Fast and Stable Human Detection Using Multiple Classifiers Based on Subtraction Stereo with HOG Features

Makoto Arie, Alessandro Moro, Yuma Hoshikawa, Toru Ubukata, Kenji Terabayashi, Kazunori Umeda

***Abstract*— In this paper, we propose a fast and stable human detection based on "subtraction stereo" which can measure distance information of foreground regions. Scanning an input image by detection windows is controlled in their window sizes and number using the distance information obtained from subtraction stereo. This control can skip a large number of detection windows and leads to reduce the computational time and false detection for fast and stable human detection. Additionally, we propose two-step boosting as a new training way of classifier with whole and upper human body models. Experimental results show that the proposal is faster and less false detection than the method described in the reference [1].**

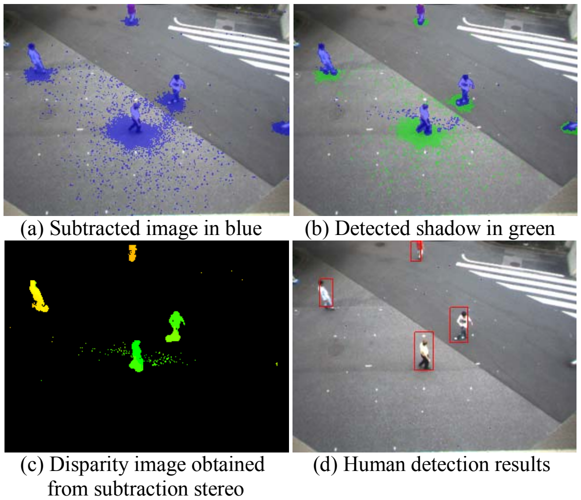
I. Introduction

Human detection can be widely used in many applications, including people counting and security surveillance in public scenes. However, detecting human is still a challenging task because of their various appearances and occlusion problem.

One of the most successful approaches of human detection is based on HOG (Histograms of Oriented Gradients) descriptor proposed by N. Dalal and B. Triggs [1], which is robust to illumination changes due to using edge information. Various features for object detection have been proposed by P. Viola and J. Jones [2], K. Levi and Y. Weiss [3], and B. Wu and R. Nevatia [4].

Human detection based on HOG features [1] requires a lot of computational time for the following three reasons: (i) computation of HOG feature, (ii) detection windows scanning whole area in an input image, (iii) multiple sizes of detection windows. Real-time processing is important for real-time applications at the same time as high performance of human detection.

Additionally, false detection increases with the number of detection windows increasing rapidly because of the above points (ii) and (iii). Therefore, a large number of detection windows are inappropriate in terms of both computational cost and detection accuracy.



(a) Subtracted image in blue (b) Detected shadow in green

(c) Disparity image obtained from subtraction stereo (d) Human detection results

Fig. 1. Overview of the proposed human detection based on subtraction stereo. Distance information of foreground regions are obtained from the subtraction stereo through shadow removal. Colors in image (c) correspond to distance from a camera. Red bounding boxes in image (d) represent the areas of detected humans.

In this paper, we propose a novel human detection which is faster and less false detection than the method described in the reference [1]. The proposed method is based on "subtraction stereo" [5] which calculates distance information focusing on only foreground regions in an input image.

The problem of many detection windows in human detection about computational cost and false detection have been solved by using distance information of foreground regions obtained from the subtraction stereo as described below. The size of detection window is determined appropriately and dynamically based on the distance information and paraperspective projection model, which can avoid multiple tests of window size and leads to reduce the number of detection windows. Calculations of HOG features and recognition processing are performed for only foreground regions extracted by subtraction stereo with shadow removal to decrease computational cost and false detection.

In addition, to improve detection accuracy, we propose two-step boosting for training a classifier with whole and upper body models. The two-step boosting is a process, which selects characteristic features of positive image by comparing features in negative image, after an AdaBoost. The paper is organized as follow. In Section II, related approaches for human detection

M. Arie, Y. Hoshikawa and T. Ubukata are with the Course of Precision Engineering, School of Science and Engineering, Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan (e-mail:{arie, hoshika, ubukata}@sensor.mech.chuo-u.ac.jp)

A. Moro is with the Department of Industrial and Information Engineering, Univ. of Trieste / CREST, JST, P. le Europe 1, 34127 Trieste, Italy (e-mail: alessandro.moro@stud.units.it).

K. Terabayashi and K. Umeda are with the Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan (e-mail:{terabayashi, umeda}@mech.chup-u.ac.jp).

are overviewed. In Section III, distance measurement of foreground regions is described with subtraction stereo and shadow detection. In Section IV, we propose a novel method for controlling the size and number of detection windows based on the distance information of foreground regions. Then, we propose two-step boosting as a new training way of classifier with whole and upper body models in Section V. In Section VI, the proposal is evaluated by comparing with the reference [1] in terms of computational time and detection accuracy. The conclusion is summarized in Section VII.

II. Related Works

Recent research on human detection has used monocular vision [1], [6] - [9] stereo vision [10] - [14], and LIDAR sensing [15]. An overview of several approaches for pedestrian detection can be found in [16].

One of the most popular recent appearance based human detection algorithms is the HOG method proposed by Dalal and Triggs [1]. They characterized human regions in an image using HOG descriptor, which are a variant of the well-known SIFT descriptor [17]. Unlike SIFT, which is sparse, the HOG descriptor offers a denser representation of an image region by tessellating it into cells which are further grouped into overlapping blocks. Zhu et. al [18] extend the HOG descriptor and utilize a cascade classifier structure to increase detection speed.

Another leading real time, monocular vision system for human detection in cars was proposed by Shashua et al [6]. The authors used a focus of attention mechanism to detect window candidates very rapidly. The window candidates (approximately 70 windows per frame on average) are classified into pedestrian or non-pedestrians using a two stage classifier.

Ess et al. [12] - [14] describe a stereo based system for three dimensional dynamic scene analysis from a moving vehicle, which integrates sparse three dimensional structure estimation with multi-cue image based descriptors to detect pedestrians. The authors show that the use of sparse three dimensional structures significantly improves of the performance of the pedestrian detector. Still, the best performance cited is 40% detection at 1.65 false positive per image frame. While the structure estimation is done in real-time, the pedestrian detections is significantly slower.

Bajracharya et al. [11] describe a real-time stereo-based system that can detect human up to 40m in highly cluttered environments. The stereo range maps are projected into a polar-perspective map that is segmented to produce clusters of pixel corresponding to upright objects. Geometric features are computed for the resulted three dimensional point clouds and used to train pedestrian classifiers. Appearance based features are not used for classification.

In [9], a pedestrian detection method based on the covariance matrix descriptor [19] is proposed and shows better performance on the INRIA dataset [1] than the HOG descriptor, but an experimental study conducted by Paisikriangkarai et al. [20] shows that the covariance matrix descriptor is slightly inferior to the HOG descriptor on the DaimlerChrysler pedestrian benchmark dataset created in [21].

III. Extraction of Foreground Regions with Distance Information

We proposed a method of fast and low false alarm scanning detection window using subtraction stereo. When we perform human detection, we scan a detection window only to foreground regions. Therefore, since a background region is not scanned, computation time and false alarm decreased. In addition, the accuracy of human detection is further improvement by performing shadow detection. The construction technique of extraction of foreground regions is as follows.

A. Subtraction Stereo

The edge information includes much information for human detection. Therefore we remove background information by the subtraction stereo. We show the algorithm of the subtraction stereo [5]. The subtraction stereo extracts foreground regions in a scene by background subtraction method. Then a disparity image is obtained by the stereo matching of each foreground regions. Also this system can measure actual heights and widths of the foreground regions. An example of disparity image is shown in Figure 1 (c).

B. Shadow Detection

Shadow detection is used to refine the foreground. The image obtained using subtraction stereo includes noise affected by the shadow. This noise seriously affects the human detection.

We have improved the shadow detection described in [22]. When we define $I(x,y)$ as the intensity of the pixel located in the two-dimensional image position (x,y) and $I'(x,y)$ as the intensity of the background pixel, the equation for the evaluation of shadow is described as

$$\theta_{(t+1,x,y)} = \begin{cases} \alpha\Psi_{(x,y)} + \beta\Lambda_{(x,y)} + (1-\alpha-\beta)\theta_{(t,x,y)}, \\ \qquad\qquad if\ \dfrac{I_{(x,y)}}{\eta} < I'_{(x,y)} \\ \infty, otherwise \end{cases} \quad (1)$$

where θ corresponds to a shadow value. This value will be applied a threshold to determine if a pixel is a shadow. A small shadow value corresponds to a shadow point. The functions Ψ and Λ show color constancy between pixels and within pixel. α, β, and η are constant weights of textures, colors, and intensity, which are determined empirically in our experiments. The details of this method are explained in [22]. The result of the shadow detection is shown in Figure 1 (b).

IV. Human Detection

The human detection procedure based on the HOG feature at subtraction image: (1) extracts foreground and computes distance of its regions using the subtraction stereo described in Section II; (2) dynamically changing detection window size by the distance; (3) compute the HOG feature of foreground regions for each scanning detection window; (4) discriminate whether human or not by an AdaBoost classifier.

A. Dynamically Changing Detection Window Size

In this section, we propose a method of dynamically changing window size using subtraction stereo. This method solves the problems the large computation time of HOG features, and false detection. The method of Dalal [1] is in order to calculate the HOG feature from an image, scanning the detection window to end to end of an image. Therefore, their computation time is high and real-time processing is difficult. Also scanning multiple detection windows in different size has high computation time.

First, the scanning detection window size is computed from the distance of the foreground regions. Second, the height of detection window size is rectified by the position of human in image. Because we assume the paraperspective projection, rectify the height of detection window size. As a reason for select the paraperspective projection while there is various perspective projection, the weak perspective projection rectifies only the height it to the distance, but the paraperspective projection rectifies it in consideration both the different in vision occurs with the position of human in image and elevation angle and height of a camera. From these, the detection window of different size is scanned at once.

For example, when the height of a camera is 1.6m and elevation angle is 0 degree, the size (height and width of pixel) in each distance from a human to a camera is shown in Figure 2. Since the distance and the size from a human to a camera have the relation of an inverse proportion from Figure 2, the constant of proportion of height and width are computed. In addition, how to see a human in distance differs on a scene. It is necessary in much time to scan many detection windows on a scale of being different. Therefore, the height and width of the scanning detection window size are rectified by eq. (2) (3) which assumed paraperspective projection.

$$height = \frac{k_h}{Y_W}(\cos\theta - y\sin\theta) \qquad (2)$$

$$width = \frac{k_w}{Z_C} \qquad (3)$$

Where, k_h, k_w are constant, Y_W is distance of world coordinate system, Z_C is distance of axis direction, θ is elevation angle of a camera, y is image coordinates which normalized the length to 1.

HOG Features

We use HOG feature for human detection. The HOG feature [1] has shown success in object detection. The HOG feature [1] has shown success in object detection. The HOG feature has been accepted as one of the best feature to capture gradient information. However, it cannot compute its information quickly, and it computes a region of background that is not relationship human detection. Therefore we extract its information at foreground of subtraction image. The HOG feature the distribution of image gradients on magnitude and different orientation. We follow the procedure in [1] to extract the HOG feature. The histogram of the gradient orientation is used for analysis of the edge orientation. It is created in constant number, which is called cell. Since the size of the detection window changes dynamically, it also changes the size of cell according to it. We predefine the number of cell 6×12 and in a square region. The number of the bins of the histogram is decided by the number of partitions of the gradient orientation. We predefine the number of orientation bin is 9. Then the histogram is normalization in regions predefine, which is called block. We predefine the block size 3×3 cell and in a square region. An eq. (4) is used for normalization.

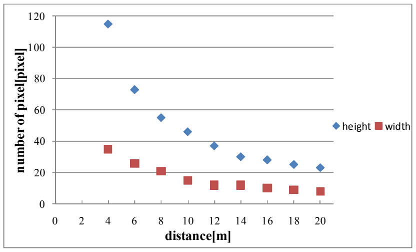


Fig. 2. The distance and the size from a human to a camera have the relation of an inverse proportion.

$$f = \frac{V}{\sqrt{\|V\|^2 + \varepsilon^2}} \qquad (4)$$

Where V is HOG feature vector, ε is a small regularization constant.

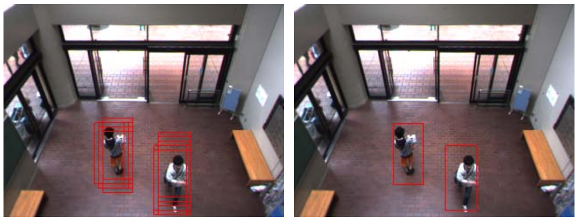
B. Post processing of Detections

For human detection, we scan the trained classifier over all sliding windows in each scale in the subtraction image. As a result, as shown in Figure 3 (a), there will be many detection around each target. To obtain the number of targets and the exact location of each target from these detection windows, postprocessing is necessary. The detection window recognized as a human is unified using mean shift clustering [23].

By postprocessing, as shown in Figure 3 (b), we can get exact number of pedestrians and obtain more accurate target locations than the way the raw detection windows.

V. Structure of Multiple Classifiers

We describe the construction method of Real AdaBoost for investigating whether the inside of a detection window is a human. In addition, we describe the construction method of a full-body and an upper part of body classifier in which two-step boosting was used. Explanation of two-step boosting is shown in the section of each classifier.



(a) Before clustering (b) After clustering

Fig. 3. Detection windows recognized as a human are unified using mean shift clustering.

A. Real Adaboost Training

We use Real Adaboost [2] to learn the classification function. This is because Real Adaboost is an effective and efficient learning algorithm for training on high-dimensional large dataset. Compared with other statistical learning approaches (e.g., SVM), which try to learn a single powerful discriminant function from all the specified features extracted from training samples, the Adaboost algorithm combines a collection of simple weak classifiers on a small set of critical features to form a strong classifier using the weighted majority vote. This means the Adaboost classifier can work very fast in the testing stage. Furthermore, Adaboost is not prone to overfitting and provides strong bounds on generalization which guarantees the comparable performance with SVM.

Gathering a representative set of negative samples is very difficult. To overcome the problem of defining this extremely large negative class, bootstrapping training is adopted. A preliminary classifier is trained on an initial training set, and then used to predict the class categories of a large set of patches randomly sampled from many added to the negative training set for the next iteration of training.

B. Two-step Boosting

We proposed two-step boosting for the method of building a classifier. When building an AdaBoost classifier [2], false detection increases only by comparing the HOG feature in a positive sample (human) and negative sample (car, building, door). We call the classifier built by this method as one-step boosting. However it is difficult to create a classifier with high accuracy of human detection, because one-step boosting does not choose the characteristic features in humans.

In the first step of two-step boosting, features representing human shape appropriately are selected as candidates of finally used features in a classifier by comparing between only positive images. Then, the candidate features are refined in the second step by comparing features in negative images in which characteristic features of negative images are not selected. We call this process as the two-step boosting. The accuracy of human detection is improved using two-step boosting. The next section explains the construction method of the full-body and the upper of body classifier using two-step boosting.

C. Multiple Classifiers

In this section, we propose a method of multiple classifiers effective in human detection using the full-body and the upper part of body samples. The full-body and the upper part of body classifiers are built from each sample, and after human detection scans the full-body classifier, the upper part of body classifier is scanned again. Only by the full-body classifier, false detection increases. Because the door and the building include the HOG feature of the straight line ingredient of an intensity gradient, they are false detected to be a human. In addition, the upper part of body classifier has high human detection accuracy. However, since the full-body classifier differs from different false detection, it cannot be used by itself. The construction technique of the full-body and the upper part of body classifier is as follows.

Full-body classifier. First, the image of a positive sample is divided into the upper part of body and the lower part of body, and the HOG feature is computed by each part. Then, in order to choose the feature having the difference with the direction of an intensity gradient by boosting, the feature in the cell of the upper part of body and the lower part of body is compared as shown in Figure 4. Finally, the full-body classifier effective in human detection is created by boosting again in the feature chosen previously and the feature acquired from a negative sample. A characteristic feature of human can be chosen by performing this two-step boosting. Although Viola [2] performs one-step boosting which used the positive and the negative samples, since our method is performed two-step boosting, it is more accurate than the method of Viola [2].

Upper part of body classifier. The positive sample of the upper part of body uses an image as shown in Figure 4. In order to choose the feature which has symmetry in the direction of intensity gradient, the feature in each cells in a positive sample is compared as shown in Figure 5. Then, the upper part of body classifier is created by boosting again like the full-body classifier in the feature obtained from a negative sample. The upper part of body classifier makes hard to detect the object (the straight line ingredient of an intensity gradient is included) which is false detection by the full-body classifier.

VI. Experiments

In this section, two groups of experiments are carried to validate our proposal method. First, we compare the human detection results between normal image and subtraction image. Second, we compare the performance of our multiple classifiers using two-step boosting with that of the HOG-AdaBoost detector [1], one state-of-the-art human pedestrian detector by the data set built uniquely. The results have been obtained using an Intel Core2 Duo CPU, 3.00 GHz with 4GB ram.

A. Data set

We verify the validity of our method by the dataset built uniquely. The positive sample of the data set is constituted in consideration of elevation angle. The number of the images of

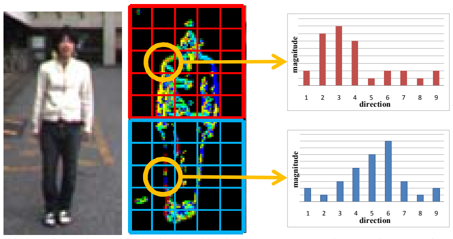


Fig. 4. The full-body classifier is built by two-step boosting. The difference with the direction of an intensity gradient feature is chosen by comparing the feature of the upper part of body and the lower part of body.

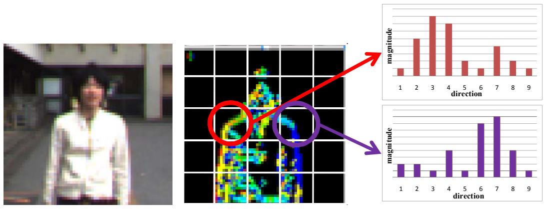


Fig. 5. The upper part of body classifier is built by two-step boosting. The symmetry with the direction of an intensity gradient feature is chosen by comparing the feature of the upper part of body.

positive sample is each 3500 images per 10 degrees in total 17,500 images from the scene which elevation angle is 0-50 degrees. Since how to see a human changes with elevation angle, the scene in each elevation angel is used. In addition, there is also a problem that a total individual is not projectable, so that an elevation angle becomes large. Since a human may not be detected from these problems, the positive sample for every elevation angle is prepared. The number of the images of negative sample is 20,000 images, and the elevation angle is not related. The size of all sample images is 64×128. Since it aims at applying to a surveillance camera, it is verifying on the scene of various elevation angle. Therefore, we do not use the data set of INRIA.

B. *Experimental results*

In the first experimental, the performance of human detection between normal image and subtraction image are compared on an original data set. Since it uses for a subtraction image determining the scan region and size of a detection window, the HOG feature in a detection window is calculated from a normal image. Therefore, when the human detection is performed by normal image and subtraction image, the classifier built the same data set is used. The classifier used in this experiment is a one-step boosting. The human detection results are evaluated by four measures, True Positive rate (TP), False Positive rate (FP), Precision (TP/(TP+FP)) and Processing Speed (PS). There are two verification methods by the normal image: (1) scanning detection window size is fixed by 30×60; (2) it size is fixed by 60×120. Then paraperspective projection is assumed and the size of detection window recognized to be a human is rectified.

As Table 1 shows, the human detection with subtraction image improved FP and PS than with normal image [1]. Because we removed the background information that it is not relationship with detection human.

TABLE I
PERFORMANCE COMPARISON BETWEEN THE PROPOSAL AND REFERENCE [1]

	TP (%)	FP (%)	Precision (%)	PS (ms)
Reference [1] (30×60)	77.2	10.5	88.0	423.4
Reference [1] (60×120)	71.2	7.8	90.1	223.8
Proposal with subtraction stereo (Full body,one-step)	**78.5**	**3.2**	**96.1**	**58.3**

TABLE II
PERFORMANCE COMPARISON BETWEEN THE PROPOSED MODELS AND TRAINING METHODS

	TP (%)	FP (%)	Precision (%)
Full body (one-step)	78.5	3.2	96.1
Upper body (one-step)	86.2	14.4	85.7
Full-body (two-step)	80.3	1.2	98.5
Upper body (two-step)	**86.9**	10.2	89.4
Full + Upper (one-step)	77.8	0.9	99.0
Full + Upper (two-step)	79.8	**0.6**	**99.2**

In addition, since the HOG feature is very small and the number of a scanning detection window become at once, the human detection is enabled to real time processing. Then, that the performance of PS improved has the method of scanning detection window only foreground region and dynamically changing detection window size by using subtraction stereo. In the scene of Figure 6, since the elevation angle is small and the height of a camera is low, the size of a detection window changes by the distance from a camera to a human. Therefore high computation time can be decreased by the method of dynamically changing detection window size. Moreover, in the scene of Figure 7, since the size of the detection window does not change so much regardless of distance from a camera to a human, high computation time can be decreased by the method of scanning detection window only foreground region. From above, computation time can be considerably decreased to any scene.

In the second experiment, we aim to evaluate the performance of our classifier based on two-step boosting. Therefore, we compare multiple classifiers built by one-step boosting which method of Viola [2], and those classifiers built by two-step boosting which we propose. Furthermore, since false detection may occur mostly only in a full-body classifier, the human detection methods using multiple classifiers are performed by both one-step and two-step boosting, and the effectivity of human detection is verified. The multiple classifiers are scanning upper part of body classifier, after scanning full-body classifier. The date set used for classifier construction in this experiment is the same as the fist experiment. The human detection results are evaluated by three measures, TP, FP, and Precision.

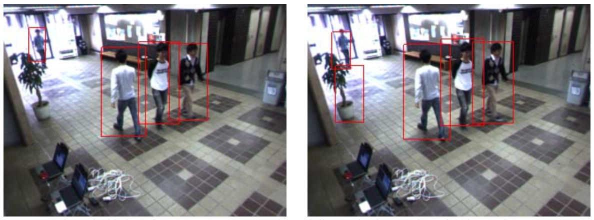


Fig. 6. The full-body (two-step) classifier (left). The full and upper part of body (two-step) classifier (right).

As Table 2 shows, the full-body classifier built by two-step boosting has the accuracy of human detection higher than this one built by one-step boosting. It is because the classifier was built in the feature which can describe a human more by performing two-step boosting. The feature which can express a human is specific to humans, such as a head, a shoulder, and a leg. The result of the human detection by the full-body classifier when using one-step and two-step boosting is shown in Figure 6. In addition, it is the same as that of the full-body classifier also about the human detection accuracy of the upper part of body classifier in one-step and two-step boosting. However, the full-body classifier has a lot that false detection like the automatic door of a Figure 7. This is because the feature in the straight line direction of an intensity gradient like a leg of human is included in construction of the full-body classifier.

VII. Conclusion

In this paper, we presented a fast and stable human detection algorithm using the subtraction stereo with HOG feature. Since the scanning region was putted down to the foreground regions by using the subtraction stereo, the accuracy and processing speed of human detection have been improved. In addition, the accuracy of the human detection has been improved using the full-body and the upper part of body classifier built by two-step boosting which we proposed.

However, when the distance from the stereo camera to a human is far, the accuracy of the human detection is lower than the camera to a human is near, which is the next problem to be tackled.

References

[1] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", *IC PR*, *Vol. 2*, pp.886-893, June 2005.

[2] P. Viola, J. Jones, "Rapid object detection using a boosted cascade of simple features", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 511-518, 2001.

[3] K. Levi, Y. Weiss, "Learning object detection from a small number of example: The importance of good feature", *IEEE Computer Vision and Pattern Recognition*, Vol. 2, pp.53-60, 2004.

[4] B. Wu, R. Nevatia, "Detection of multiple, partially occluded human in a single image by bayesian combination of edgelet part detectors", *IEEE International Conference on Computer Vision*, vol. 1, pp.90-97, 2005.

[5] K. Umeda, et al., "Subtraction Stereo –A Stereo Camera System That Focuses On Moving Regions-", *Proc. Of SPIE-IS&T Electronic Imaging, Vol*. 7239 Three-Dimensional Imaging Metrology, 723908, 2009.

[6] A. Shashua, Y. Gbalyahu, and G. Hayun, "Pedestrian detection for driver assistance systems: Single-frame classification and system level performance", in *In Proc. of the IEEE Intelligent Vehicle Symposium*, 2004.

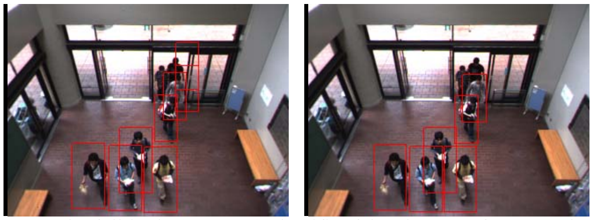


Fig. 7. The full-body (one-step) classifier false-detects a foreground region with the feature in the straight line direction of an intensity gradient like the automatic door (left). The hull-body (two-step) classifier is not detection the automatic door (right).

Moreover, in the upper part of body classifier, since there is few kind of feature to constitute, different false detection from the full-body classifier may occur. Then, the multiple classifiers built by two-step boosting have the performance of TP and FP better then one-step boosting. It is because the performance of the two-step boosting of each classifier is better than the one-step boosting. In addition, There is also a reason the kinds of false detection differ, in the full-body and the upper part of body classifier built by two-step boosting. From these results, the classifier which was most suitable in consideration of three measures is the full-body and the upper part of body classifier built by two-step boosting.

[7] P. Viola, M. Jones, and D. Snow, "Detection pedestrian using patterns of motion and appearance", in *ICCV*, 2003, pp.734-741.

[8] P. Sabzmeydani, G. Mori, "Detection pedestrians by learning shapelet features", in *CVPR*, 2007.

[9] O. Tuzel, F. Porinki, and P. Meer, "Human detection via classification on riemannian manifolds", in *CVPR*, 2007.

[10] D. M. Gavrila, S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle", *IJCV*, *vol*. 73, pp.41-59, 2007.

[11] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. H. Matthies, "Results from a real-time stereo-based pedestrian detection system on a moving vehicle", in *IEEE Workshop on People Detection and Tracking at ICRA*, 2009.

[12] A. Ess, B. Leibe, and L. Van. Gool, "Depth and appearance for mobile scene analysis", in *ICCV*, 2007.

[13] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "A mobile vision system for robust multi-person tracking", in *CVPR*, 2008, pp.734-741.

[14] A. Ess, B. Leibe, K. Schindler, and L. Van. Gool, "Moving obstacle detection in highly dynamic scenes", in *In Proceedings of ICRA*, 2009.

[15] K. Fuerstenberg, K. Dietmayer, and V. Willhoeft, "Pedestrian recognition in urban traffic using a vehicle based multilayer laserscanner", in *IEEE Intelligent Vehicle Symposium*, *vol*. 1, 2002.

[16] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark", in *CVPR*, 2009.

[17] D. G. Lowe, "Distinctive image feature from scale-invariant key-points", IJCV, vol. 60, pp.91-110, 2004.

[18] Q. Zhu, S. Avidan, M. Yeh, and K. Cheng, "Fast human detection using a cascade of histograms of oriented gradients", *In Proc. of CVPR*, 2006.

[19] O. Tuzel, F. Porinki, and P. Meer, "Region covariance: A fast descriptor for detection and classification", *In Proc. of CVPR*, 2006.

[20] S. Paisitkriangkrail, C. Shen, and J. Zhang, "An experimental study on pedestrian classification using local features", *In IEEE Inter. Symp. on Circuit and System* (ISCAS), 2008.

[21] S. Munder, D. Gavrila, "An experimental study on pedestrian classification", *PAMI*, 28(11), 2006.

[22] A. Moro, et al., "Auto-adaptive threshold and shadow detection approaches for pedestrian detection", In *Proc. AWSVCI*, pp.9-12, 2009.

[23] D. Comaniciu, P.Meer, "Mean Shift Analysis and Applications", IEEE International Conference on Computer Vision, pp.1197-1203, 1999.