Detection of traffic anomalies for a safety system of smart city

A.Z. Makhmutova[1,2], I.V. Anikin[1,2], R.N. Minnikhanov[1], T.E. Bolshakov[1,2], M.V. Dagaeva[1,2]

[1]Road Safety State Company, Orenburgskij trakt 5, Kazan, Russia, 420059
[2]Kazan National Research Technical University named after A.N. Tupolev – KAI, K. Marks str. 10, Kazan, Russia, 420111

Abstract. For modern smart city with sustainable development we need to provide reasonable level of safety and efficient management of the resources. Instant response to incidents and abnormal situations will help to provide such high bars for city residents, which requires deployment of application of intelligent information processing and data analytics into infrastructure. Closed-circuit television (CCTV) is playing a key part in assurance of city security - most of the modern large cities equip with powerful monitoring systems and surveillance cameras. Video data covers most of the city and could be efficiently used to find anomalies or trends. This hard task for non-stop video monitoring could be solved by modern achievements in machine learning and computer vision techniques, which can automate the process of video analysis and identify anomalies and incidents without human intervention. In this paper, we used computer vision methods like object detection and tracking, as well as neuron networks for classification and detection of anomalies on real time video. As a result of this work we suggested the working approach for detection of vehicle/pedestrian violating legal trajectory anomaly, which we tested on real-time video provided by surveillance cameras of the city of Kazan.

1. Introduction

Large cities have always suffered from the problems of traffic congestions and difficulties in creating safe road systems [1] [2]. Nowadays these issues require efficient solutions due to annual population growth and increasing number of vehicles. Non-optimized traffic flow raises a whole list of global issues starting from economic to social and environmental. Therefore, one of the most important goal of the modern intelligent ITS is to provide effective traffic management in the cities. For achieving it we need to obtain optimal usage of the roads and efficient traffic elements management, reach continuity of traffic and freight management and provide road safety and security.

Described task require monitoring for external and internal factors, collection of the actual data about situations on the roads. The relevant for this task data may come from many sources, which can be categorized into few groups:

1) roadway data, which is collected by different sensors, active or passive in nature, collecting data of regular traffic operations [4]. In recent years it become popular to use video surveillance cameras for roadway data collection.

2) vehicle-based data, collected by technologies, such as electronic toll tags and radionavigation-satellite services (global positioning systems (GPS), GALILEO, GLONASS, etc.), which combined with cell phone-based Bluetooth and Wi-Fi radios [5][6].

3) traveler-based data voluntarily provided by drivers, which use mobile communications and applications.

4) wide area data, which obtained by multiple sensor networks, photogrammetry and video recording from unpiloted aircraft and space-based radar or Geographic Information Systems (GIS).

5) indirect data from external systems. For example, emergency management information systems (EMIS) do not directly store information about the road environment, but can store data about incidents that occur in the city. These incidents cause traffic jams, and impede traffic in the city. Data mining, forecasting and analysis of the EMIS data can reduce response time, which will scale down or even prevent traffic jams [7].

Considering all presented above types, the question about data obtaining tools for further analysis, which will cover more tasks and efficiently assist in solving the described above problems, could be raised. Usage of video surveillance holds promise and demonstrate its profitability in various fields of modern cities, despite on its shortcomings, like dependence on environmental factors (rain, fog, brightness, etc.) and accuracy loss. First, the installation of supported tools can be performed without any additional work on the roadway. The second advantage is the price of devices and the cost of their maintenance. Additional information received from the roads is a third and significant advantage. With growth of computational abilities in the past few years, the deep learning methods make it possible to provide efficient statistical monitoring of the roads [8].

All describes possibilities lead to the idea that with the help of modern intelligent data processing algorithms and tools, we need focus our attention on video surveillance for solving the tasks in ITS area. Despite a strong connection with human decision making process, objects anomaly detection is task that can be fully automated with the help of combination of artificial intelligence (AI) like neural network and machine learning algorithms. The main goal of this paper is to give an overview of existing anomaly detection approaches in deep learning field in terms of their applicability for detection the anomalies on the surveillance cameras video on the streets of the Kazan city.

2. Anomaly detection approaches in the transport environment

It is not clearly what can be considered an anomaly and what is not. The definition depends on several factors: field of activity, type of processing data and its features, external conditions. Surveillance cameras provide visual control of a given area, which allows to constantly check so called zone of interest and highlight the *control events* consisting of certain changes in the observed area. This can be law enforcement, control over abandoned objects, crowd behavior [9]. Parking cameras can control entry for access control without involving an additional person, and monitor the number of parking spaces. In our case we are focused on detection of traffic anomalies, therefore under control events we will consider different anomalies on the road, such as:

- vehicle/pedestrian violates legal trajectory;
- traffic congestions (including traffic flow deceleration/density increase on a road section (for this task we might consider some statistical data as counting and average speed).

2.1.. Trajectory-based anomaly detection

Trajectory formation is complex and diverse task. However, this area is receiving attention of the computer vision and research community for last few years [10]. Besides transport systems it can be used in suspicious activity detection [11], sports video analysis [12], video summarization [13], synopsis generation [14]. Object's trajectory is defined by captured motion changes of moving objects. There are two possible approaches for trajectory-based anomaly detection: first approach, often used in video analysis systems on the cameras themselves, is to define the rules and highlight areas of interest and track all deviations. The second approach is based on unsupervised learning, when we give to detection system the opportunity to learn on large amount of data (video streams in our case), determine the rules automatically and then detect anomalies. Second approach is less reliable for critical systems, where the accuracy and confidence must be at the high level, but it can reveal nontrivial patterns that are not described by the rules. The main advantage of the second is its scalability and adaptation to constantly changing conditions (camera position, road conditions), which can significantly reduce manual work.

In this paper we decided to use unsupervised learning algorithms for anomaly detection. We need to keep in mind, that all steps should be processed in real-time, therefore some trade-off between accuracy and velocity of processing with privilege of the second should be in mind. First, video should be received and pre-processed. Usually video pre-processing include breakdown into frames and individual images processing. After this step we can apply object detection and tracking techniques in order to extract information about entities in the frame and their movement. Not interesting for specific problem objects should be excluded from consideration on this step. On the next step we must highlight the trajectory of objects that we have identified on the previous steps. This step is important and must be done regardless of how the learning process will be organized (by the rules or with unsupervised learning techniques). There are examples of visualization of trajectories and outlier detection on video frames on Fig. 1.

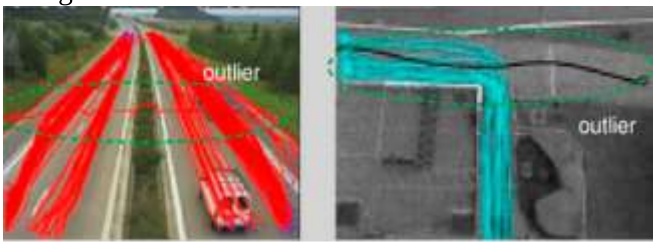


Figure 1. Example of trajectory extraction and outlier detection.

On the next step represented outliers should be registered as anomalies, because its trajectory stands out from normal. Detection of anomalies requires a comparison of the current trajectory of a moving object with the legitimate one. Therefore, before we can compare trajectories, we must identify the rules or patterns, which could be received with the machine learning technique. It is hard process due to difficulties in defining boundaries between the normal and incorrect behaviour of vehicles, especially on different lanes of the same road. Finally, after rules modelling we can identify the anomalies.

2.2. Traffic congestions detection

Prior to exploring parameters of traffic congestion it is worthy to give a definition to traffic congestion. According to Downs there is no universally accepted definition of traffic congestion [15]. Numerous definitions can be categorized into 3 groups: demand capacity related, delay-travel time related, cost related [16]. Depending on a chosen definition of traffic congestion, different measurement metrics can be used:

1) *Speed.* The average speed on any section of a road can be used to infer the state of traffic at the present point of time. This can be done by comparing present speed of traffic to off-peak period speed. The ratio of peak period to off-peak period speed is suggested as a direct measure of congestion.

2) *Travel time and delay.* Congestion is a travel time or delay in excess of the normally incurred under light or free-flow travel conditions [17]. Unacceptable congestion is travel time or delay longer than accepted norm. This norm may vary depending on geographical location, time of the day etc.

3) Level of service measures. The level of service (LOS) has been one of most popular measure of traffic congestion. It was adopted in 1985 Highway Capacity Manual [18]. LOS is subdivided into six classes ranging from A to F, which are categorized according vehicle-to-capacity ratio.

In the scope of this work 4 approaches to congestion detection were proposed:

1) Measure time of presence of vehicles in frame and compare with historical data.

2) Measure velocity of vehicles and compare with historical data.

3) Count vehicles using virtual detecting line and compare number of passing vehicles with traffic capacity of a road calculated according to regulations of Russian Federation.

4) Measure the movement in a frame and compare with calibration data.

In both detection of anomalies tasks we need to process incoming video frames and provide image processing for object detection and tracking. Therefore, we will consider these two methods in the next chapter.

3. Image processing

Segmentation is a process of subdividing an image into nonoverlapping regions so each pixel is assigned to a respective region. This step is important because here objects are separated. These objects then can be extracted for subsequent processes such as recognition or description. In case of a road traffic scene, the problem is to segment image and find the regions which will be corresponding to road and traffic participants. The problems of scene segmentation and object classification can be solved by expert systems, semantic networks and neural network systems. On the step of object detection we find and classify the object, on the step of object tracking - build its trajectory.

3.1. Object detection

Object detection is more general problem than object classification. It does not only require to decide if object is present on the image but also to find its' location. This is a difficult problem, because objects in categories such as vehicles and people can vary greatly in appearance. Variations arise not only from changes in illumination and viewpoint, but also due to nonrigid deformations and intraclass variability in shape and other visual properties. For example, people wear different clothes and take a variety of poses, while cars come in various shapes and colors. Figure 2 shows a taxonomy of object detection approaches in remote sensing image but these approaches can be generalized to object detection as a whole.

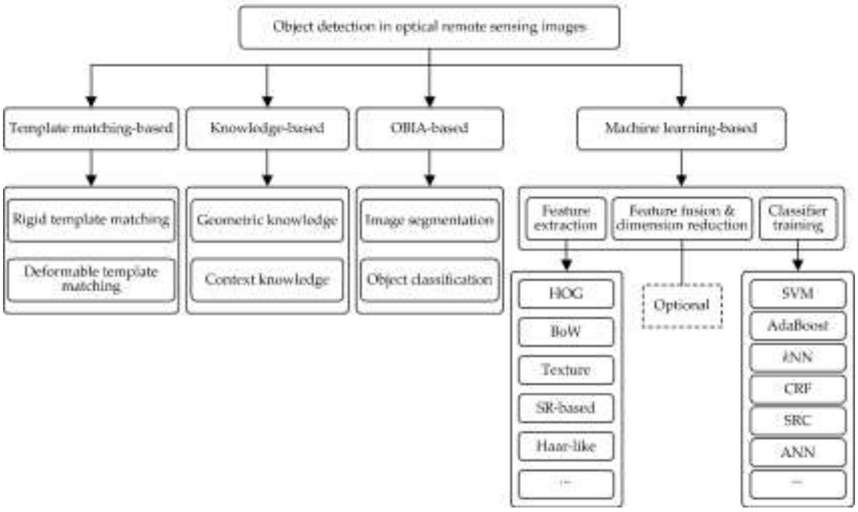


Figure 2. Taxonomy of methods for object detection.

The input of the classifier is a set of regions (sliding windows or object proposals) with their corresponding feature representations and the output is their corresponding predicted labels, i.e., object or not. Feature extraction, feature fusion and dimension reduction (optional), and classifier training play the most important roles in the performance of object detection.

Traditional techniques from statistical pattern recognition like the Bayesian discriminant and the Parzen windows were popular until the beginning of the 1990s. Since then, neural networks (ANNs) have increasingly been used as an alternative to classic pattern classifiers and clustering techniques [19]. Since 2012 after work of Krizhevsky et al.[19] majority of state of the art object segmentation, and classification are performed by Convolutional Neural Networks (CNN). These neural networks are outperforming other architectures due to data size reduction on convolution and pooling steps, which allow increasing complexity of neural network. There are four main operations in CNN: convolution, non linearity (ReLU), pooling or sub sampling, and classification (provided be fully connected layer). On the first step by applying different filters to the original image feature map is built. Then negative values are filtered by non-linear function. This function is applied pixel wise. This step removes linearity from image and allows algorithm to learn non-linear functions. On the third step pooling is performed. This step helps to reduce amount of data passed to next step substantially. Most common functions applied on this step are maximization and averaging. On the last step feature maps are passed to fully-connected layers which are responsible for classification.

Various types of CNN for object detection are differ in the way of searching extracted features of the image. There are search by regions on the image (R-CNN, [20] Fast R-CNN, [21] Faster R-CNN [22]) Single Shot MultiBox Detector (SSD) [23] and You Only Look Once (YOLO) [24] method which predict bounding boxes and probabilities for each region. After numerous test, represented in Figure 3, for our work we choose YOLO method. In YOLO network image is split into SxS grid. Each of resulting cells predicts B bounding boxes and confidences. Each cell also predicts class probabilities. Bounding boxes are combined with classes.

3.2. Object tracking

Tracking of moving objects for measuring motion parameters is an important area of application of image processing [25], [26]. Process of object tracking can be subdivided into three consequent steps - moving object detection, object classification and inter-frame tracking. On the step of object detection moving objects are separated from static background. On step of classification identified moving objects are assigned to classes based on their features. Finally, on step of tracking classified objects are identified on subsequent frames. Our YOLO model solves first two task of the object detection, but still we need to get an information about movement through the frames.

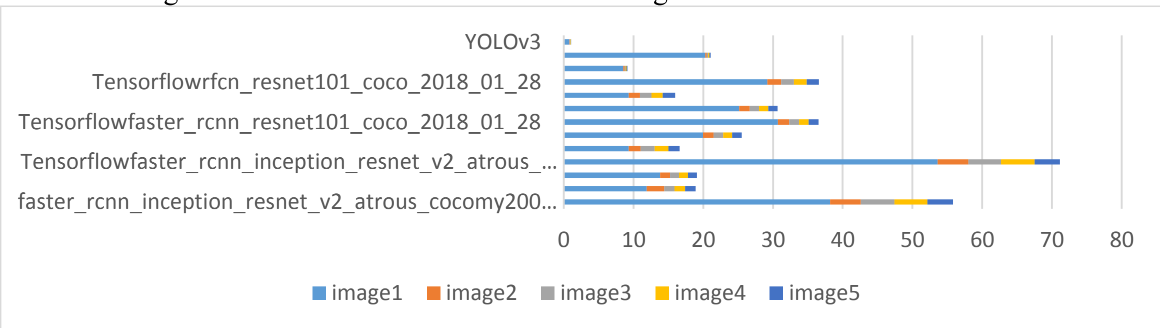


Figure 3. Comparison of calculation time of different CNN models.

Modern trackers also are not ideal. It is difficult for them to process frames with such environmental challenging factors like occlusion (an object closer to the camera overlaps the object behind it), background clutters (road has similar to vehicle color), motion blur (target region is blurred due to the motion of vehicles or the camera) and so on. Modern algorithms have to be able partially to cover these difficulties. Boosting tracker [27] is based on an online AdaBoost algorithm. The initial bounding box of an object is considered as the positive example of the object, and the rest is treated as background. Algorithm is old and outperformed by many modern algorithms. Multiple Instance Learning (MIL) tracker algorithm [28] is based on approach similar to Boosting. The difference here is that this algorithm generates multiple hypotheses in neighborhood of a center of object. All these hypotheses together with original bounding box are put into positive ‘bag’. If the prediction of main bounding box is not well-centered there is a high probability that positive ‘bag’ will contain a better prediction. Tracker does not recover from full occlusion. Kernelized Correlation Filters (KCF) tracker [29] is based on ideas of MIL and Boosting. The fact that positive bag in MIL contain boxes with large overlap allows to reduce complexity of correlation calculation from $O(n^2)$ to $O(n \log n)$. This tracker is both faster and more accurate than MIL and reports tracking failure better. It does not recover from full occlusion. Minimum Output Sum of Squared Error (MOSSE) tracker [30] uses adaptive correlation for object tracking, which produces stable correlation filters when initialized using a single frame. MOSSE tracker is robust to variations in lighting, scale, pose, and non-rigid deformations. It also detects occlusion based upon the peak-to-sidelobe ratio, which enables the tracker to pause and resume where it left off when the object reappears. We compared such qualitative signs as the smoothness of the object tracking, work with different types of objects, the number of occlusions. After numerous tests we decided to make our choice on KCF tracking algorithm. This method not only tracks objects correctly, but also does not produce so many errors like other algorithms. Its advantage is that it can work with images of low quality, resolution and frame rate. After receiving all trajectories, we can divide them into categories by type of the objects and build groups of trajectories and track anomalies (represented on Figure 4).

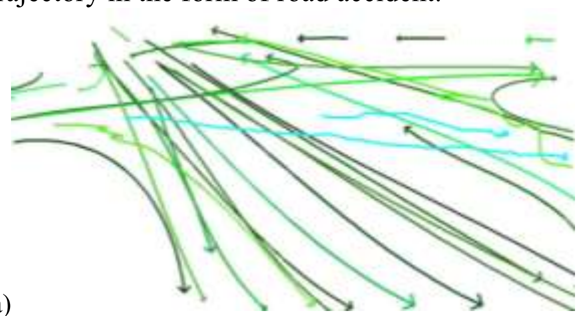
a)

b)

Figure 4. Trajectories of objects: a) separately b) on the video frame.

4. Evaluation and results

For trajectory-based anomaly detection task we used data from intersections of Kazan city, and for congestion evaluation we used data from Moscow roadside cameras. We build clusters of motion models based on DBSCAN algorithm [31], and then determine the anomalies by comparing the trajectories. Figure 5 shows the automatically detected anomaly of deviation from the reference trajectory in the form of road accident.



a)

b)

Figure.5. Identified anomaly deviations from the reference trajectory: a) two trajectories that go beyond the normal b) detected traffic accident.

For traffic congestion we tried few approaches. Approach based on time of presence was found to be impractical. This approach requires ideal detection and tracking algorithms. Neither algorithms of YOLO family nor any other modern algorithm has detection rate that would make this approach work. On tracking failures time of presence cannot be calculated correctly. Another problem of this approach is connected to performance decrease with growing number of trackers. Approach based on speed is also impractical. Furthermore, measurement of distance in pixels produces data incomparable to real distances. Without manual calibration of a camera it is impossible to get mapping from camera coordinates into real-world ones. Approach based on vehicle count gives better result in comparison to previous two. In scenes with congestion due to high level of occlusion vehicle detection is unstable as well as tracking. In Figure 6 shows congestion detection. This approach requires high level of user interaction for initialization.

5. Conclusion

As a result of this work, an approach for detection congestions and trajectories-based anomalies was developed. The developed approach is able to recognize independent reference trajectories for certain classes of objects with unsupervised learning algorithm, and identify anomalies if the spatial trajectory of the object violates to them. In the future, the expansion of the type of identified anomalies, as well as testing the system in real time is planned. There are various road rules for road lanes in the city of Kazan. For example, there are special lane for public transport and large sized vehicles. Drivers of vehicles often violate the travel ban on these lines and drive along it. We can also think about creating rules for each lane of the road, which is a very challenging task from both unsupervised learning and video processing parts.

Figure 6. Identified congestion based on vehicle count approach.

6. References

[1] Wiseman, Y. Real-time monitoring of traffic congestions / Y. Wiseman // IEEE International Conferenceon Electro Information Technology, 2017. – P. 501-505.
[2] Transport problems in cities [Electronic resource]. – Access mode: https://trimis.ec.europa.eu/sites/default/files/project/documents/20120402_173932_45110_D%201.1%20-%20Transport%20problems%20in%20cities%20-%20v3.pdf (01.11.2019).
[3] Kelly, F.J. Transport solutions for cleaner air / F.J.Kelly, T. Zhu // Science. – 2016. – Vol. 352 (6288). – P. 934-936.
[4] New Traffic Data Sources - An Overview [Electronic resource]. – Access mode: https://bitre.gov.au/events/2014/files/NewDataSources-BackgroundPaper-April%202014.pdf (04.11.2019).
[5] D'Andrea, E. Detection of traffic congestion and incidents from GPS trace analysis / E. D'Andrea, F. Marcelloni // Expert Syst. Appl. – 2017. – Vol.73. – P. 43-56.
[6] Fontaine, M.D. Investigation of the Performance of Wireless Location Technology-Based Traffic Monitoring Systems / M.D. Fontaine, B.L. Smith // J. Transp. Eng. – 2007. – Vol. 133(3). – P. 157-165.
[7] Dagaeva, M. Big spatio-temporal data mining for emergency management information systems / M. Dagaeva, A. Garaeva, I. Anikin, A. Makhmutova, R. Minnikhanov // IET Intelligent Transport Systems. – 2019. – Vol.13(11). – P. 1649-1657.
[8] Chung, J. Image-Based Learning to Measure Traffic Density Using a Deep Convolutional Neural Network / J. Chung, K. Sohn // IEEE Transactions On Intelligent Transportation Systems. – 2018. – Vol. 19(5). – P. 1670-1675.
[9] Chen, H. Crowd Abnormal Behaviour Identification Based on Integral Optical Flow in Video Surveillance Systems / H. Chen, O. Nedzvedz, S. Ye, S. Ablameyko // Informatica, Lith. Acad. Sci. – 2018. – Vol. 29(2). – P. 211-232.
[10] Ahmed, A.Trajectory-Based Surveillance Analysis: A Survey / A. Ahmed, D.P. Dogra, S. Kar, P.P. Roy // IEEE Trans. Circuits Syst. Video Techn. – 2019. – Vol. 29(7). – P. 1985-1997.
[11] Xiang, T. Video behavior profiling for anomaly detection / T. Xiang, S. Gong // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2008. – Vol. 30(5). – P. 893-908.
[12] Rehman, A. Features extraction for soccer video semantic analysis: current achievements and remaining issues / A. Rehman, T.Saba // Artificial Intelligence Review. – 2014. – P. 1-11.
[13] Zhang, T. Person-based video summarization by tracking and clustering temporal face sequences / T. Zhang, D. Wen, X. Ding // uS Patent 9,514,353, 2016.
[14] Pritch, Y. Nonchronological video synopsis and indexing / Y. Pritch, A. Rav-Acha, S. Peleg //

IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2008. – Vol. 30(11). – P. 1971-1984.
[15] Downs, A. Still stuck in traffic: coping with peak-hour traffic congestion / A. Downs – Washington, D.C: Brookings Institution Press, 2004.
[16] Aftabuzzaman, M. Measuring traffic congestion: a critical review / M. Aftabuzzaman // Australas. Transp. Res. FORUM. – 2007. – Vol. 30.
[17] Levinson, H.S. Developing a Travel Time Congestion Index / H.S. Levinson, T.J. Lomax // Transp. Res. Rec. J. Transp. Res. Board. – 1996. – Vol. 1564(1). – P. 1-10.
[18] Egmont-Petersen, M. Image processing with neural networks – a review / M. Egmont-Petersen, D. Ridder, H. Handels // Pattern Recognit. – 2002. – Vol. 35(10). – P. 2279-2301.
[19] Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // Advances in Neural Information Processing Systems. – 2012. – Vol. 25. – P. 1097-1105.
[20] Girshick, R. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation / R. Girshick, J. Donahue, T. Darrell, J.Malik // Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014. – P. 580-587.
[21] Girshick, R. Fast R-CNN / R. Girshick // Proc. of the IEEE International Conference on Computer Vision, 2015. – P. 1440-1448.
[22] Ren, S. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / S. Ren, K. He, R. Girshick, J. Sun // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39(6). – P. 1137-1149.
[23] Liu, W. SSD: Single shot multibox detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed // European Conf. on Computer Vision. Lecture Notes in Computer Science. – 2016. – P. 21-37.
[24] Redmon, J.You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016. – P. 779-788.
[25] Kale, K. Moving object tracking using optical flow and motion vector estimation / K. Kale, S. Pawar, P. Dhulekar // 4th International Conference on Reliability, Infocom Appendix (Trends and Future Directions). – 2015. – P. 1-6.
[26] Chavez-Garcia, R.O. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking / R.O. Chavez-Garcia, O. Aycard // IEEE Trans. Intell. Transp. Syst. – 2016. – Vol. 17(2). – P. 525-534.
[27] Grabner, H. Real-time tracking via on-line boosting / H. Grabner, M. Grabner, H. Bischof // Proc. BMVC. – 2016. – Vol. 1. – P. 47-56.
[28] Babenko, B. Visual Tracking with Online Multiple Instance Learning / B. Babenko, M.-H. Yang, S. Belongie // IEEE Conference on Computer Vision and Pattern Recognition, 2009.
[29] Henriques, J.F. High-Speed Tracking with Kernelized Correlation Filters / J.F. Henriques, R. Caseiro, P. Martins, J. Batista // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2014. – Vol. 37(3).
[30] Bolme, D.Visual object tracking using adaptive correlation filters / D. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2010. – P. 2544-2550.
[31] Ester, M. A density-based algorithm for discovering clusters in large spatial databases with noise / M. Ester, H.P. Kriegel, J. Sander, X. Xu // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining – AAAI Press, 1996. – P. 226-231.