

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

МАШТАЛІР СЕРГІЙ ВОЛОДИМИРОВИЧ

УДК 004.932.2:004.93'14

**МОДЕЛІ ТА МЕТОДИ ТЕМПОРАЛЬНОЇ ОБРОБКИ ВІДЕО  
ДЛЯ ІНФОРМАЦІЙНОГО ПОШУКУ**

05.13.06 – інформаційні технології

Автореферат  
дисертації на здобуття наукового ступеня  
доктора технічних наук

Харків – 2016

Дисертацією є рукопис.

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

Науковий консультант            доктор технічних наук, професор  
**Бодянський Євгеній Володимирович,**  
Харківський національний університет  
радіоелектроніки,  
професор кафедри штучного інтелекту.

Офіційні опоненти:            доктор технічних наук, професор  
**Антощук Світлана Григорівна,**  
Одеський національний політехнічний  
університет,  
директор інституту комп'ютерних систем;

доктор технічних наук, професор  
**Жолткевич Григорій Миколайович,**  
Харківський національний університет  
імені В.Н. Каразіна,  
декан факультету математики та інформатики;

доктор технічних наук, доцент  
**Рак Тарас Євгенович,**  
Львівський державний університет безпеки  
життєдіяльності Державної служби України  
з надзвичайних ситуацій,  
проректор з науково-дослідної роботи.

Захист відбудеться "08" грудня 2016 р. о 14-00 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 "07" листопада 2016 р.

Учений секретар  
спеціалізованої вченої ради

І.П. Плїсс

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Диверсифікація використання мультимедіа і гіпермедіа акцентує увагу на розвитку інформаційних технологій, комплексу методів, способів і засобів зберігання, обробки, передачі та відображення інформації, орієнтованих на поганоструктуровані або зовсім неструктуровані дані великих обсягів. Відеопотоки є яскравими представниками таких неструктурованих даних. Мінімізація їх надлишковості, забезпечення адекватності, достовірності, повноти, об'єктивності, доступності та актуальності стають ключовими питаннями пошуку компромісу між інформаційною потребою і обсягами та властивостями накопичуваних відеоданих. Можливість в умовах невизначеності витягувати необхідні відомості з відео – основна проблема релевантності візуальних інформаційних ресурсів, пов'язаних з нелінійним аналізом, багатоаспектним автоматичним індексуванням, анотуванням і реферуванням, пошуком з різними формами запитів, зокрема, 'ad exemplum' – «за зразком». Семантичний розрив між низькорівневими ознаками візуальної інформації і вимогами високорівневої інтерпретації вимагає проведення досліджень, потенційною базою яких служить розрахунок просторових (між об'єктами сцен) зв'язків і зв'язків між відеокадрами в часі. Поліструктурність відеопотоків, багаторівневе і багатозначне змістовне значення їх елементів ставлять в ряд найважливіших задачу пошуку сукупності стійких зв'язків, які забезпечують цілісність аналізу і тотожність представлення відео в цілому. Структуризація групує елементи відеоданих, керує ступенем їх огрубіння або деталізації, визначає логічні зв'язки між виділеними групами, стратифікує відеодані. Структурування відеопослідовностей забезпечує побудову «стійкого каркасу», на базі якого розв'язуються задачі синтезу інформаційних технологій обробки, аналізу, зберігання та інтерпретації відеопотоків.

В силу природи формування і зберігання відео виняткового значення набувають і можливість, і необхідність *on line* обробки. Аналіз відеопослідовностей і / або індукованих ними багатовимірних часових рядів в режимі накопичення даних з урахуванням передісторії різної глибини і короткострокового прогнозування, мабуть, єдиний природний шлях розвитку багатофакторної міждисциплінарної інтеграції моделей і методів керування (в широкому сенсі) відео. У цьому плані визначальну перспективу має вдосконалення інструментів управління колекціями відео, і в першу чергу, системами семантичного пошуку CBVIR (Content Based Video Information Retrieval). Стосовно до CBVIR великий вплив на розвиток моделей та методів обробки та розпізнавання статичної та динамічної візуальної інформації мають вітчизняні та зарубіжні вчені С.Г. Антощук, Р.А. Воробель, Г.М. Жолткевич, В.Н. Крилов, Д.Д. Пелешко, Є.П. Путятін, Т.Є. Рак, М.І. Шлезінгер, Н. Burkhardt, Р.М. Haralick, В.К.Р. Horn, G. Liu, М. Petkovic, М. Sonka, Р. Zezula та інші. Незважаючи на численні дослідження в цьому напрямку, науково-технічна проблема побудови систем валідного семантичного пошуку динамічної візуальної інформації лишається актуальною. Основною парадигмою розвитку інформаційного пошуку в колекціях

погано структурованих даних є така їх організація, коли спочатку виключаються з розгляду свідомо непотрібні об'єкти, а детальний пошук здійснюється лише серед решти.

Неперервне зростання в різних проблемно-орієнтованих областях обсягів відеоданих, що мають різні характеристики, призводить до необхідності створення нового інструментарію темпоральної обробки відеопослідовностей – засобів on line структуризації відеопотоків на основі просторової (внутрішньокадрової) і часової (міжкадрової) сегментації відеоданих. Таким чином, розробка моделей і методів темпоральної обробки відео для інформаційного пошуку є важливою в концептуальному та прикладному аспектах науково-технічної проблемою.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконана на кафедрі інформатики Харківського національного університету радіоелектроніки в рамках держбюджетних науково-дослідних робіт: «Синтез методів обробки інформації в умовах невизначеності на основі самонавчання та м'яких обчислень» (№ ДР0107U003028, виконавець), «Еволюційні гібридні системи обчислювального інтелекту зі змінною структурою для інтелектуального аналізу даних» (№ ДР0110U000458, виконавець), «Моделі і методи грануляції і інтерпретації багатовимірних даних» (№ ДР0110U002636, відповідальний виконавець), «Моделі і методи аналізу та інтерпретації динамічних процесів» (№ ДР0113U000357, відповідальний виконавець розділу), «Нейро-фаззі системи для поточної кластеризації і класифікації послідовностей даних за умов їх викривленості відсутніми та аномальними спостереженнями» (№ ДР0113U000361, виконавець) «Розробка гібридних систем та методів обчислювального інтелекту для обробки потоків нечіткої інформації за умов нестаціонарності та невизначеності» (№ ДР0116U002539, відповідальний виконавець розділу), які виконувалися відповідно до наказів Міністерства освіти і науки України за результатами конкурсного відбору науково-дослідних робіт.

**Метою дисертаційної роботи** розв'язання науково-технічної проблеми розробки моделей та методів темпоральної обробки відеопотоків для інтелектуальних інформаційних технологій пошуку погано структурованих даних.

Для досягнення поставленої мети потрібно було вирішення таких завдань:

- аналіз тенденцій розвитку методів обробки відеопотоків в аспекті їх структурування (виділення однорідних сегментів), пошуку валідних представників сегментів, дескрипторів відеокадрів для виділення перспективних засобів, що впливають на ефективність багатоаспектного індексування в бібліотеках динамічної візуальної інформації і метричного пошуку в CVBIR;

- синтез моделей та методів сегментації відео, що забезпечують створення процедур інформаційного пошуку з запитами 'ad exemplum' на основі побудови областей елімінації для довільних конфігурацій об'єктів в пошуковому просторі;

- розробка методів кластеризації і порівняння розбиття множин для підвищення валідності семантичного пошуку в великих бібліотеках відеоданих;

- побудова моделей і методів сегментації багатовимірних часових рядів, індукованих відеопотоками, шляхом виявлення зміни їх властивостей;
- створення методів пошуку ключових кадрів, що представляють сегменти відеоданих, в задачах пошуку відео на базі механізмів пошуку зображень у великих бібліотеках статичної візуальної інформації;
- розробка методів і моделей компресії відеокадрів і відео для темпоральної обробки відеопотоків в задачах багатоаспектної індексації;
- дослідження специфіки використання розроблених моделей і методів темпоральної обробки відеопослідовностей, створення і впровадження дослідницьких та спеціалізованих програмних засобів.

*Об'єкт дослідження* – процеси обробки та аналізу відеопотоків в мультимедіа інформаційних технологіях.

*Предмет дослідження* – методи просторової та часової сегментації відеопослідовностей для метричної індексації та пошуку в відеоколекціях.

*Методи дослідження* – при розробці та дослідженні методів та моделей сегментації для індексації відеопотоків використані основні положення математичного апарату розпізнавання образів та обробки зображень, аналізу багатовимірних часових рядів, штучних нейронних мереж та ідентифікації систем, а також елементи математичної статистики при проведенні та аналізі результатів експериментальних досліджень.

**Наукова новизна** дисертації полягає в постановці і вирішенні проблеми темпоральної обробки відеопотоків, що забезпечує структурування даних для інформаційного пошуку з запитами ‘ad exemplum’ в великих відеоколекціях в просторах зображень або ознак:

- вперше запропонована просторово-часова модель послідовної сегментації відео на основі відношень подібності / нерозрізненості відеокадрів, що дозволяє визначати межі сегментів з урахуванням семантики візуальної інформації та розв'язувати задачі темпоральної обробки в on line режимі;
- вперше в загальному вигляді отримані критерії елімінації об'єктів при метричному пошуку сегментів в відеоколекціях, що дозволяють скоротити час пошуку за рахунок виключення трудомістких обчислень відстаней від запитів до аналізованих об'єктів;
- вперше запропоновані методи фрагментної кластеризації зображень, які забезпечують локальну в полі зору темпоральну сегментацію відеорядів, що дозволяє здійснювати аналіз відеопослідовностей в околі просторових областей інтересу в реальному часі;
- вдосконалені методи виявлення змін властивостей індукованих відеопотоками багатовимірних часових рядів, які на основі налаштовуваних моделей, експоненціального згладжування і аналізу головних компонент на відміну від відомих методів дозволяють шукати відеосегменти при послідовному аналізі відеопотоків в різних ознакових просторах;
- отримали подальший розвиток методи сегментації багатовимірних часових рядів на основі адаптивного прогнозування відеопослідовностей, що на відміну від відомих методів дає можливість виділення сегментів відеопотоків з

урахуванням різноманітної передісторії вхідних даних;

– отримали подальший розвиток методи формування ключових кадрів відеосегментів, які на відміну від відомих використовують розбиття відеокадрів на основі діаграм Вороного, що дозволяє враховувати «просторовий зміст» зображень, підвищуючи валідність процедур пошуку;

– вдосконалені інтелектуальні методи компресії відеокадрів, що на відміну від відомих методів, забезпечує темпоральну обробку відеопослідовностей в плані їх структуризації для інформаційного пошуку.

**Практичне значення отриманих результатів** полягає в розробці методів і моделей сегментації відеопотоків, що створює передумови для багатоаспектної індексації в задачах інформаційного пошуку і забезпечує підвищення швидкодії пошуку в системах CBVIR із запитом «за зразком».

Результати дисертаційних досліджень використані в Харківській обласній державній адміністрації для системи управління архівами селекторних нарад з використанням відеозв'язку (акт від 09.10.2014 р), в ГУМВС України в Харківській області (акт від 28.01.2015 р) з метою детектування значущих фрагментів відео, в Харківському технічному центрі рентгенівського сервісу «Спектр» (акт від 14.12.2014 р) та на Заводі рентгенівського обладнання «Квант» (акт від 18.12.2014 р) для аналізу і виділення фрагментів в потоках рентгеноскопічної відеоінформації. Теоретичні і практичні результати роботи знайшли застосування в навчальному процесі Харківського національного університету радіоелектроніки при підготовці спеціалістів і магістрів за спеціальністю «Інформатика» (акт від 18.05.2016 р).

Результати теоретико-експериментальних досліджень реалізовані у вигляді прикладних і дослідницьких програмних комплексів, один з яких «Програмний комплекс просторово-часової сегментації відеоданих в задачах пошуку за запитом «за зразком» сертифікований УкрСЕПРО.

**Особистий внесок здобувача.** Усі наукові результати, що виносяться на захист, отримані здобувачем особисто. У роботах, написаних у співавторстві, дисертанту належать постановка задач, вибір і обґрунтування їх вирішення, а саме: [1] – розроблено моделі та методи структуризації та сегментації відеопотоків; [2] – запропоновано підхід до виділення характерних ознаки зображень; [3] – запропоновано метод порівняння розбиттів, що може бути використано при порівнянні кадрів відеопослідовностей при пошуку за запитом «за зразком»; [4] – запропоновано метод фрагментного стиснення зображень; [5] – виділено вектор характеристик, який дозволяє порівнювати результати інформаційного пошуку; [6] – розроблено метод побудови характеристичного вектору ознак для зображень; [7] – запропоновано метод стиснення зображень на базі інформаційного критерію; [8] – запропоновано нейромережевий алгоритм стиснення зображень; [9] – запропоновано алгоритм самонавчання карти Кохонена для задачі стиснення зображень; [10] – запропоновано метод пошуку змін у відеоряді на базі VAR моделі; [11] – запропоновано метод побудови областей достатності та необхідності, що може бути використано при порівнянні кадрів відеопослідовностей при пошуку «за зразком»; [12] – розроблено метод стиснен-

ня зображень за допомогою головних та незалежних компонент; [13] – запропоновано метод пошуку змін в потоці відеоданих; [14] – запропоновано метод детектування змін в відео даних на базі експоненційного згладжування; [15] – запропоновано метод опису відеоданих за рахунок критерію стабільності; [16] – запропоновано матричну прогноуючу модель, яку можливо використовувати для аналізу відеоданих; [17] – модифіковано опис ключових кадрів за допомогою діаграм Вороного вищих порядків; [18] – запропоновано метод детектування ключових кадрів; [19] – розроблено матричний метод адаптивного прогнозування багатовимірних часових рядів; [20] – розроблено метод кластеризації багатовимірних даних за умов їх великих обсягів; [21] – розроблено метод сегментації відеопослідовностей даних за допомогою аналізу текстур окремих кадрів; [22] – розроблено матричний метод кластеризації зображень в великих базах даних; [23] – розроблено нейромережу для прогнозування багатовимірних часових рядів; [24] – модифіковано метод  $J$ -середніх для фрагментної кластеризації зображень в великих базах даних; [25] – запропонована модифікація методу  $X$ -середніх для фрагментної сегментації зображень; [26] – розроблено метод пошуку інваріантних характеристик скелетонів бінарних зображень; [27] – розроблено метод нормалізації зображень, на основі пошуку геометричних властивостей їх окремих елементів; [28] – запропоновано метод порівняння зображень за допомогою аналізу їх сегментацій; [29] – розроблено нейромережевий метод аналізу багатовимірних часових рядів; [30] – розроблено on line метод темпоральної сегментації відео послідовності; [31] – запропоновано метод темпоральної сегментації відеопослідовностей; [32] – розроблено метод адаптивного прогнозування, що може бути використано для темпоральної сегментації відеоданих; [33] – розроблено модель та метод виділення ключової інформації в відеопослідовностях; в [34, 35, 39, 40, 43] конкретизовані і вивчені ознакові простори та методи пошуку характерних ознак; в [36] вивчені моделі і методи пошуку із запитом «за зразком»; в [37, 38, 42, 46, 47] введені нейромережеві методи стиснення відеокадрів для темпоральної обробки; в [41, 44] розглянуто методи виявлення змін властивостей багатовимірних часових рядів; в [45] описано фрагментні методи прогнозування відеорядів; в [48] описано моделі і методи структуризації відеопотоків; в [49, 51] запропоновано фрагментні методи кластеризації при темпоральній обробці відеоданих. Задекларований особистий внесок здобувача в роботах, виконаних у співавторстві, відповідає темі та змісту дисертації.

**Апробація результатів дисертації.** Основні результати дисертаційної роботи доповідались, обговорювались та були схвалені на науково-технічних конференціях та семінарах: International Conference on Graphics, Vision and Image Processing (Egypt, Cairo, 19-21 December, 2005); 14-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. WSCG'2006 (Czech Republic, Plzen, 30 January – 3 February, 2006); Міжнародній науковій конференції «Глобальні інформаційні системи. Проблеми та тенденції розвитку» (Туапсе, 3-6 жовтня 2006); 5-th International Conference I.Tech 'Information Research & Applications' (Bulgaria, Varna, 26-30 June, 2007);

міжнародних наукових конференціях «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту». ISDMCI (Євпаторія, 19-23 травня 2008 р., Євпаторія, 18-22 травня 2009 р., Залізний Порт, 28-31 травня 2014 р., Залізний Порт, 25-28 травня 2015 р., Залізний Порт, 24-28 травня 2016 р.); International Workshop ‘Computer Graphics, Vision and Mathematics’. GraVisMa 2009 (Czech Republic, Plzen, 2-5 September, 2009); 54-th Internationales Wissenschaftliches Kolloquium (IWK) ‘Information Technology Electrical Engineering – Devices and Systems, Materials and Technology for the Future’ (Germany, Ilmenau, 7-10 September, 2009), 10-th Industrial Conference ICDM 2010 ‘Advances in Data Mining. Applications and Theoretical Aspects’ (Germany, Berlin, 12-14 July, 2010); V, VI, VII міжнародних школах-семінарах «Теорія прийняття рішень» (Ужгород, 27 вересня – 1 жовтня 2010 р., Ужгород, 1-6 жовтня 2012 р., Ужгород, 29 вересня – 4 жовтня 2014), International Conference ‘Theoretical Information Systems Research’ (Taiwan, Taipei, 10-12 February, 2012); 12-th International Conference ‘Pattern Recognition and Information Processing’ (Belarus, Minsk, 28-31 May, 2014); Міжнародному науково-практичному форумі «Наука та бізнес» (Дніпропетровськ, 29-30 червня 2015 р.); Міжнародній науково-технічній конференції «Електротехнічні та комп’ютерні системи: теорія та практика» (Одеса, 20-24 липня 2015 р.).

**Публікації.** За результатами дисертаційних досліджень опубліковано 52 наукові роботи, у тому числі 1 монографія; 24 статті у наукових фахових виданнях України з технічних наук; 8 статей в закордонних журналах (загалом 18 статей в міжнародних наукометричних базах); 19 публікацій в матеріалах міжнародних науково-технічних конференцій та семінарів (3 в міжнародних наукометричних базах).

**Структура й обсяг дисертації.** Дисертація складається зі вступу, семи розділів, висновків, списку використаних джерел та додатка. Повний обсяг дисертації становить 341 сторінку; робота містить 74 рисунка (з них 3 на окремих сторінках); 9 таблиць; список використаних джерел, що включає 321 найменування та займає 36 сторінок; додаток на 8 сторінках.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**У вступі** обґрунтовано актуальність розробки нових моделей і методів темпоральної обробки відео для інформаційного пошуку – пошуку у поганоструктурованих даних, якими є відеодані, у випадку запитів ‘ad exemplum’ за рахунок побудови процедур елімінації заздалегідь неспівпадаючих із запитами даних та проведення структуризації відео в off line та on line режимах; 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, перелік публікацій за темою роботи, надано інформацію про особистий внесок автора.

**У першому розділі** в результаті ретроспективного аналізу встановлено,



що незважаючи на велику кількість робіт, пов'язаних з напрямом контекстного аналізу візуальної інформації (CBVIR), спрямованих на розв'язання таких задач як виділення та розпізнавання динамічних об'єктів під впливом будь-яких перешкод, трекінг або відстежування та прогнозування траєкторії руху об'єктів, детектування руху в полі зору та детектування будь-яких змін в відеоданих остаточне розв'язання задач, пов'язаних з аналізом відеоінформації ще далеко попереду. Це пов'язано в першу чергу з величезними обсягами інформації, яку потрібно обробити для отримання результату. Тому одними з базових напрямів є прелімінарна обробка відеопотоків, яка проводиться з метою зменшення обсягів інформації яку необхідно в подальшому аналізувати. До таких методів відносяться виділення переднього та заднього планів, виділення та класифікація областей інтересу, розпізнавання та опис об'єктів в цих областях інтересу. Однак, ці підходи мають переваги тільки у випадках окремих задач CBVIR, зокрема, коли фон в відеопотоках майже незмінний, як це відбувається в системах стеження, або існують багато інших обмежень, які дозволяють спростити виділення необхідної інформації в відеоданих. Коли ж йде мова про інформаційний пошук в відео с запитом «за зразком», то майже будь-які обмеження неможливі за рахунок того, що як запит так і валідна до нього інформація мають довільну природу. Тому були розглянуті підходи до зменшення інформації, яка потребує обробки, за рахунок побудови дескрипторів окремих кадрів з метою виділення в подальшому ключової інформації всього відеопотоку такі, як пошук меж та кутів, модель зорової уваги, кластеризація, онтологічні підходи, генетичні алгоритми, пошук різноманітної статистичної інформації, побудова та аналіз гістограм або корелограм, тощо. До характеристик/ознак, які використовуються в методах пошуку ключової інформації, слід віднести текстурні ознаки, ознаки форми/контур, колір та інші. Однак розглянуті методи пошуку ключової інформації та побудови характеристичного вектору ознак здебільшого оперують просторовою складовою відеопотоків та спрямовані на аналіз та інтерпретацію окремих кадрів відеопослідовностей, проте добре відомо, що однією з найважливіших особливостей відеоданих є їх залежність і в часовій складовій, тобто сильна кореляція між двома сусідніми кадрами дозволяє зовсім обійтись без просторового аналізу одного з них. Цей факт широко використовується в підходах до стиснення відеоданих, де серед найвідомішого слід відзначити пошук міжкадрової різниці. Проте більшість підходів до стиснення не орієнтована на контекстний аналіз відеопослідовностей. В той же час існує досить велика кількість задач аналізу відеопослідовностей, які орієнтовані на часову складову, зокрема, кольорова/текстурна класифікація, мультимодальний контекстний пошук, виділення текстової інформації в відеопотоці, семантична сумаризація, семантичний аналіз та автоматичне виділення сцен. Ще одним з підходів, який логічно використовувати в цьому випадку для інтерпретації відео, є аналіз часових рядів, зважаючи на те, що відеопослідовність є сукупністю кадрів, пов'язаних у часі. Однак слід відзначити що більшість існуючих підходів до аналізу часових рядів спрямована на одновимірний випадок, тому виникає потреба створення модифікацій на багатовимірний випадок.

В результаті проведеного аналізу сформульовані такі основні висновки. Глобальна акумуляція поганоструктурованих відеоданих вимагає розвитку інформаційних технологій в плані вирішення науково-технічної проблеми, пов'язаної зі структуризацією відеопотоків, встановленням семантичної відповідності пошукового запиту і пошукового образу в умовах апріорної невизначеності і обмежень на час прийняття рішень. Для створення перспективних CBVIR потрібна міждисциплінарна багатофакторна інтеграція методів обробки багатовимірних сигналів, розпізнавання образів, комп'ютерного зору, управління базами даних, аналізу багатовимірних часових рядів, яка в результаті продукує інструментарій інформаційних технологій для зберігання, обробки, передачі, відображення та інтерпретації поганоструктурованих або зовсім неструктурованих даних великих обсягів. Розвиток засобів on line структуризації відеопотоків на основі просторової (внутрішньокадрової) і часової (міжкадрової) сегментації відеоданих представляє певну перспективу для скорочення семантичного розриву між ознаковими уявленнями відео і цілями предметно-орієнтованої інтерпретації, що в результаті створює передумови валідності інформаційних технологій мультимедіа та гіпермедіа. На підставі аналізу стану науково-методичного апарату CBVIR сформульована проблема створення теоретичних та прикладних основ темпоральної обробки відеопотоків для інформаційного пошуку в великих відеоколекціях

**Другий розділ** присвячений синтезу моделей темпоральної сегментації відеопослідовностей та методів пошуку із запитом 'ad exemplum', заснованих на елімінації об'єктів, які не задовольняють вимогам подібності із запитом без обчислення відстаней до них.

Специфікою відеопотоків і відеоколекцій є їх практично повна первинна неструктурованість, що істотно ускладнює створення інформаційних технологій (зберігання, пошуку, обробки, аналізу та інтерпретації). «Одиницею» структуризації є відеосегмент – послідовність кадрів, нерозпізнаних з точки зору заданого критерію. Всі ієрархічні конструкції, що забезпечують семантичну індексацію відеоданих, базуються саме на побудові розбиттів або покриттів відеопослідовностей.

Відеопотік – це впорядкований за часом набір зображень (відеокадрів)  $F = \{B_1, \dots, B_i, \dots, B_N\}$ , де  $i$  – дискретний час,  $B_i(z)$  – зображення в полі зору  $D$  в момент часу  $i$ ,  $z \in D \subset \mathbb{R}^2$ . Відеопотік  $F$  індукує часовий ряд  $X = \{x_1, \dots, x_N\}$ , в якому, за звичай  $x_i \in \mathbb{R}^p$ . Темпоральна модель відеоряду може бути абсолютною, що задається множиною зображень або ознак, та відносною – в вигляді відношень між відеокадрами, наприклад, різницевою моделлю, представляється  $N \times N$  матрицею відмінностей між відеокадрами  $D = \{d_{ij}\}_{i,j=1}^N$ , де  $d : F \times F \rightarrow \mathbb{R}_0^+$  – функціонал, який відповідає щонайменше аксіомам рефлексивності  $d(B', B'') = 0 \iff B' = B''$  та симетричності  $d(B', B'') = d(B'', B')$ . В цьому випадку отримуємо функціонал подібності, що продукує покриття відеоряду. Якщо  $d$  – метрика, отримуємо функціонал нерозрізненості, тобто розбиття відео-

послідовності.

При on line обробці відео оперувати можливо тільки передісторією, що призводить до необхідності побудови специфічного інструментарію.

Нехай на декартовому квадраті  $F \times F$  множини зображень або на декартовому квадраті множини ознак  $X \times X$  (парадигматично вони є рівносильними) задано деякий набір бінарних відношень  $T = \{t_g\}$ ,  $g \in G$  ( $G$  – індексна множина), що визначає відповідність (толерантність) або нерозрізненість (еквівалентність) пар  $B_i, B_j \in F$  ( $x_i, x_j \in X$ ). Якщо еквівалентність породжує природну структурування відеопослідовностей у формі розбиття, то толерантність потребує або формує з загальних частин новий клас, або перерозподіляє відеокадри по вже існуючим класам.

Основною одиницею структурування відео є сегмент – впорядкований за часом набір зображень  $S_{ab} = \{B_a, B_{a+1}, \dots, B_{b-1}, B_b\}$ . Зображення  $B_i, B_j \in F$  є зв'язними відносно відношення  $t_g$ , якщо  $\langle B_i, B_j \rangle \in t_g$  ( $B_i t_g B_j$ ). Сегмент  $S_{ab}$  є зв'язним, якщо " $i, j \in \{a, \dots, b\} \Rightarrow B_i t_g B_j$ " та " $a > 1 \Rightarrow \langle B_{a-1}, B_a \rangle \in t_g$ ", " $b < N \Rightarrow \langle B_b, B_{b+1} \rangle \in t_g$ ".

Зв'язний сегмент позначимо  $S_{ab}^* = [B_a] t_g$ , маючи на увазі, що  $[B_a] t_g$  – клас толерантності або еквівалентності. Замість  $B_a$  можливо використовувати будь-яке зображення, але с практичної точки зору доцільно застосовувати ключовий кадр, найбільш «схожий» на всі інші зображення сегмента, оскільки перевірка відношень  $T = \{t_g\}$ ,  $g \in G$ , зазвичай, виконується з деякою похибкою. Ключовим кадром  $B_{ab}^*$  зв'язного сегменту  $S_{ab}^*$  є зображення  $B_{ab}^* = \arg \min_{B \in [B_a] t_g} \max_{i=a}^b d(B, B_i)$ , яке можливо також синтезувати.

Сегмент є сильно-зв'язним ( $+S_{ab}$ ), якщо " $B_i, B_j \in S_{ab}, g \in G \Rightarrow \langle B_i, B_j \rangle \in t_g$ "; якщо " $B_i, B_j \in S_{ab}, \{g\}, g \in G \Rightarrow B_i t_g B_j$ ", сегмент буде слабо-зв'язним ( $-S_{ab}$ ). Якщо набір  $\{g\}$  однаковий для всіх відеокадрів сегмента, маємо однорідну слабо-зв'язність.

Тим самим, структурування відеоряду – це пошук представлення відеопослідовності  $F$  в вигляді зв'язних сегментів  $S_{a(k)b(k)}^* \in \mathcal{A}$ ,  $k = \overline{1, K}$  таких, що  $F = \bigcup_{k=1}^K S_{a(k)b(k)}^*$ , " $i \neq j \in \{1, 2, \dots, K\} \Rightarrow S_{a(i)b(i)}^* \cap S_{a(j)b(j)}^* = \mathcal{A}$ ".

Фактично on line сегментація базується на прогнозуванні або ретроаналізі деякої слідкуючої функції, яку з пам'яттю на  $d$  відеокадрів в самому загальному вигляді можливо представити в формі:

$$w_d(k_0) = \max\{k > k_0 : "k' \in [k_0, k], k'' = k' - l, l \in \{1, \dots, \min\{d, k' - k_0\}\} \Rightarrow B_{k'} t B_{k''}"\}.$$

На основі отриманої структурування для довільного метричного простору  $F$  сегментів відео досліджені можливості пошуку із запитом 'ad exemplum' з

використанням off line отриманих відстаней між сегментами колекції. При цьому on line розраховуються відстані лише для обраного набору так званих опорних точок, що і забезпечує суттєве зменшення часових затрат на пошук. Пошук виконується всередині деякої конфігурації  $\mathbb{K}=\{x_1, \dots, x_n, x_1^*, \dots, x_m^*\}$ , де  $x_1, \dots, x_n$  – сегменти з відеоколекції,  $x_1^*, \dots, x_m^*$  – опорні точки. Конфігурація може представляти собою всю відеоколекцію або будь-яку її частину (наприклад, прелімінарно знайдені кластери).

Результатом метричного пошуку  $p \in n$  сегментів в конфігурації  $K \in F$  із запитом  $y \in F$ , обмеженнями  $r(y, x_l) \in D$ ,  $l \in \{i_1, i_2, \dots, i_p\}$  при відомих відстанях  $r(y, x_i^*) = d_i$ ,  $i = \overline{1, m}$  до множини опорних точок  $\{x_k^*\}_{k \in K} \in$  впорядкована за зростанням відмінностей між сегментами множина  $K^s = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\} \in K$ , для якої при  $0 < s \in p \in n$  справедливі співвідношення

$$\begin{aligned} & "x_{ij} \in K^s, "x \in K \setminus K^s, "y \in F \quad \rho(y, x_{ij}) \in D, D^3 0, \\ & \rho(y, x_{ij}) \in \rho(y, x), \rho(y, x_{ij}) \in \rho(y, x_{ij+1}), j = \overline{1, s-1}. \end{aligned}$$

Якщо  $r(x_i, x_j) = a_{ij}$ ,  $r(x_i^*, x_j^*) = b_{ij}$ ,  $r(x_i, x_j^*) = g_{ij}$ , матриця  $P$  off line відстаней для конфігурації  $\mathbb{K}=\{x_1, \dots, x_n, x_1^*, \dots, x_m^*\}$  має вигляд

$$P = \begin{array}{cccc|cccc} \emptyset & a_{12} & K & a_{1n} & | & g_{11} & K & K & g_{1m} & \emptyset \\ \circ & 0 & K & K & | & g_{21} & K & K & g_{2m} & \div \\ \circ & & O & K & | & K & K & K & K & \div \\ \circ & & & a_{n-1n} & | & K & K & K & K & \div \\ \circ & & & 0 & | & g_{n1} & K & K & g_{nm} & \div \\ \circ & & & & | & - & - & - & - & \div \\ \circ & 0 & & & | & 0 & b_{12} & K & b_{1m} & \div \\ \circ & & & & | & & 0 & K & K & \div \\ \circ & & & & | & & & O & b_{m-1m} & \div \\ \emptyset & & & & | & & & & 0 & \emptyset \end{array}$$

Для конфігурації  $\mathbb{K}=\{x_1, \dots, x_n, x_1^*, \dots, x_m^*\}$  область елімінації  $r$ -го порядку – це область  $D_r \in C_{n+m}^{2+m}$  така, що якщо

$$\left( a_{ij} \right)_{\substack{i, j = \overline{1, n} \\ i < j}}, \left( b_{ij} \right)_{\substack{i, j = \overline{1, m} \\ i < j}}, \left( g_{ij} \right)_{\substack{i = \overline{1, n} \\ j = \overline{1, m}}}, \left( d_i \right)_{i = \overline{1, m}} \in D_r,$$

то серед точок  $x_1, \dots, x_n$  можливо вказати  $r$  точок, серед яких відсутня найбли-

жча до  $y \in F$ . Тим самим, варіюючи опорні точки  $x_1^*, x_2^*, \dots, x_m^*$ , маємо можливість переходити до скорочення потужності конфігурації  $K_i$  ( $K_0 = K, K_{i+1} = K_i / \{x_{k1}, x_{k2}, \dots, x_{kr}\}$ ), що аналізується.

В розділі детально вивчена побудова областей елімінації для конфігурацій, коли з пари сегментів при різних наборах опорних точок вибирається більш схожий на запит сегмент, тобто  $K = \{x_1, x_2, x^*\}$ ,  $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$ ,  $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*, \dots, x_m^*\}$ , а також, коли вибирається найбільш схожий сегмент з  $n$  об'єктів відеоколекції при одній опорній точці, тобто  $\mathbb{K} = \{x_1, \dots, x_n, x^*\}$ . Приведена геометрична інтерпретація цих областей елімінації.

Для загального випадку  $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, \dots, x_m^*\}$  встановлено:

$$D_r = \bigcup_{t=1}^m D_r(x_t^*).$$

$$D_r(x_t^*) = V_r(x_t^*) \cap C(x_t^*) \cap L(x_t^*),$$

де

$$V_r = \bigcup_{\substack{(s_1, \dots, s_r) \in \{1, \dots, n\} \\ s_i \neq s_j, i, j=1, r}} V(x_{s_1}, \dots, x_{s_r}, x_t^*), \quad V : \begin{cases} |\delta_t - \gamma_{1p_1}| \leq |\delta - \gamma_{1s_1}|, \\ |\delta_t - \gamma_{1p_1}| \leq |\delta - \gamma_{1s_2}|, \\ \dots \\ |\delta_t - \gamma_{1p_l}| \leq |\delta - \gamma_{1s_r}|, \end{cases}$$

$$C = \bigcup_{\substack{i, j, k=0, n \\ i \neq j \neq k}} C_{ijk}, \quad C : \begin{cases} |\alpha_{ij} + \alpha_{ik}| \geq \alpha_{jk}, \\ |\alpha_{jk} + \alpha_{ik}| \geq \alpha_{ij}, \\ |\alpha_{ik} + \alpha_{ij}| \geq \alpha_{jk}, \end{cases} \quad L : \begin{cases} |\delta_t| \leq D, \\ |\delta_t + \alpha_{01}| \leq D, \\ \dots \\ |\delta_t + \alpha_{0n}| \leq D \end{cases}$$

(тут " $i \in \{1, \dots, n\}$ "  $\Rightarrow a_{0i} = a_{i0} = g_i$ ).

Нехай для фіксованої вибірки  $x_{s_1}, \dots, x_{s_r}$  з точок  $x_1, \dots, x_n$ , для якої з урахуванням  $\{x_{p_1}, \dots, x_{p_{n-r}}\} = \{x_1, \dots, x_n\} \setminus \{x_{s_1}, \dots, x_{s_r}\}$ ,  $\mu_l = |\delta_i - \gamma_{1l}|$ ,  $M_l = \delta_i + \gamma_{1l}$ ,  $(l \in L)$  сформовані два набори:  $m_{s_1}(x_i^*), \dots, m_{s_r}(x_i^*)$ ,  $M_{p_1}(x_i^*), \dots, M_{p_{n-r}}(x_i^*)$ ,  $i = \overline{1, m}$ . Тоді, якщо

$$G(x_{s_1}, \dots, x_{s_r}) \leq g(x_{s_1}, \dots, x_{s_r}),$$

де  $g(x_{s_1}, \dots, x_{s_r}) = \min_{i=1, m} \min_{j=1, r} \{m_{s_j}(x_i^*)\}$ ,  $G(x_{s_1}, \dots, x_{s_r}) = \min_{i=1, m} \min_{j=1, n-r} \{M_{p_j}(x_i^*)\}$ ,

то виключення з конфігурації  $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, \dots, x_m^*\}$  точок  $x_{s_1}, \dots, x_{s_r}$  зберігає точку найближчу до  $y$ .

Таким чином, побудова в загальному вигляді областей елімінації в дові-

льному метричному просторі забезпечує для довільних конфігурацій при довільних наборах опорних даних (центрах категоризації, кластеризації та ін.) реалізацію різних стратегій пошуку, характерною властивістю яких є виключення (безпосередньо при пошуку) необхідності трудомістких обчислень відстаней від запиту до елементів бібліотек відео.

**Третій розділ** присвячений розробці фрагментних методів кластеризації для вирішення задач сегментації як статичних зображень, так і потоків відео за умов невизначеності. Особливу групу методів кластеризації утворюють алгоритми, призначення для обробки інформації, що зберігається в надвеликих базах даних, де на перший план виходять швидкодія і простота чисельної реалізації. У даній ситуації як досить ефективно показали себе методи кластеризації, засновані на щільності розподілу даних.

Для розробки фрагментної кластеризації зображень, заснованої на щільності, може бути використана модифікація DENCLUE (DENSity-based CLUstEring), яка призначена для обробки великих об'ємів даних, формуючи кластери довільної форми при високому рівні шумів. Використані такі припущення:

– вплив кожного вектора-образу на сусідні спостереження формально може бути описано за допомогою деякої функції, зазвичай ядерної, званої функцією впливу, яка описує взаємозв'язок всіх спостережень в деякому околі даного способу;

– загальна щільність розподілу даних в багатовимірному просторі ознак формально описується як сума функцій впливу кожного спостереження;

– кластери визначаються як околиці аттракторів щільності (D-аттракторів), які є за суттю локальними максимумами загальної функції щільності розподілу даних.

Для деякої довільної точки у просторі ознак  $y$  її вплив на образ  $x$  може бути описано за допомогою функції впливу, при цьому найбільш часто в якості таких функцій використовується або прямокутна конструкція

$$f(x, y) = \begin{cases} 0, & \text{якщо } D(x, y) > \sigma; \\ 1, & \text{інакше,} \end{cases}$$

або гауссіан

$$f(x, y) = \exp\left\{-\frac{D^2(x, y)}{2\sigma^2}\right\} \quad (1)$$

Тоді для множини спостережень  $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$  загальна функція щільності може бути представлена у вигляді

$$f^X(x) = \prod_{k=1}^N f(x, x(k)). \quad (2)$$

Ця функція є сумою ядерних функцій, характеризується наявністю множини локальних екстремумів-максимумів, іменованих D-аттракторами, кожен з яких представляє окремий кластер і може бути визначений за допомогою тих або інших оптимізаційних процедур. При цьому довільна точка  $x$  притягується до

D-аттрактора  $x^*$ , якщо послідовність ітерацій

$$x^i = x^{i-1} + \eta \frac{\tilde{N}f^x(x^{i-1})}{\|\tilde{N}f^x(x^{i-1})\|}; i=1, 2, \dots; x^0 = x \quad (3)$$

сходиться до  $x^*$ .

Якщо в якості  $f^x(x)$  використовувати (1), (2), то (3) набуває вигляду

$$x^i = x^{i-1} + \eta \frac{\sum_{k=1}^N (x(k) - x^{i-1})f^{x^i}, x(k)}{\left\| \sum_{k=1}^N (x(k) - x^{i-1})f^{x^i}, x(k) \right\|}. \quad (4)$$

Кожен з D-аттракторів характеризується власною функцією щільності

$$f^{x^*}(x) = \sum_{x(k) \hat{I} \text{ near } x^*} \delta(x, x(k)),$$

де  $\text{near } x^* = \{x(k): D(x^*, x(k)) \leq \sigma_{\text{near}}\}$ , а її екстремум визначає координати центроїда кластера.

При вирішенні завдань кластеризації завжди передбачається, що кожне багатовимірне спостереження-образ описується  $n$ -вимірним вектором  $x(k)$ , а весь процес вирішення пов'язаний саме з векторними операціями. У ситуації, коли є велика колекція зображень, що підлягають кластеризації, кожне двовимірне зображення спочатку має бути піддано векторизації, далі вирішується завдання кластеризації, а її результат піддається девекторизації – переводить векторний опис в матричну форму. Істотно спростити процес кластеризації масивів можна, не переводячи їх у векторну форму, а безпосередньо, оперуючи з матрицями. Таким чином, набором вихідних образів є набір матриць  $x(k) = \{x_{i_1 i_2}(k)\}$ ,  $k = 1, 2, \dots, N$ ,  $x(k) \hat{I} \text{ } m' \times n$ . Далі, вводячи замість векторної евклідової норми її сферичний матричний аналог  $D_S^2(x, y) = Sp(x - y)(x - y)^T$ , можливо ввести матричну функцію впливу

$$f_S(x, y) = \exp\left\{-\frac{D_S^2(x, y)}{2\sigma^2}\right\} = \exp\left\{-\frac{Sp(x - y)(x - y)^T}{2\sigma^2}\right\} \quad (5)$$

та матричну функцію щільності  $f_S^x(x) = \sum_{k=1}^N f_S(x, x(k))$ .

При цьому довільна ( $m' \times n$ ) матриця-образ  $x$  притягується до матричного D-аттрактора  $x^*$ , якщо послідовність ітерацій типу (3)

$$x^i = x^{i+1} + \eta \frac{F}{\sqrt{Sp(FF^T)}}, i=1, 2, \dots; x^0 = x$$

збігається до  $x^*$ , де  $F = \frac{\partial^2 f_S(x)}{\partial x_{i_1 i_2}^2} - (m \times n)$  матриця. Якщо ж в якості матричної функції використати вираз (5), то алгоритм оптимізації може бути записано у вигляді

$$x^i = x^{i-1} + \eta \frac{\mathring{a}_{k=1}^N r(k, i-1)}{(\text{Sp}(\mathring{a}_{k=1}^N r(k, i-1))(\mathring{a}_{k=1}^N r(k, i-1))^T)^{\frac{1}{2}}},$$

де  $r(k, i-1) = (x(k) - x^{i-1})f_s(x^{i-1}, x(k))$ , що за суттю є розширенням (4) на матричний випадок.

Використання замість векторного опису його матричного аналогу дозволяє істотно підвищити швидкодію процесу обробки інформації та уникнути ряду проблем, що виникають в завданні кластеризації даних, що описуються векторами високої розмірності. Крім запропонованого підходу також можна використовувати матричні модифікації інших методів кластеризації, зокрема,  $X$ -середніх та  $J$ -середніх.

Основна ідея методу  $X$ -середніх полягає в багаторазовому застосуванні до вихідного матричного об'єму даних  $X$  алгоритма  $k$ -середніх з різними значеннями кількості кластерів і оцінюванні отриманих результатів за допомогою того або іншого критерію, заснованого на байєсівському оцінюванні. Особливістю запропонованої матричної модифікації є можливість обробки матричних сигналів у відсутності інформації про статистичні характеристики цих сигналів та кількості кластерів, яка автоматично визначається в процесі аналізу вхідного масиву даних.

Суть методу  $J$ -середніх полягає в тому, що при знаходженні деякого центроїда, в якому процедура кластеризації зупиняється, організовується процес переміщення цього центроїда в незайняті точки, що знаходяться в його околі. Такі скачки «центроїд – незайнята точка» проводяться до тих пір, поки не буде знайдений більш глибокий екстремум, відповідний меншому значенню цільової функції. Найсуттєвіші проблеми, що пов'язані з практичним використанням методу  $J$ -середніх, виникають при завданні розмірів (діаметра) околу стрибка навколо кожного центроїда. У загальному випадку малий окіл може не виявити незайняті точки з меншими значеннями цільової функції, а надто великий окіл різко збільшує обчислювальну складність алгоритму, особливо у випадках, коли ці околи перетинаються.

Необхідно особливо підкреслити, що при використанні традиційних методів кластеризації для сегментації зображень, як правило, послаблюється одна з вимог – зв'язність областей у полі зору. У зв'язку з цим для підвищення валідності сегментації необхідна постобробка. Зручним інструментом тут виявляються алгоритми бінарної морфології, оскільки кожену область можна трактувати як двоградаційні зображення. Крім того, різні характеристики (насамперед геометричні) таких областей створюють передумови для формування дескрипторів зображення в цілому. Цьому присвячені подальші частини розділу. По-



перше, проаналізовані метричні властивості розбиттів, задля чого введений функціонал на множині довільних розбиттів  $r(P, G) = \sum_{i=1}^n \sum_{j=1}^m |p_i \vee g_j| |p_i| |g_j|$  та доведено, що цей функціонал є метрикою.

По-друге, розглянуто побудову характеристичних векторів зображень/відеокадрів, задля чого проаналізовані існуючі підходи та дано рекомендації по використанню тих чи інших характеристик. Загальна класифікація ознак форми наведена в таблиці 1.

Таблиця 1 – Класифікація підходів до аналізу форми сегментів зображень

Підходи до аналізу форми областей	Методи та характеристики
Одновимірні функції	Різні функції, які будуються на основі точок контуру, зокрема, функція комплексних координат
Полігональна апроксимація	Методи злиття та розбиття
Функції взаємозв'язку областей	Опукла оболонка, контекст області, шок графі (скелетони), матриця області (представлення в полярних або декартових координатах), тощо
Моменти	Моменти меж та областей такі як, Зерніка, інваріантні, алгебраїчні, радіальні (Чебишева), гомоцентричні, ортогональні і т.д.
Просторовий масштаб	Точки перетину областей, кривизна областей.
Перетворення областей	Гармонічні функції, перетворення Фур'є, хвильові перетворення (Wavelets), радіальні перетворення, R-функції, тощо

Аналіз характеристик дозволяє стверджувати, що комплексні координати і функція відстані, при розрахунку з урахуванням центрів ваги об'єкта є інваріантними щодо зміщення, але в той же час вони не інваріантні до масштабування і обертання. Отже, можна зробити висновок, що в векторі характеристик об'єкта можна використовувати одну з функцій, а саме ту, що буде давати найкращий результат у взаємодії з іншими характеристиками.

Лінійною щодо афінних перетворень є функція площі, тому можна зробити висновок, що ймовірність помилки значення при розрахунку цієї характеристики буде мінімальною. Функція кута прямої дотичній чутлива до шуму і дискретності, тому при використанні цієї характеристики необхідно спочатку реалізувати фільтрацію зображення.

У разі, коли необхідно віднести об'єкт до якогось класу і дрібні деталі цього об'єкта не настільки важливі, можна використовувати полігональну апроксимацію. Вона згладжує контур, що «нормалізує» форму об'єкта і спрощує класифікацію. Доцільно використовувати і функцію опуклої оболонки, яка відображає відношення периметра фігури до його опуклої області, тобто допомагає визначити, наскільки об'єкт близький до заданого класу.

У той же час, якщо вести мову про завдання аналізу відеопотоків і пошуку із запитом «за зразком», вектор ознак не повинен містити всі описані параметри. Це пов'язано з тим, що крім підвищення складності обчислювальної моделі, можливе виникнення ситуації, аналогічної «прокляттю розмірності», коли деякі характеристики шляхом компенсації значень один щодо одного будуть «приховувати» особливості сегментів, тобто приводити до помилкової ідентифікації. Таким чином, можна і необхідно говорити про спільне використання метрики на розбитті з характеристиками типу площі, периметру, скелета та кута нахилу до дотичної.

**Четвертий розділ** присвячений моделям та методам сегментації відеоданих на основі аналізу індукованих часових рядів. Оскільки багатовимірні часові ряди, відповідні відеопослідовностям, як правило, містять надлишкову інформацію, їх аналіз здійснюється на основі контролю деяких узагальнених характеристик. Задача радикально ускладнюється у випадках, коли обсяг вибірки не фіксований, що притаманне послідовному покадровому аналізу відеоданих в реальному або наближеному до цього масштабі часу. У подібних зростаючих вибірках кількість сегментів є невідомою, а в силу найбільшої варіативності змісту відео і умов реєстрації ніяких апріорних припущень про властивості відеоряду зазвичай зробити не можна. Таким чином, на перший план виступає вже не просто сегментація, а послідовне виявлення моментів зміни властивостей відеоданих.

Розглянуто моделі багатовимірних часових рядів стосовно до задачі інформаційного пошуку у відеоданих, а також методи виявлення розладнань часових рядів для виділення однорідних фрагментів відеоданих. Зокрема, виявлення властивостей багатовимірних часових рядів в процесі адаптивної ідентифікації VAR моделі. При цьому вихідною інформацією для розв'язання задачі ідентифікації та виявлення змін є тільки лише  $n$ -вимірний часовий ряд  $F$ , значення якого в реальному часі надходять на адаптивний ідентифікатор.

В матричному вигляді VAR-модель записується в вигляді

$$B(i) = KB\Phi(i) + x(i), \quad (6)$$

де матриця апріорі невідомих параметрів  $K$  містить практично всю необхідну інформацію про властивості сигналу;  $B\Phi(i) = (1, B^T(i-1), \dots, B^T(i-p))^T$ .

Задача ідентифікації полягає в тому, що відповідною до реального сигналу (6) є настроювана модель

$$\bar{B}(i) = K(i-1)B\Phi(i), \quad (7)$$

матриця параметрів  $B\Phi(i)$  якої уточняється на кожному такті часу  $i$  шляхом мінімізації прийнятого критерію ідентифікації, що є функцією різниці розрахованих  $\bar{B}(i)$  та експериментальних даних  $B(i)$ . При цьому синтезована модель (7) повинна бути працездатною і в режимі прогнозування, яку можливо представити в вигляді

$$\begin{cases} \hat{K}(i) = K(i-1) + g(i)e(i)B\Phi^T(i), \\ \hat{e}(i) = B(i) - \hat{B}(i) = B(i) - K(i-1)B\Phi(i), \end{cases}$$

де  $g(i)$  – матричний коефіцієнт посилення алгоритму, визначаючий його властивості,  $e(i)$  – векторна похибка ідентифікації.

Запропоновано багатовимірну модифікацію експоненційної стохастичної апроксимації у вигляді

$$\begin{cases} K(i) = K(i-1) + \frac{e(i)B^T(i)}{br(i-1) + \|B\Phi(i)\|^2}, \\ r(i) = br(i-1) + \|B\Phi(i)\|^2. \end{cases}$$

В якості слідкувального сигналу пропонується використовувати співвідношення

$$T_j^{TL}(i) = \frac{T\phi(i)}{d_j(i)}, \quad (8)$$

де  $T\phi(i) = (1-b)e_j(i) + bT\phi(i-1)$  не загальна сума відхилень, а згладжена похибка; при цьому повинна виконуватись умова  $b \leq b$ ;  $e_j(i) = b_j(i) - \bar{b}_j(i, b)$ .

При  $b \neq b$  слідкувальний сигнал буде варіюватися між  $-1$  та  $+1$ . Для введення автоматичного зворотного зв'язку запропоновано розраховувати параметр згладжування згідно формули  $b(i) = 1 - |T^{TL}(i)|$ , а розладнання в сигналі фіксувати при суттєвих змінах  $b(i)$ .

Зростання слідкувального сигналу свідчить про збільшення розбіжності між моделлю і контрольованою послідовністю, для компенсації якої необхідна більш швидка реакція ідентифікатора, яка забезпечується більш низьким значенням параметру згладжування. Таким чином, забезпечується негативний зворотний зв'язок. В такому випадку різкі стрибки слідкувального сигналу визначаються найкраще. Саме тому ця форма може бути корисною при аналізі змін часового ряду, описуваних співвідношеннями (6) – (8).

Основним недоліком описаного підходу є значна кількість параметрів настоюваної моделі ( $n(Nn+1)$ ), що може викликати певні труднощі при великих  $n$  та високих частотах надходження інформації на обробку.

Достатньо вдалим математичним апаратом для розв'язання задачі пошуку змін одновимірних стохастичних часових рядів є експоненційне згладжування. Введемо в розгляд елементарну модель вигляду  $b_j(i) = k_j + x_j(i)$  та припустимо, що коефіцієнт  $k_j$  час від часу може змінюватися стрибком. Задачу можливо звести до оцінки поточного значення  $k_j(i)$  по  $i$  минулим спостереженням. Оскільки значення  $k_j$  з плином часу змінюється, то для отримання цієї оцінки спостереження  $b_j(i), b_j(i-1), \dots$  потрібно брати з більшою вагою ніж спостереження, отримані раніше. В принципі, цю задачу можливо успішно розв'язати за допомогою методу ковзного вікна, при цьому, якщо заданий розмір цього вікна  $t$ , то експоненційне згладжування набуває вигляду

$$t_j(i) = ab_j(i) + (1-a)t_j(i-1) = ab_j(i) + bt_j(i-1).$$

Однак в рамках проблеми аналізу відеопослідовностей доцільно ввести експоненційне згладжування багатовимірних послідовностей у вигляді

$$t(i) = AB\phi(i) + (I - A)t(i - 1),$$

де  $t(i) = (t_1(i), t_2(i), \dots, t_n(i))^T$ ,  $A = \text{diag}(a_1, a_2, \dots, a_n)$  -  $n \times n$  діагональна матриця,  $I$  -  $n \times n$  одинична матриця.

Якщо для контролю змін використовувати сигнал

$$\begin{cases} T_j^{TL}(i) = \frac{T\phi(i)}{d_j(i)}, \\ T\phi(i) = a_j e_j(i) + b_j T\phi(i - 1), \\ d_j(k) = a_j |e_j(k)| + b_j d_j(i - 1) \end{cases}$$

по аналогії, можливо ввести його векторний аналог

$$\begin{cases} T^{TL}(i) = \frac{T\phi(i)}{d_j(i)}, \\ T\phi(i) = A\phi_j(i) + (I - A)T\phi(i - 1), \\ d_j(i) = \text{diag}(a_j |e_j(i)| + (1 - a_j)d_j(i - 1)), \end{cases}$$

при цьому контролюється кожна компонента ( $n - 1$ ) – вектору  $T^{TL}(i)$ .

Виникнення змін в процесі виявляється шляхом неперервної перевірки нерівності  $\max_i (T^{TL}(i) - T^{TL}(i - 1)) \geq g$ .

Важливою проблемою при аналізі великих масивів (як за обсягом, так і за розмірністю) спостережень, заданих у формі часових рядів, є їх стиснення з метою виділення латентних факторів, що визначають внутрішню структуру контрольованого сигналу, що в кінцевому підсумку має на меті зробити вихідний часовий ряд більш просто інтерпретувемим з точки зору виявлення змін властивостей.

Для вирішення цього завдання запропоновано використовувати перетворення Карунена-Лоева. Вхідною інформацією є ( $i \times n$ ) матриця спостережень

$$B = \begin{pmatrix} b_1(1) & b_2(1) & \dots & b_n(1) \\ \vdots & \vdots & \vdots & \vdots \\ b_1(u) & b_2(u) & \dots & b_n(u) \\ \vdots & \vdots & \vdots & \vdots \\ b_1(i) & b_2(i) & \dots & b_n(i) \end{pmatrix}$$

що утворена з  $i \times n$ -вимірних векторів спостережень  $B(u) = (b_1(u), b_2(u), \dots, b_n(u))^T$  (що для задачі інформаційного пошуку у відеоданих є представленням кадру за допомогою деякого вектору характеристик) та її кореляційна ( $n \times n$ ) матриця

вигляду

$$R(i) = \frac{1}{i} \mathop{\text{a}}^i_{u=1} B^c(u) B^{cT}(u),$$

де  $B^c(u) = B(u) - \bar{B}$  центровані відносно середнього вхідні дані.

Перша головна компонента, що несе максимум інформації про контрольований сигнал, може бути знайдена шляхом максимізації локального критерію  $I_w^1(i) = \frac{1}{i} \mathop{\text{a}}^i_{u=1} (B^c(u)w^1)^2$  за допомогою стандартного методу невизначених множників Лагранжа. Далі з кожного вектора  $B^c(u)$  вилучається його проекція на першу головну компоненту і обчислюється перша головна компонента залишків, що є другою головною компонентою вихідних даних і ортонормальна першій. Третя головна компонента обчислюється шляхом проектування кожного вихідного вектора на першу і другу головні компоненти, вилучення цієї проекції з кожного  $B^c(u)$  і знаходження першої головної компоненти отриманих залишків, що є третьою компонентою вихідних даних. Решта головних компонент обчислюється рекурсивно згідно з описаною стратегією.

До теперішнього часу розроблено досить розвинене математичне і програмне забезпечення для реалізації перетворення Карунена-Лоева, об'єднане одним загальним недоліком: необхідністю апріорного завдання матриці фіксованої розмірності. Якщо ж дані надходять послідовно у реальному часі, стандартні процедури факторного аналізу стають непрацездатними. Тому, слід використовувати рекурентні процедури реального часу для знаходження власних векторів матриці  $R(i)$  шляхом послідовної обробки спостережень багатовимірного часового ряду  $B(1), B(2), \dots, B(i), B(i+1) \dots$  без обчислення самої кореляційної матриці.

Для попередньо центрованих даних алгоритм навчання має вигляд

$$\begin{cases} w^1(k+1) = w^1(k) + h(k+1)(x(k+1) - y(k)w^1(k))y(k+1), \\ y(k+1) = x^T(k+1)w^1(k), w^1(0) = 0, y(1) = x^T(1)w^1(0), \end{cases}$$

де  $h(k+1)$  – параметр кроку навчання, що обирається достатньо малим для забезпечення стійкої роботи алгоритму.

Алгоритм забезпечує нормування вектору  $\|w^1(k)\|^2 = 1$  сам вектор  $w^1(k)$  є власним вектором матриці  $R(k)$ , що відповідає максимальному власному значенню, а вихідний сигнал  $y(k)$  характеризується максимально можливою дисперсією, тобто містить максимум інформації про багатовимірний входовий сигнал  $x(k)$ .

Далі входовий сигнал  $x(k)$  піддається експоненційному згладжуванню, фільтруючому шумові компоненти, а виявлення змін властивостей відбувається за допомогою сигналу  $T^{TL}(k)$ .

Ще одним підходом до виділення сегментів відео на базі аналізу багато-

вимірних часових рядів може бути їх адаптивне прогнозування, коли помітний розрив між прогнозом і вихідними даними слугуватиме індикатором цих змін.

Спочатку розглянемо метод багатовимірної лінійної екстраполяції стосовно до задачі однокрокового прогнозування  $n$ -вимірної нелінійної нестационарної часової послідовності стосовно описаної нами моделі. Припускаючи, як і раніше,  $\bigcup_{i=1}^N B(i)$ , отримуємо, що аналізуємий ряд ( $i$ -я його компонента) можливо описати нелінійною залежністю (NARX-модель) вигляду

$$\begin{aligned} \bar{B}_i(k) &= f_i(B_i(k-1), \dots, B_i(k-n_{A,i}), B_{\mathcal{C}}(k-1), \dots, B_{\mathcal{C}}(k-n_{\mathcal{C}}), \\ &B_{\mathcal{C}}(k-1), \dots, B_{\mathcal{C}}(k-n_{\mathcal{C}}), \dots, B_{\mathcal{C}}(k-l), \dots, B_{\mathcal{C}}(k-n_{\mathcal{C}})) = \\ &= f_i(R_{i1}(k), \dots, R_{i,n_{A,i}}(k), \dots, R_{i,n_{A,i}+n_{\mathcal{C}}q}(k)), \end{aligned}$$

де  $f_i(0)$  – апріорі невідома нелінійна залежність, яка підлягає відновленню на базі спостережень;  $\bar{B}_i(k)$  – оцінка (прогноз) контролюємої послідовності  $B_i(k)$  по даним, що доступні до  $(k-1)$ -му моменту часу;  $i=1, \dots, n$ ;  $n_{A,i}$  – глибина передісторії послідовності, що враховується;  $B_{\mathcal{C}}(k-l)$  –  $p$ -та компонента багатовимірного екзогенного сигналу, що впливає на  $B_i(k)$ ;  $l=1, \dots, n_{\mathcal{C}}$ ;  $p=1, \dots, q$ . Для задачі аналізу відео модель представимо в векторно-матричному вигляді

$$\bar{B}(k) = F(B(k-1), \dots, B(k-n_{A,i}), B_{\mathcal{C}}(k-1), \dots, B_{\mathcal{C}}(k-n_{\mathcal{C}})) = F(R(k)).$$

Метод багатовимірної лінійної екстраполяції з евклідовою метрикою може бути описаний у вигляді послідовності кроків:

1. Формування передісторії прогнозованого процесу у вигляді матриць  $R_{\mathcal{C}} = (R(1), R(2), \dots, R(N))$ ,  $B = (B(1), B(2), \dots, B(N))$ .

2. Знаходження вектору вагових коефіцієнтів  $l = (l_1, l_2, \dots, l_N)^T$ , що забезпечують мінімум нормі  $\|R(N+1) - \hat{a}_{k-1}^N l(k)R(k)\|^2 = \|R_{\mathcal{C}}(N+1) - R_{\mathcal{C}}\|^2$ .

3. Формування оптимального прогнозу у вигляді лінійної комбінації  $\bar{B}(N+1) = \hat{a}_{k=1}^N l_k B(k) = \bar{B}l$ .

Мінімізація цього виразу може бути проведена за допомогою стандартного методу найменших квадратів, в результаті чого отримуємо рішення  $l = (R_{\mathcal{C}}^T R_{\mathcal{C}})^{-1} R_{\mathcal{C}}^T R(N+1)$ , яке існує тільки тоді, коли  $R_{\mathcal{C}}^T R_{\mathcal{C}}$  не вироджена. Оскільки при малих  $N$  ( $N < n n_A + q n_{\mathcal{C}}$ ) це не так, замість операції обернення пропонується використовувати операцію псевдообернення, завдяки чому отримуємо  $l = R_{\mathcal{C}} R(N+1)$ . Звідки маємо, що задача фактично зведена до знаходження ортогональної проекції вектору  $R(N+1)$  на лінійну оболонку сформовану векторами передісторії  $R_{\mathcal{C}}$ . Якщо дані на обробку надходять поточно у реальному часі, то співвідношення слід представити у вигляді  $R_{\mathcal{C}} = (R(1), R(2), \dots, R(k))$ ,  $\bar{B}_k = (B(1), \dots, B(k))$ ,  $\|R(k+1) - \hat{a}_{l=1}^k l_l R(l)\|^2 = \|R(k+1) - R_{\mathcal{C}} L_k\|^2$ ,  $L_k = (l_1, \dots, l_k)^T$ ,

$$\bar{B}(k+1) = \mathring{a}_{l=1}^k |l B(l) = \bar{B}_k \mathbf{L}_k, \mathbf{L}_k = R_{k,c}^{\dagger} R(k+1).$$

Для обробки нестационарних часових рядів, характеристики яких непередбачувано змінюються в часі, замість обробки всієї наявної вибірки, доцільно вирішувати завдання на «ковзному вікні», що складається з  $c$  останніх спостережень. При цьому вищевказані співвідношення модифікуються до вигляду

$$\begin{aligned} R_{k,c} &= (R(k-c+1), R(k-c+2), \dots, R(k)), \\ \bar{B}_{k,c} &= (B(k-c+1), B(k-c+2), \dots, B(k)), \\ \|R(k+1) - \mathring{a}_{l=k-c+1}^k |l R(l)\|^2 &= \|R(k+1) - R_{k,c} \mathbf{L}_{k,c}\|^2, \\ \mathbf{L}_{k,c} &= (|_{k-c+1}, \dots, |_k)^T, \\ \bar{B}(k+1) &= \mathring{a}_{l=k-c+1}^k |l B(l) = \bar{B}_{k,c} \mathbf{L}_{k,c}, \\ \mathbf{L}_{k,c} &= R_{k,c}^{\dagger} R(k+1). \end{aligned}$$

Зрозуміло, що застосування для прогнозування цих співвідношень суттєво спрощує використання методу, однак залишається відкритим питання обґрунтування вибору величини вікна  $c$ , яке, як правило, задається з суто суб'єктивних міркувань, що знижує ефективність підходу в цілому. Пропонований нижче метод багатовимірної екстраполяції заснований на використанні деякої близькості (відстані) останнього вектора передісторії  $R(N+1)$  до всіх попередніх даних  $R(1), \dots, R(N)$  і формуванні прогнозу  $\bar{B}(N+1)$  за допомогою цієї ж функції:

1. Розрахунок відстані між вектором  $R(N+1)$  і всіма попередніми  $R(k)$  на основі деякої функції близькості  $d(N+1, k)$  – в найпростішому випадку евклідової метрики  $d(N+1, k) = \|B_{k,c}^{\dagger} R(N+1) - B_{k,c}^{\dagger} R(k)\|$  "  $k$ .

2. Впорядкування (ранжування) цих відстаней у порядку зростання

$$d^1(N+1, k_1) < d^2(N+1, k_2) < \dots < d^N(N+1, k_N).$$

3. Відбір перших  $c$  векторів, для яких виконується умова  $d^c(N+1, k_c) \leq \epsilon$ .

4. Формування набору вагових коефіцієнтів  $|l = \frac{(d^l)^{-1}}{\mathring{a}_{l=1}^c (d^l)^{-1}}, 1 \leq l \leq c$ .

5. Розрахунок прогнозу  $\bar{B}(N+1) = \mathring{a}_{l=1}^c |l R(l)$ .

Описаний метод не вимагає великих обсягів апріорної інформації (мала навчальна вибірка), немає необхідності вирішення додаткових завдань оптимізації або псевдообернення, є вкрай простим з точки зору обчислювальної реалізації. Однак для успішного використання всіх традиційних методів прогнозування, вихідна вибірка спостережень – часовий ряд – повинна бути досить репрезентативною. При цьому, чим більше параметрів містить прогноуюча модель, тим більше за обсягом повинна бути вихідна інформація. Разом з тим в реальних задачах досить часто виникає ситуація, коли ця вибірка або мала за обсягом, або прогнозований процес є нестационарним – містить як нерегулярні тренди, так і раптові стрибки, тому його передісторія не може бути використана

для знаходження параметрів моделі.

**П'ятий розділ** присвячений розробці методів аналізу, пошуку та формуванню ключової інформації з отриманих в результаті проведення сегментації однорідних фрагментів відеопотоків.

У першому розділі були розглянуті різні підходи до виділення ключової інформації у відео. При цьому було зроблено висновок про те, що на поточний момент не існує якогось універсального підходу. Що пов'язано в першу чергу зі специфікою як самої відеоінформації, так і способам її отримання. У зв'язку з цим пропонується проводити порівняння діаграм Вороного за допомогою спеціалізованих метрик в рамках досліджуваного простору ознак.

Нехай  $D=[a,b] \times [c,d]$  ( $a,b,c,d = const$ ) – поле зору, а  $\{p_1, p_2, \dots, p_n\}$  – множина опорних точок. Діаграма Вороного є розбиттям поля зору  $D$  на опуклі багатокутники  $V = \{v(p_1) \mid D, v(p_2) \mid D, \dots, v(p_n) \mid D\}$  таким чином, що для кожної області виконується  $v(p_i) = \{z \in \mathbb{R}^2 : d(z, p_i) \leq d(z, p_j) \forall j \neq i\}$ , де  $d(0,0)$  – евклідова метрика. Іншими словами, область Вороного  $v(p_i)$ , пов'язана з опорною точкою  $p_i$  та являє собою набір точок  $Z$ , відстань кожної з яких до пов'язаної опорної точки менша або дорівнює відстані до будь-якої іншої опорної точки.

Виходячи з визначення, діаграма Вороного, що лежить в основі запропонованої процедури пошуку ключових кадрів, будується на опорних точках. Для визначення кількості і розміщення опорних точок у полі зору відеокадрів пропонується використовувати метод Харріса. Крім того, якість діаграм Вороного (і як наслідок знайдених опорних точок) передбачається покращувати шляхом перекомпошування методом нечітких  $J$ -середніх.

Реалізація  $FJM$  методу проводиться в два етапи: знаходження локальних оптимумів за допомогою стандартного  $FCM$  і знаходження більш «глибоких» мінімумів за допомогою  $FJM$ -евристики. На першому етапі реалізується послідовність кроків:

1. Завдання початкового досить довільного розбиття  $P_p = \{Cl_1, \dots, Cl_p\}$  з центроїдами  $C(1), C(2), \dots, C(p)$ ; фаззифікатора  $b$  і порогового значення  $\epsilon > 0$ , що визначає умови зупинки алгоритму.

2. Розрахунок рівнів належності  $\mu(x(k), C(l))$  з центроїдами, отриманими на попередньому етапі.

3. Перерахунок центроїдів  $C(1), C(2), \dots, C(p)$ , з рівнями належності, отриманими на попередньому етапі.

4. Оцінка сферичної норми різниці між раніше отриманими центроїдами і центроїдами, розрахованими на третьому кроці.

5. Перевірка умов зупинки. Якщо отримана норма є меншою  $\epsilon$  – закінчення роботи алгоритму; якщо ж отримана норма більше  $\epsilon$  – повернення до кроку 2 з центроїдами, отриманими на кроці 3.

Ітерації тривають до виконання умови зупинки, а отримане рішення є координатами локального оптимуму задачі нелінійного програмування.



Другий етап – це фаза стрибків, коли з отриманого локального мінімуму відбуваються випадкові рухи в його околі з метою відшукати більш глибокий екстремум.

Було показано, що спільне завдання оптимізації може бути приведено до задачі безумовної оптимізації спеціального виду цільової функції, яка в матричному випадку може бути записана у вигляді

$$E = E(x(k), C(l)) = \mathring{a} \sum_{k=1}^{\mathring{N}} \left( \mathring{a} \sum_{l=1}^p (Sp(x(k) - C(l))(x(k) - C(l))^T)^{1-b} \right)^{1-b} \quad (9)$$

для довільних значень фаззифікатора  $b$ .

Далі з будь-якого з отриманих центроїдів  $C(1), C(2), \dots, C(p)$  відбуваються стрибки, що складаються з того, що обраний центроїд замінюється будь-яким спостереженням  $x(r)$ , після чого розраховується значення цільової функції (9) у формі

$$E = E(x(k), x(r)) = \mathring{a} \sum_{k=1}^{\mathring{N}} \left( \mathring{a} \sum_{l=1}^p (Sp(x(k) - x(r))(x(k) - x(r))^T)^{1-b} \right)^{1-b}. \quad (10)$$

Якщо виявиться, що для деякого значення критерій (10) виявиться меншим ніж (9), приймається рішення про те, що знайдений новий покращений центроїд кластеру  $Cl_l$ , після чого проводиться перерахунок всіх рівнів. Такі стрибки відбуваються в околі кожного з центроїдів, отриманих на першому етапі. Якщо виявиться, що скачки в околах всіх центроїдів не привели до поліпшення значення цільової функції (10), то або приймається рішення про закінчення процесу оптимізації, або реалізуються стрибки в околі збільшеного радіусу. Такий процес може тривати до вичерпання всіх незайнятих точок. Знайдений останнім локальний екстремум припускається глобальним.

Загальна ж процедура порівняння діаграм Вороного складається з наступних етапів:

1. Визначення однотипності відеокадрів. З урахуванням розкиду значень текстури (дисперсії, ентропії) у всій відеопослідовності виконаємо установку порогового значення згідно з правилом:

$$Threshold = \begin{cases} \frac{1}{4}, & \frac{1}{K-1} \mathring{a} \sum_{k=1}^K \left( E(B_k(z)) - \frac{1}{K} \mathring{a} \sum_{k=1}^K E(B_k(z)) \right)^2 \otimes \neq, \\ \frac{1}{2}, & \frac{1}{K-1} \mathring{a} \sum_{k=1}^K \left( E(B_k(z)) - \frac{1}{K} \mathring{a} \sum_{k=1}^K E(B_k(z)) \right)^2 \otimes \frac{1}{K} \mathring{a} \sum_{k=1}^K E(B_k(z)), \\ \frac{3}{4}, & \frac{1}{K-1} \mathring{a} \sum_{k=1}^K \left( E(B_k(z)) - \frac{1}{K} \mathring{a} \sum_{k=1}^K E(B_k(z)) \right)^2 \otimes 0, \end{cases}$$

де  $K$  – загальна кількість кадрів сегменту відеопослідовності;  $E(B_k(z))$  – значення ентропії для  $k$ -го відеокадру сегменту.

Через те, що близькість змісту є відносною мірою подібності відеокадрів,

доводиться використовувати граничне значення для завдання прийняттого ступеня їх схожості. В цьому випадку граничне значення встановлюється з урахуванням змісту відео.

2. Беремо перший ( $B_k(z)$ ) та другий ( $B_{k+1}(z)$ ) кадр для порівняння. Встановлюємо  $k = 1$ .

3. Покадрове порівняння діаграм Вороного за метрикою

$$\bar{r}(B_k(z), B_{k+1}(z)) = \sum_{i=1}^l a_i \rho_{\phi} \quad \sum_{i=1}^l a_i = 1, \quad a_i \geq 0,$$

де  $\rho_{\phi}$  – деяка локальна метрика на розбиттях Вороного,  $a_i$  – важливість ознак, що використовуються в порівнянні.

Розраховуємо  $\bar{r}(B_k(z), B_{k+1}(z))$  для двох кадрів сегменту. Якщо значення  $\bar{r}(B_k(z), B_{k+1}(z))$ , виявилось меншим ніж встановлене порогове значення, в такому випадку обидва кадри  $B_k(z)$  та  $B_{k+1}(z)$  вважаються ключовими та вилучаються як  $B_r^*(z)$  та  $B_{r+1}^*(z)$  і після того переходимо до шагу 4, інакше тільки  $B_k(z)$  вилучається як  $B_r^*(z)$  та йдемо до шагу 5.

4. Перезначаємо значення  $B_k(z) = B_{k+1}(z)$ ,  $B_{k+1}(z) = B_{k+2}(z)$  та переходимо до шагу 6.

5. Залишаємо  $B_k(z) = B_k(z)$  та встановлюємо  $B_{k+1}(z) = B_{k+2}(z)$ .

6. Переходимо до шагу 3, якщо  $k + 2 \notin K$ .

Слід зазначити, що крім розмірів відео, які впливають на швидкість пошуку ключових кадрів, серед всіх етапів в рамках запропонованого методу найбільш ресурсномістким є визначення опорних точок методом Харріса за рахунок необхідності обчислення похідних. Отже, слід використовувати інші методи пошуку опорних точок (наприклад, методи SURF, SIFT або FAST) для наближення процедури пошуку ключових кадрів до реального часу. Якщо говорити про якість пошуку, перевпорядкування опорних точок за допомогою кластеризації робить інваріантним результуюче розміщення опорних точок щодо обраного методу їх вихідного розміщення, хоча і змінює число ітерацій.

**Шостий розділ** присвячений розробці методів і моделей компресії, які дозволяють скоротити обсяги даних, що є актуальним з точки зору аналізу відеопотоків тому, що розв'язання задач інформаційного пошуку, описане в попередніх розділах, передбачає зберігання не лише самих відео, але і додаткової інформації про границі сегментів і ключові кадри кожного з сегментів.

Входова інформація для методу скорочення даних задана у вигляді масиву даних, утвореного  $N$   $n$ -вимірними векторами  $x(1), x(2), \dots, x(k), \dots, x(N)$ , де  $k = 1, 2, \dots, N$  – номери спостережень у входовому масиві, а результатом рішення є набір домінантних власних значень  $\lambda_1 > \dots > \lambda_j > \dots > \lambda_m$  та відповідних їм власних значень векторів  $w_1, w_2, \dots, w_j, \dots, w_m$ ,  $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})^T$  кореляційної ( $n \times n$ ) матриці входових даних

$$\begin{cases} R(N) = \frac{1}{N} \mathop{\text{a}}\limits^N_{k=1} (x(k) - \bar{x}(N))(x(k) - \bar{x}(N))^T, \\ \bar{x}(N) = \frac{1}{N} \mathop{\text{a}}\limits^N_{k=1} x(k). \end{cases} \quad (11)$$

Власне компресія (зниження розмірності) вихідного простору реалізується за допомогою відображення  $y(k) = Wx(k)$  де  $y(k) = (y_1(k), \dots, y_m(k))^T$ ,  $W = (w_1, w_2, \dots, w_{m-1}, w_m)^T$  –  $(m \times n)$  – матриця, утворена доміантними власними векторами кореляційної матриці  $R(N)$ . У випадку, якщо обробка даних проводиться в реальному часі, для обчислення цієї матриці замість (11) можуть бути використані рекурентні співвідношення

$$\begin{cases} H(k) = \frac{k-1}{k} H(k-1) + \frac{1}{k} (x(k) - \bar{x}(k))(x(k) - \bar{x}(k))^T, \\ \bar{x}(k) = \frac{k-1}{k} \bar{x}(k-1) + \frac{1}{k} x(k) \end{cases}$$

для стаціонарного випадку та

$$\begin{cases} H(k) = \mathbf{a}H(k-1) + (1-\mathbf{a})(x(k) - \bar{x}(k))(x(k) - \bar{x}(k))^T, \\ \bar{x}(k) = \mathbf{a}\bar{x}(k-1) + (1-\mathbf{a})x(k) \end{cases}$$

для нестационарного. Тут  $0 < \mathbf{a} < 1$  – емпірично обраний параметр забування застарілої інформації. Оптимальний алгоритм самонавчання на основі зваженого інформаційного критерію може бути записаний у вигляді

$$W(k) = W(k-1) + \frac{((x(k) - W^T(k-1)y(k))^T G(k)y(k))}{\|G(k)y(k)\|^2} G(k),$$

де  $G(k) = -((A^{-1}W(k-1)H(k)W^T(k-1)A)^{-1}W(k-1)H(k) - W(k-1))$ .

Запропонована архітектура нейронної мережі, що розв'язує розглянуту проблему, містить два шари, утворених  $m$  (в першому прихованому шарі) і  $n$  (в вихідному шарі) адаптивними лінійними асоціаторами. У першому прихованому шарі, синаптичні ваги якого утворюють  $(m \times n)$  – матрицю  $W = \{w_{ji}\}$ , проводиться стиснення інформації, при цьому на його виході обчислюються значення головних компонент  $y_1, y_2, \dots, y_m$ . Вихідний шар призначений для відновлення вхідного сигналу за допомогою  $(n \times m)$  матриці синаптичних ваг  $W^T = \{w_{ij}\}$  розраховуються значення  $\bar{x}_1(k), \bar{x}_2(k), \dots, \bar{x}_n(k)$ , що є оцінками компонент вхідного сигналу:

$$\bar{x}(k) = W^T(k-1)y(k) = W^T(k-1)W(k-1)x(k).$$

Зрозуміло, що таке відновлення можливо при  $m = n$ , однак при  $m < n$  забезпечується відновлення якомога точніше в сенсі критерію

$$E(\mathbf{h}) = \|\mathcal{E}(k)\|^2 = \|x(k) - W^T(k)y(k)\|^2 = \|x(k) - W^T(k)W(k)x(k)\|^2,$$

що вигідно відрізняє конструкції, що розглядаються, від традиційних підходів.

Однак, цей підхід не позбавлений недоліків, в зв'язку з чим доцільним оптимізувати процес навчання чотиришарової гетерогенної автоасоціативної нейронної мережі з прямою передачею інформації, призначеної для зниження розмірності оброблюваних векторів-образів з урахуванням впливу перешкод і можливим порушенням опуклості цільових функцій прихованих шарів.

Архітектура введеної автоасоціативної нейронної мережі, призначеної для зниження розмірності простору вихідних образів містить чотири послідовно з'єднаних шари нейронів. На вхід (рецепторний нульовий шар) подається послідовність векторів-образів  $x(1), x(2), \dots, x(k), \dots, x(N)$ , аналогічно до попереднього  $x(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T \hat{1} \mid N$  пронормованих на гіперкуб. Перший прихований шар містить  $h^3 n$  нейронів – елементарних персептронів Розенблатта з сигмоїдальними активаційними функціями. У разі відсутності сигналу зміщення, перший прихований шар містить  $hn$  настроюваних синаптичних ваг і описується співвідношеннями вигляду:

$$O_j^{[1]}(k) = y_j^{[1]}(u_j^{[1]}(k)) = y_j^{[1]}(\sum_{i=1}^n w_{ji}^{[1]} x_i(k)), j = 1, 2, \dots, h, \quad O^{[1]}(k) = Y^{[1]}(W^{[1]}x(k)),$$

де  $O_j^{[1]}(k)$  – вихідний сигнал  $j$ -го нейрону першого прихованого шару  $x(k)$ ,  $y_j^{[1]}(0)$  – активаційна функція,  $u_j^{[1]}(k)$  – сигнал внутрішньої активації,  $w_{ji}^{[1]}$  – синаптична вага  $i$ -го входу  $j$ -го нейрона,  $O^{[1]}(k) = (O_1^{[1]}(k), \dots, O_h^{[1]}(k))^T$ ,  $Y^{[1]} = \text{diag}(Y_1^{[1]}, \dots, Y_j^{[1]}, \dots, Y_h^{[1]})$  –  $(h \times h)$  діагональна матриця активаційних функцій,  $W^{[1]} = \{w_{ji}^{[1]}\}$  –  $(h \times n)$  матриця синаптичних ваг.

Другий прихований шар містить  $m < n$  нейронів – адаптивних лінійних асоціаторів, а його вихідний сигнал  $y = (y_1, \dots, y_m)^T$  є виходом нейронної мережі в цілому і являє собою «стислий» вхідний образ  $x$ . Другий прихований шар описується співвідношеннями

$$y_j(k) = \sum_{i=1}^h w_{ji}^{[2]} o_i^{[1]}(k), j = 1, 2, \dots, m \quad y(k) = W^{[2]}O^{[1]}(k),$$

де  $W^{[2]} = \{w_{ji}^{[2]}\}$  –  $(m \times h)$  – матриця синаптичних ваг.

Третій прихований шар подібний до першого і також містить  $h$  елементарних персептронів Розенблатта, описуваних співвідношеннями

$$O_j^{[3]}(k) = y_j^{[3]}(u_j^{[3]}(k)) = y_j^{[3]}(\mathring{a} \sum_{i=1}^m w_{ji}^{[3]} y_i(k)), j = 1, 2, \dots, h, \quad O^{[3]}(k) = Y^{[3]}(W^{[3]}y(k)),$$

де  $W^{[3]} - (h \times m)$  – матриця синаптичних ваг.

Четвертий вихідний шар подібний другого прихованого і містить  $n$  нейронів – адаптивних лінійних асоціаторів.

$$\mathring{x}_j(k) = \mathring{a} \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k), j = 1, 2, \dots, n, \quad \mathring{x}(k) = W^{[4]}O^{[3]}(k),$$

де  $\mathring{x}(k) - (n \times 1)$  – вектор, що є оцінкою вхідного сигналу  $x(k)$ , що оновлюється після стиснення,  $W^{[4]} - (n \times h)$  – матриця синаптичних ваг.

Процес навчання нейронної мережі зводиться до послідовного використання процедур

$$\begin{aligned} \dot{\downarrow} Dw_j^{[4]} &= (p^{[4]}(k))^{-1} e_j(k) O^{[3]}(k), & \dot{\downarrow} Dw_j^{[3]} &= (p^{[3]}(k))^{-1} (h^{[3]}(k) d_j^{[3]}(k) y(k) + W_3, \\ \dot{\downarrow} p^{[4]}(k) &= ap^{[4]}(k-1) + \|O^{[3]}(k)\|^2, & \dot{\downarrow} p^{[3]}(k) &= ap^{[3]}(k-1) + \|y(k)\|^2, \end{aligned}$$

$$\begin{aligned} \dot{\downarrow} Dw_j^{[2]} &= (p^{[2]}(k))^{-1} (h^{[2]}(k) d_j^{[2]}(k) O^{[1]}(k) + W_2, & \dot{\downarrow} Dw_j^{[1]} &= (p^{[1]}(k))^{-1} (h^{[1]}(k) d_j^{[1]}(k) x(k) + W_1, \\ \dot{\downarrow} p^{[2]}(k) &= ap^{[2]}(k-1) + \|O^{[1]}(k)\|^2, & \dot{\downarrow} p^{[1]}(k) &= ap^{[1]}(k-1) + \|x(k)\|^2. \end{aligned}$$

для кожного висунутого вектора-образу  $x(k)$ . Для досить великої навчальної вибірки процес налаштування може бути реалізований у реальному часі. Якщо ж обсяг вхідних даних обмежений, навчання може бути організовано по епохах шляхом багаторазового «прогону» даних через нейронну мережу.

**Сьомий розділ** присвячений експериментальному аналізу методів сегментації відеоданих, заснованих на уявленні початкових даних у вигляді багатовимірного часового ряду з подальшим перетворенням за допомогою методів, запропонованих в розділах 3-5 в одновимірні часові ряди.

По-перше, було оцінено формування векторів характеристик кадрів і їх вплив на результати сегментації одного і того ж відео. На рисунку 1 наведено 3 різних варіанти вибору вектора параметрів. При цьому вони були пронормовані з тим, щоб значення можна було порівняти. Як видно з результатів, перший і другий варіанти векторів характеристик дають неоднозначні результати, за якими складно визначити наявність розладнань в послідовності відеоданих. Третій же варіант дає результати, з яких вже можливо провести виділення окремих сегментів відеопослідовностей за результатами експерименту. Слід зазначити, що в першому і другому випадку, на відміну від третього варіанту, як однієї з невраховуваних при розгляді ознак була обрана площа сегментів просторової сегментації кадрів, виходячи з чого можна стверджувати, що ця характеристика є однією з необхідних для обчислень при побудові характеристичного вектора кадрів відеопослідовності. Загалом вектор ознак містить 12 характе-

ристик, чого досить (як виявилось в результаті емпіричних досліджень) для вдалого виділення границь сегментів відеопотоків.

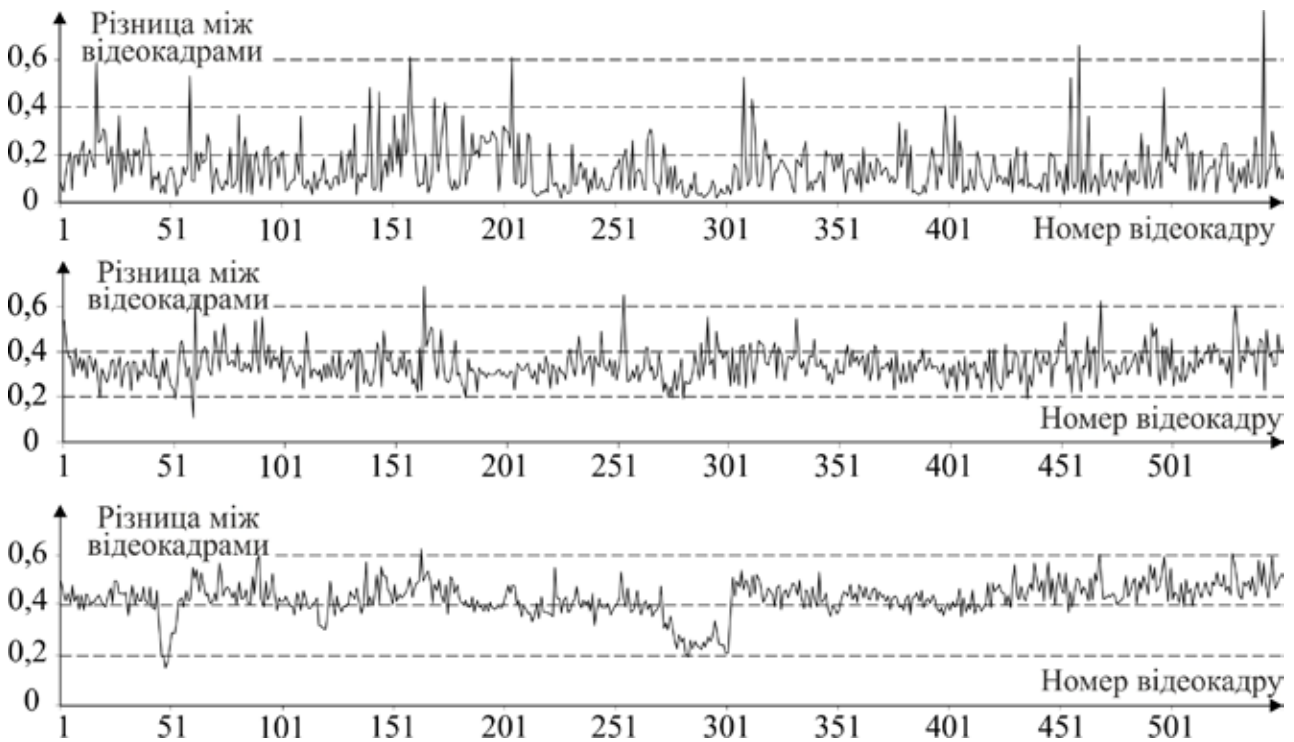


Рисунок 1 – Залежність результатів аналізу від вибору вектора характеристик (зверху – функція площі, радіуси вписаних та описаних навколо сегментів кіл, ексцентриситет; посередині – радіуси вписаних та описаних навколо сегменту кіл, кут нахилу дотичних; знизу – площа та периметр сегментів, функція площі, центроїд, функція відстані центроїду, площа опуклої оболонки)

По друге, розглянуто результати сегментації тестових відеопослідовностей. Зрозуміло, при візуальному розгляді кожних початкових відеоданих експерти легко визначають переходи між сценами. На рисунку 2 наведений приклад переходу від одного сегменту до іншого в тестових відеопослідовностях. Проте автоматичне виділення границь за допомогою реалізації методів з розділу 4 не завжди є простою задачею. Зокрема це пов'язано з тим, що при цьому іноді виникають ситуації з неявними переходами, які можуть призводити до проблем при детектуванні змін параметрів часового ряду. Така ситуація може виникати при наявності різних ефектів в відеоданих, які покликані згладити візуальні переходи для глядача, але є швидше негативним фактором при математичному аналізі подібних послідовностей.

Слід відзначити той факт, що в залежності від складності фону ми можемо отримувати більш явні показники переходів між сегментами. Так для зйомок «Авіашоу», де в якості фону виступає небо, перехід від одного сегмента відео до іншого виділяється чітко як показано на рис. 3.

Також необхідно було оцінити релевантність інформаційного пошуку,

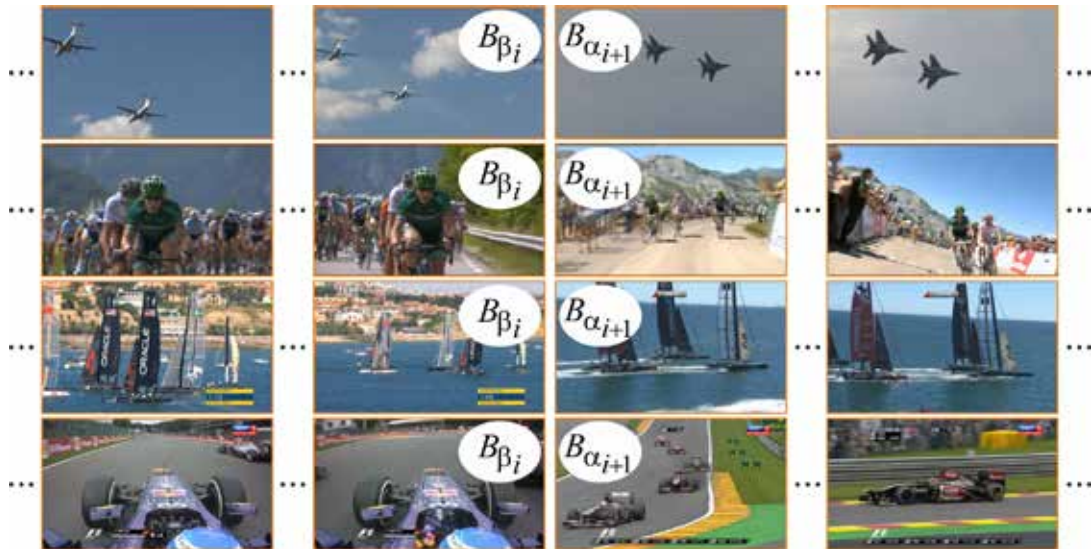


Рисунок 2 – Границі між сегментами в початкових відеопослідовностях

для чого був проведений експеримент по побудові коефіцієнту Дайса для множини відеопослідовностей взятих з циклу документальних фільмів «Знищені за мить». Цей коефіцієнт базується на класичних поняттях оцінки систем пошуку, а саме точності (*precision*) та повноти (*recall*) і обчислюється за формулою

$$F = \frac{2}{1/\textit{precision} + 1/\textit{recall}}$$

Всі відео були розбиті на декілька класів, а результат інформаційного пошуку наведений на рис. 4. Приведена лінійна діаграма показує середні значення коефіцієнтів Дайса, отримані після проведення сегментації при виділенні ключових кадрів з різних за жанром відеопослідовностей. Для кадрів, відібраних в якості ключових за допомогою діаграм Вороного з тестової відеоколекції,

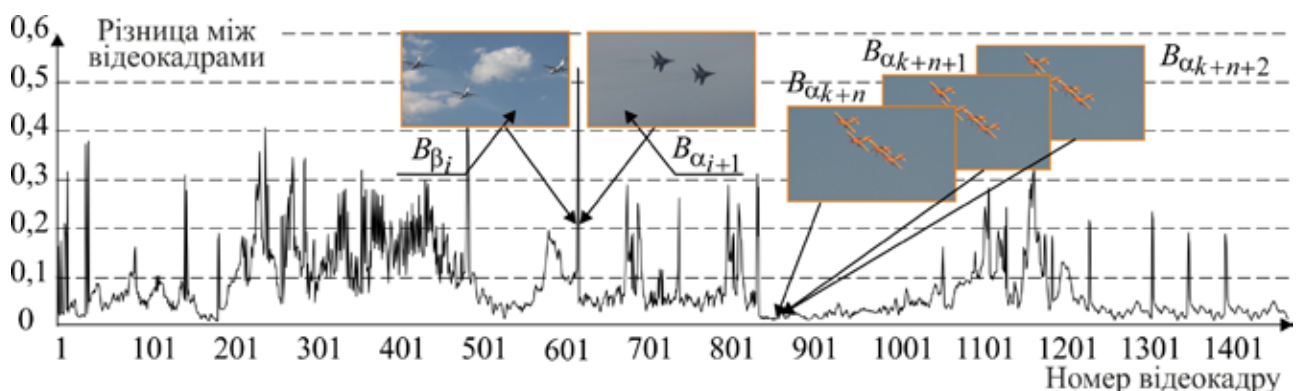


Рисунок 3 – Сегментація відеопослідовності «Авіашоу»

коефіцієнт Дайса в середньому склав 0,89, що свідчить про доволі непогані результати з точки зору релевантності технології інформаційного пошуку в відеопослідовностях.

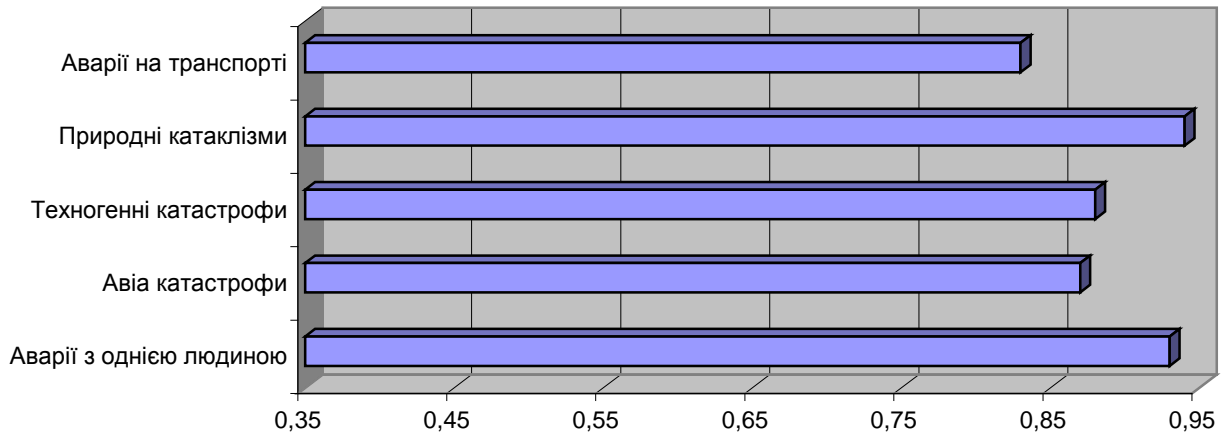


Рисунок 4 – Значення коефіцієнту Дайса для інформаційного пошуку

Наведено результати впровадження розроблених методів та моделей темпоральної обробки відеопотоків.

У додатку наведено акти впровадження теоретичних та практичних результатів дисертаційних досліджень.

## ВИСНОВКИ

Розвиток моделей та методів темпоральної обробки, аналізу та інтерпретації поганоструктурованих потоків інформації (насамперед, відеопотоків) зумовлює розвиток інтелектуальних інформаційних технологій мульти і гіпермедіа. Релевантність інформаційних ресурсів, розумний компроміс між накопиченням даних і потребами їх раціонального використання залишаються однією з ключових проблем. Тенденції перманентного розширення номенклатури вирішуваних задач, інтенсифікації обробки відеопотоків, глобальна акумуляція даних в поганоструктурованих великих відеоархівах вимагають розробки перспективних методів управління відеоданими, зводяться до вирішення складних завдань аналізу, багатоаспектного автоматичного індексування, анотування і реферування, нелінійного перегляду і пошуку з різними формами запитів. При цьому усунення семантичного конфлікту між низькорівневими моделями відео і високорівневими цілями їх предметно-орієнтованого трактування залишається ключовою проблемою. Специфіка вирішення цих завдань в СВІР визначається необхідністю *on line* обробки і аналізу дуже великих обсягів інформації в умовах суттєвої невизначеності при головуванні не просторових (у полі зору) кореляційних зв'язків, а складних відношень між відеокадри в часі в умовах обмежень на час прийняття рішень та обчислювальних ресурсів.

В результаті проведення теоретико-експериментальних дисертаційних досліджень розв'язано науково-технічну проблему темпоральної обробки відеопотоків, що забезпечує структурування даних для інформаційного пошуку з запитом 'ad exemplum' в великих відеоколекціях в просторах зображень або



ознак та отримані такі основні результати:

1. Встановлено, що для великих бібліотек відео перспективним напрямком досліджень є розробка моделей і методів темпоральної обробки відеопотоків, які забезпечують організацію семантичного пошуку із запитом «за зразком». Зміст відеоданих індукується подвійною сегментацією: в часі (on line пошук змін подій або сюжетів) і просторі (пошук областей, що корелюють з об'єктами сцен).

2. Запропоновано, обґрунтовано і досліджено моделі структуризації відеопослідовностей для задач інформаційного пошуку з запитом 'ad exemplum' на основі побудови областей елімінації для довільних конфігурацій об'єктів в довільному метричному просторі.

3. Розроблено фрагментні методи обробки візуальної інформації, які рівносно можуть використовуватися для темпоральної і просторової сегментації відеоданих з метою підвищення валідності семантичного пошуку в великих бібліотеках відеоданих.

4. Введені методи валідного порівняння скінченновимірних розбиттів відеокадрів, проаналізовані множини дескрипторів елементів розбиття і зображень, які дозволяють підвищити адекватність представлення відеорядів в ознакових просторах.

5. Синтезовано комплекс моделей і методів on line сегментації відеопослідовностей і багатовимірних часових рядів, індукованих відеопотоками, шляхом виявлення зміни їх властивостей, що забезпечує валідну структуризацію відео.

6. Запропоновано методи формування і пошуку ключових кадрів, що представляють сегменти відеоданих в задачах пошуку відео на базі механізмів пошуку зображень у великих бібліотеках статичної візуальної інформації.

7. Розроблено нейромережеві методи компресії відеокадрів і відео для темпоральної обробки відеопотоків в задачах багатоаспектної індексації, досліджені питання ефективності компресії зображень.

8. Шляхом експериментальних досліджень встановлена валідність розроблених моделей і методів темпоральної обробки відеопослідовностей, визначені умови їх раціонального використання в задачах інтелектуального інформаційного пошуку.

9. На основі запропонованих методів і моделей темпоральної обробки відеопослідовностей розроблена низка прикладних і дослідницьких програмних комплексів, які використовуються при пошуку в великих відеоархівах із запитом 'ad exemplum'.

## **СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

1. Аналіз та обробка даних засобами обчислювального інтелекту / Є.В. Бодяньський, Д.Д. Пелешко, О.А. Винокурова, С.В. Машталір, Ю.С. Іванов. – Львів: Видавництво Львівської політехніки, 2016. – 236 с.

2. Егоров, А. Сравнительный анализ методов морфологической нормализации

зации / Егоров А., Машталир С. // Радиоэлектроника и информатика. – 2005. – №4. – С. 90-95. (Входит до міжнародних наукометричних баз BASE, World Cat.)

3. Метрическое сравнение результатов сегментации изображений / Е.А. Егорова, Д.К. Киношенко, С.В. Машталир, Д.В. Шляхов // Радиоэлектроника и информатика. – 2006. – №2. – С. 55-62. (Входит до міжнародних наукометричних баз BASE, World Cat.)

4. Машталир, С.В. Нейросетевой подход к поиску симметричных подпространств при фрагментной обработке изображений в реальном времени / С.В. Машталир, Е.С. Сакало // Системні технології. – Днепропетровск: «Системні технології». – 2007. – №6 (53). – С. 153-161.

5. Луганский, А.М. Анализ классификации текстов с использованием весовых коэффициентов / А.М. Луганский, С.В. Машталир // Бионика интеллекта. – 2007. – № 2 (67). – С. 73–77.

6. Машталир, С. Метод нахождения характерных признаков формы ладони по скелетону изображения / С. Машталир, Н. Свередюк // Бионика интеллекта. – 2008. – № 1 (68). – С. 101-105.

7. Машталир, С.В. Адаптивное нейросетевое сжатие сигналов большой размерности на основе взвешенного информационного критерия / С.В. Машталир, Е.С. Сакало // Прикладная радиоэлектроника. – 2008. – Том 3. – №2. – С. 175-178.

8. Ведмедь, А.Г. Матричный алгоритм самообучения карты Кохонена в задачах обработки изображений / А.Г. Ведмедь, С.В. Машталир, Е.С. Сакало // Системи управління, навігації та зв'язку. – 2009. – Вип. 1 (9). – К.: ЦНДІНУ. – С. 188-192.

9. Робастные алгоритмы самообучения карты Кохонена в задачах обработки изображений / В.Я. Винарский, С.В. Машталир, С.Е. Сакало, К.С. Щербинин // Бионика интеллекта. – 2009. – № 1 (70). – С. 86-89.

10. Машталир, С.В. Поиск изменения сцен в видеоданных на базе анализа результатов сегментации / С.В. Машталир, С.В. Постульга, К.С. Щербинин // Бионика интеллекта. – 2010. – № 1 (72). – С. 65-69.

11. Машталир, С.В. Области достаточности и необходимости при решении задач альтернативного поиска / С.В. Машталир, К.С. Щербинин // Прикладная радиоэлектроника. – 2010. – Том 9. – №4. – С.580-583.

12. Ведмедь, А.Г. Восстановление изображений с использованием анализа главных и независимых компонент / А.Г. Ведмедь, С.В. Машталир, Е. С. Сакало // Системи обробки інформації. – Харків: ХУПС. – 2010. – Вип. 6. – С. 66-72. (Входит до міжнародних наукометричних баз BASE, Index Copernicus, SIS.)

13. Бодяньський, Є.В. Виявлення змін у потоці відеоданих на основі аналізу багатовимірних часових рядів / Є.В. Бодяньський, С.В. Машталір // Доповіді Національної академії наук України. – 2012. – №11. – С.30-33. (Входит до міжнародних наукометричних баз BIOSIS, ADAT (The Academic Database Assessment Tool.))

14. Гончаренко, М.О. Детектирование изменений сцены в потоке видео-

данных / М.О. Гончаренко, С.В. Машталир // Электротехнические и компьютерные системы. – 2012. – № 7 (83) – С. 143-147. (Входит до міжнародних наукометричних баз Ulrich's Periodicals, Index Copernicus.)

15. Власенко, Н.В. Компрессирование описания визуальных свойств путем фильтрации его компонент по критерию стабильности / Н.В. Власенко, А.С. Канунников, С.В. Машталир // Вісник НТУ ХПІ. Серія: системний аналіз, управління та інформаційні технології. – Харків: НТУ ХПІ. – 2013. – № 2 (976). – С.91-100. (Входит до міжнародних наукометричних баз Inspec, EBSCO, BASE, World Cat.)

16. Мантула, Е.В. Матричная прогнозирующая модель и ее обучение в задачах экологического мониторинга / Е.В. Мантула, С.В. Машталир // Электротехнические и компьютерные системы. – 2013. – № 10 (86). – С. 152-156. (Входит до міжнародних наукометричних баз Ulrich's Periodicals, Index Copernicus.)

17. Mashtalir, S. Stabilization of key frame descriptions with higher order Voronoi diagram / S. Mashtalir, O. Mikhnova // Бионика интеллекта. – 2013. – №1 (80). – Р. 68-72.

18. Mashtalir, S. Key frame recognition using Voronoi tessellations / S. Mashtalir, O. Mikhnova // Вісник Національного університету “Львівська політехніка”. Комп’ютерні науки та інформаційні технології. – Львів: Видавництво Львівської політехніки, 2013. – № 751. – С. 52-58. (Входит до міжнародних наукометричних баз INSPEC, OpenAIRE. ADAT, Inspec.)

19. Мантула, Е.В. Адаптивное прогнозирование временных рядов при неравноотстоящих наблюдениях / Е.В. Мантула, С.В. Машталир // Бионика интеллекта. – №2 (81), 2013. – С. 53-56.

20. Богучарский, С. Кластеризация коллекций изображений в больших базах данных на основе рекуррентной оптимизации / С. Богучарский, С. Машталир // Бионика интеллекта. – 2014. – №1 (82). – С. 43-46.

21. Богучарский, С.И. Анализ текстур в последовательности изображений на основе векторного квантования // С.И. Богучарский, С.В. Машталир // Радиотехника, информатика, управление. – 2014. – №2(31). – С. 94-99. (Входит до міжнародних наукометричних баз Ulrich's Periodicals Directory, BASE, ADAT, Index Copernicus, DOAJ, EBSCO, WorldCat та ін.)

22. Богучарский, С.И. Иерархическая агломеративная кластеризация изображений в больших базах данных / С.И. Богучарский, А.Г. Каграманян, С.В. Машталир // Системи обробки інформації. – Харків: ХУПС. – 2014. – № 8 (124). – С. 93-97. (Входит до міжнародних наукометричних баз BASE, Index Copernicus, SIS.)

23. Мантула, Е.В. Адаптивная полиномиальная нейросетевая прогнозирующая модель временных рядов и ее обучение / Е. В. Мантула, С.В. Машталир // Восточно-Европейский журнал передовых технологий. Математика и кибернетика – прикладные аспекты. – 2014. – № 2/4 (68). – С.16-20. (Входит до міжнародних наукометричних баз Ulrich's Periodicals Directory, DRIVER, BASE, Index Copernicus, WorldCat, DOAJ, EBSCO та ін.)

24. Богучарский, С.И. Модификация метода  $J$ -средних в задачах фрагме-

нтовой сегментации изображений / С.И. Богучарский, А.Г. Каграманян, С.В. Машталир // *Біоніка інтелекту*. – 2015. – №1 (84). – С. 19-22.

25. Богучарский, С.И. Модифицированный метод кластеризации Х-средних в задачах сегментации изображений / С.И. Богучарский, С.В. Машталир // *Электротехнические и компьютерные системы*. – 2015. – № 20 (96). – С. 106-110. (Входит до міжнародних наукометричних баз Ulrich's Periodicals, Index Copernicus.)

26. Kinoshenko, D. Image normalization via salient points of skeletons / D. Kinoshenko, E. Yegorova, S. Mashtalir, // *ICGST International Journal on Graphics, Vision and Image Processing. Special Issue on Features and Analysis*. – March 2006. – P. 1-7.

27. Chupikov, A. Morphological normalization of image binary cuts / Chupikov A, Mashtalir S., Yegorova E. // *Computation Imaging and Vision* / M.A. Viergever, ed. – Dordrecht: Springer. – 2006. – Vol. 32. – P. 558-564. (Входит до міжнародних наукометричних баз CrossRef, WorldCat, MS Academic Search.)

28. Image quotient set transforms in segmentation problems / D. Kinoshenko, S. Mashtalir, K. Shcherbinin, E. Yegorova // *International Journal «Information Technologies & Knowledge»*. – 2008. – Vol. 2. – No. 4. – P. 372-378. (Входит до міжнародних наукометричних баз BulDML, OpenAIRE, BASE та ін.)

29. Neural network segmentation of video via time series analysis / D. Kinoshenko, S. Mashtalir, A. Stephan, V. Vinarski // *International Journal «Information Theories & Applications»*. – 2011. – Vol. 18. – No 3. – P. 232-242.

30. On-line video segmentation using methods of fault detection in multidimensional time sequences / Ye. Bodyanskiy, D. Kinoshenko, S. Mashtalir, O. Mikhnova // *International Journal of Electronic Commerce Studies*. – 2012. – Vol. 3, No 1. – P. 1-20. (Входит до міжнародних наукометричних баз Engineering Village (EI)-GEOBASE, ABI/INFORM (ProQuest), Australian Research Council Journal List, Ulrich's periodicals directory, Index Copernicus, EIJ, DOAJ.)

31. Kinoshenko, D. Temporal video segmentation via spatial image segmentation / D. Kinoshenko, S. Mashtalir, V. Shlyakhov // *International Journal «Information Technologies & Knowledge»*. – 2013. – Vol.7. – No 3. – P. 212-219. (Входит до міжнародних наукометричних баз BulDML, OpenAIRE, BASE та ін.)

32. Mantula, E.V. Method of adaptive forecasting based on multidimensional linear extrapolation / E.V. Mantula, S.V. Mashtalir // *International Journal of Research in Engineering and Science*. – 2013. – Vol. 1. – No 4. – P. 31-37. (Входит до міжнародних наукометричних баз ANED, Index Copernicus, Jour Informatics)

33. Mashtalir, S. Key frame extraction from video: framework and advances / S. Mashtalir, O. Mikhnova // *International Journal of Computer Vision and Image Processing*. – 2014. – Vol. 4. – No 2. – P. 68-79. (Входит до міжнародних наукометричних баз ACM Digital Library, Bacon's Media Directory, Index Copernicus, Cabell's Directories, DBLP та ін.)

34. Mashtalir, S. Internal and external salient points under affine transformations. Comparative study / S. Mashtalir, K. Shcherbinin, E. Yegorova // *14-th International Conference in Central Europe on Computer Graphics, Visualization and*

Computer Vision, WSCG'2006: full papers proc, Czech Republic, Plzen, 30 January – 3 February, 2006. – Plzen: University of West Bohemia, 2006. – P. 217-223. (Входить до міжнародних наукометричних баз Scopus, DBLP, BASE, MS Academic Search.)

35. Кобылин, О.А. Анализ методов морфологической нормализации при обработке бинарных изображений / О.А. Кобылин, С.В. Машталир // Глобальные информационные системы. Проблемы и тенденции развития. 3-6 октября 2006, Туапсе. – Харьков: ХНУРЭ, 2006. – С.149-150.

36. Image partition transforms for faithful segmentation search / D. Kinoshenko, S. Mashtalir, K. Shcherbinin, E. Yegorova // Information Research & Applications: Proc. of Fifth International Conference I.Tech, Bulgaria, Varna, 26-30 June, 2007. – Sofia: ITHEA, 2007. – Vol. 2. – P. 385-389.

37. Машталир, С.В. Адаптивное нейросетевое сжатие изображений / С.В. Машталир, Е.С. Сакало // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта (ISDMCI'2008): сб. научных трудов, Евпатория, 19-23 мая 2008 г. – Херсон: ХНТУ, 2008. – Т.3, Ч.2. – С. 7-10.

38. Машталир, С.В. Методы контекстного сжатия визуальной информации / С.В. Машталир, С.В. Постульга, К.С. Щербинин // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта (ISDMCI'2009): сб. научных трудов, Евпатория, 18-22 мая 2009 г. – Херсон: ХНТУ, 2009. – Том 2. – С. 377-379.

39. Mashtalir, S. Image normalization under projective transforms / S. Mashtalir, E. Putyatin // Computer Graphics, Vision and Mathematics (GraVisMa 2009): Workshop Abstracts Proc., Czech Republic, Plzen, 2-5 September, 2009. – Plzen: University of West Bohemia, 2009. – P. 27.

40. Metrical properties of nested partitions for image retrieval / P. Grimm, Ie. Kinoshenko, S. Mashtalir, V. Shlyakhov // 54-th Internationales Wissenschaftliches Kolloquium (IWK) 'Information Technology and Electrical Engineering – Devices and Systems, Materials and Technology for the Future': Proc., Germany, Ilmenau, 7-10 September, 2009. – Ilmenau: Verlag ISLE, 2009. – ID. 14404. – 4 p.

41. Машталир, С.В. Анализ результатов сегментации видеоданных / С.В. Машталир, С.В. Постульга, К.С. Щербинин // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта. (ISDMCI'2009): сб. научных трудов, Евпатория, 17-21 мая 2010 г. – Херсон: ХНТУ, 2010. – Т. 2. – С. 431-435.

42. Fast training of neural networks for image compression / Ye. Bodyansky, P. Grimm, S. Mashtalir, V. Vinarski // Advances in Data Mining. Applications and Theoretical Aspects: Proc. of 10-th Industrial Conf. ICDM 2010, Germany, Berlin, 12-14 July, 2010. – Lecture Notes in Computer Science. – Berlin-Heidelberg-New York: Springer, 2010. – Vol. 6171. – P. 165-173. (Входить до міжнародних наукометричних баз Scopus, DBLP, ACM Digital Library, MS Academic Search та ін.)

43. Каграманян, А. Г. Свойства дифункциональности в мультиалгебраических системах / А.Г. Каграманян, С.В. Машталир, В.В. Шляхов // V міжнародна школа-семінар Теорія прийняття рішень: праці школи-семінару, Ужгород,

27 вересня – 1 жовтня 2010 р. – Ужгород: УжНУ, 2010. – С. 109-110.

44. Adaptive video segmentation via non-stationary multidimensional time series analysis / Y. Bodyanskiy, D. Kinoshenko, S. Mashtalir, O. Mikhnova // Applied and Theoretical Information Systems Research (ATISR): Proc. of Intern. Conf., Taiwan, Taipei, 10-12 February, 2012. – Taipei: Academy of Taiwan Information Systems Research, 2012. – P. 1-14.

45. Машталир, С.В. Многомерное экспоненциальное сглаживание в задаче анализа видеоданных / С.В. Машталир // VI Міжнародна школа-семінар «Теорія прийняття рішень»: праці школи-семінару, Ужгород, 29 вересня – 4 жовтня 2014 р. – Ужгород: УжНУ, 2012. – С. 136-137.

46. Машталир, С.В. Модель представления и анализа содержания видеокадров на основе диаграмм Вороного / С.В. Машталир, Е.Д. Михнова // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта. ISDMCI'2014: сб. научн. трудов междунар. научн. конф., Железный Порт, 28-31 мая 2014 г. – Херсон: ХНТУ, 2014. – С. 304-306.

47. Mashtalir, S. Video content analysis and key frame extraction method / S. Mashtalir, O. Mikhnova // Pattern Recognition and Information Processing: Proc. of the 12th Intern. Conf., Minsk, 28-31 May, 2014. – Minsk: UIIP NASB, 2014. – P. 181-186.

48. Богучарский, С.И. Формализация задачи сегментации видеоданных / С.И. Богучарский, С.В. Машталир // VII Міжнародна школа-семінар «Теорія прийняття рішень»: праці школи-семінару, Ужгород, 29 вересня – 4 жовтня 2014 р. – Ужгород, УжНУ. – 2014. – С. 43-44.

49. Богучарский, С.И. Фаззи-кластеризация изображений с учетом плотности распределения / С.И. Богучарский С.В. Машталир, Е.Д. Михнова // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта. ISDMCI'2015: сб. научн. трудов междунар. научн. конф., Железный Порт, 25-28 мая 2015 г. – Херсон: ХНТУ, 2015. – С. 251-253.

50. Машталир, С.В. Модели и методы интеллектуального анализа видеоданных / С.В. Машталир // «НАУКА И БИЗНЕС»: сб. научн. трудов I междунар. научно-практического форума, Днепропетровск, 29-30 июня 2015. – Киев: Noosphere, 2015. – С. 178-182.

51. Богучарский, С.И. Иерархическая агломеративная кластеризация в базах данных мультимедиа / С.И. Богучарский, С.В. Машталир // Междунар. науч.-техн. конф. «Электротехнические и компьютерные системы: теория и практика», Одесса, 20-24 июля 2015 г.: Электротехнические и компьютерные системы: специальный выпуск. – 2015. – 19(95). – С. 239-242. (Входит до міжнародних наукометричних баз Ulrich's Periodicals, Index Copernicus.)

52. Машталир, С.В. Анализ пространственно-временной сегментации видеопотоков / С.В. Машталир // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта. ISDMCI'2016: сб. научн. трудов междунар. научн. конф., Железный Порт, 24-28 мая 2016. – Херсон: ПП Вишемирський В.С., 2016. – С.295-297.

## АНОТАЦІЯ

Машталір С.В. Моделі та методи темпоральної обробки відео для інформаційного пошуку. – Рукопис.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Харківський національний університет радіоелектроніки, Міністерство освіти і науки України, Харків, 2016.

Дисертація присвячена вирішенню проблеми темпоральної обробки відеопотоків для інформаційного пошуку в великих відеоколекціях на базі запропонованих моделей і методів просторово-часового аналізу відео.

З урахуванням великих обсягів інформації, що треба обробити, введені критерії побудови областей елімінації заздалегідь невірної до запиту інформації. Також для підвищення швидкодії процес інформаційного пошуку розбитий на off line та online фази, що дозволяє провести великі обсяги обчислень до безпосередньо проведення запиту. З метою структуризації відеоданих введені методи темпоральної сегментації відеопослідовностей за рахунок аналізу індукованих багатовимірних часових рядів, що дозволяє провести розбиття вхідних даних на сукупність сегментів, в яких дуже великий ступінь подібності кадрів. Для цього проаналізовано побудову векторів характеристичних ознак окремих кадрів відео. Для полегшення індексації відео в окремих сегментах розроблено та модифіковано метод пошуку ключової інформації в відео, що дозволяє ще більше спростити процедури інформаційного пошуку ‘ad exemplum’. З метою скорочення кількості інформації, яку потрібно зберігати, вдосконалені інтелектуальні методи компресії відеокадрів. Обговорені результати експериментальних досліджень.

*Ключові слова:* відеодані, області елімінації, просторово-часова сегментація, ключові кадри, інформаційний пошук.

## АННОТАЦИЯ

Машталир С.В. Модели и методы темпоральной обработки видео для информационного поиска. – Рукопись.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.06 – информационные технологии. – Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины, Харьков, 2016.

В результате анализа существующих методов контекстного анализа визуальной информации выявлено (CBVIR), что, несмотря на определенные успехи при решении целых классов задач, в условиях наличия произвольной входной информации, что характерно для задач информационного поиска ‘ad exemplum’, эффективных методов анализа подобных данных на текущий момент крайне мало. Это связано, в первую очередь, с большими, а зачастую сверх большими объемами видео, требующим обработки в on line режиме и при

этом не имеющими каких-либо ограничений с точки зрения их анализа. Таким образом, разработка моделей и методов темпоральной обработки, анализа и интерпретации плохоструктурированных потоков информации позволяет развить интеллектуальные информационные технологии мультимедиа и гипермедиа. В этом случае релевантность информационных ресурсов, разумный компромисс между накоплением данных и потребностями их рационального использования также остаются одной из ключевых проблем. Тенденции непрерывного расширения классов решаемых задач, интенсификации количества обрабатываемой видеoinформации, глобальная аккумуляция данных в плохоструктурированных больших видеоархивах требуют разработки перспективных методов управления видеоданными, сводящихся к решению сложных задач анализа, многоаспектного автоматического индексирования, аннотирования и реферирования, нелинейного просмотра и поиска с различными формами запросов. При этом общие процедуры структуризации данных в больших видеоархивах должны проводиться и в off line режиме, что обеспечивает повышение эффективности информационного поиска.

Диссертация посвящена решению научно-технической проблемы темпоральной обработки видеопотоков для задач интеллектуального информационного поиска. В работе предложены, обоснованы и исследованы модели структуризации видеопоследовательностей для поиска с запросами 'ad exemplum' на основе построения областей элиминации для произвольных конфигураций объектов в произвольном метрическом пространстве, что позволяет значительно сократить время, необходимое на получение релевантных ответов на запрос за счет элиминации заведомо несоответствующих ему данных.

С целью повышения быстродействия разработаны фрагментные методы обработки визуальной информации, которые равносильно могут использоваться для темпоральной и пространственной сегментации видеоданных позволяющие как использовать различные подходы по анализу разных частей видеопоследовательностей, в том числе с выделением областей повышенного интереса как в пространстве так и во времени, что в свою очередь позволяет повысить валидность семантического поиска в больших библиотеках видеоданных.

Проанализированы и предложены методы валидного сравнения конечномерных разбиений видеокладов, за счет построения вектора дескрипторов элементов разбиений и изображений, что позволяет повысить адекватность представления видеорядов в признаковых пространствах.

Синтезирован комплекс моделей и методов on line сегментации видеопоследовательностей за счет анализа многомерных временных рядов, индуцируемых видеопотоками, путем обнаружения изменения их свойств, что обеспечивает валидную структуризацию видео, которую можно одинаково эффективно проводить как для информационного запроса, так и для элементов видеоколлекции.

Для дальнейшей структуризации сегментов видео предложены методы формирования и поиска ключевых кадров, представляющих сегменты видеоданных в задачах поиска видео на базе механизмов поиска изображений в



больших библиотеках статической визуальной информации. Также разработаны интеллектуальные методы компрессии видеокадров и видео для темпоральной обработки видеопотоков в задачах многоаспектной индексации, позволяющие в том числе сократить объемы хранимой информации после структуризации.

Путем экспериментальных исследований на множестве видеопоследовательностей установлена валидность разработанных моделей и методов темпоральной обработки видеоданных, а именно пространственно-временной сегментации с выделением ключевой информации в полученных сегментах. Определены условия рационального использования разработанных моделей и методов в задачах интеллектуального информационного поиска с запросами «по образцу». Также рассмотрены возможности использования предложенных методов анализа видеопоследовательностей для решения других классов задач и возможностей внедрения полученных результатов.

*Ключевые слова:* видеоданные, области элиминации, пространственно-временная сегментация, ключевые кадры, информационный поиск.

## ABSTRACT

Mashtalir S.V. Models and methods of temporal video processing in information retrieval. – Manuscript.

Thesis for the doctor degree of technical sciences in the speciality 05.13.06 – Information technology. – Kharkiv National University of Radio Electronics, Ministry Education and Science of Ukraine, Kharkiv, 2016.

The thesis is devoted to the development of spatio-temporal video analysis models and methods in information retrieval with query ‘ad exemplum’.

On account of very large information amount, elimination region constructions are introduced to delete information explicitly irrelevant to the query. Also, to speedup information retrieval process, information search is divided into on line phase and off line one which makes it possible to vigorous data process before execution of a query. Methods of video sequence temporal segmentation on the base of induced multivariate time series analysis are proposed for video structuring, the segmentation allows to partition video into a set of segments with large degree of frames similarity. For this, construction of video frame feature vectors is also analyzed. To facilitate the video indexing in segments method of video key frame extraction developed and modified, which can further simplify procedures for information retrieval ‘ad exemplum’. Advanced intelligent video compression techniques for raster data are proposed. Results of experimental investigations are discussed.

*Keywords:* video, elimination region, spatio-temporal segmentation, key frames, information retrieval.