Detecting pedestrian and extracting their attributes from self-mounted camera views

著者	Sakai Ryuichi, Tan Joo Kooi, Kim Hyoungseop, Ishikawa Seiji
journal or publication title	ICIC Express Letters, Part B: Applications
volume	7
number	2
page range	279-286
year	2016-02
その他のタイトル	DETECTING PEDESTRIAN AND EXTRACTING THEIR ATTRIBUTES FROM SELF-MOUNTED CAMERA VIEWS
URL	http://hdl.handle.net/10228/00007856

doi: http://dx.doi.org/10.24507/icicelb.07.02.279

ICIC Express Letters
Part B: Applications
Volume **7**, Number **2**, February **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 279–286

DETECTING PEDESTRIAN AND EXTRACTING THEIR ATTRIBUTES FROM SELF-MOUNTED CAMERA VIEWS

Ryuichi Sakai[1], Joo Kooi Tan[2], Hyoungseop Kim[2] and Seiji Ishikawa[2]

[1]Graduate School of Engineering
[2]Faculty of Engineering
Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
{ sakai; etheltan }@ss10.cntl.kyutech.ac.jp

Received July 2015; accepted September 2015

Abstract. *There are not a few visually impaired people in our society. When they want to go for a walk, it is difficult for them to perceive moving objects around them. When they walk on the sidewalk, they must pay strong attention to those coming closer to them or to moving objects. Therefore, it is indispensable for them to obtain the information on moving objects around them. This paper proposes a method of extracting moving objects, in particular, pedestrians, from self-wearable camera images. In the proposed system, a user is supposed to wear a camera and a PC and, by performing image analysis of the video taken from the camera, it acquires surrounding pedestrians' information including their peculiar characteristics. The information is fed back to the user to realize his/her safe walk. Experimental results are shown and the performance of the proposed method is evaluated. The proposed system is primarily intended to be used by visually impaired people, but it may also be used by any pedestrian who is not paying much attention to his/her surroundings.*

Keywords: Pedestrians detection, Self-mounted camera, Peculiar information, HOG, Real AdaBoost, Computer vision

1. **Introduction.** Recently, the technology on pedestrian detection has become one of the researches which are most actively done. This technology is effective as a means of walking support of visually impaired people. The reason is that it is necessary for them to avoid clashes on walking. In addition, when the problem occurs in the pedestrian side, it is necessary for the visually impaired to rush into action. To solve this problem, first of all, the recognition of his/her environment is necessary. In this paper, we propose a method of detecting pedestrians and extracting their attributes, or peculiar information, by a self-wearable camera. If a single camera is used, the cost is suppressed low compared to a system that uses two or more cameras. Moreover, if pedestrians can be detected from a self-viewpoint image where various vibrations are included, the same function will be possible with the case of a fixed viewpoint image in easier way.

Many studies on person detection have been performed including [1,2]. Studies on the tracking of a walker have also been performed including [3,4]. This is because reduction is necessary for the processing time when we apply the person detection to a video. However, most of the studies on person detection are carried out for vehicle drivers' safe driving using an in-vehicle camera, and the study on pedestrian detection using a self-mounted camera for the visually impaired people has not yet been performed. Obviously it is important for those visually impaired people to know the existence of pedestrians or moving objects around them in order to escape from crash. There is a study [5] which detects the walking direction of a pedestrian, but the practical performance is not satisfactory.

In this paper, aiming at realizing a safety tool for a visually impaired person or even a person inattentive to his/her surroundings, we propose a method of detecting pedestrians from a video captured by a camera mounted on a user and extracting some attributes (or peculiar information) of the detected pedestrians including their moving directions relative to the user. This tool will actually be a virtual eye for the safety of a user.

In the proposed method, HOG feature [6] and RealAdaBoost [7] are used for pedestrians detection at each frame. To acquire some attributes from a pedestrian, a detected pedestrian is tracked in successive frames using Euclidean distance and color information. The attributes considered in the present study are the number of the pedestrians and their walking directions. The present method concentrates on detecting walking pedestrians. Running pedestrians' detection depends on the shape of the running form. If the shape does not change much from a walking form, they may be detected by the proposed method. Detection of a person riding a bicycle is out of the scope of the present study. However, [11] may give a cue for solving the problem.

2. **Detection of Pedestrians.** HOG is the characteristic quantity that pays its attention to a gradient magnitude and direction of the brightness distribution in a local region on an image. Pedestrians are detected using the HOG and RealAdaBoost. RealAdaBoost has a characteristic that the highly precise detection is possible using a set of weak classifiers. In calculating the HOG feature, Integral Histogram [3] is used to reduce computation time. Moreover, when RealAdaBoost is used, the co-occurrence probability feature [4] is introduced, which results in high accuracy detection of a pedestrian. A pedestrian is detected based on raster scan of a window on the image concerned. This normally provides several candidate windows containing an identical person. MeanShift [5] is used for the integration of the windows.

3. **Tracking a Detected Pedestrian.** After a pedestrian is detected, the person is tracked in an image sequence. For this purpose, it is judged if the person in the window is the same person by comparing the window including the person detected in the previous kth frame with the window including the person detected in the current frame using the procedure employing Euclidean distance and the ratio of hue. It is explained in the following subsections.

3.1. **Use of Euclidean distance.** The Euclidean distance between the center coordinates of the window detected in the previous kth frame and the center coordinates of the window detected in the current frame is denoted by d. It is judged as true when d is below a certain threshold. Let us denote the center coordinates of the present frame by (x, y), and the center coordinates of the previous kth frame by (x_k, y_k). Then the judgment is done as follows,

$$d = \sqrt{(x - x_k)^2 + (y - y_k)^2} \tag{1a}$$

$$cond\_1 = \begin{cases} \text{true} & \text{if } d < Th_1 \\ \text{false} & \text{else} \end{cases} \tag{1b}$$

3.2. **Use of the ratio of hue.** Color aspect H of the HSV color specification system within the range of specification in the window is computed using the detected windows in the previous kth frame and in the present frame.

As shown in Figure 1, a window containing a pedestrian is separated into 3 by 6 squares of the same size. The appearance rate of H, denoted by $P(H)$, is computed in the center square and the sum of the appearance rate S_H of hue H within the specified range is computed by the following equation,

$$S_H = \sum_{i=H-h}^{H+h} P(H_i) \tag{2}$$

FIGURE 1. The specified color range

Let us denote the maximum value of S_H by $S_{\max}$, and H realizing $S_{\max}$ by $H_{\max}$. Then H_{abs} is defined by Equation (3a). If H_{abs} is smaller than a certain threshold Th_2, it is judged true, as given in Equation (3b), and proceed to the next step. S_{abs} is defined by Equation (4a) and, if S_{abs} is smaller than a threshold Th_3, it is judged true, as given in Equation (4b).

$$H_{\text{abs}} = \left\| H_{\max_f} - H_{\max_{f-k}} \right\| \tag{3a}$$

$$cond\_2 = \begin{cases} \text{true} & \text{if } H_{\text{abs}} < Th_2 \\ \text{false} & \text{else} \end{cases} \tag{3b}$$

$$S_{\text{abs}} = \left\| S_{\max_f} - S_{\max_{f-k}} \right\| \tag{4a}$$

$$cond\_3 = \begin{cases} \text{true} & \text{if } S_{\text{abs}} < Th_3 \\ \text{false} & \text{else} \end{cases} \tag{4b}$$

Here f is the frame number of the current frame. If all of $cond\_1$, $cond\_2$ and $cond\_3$ are true, it is judged that the person in the window of frame $f - k$ and the person in the window of frame f are the same person.

4. **Extracting Attributes.** After the same person is detected, peculiar information is extracted from the detected person. The peculiar information in the present study is the information on the detected pedestrian(s), i.e., the number of pedestrians and their traveling directions.

4.1. **Detection of the number of pedestrians.** The mean value of the number of windows detected in the past several frames is defined as the number of pedestrians.

4.2. **Detection of traveling directions.** A pedestrian's direction means the traveling direction of a pedestrian observed from a person who installs a camera and takes video of a frontal scene. Successive center coordinates of the windows containing the same person in the past several frames are paid attention to. The motion vectors connecting two successive center coordinates are averaged to give the pedestrian's direction. In the experiment, eight directions with every 45 degrees are employed as shown in Figure 2. In the figure, upper three directions indicate that a pedestrian goes away from the user (a camera holder), whereas the lower three directions show that a pedestrian approaches to

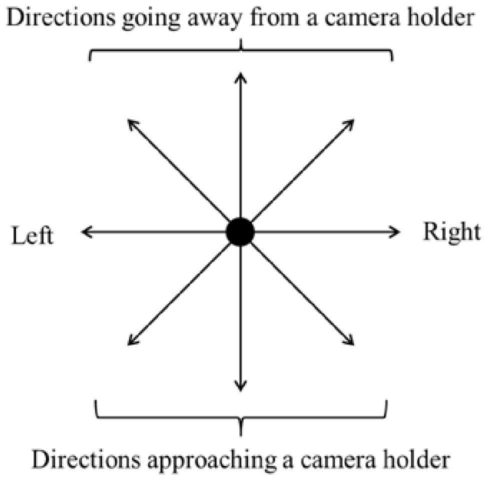

FIGURE 2. Traveling directions

the user. In particular, the center direction among the lower three directions tells that a pedestrian proceeds directly toward the user, which might cause collision.

5. **Experimental Results and Evaluation.** In this section, we describe the experimental setup, results and evaluation of the proposed method with respect to pedestrians detection, tracking and extracting attributes.

5.1. **Experimental setup.** In this study, in order to obtain self-view images, we mount a camera tight by a belt on the head of a user as shown in Figure 3. In this experiment, we use 1070 positive image samples and 6,000 negative image samples, and normalize all these images into the 48 by 96 pixels size. We perform the learning by RealAdaBoost. The BIN number is set to 64 and the number of learning times is 60. We detect pedestrians' candidate windows by using the classifier defined by the RealAdaBoost. We then integrate the detected windows, and detect pedestrians. Finally, we track the same pedestrian and extract attributes.

In order to verify the effectiveness of the proposed method, we carry out the experiment using four kinds of videos under different circumstances: (1) the case where occlusion does not occur, (2) the case where occlusion occurs, (3) the case where the number of pedestrians increases, and (4) the case where the number of pedestrians decreases. The windows show the result of the pedestrian detection. The same pedestrian is indicated by the ID number given to a window. The figure in the upper-right corner of the screen shows the number of people. The traveling direction of a pedestrian is given by an arrow

FIGURE 3. A self-mounted camera

attached to the window. Note that the traveling direction of a pedestrian is the direction relative to a camera holder.

5.2. **Experimental results.** Figure 4 shows the example of the outcome of the performed experiment. The windows show the result of the pedestrian detection.

5.3. **Evaluation.** In this experiment, we calculate two evaluation indexes, *cover* and *overlap*, using Equation (5) and Equation (6), respectively, to judge the performance of the proposed method.

$$cover = \frac{GT \cap OA}{GT} \tag{5}$$

$$overlap = \frac{GT \cap OA}{OA} \tag{6}$$

Here GT shows the region of Ground Truth Data, and the OA shows the region of a window detected in this experiment.

The case that both *cover* and *overlap* exceed threshold 0.5 is judged as true positive, TP, otherwise false positive, FP. An example of GT and OA is shown in Figure 5. The detection rate ($Precision$) of pedestrians and the false detection rate (FPR) are calculated using Equation (7) and Equation (8), respectively.

$$Precision = \frac{N_{TP}}{N_H} \times 100 \tag{7}$$

$$FPR = \frac{N_{FP}}{N_W} \times 100 \tag{8}$$

Here, N_{TP} is the total number of TP of the pedestrian; N_H is the total number of pedestrians in an evaluated frame; N_{FP} is the total number of FP; N_W is the total number of the detection window.

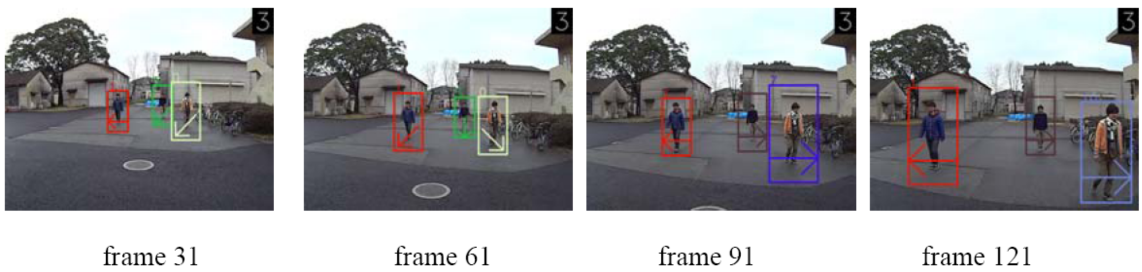

FIGURE 4. Experimental results

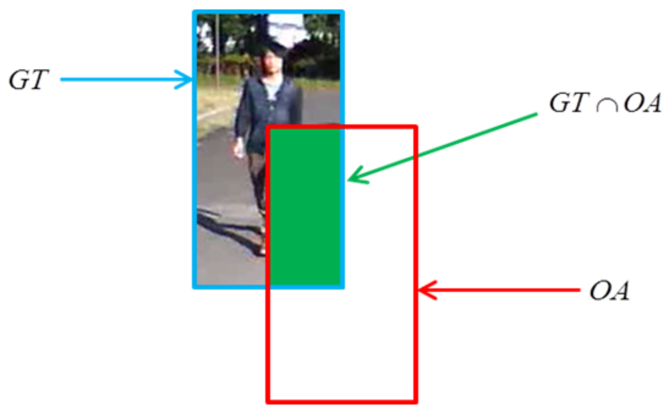

FIGURE 5. Evaluation indexes

The detection rate of the pedestrians and false detection rate in each experiment are shown in Table 1. It is noted that the evaluated frame is the frame in which an entire body of a pedestrian is captured.

The evaluation on if the tracking was carried out correctly is done in the following way. The frame is the present concern whose closest preceding frame contains a person. The number of the windows with which the tracking of a person was done correctly is denoted by N_{ID}. The accuracy of the tracking, P_{ID}, is then defined by

$$P_{ID} = \frac{N_{ID}}{{N_{ID}}^W} \times 100 \tag{9}$$

Here ${N_{ID}}^W$ is the total number of the evaluated windows. The accuracy of the tracking in each experiment is given in Table 2.

The exactness with respect to the extraction of the specified attributes is also evaluated. Let us denote the number of all the frames by N_{ALL} and the number of the frames whose pedestrians contained within them were counted correctly by N_{NUM}. Then its precision, P_{NUM}, is defined by

$$P_{NUM} = \frac{N_{NUM}}{N_{ALL}} \times 100 \tag{10}$$

Furthermore, the direction of walking of a detected pedestrian is evaluated only when the tracking of the person is correct. Two directions out of eight that are the closest to the direction the Ground Truth Data gives are regarded as correct directions for the evaluation as shown in Figure 6. If we denote the number of the windows having correct directions

Table 1. Evaluation on the number of the detected pedestrian

Experiment	The total number of the pedestrians in evaluated frames	Detection rate [%]	False detection rate [%]
1	375	87.7	8.9
2	352	73.0	9.1
3	258	74.4	28.4
4	428	61.2	39.5

Table 2. Evaluation on the tracking

Experiment	The number of the evaluation windows	Accuracy [%]
1	303	94.1
2	247	93.5
3	187	99.5
4	247	88.7

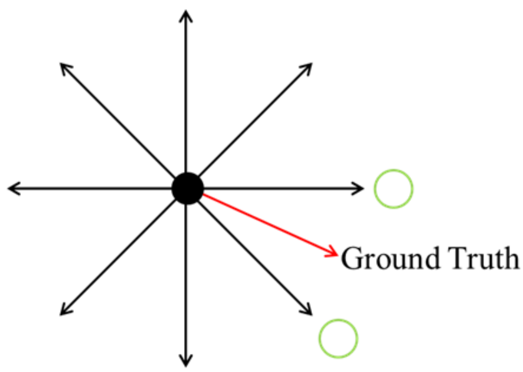

Figure 6. Evaluation of the direction

TABLE 3. Evaluation on the extraction of the peculiar information

Experiment	The number of the evaluation windows	Accuracy on the number of pedestrians [%]	Accuracy on the walk directions [%]
1	285	92.8	47.0
2	231	48.0	64.5
3	186	50.5	46.8
4	219	52.6	55.3

by N_D and the number of the windows in which person tracking is done correctly by N_{ID}, the precision with respect to the direction, P_D, is defined in the following way;

$$P_D = \frac{N_D}{N_{ID}} \times 100 \tag{11}$$

The precision on the detection of the peculiar information is given in Table 3 with respect to the number of pedestrians in front of a user and their walking directions.

6. **Conclusions.** We proposed a pedestrians detection method using a camera mounted on a user himself/herself. In the method, pedestrians in front of the user were detected employing HOG feature and RealAdaBoost. Each pedestrian was tracked using distance and color information in a given image frame sequence, and the number of the pedestrians was counted in each frame along with the detection of his/her walking direction as peculiar information. The proposed method was evaluated employing four scenes and promising results were obtained. In the pedestrian detection, the detection rate was 73.6% in average. In the tracking of a pedestrian, the accuracy was 93.3% in average. With respect to the extraction of the peculiar information, the average accuracy of the number of pedestrians was 51.2%, whereas the average accuracy of the pedestrian's directional detection was 63.6%.

In a person detection from an image, a window is raster-scanned on the image to find a window which contains a person of an appropriate size. Since a person has various sizes in an image according to the distance from the camera, various sizes of windows need to be scanned on the image, which results in a large computation time. In order to reduce the computation time, tuning of some parameters is further necessary such as the number of prepared windows, and the number of pixels with which the window is displaced.

In the proposed method, a camera is assumed to be mounted on a user. In the performed experiment, it was mounted tight by a belt at the side of a user's head. However, the captured video often contains some sway or vibration caused by the movement of the user. This disturbance should be compensated to raise respective accuracies stated above more.

Acknowledgment. This work was supported by JSPS KAKENHI Grant No. 25350477, which is greatly acknowledged.

REFERENCES

[1] P. Dollar, Z. Tu, P. Perona and S. Belongie, Integral channel features, *Proc. of the British Machine Conference*, pp.1-11, 2009.
[2] C. Guang, Y. Ding, J. Xiao and T. Han, Detection evolution with multi-order contextual co-occurrence, *Proc. of the IEEE Conference on CVPR*, pp.1798-1805, 2005.
[3] S. Tang, M. Andrilika and B. Schiele, Detection and tracking of occluded people, *International Journal of Computer Vision*, vol.26, no.9, pp.1-11, 2012.
[4] A. Andriyenko, K. Schindler and S. Roth, Discrete-continuous optimization for multi-target tracking, *Proc. of the IEEE Conference on CVPR*, pp.1926-1933, 2012.
[5] J. Tao and R. Klette, Pedestrian and direction classification using a random decision forest, *Proc. of the IEEE Conference on ICCV*, pp.230-237, 2013.

[6] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of the IEEE Conference on CVPR*, pp.886-893, 2005.

[7] R. Schapirel and Y. Singer, Improved boosting algorithms using confidence-rated predictions, *Journal on Machine Learning*, vol.37, pp.297-336, 1999.

[8] F. Porikli, Integral histogram: A fast way to extract histograms in Cartesian spaces, *IEEE International Conference on Computer Vision and Pattern Recognition*, vol.1, no.5, pp.829-836, 2005.

[9] Y. Yamauchi, T. Yamashita and H. Fujiyoshi, Pedestrian detection by boosting-based feature co-occurrence representation, *IEICE Trans. Information and Systems*, vol.J92-D, no.8, pp.1125-1134, 2009.

[10] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.603-619, 2002.

[11] H. Jung, *A Study on a Bicycle Detection System Employing Car Vision*, Ph.D. Thesis, Kyushu Institute of Technology, 2014.