INTELLIGENT VIDEO ANALYSIS OF DANGEROUS SITUATIONS

Anishchenko, Lesya N., *Bauman Moscow State Technical University, Moscow, Russia.*
Ivashov, Sergey I., *Bauman Moscow State Technical University, Moscow, Russia.*
Skrebkov, Alexey V., *Russian University of Transport (MIIT), Moscow, Russia.*

ABSTRACT

The article is devoted to development of a system for the intelligent analysis of video recordings of external surveillance cameras, which makes it possible to identify dangerous situations at railway facilities using the example of detection of falls in the track area. A method of preprocessing a video for the purpose of forming a feature space based on the use of background subtraction using the Gaussian mixture method, followed by tracking the movement of a person with the help of the Kalman filter and deformation of the shape of the mobile object as a result of applying the procrustean analysis is proposed. The selection of the optimal composition of the feature space and additional heuristics providing the isolation of episodes of falls from video recording with an average quality of the Cohen's kappa 0,62 is compared with the visual analysis by the operator.

Keywords: *railway, safety, video surveillance, intelligent video analysis, motion recognition, machine learning, form analysis.*

Background. *The task of identifying potentially dangerous persons at transport facilities is carried out according to data obtained with the help of external surveillance cameras. Developers of video surveillance systems, in particular, offer complexes using biometric identification (for example, NeoFace system, R7 glasses) [1, 2], as well as recognition of emotions in facial expressions (for example, DeepFace) [3–6]. The main disadvantage of this type of system is that information about an intruder may not be contained in the existing database.*

At the disposal of analysts there are algorithms that allow both to track things lost on platforms [7], and to identify objects along the train route [8], as well as to detect the occurrence of smoke [9], however, in the overwhelming majority of cases, data from CCTV cameras are used only for formation of archives of video recordings, and the possibilities of fixed information from remote objects are practically not used. This is due primarily to technical difficulties associated with the characteristics of intelligent video surveillance systems, namely, sensitivity of the analyzer to the conditions of illumination, presence of vibration, which is unavoidable when used in railway transport, and the like.

Separately, it is necessary to distinguish the task of ensuring safety on the aprons of low-loaded stations of commuter traffic, especially at night. However, in case of an emergency situation, a person in need may not get help from someone because of the absence of other passengers. In this connection, the urgency of using technical means that are able to send an alert signal to the station duty officer in an automated mode is increasing. Within the framework of this article, the question of such a system for recognizing human movements through the records of CCTV cameras and revealing episodes of its falls in the risk zone is considered.

Objective. *The objective of the authors is to consider intelligent video analysis of dangerous situations.*

Methods. *The authors use general scientific methods, comparative analysis, evaluation approach.*

Results.

Experimental records

In most works to identify episodes of falls using video image analysis, the evaluation of the effectiveness of the proposed classification algorithms proves to be artificially overstated due to the limitations of the database of video records used as testing and training samples. In this case, the records were obtained for the same conditions of the situation (most often in the laboratory, rather than close to reality) in the presence of uniform illumination of the analyzed area of space. In the role of subjects performing, among other things, artifacts of movement such as «fall», the same person appears, with the falls being monotonous, both in the form of movement itself and the actions preceding it, and also by the angle at which the «executor» is located relative to the camera at the time of the fall. In addition, it should be noted that almost always falls are performed on a cushioning mat, which has a contrasting color with the clothes of the subject.

In our analysis, an open database of video recordings of the Electronics and Imaging Laboratory of the National Center for Scientific Research in Chalon-sur-Son was used [12]. Its merits include the following factors:

1. Video images are obtained for different environmental conditions.

2. There is uneven illumination of the scene of the experiment, including a situation where, due to the limited dynamic range of the camera and the presence in the frame of a high brightness region (window area), the video image of the person had a small contrast compared to the situation.

3. Four subjects (3 men and 1 woman) participated in the experiments.

4. Falls of bodies occurred at different viewing angles, both from standing position and from sitting position.

5. Falls were carried out on a specially prepared cushioning base and directly on the floor.

The authors analyzed 108 records from the database, 84 of which contained a single episode of the fall. For each record, the operator has visually determined the frame numbers of the beginning and end of the episode. The duration of artifacts of the «fall» type was 22 ± 9 frames. Considering that the video recordings are made for a sampling frequency of 25 or 30 frames per second, we get the duration of the artifact equal to 0,7 ± 0,3 s.

Algorithm of video processing

The task of classification in computer vision is divided into two sub-tasks: image preprocessing and classification. The pre-processing phase is necessary to convert the visual data into a form that

Table 1

Error matrix for feature space P1

		The result of classification	
		Non-fall	Fall
True class	Non-fall	19468	871
	Fall	975	556
Cohen's kappa		0,33	
Accuracy		0,91	

is acceptable when using classification algorithms. As a result, a space of attributes is formed. In this case, it is required not only to form the space of features of the image, but also to exclude non-informational ones from consideration, while retaining all the attributes essential for the solution of the task. At the stage of classification, the classifier is trained, and as the training and testing sample, the evaluation of attributes obtained during the preprocessing of the experimental video sequence is used.

Preprocessing of data

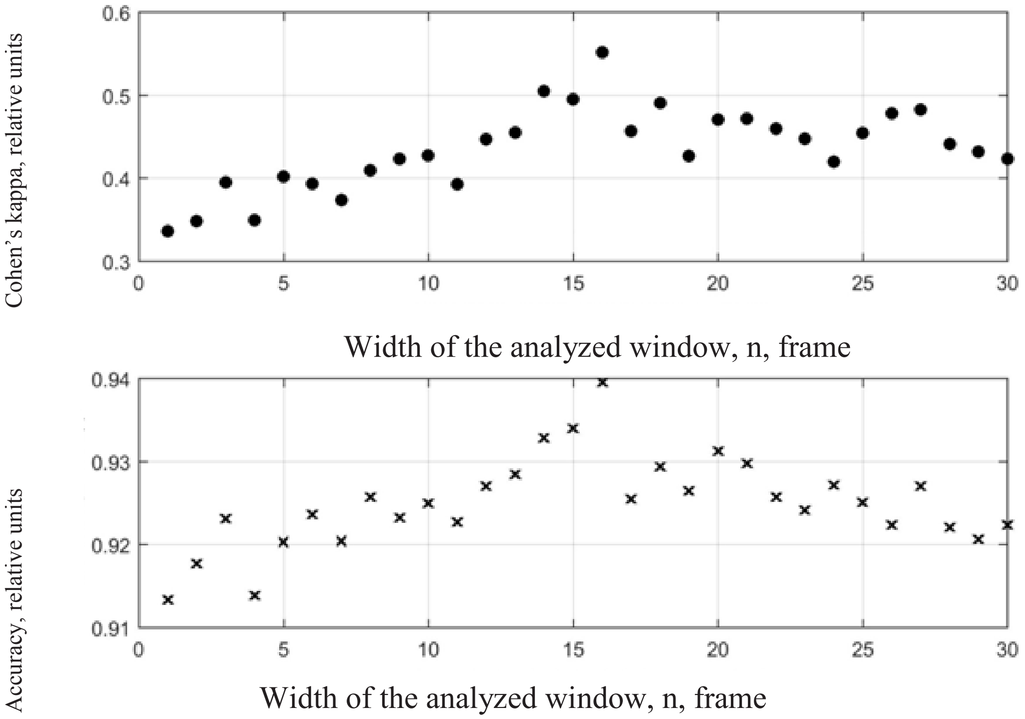
1) Selecting an object against the background of the situation. For each frame in the image, a gradual subtraction of the background was performed using the built-in Matlab functional that implements the Gaussian mixture method (the training was conducted in 20 frames). Then the segmentation of the image was carried out, as a result of which image points that were not a background were combined into blocks if their total number in a single block was not less than the threshold value. When analyzing the experimental data, an empirically selected threshold value equal to 50 pixels for a frame size of 320 by 240 is used. In case of further use of the proposed algorithm for analyzing data having a different spatial resolution, this parameter can be changed proportionally.

2) Tracking the position in the frame of the moving object. As soon as the segmentation results showed a mobile object in the frame, a Kalman filter was created, which was used to track target movements. In our work, a filter was implemented using the Matlab Configure Kalman Filter function. Variability of the tracking speed of the tracking object is taken into account with the help of an additional filter parameter (Motion Noise) [13].

When analyzing a video image, two cases of using the Kalman filter are possible:

– the mobile object is detected on the frame: then the filter predicts the position of the object in the frame of the video sequence and uses the data on the new position of the object to correct the results of the object selected by filtering the position of the object;

– the mobile object is not detected: then the position of the detected object in the frame obtained by means of the filter is formed exclusively on the basis of the analysis data of the previous frames.

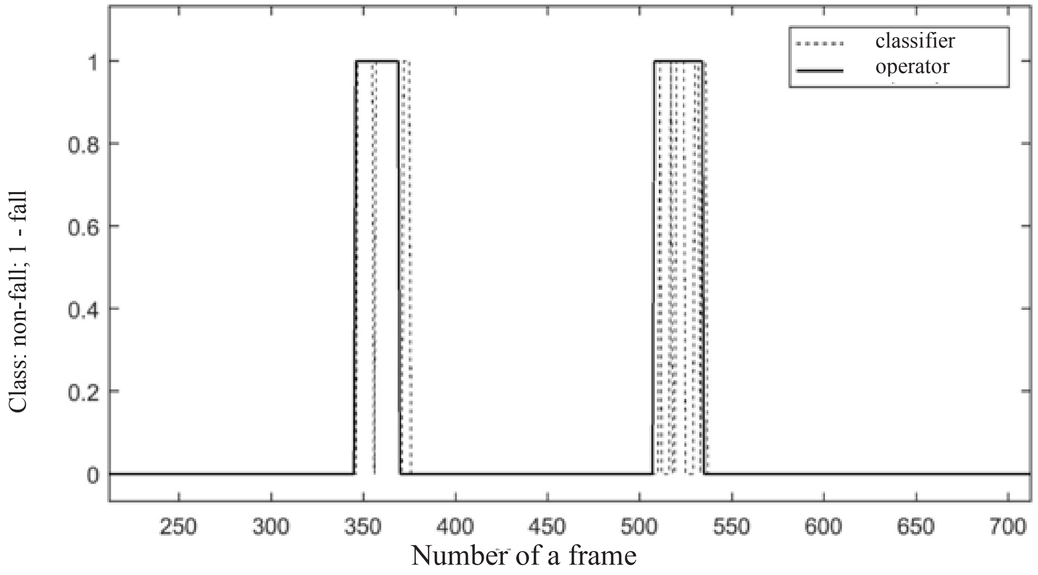


Pic. 1. Dependencies of Cohen's kappa and the accuracy on the width of the analyzed window.

Table 2

Error matrix for feature space P2

		The result of classification	
		Non-fall	Fall
True class	Non-fall	19811	528
	Fall	620	911
Cohen's kappa		0,58	
Accuracy		0,95	



Pic. 2. Comparison of the results of video recording processing by the operator and the classifier.

3) Analysis of the deformation of the shape of a mobile object. As a context descriptor of the form, in contrast to [14], where all the boundary points of the object were analyzed, only mobile characteristic boundary points were taken. For them, a procrust analysis of the form [15] was performed using the following algorithm.

At each frame of the video sequence k for the characteristic edge points of the human silhouette, a complex vector was defined:

$Z = [z_1, z_2, \ldots, z_k]$, $z_j = x_j + iy_j$,

where i – imaginary unit; x_j and y_i – coordinates of the j-th point of the image.

The centered estimation of the vector Z_C was formed by multiplying the coordinate vector Z by the centering matrix C:

$Z_C = Z \cdot C$,

$$C = I_k - \frac{1}{k}\cdot 1_k \cdot 1_k^T,$$

where I_k – unit matrix of dimension kxk; l_k – unit vector of dimension k.

For the two sequences $v = (v_1, v_2, \ldots, v_k)$ and $w = (w_1, w_2, \ldots, w_k)$ centered in the manner described above, the distance D(v, w) between them was calculated:

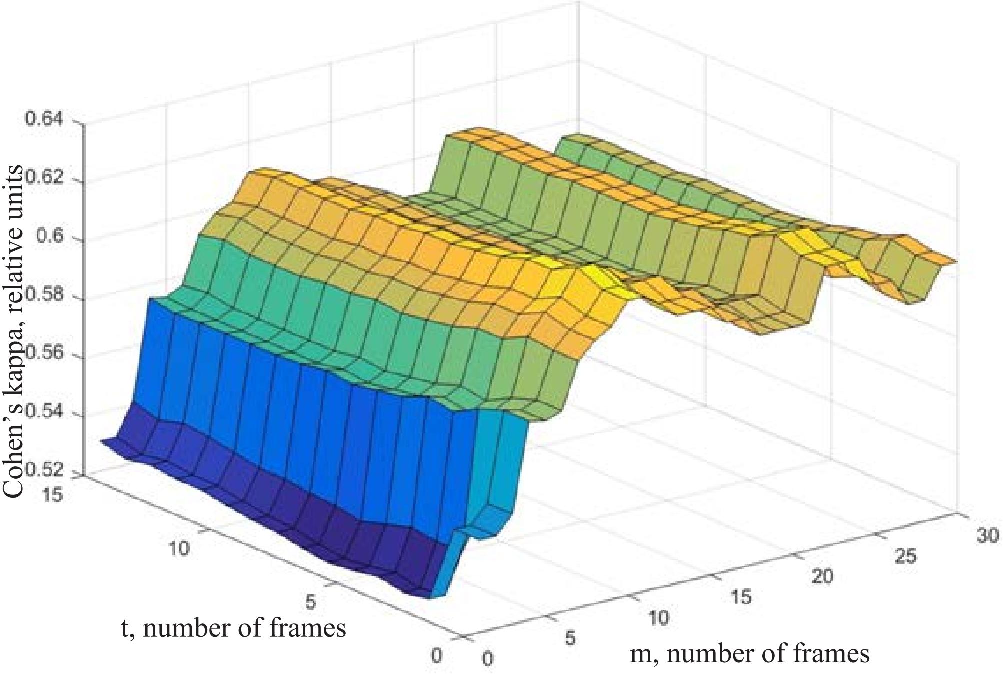
$$D(v,w) = \sqrt{1 - \frac{|v \cdot w|^2}{\|v\|^2 \cdot \|w\|^2}}.$$

The calculated distance (the procruste metric) for sequences of characteristic points of two consecutive frames is an estimate sensitive to significant deformations of the shape of the object, for example, due to a fall.

Table 3

Error matrix for feature space P2 with the use of additional heuristics

		The result of classification	
		Non-fall	Fall
True class	Non-fall	19845	494
	Fall	559	972
Cohen's kappa		0,62	
Accuracy		0,95	



Pic. 3. Dependence of Cohen's kappa value on heuristic parameters.

Characteristics extraction

After completion of the pre-processing for each of the frames of the video sequence, the following set of parameters was evaluated, forming the feature space (P1):

1) estimate of the distance D(v, w);

2) coordinates of the center of gravity of the image area corresponding to the mobile object;

3) the area of the image area corresponding to the mobile object;

4) the speedy of the center of gravity of the image region corresponding to the mobile object, along the x and y axes for two consecutive frames.

The choice of these characteristics is due to the analysis of scientific literature on this topic [11, 14, 15].

The space of features P1 in the course of the work was supplemented by a set of characteristics characterizing the speed of movement along the axis x and y of the center of gravity of the corresponding area of the image in the window by the duration n of the frames. As a result, augmented space of features (P2) was formed.

Selecting the composition of a vector

The analyzed data array was said to contain 108 video recordings (lasting from 20 seconds to 1 minute), in 84 of which there was a single episode of the fall. Each frame of the video recording on the basis of visual analysis by the operator was referred to the class of «fall» or «non-fall». The sample of the experimental data is divided into the training data (used to train the classifier and the selection of the optimal composition of the feature space) and the testing (used with the final rating of the classifier's effectiveness). The training included 75 % of the video recordings (65610), the remaining 25 % (21870 frames) comprised a testing sample.

We used a classifier based on a tree of solutions from the Matlab machine learning library. To assess the classification error, a cross-validation method was applied to k = 10 blocks.

As a measure of the quality of the classification, it is offered to use accuracy, as well as Cohen's kappa (the measure of inter-expert consent), since the sample is unbalanced. Note that the sample is called unbalanced if there is an imbalance in the classes, namely: they are presented unevenly. For example, in our case, less than 10 % of the sample corresponds to fall artifacts.

For the feature space P1, the results of applying the classifier to the test sample are shown in Table 1. The values in the cells of the error matrix correspond to the number of frames correctly or erroneously classified by the proposed algorithm.

Let's consider how the use of the supplemented feature space P2 affects the efficiency of classification. Pic. 1 shows the data on how Cohen's kappa size changes and the accuracy of the classification in the case of using the augmented space of features P2 with the window width n varying from 2 to 30 frames. The upper limit of the window width range n is selected based on the maximum duration of the incident episode of 1 s, which corresponds to 30 frames at a sampling rate of 30 frames per second.

It follows from Pic. 1, when the feature space P1 is added to the feature space P2, the Cohen's kappa grows approximately twice: from 0,31 to 0,58. In this case, the best estimates of Cohen's kappa and the accuracy correspond to the width of the analyzed window equal to 16 frames. The error matrix and the classification quality estimates for this case are shown in Table 2.

A visual comparison of the results of using the classifier and markup of video recording by the operator, shown in Pic. 2, shows that the relatively low value of Cohen's kappa due to the «fragmentation» of classification results.

In order to improve accuracy, it was offered to use additional rules (heuristics) aimed at combining frames classified as falls into a single fragment of the record if they are located at no more than m frames from each other. Also, the parameter t, corresponding to the minimum duration of the fall episode, was introduced.

Varying the values of m in the range from 1 to 30 frames and t in the range from 0 to 15 frames, an estimation of the dependence of Cohen's kappa value for the training sample on these parameters was made.

It follows from Pic. 3, the maximum value of Cohen's kappa is reached at m = 10 frames and t = 2 frames. Table 3 shows the error matrix of classification results using the proposed additional heuristics for the obtained optimal parameter values.

Thus, the use as feature space P2 for the window width n = 16 and additional heuristics allows to achieve the accuracy of the frame-by-frame classification of episodes of falls from video recording with Cohen's kappa value of 0,62.

Conclusion. *The method of intellectual analysis of video recordings of external surveillance cameras is proposed, which makes it possible to identify situations that pose a danger to life and health of people in railway transport by the example of identifying episodes of falls. An algorithm is developed for preprocessing a video for the purpose of forming a feature space based on the use of background subtraction by the Gaussian mixture method, followed by tracking the movement of a person with the help of the Kalman filter and deformation of the shape of the mobile object as a result of using the procrustean shape analysis. The proposed tree-based classification method was tested on a database of 108 video records, 84 of which contained a single episode of the fall.*

The conducted comparative study of several sets of characteristics allowed to substantiate the choice of the optimal set of characteristics and additional heuristics that ensure the allocation of episodes of falls on the video record with an average quality of the Cohen's kappa 0,62.

In the future it is planned to expand the experimental sample and add to the classifier the possibility of recognizing people who are in a state of alcoholic or narcotic intoxication by video recording, according to the peculiarities of their gait.

REFERENCES

1. Advances Recognition Systems: Rapid-Access Biometric and Credential Solution, NeoFace Express [Electronic resource] / Official website of NEC corporation, 2017. URL: https://www.necam.com/docs/?id=6c812b4d-2a12–40ed-9fea-fae81550c7aa. Last accessed 05.11.2017.

2. SMARTGLASSES7 [Electronic resource] / Official website of ODG corporation, 2017. URL: https://www.osterhoutgroup.com/pub/static/version1515417478/frontend/Infortis/ultimo/en_US/pdf/R-7-TechSheet.pdf. Last accessed 05.11.2017.

3. Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. «DeepFace: Closing the Gap to Human-Level Performance in Face Verification», Conference on Computer Vision and Pattern Recognition (CVPR), June 24, 2014.

4. Parkhi [*et al*]. Deep Face Recognition [Electronic resource]: URL: https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf. Last accessed 05.11.2017.

5. Mohammadian A., Aghaeinia H., Towhidkhah F. Video-based facial expression recognition by removing the style variations in Image Processing, IET, 2015, Vol. 9, No. 7, pp. 596–603.

6. Iosifidis A., Tefas A., Pitas I. Class-specific Reference Discriminant Analysis with application in Human Behavior Analysis, IEEE Transactions on Human-Machine Systems, 2015, Vol. 45, no. 3, pp. 315–326.

7. Maddalena L., Petrosino A. Stopped Object Detection by Learning Foreground Model in Videos, in IEEE Transactions on Neural Networks and Learning Systems, May 2013, Vol. 24, no. 5, pp. 723–735.

8. Amrutha M. P., Vince Paul. Study of Different Obstacle Detection Methods in Railway Track, International Journal of Innovative Research in Computer and Communication Engineering, Jan 2017, Vol. 5, no. 1, pp. 1204–1208.

9. Filonenko A., Hernández D. C., Jo K. H. Real-time smoke detection for surveillance, 2015, IEEE13th International Conference on Industrial Informatics (INDIN), Cambridge, 2015, pp. 568–571.

10. Rougier C., Meunier J., St-Arnaud A., Rousseau J. Fall detection from human shape and motion history using video surveillance, Proc. 21st Int. Conf. AINAW, 2007, Vol. 2, pp. 875–880.

11. Lee T., Mihailidis A. An intelligent emergency response system: Preliminary development and testing of automated fall detection, J. Telemed. Telecare, 2005, Vol. 11, no. 4, pp. 194–198.

12. Charfi I., Miteran J., Dubois J., Atri M., Tourki R. Optimised spatio-temporal descriptors for real-time fall detection: comparison of SVM and Adaboost based classification, Journal of Electronic Imaging (JEI), Vol. 22. Iss. 4, pp. 17, October 2013.

13. MathWork Documentation: Create Kalman filter for object tracking [Electronic resource] / Official website of MathWorks, 1994–2017. URL: https://www.mathworks.com/help/vision/ref/configurekalmanfilter.html. Last accessed 05.11.2017).

14. Mori G., Malik J. Estimating human body configurations using shape context matching, in Proc. Eur. Conf. Comput. Vision, 2002, Vol. 2352, pp. 150–180.

15. Rougier C., Meunier J., St-Arnaud A., Rousseau J. Robust Video Surveillance for Fall Detection Based on Human Shape Deformation, in IEEE Transactions on Circuits and Systems for Video Technology, May 2011, Vol. 21, no. 5, pp. 611–622. ●

Information about the authors:
Anishchenko, Lesya N. – Ph.D. (Eng), senior researcher of Bauman Moscow State Technical University, Moscow, Russia, anishchenko@rslab.ru.
Ivashov, Sergey I. – Ph.D. (Eng), head of the laboratory of Bauman Moscow State Technical University, Moscow, Russia, sivashiv@rslab.ru.
Skrebkov, Alexey V. – Ph.D. (Eng), associate professor of Russian University of Transport (MIIT), Moscow, Russia, skrebkov_av@mail.ru.

Article received 04.10.2017, accepted 05.11.2017.

The work was supported by the Russian Foundation for Basic Research (Grant No. 17–20–03034).