Passage detection of a train via a reference point

Lubos Rejfek[1[0000-0001-6811-7692]], Jan Pidanic[1[0000-0003-1948-3818]], Dominik Stursa[1[0000-0002-2324-162X]], Tan N. Nguyen[3[0000-0002-2286-6652]], Phuong T. Tran[2[0000-0002-1448-8882]], Zdenek Nemec[1[0000-0002-6640-8995]], and Tomas Zalabsky[1[0000-0003-2887-6955]]

[1] Faculty of Electrical Engineering and Informatics, University of Pardubice, nám. Čsl. Legií 595, Pardubice 530 02, Czech Republic
[2] Wireless Communications Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 729000, Vietnam
[3] Communication and Signal Processing Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam
lubos.rejfek@upce.cz

Abstract.
A reference point detection system for position validation of a mobile object was developed for verification of experiments. The detection is based on a classic image processing algorithm and a processing algorithm using neural networks. Both approaches are compared. High-precision concept of the system is based on a camera sensor and automatic processing of video frames for position evaluation. The designed system was tested on a real application proving correct operation.

Keywords: Reference Point, Neural Network, Object Detection, Least Squares Method.

1 Introduction

This paper describes a system for automatic monitoring of a train passage by the reference point. The requirements on the automatic development system were (1) the system will react to the specific part of the mobile object, (2) installation of the developed device will be small, portable, and non-invasive, and (3) sampling resolution must be better than 10 Hz. The described conditions exclude use of the usual systems working on the principles described in [1], [2], or [3]. It was decided to employ the automatic monitoring system based on the camera sensor. This system looks to be ideal because it meets all three conditions.

The proposed reference system is composed from a GPS antenna, camera-based sensor, SENSORAY Model 2253P (providing timing into the frames), and computer for the signal processing. A block diagram of the system is shown in **Fig. 1**. Description of the SENSORAY parameters and block diagram are in [4]. The manual shows, that an output signal from the camera must be PAL (25 fps, resolution 640x480) or NTSC (30 fps, resolution 720x576). Usable output video formats with timestamps from the

SENSORAY are avi video or MPEG-4. The PAL output is used in the final solution, the frequency 25 Hz meets the condition that the frequency must be more than 10 Hz. The part of the mobile object which must be detected is the antenna shown in **Fig. 2**, and detection of the antenna is described in chapter 2. The situation monitored by the camera is shown in **Fig. 3**. When the mobile object is in the monitored area, the algorithm detects if the antenna is in the video frame and calculates how many pixels from the reference point. This frame timestamp is recorded with information about the distance of the antenna from the reference point. The second proposed algorithm makes an analysis of the signal and picks up the times when the mobile object passed by the reference point, these algorithms are described in chapter 3. The results from the test of the system are described in chapter 4.

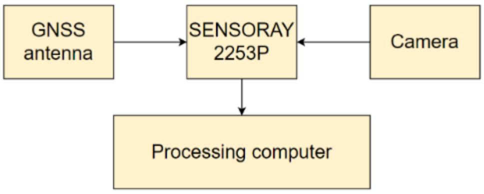


Fig. 1. Block diagram of the system.

Fig. 2. Train antenna employed for the optical detection when the object passes by the reference point.

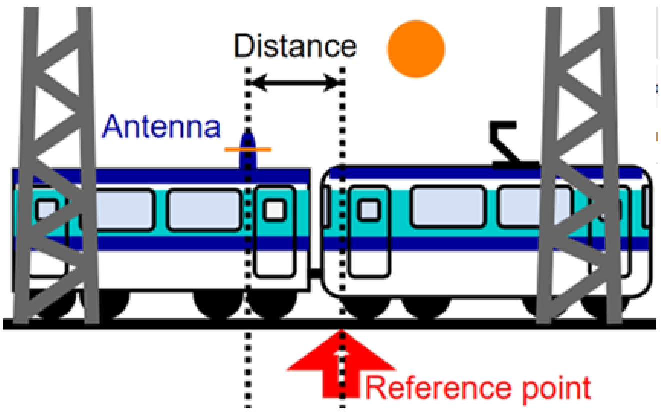


Fig. 3. Description of the situation in the video frame.

2 Antenna detection in the image

The main task is correct detection of the antenna in the image. More approaches can lead to a correct solution with different advantages and disadvantages. Two possible methods for the antenna detection and their comparisons are described below.

2.1 Image processing approach

The first step of the image processing is separation of the moving objects. This is realized as the difference between the background image and the actual frame. The background image is obtained as an average of the long sequence of frames. Small moving objects placed on the otherwise "periodic" train wagons are insignificant in this background image. The example of the differential image (actual frame minus background frame) is shown in **Fig. 4**. Two moving objects are visible in the preprocessed data; one of them is the antenna (in the top part), and the second is the mobile object which carried the antenna. The second step is detection of the objects in the frame and their classification (antenna/false alert).

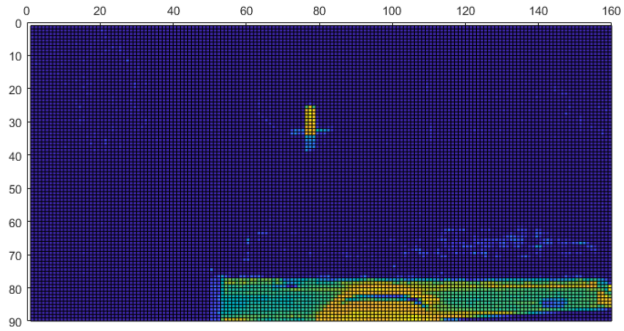


Fig. 4. Differential frame between empty frame (created as the average from more frames) and tested frame.

The middle part of the objects in the differential image can be obtained by applying the correlation function. The correlation between the differential image and template of the antenna is shown in **Fig. 5**. The equation used for the calculation of the element in the new matrix is (1), where *K* and *L* are dimensions of the window (in our case equal to the antenna template size), *AI* is the antenna image, and *DI* is the differential image. The color of pixels at the position detected by the correlation is compared with the antenna color, and if the colors match the antenna is detected. The next possible way is for the least squares method to be utilized, specifically using equation (2). The result of this method is shown in **Fig. 6**. The antenna is at the position with value near zero, and the positive sidelobes are around this zero value. The second moving object can be classified as any other object, because this object has an incorrect response after processing by (2). The problem of the image processing approach is the variable size of the antenna related to the camera distance from the reference point. This difference in the distance is caused by positioning of the system when the exact resulting distance between camera and reference point is not available and needs to be measured for every reconfiguration. Otherwise, processing must be repeated with different sizes of the antenna.

$$Y_{ij} = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} (DI_{kl} \cdot AI_{kl}) \tag{1}$$

$$Y_{ij} = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} (DI_{kl} - AI_{kl})^2 \tag{2}$$

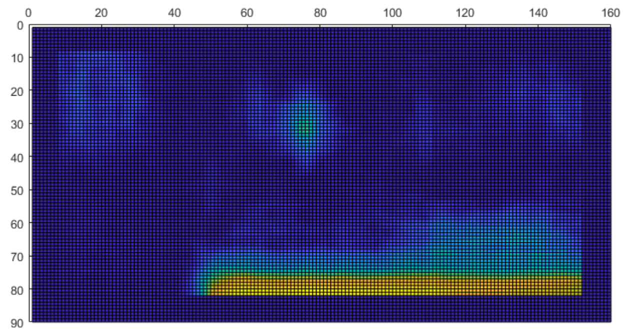


Fig. 5. Correlation of the differential frame with the image of the antenna.

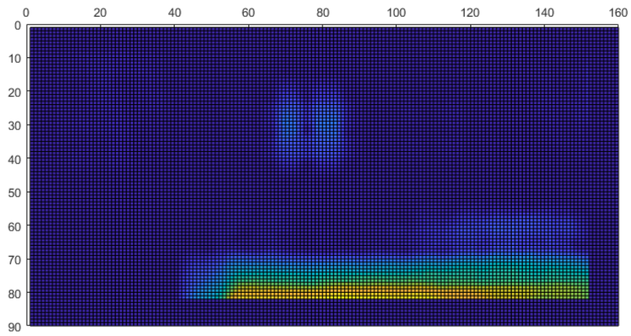


Fig. 6. Application of the least squares method for the antenna detection.

2.2 Neural network application approach

Usage of the neural network for object detection is described, for example, in [5], [6], [7], [8], or [9]. The articles show that the neural networks can be used for the introduced purpose. Neural networks are composed from the neurons, which are clustered to the layers. Description of the artificial neuron is described, for example, in [10]. The employed neural network in our application is described in **Fig. 7**. The designed neural network is composed from the input layer, the convolution layers, the pooling layer, the activation function layers, and the output layers for the classification (target/background). Description of the convolution function is in **Fig. 8**. The blue matrix shows input data, the orange matrix is filter mask, and the red matrix are processed data before application of the activation function. The input data are multiplied with the values in the filter matrix, and the sum of the multiplied numbers is sent to the activation function. Because of RGB image output, the input matrix will be 3D, it means that the convolution filter will be, in our case, also 3D. The channels are generated parallelly, which means that connection of every channel with the activation function layer is independent on the other channels. The principle of the convolution layer is also described in [11]. The matrix is split into the sub matrixes and the biggest value is taken from every sub matrix in the used max pooling layer. The principle of the pooling layers is described more in [12], where the average pooling layer is also described. The trainable parameters are in the convolution layers. The trainable weights are 3x3x3x64 in the first convolution layer, 3x3x64x64 in the second convolution layer, 4x4x64x64 in the transposed convolution layer, and 1x1x64x2 in the last convolution layer. The trainable biases are 1x1x64 in the first three convolution layers and 1x1x2 in the last convolution layer. The neural network has 104 450 trainable parameters in total. The used network is trained just for the detection of the antenna in the camera image. The examples of the data which were used for the neural network learning are shown in **Fig. 9**. The size of the templates is 32x32 pixels. The datasets were obtained on a flat landscape in central Europe. The dataset was composed from 1000 examplesand learning was realized over 300 epochs. An example of the antenna detection is shown in **Fig. 10**.

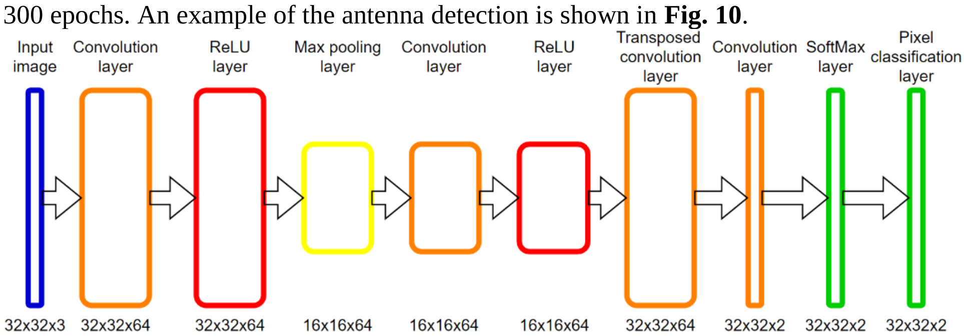


Fig. 7. Neural network for the detection of the antenna in the image.

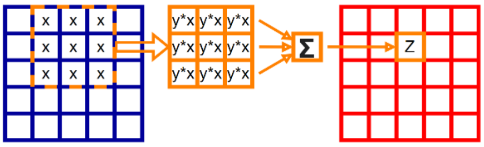


Fig. 8. Convolution function (blue – input image, orange – convolution mask, and red – array of the results).

Fig. 9. Examples of the data for the neural network training.

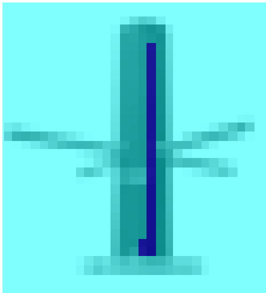


Fig. 10. Example of the automatic detection of the antenna.

Benefits of this approach are that the signal processing is easier for programming, and faster for the processing during the measurement. The distance between the camera and monitored area can be variable thanks the different sizes of the objects used for the neural network learning. After the comparison of the methods results, the neural network realization was found to be optimal for such a system.

3 Algorithm for passage detection

The final algorithm used in the system is shown in **Fig. 11**. The first step is installation and setting of the system. The second step is detection of the antenna in the image. The third step is comparison of the detected targets with the targets detected in the previous frame. The targets on the same positions in both frames are inserted into a mask. The fourth step is suppression of the static targets by the mask application. The fifth step is making the sum over the columns and removing targets composed from less than 2 points (noise suppression). If the values in the data line are only zeros, the target distance is set on the frame edge. If there are any non-zero values, the distance of the biggest value from the reference point is employed as the target distance and placed in the next position in the data line (this line is shown in **Fig. 13**). The algorithm can be continued with the next frame and after finished, it saves the line of the distances from the reference point.

The result is the graph shown in **Fig. 13**, and the passage time can be read from this graph. But this is not practical. The second algorithm which detects the local minimum values and marks the position in the data line, was made to overcome it. The first step filters the noise caused by the previous algorithm (the first observations of the static targets before their including to the mask). Next, the detected deviations are separated into the list of the events (number of the points in the list is equal to number of deviations caused when the object passes by the reference point). If any events are not detected by the algorithm, it can be finished as a blank measurement. When the values are zero, the frame shows the event when the object passes by the reference point. If no zero value is detected in the deviation, the object passed by the reference point between two video frames.

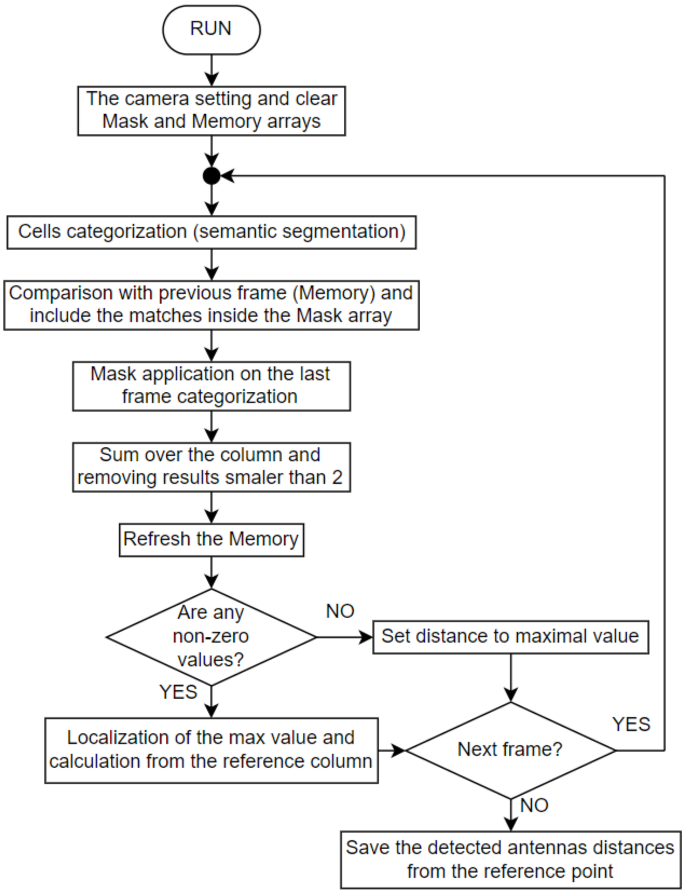


Fig. 11. Algorithm for creation of the data series.

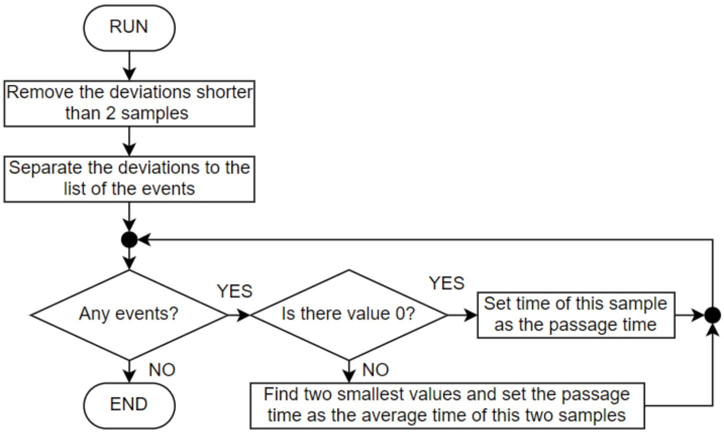
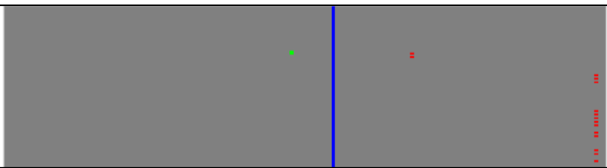
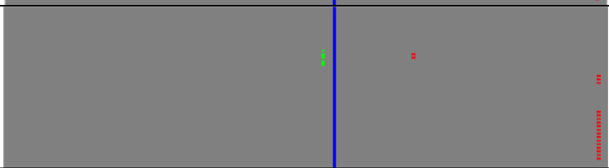
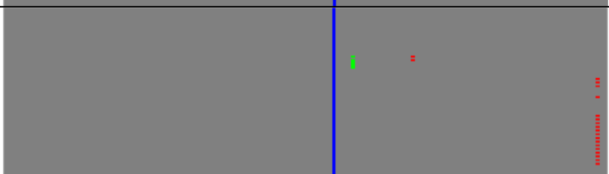
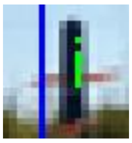
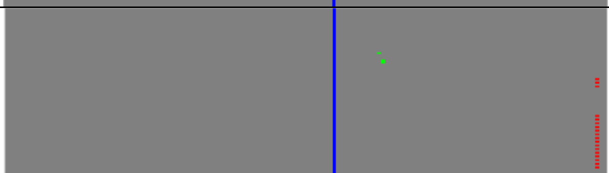
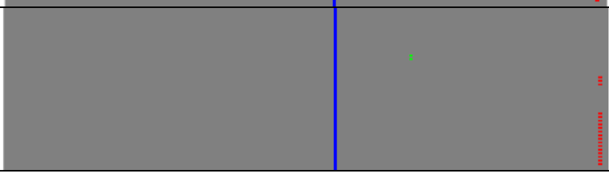


Fig. 12. Algorithm for extracting passes through the reference point.

4 Application of the algorithm on the real data

The algorithm was tested during the measurement when the train antenna was installed on a car trailer. The system was tested when the object moved in both directions. During the next tests, the distance between the reference point and the system was modified. One example of the output from the proposed first algorithm, described in **Fig. 11**, is shown in **Fig. 13**. We can see that the car passed the reference point twice. The frames 219 and 4029 show the cases when the object passed by the reference point. The classification of the frames, when the object passed by the reference point is shown in **Table 1**. The first column shows the frame number, the second column contains the frame classification, and the third column shows the antenna detected by the neural network with the fitting markers obtained from the system. The colors representing the pixel classification are gray (a background), blue (a reference point), green (the detected antenna), and red (a phantoms detected by the mask application).

Table 1. The reference point passage by the mobile object (monitored by designed system).

Frame number	Analyzed frame	Detected antenna
218		
219		
220		
221		
222		

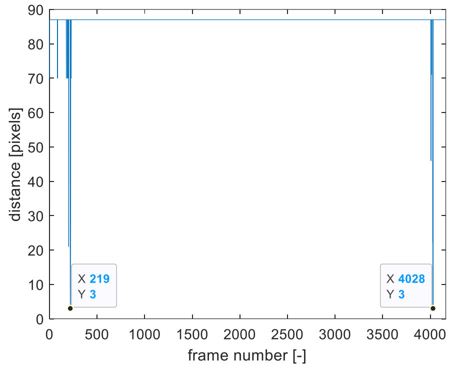


Fig. 13. Signal recorded during the measurement.

5 Conclusion

The paper describes a system developed for the detection of the reference point passage by a specific object.
The system detects the specific part of the mobile object and records the time when it passes by the reference point. Two approaches for the antenna detection (image processing and the neural network) were tested. The neural network was utilized in the final system because it worked much faster than the image processing-based concept. The neural network can be slower only in case, when just one size of the reference picture is used for the correlation, but the system must be installed in the exact distance from the reference point in this case. A ratio between the image processing approach and the neural network approach is 0.69 * n, where n is number of the used reference pictures of the antenna.
The algorithm described in the article suppresses the phantoms caused by the neural network and improves results of designed system.
The designed system was validated by real measurements and the results show that the system correctly detects when the object passes by the reference point and meets the conditions described in the introduction. The designed device is fully applicable for the intended purpose.

Acknowledgement

The work was supported by ERDF/ESF "Cooperation in Applied Research between the University of Pardubice and companies, in the Field of Positioning, Detection and Simulation Technology for Transport Systems (PosiTrans)" (No. CZ.02.1.01/0.0/0.0/17_049/0008394), thanks to Charles Hooper for the language proof reading.

References

1. DUBOVAN, J., FRNIAK, M., LITVIK, J., MARKOVIC, M. and DADO, M., 2019. Optical Sensors Utilization as a Part of Intelligent Transport Systems, 2019 Conference on Microwave Techniques, COMITE 2019 - Microwave and Radio Electronics Week, MAREW 2019 2019.
2. BEZOUSEK, P., HAJEK, M. and POLA, M., 2010. Effects of signal distortion in a FMCW radar on range resolution, Proceedings of 15th Conference Microwave Techniques, COMITE 2010 2010, pp. 113-116.
3. DANIŠOVIČ, P., ŠRÁMEK, J., HODOŇ, M. and HÚDIK, M., 2017. Testing measurements of airflow velocity in road tunnels, MATEC Web of Conferences 2017.
4. SENSORAY, USB Audio/Video Codec Model 2253 Hardware Manual version 1.2.1, published 2014, available from: http://www.sensoray.com/downloads/man_2253_hw_1.2.1.pdf
5. REDMON, J., FARHADI, A.: YOLO9000: Better, Faster, Stronger. Computing Research Repository (CoRR) (2016), http://arxiv.org/abs/1612.08242
6. ARRI, H., KHADKA, N., CHOPRA, V., KUMAR, A., KARMAKAR, S., Nayak, S., and Ospanova, A., 2022, Real Time Object Recognition with Voice Feedback using YOLOv4-tiny Deep Learning Model, Journal: Science & Education, DOI: 10.15680/IJIRSET.2022.1104137, pp.: 3932-3940.
7. KARMARKAR, R. R., HONMANE, V. N., 2021, OBJECT DETECTION SYSTEM FOR THE BLIND WITH VOICEGUIDANCE, International Journal of Engineering Applied Sciences and Technology, Issue 2, Vol. 6, pp.: 67-70 Doi.: 10.33564/IJEAST.2021.v06i02.013
8. WANG, L., LI, R., ZHANG, C., FANG, S., DUAN, C., MENG, X. and ATKINSON, P.M., 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS JPRS, 190, pp. 196-214.
9. ABDOLLAHI, A., PRADHAN, B., SHUKLA, N., CHAKRABORTY, S. and ALAMRI, A., 2021. Multi-object segmentation in complex urban scenes from high-resolution remote sensing data. Remote Sensing, 13(18),.
10. MENG, Z., HU, Y. and ANCEY, C., 2020. Using a data driven approach to predict waves generated by gravity driven mass flows. Water (Switzerland), 12(2),.
11. Chollet, F.: Deep learning with Python. Shelter Island, New York: Manning Publications Co., [2018]. ISBN 978-1617294433.
12. YINGGE, H., ALI, I. and LEE, K.-., 2020. Deep neural networks on chip - A survey, Proceedings - 2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020 2020, pp. 589-592.