Visual Simulation Method for Installation of Wayside Facilities Using Video Sequences

Nozomi NAGAMINE

Signalling Systems Laboratory, Signalling and Transport Information Technology Division

Ryuta NAKASONE

Masato UKAI

Signalling and Transport Information Technology Division

This research aims to create a simulator that would generate simulated images for use when investigating where to install wayside signals along a railway track. A simulation method has been developed, whereby the virtual line of sight of the driver with respect to wayside signals on a track is displayed by using the images of signals overlaid onto the view from the front of the train. By means of this simulation technique, the time needed for the study of on-site wayside signal positioning can be cut without imposing restrictions on train schedules or operations. This report describes the algorithms applied and explains the verifications made using test data.

Keywords: simulation, installation position, image processing, wayside signal

1. Introduction

Currently simulated wayside signals made out of painted plastic bottles or other similar devices are used to consider installation positions, because visibility has to be checked from the driver's cab. A staff member holds the dummy wayside signal at the planned position and visibility is checked from a position located at the required distance (e.g. a point 600 m away from the planned signal position). These operations require both the signal staff and train operation staff to schedule a meeting on-site after a lengthy process; this is one of the reasons why it takes considerable time before wayside signals are actually installed.

Japanese wayside signals are classified into two categories: main signals which are always turned on; and others, such as obstruction warning signals which are normally off. Obstruction warning signals flash red whenever a critical event occurs (Fig.1).

Wayside signal installations are normally subject to plans prepared several years in advance and for work to be spread over a certain period of time. At the same time, there is an urgent need to improve the safety of level crossings, and as a result, the introduction of obstruction warning signals has been rushed in Japan.

In order to deal with this problem, a simulation method has been developed [1,2,3], whereby camcorder footage taken from the driver's cab is used and the virtual line of sight of the driver of wayside signals on the track is displayed by using overlaid images. This simulation method is used to adjust and establish scheduled wayside signal positions. Simulation makes it possible to cut the time needed for determining on-site wayside signal positions without restricting train schedules or operations. This report describes the algorithms applied and explains verifications made using test data.

Fig. 1 Obstruction warning signal

2. System layout

The system can be described briefly as follows: the wayside signal position display simulator is given a range of data, including photographic images of the forward looking line of sight of the driver seen from the cab, and positional information of the planned location of the wayside signals. The simulator outputs images depicting the wayside signals overlaid on the forward-looking view from the cab.

Broadly speaking, the simulator relies on four data processes: images of the forward-looking view from the cab and kilometer matching, identification of rails from the images of the forward-looking view from the cab and conversion into coordinates, conversion of absolute wayside signal location coordinates to corresponding coordinates of the images of the forward-looking view from the cab, and display of overlaid wayside signals. These processes are described below.

- 1) Images of the forward-looking view from the cab and kilometer matching

Using image frames alone does not indicate the kilometer mark of where the train is, making the precise determination of where to install wayside signals impossible. For this reason, the images are correlated with kilometer marker points on the track.

- 2) Identification of rails from the images of the forward-looking view from the cab and conversion into coordinates

Though wayside signal positions are determined by comparing them with the position of the rails shown in the images by means of the simulator, the position of the rail in the images is not detected for every image displayed on the screen, since rail detection imposes a considerable load on the machine. Therefore, it was deemed more effective to identify the rails from the images in advance. In addition, converting the position of the rails from image-based coordinates to train position coordinates makes calculating the position of wayside signals easier. To this end, the position of the rails is detected and the coordinates are converted to an equivalent set of train-based coordinates.

- 3) Conversion of absolute wayside signal position coordinates to coordinates of images of the forward-looking view from the cab

This procedure is composed of the image, its corresponding kilometer position on the track, the position of the rails, and wayside signal positions, to calculate wayside signal positions that on the image would be within the visible range of the view looking forward from the cab.

- 4) Wayside signal display

Wayside signal images were overlaid onto the images of the view looking forwards out of the driving cab on a proportional scale. The position of the wayside signals can be changed, added, or deleted as needed.

3. Coordinate systems and conversion

The system displays the positions of wayside signals and other objects with three parameters: absolute coordinates corresponding to kilometer designations on the track, coordinates starting at the train location, and coordinates based on the view defined onscreen. The explanation of these coordinate systems and their conversion methods are given as follows.

3.1 Coordinate systems

The system calculates coordinates based on kilometer designations on the track (absolute coordinates) based on the center of the tracks at a starting kilometer marker of zero, with increments defined in terms of distance and height from the kilometer marker and track center. The degree of distance from the track center changes according to the corresponding train's direction of travel, with the rightward direction on the x-axis defined as positive non-zero values. Relating this data to points on a map produces coordinates very close to absolute coordinates.

In the train-based coordinates, meanwhile, the center of the track in front of the train and the height of the rail

surface are defined as their points of origin. The system describes positions based on the distance from the front of the train, rightward distance, and height.

In the screen-based coordinates, the center of the image taken from the fore view of the train is defined as their points of origin. The system describes positions based on the location on the horizontal and vertical axes with respect to the center of the image.

3.2 Coordinate conversion

Train-based coordinates and screen-based coordinates are mutually convertible using parameters sourced from the camera. Coordinates and their conversion methods are defined below.

The train-based coordinate system takes the following values: a target i and its position (x_i, y_i, z_i) , the camera orientation $(\theta_x, \theta_y, \theta_z)$, angle of depression around the x-axis (θ_x) , angle of roll around the y-axis (θ_y) , angle of yaw around the z-axis (θ_z) , and camera position (x_c, y_c, z_c) . For screen-based coordinates, the values are described as follows: target position (w_i, h_i) , camera focal length f , and CCD size $(size_w \times size_h)$, and screen resolution $(pic_w \times pic_h)$.

(A) Conversion from screen-based coordinates to train-based coordinates (F)

Because positions in the screen-based coordinate system are represented by two-dimensional data and positions in the train-based coordinate system are represented by three-dimensional data, the conversion lacks one dimension; however, this conversion process is implemented only when converting rail positions, so the target height can be given a value of zero when converting to a train-based coordinate system. The position in question becomes $(x_i, y_i, 0)$, which makes a satisfactory conversion possible.

Positions in the screen-based coordinate system (w_i, h_i) and positions along the rails in the train-based system have the following correspondence:

$$\begin{bmatrix} x_i \\ y_i \\ 0 \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \alpha R \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + w_i \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + h_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

R gives the camera orientation and is defined as follows:

$$R = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \quad (2)$$

(B) Conversion from train-based coordinates to screen-based coordinates (F^{-1})

In the train-based coordinate system, the position of a target in the three-dimensional space is given by (x_i, y_i, z_i) . Camera position, orientation, and focal distance are respectively defined as (x_c, y_c, z_c) , $(\theta_x, \theta_y, \theta_z)$, and f . A virtual screen of the same size as the CCD is assumed, and the variables corresponding to position values (x_i, y_i, z_i) are allowed to be variables (w_i, h_i) in the screen-based coordinate system. The underlying relationship is described by the following formula:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \alpha R \left(f \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + w_i \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + h_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \quad (3)$$

Solving (w_i, h_i) for the unknown α yields the following:

$$\begin{bmatrix} w_i \\ f \\ h_i \end{bmatrix} = \frac{1}{\alpha} R^{-1} \left(\begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \right) \quad (4)$$

The formula can then be expanded for f . This allows reciprocal conversion between the screen-based coordinate system and train-based coordinate system.

4. Correspondence between the image of the forward looking view from the cab and the kilometer mark

In the system, it is possible to make the kilometer marks (distance designations on the track) correspond to the image of the forward looking view from the cab by measuring the speed of each image frame in pixels and polling image frame/pixel-level distance data. Image frame/actual distance (kilometer marker) data is then obtained by converting this scale. This data processing routine consists of five subroutines: aerial view conversion of the image of the forward looking view from the cab, image frame/pixel-level speed measurement, image frame/pixel-level distance computation, creation of comparative data ambient equipment positions and image frames, and final creation of image frame/kilometer marker data.

4.1 Aerial view conversion of the image of the forward looking view from the cab

In order to calculate the speed of the train from the image of the forward looking view from the cab, the rate of movement of target objects in the image over a given time is computed. The accurate measurement of the rate of movement is conducted by adjusting each element so as to make the ratio between the apparent length on the image and the actual length invariable at every interval. The rail level view of the track in the image of the forward looking view from the cab is, therefore, corrected for the aerial view which is then used for computing the speed.

The aerial view conversion is achieved by specifying four points along the left and right rails (distant and proximate points) and then allowing all future conversions to be automated. In terms of camera orientation, only X-axis dip angle is assumed. Train-based coordinates are displaced, with the use of a new coordinate system in which the camera's focal point is defined as its point of origin. In this case, the image-based coordinates and camera-based coordinates have the following relationship:

$$x_i = -\frac{Hw_i}{f \sin \theta_x + h_i \cos \theta_x} \quad (5)$$

$$y_i = -\frac{H(f \cos \theta_x - h_i \sin \theta_x)}{f \sin \theta_x + h_i \cos \theta_x} \quad (6)$$

Two sets of equidistant points on the left and right sides of the rails are then defined as follows: the left-top point is set as $L_t : (x_{L_t}, y_{L_t}, -H)$; the right-top point, as $R_t : (x_{R_t}, y_{R_t}, -H)$; the left-bottom point, as $L_b : (x_{L_b}, y_{L_b}, -H)$; the right-bottom point, as $R_b : (x_{R_b}, y_{R_b}, -H)$. In a similar fashion, the corresponding positions for the camera coordinates and the corresponding onscreen positions are set as: $L_t : (w_{L_t}, h_{L_t})$, $R_t : (w_{R_t}, h_{R_t})$, $L_b : (w_{L_b}, h_{L_b})$, $R_b : (w_{R_b}, h_{R_b})$. The distance of the points along the rail thus becomes

$$y_t - y_b = H \left(\frac{f \cos \theta_x - h_b \sin \theta_x}{f \sin \theta_x + h_b \cos \theta_x} - \frac{f \cos \theta_x - h_t \sin \theta_x}{f \sin \theta_x + h_t \cos \theta_x} \right) \quad (7)$$

Further, given the fact that the left and right rails are 1.067m apart, $\cos \theta_x$ and f are obtained. Substituting these into the above equation yields $y_t - y_b$. After normalizing the four points, the ratio between x- and y-oriented lengths becomes $(y_t - y_b)/1.067$. Therefore, the four onscreen points are converted as follows, yielding an aerial view of the image.

$$(w_{L_t}, h_{L_t}) \rightarrow \left(w_{L_b}, h_b + \frac{h_t - h_b}{1.067} (w_{R_b} - w_{L_b}) \right) \quad (8)$$

$$(w_{R_t}, h_{R_t}) \rightarrow \left(w_{R_b}, h_b + \frac{h_t - h_b}{1.067} (w_{R_b} - w_{L_b}) \right) \quad (9)$$

$$(w_{L_b}, h_b) \rightarrow (w_{L_b}, h_b) \quad (10)$$

$$(w_{R_b}, h_b) \rightarrow (w_{R_b}, h_b) \quad (11)$$

Figure 2 shows an aerial conversion of an image of the forward looking view from the cab. From Fig. 2, it can be seen that the ratio between the apparent and actual distances of the rail level of the track has been corrected.

Fig. 2 On-site location after aerial view processing

4.2 Image frame/pixel-level speed calculation

After aerial conversion, the travel speed of the target object is measured with a unit of pixels. This process is described below. A scope of the movement is defined for the image, and the speed is measured as the rate of travel per frame of the video footage. Several ranges are measured in order to minimize the margin of error; after calculating several rates of travel, the ones exhibiting clear calculation errors are removed from the calculation. Pattern matching was used to keep rates of error for each scope area to a bare minimum. Of the scope areas measured, the scope exhibiting the smallest margin of error was used as a basis for computing the rate of travel. Values exceeding the threshold for the train's travel orientation or maximum speed were deemed errors and thus excluded. The value

Fig. 3 Speed calculation

obtained from the calculation of the average of the rates of travel meeting these parameters was treated as the speed per single still frame. In cases where the all the rates of travel for each scope area in a frame were outside the allowable threshold, the rate of travel for the previous frame was used as the speed instead. A program based on these parameters yielded the following speed measurements for a single-station distance as shown in Fig. 3:

4.3 Image frame/pixel-level distance calculation

Integrating the speeds for each frame then yields a range of distance data expressed in units of pixels. Distance was defined from where the footage started at 0, to obtain a distance over t frames by aggregating the speeds. Computing the distance via speed, yielded the following as shown in Fig. 4:

Fig. 4 Frame/distance (in pixels) calculation

4.4 Creation of reference data on equipment and frames

In order to correlate the image frames of the forward looking view from the cab with kilometer markers on the track, a reference data table was created based on pieces of equipment with known kilometer marker locations (e.g. level crossings) and the frame number.

4.5 Creation of image frame/kilometer mark data

Using the image frame/pixel-level distance calculation data and the table of equipment locations and frames above, the image frame/pixel-level distance data is variously expanded and compressed over the intervals shown in the reference table, creating a new set of image frame/

Fig. 5 Image frame/kilometer mark data

kilometer mark data. Distances were converted according to the table by using the ratio between the distance in pixels and the actual distance displayed on the reference table. The image frame/pixel-level distance data and ambient equipment/frame data yielded a new frame/actual distance (by kilometer marker) table, as shown in Fig. 5.

5. Detection of the rail position from images of the forward looking view from the cabs and coordinate conversion

The rail position extracted from the image-based coordinates derived from the image of the forward looking view from the cab is next converted to a train-based coordinate system. The process of extracting the position of the rails and converting this into a set of coordinates consists of the following steps: detection of the position of the rails, conversion into a set of coordinates, extraction of the track center line, correction, and creation of image frame/rail position data. The rails are first extracted from the image and the center line is then obtained based on the position of the rails. This position is converted from an image-based coordinate system to the train-based coordinate system, which is based on the actual space, and the value is stored.

In order to convert the rail position from the aforementioned image-based system to the train-based one, several nodal points are defined along the extracted rail position; these points are then converted to the train-based coordinates and the center line is obtained.

The function for the aforementioned conversion is set as F , with the inverse function being F^{-1} . The difference between the maximum and the minimum heights of the rail is so minuscule as to be safe to ignore, so the rail height is assumed to be level.

Starting from the bottom of the screen and working up vertically, nodal point values (0, 1...) are set along the rails. Using the function F , these values are converted to positions in the train-based coordinate system. In the

Fig. 6 Creation of nodal point data for onscreen rails and actual rails

Fig. 6, r corresponds to the positions in the image-based coordinate system, while R corresponds to the positions in the train-based system.

6. Displaying wayside signals on the image of the forward looking view from the cab

Next, displaying wayside signals at the appropriate locations on the image of the forward looking view from the cab requires calculating the size of the wayside signal image to be used.

6.1 Onscreen wayside signal size calculation

The data processing routines described in Section 5 solely deal with wayside signal positions. The next step involves determining in what size the wayside signals should be displayed at the desired onscreen positions. Wayside signal image data are preloaded with the parameters of pixel size and actual wayside signal size relative to the screen size. The system calculates the size on the screen using these, scaling the wayside signal pixels. In addition, the baseline position of the wayside signals is based on the center of the wayside signal image. Moreover, a maximum boundary is defined; values exceeding that boundary are deemed too far away for visual confirmation, so they are not displayed. A minimum boundary is then set. It was specified that, provided the value was within the max/min threshold, displayed images would not have less than one pixel. In order to ultimately display the wayside signal onscreen, the original pixel size is expanded or compressed until it reaches a suitable size.

6.2 Displaying wayside signals at desired positions on the image of the forward looking view from the cab

Scaling the images, as described in the prior paragraph, and replacing the pixel size at the position obtained through the routines outlined in Section 5, enables the simulator to display the wayside signals at the desired locations and in the correct proportional size. The system then overlays wayside signal images onto the image of the forward looking view from the cab irrespective of what the scene actually depicts, and therefore, there is no means of depicting wayside signals occluded by trees or other real-world obstructions.

7. Testing the validity of this simulation method

Using the actual scene from the driver's cab and the hypothetical wayside signal positions created therein, the validity of the simulation technique was tested. The image frame/kilometer mark data, image frame/track center line position data, image of the forward looking view from the cabs with overlaid wayside signals, and other results of data processing were exported as files. At the trial, video footage with the following characteristics was used: about 11km of travel distance, 20 minutes of aggregate film time, and an image size of 720 x 480 pixels.

To confirm the validity of the image of the forward looking view from the cab and kilometer marker matching performance, speed measurements were exported along with actual distance measurements, as shown in Fig. 7, and Fig. 8 respectively. These results made it possible to

Fig. 7 Results of frame/speed measurement

Fig. 8 Results of frame/actual distance calculation

Fig. 9 Displaying a wayside signal on the image of the forward looking view from the cab

obtain kilometer marker position data corresponding to each frame of the video by means of the forward looking view video footage and reference data created. In this program, the accuracy of speed measurement affects the margin of error of the output data, so further improvements of speed measurement fidelity will yield more accurate output. In addition, the more ambient equipment is included in the reference data table, the more accurate kilometer marker data can be obtained, but the discovery of ambient equipment must be done by watching the live footage and entering it into a table, so the process takes considerable time. Creating reference data more efficiently will be an issue requiring further consideration in the future.

To test the wayside signals overlaid on the image of the forward looking view from the cab, forward looking view video footage was used along with frame/kilometer

marker data, frame/track center nodal point data, wayside signal position data, wayside signal image data, and camera parameter data to export final images showing wayside signal positions for each frame, as shown below. Confirmation was obtained that the wayside signal positions in each frame were correctly converted from the rail-based coordinates to the corresponding screen-based coordinates (Fig. 9).

This demonstrates that the simulator is able to run a simulation displaying potential wayside signal installation points and can be a means to determine where ultimately signals should be installed on-site.

Further, because this simulator was designed with minimum control functionality and user-friendliness, improvement on these fronts should be explored in the future.

In addition, as mentioned above, this system cannot depict situations in which wayside signals are occluded by trees or other objects. Accurately depicting such situations would prove to be highly difficult given the current input parameters used for this system.

8. Conclusions

Wayside signal installations are normally subject to plans prepared several years in advance and for work to be spread over a certain period of time. At the same time, there is an urgent need to improve the safety of level crossings, and as a result, the introduction of obstruction warning signals has been rushed in Japan.

The present simulation technique was developed to be able to display a simulated driver's view of wayside signals along a railway line, with wayside signal images overlaid onto images depicting the forward looking view from the driver's cab. The simulation was developed in such a way that wayside signal positions could be confirmed and adjusted, thereby making the process of installing real-world wayside signals easier and more efficient. The program's performance was then tested, and results confirmed that it has more than adequate functionality for practical use.

References

- [1] Nagamine, N., Ukai, M., "A Visual Simulation Method for Considering Installation Site of Wayside Facilities using Video Sequences from the Train Cab," *IEEE Transactions on Industry Applications*, Vol. 136, No. 2, pp. 134-144, 2016 (in Japanese).
- [2] Nagamine, N., Ukai, M., "The Simulation Of An Installation Position Of Wayside Signals Using Video Sequences From The Train Cab," presented at 14th International Conference on Railway Engineering Design and Operation (COMRAIL/14), Rome, Italy, June 24-26, 2014.
- [3] Nagamine, N., Aida, M. et al., "Visibility Checking System and Installation Support for Obstruction Warning Signals," *RTRI Report*, Vol. 30, No. 1, pp. 17-22, 2016 (in Japanese).

Authors

Nozomi NAGAMINE, Ph.D.
Assistant Senior Researcher, Signalling
Systems Laboratory, Signalling and Transport
Information Technology Division
Research Areas: Image Processing, Signalling
Systems

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

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