DOI: https://doi.org/10.15276/hait.07.2024.30
UDC 004:37.091.31.3

Moving object shape detection by fragment processing

Sergii V. Mashtalir[1)]
ORCID: http://orcid.org/0000-0002-0917-6622; sergii.mashtalir@nure.ua. Scopus Author ID: 36183980100
Dmytro P. Lendel[2)]
ORCID: http://orcid.org/0000-0003-3971-1945; dmytro.lendel@uzhnu.edu.ua
[1)] Kharkiv National University of Radio Electronic, 14, Nauky Ave. Kharkiv, 61166, Ukraine
[2)] Uzhhorod National University, 3, Narodna Square. Uzhhorod, 88000, Ukraine

ABSTRACT

The development of information technologies related to the analysis of visual information is inextricably linked with the methods of extracting various features or objects to facilitate their further analysis. This is due to the growing demands for visual data from the user. At the same time exactly object detection is one of the most fundamental and challenging tasks in locating objects in images and videos. Over the past, it has gained much attention to do more research on computer vision tasks such as object classification, counting of objects, and object monitoring. At the same time, researchers almost never paid attention to the fact of the shape of a moving object, and usually left this question for further analysis. At the same time, for example, for an object classification, having an object with clear shape outlines as input would be useful. This study provides video fragment processing for moving object shape detection. Our approach is based on dividing each frame into fragments that allow the present image frame as a square matrix for a formal description. The rectangular video frame has been transformed into a square matrix by SVD (singular value decomposition), where each element is a Ky Fan norm value used as a descriptor. Scene changes in the frame will affect Key Fan norm fluctuations. Comparing the fragment norm changes with other fragment norm changes will allow us to assess how significant these changes are. If the norm value exceeds the threshold value, we can include this fragment as part of the moving object. By combining such fragments together, we will detect moving object shapes. The threshold is dynamic and depends on time. In this study, we paid attention to calculating a threshold value for a fragment's reliable identification of a moving object. We also note that the experiments were conducted for the case when there is a stationary camera (surveillance camera) and some moving object in the field of view. And in this case, it was possible to obtain a clear contour of a complex shape for a moving object. More complex cases of simultaneous movement of both the object and the camera will be considered later.

Keywords: Video stream fragmentation; Ky Fan norm; singular value decomposition; object detection; object shapes; data analysis, video processing

For citation: **Mashtalir S. V., Lendel D. P. "Moving object shape detection by fragment processing".** ***Herald of Advanced Information Technology.*** **2024; Vol. 7 No.4: 414–423. DOI: https://doi.org/10.15276/hait.07.2024.30**

INTRODUCTION

Object detection is one of the most fundamental and challenging tasks for locating objects in images and videos [1]. It has been used in various computer vision applications such as face detection and face recognition [2], pedestrian counting [3], security systems, vehicle detection [4], self-driving cars, etc.

The importance of analyzing visual information, particularly dynamic content, cannot be overstated in modern science [5]. This is due to the exponentially growing amount of such data, as well as a significant increase in their quality. The importance of analyzing visual information, particularly dynamic content, cannot be overstated in modern science. The sheer volume and enhanced resolution of this data present unique opportunities and challenges. Applications such as image search systems, social network content monitoring, classification, and recognition systems highlight the need for efficient information processing. For example, while simplistic approaches like keyword filtering on social media can yield unfair results, advanced image processing methods like object detection ensure higher precision and fairness by focusing on visual content rather than textual cues.

One of the options for reducing the amount of information is to separate the object from the background. Techniques like background subtraction [7], data segmentation [8], feature extraction [9] and keypoint detection [10] have emerged as vital tools in reducing computational complexity. Background subtraction, in particular, is instrumental in video processing by isolating moving objects (foreground) from static elements (background). This method shines in stationary camera setups, where the background remains relatively stable. Similarly, segmentation techniques allow for the simplification of visual data by grouping pixels with shared characteristics, facilitating targeted analysis.

LITERATURE REVIEW

As one of the fundamental problems of computer vision, object detection forms the basis of many other computer vision tasks, such as instance segmentation [11, 12], image captioning [13], object tracking [14], etc.

From the application point of view, object detection can be grouped into two research topics "general object detection" and "detection applications" [15] where the former aims to explore the methods of detecting different types of objects under a unified framework to simulate the human vision and cognition, and the later one refers to the detection under specific application scenarios. Detecting the objects from the video is more challenging than still images due to appearance variations in video frames, such as defocus, motion blur, truncation, occlusion, and fast motion. Numerous research have been performed with video data, but it still needs improvements in detection [16, 17], [18].

The fragment-based analysis offers a promising approach to address these challenges by dividing video frames into smaller, meaningful regions or fragments for localized and robust object detection. This method enables the analysis of individual fragments, which are then aggregated to identify and track objects across frames, even under adverse conditions. Full-duplex strategy (FSNet) uses feature maps as fragments [19], focusing on the most discriminative and relevant features while ignoring irrelevant or noisy areas. The practical approach for detecting objects with distinct color or texture boundaries is Superpixel segmentation [20]. It defines fragments as groups of pixels with similar color or texture properties, simplifying the image into coherent regions for analysis. For detecting objects with well-defined shapes or edges, such as pedestrians or vehicles HOG (histogram of oriented gradients) [21], is particularly useful. It divides the image into local gradient fragments, analyzing the orientation of edges within these fragments. Treats keypoints as fragments (ORB (Oriented FAST and Rotated BRIEF)) [22], describing their local intensity patterns. ORB excels in detecting and matching objects in frames with significant changes, such as rotation or scale variations.

SIFT (scale invariant feature transform) is a widely used algorithm in computer vision for detecting and describing local features in images [23]. It is designed to be invariant to scale and rotation, meaning it can reliably identify the same features even if the image is resized, rotated, or partially occluded.

SSD (single shot detector) and YOLO (you only look once) models [24, 25] simplified object detection by integrating the detection process into a single neural network, enabling real-time performance. More recently, transformer-based models like DETR (Detection Transformer) (2020) have gained traction for their ability to perform object detection without reliance on anchor boxes, thereby simplifying training and improving robustness.

In our approach, we propose defining fragments as geometric parts of video frames, represented as matrices with arbitrary dimensions. By skipping traditional transformation steps like matrix vectorization, our method significantly reduces computational costs, freeing up computational resources required for this transformation. In the research [26], singular value decomposition of the matrix and the Ky Fan norm are proposed for scene change analysis. In the context of motion detection, this approach was expanded [27]. Dividing the frame into 5x5 or 10x10 allowed us to identify the fragments in which motion occurred. We decided to apply this approach to create a square matrix describing the frame with subsequent scene change analysis in each fragment of the frame separately. Analysis of the effectiveness of the obtained descriptor for different video data sizes shows that the change in the descriptor for each fragment is independent of the video size and aspect ratios [28]. The rectangle frame has been transformed into a square matrix 100x100 by singular value decomposition (SVD) where each element is Ky Fan norm value as a descriptor for the object detection.

If the norm value exceeds the threshold value, we can include this fragment as part of the moving object. By combining such fragments together, we will detect moving object shapes. The threshold is dynamic and depends on time. In this study, we paid attention to calculating a threshold value for a fragment's reliable identification of a moving object.

THE PURPOSE OF THE ARTICLE

The purpose of this article is to develop a video fragment processing method for moving object shape detection, leveraging singular value decomposition for video frame analysis. Our approach introduces a systematic framework that divides each frame into smaller, manageable fragments, enabling the transformation of the frame into a square matrix for precise formal description. This segmentation simplifies the computational process and ensures that the data is structured effectively for analysis.

Since the Ky Fan norm is inherently linked to SVD, we adopted it as the primary descriptor for each fragment. The Ky Fan norm's sensitivity to variations in matrix singular values makes it an effective metric for detecting scene changes. By observing the fluctuations in the Ky Fan norm across frame fragments, we can identify areas of significant transformation indicative of motion or object presence.

The methodology involves comparing norm changes within a fragment against those of other fragments within the same frame. This relative comparison enables the quantification of the significance of changes, effectively distinguishing between background variations and genuine object motion. When the norm value of a fragment surpasses a dynamically computed threshold, the fragment is flagged as belonging to a moving object.

By aggregating contiguous flagged fragments, the proposed method reconstructs the shapes of moving objects. This approach ensures that the method is robust against minor background disturbances while remaining computationally efficient. The dynamic threshold adapts over time, enhancing the method's ability to handle variations in video conditions, such as lighting changes or background noise.

Thus, to achieve the set goal, it is necessary to solve the following tasks:

1. Calculate the Ky Fan norm for each rectangular block of each frame of the video sequence.
2. Compare the norms and, if the dynamically calculated threshold value is exceeded, indicate the rectangular block as a moving object.
3. Combine adjacent "moving" blocks and form the shape of the object.

MAIN PART. SINGULAR VALUE DECOMPOSITION. KY FAN NORM OVERVIEW

Singular value decomposition [29] is one of the most widely used and versatile techniques in computer vision. It finds applications in signal and image processing, pattern recognition, motion detection [30], object detection [31], and many other fields. In video processing, it is crucial to use a method that is insensitive to the dimensionality of the input data due to the typically non-square shape of video frames. Processing a row matrix significantly reduces computational costs by avoiding the need for vectorizing the source matrix or transforming it into a square form for further operations.

Singular value decomposition is a mathematical approach in linear algebra that decomposes a given matrix into three component matrices, unveiling its fundamental structure and characteristics. This technique is widely utilized in areas like computer vision, data compression, and recommendation systems.

A method with these characteristics is Singular Value Decomposition. It enables the direct use of the original matrix without requiring dimensional transformations.

The SVD for m×n matrix A is a factorization of the form:

$$A = U\Sigma V^{*},$$

where U is an m × m complex unitary matrix; Σ is an m×n diagonal matrix with non-negative real numbers on the diagonal, and V is an n×n complex unitary matrix.

If A is real, U and V can be guaranteed to be also real orthogonal matrices. In such contexts, the SVD is often denoted:

$$A = U\Sigma V^{T}.$$

The singular values (σ_i) describe the "energy" or importance of each corresponding dimension in the matrix.

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

$$\|A\|_k^{KF} = \sigma_1(A) + ... + \sigma_k(A), \qquad (1)$$

is a matrix norm, called the Ky Fan k-norm. The last k of the Ky Fan norms (1), the sum of all singular values, is the trace norm or nuclear norm defined by

$$\|A\|_T = \mathrm{Tr}[(A^*A)^{1/2}] = \sigma_1(A) + ... + \sigma_n(A).$$

The Frobenius/Hilbert-Schmidt norm have form

$$\|A\|_{HS} = (\mathrm{Tr}[A^*A])^{1/2} = \sqrt{(\sigma_1^2(A) + ... + \sigma_n^2(A))}.$$

SVD does not impose the requirement for the source matrix to be square, making it highly adaptable for video processing. Its ability to handle matrices of any dimensions provides flexibility in representing the source data. This versatility allows video frames to be processed directly from the original image or any combination of descriptors without the need for additional transformations.

 ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

APPLICATION OF FRAGMENT PROCESSING FOR THE OBJECT SHAPES

In this section, we will consider the results produced by the developed application. Our experiment used a video surveillance camera and natural video sources. The camera position is stable.

We treat video as a sequence of frames (Fig. 1). Each frame is converted from RGB to a grayscale model so that the value of each pixel carries only intensity information. Thus, problems associated with color rendering and perceptions are excluded from consideration. Practically, it means that we will be working in the intensity domain.

Fig. 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 a grayscale model where each frame is divided into fragments. The received matrix fragment of size is applicable for SVD transformation, so singular values are calculated. Ky Fan norm is found for each fragment. The rectangle frame has been transformed into a square matrix 100x100 by SVD where each element is Ky Fan norm value as a descriptor. We used a “heat map” effect to visualize the transformation. The Ky Fan norm values serve as values for a particular color. In effect, we got the negative of the rectangular frame transformation into a square image.

Fig. 2 presents the natural image transformation into a new square image 100x100 where the pixel is Ky Fan norm value. All our calculations will be based on this matrix. Let's call this matrix A. On Fig. 2 we see frame number 144. The rectangle frame has been transformed into square matrix A, where the pixel value is the result of SVD and the Ky Fan norm is the value.

In order to visualize the results of Ky Fan norm usage for video analysis, the Python 3.10.11 application was developed and launched on an 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.
OpenCV [32] is an open source computer vision library that is used in real time computer vision. OpenCV was developed by Intel and now supported by Willow Garage and Itseez. OpenCV is designed and optimized for real time applications, although it’s developed in C and C++ languages, it’s a crossplatform library that runs on Linux, Windows and Mac OS. The OpenCV library contains hundreds of functions that cover many areas in computer vision such as robotics, medical image processing, security.

The new object's appearance in the frame, a change in lighting or the start of existing object movement will lead to fluctuations in the Ky Fan norm in certain fragments. To assess the norm fluctuation when the scene changes in the frame, it is necessary to compare these fluctuations with the threshold value. The threshold value will be different in different video sources. Even within a single video source, the threshold will not be constant. Objects may appear or disappear in the frame, and lighting may change. Thus, the threshold is a dynamic that depends on time.

a **b**

Fig. 2. **Video frame (a) transformation to square matrix A (b) from parking surveillance camera**
***Source*: compiled by the authors**

Comparing the fragment Ky Fan norm change with other fragments' norm changes gives us the possibility to assess how significant these changes are. If there are no significant changes in the frame, then the norm fluctuations in all fragments will be more or less similar. This will be similar to the calm sea surface, which is not flat, but all the waves have almost the same amplitude.

To estimate the threshold value, we need to accumulate the norm fluctuation values for each fragment over the last X frames. The obtained statistical data will make it possible to determine the deviation of the current norm value from the middle value for the last X frames. Comparing these deviations with deviations in other fragments, we can conclude how significant they are (Fig. 3).

0%	0%	0%	0%	0%	0%	0%	0%	....	0%
0%	0%	0%	0%	0%	0%	0%	0%	...	0%
0%	0%	0%	0%	0%	0%	0%	0%	...	0%
0%	0%	3%	0%	0%	0%	0%	0%	...	0%
0%	0%	3%	3%	0%	0%	0%	0%	...	0%
0%	0%	3%	3%	3%	0%	0%	0%	...	0%
0%	0%	3%	3%	0%	0%	0%	0%	...	0%
0%	3%	3%	3%	0%	0%	0%	0%	...	0%
...	...	...	...	...	...	...	...	...	0%
0%	0%	0%	0%	0%	0%	0%	0%	...	0%

Fig. 3. **Schematic representation of a 100x100 deviation matrix**
***Source*: compiled by the authors**

The result shows schematic 100x100 deviation matrix B representation. Fragments marked in color symbolize exceeding the threshold value, with the indication of this value as a percentage. The outline of the object is drawn.

In our study, we propose an approach based on threshold value calculation by comparing the current fragment norm value deviation from the middle value for the last X frames with the middle frame norm deviation as a whole.

We calculate Ky Fan norm (1) middle value Y for the last x frames

$$Y = \frac{1}{x}\sum_{i=1}^{x} kf_i$$

where kf_i is Ky Fan norm, x is frame count. The proposed algorithm assumes the matrix presence with size 100x100, which coincides with the frames division fragment numbers. Let's call this matrix B.

The element of matrix B will be the deviation δ of the current norm value from the middle value

$$\delta_i = \frac{kf_i - Y}{Y} \cdot 100\% .$$

The matrix B values will be recalculated during the fragment processing of each subsequent frame. If the fragment's middle value exceeds the element's middle value of the matrix B, we can conclude that a scene change has occurred in this fragment. So, this threshold value Δa have form

$$\Delta a = \delta_j - \frac{1}{k}\sum_{i=1}^{k} b_i ,$$

where b_i is element of matrix B, *k* is matrix size, δ_j is current fragment deviation value. If the condition is met

$$\Delta a > 0 ,$$

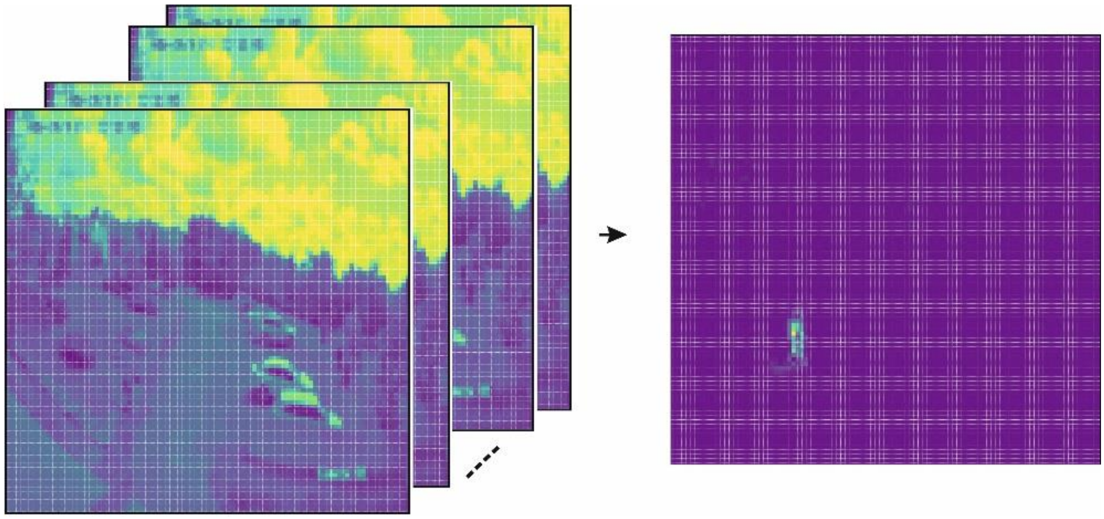
a threshold occurs for this fragment.

If there is more than one such fragment, then we will assume that we have found the object. By combining such fragments together, we will get the moving object shape (Fig. 4). In this case, as can be seen from the presented result, we form only the contour of a moving object without paying attention to other changes in the scene.

We selected frame number 144 for the detailed demonstration of the result. In certain fragments, the Ky Fan norm fluctuation exceeded the middle deviation of the norm in other fragments. The Ky Fan norm fluctuation for selected fragments (6428, 6429, 6430, 6528, 6530, 6627, 6628, 6629, 6630, 6727, 6728, 6729,6730, 6827, 6828, 6829, 6830, 6927, 6928, 6929, 6930, 7027, 7028, 7030, 7031, 7128, 7129, 7130, 7228, 7229, 7230, 7328, 7329, 7330, 7428, 7429, 7430, 7527, 7530, 7531, 7628, 7630, 7631, 7724, 7725, 7726, 7727) presented in Fig. 5 where axis x is video frames and axis y is Ky Fan norm value.

We selected only those fragments in which the norm value exceeded the threshold value. A significant change allows us to conclude that we found a moving object. In our study, the fragments number plays a meaningful way. Frame division has to be compared with the size of the moving object. This allows us not only to determine the presence of movement in a particular fragment, but by combining these fragments, we can quite accurately determine the moving object's shape.

 ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)



a b

Fig. 4 **Frame-by-frame processing;**
a – Frames have been converted into matrix A;
b – Matrix B "heat map" presentation for frame number 144 with fragments whose values exceeded the threshold value
***Source*: compiled by the authors**

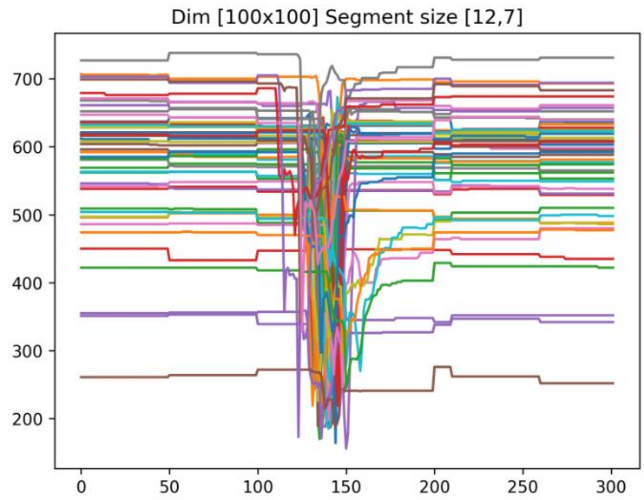


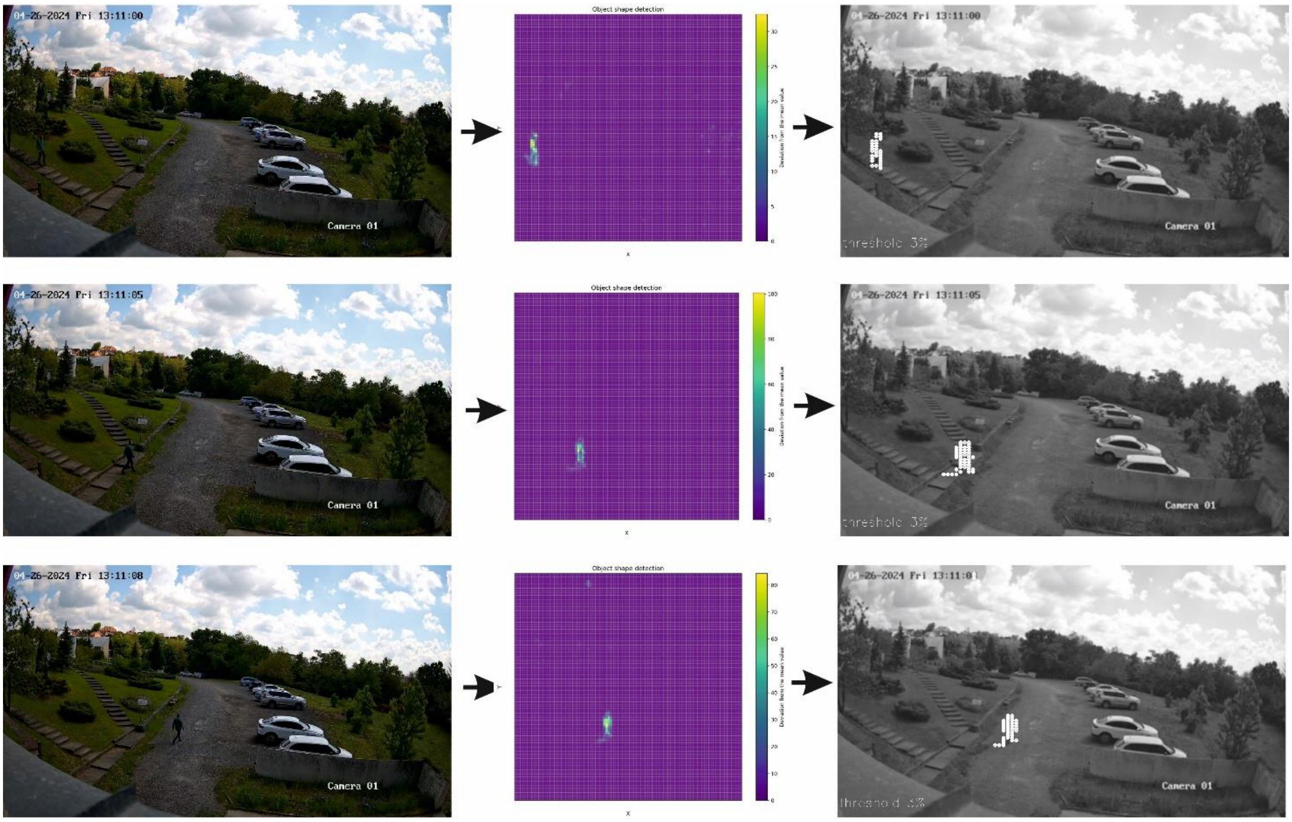
Fig. 5. **Ky Fan norm fluctuation for set of fragments. Extremum values in frame number 144**
***Source*: compiled by the authors**

Matrix B values are dynamic and depend on time. Matrix B elements depend on Ky Fan norm fluctuation for the last X frames from every fragment. The threshold value is dynamic, too. This allows us to accurately determine a moving object with contours by combining fragments with the same deviation. Fig. 6 shows a matrix B projection on the real frame. Object-detected frames marked by white spots.

The fragment approach faces the question: How many fragments should the frame be divided into? If the fragment size coincides with the frame size, then the norm fluctuation will indicate a general change in the scene in the frame. Dividing the frame on 5x5, 10x10, or 20x20 fragments allows detecting object movement and building motion tracking. Increasing the fragment numbers to 100x100 or 200x200 allows for highlighting the object and its contours.

CONCLUSIONS

The combination of fragment analysis and SVD allows for finding fragments of the frame that could serve as a region of interest and could be an object-detecting area. In this study, a Ky Fan norm matching-based SVD-covariance descriptor for object detection and object shapes was proposed. The methodology describes the SVD generation of the region covariance as the object detection feature. Experimental results demonstrate the correctness and prospect of the proposed approach. Scene changes in the frame will affect Ky Fan norm fluctuations. Comparing the fragment norm changes with other fragment norm changes will allow us to assess how significant these changes are. If the norm value exceeds the threshold value, we can include this fragment as part of the moving object. The object's size in relation to the fragment size affects the number of fragments that will be identified as the object detection region. By combining such fragments together, we will detect moving object shapes. The threshold is dynamic and depends on time. The methodology describes threshold value calculation. The low cost of SVD and the absence of additional computations make our approach simple and efficient. As a result, the proposed approach allows for a significant reduction in the size of analyzed information for further classification of



a **b** **c**

Fig. 6. **Fragment processing result. Object detection and matrix B projection on the real frame:**
a – Man walking through parking. Frame numbers are 15,144,198;
b – “**Heat map” shows threshold frames;**
c – Object-detected frames marked by white spots
Source: compiled by the authors

moving objects. In our study, the camera is stable. The movement of the camera relative to the object. significantly complicates data analysis. For future research, drone data will be used. A moving camera is relative to a moving object combined with fragment analysis for object shape detection.

REFERENCES

1. Chen, W., Luo, J. & Zhang, F. “A review of object detection: Datasets, performance evaluation, architecture, applications and current trends”. *Multimed Tools Appl*. 2024; 83: 65603–6566. DOI: https://doi.org/10.1007/s11042-023-17949-4.

2. Khan, M. A., Zaki, H. F. B. M., et al. “A systematic review on facial detection and recognition: Limitations and opportunities”. *AJoEEE*. Oct. 2024; 4 (2): 61–76. DOI: https://doi.org/10.69955/ajoeee.24.v4i2.70.

3. Miguel, J, Mendonça, P, Quelhas, A, Caldeira, J. M. L. P. & Soares, V. N. G. J. “Using computer vision to collect information on cycling and hiking trails users”. *Future Internet*. 2024; 16 (3):104. DOI: https://doi.org/10.3390/fi16030104.

4. Ma, C. & Xue, F. “A review of vehicle detection methods based on computer vision”. *Journal of Intelligent and Connected Vehicles*. 2024; 7 (1): 1–18. DOI: https://doi.org/10.26599/JICV.2023.9210019.

5. Verma, B., Zhang, L. & Stockwell, D. “Roadside video data analysis: Deep learning”. *Springer*. 2017. DOI: https://doi.org/10.1007/978-981-10-4539-4.

6. Peng, S., Wang, M. X., Shah, J. A. & Figueroa, N. “Object permanence filter for robust tracking with interactive robots”. *IEEE International Conference on Robotics and Automation (ICRA)*. Yokohama, Japan. 2024. p. 4909–4915. DOI: https://doi.org/10.1109/ICRA57147.2024.10611528.

7. Sengar, S. S. & Mukhopadhyay, S. “Foreground detection via background subtraction and improved three-frame differencing”. *Arab J Sci Eng*. 2017; 42 (8): 3621–3633. DOI: https://doi.org/10.1007/s13369-017-2672-2.

8. Mashtalir, S. & Mashtalir, V. “Spatio-temporal video segmentation”. *In Mashtalir, V., Ruban, I., Levashenko, V. (eds) “*Advances in spatio-temporal segmentation of visual data. Studies in computational intelligence*”*. 2020; 876. DOI: https://doi.org/10.1007/978-3-030-35480-0_4.

9. Xiu, H. “Video visual feature extraction and analysis based on DL-SLAM”. *PCCNT. Proceedings of the 2023 International Conference on Power, Communication, Computing and Networking Technologies*. 2023. p. 1–5. DOI: https://doi.org/10.1145/3630138.3630471.

10. Libal, U. & Łoziuk, I. L. “Comparison of keypoint detection methods for indoor and outdoor scene recognition”. *Journal of Automation, Mobile Robotics and Intelligent Systems*. 2017; 11 (4): 7–14. DOI: https://doi.org/10.14313/JAMRIS_4-2017/32.

11. Hariharan, B., Arbeláez, P., Girshick, R. & Malik, J. “Simultaneous detection and segmentation”. *In Computer Vision–ECCV. 13th European Conference*. Zurich, Switzerland. 2014. p. 297–312. DOI: https://doi.org/10.1007/978-3-319-10584-0_20.

12. Li, C., Song, D., Tong, R. & Tang, M. “Multispectral pedestrian detection via simultaneous detection and segmentation”. *arXiv preprint*. 2018. DOI: https://doi.org/10.48550/arXiv.1808.04818.

13. Staniūtė, R. & Šešok, D. “A systematic literature review on image captioning”. *Applied Sciences*. 2019; 9 (10). DOI: https://doi.org/10.3390/app9102024.

14. Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W. & Kim, T.-K. “Multiple object tracking: A literature review”. *Artificial Intelligence*. 2021; 293: 103448. DOI: https://doi.org/10.1016/j.artint. 2020.103448.

15. Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. “Object detection in 20 years: A survey”. *Proceedings of the IEEE*. 2023; 111 (3): 257–276. DOI: https://doi.org/10.48550/arXiv.1905.05055.

16. Zhu, H., Wei, H., Li, B., Yuan, X. & Kehtarnavaz, N. “A review of video object detection: datasets, metrics and methods”. *Applied Sciences*. 2020; 10 (21): 7834. DOI: https://doi.org/10.3390/app10217834.

17. Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L. & Tang, X. “New generation deep learning for video object detection: A survey”. *IEEE Transactions on Neural Networks and Learning Systems*. 2021; 33 (8): 3195–3215. DOI: https://doi.org/10.1109/TNNLS.2021.3053249.

18. Ouardirhi, Z., Mahmoudi, S. A. & Zbakh, M. ‘Enhancing object detection in smart video surveillance: A survey of occlusion-handling approaches”. *Electronics*. 2024; 13 (3): 541. DOI: https://doi.org/10.3390/electronics13030541.

19. Ji, G. P., Fu, K., Wu, Z., Fan, D. P., Shen, J. & Shao, L. “Full-duplex strategy for video object segmentation”. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. p. 4922–4933. DOI: https://doi.org/10.48550/arXiv.2108.03151.

20. Wang, M., Liu, X., Gao, Y., Ma, X. & Soomro, N. Q. “Superpixel segmentation: A benchmark”. *Signal Processing: Image Communication*. 2017; 56: 28–39. DOI: https://doi.org/10.1016/ j.image.2017.04.007.

21. Tomasi, C. “Histograms of oriented gradients”. *Computer Vision Sampler*. 2012. p. 1–6.

22. Setiawan, A., Yunmar, R. A. & Tantriawan, H. “Comparison of speeded-up robust feature (SURF) and oriented FAST and rotated BRIEF (ORB) methods in identifying museum objects using Low light intensity images”. *In IOP Conference Series: Earth and Environmental Science*. 2020; 537 (1): 012025. DOI: https://doi.org/10.1088/1755-1315/537/1/012025.

23. Bilik, S. & Horak, K. “SIFT and SURF based feature extraction for the anomaly detection”. *arXiv preprint*. 2022; *arXiv: 2203.13068*. DOI: https://doi.org/10.48550/arXiv.2203.13068.

24. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. & Berg, A. C. “SSD: Single shot multibox detector”. *In Computer Vision–ECCV. 14th European Conference*. Amsterdam, Netherlands. 2016; 9905*:* 21–37. DOI: https://doi.org/10.1007/978-3-319-46448-0_2.

25. Vijayakumar, A. & Vairavasundaram, S. “Yolo-based object detection models: A review and its applications”. *Multimedia Tools and Applications*. 2024; 83: 83535–83574. DOI: https://doi.org/10.1007/s11042-024-18872-y.

26. Koliada, M. "Ky Fan norm application for video segmentation". *Herald of Advanced Information Technology*, 2020; 1 (3): 345–351. DOI: https://doi.org/10.15276/hait.01.2020.1.

27. Mashtalir, S. & Lendel, D. "Video pre-motion detection by fragment processing". *12th International Scientific and Practical Conference "Information Control Systems and Technologies"*. 2024; 3790: 342–351, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85207837522&partnerID=40&md5=26529ee16efb141face273bd943660d5.

28. Mashtalir, S. V. & Lendel D. P. "Video fragment processing by Ky Fan norm". *Appl. Asp. Inf. Technol*. 2024; 7.1: 59–68. DOI: https://doi.org/10.15276/aait.07.2024.5.

29. Press, W. H. "Numerical recipes 3rd edition: The art of scientific computing". *Cambridge University Press*. 2007.

30. Chetverikov, D. & Axt, A. "Approximation-free running SVD and its application to motion detection". *Pattern Recognition Letters*. 2010; 31 (9): 891–897. DOI: https://doi.org/10.1016/j.patrec.2009.12.031.

31. Wu, A., Zhao, S., Deng, C. & Liu, W. "Generalized and discriminative few-shot object detection via SVD-dictionary enhancement". *Advances in Neural Information Processing Systems*. 2021; 34: 6353–6364, https://proceedings.neurips.cc/paper/2021/file/325995af77a0e8b06d1204a171010b3a-Paper.pdf.

32. Bradski, G. "Learning OpenCV: Computer vision with the OpenCV library". *O'REILLY*. 2008.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 23.09.2024
Received after revision 15.11.2024
Accepted 21.11.2024

DOI: https://doi.org/10.15276/hait.07.2024.30
УДК 004:37.091.31.3

Виявлення форми рухомого об'єкту шляхом фрагментного аналізу

Машталір Сергій Володимирович[1)]
ORCID: http://orcid.org/ 0000-0002-0917-6622; sergii.mashtalir@nure.ua. Scopus Author ID: 36183980100
Лендьел Дмитро Павлович[2)]
ORCID: http://orcid.org/ 0000-0003-3971-1945; dmytro.lendel@uzhnu.edu.ua
[1)] Харків Національний Університет Радіо Електроніки, пр. Науки 14. Харків, 61166, Україна
[2)] Ужгородський Національний Університет, пл. Народна 3. Ужгород, 88000, Україна

АНОТАЦІЯ

Розвиток інформаційних технологій, пов'язаних з аналізом візуальної інформації нерозривно пов'язаний з методами виділення різноманітних ознак або об'єктів для полегшення їх подальшого аналізу. Це пов'язано зі зростаючими вимогами до візуальних даних з боку користувача. При цьому виявлення саме об'єктів є одним із фундаментальних і найскладніших завдань визначення місцезнаходження даних на зображеннях і відео. У минулому багато уваги приділялося додатковим дослідженням задач комп'ютерного зору, таких як класифікація об'єктів, підрахунок об'єктів і моніторинг об'єктів. При цьому дослідники майже ніколи не звертали уваги на факт того, яка форма у рухомого об'єкту, а, зазвичай, залишали це питання для подальшого аналізу. В той же час для, наприклад, задачі класифікації мати в якості вхідної інформації об'єкт з чіткими абрисами форми було б корисним. Представлене дослідження забезпечує обробку фрагментів відео для визначення форми рухомого об'єкта. Наш підхід заснований на поділі кожного кадру на фрагменти, які дозволяють представити кадр зображення у вигляді квадратної матриці для формального опису. Прямокутний відеокадр був перетворений у квадратну матрицю за допомогою SVD, де кожен елемент є значенням норми Ky Fan, в якості дескриптору. Зміни сцени в кадрі впливатимуть на коливання норми Ky Fan. Порівняння змін норми фрагментів з іншими змінами норми фрагментів дозволить нам оцінити, наскільки ці зміни значні. Якщо значення норми перевищує порогове значення, ми можемо включити цей фрагмент до складу рухомого об'єкта. Комбінуючи такі фрагменти разом, ми виявимо форму рухомого об'єкту. Поріг динамічний і залежить від часу. У цьому дослідженні ми приділили увагу розрахунку порогового значення для надійної ідентифікації фрагментів рухомого об'єкта. Також зазначимо, що експерименти проводилися для випадку,

коли є нерухома камера (камера відеоспостереження) і деякий рухомий об’єкт в полі зору. І в цьому випадку вдалося отримати чіткий контур складної форми для рухомого об’єкту. Більш складні випадки одночасного руху і об’єкту, і камери буде розглянуто пізніше.

Ключові слова: фрагментний аналіз відео послідовностей; норма Ky Fan; декомпозиція сингулярного значення; виявлення об'єктів; виявлення контурів об'єктів; фналіз даних

ABOUT THE AUTHORS

Sergii V. Mashtalir - Doctor of Engineering Science, professor, Informatics department. Kharkiv National University of Radio Electronics, 14, Nauky Ave. Kharkiv, 61166, Ukraine
ORCID: https://orcid.org/0000-0002-0917-6622; sergii.mashtalir@nure.ua. Scopus Author ID: 36183980100
Research field: Image and video processing; data analysis

Машталір Сергій Володимирович - доктор технічних наук, професор кафедри Інформатики. Харківський національний університет радіоелектроніки, пр. Науки 14. Харків, 61166, Україна
ORCID: https://orcid.org/0000-0002-0917-6622, Scopus Author ID: 36183980100 sergii.mashtalir@nure.ua

Dmytro P. Lendel - PhD student, Department of Informative and Operating Systems and Technologies. Uzhhorod National University, 3, Narodna Square. Uzhhorod, 88000, Ukraine
ORCID: https://orcid.org/0000-0003-3971-1945; dmytro.lendel@uzhnu.edu.ua
Research field: Video processing; video stream segmentation

Лендьел Дмитро Павлович - аспірант кафедри Інформаційних управляючих систем та технологій. Ужгородський національний університет, площа Народна 3. Ужгород, 88000, Україна
ORCID: https://orcid.org/0000-0003-3971-1945; dmytro.lendel@uzhnu.edu.ua