

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Кваліфікаційна наукова  
праця на правах рукопису

**СТОЛБОВИЙ МИХАЙЛО ІВАНОВИЧ**

УДК 004.932.2:004.93'14

**ДИСЕРТАЦІЯ**

**ТЕХНОЛОГІЯ ВІДЕОРЕФЕРУВАННЯ НА ОСНОВІ КЛАСТЕРИЗАЦІЇ  
ДЛЯ ІНФОРМАЦІЙНОГО ПОШУКУ**

05.13.06 – інформаційні технології

технічні науки

Подається на здобуття наукового ступеня кандидата наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

М.І. Столбовий  
(підпис, ініціали та прізвище здобувача)

Науковий керівник

Машталір Сергій Володимирович, доктор технічних наук, доцент

Харків – 2019

## АНОТАЦІЯ

*Столбовий М.І.* Технологія відеореферування на основі кластеризації для інформаційного пошуку. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Харківський національний університет радіоелектроніки, Міністерство освіти і науки України, Харків, 2019.

Величезні обсяги відео, що мають різні модальності, створюють перешкоди для багатьох інформаційних технологій. Існують численні семантичні поняття, які дуже складно або навіть неможливо вилучити автоматично. Відео анотація – це механізм для створення короткого викладу всього відео і це значна частина таких програм, як індексація відео, перегляд і семантичний пошук. Визнаючи корисність анотації відео в більш широкій області управління мультимедійним контентом, слід підкреслити, що акцент робиться на часовій сегментації та кластеризації відеоданих. Одна з центральних проблем, пов'язаних з пошуком на основі контенту з запитом «за зразком», стосується послідовностей зображень, що розпізнають та ранжують витягнуті зйомки відповідно до ступеня актуальності запиту користувача. Відеосегмент (shot) ідентифікується як важливе, цікаве або необхідне, якщо це подія з певною семантичною міткою, що привертає увагу глядача. Крім того, послідовності є важливими, якщо вони можуть бути пов'язані з деякими шаблонами або мати специфічні часові та просторові атрибути. Ефективна абстракція відео контенту може бути виражена як послідовність стаціонарних репрезентативних зображень, що називаються ключовими кадрами або рухомими зображеннями (частина відео сегментів, переміщення раскадровки або сумаризація). Якщо статична відео абстракція, яка базується на ключових кадрах, досить добре відома та досліджена, то динамічна сумаризація залишається невирішеною проблемою. У той же час, динамічна сумаризація має низку переваг перед набором ключових кадрів, серед яких по-

тенційно можуть бути вказані як виразність, так і інформативність рефератів, і, як наслідок, можливість знайти більш релевантну інформацію за допомогою відео-пошукової системи є більш очевидною.

У дисертації вирішена актуальна науково-практична задача сегментації відеоданих і моделей кластеризації та методів динамічної сумаризації відео з метою створення часових візуально-пошукових технологій.

Для досягнення необхідного відклику та точності пошуку інформації та сумаризації відео, по-перше, структуруються відео послідовності, які відповідають часовій сегментації багатовимірних часових рядів, пов'язаних з відео. Сегментація виконується або в обраному просторі ознак, або в просторі зображень, точніше, на вибраному наборі змінюваних у часі субкадрів або будь-яких його зв'язках. В результаті, на основі поточного кадру та певної передісторії обробки, хронологічно пов'язані кадри є зйомками або знімками, які є основними одиницями пошуку подій, сцен, сценаріїв для багатовимірної індексації та пошуку, а також для сумаризації відеопотоку. Головною перешкодою у часовій сегментації є ідентифікація згладжених градіальних змін між кадрами, коли доступні повільні панорами зникають, розчиняються, витираються, з'являються в кадрі тощо. Остаточна достовірність як сумаризації відео, так і пошуку відеоінформації зумовлюється точністю і семантичною надійністю часової сегментації відеопотоку. На підставі тенденцій еволюції абстракції відеопотоків з метою підвищення змісту актуальності пошуку інформації встановлено багатообіцяючі перспективи динамічної сумаризації відео. Встановлено, що відомі способи сегментації, кластеризації та сумаризації відеопослідовностей не повністю відповідають сучасним вимогам і вимагають суттєвого поліпшення.

Швидко виявлення змін властивостей векторних і матричних нестационарних збурених сигналів на основі ансамблю адаптивних моделей з власними алгоритмами ідентифікації з різними глибинами пам'яті створює передумови для сегментації відеопослідовностей з урахуванням структурних і полісемічних властивостей окремих частин відеопотоків.

Впроваджені процедури нечіткої кластеризації дозволяють вирішити проблему виявлення розбіжностей у багатовимірних часових рядах як для згладжених, так і для різких змін відеоконтенту, що забезпечує високу актуальність часового аналізу відеопотоків. Виявлення змін багатовимірних послідовностей різної довжини на основі аналізу ітеративної динамічної часової деформації, пов'язаної з адаптивним перетворенням часових рядів, дозволяє отримати модифікації алгоритмів кластеризації для векторних та матричних часових послідовностей з метою покращення сумаризації.

Запропонований метод кластеризації для багатовимірних часових рядів різної довжини у випадку невідомої кількості класів і можливості їх взаємного перетину на основі гібридизації ієрархічних агломераційних і нечітких підходів на основі центроїдів з використанням метрики Левенштейна дає можливість пошуку з складними запитами.

Запропоновані підходи до налаштування параметрів адаптивної матричної моделі та виявлення зміни матричної послідовності забезпечують зменшення розриву між формальною та семантичною актуальністю відео через те, що фрагменти, що є результатом довільної просторової згортки, можуть характеризувати околиці точок, що представляють візуальну увагу.

Введено методи кластеризації відеопослідовностей з наступним угрупованням зведених часових рядів на основі гармонійно-середньої нечіткої кластеризації мають просту обчислювальну модель і підвищену стійкість до локальних деформацій часу. Для виявлення просторових і часових подій у відеопотоках з метою семантичного узагальнення відео, кластеризація часових послідовностей різної довжини з використанням представників і динамічного перетворення часової шкали здаються перспективними з практичної точки зору.

Оригінальність наукової тематики полягає у формулюванні та вирішенні науково-технічної задачі сегментації та кластеризації відео для сумаризації з метою пошуку інформації:

– вперше запропоновано метод адаптивної ітеративної динамічної часової

деформації для вирішення задач сегментації багатовимірних часових рядів з метою відеореферування, що дозволяє в реальному часі обробляти відеопослідовності за умов, коли вони можуть непередбаченим чином змінювати властивості та мати різну довжину на відміну від існуючих методів;

– вперше запропоновано гібридний метод кластеризації відеорядів різної довжини, що заснований на використанні метрики Левенштейна, що відрізняє його від відомих, та дозволяє вирішувати задачу узагальнення представлення динамічної візуальної інформації по семантичній схожості послідовностей відеокадрів за умов перетинних класів та апріорі невідомої їх кількості;

– отримали подальший розвиток адаптивні моделі потоків відеоданих, що відрізняються від відомих можливістю аналізу областей візуальної уваги та дозволяють визначати змінення часових властивостей матричних послідовностей – фрагментів відеокадрів;

– удосконалено метод ітеративної динамічної часової деформації темпоральних матричних рядів для нечіткої, на відміну від відомих, кластеризації, що дозволяє за рахунок редукції початкових рядів підвищувати швидкодію, а також стійкість до завад та збурень завдяки використанню гармонійних середніх при нечіткій сегментації відеопослідовностей.

Методи сегментації і кластеризації відеопослідовностей були реалізовані програмно і застосовувалися для керування великими колекціями неструктурованої динамічної інформації та контекстного пошуку з запитом "ad exemplum", що підтверджується відповідними актами.

Дисертаційна робота виконана на кафедрі інформатики Харківського національного університету радіоелектроніки в рамках держбюджетних науково-дослідних робіт: «Розробка гібридних систем та методів обчислювального інтелекту для обробки потоків нечіткої інформації за умов нестаціонарності та невизначеності», «Глибинні гібридні системи обчислювального інтелекту для аналізу потоків даних та їх швидке навчання», які виконувалися та виконуються відповідно до наказів Міністерства освіти і науки України за результатами конкурсного відбору науково-дослідних робіт.

Основні положення дисертації відображені у 13 наукових працях здобувача: 3 статті у наукових фахових виданнях України з технічних наук (з них 1 індексується в Web of Science та 1 в Scopus); 3 статті в закордонних журналах (з них 2 індексуються в Scopus); 7 публікацій в матеріалах міжнародних конференцій (з них 5 публікацій індексується в Scopus).

#### Перелік публікацій автора

1. Mashtalir S., Mashtalir V., Stolbovyi M. Video shot boundary detection via sequential clustering. International Journal “Information Theories and Applications”, 2017. Vol. 24, No. 1. – P. 50-59.

2. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Video shots' matching via various length of multidimensional time sequences. International Journal of Intelligent Systems and Applications. 2017. Vol. 9, No. 11. – P. 10-16.

3. Hu Z., Mashtalir, S.V., Tyshchenko, O.K., Stolbovyi, M.I. Clustering matrix sequences based on the iterative dynamic time deformation procedure. International Journal of Intelligent Systems and Applications. 2018. Vol. 10, No. 7. – P. 66-73.

4. Машталир С.В., Столбовой М.И. Адаптивные матричные модели в задаче контроля потоков видео. Радиоелектроніка, інформатика, управління. 2018. №4(47). – С. 188-194.

5. Богучарский С.И., Машталир С.В., Столбовой М.И. Быстрое обнаружение изменения свойств многомерных временных рядов на основе идентификационного подхода к ансамблю моделей. Системи обробки інформації. Харків : ХУПС. 2018. Вып. 3 (154). – С. 74-78.

6. Машталир С.В., Столбовой М.И., Яковлев С.В. Кластеризация последовательностей видеоданных на основе гармонических k-средних. Кибернетика и системный анализ. 2019. Том 55, №2. – С. 36-43.

7. Машталир С.В., Столбовой М.И. Анализ методов пространственно-временной сегментации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнарод-

дної наукової конференції. (Залізний Порт, Україна, 22–26 травня 2017). Херсон, Видавництво ПП Вишемирський В.С. 2017. С. 287-289.

8. Kobylin O., Mashtalir S., Stolbovyi M. Video clustering via multidimensional time-series analysis. Proceedings of the 9th International Conference on Information Management and Engineering ICIME 2017 (Barcelona, Spain, 9–11 October, 2017). New York : ACM International Conference Proceeding Series. 2017. P. 60-63.

9. Машталир С.В., Столбовой М.И., Яременко М.М. Использование матричных моделей анализа временных рядов в задаче кластеризации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнародної наукової конференції (Залізний Порт, Україна, 21–27 травня 2018). Херсон, Видавництво ПП Вишемирський В.С. 2018. С.83-85.

10. Mashtalir S., Mashtalir V., Stolbovyi M. Representative based clustering of long multivariate sequences with different lengths. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 545-548.

11. Mashtalir S., Mikhnova O., Stolbovyi M. Sequence matching for content-based video retrieval. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 549-553.

12. Kinoshenko D., Mashtalir S., Shlyakhov V., Stolbovyi M. Video shots retrieval with use of pivot points. Advances in Computer Science for Engineering and Education. The First International Conference on Computer Science, Engineering and Education Applications ICCSEEA 2018 (Kyiv, Ukraine, 18–20 January, 2018) / Hu Z., Petoukhov S., Dychka I., He M. (eds). Advances in Intelligent Systems and Computing. Cham : Springer. 2019. Vol. 754. P. 102-111.

13. Kinoshenko D., Kobylin O., Mashtalir S., Stolbovyi M. Metric video retrieval speedup by irrelevant data elimination. Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018), 1104100; 8 p.

## ABSTRACT

*Stolbovyi M.I.* Clustering-based video summarization technology for information retrieval. – Manuscript.

Thesis for the candidate degree of technical sciences in the speciality 05.13.06 – information technologies. – Kharkiv National University of Radio Electronics, Ministry Education and Science of Ukraine, Kharkiv, 2019.

Enormous volumes of having different modalities video produce a barrier to many information technology activities. There exist numberless semantic concepts that are very difficult or even impossible to extract automatically. Video abstraction is a mechanism for producing a short summary of a whole video and it is a significant part of such applications as video indexing, browsing and semantic retrieval. Recognizing the usefulness of video abstraction within the broader field of multimedia content management, it should be emphasized that the focus is on video temporal segmentation and clustering. One of central problems regarding the content based retrieval with queries ad exemplum concerns with image sequences understanding and ranking extracted skims according to the relevance degree to the user query. Video excerpt (shot) is identified as important, interesting, or exciting if it is an event with a specific semantic label, or captures the attention of the viewer. In addition, sequences are essential if they may be associated with some patterns, or have specific temporal and spatial attributes.

Effective abstraction of video contents can be expressed as a sequence of stationary representative images called keyframes or moving images (video skims, moving storyboard or summary sequence).

If static video abstraction which based on keyframes is sufficiently well explored, then dynamic summarization remains an unresolved problem. At the same time, dynamic summarization has a number of advantages over a keyframe set, among which potentially enhance both the expressiveness and information of the abstracts should be indicated and, as a consequence, possibilities to find more relevant



information by video search engine become more obvious.

In the thesis the actual scientific and practical problem of video data segmentation and clustering models and methods for dynamic video summarization has been solved with the aim of temporal visual information retrieval technologies development.

To achieve the required Recall and Precision of information retrieval and video summarizing, first, video sequences are structured that corresponds to the temporal segmentation of multidimensional time series associated with video. Segmentation is executed either in the chosen feature space or in the image space, more precisely, on a selected set of time-varying subframes or any its convolutions. As a result, based on the current frame and a certain prehistory processing, chronologically connected frames are skims or shots which are the basic units for searching events, scenes, scenarios for multidimensional indexing and retrieval, as well as for video stream summarizing.

The main obstacle in temporal segmentation lies in identification of smoothed gradual changes between shots when slow panning, fade, dissolve, wipe, frame appearing in the frame, etc. are available. The final validity of both video summarizing and video information retrieval is predetermined by the accuracy and semantic reliability of time video segmentation.

Based on trends of video streams abstraction evolution in order to increase the content relevance of information retrieval, the promising perspectives of dynamic video summarization are established. It is found that the known methods of segmentation, clustering and summarization of video sequences do not fully meet modern requirements and require substantial improvement.

Fast change detection of vector and matrix nonstationary noise signal properties based on an ensemble of adaptive models with their own identification algorithms with different memory depths creates prerequisites for the segmentation of video sequences taking into consideration the structural and polysemy properties of video streams separate parts.

Introduced on-line fuzzy clustering procedures allow solving the problem of detecting discrepancies in multidimensional time series for both smoothed and abrupt changes in video content, which ensures high relevance of video streams temporal analysis.

Change detection of various lengths multidimensional sequences on the basis of Iterative Dynamic Time Warping analysis, which is associated with adaptive reduction of time series, makes it possible to obtain modifications of clustering algorithms for vector and matrix time sequences with aim of summarizing pertinences growth.

The proposed clustering method for multidimensional time series of different lengths under an unknown classes number and the possibilities of their mutual intersection based on hybridization of hierarchical agglomerative and fuzzy centroid-based approaches using Levenshtein metric provides the opportunities of a valid search with complex summarizing queries.

The proposed approaches to adaptive matrix model parameters alignment and matrix sequence change detection provide reducing the gap between formal and semantic relevance due to the fact that fragments resulting from arbitrary spatial convolution can characterize the neighborhoods of points that represent visual attention.

The introduced technique for video sequences clustering with subsequent grouping of reducible time series based on harmonic-average fuzzy clustering has a simple computational model and increased stability to local time deformations.

For the detection of spatial and temporal events in video streams with the aim of semantic summarization of video, the clustering of temporal sequences with different lengths using representatives and the dynamic transformation of the time scale seems promising from a practical point of view.

The scientific thesis originality lays in the formulation and solution of the scientific and technical problem of video temporal segmentation and clustering for summarizing with the aim of information retrieval:

- for the first time, the method of adaptive iterative dynamic time warping has

been proposed for multidimensional time series segmentation with the video summarizing goals what allows on-line video sequences processing when they can unexpectedly change properties and have different lengths;

- for the first time, a hybrid method of video clustering with different lengths of skims on the base of Levenstein metric allows to obtain video abstraction by the semantic similarity of sequences when classes intersect and number of clusters is a priori unknown;

- adaptive models of video streams are developed to detect changes of matrix sequences temporal properties, to be exact subframes surrounding spatial visual attention regions;

- the method of iterative dynamic time warping of fuzzy temporal matrix series clustering is improved, which, by reducing the initial series, allows to speed up computational model as well as to increase noise immunity due to harmonic means with video sequences fuzzy segmentation.

Video sequences segmentation and clustering methods were implemented software and have been applied for control large collections of unstructured dynamic information and context-sensitive search with requests ‘ad exemplum’ that is confirmed by relevant acts.

The thesis was performed at the Department of Informatics of the Kharkiv National University of Radio Electronics in the framework of state budget research projects: "Development of hybrid systems and methods of computing intelligence for processing fuzzy information streams under unstationarity and uncertainty", "Deep hybrid systems of computational intelligence for data streams analysis and their fast learning", which were executed in accordance with the orders of the Ministry of Education and Science of Ukraine in appliance with results of the research projects competitive selection.

The major thesis outcomes are represented in 13 research papers of the applicant: 3 articles in the scientific specialized editions of Ukraine on technical sciences (1 item is indexed in Web of Science database and 1 item in Scopus database); 3 pa-

pers in foreign journals (2 items are indexed in Scopus); 7 papers in the proceedings of international conferences (5 of them are indexed in Scopus).

#### List of the applicant publications

1. Mashtalir S., Mashtalir V., Stolbovyi M. Video shot boundary detection via sequential clustering. International Journal “Information Theories and Applications”, 2017. Vol. 24, No. 1. – P. 50-59.

2. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Video shots' matching via various length of multidimensional time sequences. International Journal of Intelligent Systems and Applications. 2017. Vol. 9, No. 11. – P. 10-16.

3. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Clustering matrix sequences based on the iterative dynamic time deformation procedure. International Journal of Intelligent Systems and Applications. 2018. Vol. 10, No. 7. – P. 66-73.

4. Машталир С.В., Столбовой М.И. Адаптивные матричные модели в задаче контроля потоков видео. *Радіоелектроніка, інформатика, управління*. 2018. №4(47). – С. 188-194.

5. Богучарский С.И., Машталир С.В., Столбовой М.И. Быстрое обнаружение изменения свойств многомерных временных рядов на основе идентификационного подхода к ансамблю моделей. *Системи обробки інформації*. Харків : ХУПС. 2018. Вып. 3 (154). – С. 74-78.

6. Машталир С.В., Столбовой М.И., Яковлев С.В. Кластеризация последовательностей видеоданных на основе гармонических k-средних. *Кибернетика и системный анализ*. 2019. Том 55, №2. – С. 36-43.

7. Машталир С.В., Столбовой М.И. Анализ методов пространственно-временной сегментации видеопоследовательностей. *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнародної наукової конференції*. (Залізний Порт, Україна, 22–26 травня 2017). Херсон, Видавництво ПП Вишемирський В.С. 2017. С. 287-289.

8. Kobylin O., Mashtalir S., Stolbovyi M. Video clustering via multidimensional time-series analysis. Proceedings of the 9th International Conference on Information Management and Engineering ICIME 2017 (Barcelona, Spain, 9–11 October, 2017). New York : ACM International Conference Proceeding Series. 2017. P. 60-63.

9. Машталир С.В., Столбовой М.И., Яременко М.М. Использование матричных моделей анализа временных рядов в задаче кластеризации видеопоследовательностей. Интеллектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнародної наукової конференції (Залізний Порт, Україна, 21–27 травня 2018). Херсон, Видавництво ПП Вишемирський В.С. 2018. С.83-85.

10. Mashtalir S., Mashtalir V., Stolbovyi M. Representative based clustering of long multivariate sequences with different lengths. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 545-548.

11. Mashtalir S., Mikhnova O., Stolbovyi M. Sequence matching for content-based video retrieval. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 549-553.

12. Kinoshenko D., Mashtalir S., Shlyakhov V., Stolbovyi M. Video shots retrieval with use of pivot points. Advances in Computer Science for Engineering and Education. The First International Conference on Computer Science, Engineering and Education Applications ICCSEEA 2018 (Kyiv, Ukraine, 18–20 January, 2018) / Hu Z., Petoukhov S., Dychka I., He M. (eds). Advances in Intelligent Systems and Computing. Cham : Springer. 2019. Vol. 754. P. 102-111.

13. Kinoshenko D., Kobylin O., Mashtalir S., Stolbovyi M. Metric video retrieval speedup by irrelevant data elimination. Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018), 110410O; 8 p.

## ЗМІСТ

ВСТУП.....	16
РОЗДІЛ 1 АНАЛІЗ СТАНУ І ТЕНДЕНЦІЙ РОЗВИТКУ МЕТОДІВ СЕГМЕНТАЦІЇ І СУММАРІЗАЦІЇ ВІДЕО. ПОСТАНОВКА ОСНОВНИХ ЗАВДАНЬ ДОСЛІДЖЕНЬ.....	22
1.1 Аналіз трендів розвитку методів відеореферування.....	23
1.2 Порівняльний аналіз методів і моделей часової сегментації Відео потоків.....	29
1.3 Аналіз методів кластеризації стосовно темпоральних сегментів відеоданих.....	35
1.4 Постановка основних задач досліджень.....	40
Висновки по першому розділу.....	44
РОЗДІЛ 2 МОДЕЛІ ТЕМПОРАЛЬНОЇ СЕГМЕНТАЦІЇ ВІДЕОПОТОКІВ.....	45
2.1 Виявлення зміни властивостей потоків даних на основі ідентифікаційного підходу до ансамблю моделей.....	46
2.2 Виявлення змін в багатовимірних потоках відео на основі послідовної кластеризації.....	52
2.3 Виявлення змін властивостей багатовимірних послідовностей різної довжини.....	60
Висновки по другому розділу.....	68
РОЗДІЛ 3 МЕТОДИ ФРАГМЕНТНОЇ СЕГМЕНТАЦІЇ ВІДЕО В ЗАДАЧАХ ІНФОРМАЦІЙНОГО ПОШУКУ .....	70
3.1 Адаптивна ітеративна динамічна часова деформація при сегментації багатовимірних послідовностей.....	71
3.2 Гібридний підхід до кластеризації відеорядів різної довжини.....	80
3.3 Адаптивні матричні моделі в задаче контролю потоків відео.....	88
Висновки по третьому розділу.....	95

	15
РОЗДІЛ 4 МЕТОДИ ПОСЛІДОВНОЇ КЛАСТЕРИЗАЦІЇ ВІДЕО ДЛЯ РЕФЕРУВАННЯ.....	97
4.1 Кластеризація матричних послідовностей на основі ітеративної динамічної часової деформації.....	98
4.2 Кластеризація послідовностей відеоданих на основі гармонійних $k$ -середніх.....	104
4.3 Кластеризація відео на основі представників.....	111
Висновки по четвертому розділу.....	117
РОЗДІЛ 5 ЕКСПЕРИМЕНТАЛЬНИЙ АНАЛІЗ МЕТОДІВ СЕГМЕНТАЦІЇ ТА КЛАСТЕРИЗАЦІЇ ВІДЕО.....	118
5.1 Специфіка технології пошуку динамічної візуальної інформації на базі відеореферування.....	118
5.2 Експериментальний аналіз ознакових просторів при просторово- темпоральній сегментації відео.....	124
5.3 Дослідження особливостей кластеризації відеорядів для відеореферування.....	133
Висновки по п'ятому розділу.....	137
ВИСНОВКИ .....	139
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	141
ДОДАТОК А. Акти впровадження результатів дисертаційної роботи.....	160
ДОДАТОК Б. Список публікацій здобувача.....	164

## ВСТУП

**Актуальність теми.** Тенденції розвитку технологій інформаційного пошуку в системах гіпермедіа, коли підлягають урахуванню відео, графіка, звук, текст, структури нелінійного середовища їх сприйняття, вимагають забезпечення розумного компромісу між інформаційною потребою і властивостями наборів даних з безперервно зростаючими обсягами. Основні складності пошуку інформації, як правило, із запитами ‘ad exemplum’ («за зразком») пов’язані з відеопотоками в силу їх принципової неструктурованості і необхідності on-line обробки. Внесення суттєвих перетворень змісту і (або) форми вихідних інформаційних продуктів у вигляді анотування і реферування зі збереженням релевантності та пертінентності пошуку являє собою інструментарій узагальнення-сумаризації відеоданих. Коротке узагальнене представлення відео є безумовним базисом оптимізації реєстрації і зберігання, вилучення та віртуалізації неструктурованої візуальної інформації в колекціях, що перманентно поповнюються.

Для досягнення необхідних показників точності і повноти інформаційного пошуку і відеореферування спочатку виконується структуризація відеопослідовностей, яка відповідає сегментації багатовимірних часових рядів, асоційованих з відео. Часова сегментація виконується або в обраному просторі ознак, або в просторі зображень, точніше кажучи, на множині фрагментів відеокадрів (матричних сигналів), які змінюються в часі та відібрані на основі їх семантичних характеристик. В результаті на основі аналізу поточного відеокадру і деякої передісторії виділяються зв’язні в хронологічному порядку відеокадри – сегменти, які і є базовими одиницями для пошуку подій, сюжетів, сценаріїв для багатоаспектного індексування в задачах пошуку, а також для синтезу узагальнюючих представлень відеопотоків. Основне навантаження при побудові темпорального розбиття лягає на виявлення плавних змін між сегментами у випадках повільного панорамування, ефектів монтажу: затемнення, розчинення, розмивання, накладення, появи кадру в кадрі і т.д. Фінальна валідність і відеореферу-



вання, і інформаційного пошуку зумовлюється саме точністю і семантичною надійністю часової сегментації відео, що призводить до необхідності розвитку методів і моделей темпоральної сегментації.

Адекватність, достовірність, об'єктивність пошуку множин послідовних відеокадрів, нерозрізнених з точки заданого критерію, зумовлюють формальну, а в підсумку і змістовну релевантність кластеризації розбиття відеопотоків. В цей час не існує достатньо ефективних методів, що забезпечують рівнопотужну продуктивність для довільних типів відео навіть при об'єктно-орієнтованій сумаризації. Досить високого рівня досягло лише статичне відеореферування, яке ґрунтується на поданні сегментів в вигляді ключових кадрів. Перспективним напрямком розвитку представляється об'єктивне динамічне відеореферування в просторах відеосегментів та їх стратифіковане узагальнення на основі кластеризації наборів багатовимірних послідовностей, в тому числі і матричних – фрагментів відеокадрів. Поряд з вибором функціоналів подібності основну складність представляє порівняння послідовностей різної довжини.

Аналіз стану моделей і методів сегментації і кластеризації відеопослідовностей, узагальнення множин їх структурних елементів дає можливість стверджувати, що один з перспективних напрямків підвищення ефективності інформаційного пошуку полягає в розробці інструментів динамічного відеореферування. Основний вплив на розвиток моделей і методів обробки і розпізнавання статичної та динамічної візуальної інформації надали українські та зарубіжні вчені С.Г. Антощук, О.М. Березький, Р.А. Воробель, Г.М. Жолткевич, В.Н Крилов, Д.Д. Пелешко, Є.П. Путятін, Т.Є. Рак, М.І. Шлезінгер, Н. Burkhardt, Р.А. Geetha, V. Hlavac, E. Keogh, M. Sonka, R.C. Veltkamp, W. Wang, P. Zezula і ін. Незважаючи на численні дослідження в даному напрямку, залишається актуальною (з практичної і теоретичної точок зору) розробка методів виявлення змін властивостей багатовимірних часових рядів на основі ідентифікаційного підходу до ансамблю моделей, послідовної кластеризації відеопотоків, дослідження властивостей багатовимірних послідовностей різної довжини,

матричних (фрагментних) темпоральних моделей.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконана на кафедрі інформатики Харківського національного університету радіоелектроніки в рамках держбюджетних науково-дослідних робіт: «Розробка гібридних систем та методів обчислювального інтелекту для обробки потоків нечіткої інформації за умов нестационарності та невизначеності» (№ ДР0116U002539, виконавець), «Глибинні гібридні системи обчислювального інтелекту для аналізу потоків даних та їх швидке навчання» (№ ДР0119U001403, виконавець), які виконувалися та виконуються відповідно до наказів Міністерства освіти і науки України за результатами конкурсного відбору науково-дослідних робіт. В рамках зазначених робіт здобувачем створено моделі темпоральної сегментації відеопотоків, методи фрагментної сегментації відео в задачах інформаційного пошуку, методи послідовної кластеризації відео для реферування відеоданих.

**Мета і завдання досліджень.** Метою дисертаційної роботи є розробка інструментарію інтелектуальних інформаційних технологій реферування відео на основі кластеризації. Для досягнення поставленої мети потрібно було вирішення таких завдань:

– аналіз тенденцій розвитку методів обробки відео в аспекті семантичної структуризації у вигляді часових розбиттів та покриттів, що забезпечують динамічне відеореферування на основі кластеризації та багатоаспектну індексацію в технологіях інформаційного пошуку;

– розробка моделей та методів сегментації і кластеризації відеоданих за умов необхідності порівнянь вхідних даних різної довжини;

– розробка нечітких методів кластеризації відеоданих, які дозволяють визначати зміни властивостей фрагментів відеокадрів;

– розробка методів послідовної кластеризації відео для реферування відеоданих на множинах сегментів;

– дослідження специфіки використання розроблених моделей і методів кластеризації відеопослідовностей, створення і впровадження дослідницьких та спеціалізованих програмних засобів.

*Об'єкт дослідження* – процеси обробки і аналізу відеопотоків в інформаційних технологіях пошуку візуальної інформації.

*Предмет дослідження* – методи часової сегментації і кластеризації відеопослідовностей для відеореферування і інформаційного пошуку з запитом «за зразком».

*Методи дослідження* – при розробці та дослідженні методів і моделей нечіткої сегментації і кластеризації відеопотоків використані основні положення математичного апарату розпізнавання образів і обробки зображень, методи обчислювального інтелекту, аналізу часових рядів, а також елементи математичної статистики при проведенні і аналізі результатів експериментальних досліджень.

**Наукова новизна отриманих результатів.** Наукова новизна дисертації полягає в постановці і вирішенні науково-технічної задачі часової сегментації і кластеризації відеоданих для відеореферування та інформаційного пошуку:

– вперше запропоновано метод адаптивної ітеративної динамічної часової деформації для вирішення задач сегментації багатовимірних часових рядів з метою відеореферування, що дозволяє в реальному часі обробляти відеопослідовності за умов, коли вони можуть непередбаченим чином змінювати властивості та мати різну довжину на відміну від існуючих методів;

– вперше запропоновано гібридний метод кластеризації відеорядів різної довжини, що заснований на використанні метрики Левенштейна, що відрізняє його від відомих, та дозволяє вирішувати задачу узагальнення представлення динамічної візуальної інформації по семантичній схожості послідовностей відеокадрів за умов перетинних класів та апріорі невідомої їх кількості;

– отримали подальший розвиток адаптивні моделі потоків відеоданих, що відрізняються від відомих можливістю аналізу областей візуальної уваги та до-

звояють визначати змінення часових властивостей матричних послідовностей – фрагментів відеокадрів;

– удосконалено метод ітеративної динамічної часової деформації темпоральних матричних рядів для нечіткої, на відміну від відомих, кластеризації, що дозволяє за рахунок редукції початкових рядів підвищувати швидкодію, а також стійкість до завад та збурень завдяки використанню гармонічних середніх при нечіткій сегментації відеопослідовностей.

**Практичне значення отриманих результатів.** Запропоновані моделі та методи дозволяють виконувати кластеризацію відеоданих за їх змістом, що спрощує процедуру відеореферування, яка є одним з напрямів підвищення загальної швидкодії пошуку. Результати дисертаційних досліджень використані в ДП «УкрНТЦ «Енергосталь», м. Харків при проектуванні систем екологічного моніторингу атмосферного повітря, де необхідно обробляти значні обсяги вхідних темпоральних даних на протязі тривалого часу (акт від 10.01.2019 р.). Теоретичні і практичні результати роботи знайшли застосування в освітньому процесі Харківського національного університету радіоелектроніки при підготовці магістрів за спеціальністю «Інформатика» (акт від 16.01.2019 р.) та при виконанні науково-дослідної роботи в Харківському національному університеті радіоелектроніки (акт від 29.01.2019 р.).

**Особистий внесок здобувача.** Всі положення, що виносяться на захист, отримані здобувачем особисто. У роботах, написаних у співавторстві, дисертанту належить: в [107] – метод пошуку границь сегментів відео; [77] – метод кластеризації послідовностей різної довжини за допомогою динамічної часової деформації; [76] – метод кластеризації послідовностей різної довжини за допомогою ітеративної динамічної часової деформації; [15] – метод кластеризації послідовностей різної довжини за допомогою адаптивної динамічної часової деформації; [6] – метод побудови ансамблю моделей для швидкого виявлення змін в відеоданих; [17] – модифікація методу гармонічних k-середніх для аналізу відео послідовностей; [16] – модифікація методу порівняння методів сегмен-

тації відеопослідовностей; [94] – модифікація методу кластеризації відеопослідовностей; [18] – матричний метод кластеризації відеопослідовностей; [106] – матрична модифікація методу кластеризації відео послідовностей різної довжини; [109] – модифікація методу порівняння результатів кластеризації відеопослідовностей; [92] – метод елімінації відеоданих; [94] – вдосконалений метод елімінації відеоданих на загальний випадок.

**Апробація результатів дисертації.** Основні результати роботи доповідалися, обговорювалися та були схвалені в рамках таких міжнародних наукових конференцій: 9th International Conference on Information Management and Engineering ICIME 2017 (Barcelona, Spain, 9–11 October, 2017); 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) (Lviv, Ukraine, August 21-25, 2018); міжнародних наукових конференціях «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (ISDMCI' 2017, 2018) (Залізний Порт, Україна, 22–26 травня 2017, Залізний Порт, Україна, 21–27 травня 2018); The First International Conference on Computer Science, Engineering and Education Applications ICCSEEA 2018 (Kyiv, Ukraine, 18–20 January, 2018), The 11th International Conference on Machine Vision (ICMV 2018) (November 1-3, 2018 Munich, Germany).

**Публікації.** Основні положення дисертації відображені у 13 наукових працях здобувача: 3 статті у наукових фахових виданнях України з технічних наук (з них 1 індексується в Web of Science та 1 в Scopus); 3 статті в закордонних журналах (з них 2 індексуються в Scopus); 7 публікацій в матеріалах міжнародних конференцій (з них 5 публікацій індексується в Scopus).

# 1 АНАЛІЗ СТАНУ І ТЕНДЕНЦІЙ РОЗВИТКУ МЕТОДІВ СЕГМЕНТАЦІЇ І СУМАРИЗАЦІЇ ВІДЕО. ПОСТАНОВКА ОСНОВНИХ ЗАВДАНЬ ДОСЛІДЖЕНЬ

Безперервне, нічим не обмежуване, зростання накопичення візуальної інформації висуває підвищені вимоги до перспективних інформаційних гіпермедіа технологій, тобто комплексу методів і засобів, спрямованих на підвищення ефективності зберігання, обробки, передачі, відображення, а в підсумку і, принаймі, предметно-орієнтованої інтерпретації мультимодальної інформації. Основний акцент лягає на аналітико-синтетичну переробку початково неструктурованих відеопотоків в силу принципового розриву між інформаційною надмірністю та інформаційною потребою, з одного боку. З іншого боку, використовувані критерії формальної відповідності, як правило, не тільки не забезпечують достатнього рівня семантичного, синтаксичного і структурного аналізу, але і входять у прямий конфлікт з релевантністю і пертинентністю.

Одним з перспективних шляхів розвитку інформаційних технологій в цілому є розробка прогресивних методів пошуку, що забезпечують бажані показники повноти, точності і оперативності, а також враховують, взагалі кажучи, парадигматичні відношення в самих різних формах. В цьому плані ключову роль відіграють методи і моделі узагальнення (абстрагування, сумаризації, анотування, реферування, резюмування) відео, які забезпечують використання складноорганізованих семантичних запитів з можливостями багатоаспектного деривантного індексування. Для просування в цьому напрямку необхідна диверсифікація досліджень в плані структурування та стратифікованого упорядкування колекцій динамічної візуальної інформації з цілями отримання моделей, здатних впоратися з великою різноманітністю відеоконтенту без участі людини.

Адекватним інструментарієм вирішення завдань відеореферування є ме-

тоди сегментації і кластеризації багатовимірних часових рядів, асоційованих з відеорядами, що відображають різноманітні динамічні процеси.

Розділ присвячений аналізу сучасного стану та трендів розвитку підходів до реферування відеопотоків, сегментації та кластеризації багатовимірних часових рядів.

### 1.1 Аналіз трендів розвитку методів відеореферування

Розвиток інформаційних відеотехнологій найтіснішим чином пов'язаний з розвитком методів синтезу відеоглядів [31, 35, 46, 54, 63, 69, 78, 86, 97, 98, 117, 133, 139]. Абстрагування (узагальнення, відділення від несуттєвих властивостей і зв'язків) відео – це коротке і змістовне уявлення відео. Мета полягає не тільки в скороченні обсягу оброблюваних даних, спрощення традиційних підходів до обробки відеопотоків, але і у створенні передумов для вирішення завдань навігації, індексування та релевантного пошуку візуальної інформації зі складними семантичним запитами [3, 12, 44, 103, 118, 128, 142]. Природно, що на даному етапі відео-узагальнення обмежується або видаленням надлишкових відеопослідовностей, або вибором істотного змісту [58, 59, 63, 71, 79, 96, 111, 129].

Відомо досить багато методів сумаризації відео з різною основою класифікації [30, 34, 38, 45, 53, 58, 72, 75, 80, 84, 86, 101, 110, 113, 134, 140, 141, 150, 164], але зі спільних позицій їх можна розділити два типи відеооглядів: статичні і динамічні [30, 63-65, 69, 82, 101, 119-120, 130, 157, 158, 163].

Методи динамічного реферування, коли вихідні дані також кліпи (відеороліки, трейлери з сюжетними історіями або без них) і статичного реферування, коли в результаті отримують набір відеокадрів. Основною перевагою статичного уявлення відео є те, що всі відеокадри можна подивитися практично одночасно (паралельно) на відміну від самого короткого динамічного ролика, перегляд

якого займає певний час. Покадрова обробка і виділення множин ключових кадрів, окрім того, спирається на весь арсенал методів цифрової обробки зображень, де успіхи більш помітні, ніж в методах обробки і аналізу відео [5, 11, 19, 21, 22, 67, 112, 123, 137, 142, 146, 159].

Організація відеоданих (темпоральні моделі грануляції, кластеризації, інкапсуляції, факторизації) є суттю часової сегментації – пошук часових інтервалів, носіїв відеокadrів, з точки зору нерозрізненості на основі заданої або формуємої в процесі обробки моделі подібності змісту відеопослідовностей. При цьому подальша кластеризація відеопослідовностей – передоснова відеореферування [38, 82, 84, 114, 150].

На рис. 1.1 наведена традиційна схема статичного відеореферування. Спочатку для вихідного відеопотоку (I) виконується темпоральна сегментація. Якщо відео  $V$  представлено кінцевою послідовністю кадрів  $V = \{F_1, F_2, \dots, F_n\}$ , то результат часової сегментації, як правило, має вигляд:  $S(V) = \{v_1, v_2, \dots, v_m\}$ , де  $v_i \cap v_j = \emptyset \quad \forall i \neq j, \cup v_j = V, i, j \in 1, 2, \dots, m$ , а сегмент – послідовність хронологічно зв'язаних відеокadrів. Слід особливо підкреслити, що в силу плавних переходів між сюжетами, замість розбиття  $S(V)$  може розглядатися покриття  $S_F(V)$ , тобто виникає потреба в нечіткому аналізі [26-27, 42-43, 51, 121, 131, 149]. Подібні «розділяючі» сегменти на практиці або видаляються з розгляду як семантично незначні і такі, що вносять інформаційний шум, або приписуються до двох послідовних сегментів повністю або частинами. Отримана темпоральна сегментація (II) є базою для пошуку ключових кадрів  $\{F_1^k, F_2^k, \dots, F_r^k\}$ ,  $r \ll n$ . Кількість ключових кадрів в кожному сегменті може варіюватися і вибиратися за різними критеріями, часто суперечливими, наприклад, найтипівіший кадр або найрізніший (III). Серед обраних ключових кадрів шукаються однакові або схожі (IV), після виключення яких отримуємо результат відеореферування (V).



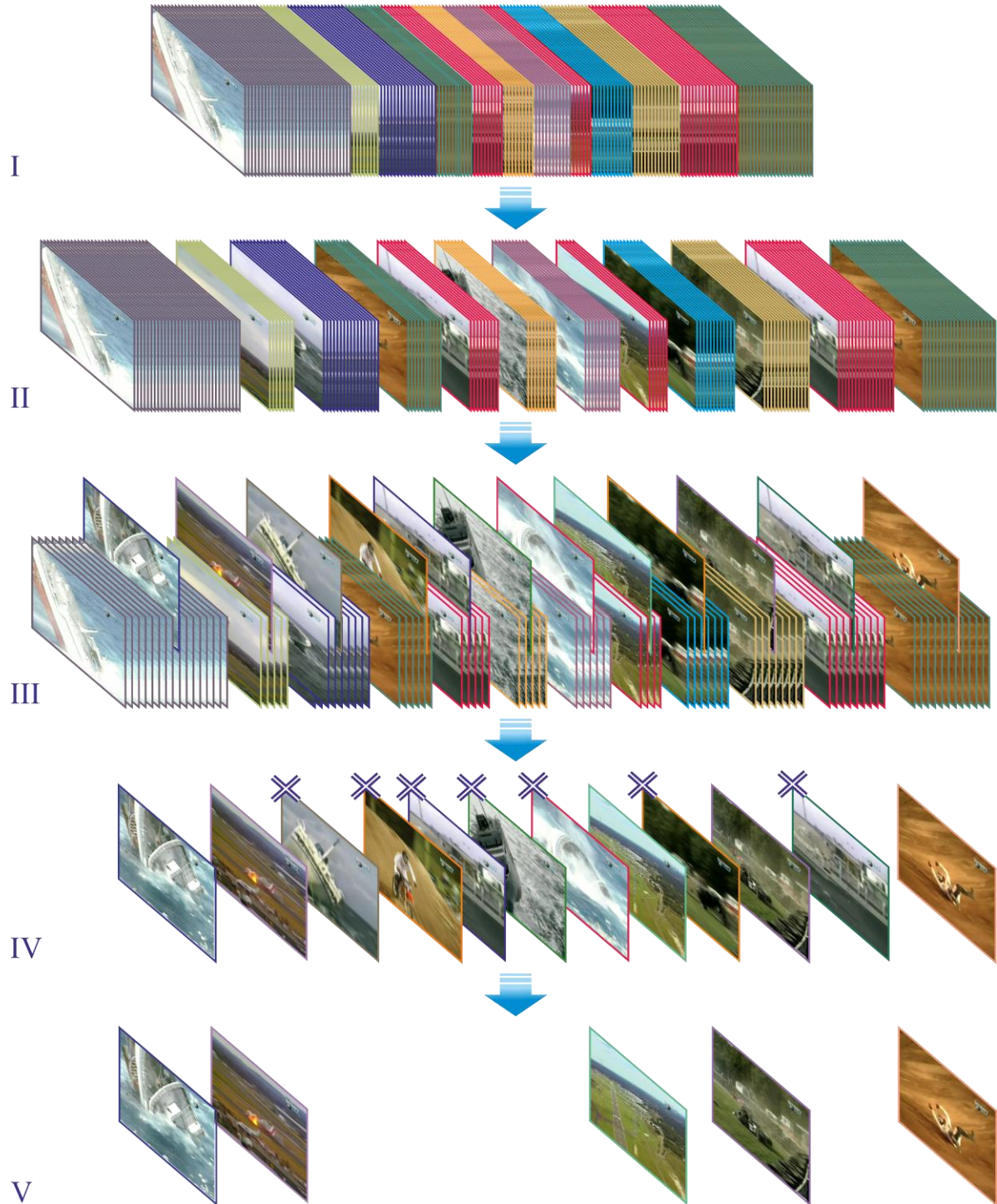


Рисунок 1.1 – Узагальнена схема статичного відеореферування

Розглянемо принципи пошуку відео на основі відеореферування із запитом «ad exemplum». Нехай  $U = \{F_1^k, F_2^k, \dots, F_N^k\}$  – деяка колекція ключових кадрів, що описують відеопотоки. Будемо вважати, що на множині  $U$  заданий на-

бір нормованих метрик

$$\rho_l(F_i^k, F_j^k) = \frac{\rho_l(F_i^k, F_j^k)}{1 + \rho_l(F_i^k, F_j^k)},$$

які у вигляді лінійної комбінації забезпечують семантично стійкий пошук відеокадрів, принаймі в заданій предметно орієнтованій області. Не порушуючи спільності, опустимо для простоти розуміння індексацію (фактично визначаючи покадровий пошук) позначимо через  $X \subseteq U$  відеодані в просторі ключових кадрів. Завдання полягає в пошуку по заданому запиту  $y \in U$  елемента (ів)  $x_r \in X$ , найкращим чином йому відповідного. Поняття «найкращим чином» зазвичай трактується як мінімум відстані  $\rho(y, x)$ ,  $y \in U$ ,  $x \in X$ . Якщо необхідно знайти  $p \in \mathbb{N}$  найсхожіших ключових кадрів  $X^p = \{x_{r_1}, x_{r_2}, \dots, x_{r_p}\} \subseteq X$ , необхідні відеокадри повинні задовільняти умові

$$\forall x_{r_j} \in X^k, \forall x \in X \setminus X^p, \forall y \in U \rho(y, x_{r_j}) \leq \rho(y, x), \rho(y, x_{r_j}) \leq \rho(y, x_{r_{j+1}}), j = \overline{1, p-1}.$$

Ретроспективний аналіз методів і алгоритмів індексації та пошуку таких, як методи просторової індексації (SAM – spatial access methods),  $R$ -дерева і їх модифікація,  $k-d-b$  ( $k$ -dimensional binary tree) дерева, лінійні квадродререва для ознакових просторів невисокої розмірності, методи метричної індексації AESA (Approximating and Eliminating Search Algorithm) і LAESA (Linear Approximating and Eliminating Search), VP-дерева (Vantage-Point tree), їх модифікації MVP-дерева, X-дерева для надвеликих колекцій зображень [69, 72, 124, 127], дозволяють стверджувати, що відеореферування виводить інформаційний відеопошук на принципіально новий рівень за рахунок формування складних семантичних запитів на основі сумаризації відеопотоків, зокрема, у вигляді наборів

ключових кадрів [58, 63-65, 79, 96, 105, 111, 129].

Найбільш розвинені методи статичного узагальнення відео на основі низькорівневих ознак. Пошук ключових кадрів (з темпоральною сегментацією або без неї при послідовному аналізі відеокадрів) базується на аналізі гістограм, текстур, текселей (мікротекстур), дескрипторів меж, ліній, кутів, точок і областей локалізації і позиціонування областей інтересу, обліку ознак форми цільового об'єкта або областей, отриманих в результаті сегментації поля зору, використанні просторових відношень для застосування морфологічних перетворень [104, 138, 144]. Перевага низькорівневого підходу полягає в тому, що подібні ознаки легко обчислювати, звідки виходить, що вони є кращий вибір для програм, що вимагають швидкого відгуку для ситуацій, коли сцени не є складними [21, 126].

В даний час для зниження семантичного розриву між низькорівневими ознаками і вимогами до смислових навантажень відео акцент зміщується до розпізнавання об'єктів, виділення подій, виявлення жестів, осіб, емоцій і т.п. [78, 97, 98, 117]. Підходи до узагальнення відео на основі функціональних можливостей високого рівня залежать, однак, від конкретного використання, потребують чутливість до змісту відео, але вимагають великих обчислювальних ресурсів та нерідко спираються на предметно-орієнтовані евристики [3, 69, 72]. Необхідно підкреслити, що облік темпоральних зв'язків у відеопотоках істотно підвищує валідність обробки і аналізу відео, і тим самим, динамічне відеореферування є перспективний шлях розвитку систем розуміння візуальної інформації в цілому.

Ще одним трендом розвитку відеореферування є комбінування різних підходів, що враховують глобальні та локальні властивості зображень. В аспекті локальності крім вже традиційних областей інтересу необхідно підкреслити області візуальної уваги [80, 82]. Однією з головних переваг моделей візуальної уваги (Visual Attention Models) є те, що вони не вимагають повного розуміння візуального змісту. Ці моделі пов'язані з когнітивними функціями, що відпові-

дають за ідентифікацію візуально різної інформації. Їхній головний недолік – вони досить складні, оскільки прагнуть до аналогій із зоровою системою людини. Крім того, подібні моделі недостатньо вивчені, існує безліч принципово різних моделей уваги, які потрібно якимось чином об'єднувати для розрахунку загальної візуальної привабливості змісту [80, 82, 126]. Проте, області візуального інтересу можна шукати в околах особливих точок, що характеризуються повторюваністю – знаходяться в одному і тому ж місці поля зору, не дивлячись на зміни точки огляду і освітленості, інформативністю – околи повинні мати помітні відмінності один від одного, локальністю – особлива точка повинна зменшувати чутливість до геометричних і фотометричних спотворень, точністю – виявлені особливі точки повинні точно локалізуватися як на оригінальному документі, так і при зміні масштабу. Час виявлення особливих точок на зображенні має бути допустимим в критичних за часом програмних засобах, а їх кількість (число виявлених особливих точок) має бути досить великим, щоб забезпечити ефективний відбір [79, 96].

Практично всі методи динамічної сумаризації так чи інакше пов'язані з покадровою обробкою відео, але одиницею аналізу є відеопослідовність. Репрезентативні послідовності є узагальненням ключових кадрів. Ясно, що їх пошук більш тісно пов'язаний з темпоральною сегментацією відеопотоків, що подаються багатовимірними часовими рядами в ознакових просторах або матричними (фрагментними) часовими рядами безпосередньо в просторі зображень [47, 49, 73, 83, 104, 125, 130].

При обробці відеопотоків фінальна пертінентність нерідко спирається на аналіз областей візуальної уваги і інтересу, які фактично є деякими локальностями поля зору. Інакше кажучи, фрагментна обробка відеоданих дозволяє знижувати (а іноді і повністю усувати) семантичний розрив між ознаками низького рівня, що обчислюються на кожному кадрі, і вимогами високого рівня при інтерпретації відео в часі на основі таких понять, як сценарії, події і т.п. На сучасному етапі одна з головних переваг фрагментної обробки полягає в тому, що

вона стійка до цілого ряду темпоральних деформацій відео – частковим затулянням об'єктів, варіацій ракурсу внаслідок зміни взаємного розташування та орієнтації об'єктів і датчика, мінливості умов спостереження (освітленість, контраст і ін.) . Фрагмент, який є околom деякої особливої точки [63-65, 111], може бути представлений або вектором ознак, наприклад, гистограмою яскравостей або напрямків градієнтів, або матрицею зображення або результатом, взагалі кажучи, довільної просторової згортки.

Резюмуючи проведений аналіз, можна зробити висновок, що розвиток підходів до узагальнення відеоданих пов'язаний з розробкою технологій динамічного відеореферування для інформаційного пошуку, яке в свою чергу, вимагає розробки, обґрунтування і дослідження перспективних методів сегментації і кластеризації відеопотоків.

## 1.2 Порівняльний аналіз методів і моделей часової сегментації відеопотоків

Для виявлення змін властивостей часових рядів відомий цілий набір відносно незалежних підходів: статистичний (кореляційний, спектральний, регресійний аналіз), адаптивний (експоненціальне згладжування, аналіз слідкуючого сигналу), заснований на методах класифікації (кластеризації, сегментації) [82, 114, 125, 144, 148]. Найбільш широкого поширення набули: моделі авторегресії змінного середнього (AR – Auto-Regressive, MA – Moving-Average, ARMA – Auto-Regressive Moving-Average, ARIMA – Auto-Regressive Integrated Moving-Average), і їх модифікації типу CARMAX (Control Auto-Regressive Moving-Average Model with auxiliary inputs, а також CARMA, CARIMA, CARIMAX), SSM (State-Space Models), SET (Spectral Expansion Technique) [82, 84, 114]. Стосовно до відеопотоків, коли, по-перше, природним чином виникає необхідність on-line обробки, під час перегляду-реєстрації послідовностей ознак або зобра-

жень лише з деякою буферною пам'яттю на передісторію, а по-друге, коли відсутня будь-яка завжди апріорна формалізована інформація про властивості багатовимірних часових рядів, індукованих відео, а ці властивості можуть радикально змінюватися при зміні сюжетів, найбільш перспективним представляється підхід, заснований на сегментації-кластеризації темпоральних послідовностей.

Нехай відео  $V = \{F_1, F_2, \dots, F_N\}$  індукує багатовимірний часовий ряд  $x(k) = (x_1(k), x_2(k), \dots, x_N(k))^T$ , де  $1 \leq k \leq N$  – дискретний час, в моменти якого проводиться фіксація відеокадра. Сегментом отриманого ряду  $S(a, b) = \{a \leq k \leq b\}$  є однорідна, с точки зору заданого критерію, послідовність  $x(a), x(a+1), \dots, x(b)$ , а завдання сегментації полягає в знаходженні сегментів, що не перетинаються  $S^c = \{S_i(a_i, b_i), 1 \leq i \leq c\}$  таких, що  $a_1 = 1, b_c = N, a_i = b_{i-1} + 1$ . Інакше кажучи, потрібно знайти  $c$  інтервалів таких, що  $S_1 < S_a < \dots < S_c$ , тобто фактично визначити їх межі  $a_i, b_i$ .

Взагалі кажучи, багатовимірні часові ряди несуть в собі надмірну інформацію, отже, їх обробка здійснюється на основі аналізу деяких узагальнених характеристик таких, як спектри, кореляційні функції, середні, дисперсії, головні компоненти і т.д., які необхідно обчислити в процесі цього аналізу, для чого існують досить глибоко-досліджені статистичні методи [11, 73, 160, 161]. Завдання істотно ускладнюється у випадках, коли обсяг вибірки  $N$  не фіксований, кількість сегментів  $c$  невідома, а самі дані  $x(k)$  надходять на обробку послідовно в реальному часі. Більш того, сегменти через природу відеопотоку (редакторські зміни сюжетів, дуже повільне панорамування) можуть перетинатися і необхідно переходити від покриттів (нечіткого аналізу) до розбиття.

Розглянемо компактні форми опису багатовимірних часових рядів для випадків, коли число спостережень не фіксоване і зростає з плинном часу, а математичні моделі таких рядів орієнтовані на виявлення моментів зміни основних характеристик.

Моделі багатовимірних послідовностей в структурному вигляді представляються наступним чином [7, 14]:

$$\sum_{l=0}^p B_l x(k-l) + Dz(k) = \eta(k),$$

де  $B_l$  – матричні коефіцієнти при внутрішньосистемних (ендогенних) змінних,

$B_0$  – невироджена матриця при ендогенних змінних поточного часу,

$D$  – матриця коефіцієнтів при позасистемних (екзогенних) змінних,

$z(k)$  – вектор екзогенних змінних, що включає і їх запізнілі значення,

$\eta(k)$  – векторний збурюючий сигнал з нульовим математичним очікуванням і обмеженими другими моментами.

Завдання полягає в ідентифікації параметрів цієї моделі. Для цього досить часто використовуються непрямий, двохкроковий або трикроковий методи найменших квадратів, які, однак, призначені для роботи тільки з фіксованою і заздалегідь заданою вибіркою. Алгоритми, призначені для роботи в послідовному режимі такі, як релаксаційний, рекурсивний або метод нерухомої точки, характеризуються недостатньо високою швидкістю збіжності для того, щоб забезпечувати ефективну обробку сигналу в режимі реального часу [30, 46, 82, 104].

Для вирішення завдання виявлення зміни властивостей багатовимірних рядів зручним, компактним і ефективним описом є векторні авторегресійні моделі (VAR-моделі), які пов'язують минулі і поточні спостереження векторного сигналу  $x(k)$  в формі

$$x(k) = B_0 + \sum_{l=1}^p B_l x(k-l) + \xi(k),$$

де  $B_0 = \{b_{0i}\}$  –  $(n \times 1)$  вектор середніх значень,

$B_l = \{b_{lij}\}$  –  $(n \times n)$  матриці параметрів,

$p$  – порядок моделі.

Виявлення зміни властивостей багатовимірного часового ряду  $x(k)$  пов'язане з контролем кожної з його компонент  $x_i(k), i = 1, 2, \dots, n$ , при цьому можливе виникнення наступних трьох ситуацій:

а) зміна середніх значень (рівнянь)  $l \leq p$  компонент

$$b_{0i}(k) = \begin{cases} b_{0i}, & \text{якщо } k < k_a, \\ b_{0i}^a, & \text{якщо } k \geq k_a, \end{cases}$$

б) зміна характеристик (дисперсій)  $l \leq p$  збурень  $\xi_i(\sigma_i^2)$

$$x_i(k) = \begin{cases} b_{0i} + \sum_{l=1}^p \sum_{j=1}^n b_{lij} x_j(k-l) + \xi_i(k), & \text{якщо } k < k_a, \\ b_{0i} + \sum_{l=1}^p \sum_{j=1}^n b_{lij} x_j(k-l) + \xi_i^a(k), & \text{якщо } k \geq k_a, \end{cases}$$

в) зміна коефіцієнтів  $b_{lij}$ , що приводить до зміни автокореляційних властивостей нестационарного часового ряду

$$\text{внутрішньосистемний } x_i(k) = \begin{cases} b_{0i} + \sum_{l=1}^p \sum_{j=1}^n b_{lij} x_j(k-l) + \xi_i(k), & \text{якщо } k < k_a, \\ b_{0i} + \sum_{l=1}^p \sum_{j=1}^n b_{lij}^0 x_j(k-l) + \xi_i(k), & \text{якщо } k \geq k_a, \end{cases}$$

де  $k_a$  – момент часу, в який відбувається вимір властивостей.

Діагностика в режимі реального часу може бути реалізована на основі настроюваних VAR-моделей, яких навчаються за допомогою тих чи інших проце-



дур адаптивної ідентифікації з кінцевою пам'яттю [7, 10].

Нерідко при аналізі великих за обсягом і за розмірністю часових рядів є завдання виділення латентних факторів, що в підсумку має на меті спростити ряд з точки зору виявлення змін його властивостей. Одним з найбільш ефективних підходів до вирішення цього завдання є метод головних компонент, який при сегментації відеорядів може бути представлений таким чином.

Стартова інформація для аналізу –  $k \times n$  матриця спостережень

$$X = \begin{pmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & & x_n(2) \\ \vdots & & & \\ x_1(u) & x_2(u) & \dots & x_n(u) \\ \vdots & & \ddots & \\ x_1(k) & x_2(k) & \dots & x_n(k) \end{pmatrix},$$

утворена з  $k$   $n$ -мірних векторів  $x(u) = (x_1(u), x_2(u), \dots, x_n(u))^T$ , і її кореляційна  $(n \times n)$  матриця виду

$$R(k) = \frac{1}{k} \sum_{u=1}^k (x(u) - \bar{x})(x(u) - \bar{x})^T = \frac{1}{k} \sum_{u=1}^k x^c(u) x^{cT}(u),$$

де  $x^c(u) = x(u) - \bar{x}$  – центровані щодо середнього вихідні дані.

Метод головних компонент полягає в проєктуванні спостережуваних вихідних даних з початкового  $n$ -вимірного простору в  $m$ -вимірний ( $n > m \geq 1$ ) вихідний і зводиться до знаходження системи  $w^1, w^2, \dots, w^m$  ортогональних власних векторів матриці  $R(k)$  таких, що  $w^1 = (w_1^1, w_2^1, \dots, w_n^1)^T$  відповідає найбільшому власному значенню  $\lambda$ , матриці  $R(k)$ ,  $w^2 = (w_1^2, w_2^2, \dots, w_n^2)^T$  – другого за величиною своїм значенням  $\lambda$  і т.д. Іншими словами, шукається розв'язання

матричного рівняння

$$(R(k) - \lambda_l I)w^l = 0$$

таке, що  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$  та  $\|w^l\|^2 = 1$ .

Таким чином, вирішується завдання переходу в простір більш низької розмірності з мінімальною втратою інформації, тобто задача послідовного пошуку множини власних векторів  $w^1, w^2, \dots, w^m$  шляхом оптимізації кожного з локальних функціоналів, які формують глобальний критерій

$$I_w(k) = \frac{1}{k} \sum_{l=1}^m \sum_{u=1}^k (x^{cT}(u)w^l)^2$$

при обмеженнях

$$\begin{cases} w^{lT} w^p = 0, \text{ при } l \neq p; \\ w^{lT} w^p = 1. \end{cases}$$

Перша головна компонента, що несе максимум інформації про відеопотік може бути знайдена шляхом максимізації локального критерію

$$I_w^1(k) = \frac{1}{k} \sum_{u=1}^k (x^c(u)w^1)^2$$

за допомогою стандартного методу невизначених множників Лагранжа.

Далі з кожного вектору віднімається його проекція на першу головну компоненту  $x^c(u)$  і обчислюється перша головна компонента залишків, що є

другою головною компонентою вихідних даних і ортонормального до першої.

Третя головна компонента обчислюється шляхом проектування кожного вихідного вектора на першу і другу головні компоненти, вирахування цієї проєкції  $x^c(u)$  з кожного і знаходження першої головної компоненти отриманих залишків, що є третьою компонентою вихідних даних. Решта головних компоненти обчислюються рекурсивно відповідно до поданої процедури.

З проведеного аналізу безпосередньо випливає, що валідність сегментації відеопотоків, їх наближення до ground-truth темпоральних розбиттів вимагають розвитку методів детектування зміни властивостей часових рядів, індукованих відеоданими. Визначення меж сегментів фактично структурує відеопотік. Важливо, що сегменти – це хронологічно зв'язані послідовності відеокадрів, тобто стандартним є наявність схожих сегментів, які рознесені в часі спостереження, що вимагає їх подальшої кластеризації та класифікації. Будучи основним елементом подальшої сумаризації відео, сегменти привносять і нові складності, зокрема, їх довжини різні, що істотно ускладнює їх компаративний аналіз.

### 1.3 Аналіз методів кластеризації стосовно темпоральних сегментів відеоданих

Концептуально кластеризація, як спосіб створення груп об'єктів таким чином, що об'єкти в одному кластері дуже схожі, а об'єкти в різних кластерах (і власне самі кластери) істотно різні, являє собою більш загальний теоретико-прикладний інструментарій аналізу даних [25, 29, 35, 57, 62, 82, 84, 87, 150-152, 155, 163] , ніж їх сегментація, оскільки остання передбачає необхідність зв'язності. При сегментації відеокадрів – в полі зору, при сегментації відеопотоків – в часі. Типи даних (інтервальні, порядкові, категоріальні, транзакційні, часові ряди і т.д.) і їх суміші безпосередньо впливають на кластеризацію, будучи, мабуть, основним фактором, який слід враховувати при виборі і модифікації

алгоритмів кластеризації [27, 29], в тому числі з урахуванням перетворень шкал, нормалізацією і трансформаціями даних [56, 70].

Проаналізуємо методи сегментації, розвиток яких може бути покладено в основу обробки відеопотоків.

Ієрархічні алгоритми кластеризації і алгоритми поділу є двома основними типами кластеризації. На відміну від алгоритмів поділу, які продукують одне розбиття даних, ієрархічні алгоритми в результаті дають послідовність вкладених розбиттів. Існує два основних типу ієрархічних алгоритмів: агломеративні (об'єднавчі) алгоритми і та дівізімні (розділяючі) алгоритми. Агломеративні алгоритми стартують з кожного окремого об'єкта, тоді як дівізімні алгоритми аналізують всі об'єкти і повторюють розбиття великих кластерів на менші частини з можливим перерозподілом по кластерам [35, 56, 114].

При аналізі відеопослідовностей велике значення мають нечіткі алгоритми кластеризації, які на відміну від чіткої кластеризації (кожна точка належить одному і тільки одному кластеру), дозволяють точці даних належати двом або більше кластерам з різними рівнями функцій належності. І теоретично, і практично найбільш розвинені нечіткі  $k$ -середні, нечіткі  $k$ -моди і  $c$ -середні [68, 87, 121, 155, 156].

Одну з ключових ролей з прикладної точки зору при кластеризації великих наборів даних і наборів даних великих розмірностей грають алгоритми, засновані на центрах (середніх, медоїдах, прототипах, модах і т.д) кластерів [87, 121, 155]. При аналізі відеопотоків можна використовувати їх ітераційні модифікації.

Досить перспективним напрямком відшукування глобальних екстремумів в задачах побудови угруповань даних є, як правило, евристичні алгоритми пошукової кластеризації, серед яких доцільно виділити підходи, засновані на генетичних алгоритмах, алгоритмах табу-пошуку, алгоритмах імітації відпалу [68, 87, 152]. Реалізація подібних алгоритмів в on-line режимі малоімовірна, але пошук глобальних екстремумів може бути використаний при відеореферуванні сегме-

нтованих послідовностей кадрів. До цього ж класу відносяться графові алгоритми кластеризації. Вони групують набір даних шляхом перетинів графа або гіперграфу, побудова яких зазвичай ґрунтується на матриці відмінностей даного набору даних [114, 151].

Грид-алгоритми кластеризації також відносяться до зазначеного типу. Їх суть: створення структури сітки, оцінка щільності для кожного осередку, сортування осередків по їх щільності, виявлення кластерних центрів і обхід сусідніх осередків [29, 57]. Основна перевага – значне зниження складності обчислень, недолік – розмиття релевантності відеореферування. Сказане повною мірою відноситься до кластеризації на основі щільності, здатної знаходити кластери довільної форми. Кластери визначаються як області підвищеної щільності елементів, розділені областями низької щільності. Такі алгоритми не в повній мірі підходять для багатовимірних наборів даних. Серед них виділяються DBSCAN (Density Based Spatial Clustering of Applications with Noise), BRIDGE, DBCLASD, (Distribution-Based Clustering of LArge Spatial Databases), DENCLUE (DENsity-based CLUstEring) [5, 87, 114].

У моделі-орієнтованих алгоритмах передбачається, що дані надходять з суміші розподілів ймовірностей, кожна з яких представляє окремий кластер. Найяскравіший представник аналізу сумішей розподілів – EM (Expectation Maximization) алгоритм [29, 152]. Кластеризація в підпросторах є відносно новою концепцією, її особливість полягає в тому, що алгоритми здатні ідентифікувати різні кластери, вбудовані в різні підпростори багатовимірних даних [57, 62], що, принаймі може бути корисним при раціональному виборі ознакових просторів при мультіаспектній індексації відеопотоків в задачах інформаційного пошуку.

Міра подібності / відмінності використовується для кількісного опису порівняння двох точок даних, двох кластерів, точки і кластеру, а в найзагальнішому вигляді – розбиття або покриття, які є результатами кластеризації. Ясно, що міри схожості / відмінності є основними елементами будь-якого алгоритму

кластеризації.

Класичний аналіз часових рядів включає в себе виявлення закономірностей (наприклад, аналіз трендів, сезонності, автокореляцію, моделі авторегресії і прогнозування). Традиційна обробка часових рядів нерідко передбачає фільтрацію, амплітудну нормалізацію, видалення базового рівня часового ряду, лінійного тренду і т.д., але для темпоральної обробки відеопотоків подібні операції не мають сенсу. З точки зору поганоструктурованих колекцій часових даних, обробки відеопотоків, інформаційного пошуку виникає ряд специфічних завдань:

– Проблема схожості: потрібно визначити, чи мають різні часові ряди однакове поведінку, а саме як за двома часовими рядами  $x = (x_1, x_2, \dots, x_r)$  і  $y = (y_1, y_2, \dots, y_t)$  можна обчислити відмінність  $d(x, y)$  або схожість  $s(x, y)$ ?

– Проблема індексації: пошук найкращого збігу з запитом до надвеликого набору даних з максимально можливим скороченням переборів.

– Проблема подібності підпоследовності: за заданим шаблоном або запитом  $Q$  до неструктурованої колекції  $S$  і показником відстані знайти місце розташування, яке найкращим чином відповідає  $Q$ .

– Проблема кластеризації: пошук природних угруповань часових рядів в базі даних по деякому показнику подібності або відмінності, наприклад, так, щоб елементи кожного кластеру мали однакову темпоральну динаміку.

– Проблема виявлення вирішуючих правил: пошук правил, що пов'язують шаблони в часі ряду з іншими шаблонами в цьому або інших часових рядах.

Одним з найважливіших завдань при обробці та інтерпретації відеопотоків є компаративний аналіз відеопоследовностей різної довжини, яке може вирішуватися за допомогою динамічної деформації часової шкали (Dynamic Time Warping).

Основна ідея DTW полягає в тому, що последовності розширюються повторюваними елементами, і відстань обчислюється між розширеними последо-

вностями. Отже, динамічна деформація часової шкали може обробляти вхідні послідовності різної довжини, найчастіше зменшуючи частоту помилок на порядок при розрахунку відстаней [36, 37, 50, 88-90, 128, 162]. Більш строго суть DTW полягає в наступному.

Припустимо, дано два часових ряду  $x = (x_1, x_2, \dots, x_r)$  та  $y = (y_1, y_2, \dots, y_s)$ , де довжини рядів  $r$  та  $s$  не обов'язково рівні. Нехай  $M$  –  $r \times s$ -матриця, в якій  $(i, j)$ -елемент – це квадрат евклідової відстані  $d(x_i, y_j) = (x_i - y_j)^2$  між точками  $x_i$  та  $y_j$ . Кожен елемент  $(i, j)$  в  $M$  відповідає вирівнюванню між точками  $x_i$  та  $y_j$ . Тоді кожне можливе відображення  $x$  в  $y$  може бути представлено як викривлений шлях в матриці  $M$ , тобто безперервною множиною матричних елементів.

Нехай  $W = w_1, w_2, \dots, w_k$  – викривлений шлях, в якому  $w_k = (i_k, j_k)$ , тоді  $\max\{r, s\} \leq K < r + s - 1$ . На викривлені шляхи накладається ряд обмежень: монотонність, неперервність, граничні умови.

– Монотонність: якщо  $w_k = (i, j)$ , то  $w_{k-1} = (i', j')$ , де  $i \geq i'$  та  $j \geq j'$ ; це означає, що викривлений шлях ніколи не йде вниз і вліво.

– Неперервність: якщо  $w_k = (i, j)$ , то  $w_{k-1} = (i', j')$ , де  $i \leq i' + 1$  та  $j \leq j' + 1$ ; це означає, що викривлений шлях проходить по сусіднім елементам, і жоден елемент послідовності не може бути опущений.

– Граничні умови: перший та останній елементи в  $W$  фіксовані, а саме  $w_1 = (1, 1)$ ,  $w_K = (r, s)$ ; якщо для викривленого шляху задане вікно розміру  $w$ , то розглядаються лише такі елементи матриці  $M$ , для яких  $|i - j| \leq w$ .

Існує експоненційно багато шляхів, які задовольняють вищевказаним умовам. Оптимальний шлях деформації серед них той, який мінімізує її вартість

$$\text{DTW}(x, y) = \frac{\sqrt{\sum_{l=1}^K d(x_{i_l}, y_{j_l})}}{K},$$

де  $(i_l, j_l) = w_l$  для  $l = 1, 2, \dots, K$ .

Оптимальний шлях може бути знайдений за допомогою динамічного програмування. Нехай  $\gamma(i, j)$  визначає DTW між послідовностями  $x_1, x_2, \dots, x_i$  та  $y_1, y_2, \dots, y_j$ , тоді оптимальний шлях можливо знайти за допомогою динамічного програмування, щоб оцінити рекурсію

$$\gamma(i, j) = d(i, j) + \min \{ \gamma(i-1, j), \gamma(i-1, j-1), \gamma(i, j-1) \}.$$

Часова складність для базової реалізації становить  $O(rs)$ . Слід особливо підкреслити, що DTW можливо апроксимувати різними шляхами: на основі стиснення часових рядів, їх передискретизацією або іншими уявленнями часового ряду.

Підводячи проміжний підсумок, слід підкреслити наступне.

Вельми специфічним завданням є обробка відеоінформації в найширшому сенсі (за програмними цілями і методами), включаючи інформаційний пошук в надвеликих колекціях відеопотоків. Основними факторами, що визначають складність цього завдання, є величезні обсяги аналізованої інформації, представлення зображень в матричній формі або не завжди в релевантних ознакових просторах, спотворення відеокадрів різного роду збуреннями і шумами. У таких завданнях, класичні традиційні методи кластеризації нерідко виявляються неефективними і вимагають свого розвитку.

#### 1.4 Постановка основних задач досліджень

Розвиток інфраструктури інформаційних технологій в аспекті диверсифікації засобів одночасного використання різних форм представлення даних, ме-



тодів, способів і засобів, що забезпечують зберігання, обробку, передачу, відображення, аналіз і інтерпретацію інформації, що включає в себе графіку, звук, зображення, відео, текст і посилання, для створення нелінійних середовищ управління мультимедіа представляє безперечний інтерес. Особлива увага приділяється відеопотокам, дослідження яких є відносно новою, що постійно розширюється, теоретико-прикладної областю та часто вимагає високорівневого аналізу контенту. Перелік виникаючих мультидисциплінарних завдань аналітико-синтетичної переробки відеоданих практично безмежний, але останнім часом все більша увага приділяється обліку парадигматических, аналітичних та асоціативних просторово-темпоральних відношень, що створює передумови для раціонального забезпечення інформаційних потреб. Одним з найважливіших механізмів в цьому напрямку є відеореферування. До теперішнього моменту запропоновані різні ідеї і методи для ефективної абстракції відео-контенту. Спираючись на розвиток комп'ютерних технологій в цілому, відеосумаризація розвивається в двох напрямках: статичному (на основі ключових кадрів) і динамічному (на базі скімінгу) відеореферуванні. Ясно, що, спираючись на досягнення методів комп'ютерного зору, статична відеореферування досягло більш високого рівня, але динамічна сумаризація має більш широку перспективу в силу обліку темпоральних зв'язків при обробці відеоконтенту.

Найпростіший динамічний метод створення відеорефератів – це рівномірний сабсемплінг, який видобуває уривки вихідного відео з фіксованою тривалістю через фіксовані інтервали. Стартуючи з «швидкого перемотування вперед» – рівномірного або адаптивного управління частотою кадрів для аналізу відеоконтенту, дослідники перенесли акцент на збереження однакової частоти кадрів відео шляхом вибору тільки «відповідних» або «важливих» відеофрагментів, наприклад, з огляду на знання про ключові кадри простий метод генерації анотацій / рефератів полягає в тому, щоб включати часові околиці ключових кадрів.

З точки зору перегляду і навігації відео, хороший відеореферат дозволяє

Користувачу отримувати максимальну інформацію про цільову відеопослідовність за зазначений час або достатню інформацію за мінімальний час. Довжина кожної відеопослідовності для узагальнення відео може бути вказана апіорі або залишатися невідомою до кінця скімінгу. Очевидно, другий варіант є більш адекватним по відношенню до змісту відео, але пов'язаний зі складними обчислювальними моделями, в яких до того ж необхідні порівняння послідовностей різних довжин.

Методи абстракції найчастіше призначені для полегшення перегляду колекцій відеоданих. Основним обмеженням перегляду відеоабстракцій є те, що він може бути ефективним тільки в тому випадку, якщо кількість цільових відеопослідовностей доволі мала. Перегляд великої кількості відеорефератів для пошуку потрібної послідовності може бути дуже трудомістким і виснажливим для Користувача. Отже, механізм пошуку відео додатковоо служить фільтром змісту колекцій відеоданих відповідно до запитів Користувача, а результатом є анотації потенційно релевантних відеопослідовностей.

На додаток до полегшення перегляду великий відео колекції, відео огляди також допомагають Користувачеві взаємодіяти з відеорядами нелінійним чином для швидкого доступу до семантично релевантного хронологічного положення в послідовності, що особливо корисно при редагуванні відео і створення програмних засобів.

Для стратифікованого відеореферування (множин ключових відеокадрів, відеосегментів, їх угруповання, що відображають події послідовності сцен) довжина ковзної частини в ієрархії спрямована вниз, і обробка кожної одиниці може виконуватися незалежно. Подібні процедури можуть реалізовуватися тільки в off-line режимі.

Процес автоматичної генерації сумаризації відео в загальному вигляді представляється наступними етапами: темпоральна сегментація, вибір релевантних сегментів, стиснення або скорочення сегментів, мультимодальна інтеграція та збирання відеоузагальнення. Всі ці етапи в тій чи іншій мірі пов'язані з

завданнями сегментації і кластеризації відеопослідовностей.

Таким чином, на відміну від традиційних коротких характеристик первинних документів, відеореферування як методика узагальнення – це короткий виклад змісту відеопотоків в сенсі деяких критеріїв смислової відповідності або іншими словами, релевантний пошуковий образів, що забезпечує інформаційну потребу Користувача. На основі, як правило, багатоаспектного деривантного індексування в результаті забезпечуються інтерактивні та автоматичні засоби аналітико-синтетичної семантичної переробки відео, швидкого пертінентного перегляду, організації нелінійного хронологічного середовища, а також інформаційного пошуку зі складними семантичними запитами з підвищенням показників точності, повноти й оперативності пошуку.

На основі викладеного можна сформулювати основні завдання дисертаційних досліджень.

– Розробити методи швидкого виявлення зміни властивостей багатовимірних часових рядів, індукованих відеопотоками, на основі ідентифікаційного підходу до ансамблю моделей.

– Розробити методи виявлення змін в потоках відео, представлених у вигляді багатовимірних часових рядів, на основі послідовної кластеризації.

– Розробити методи виявлення змін властивостей багатовимірних темпоральних послідовностей різної довжини.

– Дослідити властивості адаптивності ітеративної динамічна часової деформації в задачах сегментації-кластеризації багатовимірних послідовностей, які продукуються послідовностями відеокадрів.

– Формалізувати і вивчити гібридний підхід до кластеризації відеорядів різної довжини.

– Розробити адаптивні матричні моделі в задачі контролю потоків відео.

– Розробити метод кластеризації послідовностей відеоданих на основі гармонійних  $k$ -середніх.

– Для вирішення завдань відеореферування синтезувати процедуру клас-

теризації відео на основі представників – темпоральних сегментів.

– Запропонувати і дослідити технологію динамічного відереферування для задач інформаційного пошуку візуальної інформації з запитами ‘ad exemplum’.

### Висновки по першому розділу

1. Динамічне відереферування – узагальнення відеопотоків на основі виділення послідовностей відеокадрів, нерозпізнаних з позицій заданого критерію, пошук представників, найбільшою мірою відповідних контенту відеоданих, сумаризація знайдених сегментів – являє собою важливу науково-технічну задачу, рішення якої покращує технології інформаційного пошуку зі складними семантичними запитами.

2. Темпоральна сегментація – синтез хронологічно зв’язних розбиттів або покриттів відеопослідовності на окремі блоки, перетворена частина яких використовується при відеореферуванні – основний інструмент попередньої обробки відеопотоків на основі аналізу багатовимірних часових рядів, індукованих відео, або матричних послідовностей (фрагментів) відеокадрів, головною метою якої є структуризація відеоданих.

3. Розвиток методів кластеризації на рівні відеокадрів і їх фрагментів, а також на рівні відеопослідовностей (особливо в послідовному режимі) створює широкий спектр передумов для обробки та інтерпретації відеоданих в цілому та насамперед для аналітико-синтетичної переробки великих колекцій відеопослідовностей.

Список використаних джерел у даного розділі наведено у повну списку використаних джерел під номерами: [3, 5, 7, 8, 10 12, 14, 19, 21, 22, 25, 27, 29-31, 34 -38, 44-47, 49, 50, 53, 54, 56-59, 62-65, 67-73, 75, 78-80, 82-84, 86-90, 96-98, 101, 103-105, 110 -114, 117-121, 123, 125, 126, 128, 129, 130, 133, 134, 137-142, 144, 148, 150-152, 155-164].

## 2 МОДЕЛІ ТЕМПОРАЛЬНОЇ СЕГМЕНТАЦІЇ ВІДЕОПОТОКІВ

Первинним, а часто і першочерговим завданням інформаційних технологій обробки і аналізу відеопотоків є структурування відеорядів у вигляді сегментів (підпоследовностей, «нерозпізнаних» з точки заданого критерію) потоків даних. Сукупність сегментів представляє собою розбиття або покриття відеоследовностей, а кожен з них є основною «одиницею» обробки. Саме тому точність і валідність їх виділення зумовлюють ефективність реферування і / або анотування відеоданих, а також інформаційного пошуку з запитом в формі відеоряду [12, 32, 135]. Більш того, отримані сегменти – це вихідні дані для подальшої семантичної кластеризації. Оперування колекціями відеопотоків і окремими відеорядами виконується або в просторі зображень (як правило, наборів фрагментів), або в багатовимірних признакових просторах, які продукуються властивостями кожного кадру. Специфіка темпоральної сегментації потоків візуальної інформації [47, 66, 83, 125, 138, 145, 154] полягає в тому, що вона здійснюється в online режимі, коли можливо враховувати лише поточне спостереження і деяку передісторію .

Розділ присвячений розробці методу виявлення зміни властивостей потоків даних на основі ідентифікаційного підходу до ансамблю моделей. Увагу також сфокусовано на непрямому підході до кластеризації последовностей, який в поєднанні з методами нечіткої рекуррентної оптимізації забезпечує синтез досить простих обчислювальних моделей сегментації відеорядів. Оскільки в структуруванні відеорядів особливе місце займають питання порівняння последовностей різної довжини, запропонована сегментація на основі ітеративної динамічної часової деформації, в основі якої лежить адаптивна редукція часових рядів.

## 2.1 Виявлення зміни властивостей потоків даних на основі ідентифікаційного підходу до ансамблю моделей

При сегментації і кластеризації часових рядів для виявлення зміни властивостей перспективним видається використання «модельних» методів (МВА - Model Based Approach) [102]. Відповідно до МВА аналізуються не власне часовий ряд, який може містити великий обсяг спостережень, а деяка компактна математична модель, що досить адекватно описує властивості потоку даних. Якщо ж спостереження надходять на обробку послідовно в online режимі, ефективним є використання адаптивних моделей, параметри яких безперервно уточнюються в режимі реального часу.

В якості таких моделей досить зручно використовувати так звані векторні авторегресійні моделі (Vector AutoRegression - VAR-models) [85], що мають відносно невелику кількість параметрів, що настроюються і з точки зору обчислювальної реалізації найбільш прості для аналізу багатовимірних часових рядів [14, 16, 40, 41, 43]. Особливістю VAR моделей є те, що вони пов'язують з допомогою простих співвідношень минулі, поточні та майбутні значення векторного сигналу  $x(k)$  в формі

$$x(k) = A_0 + \sum_{l=1}^p A_l x(k-l) + \xi(k), \quad (2.1)$$

де  $x(k)$  –  $(n \times 1)$ -вимірний контролюємий сигнал в поточний момент дискретного часу  $k = 1, 2, \dots$ ;

$A_l$  –  $(n \times n)$  матриці коефіцієнтів, що оцінюються;

$p \geq 1$  – порядок моделі;

$\xi(k)$  – «білошумне» збурення, компоненти якого мають обмежену дисперсію.

Для знаходження коефіцієнтів в online режимі введемо складову мат-

рицю, утворену цими коефіцієнтами, виду  $A = (A_0 : A_1 : \dots : A_p)$  та розширений вектор входів  $x(k) = (1 : x^T(k-1) : \dots : x^T(k-p))^T$  розмірностей  $(n \times pn + 1)$  та  $(pn + 1 \times n)$  відповідно та перепишемо вираз (2.1) в компактній формі

$$x(k) = AX(k) + \xi(k), \quad (2.2)$$

де матриця параметрів  $A$  містить всю необхідну інформацію про контролюєму багатовимірну послідовність.

У відповідність опису (2.2) можна поставити адаптивну модель виду

$$\hat{x}(k) = A(k-1)X(k), \quad (2.3)$$

де  $\hat{x}(k)$  –  $n$ -вимірний вихідний сигнал моделі;

$A(k-1)$  – матриця коефіцієнтів, що настроюються, що отримана на підставі попередніх  $k-1$  спостережень.

Для уточнення матриці оцінок  $A(k)$  нестационарного сигналу  $x(k)$  в принципі можливо використовувати традиційний експоненційно зважений рекурентний метод найменших квадратів, який, однак, ймовірно нестійкий при великих розмірностях матриці параметрів  $A(k)$ , що настроюються. У зв'язку з цим використана багатовимірна модифікація [4] алгоритму зваженої стохастичною апроксимації [42] виду

$$\begin{cases} A(k) = A(k-1) + \frac{e(k)x^T(k)}{\alpha r(k-1) + \|x(k)\|^2}, \\ e(k) = x(k) - \hat{x}(k) = x(k) - A(k-1)x(k), \\ r(k) = \alpha r(k-1) + \|x(k)\|^2, \end{cases} \quad (2.4)$$

де  $0 \leq \alpha \leq 1$  – параметр забування.

Слід зазначити, що при  $\alpha = 0$ , алгоритм (2.4) співпадає з однокроковим алгоритмом прогнозування багатовимірних випадкових послідовностей [132]. Дана процедура забезпечує максимальну швидкодію процесу настройки параметрів моделі, але вона вкрай чутлива до появи різких викидів в сигналі обурення  $\xi(k)$  і стрибків параметрів, що входять в матрицю  $A$  в (2.2). При  $\alpha = 1$  (2.4) набуває властивостей стохастичної апроксимації і має виражені згладжуючі властивості. У задачі обробки стаціонарних сигналів  $\alpha = 1$  забезпечує побудову оптимального прогнозу.

Для виявлення зміни властивостей векторного сигналу  $x(k)$  в [4] було запропоновано використовувати концепцію стежачого сигналу і настроємого з його допомогою параметра  $\alpha$ . Даний підхід забезпечує надійне виявлення зміни властивостей контрольованої послідовності, але він досить інерційний, тобто сигнал розладнання генерується з деяким запізненням. У ситуаціях, коли на перший план виходить підвищена швидкодія процесу виявлення розладнань, використання стежачого сигналу може виявитися неефективним.

У зв'язку з цим, пропонується використовувати концепцію ансамблю прогнозуючих моделей, коли навчання структури (2.3) проводиться за допомогою набору алгоритмів (2.4) з різними значеннями параметра забування  $0 = \alpha_1 < \alpha_2 < \dots < \alpha_r < \dots < \alpha_s = 1$ . Таким чином, генерується набір прогнозів  $\hat{x}_1(k), \hat{x}_2(k), \dots, \hat{x}_r(k), \dots, \hat{x}_s(k)$ , при цьому  $\hat{x}_1(k)$  практично миттєво реагує на всі можливі обурення і зміни в сигналі  $x(k)$ , а по суті  $\hat{x}_s(k)$  є відфільтрованою на всьому інтервалі спостережень послідовністю.

В процесі настройки ансамблю моделей контролюється або сигнал розбіжності між прогнозами

$$T_{qr}(k) = \|\hat{x}_q(k) - \hat{x}_r(k)\|, \forall q, r = 1, 2, \dots, S \quad (2.5)$$

або сигнал розбіжності між матрицями параметрів, що настроюються



$$T_{qr}^A(k) = (Sp(A_q(k) - A_r(k))(A_q(k) - A_r(k))^T)^{1/2} \quad \forall q, r = 1, 2, \dots, S. \quad (2.6)$$

При перевищенні сигналом розладнання деякого апіорі заданої умови  $\Delta$  або  $\Delta^A$

$$T_{qr}(k) > \Delta, \quad T_{qr}^A(k) > \Delta^A, \quad (2.7)$$

приймається рішення про те, що багатовимірний сигнал  $x(k)$  зазнав зміна властивостей.

Перейдемо до адаптивного виявлення зміни властивостей багатовимірних послідовностей на основі матричних авторегресійних моделей [7, 10, 15].

У завданнях обробки потоків відео, коли кожен окремий кадр  $x(k) = \{x_{i_1, i_2}(k)\}$ ,  $i_1 = 1, 2, \dots, n$ ;  $i_2 = 1, 2, \dots, v$  являє собою по суті  $(n \times v)$ -матрицю, використання моделі (2.1) представляється неефективним, оскільки веде до різкого збільшення числа оцінюваних параметрів, а отже, до зниження швидкодії процесу виявлення розладнань. У цій ситуації доцільно ввести в розгляд матричну авторегресійну модель (Matrix AutoRegression - MAR-model) виду

$$x(k) = \sum_{l=1}^p A_l x(k-l) B_l + \xi(k), \quad (2.8)$$

де  $A_l, B_l$  –  $(n \times n)$ ,  $(v \times v)$ -матриці параметрів, що оцінюються;

$p \geq 1$  – порядок моделі;

$\xi(k)$  – «білошумне» матричне збурення.

Вводячи складові матриці  $A = (A_1 : A_2 : \dots : A_p)$ ,  $B = (B_1 : B_2 : \dots : B_p)^T$  розмірностей  $(n \times np)$ ,  $(vp \times v)$  відповідно і розширену матрицю входів

$$x(k) = \begin{pmatrix} x(k-1) & & & & 0 \\ & \ddots & & & \\ & & x(k-l) & & \\ & & & \ddots & \\ 0 & & & & x(k-p) \end{pmatrix}$$

розмірності  $(np \times vp)$ , перепишемо вираз (2.8) в компактній формі

$$x(k) = AX(k)B + \xi(k), \quad (2.9)$$

що є матричним аналогом опису (2.2). Зауважимо також, що (2.9) містить  $(n^2 + v^2)p$  параметрів, що підлягають оцінюванню.

У відповідність опису (2.9) можна поставити адаптивну модель виду

$$\hat{x}(k) = A(k-1)X(k)B(k-1), \quad (2.10)$$

де  $\hat{x}(k)$  –  $(n \times v)$ -вимірний вихідний сигнал моделі;

$A(k-1)$ ,  $B(k-1)$  – матриці коефіцієнтів, що настраюються, що отримані на підставі попередніх  $k-1$  спостережень.

Для уточнення параметрів моделі (2.10) в [4] був введений градієнтний адаптивний алгоритм ідентифікації виду

$$\begin{cases} A(k) = A(k-1) + \eta_A(k)e_A(k)B^T(k-1)x^T(k), \\ e_A(k) = x(k) - A(k-1)x(k)B(k-1), \\ B(k) = B(k-1) + \eta_B(k)x^T(k-1)A(k)e_B(k), \\ e_B(k) = x(k) - A(k)x(k)B(k-1), \end{cases} \quad (2.11)$$

де  $\eta_A(k)$ ,  $\eta_B(k)$  – параметри кроку навчання.

В [42] були також оцінені значення параметрів кроку  $\eta_A(k)$ ,  $\eta_B(k)$ , що

забезпечують максимальну швидкість збіжності процедури (2.11). При цьому отримана процедура є аналогом векторного алгоритму (2.4) при  $\alpha = 0$ , тобто вкрай чутлива до впливу зовнішніх шумів, що не дозволяє розрізнити зміни в сигналі від впливу шумів  $\xi(k)$ . У зв'язку з цим в [132] був введений матричний аналог експоненційно зваженого методу найменших квадратів. Введений алгоритм, незважаючи на високу якість ідентифікації, виявився дуже громіздким з обчислювальної точки зору, що ускладнює його використання в режимі реального часу, коли дані на обробку надходять з високою частотою.

Тим самим, доцільно ввести в розгляд матричний аналог алгоритму (2.4), який може бути записаний у вигляді

$$\left\{ \begin{array}{l} A(k) = A(k-1) + \frac{e_A(k)B^T(k-1)x^T(k)}{\alpha r_A(k-1) + Sp(B^T(k-1)x^T(k)x(k)B(k-1))}, \\ r_A(k) = \alpha r_A(k-1) + Sp(B^T(k-1)x^T(k)x(k)B(k-1)), \\ B(k) = B(k-1) + \frac{x^T(k-1)A(k)e_B(k)}{\alpha r_B(k-1) + Sp(x^T(k-1)A(k)A^T(k)x(k-1))}, \\ r_B(k) = \alpha r_B(k-1) + Sp(B^T(k-1)x^T(k)x(k)B(k-1)). \end{array} \right. \quad (2.12)$$

При  $\alpha = 1$  процедура (2.12) набуває властивостей стохастичної апроксимації, тобто реалізує фільтрацію оброблюваного сигналу, проте вона занадто інерційна, щоб швидко виявити зміни, що відбуваються.

Для виявлення змін в матричному сигналі в [40, 41] було запропоновано використовувати стежачий сигнал, однак, як уже зазначалося, така процедура занадто інерційна в задачах швидкого виявлення змін властивостей.

У зв'язку з цим подібно ансамблю VAR-моделей з різними значеннями параметра забування  $\alpha$  можливо ввести ансамбль MAR-моделей, які можливо настроїти за допомогою (2.12). Далі в розгляд можна ввести сигнал розбіжності між прогнозами

$$T_{qr}(k) = (Sp(\hat{x}_q(k) - \hat{x}_r(k)(\hat{x}_q(k) - \hat{x}_r(k))^T)^{1/2}, \quad (2.13)$$

і сигнали розбіжності між матрицями параметрів, що настроюються

$$\begin{cases} T_{qr}^A(k) = (Sp(A_q(k) - A_r(k))(A_q(k) - A_r(k))^T)^{1/2}, \\ T_{qr}^B(k) = (Sp(B_q(k) - B_r(k))(B_q(k) - B_r(k))^T)^{1/2}, \\ \forall q, r = 1, 2, \dots, S \end{cases} \quad (2.14)$$

при перевищенні контрольних рівнів

$$T_{qr}(k) > \Delta, T_{qr}^A(k) > \Delta^A, T_{qr}^B(k) > \Delta^B \quad (2.15)$$

приймається рішення про те, що в матричному сигналі  $x(k)$  відбулися зміни.

Підводячи певний проміжний підсумок, слід підкреслити, що розглянута задача швидкого виявлення зміни властивостей векторних і матричних нестационарних зашумлених сигналів. В основу покладена ідея ансамблю адаптивних моделей, кожна з яких налаштовується за допомогою власного алгоритму ідентифікації [6]. При цьому для кожної з моделей використовуються алгоритми з різною глибиною пам'яті. Рішення про розладнання приймається за допомогою контролю сигналів розбіжності між реальним процесом і виходами адаптивних моделей.

## 2.2 Виявлення змін в багатовимірних потоках відео на основі послідовної кластеризації

Традиційні online підходи, пов'язані з виявленням змін властивостей сигналів [2, 4, 108, 115-116], як правило, орієнтовані на раптово виникаючі розладнання в той час, як в відеопотоках зміни нерідко відбуваються досить

плавно (затемнення, розчинення, розмивання, накладення, поява кадру в кадрі і тому подібні технічні ефекти). Крім того слід вказати контекстні режисерські ефекти: перехід під час дії (Cutting on Action), монтажні перебивки (Cutaways), паралельний монтаж (Cross-Cut), поєднання (Match cuts). Зазвичай подібні ефекти штучні, за винятком, мабуть, повільного панорамування або плавного семантичного та візуального зміни зображень. У будь-якому випадку виникає можливість того, що окремі підпоследовності відеокадрів практично рівномірно зіставляються до двох і навіть більше сегментів.

У подібних ситуаціях кращими стають методи нечіткої сегментації часових рядів [4, 26, 27], фактично породжують покриття відеоряду, які нерідко в силу своєї обчислювальної громіздкості стають неефективними при високих частотах надходження даних на обробку, властиво відеобробці.

Звичайно ж, в даній ситуації можна було б скористатися нечіткою кластеризацією [29, 33], перш за все, методами рекурентної кластеризації [51, 121], але вони страждають від ефектів «прокляття розмірності» і «концентрації норм», породжуваних досить високою розмірністю оброблюваних сигналів.

Перспективним може виявитися підхід, заснований на непрямому підході до кластеризації послідовностей (indirect sequence clustering approach) [70], який в поєднанні з методами нечіткої рекурентної оптимізації, згаданими вище, дозволить подолати виникаючі проблеми.

Нехай вихідна інформація задана у формі багатовимірної послідовності  $x(1), x(2), \dots, x(k), \dots, x(n)$ ,  $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ , де  $k = 1, 2, \dots, N$  – поточний дискретний час.

Згідно непрямому підходу до кластеризації часових рядів на кластери розбивається не саме сигнал  $x(k)$ ,  $k = 1, 2, \dots, N$ , а деякі його характеристики, такі як середнє, дисперсія, автокореляція. Так, для  $x_i(k)$ ,  $k = 1, 2, \dots, n$  можна ввести рекурсивні оцінки:

– середнього

$$\bar{x}_i(k) = \bar{x}_i(k-1) + \frac{1}{k}(x_i(k) - \bar{x}_i(k-1)); \quad (2.16)$$

– дисперсії

$$\sigma_i^2(k) = \sigma_i^2(k-1) - \frac{1}{k} \left( (x_i(k) - \bar{x}_i(k-1))^2 - \sigma_i^2(k-1) \right); \quad (2.17)$$

– коефіцієнтів автокореляції:

$$r_i(k, \tau) = r_i(k-1, \tau) + \frac{1}{k} \left( (x_i(k) - \bar{x}_i(k))(x_i(k-\tau) - \bar{x}_i(k)) - r_i(k-1, \tau) \right), \quad (2.18)$$

де  $\tau=1, 2, \dots, \tau_{max}$ .

Нескладно ввести в розгляд векторно-матричні аналоги співвідношень (2.16-2.18):

$$\bar{x}(k) = \bar{x}(k-1) - \frac{1}{k}(x(k) - \bar{x}(k-1)), \quad (2.19)$$

де  $\bar{x}(k) = (\bar{x}_1(k), \bar{x}_2(k), \dots, \bar{x}_n(k))^T$ ,

$$R(k, \tau) = R(k-1, \tau) + \frac{1}{k} \left( (x(k) - \bar{x}(k))(x(k-\tau) - \bar{x}(k))^T - R(k-1, \tau) \right) \quad (2.20)$$

(тут  $\tau=0, 1, 2, \dots, \tau_{max}$ ) при цьому діагональні елементи симетричної матриці  $R(k, 0)$  є, по суті, оцінками дисперсій

$$r_{ii}(k, 0) = \sigma_i^2(k),$$

а внедіагональні елементи – коефіцієнтами взаємної кореляції

$$r_{ij}(k, 0), \quad i, j = 1, 2, \dots, n.$$

Для пояснення адаптивних властивостей рекурентних процедур (2.16) - (2.20), контрольовані характеристики зручніше оцінювати на основі модифікації експоненційного згладжування [122]

$$\bar{x}_i^\alpha(k) = \alpha x_i(k) + (1 - \alpha) \bar{x}_i^\alpha(k - 1), \quad 0 < \alpha < 1, \quad (2.21)$$

$$(\sigma_i^\alpha(k))^2 = \alpha \left( x_i(k) + \bar{x}_i^\alpha(k) \right)^2 + (1 - \alpha) \left( \sigma_i^\alpha(k - 1) \right)^2, \quad (2.22)$$

$$r_i^\alpha(k, \tau) = \alpha \left( x_i(k) - \bar{x}_i^\alpha(k) \right) \left( x_i(k - \tau) - \bar{x}_i^\alpha(k) \right) + (1 - \alpha) r_i^\alpha(k - 1, \tau), \quad (2.23)$$

де  $\alpha = 2 / (L + 1)$  – вектор забування, що забезпечує згладжування на ковзному вікні, що містить  $L$  останніх спостережень  $x_i(k), x_i(k - 1), \dots, x_i(k - L + 1)$ .

Таким чином, у відповідність кожній компоненті часового ряду  $x_i(k)$  може бути поставлений  $(2 + \tau_{max}) \times 1$ -вектор ознак

$$\tilde{x}_i(k) = \left( \bar{x}_i^\alpha(k), (\sigma_i^\alpha(k))^2, r_i^\alpha(k, 1) \dots r_i^\alpha(k, \tau_{max}) \right)^T,$$

який і є об'єктом кластеризації.

Для обробки  $(n + 1)$ -вимірному сигналу для оцінки вектора середніх  $\bar{x}^\alpha(k)$  і автокореляційної матриці  $R^\alpha(k, \tau)$  можна також використовувати процедуру експоненційного згладжування виду [41]

$$\bar{x}^\alpha(k) = A x(k) + (I - A) \bar{x}^\alpha(k - 1), \quad (2.24)$$

тут  $A = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$ ,  $I - (n \times n)$  – одинична матриця,

$$R_i^\alpha(k) = A \left( \left( x(k) - \bar{x}^\alpha(k) \right) \left( x(k - \tau) - \bar{x}_i^\alpha(k) \right)^T \right)_i + (I - A) R_i^\alpha(k - 1) \quad (2.25)$$

тут індекс  $i = 1, 2, \dots, n$  – позначає номер стовпця відповідної матриці.

Використання співвідношень (2.24), (2.25) веде до того, що необхідно постійно контролювати  $0,5(n^2 + n)(\sigma_{max} + 1) + n$  параметрів, що ускладнює обробку сигналів в реальному часі. У зв'язку з цим пропонується в подальшому для контролю і відповідно online кластеризації використовувати поточні значення  $((2 + \tau_{max}) \times n)$ -матриці

$$\tilde{x}(k) = (\tilde{x}_1(k), \tilde{x}_2(k), \dots, \tilde{x}_n(k)),$$

елементи якої на кожному кроці часу множаться за допомогою співвідношень (2.21) (2.23).

Тепер можна звернутися до виявлення змін властивостей компонент часового ряду.

Оскільки зміни, що виникають в процесі обробки багатовимірною сигналу  $\tilde{x}(k)$ , можуть носити як стрибкоподібний, так і плавний характер, для їх аналізу можуть бути використані алгоритми нечіткої кластеризації (з різними значеннями фаззифікаторів), засновані на використанні нечітких цільових функцій [29].

Отже, нехай об'єктом аналізу є  $i$ -я компонента матричного сигналу  $\tilde{x}(k)$ , що представлена у вигляді  $(\alpha + \tau_{max}) \times 1$  вектору

$$\tilde{x}_i(k) = \left( \bar{x}_i^\alpha(k), (x_i^\alpha(k))^2, r_i^\alpha(k, 1) \dots r_i^\alpha(k, \tau_{max}) \right)^T,$$

а в якості цільової функції використовується стандартний критерій самонав-



чання

$$E(\mu(\tilde{x}_i(k), C_i(l))) = \sum_{k=1}^N \sum_{l=1}^M \mu^\beta(\tilde{x}_i(k), C_i(l)) D^2(\tilde{x}_i(k), C_i(l)) \quad (2.26)$$

при обмеженнях

$$\sum_{l=1}^M \mu(\tilde{x}_i(k), C_i(l)) = 1, \quad \forall k = 1, 2, \dots, N, \quad (2.27)$$

$$0 < \sum_{k=1}^N \mu(\tilde{x}_i(k), C_i(l)) \leq N, \quad \forall l = 1, 2, \dots, M. \quad (2.28)$$

Тут  $\mu(\tilde{x}_i(k), C_i(l))$  – рівень нечіткої належності вектору  $\tilde{x}_i(k)$  до  $i$ -го кластеру,  $C_i(l)$  – центроїд цього кластера, що оцінюється,  $\beta$  – ненегативний параметр (фазифікатор),  $D^2(\tilde{x}_i(k), C_i(l))$  – відстань між векторами  $\tilde{x}_i(k)$  та  $C_i(l)$  в прийнятій метриці (в найпростішому і найбільш поширеному випадку – евклідової). В цьому випадку цільова функція (2.26) може бути переписана у вигляді

$$E(\mu(\tilde{x}_i(k), C_i(l))) = \sum_{k=1}^N \sum_{l=1}^M \mu^2(\tilde{x}_i(k), C_i(l)) \|\tilde{x}_i(k) - C_i(l)\|^2. \quad (2.29)$$

Оптимізація цільової функції (2.26) за допомогою стандартної техніки нелінійного програмування веде до результату, відомому як метод кластеризації нечітких  $C$ -середніх

$$\begin{cases} \mu(\tilde{x}_i(k), C_i(l)) = \frac{\|x_i(k) - C_i(l)\|^{-2}}{\sum_{j=1}^m \|\tilde{x}_i(k) - C_i(j)\|^{-2}}, \\ C_i(l) = \frac{\sum_{k=1}^N \mu^2(\tilde{x}_i(k), C_i(l)) x_i(k)}{\sum_{k=1}^N \mu^2(\tilde{x}_i(k), C_i(l))}. \end{cases} \quad (2.30)$$

Необхідно зазначити, що метод (2.30) описує пакетну процедуру обробки інформації, коли вся вибірка, що підлягає аналізу задана і не змінюється в процесі роботи. Зрозуміло, що подібна процедура не може бути використана для виявлення змін в online режимі.

Для роботи в послідовному режимі доцільно скористатися адаптивною версією (2.30) [39], що дозволяє в реальному часі вирішувати задачу кластеризації-сегментації по ходу надходження даних на обробку. У прийнятих позначеннях процедура кластеризації може бути записана у вигляді

$$\begin{cases} \mu(\tilde{x}_i(k), C_i(l, k-1)) = \frac{\|\tilde{x}_i(k) - C_i(l, k-1)\|^{-2}}{\sum_{j=1}^m \|\tilde{x}_i(k) - C_i(j, k-1)\|^{-2}}, \\ C_i(l, k) = C_i(l, k-1) + \eta(k) \mu^\beta(\tilde{x}_i(k), C_i(l, k-1)) (\tilde{x}_i(k) - C_i(l, k-1)), \end{cases} \quad (2.31)$$

де  $0 < \eta(k) < 1$  – параметр кроку пошуку, обраний зазвичай з емпіричних міркувань.

При  $\beta = 2$  процедура (2.31) призведе до тих самих результатів, що і пакетний алгоритм (2.30), однак при цьому дозволяє обробляти послідовність даних, що надходять на обробку в реальному часі. При  $\beta = 0$  (2.31) набуває форму алгоритму самонавчання Кохонена [95], що є одним з найбільш популярних при вирішенні задач кластеризації. Зауважимо також, що  $\beta = 0$ ,  $\eta(k) = k^{-1}$  відповідає рекурентній версії методу  $k$ -середніх [29], широко використовуваного в задачах обробки візуальної інформації.

Узагальнимо викладене для виявлення змін властивостей багатовимір-

ної послідовності.

Оскільки в процесі виявлення, описаному вище, одночасно реалізується  $n$  одновимірних процедур типу (2.31), з обчислювальної точки зору ефективніше обробляти відразу всю матрицю параметрів  $\tilde{x}(k)$  розмірності  $((2 + \tau_{max}) \times n)$ . Для цього доцільно скористатися матричною модифікацією методу нечітких С-середніх

$$E(\mu(\tilde{x}(k), C(l))) = \sum_{k=1}^N \sum_{l=1}^m \mu^2(\tilde{x}(k), C(l)) Sp(\tilde{x}(k) - C(l))(\tilde{x}(k) - C(l))^T, \quad (2.32)$$

де замість евклідової метрики використовується метрика норми Фробениуса. Також в розгляд вводиться  $m$  матриць-центроїдів  $C(l), l = 1, 2, \dots, m$ .

Мінімізація (2.32) при обмеженнях (2.27), (2.28) веде до аналітичного виразу

$$\left\{ \begin{array}{l} \mu(\tilde{x}(k), C(l)) = \frac{\left( Sp(\tilde{x}(k) - C(l))(\tilde{x}(k) - C(l))^T \right)^{-1}}{\sum_{j=1}^m \left( Sp(\tilde{x}(k) - C(j))(\tilde{x}(k) - C(j))^T \right)^{-1}}, \\ C(l) = \frac{\sum_{k=1}^N \mu^2(\tilde{x}(k), C(l)) \tilde{x}(k)}{\sum_{k=1}^N \mu^2(\tilde{x}(k), C(l))}. \end{array} \right. \quad (2.33)$$

Вочевидь, що вираз (2.33) не може бути використано для вирішення даної задачі в online режимі. У цьому випадку доцільно скористатися матричною модифікацією рекурентної процедури (2.31), що має в даному випадку вид

$$\left\{ \begin{array}{l} \mu(\tilde{x}(k), C(l, k-1)) = \frac{\left( Sp(\tilde{x}(k) - C(l, k-1))(\tilde{x}(k) - C(l, k-1))^T \right)^{-1}}{\sum_{j=1}^m \left( Sp(\tilde{x}(k) - C(j, k-1))(\tilde{x}(k) - C(j, k-1))^T \right)^{-1}}, \\ C(l, k) = C(l, k-1) + \eta(k) \mu^\beta(\tilde{x}(k), C(l, k-1))(\tilde{x}(k) - C(l, k-1)), \end{array} \right. \quad (2.34)$$

де параметр кроку  $\eta(k)$  вибирається з тих же міркувань, що і в (2.31).

Можна також зауважити, що при  $\beta=0$  (2.34) набуває форму чіткого матричного алгоритму самонавчання для online рішень задачі кластеризації виду [8]

$$C(l, k) = C(l, k-1) + \eta(k)(\tilde{x}(k), C(l, k-1)),$$

що є, по суті, матричним варіантом правила самонавчання за принципом Т. Кохонена «Переможець отримує все» [95].

Резюмуючи, відзначимо, що розглянута задача виявлень змін розладнань в багатовимірних (матричних) потоках даних, що послідовно надходять на обробку [107]. Введено online процедури нечіткої кластеризації, що дозволяють вирішувати завдання як для повільних, так і стрибкоподібних змін.

### 2.3 Виявлення змін властивостей багатовимірних послідовностей різної довжини

Говорячи про структурування часових рядів, слід пам'ятати, що традиційні методи кластеризації виявляються зазвичай не ефективними. Це пояснюється тим, що спостереження строго впорядковані в часі, і ніяке перемішування даних абсолютно неприпустимо. Властивості часових рядів при кластеризації в деяких ознакових просторах можуть змінюватися в часі так, що одна реалізація на різних часових ділянках може мати різні властивості. Але найскладніше – це те, що оброблювані послідовності можуть мати різну

довжину, що в принципі не дозволяє використовувати традиційні метрики, використовувані в кластерному аналізі. Розглянемо методи кластеризації-сегментації виявлення змін властивостей багатовимірних (векторних і матричних) часових рядів різної довжини, чисельно простих і таких, що дозволяють скоротити час обробки інформації, тобто забезпечують передумови для online аналізу відеопотоків.

Проведемо аналіз оцінки схожості часових рядів на основі динамічної часової деформації (Dynamic Time Warping – DTW).

У задачах кластеризації часових рядів поширення набули методи, що базуються не на використанні тієї чи іншої метрики [9, 93], а на мірі схожості, яка описує форму цих рядів [29]. В рамках цього підходу найбільшого поширення набула міра динамічної часової деформації (Dynamic Time Warping) [37, 90, 162], що дозволяє порівнювати між собою послідовності різної довжини.

Суть DTW підходу полягає в наступному: нехай є дві одновимірні послідовності різної довжини:  $X = \{x(1), x(2), \dots, x(t), \dots, x(N)\}$  і  $Y = \{y(1), y(2), \dots, y(l), \dots, y(M)\}$ ,  $N \neq M$ , між якими необхідно визначити «відстань», що дозволяє в подальшому вирішити задачу кластеризації. В розгляд вводиться  $[N + M]$  – матриця, елементами якої є відстані (зазвичай евклідові  $d(x(k), y(l))$ ) між усіма відліками часових рядів  $X$  і  $Y$ . Далі вводиться, так званий, «викривлений» шлях (Warping Path)  $W = \{w_1, w_2, \dots, w_L\}$  (тут  $\max(N, M) \leq L < M + N + 1$ ,  $w_q = (d(x(k), y(l)))_q$ ,  $q = 1, \dots, L$ ) з допомогою яких і встановлюється схожість між рядами  $X$  і  $Y$ .

На викривлений шлях накладається ряд умов:

– крайові умови  $w_1 = (d(x(1), y(1)))_1$  і  $w_L = (d(x(N), y(M)))_L$ , що означають, що викривлений шлях починається і закінчується в діагонально протилежних кутках матриці відстаней;

– умови безперервності, що позначають, що для деякого

$w_q = (d(x(k), y(l)))_q$  справедливі співвідношення  $w_{q-1} = (d(x(k'), y(l')))_{q-1}$ ,  
 $k - k' \leq 1, l - l' \leq 1$ ;

– умови монотонності, що позначають, що для  $w_q = (d(x(k), y(l)))_q$   
справедливі співвідношення  $w_{q-1} = (d(x(k'), y(l')))_{q-1}, k - k' \geq 0, l - l' \geq 0$ .

В якості міри схожості вибирається шлях, що мінімізує цільову функцію

$$DTW(x, y) = \min_W \left\{ \frac{1}{L} \sum_{l=1}^L d(x(k), y(l)) \right\} \quad (2.35)$$

тобто зв'язує всі точки аналізованих рядів.

Власне пошук оптимального шляху заснований на ідеях динамічного програмування за допомогою рекурентного співвідношення

$$D(k, l) = d(x(k), y(l)) + \min \{ D(k, l-1), D(k-1, l), D(k-1, l-1) \} \quad (2.36)$$

де  $D(k, l)$  – так звані накопичені (cumulative) відстані між спостереженнями  $x(k)$  та  $y(l)$ . Накопичена відстань  $D(N, M)$  і є динамічна часова відстань, тобто

$$DTW(x, y) = D(N, M). \quad (2.37)$$

Даний підхід був розглянутий в [26] на разі багатовимірних часових рядів, де компонент  $x(k), y(l) - (n \times 1)$  – векторні спостереження. У цьому випадку в якості відстані між  $x(k), y(l)$  використовується евклідова норма

$$d^2(x(k), y(l)) = \|x(k) - y(l)\|^2,$$

при цьому вирази (2.35) - (2.37) також залишаються незмінними.

Використання квадратичних норм в задачах обробки часових рядів обмежується тим, що отримані результати дуже чутливі до викидів і різких змін. У зв'язку з цим в [88] була введена динамічна часова деформація на основі похідних (Derivative Dynamic Time Warping – DDTW), де замість відстані  $d(x(k), y(l))$  використовуються оцінки на основі різниць  $d(\tilde{x}(k), \tilde{y}(l))$ , де

$$\tilde{x}(k) = \frac{(x(k) - x(k-1)) + ((x(k+1) - x(k-1))/2)}{2}, \quad (2.38)$$

$$\tilde{y}(k) = \frac{(y(k) - y(k-1)) + ((y(k+1) - y(k-1))/2)}{2}. \quad (2.39)$$

Зрозуміло, що всі перетворення, описані вище, справедливі і для скалярних, векторних та матричних сигналів (2.38), (2.39).

І хоча DTW в загальному випадку не є метрикою, а тільки мірою схожості, в [128] підкреслюється, що з точки зору саме кластеризації ця міра є найкращою в сенсі одержуваної точності. Основним же недоліком DTW є висока складність обчислювальної реалізації, особливо при обробці досить «довгих» часових рядів. У подібних випадках використання DTW є неефективним.

Розглянемо кластеризацію часових рядів на основі ітеративної динамічної часової деформації (IDTW).

Громіздкість стандартного DTW змусила шукати альтернативні підходи до оцінки близькості часових рядів, при цьому одним з найбільш ефективних прийомів є ітеративна динамічна часова деформація [50, 76]. В її основі лежить ідея редукції початкових послідовностей, коли ділянки цих послідовностей замінюються відповідними середніми значеннями. При цьому початкові послідовності розбиваються на приблизно рівні інтервали, при цьому кількість цих інтервалів для рядів різної довжини однаково.

Отже, нехай є дві послідовності  $X$  і  $Y$  довжиною  $N$  і  $M$  спостере-

жень відповідно, при цьому в загальному випадку ряди  $X$  і  $Y$  можуть бути скалярами, векторами або матричними рядами.

У кожному з рядів виділимо спостереження  $x(1), x(\lfloor N/4 \rfloor), x(\lfloor N/2 \rfloor), x(\lfloor 3N/4 \rfloor), x(N), y(1), y(\lfloor M/4 \rfloor), y(\lfloor M/2 \rfloor), y(\lfloor 3M/4 \rfloor), y(M)$  ( $\lfloor \cdot \rfloor$  – ціла частина числа) і обчислимо середні на ділянках

$$\begin{cases} x(1) - x(\lfloor N/4 \rfloor), \\ x(\lfloor N/4 \rfloor + 1) - x(\lfloor N/2 \rfloor), \\ x(\lfloor N/2 \rfloor + 1) - x(\lfloor 3N/4 \rfloor), \\ x(\lfloor 3N/4 \rfloor + 1) - x(N), \end{cases} \quad \begin{cases} y(1) - y(\lfloor M/4 \rfloor), \\ y(\lfloor M/4 \rfloor + 1) - y(\lfloor M/2 \rfloor), \\ y(\lfloor M/2 \rfloor + 1) - y(\lfloor 3M/4 \rfloor), \\ y(\lfloor 3M/4 \rfloor + 1) - y(M). \end{cases}$$

Таким чином, в розгляд вводяться скорочені послідовності  $\bar{x}(1), \bar{x}(2), \bar{x}(3), \bar{x}(4); \bar{y}(1), \bar{y}(2), \bar{y}(3), \bar{y}(4)$  між якими нескладно оцінити схожість на основі стандартної DTW з  $(4 \times 4)$  – матрицею відстаней. Зрозуміло, що такі скорочені ряди можуть бути кластеризовані за допомогою будь-якого з відомих алгоритмів.

Природно, що така апроксимація є занадто грубою, а якість кластеризації природно виявляється низьким. Тому на наступному етапі кожна з сформованих ділянок ділиться навпіл, тобто наприклад, замість  $x(1) - x(\lfloor N/4 \rfloor)$  формується два інтервали  $x(1) - x(\lfloor N/8 \rfloor), x(\lfloor N/8 \rfloor + 1) - x(\lfloor N/4 \rfloor)$  і т.п., на яких вводяться свої середні  $\bar{\bar{x}}(1), \bar{\bar{x}}(2), \dots, \bar{\bar{x}}(8); \bar{\bar{y}}(1), \bar{\bar{y}}(2), \dots, \bar{\bar{y}}(8)$  схожість між якими оцінюється на підставі  $(8 \times 8)$  – матрицею відстаней.

Зрозуміло, що на наступному етапі виникають  $(16 \times 16), (32 \times 32)$  і т.п. матриці відстаней так, що задача істотно ускладнюється. Тому з обчислювальної точки зору, набагато простіше організувати процес дроблення початкового ряду не на 4, 8, 16, ... ділянок, а, наприклад, 4, 5, 6 ... інтервалів, що різко спрощує задачу.



Сформувавши масив скорочених рядів, кожен з яких містить однакову кількість відліків, нескладно вирішити далі задачу їх кластеризації на основі запропонованої модифікації методу  $k$ -середніх, алгоритм якого тоді виглядає наступним чином.

Отже, нехай є масив часових послідовностей  $x_1, x_2, \dots, x_q, \dots, x_Q$ , кожна з яких містить  $N_1, N_2, \dots, N_q, \dots, N_Q$  спостережень.

1. Після першого етапу редукції отримуємо ті ж  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_Q$  при цьому кожен з них містить тільки по 4 відліку.

2. Далі досить довільним чином формується  $m$  центроїд  $C_1, C_2, \dots, C_m$  в формі послідовностей, що містять таке ж число відліків, що і в редукованих рядах  $\bar{x}_q$ .

3. На наступному етапі кожен  $\bar{x}_q$  приписується одному з центроїдів за правилом  $\bar{x}_q \in Cl_j$  якщо  $IDTW(\bar{x}_q, c_j) < IDTW(x_q, c_l)$  для  $q = 1, 2, \dots, Q; j, l = 1, 2, \dots, m$ .

4. Після цього проводимо перерахунок початкових центроїдів згідно виразу  $C_j = \frac{1}{N_j} \sum_{\bar{x}_q \in Cl_j} \bar{x}_q$ , тобто виробляється усереднення всіх скорочених рядів по кожному з кластерів. При цьому може виникнути ситуація, коли центроїди, що сформувалися, мають більшу кількість відліків в порівнянні з рядами, що обробляються. У цьому випадку проводиться редукція самих центроїдів за тими ж правилами, що застосовуються для початкових рядів.

5. Процес уточнення центроїдів проводиться до тих пір, поки норма (евклідова або сферична) відхилення уточнених центроїдів  $C_j^{new}$  від тих, що були отримані минулої ітерації алгоритму  $C_j^{old}$  для всіх  $j$  не буде перевищувати деякого заданого порогу  $\varepsilon$ .

У цьому випадку алгоритм або закінчує свою роботу, або виробляється глибша редукція часових рядів з великою кількістю відліків.

Як видно, запропонований підхід набагато простіший в реалізації в порівнянні з відомими методами кластеризації часових рядів різної довжини. Перейдемо тепер до адаптивної редукції часових рядів.

Розглянута вище процедура редукції часових рядів містить досить великий рівень суб'єктивізму і жодним чином не пов'язана з властивостями аналізованих послідовностей. У зв'язку з цим більш ефективно замінити ряди їх середніми на стаціонарних ділянках, коли їх статистичні властивості практично не змінюються. При цьому доцільно скористатися методами виявлення зміни середніх значень, серед яких, одним з найбільш ефективних є метод, заснований на критерії Пейджа-Хінклі [115].

Отже, нехай є послідовність  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  і передбачається, що її середнє значення має значення  $\bar{x}_1$  при  $1 \leq k \leq r$  та при  $r \leq k$ . Для того, щоб скористатися критерієм Пейджа-Хінклі, необхідно задати мінімальну величину стрибка  $\delta$ , що визначає «зрив» стаціонарності контрольованого процесу, але оскільки знак цього стрибка апріорі невідомий, то процедура виявлення змін повинна містити дві частини. Перша частина процедури призначена для виявлення зменшення середнього значення і може бути записана у вигляді

$$\begin{cases} T_-(0)=0, \\ T_-(k)=\sum_{\tau=1}^k (x(\tau) - \bar{x}_1 + 0.5\delta) \\ T_-^*(k)=\max_{0 \leq \tau \leq k} T_-(\tau). \end{cases} \quad (2.40)$$

Рішення про те, що контрольоване середнє істотно зменшилося, приймається в момент  $k = r$ , коли виконується нерівність

$$T_-^*(r) - T_-(r) \geq \chi \quad (2.41)$$

де  $\chi$  – деякий поріг, обраний з емпіричних міркувань.

Процедура для виявлення збільшення середнього має вигляд

$$\begin{cases} T_+(0)=0, \\ T_+(k)=\sum_{\tau=1}^k (x(k)-\bar{x}_1-0.5\delta) \\ T_+^*(k)=\max_{0\leq\tau\leq k} T_+(\tau). \end{cases} \quad (2.42)$$

Рішення про збільшення середнього приймається в момент  $k = r$ , коли виконується нерівність

$$T_+^*(r) - T_+(r) \geq \chi \quad (2.43)$$

Для того, щоб реалізувати процес контролю в послідовному режимі, можна скористатися адаптивною модифікацією (2.40) - (2.43), що має вигляд

$$\begin{cases} T_-(0)=0, \\ T_-(k)=T_-(k-1)+x(k)-\bar{x}_1+0.5\delta \\ T_-^*(k)=T_-(k)+0.5\left(1-\frac{T_-(k)-T_-(k-1)}{\sqrt{(T_-(k)-T_-(k-1))^2}}(T_-(k-1)-T_-(k))\right), \\ T_-^*(k)-T_-(k) \geq \chi \end{cases} \quad (2.44)$$

та

$$\begin{cases} T_+(0)=0, \\ T_+(k)=T_+(k-1)+x(k)-\bar{x}_1-0.5\delta, \\ T_+^*(k)=T_+(k)+0.5\left(1+\frac{T_+(k)-T_+(k-1)}{\sqrt{(T_+(k)-T_+(k-1))^2}}\right)(T_+(k-1)-T_+(k)), \\ T_+(k)-T_+^*(k)\geq\chi \end{cases} \quad (2.45)$$

Таким чином, використана процедура адаптивної редукції часових рядів (2.40) - (2.45) дозволяє підвищити ефективність процесу ітеративної динамічної часової деформації (IDTW) і тим самим підвищити якість кластеризації часових рядів різної довжини.

#### Висновки по другому розділу

1. Швидке виявлення змін властивостей векторних і матричних нестационарних зашумлених сигналів на основі ансамблю адаптивних моделей з власними алгоритмами ідентифікації з різною глибиною пам'яті створює передумови для прийняття в реальному часі рішення про розладнання на базі контролю сигналів розбіжності між реальним процесом і виходами адаптивних моделей.

2. Введені online процедури нечіткої кластеризації дозволяють вирішувати завдання виявлення розладнань в багатовимірних потоках даних як для повільних, так і стрибкоподібних змін властивостей, що стосовно відео забезпечує високу релевантність темпорального аналізу відредагованих відеопотоків.

3. Виявлення змін властивостей багатовимірних послідовностей різної довжини на основі ітеративної динамічної часової деформації, в основі якої лежить адаптивна редукція часових рядів, дозволяє отримувати модифікації алгоритмів кластеризації векторних і матричних послідовностей з метою їх реферування.

Список використаних джерел у даного розділі наведено у повну списку

використаних джерел під номерами: [2, 4, 6-10, 12, 14-16, 26-27, 29, 32-33, 37, 39-43, 47, 66, 50-51, 70, 76-77, 83, 85, 88, 90, 93, 95, 102, 107-108, 115-116, 121-122, 125, 128, 132, 135,138, 145, 154, 162].

### 3 МЕТОДИ ФРАГМЕНТНОЇ СЕГМЕНТАЦІЇ ВІДЕО В ЗАДАЧАХ ІНФОРМАЦІЙНОГО ПОШУКУ

Адекватність реферування, анотування, сумаризації погано структурованих даних великих обсягів, зумовлюючи точність і повноту інформаційного пошуку, справляють істотний вплив на формальну, а в підсумку і змістовну релевантність. При обробці відеопотоків фінальна пертинентність спирається на аналіз областей візуальної уваги і інтересу, які фактично є деякими локальними поля зору. Інакше кажучи, фрагментна обробка відеоданих дозволяє знижувати (а іноді і повністю усувати) семантичний розрив між ознаками низького рівня, що обчислюються на кожному кадрі, і вимогами високого рівня інтерпретації відео в часі на основі таких понять, як сценарії, події і т.п. Але на сучасному етапі головна перевага фрагментної обробки полягає в тому, що вона стійка до цілого ряду темпоральних деформацій відео - частковим затулянням об'єктів, варіацій ракурсу внаслідок зміни взаємного розташування і орієнтації об'єктів і датчика, мінливості умов спостереження (освітленість, контраст і ін.). Фрагмент, який є околom деякої особливої точки, можливо представити або вектором ознак, наприклад, гістограмою яскравостей або напрямків градієнтів, або матрицею у вигляді зображення або результату, взагалі кажучи, довільної просторової згортки. Фрагментна сегментація відео, тим самим, дає можливість отримувати багатовимірні часові ряди, що впливають на зближення формальної і семантичної релевантності.

Розділ присвячений побудові моделі матричних часових рядів на основі ітеративної динамічної часової деформації, кластеризації часових рядів на основі самоорганізованих карт Кохонена, ієрархічної кластеризації рядів різної довжини, включаючи випадки перетинних класів. Також обговорюється налаштування параметрів адаптивної матричної моделі та виявлення змін властивостей матричної послідовності.

### 3.1 Адаптивна ітеративна динамічна часова деформація при сегментації багатовимірних послідовностей

Сегментація-кластеризація часових рядів, індукованих відеопотоками, істотно ускладнюється, коли підлягають аналізу послідовності мають різну довжину, що приводить до неможливості використання метрик, що традиційно використовуються в кластерному аналізі. У подібних ситуаціях зазвичай застосовують заходи подібності-схожості, серед яких в першу чергу, можна виділити DTW – так звану динамічну часову деформацію [15, 36-37, 50, 88-90, 128, 162], що дозволяє оцінити «відстань» між довільними вибірками з відеоряду.

Нехай задані дві одновимірні послідовності різної довжини  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  і  $Y = \{y(1), y(2), \dots, y(k), \dots, y(M)\}$ ,  $N \neq M$ , між якими необхідно оцінити відстань, придатну для вирішення завдання кластеризації за допомогою тих чи інших відомих підходів. Для цього в розгляд вводиться  $(N \times M)$  – матриця відстаней (в довільній метриці) з елементами  $d(x(k), y(l))$   $k = 1, 2, \dots, N; l = 1, 2, \dots, M$  між усіма елементами вибірок. На цій матриці розраховуються так званий «викривлений шлях», сформований послідовністю  $W = \{w_1, w_2, \dots, w_L\}$ ,  $w_q = d(x(k), y(l))_q$ ,  $q = 1, 2, \dots, L$ ,  $\max(N, M) \leq L < M + N + 1$ , який є основою для оцінки подібності між рядами  $X$  та  $Y$ . Як «відстань» між рядами виступає «викривлений шлях», який доставляє мінімум критерію

$$DTW(x, y) = \min_W \left\{ \frac{1}{L} \sum_{q=1}^L d(x(k), y(l))_q \right\}. \quad (3.1)$$

Мінімізація критерію (3.1) проводиться за допомогою процедури динамічного програмування на основі рекурентного співвідношення  $D(k, l) = d(x(k), y(l)) + \min\{D(k, l-1), D(k-1, l), D(k-1, l-1)\}$ , де  $D(k, l)$  – куму-

лятивна відстань між елементами послідовностей  $x(k), y(l)$ . При цьому  $DTW(x, y) = D(N, M)$ .

Цей підхід в [36] був поширений на багатовимірні послідовності  $x(k), y(l) \in \mathbb{R}^N$ . При цьому в якості відстані між спостереженнями використовувалася евклідова метрика

$$d(x(k), y(l)) = \|x(k) - y(l)\|^2. \quad (3.2)$$

У разі обробки матричних сигналів, зокрема, відеоданих, коли  $x(k) = \{x_{i_1, i_2}(k)\} \in \mathbb{R}^{n \times v}$ ,  $y(l) = \{y_{i_1, i_2}(l)\} \in \mathbb{R}^{n \times v}$ , доцільно ввести в розгляд норму Фробеніуса

$$d(x(k), y(l)) = \left( \text{Sp}(x(k) - y(l))(x(k) - y(l))^T \right)^{1/2}, \quad (3.3)$$

узагальнюючу метрику (3.2).

Основні проблеми, пов'язані з використанням  $DTW$ , виникають в задачах, пов'язаних з обробкою довгих часових рядів, що характерно для аналізу відеопотоків. В цьому випадку динамічне програмування в умовах великих розмірностей нашоухується на значні обчислювальні труднощі, що робить цей підхід неієвим.

У зв'язку з цим, доцільною представляється розробка модифікацій  $DTW$ -підходів, які були б орієнтовані на обробку часових рядів значної довжини. Розглянемо побудову математичної моделі часових рядів на основі ітеративної динамічної часової деформації.

Обчислювальну складність можна ієтотно скоротити, скориставшись ієдеями ітеративної динамічної часової деформації [50]. Її суть полягає в редукції аналізованих рядів за допомогою тих чи інших оцінок окремих інтервалів послідовностей. При цьому, якщо задано два ряди  $X$  і  $Y$ , що містять



$N$  і  $M$  спостережень відповідно, з них виділяються, наприклад, по п'ять точок:

$$x(1), x\left(\left\lfloor \frac{N}{4} \right\rfloor\right), x\left(\left\lfloor \frac{N}{2} \right\rfloor\right), x\left(\left\lfloor \frac{3N}{4} \right\rfloor\right), x(N) \quad \text{і}$$

$y(1), y\left(\left\lfloor \frac{M}{4} \right\rfloor\right), y\left(\left\lfloor \frac{M}{2} \right\rfloor\right), y\left(\left\lfloor \frac{3M}{4} \right\rfloor\right), x(M)$  (тут  $\lfloor \odot \rfloor$  – ціла частина числа),

після чого на інтервалах

$$x(1) - x\left(\left\lfloor \frac{N}{4} \right\rfloor\right), x\left(\left\lfloor \frac{N}{4} \right\rfloor + 1\right) - x\left(\left\lfloor \frac{N}{2} \right\rfloor\right), x\left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) - x\left(\left\lfloor \frac{3N}{4} \right\rfloor\right), x\left(\left\lfloor \frac{3N}{4} \right\rfloor + 1\right) - x(N),$$

$$y(1) - y\left(\left\lfloor \frac{M}{4} \right\rfloor\right), y\left(\left\lfloor \frac{M}{4} \right\rfloor + 1\right) - y\left(\left\lfloor \frac{M}{2} \right\rfloor\right), y\left(\left\lfloor \frac{M}{2} \right\rfloor + 1\right) - y\left(\left\lfloor \frac{3M}{4} \right\rfloor\right), y\left(\left\lfloor \frac{3M}{4} \right\rfloor + 1\right) - x(M)$$

розраховуються середні  $\bar{x}(1), \bar{x}(2), \bar{x}(3), \bar{x}(4), \bar{y}(1), \bar{y}(2), \bar{y}(3), \bar{y}(4)$ , які і утворюють скорочені послідовності. Ці спостереження формують  $(4 \times 4)$ -матрицю відстаней, на основі якої розраховується кумулятивна відстань

$$DTW(x, y) = DTW(4, 4). \quad (3.4)$$

Якщо оцінки, одержані за допомогою (3.4), не задовільні за точністю, на основі редукованих рядів проводиться розбиття інтервалів навпіл, потім формуються середні  $\bar{x}(1), \bar{x}(2), \dots, \bar{x}(8), \bar{y}(1), \bar{y}(2), \dots, \bar{y}(8)$ , і розраховується  $DTW(8, 8)$ . Процес редукції рядів проводиться до досягнення необхідного результату, при цьому кількість інтервалів зростає в геометричній прогресії.

З обчислювальної точки зору набагато зручніше нарощувати число інтервалів в арифметичній прогресії, наприклад,  $g = 4, 5, \dots, R \leq \min\{N, M\}$ . Таким чином, у відповідність кожному ряду ставиться послідовність, яка містить  $g + 1$  відліків.

Повертаючись до задачі кластеризації часових послідовностей  $x_1, x_2, \dots, x_q, \dots, x_Q$ , кожна з яких містить  $N_1, N_2, \dots, N_q, \dots, N_Q$  спостережень відповідно, можна звести її до задачі обробки послідовностей  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_q, \dots, \bar{x}_Q$ , при цьому кожна послідовність містить  $g + 1$  відліків.

Такий підхід ефективний тільки в ситуації, коли властивості рядів не

змінюються на кожному з сформованих інтервалів. Оскільки в реальних задачах обробки відео це далеко не завжди так, кластеризація на основі середніх може виявитися неефективною.

З позицій загальної задачі кластеризації часових рядів розглянутий підхід відноситься до так званого підходу, заснованого на ознаках. У завданнях обробки нестационарних послідовностей більш перспективним представляється підхід, заснований на математичних моделях часових рядів [18, 20, 102]. Необхідно підкреслити, що цей підхід розвинений тільки для послідовностей однакової довжини. У зв'язку з цим вважаємо за доцільне об'єднати даний підхід з ітеративною динамічною часовою деформацією.

Отже, нехай є набір часових послідовностей різної довжини, кожна з яких розбита на  $p$  інтервалів, а в кожному з рядів  $x_q$  ці інтервали можуть містити різну кількість спостережень.

Позначимо  $x_{qr}(\tau)$  спостереження  $q$ -го ряду в  $r$ -м інтервалі, де  $\tau$  – номер даного спостереження в цьому інтервалі, який змінюється від одиниці до числа, що визначає довжину цього інтервалу.

У відповідність кожному з спостережень поставимо лінійну адаптивну модель виду

$$x_{qr}(\tau) = a_{qr} + b_{qr}\tau = A_{qr}^T \bar{\tau}, \quad (3.5)$$

$$\text{де } A_{qr}^T = (a_{qr}, b_{qr})^T, \bar{\tau} = (1, \tau)^T.$$

Налаштування параметрів моделі проводиться за допомогою рекурентного методу найменших квадратів:

$$\begin{cases} A_{qr}^T(\tau) = A_{qr}^T(\tau-1) + P_{qr}(\tau)((x_{qr}(\tau) - A_{qr}^T(\tau-1)\bar{\tau})\bar{\tau}), \\ P_{qr}(\tau) = P_{qr}(\tau-1) - \frac{P_{qr}(\tau-1)\bar{\tau}\bar{\tau}^T P_{qr}(\tau-1)}{1 + \bar{\tau}^T P_{qr}(\tau-1)\bar{\tau}}. \end{cases} \quad (3.6)$$

В результаті такої обробки даних кожному  $q$ -му ряду  $x_q$  ставиться у відповідність  $(2g \times 1)$ -вектор  $A_q$ , на основі якого в подальшому і проводиться власне кластеризація. Зауважимо також, що якщо на ділянках розбивки ряди стаціонарні, то параметри  $b_{qr}$  близькі до нуля, тобто повертаємося до ітеративної динамічної часової деформації.

У разі багатовимірних часових рядів  $x_q \in \mathbb{R}^n$ , у відповідність кожному ряду ставиться модель

$$\bar{x}_{qr}(\tau) = A_{qr}^T \bar{\tau}, \quad (3.7)$$

де  $A_{qr}$  –  $(n \times 2)$ -матриця параметрів, що підлягають оцінюванню.

Для налаштування параметрів моделі (3.7) може бути використаний рекурентний алгоритм виду

$$\begin{cases} A_{qr}^T(\tau) = A_{qr}^T(\tau-1) + ((x_{qr}(\tau) - A_{qr}^T(\tau-1)\bar{\tau})\bar{\tau}^T P_{qr}(\tau)), \\ P_{qr}(\tau) = P_{qr}(\tau-1) - \frac{P_{qr}(\tau-1)\bar{\tau}\bar{\tau}^T P_{qr}(\tau-1)}{1 + \bar{\tau}^T P_{qr}(\tau-1)\bar{\tau}}. \end{cases} \quad (3.8)$$

Зауважимо, що співвідношення (3.7) і (3.8) були використані в [4, 40, 41] для online сегментації відеопотоків.

Таким чином, кожному ряду  $x_q$  ставиться у відповідність  $r$  матриць, які можуть бути об'єднані в складову  $(n \times 2g)$ -матрицю  $A_q$ .

І, нарешті, якщо мова йде про задачі  $x_{qr}(\tau) = \{x_{qri_1 i_2}(\tau)\} \in \mathbb{R}^{n \times v}$ , у відповідність кожної послідовності може бути поставлена математична модель

$$\hat{x}_{qr}(\tau) = A_{qr}^T \bar{\tau} B_{qr}, \quad (3.9)$$

де  $A_{qr}$  –  $(n \times 2)$ –матриця,

$B_{qr}$  –  $(1 \times \nu)$ –вектор параметрів.

Для налаштування параметрів моделі (3.9) може бути використана модифікація алгоритму, введеного в [13]. Для цього у відповідність (3.9) поставимо модель виду

$$\hat{x}_{qr}(\tau) = A_{qr}^T \bar{\tau}_{qr}^B, \quad (3.10)$$

де  $\bar{\tau}_{qr}^B = \bar{\tau} B_{qr}$  –  $(2 \times \nu)$ –матриця.

Оцінка найменших квадратів матриці  $A_{qr}$  в (3.10) може бути записана у вигляді

$$A_{qr} = p_{A_{qr}} R_{A_{qr}}^{-1},$$

де

$$\begin{cases} p_{A_{qr}} = \sum_{\tau} x_{qr}(\tau) \bar{\tau}_{qr}^{BT}, \\ R_{A_{qr}} = \sum_{\tau} \bar{\tau}_{qr}^B \bar{\tau}_{qr}^{BT} \end{cases}$$

або в рекурентній формі

$$\begin{cases} p_{A_{qr}}(\tau) = p_{A_{qr}}(\tau-1) + x_{qr}(\tau) \bar{\tau}_{qr}^{BT}(\tau), \\ R_{A_{qr}}(\tau) = R_{A_{qr}}(\tau-1) + \bar{\tau}_{qr}^B(\tau) \bar{\tau}_{qr}^{BT}(\tau). \end{cases} \quad (3.11)$$

Аналогічно до попереднього введемо також модель

$$\hat{x}_{qr}(\tau) = \bar{\tau}_{qr}^A B_{qr}, \quad (3.12)$$

де  $\bar{\tau}_{qr}^A$  –  $(n \times 1)$ -вектор.

Оцінка найменших квадратів вектора-рядка  $B_{qr}$  в (3.10) може бути записана у вигляді

$$B_{qr} = R_{B_{qr}}^{-1} p_{B_{qr}},$$

де

$$\begin{cases} R_{B_{qr}} = \sum_{\tau} \bar{\tau}_{qr}^{AT} \bar{\tau}_{qr}^A, \\ p_{B_{qr}} = \sum_{\tau} x_{qr}^T(\tau) \bar{\tau}_{qr}^{AT} \end{cases}$$

або в рекурентній формі

$$\begin{cases} p_{A_{qr}}(\tau) = p_{B_{qr}}(\tau-1) + x_{qr}^T(\tau) \bar{\tau}_{qr}^{AT}(\tau), \\ R_{A_{qr}}(\tau) = R_{B_{qr}}(\tau-1) + \bar{\tau}_{qr}^{AT}(\tau) \bar{\tau}_{qr}^A(\tau). \end{cases} \quad (3.13)$$

В результаті ряду  $x_q$  ставиться у відповідність  $(n \times 2g)$ -матриця  $A_q$  і  $(vg \times 1)$ -вектор  $B_q^T$ , що підлягають кластеризації.

Тим самим, для одновимірних рядів об'єктом кластеризації є  $(2g \times 1)$ -вектори  $A_q$ , багатовимірних рядів –  $(n \times 2g)$ -матриці  $A_q$ , і, нарешті, для матричних рядів –  $(n \times 2g)$ -матриця  $A_q$  і  $(vg \times 1)$ -вектор  $B_q^T$ .

Перейдемо до кластеризації часових рядів на основі самоорганізованих карт Кохонена.

Для кластеризації наборів векторів  $A_q$ ,  $B_q^T$  (для одновимірних і багато-

вимірних послідовностей) і матриць  $A_q$  (для багатовимірних і матричних сигналів) скористаємося підходом, що базується на прототипах-центроїдах. При цьому весь набір даних розбивається на  $m$  кластерів з центроїдами  $\tilde{n}_j$ ,  $j=1,2,\dots,m$ , які необхідно визначити. Надалі, без втрати загальності, вектори, що підлягають кластеризації, будемо позначати  $B_q$ , а матриці –  $A_q$ .

Для вирішення задачі кластеризації скористаємося кластерувальними нейронними мережами Кохонена відомими також як самоорганізованні карти [95]. Розглянемо спочатку процес кластеризації-самоорганізації для векторних сигналів  $B_q$ ,  $q=1,2,\dots,Q$ . Процедура самоорганізації складається з двох послідовних етапів: конкуренції і синаптичної адаптації і починається з досить довільного завдання центроїдів  $c_j^B(0)$ , які далі повинні уточнюватися в процесі адаптації.

Робота системи починається з подачі на обробку першого вектора  $B_1$  для порівняння з усіма  $c_j^B(0)$  в евклідовій метриці

$$d(B_1, c_j^B(0)) = \|B_1 - c_j^B(0)\|^2.$$

Далі визначається центроїд-«переможець»  $c_j^{B*}$ , найближчий в сенсі відстані (3.2) до вектору  $B_1$  такий, що

$$d(B_1, c_j^{B*}(0)) = \min_j d(B_1, c_j^{B*}(0)) \quad (3.14)$$

У цьому полягає суть конкуренції – визначення для кожного  $B_q$  свого переможця  $c_j^{B*}(q-1)$ .

На етапі адаптації проводиться уточнення положення центроїда-переможця за допомогою рекурентного співвідношення (правило самонавчання «Переможець отримує все»)

$$c_j^B(1) = \begin{cases} c_j^{B^*}(0) + \eta(1)(B_1 - c_j^{B^*}(0)), & \text{якщо } c_j^{B^*}(0) \text{ – переможець,} \\ c_j^B(0) & \text{для інших центроїдів.} \end{cases}$$

Тут  $\eta(1)$  – параметр кроку налаштування, що монотонно зменшується в процесі адаптації та обираний зазвичай з емпіричних міркувань.

Після подачі на обробку  $q$ -го вектору  $B_q$  і визначення переможця  $c_j^{B^*}(q-1)$  процес адаптації реалізується згідно співвідношенню

$$c_j^B(q) = \begin{cases} c_j^{B^*}(q-1) + \eta(q)(B_q - c_j^{B^*}(q-1)), & \text{якщо } c_j^{B^*}(q-1) \text{ – переможець,} \\ c_j^B(q-1) & \text{для інших центроїдів.} \end{cases} \quad (3.15)$$

Процес уточнення проводиться до вичерпання всієї вибірки  $B_1, B_2, \dots, B_Q$ . Якщо центроїди не застабілізовано так, що

$$\sum_{j=1}^m \|c_j^B(Q) - c_j^B(Q-1)\| \leq \varepsilon, \quad (3.16)$$

то після  $B_Q$  на обробку знову подаються  $B_1, B_2, \dots$ . Цей процес триває до виконання умови (3.16).

Результат кластеризації полягає в отриманні  $m$  стійких значень центроїдів  $c_j^B$ . Необхідно зауважити, що цей результат збігається з результатом кластеризації на основі широко розповсюдженого методу  $k$ -середніх, однак вимагає значно менших обчислювальних витрат.

Для вирішення завдання кластеризації матриць  $A_1, A_2, \dots, A_Q$  можна скористуватися матричною модифікацією самоорганізовної карти [8]. При цьому етап конкуренції реалізується за допомогою метрики (3.3)

$$d(A_q, c_j^A(q-1)) = \left( Sp(A_q - c_j^A(q-1))(A_q - c_j^A(q-1)) \right)^{1/2},$$

$$d(A_q, c_j^{A^*}(q-1)) = \min_j d(A_q, c_j^A(q-1)),$$

а етап адаптації - за допомогою рекурентного співвідношення

$$c_j^A(q) = \begin{cases} c_j^{A^*}(q-1) + \eta(q)(A_q - c_j^{A^*}(q-1)), & \text{якщо } c_j^{A^*}(q-1) \text{ - переможець,} \\ c_j^A(q-1) & \text{для інших центроїдів.} \end{cases} \quad (3.17)$$

Як видно, процедури адаптації (3.15) і (3.17) структурно збігаються, що в принципі дозволяє навіть для векторних послідовностей використовувати співвідношення (3.17) тим більше, що фробеніусова норма є узагальненням евклідової.

### 3.2 Гібридний підхід до кластеризації відеорядів різної довжини

При структуризації відеопотоків асоційовані часові ряди є багатовимірними послідовностями: або векторними (в просторі ознак), коли  $x_q(k) = (x_{q1}(k), \dots, x_{qi}(k), \dots, x_{qn}(k))^T \in \mathbb{R}^N$ , або матричними (в просторі зображень або їх фрагментів), коли  $x_q(k) = \{x_{qi_1, i_2}(k)\} \in \mathbb{R}^{n \times v}$ ,  $i_1 = 1, 2, \dots, n, i_2 = 1, 2, \dots, v$ . Складність задачі пов'язана з тим, що спостереження



в кожній з вибірок  $x_q(k)$  строго впорядковані і не можуть бути схильні до переміщення.

Результатом рішення задачі кластеризації є розбиття вихідного набору даних на  $m$  кластерів-класів  $Cl_1, Cl_2, \dots, Cl_m$  в кожному з яких містяться дані близькі один до одного в сенсі застосовуємої метрики.

Слідуючи [87], методи кластеризації можуть бути в найзагальнішому випадку умовно розділені на два великі класи: ієрархічні і засновані на розбитті. Більш популярними є методи, засновані на розбитті, завдяки своїй математичній строгості, наочності отриманих результатів, розвиненому програмному забезпеченню. Ключовим поняттям тут є центроїди-прототипи – представники-зразки, навколо яких групуються дані кожного кластера. Ієрархічні методи в свою чергу діляться на агломеративні та дивізімні. Вони автоматично визначають оптимальне число кластерів шляхом попарного об'єднання даних і послідовного «вирощування» кластерів або шляхом послідовного дроблення вихідної вибірки на підкластери до досягнення необхідної точності в сенсі прийнятого критерію якості кластеризації. Так чи інакше, обидва підходи мають свої переваги і недоліки, при цьому зрозуміло, що метод, який об'єднує в собі переваги обох підходів був би вельми привабливий з точки зору користувача-практика.

Ситуація істотно ускладнюється, коли оброблювані послідовності  $x_1, x_2, \dots, x_q, \dots, x_Q$  мають різну довжину  $N_1, N_2, \dots, N_q, \dots, N_Q$ . Зрозуміло, що про розрахунок центроїдів не може бути й мови, оскільки не ясно якою має бути довжина кожного центроїда і наскільки по довжині відрізняються центроїди кожного кластера. У цій ситуації більш кращим представляється використання ієрархічного агломеративного підходу, ключовим моментом якого є вибір міри схожості або відмінності між будь-якими двома рядами з набору  $x_1, x_2, \dots, x_q, \dots, x_Q$ , на основі якої буде проводитися об'єднання окремих послідовностей в кластерах.

Проаналізуємо оцінки схожості-відмінності між багатовимірними часовими рядами різної довжини.

Розглянемо ситуацію, коли є два часових ряди  $X = \{x(1), \dots, x(k), \dots, x(N)\}$  (або вектор  $x(k) \in \mathbb{R}^n$ , або матриця  $x(k) \in \mathbb{R}^{n+v}$ ),  $Y = \{y(1), \dots, y(l), \dots, y(M)\}$  (вектор  $y(k) \in \mathbb{R}^n$ , або матриця  $y(k) \in \mathbb{R}^{n+v}$ )  $N \neq M$ , і необхідно визначити схожість (відстань) між цими послідовностями. При цьому можна оцінити відстань між будь-якими двома спостереженнями цих рядів або за допомогою евклідової

$$d^2(x(k), y(l)) = \|x(k) - y(l)\|^2 = \sum_{i=1}^n (x_i(k) - y_i(l))^2, \quad x(k), y(l) \in \mathbb{R}^n,$$

або сферичної метрики

$$d^2(x(k), y(l)) = Sp(x(k) - y(l))(x(k) - y(l))^T, \quad x(k), y(l) \in \mathbb{R}^{n+v}.$$

Для оцінки схожості між часовими рядами в даний час найбільш популярною є так звана динамічна часова деформація (DTW) [37, 89] і її різні модифікації [50, 88, 90]. В основі DTW лежить поняття викривленого шляху (warping path) між точками  $x(1), y(1)$  і  $x(N), y(M)$  для знаходження якого використовується процедура динамічного програмування, досить громіздка з обчислювальної точки зору. В результаті реалізації цієї процедури знаходиться оптимальний викривлений шлях, який задає схожість між даними  $X$  і  $Y$ .

При цьому слід пам'ятати, що DTW задає міру схожості, але не є метрикою (не виконується нерівність трикутника), хоча в її основі лежить евклідова та фробеніусова відстані.

Для оцінки відстані між часовими рядами різної довжини може бути використана метрика Левенштейна [99], яка стосовно аналізованої тут задачі може бути розрахована за допомогою співвідношень

$$d_L(x(k), y(l)) = \begin{cases} \max(k, l), & \text{если } \min(k, l) = 0 \\ \min \begin{cases} d_L(x(k-1), y(l)) + 1, \\ d_L(x(k), y(l-1)) + 1, \\ d_L(x(k-1), y(l-1)) + 1(x(k), y(l)), \end{cases} & \end{cases}$$

де  $d_L(x(k), y(l))$ , відстань між першими  $k$  спостереженнями  $X = \{x(1), \dots, x(k), \dots, x(N)\}$  та  $l$  спостереженнями  $Y = \{y(1), \dots, y(l), \dots, y(M)\}$  початкових рядів

$$1(x(k), y(l)) = \begin{cases} 0, & \text{якщо } x(k) = y(l) \\ 1, & \text{інакше.} \end{cases}$$

Загальна ж відстань між рядами задається співвідношенням

$$D_L(N, M) = d_L(x(N), y(M)).$$

Нескладно помітити значну схожість між розрахунком DTW і відстанню Левенштейна, в основі якого лежить процедура динамічного програмування. При цьому, однак,  $D(N, M)$  є лише мірою подібності, а  $D_L(N, M)$  – метрикою в повному розумінні цього слова.

Розглянемо тепер процедуру ієрархічної кластеризації багатовимірних часових рядів різної довжини.

На підставі розглянутих мір подібності може бути побудована процедура ієрархічної агломеративної кластеризації, в основу якої покладено досить ефективний і простий з обчислювальної точки зору алгоритм Евєрітта [57, 87]. Отже, нехай є масив  $x_1, x_2, \dots, x_q, \dots, x_Q$ , утворений  $Q$  багатовимірних

ми послідовностями різної довжини. На підставі цього масиву будується  $(Q \times Q)$ -матриця відстаней  $D_L(x_q, x_p)$  між усіма рядами, які утворюють цей масив. Зрозуміло, що ця матриця є симетричною з нульовими діагональними елементами. Всі інші елементи матриці - невід'ємні.

На першому етапі процедури формується кластерів  $Q/2$ , якщо  $Q$  – парне або  $(Q+1)/2$  кластерів, якщо  $Q$  – непарне, шляхом попарного об'єднання найближчих один до одного рядів  $x_q, x_p \forall q, p$ . Таким чином формується набір кластерів першого рівня  $Cl_1^{[1]}, Cl_2^{[1]}, \dots, Cl_j^{[1]}, \dots, Cl_{Q/2}^{[1]} (Cl_{(Q+1)/2}^{[1]}$  – в цьому випадку цей кластер складається з однієї послідовності).

На другому етапі, на підставі матриці відстаней, оцінюється відстань між усіма сформованими кластерами  $Cl_j^{[1]}$ , при цьому в якості цієї відстані береться мінімальна відстань між спостереженнями з різних класів

$$D_L(Cl_j^{[1]}, Cl_r^{[1]}) = \min_{\substack{x_q \in Cl_j^{[1]}, \\ x_p \in Cl_r^{[1]}}} D_L(x_q, x_p).$$

Далі попарно об'єднуються найближчі один до одного кластери, в результаті чого формується масив кластерів другого рівня  $Cl_j^{[2]}$ .

В результаті процес триває до формування набору кластерів  $m$ -го рівня  $Cl_j^{[m]}$ , що задовольняє прийнятним вимогам до якості кластеризації.

Якість кластеризації в розглянутій задачі зручно оцінювати за допомогою індексу Данна (Dunn index) [56], який повністю заснований на вихідній матриці відстаней. При цьому на кожному рівні оцінюється діаметр сформованих кластерів, в якості якого береться максимальна відстань між рядами, що входять в цей кластер:

$$DiamCl_j^{[m]} = \max_{\substack{x_q \in Cl_j^{[m]}, \\ x_p \in Cl_r^{[m]}}} D_L(x_q, x_p).$$

Індекс Данна в розглянутій задачі для  $m$ -го рівня кластеризації може бути записаний у формі

$$Du(Q^{[m]}) = \min_{j=1, \dots, Q^{[m]}} \left( \min_{r=j+1, \dots, Q^{[m]}} \frac{D_L(Cl_j^{[m]}, Cl_r^{[m]})}{\max_{j=1, \dots, Q^{[m]}} DiamCl_j^{[m]}} \right),$$

де  $Q^{[m]}$  – кількість сформованих кластерів на  $m$ -м рівні.

Чим більше значення індексу Данна, тим вище якість кластеризації. Зрозуміло, що якщо на  $m$ -му рівні значення індексу припиняє збільшуватися, тобто

$$Du(Q^{[m]}) \leq Du(Q^{[m-1]}),$$

процес кластеризації може бути зупинений, а значення  $Q^{[m]}$  вважається істинним числом кластерів у початковій вибірці  $x_1, x_2, \dots, x_q, \dots, x_Q$ .

Більший інтерес представляє аналіз кластеризації часових рядів різної довжини в разі перетинних класів.

Семантично пов'язані послідовності в відеопотоках нерідко рознесені в часі, а кластери можуть перекривати один одного (принаймні при використанні різних ефектів редагування відео). Ця ситуація є предметом розгляду нечіткого кластерного аналізу часових рядів, але при цьому завжди передбачається, що ряди мають однакову довжину. Зауважимо також, що відомі методи нечіткого кластерного аналізу використовують прототипи-центроїди, в якості яких використовують зважені середні спостережень в кожному з сформованих класів.

Завдання нечіткої кластеризації в прийнятих тут позначеннях може бути сформульована як задача мінімізації цільової функції

$$E(\mu(x_q, c_j^{[m]})) = \sum_{q=1}^Q \sum_{j=1}^{Q^{[m]}} \mu^\beta(x_q, c_j^{[m]}) D(x_q, c_j^{[m]})$$

при обмеженнях

$$\sum_{j=1}^{Q^{[m]}} \mu(x_q, c_j^{[m]}) = 1,$$

де  $\mu(x_q, c_j^{[m]}) > 0$  – рівень нечіткої належності послідовності  $x_q$  до кластеру  $Cl_j^{[m]}$ , центроїдом якого є  $c_j^{[m]}$ ;

$\beta > 0$  – параметр фаззифікації зазвичай рівний двом в задачах, що використовують евклідову (сферичну) метрику;

$D(x_q, c_j^{[m]})$  – відстань між  $x_q$  та центроїдом  $c_j^{[m]}$  в тій же метриці.

Зрозуміло, що в розглянутій тут задачі говорити про прототипи-центроїди не доводиться, а замість них можна використовувати так звані представники-зразки, які є по суті одним з спостережень, найбільш характерним для конкретного кластера.

В якості такого представника  $x_j^{[m]}$  для кластера  $Cl_j^{[m]}$  може бути взята послідовність  $x_q$  мінімально віддалена від усіх інших рядів, що входять в цей кластер, тобто

$$x_j^{[m]} = x_q \in Cl_j^{[m]} = \arg \min_{\substack{x_q \in Cl_j^{[m]} \\ x_p \in Cl_j^{[m]}}} \sum_{q,p=1}^{Q_j^{[m]}} D_L(x_q, x_p),$$

де  $Q_j^{[m]}$  – число спостережень в  $j$ -м кластері  $m$ -го рівня.

Тоді рівень належності спостереження  $x_p$  до кластеру  $Cl_j^{[m]}$  може бути розрахований за допомогою простого співвідношення

$$\mu(x_p, x_j^{[m]}) = \frac{D_L^{-1}(x_p, x_j^{[m]})}{\sum_{S=1}^{Q^{[m]}} D_L^{-1}(x_p, x_S^{[m]})}.$$

Зрозуміло, що якщо в процесі ієрархічної кластеризації послідовність  $x_p$  була віднесена до кластеру  $Cl_j^{[m]}$ , то і значення  $\mu(x_p, x_j^{[m]})$  має бути максимальним серед всіх  $x_S^{[m]}$ ,  $S = 1, 2, \dots, Q^{[m]}$ .

Далі розглянемо загальний покроковий алгоритм процесу ієрархічної агломеративної нечіткої кластеризації багатовимірних часових рядів нечіткої довжини:

1. Задаємо початкові ряди  $x_1, x_2, \dots, x_q, \dots, x_Q$  і розраховуємо  $(Q \times Q)$  – матрицю відстаней.
2. Формуємо масив кластерів першого рівня на основі попарного об'єднання вихідних рядів.
3. Формуємо масив кластерів наступного рівня на основі попарного об'єднання кластерів попереднього рівня.
4. Розрахунок індексу Данна для оцінки якості кластеризації.
5. Якщо якість не задовольняє, переходимо до кроку 3 інакше крок 6.
6. Розрахунок представників-зразків для кожного кластера останнього рівня.
7. Розрахунок рівнів належності кожного спостереження  $x_q$  до кластеру  $Cl_j^{[m]}$ .

Відзначимо простоту обчислювальної реалізації запропонованого підходу, що зводиться до послідовності елементарних арифметичних операцій над часовими рядами, що аналізуються.

### 3.3 Адаптивні матричні моделі в задаче контролю потоків відео

Розглянемо тепер метод контролю і аналізу потоків матричних сигналів, що не апелює до статистичної природі оброблюваних даних, що типово для практично будь-якої обробки відеопотоків.

Нехай аналізована відеоінформація задана у вигляді багатовимірної матричної послідовності  $x(1), x(2), \dots, x(k), \dots, x(N), \dots$ , де  $x(k) = \{x_{i_1 i_2}(k)\}$ ,  $i_1 = 1, 2, \dots, n$ ,  $i_2 = 1, 2, \dots, v$  – матриця параметрів, що характеризує окремий фрагмент (кадр) вхідної послідовності;  $k = 1, 2, \dots, N, \dots$  – поточний дискретний час.

Для опису вхідних сигналів, представлених у вигляді багатовимірної матричної послідовності, в [1, 15] була введена прогнозуюча матрична авторегресійна модель виду

$$x(k) = \sum_{h=1}^{n_A} A^h x(k-h) B^h + \xi(k), \quad (3.18)$$

де  $A^h, B^h$  –  $(n \times n)$  та  $(v \times v)$  – матриці параметрів, що підлягають оцінюванню,

$n_A$  – розмірність,

$v$  – порядок авторегресії,

$\xi(k)$  –  $(n \times v)$  – матричний білий шум.

Вводячи далі складові матриці параметрів  $\tilde{A} = (A^1 : A^2 : \dots : A^{n_A})$  розмірності  $n \times n n_A$ ,  $\tilde{B} = (B^1 : B^2 : \dots : B^{n_A})^T$  розмірності  $v n_A \times v$  та



$$\tilde{x}(k-1) = \begin{pmatrix} x(k-1) & \dots & 0 \\ & \ddots & \\ \vdots & & x(k-h) & \vdots \\ 0 & \dots & & \ddots & x(k-n_A) \end{pmatrix}$$

розмірності  $nn_A \times vn_A$ , вираз (3.17) записується в компактному вигляді

$$x(k) = \tilde{A}\tilde{x}(k-1)\tilde{B} + \xi(k),$$

у відповідність до якого може бути поставлена адаптивна матрична модель виду

$$\hat{x}(k) = \tilde{A}(k-1)\tilde{x}(k-1)\tilde{B}(k-1), \quad (3.19)$$

де  $\tilde{A}(k-1)$ ,  $\tilde{B}(k-1)$  – матриці параметрів, що настраюються, одержані на підставі попередніх  $(k-1)$  спостереженнях  $\tilde{x}(1), \tilde{x}(2), \dots, \tilde{x}(k-1)$ .

Для налаштування параметрів адаптивної матричної моделі (3.19) в [13] був запропонований оптимальний градієнтний алгоритм ідентифікації, що мінімізує однокрокові критерії

$$E(\tilde{x}(k), \tilde{A}) = Sp(x(k) - \tilde{A}(k-1)\tilde{x}(k-1)\tilde{B}(k-1)) \times \\ \times (x(k) - \tilde{A}(k-1)\tilde{x}(k-1)\tilde{B}(k-1))^T = SpV^A(k)V^{AT}(k) = \|V^A(k)\|^2, \quad (3.20)$$

$$E(\tilde{x}(k), \tilde{B}) = Sp(x(k) - \tilde{A}(k)\tilde{x}(k-1)\tilde{B}(k-1)) \times \\ \times (x(k) - \tilde{A}(k)\tilde{x}(k-1)\tilde{B}(k-1))^T = SpV^B(k)V^{BT}(k) = \|V^B(k)\|^2, \quad (3.21)$$

де  $\|\circ\|$  – матрична норма Фробеніуса.

Процедура градиентної мінімізації критеріїв (3.20), (3.21) може бути записана у вигляді

$$\begin{cases} \tilde{A}(k) = \tilde{A}(k-1) + \eta_A(k) V^A(k) \tilde{B}^T(k-1) \tilde{x}^T(k-1), \\ \tilde{B}(k) = \tilde{B}(k-1) + \eta_B(k) \tilde{x}^T(k-1) \tilde{A}(k) V^B(k), \end{cases} \quad (3.22)$$

де  $\eta_A(k)$ ,  $\eta_B(k)$  – параметри кроку пошуку.

Оптимізація процедури (3.22) по швидкодії веде до алгоритму

$$\begin{cases} \tilde{A}(k) = \tilde{A}(k-1) + (SpV^A(k) \tilde{B}^T(k-1) \tilde{x}^T(k-1) \tilde{x}(k-1) \tilde{B}(k-1) V^{AT}(k)) \times \\ \times (SpV^A(k) \tilde{B}^T(k-1) \tilde{x}^T(k-1) \tilde{x}(k-1) \tilde{B}(k-1) \tilde{B}^T(k-1) \tilde{x}^T(k-1) \tilde{x}(k-1) \times \\ \times \tilde{B}(k-1) V^{AT}(k))^{-1} V^A(k) \tilde{B}(k-1) \tilde{x}^T(k-1), \\ \tilde{B}(k) = \tilde{B}(k-1) + (SpV^{BT}(k) \tilde{A}(k) \tilde{x}(k-1) \tilde{x}^T(k-1) \tilde{A}^T(k) V^B(k)) \times \\ \times (Sp \tilde{A}(k) \tilde{x}(k-1) \tilde{x}^T(k-1) \tilde{A}^T(k) V^B(k) V^{BT}(k) \tilde{A}(k) \tilde{x}(k-1) \times \\ \times \tilde{x}^T(k-1) \tilde{A}^T(k))^{-1} \tilde{x}^T(k-1) \tilde{A}^T(k) V^B(k). \end{cases} \quad (3.23)$$

Алгоритм (3.23), будучи матричною модифікацією алгоритму Качмажа і володіючи максимальною швидкодією в класі градієнтних процедур, не володіє властивостями, що фільтрують, а, отже, чутливий до дії перешкод, що обмежує його можливості при роботі з зашумленими сигналами.

Ефективними згладжуючими властивостями володіють процедури ідентифікації, засновані на методі найменших квадратів, однак він не може бути використаний у вихідній формі для моделі типу (3.19).

Для подолання цього утруднення поставимо у відповідність моделі (3.19) вираз

$$\hat{x}(k) = \tilde{A} x^B(k),$$

(тут  $x^B(k) = \tilde{x}(k-1)\tilde{B}(k-1)$ ) та введемо критерій ідентифікації

$$E(x^B(k), \tilde{A}) = \sum_{\tau=1}^k Sp(x(\tau) - \tilde{A}(k)x^B(\tau))(x(\tau) - \tilde{A}(k)x^B(\tau))^T,$$

мінімізація якого по  $\tilde{A}(k)$  призводить до оцінки [1]

$$\tilde{A}(k) = p_{\tilde{A}}(k)R_{\tilde{A}}^{-1}(k), \quad (3.24)$$

де

$$p_{\tilde{A}}(k) = \sum_{\tau=1}^k x(\tau)x^{B^T}(\tau), \quad (3.25)$$

$$R_{\tilde{A}}(k) = \sum_{\tau=1}^k x^B(\tau)x^{B^T}(\tau). \quad (3.26)$$

Розрахунок матриць (3.25), (3.26) нескладно організувати в рекуррентній формі, при цьому

$$\begin{cases} p_{\tilde{A}}(k) = x(k)x^{B^T}(k) + p_{\tilde{A}}(k-1), \\ R_{\tilde{A}}(k) = x^B(k)x^{B^T}(k) + R_{\tilde{A}}(k-1), \end{cases}$$

або

$$\begin{cases} p_{\tilde{A}}(k) = \frac{1}{k}x(k)x^{B^T}(k) + \frac{k-1}{k}p_{\tilde{A}}(k-1), \\ R_{\tilde{A}}(k) = \frac{1}{k}x^B(k)x^{B^T}(k) + \frac{k-1}{k}R_{\tilde{A}}(k-1). \end{cases} \quad (3.27)$$

Аналогічно до попереднього введемо модель

$$\hat{x}(k) = x^A(k) \tilde{B}$$

(тут  $x^A(k) = \tilde{A} \tilde{x}(k-1)$ ) та критерій ідентифікації

$$E(\tilde{B}, x^A(k)) = \sum_{\tau=1}^k Sp(x(\tau) - x^A(\tau) \tilde{B}(k))(x(\tau) - x^A(\tau) \tilde{B}(k))^T$$

мінімізація якого по  $\tilde{B}(k)$  веде до співвідношень аналогічним (3.24)–(3.26)

$$\tilde{B}(k) = R_{\tilde{B}}^{-1}(k) p_{\tilde{B}}(k), \quad (3.28)$$

$$\text{де } p_{\tilde{B}}(k) = \sum_{\tau=1}^k x^T(\tau) x^A(\tau),$$

$$R_{\tilde{B}}(k) = \sum_{\tau=1}^k x^{AT}(\tau) x^A(\tau).$$

У рекурентній формі маємо

$$\begin{cases} p_{\tilde{B}}(k) = x^T(k) x^A(k) + p_{\tilde{B}}(k-1), \\ R_{\tilde{B}}(k) = x^{AT}(k) x^A(k) + R_{\tilde{B}}(k-1), \end{cases}$$

$$\begin{cases} p_{\tilde{B}}(k) = \frac{1}{k} x^T(k) x^A(k) + \frac{k-1}{k} p_{\tilde{B}}(k-1), \\ R_{\tilde{B}}(k) = \frac{1}{k} x^{AT}(k) x^A(k) + \frac{k-1}{k} R_{\tilde{B}}(k-1). \end{cases} \quad (3.29)$$

В очевидь, що використання подібного підходу веде до того, що згладжуються не тільки випадкові обурення, а й зміни в корисному сигналі, які необхідно виявити.

У подібній ситуації доцільно скористатися ідеями експоненціального згладжування Р.Брауна, які займають місце між однокроковими процедурами ідентифікації і класичним методом найменших квадратів.

Вводячи параметр згладжування  $0 < \alpha < 1$  в співвідношення (3.27), (3.29), остаточно приходимо до процедури налаштування адаптивної матричної моделі:

$$\begin{aligned} \tilde{A}(k) &= p_{\tilde{A}}(k) R_{\tilde{A}}^{-1}(k), \\ \left\{ \begin{array}{l} p_{\tilde{A}}(k) = \alpha x(k) x^{B^T}(k) + (1-\alpha) p_{\tilde{A}}(k-1), \\ R_{\tilde{A}}(k) = \alpha x^B(k) x^{B^T}(k) + (1-\alpha) R_{\tilde{A}}(k-1). \end{array} \right. \end{aligned} \quad (3.30)$$

$$\begin{aligned} \tilde{B}(k) &= R_{\tilde{B}}^{-1}(k) p_{\tilde{B}}(k), \\ \left\{ \begin{array}{l} p_{\tilde{B}}(k) = \alpha x^T(k) x^A(k) + (1-\alpha) p_{\tilde{B}}(k-1), \\ R_{\tilde{B}}(k) = \alpha x^{A^T}(k) x^A(k) + (1-\alpha) R_{\tilde{B}}(k-1). \end{array} \right. \end{aligned} \quad (3.31)$$

Контроль за змінами властивостей матричного сигналу зручно організувати базуючись на ідеях експоненціального згладжування. Найбільш відповідною основою для цього є використання, так званого, слідкуючого сигналу, найбільш ефективним з яких з точки зору процесу виявлення властивостей є сигнал Д. Трігга-А. Ліча [143], який відомий в скалярній формі, орієнтований на аналіз одновимірних послідовностей. В [23] введена модифікація методу Трігга-Ліча, орієнтована на аналіз векторних часових рядів, при цьому необхідно відзначити, що процедура вимагає використання набору параметрів, що задаються ззовні, що природно вносить в процес контролю значний елемент суб'єктивізму. Зрозуміло, що використання цього підходу для контролю матричних процесів навряд чи може забезпечити необхідну якість контролю.

Базовий слідкуючий сигнал Трігга-Ліча в одновимірному випадку може бути записаний у вигляді [143]

$$T^{TL}(k) = \frac{T'(k)}{T''(k)}, \quad (3.32)$$

де

$$T'(k) = \alpha v(k) + (1-\alpha)T'(k-1) \quad (3.33)$$

$$T''(k) = \alpha |v(k)| + (1-\alpha)T''(k-1) \quad (3.34)$$

(тут  $v(k) = x(k) - \hat{x}(k)$  – помилка прогнозування на  $k$ -м такті контролю сигналу  $x(k)$ ), при цьому факт змін властивостей сигналу  $x(k)$  в кожен момент часу фіксується в ситуації коли

$$T^{TL}(k) - T^{TL}(k-1) \geq \Delta, \quad (3.35)$$

де  $\Delta$  – деякий поріг, який визначає зміну властивостей.

Введемо в розгляд апостеріорну помилку настройки моделі (3.19)

$$v(k) = x(k) - \tilde{A}(k)\tilde{x}(k-1)\tilde{B}(k) = \left\{ v_{i_1 i_2}(k) \right\},$$

що задається в вигляді  $(n \times v)$  – матриці відхилень виходів моделі від реального сигналу.

Як слідкуючий сигнал будемо розглядати вираз (3.32), де замість (3.33), (3.34) введемо оцінки

$$\begin{cases} T'(k) = \alpha \sum_{i_1=1}^n \sum_{i_2=1}^v v_{i_1 i_2}(k) + (1-\alpha)T'(k-1), \\ T''(k) = \alpha \sum_{i_1=1}^n \sum_{i_2=1}^v |v_{i_1 i_2}(k)| + (1-\alpha)T''(k-1), \end{cases}$$

які також можуть бути записані у формі

$$\begin{cases} T'(k) = \alpha(nv)^{-1} \sum_{i_1=1}^n \sum_{i_2=1}^v v_{i_1 i_2}(k) + (1-\alpha)T'(k-1), \\ T''(k) = \alpha(nv)^{-1} \sum_{i_1=1}^n \sum_{i_2=1}^v |v_{i_1 i_2}(k)| + (1-\alpha)T''(k-1) \end{cases}$$

зручнішою для online контролю.

Власне контроль сигналу проводиться на основі нерівності (3.35) і в разі виявлення істотної зміни властивостей проводиться оновлення параметрів згладжування згідно співвідношенню  $\alpha = |T^{TL}(k)|$ .

### Висновки по третьому розділу

1. Запропонована процедура кластеризації часових рядів (скалярних, векторних, матричних), в основі якої лежать модифікації ітеративної динамічної часової деформації і кластерувальні мережі Кохонена, що дозволяє обробляти великі відеопотоки на основі простих обчислювальних моделей.

2. Запропонований метод кластеризації багатовимірних часових рядів різної довжини в умовах невідомої кількості класів і можливості їх взаємного перетину на основі гібридизації ієрархічного агломеративного і нечіткого, заснованого на центроїдах, підходів з використанням метрики Левенштейна створює достатні передумови для online сегментації відео і offline кластеризації в цілях відеореферування.

3. Запропоновані підходи до налаштування параметрів адаптивної матричної моделі та виявлення змін властивостей матричної послідовності забезпечують можливості підвищення семантичної релевантності фрагментної сегментації відеопотоків.

Список використаних джерел у даного розділі наведено у повну списку використаних джерел під номерами: [1, 13, 15, 18, 20, 23, 36-37, 50, 56-57, 87-90, 95, 102, 128, 143, 162].



## 4 МЕТОДИ ПОСЛІДОВНОЇ КЛАСТЕРИЗАЦІЇ ВІДЕО ДЛЯ РЕФЕРУВАННЯ

Повнота, точність і оперативність інформаційного пошуку відео в чому визначається адекватністю, достовірністю, об'єктивністю відеореферування. Поліструктурність і багатозначність окремих частин відеопотоків призводять до необхідності узагальнення змісту шляхом переходу від online сегментації відеорядів до offline кластеризації (в тому числі стратифікованою). Такі види лінійної та нелінійної (в часі) обробки створюють передумови для виявлення розумного компромісу між інформаційною потребою і обсягами і властивостями збережених даних для багатоаспектного індексування, що забезпечує релевантний пошук подій (сюжетів, сценаріїв) із запитамі в формі відеопослідовностей, наборів ключових кадрів або їх різних ознакових описів – багатовимірних часових рядів.

Локальна просторова (внутрікадрова) і часова (міжкадрова) обробка відеоданих крім спрощеного пошуку формальної відповідності дає можливість збільшувати ступінь смислової відповідності за рахунок виділення найбільш значущих областей візуальної уваги або інтересу. Саме вони у вигляді частини зображення або розподілу ознак по заданій області, наприклад, градієнтів формують фрагменти, в досить адекватно відображаючому зміст вигляді. Крім того, можливість вибору фрагментів усуває проблеми, пов'язані зі зміною ракурсу або частковими затуленнями об'єктів.

Розділ присвячений розробці методів фрагментної (матричної) кластеризації незв'язаних в часі послідовностей, визначених на стадії сегментації відеорядів. Увага сфокусована на кластеризації матричних послідовностей на основі ітеративної динамічної часової деформації, гармонійних k-середніх, представників сегментів.

#### 4.1 Кластеризація матричних послідовностей на основі ітеративної динамічної часової деформації

Задача кластеризації займає особливе місце в аналізі відеопослідовностей, коли більшість з відомих методів втрачає свою ефективність або навіть працездатність. Це пояснюється перш за все там, що на перший план виходять підвищені вимоги до швидкодії в умовах великих обсягів оброблюваних даних, а власне об'єктами кластеризації нерідко є матричні (фрагментні) часові ряди  $X_1, \dots, X_q, X_r, \dots, X_Q$ , кожен з яких містить  $N_1, \dots, N_q, N_r, \dots, N_Q$  спостережень так, що  $X_r = \{x_r(1), x_r(2), \dots, x_r(k), \dots, x_r(N_r)\}$ , а кожне окреме спостереження являє собою  $(n \times m)$ -матрицю  $x_r(k) = \{x_{ri_1, i_2}(k)\}$   $i_1 = 1, 2, \dots, n$ ,  $i_2 = 1, 2, \dots, m$ .

Однак основні проблеми, що виникають при кластеризації часових рядів, визначаються не стільки високою розмірністю оброблюваних даних, скільки тим, що ці ряди мають різну довжину. Так, наприклад, якщо ряд  $X_q$  містить  $N_q$   $(n \times m)$  матриць відеокадрів, то ряд  $X_r$  – відповідно  $N_r$  матриць тієї ж розмірності, що не дозволяє для визначення близькості між ними розрахувати різницю  $\Delta_{q,r} = X_q - X_r$  і відповідно скористатися тими чи іншими нормами цієї різниці [55]. Альтернативою традиційним нормам може служити так звана динамічна часова деформація – DTW [36, 50, 88-90, 128, 162], в основі якої лежить евклідова норма, розрахована між двома векторними спостереженнями. Зрозуміло, що в разі матричних спостережень може бути використана сферична норма, що є узагальненням евклідової.

Тим самим, якщо задані дві послідовності  $X_q$  і  $X_r$  такі, що  $X_q = \{x_q(1), x_q(2), \dots, x_q(k), \dots, x_q(N_q)\}$  і  $X_r = \{x_r(1), x_r(2), \dots, x_r(k), \dots, x_r(N_r)\}$ ,  $N_q \neq N_r$  за допомогою динамічної часової деформації можна оцінити їх схо-

жість або відмінність. Для цього в розгляд вводиться  $(N_q \times N_r)$ -матриця відстаней з елементами

$$d(x_q(k), y_r(l)) = \left( Sp(x_q(k) - y_r(l))(x_q(k) - y_r(l))^T \right)^{1/2}.$$

У цій матриці визначається «викривлений шлях», який визначається послідовністю елементів  $W = \{w_1, w_2, \dots, w_L\}$ ,  $w_t = d(x_q(k), y_r(l))_t$ ,  $\max(N_q, N_r) \leq L < N_q + N_r + 1, t = 1, 2, \dots, L$  при цьому чим він менше, тим ближчі один до одного  $X_q$  і  $X_r$ . Цей шлях визначається мінімізацією критерію

$$DTW(x_q, y_r) = \min_W \left\{ \frac{1}{L} \sum_{t=1}^L d(x_q(k), y_r(l))_t \right\},$$

яка реалізується на основі рекурентної процедури динамічного програмування

$$D(k, l) = d(x_q(k), y_r(l)) + \min \{ D(k, l-1), D(k-1, l), D(k-1, l-1) \}$$

де  $D(k, l)$  – кумулятивні відстані між елементами послідовностей  $x_q(k), x_r(l)$ .

Очевидно, що

$$DTW(x_q, y_r) = D(N_q, N_r). \quad (4.1)$$

Підхід до кластеризації часових рядів на основі DTW для досить довгих матричних послідовностей втрачає свою ефективність, оскільки вимагає великих обчислювальних витрат. Розглянемо модифіковану ітеративну динамічну часову деформацію.

Ідея ітеративної динамічної часової деформації полягає в редукуванні початкових рядів [50], при цьому аналізовані часові ряди спочатку розбиваються на чотири ділянки, а на кожній з цих ділянок ряд замінюється своїми середніми значеннями. Таким чином, ряди  $X_q$  і  $X_r$  замінюються послідовностями  $\bar{X}_q$ ,  $\bar{X}_r$  кожна з яких містить по чотири ділянки  $\bar{x}_q(1), \bar{x}_q(2), \bar{x}_q(3), \bar{x}_q(4)$  та  $\bar{x}_r(1), \bar{x}_r(2), \bar{x}_r(3), \bar{x}_r(4)$ . Близькість між рядами  $\bar{X}_q$ ,  $\bar{X}_r$  далі оцінюється на підставі  $(4 \times 4)$ -матриці відстаней, при цьому

$$DTW(\bar{x}_q, \bar{x}_r) = DTW(4,4).$$

Якщо отримані результати не задовільні, кожен з ділянок  $\bar{x}_q(k), \bar{x}_r(k)$  ділиться навпіл і оцінюється відстань

$$DTW(\bar{x}_q, \bar{x}_r) = DTW(8,8).$$

Процес дроблення ділянок навпіл триває до отримання необхідної точності результату. При цьому кількість аналізованих ділянок зростає в геометричній прогресії.

З обчислювальної точки зору набагато простіше дробити вихідні ряди так, щоб кількість ділянок збільшувалася в арифметичній прогресії так, щоб в результаті редукції були отримані послідовності  $\bar{X}_1, \dots, \bar{X}_q, \bar{X}_r, \dots, \bar{X}_Q$ , кожна

з яких містить  $S$  ділянок, при цьому  $\bar{X}_q = \{\bar{x}_q(1), \bar{x}_q(2), \dots, \bar{x}_q(S)\}$ ,

$\bar{X}_r = \{\bar{x}_r(1), \bar{x}_r(2), \dots, \bar{x}_r(S)\}$ . І хоча ряди  $\bar{X}_q$ ,  $\bar{X}_r$  містять однакову кількість ділянок, використовувати для оцінки відстані між ними сферичну норму

$D(\bar{x}_q, \bar{x}_r) = \left( Sp(\bar{x}_q - \bar{x}_r)(\bar{x}_q - \bar{x}_r)^T \right)^{1/2}$  некоректно, оскільки ділянки цих рядів

мають різну довжину.

У зв'язку з цим може бути використана модифікація оцінки (4.1) виду

$$DTW(\bar{x}_q, \bar{y}_r) = D(S, S)_\xi = \bar{D}(\bar{x}_q, \bar{y}_r). \quad (4.2)$$

Зрозуміло, що замість стандартного середнього арифметичного для кожної з ділянок часових рядів може бути використана будь-яка інша оцінка, яка використовується при аналізі часових послідовностей [42, 55, 94, 102, 147].

Розглянемо кластеризацію редукованих матричних часових рядів, індукованих відеопотоками.

Для вирішення власне задачі кластеризації може в принципі бути використаний будь-який з відомих алгоритмів кластеризації, в яких метрика, яка визначає близькість оброблюваних рядів, замінюється мірою близькості (4.2). Стосовно до часових рядів найбільш популярними і ефективними є методи, засновані на прототипах-центроїдах [116], в якості яких зазвичай використовуються арифметичні середні, медіани і медоїди. В [26, 77] для задач аналізу відеопослідовностей були використані саме такі прототипи. При цьому, якщо оброблюваний масив відеорядів необхідно розбити на  $m$  кластерів  $Cl_1, Cl_2, \dots, Cl_j, \dots, Cl_m$  з центроїдами  $C_1, C_2, \dots, C_j, \dots, C_m$ , правило віднесення ряду  $\bar{X}_q$  до кластеру  $Cl_j$  може бути записано у вигляді  $\bar{X}_q \in Cl_j$ , якщо  $\bar{D}(\bar{x}_q, C_j) < \bar{D}(\bar{x}_q, C_f)$  для всіх  $q = 1, 2, \dots, Q; j, f = 1, 2, \dots, m$ .

При цьому початкові центроїди  $C_j, C_f$  вибираються досить довільно. Власне процес кластеризації складається з послідовності ітерацій, на яких перераховуються координати центроїдів.

Так, наприклад, в найбільш популярному методі кластеризації  $k$ -середніх в процесі налаштування центроїдів мінімізується цільова функція

$$E(\bar{x}_q, C_j) = \sum_{q=1}^Q \sum_{j=1}^m u_j(q) \bar{D}^2(\bar{x}_q, C_j), \quad (4.3)$$

а центроїди-середні на кожній ітерації процедури обчислюються згідно співвідношенню

$$C_j = \frac{1}{N_j} \sum_{\bar{X}_q \in Cl_j} \bar{X}_q = \frac{\sum_{q=1}^Q u_j(q) \bar{X}_q}{\sum_{q=1}^Q u_j(q)}, \quad (4.4)$$

де  $u_j(q)$  – чітка функція належності для ряду  $j$ -му кластеру така, що

$$u_j(q) = \begin{cases} 1, & \text{якщо } \bar{X}_q \in Cl_j, \\ 0, & \text{інакше.} \end{cases}$$

Разом з цим відомо [25, 29, 151], що методи, засновані на середніх або медіанах, вкрай чутливі до невдалого вибору початкових центроїдів так, що процес кластеризації потребуватиме більшої кількості ітерацій, що абсолютно неприйнятно при обробці великих масивів відеоспостережень.

Альтернативою може служити підхід до кластеризації, заснований на використанні гармонійних середніх [17, 68, 155-156], при цьому для послідовності скалярних спостережень  $x_1, \dots, x_q$  гармонійне середнє може бути розраховане у вигляді

$$\bar{x} = \frac{Q}{\sum_{q=1}^Q x^{-1}(q)}.$$

Алгоритм, введений В. Zhang з співавторами [156] для масиву векторних спостережень, може бути модифікований для кластеризації матричних об'єктів  $\bar{X}_1, \dots, \bar{X}_q, \bar{X}_r, \dots, \bar{X}_Q$ , де кожен  $\bar{X}_q$  має розмірність  $n \times ms$ , а відстань між  $\bar{X}_q$  і центроїдом  $C_j$  оцінюється виразом  $\bar{D}(\bar{x}_q, C_j)$  типу (4.2). При цьому в процесі обробки даних мінімізується цільова функція

$$E(\bar{x}_q, C_j) = \sum_{q=1}^Q \frac{m}{\sum_{j=1}^m \bar{D}^2(\bar{x}_q, C_j)}, \quad (4.5)$$

яка також як і (4.3) ґрунтується на відстані  $\bar{D}^2(\bar{x}_q, C_j)$ .

Мінімізація (4.5) по матриці  $C_j$  веде до оцінки

$$C_j = \frac{\sum_{q=1}^Q \bar{D}^{-4}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2} \bar{x}_q}{\sum_{q=1}^Q \bar{D}^{-4}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2}}, \quad (4.6)$$

яка подібно (4.4) також обчислюється в результаті послідовності ітерацій, однак їх кількість практично не залежить від початкових умов. Тут же автор вводить модифікацію (4.6) у вигляді

$$C_j = \frac{\sum_{q=1}^Q \bar{D}^{-\alpha}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2} \bar{x}_q}{\sum_{q=1}^Q \bar{D}^{-\alpha}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2}} \quad (4.7)$$

де параметр  $\alpha$  має сенс близький до фаззифікатору, використовуваному в процедурах нечіткої кластеризації.

Нескладно ввести в розгляд нечіткі модифікації процедур (4.6), (4.7) на основі підходу, що розглядається в [155]. При цьому нечіткий рівень належності  $q$ -го ряду  $j$ -му кластеру може бути розрахований відповідно до співвідношення

$$u_j(q) = \frac{\bar{D}^{-2}(\bar{x}_q, C_j)}{\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j)}$$

для центроїдів (4.6) та

$$u_j(q) = \frac{(\bar{D}^{-2}(\bar{x}_q, C_j))^{\frac{1}{1-\beta}}}{\sum_{j=1}^m (\bar{D}^{-2}(\bar{x}_q, C_j))^{\frac{1}{1-\beta}}}$$

для (4.7). Тут  $\beta > 0$  – параметр фаззифікації, що визначає «розмитість» меж між сформованими кластерами.

Необхідно відзначити, що з обчислювальної точки зору отримання оцінок (4.6), (4.7) не складніше оцінювання центроїдів  $k$ -середніх, проте вимагає меншої кількості ітерацій.

## 4.2 Кластеризація послідовностей відеоданих на основі гармонійних $k$ -середніх

При різних апріорних припущеннях про природу даних особливе місце в задачах кластеризації займають методи, засновані на прототипах-центроїдах, завдяки, перш за все, ясній фізичній інтерпретації і простоті обчислювальних моделей. Всі вони базуються на тих чи інших алгоритмах оп-



тимізації цільових функцій. Тут, перш за все, слід відзначити процедури  $k$ -means,  $k$ -medoids,  $k$ -representatives, що існують як в *crisp* так і в *fuzzy* варіантах і породили безліч модифікацій. Ці процедури відрізняються один від одного видом використовуваної метрики, що визначає відстань між спостереженнями з початкового набору даних і можливими прототипами центроїдів. Найбільш популярними тут є процедури, пов'язані з  $k$ -means і евклідовою метрикою, що дозволяє звести початкову задачу до вирішення стандартної проблеми квадратичної оптимізації. При цьому передбачається, що початковою інформацією є масив векторних спостережень

$$X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}, \quad x(k) = (x_1(k), \dots, x_i(k), \dots, x_N(k))^T,$$

де  $k$  – номер спостережень у початковому наборі даних. При цьому власне задача кластеризації зводиться до пошуку набору центроїдів  $c_1, \dots, c_j, \dots, c_M$ ;

$c_j = (c_{j1}, \dots, c_{ji}, \dots, c_{jN})^T$ , що забезпечують мінімум цільової функції

$$\begin{aligned} E(x(k), c_j) &= \sum_{k=1}^N \sum_{j=1}^M u_j(k) \|x(k) - c_j\|^2 = \\ &= \sum_{k=1}^N \sum_{j=1}^M \mu_j(k) (x(k) - c_j)^T (x(k) - c_j), \end{aligned} \quad (4.8)$$

де  $\mu_j(k)$  – рівень чіткої належності  $k$ -го спостереження  $j$ -му кластеру. Цей рівень задається виразом

$$\mu_j(k) = \begin{cases} 1, & \text{якщо } x(k) \text{ належить } j\text{-му кластеру;} \\ 0, & \text{інакше.} \end{cases} \quad (4.9)$$

Задача нечіткої кластеризації відрізняється тим, що рівень належності приймає значення в інтервалі

$$0 \leq \mu_j(k) \leq 1 \quad (4.10)$$

та відповідає умові

$$\sum_{j=1}^M \mu_j(k) = 1, \forall k = 1, 2, \dots, N. \quad (4.11)$$

Природно, що задача кластеризації різко ускладнюється у разі, якщо обробці підлягають не вектори, а відповідні фрагментам зображень матриці  $x(k) = \{x_{i_1, i_2}(k)\}, i_1 = 1, 2, \dots, n; i_2 = 1, 2, \dots, v$ . Така ситуація виникає в задачах обробки зображень, коли кластеризації підлягають окремі кадри відеопослідовності. Звичайно, для матриці  $x(k)$  можна здійснити векторизацію, перетворивши в  $(nv \times 1)$  вектор. Однак при цьому в процесі оптимізації можуть виникнути труднощі, пов'язані з, так званим ефектом «прокляття розмірності» і втрати просторових кореляційних зв'язків. Саме тому природно піддавати кластеризації безпосередньо матриці, використовуючи матричний вигляд норм, узгоджених з векторними формами цих норм, наприклад, сферичну норму Фробеніуса. Розглянемо тепер матричний метод *harmonic k-means*.

Розвитком методу *k-means*, також використовують евклідову норму, є *harmonic k-means* (НКМ) [17, 68, 152, 156], в основі якого лежить поняття гармонійного середнього. Так, для скалярного набору  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  в якості гармонійного середнього використовується вираз

$$\bar{x}_H = \frac{N}{\sum_{k=1}^N x^{-1}(k)}. \quad (4.12)$$

В [62] показано, що процедура НКМ призводить до результатів близьким до тих, що забезпечуються  $k$ -means, проте ця процедура набагато менш чутлива до початкового завдання центроїдів-прототипів, що особливо важливо при обробці великих масивів інформації. Якщо цільова функція кластеризації векторних спостережень для  $k$ -means має вигляд (4.8), то для НКМ вона приймає форму

$$\begin{aligned}
 E(x(k), c_j) &= \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{\|x(k) - c_j\|^2}} = \\
 &= \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{(x(k) - c_j)^T (x(k) - c_j)}}. \tag{4.13}
 \end{aligned}$$

Якщо ж обробляється вибірка  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  утворена  $(n \times v)$  матрицями  $x(k) = \{x_{i_1, i_2}(k)\}$ , то використання фробеніусової норми в (4.13) відповідно призводить до виразу

$$\begin{aligned}
 E(x(k), c_j) &= \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{\|x(k) - c_j\|^2}} = \\
 &= \sum_{k=1}^N \frac{M}{\sum_{j=1}^M \frac{1}{Sp((x(k) - c_j)(x(k) - c_j)^T)}} \tag{4.14}
 \end{aligned}$$

де  $Sp(\circ)$  – символ сліду матриці,  $c_j, j=1, 2, \dots, M$  – матричні центроїди-прототипи.

Оптимізація цільової функції (4.13) за допомогою підходу, запропонованого в [68], дозволяє отримати оцінку виду

$$\begin{aligned}
c_j &= \frac{\sum_{k=1}^N \|x(k) - c_j\|^{-4} \left( \sum_{j=1}^M \|x(k) - c_j\|^{-2} \right)^{-2} x(k)}{\sum_{k=1}^N \|x(k) - c_j\|^{-4} \left( \sum_{j=1}^M \|x(k) - c_j\|^{-2} \right)^{-2}} = \\
&= \frac{\sum_{k=1}^N Sp(A(k, j)A^T(k, j))^{-2} \left( \sum_{j=1}^M Sp((A(k, j)A^T(k, j))^{-1})^{-2} x(k) \right)}{\sum_{k=1}^N Sp(A(k, j)A^T(k, j))^{-2} \left( \sum_{j=1}^M Sp((A(k, j)A^T(k, j))^{-1})^{-2} \right)}, \quad (4.15)
\end{aligned}$$

де  $A(k, j) = (x(k) - c_j)$

В роботі [25] запропоновано також оцінювати рівень нечіткої належності на основі (4.15) у вигляді

$$\mu_j(k) = \frac{\|x(k) - c_j\|^{-4}}{\sum_{j=1}^M \|x(k) - c_j\|^{-4}} = \frac{Sp((x(k) - c_j)(x(k) - c_j)^T)^{-2}}{\sum_{j=1}^M Sp((x(k) - c_j)(x(k) - c_j)^T)^{-2}}. \quad (4.16)$$

Дана оцінка близька до оцінки належності на основі FCM [152]

$$\mu_j(k) = \frac{\|x(k) - c_j\|^{-2}}{\sum_{j=1}^M \|x(k) - c_j\|^{-2}}. \quad (4.17)$$

Нескладно бачити, що з обчислювальної точки зору НКМ лише незначно складніше стандартного FCM.

Перейдемо до кластеризації матричних послідовностей на основі harmonic  $k$ -means.

Ясно, що задача кластеризації різко ускладнюється, якщо об'єктом дослідження є не окремі зображення, а їх послідовності, наприклад, представлені у вигляді відеоряду. Задача сегментації і кластеризації для векторних послідовностей розглянута в [25, 102]. При цьому передбачалося, що в розпорядженні дослідника є єдиний часовий ряд, який повинен бути розбитий на однорідні в деякому сенсі ділянки за допомогою тих чи інших алгоритмів кластеризації, включаючи відповідно і нечітку. Якщо ж обробці підлягає безліч рядів, ситуація різко ускладнюється, оскільки на перший план виходить питання оцінки відстані між послідовностями різної довжини. Тут може бути використана процедура Dynamic Time Warping (DTW) [37], яка, однак, стає неефективною, якщо оброблювані сигнали містять досить багато відліків. Таким чином, кластеризації можуть бути піддані тільки короткі послідовності.

У цій ситуації дуже корисною може виявитися процедура Iterative Deepening Dynamic Time Warping (IDDTW) [50], яка, однак, вимагає істотної модифікації в разі, якщо об'єктом досліджень є досить довгі матричні часові ряди.

Стандартна DTW дозволяє визначити «відстань / схожість» між двома часовими (а в нашому випадку ще й матричними) послідовностями різної довжини  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  і  $Y = \{y(1), y(2), \dots, y(l), \dots, y(M)\}$ ,  $N \neq M$   $x(k) = \{x_{i_1, i_2}(k)\}$ ,  $y(l) = \{y_{i_1, i_2}(l)\}$ , на підставі відстані, в нашому випадку сферичної

$$d(x(k), y(l)) = (Sp(x(k) - y(l))(x(k) - y(l))^T)^{1/2} \quad (4.18)$$

між двома окремими кадрами рядів  $X$  та  $Y$ .

Таким чином, вихідною інформацією для розрахунків є  $(N \times M)$  матриця відстаней (4.18) між усіма відліками двох рядів.

В якості запобіжного схожості вибирається так званий «optimal warping path»  $D(N, M)$  який може бути знайдений за допомогою процедури динамічного програмування

$$D(k, l) = d(x(k), y(l)) + \min \{D(k-1, l), D(k, l-1), D(k-1, l-1)\}, \quad (4.19)$$

де  $k = 1, 2, \dots, N; l = 1, 2, \dots, M$ .

Зрозуміло, що при великих  $N$  та  $M$  методи динамічного програмування стають неефективними в силу великої розмірності.

Подолати ці труднощі можливо за допомогою IDDTW, проте використання стандартної процедури [50] також стикається з низкою проблем. При цьому вихідні ряди розбиваються на чотири ділянки (відліки), а на кожному з яких ці послідовності замінюються своїми середніми значеннями. Далі проводиться стандартне оцінювання відстані між матричними рядами, що містять по чотири відліку. Якщо отримана оцінка відстані незадовільна, в розгляд вводиться 8 ділянок, далі шістнадцять і т.д. Це означає, що кількість відліків зростає в геометричній прогресії.

З обчислювальної точки зору набагато зручніше розглядати не  $2^N$ , де  $N = 2, 3, \dots$  ділянок оброблюваних рядів, а - 4, 5, 6, 7, ... Далі для оцінки відстані можна використовувати як вираз (4.19), так і істотно більш просту оцінку (4.18).

Таким чином, якщо обробці підлягає  $Q$  часових рядів  $x_1, \dots, x_q, \dots, x_Q$  кожен з яких містить відліків  $N_1, \dots, N_q, \dots, N_Q$ , отримуємо набір матричних послідовностей  $\bar{x}_1, \dots, \bar{x}_q, \dots, \bar{x}_Q$ , що містять фіксовану кількість  $K$  відліків, тобто  $\bar{x}_q = \{\bar{x}_q(1), \dots, \bar{x}_q(k), \dots, \bar{x}_q(K)\}$ ,  $\bar{x}_q(k) = \{\bar{x}_{qi, i_2}(k)\}$ . При цьому відстань між будь-якими двома відліками різних рядів задається з використанням виразу (4.18)

$$d^2(\bar{x}_q(k), \bar{y}_p(l)) = Sp(\bar{x}_q(k) - \bar{y}_p(l))(\bar{x}_q(k) - \bar{y}_p(l))^T. \quad (4.20)$$

Розглядаючи  $M$  центроїдів  $c_j$ ,  $j = 1, 2, \dots, m$  у вигляді  $(n \times vk)$ -матриць з елементами, що підлягають оцінюванню, можна ввести процедуру кластеризації матричних часових рядів довільної довжини у вигляді

$$\left\{ \begin{array}{l} c_j = \frac{\sum_{q=1}^Q d^{-4}(\bar{x}_q, c_j) \left( \sum_{j=1}^M d^{-2}(\bar{x}_q, c_j) \right)^{-2} \bar{x}_q}{\sum_{q=1}^Q d^{-4}(\bar{x}_q, c_j) \left( \sum_{j=1}^M d^{-2}(\bar{x}_q, c_j) \right)^{-2}}, \\ \mu_j(q) = \frac{d^{-4}(\bar{x}_q, c_j)}{\sum_{j=1}^M d^{-4}(\bar{x}_q, c_j)}, \end{array} \right. \quad (4.21)$$

що є узагальненням методу  $k$ -harmonic means на випадок довгих матричних послідовностей, що містять різну кількість відліків.

### 4.3 Кластеризація відео на основі представників

Для оцінки подібності двох багатовимірних часових рядів, індукованих відеорядами, будемо як і раніше користуватися модифікацією популярного методу динамічної часової деформації (динамічної трансформації часової шкали), який отримав широке поширення для оцінки відстані між послідовностями різної довжини. Суть його полягає в наступному. Введемо в розгляд дві багатовимірних послідовності  $X = \{x(1), x(2), \dots, x(k), \dots, x(M)\}$  і  $Y = \{y(1), y(2), \dots, y(l), \dots, y(N)\}$ ,  $N \neq M$ , де  $x(k)$ ,  $y(k)$  або вектори  $x(k) = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_n(k))$ ,  $y(l) = (y_1(l), y_2(l), \dots, y_i(l), \dots, y_n(l))$  в деякому просторі ознак  $\Omega \subset \mathbb{R}^n$ , або матриці (фрагменти відеокадрів) в просто-

рі сигналів  $x(k) = \{x_{i_1 i_2}(k)\}$ ,  $y(l) = \{y_{i_1 i_2}(l)\}$ ,  $x(k), y(l) \in \mathbb{R}^{h \times v}$ . Далі в розгляд вводиться  $(N \times M)$  матриця відстаней в прийнятій локальній (поелементній) метриці з елементами  $d(x(k), y(l))$ ,  $k = 1, 2, \dots, N$ ,  $l = 1, 2, \dots, M$  між усіма елементами аналізованих послідовностей. На підставі цієї матриці будується «викривлений шлях» у вигляді ряду відстаней  $W = \{w_1, w_2, \dots, w_q, \dots, w_L\}$ ,  $w_q = d(x(k), y(l))_q$ ,  $\max\{N, M\} \leq L \leq M + N - 1$ ,  $q = 1, 2, \dots, L$ , який і визначає подібність між  $X$  та  $Y$  на основі кумулятивної відстані

$$D(k, l) = d(x(k), y(l)) + \min\{D(k, l-1), D(k-1, l), D(k-1, l-1)\}. \quad (4.22)$$

Результуючий «викривлений шлях» може бути представлений у вигляді  $DTW(X, Y) = D(N, M)$ , який, взагалі кажучи, являє собою міру подібності, але не метрику.

Як відстань між елементами  $X$  та  $Y$  зазвичай використовується евклідова метрика

$$d(x(k), y(l)) = \|x(k) - y(l)\|_2, \quad (4.23)$$

яка при обробці фрагментів відеокадрів набуває вигляду метрики норми Фробеніуса

$$d(x(k), y(l)) = (\text{Sp}(x(k) - y(l))(x(k) - y(l))^T)^{\frac{1}{2}}. \quad (4.24)$$

У ситуаціях, коли оброблювані дані зашумлені і мають викиди, доцільно використання манхеттенської метрики, що володіє робастними властивостями. При цьому (4.23) відповідає вираз

$$d(x(k), y(l)) = \sum_{i=1}^n |x_i(k) - y_i(l)|, \quad (4.25)$$



а для (4.24) маємо

$$d(x(k), y(l)) = \sum_{i_1=1}^n \sum_{i_2=1}^v |x_{i_1}(k) - y_{i_2}(l)|. \quad (4.26)$$

Звернемося тепер власне до задачі кластеризації. Введемо в розгляд набір багатовимірних часових рядів  $X_1, X_2, \dots, X_q, \dots, X_Q$ , які повинні бути згруповані в  $m$  кластерів. Передбачається, що кожен з кластерів містить різну кількість спостережень  $N_1, N_2, \dots, N_q, \dots, N_Q$ , відповідно. Безпосередня кластеризація початкових послідовностей представляється малоефективною, оскільки розрахунок (4.22) на основі динамічного програмування на довгих рядах, продукованих відеоданими, пов'язаний з високою обчислювальною складністю. Крім того, послідовності, що оброблюються, як правило, нестационарні, тобто різні сегменти можуть відноситись до різних класів.

У подібних ситуаціях можна скористатися «віконним» підходом, коли кожен ряд  $X_q$ ,  $q=1, 2, \dots, Q$  розбивається на  $P$  ділянок-«вікон», в результаті чого виникає набір нових рядів

$$WX_{q1}, X_{q2}, \dots, WX_{qp}, \dots, WX_{qP}, \quad p=1, 2, \dots, P,$$

які розглядаються далі як незалежні послідовності, що підлягають кластеризації. При цьому, якщо вікна  $WX_{qp}$  одного ряду  $X_q$  потрапляють в різні кластери, що свідчить про нестационарність початкової послідовності, і отримані сегменти можуть класифікуватися як незалежні один від одного сигнали.

Таким чином, в розгляд вводиться  $QP$  сигналів  $WX_{qp}$ , кожен з яких містить  $N_{pq}$  відліків, а кінцевою метою є розбиття цих рядів в режимі самонавчання на  $m$  однорідних у сенсі (4.22) - (4.26) класів.

Звернемося тепер власне до кластеризації на основі представників. Серед відомих алгоритмів кластеризації, як підкреслювалося вище, найбільш широкого поширення набули методи, засновані на прототипах-центроїдах, завдяки простоті обчислювальних моделей і наочності отриманих результатів. Однак ці методи малоприсадибні до кластеризації багатовимірних часових рядів, які продукуються відеопотоками, оскільки оброблювані образи у відомих підходах мають однакову розмірність, а в нашому випадку основний фактор – різні довжини ряду:  $WX_{qp}$  має  $N_{pq}$  елементів.

У подібних ситуаціях кращими можуть виявитися підходи, засновані на використанні зразків-представників, де замість обчислюемого прототипу-центроїду вибирається один з векторів (матриць)-образів, наявних в оброблюваній вибірці

$$WX_{11}, \dots, X_{1p}, \dots, WX_{21}, \dots, WX_{qp}, \dots, WX_{QP},$$

що містить  $QP$  часових сегментів.

Робота алгоритму починається з вибору початкових представників  $WX_{qp}^1(0), WX_{qp}^2(0), \dots, WX_{qp}^m(0)$ , де  $m$  – задане число кластерів. Як  $WX_{qp}^1(0)$  вибирається елемент вибірки найбільш віддалений від усіх інших, тобто  $DTW(WX_{qp}^1(0), WX_{qp}(0)) > DTW(WX_{rs}, WX_{qp})$ ,  $\forall q, p \in \{1, 2, \dots, Q\}$ ,  $\forall r, s \in \{1, 2, \dots, P\}$ .

Далі  $WX_{qp}^1(0)$  тимчасово виключається з вибірки і вибирається другий представник, як і раніше, найбільш віддалений від залишених у вибірці елементів, тобто

$$DTW(WX_{qp}^1(0), WX_{qp}(0)) > DTW(WX_{qp}^2(0), WX_{qp}(0)) > DTW(WX_{rs}, WX_{qp}).$$

Ця процедура вибору повторюється  $m$  раз до формування початкових образів таких, що

$$DTW(WX_{qp}^1(0), WX_{qp}(0)) > \dots > DTW(WX_{qp}^m(0), WX_{qp}(0)) > DTW(WX_{rs}, WX_{qp}).$$

Дана процедура вибору початкових образів ефективна, якщо вихідні дані не містять аномальних викидів. В іншому випадку початкові представники можуть бути обрані випадковим чином, подібно до того, як це трапляється в жадібних алгоритмах кластеризації, наприклад, в класичному алгоритмі  $k$ -середніх. Стосовно до обробки відеопотоків подібні викиди можуть являти собою, наприклад, часові відрізки вставленої рекламної продукції, кластеризація навколо яких семантично безглузда. Інакше кажучи, попередня обробка структурованої вибірки  $WX_{11}, \dots, X_{1p}, \dots, WX_{21}, \dots, WX_{qp}, \dots, WX_{QP}$  набуває особливого значення.

На другому етапі  $QP - m$  образів, що залишилися, розподіляються по кластерам  $Cl_j$ ,  $j = \overline{1, m}$  згідно співвідношенню:  $WX_{qp} \in Cl_j$ , якщо

$$DTW(WX_{qp}, WX_{qp}^j(0)) < DTW(WX_{qp}, WX_{qp}^l(0)) \quad \forall j \neq l \in \{1, 2, \dots, m\}.$$

Таким чином, всі наявні образи групуються в околах кожного з початкових представників  $WX_{qp}^j(0)$ .

У кожній з сформованих груп на третьому етапі розраховується новий представник, в якості якого вибирається спостереження з мінімально сумарною відстанню до всіх точок початкового кластера. Для уточненого представника справедливі співвідношення:  $WX_{qp} = WX_{qp}^j(1)$ , якщо

$$\sum DTW(WX_{qp} \in Cl_j, WX_{qp}^j(1)) < \sum DTW(WX_{qp} \in Cl_j, WX_{rs} \in Cl_j).$$

Після виявлення  $m$  уточнених представників  $WX_{qp}^j(1)$ ,  $j = \overline{1, m}$  відбувається повернення до другого етапу, де представники приписуються знову сформованим представникам згідно з правилом:  $WX_{qp} \in Cl_j$ , якщо

$$DTW(WX_{qp}, WX_{qp}^j(1)) < DTW(WX_{qp}, WX_{qp}^l(1)) \quad \forall j \neq l \in \{1, 2, \dots, m\}.$$

Цей процес повторюється до тих пір, поки всі представники не перестануть змінюватися, тобто для всіх  $j$  від 1 до  $m$  виконується умова стабілізації  $WX_{qp}^j(\alpha + 1) = WX_{qp}^j(\alpha)$ , де  $\alpha = 0, 1, 2, \dots$  є номером ітерації.

Початкові ряди  $X_q$  можна отримати з  $WX_{qp}$  в кінці процедури. Якщо для будь-якого елемента  $q$ , всі елементи належать одному кластеру, це означає, що ряд  $X_q$  є стаціонарним і не змінює властивостей на часовому інтервалі від 1 до  $N_q$ . В іншому випадку маємо, що на таких сегментах відбулися зміни у властивостях часових рядів, і потрібний більш докладний аналіз відповідних сегментів ряду.

Звертаючись до досліджуваних даних, а саме до відеопотоків, необхідно відзначити одну важливу для застосунків деталь. Раціональна ідентифікація візуального змісту в основному у випадках ефектів редагування (плавне зникнення/прояв, розчинення, збільшення розмиття і різкості відеокадрів і т.п.) в силу меншовартості для реферування відео (але не для сегментації часових рядів) вимагає досить «тонких» алгоритмів аналізу змісту відеопотоку. Тому існують достатні підстави для видалення відеопереходів на етапі класифікації, і серйозну увагу слід приділити виявленню послідовностей кадрів, які, мабуть, належать двом послідовним у часі темпоральним сегментам.

## Висновки по четвертому розділу

1. Введена процедура кластеризації послідовностей відеоданих, заснована на використанні модифікованої ітеративної динамічної часової деформації та подальшої кластеризації редукованих часових рядів на основі матричної нечіткої кластеризації методу гармонійних  $k$ -середніх характеризується простотою (в порівнянні з прототипами) чисельної реалізації, що принципово для обробки відеопослідовностей.

2. На відміну від традиційних підходів обробки відео введені процедури не критичні до початкового вибору прототипів-центроїдів і дозволяє класифікувати дані в умовах класів, які перекриваються, що може бути використано при аналізі відеоданих з нечітко вираженими переходами між сценами.

3. Для виявлення просторових і часових подій в відеопотоках з метою реферування відео в цілому перспективним підходом є кластеризація послідовностей різної довжини на основі представників (безпосередньо підмножин відео або набору ключових кадрів) з використанням динамічної трансформації часової шкали.

Список використаних джерел у даного розділі наведено у повну списку використаних джерел під номерами: [17, 25-26, 29, 36-37, 42, 50, 55, 62, 68, 77, 88-90, 94, 102, 116, 128, 147, 151-152, 155-156, 162].

## 5 ЕКСПЕРИМЕНТАЛЬНИЙ АНАЛІЗ МЕТОДІВ СЕГМЕНТАЦІЇ ТА КЛАСТЕРИЗАЦІЇ ВІДЕО

Процес відбору, відображення, аналізу і збереження візуальної інформації, визначеної запитом – суть технології інформаційного пошуку. Пошук за критеріями формальної і смислової відповідності згідно зі зростом обсягів накопичуваної візуальної інформації представляє важливий напрям між-діслінарних досліджень. Основна спонукальна причина розвитку моделей і методів пошуку – мінімізація розриву між інформаційною потребою та інформаційним запитом, тобто стабільне забезпечення релевантності вилучення візуальної інформації. У випадках роботи з відеопотоками – типовими представниками складної неструктурованої інформації великих і надвеликих обсягів – на перший план виходять процедури просторово-часової on-line обробки. У такій обробці особливу роль відіграють методи сегментації та кластеризації відеопослідовностей, представлених або багатовимірними часовими рядами в просторах ознак, або безпосередньо матричними рядами, відповідними фрагментами відеокадрів.

Розділ присвячений обговоренню специфіки обробки відеопотоків в задачах відеореферування для інформаційного пошуку.

### 5.1 Специфіка технології пошуку динамічної візуальної інформації на базі відеореферування

Існують різні стратегії інформаційного пошуку, які залежать від типу пошукового завдання, критеріїв видачі та характеру діалогу між споживачами інформації [49, 84, 86, 141, 142, 153, 158, 164]. У найзагальнішому вигляді технологія інформаційного пошуку складається з наступних етапів: уточнення ін-

формаційної потреби і формулювання запиту, визначення сукупності носіїв-власників інформації, вилучення необхідної інформації, пред'явлення отриманої інформацією та оцінки результатів пошуку, щонайменш, двох його складових: повноти (нічого не втрачено) і точності (запиту, не знайдено нічого зайвого) [66, 67, 74, 91, 92, 103, 159]. Ясно, що зворотний зв'язок дозволяє поліпшувати результати пошуку, як мінімум, за рахунок уточнення запиту.

Стосовно до відеопотоків виникає цілий ряд специфічних моментів, багато в чому зумовлюючих результат пошуку і його валідність. На рис. 5.1 приведена узагальнена схема відеопошуку. Вже на стадії отримання відеоряду з метою зближення on-line і real time реалізацій нерідко проводиться проріджування відеокадрів, як правило, на довжину відеобуферу. Тоді при виникненні явних змін властивостей відеоряду на основі дихотомії виконується ретроаналіз, стартуючи з наступного після відкидання відеокадру. Тут же можуть застосовуватися традиційні для відеопослідовностей прийоми обробки окремих відеокадрів. Перш за все слід вказати передискретизацію та переквантування, результати яких не тільки дозволяють скоротити час подальшої обробки, але і забезпечують раціональні по точності можливості обчислення дескрипторів. Саме на основі цих дескрипторів синтезуються часові ряди, темпоральна сегментація яких створює елементи подальшої обробки – кластеризації та сумаризації відеоданих. На цьому ж етапі (але частіше при темпоральній сегментації) можуть бути виключені з розгляду «довгі» межі між сегментами, що виникають за рахунок ефектів розчинення, затемнення, прояви кадру в кадрі, згортання і тому подібне. Видалення подібних сегментів не позначається на відеореферірованні, але може впливати на пошук сегментів.

Отже, після отримання послідовності відеокадрів для кожного з них виконується обчислення ознак. Конкретизація вибору низькорівневих ознак, з одного боку, має на меті зниження семантичного розриву з вимогами семантики, з іншого, спрямована на мінімізацію обчислювальної складності.

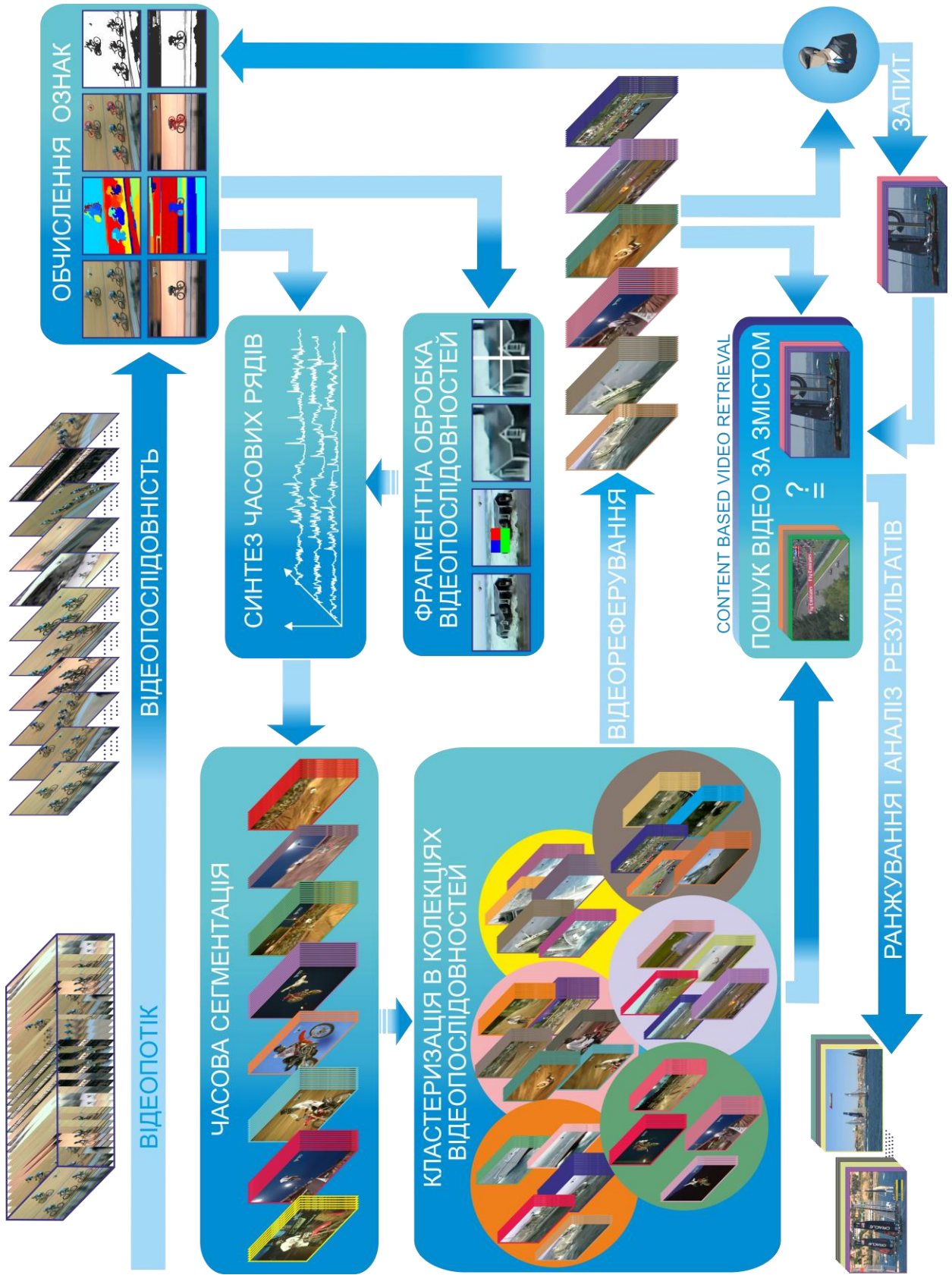


Рисунок 5.1 – Схема інформаційного пошуку відео на основі реферування



Ознаки обчислюються в просторі зображень, їх сегментації, на решітці або фрагментах, що описують околиці особливих точок, на носіях об'єктів, що відповідають бінаризації відеокадрів. В даний час найбільш перспективними є два підходи, спрямовані на використання в глобальному разі сегментації зображень як ознак, а в локальному – використання околиць локальних точок, що, взагалі кажучи, є проміжним етапом отримання валідних прізнакових описів для усунення семантичного конфлікту між низькорівневими ознаковими описами зображень, мірами подібності та високорівневими концепціями їх сприйняття. Перспективність фрагментної обробки (облік локальностей) послідовностей відеокадрів полягає в тому, що області візуального інтересу представляють собою підмножини особливих точок, інструментарій виділення яких вже досить добре розвинений.

Багатовимірні часові ряди (в тому числі і матричні) в роботі розглядаються в двох аспектах: абсолютному і відносному моделюванні властивостей часового ряду. У першому випадку фіксуються значення ознак і після попередньої нормалізації темпоральна сегментація ведеться на основі обчислюваної on-line ознакової відмінності відеокадрів. Головне достоїнство полягає в простому підключенні передісторії, що особливо важливо при малопомітних (плавних) змінах властивостей часового ряду. У іншому випадку сегментація фактично проводиться на відмінностях відеокадрів, що нерідко дає прямий результат пороговим відсіканням. У більш складних випадках при таких моделях ефективними стають непрямі методи сегментації-кластеризації, оскільки, як правило, статистичні властивості відеокадрів постійні всередині окремих сегментів і відрізняються від сегменту до сегменту.

На рис. 5.2. і 5.3 приведені фрагменти відносного часового ряду для одних і тих самих вихідних відеопотоків. У першому випадку – усереднення по різниці градієнтів в околицях 10 відповідних особливих точок, у другому – різниця між двома розбиттями, отриманими графовим методом. Видно, що в обох випадках піки в основному лежать на межах сегментів, але поведінка ряду оціню-

ється краще на вибіркових моментах всередині сегментів.

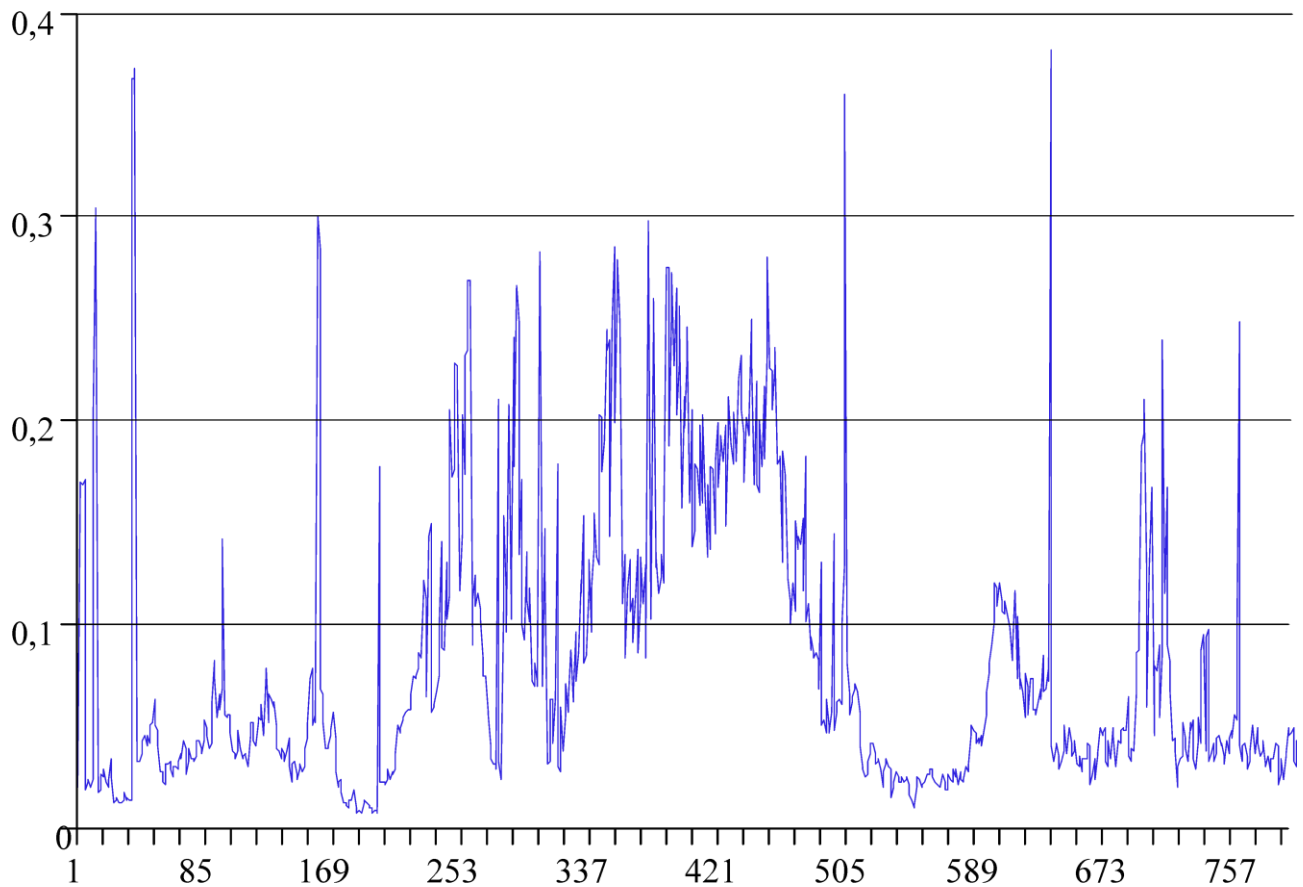


Рисунок 5.2 – Відрізок відеоряду (усередненні фрагментні ознаки)

Наступні етапи відеореферування – сегментація і кластеризація описані в наступних підрозділах. Тут зупинимося на підходах до пошуку із запитами – сегмент або відеоузагальнення. Інформаційний пошук – зазвичай ітераційний процес (з уточненням запиту), що базується на редукції простору перебору, а також в аспекті скорочення розмірності даних. Він може поділятися на пошук в фіксованому просторі ознак, ієрархічно упорядкованому просторі атрибутів, в альтернативних просторах або динамічному (змінюється в процесі пошуку) при знаковому просторі.

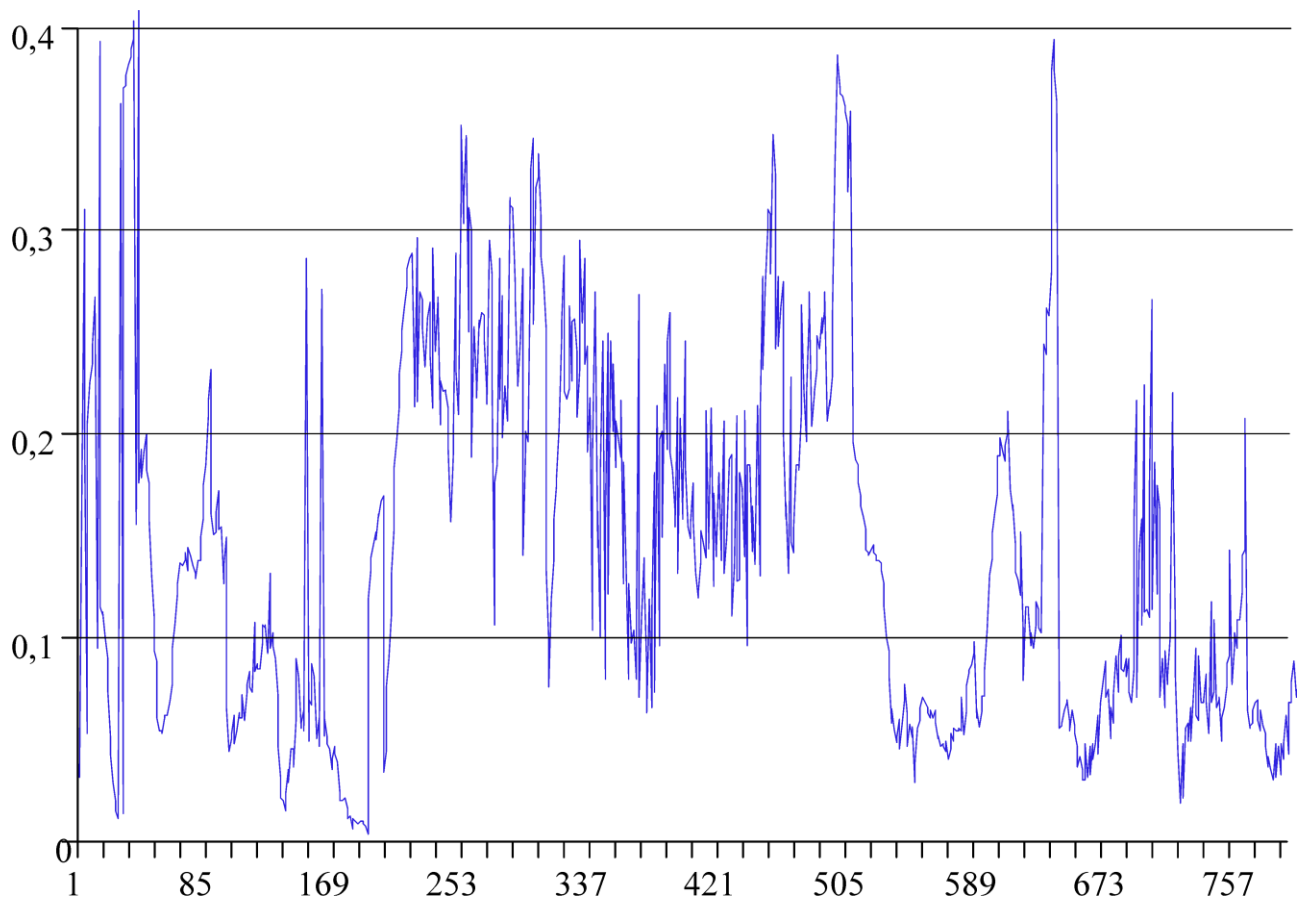


Рисунок 5.3 – Відрізок відеоряду (глобальні ознаки, данні з рис. 5.2)

Для підвищення валідності пошуку з використанням сумаризації відеопотоків в технології вилучення візуальної інформації доцільно застосовувати засновані на навчанні методи асоціації низькорівневих ознак з високорівневими поняттями, генерація смислових шаблонів для підтримки семантичного пошуку, онтології об'єктів для визначення високорівневих концепцій, зворотній зв'язок з Користувачем для уточнення запиту.

Пошук з відеоузагальненням має одну важливу особливість в порівнянні з традиційними моделями, а саме: запити можуть формуватися традиційно у вигляді сегментів, а й власне відеореферат, згенерований по спостережуваному відео, може використовуватися в якості запиту. Механізм пошуку зазвичай залишається без зміни: пошук виконується на множині сегментів. Однак для підвищення перцептивності відеоузагальнення може сприйматися як запит. Тоді в

наборах відеорядов необхідно проводити своєрідну каталогізацію, яка фактично відповідає стратифікованій структуризації колекції відеопотоків.

Все вищесказане стосується більшою мірою до динамічного відеореферування, однак статичний підхід в технології інформаційного пошуку не можна скидати з рахунків. По-перше, статичне відеореферування більш швидкодіючий і, отже, може використовуватися для оціночного аналізу, але що більш важливо – в даний час вже є засоби категоризації зображень, чого не можна сказати про відеодані. Аналіз категоризованих ключових кадрів помітно підвищує семантичність пошуку динамічної візуальної інформації.

## 5.2 Експериментальний аналіз ознакових просторів при просторово-темпоральній сегментації відео

Експерименти проводили з відеорядами «новинного» змісту, тривалість кожного з них становила не менше хвилини, тобто не менш як 1500 кадрів. Використовувався медіаконтейнер AVI з відношенням сторін 16:9 і 4:3, частотою 25 кадр/сек. Роздільна здатність фіксувалася на рівні  $1280 \times 720$  та  $720 \times 576$  пікселів відповідно. Кожен з кадрів кожного відеоряду представлявся у форматі TIFF без стиснення з метою усунення впливу можливих втрат за рахунок згладжування кольорів, що особливо важливо для визначення ефективності часової сегментації. Валідність сегментації багатовимірних часових рядів, асоційованих з відеопотоками, визначалася за рахунок порівняння з ground-truth сегментацією, виконуваною інтерактивно on-line з подальшим покадровим уточненням. Якість кластеризації і відеореферування визначались з позицій пертіненності.

Вибір ознакових просторів для сегментації та кластеризації відеоданих є принциповим питанням, відповідь на який забезпечує не тільки можливості побудови розбиття і покриття відеорядів, але в підсумку і валідність інформаційного пошуку. Для побудови ознакових просторів, які фактично продукують ча-

сові ряди, що підлягають сегментації, використовуються властивості кольору, текстури, форми окремих областей та їх сукупностей, взаємні координатні опису, множини характерних точок. Кольорові ознакові описи зазвичай асоціюються з аналізом гістограм, а для обліку просторових характеристик розподілу кольору – вектори когерентності кольору, різновиди коррелограмм (автокоррелограмми, зв'язкові коррелограмми).

Текстури, що характеризують просторовий розподіл кольорів, ґрунтуються на статистичному підході, на відповідність сприйняття людини, аналізі спектральних характеристик (вейвлети, фільтри Габора, розкладання в базисі Уолша та ін.).

Ознаки форми представляють собою найбільш перспективний інструмент для усунення семантичного конфлікту, проте їх використання істотно залежить від точності алгоритмів бінаризації або в загальному випадку – просторової сегментації відеокадрів.

Серед ознакових описів форми слід виділити моментні інваріанти, опису контурів кутами повороту, частотні ознаки Фур'є, інваріантні геометричні моделі форм. Існує цілий ряд параметрів областей, які легко описати у вигляді кількісних характеристик, достатніх для повноцінного рішення про зміну контенту відеокадру.

У більшості ж випадків ці характеристики можуть використовуватися як деякі фільтруючі елементи, що відсівають незначні області. До таких (інтуїтивно зрозумілих) характеристик варто віднести площу, периметр, центр області інтересу, а також радіуси максимальних вписаних та мінімальних описаних кіл навколо області інтересу, а також довжини півосей еліпсу із заданою площею та периметром, діаметр Мартіна, кут мінімальних моментів, параметр Вейделла, кутовий параметр Фішера, проекцію Пентланда, параметри округлості Пірарда і Венворта і цілий ряд інших.

Слід особливо підкреслити, що прагнення до зниження семантичного конфлікту дозволяє в якості основної бази визначення ознак виділити фрагмен-

тну обробку відеокадрів, а також їх сегментацію як ознаки відеокадру [48, 73, 83]. Однак фіксація фрагментів хоча і дає можливість виконувати темпоральну сегментацію, точність її гарантована не завжди. На рис. 5.4 наведено приклад (часові сегменти вказані одним кольором) елементів відео з проріджуванням в 30 кадрів. Звідси безпосередньо випливає, що фрагменти повинні вибиратися адаптивно з урахуванням їх попереднього в часі положення, з можливістю виключення частини з них з розгляду в процесі on-line обробки. Крім того, необхідно враховувати «перезавантаження» набору фрагментів навіть в межах одного відеосегменту. З теоретичних положень роботи слід, що ознака – деяка згортка фрагменту.

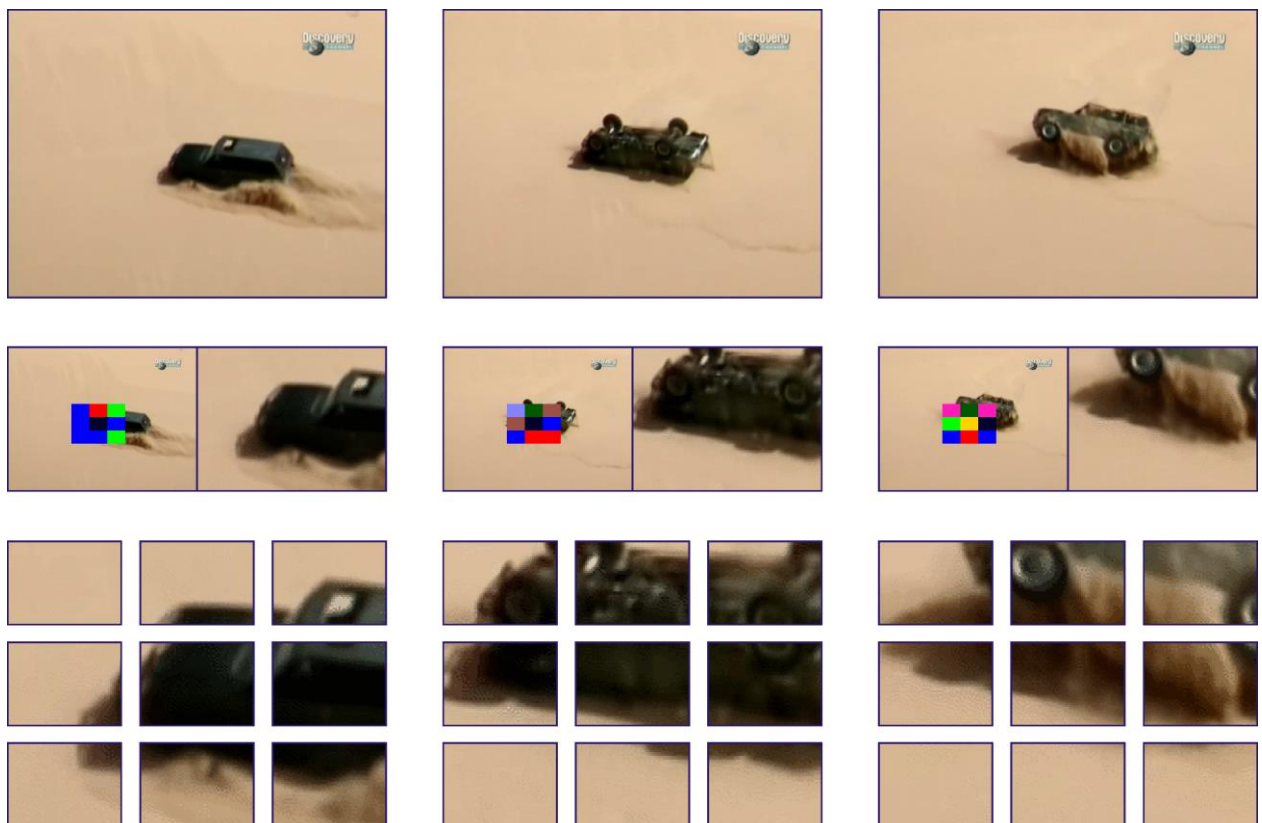


Рисунок 5.4 – Фрагментна темпоральна сегментація з фіксацією положення

Завдання вибору центрів фрагментів може вирішуватися з використанням добре опрацьованих інструментів виділення особливих точок. Основні вимоги,

що пред'являються до виділення точок на зображеннях зводяться до наступного [24, 66, 130, 135, 137,146].

Особлива точка повинна знаходитися в одному і тому ж місці сцени, незважаючи на зміни взаємного розташування об'єкту та відеодатчиків. Це властивість повторюваності (repeatability) стосовно відео означає, що на хронологічно суміжних відеокадрах положення точок всередині сегменту має приблизно збігатися.

На одному відеокадрі околиці особливих точок повинні мати великі відмінності один від одного. Властивість відмінності або інформативності (distinctiveness, informativeness) забезпечують і детектування, і зіставлення особливих точок.

Особлива точка (аж до суперпіксела) повинна займати невелику область. Це властивість локальності (locality) нівелює геометричні й фотометричні спотворення. Що стосується кількості (quantity) виявлених особливих точок, воно повинно бути велике, але адаптивно визначатися з використанням простого і інтуїтивного зрозумілого порога.

Стосовно до відеосегментації можна вибирати особливі точки на решітці так, щоб їх число було приблизно однаково в кожному фрагменті, а щільність розташування особливих точок повинна відображати контент відеокадру.

Властивість точності (accuracy) відображає необхідність точної локалізації точок при геометричних перетвореннях площині, насамперед масштабних трансформацій. Нарешті, ефективність (efficiency) – час виявлення особливих точок на зображенні має бути допустимим в критичних за часом програмах.

Важливо підкреслити, що області візуальної уваги, які багато в чому визначають ground-truth темпоральну сегментацію, є підмножинами областей інтересу, які, в свою чергу, визначаються наборами особливих точок.

Особливі точки знаходилися за допомогою алгоритмів SURF (Speeded Up Robust Features, SIFT (Scale-Invariant Feature Transform), BRISK (Binary Robust Invariant Scalable Keypoints), KAZE та Harris Corner Detector.

На рис. 5.5 показані приклади пошуку меж сегментів на основі аналізу фрагментних послідовностей (з вибором з 20 найбільш потужних відгуків) з виділенням особливих точок алгоритмом SURF. Тут і далі особливі точки показані червоними хрестами на зображеннях.



Рисунок 5.5 – Приклади границь сегментів часової сегментації

Ще одна перевага використання особливих точок, точніше кажучи, адаптивного фрагментного підходу, полягає в можливості додаткового аналізу темпорального сюжету в рамках одного сегменту. Особливі точки, які виділяються різними алгоритмами, як правило, варіативні, але раціональний вибір відповідних попереднім кадрам точок в їх околах на наступних відеокадрах забезпечує адекватний синтез матричних часових рядів. Поряд з цим, відбір точок додатково дозволяє враховувати і часткові затулювання, і рух об'єктів, не порушуючи валідності темпоральної сегментації. На рис. 5.6 з проріджуванням в 10 кадрів наведено приклад руху об'єкта в рамках одного часового сегменту.



Рисунок 5.6 – Приклад руху об'єкту всередині одного сегменту

Інша перевага локальності – ігнорування часткового затулення на відео ілюструється рис. 5.7.





Рисунок 5.7 – Можливість часткових затулень всередині відеосесегменту

Як неважко помітити, часткові затулення в даному випадку не мають жодного впливу на темпоральну сегментацію. Проте на практиці далеко не завжди все так благополучно. На рис. 5.8 показані особливі точки, виділені SURF (зліва) і KAZE (праворуч). Видно, що в мультимедійній ситуації алгоритми працюють по-різному. Подібні ситуації повинні не впливати на генерацію матричних часових рядів.



Рисунок 5.8 – Вплив накладання графіки на виявлення особливих точок

На рис. 5.9 показані результати часової сегментації відеоряду з 1500 відеокадрів, тобто тривалістю в одну хвилину. На графіку (внизу) досить чітко виділяються межі сегментів, вгорі показані власне граничні відеокадри.

Тут слід звернути увагу на те, що відеокадри 131-165 виключені з матеріалів для відеореферування, оскільки в початковому відео представляли собою зміну сюжетів повільним «проявом» граничних кадрів. Відзначимо, що результати експериментальних досліджень підтверджують перспективність фрагментної обробки.

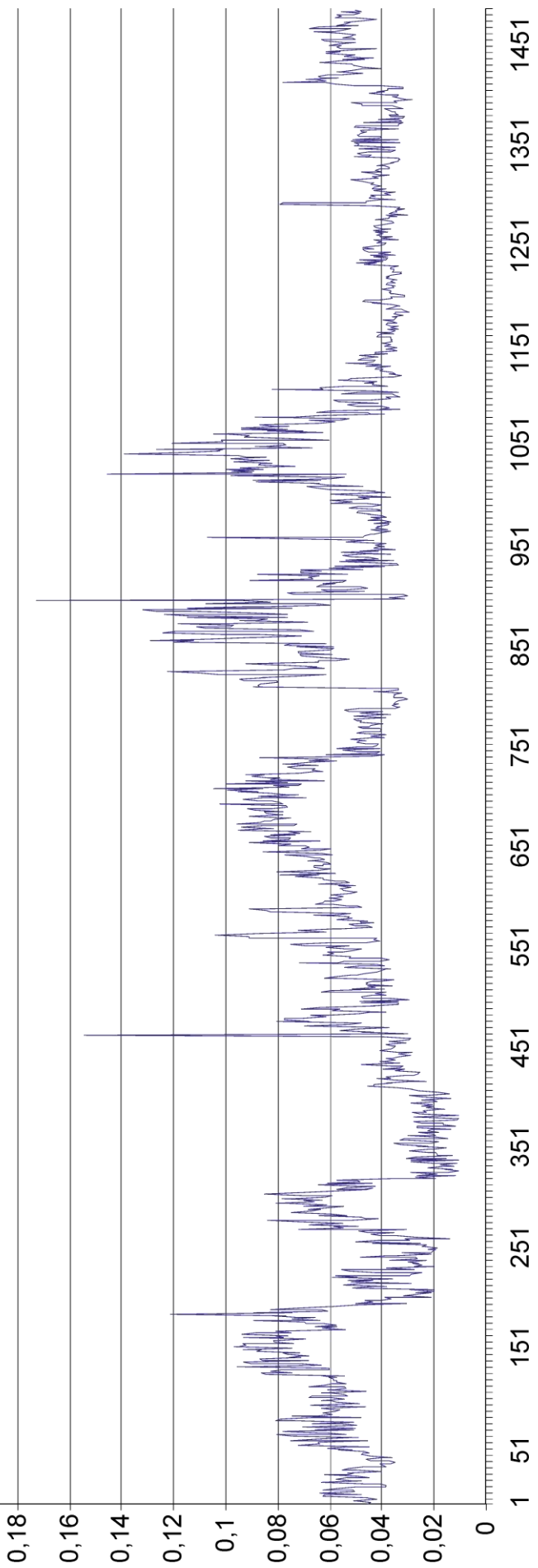


Рисунок 5.8 – Результат темпоральної сегментації відеопотоку

Розбиття і покриття поля зору з елементами, коррелюючими з об'єктами сцен, – один з багатообіцяючих шляхів моделювання змісту відеокадрів для виявлення змін властивостей відеорядів. Сегментацію можна, що стало вже традиційним, використовувати для обчислення векторів ознак форми по окремим областям, але можливі неточності детектування областей, а головне – пошук, в умовах суттєвої невизначеності, розумного компромісу між надмірною і недостатньою сегментацією, знижують ефективність даного підходу.

Як вже зазначалося вище, результати просторової сегментації можуть виступати як ознака середнього рівня, оскільки розбиття як «множина множин» відображає сукупну геометричну інформацію, що включає в себе взаємне розташування всіх значущих областей. Крім того, аналіз вкладених факторно-множин в результаті може забезпечити обробку окремих областей, що корелюють з шуканим об'єктом, незалежно від фонових компонентів відеокадрів.

Для порівняння сегментацій використана нормована метрика [9]

$$\rho(X, Y) = \frac{\sum_{\alpha} \sum_{\beta} \text{card}([x]_{\alpha} \Delta [y]_{\beta}) \text{card}([x]_{\alpha} \cap [y]_{\beta})}{1 + \sum_{\alpha} \sum_{\beta} \text{card}([x]_{\alpha} \Delta [y]_{\beta}) \text{card}([x]_{\alpha} \cap [y]_{\beta})},$$

де  $X = \{[x]_{\alpha}\}$ ,  $Y = \{[y]_{\beta}\}$  – результати сегментації двох послідовних або з проріджуванням відеокадрів ( $X$  або  $Y$  також може бути ground-truth сегментацією);

$[x]_{\alpha} \Delta [y]_{\beta}$  – симетрична різниця;

$X = \cup_{\alpha} [x]_{\alpha}$ ,  $\forall \alpha' \neq \alpha'' \Rightarrow [x]_{\alpha'} \cap [x]_{\alpha''} = \emptyset$ .

На рис. 5.10 наведені результати графової сегментації відеокадрів трьох різних відеорядів. Якщо зміст рис. 5.10, а) легко інтерпретується по сегментації в двох з чотирьох випадків, рис. 5.10, б) завжди правильно трактується, а по рис. 5.10, в) інтерпретації немає, використання метрики для розбиття завжди



дає результат порівняння, що забезпечує темпоральну сегментацію відеорядів.

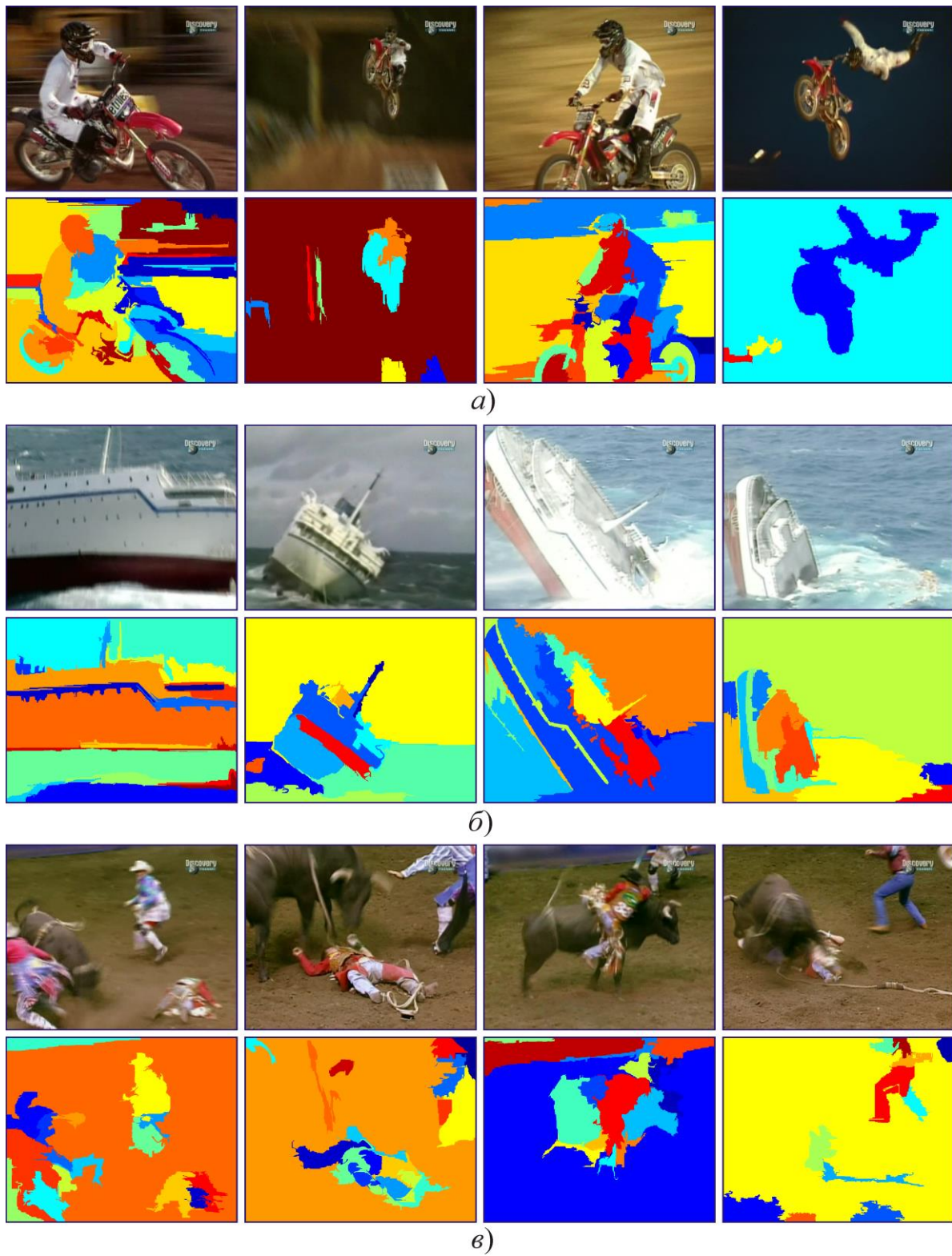


Рисунок 5.10 – Результати просторової сегментації відеокадрів

### 5.3 Дослідження особливостей кластеризації відеорядів для відеореферування

Рішення завдання отримання часових рядів, які адекватно відображають семантичні властивості відеопослідовностей, пов'язане не тільки і не стільки з обчислювальною складністю, скільки з необхідністю отримувати темпоральні ряди, пов'язані з «змістом» відео, в усякому разі, в плані однорідності відеосегментів [34, 52, 73, 81, 108, 149, 160, 162]. Ця однорідність (в сенсі заданого критерію) визначає структуру відеоряду, а в підсумку зумовлює якість автоматичного динамічного відеореферування для інформаційного пошуку із запитом «за зразком». Тим самим, при повнокадровому аналізі навантаження лягає на валідність просторової сегментації хронологічно пов'язаних зображень. При фрагментній обробці важливим моментом є їх вибір і встановлення міжкадрового зв'язку. Зв'язок в основному визначається розташуванням околиць особливих точок в полі зору шляхом аналізу близькості їх характеристик, наприклад, центрів тяжіння. Це легко пояснити високою частотою надходження зображень, що дає малі зміни в рамках одного відеосегменту. Такий підхід дозволяє (що найменш, на множині відео, використаних в експериментальних дослідженнях) знаходити ті ж семантичні межі сегментів, які були стійко виділені інтерактивно.

На рис 5.11 наведені приклади сегментації відеорядів. У лівому верхньому ряду розташоване аналізоване відео (на ілюстрації показаний окремий відеокадр). Нижче наводяться результати виділення особливих точок, відбір найбільш репрезентативних серед яких забезпечується фрагментна обробка. Крім того тут наводяться результати просторової сегментації кожного відеокадра. Наступний стовпець відображає власне результати темпоральної сегмен-

тації: послідовно змінюються рядки-відеосегментів, доповнені ключовим кадром і «усередненим» відеокадром.

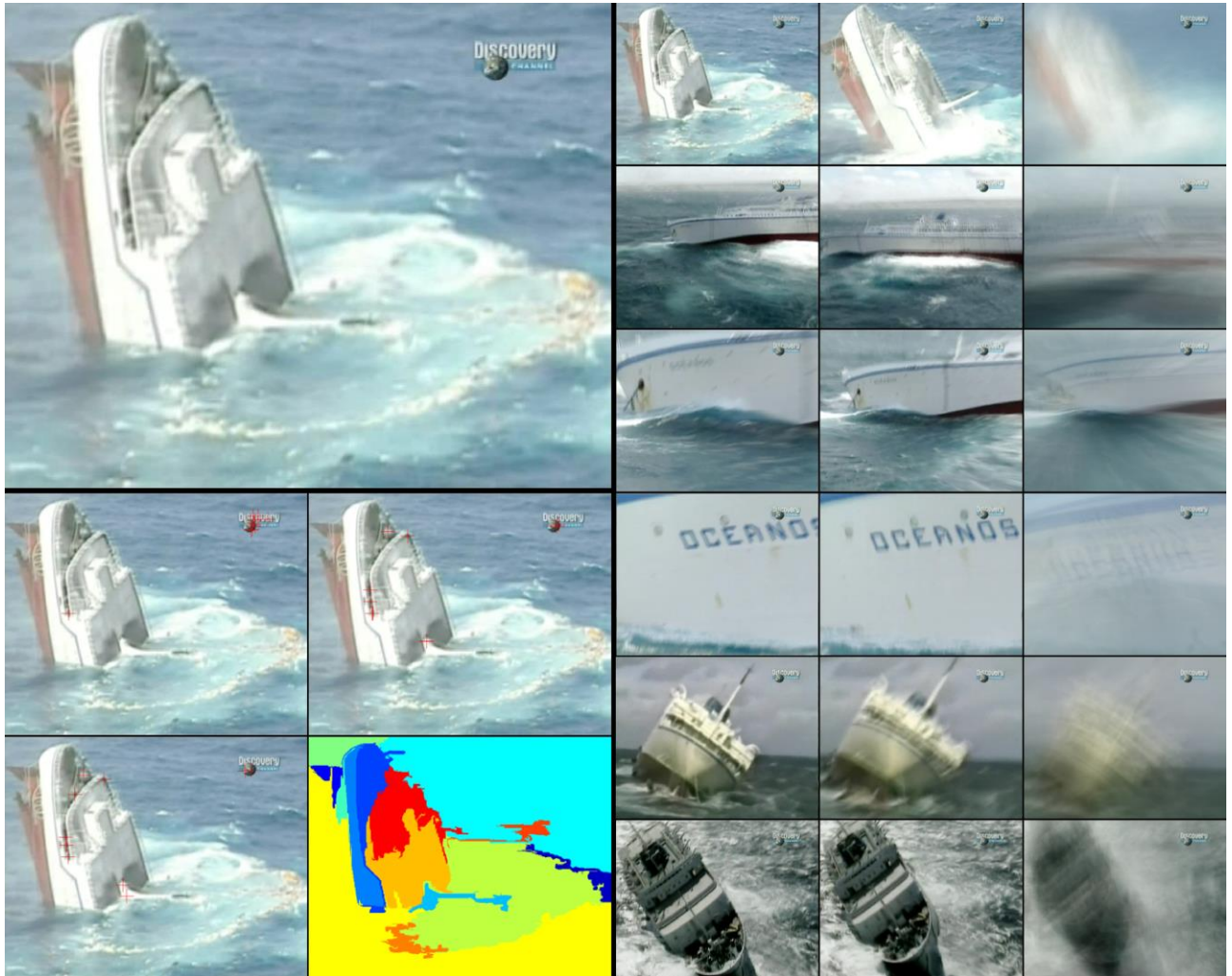


Рисунок 5.11 – Результати темпоральної сегментації

Аналогічні результати (для інших алгоритмів виділення особливих точок і параметрів просторової сегментації) наведені на рис.5.12.

При on-line обробці схожі в часі відеокадри, подібні за формою або структурою індуковані багатовимірними і, зокрема, матричними часовими рядами досить успішно дозволяють семантично сегментувати відеопослідовності саме в силу хронологічної та змістовної зв'язності. Завдання ефективної off-line кластеризації ще далеке від свого рішення внаслідок різних довжин відеосег-



ментів з довільним змістом [94, 106, 109, 152, 161].

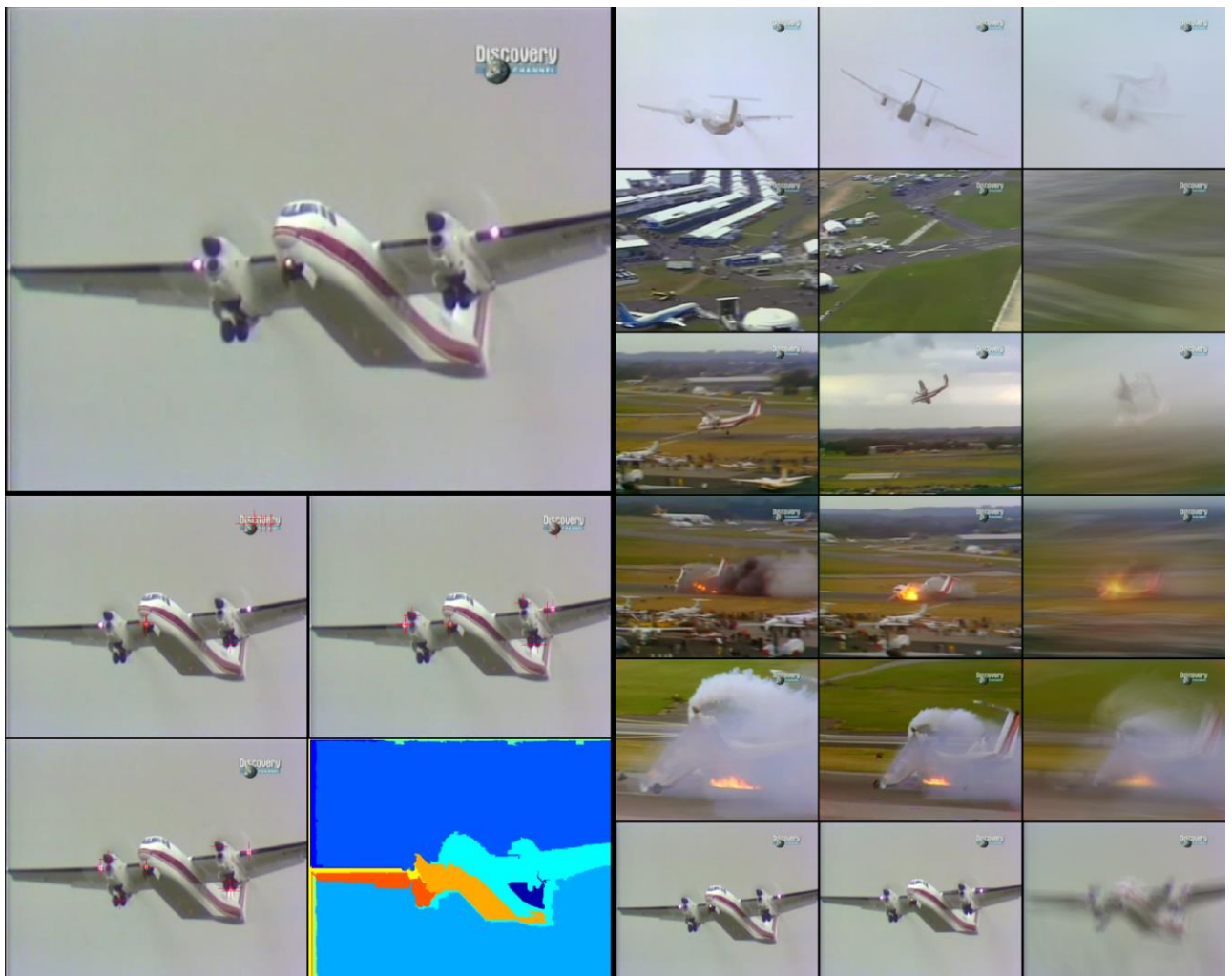


Рисунок 5.12 – Результати темпоральної сегментації

Крім того, слід підкреслити, що алгоритми кластеризації можуть бути тільки медоїдного типу і зазвичай повинні допускати мультіаспектність (в залежності від спостережуваних подій, сюжетів, сценаріїв). Використовувані в дисертації DTW та метрика Левенштейна безумовно є в даний час найбільш перспективними для динамічного відеореферування, проте їх обчислювальна складність і часто принципова неясність при інтерпретації компаративного аналізу в ході кластеризації вимагають пошуку нових підходів.

На рис. 5.13 показаний відеокадр кластеризації відеосегментів, на якому

представлені: для ключового представника кожного кластеру ключовий відео-кадр та усереднене зображення всього сегменту. Праворуч показані всі предста-



Рисунок 5.13 – Результати кластеризації відеосегментів

вники даного кластеру. Необхідно відзначити, що на відміну від наведених на рис. 5.13 представників, вони частіше виявляються семантично неоднорідними. Це ще раз підкреслює необхідність акцентування досліджень на «семантичних» метриках і заходи порівняння відеорядів.



Підводячи підсумки експериментальних досліджень, слід особливо підкреслити, що відеодані – одне з домінуючих джерел доступних і релевантних даних в традиційних і нових інформаційних технологіях. Однак глобальна безперервно наростаюча акумуляція даних в поганотруктурованих відеоархівах породжує цілий клас складних завдань, пов'язаних з встановленням семантичної відповідності в умовах апріорної невизначеності та обмежень на час прийняття рішень в умовах, як правило, недостатності обчислювальних ресурсів. До таких завдань відносяться синтез заходів і метрик, призначених для порівняння відеоузагальнень різної тривалості з достатнім рівнем моделювання сприйняття візуальної інформації людиною, як мінімум, в предметно-орієнтованій області, усунення семантичного конфлікту між низькорівневими моделями відео і високорівневими цілями їх трактування, оцінка ефективності багатоаспектного асоціативного пошуку на базі стратифікованої сумаризації. Для цього потрібно міждисциплінарна багатофакторна інтеграція методів обробки багатовимірних часових рядів, розпізнавання образів, комп'ютерного зору в рамках вдосконалення інформаційних технологій пошуку.

#### Висновки по п'ятому розділу

1. Динамічне відеореферування для інформаційних технологій контекстного пошуку надає можливості підвищення рівня семантики за рахунок використання хронологічно не зв'язаних узагальнених запитів, сформованих на основі знайдених структурних елементів відеопослідовностей.

2. З результатів експериментальних досліджень on-line обробки асоційованих з відеоданими зростаючих багатовимірних часових рядів, особливо виділяючи матричні темпоральні послідовності, впливає що найбільшу валідність забезпечують просторова сегментація відеокадрів і набори фрагментів з центрами в особливих точках, що переважно

характеризують зони візуальної уваги,

3. Сегментація часових рядів, індукованих відеоданими, кластеризація отриманих сегментів з урахуванням локальних або глобальних властивостей сюжетів з урахуванням диверсифікації вирішуваних завдань і інтенсифікації обробки відеопотоків є досить уніфікованим інструментарієм, який може бути покладений в основу різних програмних комплексів.

Список використаних джерел у даного розділі наведено у повну списку використаних джерел під номерами: [9, 24, 34, 48, 49, 52, 73, 81, 83, 84, 86, 108, 130, 135, 137, 141, 142, 146, 149, 153, 158, 160, 162, 164].

## ВИСНОВКИ

В дисертаційній роботі для розвитку технологій інформаційного пошуку вирішена актуальна науково-практична задача розробки моделей і методів сегментації та кластеризації відео них для динамічного відеореферування. Проведені дослідження дозволяють зробити такі висновки.

1. На основі аналізу трендів розвитку методів реферування, анотування і сумаризації відеопотоків з метою підвищення змістовної релевантності інформаційного пошуку встановлена перспективність динамічного відеореферування. Показано, що відомі методи сегментації, кластеризації та узагальнення відеопослідовностей не в повній мірі задовольняють сучасним вимогам і потребують суттєвого удосконалення.

2. Оперативне виявлення змін властивостей векторних і матричних нестационарних збурених сигналів на основі ансамблю адаптивних моделей з власними алгоритмами ідентифікації з різною глибиною пам'яті створює передумови для сегментації відеорядів з урахуванням поліструктурності і багатозначності окремих частин відеопотоків.

3. Введені on-line методи нечіткої кластеризації дозволяють вирішувати завдання виявлення розладнань в багатовимірних часових рядах як для повільних, так і стрибкоподібних змін змісту відео, що забезпечує високу релевантність темпорального аналізу відеопотоків.

4. Вперше для виявлення змін властивостей багатовимірних послідовностей різної довжини на основі ітеративної динамічної часової деформації, в основі якої лежить адаптивна редукція часових рядів, дозволяє отримувати модифікації алгоритмів кластеризації векторних і матричних послідовностей з метою їх реферування з підвищеною пертінентністю.

5. Вперше запропонований метод кластеризації багатовимірних часових рядів різної довжини в умовах невідомої кількості класів і можливості їх взаємного перетину на основі гібридизації ієрархічного агломеративного і

нечіткого, заснованого на центрах ваги, підходів з використанням метрики Левенштейна забезпечує можливість валідного пошуку зі складними реферованими запитами.

6. Отримали подальший розвиток методи налаштування параметрів адаптивної матричної моделі та виявлення змін властивостей матричної послідовності забезпечують скорочення розриву між формальною і семантичною релевантністю за рахунок того, що фрагменти, які є результатом довільної просторової згортки, характеризують околиці точок, що забезпечують візуальну увагу.

7. Удосконалено метод кластеризації послідовностей відеоданих, заснована на використанні модифікованої ітеративної динамічної часової деформації і подальшої кластеризації редукованих часових рядів на основі матричної нечіткої кластеризації на базі гармонійних  $k$ -середніх володіє простою обчислювальною моделлю і підвищеною стійкістю до локальних часових деформацій.

8. Для виявлення просторових і часових подій в відеопотоках з метою семантичного реферування відео перспективною виглядає кластеризація послідовностей різної довжини на основі представників з використанням динамічної трансформації часової шкали.

9. Розроблені методи сегментації і кластеризації відеопослідовностей для вирішення задач керування, у тому числі в інтерактивному режимі, великими колекціями неструктурованої динамічної інформації та контекстного пошуку з запитами «за зразком» реалізовані програмно та використані у розробках ДП «УкрНТЦ «Енергосталь», наукових розробках та освітньому процесі ХНУРЕ, що підтверджено відповідними актами.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аналіз та обробка даних засобами обчислювального інтелекту: монографія / Є.В. Бодянський, Д.Д. Пелешко, О.А. Винокурова, С.В. Машталір, Ю.С. Іванов. – Львів: Видавництво Львівської політехніки, 2016. – 236 с.
2. Андерсен, Т. Статистический анализ временных рядов; пер с англ. под ред. И.Г. Беляева. – М.: Мир, 1976. – 755 с.
3. Антощук, С.Г., Годовиченко Н.А. Оценка эффективности моделей событий в задачах семантического анализа видеопотока. Електротехнічні та комп'ютерні системи. – 2014. – №13(89). – С. 144-151.
4. Бодянський Є.В., Машталір С.В. Виявлення змін у потоці відеоданих на основі аналізу багатовимірних часових рядів. Доповіді Національної академії наук України. – 2012. – №11. – С. 30-33.
5. Богучарский С.И., Машталир С.В. Анализ текстур в последовательности изображений на основе векторного квантования. Радиоэлектроника, информатика, управление. – 2014. – №2(31). – С. 94-99.
6. Богучарский С.И., Машталир С.В., Столбовой М.И. Быстрое обнаружение изменения свойств многомерных временных рядов на основе идентификационного подхода к ансамблю моделей. Системи обробки інформації. – 2018. – № 3 (154). – С. 74-78 .
7. Бокс Дж., Дженкинс Г. Анализ временных рядов. – М.: Мир, 1974. – 406 с.
8. Ведмедь А.Г, Машталир С.В., Сакало Е.С. Матричный алгоритм самообучения карты Кохонена в задачах обработки изображений. Системи управління, навігації та зв'язку. – 2009. – № 1(9). – С.188–192.
9. Егорова Е.А., Киношенко Д.К., Машталир С.В., Шляхов Д.В. Метрическое сравнение результатов сегментации изображений. Радиоэлектроника и информатика. – 2006. – №2. – С. 55-62.
10. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и вре-

менные ряды; пер с англ. под ред. А.Н Колмогорова, Ю.Н. Прохорова. – М.: Наука, 1976. – 736 с.

11. Максимів О., Рак Т., Пелешко Д. Зменшення кількості хибних викликів під час розв'язання задачі детектування полум'я у відеопотоці з використанням глибоких згорткових нейронних мереж. Вісник Національного університету «Львівська політехніка». – 2017. – № 864. – С. 268-276.

12. Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск; пер с англ. Д. А. Ключин. – М.: Вильямс, 2011. – 528 с.

13. Мантула Е.В., Машталир С.В. Матричная прогнозирующая модель и ее обучение в задачах экологического мониторинга. Электротехнические и компьютерные системы. – 2013. – №10(86). – С. 152-156.

14. Машталир С.В., Постульга С.В., Щербинин К.С. Поиск изменения сцен в видеоданных на базе анализа результатов сегментации. Бионика интеллекта. – 2010. – № 1. – С. 65-69.

15. Машталир С.В., Столбовой М.И. Адаптивные матричные модели в задаче контроля потоков видео. Радиоелектроніка, інформатика, управління. – 2018. – № 4(47). – С. 188-194.

16. Машталир С.В., Столбовой М.И. Анализ методов пространственно-временной сегментации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали міжнародної наукової конференції. (Залізний Порт, Україна, 22–26 травня 2017). Херсон: Видавництво ПП Вишемирський В.С. –2017. – С.287-289.

17. Машталир С.В., Столбовой М.И., Яковлев С.В. Кластеризация последовательностей видеоданных на основе гармонических k-средних. Кибернетика и системный анализ. – 2019. – Том 55, № 2. – С. 36-43.

18. Машталир С.В., Столбовой М.И., Яременко М.М. Использование матричных моделей анализа временных рядов в задаче кластеризации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту: Матеріали міжнародної наукової конференції

(Залізний Порт, Україна, 21–27 травня 2018). Херсон: Видавництво ПП Вишемирський В.С. – 2018. – С.83-85.

19. Михнова Е.Д. Анализ видеоданных на основе диаграмм Вороного различного порядка. Збірник наукових праць ХУПС. – Харків: ХУПС. – 2014. – № 1(38). – С. 142-145.

20. Орлов Ю.Н., Осминин К.П. Нестационарные временные ряды. Методы прогнозирования с примерами анализа финансовых и сырьевых рынков. – М.: Либроком, 2011. – 384 с.

21. Пелешко Д.Д., Иванов Ю.С. Супровід рухомих об'єктів у локальних системах відеоспостереження міської інфраструктури. Вісник Львівського державного університету безпеки життєдіяльності: збірник наукових праць. – Львів: ЛДУБЖ. – 2013. – № 8. – С. 37-41.

22. Путятин Е.П., Аверин С.И. Обработка изображений в робототехнике. – М.: Машиностроение, 1990. – 320 с.

23. Чуев Ю.В., Михайлов Ю.Б., Кузьмин В.И. Прогнозирование количественных характеристик процессов. – М.: Сов. Радио, 1975. – 400 с.

24. Abdelwahab M.A., Abdelwahab M.M., Uchiyama H., Shimada A., Taniguchi R. Video object segmentation based on superpixel trajectories. In: Campilho A., Karray F. (eds.) Image Analysis and Recognition. ICIAR 2016. Lecture Notes in Computer Science, Vol. 9730. Cham: Springer. – 2016. – P. 191-197.

25. Abonyi J. and B. Feil, Cluster analysis for data mining and system identification. Basel: Birkhäuser, 2007. – 306 p.

26. Abonyi J., Feil B., Nemett S., Arva P. Fuzzy clustering based segmentation of time series. Lecture Notes in Computer Science. Berlin: Springer, 2003. – Vol. 2810. – P. 275–285.

27. Abonyi J., Feil B., Nemett S., Arva P. Modified Gath-Geva clustering for fuzzy segmentation of multivariable time series. Fuzzy Sets and Systems. – 2005. – Vol.149, No. 1. – P. 39–56.

28. Advances in image and video segmentation / Zhang Y.-J. (ed.). – Hershey-London-Melbourne-Singapore: IRM Press, 2006. – 457 p.
29. Aggarwal C.C., Reddy C.K. Data clustering: Algorithms and applications. Boca Raton: CRC Press, 2014. – 648 p.
30. Almeida J., Leite N. J., Torres R.S. VISON: video summarization for online applications. *Pattern Recognition Letters*. – 2012. – Vol. 33, No. 4. – P. 397-409.
31. Antani, S., Kasturi R., Jain R. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition*. – 2002. – Vol. 35, No. 4. – P. 945-965.
32. Asghar M.N., Hussain F., Manton R. Video indexing: a survey. *International Journal of Computer and Information Technology*. – 2014. – Vol. 3, No. 1. – P. 148-169.
33. Badavas P. C. Real-time statistical process control. – Englewood Cliffs: Prentice-Hall, 1993. – 232 p.
34. Bajaj D., Sharma S. Comparative analysis of shot boundary detection algorithms for video summarization. *CSI Transactions on ICT*. – 2016. – Vol. 4, No. 2–4. – P. 265–269.
35. Baraldi L., Grana C., Cucchiara R. Shot and scene detection via hierarchical clustering for re-using broadcast video. In: Azzopardi G., Petkov N. (eds.) *Computer Analysis of Images and Patterns. CAIP 2015. Lecture Notes in Computer Science*, Cham: Springer. – 2015. – Vol. 9256. – P. 801-811.
36. Begum N., Ulanova L., Wang J., Keogh E. Accelerating dynamic time warping clustering with a novel admissible pruning strategy. *Proc. of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. NY: ACM. –2015. – P. 49-58.
37. Berndt D., Clifford S. Using dynamic time warping to find patterns in time series. *Proc. KDD Workshop– Palo Alto: AAAI Press*. – 1994. – P. 359-370.
38. Bhaumik H., Bhattacharyya S., Chakraborty S. Redundancy elimination in video summarization. In: Awad A., Hassaballah M. (eds.) *Image Feature Detec-*



tors and Descriptors. *Studies in Computational Intelligence*, Cham: Springer. – 2016. – Vol. 630. – P. 173-202.

39. Bodyanskiy Ye. *Computational intelligence techniques for data analysis*. *Lecture Notes in Informatics*. Bonn: GI, –2005. – Vol. P-72. – P. 15-36.

40. Bodyanskiy Y., Kinoshenko D., Mashtalir S., Mikhnova O. Adaptive video segmentation via non-stationary multidimensional time series analysis. *Applied and Theoretical Information Systems Research: Proc. of Intern. Conf.*, Taipei, 10-12 Feb. 2012. – 2012. – P. 1-14.

41. Bodyanskiy Ye., 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. – Vol. 3, No. 1. – P. 1-20.

42. Bodyanskiy Ye., Kolodyazhniy V., Stefan A. An adaptive learning algorithm for a neuro-fuzzy network. *Lecture Notes in Computer Science*. Berlin-Heidelberg New York: Springer. – 2011. – Vol. 2206. – P. 68-75.

43. Bodyanskiy Ye., Popov S., Rybalchenko T. Multilayer neuro-fuzzi network for short term electric load forecasting. *Lecture Notes in Computer Science*. Berlin-Heidelberg: Springer-Verlag, – 2008. – Vol. 5010. – P. 339-348.

44. Bohm C., Berchtold S., Keim D.A. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. *ACM Computing Surveys*. – 2001. – Vol. 33, No. 3. – P. 322-373.

45. Castelo-Fernández C., Calderón-Ruiz G. Automatic video summarization using the optimum-path forest unsupervised classifier. In: Pardo A., Kittler J. (eds.) *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. CIARP 2015. *Lecture Notes in Computer Science*. Cham: Springer. – 2015. – Vol. 9423. – P. 760-767.

46. Chen L., Zhou Y.P., Chiu D.M. Smart streaming for online video services. *IEEE Transactions on Multimedia*. – 2015. – Vol. 17, No. 4. – P. 485–497.

47. Chen L., Stentiford F.W.M. Video sequence matching based on temporal

ordinal measurement. *Pattern Recognition Letters*. – 2008. – Vol. 29, No. 13. – P. 1824-1831.

48. Cheng M.M., Zhang G.X., Mitra N.J. et al. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2015 – Vol. 37, No. 3. – P. 569-582.

49. Choroś K. Automatic categorization of shots in news videos based on the temporal relations. In: Núñez M., Nguyen N., Camacho D., Trawiński B. (eds.) *Computational Collective Intelligence. Lecture Notes in Computer Science*. Cham: Springer. – 2015. – Vol. 9330. – P. 13-23.

50. Chu S., Keogh E.I., Hart D., Pazzani M. Iterative deepening dynamic time warping for time series. *Proceedings of the 2002 SIAM International Conference on Data Mining*. – 2002. – P. 195-212.

51. Chung F. L., Lee T. Fuzzy competitive learning. *Neural Networks*. – 1994. – Vol. 7, No. 3. – P. 539 – 552.

52. Dadashi R., Kanan H.R. AVCD-FRA: a novel solution to automatic video cut detection using fuzzy-rule-based approach. *Computer Vision and Image Understanding*. – 2013. – Vol. 117, No. 7. – P. 807-817.

53. Dash A., Albu A.B. A Domain independent approach to video summarization. In: Blanc-Talon J., Penne R., Philips W., Popescu D., Scheunders P. (eds.) *Advanced Concepts for Intelligent Vision Systems. ACIVS 2017. Lecture Notes in Computer Science*. Cham: Springer. – 2017. – Vol. 10617. – P. 431-442.

54. Datta R. et al. Image retrieval: ideas, influences, and trends of the new age *ACM Computing Surveys (CSUR)*. – 2008. – Vol. 40. – No. 2. – P. 51-60.

55. Deza M.M., Deza E. *Encyclopedia of distances*. Dordrecht, Heidelberg, London, New York: Springer, 2009. – 583 p.

56. Dunn, J.C. A Fuzzy Relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*. 1973. – Vol. 3, No.3. – P. 32–57.

57. Everitt B., Landau S., Leese M. *Cluster analysis*. Chichester: Wiley, 2011. –

321 p.

58. Fei, M., Jiang, W. & Mao, W. A novel compact yet rich key frame creation method for compressed video summarization. *Multimedia Tools and Applications*. – 2018. – Vol. 77, No. 10. – P. 11957–1197.
59. Feng L., Bhanu B. Semantic concept co-occurrence patterns for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. –2016. – Vol. 38, No. 4. – P.785-799.
60. Fragkiadaki K., Arbeláez P., Felsen P., Malik J. Learning to segment moving objects in videos. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, Massachusetts, June 8-10. – 2015. – P. 4083-4090.
61. Gaikwad B.P., Manza R.R., Manza G.R. Automatic video scene segmentation to separate script and recognition. In: Satapathy S., Biswal B., Udgata S., Mandal J. (eds.) *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Advances in Intelligent Systems and Computing*. Cham: Springer. – 2015. – Vol. 328. – P. 225-235.
62. Gan G., Ma C., and Wu J., *Data clustering: theory, algorithms, and applications*. ASA-SIAM Series on Statistics and Applied Probability, Philadelphia: SIAM, Alexandria: ASA, 2007. – 466 p.
63. Gao, Z., Lu, G. & Yan, P. Key-frame selection for video summarization: an approach of multidimensional time series analysis. *Multidimensional Systems and Signal Processing*. – 2018. – Vol. 29, No. 4. – P. 1485-1505.
64. Gao, Z., Lu, G., Lyu, C. et al. Key-frame selection for automatic summarization of surveillance videos: a method of multiple change-point detection. *Machine Vision and Applications*. – 2018. – Vol. 29, No. 7. – P. 1101–1117.
65. Gao, Z., Lu, G., Yan, P. et al. Retrospective analysis of time series for frame selection in surveillance video summarization. *Signal, Image and Video Processing*. – 2017. – Vol. 11, No. 4. – P. 581–588.
66. Geetha P.A., Narayanan V. A survey of content-based video retrieval. *Jour-*

nal of Computer Science. – 2008. – Vol.4, No. 6. – P. 474-486.

67. Grigorova A., De Natale F.G.B., Dagli C., Huang T.S. Content-based image retrieval by feature adaptation and relevance feedback. *IEEE Transactions on Multimedia*. – 2007. – Vol. 9, No. 6. – P. 1183-1192.

68. Güngör Z., Ünler A. K-harmonic means data clustering with tabu-search method. *Applied Mathematical Modelling*. – 2008. – Vol. 32, No. 6. – P. 1115-1125.

69. Hesabi Z., Tari Z., Goscinski A., Fahad A., Khalil I., Queiroz C. Data summarization techniques for big data – A survey. In: Khan S., Zomaya A. (eds.) *Handbook on Data Centers*. NY: Springer. – 2015. –P. 1109-1152.

70. Hoepfner F, Klawonn F. Fuzzy clustering of sampled functions. *Proc. 19th Int. Conf. of the North American Fuzzy Information Processing Society (NAFIPS)*, Atlanta, USA. – 2000. – P. 251 – 255.

71. Hong C., Li N., Song M., et al. An efficient approach to content-based object retrieval in videos. *Neurocomputing*. – 2011. – Vol. 74, No. 17. – P. 3565-3575.

72. Hu T.L., Li Z.C., Xing M., Su W., Tang J.H. Unsupervised video summaries using multiple features and image quality. *Proc. of 2017 IEEE Third International Conference on Multimedia Big Data*, Los Alamitos: CPS – 2017. – P. 117–120.

73. Hu Y.T., Huang J.B., Schwing A.G. Unsupervised video object segmentation using motion saliency-guided spatio-temporal propagation. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds.) *Computer Vision – ECCV 2018*. *ECCV 2018. Lecture Notes in Computer Science*. Cham: Springer. – 2018. – Vol. 11205. – P. 813-830.

74. Hu Y.T., Huang J.B., Schwing A.G. VideoMatch: matching based video object segmentation. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds.) *Computer Vision – ECCV 2018*. *ECCV 2018. Lecture Notes in Computer Science*. Cham: Springer. 2018. – Vol. 11212. – P. 56-73.

75. Hu, T. & Li, Z. Video summarization via exploring the global and local importance. *Multimedia Tools and Applications*. – 2018. – Vol. 77, No. 17. – P. 22083–22098.
76. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Clustering matrix sequences based on the iterative dynamic time deformation procedure. *International Journal of Intelligent Systems and Applications*. –2018. – Vol. 10, No. 7. – P. 66-73.
77. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Video shots' matching via various length of multidimensional time sequences *International Journal of Intelligent Systems and Applications*. – 2017. – Vol. 9, No. 11. – P. 10-16.
78. Inoue M., Yasuhara R. Considering conversation scenes in movie summarization. In: Tseng YH. et al. (eds.) *Information Retrieval Technology. AIRS 2018. Lecture Notes in Computer Science*. Cham: Springer. – 2018. – Vol. 11292. – P. 164-170.
79. Ioannidis A., Chasanis V., Likas A. Weighted multi-view key-frame extraction. *Pattern Recognition Letters*. – 2016. – Vol. 72. – P. 52-61.
80. Jacob, H., Pádua, F.L.C., Lacerda, A. et al. A video summarization approach based on the emulation of bottom-up mechanisms of visual attention. *Journal of Intellectual Information Systems*. – 2017. – Vol. 49, No. 2. – P. 193-211.
81. Jang W.D., Kim C.S. Streaming video segmentation via short-term hierarchical segmentation and frame-by-frame Markov random field optimization. In: Leibe B., Matas J., Sebe N., Welling M. (eds.) *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*. Cham: Springer. – 2016. – Vol. 9910. – P. 599-615.
82. Ji Q.-G., Fang Z.-D., Xie Z.-H., Lu Z.-M. Video abstraction based on the visual attention model and online clustering. *Signal Processing: Image Communication*. – 2013. – Vol. 28, No. 3. – P. 241-253.
83. Jiang H., Zhang G., Wang H., Bao H. Spatio-temporal video segmentation of static scenes and its applications. *IEEE Transactions on Multimedia*. – 2015. –

Vol. 17, No. 1. – P. 3-15.

84. John A.A., Nair B.B., Kumar P.N. Application of clustering techniques for video summarization – an empirical study. In: Silhavy R., Senkerik R., Kominkova-Oplatkova Z., Prokopova Z., Silhavy P. (eds.) *Artificial Intelligence Trends in Intelligent Systems. CSOC 2017. Advances in Intelligent Systems and Computing*. Cham: Springer. – 2017. – Vol. 573. –P. 494-506.
85. Juselius K. *The cointegrated VAR-model: methodology and applications*. NY: Oxford University Press Inc., 2006. – 457 p.
86. Kannan R., Ghinea G., Swaminathan S. What do you wish to see? A summarization system for movies based on user preferences. *Information Processing & Management*. – 2015 – Vol. 51. – P. 286-305.
87. Kaufman L., Rousseeuw P.J. *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons, 2009. – 342 p.
88. Keogh E. Pazzani M. Derivative dynamic time warping. *Proceedings of the 2001 SIAM International Conference on Data Mining. Series: Proceedings in Applied Mathematics, No 105*. – Society for Industrial & Applied, CD-ROM. – 2001. – P.1-11.
89. Keogh E.J., Pazzani M.J. Scaling up dynamic time warping to massive datasets. In: Żytkow J.M., Rauch J. (eds.) *Principles of Data Mining and Knowledge Discovery. PKDD 1999*. Berlin, Heidelberg: Springer. *Lecture Notes in Computer Science*. –1999. – Vol. 1704. – P. 1-11.
90. Keogh E., Pazzani M. Scaling up dynamic time warping for data mining applications. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. – NY: ACM. – 2000. – P.285-289.
91. Kinoshenko D., Kobylin O., Mashtalir S., Stolbovyi M. Metric video retrieval speedup by irrelevant data elimination *Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018)*, 110410O. –2018. – 8p.
92. Kinoshenko D., Mashtalir S., Shlyakhov V., Stolbovyi M. Video shots retrieval with use of pivot points. *Advances in Computer Science for Engineering*

- and Education. The First International Conference on Computer Science, Engineering and Education Applications ICCSEEA 2018 (Kiev, Ukraine, 18–20 January, 2018) / Hu Z., Petoukhov S., Dychka I., He M. (eds.). *Advances in Intelligent Systems and Computing*. Cham: Springer. – 2019. – Vol. 754. – P. 102-111.
93. Kinoshenko D., Mashtalir V., Shlyakhov V. A partition metric for clustering features analysis. *International Journal “Information Theories and Applications”*. – 2007. – Vol. 14, No .3. – P. 230-236.
94. Kobylin O., Mashtalir S., Stolbovyi M. Video clustering via multidimensional time-series analysis. *Proceedings of the 9th International Conference on Information Management and Engineering ICIME 2017 (Barcelona, Spain, 9–11 October, 2017)*. NY: ACM International Conference Proceeding Series. – 2017. – P. 60-63.
95. Kohonen T. *Self-Organizing Maps*. – Berlin: Springer-Verlag, 1995 – 362 p.
96. Kumar G.S.N., Reddy V.S.K., Srinivas Kumar S. Video shot boundary detection and key frame extraction for video retrieval. In: Bhateja V., Tavares J., Rani B., Prasad V., Raju K. (eds.) *Proceedings of the Second International Conference on Computational Intelligence and Informatics*. *Advances in Intelligent Systems and Computing*. Singapore: Springer. – 2018. – Vol. 712. – P. 557-567.
97. Kumar K., Shrimankar D.D., Singh N. V-LESS: A Video from linear event summaries. In: Chaudhuri B., Kankanhalli M., Raman B. (eds.) *Proceedings of 2nd International Conference on Computer Vision & Image Processing*. *Advances in Intelligent Systems and Computing*, Singapore: Springer. – 2018. – Vol. 703. – P. 385-395.
98. Kundu A., Janwe N. A survey on video segmentation the future roadmap. *International Journal of Modern Trends in Engineering and Research*. –2015, – Vol. 2, No. 3. – P. 527-534.
99. Levenshtein V. Binary codes capable of correcting deletions, insertions, and reversals. *Dokl. Akad. Nauk SSSR*, – 1965. – Vol. 163, No. 4. – P. 845–848.
100. Li A., Lin M., Wu Y., Yang M.-H., and Yan S. NUS-PRO: A new visual

- tracking challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – Vol. 38, No. 2. – 2016. – P. 335-349.
101. Li Y. & Merialdo B. Multimedia maximal marginal relevance for multi-video summarization. *Multimedia Tools and Applications*. – 2016. – Vol. 75, No. 1, – P. 199-220.
102. Liao T.W. Clustering of time series data. *Pattern Recognition*. – 2005. – Vol. 38. No.11. – P. 1857-1874.
103. Liu Y. , Zhanga D., Lua G., Ma W.-Y. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*. – 2007. – Vol. 40, No. 1. – P. 262-282.
104. Liwicki S., Zafeiriou S.P, Pantic M. On-line kernel slow feature analysis for temporal video segmentation and tracking. *IEEE Transactions on Image Processing*. – 2015. – Vol. 24, No.10. – P. 2955-2970.
105. Maksymiv O., Rak T., Peleshko D. Video-based flame detection using LBP-based descriptor: influences of classifiers variety on detection efficiency. *International Journal of Intelligent Systems and Applications*. – 2017. – Vol. 9, No. 2. – P. 42-48.
106. Mashtalir S., Mashtalir V., Stolbovyi M. Representative based clustering of long multivariate sequences with different lengths. *Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*. (Lviv, Ukraine, August 21-25). – 2018 – P. 545-548.
107. Mashtalir S., Mashtalir V., Stolbovyi M. Video shot boundary detection via sequential clustering. *International Journal “Information Theories and Applications”*, – 2017. – Vol. 24, No. 1. – P. 50-59.
108. Mashtalir S., Mikhnova O. Detecting Significant Changes in Image Sequences. *Multimedia Forensics and Security*. (Hassanien A., et al. eds.). *Intelligent Systems Reference Library*. Cham: Springer. – 2017. – Vol. 115. – P. 161-191.
109. Mashtalir S., Mikhnova O., Stolbovyi M. Sequence matching for content-based video retrieval. *Proceedings of the 2018 IEEE Second International Confer-*



- ence on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25). – 2018. – P. 549-553.
110. Mei S, Guan G, Wang Z et al Video summarization via minimum sparse reconstruction. *Pattern Recognition*. – 2015. – Vol.48, No. 2. – P.522–533.
111. Mizher M.A.A., Ang M.C., Abdullah S.N.H.S., Ng K.W. Action key frames extraction using L1-norm and accumulative optical flow for compact video shot summarisation. In: Badioze Zaman H. et al. (eds.) *Advances in Visual Informatics. IVIC 2017. Lecture Notes in Computer Science*. Cham: Springer. – 2017. – Vol. 10645. – P. 364-375.
112. Münzer B., Schoeffmann K., Böszörmenyi L. Content-based processing and analysis of endoscopic images and videos: a survey. *Multimedia Tools and Applications*. – 2018. – Vol. 77, No. 1. – P. 1323-1362.
113. Mussel Cirne M.V., Pedrini H. VISCOM: A robust video summarization approach using color co-occurrence matrices. *Multimedia Tools and Applications*. – 2018. – Vol. 77, No. 1. – P. 857–875.
114. 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*. Singapore: Springer. – 2019. – Vol. 65. – P. 23-31.
115. Nikiforov I.V. Sequential detection of changes in stochastic process. *Proc. 9-th IFAC/IFORS Symp. Identification and System Parameter Estimation – Budapest, 1991*. – Vol.1. – P.11-19.
116. Nisbet R., Miner G., Yale K. *Handbook of statistical analysis and data mining applications*. San Diego: Elsevier Inc., 2018. – 822 p.
117. Nixon L., Apostolidis E., Markatopoulou F., Patras I., Mezaris V. Multi-modal video annotation for retrieval and discovery of newsworthy video in a news verification scenario. In: Kompatsiaris I., Huet B., Mezaris V., Gurrin C., Cheng WH., Vrochidis S. (eds.) *MultiMedia Modeling. MMM 2019. Lecture*

Notes in Computer Science. Cham: Springer. – 2019. – Vol. 11295. – P. 143-155.

118. Otani M., Nakashima Y., Rahtu E., Heikkilä J., Yokoya N. Video summarization using deep semantic features. In: Lai SH., Lepetit V., Nishino K., Sato Y. (eds.) Computer Vision. – ACCV 2016. ACCV 2016. Lecture Notes in Computer Science. Cham: Springer. – 2017. – Vol. 10115. – P. 361-377.

119. Pal G., Rudrapaul D., Acharjee S., Ray R., Chakraborty S., Dey N. Video shot boundary detection: A review. In: Satapathy S., Govardhan A., Raju K., Mandal J. (eds.) Emerging ICT for Bridging the Future, -Proceedings of the 49th Annual Convention of the Computer Society of India CSI Vol . 2. Advances in Intelligent Systems and Computing. Cham: Springer. – 2015. – Vol. 338. – P. 119-127.

120. Papadopoulos D.P., Kalogeiton V.S., Chatzichristofis S.A., Papamarkos N. Automatic summarization and annotation of videos with lack of metadata information. Expert Systems with Applications. – 2013. – Vol. 40, No. 14. – P. 5765-5778.

121. Park D.C., Dagher I. Gradient based fuzzy c-means (GBFCM) algorithm. IEEE Int. Conf. on Neural Networks. – 1984. – P. 1626-1631.

122. Pau L.F. Failure diagnosis and performance monitoring. NY: Marcel Dekker Inc., 1981. – 427 p.

123. Peleshko D., Rak T., Izonin I. Image superresolution via divergence matrix automatic detection of crossover. International Journal of Intelligent Systems and Applications. – 2016. – Vol. 8, No. 12. – P. 1-8.

124. Petkovic M., Jonker W. Content-based video retrieval: a database perspective (Multimedia systems and applications). – Boston-Dordrecht-London: Kluwer Academic Publishers, 2004. – 168 p.

125. Porter S., Mirmehdi M., Thomas B. Temporal video segmentation and classification of edit effects. Image and Vision Computing. – 2003. – Vol. 21, No.13-14. – P. 1097-1106.

126. Qin Z., Shelton C.R. Event detection in continuous video: An inference in point process approach. *IEEE Transactions on Image Processing*. – 2017. – Vol. 26, No. 12. – P. 5680-5691.
127. Qu, W., Zhang, Y., Wang, D. et al. Semantic movie summarization based on string of IE-RoleNets. *Computational Visual Media*. – 2015. – Vol. 1, No. 2. – P. 129-141.
128. Rakthanmanon T., B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. SIGKDD 2012*. NY: ACM. – 2012, P.2 62-270.
129. Rashmi B.S., Nagendraswamy H.S. Shot-based keyframe extraction using bitwise-XOR dissimilarity approach. In: Santosh K., Hangarge M., Bevilacqua V., Negi A. (eds.) *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2016. Communications in Computer and Information Science*. Singapore: Springer. – 2017. – Vol. 709. – P. 305-316.
130. Ren J., Ren J. Feature pooling using spatio-temporal constrain for video summarization and retrieval. In: Park J., Jin H., Jeong Y.S., Khan M. (eds.) *Advanced Multimedia and Ubiquitous Engineering. Lecture Notes in Electrical Engineering*. Singapore: Springer. – 2016. – Vol. 393. – P. 381-387.
131. 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. ECCV 2018. Lecture Notes in Computer Science*. Cham: Springer. 2018. – Vol. 11216. – P. 358-37.
132. Rudenko O.G., Bodyanskii E.V., Pliss I.P. Adaptive algorithm for prediction of random sequences. *Soviet automatic control*. – 1979. – Vol.12, No. 1. – P. 46-48.
133. Sao N., Mishra R. A survey based on video shot boundary detection techniques. *International Journal of Advanced Research in Computer and Communica-*

- tion Engineering. – 2014. – Vol. 3, No. 4. – P. 6279–6282.
134. Senthil Murugan A., Suganya Devi K., Sivaranjani A. et al. A study on various methods used for video summarization and moving object detection for video surveillance applications. *Multimedia Tools and Applications*. – 2018. – Vol. 77, No. 18. – P. 23273-23290.
135. Schonfeld D., Shan C., Tao D., Wang L. Video search and mining. *Studies in Computational Intelligence*. – Berlin: Springer. – 2010. – Vol. 287. – 388 p.
136. Singh R.D., Aggarwal N. Novel research in the field of shot boundary detection – A survey. In: El-Alfy ES., Thampi S., Takagi H., Piramuthu S., Hanne T. (eds.) *Advances in Intelligent Informatics. Advances in Intelligent Systems and Computing*. Cham: Springer. – 2015. – Vol. 320. – P. 457-469.
137. Sonka M., Hlavac V., Boyle R. *Image processing, analysis, and machine vision*. – 3rd ed. – Toronto: Thomson, 2007. – 850 p.
138. Sousa de 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 2016. Lecture Notes in Computer Science*, Cham: Springer. – 2017. – Vol. 10125. – P. 225-232
139. Souza C., Pádua F., Nunes C., Assis G., Silva G. An unified approach to content-based indexing and retrieval of digital videos from television archives. *Artificial Intelligence Research*. – 2014. – Vol. 3, No. 3. – P.49-61.
140. Sun M., Farhadi A. Taskar B., Seitz S. Summarizing unconstrained videos using salient montages. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – Vol. 39, No. 11. – 2017. – P. 2256-2269.
141. Thingom C., Yeon G. Multi-view video summarization. In: Satapathy S., Mandal J., Udgata S., Bhateja V. (eds.) *Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing*. New Delhi: Springer. – 2016. – Vol. 435. – P. 461-473.
142. Tousch A.-M., Herbin S., Audibert J.-Y. Semantic hierarchies for image an-

- notation: a survey. *Pattern Recognition*. – 2012. – Vol. 45, No. 1. – P. 333-345.
143. Trigg D.W., Leach A.G. Exponential smoothing with an adaptive response rate. *Operational Research Quarterly*. – 1967. – Vol. 18, No. 1. – P. 53-59.
144. Tziakos I., Cavallaro A., Xu L.-Q. Video event segmentation and visualisation in non-linear subspace. *Pattern Recognition Letters*. – 2009. – Vol. 30, No. 2. – P. 123-131.
145. Vázquez-Martín R., Bandera A. Spatio-temporal feature-based keyframe detection from video shots using spectral clustering. *Pattern Recognition Letters*. – 2013. – Vol. 34, No. 7. – P. 770-779.
146. Veltkamp R.C., Burkhardt H., Kriegel H.-P. State-of-the-art in content-based image and video retrieval. *Computational Imaging and Vision*. – Netherlands: Kluwer Academic Publishers. – 2001. – 343 p.
147. Wang X., Mueen A., Ding H., Trajcevski G., Scheuermann P., Keogh E. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*. – 2013. – Vol. 26, No.2. – P. 275–309.
148. Wang W., Shen J., Yang R., Porikli F. Saliency-aware video object segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2018. – Vol. 40, No. 1. – P. 20-33.
149. Wu L., Zhang S., Jian M., Zhao Z., Wang D. Shot boundary detection with spatial-temporal convolutional neural networks. In: Lai JH. et al. (eds.) *Pattern Recognition and Computer Vision. PRCV 2018. Lecture Notes in Computer Science*. Cham: Springer. – 2018. – Vol. 11257. – P. 479-491.
150. Wu, J., Zhong, S., Jiang, J. et al. A novel clustering method for static video summarization. *Multimedia Tools and Applications*. – 2017. – Vol. 76, No. 7. – P. 9625-9641.
151. Xu R., Wunsch D. *Clustering*. Hoboken: John Wiley & Sons, Inc., 2008. – 357 p.
152. Xu R., Wunsch D. Survey of clustering algorithms. *IEEE Transactions on*

Neural Networks. – 2005. – Vol. 16, No. 3. – P. 645-678.

153. Xue J.R., Fang, J.W. & Zhang, P. A survey of scene understanding by event reasoning in autonomous driving. *International Journal of Automation and Computing*. – 2018. – Vol. 15, No. 3. – P. 249–266.

154. Yang M.Y., Reso M., Tang J., Liao W., Rosenhahn B. Temporally object-based video co-segmentation. In: *Bebis G. et al. (eds.) Advances in Visual Computing. ISVC 2015. Lecture Notes in Computer Science*. Cham: Springer. – 2015. – Vol. 9474. – P. 198-209.

155. Yeh W.-C., Jiang Y., Chen Y.-F., Chen Z. Correction: a new soft computing method for k-harmonic means clustering. *PLoS ONE*. – 2017. – Vol. 12(1).

156. Zhang B., Hsu M., Dayal U. K-harmonic means – a spatial clustering algorithm with boosting. In: *Roddick J.F., Hornsby K. (eds.) Temporal, Spatial, and Spatio-Temporal Data Mining. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Verlag. – 2001. – Vol. 2007. – P.31-45.

157. Zhang L., Xia Y., Mao K., Ma H. An effective video summarization framework toward handheld devices. *IEEE Transactions on Industrial Electronics*. – 2015. – Vol. 62, No. 2. – P.1309-1316.

158. Zhang S., Zhu Y., Roy-Chowdhury A.K. Context-aware surveillance video summarization. *IEEE Transactions on Image Processing*. – 2016. – Vol. 25, No. 11. – P. 5469-5478.

159. Zholtkevych G.M., Dorozhinsky V.V., Khadikov A.V. “Real-time” event processing and machine learning methods. *Systems of Arms and Military Equipment*. – 2016. – No. 2(46). – P.79-83.

160. Zholtkevych G., El Zein H.K. Two approaches to modeling logical time in cyber-physical systems. *Information and Communication Technologies in Education, Research and Industrial Applications (Ermolaev V. et al. eds.)*. *Communications in Computer and Information Science*. Cham: Springer. – 2018. – Vol. 826. – P. 21-40.

161. Zholtkevych G., Polyakova L.Yu., El Zein H.K. Category methods for

modelling logical time based on the concept of clocks. *Information and Communication Technologies in Education, Research and Industrial Applications* (Ermolaev V. et al. eds.). *Communications in Computer and Information Science*. Cham: Springer. – 2019. – Vol. 1007. – P. 89-101.

162. Zhou F., De la Torre F. Generalized canonical time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2016. – Vol. 38, No. 2. – P. 279-294.

163. Zhou H., Sadka A.H., Swash M.R. et al. Feature extraction and clustering for dynamic video summarization. *Neurocomputing*. – 2010. – Vol. 73, No. 10-12. – P. 1718-1729.

164. Zhu X., Loy C.C., Gong S. Learning from multiple sources for video summarization. *International Journal of Computer Vision*. – 2016. – Vol. 117, No. 3. – P. 247-268.

## ДОДАТОК А

Акти впровадження результатів дисертаційної роботи



«Затверджено» :

Генеральний директор  
ДП «УкрНТЦ «Енергосталь»,  
доктор технічних наук, професор,  
Заслужений діяч науки і техніки України,  
Лауреат Державної премії України  
в галузі науки і техніки

Д.В. Сталінський

« 10 »

2019р.

АКТ

про впровадження результатів дисертаційної роботи  
*«Технологія відеореферування на основі кластеризації для інформаційного пошуку»*  
аспіранта кафедри інформатики  
Харківського національного університету радіоелектроніки  
Столбового Михайла Івановича

Комісія в складі: заступника генерального директора Мантули В.Д.; заступника директора структурного підрозділу Кривоносова А.І. провела роботу по встановленню використання результатів дисертаційної роботи «Технологія відеореферування на основі кластеризації для інформаційного пошуку» аспіранта кафедри інформатики ХНУРЕ Столбового М.І. Відеореферування на основі кластеризації прийняте для використання при проектуванні систем екологічного моніторингу атмосферного повітря, де необхідно обробляти значні обсяги вхідних даних на протязі тривалого часу, з виявленням найбільш важливих інгредієнтів забруднення. Запропоновані методи дозволяють не тільки отримувати цілісне, можливо стратифіковане, уявлення про потік даних отриманих від постів спостереження моніторингу, а і виявити різного роду аномалії в отриманих даних.

Для підвищення валідності інтерпретації результатів важливу роль відіграють опановані дисертантом алгоритми аналізу темпоральних сегментів різної довжини, що зумовлює велику варіативність вхідних потоків даних.

Розглянувши матеріали дисертаційної роботи і апробаційні результати використання інформаційної технології реферування відеопослідовностей, комісія відзначає, що застосування багаторазової кластеризації забезпечує переваги ієрархічного підходу, тобто можливість отримання різного ступеня скорочення описів вхідних часових рядів.

Комісія наголошує про доцільність впровадження результатів дисертаційної роботи в науково-дослідних та проектних роботах ДП «УкрНТЦ «Енергосталь». Безумовний інтерес представляє подальше збільшення достовірності сегментації відеоданих в часі, що містять різні категорії візуальної інформації, і підвищення точності оцінювання кількісних характеристик об'єктів на кожному зображенні і відеопотоці в цілому.

Заступник генерального директора,  
Заслужений металург України



В.Д. Мантула

Заступник директора структурного підрозділу,  
доктор технічних наук, професор,  
Лауреат Державних премій СРСР



А.І. Кривоносов

ЗАТВЕРДЖУЮ  
Проректор з НМР Харківського  
національного університету  
радіоелектроніки

І. В. Рубан

\_\_\_\_\_ 2019 р.

### АКТ

про використання в навчальному процесі результатів дисертаційної роботи  
«Технологія відеореферування на основі кластеризації для інформаційного пошуку»  
аспіранта кафедри інформатики  
Харківського національного університету радіоелектроніки  
Столбового Михайла Івановича

Комісія у складі голови – декана факультету ІТМ д.ф.-м.н., проф. Дорошенка В.О. та членів: зав. каф. інформатики д.т.н., проф. Пуятіна Є.П.; проф. каф. інформатики д.т.н., проф. Гороховатського В.О. провела роботу по встановленню використання результатів дисертаційної роботи «Технологія відеореферування на основі кластеризації для інформаційного пошуку» аспіранта кафедри інформатики Столбового М.І. в навчальному процесі при підготовці спеціалістів та магістрів за спеціальністю 122 – Комп'ютерні науки (спеціалізація «Інформатика»).

Розглянувши матеріали роботи та зміст навчального процесу на кафедрі інформатики, комісія відзначає, що при проведенні лекційних занять та лабораторного практикуму за курсами «Класифікація та кластеризація даних», «Теорія розпізнавання образів», «Структурний аналіз та інтелектуальне оброблення багатовимірних даних» використовуються наступні результати дисертаційних досліджень:

- матричні модифікації адаптивних метричних методів кластеризації відеоданих;
- моделі кластеризації потоків даних різної довжини за допомогою ітеративної динамічної деформації;
- використання запропонованого ансамблю моделей для швидкого виявлення змін багатовимірних часових рядів.

Комісія відзначає, що використання результатів дисертаційних досліджень дає можливість покращення якості вивчення новітніх моделей та методів обробки та інтерпретації відеоданих.

Голова комісії  
декан факультету ІТМ д.ф.-м.н., проф.

В.О. Дорошенко В.О. Дорошенко

Члени комісії:  
зав. каф. інформатики, д.т.н., проф.

Є.П. Пуятін Є.П. Пуятін

проф. каф. інформатики, д.т.н., проф.

В.О. Гороховатський В.О. Гороховатський



ЗАТВЕРДЖУЮ

Проректор з наукової роботи  
Харківського національного  
університету радіоелектроніки М. В. Неофітний

«25» 01/2019 р.



## АКТ

про впровадження результатів кандидатської дисертаційної роботи  
аспіранта кафедри інформатики**Столбового Михайла Івановича****«Технологія відеореферування на основі кластеризації для  
інформаційного пошуку»**у науково-дослідну роботу № 307 Харківського національного університету  
радіоелектроніки

Комісія у складі голови – д.т.н., проф. Бодянського Є. В., наукового керівника НДР №307, професора кафедри штучного інтелекту, наукового керівника ПНДЛ АСУ ХНУРЕ; та членів – к.т.н., с.н.с. Плісс І. П., відповідального виконавця НДР №307, Хоменко Н. Ю., начальника наукового інформаційно-аналітичного відділу ХНУРЕ, склала цей акт про те, що результати дисертаційної роботи Столбового М.І. використовуються у Харківському національному університеті радіоелектроніки при виконанні держбюджетної науково-дослідної роботи № 307 «Динамічний інтелектуальний аналіз послідовностей нечіткої інформації за умов суттєвої невизначеності на основі гібридних систем обчислювального інтелекту» № ДР 0116U002539, яка виконується згідно з наказом Міністерства освіти і науки України за результатами конкурсного відбору проектів наукових досліджень.

Розроблено матричні моделі кластеризації багатовимірних даних різної довжини за допомогою використання ітеративної динамічної часової деформації, що дозволяє значно покращити можливості матричних метричних методів порівнянь багатовимірних даних. За допомогою розроблених Столбовим М.І. моделей кластеризації здійснюється онлайн обробка відео потоків з метою їх реферування та з'являються нові можливості до розв'язання задачі інформаційного пошуку.

Результати використані в науково-технічному звіті по НДР №307.

Голова комісії:



Є. В. Бодянський

Члени комісії:



І. П. Плісс



Н. Ю. Хоменко

## ДОДАТОК Б

## Список публікацій здобувача

1. Mashtalir S., Mashtalir V., Stolbovyi M. Video shot boundary detection via sequential clustering. International Journal “Information Theories and Applications”, 2017. Vol. 24, No. 1. – P. 50-59.
2. Hu Z., Mashtalir S.V., Tyshchenko O.K., Stolbovyi M.I. Video shots' matching via various length of multidimensional time sequences. International Journal of Intelligent Systems and Applications. 2017. Vol. 9, No. 11. – P. 10-16.
3. Hu Z., Mashtalir, S.V., Tyshchenko, O.K., Stolbovyi, M.I. Clustering matrix sequences based on the iterative dynamic time deformation procedure. International Journal of Intelligent Systems and Applications. 2018. Vol. 10, No. 7. – P. 66-73.
4. Машталир С.В., Столбовой М.И. Адаптивные матричные модели в задаче контроля потоков видео. Радиоелектроніка, інформатика, управління. 2018. №4(47). – С. 188-194.
5. Богучарский С.И., Машталир С.В., Столбовой М.И. Быстрое обнаружение изменения свойств многомерных временных рядов на основе идентификационного подхода к ансамблю моделей. Системи обробки інформації. Харків : ХУПС. 2018. Вып. 3 (154). – С. 74-78.
6. Машталир С.В., Столбовой М.И., Яковлев С.В. Кластеризация последовательностей видеоданных на основе гармонических k-средних. Кибернетика и системный анализ. 2019. Том 55, №2. – С. 36-43.
7. Машталир С.В., Столбовой М.И. Анализ методов пространственно-временной сегментации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнародної наукової конференції. (Залізний Порт, Україна, 22–26 травня 2017). Херсон, Видавництво ПП Вишемирський В.С. 2017. С. 287-289.
8. Kobylin O., Mashtalir S., Stolbovyi M. Video clustering via multidimensional time-series analysis. Proceedings of the 9th International Conference on Information Management and Engineering ICIME 2017 (Barcelona, Spain, 9–11 October, 2017).

New York : ACM International Conference Proceeding Series. 2017. P. 60-63.

9. Машталір С.В., Столбовой М.И., Яременко М.М. Использование матричных моделей анализа временных рядов в задаче кластеризации видеопоследовательностей. Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : Матеріали міжнародної наукової конференції (Залізний Порт, Україна, 21–27 травня 2018). Херсон, Видавництво ПП Вишемирський В.С. 2018. С.83-85.

10. Mashtalir S., Mashtalir V., Stolbovyi M. Representative based clustering of long multivariate sequences with different lengths. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 545-548.

11. Mashtalir S., Mikhnova O., Stolbovyi M. Sequence matching for content-based video retrieval. Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP). (Lviv, Ukraine, August 21-25, 2018). P. 549-553.

12. Kinoshenko D., Mashtalir S., Shlyakhov V., Stolbovyi M. Video shots retrieval with use of pivot points. Advances in Computer Science for Engineering and Education. The First International Conference on Computer Science, Engineering and Education Applications ICCSEEA 2018 (Kyiv, Ukraine, 18–20 January, 2018) / Hu Z., Petoukhov S., Dychka I., He M. (eds). Advances in Intelligent Systems and Computing. Cham : Springer. 2019. Vol. 754. P. 102-111.

13. Kinoshenko D., Kobylin O., Mashtalir S., Stolbovyi M. Metric video retrieval speedup by irrelevant data elimination. Proc. SPIE 11041, Eleventh International Conference on Machine Vision (ICMV 2018), 110410O; 8 p.