Frontal Obstacle Detection Using Background Subtraction and Frame Registration

Ryuta NAKASONE  Nozomi NAGAMINE
Signalling Systems Laboratory, Signalling and Transport Information Technology Division

Masato UKAI
Signalling and Transport Information Technology Division

Hiroki MUKOJIMA  Daisuke DEGUCHI  Hiroshi MURASE
Nagoya University

Systems such as Automatic Train Protection and moving block sections help prevent trains colliding, however collisions with unexpected obstacles in front of a train can only be avoided if seen by the driver. In an effort to reduce the possibility of this type of collision and to improve passenger safety, an obstacle detection method has been proposed using a monocular camera and image processing. The proposed method can detect obstacles by comparing live images from the camera with images obtained by other trains operating earlier along the same route. The difference between the two sets of images are defined as obstacles. The performance of the method was verified by conducting experiments using rolling stock and imitation obstacles.

Keywords: obstacle detection, image processing, background subtraction

1. Introduction

Train collisions are prevented with systems such as Automatic Train Protection (ATP) but collisions with foreign objects on the track can only be prevented if a driver sees the obstacle. Preventive solutions such as Rock-fall Detection Devices and Obstruction Warning Signals have been installed, but they are only effective in limited situations and so the need for an onboard “anti-collision system” is high.

This paper proposes a method to detect obstacles on the track by means of train front cameras and image processing. Background subtraction is used as the main method for obstacle detection. Differences between live images from the camera and images collected from other driving cabs, which have previously run the same routes, are assumed to be potential obstacles. The main reason for selecting this particular method of detection which does not require the definition of each detected object, is to avoid having to add a machine learning function to the system. Background subtraction in theory is an effective method to detect objects in a specific area captured on camera. Using this approach, moving objects are detected by calculating the difference between a live frame and a reference frame, often called the “background image/model”. Normally this method consists of a fixed camera detecting a moving object in the Region of Interest (ROI). Applying this to a moving train therefore presents some difficulties.

Potential obstacles on the track that a train operating daily throughout the year may run into include fallen rocks and fallen trees due to severe weather, wild animals crossing the tracks, and more commonly, human beings such as railway workers and pedestrians at railroad crossings. The Japan Transport Safety Board, an organization under the

Table 1: Accidents resulting in derailment

<table>
<thead>
<tr>
<th>Year</th>
<th>Weather</th>
<th>Speed (km/h)</th>
<th>Distance to object (m)</th>
<th>Type of object</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>Cloud</td>
<td>62</td>
<td>130</td>
<td>Trees</td>
</tr>
<tr>
<td>2006</td>
<td>Snow</td>
<td>60</td>
<td>100</td>
<td>Trees</td>
</tr>
<tr>
<td>2014</td>
<td>Snow</td>
<td>70</td>
<td>50</td>
<td>Trees</td>
</tr>
<tr>
<td>2014</td>
<td>Rain</td>
<td>50</td>
<td>60</td>
<td>Trees (Landslide)</td>
</tr>
<tr>
<td>2002</td>
<td>Rain</td>
<td>120</td>
<td>100</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2003</td>
<td>-</td>
<td>43</td>
<td>10</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2004</td>
<td>-</td>
<td>45</td>
<td>80</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2005</td>
<td>Sunny</td>
<td>50</td>
<td>100</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2011</td>
<td>-</td>
<td>35</td>
<td>27</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2012</td>
<td>Sunny</td>
<td>20</td>
<td>7</td>
<td>Fallen rocks</td>
</tr>
<tr>
<td>2012</td>
<td>-</td>
<td>60</td>
<td>30</td>
<td>Fallen rocks</td>
</tr>
</tbody>
</table>
Ministry of Land Infrastructure Transport and Tourism, publishes reports on major railway accidents and incident that are open to the public. By looking into these reports, we found that there have been more than 10 collisions with fallen rocks and fallen trees which resulted in train derailment, since 2001. Table 1 shows the year each accident occurred, the weather at the time, and speed and distance of the train from the object when it was seen by the driver, and the type of the obstacle.

The general use of cameras and image processing has started across a number of industrial fields. Especially in the automotive industry, practical use of cameras for Active Driver Assist (ADA) systems has had a huge impact on improving driver safety. Also in the railway industry, cameras are being introduced for example in driver recorders, and further use of captured images is expected. Estimating and defining every single object the train may collide with is unrealistic, and so we have chosen a method for obstacle detection which does not require target objects to be defined.

2. Technical survey

The performance of Active Driver Assist (ADA) in the automotive industry has improved. This section describes a study that was conducted to identify the advantages and disadvantages of applying this system to trains. The two key technologies used for ADA are estimation of user position and obstacle detection. Confiming user position within a 3D environment and detecting obstacles in their path enables the user to navigate through urban areas without colliding with the back of a truck or to change lanes smoothly. The main sensors in the collision detection device are laser sensors that use an Infra-red or a millimeter wavelength, and/or passive sensors such as cameras.

2.1 LiDAR

Expectations in relation to the potential of Light Detection and Ranging (LiDAR) technology have risen in recent years. This is a type of radar which uses pulse laser instead of radio waves to measure distance. It can detect a wide range of materials including non-metallic objects and its narrow beam can map features to a high resolution.

There are mainly two types of LiDAR devices, a built-in sensor with a limited field of view and a fully rotational sensor with a field of view of 360 degrees. Both sensors are able to map areas to a distance of up to 150 m to a resolution of a few centimeters, however observation data is dispersive, which means that they must often be used together with a camera to determine target objects.

2.2 Computer stereo vision

One of the most successful ADA systems in Japan is “Eyesight”, developed by the Subaru Corporation, which uses computer stereo vision. Computer stereo vision extracts 3D information from images observed by two strictly aligned cameras. The cameras are displaced horizontally from one another, and calculating the parallax of the same object allows us to detect the distance to the target. Although this is a promising technology, it is difficult to calculate the parallax of objects that are far away from the observer.

2.3 Deep neural network

In the field of research, methods using Deep Neural Networks (DNN) have enabled high gains in accuracy in pedestrian detection. This is a machine learning algorithm using layers of nonlinear processing units to extract features without supervision. However, DNN models are known to be slow when used for slide window classifiers and the target object must be defined.

2.4 Fundamental features

In sum, an obstacle detection system needs to be able to:

1. Detect objects from as far away as possible.
2. Detect obstacles of different shape and size.
3. In case of failure, identify the causes for the non-identification.

The biggest difference between automotive and railway vehicles is breaking distance. While a car running at a speed of 100 km/h stops probably at a distance of 100 m, a train running in the same conditions would need a distance of 600 m or more to stop. Under the methods mentioned previously in 2.1 and 2.2, the distance range of detection is limited, and under the method in 2.3, there are limitations regarding the target object. This paper therefore proposes a method using a single camera and background subtraction by means of image processing.

3. Proposed method

This section describes method and approach used to detect obstacles in front of a train using image processing. It is assumed that the train is running on a single track, and images collected from trains passing the same section earlier are used as background images. These images serve as a reference to the live feed from the train front camera and are able to detect differences on the track ahead.

To detect obstacles by comparing input and reference images, pixel level alignment is needed. In the present method, first a reference frame, an image captured nearest to the current train’s position, is extracted from the database using image sequence matching [1]. Pixel level alignment makes it possible to adjust small differences in the images and obtain perfectly matched images. Finally, multiple image subtraction methods are applied to compute the difference between the two frames, that is, the absence or presence of obstacles in front of the train.

3.1 Selecting the reference frame

Finding reference frames from pre-collected images in the database must be reliable and efficient. Train front view cameras always follow the same trajectory when passing the same route. This results in a very short base-line length between the cameras shooting real time images and the reference image sequences. Figure 1 shows the close and distant train front views of the current and the reference image sequences. Figure 2 shows the difference in view in captured locations A and B.

When the live frame is captured at a position near to
the reference frame, the relative angle from the x-axis becomes smaller, compared to frames captured away from it. Using this measure as the frame correspondence rate, it is possible to compute the value regardless of the base-line length between the cameras.

3.2 Pixel level alignment

In order to obtain accurate image alignment with the method proposed, pixel-wise image registration is performed against the two corresponding images obtained in the previous step. Here, DeepFlow [2] is used to calculate the deformation field in each image, and by applying the deformation field, the two images are completely aligned. Figure 3 shows the image difference between the original frames and Fig. 4 shows the image difference after pixel-wise alignment. In these images, darker pixels indicate larger image errors.

Here, two types of image subtraction method are combined to solve this problem. The first one is Normalized Vector Distance (NVD), for detecting differences in color, and the second one is Radial Reach Filter (RRF) [3], for detecting differences in texture. NVD is calculated as,

$$ NVD(a, b) = \frac{a - b}{||a|| \cdot ||b||} $$

(1)

Here, a and b are image patches represented by vectors consisting of RGB channels. For noise suppression, a Gaussian filter is applied to both images obtained by NVD and RRF. Then, two binary images are processed by applying a threshold and the extracted pixels are considered as candidates for obstacles. Finally, morphological operations (opening and closing) and connected-component labeling are applied, and the bounding box of each component is extracted.

Figure 5 shows examples of image differences calculated by NVD and RRF. The black pixels indicate large image differences which indicate possible obstacles.

4. Evaluation

The performance of the proposed method was evaluated through an experiment using a test car and a test track at the Railway Technical Research Institute, Japan. A camera was mounted at the front of a cab at a height of 2.5 m. The size of the images captured was 1,920 by 1,440 pixels. In this experiment, the railway trolley was controlled manually. The test track was made of a straight section approximately 250 m in length leading up to the obstacle and curved from there on coming from the opposite direction.

4.1 Evaluation method

The correct bounding boxes for all the obstacles were annotated manually. If the detected bounding box covered more than 10 % of an annotated obstacle area, it was considered to have been detected correctly. Otherwise, it was considered to be a false detection. Also when the following condition was not satisfied, it was considered to be a false
Here, w and h are the width and the height of the manually annotated obstacle, respectively. The Δx and Δy are differences in size between the detected obstacle and the annotated obstacle in x (horizontal) and y (vertical) coordinates respectively. The relationships of the bounding boxes are shown in Fig. 6.

The detection accuracy was evaluated by the following criteria:

\[
\text{Detection rate} = \frac{\# \text{ of detected obstacles}}{\# \text{ of obstacles}}
\]  

The performance of the method proposed was evaluated based on the criteria.

### 4.2 Test results

Images from 4 runs without obstacles were collected for the experiment to serve as a reference. 17 runs with a total of 5,000 frames were then made, with various objects placed as obstacles. Figure 7 shows examples of potential obstacles. The image frames containing obstacles were used for evaluation. Applying the proposed method achieved a detection rate of 80% at distances of up to 200 m, regardless of the shape of the obstacle. The farthest obstacle successfully detected was 232 m away from the camera. The detection rate for each distance from the obstacles are shown in Fig. 8. However, there were several cases of false detection due to movement of shadows, and difference in the period of time between the reference frame and the obtained image. Other metrics for image subtraction will be investigated to try to overcome this problem.

### 5. Conclusion

This paper proposes a method for detecting obstacles on railway tracks by means of train front cameras and image processing. The frame correspondence rate can be calculated regardless of the short base-line length between detection.

\[
2\Delta x \leq w \quad \text{and} \quad 2\Delta y \leq h
\]  

Fig. 5 Obstacle candidates

Fig. 6 Positive detection

Fig. 7 Test targets
cameras. The pixel level alignment and two different image subtraction methods are effective for detecting both human and non-human obstacles on railway tracks. Furthermore, in this method pre-definition of target obstacles is not necessary. Future work will include introduction of other background subtraction methods and evaluation for various lighting conditions, seasons and weather.

Fig. 8 Detection rate

<table>
<thead>
<tr>
<th>Detection Rate (%)</th>
<th>Distance to Obstacle (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>60</td>
<td>150</td>
</tr>
<tr>
<td>80</td>
<td>200</td>
</tr>
<tr>
<td>100</td>
<td>250</td>
</tr>
<tr>
<td>100</td>
<td>300</td>
</tr>
</tbody>
</table>

Non-human

Human

References


Authors

Ryuta NAKASONE
Researcher, Signalling Systems Laboratory, Signalling and Transport Information Technology Division
Research Areas: Image Processing

Nozomi NAGAMINE, Dr. Eng.
Chief Researcher, Signalling Systems Laboratory, Signalling and Transport Information Technology Division
Research Areas: Image Processing

Masato UKAI
Principal Researcher, Signalling and Transport Information Technology Division
Research Areas: Image Processing

Hiroki MUKOJIMA
Nagoya University

Daisuke DEGUCHI
Nagoya University

Hiroshi MURASE
Nagoya University