PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Finding a rush-out human employing a human body direction detector

Yuta Ono, Joo Kooi Tan, Akitoshi Sato

Yuta Ono, Joo Kooi Tan, Akitoshi Sato, "Finding a rush-out human employing a human body direction detector," Proc. SPIE 11049, International Workshop on Advanced Image Technology (IWAIT) 2019, 110491C (22 March 2019); doi: 10.1117/12.2521597

Event: 2019 Joint International Workshop on Advanced Image Technology (IWAIT) and International Forum on Medical Imaging in Asia (IFMIA), 2019, Singapore, Singapore

Finding a Rush-out Human Employing a Human Body Direction Detector

Yuta Ono*[a], Joo Kooi Tan**[b], Akitoshi Sato[a]
[a]Graduate School of Engineering, Kyushu Institute of Technology, Fukuoka, Japan;
[b]Faculty of Engineering, Kyushu Institute of Technology, Fukuoka, Japan

ABSTRACT

Recently, along with rapid development of the image processing technology, image processing has been adopted in various fields for various purposes. Development of an intelligent machine that mounts a camera as an eye is a thriving technology, and it is employed not only in industrial fields but also in the fields involving ordinary citizens. Especially, development of Intelligent Transportation Systems is very active and many methods of detecting human and automobiles have been proposed using laser radars, LIDARs or in-vehicle cameras. However, they remain only on the detection of the presence of such objects and the methods to detect rush-out objects into a road have not been developed yet.

In this paper, a method is proposed which detects a human from an image with his/her body direction information. This intends to detect a human who might rush out into a road in front of an ego-car. In order that the human model used for extracting the feature may capture the appearance of human rush-out properly, directional human models and classifiers are introduced. The proposed method was examined its performance experimentally and the effectiveness of the method was shown/ satisfactory results were obtained.

Keywords: MSC HOG, human detection, body direction, rush-out

1. INTRODUCTION

In recent years, along with the development of image processing technology and the evolution of imaging equipment, image processing has been adopted in various fields. In the in-vehicle field, it is used for understanding the surroundings of an ego-car for automatic driving, and in the industrial field, it contributes to automating picking objects and visual inspection in production lines. In addition, the role played by a camera as an eye of a robot is also large and human recognition and action recognition to realize coexistence between people and robots, especially by service robots, is an important issue. As the methods for human recognition, training using features [1,2,3,4,5], foreground extraction using sequential background updating [6], deep learning [7,8] have been proposed. However, these proposals remain only on the detection of the presence of human or automobiles and the methods to detect rush-out of a human into a road have not been developed yet.

In this paper, a method is proposed which detects a human with his/her body direction information from in-vehicle images employing MSC-HOG, an improved version of the HOG, and Real AdaBoost. Moreover, as 'rush-out' affects selecting the cells used in MSC-HOG, we introduce a directional human model. The MSC-HOG feature is an extension of the HOG feature. It provides a better feature than the HOG feature, since the cell size and the distribution of the cells in a window are much more flexible compared to the HOG. The MSC-HOG feature is computed at all the cells in a window (which scans a given image) and collected into a single feature vector representing the texture feature of the window. The feature vector is fed into a human discriminator obtained from the Real AdaBoost algorithm to judge if the window contains a human. At this time, feature extraction and construction of a classifier are performed in consideration of co-occurrence of extracted features and variation of learning data. Moreover, it is necessary for the human model used for feature extraction to focus on the shape of a human and the sign of a human's rush-out. Therefore, we create a human model providing an average shape of a human posture with various body directions, and based on the directional human model, the MSC-HOG feature is computed. To confirm the effectiveness of the proposed method based on the MSC-HOG feature and the directional human model, we conduct an experiment of detecting a human crossing the road using in-vehicle camera videos.

*ono.yuta434@mail.kyutech.jp, **etheltan@cntl.kyutech.ac.jp

International Workshop on Advanced Image Technology (IWAIT) 2019,
edited by Qian Kemao, et al., Proc. of SPIE Vol. 11049, 110491C
 · doi: 10.1117/12.2521597

2. DETECTION METHODOLOGY

2.1 Outline of the proposed method

In the proposed method, feature extraction is performed on an input image, and the extracted feature is applied to a classifier, thereby detecting whether or not a human exists in the image. MSC-HOG feature [9] is used to extract features and Real AdaBoost is employed for getting classifiers. Next, in order to perform rush-out detection, we create a human model based on a human's feature on body directions. We extract the features of a human who is facing a road from the right-hand side or from the left-hand side and make a human model, considering body direction. We then use the directional human model to detect the rush-out action. The flowchart of the proposed detection method is shown in Fig. 1.

2.2 MSC-HOG feature

The MSC-HOG is an extended technique of the original HOG. In the original HOG, the cells defined in a detection window have an identical size and arranged in a constant interval. On the other hand, in the proposed MSC-HOG feature, a number of cells with different widths and heights are placed randomly along the outline of a human model, allowing overlap. Hence, the distribution of the cells in a window are much more flexible and effective compared to the original HOG feature. This may be obvious because the MSC-HOG can define a rectangle cell that fits a human leg. An example of a cell placement in a detection window is given in Fig. 2, where (a) is the case of HOG, whereas (b) the MSC-HOG case.

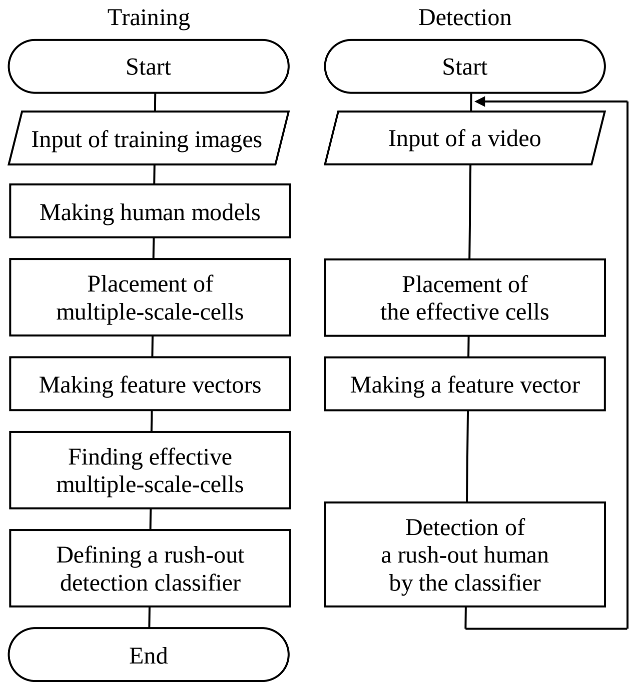
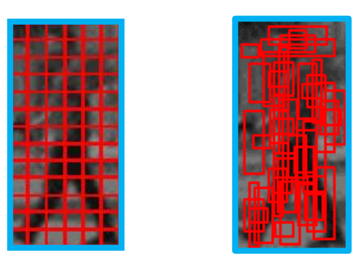


Figure 1. Flowchart of the proposed method.



(a) (b)

Figure 2. Examples of cells placed in the detection window (blue rectangle): (a) HOG feature, (b) MSC-HOG feature.

2.3 Directional human model

Before calculating MSC-HOG feature, human models are made for cell placement. Cells having different sizes are placed randomly in a detection window, as shown in Fig. 2(b), referring to the human model.

A human model is made in the following way. A number of human images are collected. A luminance gradient intensity image of every human image is calculated from Eqs.(2.1), (2.2), and the average image of all the gradient intensity images is computed as a human model. Cells are densely arranged at or around the pixels with high luminance values on the model as illustrated in Fig. 2(b).

2.4 Calculation of a feature vector

The MSC-HOG feature is calculated in the following steps.

(i) Calculating the gradient vectors in a cell

With each cell, the magnitude m and direction θ of a gradient vector of a pixel at (x,y) is calculated by the following equation;

$$m(x,y)=\sqrt{\left\{f_x(x,y)\right\}^2+\left\{f_y(x,y)\right\}^2} \tag{2.1}$$

$$\theta(x,y)=\tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2.2}$$

$$f_x(x,y)=f(x+1,y)-f(x-1,y) \tag{2.3}$$

$$f_y(x,y)=f(x,y+1)-f(x,y-1) \tag{2.4}$$

Here (x,y) is the coordinates of a pixel in an image, $f(x,y)$ is an image intensity at (x,y), $f_x(x,y)$ and $f_y(x,y)$ are the gradients in the horizontal and the vertical direction, respectively.

(ii) Creating a gradient histogram

The direction (0 ° to 180 °) of the gradient vector given by Eq.(2.2) is separated into 9 directions with every 20 degrees. The magnitude of a gradient vector given by Eq.(2.1) is summed up with every direction within a cell to make a histogram $\boldsymbol{q}_i'$ of oriented gradients of the cell c_i. Histogram $\boldsymbol{q}_i'$ is the feature vector of the *i*-th cell containing nine components and is normalized as shown in Eqs.(2.5), (2.6) to yield vector $\boldsymbol{q}_i$ ($\|\boldsymbol{q}_i\|=1$). Its component q_{ij} is the normalized *j*-th histogram value of the *i*-th cell.

$$\boldsymbol{q}_i=\left\{q_{ij}\right\}=\frac{\boldsymbol{q}_i'}{\|\boldsymbol{q}_i'\|+\varepsilon} \tag{2.5}$$

$$q_{ij}'=\sum_{\substack{\theta(x,y)\in bin_j\\(x,y)\in c_i}} m(x,y) \tag{2.6}$$

(iii) Making a MSC-HOG feature vector

The histogram obtained in (ii) defines a 9-dimensional feature vector. It is derived from each cell and normalized so that its norm is 1. An overall feature vector is defined by collecting all the normalized 9-dimensional vectors, representing a MSC-HOG feature vector. Given *m* cells in a detection window, the MSC-HOG feature vector $\boldsymbol{x}$ is defined as follows;

$$\boldsymbol{x}=\left(\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots,\boldsymbol{q}_i,\ldots,\boldsymbol{q}_m\right),\ \boldsymbol{q}_i=\left(q_{i1},q_{i2},\ldots,q_{i9}\right) \tag{2.7}$$

2.5 Finding the cells effective for human detection

Each dimension of the feature vector created in the previous section represents the gradient magnitude of a certain direction in a certain cell. Therefore, a dimension effective for discrimination obtained by learning a feature vector calculated from learning data means a cell effective for human detection. Real AdaBoost is used for the learning, and strong classifier $H_d(\boldsymbol{x})$ of each direction is defined as follows;

$$H_d(\boldsymbol{x}) = \sum_{t=1}^{T} h_{d,t}(\boldsymbol{x}) \tag{2.8}$$

Here T is the number of training round, $h_{d,t}(\boldsymbol{x})$ is a weak classifier of each direction at round t, $d \in \{F, L, R\}$ is the direction, i.e., front, left or right.

2.6 Introduction of co-occurrence feature

For the cells and gradient directions judged valid for identification, their co-occurrence relation is considered for the identification. Feature vector $\boldsymbol{co}$ considering co-occurrence are given as follows;

$$\boldsymbol{co} = (co_k) \quad (k = 1, 2, \cdots, T^2) \tag{2.9}$$

$$\boldsymbol{v} \equiv (q^{(1)}, q^{(2)}, ..., q^{(T)}) \tag{2.10}$$

$$co_{i(T-1)+j} = q^{(i)} + q^{(j)} \quad (i, j = 1, 2, \cdots, T) \tag{2.11}$$

Where $\boldsymbol{v}$ is a feature vector dimensions corresponding to weak classifiers selected by learning. In the case of constructing classifiers considering co-occurrence, learning is performed again using the created co-occurrence feature vector.

2.7 Body direction detector

Using the constructed classifier in each direction, we finally construct a classifier $H_{whole}(\boldsymbol{x})$ that contains body direction elements.

$$H_{whole}(\boldsymbol{x}) = \max\left(\frac{H_F(\boldsymbol{x}) - th_F}{\sigma_F}, \frac{H_L(\boldsymbol{x}) - th_L}{\sigma_L}, \frac{H_R(\boldsymbol{x}) - th_R}{\sigma_R}\right) \tag{2.12}$$

$$\sigma_d^2 = \frac{1}{N}\sum_{i=1}^{N}\left(H_d(\boldsymbol{x}_i) - m_d\right)^2 \tag{2.13}$$

Here σ_d^2 is the factor for normalization: If $m_d = \sum_{i=1}^{N} H_d(\boldsymbol{x}_i)/N$, σ_d^2 means the variance of $H_d(\boldsymbol{x})$: If $m_d = th_d$, σ_d^2 means the second moment of $H_d(\boldsymbol{x})$ around th_d. If $H_{whole}(\boldsymbol{x}) \geq th_{whole}$, the detector recognizes the detected window as a human and judges the body direction d^* by the following equation.

$$d^* = \arg\max_d \left(\frac{H(\boldsymbol{x}) - th}{\sigma}\right)_d \tag{2.14}$$

3. EXPERIMENTAL RESULTS

In order to confirm the effectiveness of the proposed method based on the MSC-HOG feature and directional human models, we conducted an experiment on detecting a human/pedestrian crossing the road using in-vehicle camera videos. There are two kinds of experiments: First, pedestrian classification experiment using still images is carried out performed: Next, pedestrian detection experiment based on a video is conducted performed.

3.1 Experimental environment and the measure of evaluation

As the datasets for training, we choose Daimler Pedestrian Path Prediction Benchmark Dataset [10] and INRIA Person Dataset [1]. With respect to Daimler Pedestrian Path Prediction Benchmark Dataset, pedestrian regions are cut out from each scene in the video using annotations, and used as learning images. As experimental videos, we choose Daimler Pedestrian Path Prediction Benchmark Dataset. The number of the employed videos is 32. The detection accuracy is evaluated using *recall*, *precision*, and *F-measure* defined by

$$recall = \frac{TP}{TP + FN} \times 100[\%] \tag{3.1}$$

$$precision = \frac{TP}{TP + FP} \times 100[\%] \tag{3.2}$$

$$F\text{-}measure = \frac{2 \cdot recall \cdot precision}{recall + precision} \tag{3.3}$$

Here *TP* is the number of the frames in which pedestrians were detected as pedestrians correctly, *FN* is the number of the frames where pedestrians were not detected as pedestrians incorrectly, and *FP* is the number of the frames where non-pedestrians were detected as pedestrians incorrectly.

3.2 Result on still images

Using the constructed classifier, a classification experiment was performed on the still images cut out from the dataset. The images used for identification are pedestrian images facing front, left and right, and the background images. Examples of the classification result are shown in Fig. 3, and a confusion matrix on the experimental result is given in Table 1.

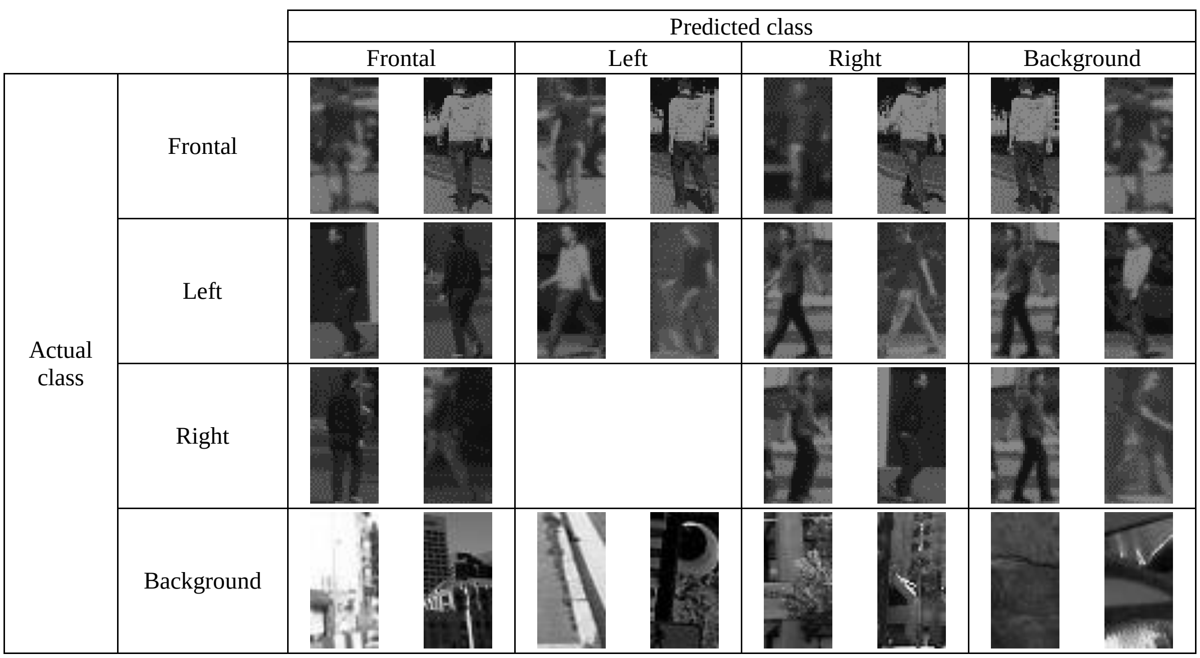


Figure 3. Classification results on dataset still images.

Table 1. Confusion matrix on body direction detection.

		Predicted class			
		Frontal	Left	Right	Background
Actual class	Frontal	1,727	81	14	46
	Left	28	2,154	3	43
	Right	11	0	2,189	28
	Background	16	20	31	2,933

(a)

(b)

Figure 4. Detection results on videos: (a) Crossing from right to left, (b) crossing from left to right.

3.3 Result on videos

Detection experiment was carried out by applying the constructed classifier to the videos. In-vehicle images used for the experiment include those in which the ego-car is stationary or moving, and those in which pedestrians cross the road or stop before crossing. Some of the experimental results are shown in Fig. 4. The average values of *recall*, *precision* and *F-measure* in all 32 scenes were 55.3, 86.9 and 67.6, respectively.

4. CONCLUSION

In this paper, we proposed a method of detecting a rush-out human and classifying body direction from videos. In order to construct a rush-out human detector, we employed the MSC-HOG feature based on a human model which provides an average human shape of a facing posture and a sideway posture. The Real AdaBoost algorithm was applied to the feature vectors to define a classifier obtain the detector. For body direction classification, co-occurrence MSC-HOG feature and directional normalized classifier were introduced. We conducted experiments using INRIA Person dataset and Daimler Pedestrian Path Prediction Benchmark Dataset, and obtained the results showing effectiveness of the presented method.

ACKNOWLEDGMENT

This research was supported by JSPS Kakenhi, Grant Number 16K01554.

REFERENCES

[1] Dalal, N. and Triggs, B.: “Histograms of oriented gradients for human detection”, IEEE Computer Vision and Pattern Recognition, 886-893 (2005).

[2] Viola, P. and Jones, M.: “Rapid object detection using a boosted cascade of simple features”, Computer Vision and Pattern Recognition, 511-518 (2001).

[3] Schapire, R. E. and Singer, Y.: “Improved boosting algorithms using confidence-rated predictions”, Machine Learning, 297-336 (1999).

[4] Vapnik, V. and Lerner, A.: “Pattern recognition using generalized portrait method”, Automation and Remote Control, 774-780 (1963).

[5] Sakai, R., Tan, J. K., Kim, H. and Ishikawa, S.: “Detecting pedestrian and extracting their attributes from self-mounted camera views”, ICIC Express Letters, Part B: Applications, Vol.7, No.2, 279-286 (2016).

[6] Tan, J. K., Ishikawa, S., Sonoda, S., Miyoshi, M. and Morie, T.: “Moving objects segmentation at a traffic junction from vehicular vision”, ECTI Trans. on Computer and Information Technology, Vol.5, No.2, 73-88 (2011).

[7] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: “Rich feature hierarchies for accurate object detection and semantic segmentation”, Proceedings of the IEEE conference on computer vision and pattern recognition, 580-587 (2014).

[8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: “You only look once: Unified, real-time object detection”, Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788 (2016).

[9] Ehara, Y., Tan, J. K., Kim, H. and Ishikawa, S., “Human detection from an image employing reduced number of weak-classifiers”, Proc. of the IIEEJ Image Electronics and Visual Computing Workshop, USB_5 pages (2012).

[10] Kooij, J. F. P., Schneider, N., Flohr, F. and Gavrila, D. M., “Context-based pedestrian path prediction”, European Conference on Computer Vision, 618-633 (2014).