ADVANCED THERMAL CAMERA BASED SYSTEM FOR OBJECT DETECTION ON RAIL TRACKS

by

Milan G. PAVLOVIĆ[a*], Ivan T. ĆIRIĆ[b], Danijela RISTIĆ-DURRANT[c], Vlastimir D. NIKOLIĆ[b], Miloš B. SIMONOVIĆ[b], Milica V. ĆIRIĆ[d], and Milan S. BANIĆ[b]

[a] College of Applied Technical Sciences, Nis, Serbia
[b] Faculty of Mechanical Engineering, University of Nis, Nis, Serbia
[c] Institute of Automation, University of Bremen, Bremen, Germany
[d] Faculty of Civil Engineering and Architecture, University of Nis, Nis, Serbia

Original scientific paper
https://doi.org/10.2298/TSCI18S5551P

In this paper, an advanced thermal camera-based system for detection of objects on rail tracks is presented. Developed system is powered by advanced image processing algorithm, in order to achieve greater reliability and robustness, and tested on set of infrared images captured at night conditions. The goal of this system is to detect objects on rail tracks and next to them and estimate distances between camera stand and detected objects. For that purpose, different edge detection methods are tested, and finally Canny edge detector is selected for rail track detection and for determination of region of interest, further used for analysis in object detection process. In determined region of interest, region-based segmentation is used for object detection. For estimation of distances between camera stand and detected objects, homography based method is used. Validation of estimated distances is done, in respect to real measured distances from camera stand to objects (humans) involved in experiment. Distances are estimated with a maximum error of 2%. System can provide reliable object detection, as well as distance estimation, and for improved robustness and adaptability, artificial intelligence tools can be used.

Key words: *thermal imaging, image processing, rail detection, edge detection, object detection, region-based segmentation, distance estimation, homography*

Introduction

Large capacity and characteristic infrastructure make the railway very important in the field of goods and passenger transport. Railway infrastructure is very complex, so many problems come in intersection with other types of traffic. In case of an intersection of railway, road and pedestrian traffic at railway level crossings, there are many possibilities for accidents, and those places are marked as a weak safety point. However, even more dangerous places are *unmarked crossings,* where pedestrians cross the rail tracks based on own habit and needs. Those places are completely unsafe for crossing because they are usually not quite visible, at night with poor illumination or completely without it, and there are no any marks, so the train driver does not expect any pedestrian there.

* Corresponding author, e-mail: milanpavl@gmail.com

On railway level crossings, cause of the accident can be consequence of malfunction of equipment – train or/and railway infrastructure, fault of the warning system (flashing lights, warning tones, and boom gates) on railway level crossings, *etc.* However, accidents can be caused by human – train/car/truck driver, pedestrians or other objects. In case of pedestrians, using of the railway level crossing despite the activated warning system or using of *unmarked crossings*, significantly increases risk for accidents with fatal end. From the train's driver point of view, the existence of any living or/and non-living objects on a level crossing, when a train is coming, represent an obstacle. There are many monitoring systems, which can detect objects near or at level crossing. Some of them are based on objects detection principle, while others are conventional methods which are based on optical principles (optical beam, ultrasonic 3-D laser radar) [1]. On the other side, there are image processing methods, which are based on conventional monochromatic, RGB cameras, and stereo camera [1]. The task of all those systems is to monitor the railway crossing and, in case of possible obstacle existence, inform the train driver and timely decrease possibilities for an accident. Some of those systems are dependent of different weather conditions, while the other activates false alarms caused by shadows, large drops of rain, or snowflake [1].

However, a common characteristic of all those systems is that they can operate only in the day and good-light conditions. Due to different illumination and weather conditions, as well as in low-light conditions, visibility of objects can be reduced, and those systems do not give satisfactory results of detection. In these very specific conditions, thermal imaging system can be used, because its operating range is in the invisible infrared region of the spectrum. There are no many fully developed methods for detection living and non-living objects, with thermal imaging systems [2]. For detection and tracking of objects, Shaik and Iftekharuddin [3] proposed algorithm based on frequency domain correlation and Bayesian probabilistic techniques, respectively. This algorithm showed good results for real-time detection and tracking of static or moving targets, but it is sensitive on background noise. On the other side, distinguishing the tire's thermal energy reflection area on the road surface from the other areas is used for detection of vehicles is presented in [4]. This method, in combination with previously developed method, gave high vehicle detection accuracies under various environmental conditions. However, thermal imaging systems are used also for human detection. In [5], applicability of local-feature based object detector for the case of people detection in thermal data is presented. In the task of person detection in different real-world scenarios, SURF feature detector and descriptor is adopted, and its performance is evaluated, and results showed how this local-feature based detector can be used to recognize specific object parts, *i. e.*, body parts of detected people. On the other side, for detection of humans in thermal images, a new method based on the shape context descriptor (SCD) with the Adaboost cascade classifier framework is presented in [6]. The SCD is used for local edge feature extraction, while boosting algorithms with a cascade framework is used for human classification. Results showed that shape context features with boosting classification provide a significant improvement on human detection in thermal images.

However, the complexity of object detection rises when thermal imaging system is mounted on moving platform or some vehicle. An approach to real-time human detection through processing video captured by an infrared (IR) camera mounted on the autonomous mobile platform is presented in [7]. Advanced control of mobile robot with goal to recognize human in indoor environment and allow adequate human-robot interaction, is presented in [8]. Its operation is based on intelligent advanced segmentation and classification of detected regions of interests in every frame acquired by IR camera. Furthermore, Bertozzi *et al.* [9] used

stereo system for the detection of pedestrians using two far-IR cameras mounted on test vehicle. In situations where pedestrians are very close to each other and at the same distance from the vision system, they are often detected as a single pedestrian, as well as, presence of objects with a similar size and shape of a pedestrian represent the most frequent cause of misdetection. However, a system with monocular IR camera mounted on test vehicle – car, presented in [10], is based on a multi-resolution localization of warm symmetrical objects with specific size and aspect ratio with goal to detect pedestrians. Experiments showed that the proposed system is able to detect one or more pedestrians, but only in the range of 7-43.5 m.

In the field of railways, there are no many uses of thermal imaging system in any purposes. However, Forthand and Zamjatnins [11] used thermal camera mounted on the train's roof and analyzed possible dangerous situation when trains operates at night and bad weather conditions, especially when some objects can be on rail tracks. It is concluded that, with utilization of IR camera, the driver can recognize possible object on the railway long before the object is illuminated by the train's headlights, but can not determine if that object is a human or not. Furthermore, a method for detecting objects on the rail tracks in front of a moving train using a monocular thermal camera is proposed in [12]. This method is based on localization of the rail tracks and anomaly detector which detect objects that do not look like rail tracks where are expected to be. However, the method was tested only on simulated objects square and rectangle shaped, but results showed that it can be used only on a limited range due to its assumption of constant curvature.

This paper presents an advanced system based on thermal camera for detection of objects on rail tracks. In order to achieve greater reliability and robustness, advanced image processing algorithm is implemented and tested on a set of images captured with thermal camera (infrared camera) at night conditions. Developed algorithm has three steps: detecting and localization of rail tracks, object detection and estimation of distances between camera stand and detected objects. At first, edge detector with threshold is used for detection and localization of rail tracks and, based on that, determination of region of interest (ROI). Five edge detectors are used, in order to define the best detection method. Canny edge detector showed the best results, and ROI is determined as region where rail tracks are detected with a certain space next to them. The goal was to detect objects on rail tracks and/or in their close vicinity. After that, region-based segmentation with defined optimal threshold range, is used for object detection. During experiment, in determined ROI, four object/pedestrians are detected. Estimation of distances between camera stand and detected objects/pedestrians is done by using homography based method.

Rail track detection

An image can be defined by a set of regions that are connected and non-overlapping, such that the pixels in each partitioned region possess an identical set of properties or attributes [13]. These sets of properties of the image may include gray levels, contrast, spectral values, textural properties, *etc.* which can be result of variations in reflectance, illumination, color, shade, texture, orientation and depth of scene surfaces. Abrupt changes in gray level can be used for partition an image. The principal areas of interest within this category are the detection of lines and edges in an image. An edge is a point or set of points that creates a curve that follows a path of significant local change in gray level and, because of that, carry a lot of information about the various regions in the image. In addition to the edges, lines also can be detected from the abrupt change in the gray level. However, there is an important difference

between edges and lines – an edge essentially defines boundaries between two regions while, line may be embedded inside a single uniformly homogenous region [13, 14].

The edge detection is process with goal to detect significant local changes in the pixel intensity in an image. The change in intensity level is measured by the gradient of the image so, since an image $f(x, y)$ is a 2-D function, its gradient is a vector [13]:

$$\begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \dfrac{\mathrm{d}f}{\mathrm{d}x} \\ \dfrac{\mathrm{d}f}{\mathrm{d}y} \end{bmatrix} \tag{1}$$

The magnitude of the gradient may be computed [13]:

$$\begin{aligned} G[f(x, y)] &= \sqrt{G_x^2 + G_y^2} \\ G[f(x, y)] &= |G_x| + |G_y| \\ G[f(x, y)] &= \max\{|G_x|, |G_y|\} \end{aligned} \tag{2}$$

Equations (2) represent three possible ways for the computation of the magnitude of the gradient. However, the direction of the gradient is (θ is measured with respect to x-axis) [13]:

$$\theta(x, y) = \tan^{-1}\frac{G_y}{G_x} \tag{3}$$

Gradient operators calculate the change in gray level intensities, as well as the direction of change. This calculation is performed by the difference in values of the neighboring pixels, *i. e.* the derivate along the x-axis and the derivate along y-axis. The gradients in a 2-D image are approximated by [13]:

$$G_x = f(i + 1, j) - f(i, j), \qquad G_y = f(i,j + 1) - f(i, j) \tag{4}$$

For obtaining the x-direction gradient and y-direction gradient, gradient operators require two masks. However, these two gradients are combined in order to obtain a vector quantity whose magnitude represents the strength of the edge gradient at a point in the image and angle represents the gradient angle.

The goal of ideal edge detector is to detect an edge point precisely in the sense that a true edge point in an image should not be missed, but a false edge point should not be wrong detected. However, since quality of detection is dependent on lighting conditions, the presence of objects of similar intensities, density of edges in the scene, and noise, different operator-based edge detectors are used for different purposes. The edge detectors, presented in this paper, are based on a single derivative – Roberts, Sobel, Prewitt and Canny, as well as Laplacian of Gaussian (LoG), which is edge detector based on second derivative [13].

The Roberts edge detector [13] is based on 2 × 2 gradient operator, which is used to approximate the gradient magnitude of an image through discrete differentiation, which is achieved by computing the sum of the squares of the differences between diagonally adjacent pixels. This approximation of gradient magnitude is given [13]:

$$G[f(i, j)] = [f(i, j) - f(i + 1, j + 1)] + [f(i + 1, j) - f(i,j + 1] \tag{5}$$

The convolution mask for Robert's operator is a 2 × 2 mask, fig. 1, and, since only four input pixels need to be examined to determine the value of each output pixel, and only subtractions and additions are used in the calculation, this detector is fast but very sensitive to noise.

(a)

1	1
–1	–1

(b)

1	2	1
0	0	0
–1	–2	–1

(c)

1	0	–1
2	0	–2
1	0	–1

Figure 1. Robert mask (a), Sobel mask for calculation gradient G_x (b), and gradient G_y (c)

The Sobel edge detector [13] is a 3 × 3 neighborhood based gradient operator, which calculates an approximation of the gradient of the image intensity function. This detector finds the direction of the largest increase from light to dark and the rate of change in that direction. The convolution masks, figs. 1(b) and 1(c), for the Sobel operator are separately applied on the input image.

The reason for that is to yield two gradient components G_x and G_y in the horizontal and vertical orientations, respectively [13]:

$$G_x = [f(i-1, j-1) + 2f(i-1, j) + f(i-1, j+1)] - [f(i+1, j-1) + 2f(i+1, j) + f(i+1, j+1)] \quad (6)$$

$$G_y = [f(i-1, j-1) + 2f(i, j-1) + f(i+1, j-1)] - [f(i-1, j+1) + 2f(i, j+1) + f(i+1, j+1)] \quad (7)$$

The gradient magnitude is calculated [13]:

$$G[f(x, y)] = \sqrt{G_x^2 + G_y^2} \quad (8)$$

The Prewitt edge detector [13] is based on Prewitt operator defined by a set of eight masks – four of which are shown in fig. 2. Other mask is generated by rotation of 90°, successively [13]. On that way, the mask produces maximal response through giving the direction of the gradient.

1	1	1
0	0	0
–1	–1	–1

0	1	1
–1	0	1
–1	–1	0

–1	0	1
–1	0	1
–1	0	1

–1	–1	0
–1	0	1
0	1	1

Figure 2. Prewitt mask in 90° successive rotations

For Prewitt operator, calculating of the magnitude and direction of the edge gradients can be done as in case of Sobel operator.

Canny edge detector [13] is based on operator that uses multi-stage algorithm, which includes: detection of edge with low error rate, the edge point detected should be well localized (located edges must be as close as possible to the true edges), and single edge point response (detector should return only one point for each true edge point). The Canny edge detection algorithm consists of the following basic steps [13, 15]:

– smooth the input image with a Gaussian filter in order to remove the noise,
– calculate the gradient magnitude and angle images,

– apply non-maximum suppression to the gradient magnitude image,
– use double threshold to determine potential edges, and
– connectivity analysis to finalize detection by suppressing all the other edges that are weak and not connected to strong edges.

The LoG detector [13] is based on operator which first applies Gaussian smoothing, followed by the Laplacian operator. The Gaussian function reduces the noise and the resultant Laplacian mask minimizes the probability of false edge detection. After that, a finally checks for zero crossings are performed. The LoG function for convolution is defined [13]:

$$LOG(x, y) = \frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right] \mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{9}$$

The 5 × 5 convolution mask for LoG edge detector is shown in eq. (10):

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \tag{10}$$

The main task for edge detector was to detect rail tracks, on IR images captured at night conditions. Detection should be as good as possible, with small noise, or completely without it. For that purpose, five edge detectors – Roberts, Sobel, Prewitt, Canny, and LoG are used, in order to perform edge detection on variety of scenarios, analyze results, and choose the best one. For each edge detector, threshold is set. The most complex scenario includes rail tracks, humans and the environment, as three completely different kinds of objects for potential detection, and is shown in fig. 3, as well as results of different edge detectors.

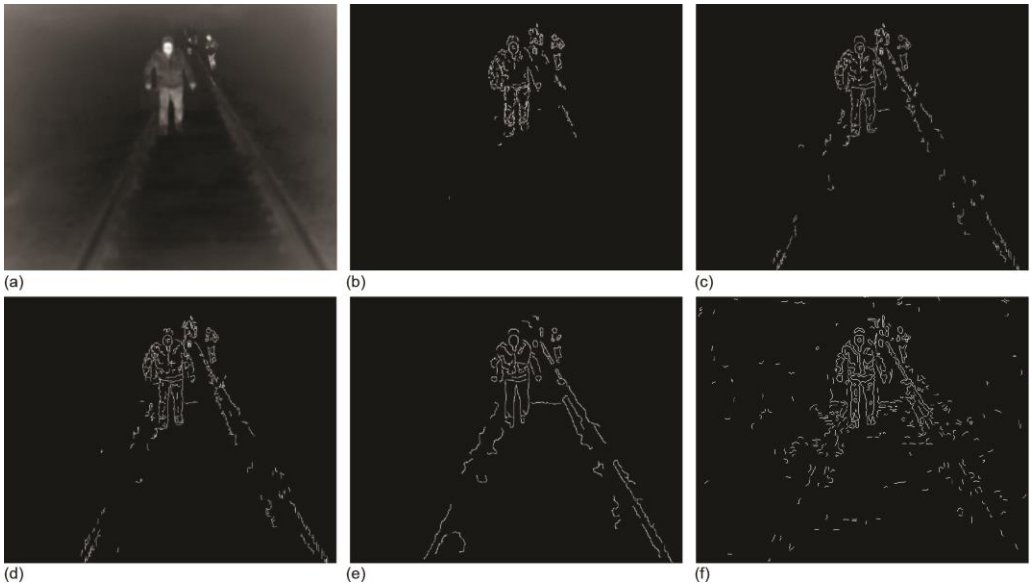

Figure 3. One tested scenario (a), Roberts (b), Sobel (c), Prewitt (d), Canny (e), and LoG (f)

Results showed that first edge detector – Roberts, gave very bad results, because rail tracks, as one of the main tasks, are not detected at all. Sobel and Prewitt edge detectors showed similar results, as expected, with many discontinuities in edges from rail tracks and noise in human detection, which can be problem in further analysis. The LoG edge detector provided results with lot of noise and rail tracks are not detected in first part of the image. However, the Canny edge detector showed the best results in detection of rail tracks and humans. The detection was almost without noise, but there were some small discontinuities in rail tracks and humans.

After applying Canny edge detector, detected edges of rail tracks were extracted and analyzed. Based on positions of detected rail tracks, ROI is determined, in order to perform analysis for object detection. The ROI was determined as region where rail tracks are detected with a certain space next to it, because the goal is to detect objects on railway crossing, *i. e.* on rail tracks and/or in their close vicinity.

Object detection

For detection of objects, region-based segmentation method, which is based on partitioning an image into regions that are similar according to a set of predefined criteria, was used. This method can be used as thresholding, region growing, and region splitting and merging [16-18]. However, in this case, thresholding and region growing for object detection was used.

Thresholding is simple and fast for implementation and, because of its intuitive properties, has central position in applications of image segmentation. The thresholding operation involves identification of a set of optimal thresholds, based on which the image is partitioned into several meaningful regions. There are several different methods for choosing a threshold. Users can manually choose a threshold value, or a thresholding algorithm can compute a value automatically, which is known as automatic thresholding. A simple method would be to choose the mean or median value, the rationale being that if the object pixels are brighter than the background, they should also be brighter than the average.

In this case, optimal threshold range was defined, in order to detect only objects/pedestrians. However, in next step, neighboring regions of a binary image were connected by morphological closing of the gaps between the regions and by smoothening their outer edges with a disc-shaped structuring element. After that, problem with noise was solved by removing all the segmented areas smaller than 40 pixels. This method was tested on set of scenarios, and one of them is shown in fig. 4(a). During experiment, in previously determined ROI, four object/pedestrians were detected, as shown in fig. 4(b). Those detected objects/pedestrians (marked with yellow rectangles) are located on rail tracks (marked with green color) and near them from inside and outside, *i. e.* in their close vicinity, fig. 4(c).

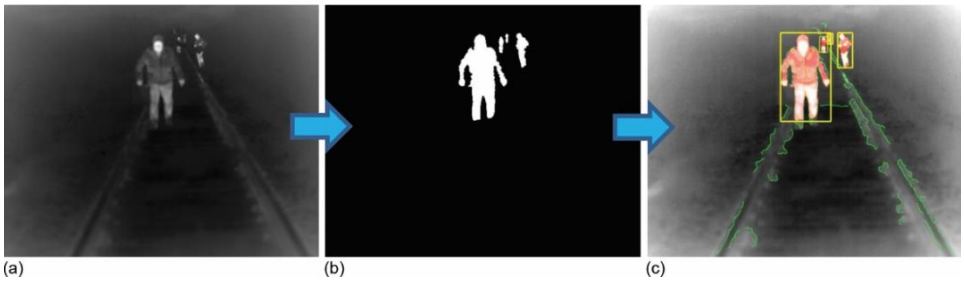

Figure 4. Input image (a), detected objects (b), and located objects (c)

Distance estimation of objects on rail tracks

After detection of objects/pedestrians, distance estimation from camera stand to objects was performed by using of homography method. The homography is also called projectivity, which is defined as *invertible mapping h from P^2 to itself such that three points x_1, x_2, and x_3 lie on the same line if and only if h(x_1), h(x_2), and h(x_3) do* [19]. However, in this way, projectivity is defined in terms of geometry. So, an equivalent algebraic definition of projectivity is that *a mapping h: $P^2 \rightarrow P^2$ is a projectivity if and only if there exists a non-singular 3 × 3 matrix H such that for any point in P^2 represented by a vector x it is true that h(x) = Hx* [19]. So, a planar projective transformation, which was used for estimation of distance, represents linear transformation on homogeneous 3-vectors represented by a non-singular 3 × 3 matrix [19]:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{11}$$

or shorter

$$\boldsymbol{x'} = \mathrm{H}\boldsymbol{x}$$

Matrix H is so-called a homogeneous matrix, since as in the homogeneous representation of a point, and there are eight independent ratios amongst the nine elements of H. So, one element of matrix H can have arbitrary value, and usually is $h_{33} = 1$. Based on that, distance estimation was included in two phases – calculation of homography matrix H and mapping of pixels from one to another image.

At first, estimation started with mapping one point from the plane in the real environment to the plane in the image. The co-ordinates (x, y) of this point can be given explicitly in relation to an arbitrarily chosen co-ordinate system. By expanding with number 1, an equation in homogeneous co-ordinates $(x, y, 1)^T$ is obtained. A homography would give a point whose homogeneous co-ordinates are in the general case $(x_1', x_2', x_3')^T$. Now, the equation can be written in the following form:

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ y \\ 1 \end{pmatrix} \tag{12}$$

In order to get the non-homogeneous co-ordinates of the mapped point, *i. e.* its real co-ordinates in the image (x', y'), it is necessary to divide the first and second equations with the third equation. The following equations are obtained:

$$x' = \frac{x_1'}{x_3'} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}, \qquad y' = \frac{x_2'}{x_3'} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{13}$$

One point gives two linear equations according to the elements of the homography matrix, H. There are eight degrees of freedom, which means that the co-ordinates for four points (in the image and in the real environment) are required to find the elements of the matrix H, with the restriction that none of the three are collinear. The last element of the H matrix is, as previously said, usually taken to be equal to the number 1, $h_{33} = 1$, in order that the output image will be in proportion to the selected length unit. So, if the co-ordinates of four

points are known, the linear equation system can be written in matrix form, as is presented in eq. (14):

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1x_1' & -y_1y_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1y_1' & -y_1x_1' \\ x_1 & y_1 & 1 & 0 & 0 & 0 & -x_2x_2' & -y_2y_2' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_2y_2' & -y_2x_2' \\ x_1 & y_1 & 1 & 0 & 0 & 0 & -x_3x_3' & -y_3y_3' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_3y_3' & -y_3x_3' \\ x_1 & y_1 & 1 & 0 & 0 & 0 & -x_4x_4' & -y_4y_4' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_4y_4' & -y_4x_4' \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4' \end{bmatrix} \tag{14}$$

or shorter

$$A\vec{h} = \vec{b} \tag{15}$$

where the matrix A is dimension of 8 × 8 and represents the matrix of the system, $\vec{h}$ – the vector of elements of the homography matrix H (homogeneity vector) with dimension of eight, and $\vec{b}$ – the vector of co-ordinates of the mapped points with dimension of eight. Vector $\vec{h}$ can be obtained:

$$\vec{h} = A^{-1}\vec{b} \tag{16}$$

For calculation of homography matrix H, four points – vertex of blue quadrilateral was used, fig. 5(a). Co-ordinates of these points are known in real world and on image. The co-ordinates in real world are calculated based experiment data, *i. e.* on known positions of people on rail tracks relative to the camera stand, while co-ordinates in image are determined within an image.

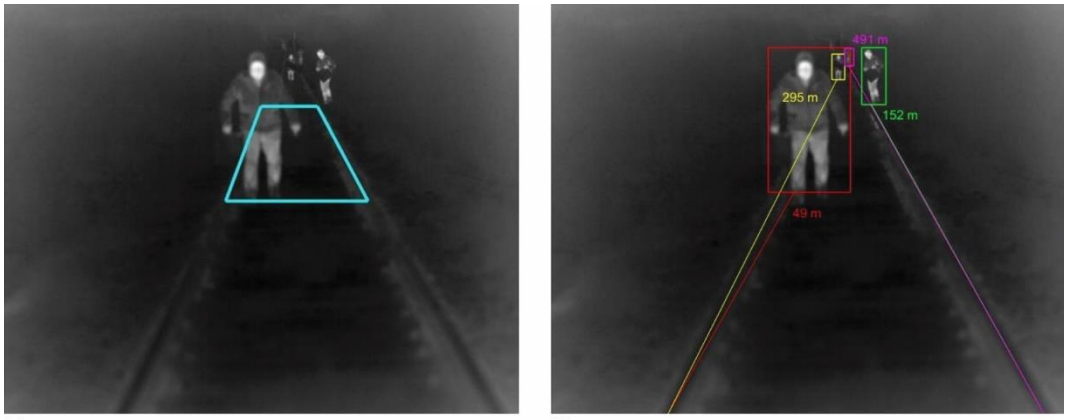

Figure 5. Points for calculation of matrix H (a), and estimated distances (b)

Using an inverse matrix H, an estimation of the distance between camera stand and previously detected objects/pedestrians was performed. Estimated distances from camera stand to detected objects/pedestrians, fig. 5(b) were 49 m (marked with red color), 152 m (marked with green color), 295 m (marked with yellow color), and 491 m (marked with magenta color). For validation of estimation, real measured distances from camera stand to objects-humans involved in experiment, were measured along rail tracks: 50 m, 150 m, 300 m, and 500 m, respectively. The error in estimation is a consequence of uncertainty in homography matrix H calculation as well as object detection (detection of exact point on the rail tracks where the object was). However, with a maximum error of 2%, using of homography based method for estimation of distance, gave acceptable results.

Conclusion

Railway infrastructure is widely used in transport, because of its capacity and infrastructure. Its complexity leads to intersecting with other types of traffic, especially road, and pedestrian. Because of that, there are many potential places where traffic accident can occur, peculiarly on a marked railway level crossing. However *unmarked crossings* represent special dangerous places because there are no signs and enough illumination for timely observing of objects and pedestrians.

For increasing safety and avoiding accidents, many monitoring systems can be used but, because of infrastructure, their usage is limited only on marked level crossings. Also, the most of them are dependent on weather and environment conditions, the required amount of illumination, and other factors so, they can operate only in the day and good-light conditions. On the other side, considering that pedestrians very often use *unmarked crossings*, many accidents happen because train driver, at night and in low-light conditions, cannot timely notice pedestrians or other objects on crossing, but only when they are illuminated by train headlights.

In this paper, an advanced system for detection of objects on rail tracks based on thermal camera is presented. The advanced image processing algorithm was implemented in three steps and tested on set of IR images captured at night conditions. First, for detection of objects, ROI should be determined. The ROI should include rail tracks, because objects which are on rail tracks and/or in their close vicinity should be detected. For detection of rail tracks, edge detection was used. However, in order to obtain the best results in localization and extraction of rail tracks in whole scene, five edge detectors with set threshold were tested. The best result showed Canny edge detector and, based on detected rail tracks, ROI was determined as region where rail tracks are detected with a certain space next to them. Although some objects were detected using the Canny edge detector, these results are not usable for further analysis. Because of that, for detection of objects, region-based segmentation with defined optimal threshold range, was used. In experiment, in previously determined ROI, four objects/pedestrians were detected and located on rail tracks and near them from inside and outside. Based on showed results, it can be concluded that edge detection and region-based segmentation showed good results in process of determination of ROI and object detection. For estimation of distances between four detected objects/pedestrians and camera stand, homography based method was used. For validation of estimation, real distances from camera stand to objects-humans, involved in experiment, were measured along rail tracks. This method gave acceptable results, with a maximum error of 2% in estimation.

Developed algorithm showed good results in detecting of rail tracks and objects/pedestrians, as well as estimation of distances between camera stand and detected objects/pedestrians. However, in spite of good results, in order to improve robustness and adapt-

ability, tools from artificial intelligence domain can be used in future work for determining of adaptive parameters for object detection and estimation of distances.

Acknowledgment

This research has been done in framework of Horizon 2020 Shift2Rail project *Smart Automation of Rail Transport – SMART*.

References

[1] Pavlović, M., *et al.*, Methods for Detection of Obstacles on the Railway Level Crossing, *Proceedings*, 17th Scientific-Expert Conference on Railways RAILCON '16, Nis, Serbia, 2016, pp. 121-124
[2] Pavlović, M., *et al.*, Application of Thermal Imaging Systems for Object Detection, *Proceedings*, 13th International Conference on Accomplishments in Mechanical and Industrial Engineering, Banja Luka, Republic of Srpska, BiH, 2017, pp. 653-662
[3] Shaik, J., Iftekharuddin, K. M. Detection and Tracking of Targets in Infrared Images Using Bayesian Techniques, *Optics & Laser Technology, 41* (2009), 6, pp. 832-842
[4] Iwasaki, Y., *et al.*, Robust Vehicle Detection under Various Environmental Conditions Using an Infrared Thermal Camera and Its Application to Road Traffic Flow Monitoring, *Sensors, 13* (2013), 6, pp. 7756-7773
[5] Jueungling, K., Arens, M., Feature Based Person Detection beyond the Visible Spectrum, *Proceedings*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Miami, Fla., USA, 2009, pp. 30-37
[6] Wang, W., *et al.*, Improved Human Detection and Classification in Thermal Images, *Proceedings*, 17th IEEE International Conference on Image Processing, Beijing, 2010, pp. 2313-2316
[7] Fernandez-Caballero, A., *et al.*, Optical Flow or Image Subtraction in Human Detection from Infrared Camera on Mobile Robot, *Robotics and Autonomous Systems, 58* (2010), 12, pp. 1273-1281
[8] Ćirić, I., *et al.*, Computationally Intelligent System for Thermal Vision People Detection and Tracking in Robotic Applications, *Proceedings*, 11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Nis, Serbia, 2013, pp. 587-590
[9] Bertozzi, M., *et al.*, Pedestrian Detection by Means of Far-Infrared Stereo Vision, *Computer Vision and Image Understanding, 106* (2007), 2-3, pp. 194-204
[10] Broggi, A., *et al.*, A Multi-Resolution Approach for Infrared Vision-Based Pedestrian Detection, *Proceedings*, IEEE Intelligent Vehicles Symposium, Parma, Italy, 2004, pp. 2313-2316
[11] Forth, A., Zamjatnins, F., Night-Vision Device for Railway Vehicles for Improving Safety, *Signal+Draht, 107* (2015), pp. 38-43
[12] Berg, A., *et al.*, Detecting Rails and Obstacles Using a Train-Mounted Thermal Camera, *Proceedings*, 19th Scandinavian Conference SCIA 2015, Copenhagen, Denmark, 2015, pp. 492-503
[13] Acharya, T., Ray, A. K., *Image Processing: Principles and Applications*, John Wiley & Sons Inc., Hoboken, N. Y., USA, 2005.
[14] Nadernejad, E. , *et al.*, Edge Detection Techniques: Evaluations and Comparisons, *Applied Mathematical Sciences, 31* (2008), 2, pp. 1507-1520
[15] Moeslund, T., *Canny Edge Detection*, Laboratory of Computer Vision and Media Technology, Aalborg University, Aalborg, Denmark, 2009
[16] Gonzales, R. C., Woods, R. E., *Digital Image Processing*, Pearson Education, Hoboken, N. Y., USA, 2008
[17] Ćirić, I., *et al.*, Intelligent Optimal Control of Thermal Vision-Based Person-Following Robot Platform, *Thermal Science, 18* (2014), 3, pp. 957-966
[18] Ćirić, I., *et al.*, Thermal Vision Based Intelligent System for Human Detection and Tracking in Mobile Robot Control System, *Thermal Science, 20* (2016), Suppl. 5, pp. S1553-S1559
[19] Hartley R., Zisserman A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, USA, 2004

Paper submitted: May 4, 2018
Paper revised: August 18, 2018
Paper accepted: August 21, 2018