ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20210223081612
Preliminary communication

AI Powered Obstacle Distance Estimation for Onboard Autonomous Train Operation

Ivan ĆIRIĆ*, Milan PAVLOVIĆ, Milan BANIĆ, Miloš SIMONOVIĆ, Vlastimir NIKOLIĆ

Abstract: This paper proposes a novel method for an AI powered improvement of the estimation of a distance between the camera and an imaged object using image-plane homography. The method exploits the homography between two planes, the image plane and the rail tracks plane, and an artificial neural network that reduces the estimation error based on collected experimental data. The SMART multi-sensory onboard obstacle detection system has 3 vision sensors – an RGB camera, a thermal vision camera and a night vision camera, in order to achieve greater reliability and robustness. Although the methodology presented in this paper is applicable for each vision sensor, the proposed method was tested with the thermal camera and in impaired visibility scenarios. The validation of estimated distances is done with respect to real measured distances from the camera stand to the objects (humans) involved in the experiments. Distances are estimated with a maximum error of 2% and the proposed AI powered system can provide a reliable distance estimation in impaired visibility conditions.

Keywords: Artificial neural network; Autonomous train operation; Distance estimation; Homography; Image processing; Machine vision

1 INTRODUCTION

The quality and cost competitiveness of railway freight transportation can be considerably improved by following the trend of automation in order to achieve a cost-effective, flexible and attractive service. Today, automation and autonomous operation have become common in road, air and marine transportation. Modern harbours have Automated Guided Vehicles (AGVs) that carry shipping containers from cranes to trackside, warehouses, distribution centres, while auto-pilots are standard on air carriers and huge cargo ships, and there is no need for a large number of on-board personnel. The development of autonomous cars and trucks is already in a serious phase. Also, the development of autonomous systems in rail transport has been present mainly in the area of public transport services (driverless metro lines, light rail transit (LRT), people movers, and automated guided transit (AGT)). The basic idea was to use a certain level of automation in order to transfer operation tasks from the driver to the train control system (e.g., ERTMS). According to The International Electrotechnical Commission (IEC) standard 62290-1, Autonomous Train Operation (ATO) is part of a highly automated system with reduced driver supervision [1].

For the fully autonomous train operation all the activities and responsibilities of train operators need to be taken over by several systems that can sense the environment and overlook the scene, detect potentially dangerous objects on the train's path and react accordingly and in the right way [2-6].

The obstacle detection system, the main part of the ATO system, will need to monitor the environment according to freight specific and general use cases, e.g., EN62267 and/or relevant projects working in the field of automation. In order to fulfil strict railway standards and regulations, an Obstacle Detection System (ODS) should work in a challenging environment and in hard visibility conditions. ODS represents a machine vision system with hardware and software solutions (Fig. 1), so as to provide reliable information of obstacle existence on the railway and/or its close vicinity, and to estimate the distance from the system to the detected obstacle [7]. The system needs to operate in real time and in different light conditions (day, low-light and at night), detect obstacles at very long ranges, e.g., up to 2 km, work reliably in troubling weather conditions, including heavy winter and desert-like situations, and maintain reliability while a train moves within the speed range from 0 km/h up to 180 km/h.

A common characteristic of most ODSs is that they can operate only in 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, a thermal imaging system can be used because its operating range is in the invisible infrared region of the spectrum. There are not many fully developed methods for the detection of living and non-living objects with thermal imaging systems [7].

The complexity of object detection rises when a thermal imaging system is mounted on a moving platform or some vehicle. An approach to real-time human detection through processing a video captured by an IR (infrared) camera mounted on the autonomous mobile platform is presented in [8]. Advanced control of a mobile robot with the goal to recognize a human in an indoor environment and allow adequate human-robot interaction is presented in [9]. Its operation is based on the intelligent advanced segmentation and classification of detected regions of interest in every frame acquired by an IR camera. Furthermore, the authors in [10] used a stereo system for the detection of pedestrians employing two far-infrared cameras mounted on a 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. Furthermore, the presence of objects with a similar size and shape to a pedestrian represents the most frequent cause of misdetection.

A system with a monocular IR camera mounted on a test vehicle-a car, presented in [11], is based on a multi-resolution localization of warm symmetrical objects with a specific size and aspect ratio with the 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 to 43.5 m.

In the field of railways, there are not many uses of thermal imaging systems for any purposes. However, in

[12] the authors used a thermal camera mounted on the train's roof and analysed a possible dangerous situation when trains operate at night and in bad weather conditions, especially when some objects can be found on rail tracks. It is concluded that, with the utilization of an IR camera, the driver can recognize a possible object on the railway long before the object is illuminated by the train's headlights, but cannot determine if that object is 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 [13]. This method is based on the localization of rail tracks and the anomaly detector that detects objects that do not look like rail tracks where they are expected to be. However, the method was tested only on square and rectangle-shaped simulated objects, but the results showed that it can be used only on a limited range due to its assumption of a constant curvature.

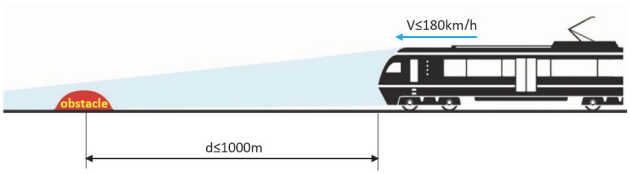

Figure 1 Concept of ODS [4]

In this paper, image-plane homography for object distance estimation from the thermal camera described in [14] is improved using an artificial neural network. The goal of the application is obstacle detection and tracking in railways, so the considered image-plane homography concerns two planes, the image plane and the rail tracks plane. In an ideal case where assumption is that entire visible rail tracks are in a single plane and the image is undistorted and rectified, homography would give very accurate distance estimation. However, due to the image distortion, and the fact that rail tracks can have up to 2% inclination and declination, the accuracy of the distance estimation varies depending on object position both in real world and image plane. An additional problem is introduced through imprecise coordinates of the bounding box of the object recognized on or near the rail tracks. Therefore, the idea of using artificial intelligence based algorithm, namely multilayer perceptron (MLP), for the distance estimation was the next logical step. The MLP distance estimator can deal with nonlinearities between input and output data. As in homography estimation, bounding box features represent input data, while output data is the estimated distance. Since the distance of the detected object plays a major role in the accuracy of the estimation, the novelty of this research lies in the selection of the homography estimated distance as an additional input of MLP.

Adequate MLP input selection greatly influences MLP performance, i.e. accuracy and speed. Based on comparison of various input selections for the MLP distance estimator, presented in this paper, one network topology is selected as the optimal for the real world application. The proposed MLP distance estimator should improve the accuracy of the traditional image-plane homography method and can be used for distance estimation in both daylight and low to no-light conditions, while also applicable for long and short range obstacle distance estimation. Unlike deep learning Multi DisNET [15], the proposed method is based on the multilayer perceptron (MLP) and therefore does not need a large training data set.

2 RELATED WORK

The estimation of distances from the ODS to the objects (obstacles, or targets in military applications) has a very important role in different safety critical applications such as mobile robotic applications, autonomous vehicles and Intelligent Transportation Systems (ITS). There are two types of methods for measuring and estimating the distance between two objects active and passive [7]. Active methods are based on sensors such as ultrasonic, radar and laser scanners, which use different types of signals in order to measure distances. Passive methods receive information about the object's position in the environment using a camera for passive measurement of the scene [16].

Vision-based systems use passive methods to provide highly valuable information about the environment and they are usually grouped into two classes: monocular and stereo-vision systems. A stereo vision-based system uses two cameras and so-called triangulation to obtain 3D coordinates of an object and thus to estimate the object distance [17]. A monocular vision-based system uses a single camera for capturing images and exploiting the geometry of the scene for distance estimation. In [18], a single camera, which has dual off-axis apertures that are covered with colour filters, was used to estimate the distance by finding the relative shifts between the projections of a point on an object through the two apertures. There is another use of the relationship between the physical distance of an object and its pixel height, which was exploited to find a mapping employed in the estimation of the distance of an object using a single image captured by a single camera, as presented in [19]. With this method, the achieved accuracy was as high as 98.76%, but with limited application. Also, a model for the relationship between the resolution of an object of interest and the distance from the camera as a growth series was proposed in [20]. The model was tested on random image samples captured by a single camera, and the results showed a very minimal error, in the range of 0.5 - 1%. Furthermore, the use of a single camera for the estimation of a distance from the camera to the human face was presented in [21]. The formula for distance estimation was based on the relationship between the pixel area and distance, and was derived from the pinhole camera model, camera calibration and area mapping. The results showed that the accuracy of measurement was above 95%. On the other hand, for automotive application, the ratio of the real distance between a host vehicle and a vanishing point, measured in metres, and the pixel distance from the host vehicle to the front vehicle, measured in pixels, was used for the estimation of the distance between the host and front vehicles, in [22]. The combination of two distance estimation methods, one based on the fact that the vehicle distance and the vehicle width were inversely proportional and the other position-based with the mapping of the detected vehicle onto a 3D space, was proposed in [23]. In testing on 1000 sequential images with the 640 × 480 resolution captured by a single camera, the proposed

method achieved 94.9% of accuracy in total, but was only applicable to daylight conditions. In [24], the focal length in pixels, the single camera height from the ground and the *y* coordinate of the point on the bottom line of the bounding box circumscribing each detected vehicle were used for distance estimation, as well as for computing a homography to increase accuracy. Improving the accuracy of distance estimation between two vehicles by monocular vision based on the vehicle pose information was presented in [25]. The results showed that this system was suitable for distance estimation when vehicles were within 30 metres. Some authors used a single camera and the Inverse Perspective Mapping (IPM) method for the estimation of distance, in order to provide a transformation of a forward-facing image to a top-down "bird's eye" view, in which there was a linear relationship between distances in the image and in the real world [26], and to remove the perspective effect in the HSV colour map [27]. The experiments showed good results, but the error increased with long distances. On the other hand, in [28], the method for distance estimation was proposed based on known parameters the camera field-of view, the height of the camera above the road level, the camera angle, etc.. Compared to IPM, this method showed better results in the short range, while both approaches showed a very similar error for medium distances (in a range of about 22 to 27 m), and the error level increased significantly (up to 9%) for the IPM method for far distances. The utilization of the homography method for the estimation of distances between a single thermal camera and detected objects in railway applications was presented in [14]. Objects (humans) were placed on and near the rail tracks and satisfactory results were achieved on a range from 50 m to 500 m with the maximal estimation error of 2%. There are methods that use a combination of a monocular camera and artificial intelligence tools for distance estimation. In [29], a detected rectangular bounding box for micro Unmanned Aerial Vehicle (mUAV) was used for distance estimation from the camera to the object. The authors made a training set of 35570 pairs of the width and height of the bounding box, and its known distance, and used it to train a support vector regressor (SVR). The distance estimation was evaluated by means of indoor videos only, and gave good results. The distance estimation system based on a Multi Hidden-Layer Neural Network, which was used to learn and predict the distance between the object and the monocular camera, was presented in [15]. This system was trained using a supervised learning technique where the input features were manually calculated parameters of the object bounding boxes resulted from the YOLO object classifier, and outputs were the accurate 3D laser scanner measurements of the distances to objects in the recorded scene. The evaluation of the system was done on the RGB images of railway scenes in a range from 100 m to 300 m and RGB images of a road scene in the range from 6 m to 30 m.

3 DATA ACQUISITION AND OBSTACLE DETECTION FOR DISTANCE ESTIMATION

The acquisition of the data used for the MLP distance estimation training and testing was done using the SMART on board obstacle detection system for ATO [30] (Fig. 2), that consists of three monocular RGB cameras, one IR camera, one night vision camera and one laser scanner. All the sensors were mounted on a specially designed vibration resistant metal housing.

Figure 2 ODS mounted on a train for dynamic testing

This setup was tested in field tests performed on rail tracks at different times of day and night. For obstacle detection in low-light conditions the thermal camera and the night vision system were used.

For the real time online processing ROS Indigo Igloo Full Desktop was used, while using OpenCV library, CUDA 8.1 and working on Ubuntu 14.04 64-bit Qt 4.8.1. For offline ANN training and testing and GA optimization MathWorks MATLAB was used as well.

In order to evaluate the proposed method for long-range object (obstacles) distance estimation, field tests were performed on a Serbian railway test-site (at the location of Babinpotok village, near the city of Prokuplje) approved for the experimental use by the Serbian Railway authorities (Fig. 3).

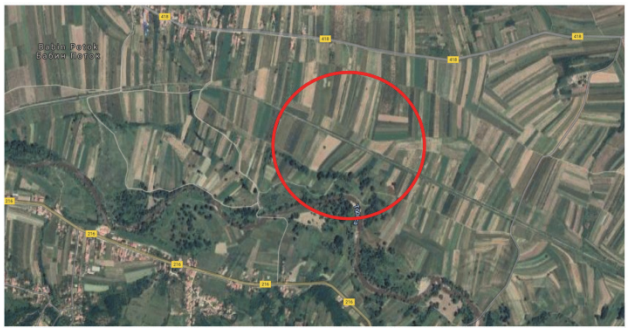

Figure 3 Testing site

The experimental set-up for the field tests presented in this paper used SMART ODS placed on a static test-stand on a level crossing so as to view the straight rail tracks in the length of about 1200 m. During the tests, humans acted as moving obstacles in the vicinity of the rail tracks and for various distance ranges. The thermal camera was used for recording the rail tracks scene with the objects (persons) on the rail tracks in low-light (night) conditions (the intensity of illumination was 0, measured with a luxmeter).

For the calculation of homography matrix, two objects (persons) were positioned on the opposite rail tracks, 50 m and 1000 m away from the test-stand. During the experiments, three persons, members of the research team, including the authors of this paper, imitated potential static obstacles on the rail tracks located at the distances of 950

m, 940 m, 930 m etc., moving10 metres towards camera, up to 50 m from the camera test-stand. The real distances from the camera stand to the "obstacles" were pre-measured and marked on the rail track using both a laser distance meter and GPS, and additionally checked during tests using GPS sensors by the persons involved in the training/testing data acquisition.

After data acquisition, object detection was done in order to provide class and bounding box of the detected object.

According to Johnson's criteria [31, 32], a distinction needs to be made between degrees of "seeing" a target in thermal imaging: Detection, Recognition and Identification (DRI). Johnson's criteria give a 50% probability of an observer discriminating an object to the specified level and 2 pixels per metre for detection are needed, 8 pixels/metre are needed for recognition and 16 pixels/metre are needed for the identification of a person.

The used FLIR TAU 2 640 × 480 is an uncooled vanadium oxide microbolometer sensor with 17 μm pixel size, which, together with a 100 mm lens, gives a field of view of 6.2° × 5° and an intrinsic field of view of 0.17 mrad [33]. Equivalently, at a 1000 m distance, each pixel covers a 17 cm × 17 cm square. This means that this system gives 5.88 pixels per metre at 1000 m, which is good enough for detection but not good enough for the recognition of potential obstacles.

The accuracy of distance estimation is correlated with an image processing algorithm for obstacle detection. Namely, small errors of bounding box coordinates can result in a significant error for long range distance estimation. During the implementation of the SMART project [14, 15, 30, 34], two approaches were used. The first approach is based on the traditional region-based segmentation and edge detection, while various soft computing methods were implemented for segmentation improvement [7, 14, 34]. The basic machine vision algorithm is presented in Fig. 4, and the image processing results from [7, 14] (Fig. 5) were used for the development of the distance estimation algorithm in this paper.

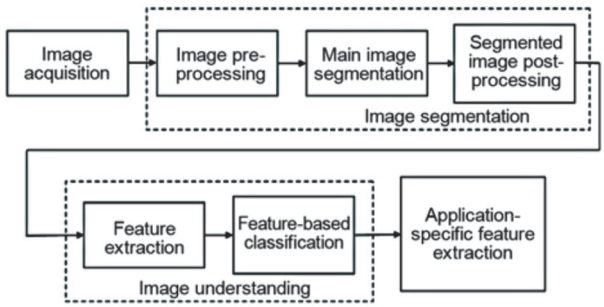

Figure 4 Machine vision algorithm

Another approach used by the authors for obstacle detection is based on a state-of-the-art computer vision object detector YOLO (You Only Look Once), trained with a COCO dataset [15]. YOLO is a fast and accurate object detector based on a Convolution Neural Network (CNN) and its outputs are bounding boxes of detected objects in the image and labels of the classes detected objects belong to.

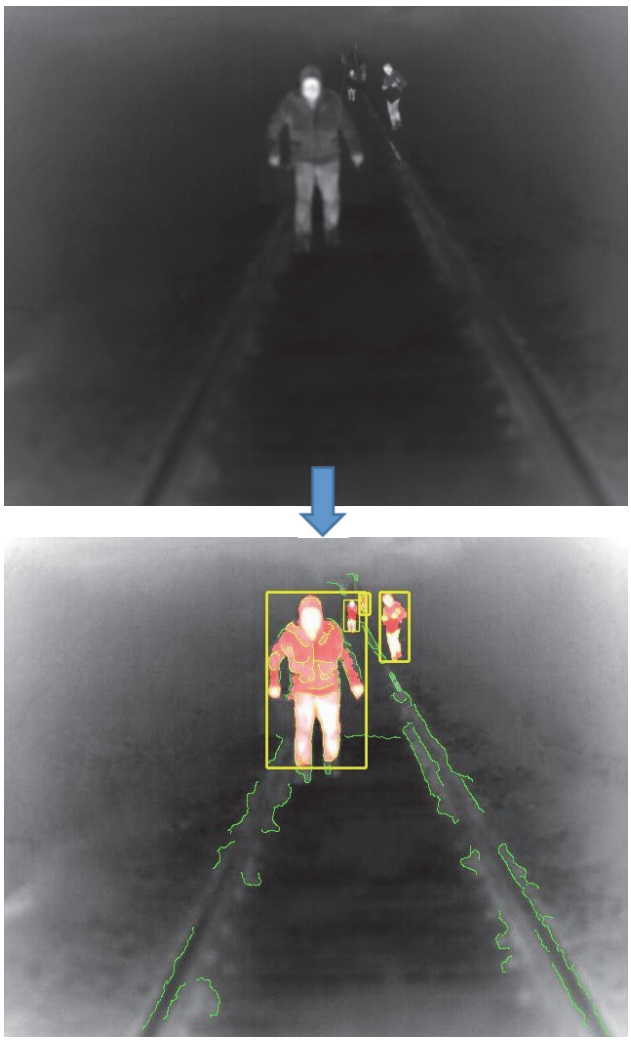

Figure 5 Image processing results [12]

In order to prepare distance estimation training and testing data, the experimental data acquired was processed using image processing algorithms developed and tested in [7, 14, 34].

4 HOMOGRAPHY BASED DISTANCE ESTIMATION

The distance estimation results presented in [7, 14] using image-plane homography were used as a starting point for the research presented in this paper. The so-called homography [35] offers the possibility of mapping the image plane and the corresponding world plane. As a result of that mapping, the world 3D coordinates of each point in the imaged world plane can be calculated. As rail tracks are located in a plane with respect to the locomotive frontal profile, homography mapping is meaningful in order to get an estimation of the distance d_h from an on-board mono camera to an object point on the rail tracks.

This estimation of object distance involves two phases: calculation of homography matrix $\boldsymbol{H}$ and mapping of points from one plane to another, the rail tracks to the camera image plane.

A point $\boldsymbol{x}$ from the rail tracks plane is mapped to a point in the image $\boldsymbol{x}'$ according to:

$$\boldsymbol{x}' = \boldsymbol{H}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x}$ is the homogeneous vector of the point from the real-world plane, $\boldsymbol{x'}$ is the homogeneous vector of the corresponding point in the image plane and $\boldsymbol{H}$ is the 3 × 3 homography matrix.

For the calculation of homography matrix $\boldsymbol{H}$, four points were used, for which coordinates in the real world and the image were known. The calculation of the coordinates in the real world was done based on the experimental data, where the objects (persons standing on the rail tracks) were at distances of 50 m and 1000 m from the camera. In order to calculate the corresponding image coordinates, the railtrack detection was done using the edge detection technique and the detection of the objects on the rail tracks was performed using the region-based image segmentation. Intersection points of the detected objects and the detected rail tracks were considered as image pixels corresponding to real world coordinates. The same methodology was used in [14], but for the short-range distances (people were standing at the distances of 50 m and 150 m).

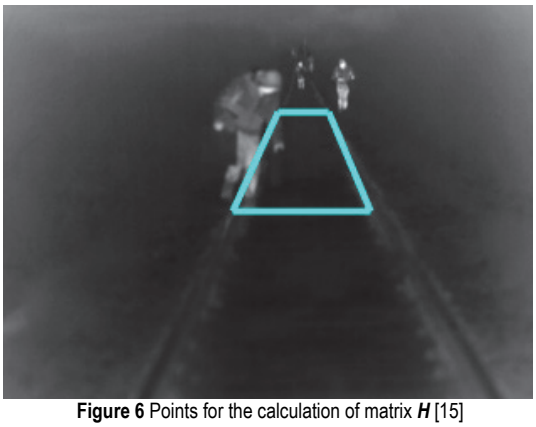

Figure 6 Points for the calculation of matrix $\boldsymbol{H}$ [15]

Calculated homography matrix $\boldsymbol{H}$ is:

$$\boldsymbol{H} = \begin{bmatrix} 5.671 & 0.257 & 260 \\ 0 & 0.115 & 1086 \\ 0 & 0.00075 & 1 \end{bmatrix}$$

Using the inverse of homography matrix $\boldsymbol{H}$, the estimation of the distance between the camera and any real-world point from the rail tracks plane d_h can be calculated as:

$$\boldsymbol{x} = \begin{bmatrix} x \\ y = d_h \\ 1 \end{bmatrix} = \boldsymbol{H}^{-1}\boldsymbol{x'} \tag{2}$$

The results of the proposed method for randomly selected distances are shown in Tab. 1.

Table 1 Estimation for randomly selected distances

Measured distance / m	65	215	375	585	775	910
Homography estimated distance-person I / m	71	200	302	470	690	924
Homography estimated distance-person II / m	67	204	332	496	714	924
Homography estimated distance-person III / m	68	207	398	504	709	924

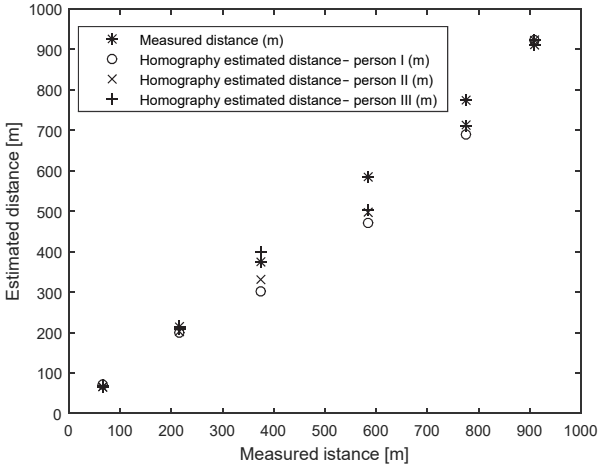

Figure 7 Estimation of distances presented in Table 1

The estimation results achieved by the homography based method can be useful, but estimation accuracy varies and this method cannot successfully incorporate the nonlinearities into the estimation model. The nonlinearities are mostly the result of the thermal image distortion and poor image processing.

6 AI-POWERED DISTANCE ESTIMATION SYSTEM

The presented homography based distance estimation lacks in precision and cannot deal with nonlinearities. The increase in accuracy with the current dataset can be achieved by alternative approach-based on artificial intelligence. The proposed system uses an artificial neural network-Multilayer perceptron, and the research was done into selecting the optimal input selection out of 5 possible inputs.

Multilayer perceptron (MLP) is a feedforward artificial neural network (ANN), consisting of an input layer, output layer, and one or more hidden layers. The neurons in each layer are connected using weighted connections, linking each neuron of the current layer to all the neurons of the following layer. Every neuron sums weighted values of the neurons from the previous layer, when activated by the current neuron activation function. Since the MLP without the activation function can perform linear mappings only, applying the activation function (usually ReLU, Sigmoid, Tanh and Identity) to the layers can add a non-linear property allowing the model to approximate highly non-linear functions. The activation function is a mathematical function that serves to normalize values to a given range or to completely eliminate undesired values. The number of input neurons equals the number of variables in each data point amongst input values. The output value is determined by input values, the architecture of the model, and the connection weights.The MLP is trained using error backpropagation. The training process can be considered as adjusting the model weights and biases with the goal of minimizing a cost function [36-38]. The most important part of network design and performance is the adequate selection of the inputs and the preparation of the training set. Another factor in MLP performance is the used architecture, defined by the hyperparameters, such as the number of hidden layers, the number of neurons in each layer, the

activation function of the neurons and the backpropagation algorithm.

The structure of the homography-ANN based distance estimation system is presented in Fig. 8.

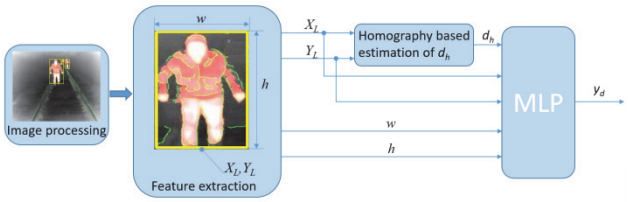

Figure 8 Structure of the AI powered distance estimation system

The estimated distance y_d is an output. Possible inputs are: homography based estimated distance d_h, X_L and Y_L coordinates in pixels of the centre of the bounding box edge that is in the rail track plane in the real world and the bounding box width w and height h in pixels.

The initial ANN in this paper consists of five inputs (neurons) in the input layer, one hidden layer with 10 perceptrons (neurons), and an output layer with one output. This type of the ANN MLP (with one hidden layer) is usually called plain vanilla. The inputs are: homography based estimated distance d_h, X_L and Y_L coordinates in pixels of the centre of the bounding box edge that is in the rail track plane in the real world and the bounding box width w and height h in pixels used for "feeding" the ANN, and the output is the predicted estimated distance y_d used for back-propagation process. This ANN model is trained on the set of data points derived from experiments and obstacle detection explained in Chapter 3.

The cost function is created using the mean squared error (MSE) and it is minimized using the Levenberg-Marquardt algorithm [39]. The activation function for all the neurons in the hidden layer is a Sigmoid function, and the activation function for the output layer is a linear function.

Based on the initial 5 input MLP (MLP5), additional topologies were considered, where only the input layer is changed, while hidden layer had 10 neurons, output is estimated distance y_d, activation functions for all the neurons in the hidden layer is a Sigmoid function, and the activation function for the output layer is a linear function and backpropagation was done using the mean squared error (MSE) and Levenberg-Marquardt algorithm. The MLP with 2 inputs (h and w, MLP2), MLP with 3 inputs (X_L, Y_L, d_h, MLP3) and MLP with 4 inputs (h, w, X_L, Y_L , MLP4) were all trained with the same training data set as MLP5. For the prediction comparison 3 more parameters, alongside with the Mean Squared Error (MSE), were used-the Root MeanSquared Error (RMSE), Mean Absolute Error (MAE) and R^2 score.

7 RESULTS

The data set prepared using the experimental data explained in Chapter 3 contained a set of 273 samples (3 persons detected every 10 metres on a range from 50 m to 950 m) was used for training, validation and testing of MLP2, MLP3, MLP4 and MLP5, while there was no batch processing and batch normalization performed.

The set of 273 samples was randomly divided so 191 samples (70%) were used for training, 41 samples (15%) were used for validation and 41 samples (15%) for testing. During the training there was no deactivation of the neurons (dropout) since the number of neurons in the hidden layer was determined using the performance metrics.

In order to evaluate the performance of the obtained models, the Mean Squared Error (MSE), Root MeanSquared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) were used. R^2score can be described as a statistical measure which is defined in the range between 0.0 and 1.0, where a value of 1.0 represents a perfect fit of the model and vice-versa [36, 37].The R^2score can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)}{\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i\right)^2} \tag{3}$$

The performance metrics can be calculated both for Homography based estimation and MLP based estimation, and the results are presented in Tab. 2.

Table 2 Performance comparison for various distance estimation methods

Distance estimation method, inputs	No. of epochs	MSE	MAE	RMSE	R^2 score
Homography, X_L, Y_L	--	271500	450	521.06	0.7795
2-input MLP (MLP2), h, w	58	2024.2	30.63	44.99	0.9707
3-input MLP (MLP3), X_L, Y_L, d_h	53	693.76	15.87	26.34	0.9899
4-input MLP (MLP4), h, w, X_L, Y_L	71	1592.5	25.77	39.9	0.9769
5-input MLP (MLP5), h, w, X_L, Y_L, d_h	75	233.83	11.44	15.29	0.9966

All the methods presented in this paper are used for the 3D reconstruction of the objects on the railtrack plane from the 2D thermal image. The low R^2 score of the homography based distance estimation method of only 0.7795 is a result of the used non-calibrated thermal image. Namely, the camera lens and sensors produce distortion and a distorted image needs to be undistorted and rectified through a calibration procedure for industrial use (like measurement, stereo-vision, etc.), and the chessboard pattern (printed on paper) is most commonly used for camera calibration. This would be beneficial for the thermal image as well, especially since a 100 mm lens was used, but the chessboard pattern for the used long range thermal camera needs to be very large and made of two materials with a big difference in reflectivity. This would make camera calibration very expensive and complicated. Thermal image distortion introduces nonlinearity that the homography method cannot deal with, and AI powered methodology can overcome this problem.

The two input MLP follows the same idea as DisNET [15] and gives a good estimation of distance. Introducing additional inputs to MLP4 and MLP5 improves network accuracy. DisNET and Multi-DisNET [15] use the class of the object as an additional input and therefore represent a much more complex structure, that needs a much larger dataset for training at the same time. The proposed MLP2,

MLP4 and MLP5 are only applicable for distance estimation for one class of objects-persons. Still, for training of the proposed MLPs no big data set was needed. As expected, 5 - input MLP has the highest R^2 score of 0.9966.

The proposed 3 - input MLP does not rely on the bounding box height or width, so the class of the object detected is of no interest for distance estimation. MLP3 can be used for distance estimation regardless of the class of the object, while having high quality regression with an R^2 score of 0.9899.

It is possible to additionally improve MLP2, MLP4 and MLP5 if the class of the detected object is introduced as an additional input, but this would require the preparation of a much larger training data set.

Additional improvement of the MLP performance can be done through hyperparameter tuning. Namely, adjustment of number of neurons in the hidden layer was done and MLP performance comparison is presented in Tab. 3.

Table 3 Performance comparison for various number of neurons in hidden layer

No. of neurons in hidden layer	No. of epochs	MSE	MAE	RMSE	R^2 score
5	114	416.78	13.04	20.41	0.9939
8	11	591.21	13.25	24.31	0.9914
10	53	693.76	15.87	26.34	0.9898
12	29	563.90	15.96	23.75	0.9918
15	34	412.06	12.40	20.30	0.9939

The results are showing that the networks with 5 and 15 neurons in hidden layer have outperformed the other 3 structures. The diagram of estimation error for every tested structure is presented in Fig. 9.

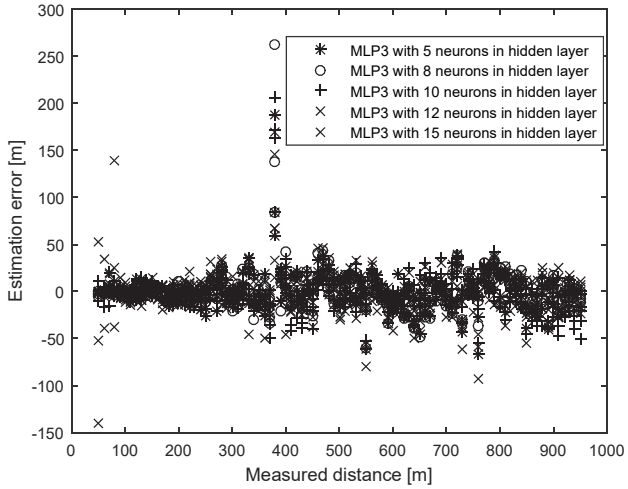

Figure 9 Absolute error of the estimated distances for MLP3 with various number of neurons in hidden layer

The performance of the proposed AI powered distance estimation methods highly depends on the accurate object detection. If the bounding box of the object detected does not have a "low" edge lying in the rail track plane, the estimation of the distance can become highly inaccurate, regardless of the method. This may happen if there is an occlusion between multiple objects, if the object detection algorithm only partly detects the object, or if the object in the real world is not in the rail track plane (such as a bird or a drone).

In order to evaluate the robustness of the proposed MLP5 and MLP3 methods and possibilities of their application in various weather conditions, the evaluation experiments were performed at another location, a level crossing in Žitorađa village, near the city of Niš, Republic of Serbia. The experimental set-up was the same as in the above described field tests, consisting of aODS with thermal camera on a test-stand. The experiments were done in night conditions, with the intensity of illumination of 0. The evaluation results are shown in Fig. 10.

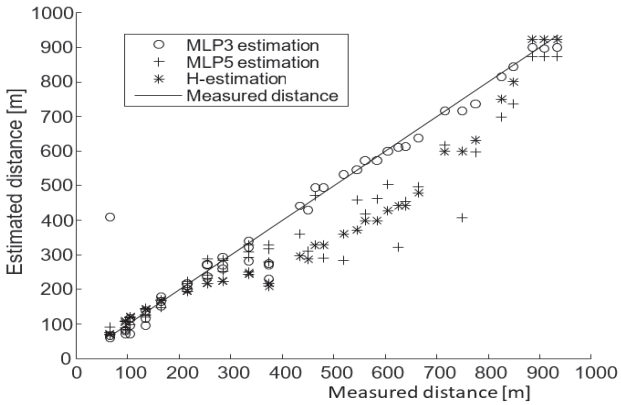

Figure 10 Distance estimation results for robustness validation

The conditions for the evaluation were complex and much different compared to the scenario used for the training data preparation - different location, different railtrack slope, slightly different camera stand position, different weather conditions (outside temperature, humidity and precipitation). Regardless of larger estimation errors (Fig. 11), the presented evaluation results show that both the MLP3 and MLP5 methods are applicable for distance estimation.

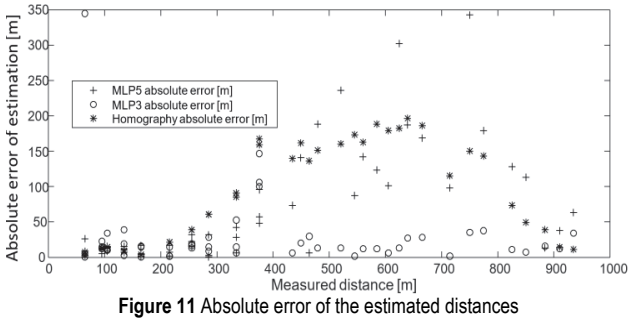

Figure 11 Absolute error of the estimated distances

The change in the scenario showed that simpler MLP3 outperformed MLP5 (Fig.10, Fig. 11), and since the main structural difference is the bounding box width and height used as an additional input of MLP5, it is presumed that the inaccurate bounding box determined through the obstacle detection process caused the estimation error of MLP5. The evaluation results once again show that obstacle detection plays a major role in accurate distance estimation and the main focus of further research should be on the improvement of the OD algorithm.

8 CONCLUSION

In this paper, the image-plane homography for object distance estimation from a thermal camera described in

[14] is improved using an artificial intelligence approach. The goal of the application is obstacle detection and tracking in railways, and the proposed Multi-Layer Perceptron based estimation approaches with 3 and 5 inputs proved valid for future research. The basic point of the research is the image-plane homography that concerns two planes, the image plane and the rail tracks plane, and is applicable for both long and short range obstacle distance estimation. The proposed methods are to be used for distance estimation in day and night light conditions, while improving the accuracy of the traditional image-plane homography method. The proposed methods use a multilayer perceptron and do not need a large training data set. While the MLP5 estimator is more accurate, at this point it can only be used to estimate the distance from the ODS system to persons in the vicinity of the railtracks. On the other hand, the proposed MLP3 estimator still has a high quality regression but it can estimate the distance of any object on or near the railtracks with high accuracy. The adjustment of the number of neurons in the hidden layer can slightly improve accuracy and MLP3 with 5 and 15 neurons in hidden layer have shown the best performance according to MSE, MAE, RMSE and R^2score. The novelty of this approach lies in its usage of the traditional homography method results as an additional input of MLP.

The evaluation results show that the object detection algorithm plays a major role in distance estimation, and further research can be focused on the improvement of the detection and segmentation algorithms. This can be done with the help of various AI methodologies, but possible further studies in both image processing and distance estimation can be done if the problem is addressed as a time series problem. If the states acquired in the previous frames are used, they can both help in estimating the distance and improving the accuracy of object detection.

Acknowledgment

This research was financially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-9/2021-14/200109). Data used in this paper were acquired during the realisation of the HORIZON 2020 S2R-OC-IP5-01-2015 project "SMart Automation of Railway Transport - SMART".

9 REFERENCES

[1] Pieriegud, J. (2018). *Digital Transformation of Railways.*Siemens Sp. z o.o. ISBN 978-83-950826-0-3.

[2] Dimitrov, L., Purgic, S., Tomov, P., & Todorova, M. (2018). Approach for Development of Real-Time Marshalling Yard Management System. *Proceedings ofInternational Conference on High Technology for Sustainable Development, HiTech 2018*. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/HiTech.2018.8566369

[3] Tomov, P. & Dimitrov, L. (2019). The role of digital information models for horizontal and vertical interaction in intelligent production. *Facta Universitatis-Series Mechanical Engineering*, *17*(3), 397-404, ISSN: 0354-2025 https://doi.org/10.22190/FUME190422037T

[4] Zhao, X. F., Liu, H. Z., Lin, S. X., & Chen, Y. K. (2020). Design and Implementation of a Multiple AGV Scheduling Algorithm for a Job-Shop. *Int. Journal of Simulation Modelling. 19*(1), 134-145. https://doi.org/10.2507/IJSIMM19-1-CO2

[5] Wang, D., Tan, K., Dong, Y., Yuan, G., & Du, X. (2020). Estimating the position and orientation of a mobile robot using neural network framework based on combined square-root cubature Kalman filter and simultaneous localization and mapping. *Advances in Production Engineering & Management, 15*(1), 31-43. https://doi.org/10.14743/apem2020.1.347 APEM

[6] Mousavi, M., Yap, H. J., Musa, S. N., & Dawal S. Z. M. (2017). A Fuzzy Hybrid GA-PSO Algorithm for Multi-Objective AGV Scheduling in FMS. *Int. Journal of Simulation Modelling. 16*(1), 58-71. https://doi.org/10.2507/IJSIMM16(1)5.368

[7] Pavlović, M. (2020). *Application of intelligent machine vision systems for autonomous train operation* (Doctoral dissertation), Retrieved from PHAIDRA.

[8] Fernández-Caballero, A., Castillo, J. C., Martínez-Cantos, J., & Martínez-Tomás, R. (2010). Optical flow or image subtraction in human detection from infrared camera on mobile robot. *Robotics and Autonomous Systems*, *58(12)*, 1273-1281. https://doi.org/10.1016/j.robot.2010.06.002

[9] Ćirić, I. T., Ćojbašić, Ž. M., Ristić-Durrant, D. D., Nikolić, V. D., Ćirić M. V., Simonović, M. B., & Pavlović, I. R. (2016). Thermal vision based intelligent system for human detection and tracking in mobile robot control system. *Thermal Science, 20*, S1553-S1559. https://doi.org/10.2298/TSCI16S5553C

[10] Bertozzi, M., Broggi, A., Caraffi, C., Del Rose, M., Felisa, M., & Vezzoni, G. (2007). Pedestrian detection by means of far-infrared stereo vision. *Computer Vision and Image Understanding, 106*, 194-204. https://doi.org/10.1016/j.cviu.2006.07.016

[11] Broggi, A., Fascioli, A., Carletti, M., Graf, T., & Meinecke, M. (2004). A Multi-resolution Approach for Infrared Vision-based Pedestrian Detection. *Proceedings of the IEEE Intelligent Vehicles Symposium* (pp. 2313-2316).Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/IVS.2004.1336347

[12] Forth, A. & Zamjatnins, F. (2015). Night-vision device for railway vehicles for improving safety. *Signal+Draht, 107*, 38-43. Retrieved from https://eurailpress-archiv.de/SingleView.aspx?show=22079&lng=en

[13] Berg, A., Öfjäll, K., Ahlberg, J., & Felsberg, M. (2015). Detecting Rails and Obstacles Using a Train-Mounted Thermal Camera. *Proceedings of the 19th Scandinavian Conference, SCIA 2015*, 492-503. https://doi.org/10.1007/978-3-319-19665-7_42

[14] Pavlović, M. G., Ćirić, I. T., Ristić-Durrant, D., Nikolić, V. D., Simonović, M. B., Ćirić, M. V., & Banić, M. S. (2018). Advanced Thermal Camera Based System for Object Detection on Rail Tracks. *Thermal Science*, *5*(22), S1551-S1561. https://doi.org/10.2298/TSCI18S5551P

[15] Haseeb, M. A., Ristić-Durrant D., Banić, M., & Stamenković, D. (2019). Multi-DisNet: Machine Learning-Based Object Distance Estimation from Multiple Cameras.In *Proceedings of International Conference on Computer Vision Systems*, 457-469. https://doi.org/10.1007/978-3-030-34995-0_41

[16] Salman, Y. D., Ku-Mahamud, K. R., & Kamioka, E. (2017). Distance measurement for self-driving cars using stereo camera. *Proceedings of 6th International Conference on Computing and Informatics ICOCI 2017*, 235-242.

[17] Leu, A., Ristić-Durrant, D., & Gräser, A. (2011). A robust markerless vision-based human gait analysis system. *Proceedings of 6th IEEE international symposium on applied computational intelligence and informatics (SACI)*, 415-420. https://doi.org/10.1109/SACI.2011.5873039

[18] Lee, S., Hayes, M. H., & Paik J. (2013). Distance estimation using a single computational camera with dual off-axis color filtered apertures. *Optics Express*, *21*(20), 23116-23129. https://doi.org/10.1364/OE.21.023116

[19] Rahman, A., Salam, A., Islam, M., & Sarker, P. (2008). An Image Based Approach to Compute Object Distance. *International Journal of Computational Intelligence Systems*, *1*(4), 304-312. https://doi.org/10.1080/18756891.2008.9727627

[20] Deepu, R., Murali, S., & Raju, V. (2013). A mathematical model for the determination of distance of an object in a 2D image. In *Proceedings of International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV),* The Steering Committee ofThe World Congress in Computer Science, Computer Engineering and AppliedComputing (WorldComp).

[21] Dong X., Zhang, F., & Shi, P. (2014). A novel approach for face to camera distance estimation by monocular vision. *International Journal of Innovative Computing, Information and Control, 10*(2), 659-669. Retrieved from http://www.ijicic.org/ijicic-13-01001.pdf

[22] Deng-Yuan, H., Chao-Ho, C., Tsong-Yi, C., Wu-Chih, H., & Kai-Wei, F. (2017). Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburbroads. *Journal of Visual Communication and Image Representation*, *46*, 250-259. https://doi.org/10.1016/j.jvcir.2017.04.006

[23] Giseok, K. & Jae-Soo, C. (2012). Vision-Based Vehicle Detection and Inter-Vehicle DistanceEstimation for Driver Alarm System.*Optical Review, 19*(6), 388-393. https://doi.org/10.1007/s10043-012-0063-1

[24] Hoi-Kok, C., Wan-Chi, S., Steven, L., Lawrence, P., & Chiu-Shing, N. (2012). Accurate DistanceEstimation Using Camera Orientation Compensation Technique for Vehicle DriverAssistance System. *Proceedingsof IEEE International Conference on Consumer Electronics (ICCE)*. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ICCE.2012.6161840

[25] Qi, S. H., Li, J., Sun, Z. P., Zhang, J. T., & Sun, Y. (2019). Distance Estimation of Monocular Based on Vehicle Pose Information. *Journal of Physics: Conference Series, 1168*, 1-8. https://doi.org/10.1088/1742-6596/1168/3/032040

[26] Tuohy, S., O'Cualain, D., Jones, E., & Glavin, M. (2010). Distance Determination for an Automobile Environment using Inverse Perspective Mapping in OpenCV. In*Proceedings ofIET Irish Signals and Systems Conference (ISSC 2010)*.Institute of Engineering and Technology. https://doi.org/10.1049/cp.2010.0495

[27] Adamshuk, R., Carvalho, D., Neme, J. H., Margraf, E., Okida, S., Tusset, A., Santos, M. M., Amaral, R., Ventura, A., & Carvalho, S. (2017). On the Applicability of Inverse PerspectiveMapping for the Forward Distance Estimation based on the HSV Colormap. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT)*. Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ICIT.2017.7915504

[28] Rezaei, M., Terauchi, M., & Klette, R. (2015). Robust Vehicle Detection and Distance Estimation Under Challenging Lighting Conditions. *IEEE Transactions On Intelligent Transportation Systems*, *16*(5), 1-20. https://doi.org/10.1109/TITS.2015.2421482

[29] Gökçe, F., Üçoluk, G., & Şahin, S. K. E. (2015). Vision-based detection and distance estimation of micro unmanned aerial vehicles.*Sensors, 15*(9), 23805-23846. https://doi.org/10.3390/s150923805

[30] Ristić-Durrant, D. (2019). *Deliverable D2.1. Report in selected sensors for multi-sensory system for obstacle detection*, Retrieved from Project SMART Smart Automation of Rail Transport (Project reference - 730836) website http://www.smartrail-automation-project.net/

[31] Johnson, J. (1958) Analysis of image forming systems,*Image Intensifier Symposium, AD 220160*, Warfare Electrical Engineering Department, U.S. Army Research and Development Laboratories, Ft. Belvoir, Va., pp. 244–273.

[32] Sjaardema, T., Smith, C., & Birch, G. (2015). *History and Evolution of the Johnson Criteria* (Tech. Rep.No. SAND2015-6368), Retrieved from Sandia National Laboratories website: https://www.osti.gov/biblio/1222446. https://doi.org/10.2172/1222446

[33] FLIR TAU 2. Retrieved February 10, 2021, from https://www.oemcameras.com/flir-tau2-640-100mm-7-5-thermal-imaging-camera-core.htm

[34] Pavlović, M., Nikolić, V., Simonović, M., Mitrović, V., & Ćirić, I. (2019). Edge DetectionParameter Optimization Based on the Genetic Algorithm for Rail Track Detection. *FACTA UNIVERSITATIS, Series: Mechanical Engineering*, *17*(3), 333-344. https://doi.org/10.22190/FUME190426038P

[35] Hartley, R. & Zisserman, A. (2004).*Multiple View Geometry in Computer Vision*. New York, USA: Cambridge University Press. https://doi.org/10.1017/CBO9780511811685

[36] Car, Z., BaressiŠegota, S., Anđelić, N., Lorencin, I., & Mrzljak, V. (2020). Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron. *Computational and Mathematical Methods in Medicine, 20*. https://doi.org/10.1155/2020/5714714

[37] Šegota, S. B., Lorencin, I., Anđelić, N., Mrzljak, V., & Car, Z. (2020). Improvement of Marine Steam Turbine Conventional Exergy Analysis by Neural Network Application. *Journal of Marine Science and Engineering*, *8*(11), 884. https://doi.org/10.3390/jmse8110884

[38] Gumbarevic, S., Milovanovic, B., Gaši, M., & Bagaric, M. (2020). Application of Multilayer Perceptron Method on Heat Flow Meter Results for Reducing the Measurement Time. *Engineering Proceedings*, *2*(29). https://doi.org/10.3390/ecsa-7-08272

[39] Sapna, S., Tamilarasi, A., & Kumar, M. P. (2012). Backpropagation learning algorithm based on Levenberg Marquardt Algorithm. *Computer Science & Information Technology(CS and IT)*, 2, 393-398. https://doi.org/10.5121/csit.2012.2438

Contact information:

Ivan ĆIRIĆ, PhD, Assistant professor
(Corresponding author)
Faculty of Mechanical Engineering,
University of Niš,
Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: ivan.ciric@masfak.ni.ac.rs

Milan PAVLOVIĆ, PhD, Lecturer
Academy of Technical-Educational Vocational Studies Niš
Aleksandra Medvedeva 20, 18000 Niš, Serbia
E-mail: milan.pavlovic@akademijanis.edu.rs

Milan BANIĆ, PhD, Assistant professor
Faculty of Mechanical Engineering,
University of Niš,
Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: milan.banic@masfak.ni.ac.rs

Miloš SIMONOVIĆ, PhD, Assistant professor
Faculty of Mechanical Engineering,
University of Niš,
Aleksandra Medvedeva 14, 18000 Niš, Serbia
milos.simonovic@masfak.ni.ac.rs

Vlastimir NIKOLIĆ, PhD, Full professor
Faculty of Mechanical Engineering,
University of Niš,
Aleksandra Medvedeva 14, 18000 Niš, Serbia
E-mail: vlastimir.nikolic@masfak.ni.ac.rs