Improvement of a traffic sign detector by retrospective gathering of training samples from in-vehicle camera image sequences

Daisuke Deguchi, Keisuke Doman, Ichiro Ide and Hiroshi Murase

Graduate School of Information Science, Nagoya University
Furo-cho Chikusa-ku Nagoya, Aichi 464–8601, Japan

Abstract. This paper proposes a method for constructing an accurate traffic sign detector by retrospectively obtaining training samples from in-vehicle camera image sequences. To detect distant traffic signs from in-vehicle camera images, training samples of distant traffic signs are needed. However, since their sizes are too small, it is difficult to obtain them either automatically or manually. When driving a vehicle in a real environment, the distance between a traffic sign and the vehicle shortens gradually, and proportionally, the size of the traffic sign becomes larger. A large traffic sign is comparatively easy to detect automatically. Therefore, the proposed method automatically detects a large traffic sign, and then small traffic signs (distant traffic signs) are obtained by retrospectively tracking it back in the image sequence. By also using the retrospectively obtained traffic sign images as training samples, the proposed method constructs an accurate traffic sign detector automatically. From experiments using in-vehicle camera images, we confirmed that the proposed method could construct an accurate traffic sign detector.

1 Introduction

In recent years, ITS (Intelligent Transport Systems) technologies have become widely available in our driving environment. In particular, understanding of the road environment in ITS is one of the most important technologies for a safe driving assistance system. Since traffic sign detection and recognition are key components for understanding the road environment, several methods have been proposed [1–4]. Bahlmann et al. proposed a method for detecting traffic signs from in-vehicle camera images [3]. They employed a cascaded AdaBoost classifier [5] for rapid detection, and color Haar-like feature is used for improving the accuracy of the detection. Although their method is accurate and fast enough, it requires a tremendous number of traffic sign images for training the AdaBoost classifier. Doman et al. solved this problem by generating training samples according to image degradation models [4]. Although this method can generate numerous training samples, it is still difficult to generate various appearances actually observed in the real environment as shown in Fig. 1. For constructing a traffic sign detector easily and accurately, it is necessary to obtain a large number of training samples from real environment without manual intervention. Also, if a traffic sign detector is constructed before applying it to an unknown environment, it is required to reconstruct the detector by using new training samples obtained in the environment. Wöhler tried to solve these problems by constructing a pedestrian detector by obtaining training samples automatically from in-vehicle camera images [6]. In this method, pedestrians were detected by using a previously constructed detector, and training samples were obtained by tracking them forward in the time space. However, to exclude false positives from training samples, this method requires that an initial detector should be relatively accurate. Therefore, it still requires a large number of training samples for constructing the initial detector. To solve this problem, this paper introduces knowledge about appearance changes of traffic signs when driving a vehicle.

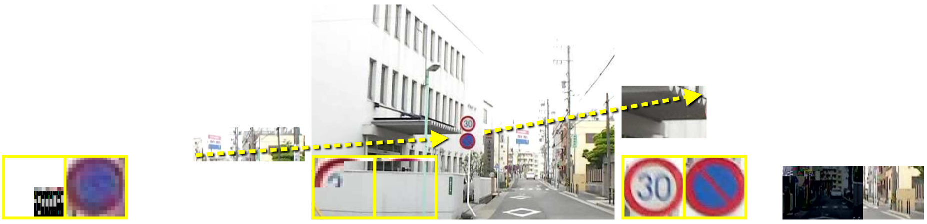
Training samples of distant traffic signs are required for constructing an accurate traffic sign detector that can detect distant traffic signs from in-vehicle camera images. However, since their sizes are too small in in-vehicle camera images, it is difficult to obtain them either automatically or manually. When driving a vehicle in a real environment, the distance between a traffic sign and the vehicle shortens gradually, and proportionally, the size of the traffic sign becomes larger. Therefore, if we can know the position of the large traffic sign, small traffic signs (distant traffic signs) can be obtained by tracking it back in the image sequence. Based on this idea, the proposed method greatly reduces the number of initial training samples, and then constructs an accurate traffic sign detector by gathering training samples retrospectively from in-vehicle camera image sequences. To use the traffic sign detector in a real environment, not only precision but also recall of the detector should be high. Therefore, the aim of the work presented in this paper is to construct a traffic sign detector having a high F-measure.

Section 2 describes the details of the proposed method. Then, experiments using in-vehicle camera images are shown in section 3. We discuss the results in section 4. Finally, we will conclude this paper in section 5.

2 Method

This paper proposes a method for constructing an accurate traffic sign detector by gathering training samples retrospectively from in-vehicle camera images. To construct an accurate traffic sign detector, traffic sign images for training should be gathered in various sizes from small (low resolution) through to large (high resolution). However, as shown in Fig. 2(a), it is difficult and time consuming to obtain numerous small traffic sign images (distant traffic signs) segmented accurately, since their sizes are small. On the other hand, large traffic sign images (close traffic signs) shown in Fig. 2(c) can be segmented accurately, and it is comparatively easy to recognize them automatically. Also, if the position of a large traffic sign is obtained, it is easy to track small traffic signs from it. Therefore, based on these ideas, the proposed method employs two strategies for gathering various traffic sign images: (1) find large traffic signs (high resolution),

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



(a) Distant traffic signs (b) Middle traffic signs (c) Close traffic signs

Fig. 2. Appearances observed at distant, middle and close traffic signs from a vehicle.

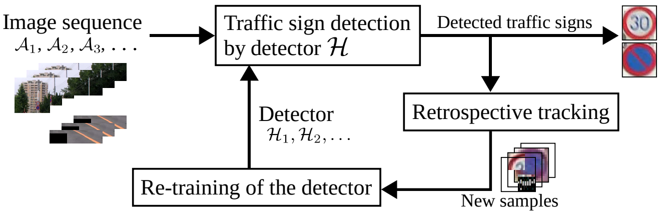
and (2) retrospective tracking from a large traffic sign to a small one. Then, the proposed method constructs a traffic sign detector by using samples obtained automatically. Figure 3 shows very common and important traffic signs when driving a vehicle in Japan. Therefore, we consider these traffic signs as our targets in this paper.

The proposed method consists of two parts: (1) retrospective gathering of traffic sign images from in-vehicle camera images, and (2) construction of a traffic sign detector by using them. The following sections describe details of these two parts.

2.1 Retrospective gathering of traffic sign images

Figure 4 shows a flowchart of our proposed method. The proposed method employs a nested cascade of a Real AdaBoost classifier for the detection of large traffic signs [11, 12]. Then, retrospective tracking is used for gathering small traffic sign images automatically. The following sections describe details of these steps.

Fig. 3. Target traffic signs.



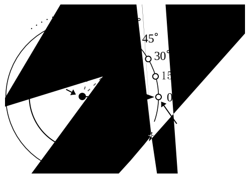


Fig. 4. Flowchart of the proposed method.

(a) (b)

Fig. 8. Examples of detection results by the proposed method (LRP). (a) there is an object similar to the target traffic signs, which is located above the traffic signs but correctly not detected, and (b) although a traffic sign is occluded by a pole, the proposed method succeeded to detect it.

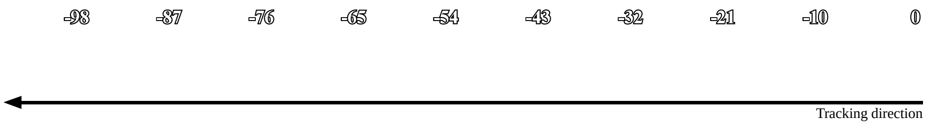
proposed method (LRP), the proposed method (HAAR), and the conventional method in precision, recall and F-measure. Examples of the detection results by the proposed method (LRP) are shown in Fig. 8.

When using Intel Xeon W5590 3.33 GHz 2, the finally constructed detector required 0.122 sec. (8.2 fps) in average for detecting traffic signs from an image. This means that the proposed method can detect traffic signs every 2 meters when the vehicle moves at 60 km/h.

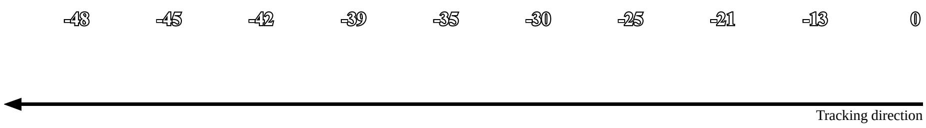
4 Discussions

As mentioned earlier, both precision and recall of a constructed traffic sign detector should be high. That is, it is required that the constructed detector should have high F-measure reflecting both precision and recall. From this point of view, as can be seen from Table 1, the proposed method could construct an accurate traffic sign detector (0.955 in F-measure) automatically by obtaining various traffic sign images from only thirteen large traffic sign images inputted manually. The accuracy of the constructed detector gradually improved by applying the proposed method iteratively. Also, as shown in Fig. 7, this can be observed from the comparison of the proposed method and the conventional method. Although the precision of the proposed method slightly degrades compared to that of the conventional method, the proposed method could obtain much higher recall rate. Therefore, F-measure was greatly improved by the proposed method. From these results, since only a small number of training samples is required as an input for the proposed method, this can greatly reduce the cost for constructing a detector. Therefore, the proposed method will be quite useful for improving the accuracy of a traffic sign detector without manual intervention.

To evaluate the effectiveness of the LRP features, we compared LRP features and Haar-like features in precision, recall, and F-measure, shown in Fig. 7.



(a) a contrast of the traffic sign is relatively high



(b) a part of traffic sign is occluded by leaves

Fig. 9. Results of retrospective tracking of a traffic sign. Relative frame number is shown at the top right of each image.

To construct an accurate traffic sign detector, training samples obtained by the method must be labeled correctly. In the case of the method using Haar-like features, some false positives are included in the training samples obtained automatically by the proposed method. Therefore, the precision of the constructed detector gradually decreased. On the other hand, in the case of using LRP features, since few false positives are included in the obtained training samples, the precision of the proposed method (LRP) is much higher than the proposed method (HAAR). However, the proposed method (LRP) still gathered a small number of false positives for training samples. We intend to improve the performance of automatic gathering of training samples in our future work.

Figure 9 shows examples of retrospective tracking of traffic signs proposed in this paper. As shown in Fig. 9(a), it can be confirmed that the proposed method could obtain traffic sign images in various resolutions from low to high. Although a part of a traffic sign in Fig. 9(b) is occluded by leaves, some edges of the traffic sign can still be observed. Since these edges were extracted, the proposed method was able to track it correctly. However, the method failed to track traffic signs when their resolution was too poor. We intend to deal with this problem in our future work.

5 Conclusions

This paper proposed a method for constructing an accurate traffic sign detector by automatic gathering of various traffic sign images based on retrospective tracking. First, the proposed method detects large (high resolution) traffic signs from in-vehicle camera images. Then, retrospective tracking is applied for obtaining small traffic sign images. By applying these steps, the proposed method allows us to automatically gather real traffic sign images in various sizes from a small one to a large one. Finally, a traffic sign detector is constructed by using the gathered traffic sign images. We evaluated 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. Future works include: (i) improvement of the tracking of small traffic signs, (ii) evaluation by applying the method to many more cases.

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 JST CREST. MIST library (http://mist.murase.m.is.nagoya-u.ac.jp/) was used for developing the proposed method.

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 **8(2)** (2007) 264–278.
2. G. Loy and N. Barnes: Fast shape-based road sign detection for a driver assistance system. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems **1** (2004) 70–75
3. 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) 255–260.
4. 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.
5. P. Viola and M. Jones: Robust real-time face detection, International Journal of Computer Vision **57(2)** (2004) 137–154.
6. C. Wöhler: Autonomous in situ training of classification modules in real-time vision systems and its application to pedestrian recognition, Pattern Recognition Letters **23(11)** (2002) 1263–1270.
7. Y. Cheng: Mean shift, mode seeking, and clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence **17(8)** (2005) 790–799.
8. I. D. Coope: Circle fitting by linear and nonlinear least squares, Journal of Optimization Theory and Applications **76(2)** (1993) 381–388.
9. G. J. Burghouts and J. M. Geusebroek: Performance evaluation of local colour invariants, Computer Vision and Image Understanding **113(1)** (2009) 48–62.
10. 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 **5337** (2008) 239–248.
11. R. E. Schapire and Y. Singer: Improved boosting algorithms using confidence-rated predictions, Machine Learning **37(3)** (1999) 297–336.
12. 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 **2** (2004) 415–418.