MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

On-road Obstacle Detection by Comparing Present and Past In-vehicle Camera Images

Haruya Kyutoku†, Daisuke Deguchi†, Tomokazu Takahashi‡,
Yoshito Mekada§, Ichiro Ide†, Hiroshi Murase†

† Graduate School of Information Science, Nagoya University, Japan
*) E-mail: kyutoku@murase.m.is.nagoya-u.ac.jp
‡ Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan
§ School of Information Science & Technology, Chukyo University, Japan

Abstract

We propose a method for detecting general obstacles on a road by subtracting present and past in-vehicle camera images. Compared to the existing learning-based methods that could detect only specific obstacles, the proposed method based on image-subtraction could detect any kind of obstacles. To achieve this, the proposed method first realizes a frame-by-frame correspondence between the present and the past in-vehicle camera image sequences, then performs a road surface registration between the corresponded frames. Obstacles are detected by using the difference of the road surface regions. To demonstrate the effectiveness of the proposed method, experiments were conducted using several image sequences captured by an actual in-vehicle camera. The experimental results showed that the proposed method could detect general obstacles accurately at a distance enough to safely avoid them.

1 Introduction

Research and commercialization of driving support technologies have become active research topics. We especially focus on a technique to detect forward obstacles from in-vehicle camera images. This technique is essential to realize a collision-warning system that could significantly reduce the number of traffic accidents. Accordingly, a number of technologies to detect forward obstacles have been developed. However, most of them could detect only specific obstacles such as pedestrians and cars. Detecting general obstacles that cannot be learned beforehand is a challenging task because in reality various obstacles exist on a road.

Recently, in-vehicle camera image database[1] is becoming popular. In addition, the speeding up of wireless telecommunications and the growth of storage capacity are remarkable. This would allow a system to collect and store in-vehicle camera images taken in the past, and use the information for driving supports.

In light of the above background, we propose a method for detecting general forward obstacles based on an image subtraction technique between present and past in-vehicle camera images that are captured at the same location.

This paper is organized as follows: Section 2 describes related work. Section 3 details the proposed method. Experimental results are shown and discussed in Section 4. Section 5 summarizes this paper.

2 Related Work

Various techniques for detecting obstacles on the road have been proposed such as [1]. They use various sensors including millimeter-wave radar, stereo or monocular camera, or infra-red camera. Currently, the millimeter-wave radars are installed only on expensive cars, and its spatial resolution is too low to detect a small obstacle. The stereo cameras still have some difficulties regarding their calibration [2] and the detection of point correspondences between images with satisfactory accuracy. The technologies using an infra-red camera are specialized only for pedestrian detection, which assists the visibility in night-time driving.

Researches with a monocular camera have been actively attempted. However, most of them detect only specific obstacles such as pedestrians and cars using an object learning-based method [3][4]. The proposed method also uses a monocular camera, but aims to detect general obstacles by comparing present and past in-vehicle camera images. Thus, there is no need to learn image features of the obstacles beforehand.

The proposed method needs a past in-vehicle camera image sequence including no obstacles that were captured along the road that the vehicle currently runs. There have been some techniques [5] for construction and updating a street image database that contains in-vehicle camera images corresponding to each road on a street map. We could obtain the past in-vehicle camera image sequence from such databases. Meanwhile, methods have been developed to remove obstacles from a street image database [6]. These techniques could be used to synthesize a past in-vehicle camera image sequence with no obstacles.

3 Detection of General Obstacles on a Road by Subtraction with Past In-Vehicle Camera Images

The obstacles are detected by using the present in-vehicle camera image sequence $F = \{f_t\}$ and the past one $G = \{g_t\}$. Here, $f_t(m, n)$ and $g_t(m, n)$ represent the pixel values at coordinates (m, n) of the t-th frame in each sequence. The differences of running speeds and running positions exist between F and G. Therefore, to detect an obstacle based on the subtraction of two images, the proposed method performs the following three processes:

1. Finding a corresponding frame

   For each t_i-th frame in the present sequence, find the corresponding frame $g_{t'_i}$ that was captured at the nearest position to that of frame f_{t_i}. Here, t_i is the target frame number for obstacle detection.

2. Registration of road surfaces

   For each pixel (x, y) in the road surface region in f_{t_i}, find the same location (x', y') in $g_{t'_i}$.

[1] Google Street View, `http://maps.google.com/`

Figure 1. Corresponding frames in the image sequences $\{f_t\}$ and $\{g_t\}$.

Figure 2. Direct subtraction of frames.

Figure 3. Subtraction after registration.

3. Obstacle detection based on image subtraction of road surfaces

   Calculate the difference between $f_{t_i}(x, y)$ and $g_{t'_i}(x', y')$, then detect regions with high difference as obstacles.

As for Step 2, it is difficult to detect obstacles by directly applying subtraction between $f_{t_i}(x, y)$ and $g_{t'_i}(x, y)$ because of the difference of the running positions between present and past sequences. Figure 1 shows the corresponding frames obtained by Step 1. The result of the direct subtraction of them is shown in Fig. 2, where we can see a spatial gap. To absorb the gap, the proposed method performs Step 2. The subtraction result after Step 2 is shown in Fig. 3.

3.1 Finding a corresponding frame

The proposed method needs to find frame $g_{t'_i}$ in G which was captured at the nearest location to that of frame f_{t_i}. To achive this, we apply a Dynamic Time Warping (DTW) method between image sequences F and G. This absorbs the difference of running speeds between the image sequences so that each corresponding pair of frames is captured at almost the same location. As a penalty for the frame correspondence used in the DTW method, we define a novel penalty measurement of a positional relationship between present and past cameras using epipolar geometry.

When camera directions are almost the same, the position of an epipole in an image is strongly related to the position between two cameras as shown in Fig. 4. The position of an epipole goes away from the vanishing point when two cameras become closer. Figure 5 shows the positions of epipoles in actual images.

Based on the above analysis, the proposed method uses the reciprocal of the distance between the epipole (e_x, e_y) and the vanishing point (v_x, v_y) as the penalty. The penalty $p(i, j)$ for a correspondence between f_i and g_j is defined as

$$p(i, j) = \frac{1}{|e_x - v_x| + \alpha}, \quad (1)$$

where α is a positive constant. The correspondences between the present and the past in-vehicle camera images are obtained by the DTW method.

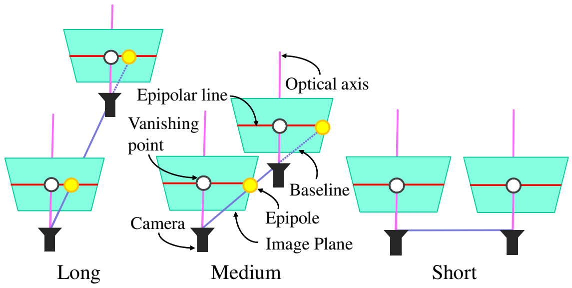


Figure 4. Relation between the positions of two cameras and epipoles.

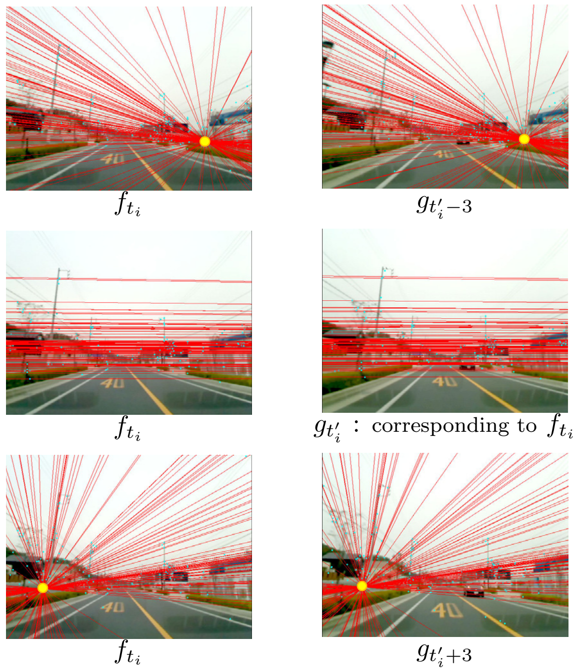


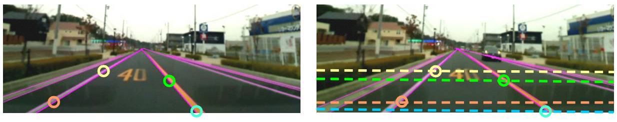
Figure 5. The position of epipoles in actual images (Line: Epipolar-Line, Circle: Epipole).

The position of the epipole is calculated from a fundamental-matrix obtained by the RANSAC algorithm [7]. The corresponding points between present and past frames are obtained by using the SIFT feature [8]. Assuming that the camera is directed to the traveling direction of the vehicle, the proposed method considers the center of the image as the vanishing point.

3.2 Registration of road surfaces

The proposed method performs registration of road surface regions between frames f_{t_i} and $g_{t'_i}$. It is assumed that the road surface is flat, and the obstacle detection area is restricted to the road surface region only. The proposed method makes correspondence of pixels in the road surface regions between the present and the past images by a projective transformation.

To obtain the projective transformation matrix, it needs four corresponding point pairs on the road surfaces between two frames. However, it is difficult to detect them with sufficient accuracy, because a road surface is usually texturless. Therefore, using the property that a corresponding point exists on its epipolar line in another frame, the proposed method determines the corresponding points as intersections of the epipolar lines with the road boundary lines. Figure 6 shows the determination of the corresponding points. The position of road boundary lines are obtained by using the Hough transform method. First, four points



(a) Present in-vehicle camera image (b) Past in-vehicle camera image

Figure 6. Determination of the correspondence points on the road surface.

are selected from the road boundary lines. Then, for each point, the epipolar line in another frame is obtained by using the fundamental-matrix obtained in Section 3.1. Finally, the intersections of corresponding road boundary lines with epipolar lines are determined as the corresponding points.

Figure 3 shows the result of the subtraction of the frames shown in Fig. 1 after the road surface registration. We can see that the spatial gap in the road surface becomes smaller than that in Fig. 2.

3.3 Detection of obstacles by subtraction of road surfaces

The proposed method detects obstacles by applying subtraction between the road surfaces after the registration. Then, the regions with high differences are detected as obstacles. To measure the difference, two types of image features; the brightness and the saturation of the HSV color space are employed.

Initially, an in-vehicle camera image is obtained as a color image. This is converted into a brightness image and a saturation image. Then, the proposed method calculates the absolute value of the difference for each pixel. The hue value is not used because the road surface is almost gray and therefore not stable. Regions with high differences are detected as obstacles.

The road surface registration often contains some errors. The neighborhood search using a normalized cross-correlation is performed to compensate for this. Then, the regions with a similarity higher than a predefined value are removed as false detections.

4 Experiment

To demonstrate the effectiveness of the proposed method, two kinds of experiments were conducted using actual in-vehicle camera images. The first experiment confirms the accuracy of the frame correspondence by the DTW method. The second experiment evaluates the accuracy of the obstacle detection by the proposed method based on image-subtraction.

4.1 Finding corresponding frames

4.1.1 Experimental condition

To acquire images used for the experiment, we mounted a high-end web camera "Logicool Qcam® Pro 9000" on a windshield of a car. The image size was 640 × 480 pixels and the frame rate was 15 fps. 35 pairs of sequences were used in this experiment. For example, sequence A is composed of 2,010 frames while sequence B is composed of 1,435 frames, for the same road section ranging approximately 530 m. The sequences included differences in the running speeds and running positions due to a stop on a traffic light and to avoid obstacles.

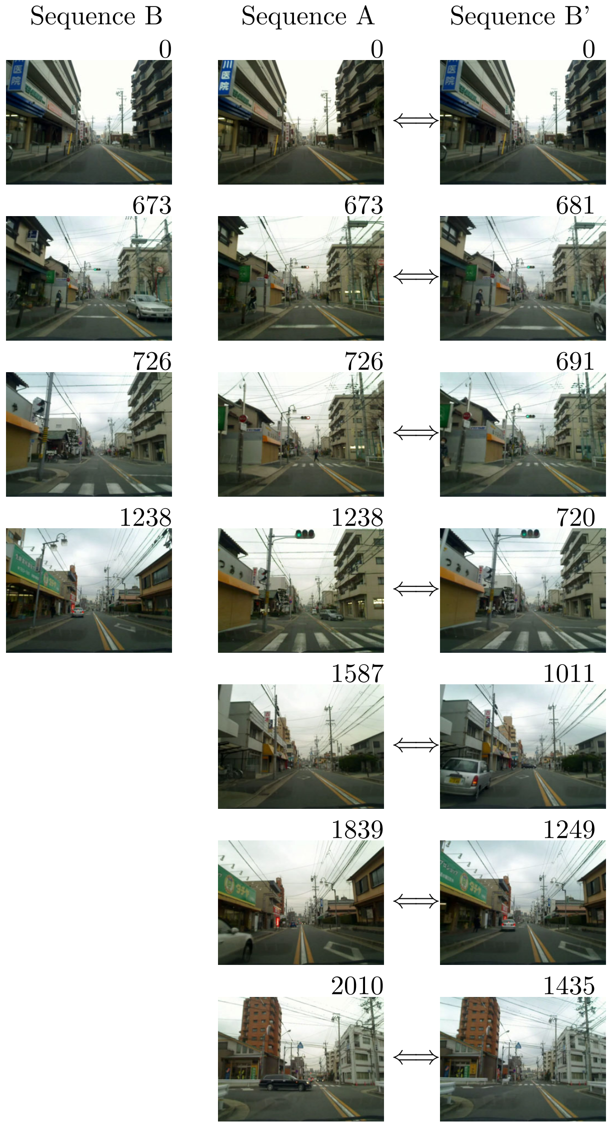


Figure 7. Result of frame correspondence between image sequences A and B. The number on the bottom left of each image represents the frame number.

4.1.2 Results

Figure 7 shows frame samples of the input sequences A and B, and the output sequence B'. Even if the frame numbers of A and B are the same, their captured locations are different because of the difference of running speeds.

From the images of A and B', it was confirmed that the DTW method makes accurate frame correspondences even if the running speeds and positions are different. The matching accuracy was good enough to detect obstacles for all sequences.

4.2 Detection of obstacles

4.2.1 Experimental condition

Two kinds of experimental data were captured by using the same camera as in Section 4.1.1. We used seven sequences including obstacles on a road section ranging approximately 150 m. The total number of frames in these sequences was 1,542, and the number of frames including an obstacle in the distance from 12 to 60 m was 735. The obstacles were a pedestrian crossing the road, a street-parking vehicle, a forward vehicle, a pylon, a cardboard box, and a ball. Besides

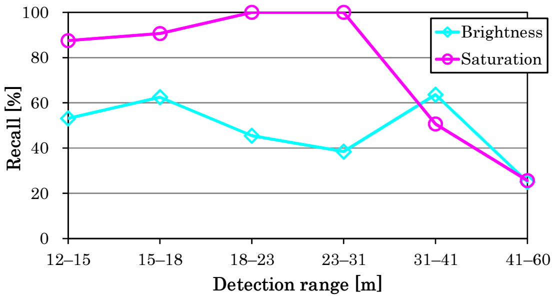


Figure 8. Recall rates of the obstacle detection.

these sequences, another one captured at the same road section including no obstacles was used as the past image sequence.

We evaluated the detection accuracy by the following criterion:

Recall rate = # of true-positives / # of obstacles .

Either the brightness image or the saturation image was used for the detection, and the recall rates were calculated with respect to the distance to the obstacles. Here, the thresholds for the detection were determined so that the average number of false-positives per frame should be less than 0.003.

4.2.2 Result and Discussion

The recall rate for each detection range is shown in Fig. 8. The examples of the detected result of distant obstacles are shown in Fig. 9.

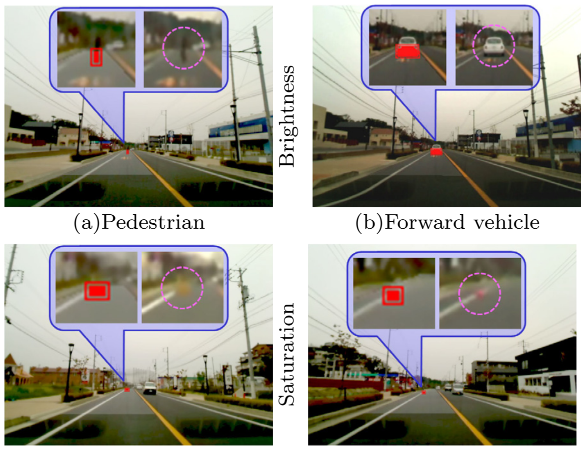
Figure 8 indicates that the obstacles closer than 31 m were able to be detected accurately by using the saturation value. In addition, the proposed method could detect small obstacles such as a ball or a pylon at a distance of more than 31 m. The obstacles with low saturation such as a vehicle or a pedestrian were successfully detected by using the brightness value. Therefore, combining these two criteria may improve the detection accuracy.

From Fig. 9, it was confirmed that various distant obstacles could be detected by the proposed method. Especially, Fig. 9(d) demonstrates the remarkable ability of the proposed method. The proposed method successfully detected the ball with a diameter of 20 cm at a distance of 48 m. This means that small obstacles could be detected far enough to be avoided safely even when running at a speed of 60 km/h.

5 Summary

We proposed a method for detecting general obstacles on a road by subtracting present and past in-vehicle camera images. Compared to the existing learning-based methods that could detect only specific obstacles, the proposed method based on image-subtraction could detect any kind of obstacles. To achieve this, the frame by frame correspondences between the present and the past in-vehicle camera image sequences was obtained. Then the road surface registration between the corresponded frames was carried out. Finally, obstacles were detected by using the subtraction of the road surface regions.

To demonstrate the effectiveness of the proposed method, an obstacle detection experiment was carried out by applying the proposed method to several image sequences actually captured by an in-vehicle camera.



(a)Pedestrian (b)Forward vehicle
(c)Cardboard box (d)Ball

Figure 9. Detection results for obstacles approximately 50m ahead. The rectangles represent the detected regions. The pair of images in the balloon in each image is the closeup of the detected region (left) and its original image (right).

The experimental results showed that the proposed method could detect general obstacles accurately at a distance far enough to safely avoid them.

Future work includes the integration of the different criteria, and the evaluation in different lighting conditions including different time and weathers.

Acknowledgement

Parts of this research were supported by JST CREST and MEXT, Grant-in-Aid for Scientific Research. This work was developed based on the MIST library (`http://mist.murase.m.is.nagoya-u.ac.jp/`).

References

[1] F. Liu, J. Sparbert, and C. Stiller, "IMMPDA Vehicle Tracking System using Asynchronous Sensor Fusion of Radar and Vision," Proc. IEEE Intelligent Vehicles Symposium, pp.168–173, 2008.

[2] J. M. Collado, C. Hilario, A. de la Escalera, and J. M. Armingol, "Self-calibration of an On-Board Stereo-vision System for Driver Assistance Systems," Proc. IEEE Intelligent Vehicles Symposium, pp.156–162, 2006.

[3] K. Nishida and T. Kurita, "Boosting with Cross-Validation Based Feature Selection for Pedestrian Detection," Proc. IEEE World Congress on Computational Intelligence, pp.1251–1257, 2008.

[4] T. Mitsui and H. Fujiyoshi, "Object Detection by Joint Features based on Two-stage Boosting," Proc. Visual Surveillance 2009, pp.1169–1176, 2009.

[5] J. Sato, T. Takahashi, I. Ide, and H. Murase, "Change Detection in Streetscapes from GPS Coordinated Omni-directional Image Sequences," Proc. 18th Int. Conf. on Pattern Recognition, vol.4, pp.935–938, 2006.

[6] J. Böhm, "Multi-image Fusion for Occlusion-free Façade Texturing," Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences., vol.35, no.B5, pp.867–872, 2004.

[7] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Comm. ACM, vol.24, no.6, pp.381–395, 1981.

[8] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. Journal on Computer Vision, vol.60, no.2, pp.91–110, 2004.