MOVING CAMERA BACKGROUND-SUBTRACTION FOR OBSTACLE DETECTION ON RAILWAY TRACKS

Hiroki Mukojima[1]   Daisuke Deguchi[2]   Yasutomo Kawanishi[1]   Ichiro Ide[1]
Hiroshi Murase[1]   Masato Ukai[3]   Nozomi Nagamine[3]   Ryuta Nakasone[3]

[1]Graduate School of Information Science, Nagoya University
[2]Information Strategy Office, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, Japan
[3]Railway Technical Research Institute,
2-8-38 Hikari-cho, Kokubunji-shi, Tokyo, Japan

ABSTRACT

We propose a method for detecting obstacles by comparing input and reference train frontal view camera images. In the field of obstacle detection, most methods employ a machine learning approach, so they can only detect pre-trained classes, such as pedestrian, bicycle, etc. This means that obstacles of unknown classes cannot be detected. To overcome this problem, we propose a background subtraction method that can be applied to moving cameras. First, the proposed method computes frame-by-frame correspondences between the current and the reference (database) image sequences. Then, obstacles are detected by applying image subtraction to corresponding frames. To confirm the effectiveness of the proposed method, we conducted an experiment using several image sequences captured on an experimental track. Its results showed that the proposed method could detect various obstacles accurately and effectively.

Index Terms— Railway safety, Object detection, Subtraction techniques

1. INTRODUCTION

Railway accidents caused by obstacles are one of the most important issues that should be solved. There is a demand for obstacle detection systems, and accordingly, a surveillance system for detecting obstacles in level crossings has been developed [1]. However, the area that can be monitored by this system is restricted due to a fixed camera.
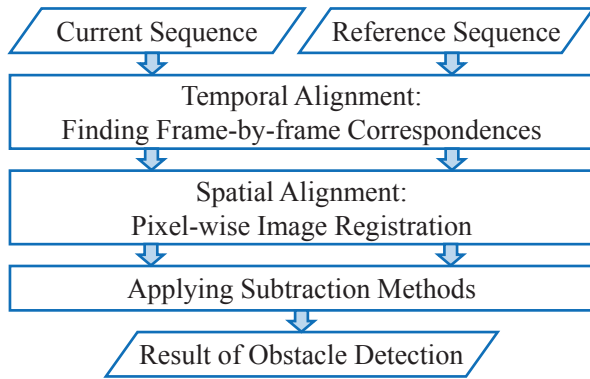
On the other hand, various sensing devices can be used for obstacle detection by being mounted on the front of a train. Since these devices do not require large modifications to the current railway system, especially to ground-side facilities, they may be easily introduced. Therefore, obstacle detection methods using frontal view sensors are expected [2, 3, 4, 5, 6, 7]. However, in the case of railway, distant obstacles must be detected since the braking distance of a train is very long.

Therefore, using millimeter-wave RADAR and LIDAR is not an option due to their low resolutions. In addition, using multiple sensors increases the cost. From this point of view, a train frontal view camera can be considered as the option for obstacle detection in a railway system.

Object detection by camera is one of the most active research areas in the computer vision field, and numerous methods have been proposed [6, 7, 8, 9, 10]. Most methods employ a machine learning approach, and they can detect pre-trained objects, such as pedestrian, bicycle, etc. However, unknown objects cannot be detected by these methods. Although background subtraction could be a solution, it cannot be simply applied to a train frontal view camera, since it moves together with the train. Therefore, it is important to develop a method for forward obstacles detection based on background subtraction that can be applied to a train frontal view camera moving together with the train.

Although few research groups have tackled background subtraction for moving cameras, most of them use a single image sequence and only moving objects can be detected [11, 12]. Meanwhile, Kyutoku et al. proposed a method for detecting general obstacles by a car mounted camera by subtracting the current image sequence from the reference (database) image sequence [13]. By assuming that these two image sequences are captured on slightly different driving paths, this method succeeded to accurately align two image sequences with the metric. This assumption requires sufficient base-line length between cameras capturing the two image sequences to compute the metric between the sequences. However, in the case of railway, sufficient base-line length cannot be obtained since trains always run on the same tracks. In addition, since this method only aligns road surfaces between two image sequences, a large registration error will occur outside of it. Thus, distant / small obstacles cannot be distinguished accurately due to noise caused by the image registration error.

Therefore, we propose a moving camera background subtraction method,which method detects obstacles by compar-

Fig. 1. Framework of the proposed method.

Fig. 2. Differences of view by varying capturing locations.

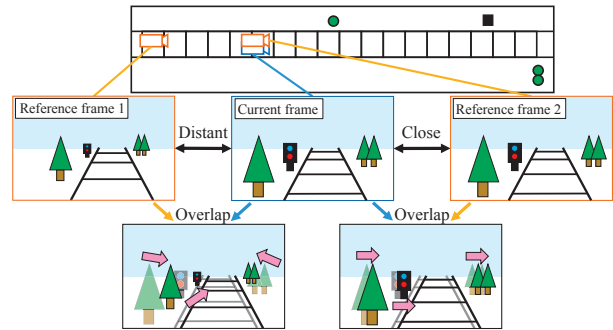ing input and reference images. The contributions of this paper are:

1. Introduction of a new metric that can align two image sequences even if the base-line length between two cameras is small.

2. Detection of arbitrary distant obstacles by pixel-wise image registration and integration of multiple image subtraction mechanisms.

2. MOVING CAMERA BACKGROUND SUBTRACTION FOR OBSTACLE DETECTION

To detect obstacles by subtracting two image sequences, pixel-level alignment is needed. In the case of a train frontal view camera, since an image sequence is captured from a moving train, two image sequences must be aligned both spatially and temporally. To solve this, the proposed method first finds a reference frame captured at the most similar location to the current frame by image sequence matching. Then, it performs pixel-wise registration between the current frame and its corresponding reference frame. Finally, multiple image subtraction methods are applied to compute the image difference between the two frames, and obstacles are detected by integrating their outputs. Figure 1 shows the framework of the proposed method.

2.1. Temporal alignment: Computation of frame-by-frame correspondences

In the case of railway, train frontal view cameras always take the same trajectory since trains run on the same track. This results in a very short base-line length between cameras of the current and the reference image sequences. To cope with this situation, the proposed method introduces a new metric to align the two image sequences. Figure 2 shows close and distant train frontal views of the current and the reference image sequences.

Let the current and the reference image sequences be $F = \{f_1, f_2, ..., f_p\}$ and $G = \{g_1, g_2, ..., g_q\}$, respectively. Here, f_i denotes the i-th frame of the current image sequence, and g_j denotes the j-th frame of the reference image sequence. First, the proposed method computes the frame-by-frame correspondences between sequences F and G. Next, the distance $d(i, j)$ between the current frame f_i and the reference frame g_j is calculated as the variance of angles between corresponding key-points as,

$$d(i,j) = \begin{cases} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} h(\theta_{ijk} - m_{ij}) & (n_{ij} \geq \alpha) \\ \infty & (\text{otherwise}) \end{cases}, \quad (1)$$

$$h(x) = \min \{x^2, (x + 2\pi)^2, (x - 2\pi)^2\}, \quad (2)$$

where n_{ij} is the number of corresponding key-point pairs between f_i and g_j, θ_{ijk} is the angle of the k-th key-point pair represented by the polar coordinate system, m_{ij} is the mean of θ_{ijk}, and α is a positive constant. Here, the angle is represented by relative angle from the x-axis. In this equation, if the current frame is captured at a camera position close to the reference frame, the variance becomes small. Moreover, it can be computed regardless of the base-line length between two cameras. Finally, frame correspondences (f_i, g_j) between the current and the reference image sequences are obtained by applying Dynamic Time Warping to minimize $d(i, j)$. Figure 3 shows an example of corresponding frames of the current and the reference image sequences.

2.2. Spatial alignment: Computation of pixel-wise image registration for temporally aligned frames

To obtain accurate image alignment, the proposed method performs pixel-wise image registration against corresponding frames f_i and g_j obtained in the previous step. Here, DeepFlow [14] is used for calculating the deformation field from g_j to f_i. Then, completely aligned image g'_j is obtained by applying the deformation field to g_j. Figure 4 shows the absolute image difference between the frames in Fig. 3. Figure 4(a) shows the image difference between the original frames $|f_i - g_j|$, and Fig. 4(b) shows the image difference

(a) Current frame f_i (b) Reference frame g_j
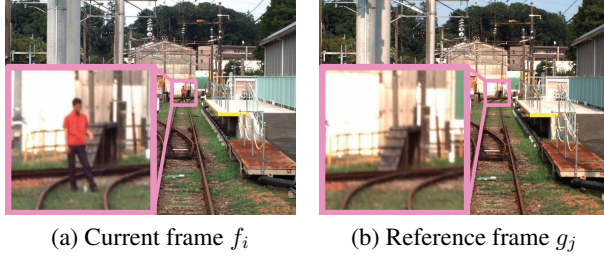
Fig. 3. Corresponding frames (f_i, g_j) in the current and the reference image sequences.

(a) Before alignment   (b) After alignment
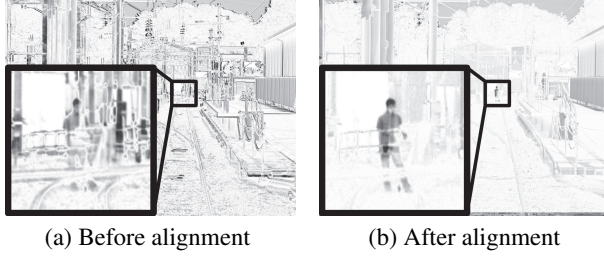
Fig. 4. Examples of image difference between the corresponding frames in Fig. 3.

after pixel-wise alignment $|f_i - g'_j|$. In these images, darker pixels indicate larger image errors.

2.3. Image subtraction for completely aligned images

Robustness against lighting conditions is one of the most important issues when developing a system for railways, since it needs to handle various environments. Here, multiple image subtraction metrics are combined to solve this problem.

First, two types of image subtraction metrics are calculated from f_i and g'_j. The first one is Normalized Vector Distance (NVD), and is calculated as,

$$\text{NVD}(\mathbf{a}, \mathbf{b}) = \left\| \frac{\mathbf{a}}{\|\mathbf{a}\|} - \frac{\mathbf{b}}{\|\mathbf{b}\|} \right\|. \quad (3)$$

Here, $\mathbf{a}$ and $\mathbf{b}$ are image patches represented in vectors consisting of RGB channels. The second one is Radial Reach Filter (RRF) proposed by Satoh et al. [15]. RRF is calculated by comparing the intensity of each RGB channel between the target pixel and its surroundings.

Next, to reduce noise, Gaussian filter is applied to difference images obtained by NVD and RRF. Then, two binary images d^1_{ij} and d^2_{ij} are obtained by thresholding. Here, the threshold T for the binarization is determined as,

$$T = \mu_{ij} + n\sigma_{ij}, \quad (4)$$

where μ_{ij} and σ_{ij} are the average and the variance of each difference image, respectively. Finally, the extracted pixels are considered as candidates of obstacles.

(a) NVD   (b) RRF
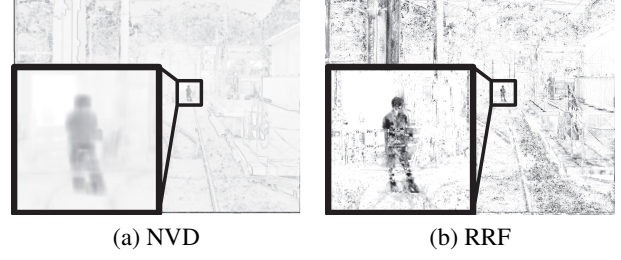
Fig. 5. Image difference calculated by NVD and RRF.

(a) NVD   (b) RRF
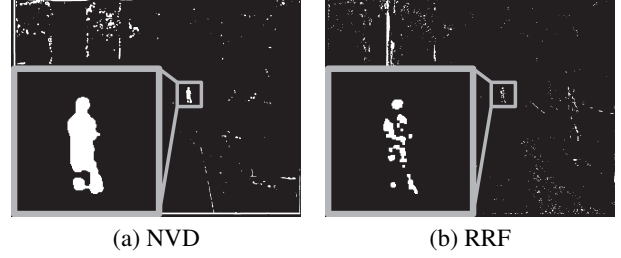
Fig. 6. Obstacles detected by NVD and RRF.

Figure 5 shows examples of image differences calculated by NVD and RRF, and Fig. 6 shows the obstacles detected by thresholding the difference images in Fig. 5. In Fig. 5, the black pixels indicate large image differences and in Fig. 6, the white pixels indicate obstacle candidates.

To reduce the false positives of obstacles, the proposed method integrates the above NVD and RRF images. Before integration, to reduce salt and pepper noise caused by ballast, grasses, and shadow borders, morphological operations (opening and closing) are applied to both d^1_{ij} and d^2_{ij}.

Then, connected-component labeling is applied, and the bounding-box of each connected-component is extracted. Here, there is a relation between the size of a bounding-box and the distance between the camera and the obstacle; A distant obstacle is observed as a small bounding-box, and a closer obstacle is observed as a large bounding-box. Therefore, the proposed method removes bounding-boxes that do not satisfy this relation. To evaluate this relation, the proposed method refers to the gauge (the width of the track) measured from the image, whose actual width is fixed and known. Then, the size of each bounding-box is compared with the gauge in the image, and small bounding-boxes are removed. In addition, the proposed method removes bounding-boxes outside the track because they do not obstruct the train. Here, the method by Nasuu et al. [16] is used for detecting tracks from a train frontal view image. These steps are applied to obstacle candidates of NVD and RRF, separately.

Finally, the proposed method outputs obstacles detected by both NVD and RRF. Here, it outputs only overlapped bounding-boxes between NVD and RRF.
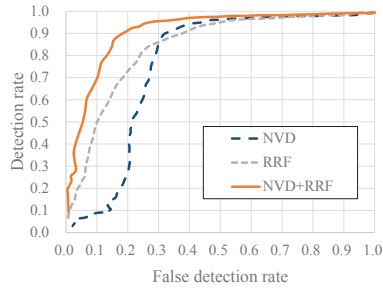
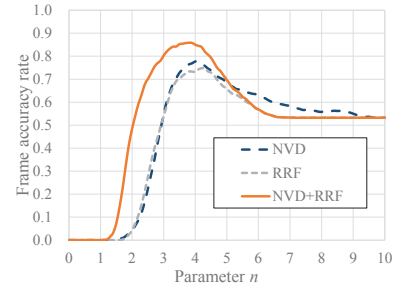**Fig. 7**. Example of a detected obstacle.　　**Fig. 8**. ROC curves.　　**Fig. 9**. Frame accuracy rate.

3. EXPERIMENTS AND DISCUSSIONS

To evaluate the effectiveness of the proposed method, we prepared train frontal view images captured on a test line in the premises of the Railway Technical Research Institute, Japan. Grasshopper3 (Point Grey Research, Inc.) was mounted on 2.5 m height of the front view of a railway trolley. The size of captured images was $1,920 \times 1,440$ pixels, and the frame rate was 10 fps. The focal length of the camera was 25 mm, and the pixel pitch was $4.54\,\mu$m. In this experiment, the railway trolley was controlled manually. A total of 2,117 frames were contained in the dataset which was constructed by extracting frames in five frames interval from the recorded five videos. No obstacle existed in three videos, and the other two videos included a pedestrian and a box as obstacles, respectively. Bounding-boxes of all obstacles were annotated manually. One of the videos including no obstacle was used as the reference image sequence, and the other videos were used as the current image sequences.

The result was evaluated by changing parameter n that controls the threshold in Eq. (4). The detection accuracy was evaluated by the following criteria:

$$\text{Detection rate} = \frac{\text{\# of detected obstacles}}{\text{\# of all obstacles}}$$

$$\text{False detection rate} = \frac{\text{\# of false detections}}{\text{\# of all detections}}$$

$$\text{Frame correctness rate} = \frac{\text{\# of correct frames}}{\text{\# of all frames}}$$

Here, if the detected bounding-box covered more than 10 % of an annotated obstacle region, we considered that it was detected correctly. Otherwise, it was considered as a false positive. Δx and Δy are the differences of the size between the detected obstacle and the annotated obstacle in x (horizontal) and y (vertical) coordinates. If the following condition was not satisfied, it was considered as a false positive.

$$2\Delta x \le w \text{ and } 2\Delta y \le h \tag{5}$$

Here, w and h are the width and the height of the manually annotated obstacle, respectively. Frame correctness was cal-

culated by counting the number of frames where obstacles were detected correctly.

Based on these criteria, the proposed method was compared with methods using either NVD or RRF. Figure 8 shows the ROC curve of each method. As seen here, the proposed method could achieve the highest accuracy. In addition, from Fig. 9, we can see that the proposed method obtained the best frame correctness of 85.8 % when $n = 3.8$.

Although the proposed method could obtain the best, it sometimes failed to detect obstacles. This could be explained that since the proposed method integrated NVD and RRF for detecting obstacles, the performance degradation of NVD or RRF affected the result of the proposed method. To overcome this problem, we will investigate other metrics for image subtraction.

4. CONCLUSIONS

This paper proposed a method of moving camera background subtraction for forward obstacle detection from a train frontal view camera. To detect general obstacles, frame-by-frame correspondences between the current and the reference image sequences of train frontal view were computed based on the angle difference of corresponding key-points. After, pixel-wise image registration, obstacles were detected by integrating two kinds of subtraction methods.

To demonstrate the effectiveness of the proposed method, experiments were conducted by capturing train frontal view image sequences on an experimental track. Its results showed the effectiveness of the proposed method.

Future works include introduction of a background modeling method and evaluation in various lighting conditions, seasons, and weathers.

Acknowledgement

5. REFERENCES

[1] Y. Sheikh, Y. Zhai, K. Shafique, and M. Shah, "Visual monitoring of railroad grade crossing," in *Proc. SPIE, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III*, 2004, vol. 5403, pp. 654–660.

[2] S. Sugimoto, H. Tateda, H. Takahashi, and M. Okutomi, "Obstacle detection using millimeter-wave RADAR and its visualization on image sequence," in *Proc. 17th Int. Conf. on Pattern Recognition (ICPR 2004)*, 2004, vol. 3, pp. 342–345.

[3] C. Premebida, O. Ludwig, and U. Nunes, "Exploiting LIDAR-based features on pedestrian detection in urban scenarios," in *Proc. 12th IEEE Int. Conf. on Intelligent Transportation Systems*, 2009, pp. 1–6.

[4] S. Mockel, F. Scherer, and P. F. Schuster, "Multi-sensor obstacle detection on railway tracks," in *Proc. 2003 IEEE Intelligent Vehicles Symp. (IV2003)*, 2003, pp. 42–46.

[5] M. Ruder, N. Mohler, and F. Ahmed, "An obstacle detection system for automated trains," in *Proc. 2003 IEEE Intelligent Vehicles Symp. (IV2003)*, 2003, pp. 180–185.

[6] 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 (CVPR2005)*, 2005, vol. 1, pp. 886–893.

[7] 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 (CVPR2008)*, 2008, pp. 1–8.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR2001)*, 2001, vol. 1, pp. 511–518.

[9] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. 1999 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR1999)*, 1999, vol. 2, pp. 246–252.

[10] O. Barnich and M. V. Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2010.

[11] A. Elqursh and A. Elgammal, "Online moving camera background subtraction," in *Proc. 12th European Conf. on Computer Vision (ECCV2012)*, 2012, vol. 6, pp. 228–241.

[12] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Proc. 12th IEEE Int. Conf. on Computer Vision (ICCV2009)*, 2009, pp. 1219–1225.

[13] H. Kyutoku, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "Subtraction-based forward obstacle detection using illumination insensitive feature for driving-support," in *Proc. Computer Vision in Vehicle Technology: From Earth to Mars (CVVT2012), Lecture Notes in Computer Science*, 2012, vol. 7584, pp. 515–525.

[14] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. 14th IEEE Int. Conf. on Computer Vision (ICCV2013)*, 2013, pp. 1385–1392.

[15] Y. Satoh, H. Tanahashi, C. Wang, S. Kaneko, S. Igarashi, Y. Niwa, and K. Yamamoto, "Robust event detection by Radial Reach Filter (RRF)," in *Proc. 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002)*, 2002, vol. 2, pp. 623–626.

[16] B. T. Nassu and M. Ukai, "A vision-based approach for rail extraction and its application in a camera pantilt control system," *IEEE Trans. on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1763–1771, 2012.