TOWARDS AUTONOMOUS OBSTACLE DETECTION IN FREIGHT RAILWAY

Danijela RISTIĆ-DURRANT[1]
Ivan ĆIRIĆ[2]
Miloš SIMONOVIĆ[3]
Vlastimir NIKOLIĆ[4]
Adrian LEU[5]
Branislav BRINDIĆ[6]

Abstract – *In order to increase the quality of rail freight, as well as its effectiveness and capacity, a level of automation of railway cargo haul at European railways needs to be increased. An important part of the autonomous operation of cargo haul at European railways shall be a complete, safe and reliable obstacle detection system to be used mainly for initiation of long distance forward-looking braking and short distance wagon recognition for shunting onto buffers. Such obstacle detection system shall be integrated into Autonomous Train Operation (ATO) module and it shall be a multi sensory system in order to provide fail safe and reliable obstacle detection at short (up to 20 m) and long range (up to 1000 m) during day and night operation, as well as operation during impaired visibility (such as in the case of fog and bad weather condition). In this paper, the state-of-the art in obstacle detection in railway as well as problems and lacks that should be overcome to achieve such autonomous obstacle detection are presented.*

Keywords – Rail Freight, Obstacle Detection, Machine Vision, Autonomous Train Operation

1. INTRODUCTION

One of the main goals of modern rail freight is to increase the quality, as well its effectiveness and capacity trough development of Autonomous Train Operation (ATO) module with autonomous obstacle detection as one of its key elements. Novel obstacle detection system should be capable to safely and reliably detect obstacles within existing mainline infrastructure at distance of 1000 m, which is sufficient to initiate forward-looking braking. In order to provide fail-safe and reliable obstacle detection, the obstacle detection system shall benefit from a fusion of signals received from multiple sensors such as thermal camera, radar, camera augmented with image intensifier, stereo vision system and lased range finder [Shift2Rail 2016]. Having in mind multi-sensor system, a novel software algorithms used for recognition of obstacles according to the railway demands and safety procedures need to be developed.

A novel fail safe and reliable system for obstacle detection on railway mainlines as well as short distance wagon/buffer recognition for shunting onto buffers, shall be integrated into planned Autonomous Train Operation (ATO) module over standardized interface. The system needs to include hardware and software solutions for obstacle detection at 1000 m in a rail-specific safety framework. Based on captured sensor data, software algorithms should recognize complex patterns ahead of locomotive relevant for rail applications in real-time.

Obstacle detection will significantly contribute to a Grade of Automation 4 (GoA 4) vision of cargo haul trains [Shift2Rail 2016], which will have an overall important impact on increase of quality of rail freight in terms of punctuality, reliability and flexibility, as well its effectiveness and capacity. Such change in rail cargo transport quality and capacity will make rail transport more competitive in regarding the road transport thus enabling the 60% cut in transport emissions by the middle of the century [Shift2Rail 2016].

[1] Institute of Automation, University of Bremen, Otto-Hahn-Allee 1, Bremen, Germany, ristic@iat.uni-bremen.de
[2] University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, Serbia, ciric.ivan@masfak.ni.ac.rs
[3] University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, Serbia, smilos@masfak.ni.ac.rs
[4] University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, Serbia, vnikolic@masfak.ni.ac.rs
[5] Institute of Automation, University of Bremen, Otto-Hahn-Allee 1, Bremen, Germany, aleu@iat.uni-bremen.de
[6] Harder Digital Sova, Bulevar cara Konstantina 80-82, 18000 Nis, Serbia, branislav.brindic@hdsova.rs

2. STATE OF THE ART

Several papers regarding the obstacle detection in railways with the purpose of improving the safety were published.. In majority of the literature, the main focus is on the existence of objects on the tracks in critical areas such as zones close to bridges, level crossings, and in the stations. However, the incidents on the railroad are not only limited only to those areas and they might happen in any spot along the track.

In general, there are two types of proposals to address this problem: a) sensory systems are stationary and placed in situ (on-site), and b) sensory systems are mobile and placed on-board trains. As it was claimed by many of the scholars in the past, in situ systems turned out to show more promising results as if a high speed train running at 350 km/h needs a distance of 2.5 km to come to a complete halt, the detection of an obstacle within this range used to be almost impossible so that multi-sensory systems placed on board trains were rejected [Garcia-Dominguez et.al. 2008]. Although it is easier to develop the in situ systems, which do not face complexity of the mobile sensors such as changing lighting, changing background, and process speed, they have their own drawbacks such as time-consuming installation and calibration, lack of the ability to detect small obstacles, etc. [Amaral et al. 2016]. By the advent of the autonomous vehicle era, making use of on-board systems seemed inevitable and taking advantage of a combination of sensors with advanced technology made their development even less challenging [Uribe et al. 2012].

Several works on detection of obstacles using in situ sensors were published. In the work of [Ohta 2005], the major aim is to detect pedestrians or small vehicles (like wheelchairs) so to avoid their collision with trains. For this purpose, two stereo camera sets (four cameras) were installed at level crossing. There are advantages in using stereo cameras such as their ability to deliver depth information or wide sensing area. Moreover, in comparison with methods such as ultrasonic or optical beam methods less number of sensors is needed. Image processing techniques are applied for obstacle detection. Since the cameras are stationary, background subtraction is successfully applied to detect obstacles. Nevertheless, the change in illumination affects the performance of the proposed system. Adverse weather conditions can also deteriorate the system's functioning.

In [Oh et al. 2007] the focus is on safety concerns in railway platforms, where many pedestrians commute. A big group of different types of sensors provides input data for the central data fusion unit. These sensors are stereo cameras, infrared sensors and thermal cameras, mounted in different points of the train station. Two main classes of hazardous situations are identified: fallen objects (which imply pedestrian falling on the railroad) and intense change in illumination (which may mean fire). Four tasks are defined in processing steps: train detection, object detection, object recognition and object tracking. The reference claims perfect train detection performance in practice. However, as in [Ohta 2005], illumination change is mentioned as a factor which affects stereo camera efficiency.

In [Garcia-Dominguez et al. 2008], the authors use a combination of infrared, ultrasonic, and vision, to overcome the weak points of each sensor with another one. With this configuration, the areas near the bridges are controlled for safety of railway and obstacle avoidance. The placement of sensors is so that a set of infrared and ultrasonic emitters and receivers are placed on both sides of the railway and form a "barrier" in controlled areas. Moreover, the vision sensor is based on a camera, mounted above the barrier. Principal Component Analysis (PCA) has been utilized in this system which has two tasks: to prevent false alarms by IR-US barrier, and to use a vision sensor for moving object detection.

The work of [Akpinarat et al. 2012] is quite similar to [Garcia-Dominguez et.al. 2008], in the sense that both use a barrier of IR-US sensors on two sides of the railroad to cover a rectangular area and to cover the drawbacks of each individual sensor. However, they take a step forward in applying fuzzy controller and Dempster–Shafer evidential theory [Kong et al. 2009] for fusion of information. These improvements are to achieve fewer false alarms in practical applications.

Using a tilting 2D laser scanner to provide a 3D point cloud and investigation of its application in railway level crossings is the subject of reference [Amaral et al. 2016]. The challenge to meet in this work is detection of small obstacles, which may remain hidden from former detection systems. As in [Ohta 2005], and using the fact that the sensory system is stationary, background subtraction is applied to detect moving objects. The hardware used for scanning is a LIDAR sensor from SICK, LMS 111 and the proposed setup is able to detect objects of size 10x10x10 cm. The paper reports high quality of object detection and zero false alarms in evaluation.

Unlike the mentioned works, the following studies have investigated the use of on board sensors. There are three sensor selection strategies for design and implementation of on-board systems, especially for autonomous train driving systems:

- Active: Emitting devices and associated sensors are placed on the train's head for scanning the near area which has the drawbacks of the short range of action and low accuracy on curve zones;
- Passive: Use less costly video cameras in front of the train and rely on image processing algorithms which have high computational requirements as

they need to process the acquired images in real-time or close to real-time but give more important data for object identification.

- Mixed strategy: Integrating the multiple sources of information to make up disadvantages of existing methods and split the working range to use both passive and active strategies.

In the autonomous driving for cars several related works are done by the combination of sensors. The most common approach to vehicle detection is using active sensors such as radar-based (i.e., millimeter-wave), laser-based (i.e., LIDAR), and acoustic-based, since they can measure certain quantities (e.g., distance) directly without requiring powerful computing resources and they can "see" at least 150 meters ahead in fog or rain, where average drivers can see through only 10 meters or less. In the literature, the majority of methods reported to detect the cars and obstacles follow two basic steps: 1) HG where the locations of possible object in an image are hypothesized which is done using a)knowledge-based methods employ a priori knowledge to hypothesize object locations in an image; b)stereo-based approaches take advantage of the Inverse Perspective Mapping (IPM) to estimate the locations of vehicles and obstacles in images; c) Motion-based methods detect vehicles and obstacles using optical flow; and 2) HV where tests are performed to verify the presence of object in an image. A comprehensive overview can be found in [Sun et al., 2006].

In Boss system [Urmson et al. 2008], the team preferred to use active sensors for obstacle detection in urban environment, since the direct measurement of range and target velocity was more important than getting richer, but more difficult to interpret, data from a vision system. They used a combination of several radar and scanning laser sensors with a variety of ranges up to 300 m. For each sensor type a specialized sensor layer is implemented, therefore the architecture enables new sensor types to be added to the system with minimal changes to the fusion layer. LIDAR scans are processed to building a 3D point cloud of curbs. Then by mitigation of false positives due to occlusions and sparse data, clustering the points initially based on distance between consecutive points, and using the dense/sparse labelling from preprocessing, sparse points are removed from the classification list, the resulting list represents the locations of the likely road and surrounding geometric cues.

The work of [Alessandretti et al. 2007] investigated how the results of different sensors could be fused together, benefiting from the best performance of each sensor, since different radar performances suggest different fusion levels. The methods differ mainly in the fusion level: Low-level fusion combines several sources of raw data to produce new raw data that are expected to be more informative and synthetic than the inputs. In intermediate-level fusion, various features such as edges, corners, lines, texture parameters, etc., are combined into a feature map that is then used by further processing stages. In the high-level fusion, each source of input yields a decision, and all the decisions are combined. They used vehicle detection algorithm based on symmetry and uses radar data in order to localize areas of interest. Data fusion operates at a high level: The vision system is used to validate radar data and to increase the accuracy of the information they provide. A long-range radar with a 77-GHz frequency that is not capable of data classification is used to detect non-vehicle objects (mainly guard rails). Two scanned radars with a 24-GHz frequency mounted above the front bumper and connected to a dedicated Electronic Control Unit (ECU) are used to detect obstacles up to 40 m with an accuracy of 0.25 m. Radar and camera calibration are performed separately, and then the two outputs are double-checked together using an obstacle in a known position.

The following works tried the mentioned strategies. [Piva et al. 2003] studies an image stabilization technique that can be used on board railway surveying systems. The general approach for image stabilization is motion estimation and compensation, proceeded by the image composition. The performance of the stabilizer is investigated on REOST (Railway Electro Optical System for Safe Transportation) system in which, a camera and a stabilizer have been mounted at the top of the locomotive and digital image stabilization. In this stabilization method, each image is characterized by two one-dimensional curves. Image instability is estimated and compensated by comparing two consecutive frames and the difference in their characteristic curves. Achieving real-time application with 14.2 frames per second is an advantage for the proposed solution.

[Uribe et al. 2012] focused on obstacle detection and tracking from the view of a camera mounted on the locomotive. This aim is achieved by rail detection and investigation of an area in the neighborhood of the rails for potential obstacles (constant or moving), as well as by detection of moving objects approaching far from the rails. Rail detection is achieved by the Hough transform, which is a well-known method for line detection. Detection of moving objects is done with Lucas-Kanade [Lucas et al. 1981] optical flow method. The image features for which optical flow is measured are obtained by Shi-Tomasi method [Shi et al. 1994]. This method chooses regions of the image, which have more chance to be repeated in consecutive frames. Nevertheless, some workarounds are needed to be able to discriminate between background and moving object pixels. The paper concurs the need of more sophisticated methods for practical application, although reports satisfactory results for evaluation of the proposed system.

In [Rüder et al. 2003], they used three progressive scan CCD video cameras as main sensors. Two are used in a stereo setup for monitoring the near range up to a distance of about 50 meters and one camera in the far range. For the far range a multi beam infrared radar senor originating from an adaptive highway cruise control system for cars is used with a detection range of 150 meters. The system receives measurements of velocity and position data along the track from the vehicle control computer as well as from a differential GPS receiver. A combination of several algorithms such as track gaps, properties of edge elements, grey value variance and correlation along the track, optical flow, statistics of textures and stereo by inverse perspective mapping were used to stabilize the results. The system detected objects of the size of 0.4 m2 up to a distance of 250 meters. A larger object would be detected from a greater distance.

Dynamic programming is used by [Kaleli et al. 2009] in order to extract the train course and left and right rails simultaneously in front of the train using the Hough transform and compute the optimal path which gives the minimum cost to extract the railroad track space. Discarding background moving elements the algorithm finds candidate dangerous objects, tracks their trajectories and foresees their paths for determining if exist a course to collision. If an object could pose danger to the safety of the train a warning with anticipation is shown.

[Möckel et al. 2003] used look-ahead sensors such as Video cameras (optical passive), LIDAR (optical active) and RADAR (electromagnetic active) and acquired up to 400 m look-ahead range and with driving speed up to 120 km/h. To get maximum confidence in interpretation the concept of redundancy to choose a multi-sensor approach with stand-alone sensor units is used. The fusion of active and passive optical sensors and a railway track data base leads to very robust system performance. In small and medium distances, the detection probabilities of the technical system seem to be higher than those of the human being. However, for large distances the human driver's detection capabilities are better in comparison to the technical system.

[Inaba 2011] employed the millimeter wave high resolution radar for the range up to 300 meters. A high level of detection performance achieved by combining in a complementary way a radar sensor, offering long-distance detection performance, and an optical image sensor. A pan/tilt pan-head was developed with an automatic tracking function for the point of regard, and vibration isolation measures to minimize camera shake using Gyrostabilizer. Rail detection is performed by matching edge matching features extraction using the Chamfer distance and by pre-smoothing the input image with a 3x3 mean filter and using a symmetrical local. For long distance, the edges are extracted by pre-smoothing the input image with a 3x3 Gaussian filter and hysteresis thresholding the responses to a 7x7 Sobel operator. [Masato 2006] applied a passive detection method using a single-lens on-board camera (low-brightness CCD) to detect obstacles in front of the train in day and night. A telephoto lens is used to monitor long-distance perspectives and compensate the probable blur effects from vibration using optical flow. When an interruption in rail continuity was detected, system assumes the possibility of a stationary obstacle resting on or beside the rail(s), then it isolates the region in which the obstacle existed by analyzing the average and dispersion of brightness and the brightness profiles projected horizontally and vertically. Other characteristic points of the obstacle were extracted from the isolated region and analyzed on time-spatial photographic images.

[Berg et al. 2015] used a FLIR SC655 thermal camera to detect obstacles on or near the rails in front of the train at a limited range. The system acquires thermal images, computing the scene geometry (homography from pixel to ground coordinates using inverse perspective mapping), detecting the railway (possible rail locations), detecting foreground objects as well as anomalies, tracking possible obstacles, and giving alarms to the driver conditioned on certain criteria. In the study by [Kruse et al. 2003], it is claimed that there is no single sensor system available which has the capability to fulfil all given requirements, therefore, a low cost multi sensor system comprising a tele-camera, a far distance 24 GHz radar Sensor, a survey camera and a near distance radar network is designed. Two video cameras are situated behind the front windshield to be protected against weather impact. The acquired data from the sensors are combined using a Kalman-Filter based intelligent fusion method. The images of the cameras determine the track's curvature and boundaries. Using the gainful data fusion a precise estimation of the state (position and velocity) of the detected objects in front of the train is acquired. Then, if a detected object is in the driveway, target tracking is activated. On the basis of rail information now the decision is made whether one or more of the detected objects are situated in the rails. In this case, an alarm message is sent to the vehicle control unit and a corresponding reaction like whistling or braking follows. The measurement results show a high potential of the considered multi sensor system. In [Weichselbaum et al. 2013], stereo vision engine S3e with hybrid SW–HW dense stereo matching and Riegl laser scanner is used to precise 3D sensor data. The effective evaluation and fusion of their signals contributes to a useful recognition performance of obstacles inside the tracks clearance volume with at least a size of 0.3 m x 0.3 m x 0.3 m at a distance from 10 m up to 80 m ahead with a latency time less than 300 ms from capture time to obstacle report time. For the future work they suggested applying a stereo vision system with thermal infrared cameras to work reliably in all weather and light conditions.

When considering particularly stereo-vision based systems either in railways or in car driver assistance systems, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene, in a manner similar to human binocular vision. In general, two types of methods are used to interpret the data obtained from stereo vision sensors in driver assistant systems: Disparity maps and Inverse Perspective Mapping (IPM).

Although computation of disparity map is time consuming, reference [Mandelbaum et al. 1998] shows that it can be done in real-time. A review on different methods of depth estimation using stereo camera system can be found in [Mohan et al. 2015]. In general, in disparity maps, position of pixels on both cameras that represent the same object are compared. The third component of pixel vector (z coordinate) can be calculated by comparing changes in pixels position, while inverse perspective mapping method is based on transforming the image and removing the perspective effect from it [Kovacic et al. 2013). An early example of applying inverse perspective mapping can be found in [Zhao et al. 1993], where the contours of the objects which stand above the ground are extracted from the images. Stereo cameras are mostly used for pedestrian and vehicle detection [Kovacic et al. 2013). Other tasks like lane detection and sign detection, though, can also be tackled with stereo cameras. But pedestrian detection and vehicle detection fully reflect the necessity for three dimensional understanding of the environment and the capabilities of utilization of stereo system to fulfill this need.

Since the advent of driver assistant systems, a great amount of work has been devoted to investigate the usefulness of stereo cameras in them. At Carnegie Mellon University’s NavLab, stereo vision is combined with neural network pattern classification of texture features [Zhao et al. 2000]. ARGO system, developed in the University of Pavia, combines stereo camera technology with template matching for detection of shoulders of pedestrians [Broggi et al. 2000]. In another example, and being developed almost at the same time with NavLab and ARGO, Daimlerchrysler applies chamfer matching (contour-based shape matching) on the data obtained from stereo camera to detect pedestrians [Gavrila 2001]. In 1999, Commission of the European Communities sponsors a joint work between some well-known companies which used to work on pedestrian safety systems to bring their work together. Some major car manufacturers like Daimlerchrysler and MAN, along with sensor providers like Siemens, were involved. The deadline of the project was defined to be in 2003 and the resulting work is documented in [Cicilloni 2003]. The work on obstacle and especially pedestrian detection systems continues till today. An example is reference [Nedevschi et al. 2009] in which, motion features and pattern matching based on chamfer distance are used for pedestrian detection. Reference [Kim et al. 2015] is one of the latest works that proves the popularity of stereo cameras in pedestrian detection. In this paper, the process of pedestrian detection is divided into three steps: First step is detection of region of interest using dense depth information. In this phase, using the size and the depth of each ROI, the system decides if the ROI can belong to a pedestrian or not. If the ROI is likely to belong to a human, Histogram of Oriented Gradients (HOG) features are extracted for it and are fed into a Support Vector Machine (SVM) classifier, which decides if the object is pedestrian or not.

Stereo cameras are used extensively in vehicle detection. Most of the applied methods are based on motion detection [Sivaraman et al. 2013], although there are examples of utilization of appearance-based approaches with stereo cameras as well. One is [Badino & Franke 2007], where appearance features are used for scene segmentation and obtaining an understanding of free space. Or in reference [Chang et al., 2005] features like size and height and width are employed to detect vehicles. In some more examples, like [Bak et al. 2010] and [van der Mark et al. 2007], Euclidean distance is measured to cluster point clouds into objects.

The use of occupancy grids [Moravec et al. 1985] for stereo vision object detection is another interesting field. Occupancy grids represent the environment by a grid and calculates the probability of each location in grid belonging to an obstacle. References [Perrollaz et al. 2010] and [Perrollaz et al. 2010] are two examples which use scene tracking and recursive Bayesian filtering to compose occupancy grid for each frame and then employing clustering methods for object detection.

Regarding the range of stereo-cameras reference [Okutomi et al. 1993] points out that there are disadvantages with using two cameras for stereo vision and suggests using a third camera. Nevertheless, according to [Sun et al., 2006] and [Kovacic et al. 2013], the system of only two cameras is preferred in future works due to the computational cost of trinocular-based system.

3. SOLUTIONS AND METHODOLOGY

State-of-the-art research has shown that significant work has been done in fusing different sensors to achieve more reliable object detection. Different combinations of sensors, such as stereo vision, mono cameras, radar and laser, were used. These combinations were mostly used for day vision; noticeable little work has been published on night vision for obstacle detection. Also, so far used combinations of sensors achieved relatively short range obstacle detection. To advance state-of-the-art, a system consisting of different sensors that would enable reliable object detection in short as well as in long range by 1000 meters is needed. Starting from the usual use of stereo-vision as preferable sensor for object detection in vehicle driver assistants systems, the required obstacle detection system shall be developed so that stereo vision system will be fused with other sensors like thermal camera, laser range finder, radar or camera augmented with image intensifier. Such system shall particularly advance state-of-the-art with respect to enabling short and long range obstacle detection in different conditions. Namely system shall be possible to use by day, night and difficult weather conditions. For that purpose, a multi-stereo system and laser shall be fused with a night vision system.

3.1. Night vision system

Night vision systems can use various types of technologies to allow users to see in low or no light scenarios. Each of these technologies has inherent advantages and disadvantages. Integrated night vision systems (INVSs) combine image output from two or more different types of night vision sensors into one composite (fused) image in order to take advantage of the strengths of each type of sensor [SAVER 2013]. The most common form of sensor fusion used for emergency responder applications is the coupling of an image intensifier (I2) with a thermal imaging sensor. Generally, I2 sensors provide an image of the surrounding environment under low-light scenarios, while thermal imaging sensors allow for the identification of objects and targets of interest by showing the thermal signatures of the objects in the environment [Couture & Plotsker 2005]. INVSs that combine these two technologies provide emergency responders with enhanced detection and recognition capabilities in fog, rain, and smog, as well as in poorly illuminated conditions. Those type of systems are very developed specially for military applications and have reasonable price.

Another fusion system with capability to produces true colour imagery is combination of a visible/near infrared (VNIR) colour EMCCD camera and thermal long-wave infrared (LWIR) microbolometer camera [Kriesel & Gat 2010]. The system can run in true color mode in day light down to quarter moon conditions, below this light level the system can functioning in monochrome mode. Those types of systems are relatively new on market and have still high price. Ultraviolet (UV), visible, and infrared (IR) light make up a portion of the electromagnetic spectrum. Only a narrow portion of the electromagnetic spectrum is visible to the unaided human eye. The ability to create an image using portions of the spectrum outside the visible range allows for the visualization and identification of objects that would be difficult or impossible using the unaided human eye [FLIR Systems 2016].

I2 night vision devices operate in the visible and near infrared (NIR) electromagnetic spectrums, while thermal imaging sensors operate in the mid- and long-wave infrared (MWIR and LWIR) spectrums. In addition to creating images from a broader portion of the electromagnetic spectrum, INVSs that utilize I2 sensors require a small amount of ambient illumination to operate effectively. Light entering an I2 sensor is amplified thousands of times to produce a visible image for the end-user. INVSs with thermal imaging sensors are capable of showing temperature differences as small as 2°–3° C between an object and its environment, thus allowing users to readily pick out objects with different heat signatures [Mitianoudis & Stathaki 2008]. Thermal sensors also provide the end user with improved visibility through many weather conditions that typically limit visibility such as fog and haze.

Fusion of images from each night vision sensor can be accomplished either optically or digitally. Optical fusion relies upon an optical combination of sensor images, while digital fusion employs digital signal processing to combine sensor images. Optical fusion in INVSs can be accomplished using a beam combiner to fuse images from two sensors into one image. Digital fusion INVSs uses digital signal processing to combine the images from each sensor into a single image.

Trend of night vision technologies in this moment is based on two actual opto-electronic technologies - image intensifiers and thermal vision. Every of those technology has own benefits and own limitations. The fusion system benefits are:

- redundancy, since we have two night vision channels
- fusion picture has an emphasized contour of the object detected from thermal image.

This approach is very rational and not expensive solution compare for high level task of night vision recognition

4. CONCLUSION

The primary goal of increasing the quality of rail freight, as well its effectiveness and capacity, is in line with European transport strategy 2011-2021 (Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system). Initiatives proposed have a goal to shift 30% of road freight over 300 km to other transport modes such as rail or waterborne transport by 2030, and more than 50% by 2050. In order to achieve such vision, developing of new infrastructure will be necessary, as well as increasing of efficiency and throughput of the existing infrastructure.

One of the ways to increase of efficiency of existing infrastructure is to automate the cargo haul. Trains are more suited for autonomous operation than other types of vehicles (especially road vehicles) as they are moving on a fixed and known track. Nevertheless, most of the innovation in autonomous vehicles is occurring on the road due to lack of innovation in railway automation as a consequence of railway heavy regulation and extreme focus on safety. Noted regulations and safety focus hampers the implementation of recent innovations in autonomous driving and pushes cargo transport to road with much higher risks and casualty levels.

Very important part of the autonomous train operation is the novel obstacle detection system for usage in automation of cargo haul. The developed obstacle detection system would make an important contribution to automated cargo haul trains necessary for implementation of EU transport strategy. Autonomous trains are significantly more efficient, safer, cheaper and have less downtime than human-operated trains. Shift2Rail Multi -annual Action Plan [Shift2Rail 2015] outlines numerous advantages of cargo haul automation which include the benefits for quality of service (due to better punctuality), increase of capacity (10 – 50%), reduced system costs (20% energy saving), reduction of operation costs (50% reduction of cost for drivers) and overall efficiency increase of 10 %. All of these benefits will result in a system cost reduction in the three-digit million Euro range, as well as great customer benefits [Shift2Rail 2015]. The cargo haul automation is of highest priority for the future of European rail freight and it is one of innovations which will lead to turnaround necessary for achievement of transport strategy goals.

ACKNOWLEDGEMENT

This research was supported by the German-Serbian bilateral project funded by Serbian Ministry of Education, Science and Technological Development and DAAD named "Robust Action Recognition for Human-Robot Synergy".

REFERENCES

[1] Akpinar, B & Gülal, E 2012, `Multisensor railway track geometry surveying system`, *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 190–197.

[2] Alessandretti, G, Broggi, A & Cerri, P 2007, Vehicle and guard rail detection using radar and vision data fusion`, *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 95–105.

[3] Amaral, V, Marques, F, Lourenço, A, Barata, J & Santana, P 2016, `Laser-based obstacle detection at railway level crossings`, *Journal of Sensors*, pp. 1–11.

[4] Badino, H, Uwe F & Rudolf M 2007, Free space computation using stochastic occupancy grids and dynamic programming`, *Workshop on Dynamical Vision, ICCV, Rio de Janeiro, Brazil*. vol. 20.

[5] Bak, A, Samia, B & Didier, A 2010, `Detection of independently moving objects through stereo vision and ego-motion extraction`, *Intelligent Vehicles Symposium IEEE*.

[6] Berg, A, Öfjäll, K, Ahlberg, J & Felsberg, M 2015, `Detecting rails and obstacles using a train-mounted thermal camera`, *In Image Analysis* pp. 492–503.

[7] Broggi, A, et al. 2000, `Shape-based pedestrian detection`, *Proceedings of the IEEE Intelligent Vehicles Symposium*.

[8] Chang, P, et al. 2005, `Stereo-based object detection, classification, and quantitative evaluation with automotive applications, *Computer Vision and Pattern Recognition-Workshops, CVPR Workshops. IEEE Computer Society Conference on. IEEE*.

[9] Cicilloni, R, Project IST – 1999 – 10107, final report, Information Society Technologies: Systems and Services for the Citizen, 2003

[10] Couture M & Plotsker V,`Design Challenges and Consideration for Image Fusion in Multi-Spectral Optical Systems`, OASYS Proceedings SPIE 5783, Infrared Technology and Applications XXXI.

[11] FLIR Systems. Available from: http://www.flir.com/home/. [12 January 2016].

[12] Garcia-Dominguez, JJ, Urena, JU, Hernandez-Alonso, A, Mazo-Quintas, M, Vazquez, JF & Diaz, MJ 2008, `Multi-sensory system for obstacle detection on railways`, In Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE (pp. 2091–2096).

[13] Gavrila, DM 2001, `Sensor-based pedestrian protection`, *Intelligent Systems, IEEE* 16.6, pp. 77-81.

[14] Inaba, MUTNNW 2011, `Obstacle detection on railway track by fusing radar and image sensor`, *In World Congress Railway Research*, *Tokyo: The University of Electro-Communications*, p. 12.

[15] Kaleli, F & Akgul YS 2009, `Vision-based railroad track extraction using dynamic programming`, *In Intelligent Transportation Systems, ITSC*'09, 12*th International IEEE Conference on*, pp. 1–6.

[16] Kim, J & Maral M 2015, `Stereo-based region of interest generation for real-time pedestrian detection`, *Peer-to-Peer Networking and Applications* 8.2, pp. 181-188.

[17] Kong, Q-J, Li, Z, Chen, Y & Liu, Y 2009, `An approach to urban traffic state estimation by fusing multisource information`, *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 3,pp. 499–511.

[18] Kovacic, K, Edouard, I & Hrvoje, G, `Computer vision systems in road vehicles: a review`, *the Proceedings of the Croatian Computer Vision Workshop*.

[19] Kriesel JM & Gat N 2010, `True color night vision (TCNV) fusion system using VNIR EMCCD and LWIR microbolometer camera`, *Proceedings SPIE* 7697, *Signal Processing, Sensor Fusion, and Target Recognition* XIX.

[20] Kruse, F, Milch, S & Rohling, H 2003, `Multi sensor system for obstacle detection in train applications`, *Proc. of IEEE Tr.*, pp. 42–46.

[21] Lucas, BD, Kanade, T et al. 1981, `An iterative image registration technique with an application to stereo vision`, *In IJCAI*, vol. 81, pp. 674–679.

[22] Mandelbaum, R et al. 1998, `Real-time stereo processing, obstacle detection, and terrain estimation from vehicle-mounted stereo cameras`, *Applications of Computer Vision*, WACV'98, *Proceedings, Fourth IEEE Workshop on*. IEEE.

[23] Masato, U 2006, `A new system for detecting obstacles in front of a train`, *Railway Technology Avalanche,* p. 73.

[24] Mitianoudis, N & Stathaki, T 2008, `Enhancement of multiple sensor images using joint image fusion and blind restoration`, *Image Fusion, Algorithms and Applications*, pp. 299-326.

[25] Möckel, S, Scherer, F & Schuster, PF 2003, `Multi-sensor obstacle detection on railway tracks`, *In Intelligent Vehicles Symposium*, *Proceedings, IEEE,* pp. 42–46.

[26] Mohan, D & Ranjith AR 2015, `A review on depth estimation for computer vision applications`, positions 4.11.

[27] Moravec, HP & Alberto E 1985, High resolution maps from wide angle sonar`, *Robotics and Automation. Proceedings*, *IEEE International Conference on*. vol. 2.

[28] Nedevschi, S, Silviu, B & Corneliu T 2009, `Stereo-based pedestrian detection for collision-avoidance applications`, *Intelligent Transportation Systems*, *IEEE Transactions on* 10.3, pp. 380-391.

[29] Oh, KL 2007, `A monitoring system with ubiquitous sensors for passenger safety in railway platform`.

[30] Ohta, M 2005, `Level crossings obstacle detection system using stereo cameras`, *Railway Technical Research Institute*, *Quarterly Reports*, vol. 46, no. 2.

[31] Okutomi, M & Takeo, K 1993, `A multiple-baseline stereo`, *Pattern Analysis and Machine Intelligence*, *Transactions on* 15.4, pp. 353-363.

[32] Perrollaz, M, Anne, S & Didier, A 2010, `Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection`, *Intelligent Vehicles Symposium* (IV), *IEEE*.

[33] Perrollaz, M, Yoder, JD & Christian, L 2010, `Using obstacles and road pixels in the disparity-space computation of stereo-vision based occupancy grids`, *Intelligent Transportation Systems* (ITSC), 13*th International IEEE Conference on*.

[34] Piva, S, Zara, M, Gera, G & Regazzoni, CS 2003, `Color-based video stabilization for real-time on-board object detection on high-speed trains`, *In Proceedings - IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 299–304.

[35] Rüder, M, Mohler, N & Ahmed, F 2003, An obstacle detection system for automated trains. *In Intelligent Vehicles Symposium*, *Proceedings*, *IEEE*, pp. 180–185.

[36] SAVER 2013,`Handbook of night vision technologies`, System Assessment and Validation for Emergency Responders.

[37] Shi, J & Tomasi, C 1994,` Good features to track`, In Computer Vision and Pattern Recognition, Proceedings CVPR'94., IEEE Computer Society Conference on, pp. 593–600.

[38] Shift2Rail Joint Undertaking, Multi-Annual Action Plan, Brussels, November 2015,

[39] Sivaraman, S & Mohan MT 2013, `A review of recent developments in vision-based vehicle detection`, *Intelligent Vehicles Symposium*.

[40] Sun, Z, Bebis, G & Miller, R 2006, `On-road vehicle detection: A review`, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, vol. 28, no. 5, pp. 694–711.

[41] Uribe, JA, Fonseca, L & Vargas, JF 2012, `Video based system for railroad collision warning`, *In Security Technology* (ICCST), IEEE *International Carnahan Conference on*, pp. 280–285.

[42] Urmson, C, Anhalt, J, Bagnell, D, Baker, C, Bittner, R, Clark, MN & others. 2008, `Autonomous driving in urban environments: Boss and the urban challenge`, *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466.

[43] Van der Mark, W, Johan, C, Heuvel, VD & Frans, CAG 2007, `Stereo based obstacle detection with uncertainty in rough terrain`, *Intelligent Vehicles Symposium*, *IEEE*.

[44] Weichselbaum, J, Zinner, C, Gebauer, O & Pree, W 2013, `Accurate 3D-vision-based obstacle detection for an autonomous train. *Computers in Industry*, vol. 64, no. 9, pp. 1209–1220.

[45] Zhao, G-W & Shinichi, Y 1993, `Obstacle detection by vision system for an autonomous vehicle`, *Intelligent Vehicles Symposium*, *Tokyo, Japan*, *Proceedings of the Intelligent Vehicles*' 93 *Symposium*.

[46] Zhao, L & Charles ET 2000, `Stereo-and neural network-based pedestrian detection`, *Intelligent Transportation Systems*, *IEEE Transactions on* 1.3, pp. 148-154.