Increasing Road Safety Based on Machine Vision: A Case Study

Aron David Agg
Doctoral School of Multidisciplinary Engineering Sciences (MMTDI)
Széchenyi István University (SZE)
Győr
Email: aggaaron@gmail.com

***Abstract*— The traffic surveillance systems and Intelligent transportation systems (ITS) are important and dynamically developing field of transport infrastructure. The number of cameras placed in road environment has grown exponentially over the past decade. Meanwhile image processing and machine vision-based methods became widely used. Our paper reviews the most commoly used image processing techniques in traffic surveillance systems and it gives a complex picture of their application possibilities. In the second part of the paper, I examine conceptual structure, operation and application possibilities of a specific traffix monitoring system as a case study. We focus on the difference between human vision and machine vision and review the possibility of human-machine and machine-machine communication with the traffic monitoring system.**

Keywords—intelligent traffic system (ITS); traffic surveillance system; traffic analysis; urban traffic; vehicle detection;

I. INTRODUCTION

Today's camera systems are part of most commonly used tools and devices. Most vehicles are able to use cameras for driving and reversing, not to mention the large number of cameras used in vehicles with self-driving and driving assistance functions. Cameras wouldn't play as extremely important role, if they didn't use different image processing algorithms often with artificial intelligence. These techniques allow the efficient utilization of the visual information. Decreasing prices of camera hardwares created opportunities for the cost-effective applicability of cameras in traffic environment.

Cameras located along roads can perform several different surveillances, monitoring, security and safety tasks. We can divide cameras into two groups according to their location [1,2,9]. The first gourp includes cameras located in urban environments, the second group includes cameras located along highways and expressways. Depending on which class they belong to, they also differ in the way they are used and the information they can obtain. The most common tasks are vehicle counting, accident detection, speed measuring, automatic licence plate recognition (ALPR), congestion detection along motorways. The composition of traffic is completely different in urban environment. There are two-wheelers with higher number and cyclists and pedestrians also pose other problems. The main tasks of cameras located in urban environment are the detection of accidents, vehicle counting, traffic management, vehicle identification and detection of traffic violations.

As the tasks are different for cameras placed in different locations, the image processing methods are different too. In the case of cameras used along motorways the traffic is usually arranged in ordely well-defined lanes and vehicles are traveling in parallel at almost steady speed. Therefore, even more simple algorithms are able to perform e.g. traffix counting and vehicle callification tasks or the detection of abnormal traffic situations with more accurate result. In contrast, in an urban environment the trajectory of road users is characterized by frequent accelearions and decelerations, and there are random changes of directions at intersections. The traffic located in urban environment have more transport actors e.g. pedestrians, cyclists, scooters often with electric driving. These make the detection, tracking and classifying of the vehicles very difficult in urban environments. Due to the differences between the two environments different camera types are installed. Along highways, cameras with lower resolutions are enough, but applications in urban environments need more accurate and faster devices e.g. for vehicle identification, ALPR, surveillance of pedestrians etc. [1]

This paper aims to provide a comprehensive review of image processing methods in traffic surveillance systems and present an application of these systems in the field of traffic safety as a case study. In chapter 2. We summarize the most common operating priciples of machine vision techniques in traffic surveillance systems. Chapter 3. peresnts a case study of a traffic surveillance system focussing on traffic safety through human-machine and machine-machine communication. In this chapter we positioned the possible forms of communication on the field of CogInfoCom. The last chapter summarizes the conclusions of this paper about the usability and outlooks to future research.

II. MACHINE VISION METHODS IN TRAFFIC SURVEILLANCE SYSTEMS

In this section, we introduce the parts of vision-based traffic analysis systems through a review of some important studies.

Buch et al. [1] in 2011 carried out a comprehensive review about image processing techniques and machine vision methonds in traffic surveillance systems. At the beginning of their paper, they listed three areas of applicability: vehicle counting, ALPR and incident detection. Befor camera-based solutions they solved the vehicle counting tasks with some sensors, e.g. inductive loops in the road pavement. The drawback of this solution is that it is not able to classify the vehicles. In contrast, vision-based vehicle counting allows counting and classifying in one step. This application is a relevant field of traffic surveillance systems in both areas, urban and highway.

The objective of ALPR is the identification of vehicles. The most common area of application is drive-through at entry and exit points. In these applications it is possible to use the best camera specification and position. More generally, ALPR is more challenging, in terms of vehicle detection and identification on motorways.

A third application area mentionded by the study is the incident detection. This is a broader use, as most applications that do not include the previous two groups can be categorized here, such as accident or quasi-accident detection, stopped vehicle detection, speed control, violation detection in urban environments or moving object tracking. These methods can be used separately, but if combined in a common application, it would be a complex automated traffic surveillance system. An important element of this study is the classification of the machine vision techniques. They are grouped the methods in two classes **top-down** and **bottom-up** approaches. The first step of the top-down approach is to create a model about the background on the video. Then it is able to detect the moving objects on the frames based on this background model. There are more techniques for modelling of the background:

- **Frame Differencing** [10] As the difference between each of the following frames, moving objects will appear on the result image, beacause the traffic cameras are under constant observation conditions in almost all situations.
- **Background Subtraction** [7] This algorithm creates a mean background model from the N pieces of the following frames. There are different types, the general trait of these algorithms is the sensitivity to the changing of the environvment (e.g. parked vehicles, day/night changing).
- **GMM** [14] This is an improved version of the Background Subtraction method mentioned before. It is less sensitive to light and environment changes, but requires a relatively large amount of memory and computing capacity.
- **Graph Cuts** [4] this method solves the problem of foreground separation by a Markov random field approach.
- **Shadow Removal** [11] Based on the detection of moving shadows method.

The model of the background allows the top-down methods to detect the moving objects. The next step is the classification of the detected objects. Calssification groups the detected instancies. To the classification it is nessesary to get some properties about the object. These properties are called fetures. An example to the classification is vehicle counting. When it is necessary to count the different vehicle types separately (e.g. cars, motorbikes, trucks, buses, bikes, pedestrians), then it uses the classification. The traditional classifiers are based on creating a vector from the features and it can represent the type of the object. The goal of this technique is to change the large array (image) to a one-dimensional vector. An important property is that this vector is able to well represent the original image and its type. This feature vector can be created from the image detail that contains the object, this is called **region based technique** [5].

Another approach is based on the edges of the instance. The contours of the objects can represent the original object well this method is called **Contour based technique** [3].

After creating feature vectors with different machine learning methods, we can define a classifier, it is able to generate a label from the input feature vector. To the training we have to create a training set from the data. The vectors can be placed in an N-dimensional space where the Manhattan or Euclidean distance of the coordinates can be measured easily [15]. At this point it is important to mention the **dimension reduction methods**. Not all dimensions of the feature vectors are required, not all coordinates contain useful information. Exploring these can reduce computational capacity and result in greater accuracy [16].

It is able to classify the vectors in N-dimensional space with different techniques. **Clustering** defines groups based on the relative distances of the vectors from the training dataset. [36].

The **NN or KNN algorithms** are other types of the classifying methods. After placing the training set in the N-dimensional space, at the coming of a new unlabeled feature vector, this method places it in the place and finds the nearest neighbor vector or the K pieces of the nearest neighbor vectors and the new instance takes the label of its neighbors [17].

The goal of the **SVM (support vector machine)** is to serach planes or hyperplanes in the N-dimensional space. These planes separate different vector types with the least possible error [28].

Classification by probabilistic methods is the last option. Based on the visual information and the probability of occurrence of vehicle types. A probability estimate is given about the probabilities of the given object belonging to classes [19].

In contrast, the Bottom-Up approach starts with the detection of objects in the first step and categorizes the objects into different categories at the same time as the detection, so it performs the recognition in one step. To detect objects, it is necessary to identify some feature vectors in the image. The study divided them into four groups:

- **Keypoint detection methods:**

  - **Basic patch based**: Based on the values of the pixels or the histogram of the image. It is sensitive to the noise. [38]
  - **SIFT**: The generated features are invariant to the different geometry transformation and the local illumination changes of the image. [20]
  - **SURF**: Same as SIFT, but it uses higher classification speed. [21]
  - **HOG**: It generates a grid on the input image to calculate the feature vector. The number of cells influences the accuracy of the algorithm, so it increases the computation time. [22]
  - **Other methods**: Alongside the most common used methods metioned before, there are other algorithms. These are often the combinations of the previous four methods.
- **Boosting:** This method uses combination of simple classifiers to improve their performance, e.g. AdaBoost [23]
- **Explicit Shap**e: This technique recognizes the parts of the object that must be recognized. Some types of this method: K-fans [24], ISM [25], Alphabets [26].
- **Object recognition without Explicit Shape Structure:** It can filter the large amount of features from image. The training method selects the most effective features with AdaBoost classifier. [30]

Al-Smadi et al. [9] reviewed the machine vision techniques used in ITSs. First part of their study deals with object detection methods able to detect vehicles. Their paper introduces image differencing, background modeling and background subtraction furthermore **predictive background modeling** [31] and **optical flow** [32].

On the other side of the detection techniques there are the appearance-based methods. Buch et al. [1] also dealt with these types of methods. An addition is the user of **Haar-like features** [33]. Based on these features e.g. an Adaboost classifier can be well trainable. These features are sensitive to the vertical, horizontal and symmetrical positions. This feature enables this method to be used in real time application with limited size of training data.

Detection is followed by chapters reviewing recognition and classification. The purpose of recognition is to find the connection between and object in real world and its projection on the plain of the image. The detection objects need to be sorted and classified for the reasons mentioned above.

Husain et al. [8] in 2020 visited the vehicle detection techniques used in the Intelligent Transportation Systems. This study is special because it deals with analysis in foggy environment. It takes a look at the techniques and methods included in the previous two studies. From 2010s the machine learning and neural network applications has become increasingly common in the filed of ITSs. Husain et al. [8] focused on the image processing methods used for ITSs. The machine learning methods and deep neural networks are also part of artificial intelligence. Machine learning methods can be divided into three groups:

- **Supervised learning**: Based on training dataset in which we know the input values and the output value to be assigned to the input, the model learns the correct predictions.
- **Unsupervised learning:** The machine clusters the input data, we don't know the output value to the input.
- **Reinforcrment learning:** The machine stores the experiences of previous runs, so that it is able to improve the performance of the model. In this case, the sysem learns by running or simulating a specific event over and over again.

Deep neural network can have three basic layer types: input layer(s), intermediate hidden layer(s) and output layer(s). Compared to classical neural networks, the neurons of the most commonly used deep neural networks use an activation function. During the training procedure the connected neurons change the weights on their inputs to achive a more accurate output value. The average size of networks has multiplied and the complexity of their structures has also increased in the recent decade. These have resulted in an increase in cuputing capacity; however, we can solve this problem by using GPUs instead of CPUs.

The study [8] summarizes the state-of-the-art neural network-based methods for vehicle detection and recognition.

- **Fast region-CNN** [6]: Compared to previous convolutional neural networks, it has taken several innovative approaches to reduce training and prediction time in addition to increasing the accuracy.
- **Faster region-CNN** [13]: Combining the previous method with a Region Proposal Network (RPN), futher increasing the computational speed.
- **YOLO** (You Only Look Once) [12]: Redmon et al. created a completely new approach to object detection that can provide real-time object detection with appropriate hardwares.
- **SSD** (Single Shot Detector) [35]: Liu et al. in the same year also published their real-time object detection network. This model achieved 80% mean average precision on Pascal VOC2012 dataset.

In traffic surveillance systems and ITSs in addition to vehicle detection the other horizontal element is vehicle tracking. The description of an object can usually be specified by its coordinates, area, and type. After detection and classification, it is necessary to record their trajectory to analyze the movement of traffic users. The trajectory is created by following the movement and trajectory of the objects.

Buch et al. collected several object tracking methods. A common feature of these tracking methods is to find the relationship between objects detected on consecutive images. Then connect the objects and their features. This sequential method creates a historical route for the movement of the vehicle. Tracking can also be based on the different features of the object. E.g. it is able to calculate distance of appearing objects based on central coordinates. Other approach is to calculate the overlap of the objects on the different frames to connect them. Some more complex vehicle tracking algoriths: Kalman filter [27], S-T MRF (spatial-temporal Markov random field) [28], Graph correspondence [29].

Vehicle tracking techniques can be categorized by treating model-based tracking mathods, region-based tracking methods and feature-based tracking methods separately [2].

Model-based tracking methods fit a two or three-dimensional model to the detected vehicle. This allows us to specify its relative orientation. This is a robust tracking method, but requires a lot of computation.

The region-based tracking method approximates the contour of the vehicle with a more simple geometrical shape, which can be e.g. a circle, ellipse, triangle, etc. Then, it is possible to calculate further with the paramters of this simple shape.

Feature-based tracking includes the previously mentioned Kálmán filter-based tracking.

III. Case study

The last section is about the commonly used machine vision techniques in traffic surveillance systems. It is clear that different applications require different image processing methods and this is one of the reasons why automation of vision tasks can be very complex, even if we focus on a relatively small segment of human vision, the perception of moving objects in traffic area. This complexity causes that in most cases the range of tasks that can be automated by machine vision in transport is limited. Through the evolution humans have come to be able to perceive and interpret the environment visually by their eyes at very high accuracy and with very low latency. This ability makes us able to recognize newer and newer situations never seen before, and we can associate them with a previous similar situation. If we try to compare human vision and neural networks used in machine vision, we will soon realize that human perception adaptation is much more flexible. Under the training of deep neural networks, in most cases the training means getting to the optimal weights of the edges and parameters of the network. Howoever, the structure of the network limits its application possibilities. No topology can be applied in general, while the human brain can function much more generally compared to artificial neural networks. [37]

In this section we discuss the application possibilities of a special system that may be able to increase traffic safety through human-machine and machine-to-machine communication under appropriate conditions.

There is a common traffic situation that can give rise to a quasi-accident or even a real accident when at an intersection two road users cross eachother's path, and it is not possible to notice the other road user before arriving to the intersection. This situation can be dangerous for pedestrians too, when they cross a route of some kind of vehicle. The regularization of unseenable junctions can be done through site-specific traffic rules. E.g. at crossroads we can define a priority route, and the other ways have obligation to give priority. This traditional method works as long as we know that all participants in the traffic know and follow this prescribed rule. The main problem with this is the presence of pedestrians and cyclists, for whom knowledge of traffic rules is not a requirement. In such cases traffic situations may arise where a pedestrian or cyclist detects an oncoming vehicle late and an accident occurs.

Another possible commonly used accident prevention technique is the use of traffic mirrors. Due to their shape, these bulging mirrors are able to provide a large viewing angle, however. this feature is their disadvantage too. The image perceptible in the mirror distort the image of the real environment in the direction of the curvature of the mirror. Interpreting a distorted image makes it difficult for even human vision to interpret what is seen, especially when a moving object needs to be detected while moving.

Traffic mirrors often help to avoid accidents by providing additional information to road users about directions that are impossible to see. Starting from the image processing methods described above, it can be recognized that all the tasks performed by the mirror can be automated. The first step is to detect and track objects moving toward the junction. Based on the tracking, we can define the trajectory of the object. In the simplest case we can create rules based on the trajectory, with which objects in dangerous distance to the junction can be detected in time. In this case signals can be generated in the other direction to the junction. But what are these signals. An exact answer to this question can only be given if we know the type and category of traffic participants in the junction. So we have categorized the road users in advance:

- Pedestrians
- Vehicles; human power or machine assistance, and travelling in the same area as pedestrians: e.g. bike, scooter, electrical scooter, electrical bike
- Motorbike
- Vehicles: e.g. cars, buses, trucks

In the first view we collected the most common furms of human-driven transport types that can occur in urban traffic. In the traditional transport system, the dynamic form of traffic control is the use of traffic lights. This method is also suitable for intelligent accident prevention devices, because the drivers know traffic signs. Traffic users can be informed about coming dangerouos situation with traditional traffic light signal e.g. red light signal. In this situation the traffic light signal realizes a one-way communication between pedestrian or driver and the traffic surveillance machine. One-way communication is sufficient for transport in this specifically application, as in this situation the task is to provide additional information for those approaching.

If we approach this communication from the point of view of CoginfoCom, than it is able to define participants of the communication. First of all there is the traffic surveillance system and it is an artificially cognitive entity in this context. Secondly the traffic participants are human entities in the communication. In this case the mode of communication is inter-cognitive communication, because it is between humans and an artificially cognitive system. The type of the communication is sensor-bridging communication and it is icon-based, what need to interpret by the human entities. [39]

It is able to define general regularities, because we are observing an automated system. E.g. in one of the junction's ways the vehicles must be stopped by the machine and the other one gets green light, when it detects collosion hazard. If we define not just traditional signals as a new complex set of rules that can be changed by managing more complex traffic situations. The disadvantage of these is that road users have to be prepared to understand the new signs.

Intralogistics shows many similarities to the previously described problems in terms of traffic accindents. There are narrow traffic routes that often intersect between high and opaque shelves in logistical areas. As a result, devices to those outlined earlier are beginning to spead in logistics [34]. In addition, the number of autonomous material handling vehicles have increased in the last few years. To achieve safe traffic, these autonomos vehicles are equipped with high variety of sensors. By these the vehicle is able to locate itself accurately and real time within the logistics area. Positioning is not only important from a safety conditions point of view; it is also necessary to perform logistical tasks.

This trend is expected to be reflected in road transport in the near future. In terms of previous topics this presents new opportunities. In order to be able to discuss this as well, we added additional categories to the last grouping:

- Autonomous passenger vehicles
- Autonomous freight vehicles

If we think of passenger transport, either public transport, demand-based transport services or even means of transport for tourism can be implemented with autonomous vehicles. In the case of freight transport, the use of material supply vehicles on both road and pedestrian frequencies may arise. In each case, they must operate in the area affected by the traditional transport operators discussed earlier. For this reason, an accident can easily occur, as the robot's sensing is also limited in these locations. The device We are discussing may also be able to provide additional information for autonomous vehicles and robots. In this case, the goal is clearly to implement machine-to-machine communication. Autonomous devices may also be able to perceive visual signals of convetional traffic, however, it is much easier to transmit this information through a communication interface, as they are able to interpret it and perform an intervention operation. This intervention can be a slowdown, a stop, or even an evasive movement.

In this case we can describe the details of the communication again. The participants are obviously defineable, there are two cognitive entities, the traffic surveillance system and the autonomous vehicle. Based on the participants this is an intra-cognitive communication mode. Type of this communication can be two different. One hand it can be the same as the first case with humans. When the traffic surveillance system shows traffic signs, than the vehicle have to interpret them. This is the sensor-bridging case. On the other hand the autonomous vehicles use often vision based methods to getting information from their environment. When the vehicle access the image information of the traffic surveillance system or it can gets the coordinates and trajectory of the detected objects, it can be sensor-sharing communication. [39]

If we look further at the possibility of creating transport routes used by purely autonomous vehicles and where the number of other road users is zero, the vehicles will be able to exchange the information between themselves that is necessary to avoid accidents. Until this is achived, it will be necessary to use methods similar to the one mentioned above.

IV. Conclusions

In this paper we summarized the most commonly used image processing methods for traffic surveillance systems and gave a comprehensive review from the complexity of machine vision field in this topic. In the second half of the paper, we reflected on the applied methods and discussed the applicability of possible accident prevention devices in the filed of traffic. We mentioned several boundary conditions for the application of the method, such as the types of road users and the forms of communication that can be estabilished with them and we described their positions in the field of CogInfoCom. Finally, considering the concept further in terms of autonomous vehicles, we also discussed the basics of a possible communication method. The feasibility of the system is supported by the specifications and real-world performance tests of the discussed image processing methods.

Based on the topics covered in this paper, it is clear that there is scope for multi-directional research in this area. Properly performing the detection task and object tracking are crucial points for the operation of the system. A separate research gap could be the functional and feasibility study of optimal human-machine and machine-machine communication. Based on this study some research is able to aim the field of sensory substitution or icon-based approaches and applications.

It is an observable fact that the vision and neural network based solutions passed the trough of the disillusionment phase of the technological hype cycle and them appeard in a wide field of applications. Them solve successfully the practicall obstacles and there are many possible use cases.

References

[1] N. Buch, S. A. Velastin and J. Orwell, "A Review of Computer Vision Techniques for the Analysis of Urban Traffic," IEEE Transactions on Intelligens Transportation Systems 12, p. 920—939., 2011.

[2] M. Al-Smadi, K. Abdulrahim and R. A. Salam, "Traffic surveillance: A review of vision based vehicle detection, recognition and tracking," International Journal of Applied Engineering Research, pp. 713-726., 2016.

[3] W. Hu, X. Xiao, D. Xie, T. Tan and S. Maybank, "Traffic accident prediction using 3-D model based vehicle tracking," IEEE Trans. Veh. Technol., vol. 53, no. 3, pp. 677-694., 2004.

[4] Y. Boykov and O. Veksler, "Graph Cuts in Vision and Graphics: Theories and Applications BT - Handbook of Mathematical Models in Computer Vision," Handbook of Mathematical Models in Computer

Vision, pp. 79-96., 2006.

[5] N. Buch, J. Orwell and S. A. Velastin, "Detection and classification of vehicles for urban traffic scenes," Proc. Int. Conf. VIE, pp. 182-187., 2008.

[6] R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision, pp. 1440-1448., 2015.

[7] S. Gupte, O. Masoud, R. F. K. Martin and N. P. Papanikolopoulos, "Detection and classification of vehicles," IEEE Trans. Intell. Transp. Syst., vol. 3, no. 1, p. 37–47., 2002.

[8] A. A. Husain, T. Maity and R. K. Yadav, "Vehicle detection in intelligent transport system under a hazy environment: a survey," IET Image Processing, pp. 1-10., 2020.

[9] M. R. Ibrahim, J. Haworth and T. Cheng, "International Journal of Applied Engineering Research," Cities 96, 2020.

[10] K. Park, D. Lee and Y. Park, Video-based detection of street-parking violation, Proc. Int. Conf. Image Process. CVPR, 2007.

[11] A. Prati, I. Miki, R. Cucchiara and M. M. Trivedi, "Comparative evaluation of moving shadow detection algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 7, pp. 918-923., 2003.

[12] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788., 2016.

[13] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-10., 2016.

[14] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., vol. 2, p. 246–252., 1999.

[15] T. Gao, Z. G. Liu, W. C. Gao and J. Zhang, "Moving vehicle tracking based on SIFT active particle choosing," Proc. Adv. Neuro-Inf. Process., vol. 5507, no. Lect. Notes Comput. Sci., pp. 695-702., 2009.

[16] W. Zhang, B. Yu, G. J. Zelinsky and D. Samaras, "Object class recognition using multiple-layer boosting with heterogeneous features," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog, vol. 2, pp. 323-330., 2005.

[17] J. W. Hsieh, S. H. Yu, Y. S. Chen and W. F. Hu, "Automatic traffic surveillance system for vehicle tracking and classification," IEEE Trans. Intell. Transp. Syst., vol. 7., no. 2., pp. 175-187., 2006.

[18] K. R. Muller, S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf, "An introduction to kernel-based learning algorithms," IEEE Trans. Neural Netw., vol. 12., no. 2., pp. 181-202., 2001.

[19] X. Song and R. Nevatia, "Detection and tracking of moving vehicles in crowded scenes," Proc. IEEE WMVC, p. 4., 2007.

[20] D. G. Lowe, "Object recognition from local scale-invariant features," Proc. ICCV, vol. 2, pp. 1150-1157., 1999.

[21] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded-up robust features," Proc. ECCV, vol. 3951, pp. 404-417., 2006.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. IEEE Comput. Soc. Conf. CVPR, vol. 1, pp. 886-893., 2005.

[23] Y. Freund, "Boosting a weak learning algorithm by majority," Inf. Comput., vol. 121., no. 2., pp. 256-285., 1995.

[24] D. Crandall, P. Felzenszwalb and D. Huttenlocher, "Spatial priors for part-based recognition using statistical models," Proc. IEEE Conf. Comput. Vis. Pattern Recog., vol. 1., pp. 10-17., 2005.

[25] N. Cornelis, B. Leibe, K. Cornelis and L. V. Gool, "Three-dimensional urban scene modeling integrating recognition and reconstruction," Int. J. Comput. Vis., vol. 78., no. 2., pp. 121-141., 2008.

[26] I. Creusen, R.Wijnhoven and P. H. N. deWith, "Applying feature selection techniques for visual dictionary creation in object classification," Proc. Int. Conf. IPCV Pattern Recog., pp. 722-727., 2009.

[27] J. Wang, G. Bebis and R. Miller, "Robust video-based surveillance by integrating target detection with tracking," Proc. CVPRW, p. 137., 2006.

[28] S. Kamijo, K. Ikeuchi and M. Sakauchi, "Vehicle tracking in low-angle and front-view images based on spatiotemporal Markov random field model," Proc. 8th World Congr. Intell. Transp. Syst., 2001.

[29] S. Gupte, O. Masoud, R. F. K. Martin and N. P. Papanikolopoulos, "Detection and classification of vehicles," IEEE Trans. Intell. Transp. Syst., vol. 3., no. 1., pp. 37-47., 2002.

[30] A. Opelt, A. Pinz, M. Fussenegger and P. Auer, "Generic object recognition with boosting," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28., no. 3., pp. 416-431., 2006.

[31] K. Karmann, A. Brandt and R. Gerl, "Moving object segmentation based on adaptive reference images," Proc. of 5th European Signal Processing Conference, vol. 2., pp. 951-954., 1990.

[32] Y. Liu, Y. Lu, Q. Shi and J. Ding, "Optical flow based urban road vehicle tracking," IEEE 9th International Conference Computational Intelligence and Security (CIS), pp. 391-395., 2013.

[33] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1., pp. 501-511., 2001.

[34] S. Trab, E. Bajic, A. Zouinkhi, A. Thomas, M. Abdelkrim, H. Chekir és R. Ltaifer, „A communicating object's approach for smart logistics and safety issues in warehouses," Concurrent Engineering Research and Applications , pp. 53-67., 2017.

[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu and A. C. Berg, "SSD: Single Shot MultiBox Detecto," Lecture Notes in Computer Science, pp. 21-37., 2016.

[36] S. Atev, G. Miller, N. P. Papanikolopoulos, "Clustering of vehicle trajectories," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 3, pp. 647-657., 2010.

[37] O. Irsoy and E. Alpaydin, "Continuously Constructive Deep Neural Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 31., no. 4., pp. 1124-1133., 2020.

[38] S. Agarwal, A. Awan and D. Roth, "Learning to detect objects in images via a sparse part-based representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 11, pp. 1475-1490., 2004.

[39] P. Baranyi, A. Csapo and G. Sallai, "Cognitive Infocommunications (CogInfoCom)," Springer International Publishing Switzerland, 2015.