Intelligent Traffic Sign Detector: Adaptive Learning Based on Online Gathering of Training Samples

Daisuke Deguchi, Mitsunori Shirasuna, Keisuke Doman, Ichiro Ide and Hiroshi Murase

Graduate School of Information Science, Nagoya University

Furo-cho Chikusa-ku Nagoya, Aichi 464–8601, Japan

`ddeguchi@is.nagoya-u.ac.jp`

Abstract— This paper proposes an intelligent traffic sign detector using adaptive learning based on online gathering of training samples from in-vehicle camera image sequences. To detect traffic signs accurately from in-vehicle camera images, various training samples of traffic signs are needed. In addition, to reduce false alarms, various background images should also be prepared before constructing the detector. However, since their appearances vary widely, it is difficult to obtain them exhaustively by manual intervention. Therefore, the proposed method simultaneously obtains both traffic sign images and background images from in-vehicle camera images. Especially, to reduce false alarms, the proposed method gathers background images that were easily mis-detected by a previously constructed traffic sign detector, and re-trains the detector by using them as negative samples. By using retrospectively tracked traffic sign images and background images as positive and negative training samples, respectively, the proposed method constructs a highly accurate traffic sign detector automatically. Experimental results showed the effectiveness of the proposed method.

I. INTRODUCTION

In recent years, ITS (Intelligent Transport System) technologies play an important role to make our driving environment safe and comfortable. In ITS, in-vehicle camera is now widely used for recognition and understanding of the road environment. There are various objects in our road environment, and it is necessary to develop methods to detect and to recognize them accurately. Since traffic signs are very important information for driving, several research groups proposed methods to detect them automatically [1], [2], [3]. Nowadays, such traffic sign detection systems are also commercially available (BMW, Mercedes-Benz, etc.). However, they can only detect limited types of traffic signs, such as speed limit and overtaking restrictions.

One of the successful methods for detecting objects from an image is the method employing a cascaded AdaBoost classifier that is widely used for face detection [4]. Bahlmann et al. used this approach for traffic sign detection from in-vehicle camera images, and they showed that accurate traffic sign detection could be achieved. Although this approach is quite accurate and fast enough, it requires a large number of traffic sign images for training the classifier. Usually, these traffic sign images are manually obtained by specifying their clipping rectangle in in-vehicle camera images. Since this task is very expensive and time consuming, it is difficult to apply this approach to detect various traffic signs. To solve this problem, Doman et al. proposed a method for construct-

Fig. 1. Examples of various appearances of traffic signs.

ing the AdaBoost classifier by generating various traffic sign images based on image degradation models [3]. Although this method can greatly reduce the cost for preparing traffic sign images, it is difficult to generate various appearances of traffic signs actually observed in a real environment (Fig. 1). To construct a traffic sign detector easily and accurately, it is necessary to obtain a large number of training samples from a real environment without manual intervention. Wöhler et al. [5] and our group [6] tried to solve these problems by obtaining training samples automatically from in-vehicle camera images. These approaches are effective to construct an accurate detector with less manual intervention. However, these are methods for obtaining positive training samples only, such as pedestrians and traffic signs. Therefore, these methods still require a large number of negative samples, that is, background images. To construct an accurate traffic sign detector, the negative samples should have various appearances, which however, is difficult and time consuming to obtain manually. Therefore, this paper proposes a method for constructing an accurate traffic sign detector automatically by obtaining positive and negative samples simultaneously from in-vehicle camera images. This paper focuses on regulatory signs with a red ring surrounding them, such as those in Fig. 2, because they are very common and important when driving a vehicle in Japan.

Section II describes the details of the proposed method. Then, section III shows the experimental setup and results using in-vehicle camera images. The results are discussed in section IV. Finally, we will conclude this paper in section V.
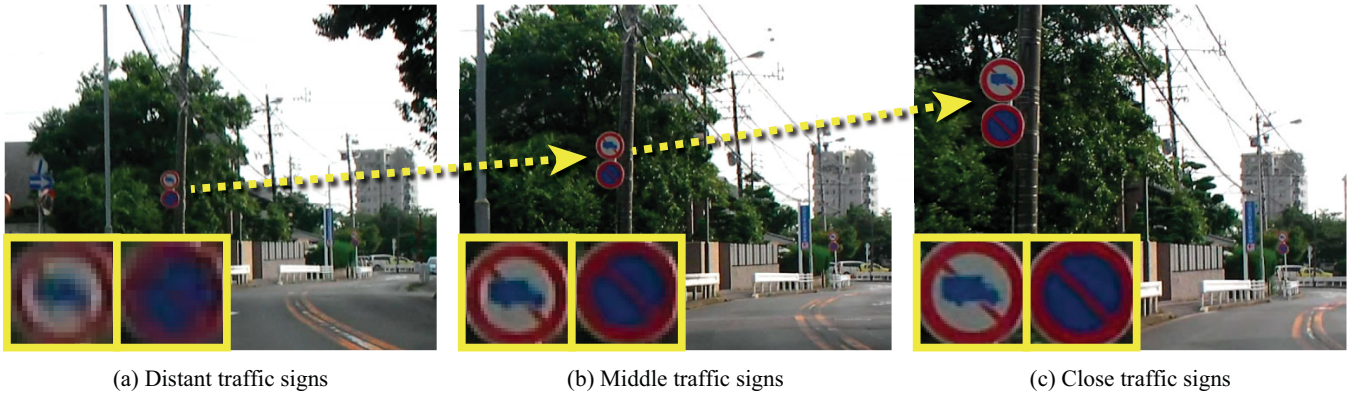
(a) Distant traffic signs   (b) Middle traffic signs   (c) Close traffic signs

Fig. 3. Appearances observed at distant, middle and close traffic signs from a vehicle.

Fig. 2. Target traffic signs.

II. METHOD

This paper proposes a method for constructing a highly accurate traffic sign detector automatically by gathering positive and negative training samples simultaneously from in-vehicle camera images. To construct an accurate traffic sign detector, various traffic sign images (positive samples) and background images (negative samples) should be prepared, and they must include various appearances observed in a real environment, such as size, illumination, and texture changes. However, it is very difficult to obtain such samples comprehensively. Also, since backgrounds vary widely, gathering of those images is an especially difficult problem. To solve these problems, by extending the idea we proposed in [6], this paper proposes a method for constructing an accurate traffic sign detector by gathering traffic sign images and background images simultaneously from in-vehicle camera images.

A. Basic idea

As shown in Fig. 3(a), it is difficult to detect and segment small traffic sign images (distant traffic signs) accurately and automatically. On the other hand, large traffic sign images (close traffic signs) can be segmented accurately (Fig. 3(c)), and it is comparatively stable to detect them automatically. Also, appearances of traffic signs change gradually in successive frames. Based on these characteristics, the proposed method obtains traffic sign images and background images simultaneously, under the following assumptions:

(i) Large traffic signs can be detected stably in successive frames.

(ii) Large traffic signs can be detected even if the position of the detection window changes slightly.

(iii) If the position of a large traffic sign is given, it is easy to track back smaller traffic signs retrospectively from it.

From the assumptions (ii) and (iii), traffic sign images can be obtained automatically by applying two steps: (1) find large traffic signs (high resolution), and (2) retrospectively track from a large traffic sign to smaller ones.

On the other hand, from the assumptions (i) and (ii), background images can be obtained as a large image that is not detected in successive frames nor in adjacent detection windows. Therefore, the proposed method defines a likelihood of a background by measuring the overlap-ratio of the detected traffic signs temporally and spatially. From the fact that these are false alarms obtained by the detector, it is expected that they should become good negative samples for training to reduce false alarms of the detector we train it again. Therefore, the proposed method uses them as negative samples for re-training the detector.

Finally, the proposed method constructs a traffic sign detector by using traffic sign images and background images obtained automatically as positive and negative training samples, respectively.

Fig. 4 shows the flowchart of our proposed method. The proposed method consists of three parts: (1) retrospective gathering of traffic sign images from in-vehicle camera images, (2) gathering of background images by evaluating temporal and spatial overlap-ratio of the detected traffic signs, and (3) construction of a traffic sign detector by using both traffic sign images and background images. By repeating this process, the proposed method improves the accuracy of the traffic sign detector iteratively. The following sections describe details of these two parts.

B. Retrospective gathering of traffic sign images

The proposed method gathers traffic sign images as positive samples for training the traffic sign detector. This step consists of three parts: (1) detection of traffic sign candidates, (2) detection of a large traffic sign, and (3) retrospective tracking of traffic signs. In the proposed method, a nested
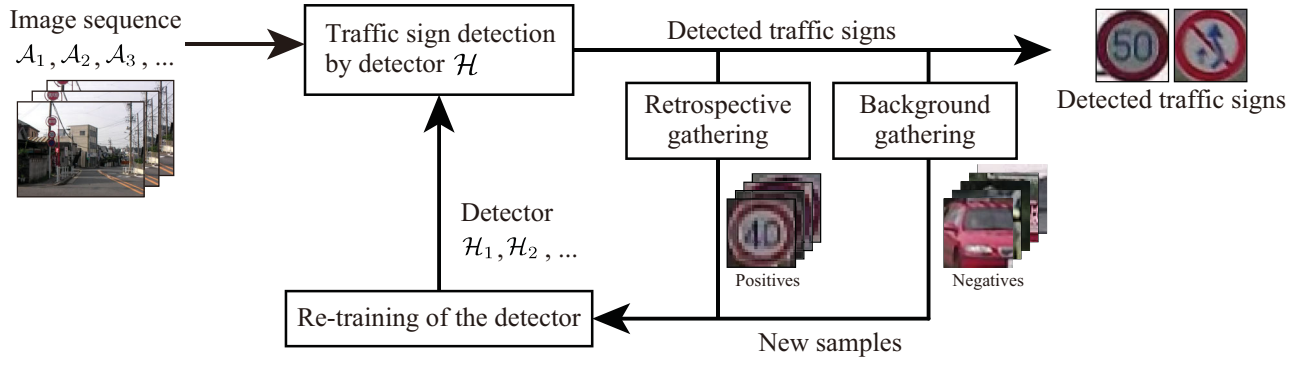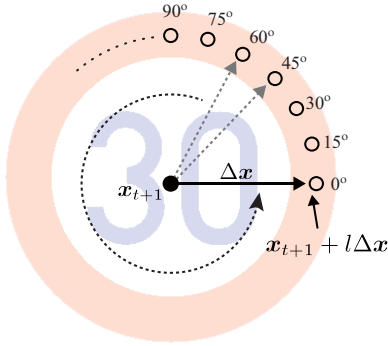
Fig. 4. Flowchart of the proposed method.

Fig. 5. Edge detection of a traffic sign.

cascade of a Real AdaBoost classifier is used to detect traffic signs [10], [11]. The details of the detector are explained in section II-D. The proposed method uses this detector for obtaining large traffic signs. Then, small traffic signs are obtained automatically by retrospective tracking. The following sections describe details of these steps.

1) Detection of traffic sign candidates: This step searches traffic sign candidates from an in-vehicle camera image. First of all, traffic sign candidates are searched by scanning detection windows inside the entire region of an in-vehicle camera image. Each detection window is evaluated by using a previously constructed traffic sign detector $\mathcal{H}_k$, and decided whether it is a traffic sign or not. Then, the detected traffic sign candidates are merged by Mean shift clustering [7].

2) Selection of a large traffic sign: From the assumption (ii), multiple traffic sign candidates can be obtained around a large traffic sign. Based on this characteristic, the proposed method removes false positives (i.e. not a traffic sign) by evaluating the number of merged candidates. Finally, large traffic signs are obtained by selecting large size detected candidates, and these are used as initial positions of the traffic sign in the retrospective tracking.

3) Retrospective tracking of traffic signs: To obtain smaller traffic sign images, this step tracks back in the image sequence from the position of the large traffic sign detected in the previous step. This process is formulated as a process of finding the position and the size of the t-th traffic sign by

using those of the $(t + 1)$-th one.

Before tracking traffic signs, the proposed method enhances red components of an input image by calculating the red value of normalized color (Eq.(3)), since our target traffic signs have red edges as shown in Fig. 2. Then, we obtain an image $\mathbf{F}_t$ by applying a Gaussian filter to the image, and use this image $\mathbf{F}_t$ for tracking smaller traffic signs.

First, by increasing l, the proposed method finds edge pixels of the t-th traffic sign as a pixel that satisfies

$$\nabla \mathbf{F}_t(\boldsymbol{x}_{t+1} + l\Delta\boldsymbol{x}) \cdot \Delta\boldsymbol{x} < 0, \tag{1}$$

where $\nabla \mathbf{F}_t(\boldsymbol{x})$ is the gradient of the image intensity at position $\boldsymbol{x}$, and "·" is the inner product of vectors. As shown in Fig. 5, this search process is repeated with rotating the search direction $\Delta\boldsymbol{x}$. Finally, the proposed method computes the position and the size of the traffic sign at time t by fitting a circle to those edge pixels [8]. Here, RANSAC approach is employed to avoid the outliers of inappropriate edge pixels.

By repeating these steps as $t \leftarrow t-1$, the proposed method tracks back smaller traffic signs in the image sequence.

C. Gathering of background images

This step gathers background images as negative samples for training the traffic sign detector. From the assumptions (i) and (ii), large traffic signs can be found as a large image that is detected both in successive frames and in adjacent detection windows. That is, a large image, not detected in successive frames nor in adjacent detection windows, can be considered as a background. Based on this idea, the proposed method obtains background images by evaluating the following criteria.

First, the proposed method applies the step II-B.1) to an in-vehicle camera image sequence, and obtains traffic sign candidates. Then, the proposed method evaluates two types of overlapping criteria N_1 and N_2. N_1 is related to the assumption (i), and it is calculated as the number of overlapped candidates between successive frames. Meanwhile, N_2 corresponds to the assumption (ii), and this is calculated by the number of merged candidates in the step II-B.1). By using N_1 and N_2, background images are obtained as a candidate that satisfies

$$\frac{1}{1 + e^{-N_1}} \times \frac{1}{1 + e^{-N_2}} < T_1. \tag{2}$$

(a) Input	(b) Gray	(c) Red	(d) Green
(e) Blue	(f) Eq.(3)	(g) Eq.(4)	(h) Eq.(5)

Fig. 6.   Examples of color feature images for computing LRP features.

D. Construction of a traffic sign detector

In this paper, a nested cascade of a Real AdaBoost classifier [10], [11] is used as the traffic sign detector. The classifier is trained by using positive and negative samples obtained in the steps II-B and II-C, respectively. LRP (Local Rank Pattern) features [9] are used for training the Real AdaBoost classifier. Here, as used in the previous works [2], [3], we use seven types of color feature images for computing LRP features. Fig. 6 shows examples of these color feature images. Pixel values of these images are calculated as gray scale value (f_1), RGB values ($f_2 \sim f_4$), and normalized RGB values ($f_5 \sim f_7$). Here,

$$f_5(\boldsymbol{x}) = \frac{r(\boldsymbol{x})}{r(\boldsymbol{x}) + g(\boldsymbol{x}) + b(\boldsymbol{x})}, \qquad (3)$$

$$f_6(\boldsymbol{x}) = \frac{g(\boldsymbol{x})}{r(\boldsymbol{x}) + g(\boldsymbol{x}) + b(\boldsymbol{x})}, \qquad (4)$$

$$f_7(\boldsymbol{x}) = \frac{b(\boldsymbol{x})}{r(\boldsymbol{x}) + g(\boldsymbol{x}) + b(\boldsymbol{x})}, \qquad (5)$$

where $r(\boldsymbol{x})$, $g(\boldsymbol{x})$ and $b(\boldsymbol{x})$ represent red, green and blue values at a pixel $\boldsymbol{x}$, respectively.

III. EXPERIMENT

To evaluate the accuracy of the proposed method, we used in-vehicle camera images. This section shows details of the data, experimental setup, and evaluation and results.

A. Data

We captured in-vehicle camera images by using SANYO Xacti DMX-HD2 mounted on the windshield. The image was captured in the size of 640×480 pixels (30 fps). The captured images were divided into five image sequences ($\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{A}_4$), and the total number of images were 3,907. These images were used as inputs of the proposed method. On the other hand, we prepared another 2,967 images containing at least one traffic sign between 15×15 pixels and 45×45 pixels, and these images were used for evaluation. Here, this evaluation data contained 4,886 traffic signs. Also, we prepared M (= 1 or 10) images (640×480 pixels) containing no traffic sign, and these images were used for constructing an initial traffic sign detector $\mathcal{H}_0$.

TABLE I
DETECTION RATE OF THE ITERATIVELY CONSTRUCTED DETECTORS
$\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_4$ BY THE PROPOSED METHOD ($M = 1$).

Detector	Precision	Recall	F-measure
$\mathcal{H}_0$	0.139	0.917	0.242
$\mathcal{H}_1$	0.447	0.868	0.590
$\mathcal{H}_2$	0.745	0.903	0.816
$\mathcal{H}_3$	0.848	0.904	0.875
$\mathcal{H}_4$	0.856	0.901	0.878

B. Experimental setup

In this experiment, five traffic sign detectors were constructed one by one according to the following steps. At first, thirteen large traffic signs were manually selected from the image sequence $\mathcal{A}_0$, and 500 traffic sign images were simply generated by changing their clipping positions and rotation angles. Then, an initial traffic sign detector $\mathcal{H}_0$ was constructed by using these images. Here, negative samples were obtained from M in-vehicle camera images containing no traffic sign by changing their clipping positions randomly, and 5,000 negative samples were used for training classifiers in each stage of the cascade. Next, according to the processes described in sections II-B and II-C, traffic sign images and background images were gathered simultaneously from the image sequence $\mathcal{A}_1$ by using the constructed detector $\mathcal{H}_0$. Then, images used in $\mathcal{H}_0$ and images obtained in the above step were used for constructing a second traffic sign detector $\mathcal{H}_1$. $\mathcal{H}_2$, $\mathcal{H}_3$, and $\mathcal{H}_4$ were also constructed iteratively by the same manner as the above.

C. Evaluation

To evaluate the simultaneous gathering of training samples proposed in this paper, we compared the proposed method with the previous method [6]. We performed this evaluation in two conditions by changing the number of initial in-vehicle camera images used for obtaining negative samples. The first condition used only one in-vehicle camera image ($M = 1$), and the second one increased it to ten images ($M = 10$). We used $T_1 = 0.4$ for gathering background images described in section II-C. The experiment shown in section III-B was performed ten times, and we evaluated the accuracy of the constructed detectors in precision and recall rates with corresponding F-measures.

Table I shows the results of the iteratively constructed detectors $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_4$ of the proposed method in precision and recall rates with corresponding F-measures ($M = 1$). Here, F-measure was calculated as

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \qquad (6)$$

Fig. 7 shows the comparison of the performance between the proposed method and the previous method in precision, recall rates and F-measures. Fig. 8 shows examples of the detected traffic signs by the proposed method and the previous method.

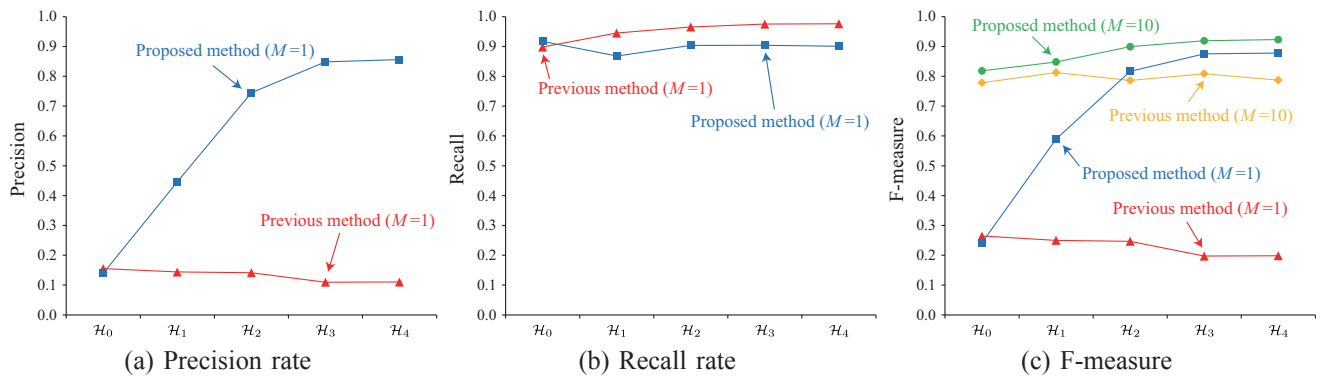(a) Precision rate   (b) Recall rate   (c) F-measure

Fig. 7. Results of detectors $\mathcal{H}_0 \sim \mathcal{H}_4$ iteratively constructed by the proposed method and the previous method in precision, recall rates and F-measure. Here, M is the number of in-vehicle camera images not containing traffic signs that were input initially for obtaining negative samples.

(a) Detection results of the proposed method.
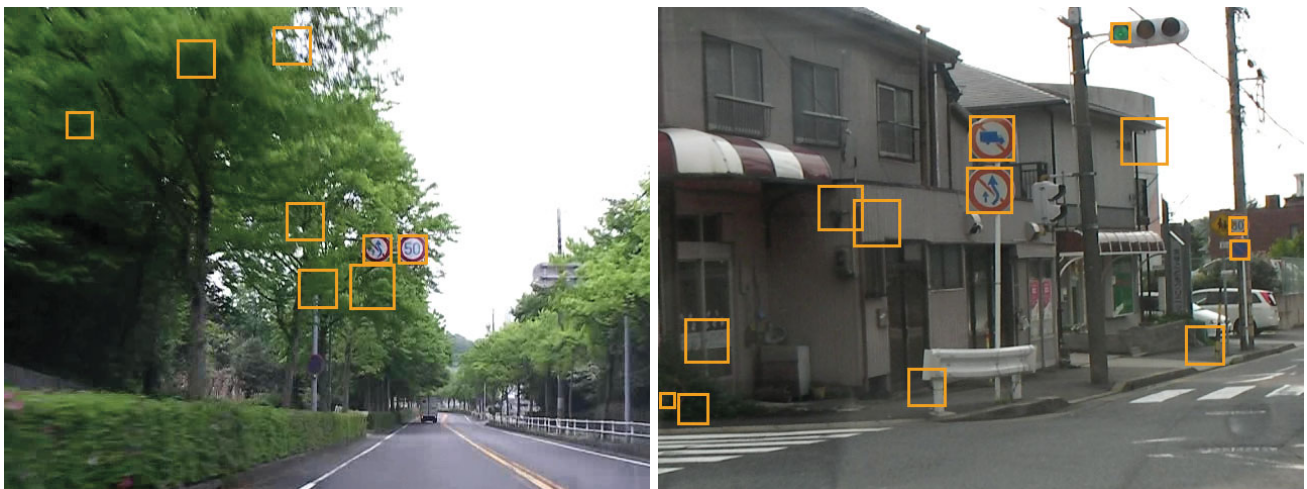
(b) Detection results of the previous method [6].

Fig. 8. Examples of detection results using the detector $\mathcal{H}_4$ constructed by the proposed method and the previous method. (a) shows results of the proposed method, and (b) shows results of the previous method. The previous method does not gather background images as negative samples for training the detector. Therefore, many mis-detections occur in comparison with the results of the proposed method.

IV. DISCUSSIONS

If we want to make use of a traffic sign detector in a real road environment, it is required that both precision and recall of the detector should be high. That is, F-measure calculated by Eq.(6) should be high. From this point of view, as seen in Fig.7(c), the proposed method could achieve high F-measure by increasing the number of image sequences. Especially, since the proposed method retrains the detector by using background images that were easily mis-detected by a previously constructed detector, the proposed method could obtain F-measure of 0.878 at maximum. In contrast, the previous method could not improve the F-measure even if the number of image sequences increased. Since the previous method does not employ the mechanism of background image gathering, it obtains negative samples from only initial in-vehicle camera images containing no traffic sign, to train the detector. Consequently, the previous method could not train enough variations of backgrounds, especially in the case of $M = 1$. Although the previous method could keep high recall rate by using retrospective gathering for obtaining traffic sign images of positive samples, its precision did not improve. Therefore, it is confirmed that the proposed method could improve the accuracy by obtaining both traffic sign images and background images simultaneously.

We compared the accuracy of the detectors when increasing the number of initial in-vehicle camera images not containing traffic signs. Fig.7(c) shows results of the proposed method and the previous method in F-measure when using one or ten in-vehicle camera images for initial inputs. As seen in Fig.7(c), the proposed method achieved higher F-measure than the previous method, even if we used only one in-vehicle camera image not containing traffic signs as an initial input ($M = 1$). Since the proposed method automatically obtains negative samples from in-vehicle camera image sequences, the proposed method can obtain various negative samples that are not included in the initial training samples, as shown in Fig. 9. Therefore, the proposed method could improve the accuracy of the traffic sign detector iteratively by taking in-vehicle camera image sequences. Since the traffic sign detector is usually used in an unknown environment, the proposed method will become a strong tool for adapting the detector to that environment. However, the proposed method sometimes gathered inappropriate background images, such as images partially containing a traffic sign. This degrades the accuracy of the traffic sign detector. We intend to improve the mechanism of background gathering for avoiding the above problem in our future work.

V. CONCLUSIONS

This paper proposed an intelligent traffic sign detector using adaptive learning based on online gathering of various traffic sign images and background images. The proposed method applies retrospective tracking for obtaining smaller traffic sign images. At the same time, background images are obtained by evaluating temporal and spatial overlap-ratio of detected traffic signs. Finally, the proposed method constructs a traffic sign detector by using them. We evaluated

Fig. 9.   Examples of background images obtained by the proposed method.

the accuracy and the effectiveness of the proposed method by applying it to actual in-vehicle camera images. Experimental results showed that the proposed method could improve the accuracy of the traffic sign detector satisfactorily, even if we use only one in-vehicle camera image not containing traffic signs as an initial input for obtaining negative samples. Future works include: (i) improvement of the retrospective tracking especially for small traffic signs, (ii) improvement of the mechanism for gathering background images, and (iii) evaluation by applying the method to a larger dataset.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jiménez, H. Gómez-Moreno, and F. López-Ferreras, "Road-sign detection and recognition based on support vector machines," IEEE Transactions on Intelligent Transportation Systems, Vol.8, No.2, 2007, pp.264–278.

[2] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," Proceedings of IEEE Intelligent Vehicles Symposium, 2005, pp.255–260.

[3] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "Construction of cascaded traffic sign detector using generative learning," Proceedings of International Conference on Innovative Computing, Information and Control, 2009, ICICIC-2009-1362.

[4] P. Viola and M. Jones, "Robust real-time face detection," International Journal of Computer Vision, Vol.57, No.2, 2004, pp.137–154.

[5] 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, 2002, pp.1263–1270.

[6] D. Deguchi, K. Doman, I. Ide, and H. Murase, "Improvement of a traffic sign detector by retrospective gathering of training samples from in-vehicle camera image sequences," Proceedings of ACCV2010 Workshop on Computer Vision in Vehicle Technology: From Earth to Mars, 2010, pp.1–10.

[7] Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17, No.8, 2005, pp.790–799.

[8] I. D. Coope, "Circle fitting by linear and nonlinear least squares," Journal of Optimization Theory and Applications, Vol.76, No.2, 1993, pp.381–388.

[9] M. Hradis, A. Herout, and P. Zemcik, "Local rank patterns — novel features for rapid object detection," Proceedings of the International Conference on Computer Vision and Graphics, Lecture Notes in Computer Science, Vol.5337, 2008, pp.239–248.

[10] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," Machine Learning, Vol.37, No.3, 1999, pp.297–336.

[11] C. Huang, H. Ai, B. Wu, and S. Lao, "Boosting nested cascade detector for multi-view face detection," Proceedings of the International Conference on Pattern Recognition, Vol.2, 2004, pp.415–418.