2013 IEEE Intelligent Vehicles Symposium (IV)
June 23-26, 2013, Gold Coast, Australia

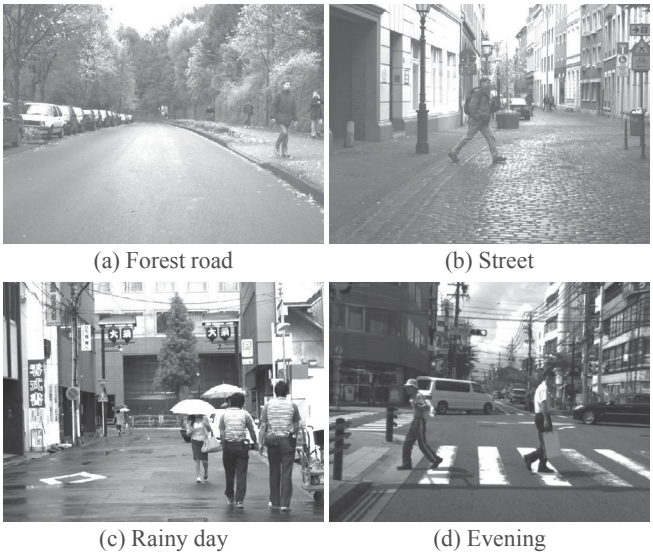
Pedestrian detection by scene dependent classifiers with generative learning

Hidefumi Yoshida[1], Daichi Suzuo[1], Daisuke Deguchi[2], Ichiro Ide[1], Hiroshi Murase[1], Takashi Machida[3] and Yoshiko Kojima[3]

Abstract— **Recently, pedestrian detection from in-vehicle camera images is becoming an crucial technology for Intelligent Transportation Systems (ITS). However, it is difficult to detect pedestrians accurately in various scenes by obtaining training samples. To tackle this problem, we propose a method to construct scene dependent classifiers to improve the accuracy of pedestrian detection. The proposed method selects an appropriate classifier based on the scene information that is a category of appearance associated with location information. To construct scene dependent classifiers, the proposed method introduces generative learning for synthesizing scene dependent training samples. Experimental results showed that the detection accuracy of the proposed method outperformed the comparative method, and we confirmed that scene dependent classifiers improved the accuracy of pedestrian detection.**

I. INTRODUCTION

In recent years, pedestrian detection has become one of the most interesting topics in the computer vision field, and it is also becoming one of the most crucial technology for surveillance systems and for Intelligent Transportation Systems (ITS). Especially, for ITS, pedestrian detection plays an important role to prevent traffic accidents. Therefore, many research groups have tackled this problem, and several methods have been proposed [1], [2], [3], [4]. One of the most successful methods to detect pedestrians from an image is a method that employs Histogram of Oriented Gradients (HOG) and Support Vector Machine (SVM) [5]. Enzweiler et al. [2] reported that pedestrian detection using HOG and SVM could detect pedestrians accurately through experiments using in-vehicle camera images. Although this method can detect pedestrians accurately, it requires a tremendous number of training samples with widely variable appearance gathered manually to construct the classifier. Thus, the accuracy of pedestrian detection is highly dependent on the quality and the variety of samples used for training the classifier. However, as seen in Fig. 1, the appearance of a pedestrian and its background differs widely depending on the scenes, such as an urban area or a suburban area, a sunny day or a rainy day, and so on. Therefore, it is difficult to construct a classifier that can detect pedestrians accurately in various scenes by obtaining training samples exhaustively. To overcome this problem, Wöhler [6] proposed a method that obtains pedestrian images automatically from in-vehicle camera images by combining detection and tracking of pedestrians. This method enables us to construct an accurate classifier with a small number of manual interventions. The pedestrian images for training are successively extended by repeating the detection and tracking in this approach. However, pedestrian images can be obtained only when pedestrians are detected. Therefore, to gather various pedestrian images, it requires the collection of a large amount of in-vehicle camera image sequences. In general, gathering the pedestrian images becomes all the more difficult when this approach is applied to places in which people are seldomly present.



(a) Forest road (b) Street

(c) Rainy day (d) Evening

Fig. 1. Examples of the appearance changes of pedestrians in different scenes. (a) and (b) are from the "Daimler Mono Pedestrian Detection Benchmark Data Set" [2] taken in Germany, and (c) and (d) were taken by the authors in Japan.

On the other hand, it is possible to construct a classifier that works accurately only in a specific scene by using a relatively small number of training samples. This is because the variety of appearances of pedestrians is not so wide in a specific scene. Wang et al. [7] applied this idea of adaptation to pedestrian detection using a stationary camera. In the case of the road environment (using an in-vehicle camera), scene is one of the important factors that affects the appearance of the pedestrian combined with its background. Based on

[1]Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan yoshidah@murase.m.is.nagoya-u.ac.jp, suzuod@murase.m.is.nagoya-u.ac.jp, ide@is.nagoya-u.ac.jp, murase@is.nagoya-u.ac.jp

[2]Information and Communications Headquarters, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan ddeguchi@nagoya-u.jp

[3]Toyota Central Research & Development Laboratories, Inc., 41-1, Yokomichi, Nagakute, Aichi 480-1192, Japan

these ideas, this paper proposes a method that constructs scene dependent classifiers to improve the accuracy of pedestrian detection from in-vehicle camera images. The scene information that is a scene category obtained from the GPS location is used for the selection of the classifier optimized for each scene. There is no conventional approach to improve the accuracy of pedestrian classifiers via scene information. In addition, this paper introduces an approach of generative learning [8], [9], [10], [11] to synthesize a large number of scene dependent training samples.

The main characteristics of this paper are the following two points.

1) Construction of scene dependent classifiers for pedestrian detection from in-vehicle camera images.
2) Introduction of generative learning to construct classifier for generating various training samples adaptive to each scene category.

This paper consists of five parts. At first, details of scene dependent classifiers are described in section II. Section III describes an overview of the proposed method and detailed procedures for detecting pedestrians by using scene dependent classifiers. Then, section IV describes an experiment using in-vehicle camera images. Finally, we conclude the paper in section V.

II. SCENE DEPENDENT CLASSIFIER

In this paper, the proposed method considers three scene categories; a residential scene, a forest road scene, an urban scene. Each in-vehicle camera image is classified into one of these scene categories based on its appearance. Since the geographic location of each in-vehicle camera image can be obtained by GPS, the location on a map can be associated with its scene.

Figure 2 illustrates the basic concept of scene dependent classifiers. A scene dependent classifier is a classifier associated with a scene category, which is optimized so that it can accurately detect pedestrians adaptive to the scene category. The proposed method represents the relationship between classifiers and scene categories in the form of a map namely, "Classifier map". This map is constructed prior to the pedestrian detection. Here, locations in the map are classified by their appearances, and they are associated with scene categories. When detecting pedestrians at a certain location, the corresponding scene dependent classifier is selected according to the map.

III. PEDESTRIAN DETECTION BASED ON SCENE DEPENDENT CLASSIFIERS

In this section, we propose a method to construct scene dependent classifiers to improve the accuracy of pedestrian detection. To improve the detection accuracy, a classifier should be optimized to the actual scene category that the classifier will be applied, and the constructed classifiers should be used according to the scene category where a vehicle is running. However, it is very difficult to obtain various samples exhaustively for each scene category. To solve these problems, the proposed method takes a generative learning approach. Here, training samples are generated using pedestrian textures and non-pedestrian images obtained beforehand in each scene.

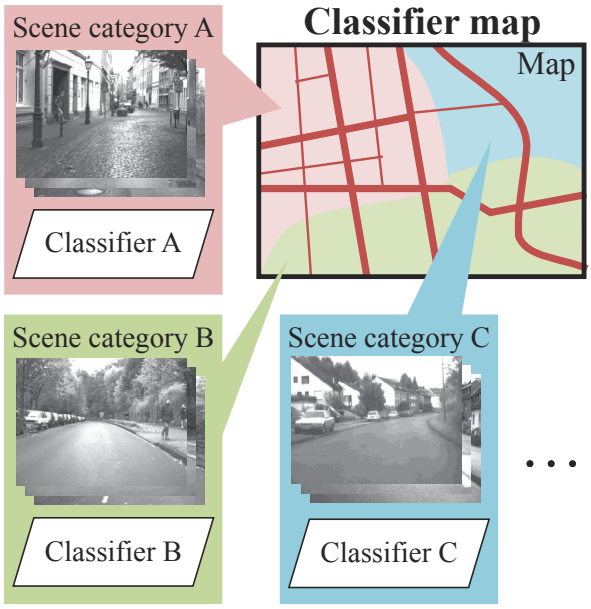


Fig. 2. The basic concept of scene dependent classifiers. In this figure, the "Classifier map" contains several scene categories with a classifier associated with each of them.

A. Overview

As seen in this figure, the proposed method consists of two phases; (i) the training phase and (ii) the detection phase.

In the training phase, the pedestrian images are generated by the generative learning approach [9], [11]. This approach generates pedestrian textures with various appearances including shape change and texture change. Then, a pedestrian texture is super-imposed on a non-pedestrian image (background image) clipped from an in-vehicle camera image. Here, the proposed method uses non-pedestrian images obtained in each scene for training a scene dependent classifier. A non-pedestrian image of each scene is not only used as a background image for generating a positive sample (pedestrian image) but also used as a non-pedestrian image for a negative sample. Thus, each classifier is optimized for each scene. Finally, the proposed method constructs the scene dependent classifiers by using the generated pedestrian images and non-pedestrian images, and then associated with the map.

In the detection phase, a scene dependent classifier is used to detect pedestrians from in-vehicle camera images. The classifier is selected from the classifier map according to the actual location obtained by GPS. The details of these two phases are described below.

Figure 3 shows the flowchart of the proposed method.

B. Training phase

As seen in Fig. 4, according to its scene, the proposed method constructs scene dependent classifiers by using non-pedestrian images gathered from in-vehicle camera images.

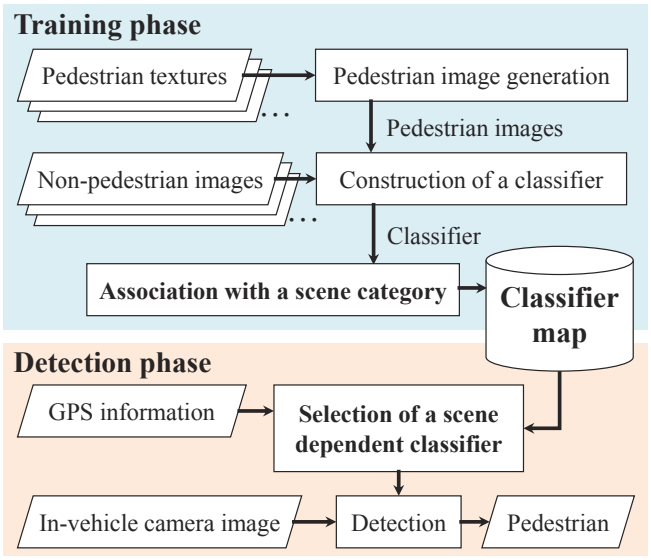


Fig. 3. Process flow of the proposed method.

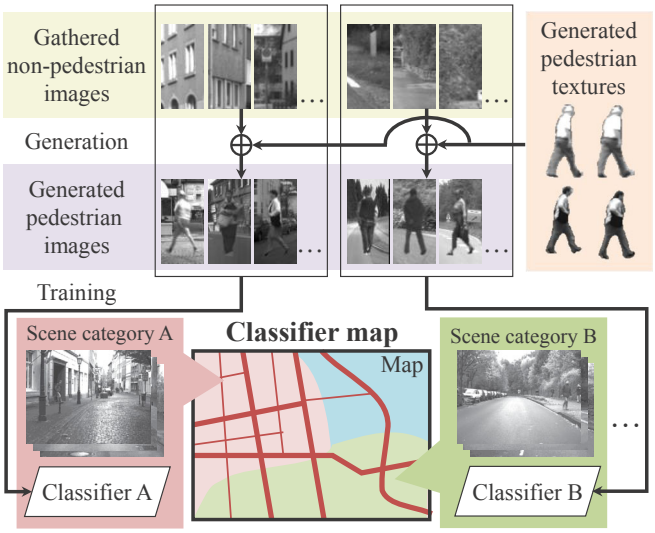


Fig. 4. Construction of the scene dependent classifiers. Non-pedestrian images are gathered from each scene, and pedestrian images are generated by using pedestrian textures and these non-pedestrian images. Classifiers are constructed by using both pedestrian images and non-pedestrian images. Finally, the classifiers are associated with the map based on its scene category.

To construct a classifier, pedestrian images are generated by using a small number of pedestrian textures and these non-pedestrian images. Here, pedestrian textures are prepared manually prior to this training phase, and they are used for constructing each classifier. Finally, each classifier is trained with appearances of pedestrians in each scene category.

1) Gathering of non-pedestrian images: Non-pedestrian images are gathered from in-vehicle camera images captured prior to the training phase. To construct scene dependent classifiers, these in-vehicle camera images must be classified into scenes. Here, these non-pedestrian images are used for negative samples and for background images of pedestrian images (positive samples). Especially, background images are selected so that a pedestrian does not float in the sky

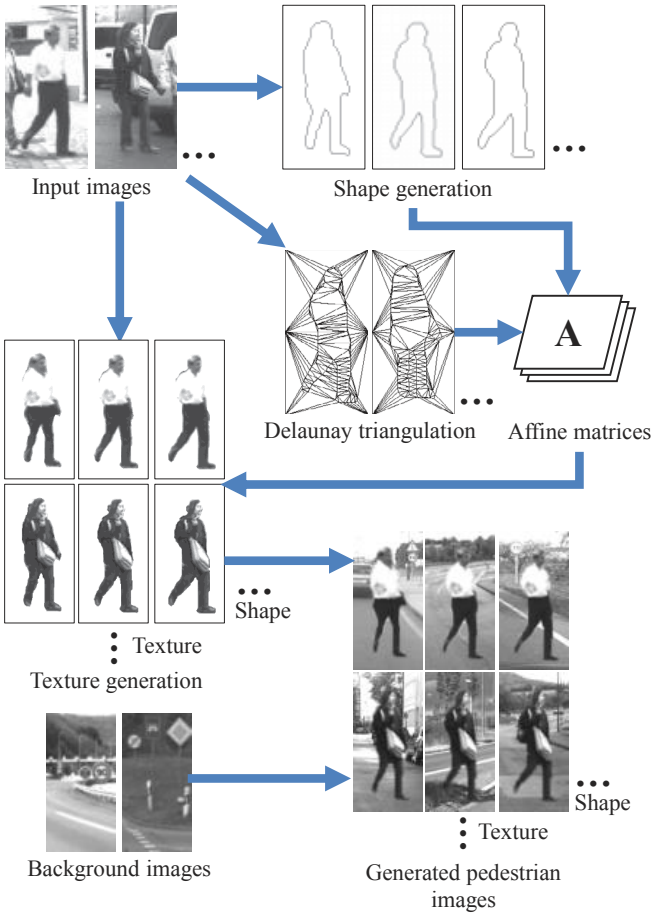


Fig. 5. Overview of the pedestrian image generation.

nor lies on the road surface by choosing appropriate clipping positions. This is done by referring to the vertical position of each clipped image.

2) Pedestrian image generation: To construct a scene dependent classifier, it is necessary to prepare pedestrian images depending on the scene category. The proposed method generates these pedestrian images by using the generative learning approach [9], [11]. It generates various pedestrians by changing contour shapes, textures and backgrounds as shown in Fig. 5.

The shape generation is performed by employing Statistical Shape Models (SSM) [12] as a generation model. By using SSM, the synthesized new shape $\boldsymbol{y}$ is represented as

$$\boldsymbol{y} = \bar{\boldsymbol{v}} + \mathbf{P}\boldsymbol{b}, \tag{1}$$

where $\bar{\boldsymbol{v}}$ is the mean vector corresponding to the pedestrian contours, and $\mathbf{P}\boldsymbol{b}$ represents the shape perturbation. Matrix $\mathbf{P}$ consists of eigenvectors obtained by applying PCA to pedestrian shapes in each pose class, and these eigenvectors are selected so that the cumulative contribution ratio of corresponding eigenvalues exceeds 99%.

Then, the texture generation is performed by applying a procedure similar to that in the shape generation. First, the proposed method applies the Delaunay triangulation algorithm to the points placed on the contour of a pedestrian, and then obtains a set of triangles as shown in Fig. 5. Then, the

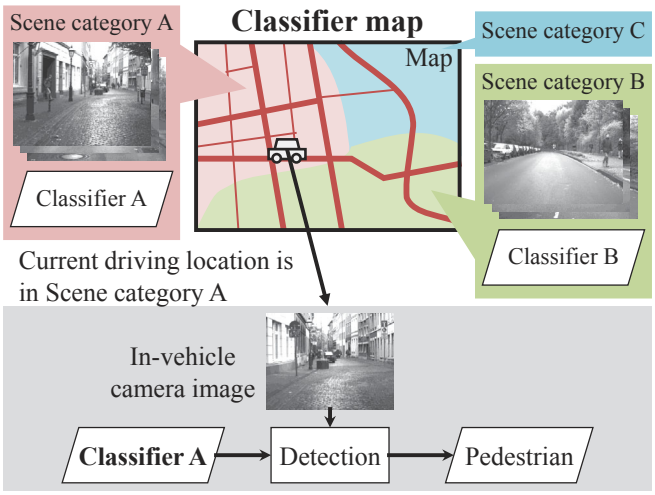


Fig. 6. Selection of the scene dependent classifier. In this figure, the "Classifier Map" contains three scene categories A, B and C. When a vehicle is at a location corresponding to scene category A, the proposed method selects classifier A, and uses it for the pedestrian detection.

proposed method computes an affine transformation matrix **A** for each triangle by referring to the result of the shape generation. This transformation transforms vertices of each triangle from an input pedestrian image to the synthesized shape. Then, the texture inside each triangle is mapped onto the synthesized shape by using this transformation matrix **A**. Finally, the proposed method applies this texture mapping process for all triangles obtained by the Delaunay triangulation algorithm. After applying the above process, variously textured pedestrian images for the same shape is obtained. By using these images, the proposed method synthesizes various textures for each shape. First, the proposed method represents intensities of each image as an intensity vector. Then, by applying the SSM algorithm to the intensity vectors, a new pedestrian texture is obtained.

As the last step of the synthesis, the proposed method combines the synthesized pedestrian image with various background images. The background images are extracted from in-vehicle camera images containing no pedestrian by changing the parameters such as the clipping position and the size of the clipping rectangle. To synthesize pedestrian images whose appearances can be observed in the scene, the proposed method applies non-pedestrian images extracted in the previous step as a background of the generated pedestrian images. Since we can assume that a pedestrian does not float in the sky nor lie on the road, the proposed method sets the parameters for background extraction so that an image is not composed of only the sky or a road surface. Finally, the proposed method uses feathering along the contour for synthesizing a pedestrian image super-imposed on a background image to make it natural.

3) Construction of a scene dependent classifier: For the discrimination, this paper uses non-linear SVM with Gaussian radial basis function kernel, and it is used for the pedestrian detection. For training the classifier, cuSVM [13] is used, which is an implementation of SVM using GPU. As reported in [14], [15], the GPU accelerated classifier is now widely used for fast pedestrian detection. HOG (Histogram of Oriented Gradients) [5] are used as features for training the classifier, and calculated from pedestrian images and non-pedestrian images. Finally, the classifier is associated with the map according to its scene.

C. Detection phase

The basic concept of selection of the scene dependent classifier is shown in Fig. 6. The input of this phase is an in-vehicle camera image with GPS location. The proposed method selects a classifier from the classifier map according to the current location and detects pedestrians by using the selected classifier.

1) Selection of the scene dependent classifier: The proposed method selects a classifier with the scene category according to the current driving location from the classifier map. When driving a vehicle, the proposed method detects pedestrians by using the selected classifier. Since the classifier is trained using samples obtained from the scene, it is expected to detect pedestrians more accurately than using the others.

2) Pedestrian detection: The proposed method detects pedestrians from an in-vehicle camera image by using the scene dependent classifier selected in the previous step. Here, pedestrian detection is performed by sliding a detection window over the entire region of an image, and each detection window is evaluated by applying the selected classifier. Finally, if a score of the classifier is larger than a threshold, the proposed method outputs the detection window as a pedestrian.

IV. Experiments

We evaluated the performance of the proposed method by using in-vehicle camera images captured in the three scene categories; residential scene, forest road scene and urban scene. The following sections describe details of the dataset, experimental setup, evaluation, and results.

A. Dataset

In this experiment, the proposed method was evaluated by using the "Daimler Mono Pedestrian Detection Benchmark Data Set"[1] [2]. We manually gathered 50 pedestrian images, and 100 in-vehicle camera images including no pedestrian for each scene from the training set of this dataset (Fig. 7).

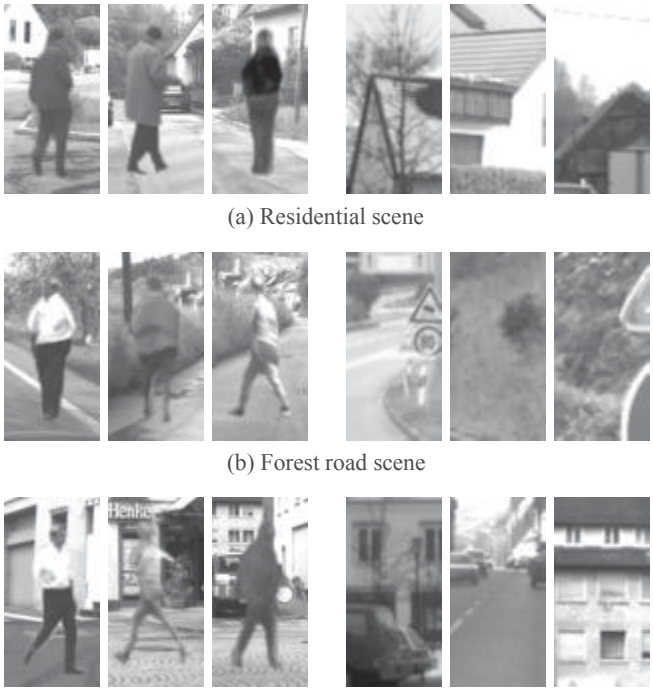
For validation, we prepared 940 in-vehicle camera images from the test set of the dataset, where 953 pedestrians were present in the images. The resolution of the in-vehicle camera images was 640×480 pixels and the training images was 48×96 pixels.

B. Experimental setup

In this experiment, since GPS information was not available in the dataset, we manually labeled the image sequence with three scene categories according to their appearances.

[1] http://www.gavrila.net/Research/Pedestrian_Detection/Daimler_Pedestrian_Benchmark_D/Daimler_Mono_Ped__Detection_Be/daimler_mono_ped__detection_be.html

Fig. 7. Examples of pedestrian images in the Daimler dataset [2].



(a) Residential scene

(b) Forest road scene

(c) Urban scene

Fig. 8. Examples of the generated pedestrian images and non-pedestrian images for each scene class.

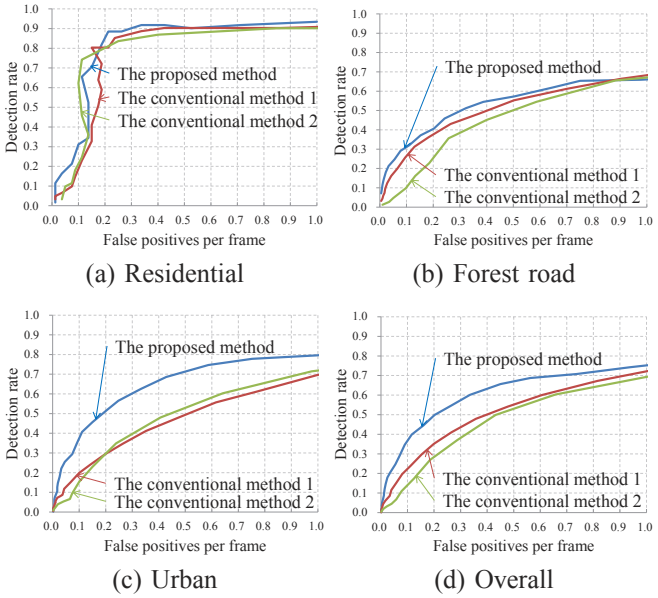


Fig. 9. ROC curves of the proposed method and the comparative methods. (d) shows the overall performance of (a), (b) and (c).

C. Evaluation

To evaluate the performance of the proposed scene dependent classifiers, we compared the proposed method with conventional methods.

The proposed method constructed a classifier for each scene of the image sequence taken from an in-vehicle camera, and the classifier corresponding to the target scene was used for the pedestrian detection. Every classifier was trained by using 5,000 pedestrian images and 5,000 non-pedestrian images. These 5,000 pedestrian images were generated from 50 pedestrian images. As non-pedestrian images, 5,000 images were randomly clipped from 100 in-vehicle camera images including no pedestrian described in section IV-A. Figure 8 shows examples of the generated pedestrian images and the gathered non-pedestrian images. On the other hand, the conventional methods constructed one classifier for the entire input image sequence. The conventional method 1 was constructed by using 5,000 generated pedestrian images and 5,000 non-pedestrian images of the Daimler dataset. Meanwhile, the conventional method 2 was constructed by using 50 pedestrian images and 5,000 non-pedestrian images from the Daimler dataset. Here, the 50 pedestrian images were the same as those used in the proposed method.

The detection rate was measured by evaluating the overlap between the detection result and the ground-truth labeled manually. Figure 9 shows the ROC curves drawn by changing the threshold of the classifier score described in section III-C.

D. Results and Discussion

Figure 9 shows the ROC curves of the proposed method and the conventional methods. As seen here, the proposed method outperformed the conventional methods. Since the proposed method constructed a classifier optimized for each scene and selected the classifier according to the scene, it achieved higher performance in comparison with the conventional methods. Figure 10 shows examples of the detection results in different scenes. The proposed method could reduce false negative by using pedestrian images synthesized depending on a scene. On the other hand, false positive alarms were reduced by learning non-pedestrian images gathered from a corresponding scene.

In the method proposed in this paper, we constructed classifiers at a certain moment according to a snapshot of a road scene. However, the appearance of the scene may change with time. Thus, it may be difficult to detect pedestrians accurately without the reconstruction of the classifier. To solve this problem, the proposed method should be extended by introducing an online reconstruction process of the scene dependent classifiers.

Fig. 10. Comparison of the detection results.

V. Conclusions

This paper proposed a new concept of scene dependent classifiers to improve the accuracy of pedestrian detection. The proposed method constructs a scene dependent classifier for scene categories and selectively applies a classifier according to the scene for detection. To construct a scene dependent classifier, the proposed method generated various training samples using images obtained at each scene category. We evaluated the accuracy and the effectiveness of the proposed method by applying it to in-vehicle camera images. Experimental results showed that the proposed method improved the accuracy of the pedestrian detection in various scenes. Future work will include (i) extension of the proposed method to other scene categories such as weather and illumination, (ii) introduction and evaluation of the reconstruction process of scene dependent classifiers, (iii) automatic clustering of non-pedestrian images according to the appearance obtained from an aerial image, and (iv) evaluation by a larger dataset.

Acknowledgements

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 a CREST project of JST. This work was developed using the MIST library (http://mist.murase.m.is.nagoya-u.ac.jp/).

References

[1] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2009*, June 2009, pp. 304–311.

[2] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and expreriments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.

[3] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, July 2010.

[4] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, Apr. 2012.

[5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005*, vol. 1, June 2005, pp. 886–893.

[6] C. Wöhler, "Autonomous in situ training of classification modules in real-time vision systems and its application to pedestrian recognition," *Pattern Recognition Letters*, vol. 23, no. 11, pp. 1263–1270, Sep. 2002.

[7] M. Wang and X. Wang, "Automatic adaptation of a generic pedestrian detector to a specific traffic scene," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2011*, June 2011, pp. 3401–3408.

[8] H. Murase, "Learning by a generation approach to appearance-based object recognition," in *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 1, Aug. 1996, pp. 24–29.

[9] M. Enzweiler and D. M. Gavrila, "A mixed generative-discriminative framework for pedestrian classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2008*, June 2008, pp. 1–8.

[10] J. Marín, D. Vázquez, D. Gerónimo, and A. M. López, "Learning appearance in virtual scenarios for pedestrian detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2010*, June 2010, pp. 137–144.

[11] H. Yoshida, D. Deguchi, I. Ide, H. Murase, K. Goto, Y. Kimura, and T. Naito, "Integration of generative learning and multiple pose classifiers for pedestrian detection," in *Proceedings of the International Conference on Conmputer Vision Theory and Applications*, vol. 1, Feb. 2012, pp. 567–572.

[12] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models. Their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[13] A. Carpenter, "cuSVM: A CUDA implementation of support vector classification and regression." [Online]. Available: http://patternson-ascreen.net/cuSVM.html

[14] S. Bauer, S. Kohler, K. Doll, and U. Brunsmann, "FPGA-GPU architecture for kernel SVM pedestrian detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2010*, June 2010, pp. 61–68.

[15] L. Zhang and R. Nevatia, "Efficient scan-window based object detection using GPGPU," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2008*, June 2008, pp. 1–7.