, we take the approach using static features. Note that in this paper, we define pedestrian's orientation as eight directions in a similar way as in other works [3] (Fig. 2).

The pedestrian orientation classification problem has been attempted by many research groups. This problem is challenging due to the various appearances of pedestrians and various surrounding environments. Especially, since the background texture affects the classification significantly, it has prevented accurate classification in previous methods. Although separating the human body from the background should solve this problem, it is difficult due to textural complexity. Actually, background removal is not only a problem in pedestrian orientation classification, but also a challenging problem in the image recognition field in general.

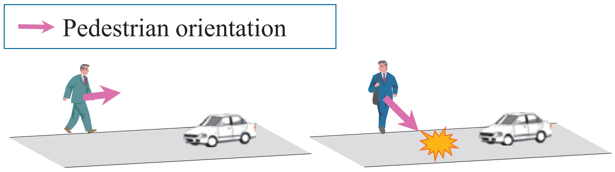


Fig. 1. Relationship between pedestrian orientation and collision with vehicle. When a pedestrian is facing towards the road (right), there is a risk of collision with a vehicle.

Meanwhile, a new RGB-ToF camera has been developed thanks to progress in sensor technology. The newly available single-chip RGB-ToF camera can capture both RGB and depth images along the same optical axis. This means that the RGB and depth values can be obtained from the exactly same point at the same moment. Segmentation from depth image will be equivalent to segmentation from RGB image, where the former is easier than the latter since background texture is negligible in depth image. For this reason, human body cropping becomes easier by utilizing a single-chip RGB-ToF camera, hence, the classification of pedestrian orientation could become easier and more accurate.

In summary, the contribution of this paper is the improvement of the accuracy of pedestrian orientation classification by background removal utilizing the newly available single-chip RGB-ToF camera.

II. RELATED WORK

Various methods have been proposed for pedestrian orientation classification. Many works that combine image feature extraction and supervised learning have been reported. Weinrich et al. used Histogram of Oriented Gradients (HOG) features [4] and a decision tree with Support Vector Machines (SVMs) for the classification [5], and classified pedestrian's orientation into eight directions. The HOG feature was used to represent the appearance of a pedestrian. The pedestrian's orientation was classified by using SVM learned the differences in body shape for each orientation. Meanwhile, Tao and Klette proposed a method using random forest as a classifier and used body parts selectively for training [3]. Those methods classify pedestrian's orientation with high

[1] F. Shinmura is with the Institute of Innovation for Future Society, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. shinmuraf@murase.m.is.nagoya-u.ac.jp

[2] Y. Kawanishi, I. Ide and H. Murase are with the Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. {kawanishiy, ide, murase}@is.nagoya-u.ac.jp

[3] D. Deguchi is with the Information & Communications, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. ddeguchi@nagoya-u.jp

[4] H. Fujiyoshi is with the Department of Robotics Science and Technology, Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-8501, Japan. hf@cs.chubu.ac.jp

978-1-4673-7265-7/15/$31.00 ©2015 IEEE

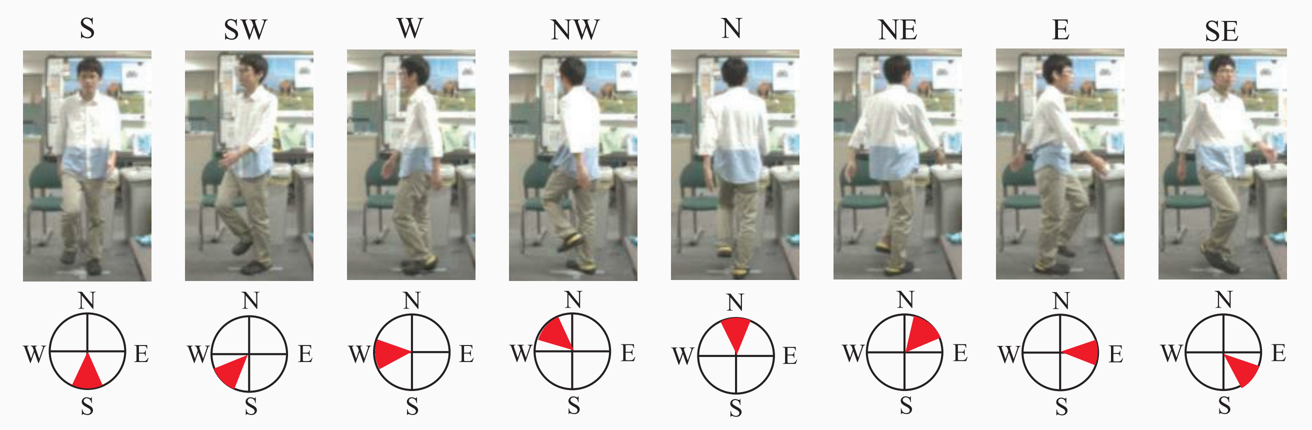


Fig. 2. Example of pedestrian orientation.

accuracy by improving both image feature and classifier training.

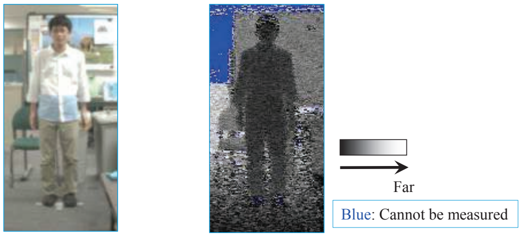
However, the accuracy of the classification is degraded due to various environments. To circumvent this problem, Hattori et al. roughly separated foreground regions from the background using disparity values obtained from stereo images [7], and used them for pedestrian detection to remove background regions. However, since foreground segmentation from disparity values was not accurate, they also used features of background regions.

Also, there is another approach that realizes the pedestrian detection and the orientation classification simultaneously. Goto et al. proposed an appearance feature named FIND, and their method detected a pedestrian for each orientation by a cascade detector using the proposed FIND feature and HOG feature [6]. However, since the features trained for the pedestrian detection and that for the orientation classification are different, the orientation classification by training the latter is expected to be more accurate.

III. OVERVIEW OF THE PROPOSED METHOD

This proposes a method to classify a pedestrian's orientation from the pedestrian detected beforehand. The proposed method takes an approach that reduces the influence of backgrounds utilizing an RGB-ToF camera. The proposed method separates foreground regions from the background referring to depth values obtained from an RGB-ToF camera. As RGB and depth images are captured along the same optical axis, foreground regions can be separated more accurately than using stereo images. Therefore, the influence of the background regions are expected to be reduced more effectively.

We utilize a newly available single-chip coaxial RGB-ToF (RGB-D) camera that uses Panasonic MN34901TL sensor for pedestrian orientation classification. This camera can coaxially acquire an RGB image and an infrared image at the frame rate of 24 fps. It can also measure the target depth by the Time-of-Flight (ToF) principle using infrared light. This camera allows us to use spatially aligned RGB and depth data simultaneously. The actual images acquired by this camera is shown in Fig. 3.



(a) RGB image. (b) Depth image.

Fig. 3. Example of images acquired from the RGB-ToF camera.

The process flow of the proposed method is shown in Fig. 4. Assuming that a pedestrian is detected beforehand, the cropped pedestrian image is input to the process flow. The training phase consists of the following four steps:

1) Noise reduction of the depth image.
2) Cropping of human body using the depth image.
3) Computation of HOG feature in the human body.
4) Construction of a multi-class SVM for classifying the pedestrian's orientation.

The classification phase also consists of the following four steps: The same three steps as the training phase (from step 1 to 3), and the classification of the pedestrian's orientation.

The output of the proposed method is the pedestrian's orientation classified in eight directions as shown in Fig. 2.

As a preprocessing step, the proposed method first reduces noise observed in the depth image since ToF sensor is usually affected sensor noise. In order to crop the human body from a depth image, the outline of the human body should be preserved after noise removal. Since noise in a depth image is large, it is difficult to reduce it by a conventional smoothing filter. Therefore, the proposed method employs the cross-bilateral filter [8] using both RGB and depth images for smoothing. This filter uses the RGB image that is spatially aligned with the depth image, because we observed relatively smaller noise in the RGB image than that in the depth image.

Next, the proposed method crops the human body from

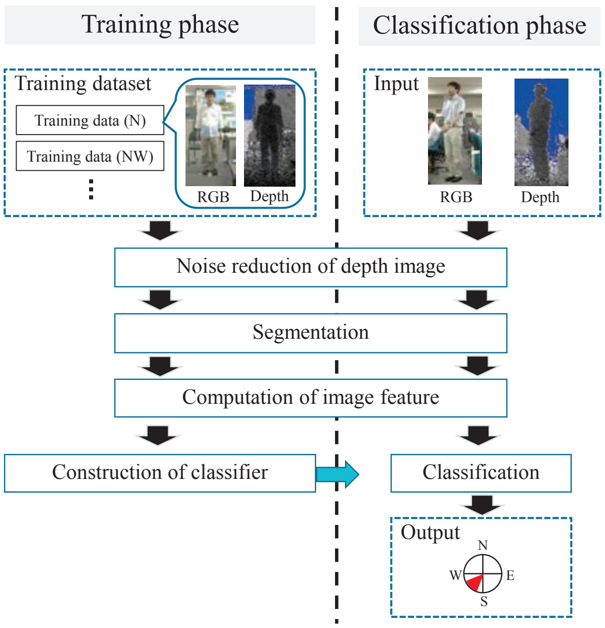


Fig. 4. Process flow of the proposed method.

the depth image. Since there is enough difference between in-pedestrian depth and pedestrian-to-background, it is easy to separate pedestrians and the background by the depth information.

Then, HOG feature is computed from the RGB image of the human body. Although the HOG feature can represent the shape of a human body, it is strongly affected by the background textures. However, these could be reduced since the background regions are removed in the previous process.

Finally, pedestrian orientation is classified into eight directions (S, SW, W, NW, N, NE, E, SE). Accordingly, a multi-class SVM is used as the classifier. Details of the training and classification phase are described in the following sections.

IV. Training Phase

In the training phase, the proposed method constructs a classifier for pedestrian orientation classification, as shown in the left side of Fig. 4.

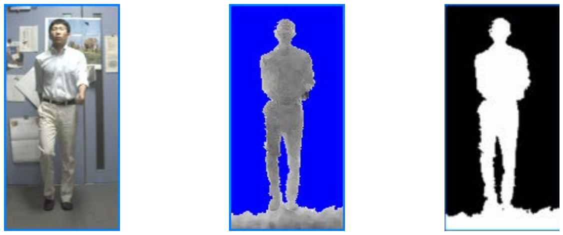
A. Noise reduction of the depth image referring to the RGB image

Noise reduction using coaxial RGB-D characteristics can be formulated as a cross-bilateral filter [8] as

$$F(\boldsymbol{x}) = \frac{\displaystyle\sum_{\boldsymbol{x}' \in N(\boldsymbol{x})} w_d(\boldsymbol{x}, \boldsymbol{x}') w_v(g(\boldsymbol{x}), g(\boldsymbol{x}')) f(\boldsymbol{x})}{\displaystyle\sum_{\boldsymbol{x}' \in N(\boldsymbol{x})} w_d(\boldsymbol{x}, \boldsymbol{x}') w_v(g(\boldsymbol{x}), g(\boldsymbol{x}'))}, \tag{1}$$

$$w_d(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{2\sigma_1^2}), \tag{2}$$

$$w_v(g(\boldsymbol{x}), g(\boldsymbol{x}')) = \exp(-\frac{(g(\boldsymbol{x}) - g(\boldsymbol{x}'))^2}{2\sigma_2^2}), \tag{3}$$



(a) Input RGB image. (b) Input depth image. (c) Segmentation result.

Fig. 5. Example of segmentation result.

where $\boldsymbol{x}$ is the coordinates of a target pixel, $N(\boldsymbol{x})$ denotes neighborhood pixels of $\boldsymbol{x}$, σ_1 and σ_2 are smoothing parameters, respectively. Functions $f(\cdot)$ and $g(\cdot)$ represent the pixel values of the depth image and its corresponding RGB image. Function $w_d(\cdot)$ is the weight function assigned according to the spatial distance, and $w_v(\cdot)$ is the weight function assigned according to the difference of pixel values, as defined in (2) and (3).

B. Segmentation of the human body region in the depth image

When a bounding box of a human body is obtained, the depth value at the center of the bounding box must be the reference depth value of the human body. Pixels with depth values similar to the reference are considered as part of the human body. Since the RGB and depth images are coaxially acquired by the RGB-ToF camera, the segmentation result obtained from the depth image can be applied to the RGB image directly. Fig. 5 shows an example of a segmentation result.

C. Computation of the HOG feature with background removal

HOG feature is used as the image feature. It is computed from the RGB image as in [4]; The magnitude of an image gradient is computed as

$$m(\boldsymbol{x}) = \sqrt{\left(\frac{\mathrm{d}}{\mathrm{d}x} g(\boldsymbol{x})\right)^2 + \left(\frac{\mathrm{d}}{\mathrm{d}y} g(\boldsymbol{x})\right)^2}, \tag{4}$$

where $\boldsymbol{x}$ is the image coordinates of the pixel in focus, $m(\boldsymbol{x})$ is the gradient magnitude at $\boldsymbol{x}$. $\frac{\mathrm{d}}{\mathrm{d}x} g(\boldsymbol{x})$ and $\frac{\mathrm{d}}{\mathrm{d}y} g(\boldsymbol{x})$ are horizontal and vertical intensity gradients at $\boldsymbol{x}$, respectively.

The proposed method eliminates voting from background pixels by modifying (4) as

$$m'(\boldsymbol{x}) = B(\boldsymbol{x}) \sqrt{\left(\frac{\mathrm{d}}{\mathrm{d}x} g(\boldsymbol{x})\right)^2 + \left(\frac{\mathrm{d}}{\mathrm{d}y} g(\boldsymbol{x})\right)^2}, \tag{5}$$

$$B(\boldsymbol{x}) = \begin{cases} 1 & (\text{if } \boldsymbol{x} \text{ is theforeground}) \\ 0 & (\text{otherwise}). \end{cases} \tag{6}$$

D. Construction of the pedestrian orientation classifier

A multi-class SVM classifier for eight directions is constructed. The classifier consists of multiple one-against-one sub-classifiers trained by LIBSVM [9]. Training data for each orientation were prepared, and the classifier learns the features extracted from them.

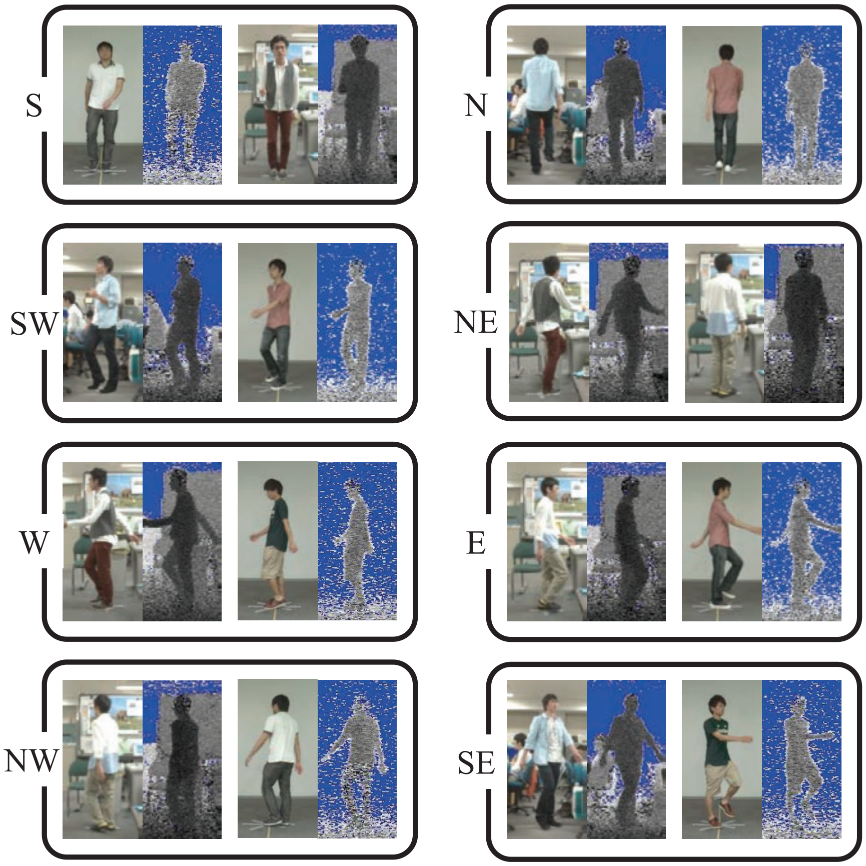


Fig. 6. Example of RGB and depth images used in the experiment.

TABLE I

EXPERIMENTAL RESULTS

Method	Accuracy
With background removal (Proposed method)	84.7 %
Without background removal	68.0 %

V. CLASSIFICATION PHASE

In the classification phase, the pedestrian's orientation is predicted from an input image, as shown in the right side of Fig. 4.

By applying the same procedure as in the training phase, noise of depth image is reduced, human body is cropped from the depth image, and image features are computed from input RGB and depth images by applying background removal. Finally, the pedestrian's orientation is classified from the computed features by using the classifier constructed as described in Section IV.

In order to classify into eight directions, 28 one-against-one classifiers are applied to an input data, and then a voting process is applied [9]. A class with the maximum number of votes is selected as the output.

VI. EXPERIMENT

An experiment on pedestrian orientation classification was conducted in order to evaluate the effectiveness of the proposed method.

A. Dataset

For the experiment, we prepared RGB and depth images captured indoors by the single-chip RGB-ToF camera from approximately four meters away. The resolutions of the RGB and depth images were 640×480 pixels and 320×240 pixels, respectively. Bounding boxes of the human body were manually annotated. The sizes of the cropped RGB images were from 175×350 to 225×450 pixels, and those of the depth images were half of the RGB images. In total 12,800 image pairs including eight orientations of eight pedestrians were prepared. Fig. 6 shows an example of the prepared images. The pedestrians were captured while walking and standing for each of the eight orientations.

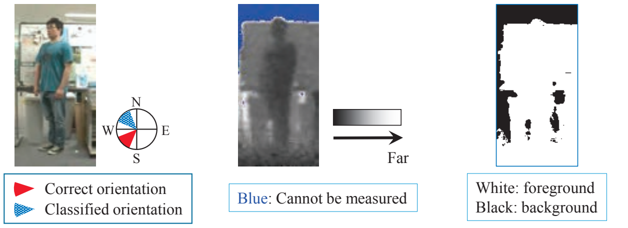
B. Results

The recognition rate of pedestrian orientation was used as an evaluation criteria. Seven pedestrians (11,200 images) were used for training, and another pedestrian (1,600 images) was used for evaluation. The experiment was repeated eight times by changing the pedestrian in a leave-one-out manner, and then the average recognition rate was computed in a

TABLE II

CONFUSION MATRIX OF THE RESULTS OBTAINED BY THE PROPOSED METHOD.

		Classified orientation							
		S	SW	W	NW	N	NE	E	SE
Correct orientation	S	**84.3 %**	5.3 %	0.0 %	0.8 %	0.9 %	8.3 %	0.1 %	0.4 %
	SW	0.1 %	**89.6 %**	7.2 %	3.1 %	0.0 %	0.0 %	0.0 %	0.1 %
	W	0.0 %	10.8 %	**88.1 %**	1.2 %	0.0 %	0.0 %	0.0 %	0.0 %
	NW	1.5 %	0.2 %	0.6 %	**94.6 %**	2.9 %	0.0 %	0.2 %	0.0 %
	N	2.2 %	0.0 %	0.0 %	5.6 %	**84.2 %**	7.9 %	0.0 %	0.1 %
	NE	3.8 %	0.3 %	0.1 %	0.1 %	1.8 %	**85.4 %**	5.8 %	3.1 %
	E	1.4 %	0.0 %	2.3 %	0.1 %	0.0 %	12.4 %	**82.8 %**	1.0 %
	SE	11.9 %	0.0 %	0.1 %	1.3 %	0.1 %	13.8 %	3.9 %	**69.1 %**



(a) Classification result. (b) Input depth image. (c) Segmentation result.

Fig. 7. Example of images where pedestrian orientation was incorrectly classified.

cross validation manner.

In order to confirm the effectiveness of the proposed background removal approach for pedestrian orientation classification, the proposed method was compared with a method without background removal. The result of this experiment is summarized in Table I. In addition, the confusion matrix of the results is shown in Table II.

C. Discussion

As shown in Table I, the proposed method achieved higher recognition rate than the comparative method. This indicates the effectiveness of the background removal for improving the classification accuracy. As shown in Table II, the classification result of SE was worse than that of the other directions. This was caused by the personal differences in the used dataset. The dataset was constructed with eight persons, however, it was small for training. The construction of a larger dataset is part of our future work.

Fig. 7 shows an example of classification results when the proposed method failed. Fig. 7(a) is the input RGB image and the classification result, Fig. 7(b) is the input depth image, and Fig. 7(c) is the segmentation result from the input depth image. White regions in Fig. 7(b) represent the human body, and black regions represent the background. From this, we can see that the cropping of human body was inaccurate. Hence features of background textures were extracted and affected the classification. This occurred when the distance between the pedestrian and other objects were close. Knowledge about the rough body shape can be useful to prevent such a problem.

VII. CONCLUSIONS

This paper proposed a method for classifying a pedestrian's orientation using a single-chip RGB-ToF camera. The human body was separated from the background in input images using coaxial RGB and depth images. We confirmed the improvement of the accuracy on pedestrian orientation classification by the proposed background removal approach through an experiment. Our future work will include experiments in outdoor environments.

ACKNOWLEDGMENT

Parts of this research were supported by JST, Nagoya University COI and MEXT, Grant-in-Aid for Scientific Research.

REFERENCES

[1] G.P. Stein, Y. Gdalyahu, and A. Shashua, Stereo-Assist: Top-down Stereo for Driver Assistance Systems, In Proceedings of 2010 IEEE Intelligent Vehicles Symposium, June 2010, pp. 723–730.

[2] F. Flohr, M. Dumitru-Guzu, J.F.P. Kooij, and D.M. Gavrila, Joint Probabilistic Pedestrian Head and Body Orientation Estimation, In Proceedings of 2014 IEEE Intelligent Vehicles Symposium, June 2014, pp. 617–622.

[3] J. Tao and R. Klette, Part-based RDF for Direction Classification of Pedestrians, and a Benchmark, In Proceedings of Workshop on Intelligent Vehicles with Vision Technology in the 12th Asian Conference on Computer Vision, Nov. 2014, w11-p2.

[4] N. Dalal and B. Triggs, Histograms of Oriented Gradients for Human Detection, In Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2005, pp. 886–893.

[5] C. Weinrich, C. Vollmer, and H.-M. Gross, Estimation of Human Upper Body Orientation for Mobile Robotics Using an SVM Decision Tree on Monocular Images, In Proceedings of 2012 IEEE/RS International Conference on Intelligent Robots and Systems, Oct. 2012, pp. 2147–2152.

[6] K. Goto, K. Kidono, Y. Kimura, and T. Naito, Pedestrian Detection and Direction Estimation by Cascade Detector with Multi-classifiers Utilizing Feature Interaction Descriptor, In Proceedings of 2011 IEEE Intelligent Vehicles Symposium, June 2011, pp. 224–229.

[7] H. Hattori, A. Seki, M. Nishiyama, and T. Watanabe, Stereo-based Pedestrian Detection using Multiple Patterns, In Proceedings of British Machine Vision Conference 2009, Sep. 2009.

[8] G. Pestschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, Digital Photography with Flash and No-Flash Image Pairs, ACM Transaction on Graphics, 23(3), 2004, pp. 664–672.

[9] C.-C. Chang and C.-J. Lin, LIBSVM: A library for Support Vector Machines, ACM Transactions on Intelligent Systems and Technology, 2(3), 2011, pp. 2:27:1–27:27, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.