

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

СТОЛБОВИЙ МИХАЙЛО ІВАНОВИЧ

УДК 004.932.2:004.93'14

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

05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2019

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки України.

Науковий керівник доктор технічних наук, доцент
Машталір Сергій Володимирович,
Харківський національний університет
радіоелектроніки,
професор кафедри інформатики.

Офіційні опоненти: доктор технічних наук, професор
Жолткевич Григорій Миколайович,
Харківський національний університет
ім. В.Н. Каразіна, МОН України,
декан факультету математики та інформатики;

доктор технічних наук, доцент
Рак Тарас Євгенович,
IT Step University, м. Львів, МОН України,
проректор з науково-педагогічної роботи.

Захист відбудеться “_____” _____ 2019 р. о _____ годині на засіданні спеціалізованої вченої ради Д 64.052.08 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Науки, 14.

З дисертацією можна ознайомитись у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Науки, 14.

Автореферат розісланий “_____” _____ 2019 р.

Учений секретар
спеціалізованої вченої ради

І.П. Плїсс

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

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 запитамі ‘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 пертіненності 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 відореферування 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 відореферування, 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 відореферування, яке ґрунтується на поданні сегментів в вигляді ключових кадрів. Перспективним напрямком розвитку представляється об’єктивне динамічне відореферування 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. Антошук, 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 нестаціонарності 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 запитами «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 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 редукації 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 [1] – метод пошуку границь сегментів відео; [2] – метод кластеризації послідовностей різної довжини за допомогою динамічної часової деформації; [3] – метод кластеризації послідовностей різної довжини за допомогою ітеративної динамічної часової деформації; [4] – метод кластеризації послідовностей різної довжини за допомогою адаптивної динамічної часової деформації; [5] – метод побудови ансамблю моделей для швидкого виявлення змін в відеоданих; [6] – модифікація методу гармонічних k-середніх для аналізу відео послідовностей; [7] – модифікація методу порівняння методів сегментації відеопослідовностей; [8] – модифікація методу кластеризації відеопослідовностей; [9] – матричний метод кластеризації відеопослідовностей; [10] – матрична модифікація методу кластеризації відео послідовностей різної довжини; [11] – модифікація методу порівняння результатів кластеризації відеопослідовностей; [12] – метод елімінації відеоданих; [13] – 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).

Структура й обсяг дисертації. Дисертація складається зі вступу, п'яти розділів, висновків, списку використаних джерел та додатка. Повний обсяг дисертації становить 155 сторінок; робота містить 14 рисунків (з них 2 на окремих сторінках); список використаних джерел, що включає 164 найменування та займає 19 сторінок; 2 додатка на 6 сторінках, анотації на 12 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність розробки нових методів сегментації та кластеризації відеопослідовностей для розв'язання задачі автоматичного відеореферування, що призначений для скороченого представлення візуальної інформації та організації інформаційного пошуку; розглянуто сучасний стан проблеми, визначені мета, об'єкт, предмет і методи дослідження, наведені задачі, що розв'язуються у дисертаційній роботі, зв'язок з науковими темами, наведено наукову новизну та практичне значення отриманих результатів, надано інформацію про особистий внесок автора.

У першому розділі в результаті проведеного аналізу встановлено, що темпоральна сегментація і кластеризація відеорядів є одним з ключових завдань обробки та інтерпретації відеопотоків в цілому. Сегментація в часі створює передумови для семантичної структуризації відеоданих у вигляді побудови розбиття або покриття багатовимірних часових рядів, асоційованих в ознакових просторах з відеопослідовностями, або власне матричних часових рядів, що безпосередньо відображають зміст фрагментів кожного відеокадру. Отримані сегменти, невиразні з точки зору заданих критеріїв зв'язні хронологічно послідовності відеокадрів, є базисним набором елементів відео, на множині яких будуються скорочені репрезентативні відеоузагальнення. Якщо для статистичного відеореферування (на основі ключових кадрів) розроблений, обґрунтований і апробований цілий комплекс методів, що забезпечують достатні показники повноти, точності і оперативності інформаційного пошуку на базі мультидисциплінарних інформаційних технологій CBIR (Content-Based Image Retrieval) систем, то в динамічній сумаризації лише робляться перші кроки. Отримання ідентифікаційних метаданих (в семантичному, синтаксичному та структурному аспектах) про скорочених представленнях відео пов'язано не тільки з полісемізом візуальної інформації, але і з цілим рядом інструментальних складнощів. У цьому плані слід підкреслити відмінність носіїв об'єктів порівняння і при кластеризації, і безпосередньо при інформаційному пошуку, перш за все істотну варіативність довжин послідовностей даних, що помітно звужує спектр потенційно корисних методів.

Певну перспективу представляє розвиток і модифікація on-line та off line методів сегментації і кластеризації шляхом їх адаптації до виявлення змін властивостей багатовимірних часових рядів різної довжини, а також до фрагментної обробки двовимірних полів (зображень, текстур, матриць градієнтів і т.п.) , що змінюються в часі, що в кінцевому підсумку забезпечить створення конкурентоспроможного інструментарію відеоаналізу. Виходячи з цього, сформульовано мету та задачі дослідження.

Другий розділ присвячений розробці методу виявлення зміни властивостей потоків даних на основі ідентифікаційного підходу до ансамблю моделей для сегментації і кластеризації відеоданих за умов необхідності порівнянь вхідних даних різної довжини. Увагу також сфокусовано на непрямому підході до кластеризації послідовностей, який в поєднанні з методами нечіткої рекурент-

ної оптимізації забезпечує синтез досить простих обчислювальних моделей сегментації відеорядів.

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 $k = 1, 2, \dots, N$ – точний дискретний час.

Згідно із непрямим підходом 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, автокореляції. Так, для векторно-матричних аналогів, ці характеристики виглядають наступним чином

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

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

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

(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 надання адаптивних властивостей рекурентним процедурам, контрольовані характеристики зручніше оцінювати на основі модифікації експоненційного згладжування:

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

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

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

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

Завдяки цьому, 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.

В свою чергу, для обробки $(n+1)$ -вимірною сигналу для оцінки вектору середніх $\bar{x}^\alpha(k)$ і автокореляційної матриці $R^\alpha(k, \tau)$ можна також використовувати процедуру експоненційного згладжування виду

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

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

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

Далі звертаємось до власне виявлення змін властивостей компонент часового ряду. При цьому 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

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

Запропоновано використовувати обчислення відразу всієї матриці параметрів $\tilde{x}(k)$ розмірності $((2 + \tau_{max}) \times n)$. Для цього доцільно скористатися матричною модифікацією методу нечітких С-середніх

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

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

$$0 < \sum_{k=1}^N \mu(\tilde{x}_i(k), C_i(l)) \leq N, \quad \forall l = 1, 2, \dots, M. \quad (3)$$

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

Мінімізація (1) при обмеженнях (2), (3) веде до аналітичного виразу

$$\left\{ \begin{array}{l} \mu(\tilde{x}(k), C(l)) = \frac{\left(Sp(\tilde{x}(k) - C(l))(\tilde{x}(k) - C(l))^T \right)^{-1}}{\sum_{j=1}^m \left(Sp(\tilde{x}(k) - C(j))(\tilde{x}(k) - C(j))^T \right)^{-1}}, \\ C(l) = \frac{\sum_{k=1}^N \mu^2(\tilde{x}(k), C(l)) \tilde{x}(k)}{\sum_{k=1}^N \mu^2(\tilde{x}(k), C(l))}. \end{array} \right. \quad (4)$$

Вочевидь, що вираз (4) не може бути використано для вирішення даної задачі в on-line режимі. У цьому випадку доцільно скористатися матричною модифікацією рекурентної процедури, що має в даному випадку вигляд

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

де параметр кроку $\eta(k)$, обраний зазвичай з емпіричних міркувань, $\beta > 0$ – параметр фаззифікації.

Резюмуючи, відзначимо, що розглянута та вирішена задача виявлень змін розладнань в багатовимірних (матричних) потоках даних, що послідовно надходять на обробку. Введено on-line процедури нечіткої кластеризації, що дозволяють вирішувати завдання як для повільних, так і стрибкоподібних змін.

Третій розділ присвячений побудові моделі матричних часових рядів на основі ітеративної динамічної часової деформації *DTW* (*Dynamic Time Warping*) та кластеризації цих рядів на основі самоорганізованих карт Кохонена, ієрархічної кластеризації рядів різної довжини, включаючи випадки перетинних класів, що дозволяють визначати зміни властивостей фрагментів відеокадрів. Також обговорюється налаштування параметрів адаптивної матричної моделі та виявлення змін властивостей матричної послідовності.

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 наштовхується на значні обчислювальні труднощі, що робить цей підхід неідеальним.

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

Обчислювальну складність можна істотно скоротити, скориставшись ідеями ітеративної динамічної часової деформації. Її суть полягає в редукації 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: $x(1), x(\lfloor \frac{N}{4} \rfloor), x(\lfloor \frac{N}{2} \rfloor), x(\lfloor \frac{3N}{4} \rfloor), x(N)$ і $y(1), y(\lfloor \frac{M}{4} \rfloor), y(\lfloor \frac{M}{2} \rfloor), y(\lfloor \frac{3M}{4} \rfloor), x(M)$ (тут $\lfloor \circ \rfloor$ – ціла частина числа), після чого на інтервалах

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

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

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 (4×4) -матрицю відстаней, на основі якої розраховується кумулятивна відстань

$$DTW(x, y) = DTW(4, 4).$$

Якщо оцінки, одержані за допомогою цієї формули, 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 редукації 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 $g + 1$ відкликів.

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 $g + 1$ відкликів.

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 τ – 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

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

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

Налаштування параметрів моделі проводиться за допомогою рекурентного методу найменших квадратів:

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

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

$$\bar{x}_{qr}(\tau) = A_{qr}^T \bar{\tau}, \quad (5)$$

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

Для налаштування параметрів моделі (5) може бути використаний рекурентний алгоритм виду

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

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, яка стосовно аналізованої тут задачі може бути розрахована за допомогою співвідношень

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

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

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

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 яких 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.

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 Еверітта. Отже, нехай є масив $x_1, x_2, \dots, x_q, \dots, x_Q$, утворений Q багатовимірними 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 $Q/2$ кластерів, 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 прийнятим вимогам до якості кластеризації.

Далі розглянемо загальний покроковий алгоритм процесу ієрархічної агломеративної нечіткої кластеризації багатовимірних часових рядів різної 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. Розраховуємо індекс Данна для оцінки якості кластеризації.
5. Якщо якість не задовольняє, переходимо до кроку 3 інакше крок 6.
6. Розраховуємо представників-зразків для кожного кластера останнього рівня.
7. Розраховуємо рівні належності кожного спостереження x_q до кластеру $Cl_j^{[m]}$.

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

Альтернативою може служити підхід до кластеризації відео для реферування відеоданих на множинах сегментів, що розглядається в **четвертому розділі**, заснований на використанні гармонійних середніх, 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 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

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

При цьому в процесі обробки даних мінімізується цільова функція

$$E(\bar{x}_q, C_j) = \sum_{q=1}^Q \frac{m}{\sum_{j=1}^m \bar{D}^2(\bar{x}_q, C_j)}, \quad (6)$$

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

Мінімізація (6) по матриці C_j веде до оцінки

$$C_j = \frac{\sum_{q=1}^Q \bar{D}^{-4}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2} \bar{x}_q}{\sum_{q=1}^Q \bar{D}^{-4}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2}}, \quad (7)$$

яка обчислюється в результаті послідовності ітерацій, однак їх кількість практично не залежить від початкових умов. Тут же введено модифікацію (7) у вигляді

$$C_j = \frac{\sum_{q=1}^Q \bar{D}^{-\alpha}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2} \bar{x}_q}{\sum_{q=1}^Q \bar{D}^{-\alpha}(\bar{x}_q, C_j) (\sum_{j=1}^m \bar{D}^{-2}(\bar{x}_q, C_j))^{-2}} \quad (8)$$

де параметр α має сенс близький до фаззіфікатору, використовуваному в процедурах нечіткої кластеризації.

Нескладно ввести в розгляд нечіткі модифікації процедур (7), (8). 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 (7) 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 (8). Тут $\beta > 0$ – параметр фаззіфікації, що визначає «розмитість» меж між сформованими кластерами.

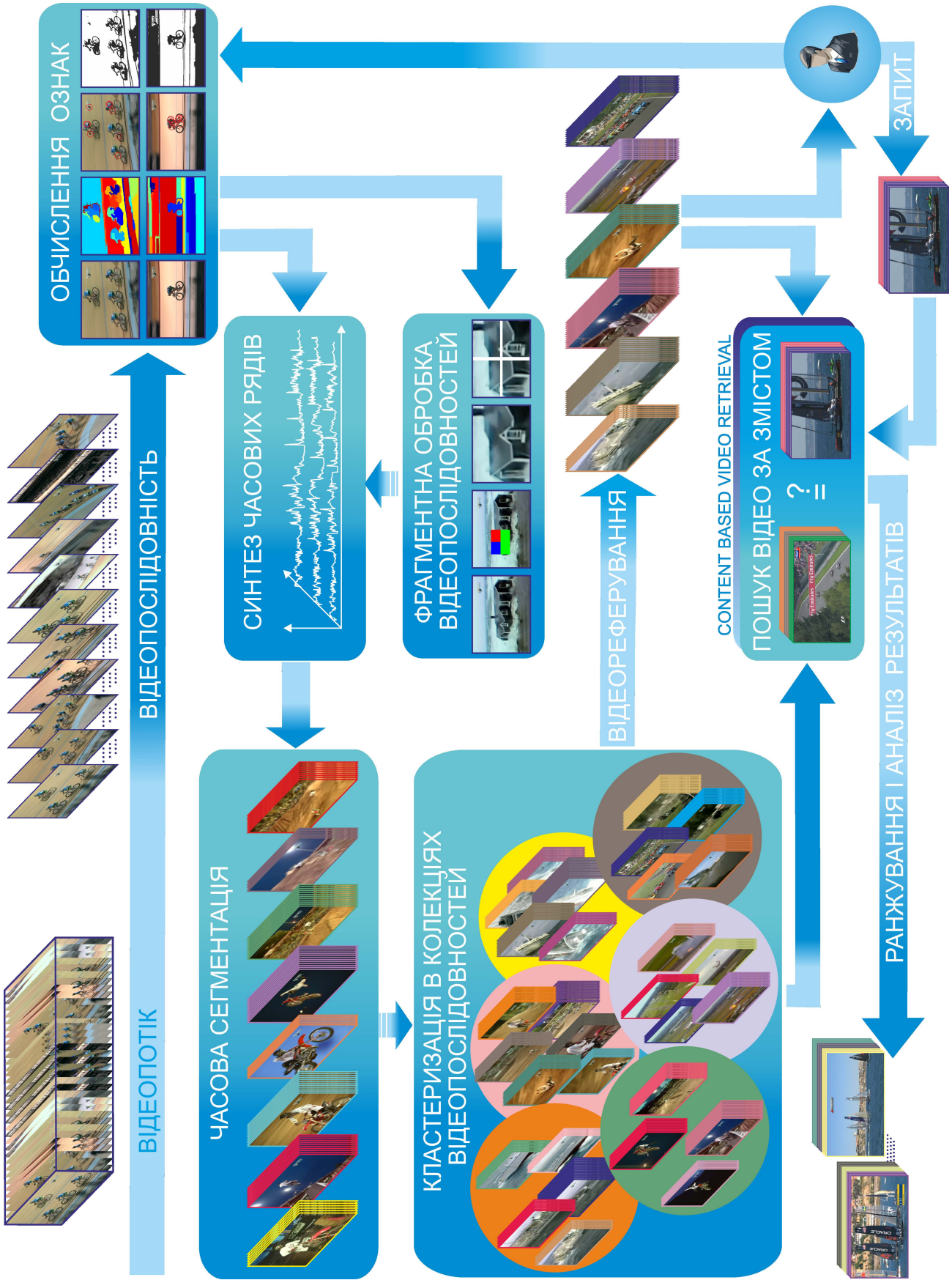


Рисунок 1 – Схема відеореферування для інформаційного пошуку

Необхідно відзначити, що з обчислювальної точки зору отримання оцінок (7), (8) не складніше оцінювання центроїдів k -середніх, проте вимагає меншої кількості ітерацій.

П'ятий розділ присвячений експериментальному аналізу запропонованих моделей та методів сегментації і кластеризації відеорядів, що забезпечують технологію динамічного відеореферування для інформаційного пошуку. На рис. 1 представлена узагальнена схема обробки та аналізу відеоданих в задачах контекстного пошуку. Спочатку з відеопотоку виділяється послідовність кадрів, для кожного з яких в on-line режимі обраховуються глобальні і/або локальні ознаки, на основі яких виникають багатовимірні часові ряди. Визначення глобальних ознак базується на просторовій сегментації відеокадрів, обчисленні характеристик форми кожної з отриманих областей розбиття поля зору. При цьому враховуються і колірні властивості у вигляді моментних, гістограмних і текстурних характеристик. Використано різні набори дескрипторів форм областей, індукованих просторовою сегментацією. Локальні ознаки являють собою фрагменти (власне зображення або деякі згортки) з центрами в особливих точках, які характеризують області візуального інтересу або уваги. Особливі точки знаходилися за допомогою алгоритмів SURF, SIFT, BRISK, KAZE та Harris Corner Detector. На етапі темпоральної on-line сегментації можливо проріджування кадрів на будь-яку кількість елементів послідовності, що не перевищує довжину відеобуферу, що забезпечує дихотомічне уточнення меж сегментів на основі передісторії. У випадках виникнення покриттів відеопослідовності внаслідок ефектів затемнення, розчинення, розмивання, накладення, з'явилися кадри в кадрі. Як межа сегменту вибиралася або середина часового переходу, або перехід був окремим сегментом, але виключався з подальшого аналізу. На етапі кластеризації використаний весь запропонований набір методів кластеризації відеопослідовностей, а при порівняннях відеорядів застосовувалися DTW, метрики Левенштейна і Фробеніуса. Для відеореферування вибиралися представники кластерів, що найменш відрізняються від всіх сегментів даного кластеру.

Експеримент 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 1280×720 та 720×576 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.

На рис. 2 показані приклади пошуку меж сегментів на основі аналізу фрагментних послідовностей (з вибором з 20 найбільш потужних відгуків) з виділенням особливих точок алгоритмом SURF.



Рисунок 2 – Приклади границь сегментів часової сегментації

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 затуляння, і рух об'єктів, не порушуючи валідності темпоральної сегментації. На рис. 3 з проріджуванням в 10 кадрів наведено приклад руху об'єкта в рамках одного часового сегмента.

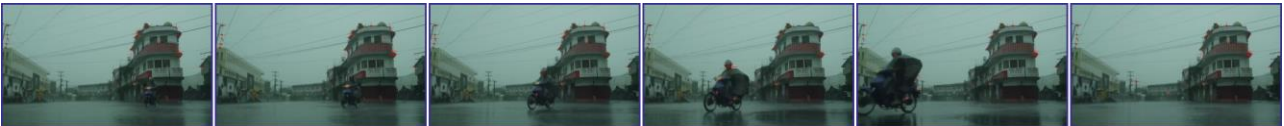


Рисунок 3 – Приклад руху об'єкту всередині одного сегменту

Якість динамічного відеореферування оцінювалося візуально.

У додатках наведено список опублікованих праць за темою дисертації, акти впровадження теоретичних і практичних результатів дисертаційних досліджень.

ВИСНОВКИ

В дисертаційній роботі для розвитку технологій інформаційного пошуку вирішена актуальна науково-практична задача розробки моделей і методів сегментації та кластеризації відеоданих 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 зашумлених 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 релевант-

ність 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 запитом «за зразком» реалізовані програмно та використані у розробках ДП «УкрНТЦ «Енергосталь», наукових розробках та освітньому процесі ХНУРЕ, що підтверджено відповідними актами.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

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. (Індексується в Scopus)

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. (Індексується в Scopus)

4. Машталир С.В., Столбовой М.И. Адаптивные матричные модели в задаче контроля потоков видео. *Радиоэлектроника, информатика, управління*. 2018. №4(47). – С. 188-194. (*Індексується в Web of Science*)
5. Богучарский С.И., Машталир С.В., Столбовой М.И. Быстрое обнаружение изменения свойств многомерных временных рядов на основе идентификационного подхода к ансамблю моделей. *Системы 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. (*Індексується в Scopus*)
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. (*Індексується в Scopus*)
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. (*Індексується в Scopus*)
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. (*Індексується в Scopus*)
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. (*Індексується в Scopus*)
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; 8p.* (*Індексується в Scopus*)

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.

Дисертація присвячена розвитку технологій інформаційного пошуку в плані розробки моделей і методів сегментації і кластеризації відеоданих для динамічного відеореферування.

Введено методи і моделі виявлення змін властивостей векторних і матричних нестационарних зашумлених сигналів на основі ансамблю адаптивних моделей з власними алгоритмами ідентифікації з різною глибиною пам'яті, on-line процедури нечіткої кластеризації, що забезпечують часову сегментацію як для повільних, так і стрибкоподібних змін змісту відео. На основі гібридизації ієрархічного агломеративного і нечіткого, заснованого на центрах ваги, підходів для динамічного відеореферування запропонований метод кластеризації багатомірних часових рядів різної довжини в умовах невідомої кількості класів і можливості їх взаємного перетину. Для відеореферування запропонована кластеризація послідовностей відеоданих, яка заснована на використанні модифікованої ітеративної динамічної часової деформації і подальшої кластеризації редукованих часових рядів на основі матричної нечіткої кластеризації на базі гармонійних k -середніх.

Результати теоретико-експериментальних досліджень реалізовані і впроваджені у вигляді прикладних і дослідницьких програмних комплексів.

Ключові слова: відео, відеореферування, багатомірні часові ряди, сегментація, кластеризація.

АННОТАЦИЯ

Столбовой М.И. Технология видеореферирования на основе кластеризации для информационного поиска. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины, Харьков, 2019.

Диссертация посвящена развитию технологий информационного поиска в плане разработки моделей и методов сегментации и кластеризации видеоданных для динамического видеореферирования.

Введены методы и модели обнаружения изменений свойств векторных и матричных нестационарных зашумленных сигналов на основе ансамбля адаптивных моделей с собственными алгоритмами идентификации с разной глубиной памяти, on-line процедуры нечеткой кластеризации, обеспечивающих временную сегментацию как для медленных, так и скачкообразных изменений содержания видео. На основе гибридизации иерархического агломеративного и

нечеткого, основанного на центроидах, подходов для динамического видеореферирования предложен метод кластеризации многомерных временных рядов разной длины в условиях неизвестного количества классов и возможности их взаимного пересечения. Для видеореферирования предложена кластеризация последовательностей видеоданных, основанная на использовании модифицированной итеративной динамической временной деформации и последующей кластеризации редуцируемых временных рядов на основе матричной нечеткой кластеризации на базе гармонических k -средних.

Результаты теоретико-экспериментальных исследований реализованы и внедрены в виде прикладных и исследовательских программных комплексов.

Ключевые слова: видео, видеореферирование, многомерные временные ряды, сегментация, кластеризация.

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 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 segmenta-

tion and clustering models and methods for dynamic video summarization has been solved with the aim of temporal visual information retrieval technologies development.

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 established 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.

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.

Keywords: video, video retrieval, multidimensional time series, segmentation, clustering.