Video pre-motion detection by fragment processing

Sergii Mashtalir[1,†] and Dmytro Lendel[2,†]

[1] Kharkiv National University of Radio Electronic, 14, Nauky Ave., Kharkiv, Ukraine
[2] Uzhhorod National University, 3, Narodna Square, Uzhhorod, Ukraine

Abstract

With the rapid development of technology in recent years, the use of cameras and the production of video and image data have significantly increased. Therefore, there is a great need to develop and improve video surveillance techniques to their maximum extent, particularly in terms of their speed, performance, and resource utilization. In this study, we focused on the formalization of video frame descriptions in the context of solving video motion detection and motion tracking. Our approach is based on dividing each frame into blocks that allows to present image frame as a square matrix for a formal description. The frame block is a matrix of arbitrary dimensions. The ability to skip the step of matrix transformation to a square dimension or vectorization using some descriptor allows to reduce computational costs, freeing up computational resources required for this transformation. In our study, we used Ky Fan norm value as image frame block descriptor. The Ky Fan norm is built on top of matrix singular values. A singular decomposition does not impose restrictions on either the dimension or the character of the elements of the original matrix. Ky Fan norm fluctuations do not depend on video frame size. The decision about the presence of changes in the context of motion detection is made based on a comparison of array consecutive images descriptors, so the values of the Ky Fan norm. Changing the Ky Fan norm in neighboring blocks allows to build a motion tracking.

Keywords

Video stream fragmentation; Ky Fan norm; Singular value decomposition; Motion detection, Motion tracking, Data Analysis

1. Introduction

The video data amount and its quality are increasing every year. Processing a large volume of information is a challenge for modern information systems in almost all classes of tasks. Probably, since the surveillance of cameras appearance, motion detection and motion tracking remain the most relevant ones. Modern approaches consider motion detection and motion tracking for various tasks: traffic flow control [1-3], security cameras motion detection [4], object tracking [5], multiple object tracking [6], etc. In the context of solving the motion detection issue, the ability to skip the step of matrix transformation to a square dimension or vectorization using some descriptor allows reducing computational costs, freeing up computational resources required for this transformation, and making it look worthwhile.

When dealing with natural data, challenges cannot be avoided. The quality of algorithms is affected by such natural phenomena as rain, snow, or changes in lighting. They occlude background information and can significantly impair visibility, which makes motion detection difficult. Most of the existing methods rely heavily on synthetic training data, and thus raise the domain gap problem that prevents the trained models from performing adequately in real testing cases [7,8].

ICST-2024: Information Control Systems & Technologies, September 23-25, 2023, Odesa, Ukraine.
* Corresponding author.
sergii.mashtalir@nure.ua (S. Mashtalir); dmytro.lendel@uzhnu.ua (D. Lendel)
0000-0002-0917-6622 (S. Mashtalir); 0000-0003-3971-1945 (D. Lendel)

Modern approaches in the video data analyzing process also focus on the privacy aspects. In the cloud era, a large amount of data is uploaded to and processed by public clouds. The risk of privacy leakage has become a major concern for cloud users. Cloud-based video surveillance deals with motion detection, which may reveal the privacy of people in a surveillance video. Privacy-preserving video surveillance allows motion detection while protecting privacy. Motion detection method on encrypted and HEVC-compressed videos has been presented [9]. It adopts a novel approach that exploits inter-prediction reference relationships among coding blocks to detect motion regions.

Different approaches for detecting and matching objects and images with set of other images have been proposed in literature [10] such as optical flow [11,12,13], image segmentation [14,15,16] and region based methods [17,18,19], but techniques based on local features are the most popular for detection, recognition and tracking applications. Feature extraction is mainly a two-step process, detection and description of an interest point. Where detection means to locate image points with some distinguishable property and description contains the information (such as derivatives) of the neighborhoods of points which provide a mean of establishing point to point correspondences and improves matching results. Both of these steps in feature extraction play a vital role in upgrading the performance of the algorithm. The efficiency and accuracy of the technique lies in accurately detecting feature points and efficiently generating its descriptor.

Matching two images together and finding exact correspondence between their feature points are the most challenging tasks, but also the most fundamental ones for any recognition and tracking application. In order to find the local information from images which should be sufficient enough to give information regarding the moving object's pose is the prime objective of this application.

The SVD [20] is a powerful and robust mathematical tool often used in signal and image processing, computer vision, pattern recognition, fragment analyses and other areas. Recently, it has been successfully applied to adaptive background modelling and motion detection in image sequences [21]. For greyscale sequences, the data matrix A of size m × n is formed by m consecutive frames, where m is the size (depth) of the temporal data window, n the number of pixels in a frame. Each frame is read row-by-row and stored in a row of A. For color images, each pixel is typically represented in A by three values, e.g., the RGB codes. SVD-based moving object detection uses the residual error for a few largest singular values as the measure of change in the observed scene.

SVD solves is one of its most important visual characteristics - periodicity of an object. Recently, several low-rank/sparse matrix decomposition techniques indicated that a relationship exists between the frequency components of the motion matrix and its decomposition components. This relationship was mostly identified based on empirical evidence without proper analysis, which led to an unclear understanding and poor utilization. Approach [22] attempts to establish the relationship between the periodic components in the motion matrix and its singular value decomposition (SVD) components. The transformation of the periodic components of the motion matrix through QR factorization and Golub–Kahan bidiagonalization, which are the two essential steps of SVD, was thoroughly discussed and analyzed.

This approach [23] proposes a moving object detection algorithm using corner point matching based on singular value decomposition to deal with the problem of the effect because of the changes of light and background. Firstly, the Kalman filtering is used to predict the target center and area; Secondly, corner points are detected in the target area by Harris corner detector; finally, corner matching between the corners of current frame and the corners of target template is based on the improved singular value decomposition algorithm.

In the research [24], singular value decomposition of the matrix and the Ky Fan norm are proposed for scene change analysis. The obtaining an abbreviated description of video frames allows to reduce both time and computational costs when further solving a whole range of video analysis problems. Analysis of the effectiveness of the obtained descriptor for different video data sizes, showing that the change in the descriptor for each block is independent of the video size and aspect ratios [25].

It should be noted that despite the large number of researches in the field of motion detection and tracking, at the moment there are practically no approaches based on fragment analysis of video streams. So, in proposed approach we decided to apply fragment analysis in combination with SVD. This combination can get a fairly effective pre-motion detector. A scene change in the individual block will be associated with a Ky Fan norm changes. If the norm value exceeds the threshold value, we can conclude motion detection in a specific segment. Ky Fan norm changing in neighboring fragments will allow to select a zone of interest and build object motion tracking.

2. Singular value decomposition, Ky Fan norm overview

The singular value decomposition is an extremely useful tool across computer vision. One of the reasons for this is the singular value decomposition can be used to show the strength of the relationship between data sets. A common tool for calculating these relationships is principal component analysis. Principal component analysis takes a data matrix A and forms a new matrix M of vectors ordered according to their variance. It is found through the formula:

$$M = AW, \tag{1}$$

where W is a matrix composed of the eigenvectors of A*A. A quick look at the singular value decomposition of A*A shows that the matrix of “right singular vectors” [26] of A, V, is the matrix of eigenvectors for A*A. Thus a singular value decomposition is a very simple way of finding the principal component analysis.

The singular value decomposition is also a very handy tool for estimation of inverses of singular matrices. A matrix is nonsingular matrix if it has all nonzero singular values. In this case the inverse is very easy to calculate and can be found by simply performing the following calculation on the singular value decomposition:

$$A = USV^*, A^{-1} = VS_0^{-1}U^* \tag{2}$$

where S^{-1} can be calculated easily by taking the inverse of each singular value. However, if the singular value of zero appears in the singular value decomposition of the matrix, then the matrix is singular, and the inverse is approximated by

$$A^{-1} = VS_0^{-1}U^* \tag{3}$$

where S_0^{-1} has entries of the inverse of the singular value when the singular value is greater than some small threshold value and 0 otherwise.

The SVD is related to many common matrix norms and provides an efficient method to calculate them. It follows from our existence the sum first k singular values:

$$||A||_k^{KF} = \sigma 1(\mathrm{A}) + .. + \sigma k(\mathrm{A}) \tag{4}$$

is a matrix norm, called the Ky Fan k-norm.

SVD does not require source matrix to be square which makes it easily applicable for video processing. The point is that support of matrices of any dimension gives flexibility in source data representation. Technical to represent video frames can be based even on source image as well as any composition of descriptors without additional transformations.

3. Application of Ky Fan norm fragment analyses for the motion tracking

In this section we will consider results produced by the developed application. In our experiment we used a video surveillance camera HIK Vision model DS-2CDD2047G2H-LIU, with firmware of

5.7.13 build 230706, video format of 1280x720, and frame rate of 25. The camera was installed on office building parking and worked the whole day.

The first step is to represent the sourced videos as a sequence of frames. An example of such a representation is shown in Figure 1.

Each frame is converted from RGB to grayscale model so that the value of each pixel carries only intensity information. Thus, problems associated with color rendering and color perceptions are excluded from consideration. Practically means that we will be working in the intensity domain.

As a result, the change in illumination will affect our experiment and we will measure this change by SVD. In the context of solving video surveillance, we can detect new object in the frame and follow object motion by fluctuation of singular values in the fragments.

Figure 1: Video source as a sequence of frames. Source: compiled by the authors.

The result of frame-by-frame processing is a new video source in grayscale model with marked blocks, which is shown on Figure 2. Each block contains the following data: Ky-Fan norm value, Ky-Fan norm middle value for the last X frames and deviation from threshold. One block zoomed in can illustrate the result in a better way. White spot is marking fragment that was determined as a motion area.

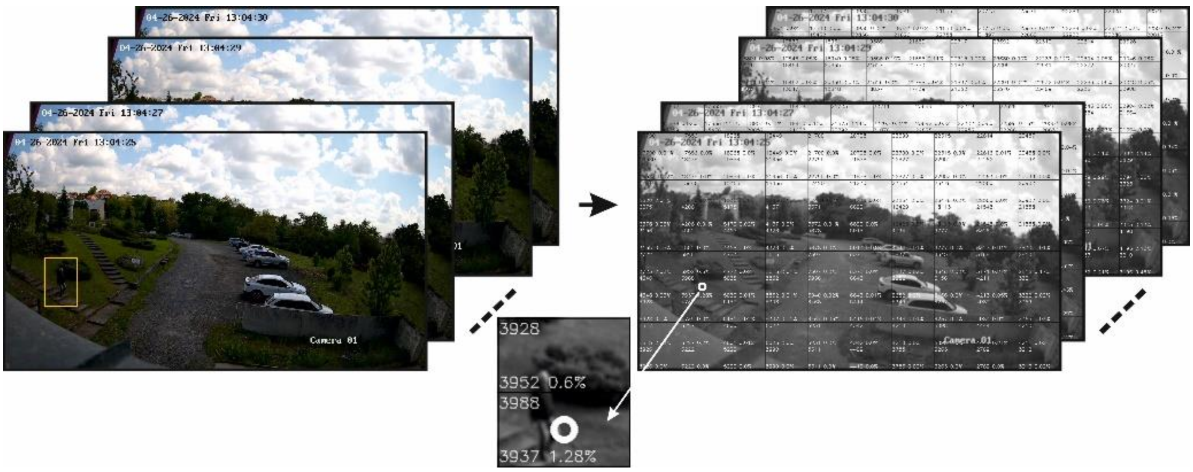

Figure 2: The result of frame-by-frame processing is a new video source in grayscale model with marked blocks with Ky-Fan norm value, Ky-Fan norm middle value for each block. Source: compiled by the authors.

3.1. Motion Detection

Every frame has been divided into 5x5, 10x10 and 20x20 blocks. We received matrix 5x5, 10x10 and 20x20. Received matrix block of a size is applicable for SVD transformation so singular values are calculated.

As a result, Ky Fan norm is found for each block. The choice of the block count depends on the size of the interest area.

The results are given for 10x10 (100 blocks) to better illustrate the result. Now we consider results of Ky Fan norm application for motion detecting.

Our approach is based on comparing the Ky-Fan norm value with the middle value for the last x frames.

If the difference exceeds threshold, we consider the scene changed or movement has been occurred in the fragment.

$$\Delta\sigma = f(x) - y \tag{5}$$

$f(x)$- middle value of last x frames, y- Ky-Fan value, $\Delta\sigma$ - motion detection sensitivity.

We selected fragment number 83 for the result demonstration in details. Fragment order is from left to right and from top to bottom. Fragment number 83 marked by yellow frame Figure 3. Motion detecting result is shown in Figure 4.

We compared Ky Fan norm value to the middle value for the last 20 frames. The experiment has established that the deviation from the threshold value should be within 1-3% for object motion detection.

On the frame number 90 Ky Fan norm value exceed threshold and movement has been detected. Ky Fan norm values for last 20 frames shown in Table 1.

Figure 3: Motion detecting result for fragment 83 marked by yellow frame. Source: compiled by the authors.

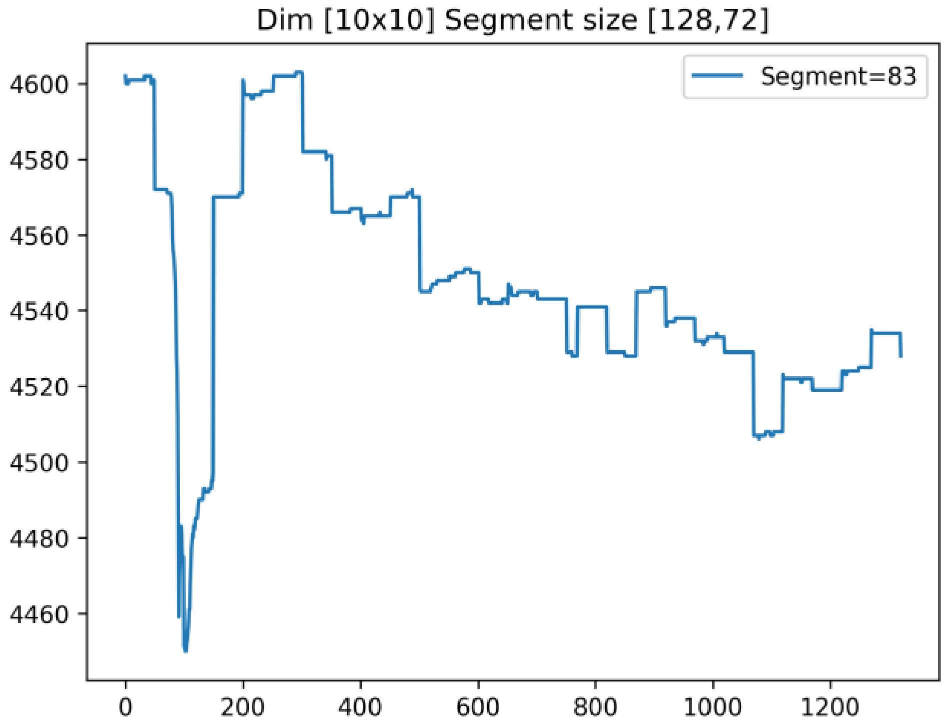

Figure 4: Ky-Fan norm fluctuation for fragment 83. Y is Ky Fan norm value, X is frame numbers. Source: compiled by the authors.

Table 1
Ky Fan norm values for last 20 frames

Frame number	Ky Fan norm	Frame number	Ky Fan norm
71	4572	81	4559
72	4571	82	4556
73	4571	83	4555
74	4571	84	4553
75	4571	85	4550
76	4571	86	4547
77	4571	87	4539
78	4571	88	4528
79	4570	89	4524
80	4567	90	4512

3.2. The tracking method

We can detect Ky Fan norm fluctuation in any fragment and if the value will exceed threshold, we can confirm movement in this fragment.

By combining the fragments in which the norm value has changed, we can build a graph of the object's movement.

As a result, we can track the way of the person on the parking Figure 5. In our video source the man stair down, cross parking and reach the car. White spots marked fragments with extreme values.

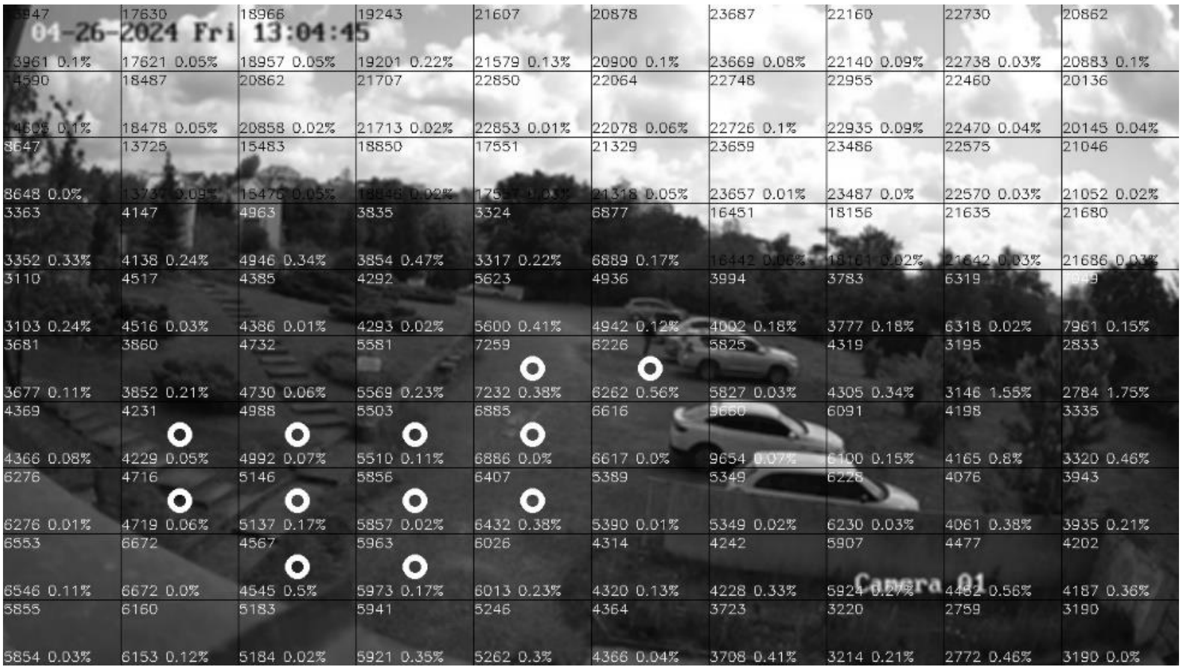

Figure 5: Motion tracking. Source: compiled by the authors.

3.3. Avoid artifacts

Natural lighting is variable: the sun's rays, the movement of clouds lead to changes in the illumination of different fragments. Unstable illumination will provide motion detection in fragments without scene changes or object motion. Our approach has to exclude such fragments from the motion graphs. We need select only neighboring fragments because the motion object will cross only fragment with common border. Any other fragments have to be excluded from our track. We can apply square filter 3x3, which will determine neighboring fragments with motion tracking graph Figure 6. Fragment 38 will exclude from motion track because no neighboring fragments in filter frame. Fragment 84 will apply in motion track because fragments 83 and 73 are neighbors.

Figure 6: Square filter 3x3, which will determine neighboring fragments with motion tracking graph. Source: compiled by the authors.

Motion tracking before filtering is show on Figure 7. Filtering result is shown in the Figure 8. It should be noted that filtering based on neighboring fragments can be applied when there is one moving object in the frame.

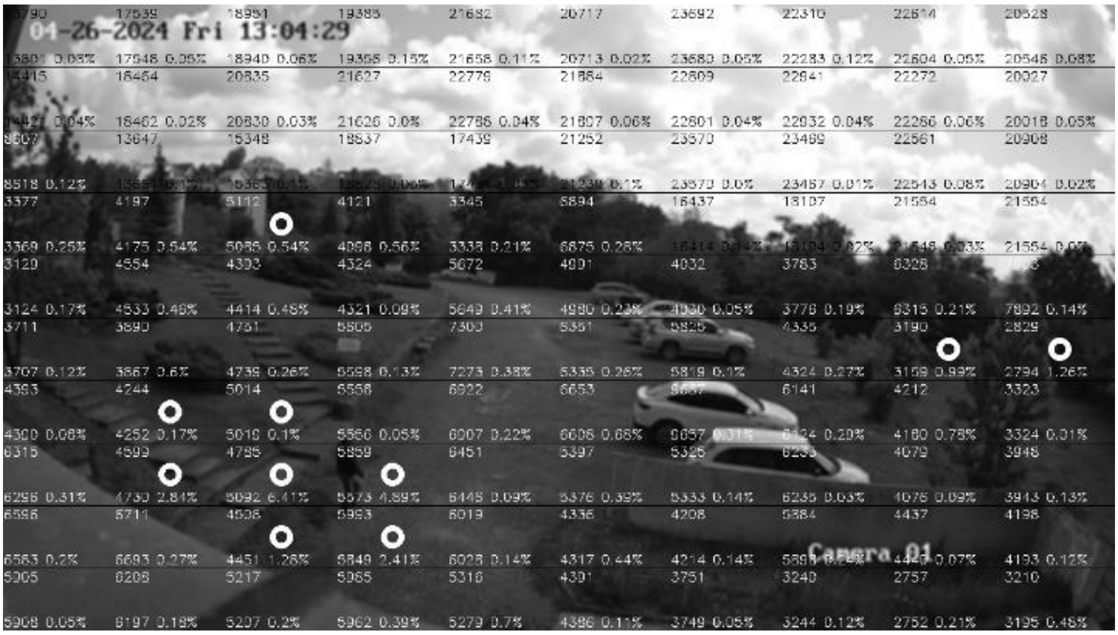

Figure 7: Motion tracking before filtering. Source: compiled by the authors.

In order to visualize results of Ky Fan norm usage for video analysis Python 3.10.11 application was developed and launched on Intel Core i5 processor with 16 Gb RAM and Windows OS installed. The application has dependencies from two open-source libraries with Apache license: OpenCV version 4.7.0 and numpy version 1.24.3.

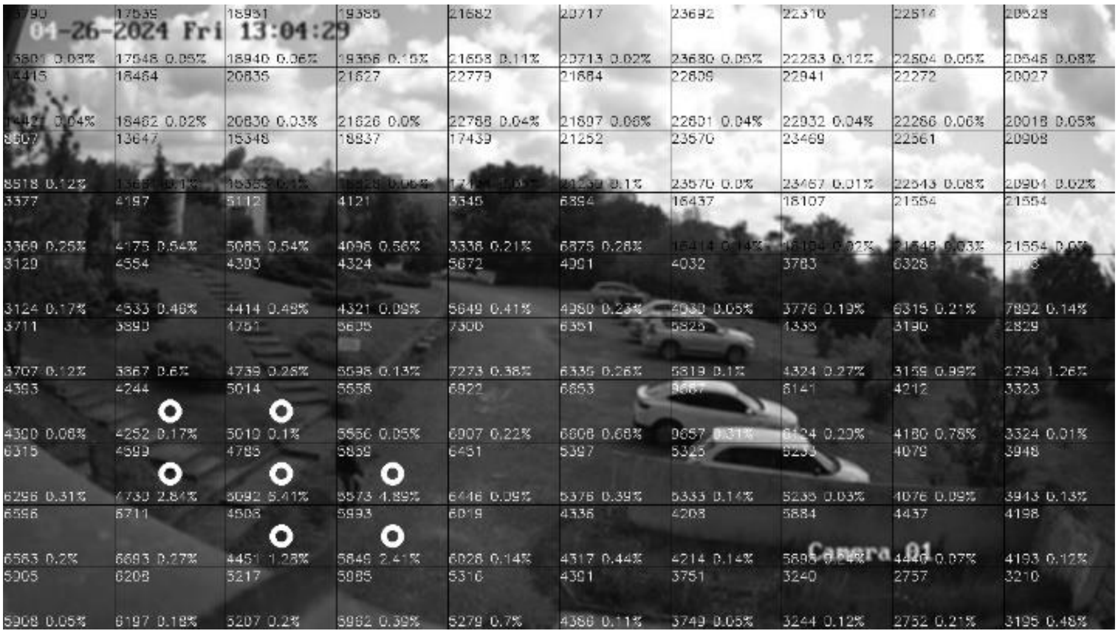

Figure 8: Motion tracking after filtering. Source: compiled by the authors.

4. Conclusions

Proposed approach will not answer the question of which object is moving, so there is no talk about classifying moving objects. The combination of fragment analysis and SVD allows finding fragments of the frame that could serve as a region of interest, and to which a more reliable algorithm with analysis only the necessary information of a much smaller size compared to the initial data could be applied. In this study a Ky Fan norm matching based SVD-covariance descriptor for motion detection and motion tracking was proposed. Methodology describes the SVD generation of the region covariance as the tracking feature. Experimental results demonstrate the effectiveness and prospect of proposed approach. The choice of fragments count depends on the size of the motion detection interest area. The object's size in relation to the fragment size affects the number of fragments that will be identified as motion detection region. If object's size covers several fragments, than motion detection will be established in all these fragments. In addition, it should be noted that the number of fragments will affect the motion detection threshold value. The low cost of SVD and the absence of additional computations make our approach simple and efficient, which was named as "pre-motion detector". As a result, the proposed approach allows to significantly reduce the size of analyzed information for further classification of moving objects. For further research the tracking performance will be further improved by combining probabilistic frames.

References

[1] H. Ghahremannezhad, H. Shi, C. Liu, Object Detection in Traffic Videos: A Survey, IEEE Trans. Intell. Transp. Syst. (2023) 1–20. doi:10.1109/tits.2023.3258683.

[2] Y. Wu, Y. Ye, C. Zhao, Coherent Motion Detection with Collective Density Clustering, y: MM '15: ACM Multimedia Conference, ACM, New York, NY, USA, 2015. doi:10.1145/2733373.2806227.

[3] C. Fu, Q. Li, M. Shen, K. Xu, Frequency Domain Feature Based Robust Malicious Traffic Detection, IEEE/ACM Trans. Netw. (2022) 1–16. doi:10.1109/tnet.2022.3195871.

[4] Parihar, S. S., & Khaskalam, A. A review paper on the different types motion detection techniques. International Research Journal of Modernization in Engineering Technology and Science 6 2 (2024) 1337-1342.

[5] Challa, S. Fundamentals of object tracking. Cambridge University Press, 2011. https://doi.org/10.1017/CBO9780511975837

[6] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, T.-K. Kim, Multiple object tracking: A literature review, Artif. Intell. (2020) 103448. doi:10.1016/j.artint.2020.103448.

[7] W. Yang, R. T. Tan, S. Wang, A. C. Kot, J. Liu, Learning to Remove Rain in Video With Self-Supervision, IEEE Trans. Pattern Anal. Mach. Intell. (2022) 1–18. doi:10.1109/tpami.2022.3186629.

[8] C. Chen, H. Li, Robust representation learning with feedback for single image deraining, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 7742-7751. doi:10.48550/arXiv.2101.12463

[9] C. Liu, X. Ma, S. Cao, J. Fu, B. B. Zhu, Privacy-preserving Motion Detection for HEVC-compressed Surveillance Video, ACM Trans. Multimedia Comput., Commun., Appl. 18.1 (2022) 1–27. doi:10.1145/3472669.

[10] Kanwal, Nadia. Motion Tracking in Video using the Best Feature Extraction Technique. 2009.

[11] M. K. Hossen, S. H. Tuli, A surveillance system based on motion detection and motion estimation using optical flow, in: 2016 International Conference on Informatics, Electronics and Vision (ICIEV), IEEE, 2016. doi:10.1109/iciev.2016.7760081.

[12] S. Karpuzov, G. Petkov, S. Ilieva, A. Petkov, S. Kalitzin, Object Tracking Based on Optical Flow Reconstruction of Motion-Group Parameters, Information 15 6 (2024) 296. doi:10.3390/info15060296.

[13] Z. Pan, D. Geng, A. Owens, Self-Supervised Motion Magnification by Backpropagating Through Optical Flow, in: NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems 13, 2023, pp. 253–273. doi:10.48550/arXiv.2311.17056

[14] X. Lu, R. Manduchi, Fast image motion segmentation for surveillance applications, Image Vis. Comput. 29 2 3 (2011) 104–116. doi:10.1016/j.imavis.2010.08.001.
[15] D. Zhang, G. Lu, Segmentation of moving objects in image sequence: A review. Circuits Systems and Signal Process 2 0(2001) 143–183. doi:10.1007/BF01201137
[16] C. Li, S. Q. Zheng, B. Prabhakaran, Segmentation and recognition of motion streams by similarity search, ACM Trans. Multimedia Comput., Commun., Appl. 3 3 (2007) 16. doi:10.1145/1236471.1236475..
[17] J. B. Kim, H. J. Kim, Efficient region-based motion segmentation for a video monitoring system, Pattern Recognit. Lett. 24 1 3 (2003) 113–128. doi:10.1016/s0167-8655(02)00194-0.
[18] LU, Nan, et al. An Improved Motion Detection Method for Real-Time Surveillance. IAENG International Journal of Computer Science 35 (2008) 10.
[19] A. Zahra, M. Ghafoor, K. Munir, A. Ullah, Z. Ul Abideen, Application of region-based video surveillance in smart cities using deep learning, Multimedia Tools Appl (2021). doi:10.1007/s11042-021-11468-w.
[20] H. W. Press, A. S. Teukolsky, T. W. Vetterling, P. B. Flannery, Numerical recipes in C : the art of scientific computing. 2002.
[21] D. Chetverikov, A. Axt, Approximation-free running SVD and its application to motion detection, Pattern Recognit. Lett. 31 9 (2010) 891–897. doi:10.1016/j.patrec.2009.12.031.
[22] N. Kamel, I. Kajo, Y. Ruichek, On Visual Periodicity Estimation Using Singular Value Decomposition, J. Math. Imaging Vis. 61 8 (2019) 1135–1153. doi:10.1007/s10851-019-00894-z.
[23] K. Meng, G. Minggang, C. Tao, The corner matching based on improved singular value decomposition for motion detection, in: Proceedings of the 31st Chinese Control Conference, 2012, pp. 3727-3732.
[24] M. Koliada, Ky fan norm application for video segmentation. Herald of Advanced Information Technology 1 3 (2020) 345-351.
[25] S. V. Mashtalir, D. P. Lendel, Video fragment processing by Ky Fan norm, Appl. Asp. Inf. Technol. 7.1 (2024) 59–68. doi:10.15276/aait.07.2024.5.
[26] R. Szeliski Computer Vision: Algorithms and Applications. 2nd Edition. Springer, 2022.