# RECOGNITION THE INSTALLATION PLACE OF SIGNAL EQUIPMENT USING ONE HANDY CAMERA

HIROKI MUKOJIMA & NOZOMI NAGAMINE
Railway Technical Research Institute, Japan

## ABSTRACT

Signalling and communications facilities in the railway are installed not in one place but scattered out along track-side between adjacent stations. For this reason, a great deal of labour is currently required in maintenance work for performing individual inspections, and in facility management work for ascertaining the installed positions and their types. For example, when repairing or improving for signal equipment, we have to update a database such as a management ledger based on the drawings. However, since the workers manually update the ledgers, there is a concern that input or deletion omission possibly occurs. In order to reduce human errors and the workload in maintenance, there is a requirement for a system that can automatically recognize and inspect the equipment without going to the site. Although there are methods to grasp the position and state of the equipment using distinctive sensors such as a LiDAR sensor and a stereo camera, it is necessary to prepare a dedicated vehicle, expensive sensors, or both. Therefore, we are developing a system that supports the maintenance work of signal equipment using only a handy camera. To use the system, all you need is a camera and a camera mount, such as a tripod. Our system is that assists ledger management by recognizing signal and communication equipment from the video obtained by the handy camera and estimating the location of the equipment. This paper describes the outline of our system and the fundamental elemental technologies for building it.

*Keywords: deep learning, handy camera, image processing, maintenance, signal equipment.*

## 1 INTRODUCTION

In the railway, due to the nature of its role, some pieces of equipment related to signal communication are not installed in one place spatially but are distributed and installed along track-side between adjacent stations. As a result, many workers are now working for maintenance, such as individual inspection for each facility, management work, recording the position, and type of facilities in the entire line section.

For example, when repairing or improving signal equipment, a database such as a facility management ledger is updated based on the drawings. However, if manual updating is performed based on the drawings on paper, there might be possible omissions in input or deletion. Also, in the work of construction design, it is necessary to consider how to arrange a cable route and calculate the number of troughs to be opened and closed back on the route. Such detailed information necessary for studying the design is often not registered in the database, so the designer has to go to the site and check it.

In order to reduce human error and workload in the maintenance work described above, it is required that a system can automatically recognize equipment and consider construction design without going to the site. For such a system, distinctive sensors are often used, such as a LiDAR sensor that irradiates a laser beam to sense the position of the object and a stereo camera that obtains three-dimensional information of the object based on the parallax of the image. Several methods for grasping the location and condition of equipment have been proposed [1]–[3]. However, because they use a dedicated vehicle, expensive sensors, or both, the cost may be awkward to use in rural areas.

© 2020 WIT Press, www.witpress.com
ISSN: 2058-8305 (paper format), ISSN: 2058-8313 (online), http://www.witpress.com/journals
DOI: 10.2495/TDI-V4-N3-264-273

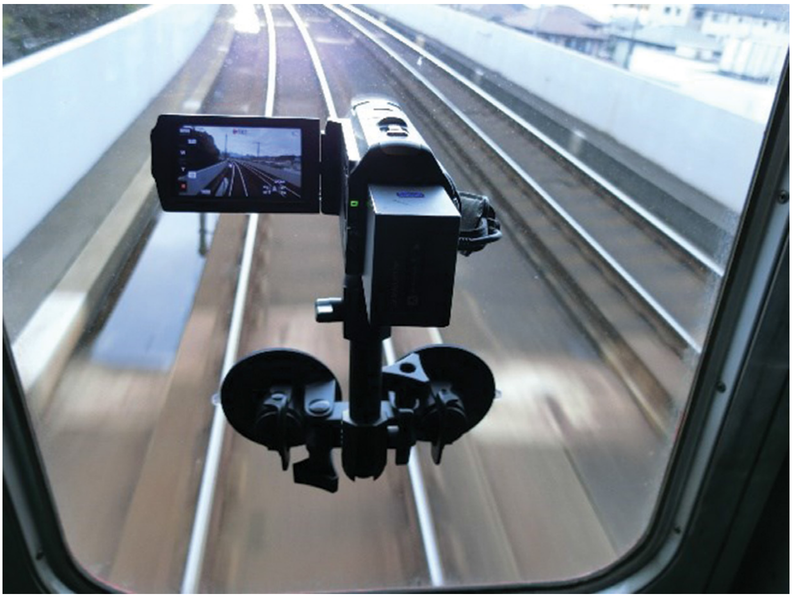


Figure 1: Shooting a video sequences on the front of commercial train *(source: Nagamine et al., 2020) [5].*

Therefore, we are developing a system that supports maintenance work for signal equipment using only one commercially available handy camera [4], [5]. To use our system, all users need is a video camera and a camera platform such as a tripod as hardware. The proposed system is effective for line sections, where the scale is small, and expensive systems cannot obtain the merit of introduction.

This system recognizes signal communication equipment from images of the front of the train acquired by a handy camera and estimates the location of the equipment to assist in checking ledgers and drawings. In this paper, we describe the outline of the system and element techniques for constructing it. Our method converts front images into birds-eye view images, estimates the distance for each frame, and recognizes names of some pieces of equipment from the forward images.

## 2 SYSTEM OVERVIEW

The proposed system requires only one handy camera. The user places the handy camera in a position to capture a picture from the front of a train in the same way as the driver sees. For installation, use a camera mount that can be installed on the front glass with a suction cup, etc., and fix it so as to be stable. Figure 1 shows an example of an installation.

Figure 2 shows the flow of our system process after shooting. The captured view image files from train cab are processed by a data generation unit to generate data used in our viewer application. The data is input to the viewer application together with the images, and the equipment information is displayed. The data generation unit consists of three processes using the cab view images: a unit that generates top view images, a unit that estimates the shooting position of each frame, and a unit that extracts the signal equipment.

The top view image generation unit takes each frame of the images as an input and converts the region of the track in it into top view images. In a front image, things in the top part of

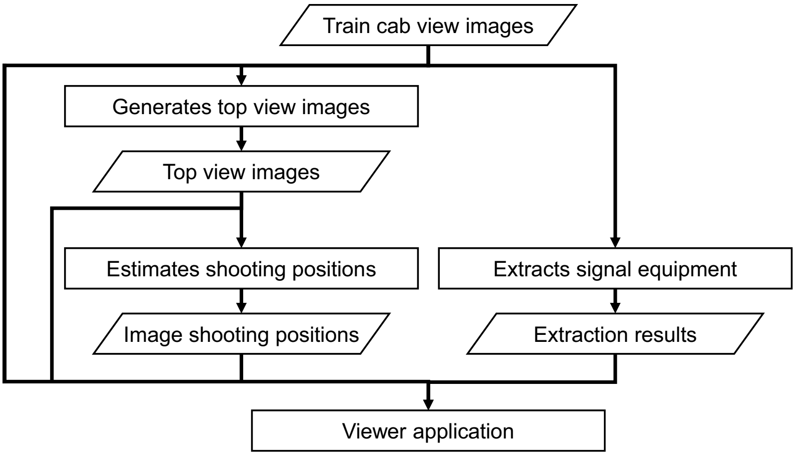


Figure 2: Processing flow of our system.

the picture means far from the train, which makes the pixels for the rail width smaller in the top part. However, by converting to a top image, the pixel size becomes the same regardless of the distance, making it easy to handle when measuring distance.

For the unit that estimates the distance from cab view images, the flow speed of each frame is measured from the top view images, and the amount of pixels movement concerning the frame is estimated. Then, by integrating the estimated velocities, the cumulative movement amount for each frame is estimated. The shooting position of each frame is estimated by dividing the cumulative movement amount of pixels concerning the total travelled distance. Since the actual distance concerning the moving amount of the pixels is computed, the speed for each frame can be estimated at the same time.

The signal equipment recognition unit detects the signal equipment by a deep learning from each frame of a video. It also recognizes the installation distance of the equipment by using the position found by the distance estimation unit.

The details of the processing unit and the viewer application will be described below.

## 3 GENERATING A TOP VIEW IMAGE

In the moving images taken towards the front of the train, the objects far from the camera are taken small and the near objects become large. For example, as shown in Fig. 3, the sizes of the rail width vary depending on the distance from the camera although they should be the same. It is not intuitive for users to measure a distance between facilities in an image where the actual scales change depending on the object positions. Therefore, for ease of handling, a top view image of the track plane is generated so that the relationship between the pixel and the actual scale is the same at any position in the image.

A top view image can be generated by using projective transformation. The projective transformation is represented by a matrix that transforms coordinates $\left(x, y\right)$ into coordinates $\left(x', y'\right)$ as follows.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} \sim \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{1}$$

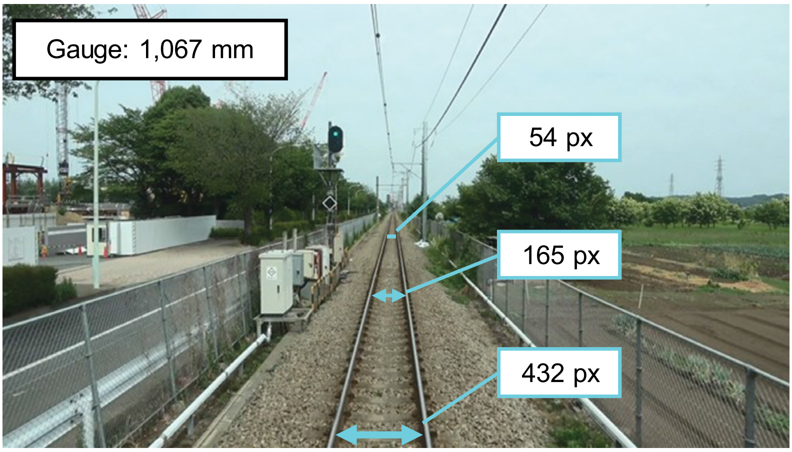


Figure 3: Example of difference in gauge on a train cab view image.

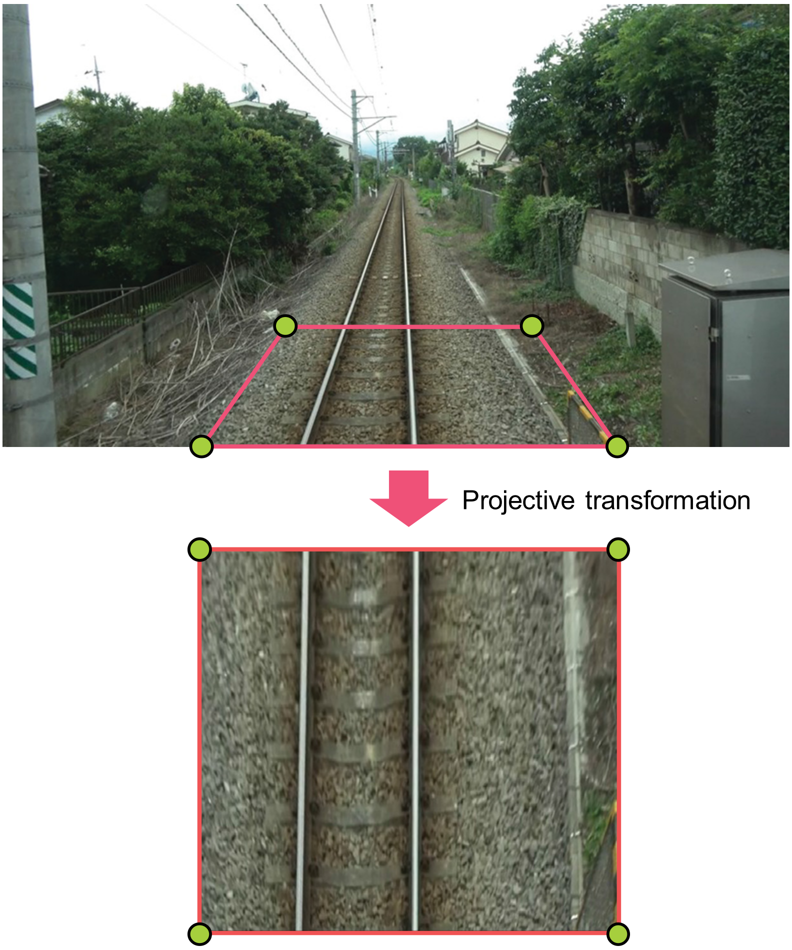


Figure 4: Projective transformation of trapezoid to rectangle.

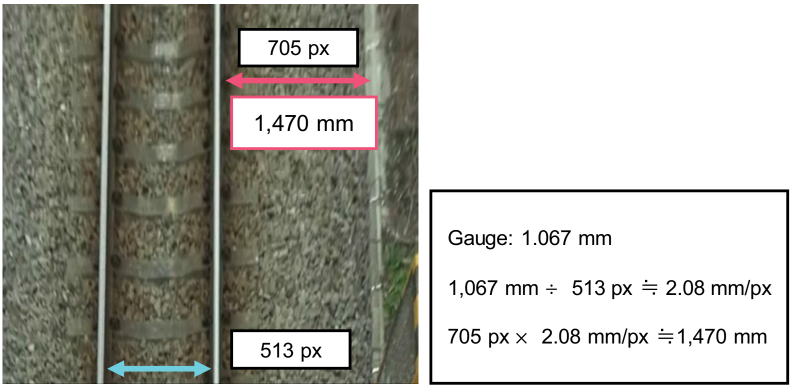

705 px
1,470 mm
513 px
Gauge: 1.067 mm
1,067 mm ÷ 513 px ≒ 2.08 mm/px
705 px × 2.08 mm/px ≒ 1,470 mm

Figure 5: Example of distance measurement on a top view image.

In Eqn. (1), the number of effective parameters is eight, so the transformation matrix can be found if there are correspondences of four sets of points with $h_{33} = 1$.

As described above, four sets of corresponding points are required to convert the track plane, which has been transformed into a trapezoidal shape, into a top view by the projective transformation. Therefore, it is important that a method of correctly selecting four points from a cab view image. Assuming that the camera is mounted at a roll angle of 0° with respect to the vertical direction, the four points forming the trapezoid shown in Fig. 4 become a rectangle when viewed from the top view because of the condition that the rails are parallel in the straight line. After that, the transformation matrix can be obtained by setting the size of the transformed rectangle so that railway sleepers are transformed with the correct length and breadth ratio. The distance can be measured as shown in Fig. 5 by converting the scale of pixel into actual distance using the fact that the gauge is 1,067 mm on a top view image.

Since the projective transformation matrix changes depending on the relationship between the camera mounting position and the track plane, it must be calculated every time when the camera installation conditions change. However, if the user does not move the camera installation position in one video taken, since the positional relationship with the track plane will not change, the transformation matrix at any one frame in the video can be use.

## 4 DISTANCE ESTIMATION FROM VIDEO

We estimate the speed and travel distance of the train, from the video using the method [3] [4] [6] that we have proposed so far to estimate the speed and position of the train from the video in front of the train. The amount of pixel movement, ‘hereafter, optical flow’, is calculated between each frame of the moving image. The actual distance per pixel movement is estimated from the calculated cumulative movement amount of pixels and the travelling distance of the train. After that, by applying the optical flow of each frame, the estimated actual distance per pixel is estimated as the speed and the shooting position in each frame.

### 4.1 Optical flow estimation

In order to estimate the train speed and position from train cab images, our method estimates optical flow in each frame. If cab images are used as they are, the correspondence between the pixels and the actual scale changes depending on the position on the image, as described

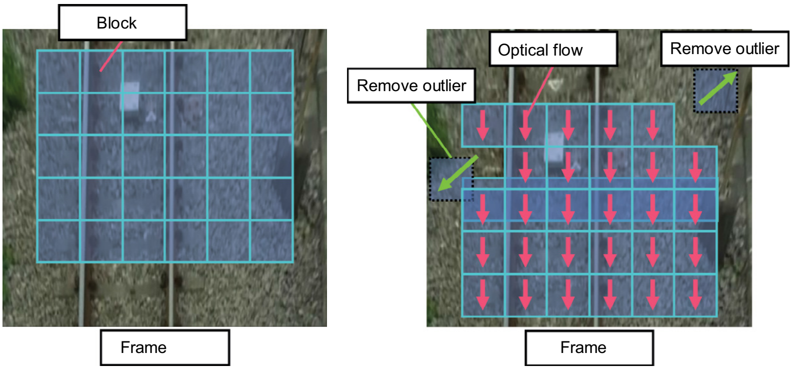


Figure 6: Concept of estimation method of optical flow.

in Chapter 3, we use the top view image generated in Chapter 3, where, the optical flow is obtained by the block matching method. Then, the outliers are removed from each block. The concept of processing is shown in Fig. 6. If the image sequence of a moving image is expressed as frame *n* (*n* = 1,2,3, ..., *N*) and the movement vector of the pixel in block $i$ between frames $T-1$ and is expressed as $\boldsymbol{d}_t^i$, a set $C_t$ of movement vectors within the range of thresholds $th_{\min}$ and $th_{\max}$ is obtained as follows.

$$C_t = \left\{ \boldsymbol{d}_t^i | th_{\min} < \| \boldsymbol{d}_t^i \| < th_{\max} \right\}. \tag{2}$$

Also, using the mean value $\text{Mean}(C_t)$ of the set $C_t$ and the standard deviation $\text{Std}(C_t)$, the elements of $C_t$ are standardized, and the set $C_t'$ whose standardized value is less than the threshold $th_s$ is calculated as follows.

$$\hat{d}_i(t) = \frac{\| \boldsymbol{d}_t^i \| - \text{Mean}(C_t)}{\text{Std}(C_t)}. \tag{3}$$

$$C_t' = \{ \boldsymbol{d}_t^i \, | \, | \hat{d}_i(\text{t}) | < th_s, \boldsymbol{d}_t^i \in C_t \}. \tag{4}$$

Finally, the pixel movement amount $d(t)$ between the frames $t-1$ and $t$ is calculated as follows.

$$d(t) = \text{Mean}(C_t') \tag{5}$$

Here, $d(1) = 0$ is set because calculation cannot be performed when calculation $t = 1$, which is the first frame. Also, note that the unit of $d(t)$ is pixels per frames. Figure 7 shows an example of the result of obtaining $d(t)$ for each frame.

## 4.2 Estimation of frame shooting position using optical flow

The shooting position of each frame is estimated using computed optical flow. Users provide reference positions for one or more sections of cab view images. For example, the locations of two stations. The reference positions given by the users are $r_a$ [km] for the $a$ th frame captured at point *A* and $r_b$ [km] for the $b$ th frame captured at point *B*. At this time, the travel

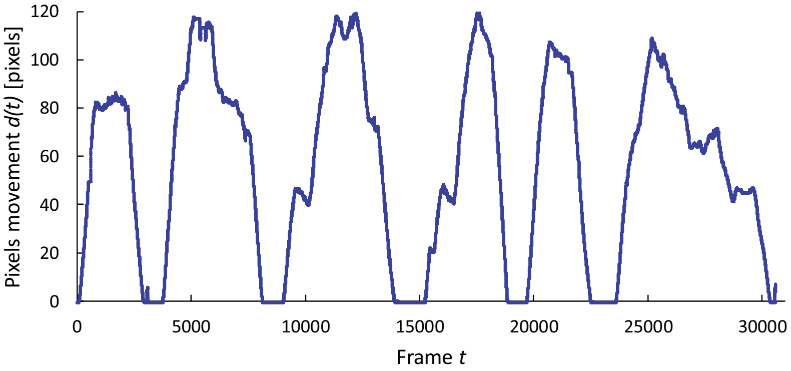


Figure 7: An example of the result of obtaining pixel movement in each frame.

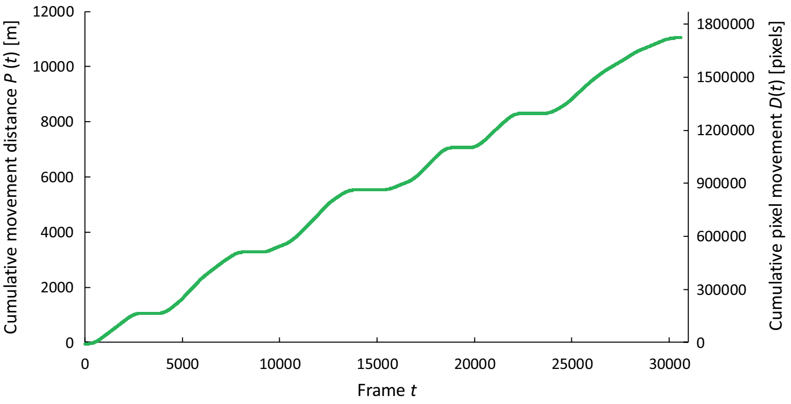


Figure 8: An example of calculating the cumulative pixel movement amount and cumulative movement distance in each frame.

distance between $A$ and $B$ is equal to $|r_b - r_a|$, and the position in any other frame is estimated by linear interpolation. When the cumulative amount of pixel movement in the $t$ th frame is $D(t)$, the shooting position $P(t)$ for the $t$th frame can be obtained as follows.

$$P(t) = r_a + \frac{r_b - r_a}{D(b) - D(a)}\big(D(t) - D(a)\big). \tag{6}$$

$$D(x) = \sum_{i=1}^{x} d(i). \tag{7}$$

Figure 8 shows an example of calculating the cumulative pixel movement amount $D\,(t)$ and cumulative movement distance $P(t)$ in each frame. Since the pixel and the actual distance can be converted using Eqn. (6), the train speed can also be calculated from the frame rate and the amount of pixel movement between frames.

## 5 SIGNAL EQUIPMENT RECOGNITION AND POSITION ESTIMATION

### 5.1 Signal equipment recognition by deep learning

Our system uses Yolov3 [7], an object recognition method based on deep learning, to detect signal equipment from train cab view images. We manually annotated 18 classes of railway equipment shown in Table 1 and Fig. 9 and conducted transfer learning against the pre-learning model. The number of added annotation data is 2,782 frames.

### 5.2 Estimation of equipment position

The shooting position obtained in Chapter 4 is added to each frame of cab view images. Installation locations of the equipment are estimated from this information. Once the equipment appears in a frame, it appears in successive frames, so it is necessary to decide which frame position to use. Therefore, the detection target is tracked between consecutive frames, and the frame whose -coordinate is closest to the 20% position from the bottom edge of the image is adopted, and the frame shooting position is used as the equipment position.

## 6 VIEWER APPLICATION

Our system integrates and displays the data obtained in Chapters 3–5. The camera used for the system is Sony FDR-AX55. The shutter speed was set to about 1/500 to 1/1,000 s as a shooting parameter, and 3,840 × 2,160/30 fps progressive was set at XAVC 4K (100 Mbps) as a parameter for determining image quality. By setting the shutter speed low, the blurring of the track surface can be reduced. In addition, the references are the position of the station at the start and end of shooting.

Figure 10 shows the operation screen of the system. The run curve and running position curve for the frontal image are displayed, and the track bar displays the current frame position. The converted track image is displayed on the upper right of the screen. Also, the equipment list is displayed in another window. The user can jump to the frame where the target equipment is displayed by clicking the list.

Table 1: Railway equipment classes.

| No. | Class | No. | Class |
|---|---|---|---|
| 1 | Signal | 10 | Emergency button |
| 2 | Ground coil | 11 | Signal call position marker |
| 3 | Emergency train stop warning light | 12 | Kilometres post |
| 4 | Speed limit sign | 13 | Gradient post |
| 5 | Obstruction warning signal | 14 | Electrification mast |
| 6 | Stop marker | 15 | Level crossing |
| 7 | Crossing gate | 16 | Impedance bond |
| 8 | Location box | 17 | Level crossing warning light |
| 9 | Point machine | 18 | Track-side telephone |

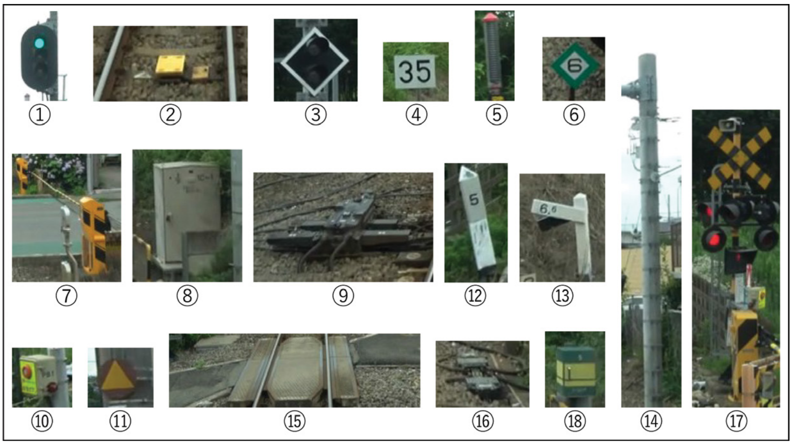
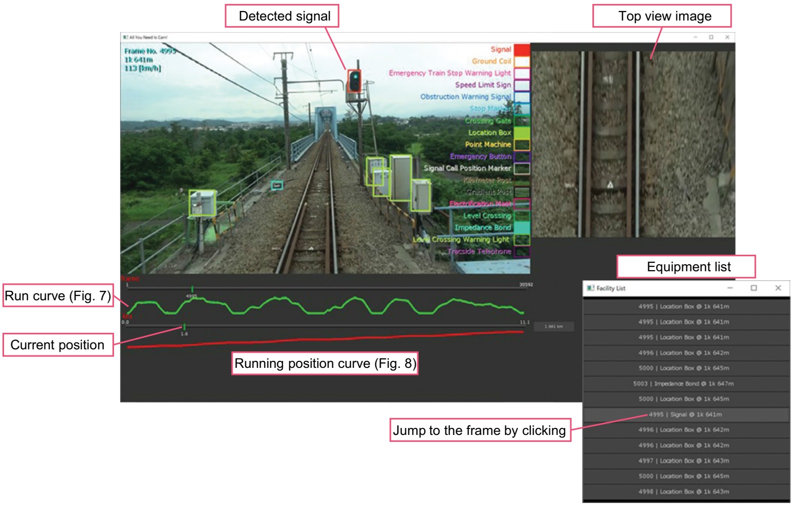


Figure 9: Image example of each equipment corresponding to the numbers in Table 1.



4995 | Location Box @ 1k 641m
4995 | Location Box @ 1k 641m
4995 | Location Box @ 1k 641m
4996 | Location Box @ 1k 642m
5000 | Location Box @ 1k 645m
5003 | Impedance Bond @ 1k 647m
5000 | Location Box @ 1k 645m
4995 | Signal @ 1k 641m
4996 | Location Box @ 1k 642m
4996 | Location Box @ 1k 642m
4997 | Location Box @ 1k 643m
5000 | Location Box @ 1k 645m
4998 | Location Box @ 1k 643m

Figure 10: Our system screens.

## 7 CONCLUSION

As an inexpensive system to reduce the maintenance of signal and communication equipment, we propose a system that uses a handy camera and image processing. It is applied to the train cab view images on the actual track to confirm the system operation, and it is confirmed that the system as whole works without problems.

In the future, it will be applied to various line sections and it will be confirmed that the system will operate normally regardless of the line sections. In addition, the system will be configured to improve the accuracy of position estimation, add learning classes, and improve the accuracy of equipment position estimation. With respect to position estimation, features such as kilometres posts and gradient posts whose positions are precise can be extracted as equipment. Therefore, it is possible to apply this information to position estimation from cab view images to improve accuracy. As a functional addition to the system, a facility ledger output function based on the extracted facility information and a simple drawing generation function will be added.

The system we ultimately aim to learn the changes in the appearance of each equipment and estimate the degradation level of equipment. It is planned to be useful for the appearance inspection in which a person visually judges the degradation.

## REFERENCES

[1] Leslar, M., Perry, G. and McNease, K., Using Mobile LIDAR to Survey a Railway line for Asset Inventory. *Proceedings of the ASPRS 2010 Annual Conference*, pp. 26–30, 2010.

[2] Arastounia, M., Automated recognition of railroad infrastructure in rural areas from LIDAR data. *Remote Sensing*, **7**(11), pp. 14916–14938, 2015.

[3] Harmsen, F., Hintze, P. & Elstner, J., What, where, when, why? – automated capture of railway infrastructure data. *Signalling & Datacommunication,* **111**(12), 2019.

[4] Nagamine, N. & Mukojima, H., Signal equipment recognition method from camcorder video sequences, *The papers of Technical Meeting on "Transportation and Electric Railway", IEE Japan, TER-20-25*, pp. 121–126, 2020 (in Japanese).

[5] Nagamine, N. & Mukojima, H., Generation Method of Continuous Bird's-eye View Image from Camcorder Video, *The papers of Technical Meeting on "Transportation and Electric Railway", IEE Japan, TER-20-56*, pp. 121–126, 2020 (in Japanese).

[6] Nagamine, N. & Ukai, M., The simulation of an installation position of wayside signals using video sequences from the train cab, *WIT Transactions on The Built Environment*, vol. 155, WIT Press: Southampton and Boston, pp. 97–109, 2015.

[7] Redmon, J. & Farhadi, A., YOLOv3: an incremental improvement, *arXiv preprint arXiv:1804.02767*, 2018.