Detector Ensemble based on False Positive Mining for Pedestrian Detection

Yuki Suzuki
Graduate School of Info. Sci.,
Nagoya University, Japan
suzukiy@murase.m.is.nagoya-u.ac.jp

Daisuke Deguchi
Info. Strategy Office,
Nagoya University, Japan
ddeguchi@nagoya-u.jp

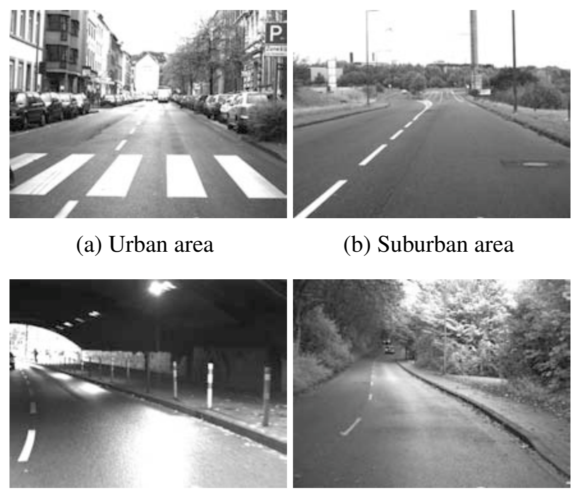
Yasutomo Kawanishi
Graduate School of Info. Sci.,
Nagoya University, Japan
kawanishi@is.nagoya-u.ac.jp

Ichiro Ide
Graduate School of Info. Sci.,
Nagoya University, Japan
ide@is.nagoya-u.ac.jp

Hiroshi Murase
Graduate School of Info. Sci.,
Nagoya University, Japan
murase@is.nagoya-u.ac.jp

Abstract

In recent years, the demands for Advanced Driving Assistance Systems (ADAS) is increasing, and pedestrian detection methods using an in-vehicle camera have been widely studied. In the case of pedestrian detection using an in-vehicle camera, since road environment varies widely, it is very difficult to do so accurately by a single classifier. This wide variety could also be understood as large intra-class variation of backgrounds, which leads to the increase of over-detections. To overcome this problem, this paper proposes a method of pedestrian detection using environment clustering based on false detection tendencies. By analyzing the false detection tendency in each environment, the proposed method creates classifiers that can cope with false detections observed in the specific environment. In addition, detector ensemble is introduced to extend this idea for handling multiple environments at the same time. To evaluate the effectiveness of the proposed method, experiments were conducted on the Daimler mono benchmark datasets. Results showed that the proposed method outperformed the conventional methods.



(a) Urban area (b) Suburban area

(c) Tunnel (d) Forest area

Figure 1. Difference of appearance in in-vehicle camera images for several road environments.

1. Introduction

In recent years, to prevent collision between vehicles and pedestrians, collision avoidance systems have been actively developed as one of the Advanced Driving Assistance Systems (ADAS) [1]. In particular, many research groups and automotive industries have proposed pedestrian detection methods using an in-vehicle camera [2]. Since an in-vehicle camera is inexpensive and its resolution is relatively high, it is becoming an important sensor for obtaining information on the environment surrounding a vehicle.

Machine learning technique is usually employed to detect pedestrians from an in-vehicle camera. For example, Dalal et al. proposed a pedestrian detection method by combining the Histograms of Oriented Gradients (HOG) descriptor and the SVM classifier [3]. Felzenszwalb et al. extended this idea to cope with various appearances of pedestrians (e.g. posture changes) by parts-based model using the latent SVM classifier [4].

In order to take a machine learning approach, a huge amount of training data needs to be prepared for the classifier to be able to deal with various appearances. However, in general, it is well known that the accuracy of a classifier decreases as the intra-class variation increases. In the case of pedestrian detection using an in-vehicle camera, it is very difficult to handle these variations with a single classifier, because the appearance of road environment varies largely according to geo-locations, illumination conditions, weather conditions and so on (Figure 1). This also leads to the increase of intra-class variation of backgrounds (negatives), which leads to the increase of over-detections. One reasonable solution to solve this problem is applying a pedestrian detection method that is capable to adapt with environment changes [5, 6]. Siva et al. proposed an object detection method using several background images extracted from a database depending on the query image (similar to the query image) [5]. Suzuo et al. proposed a method that classifies environments based on their appearances frame by frame, and constructs a classifier for each environment [6]. Since these methods classify environments based on their appearances, they work well when each environment consists of a simple appearance. However, both methods struggle when an environment contains a variety of appearances at the same time (e.g. buildings in a forest area).

On the other hand, when we observe detection results carefully, characteristics of false detections are found in each environment, such as trees, utility poles, and so on. They show that false detections are observed commonly between various environments. Based on this observation, false detection can be a key to classify environments. Accordingly, this paper proposes a method to improve the detection accuracy by classifying road environments based on the tendency of false detection and constructing an ensemble of detectors corresponding to false detection tendencies. Here, each detector detects pedestrians in the sliding window manner by evaluating a classifier.

Contributions of this paper are as follows:

- This paper introduces a novel concept of the tendency of considering false detections for clustering road environments. Using this information, the proposed method classifies road environments that are hard to be dealt with by the classifier containing various appearances, and adapts the detector to each of the tendencies. It shows more robustness when dealing with various appearances observed in an environment at the same time.

- The proposed method constructs the classifiers considering the tendency of false detections. This improves the detection accuracy when an environment consists of several background images that have multiple false detection tendencies.

2. Detector ensemble based on false detections

As the major problem in pedestrian detection, the degradation of the detection accuracy is caused by the varying appearance observed in an road environment. As explained in Section 1, it is difficult to handle this problem by a single detector. On the other hand, the detection accuracy can be increased by using detectors trained for particular scenes (such as forest area, urban area, and so on) [6]. For example, to detect a pedestrian in a forest scene, Suzuo et al. used a detector trained with forest scenes, and in an urban area, a detector trained with urban scenes. However, the problem arises when a scene consists of multiple appearances at the same time. For example, an environment where a part of an image looks like a forest area, but the other parts look like an urban area. In this case, their method cannot classify road environments properly, and the detection accuracy decreases. Figure 2 shows several false detections observed in a certain road environment. In the forest scene, false detections were observed not only in trees but also in traffic signs and building windows. Since there are multiple appearances in the scene, many false detections can occur. To deal with this problem, this paper extends the idea of Suzuo et al.'s road environment clustering method by introducing a novel criteria based on false detection tendency. Since false detections are backgrounds incorrectly detected by a classifier, they can be considered as backgrounds that is hard to be dealt with by the detector. Motivated by this deduction, the environments containing these backgrounds can be determined by false detections. Figure 3 shows examples of false detections by a HOG+SVM detector from

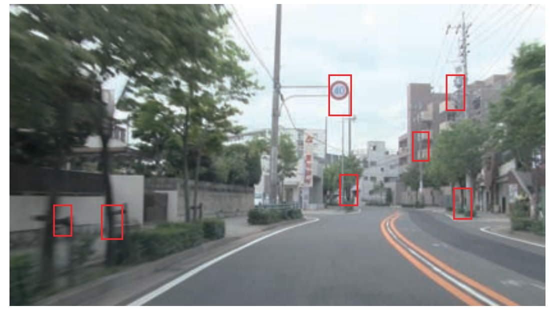


Figure 2. Example of various false detections in a road environment.

Figure 3. Examples of false detection.

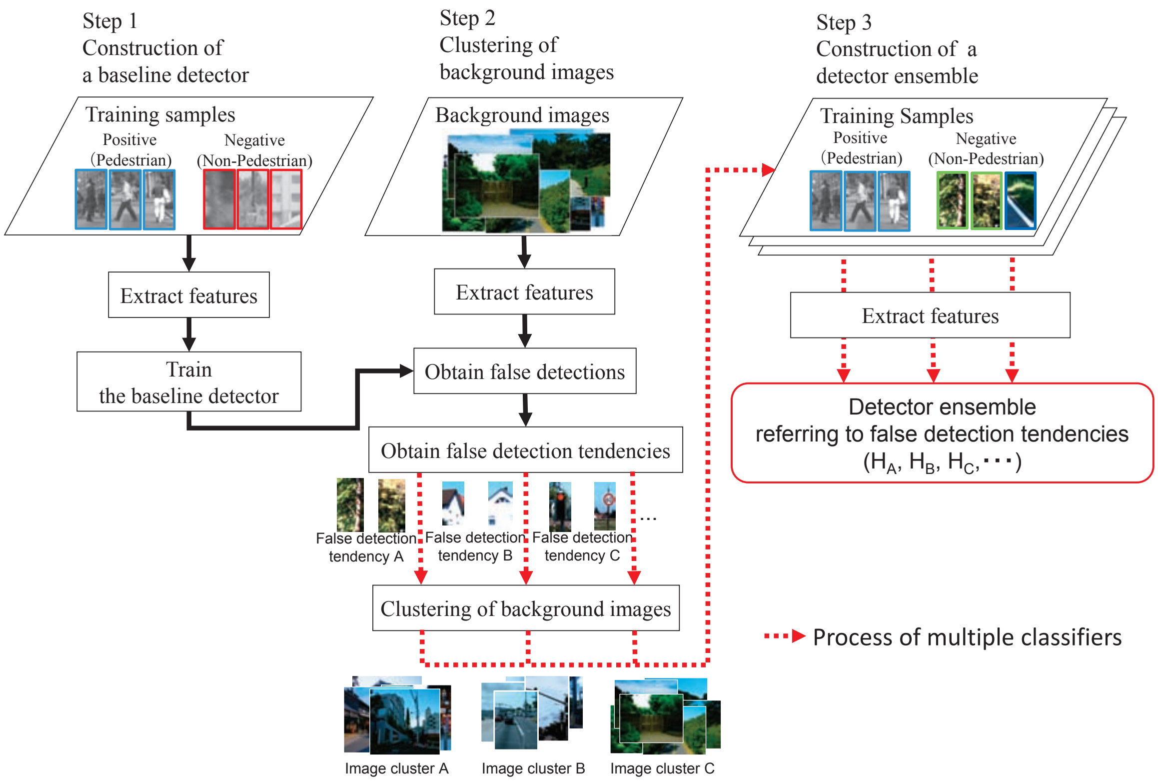


Figure 5. Training process flow of the adaptive classifier referring to false positive tendency.

(a) Traffic signs in an urban area. (b) Traffic signs in a forest area.

Figure 4. Example of false detections in road environments with different appearances.

an in-vehicle camera image captured in a road environment. Various false detections are observed, such as trees, building windows, traffic signs, traffic signals, and utility poles. Figure 4 shows examples of false detections obtained in different scenes (urban and forest areas). Although the appearances of the entire image are different, similar false detections (traffic signals, traffic signs, trees, and utility poles) are observed. That is, occurrence of false detections does not depend on the scene, and the tendency of their occurrences can be shared between different scenes.

From the above points of view, if the tendency of false detections is considering for the clustering of environments, we can expect three advantages. First, it allows us to divide environments according to backgrounds that are hard to be dealt with by the detector. Second, since the appearance of the entire image does not affect the results of environment clustering, the problems caused by the use of the entire image for appearance clustering can be solved. Finally, there is a possibility that an additional classifier can be trained specifically for backgrounds that are hard to be dealt with by a certain detector.

The following section explains the detailed process for constructing detectors using the tendency of false detections.

3. Pedestrian detection based on false detection tendency

The proposed method consists of a training phase and a detection phase. The following sections describe each part in detail.

3.1. Training phase

Figure 5 shows the training flow of the proposed method. As seen in the image, the training phase consists of the following three steps.

Step 1: Construction of a baseline detector.

Step 2: Clustering of background images based on false detection tendency.

Step 3: Construction of a detector ensemble using the clustering results.

3.1.1 Step 1: Construction of a baseline detector

In this step, a baseline detector is constructed using the HOG feature with the SVM classifier (HOG+SVM classifier). When constructing a HOG+SVM classifier, pedestrian and non-pedestrian images are prepared. Here, non-pedestrian images are sampled randomly from background images of all road environments.

3.1.2 Step 2: Clustering of background images based on false detection tendency

Since a baseline detector is trained using various appearances of all road environments, false detections will occur where the background includes appearances that are hard to be dealt with by the detector. The proposed method gathers these false detections, and considers them for clustering environments.

First of all, by using the baseline detector, false detections are gathered from background images of all road environments. Then, k-means clustering is applied to the false detections, and false detection tendency is represented by the center of each cluster. Here, Euclidean distance between PCA-HOG [7] features is used as a metric of k-means clustering.

Next, the proposed method computes the one-to-many relationships between a background image and false detection tendencies. The proposed method counts the number of false detections corresponding to each false detection tendency. Here, the distance between a false detection and false detection tendency (cluster center) is measured, and false detection tendencies whose distance is smaller than a certain threshold are selected.

3.1.3 Step 3: Construction of a detector ensemble using the clustering results

In this step, the proposed method constructs detectors for each false detection tendency. Here, each detector is trained using background images (negatives) associated with the false detection tendency. Pedestrian images (positives) used in this step are exactly the same as those of a baseline detector. Finally, multiple detectors are combined to form a detector ensemble. Details about the detector ensemble are explained in the next section.

3.2. Detection phase

First of all, all detectors including the baseline detector are applied to an input image. In this process, multiple detection windows are placed over the entire region of an image by changing the scale and the location. Here, each detector evaluates the detection windows separately, and the proposed method outputs the detection window voted by more than half of the detectors.

(a) Conventional method (b) Proposed method

Figure 6. Example of detected pedestrians

4. Evaluation

To evaluate the effectiveness of the proposed method, we conducted an experiment. We compared the proposed method with the detection method by Dalal et al. [3]. The following sections describe details about the test dataset and parameters for constructing the detectors.

4.1. Experimental data

In this experiment, Daimler mono benchmark dataset [8] is used for evaluation. This dataset contains gray-scale images captured by a single in-vehicle camera. As the test data, 8,000 images were selected from this database.

We used pedestrian images provided for training in the dataset as pedestrian image, and as non-pedestrian images, images extracted from parts of the training images not labeled as pedestrian. These non-pedestrian images were used for both clustering of false detections and training detectors.

4.2. Parameters for constructing detectors

Detection target in this experiment was a pedestrian whose size was larger than or equal to 48×96 pixels. Di-

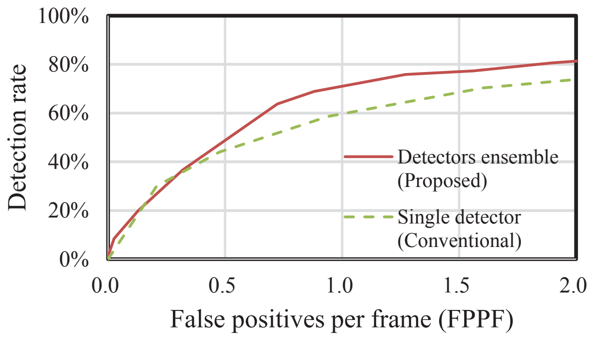


Figure 7. Detection accuracy in FROC curve.

mensions of the HOG feature was 6,024, and its cell size and its block size were 5 cells and 6 pixels, respectively. The number of the clusters that the false detection tendencies were clustered was 5. In the experiment, LIBLINEAR[1] and its default parameters were used for constructing detectors.

4.3. Results and discussions

To evaluate the effectiveness of the proposed method, we compared it with a conventional pedestrian detector [3]. By following the criteria used in the PASCAL VOC Dataset [9], the detection is considered as correct when the overlapped ratio was greater than or equal to a certain threshold (0.3 was used in the experiment). Figures 6 and 7 show a comparison of the proposed method and the conventional method in FROC-curve. As seen in the graph, the proposed method outperformed the conventional method, which indicates that the proposed method reduced false detections in comparison with the conventional method. Especially, the proposed method outperformed the conventional method by 9% when FPPF was 1.0. Although the proposed method used HOG+SVM as a baseline detector, an arbitrary detection method (such as DPM [4]) can be employed in the proposed framework. Therefore, we will consider introducing other detection methods in the future.

5. Conclusion

This paper proposed a method of constructing an ensemble of multiple pedestrian detectors using false detection tendencies. Since false detections are incorrectly detected backgrounds, we can then determine the environments that are hard to be dealt with by the classifier. The proposed method uses this character for clustering road environments, and constructs a classifier for each environment. Finally, the constructed classifiers are combined to form a detector ensemble. We evaluated the accuracy and the effectiveness of the proposed method by applying it to the Daimler dataset. Experimental results showed that the proposed method outperformed the conventional method. Future works include: (i) evaluations by a larger dataset, (ii) introduction of Deformable Part Model and Deep-Learning for improving the detection accuracy.

[1] http://www.csie.ntu.edu.tw/~cjlin/liblinear/

Acknowledgement

Parts of this research were supported by a Grant-In-Aid for Young Scientists from MEXT, a Grant-In-Aid for Scientific Research from MEXT, and Nagoya University COI.

References

[1] D. Geronimo, M. A. Lopez, D. A. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," IEEE Trans. on Pattern Analysis and Machine Intelligence, 32(7):1239–1258, 2010. 1

[2] P. G. Stein, G. Yoram, and S. Amnon, "Stereo-assist: Top-down stereo for driver assistance systems." In Proc. IEEE Intelligent Vehicles Symposium 2010, pages 723–730, 2010. 1

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pages 886–893, 2005. 1, 4, 5

[4] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," In Proc. 2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pages 1–8, 2008. 1, 5

[5] P. Siva, R. Chris, X. Tao, and A. Lourdes, "Looking beyond the image: Unsupervised learning for object saliency and detection," In Proc. 2013 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pages 3238–3245, 2013. 2

[6] D. Suzuo, D. Deguchi, I. Ide, H. Murase, H. Ishida, and Y. Kojima, "Environment adaptive pedestrian detection using in-vehicle camera and GPS," In Proc. Int. Conf. on Computer Vision Theory and Applications 2014, 354–361, 2014. 2

[7] T. Kobayashi, A. Hidaka, and T. Kurita, "Selection of Histograms of Oriented Gradients features for pedestrian detection," Neural Information Processing, 14th Int. Conf. ICONIP2007, Kitakyusyu. Japan, Nov.13–16, 2007, Revised Selected Papers, Part II, Lecture Notes in Computer Science, 4985, pages 598–607, 2008. 4

[8] M. Enzweiler and G. M. Dariu, "Monocular pedestrian detection: Survey and experiments," IEEE Trans. on Pattern Analysis and Machine Intelligence, 31(12):2179–2195, 2009. 4

[9] M. Everingham, L. Van Gool, C. K. Williams, and J. Winn, "The PASCAL visual object classes (VOC) challenge," Int. Journal of Computer Vision, 88(22):303–338, 2010. 5