DOI: https://doi.org/10.15276/aait.07.2024.5
UDC 004.93'1

Video fragment processing by Ky Fan norm

Sergii V. Mashtalir[1)]
ORCID: http://orcid.org/0000-0002-0917-6622; sergii.mashtalir@nure.ua. Scopus Author 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

In this study, we focused on the formalization of video frame descriptions in the context of solving video segmentation problem. Since native video data can have various sizes, dividing each frame into blocks allows 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. We conducted a comparative analysis of the effectiveness of the obtained descriptor for different video data sizes and with different aspect ratios, showing that the change in the descriptor for each block is independent of the video size and aspect ratios. Changes in the descriptors for each block from frame to frame are identical for video data of varying sizes. This means that as a result of such fragment transform, a square matrix of a fixed size is created, regardless of the output video size. This makes it possible to unify further processing of the video, which can be useful for the task of information search in large video databases under the conditions of providing a query "ad exemplum". In this case, we can analyze the existing database in offline mode and match each video with a fixed square matrix of descriptors, which will significantly reduce the time and amount of resources when matching with the query. Also, this approach can be effectively used to analyze video data for the motion detection and scene change tracking.

Keywords: Video stream fragmentation; Ky Fan norm; Singular value decomposition

For citation: **Mashtalir S. V., Lendel D. P. "Video fragment processing by Ky Fan norm".** ***Applied Aspects of Information Technology.*** **2024; Vol. 7 No. 1: 59–68. DOI: https://doi.org/10.15276/aait.07.2024.5**

INTRODUCTION

Video data segmentation is an area of current interests in the field of video processing and computer science. It affects dealing with big data storages, streaming video analysis, processing data from surveillance cameras and so on. Recent research has focused on fast and real time video data processing [1, 2], scene change detection [3, 4], semantic segmentation of video data [5]. Modern approaches suggest mathematical apparatus like time series, cluster analysis, and can be successfully applicable to detect scene changes and key frames [6, 7].

The quality of video data is increasing, as a result, increase in the data amount in each frame, complicating data processing. In the context of scene change, there is not always a need to analyze the entire frame. It is practical to divide the frame into blocks with subsequent analysis of each block separately. This approach allows determining the region of interest that deserves attention.

Parallel calculation provides a time advantage, but the calculation complexity remains [8]. Reducing the block size for analysis offers advantages not only in time but also in calculation speed.

When it comes to solving video stream segmentation and identification of key frames, attention should be paid to the formation of descriptors and the method of comparing two descriptors. This topic is wide enough and is actively developed. Newest achievements and novel mechanisms are published in [9, 10], [11, 12]. Descriptor-based approaches allow to reduce dimension of a multidimensional source data significantly, on the other hand – its calculation still require full processing of video frame (block). So, having in place meaningful metric retrieved from multidimensional source data directly rather than vectorized representation allow to skip one step of the processing comparing to traditional approach [13]. At the same time, such approaches allow you to compare videos of different lengths [14, 15] however, questions remain with comparing videos of different dimensions. When dealing with native video data, we will one way or other encounter transformations of non- square matrices. Almost all available techniques use summarizing methods to equalize the size of the matrices which lead to loss some parts of information. Approach present a non-square determinant kernel based on Radic's definition for efficient video data search [16].

Real video data is non-square. If we consider dividing the frame into blocks of 10x10 or 20x20, we get a frame square matrix description, where each element is a non-square matrix. Ability to skip the step of matrix transformation to square dimension or vectorization using a descriptor allows reducing computational costs and forming a frame square matrix description.

Special attention should be pay to the matrix decomposition, which allows representing the original matrix`s block in a form that that has the required characteristic, such as symmetry, orthogonality, and etc. In machine learning, matrix decomposition is actively used to detect hidden patterns and connections in the output data. The most used matrix decompositions that allow one to reduce dimension are the principal component method, nonnegative matrix decomposition, and singular decomposition. The use of matrices opens wide possibilities in the application of existing methods and approaches [17, 18]. In video processing, it is essential to have a technique that is insensitive to the size of the data source due to the non-square nature of the video frame (block).

In the research [19], singular value decomposition of the matrix and the Ky Fan norm are proposed for scene change analysis. We decided to apply this approach to create a square matrix describing the frame with subsequent scene change analysis in each block of the frame separately. We analyzed changes in the Ky Fan norm from frame to frame in the context of the video frame size.

THE PURPOSE OF THE ARTICLE

Considering the fact that in the most general case, the video entering the input of any system for further analysis can be of arbitrary size and aspect ratio, there is a need to reduce it to some unified compact description so that it can be easily compared with other video in the future (with a different resolution or aspect ratio) to find matches. However, one of the shortcomings in existing matrix approaches for video comparison is the need for matrix consistency.

Thus, the purpose of this article is to develop a video preprocessing method to obtain a compact video description based on singular value decomposition (SVD) applied to rectangular video blocks with a square matrix of descriptors as result. Since Ky Fan norm is built on top of SVD, it was taken as a descriptor for individual video block. Video frame is represented as a sequence of image blocks which in turn are matrices. For each block matrix SVD is applied so that Ky Fan norm can be retrieved. The approach is based not only on fixing scene changing in the blocks, but also on the analysis of the change in the descriptor value depending on the size of the video data.

MAIN PART
SINGULAR VALUE DECOMPOSITION. KY FAN NORM OVERVIEW

The singular value decomposition is the most common and useful decomposition in computer vision. The goal of computer vision is to explain the three dimensional world through two dimensional pictures. In the real world, most of these pictures will produce both square and non-square singular matrices and transformations. Inverting transformations from two dimensions to three dimensions will therefore not be completely accurate, but can be estimated quite well through singular value decomposition. Singular value decomposition will also allow us to establish a sense of order in objects and is therefore useful whenever attempting to compare.

In video processing it is worthwhile to have technique insensitive to the dimension of a source data due to the non-square nature of video frame. Dealing with row matrix allows reduce computational costs significantly because in this case we can avoid vectorization of a source matrix or transformation into another form like square one for further processing.

Possible technique with such characteristics is singular values decomposition. It allows use initial matrix without dimension transformations. The singular value decomposition (SVD) for $m \times n$ matrix A is a factorization of the form

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

where U is an $m \times m$ complex unitary matrix, Σ is an $m \times n$ diagonal matrix with non-negative real numbers on the diagonal, and V is an $n \times n$ 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 SVD is related to many common matrix norms and provides an efficient method to calculate them. It follows from our existence proof for the SVD that

$$\|A\|_2 = (\rho(A^*A))^{1/2} = \sigma 1(A),$$

and this equivalence provides the standard method to calculate the matrix 2-norm, e.g. in Matlab's norm (A,2). More generally, the sum of the first k singular values

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

is a matrix norm, called the Ky Fan k-norm. In practice, the parameter is selected as 85% of the total number of eigenvalues.

The last of the Ky Fan norms, the sum of all singular values, is the trace norm or nuclear norm defined by

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

The Frobenius/Hilbert-Schmidt norm

$$\|A\|_{HS} = (Tr[A^*A])^{1/2} = (\sigma 1\ 2\ (A) + ... + \sigma n\ 2\ (A))1/2.$$

More information related to the SVD approach and derivation can be found in [20].

Singular value decomposition 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.

Singular value decomposition metrics and norms are the standard approach to video analysis, but non- square norms is quite a modern trend.

APPLICATION OF KY FAN NORM FOR THE FRAGMENT PROCESSING

In this section we will consider results produced by developed application. The experiments were carried out for 10 videos with various sizes and various technical characteristics related to different domains: video from surveillance cameras, modeling of geometric surfaces, animation clip. In our experiments, we used videos of the following formats: 825x480, 1280x720 and 1920x1080. First step is to represent source videos as a sequence of frames. An example of such a representation is shown in Fig. 1. Then 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.

An alternative to this approach is to use a transition to another color space, for example HSV, which is also focused primarily on the intensity value. The result of frame-by-frame processing is a new video source in grayscale model with marked blocks and values of Ki Fan norms for each block is shown on Fig. 2 with one block zoomed in to better illustrate the result.

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

Every frame of each different size videos has been divided on 5x5, 10x10 and 20x20 blocks. We received matrix 5x5, 10x10 and 20x20 for same video sources: 825x480, 1280x720 and 1920x1080. 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 and difference between them is taken. Now we consider results of Ky Fan norm application for video fragmentation for several test videos.

When dividing the 5x5 frame, we received the following block sizes 170x96, 256x144, 384x216 for video formats 825x480, 1280x720 and 1920x1080. An example of a 5 by 5 split of one frame from a video of different sizes is shown in Fig. 3. The total number of blocks is 25 per frame. At the same time, you can see that the values of the Ky Fan norm differ both for individual blocks within the frame and for the same blocks on frames of different resolutions. And if the first difference is not in doubt, then the differences in the norm values for video frames of different dimensions forced us to study their changes in more detail to obtain an answer about the possibility of carrying out such a unification of videos of different dimensions.

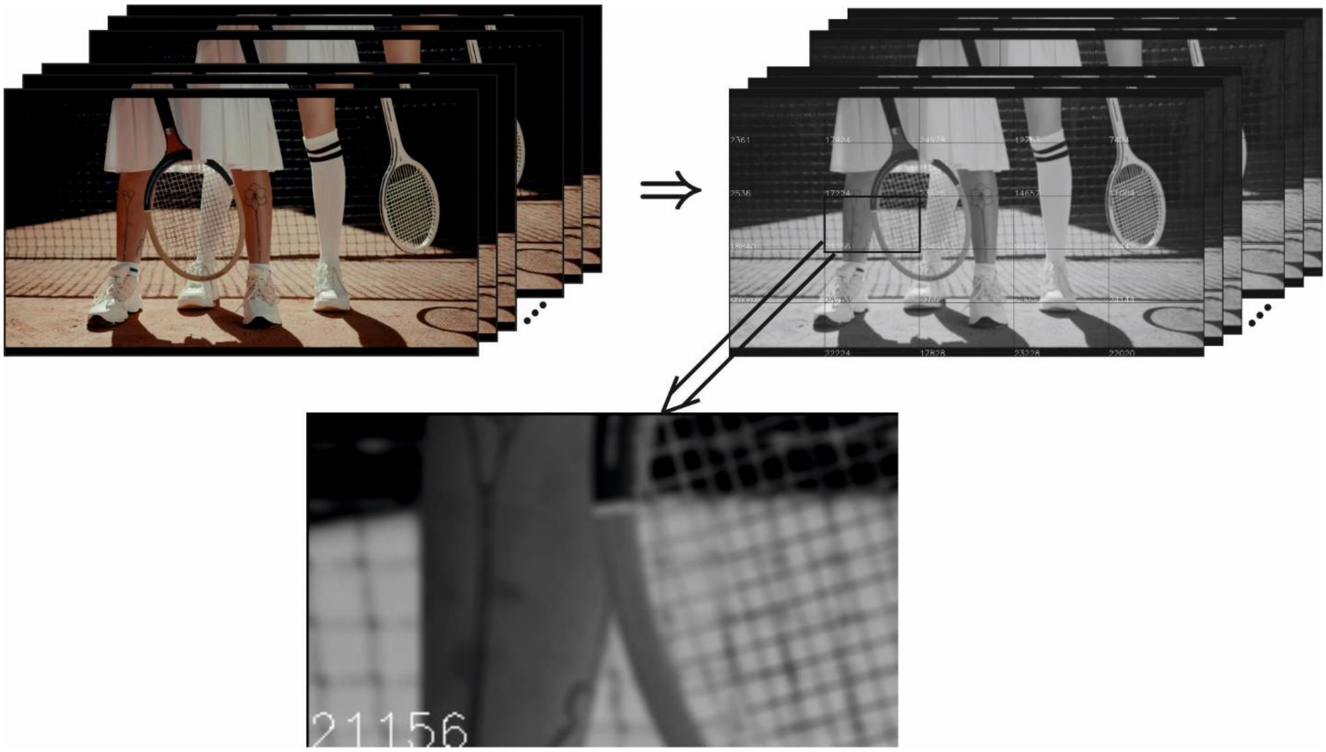
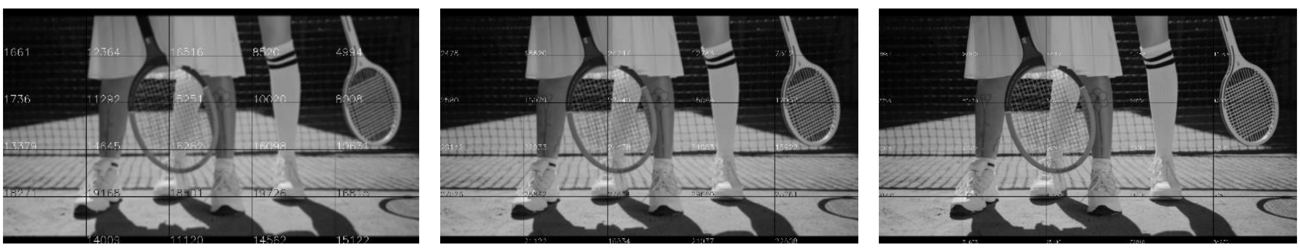


Fig. 2. **The result of frame-by-frame processing is a new video source in grayscale model with marked blocks and values of Ky Fan norms for each block**
***Source:* compiled by the authors**

For this purpose, graphs of Ky Fan norm values were constructed for all video sequences individual blocks of all sizes considered. The Fig. 4 shows the Ky Fan norm correlations for each format (a) 825x480, (b) 1280x720, and (c) 1920x1080. The X-axis is the frame numbers, the Y-axis is the Ky Fan's norm value. Blocks of the frame are marked in color. From the given data, we can conclude that the correlation of the Ky Fan norm does not depend on the format and size of the video source.

In other words, the difference in blocks on the same frames of different sizes is associated only with the number of pixels when calculating the Ky Fan norm and, accordingly, differ only by the corresponding coefficient, which does not affect the trends of changes that occur in the video itself. This can be seen both for blocks in which movements occurred and, accordingly, there were changes, and for essentially background blocks in which there were no changes throughout the entire video fragment.

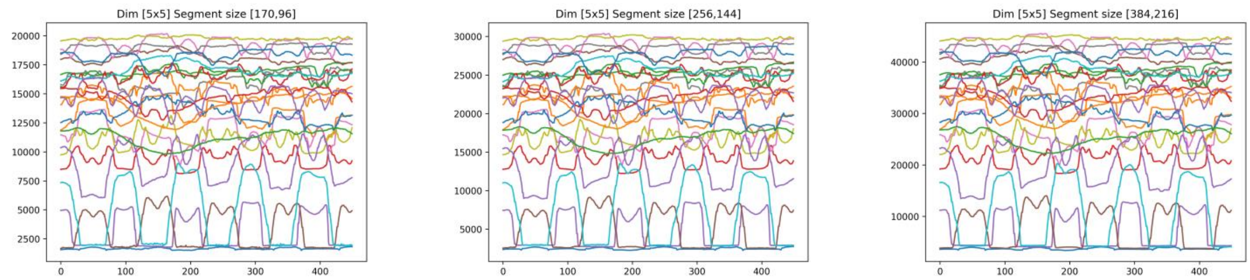
Next, an experiment was carried out with a dividing a 10x10 frame, we received the following block sizes 85x48, 128x72, 192x108 for video formats 825x480, 1280x720 and 1920x1080 as its shown on Fig. 5.



a b c

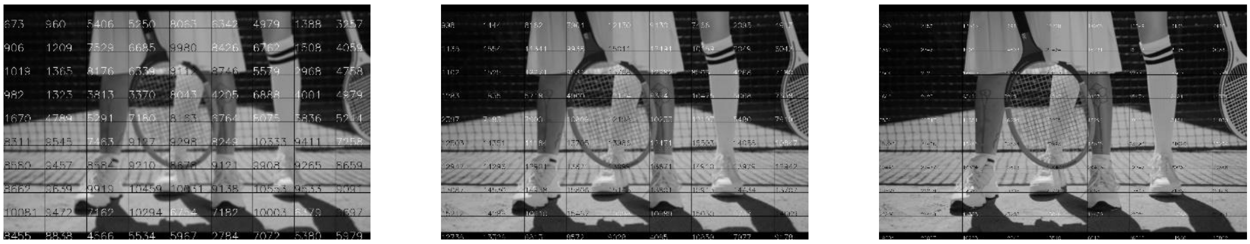
Fig. 3. **Frame number 100 divided by 25 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**

 ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)



a b c

Fig. 4. **Ky Fan norm fluctuations for frames divided by 25 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**



a b c

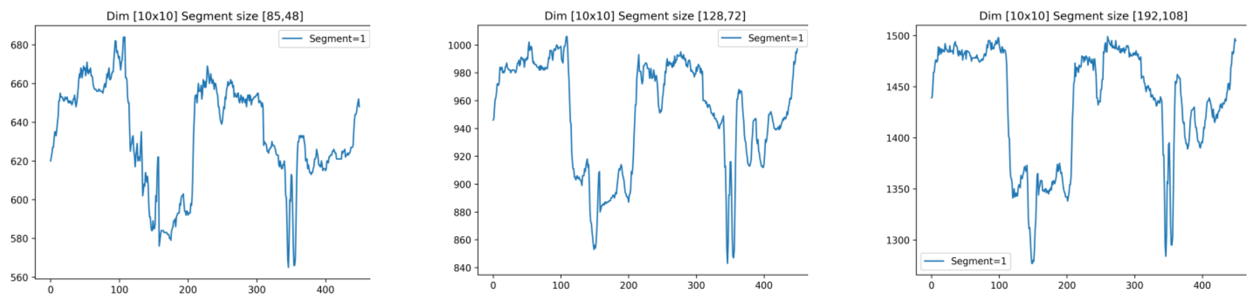
Fig. 5. **Frame number 100 divided by 100 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**

The total number of blocks is 100 per frame. The Fig. 6 shows the segment 1 Ky Fan normal correlations for each format (a) 825x480, (b) 1280x720, and (c) 1920x1080. Since the number of blocks is significant, the first block (upper left corner) was chosen for clarity. From the given data, we can concluded that the Ky Fan`s norm correlation is preserved with small deviations.

Next, when we dividing videos on a 20x20 frame, we received the following block sizes 42x24, 64x36, 96x54 for video formats 825x480, 1280x720 and 1920x1080. The result of this partition is shown in Fig. 7. The total number of blocks is 200 per frame. The Fig. 8 shows the Ky Fan normal correlations for each format (a) – 825x480, (b) – 1280x720, and (c) – 1920x1080.

For clarity, the first block (upper left corner) was selected. From the given data, we can conclude that the correlation of the form of Ky Fan has significant deviations.

Also, in our experiments, we used videos of the formats: 640x480, 960x720 and 1440x1080 to look at the result when analyzing video with an aspect ratio of 4:3. To do this, we also divided the video of these formats into blocks; an example of such a division into 25 fragments (5 by 5) is shown in Fig. 9.

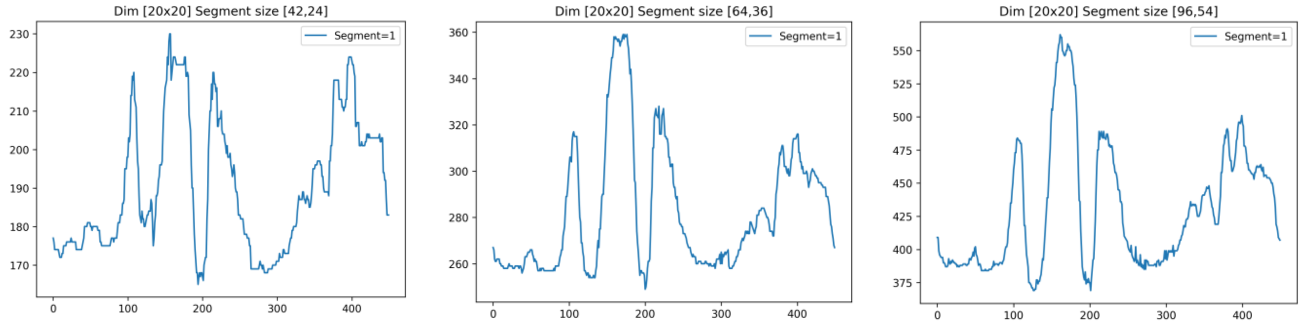


a b c

Fig. 6. **Segment 1 Ky Fan norm fluctuations for frames divided by 100 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**

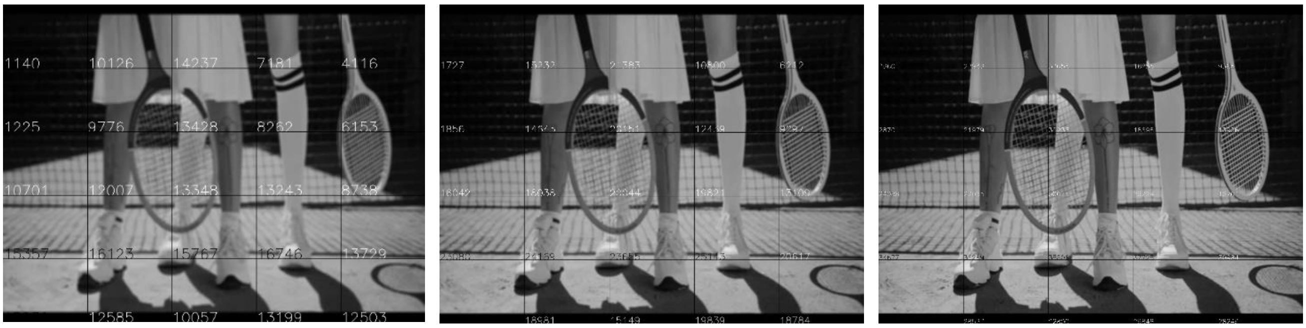
a b c

Fig. 7. **Frame number 100 divided by 200 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**



a b c

Fig. 8. **Segment 1 Ky Fan`s norm fluctuations for frames divided by 200 blocks:**
a – 852x480; b – 1280x720; c – 1920x1080
***Source:* compiled by the authors**



a b c

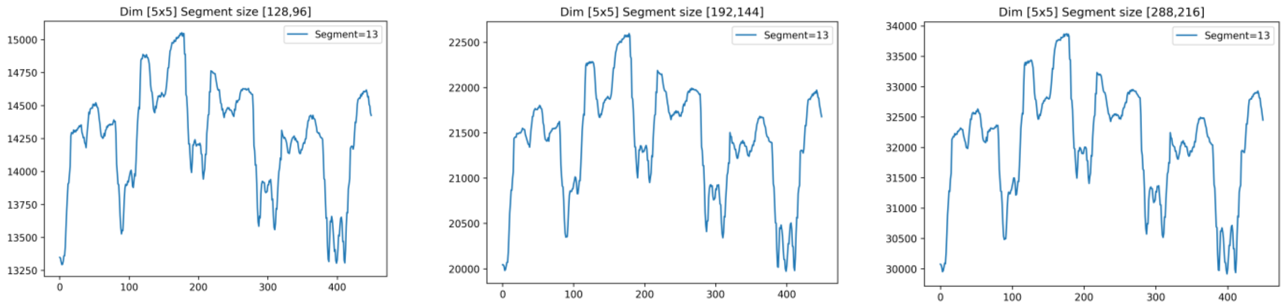
Fig. 9. **Frame number 100 divided by 25 blocks (a) 640x480, (b) 960x720 (c) 1440x1080**
***Source:* compiled by the authors**

The Fig. 10 shows the Ky Fan norm correlations for each format for frame 13. As you can see, changing the aspect ratio does not affect the selection of descriptors using the Ky Fan norm.

Table 1 shows the norm results for the first 30 frames of the video in three different dimensions. As can be seen from the obtained values, the trends of change are preserved, and the difference in value is associated with the different number of eigenvalues after SVD, which we take to calculate the Ky Fan norm.

As can be seen from Table 2, if we take the ratio of values for video resolutions of 640x480 and 960x720, then the number of pixels is related as 2 to 3. At the same time, the ratio of Ky Fan norms for fragments is also preserved and is equal to approximately 0.66 with an error of 3-4 decimal places.

It is also easy to trace the same exact relationship (2 to 3 for video and 0.66 for Ky Fan norms) for videos with sizes 960x720 and 1440x1080 with the corresponding norm value.



a b c

Fig. 10. **Segment 13 Ky Fan norm fluctuations for frames divided by 25 blocks (a) 640x480, (b) 960x720 (c) 1440x1080**

***Source:* compiled by the authors**

Table 1. **Segment 13 Ky Fan norm fluctuations values for first 30 frames divided by 25 blocks**

640x480	13348	13343	13315	13293	13295	13317	13360	13360	13427	13569
	13805	13906	13912	13975	14039	14268	14295	14288	14287	14314
	14311	14300	14315	14312	14328	14329	14345	14339	14349	14353
960x720	20044	20043	20026	19981	19990	20028	20070	20070	20184	20390
	20725	20868	20866	20992	21068	21436	21448	21453	21451	21492
	21487	21496	21501	21496	21507	21523	21550	21539	21538	21533
1440x1080	30075	30067	30023	29952	29971	30028	30093	30083	30270	30566
	31072	31283	31272	31474	31573	32129	32153	32178	32174	32225
	32222	32216	32263	32259	32282	32289	32317	32308	32293	32291

***Source:* compiled by the authors**

Table 2. **Segment 13 Ky Fan norm correlations for first 30 frames divided by 25 blocks**

640x480/960x720	0.6659	0.6657	0.6648	0.6652	0.6651	0.6649	0.6657	0.6657	0.6652	0.6655
	0.6661	0.6664	0.6667	0.6657	0.6664	0.6656	0.6665	0.6660	0.6660	0.6660
	0.6660	0.6652	0.6658	0.6658	0.6662	0.6657	0.6657	0.6657	0.6658	0.6666

***Source:* compiled by the authors**

Thus, it has been experimentally established that an increase in the number of blocks in combination with the video source size leads to a decrease in the output matrix dimensionality for the SVD transformation and, as a result, a significant deviation in the Ky Fan norm correlation. If the size of the input matrix is lower than the threshold value, then the standard deviation becomes significant. Changing the video size leads to an increase in the norm value by the corresponding coefficient.

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.

CONCLUSIONS

Since native video data can have various sizes, dividing each frame into blocks allows present image frame as an equal square matrix for a formal description. The frame block is a matrix of arbitrary dimensions. Rectangular fragment analysis of video frames using Ky Fan norm and obtaining a square matrix of descriptors is the scientific novelty of this work. From a practical point of view, obtaining an abbreviated description of video frames allows you to reduce both time and computational costs when further solving a whole range of video analysis problems. We conducted a comparative 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. However, it has been experimentally shown that the proposed approach to obtaining descriptors depends on how many fragments the video frames are divided into. As the number of fragments increases, more deviations appear in the trends of video sequences for individual frame fragments, which can lead to inaccuracies in further analysis. In the future, we plan to conduct research on the obtained fragmentary descriptions of video frames to solve problems related to motion and changes detection. As an intermediate result, it should be said that our approach allows you to formally describe the frame in the form of a square matrix and apply for it other algorithms for video analysis.

REFERENCES

1. Apostolidis, E. & Mezaris, V. “Fast shot segmentation combining global and local visual descriptors”. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence: Italy. 2014. p. 6583–6587, https://www.scopus.com/authid/detail.uri?authorId=55893131400.
DOI: https://doi.org/10.1109/ICASSP.2014.6854873.

2. Xu, W., Shen, Y., Lin, Q., Allebach, J. P. & Zhu, F. “Exploiting temporal information in real-time portrait video segmentation”. *HCMA '23: Proceedings of the 4th International Workshop on Human-Centric Multimedia Analysis*. 2023. p. 33–39, https://www.scopus.com/authid/detail.uri?authorId=7005836978. DOI: https://doi.org/10.1145/3606041.3618067.

3. Amini, E. & Jassbi, S. J. “A quick algorithm to search and detect video shot changes”. *International Journal of Computer Applications (0975 8887)*. 2015; 115 (3): 1–4, https://www.scopus.com/authid/detail.uri?authorId=24328853600. DOI: https://doi.org/10.5120/20128-2213.

4. Rotman, D., Yaroker, Ye., Amran, E., Barzelay, U. & Ben-Ari R. “Learnable optimal sequential grouping for video scene detection”. *MM '20: Proceedings of the 28th ACM International Conference on Multimedia*. October 2020. p. 1958–1966. DOI: https://doi.org/10.1145/3394171.3413612.

5. Li, Y., Shi, J. & Lin, D. “Low-latency video semantic segmentation”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018. p. 5997–6005, https://www.scopus.com/authid/detail.uri?authorId=8662530900.
DOI: https://doi.org/10.48550/arXiv.1804.00389.

6. Nayak, T. & Bhoi, N. “Comparative analysis of different clustering techniques for video segmentation”. In: Saini, H., Singh, R., Kumar, G., Rather, G., Santhi, K. (Eds.). “Innovations in electronics and communication engineering”. *Lecture Notes in Networks and Systems*. *Springer.* Singapore. 2019; 65: 23–31, https://www.scopus.com/authid/detail.uri?authorId=24722693600.
DOI: https://doi.org/10.1007/978- 981-13-3765-9_3.

7. Sousa e Santos, A. C. & Pedrini, H. “Video temporal segmentation based on color histograms and cross-correlation”. In: Beltrán-Castañón, C., Nyström, I., Famili, F. (eds) “Progress in pattern recognition, image analysis, computer vision, and applications”. *CIARP*. *Lecture Notes in Computer Science*. *Springer*, Cham. 2017; 10125, https://www.scopus.com/authid/detail.uri?authorId=6506736837.
DOI: https://doi.org/10.1007/978-3-319-52277-7_28.

8. Abdollahi, N., Jafari, M., Bayat, M., Amiri, A. & Fathy, M. “An efficient parallel algorithm for computing determinant of non-square matrices”. *International Journal of Distributed and Parallel Systems (IJDPS)*. 2015; 6 (4): 1–13. DOI: https://doi.org/10.48550/arXiv.1508.01326.

9. Huang, J.-H. & Worring, M. “Query-controllable video summarization”. *ICMR '20: Proceedings of the International Conference on Multimedia Retrieval*. 2020. p. 242–250, https://www.scopus.com/authid/detail.uri?authorId=56133401600.
DOI: https://doi.org/10.1145/3372278.3390695.

10. Kansal, K., Kansal, N., Bavana, S., Vamshi, B.K. & Goyal N. "A systematic study on video summarization: Approaches, challenges, and future directions". *NarSUM '23: Proceedings of the 2nd Workshop on User-centric Narrative Summarization of Long Videos*. 2023. p 65–73, https://www.scopus.com/authid/detail.uri?authorId=57207108506.
DOI https://doi.org/10.1145/3607540.3617139.

11. Nixon, L., Apostolidis, E., Markatopoulou, F., Patras, I. & Mezaris, V. "Multimodal video annotation for retrieval and discovery of newsworthy video in a news verification scenario". In: Kompatsiaris, I., Huet, B., Mezaris, V., Gurrin, C., Cheng, W. H., Vrochidis, S. (Eds.). *Multi Media Modeling*. *Lecture Notes in Computer Science*, *Springer*. Cham. 2019; 11295: 143–155. DOI: https://doi.org/10.1007/978-3-030-05710-7_12.

12. Rochan, M., Ye, L. & Wang, Y. "Video summarization using fully convolutional sequence networks". In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.). *Computer Vision – ECCV 2018*. *Lecture Notes in Computer Science*. *Springer*. Cham. 2018; 11216: 358–367, https://www.scopus.com/authid/detail.uri?authorId=37020746300.
DOI: https://doi.org/10.1007/978-3-030-01258-8_22.

13. Bodyanskiy, Y., Kinoshenko, D., Mashtalir, S. & Mikhnova, O. "On-line video segmentation using methods of fault detection in multidimensional time sequences". *International Journal of Electronic Commerce Studies*. 2012; 3 (1): 1–20. DOI: https://doi.org/10.7903/ijecs.1010.

14. Hu, Z., Mashtalir, S. V., Tyshchenko, O. K. & Stolbovyi, M. I. "Video shots' matching via various lengths of multidimensional time sequences International". *Journal of Intelligent Systems and Applications*, 2017; 9 (11): 10–16, https://www.scopus.com/authid/detail.uri?authorId=36183980100.
DOI: https://doi.org/10.5815/ijisa.2017.11.02.

15. Mashtalir, S., Mikhnova, O. & Stolbovyi, M. "Multidimensional sequence clustering with adaptive iterative dynamic time warping". *International Journal of Computing*. 2019; 18 (1): 53–59, https://www.scopus.com/authid/detail.uri?authorId=36183980100.
DOI: https://doi.org/10.47839/ijc.18.1.1273

16. Jafari, M., et al. "Generalization of determinant kernels for non-square matrix and its application in video retrieval". *International Journal of Scientific Research in Computer Science and Engineering*. 2015; 3 (4): 1–6. ISSN: 2320-7639.

17. Agarwal, A., Amjad, M.J., Shah, D. & Shen, D. "Model Agnostic Time Series Analysis via Matrix Estimation" *ACM SIGMETRICS Performance Evaluation Review*. 2019; 1: 85–86. DOI: https://doi.org/10.1145/3376930.3376984.

18. Zhang, X. "Matrix analysis and application". New York: *Cambridge University Press*. 2017.

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

20. Kahu, S. & Rahate, R. "Image compression using singular value decomposition". *International Journal of Advancements in Research & Technology*. 2013; 2 (8): 244-248. ISSN 2278-7763.

Conflicts of Interest: the authors declare no conflict of interest

Received 05.01.2024
Received after revision 18.03.2024
Accepted 22.03.2024

DOI: https://doi.org/10.15276/aait.07.2024.5
УДК 004.93'1

Фрагментна обробка відео за нормою Кі Фана

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] Харківський Національний Університет Радіоелектроніки, проспект Науки 14. Харків, 61166, Україна
[2] Ужгородський Національний Університет, площа Народна 3. Ужгород, 88000, Україна

АНОТАЦІЯ

У даному дослідженні ми зосередили нашу увагу на способі формалізації опису блоків відеокадрів в контексті рішення задачі відео фрагментації. Оскільки відеодані можуть бути різного розміру, розбиття на блоки кожного кадру дає можливість формально описати кадр як квадратну матрицю. Блок кадру є матриця довільної розмірності. Можливість пропустити крок приведення такої матриці до квадратної, або векторизація за допомогою деякого дескриптора дозволяє знизити обчислювальні витрати, визволяючи обчислювальні ресурси необхідних для цього перетворення. В цьому дослідженні ми використовуємо норму Кі Фана в якості дескриптора блоку кадру. Норма Кі Фана побудована на основі сингулярних чисел матриці. Сингулярний розклад не має обмежень ні до розмірності, ні до характеру елементів вихідної матриці. Ми провели порівняльний аналіз ефективності отриманого дескриптора для відео даних різного розміру і з різним відношенням сторін який показав, що зміна дескриптора кожного блоку не залежить ні від розміру відео, ні від відношення сторін. Зміни дескрипторів кожного блоку від кадру до кадру є ідентичними для відеоданих різного розміру. Це означає, що в результаті такого фрагментного перетворення отримується квадратна матриця фіксованого розміру незалежно від розміру вихідного відео. Це дозволяє уніфікувати подальшу обробку відео, що може бути корисним для задачі інформаційного пошуку в великих базах відео даних за умов надання запиту «за зразком». В цьому випадку ми в режимі оффлайн можемо проаналізувати існуючу базу і співставити кожному відео фіксовану квадратну матрицю дескрипторів, що дозволить значно зменшити час та кількість ресурсів при співставленні із запитом. Також даний підхід може бути ефективно використаний для аналізу відео даних з метою детектування руху і відстеження зміни сцени.

Ключові слова: фрагментація відеопотоку; Кі Фан норма; декомпозиція сингулярного значення

ABOUT THE AUTHORS

Sergii Volodymyrovych Mashtalir - Doctor of Engineering Science, Professor of 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 Харківського національного університету радіоелектроніки, проспект Науки 14. Харків, 61166, Україна

Dmytro P. Lendel - PhD student of 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 в Ужгородському національному університеті (УжНУ), площа Народна, 3. Ужгород, 88000, Україна