

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

МІХНОВА ОЛЕНА ДМИТРІВНА

УДК 004.932.2:004.93'14

**МОДЕЛЬ ТА МЕТОД ПОШУКУ КЛЮЧОВИХ КАДРІВ ДЛЯ
РЕФЕРУВАННЯ ВІДЕОДАНИХ**

05.13.23 – системи та засоби штучного інтелекту

АВТОРЕФЕРАТ
дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2014

Дисертацією є рукопис.

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

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

Офіційні опоненти: доктор технічних наук, професор
Іванов Володимир Георгійович,
Національний юридичний університет
імені Ярослава Мудрого,
завідувач кафедри інформатики
і обчислювальної техніки;

доктор технічних наук, професор
Пелешко Дмитро Дмитрович,
Національний університет
«Львівська політехніка»,
професор кафедри інформаційних
технологій видавничої справи.

Захист відбудеться “ ____ ” _____ 2014 р. о ____ годині на засіданні спеціалізованої вченої ради Д 64.052.01 у Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

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

Автореферат розісланий “ ____ ” _____ 2014 р.

Учений секретар
спеціалізованої вченої ради Д 64.052.01

О.А. Винокурова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Поширення обсягів відеоданих у глобальній мережі, використання камер спостереження, відеореєстраторів та відеоендоскопів у різних сферах діяльності призвело до необхідності створення принципово нових методів обробки динамічно змінюваних графічних даних великого обсягу. Відеодані являють собою найбільш інформативний тип мультимедіа, що поєднує графічну, звукову, а іноді й текстову інформацію. Окрім того, ця інформація постійно змінюється в часі, що є однією з основних проблем, адже велике значення має швидкість обробки відеоданих, яка погіршується з ускладненням процедури обробки.

До основних напрямків обробки відеоданих відносять: сегментацію на об'єкти та спостереження за ними у динаміці, архівацію та класифікацію за жанрами, індексацію та пошук за запитом, анотування та реферування відео, виявлення границь сцен, виявлення повторів і реклами. Останнім часом актуальними стають задачі аналізу відео з урахуванням змісту (CBVR – Content Based Video Retrieval), так званий семантичний або високорівневий аналіз, який здійснюється за допомогою методів штучного інтелекту. Однак, в процесі створення систем CBVR виникають труднощі, пов'язані зі значним відривом традиційних низькорівневих ознак, що використовуються для аналізу зображень протягом останніх десятиліть, від високорівневого семантичного представлення змісту відео. Ще одна проблема виникає через суб'єктивність сприйняття відеоматеріалу користувачами й обсягами цього матеріалу, обробка якого вимагає просторово-часової сегментації з урахуванням різноманітних властивостей та наборів ознак.

Незважаючи на значний прогрес в області комп'ютерного зору та розпізнавання образів, пов'язаний з іменами таких видатних українських і закордонних вчених як С.Г. Антошук, Л.Г. Ахметшина, О.М. Березький, Р.А. Воробель, В.Г. Іванов, Ю.П. Кондратенко, В.М. Крилов, Н.О. Обухова, Ю.О. Скобцов, Д.Д. Пелешко, Є.П. Путятін, О.Н. Романюк, М.І. Шлезінгер, M. Agrawala, K.G. Derpanis, R.M. Haralick, V. Hlavac, J.B.T.M. Roerdink, D. Salesin, D. Schonfeld, P. Soille, M. Sonka, R. Szeliski, досі складно знайти метод, який би однаково ефективно виконував обробку різних за жанром відеоматеріалів, що отримуються від різних за якістю камер зі своїм налаштуванням, знятих під різним ракурсом і освітленням. Як правило, кожен метод обробки відео орієнтований на вузький клас задач, наприклад, тільки на стаціонарний фон та відносно великі об'єкти, що рухаються. При цьому клас об'єктів для розпізнавання (автомобілі, люди, номерні знаки тощо) також відомий апріорно.

Відносно новим напрямком обробки відео, особливо для вітчизняної науки, є статичне та динамічне реферування. У результаті статичного реферування відео користувач одержує стаціонарні репрезентативні зображення (ключові кадри), які характеризують вихідний відеоматеріал та якнайкраще дають уявлення про нього, тоді як у результаті динамічного реферування створюється відеоролик, що є коротшим за вихідний. Проаналізувавши існуючі зарубіжні підходи до пошуку ключових кадрів, можна зробити висновок про те, що деякі групи підходів успішно вилучають кадри зі значним рухом, інші можуть бути ефективно застосовані до

відео з незначною зміною фона, деякі доцільно використовувати тільки для відео з незмінним освітленням, інакше значимі за змістом кадри можуть бути пропущені або, навпаки, вилучені кадри з високим рівнем кореляції між собою.

Також, негативний вплив на результат спричиняє той факт, що число ключових кадрів у багатьох методах задається апріорно, а кадри вилучаються з попередньо сегментованого на сцени відеоматеріалу, обмежуючи кількість відібраних кадрів і встановлюючи рамки пошуку без урахування змісту. Таким чином, актуальною проблемою СВВР є дослідження методів пошуку ключових кадрів, виявлення можливостей підвищення якості та прискорення відбору значимих кадрів за умови універсальності процедури для відеоданих різного типу.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана на кафедрі інформатики Харківського національного університету радіоелектроніки в рамках держбюджетної науково-дослідної роботи: «Еволюційні гібридні системи обчислювального інтелекту зі змінною структурою для інтелектуального аналізу даних» (№ ДР 0110U000458, виконавець), а також господарсько-договірної науково-дослідної роботи «Розробка спеціалізованої підсистеми візуального аналізу вимірювальних даних» (№ ДР 0110U004788, виконавець). В рамках зазначених робіт здобувачем створено модель представлення змісту відеокadrів на основі математичного апарата діаграм Вороного, розроблено процедуру аналізу однотипності змісту відео за допомогою інтелектуальної обробки низькорівневих ознак, процедуру сегментації відеокadrів на основі діаграм Вороного вищих порядків, метод пошуку ключових кадрів, метрики порівняння розбиттів Вороного для кожного відеокадру.

Мета та задачі дослідження. Мета дисертаційної роботи полягає в розробці універсальної моделі та методу пошуку ключових кадрів у різних за змістом відеоданих, просторово-часовій сегментації відеоданих, інтелектуальній обробці низькорівневих ознак та аналізу змісту відео.

Поставлена мета передбачає розв'язання наступних задач:

– аналіз існуючих методів обробки динамічної візуальної інформації та трендів їх розвитку, що забезпечують базис семантичного аналізу в плані ефективного пошуку ключових кадрів для відеореферування та пошуку відеоматеріалів з урахуванням їх змісту;

– розробка моделі представлення відео апроксимацією областей, продукція яких виконується за допомогою «особливих» точок зображень, яка з урахуванням переваг виділення низькорівневих ознак на основі опорних точок дозволяє на більш високому рівні врахувати «зміст» відеокadrів;

– синтез методу пошуку ключових кадрів в умовах неоднорідного у часі змісту відеопослідовностей, який повинен забезпечувати надійне в аспекті значень мір повноти й точності представлення відеосегментів;

– дослідження специфіки використання запропонованої моделі та методу, апробація методу пошуку ключових кадрів на тестових відеоколекціях відкритого доступу, розв'язання за його допомогою практичних задач і впровадження результатів досліджень на промисловому підприємстві та в навчальний процес.

Об'єкт дослідження: процеси контекстної інтерпретації динамічної візуальної інформації.

Предмет дослідження: методи аналізу та представлення відео з метою реферування, метричної індексації та пошуку в відеоколекціях.

Методи дослідження: математичний апарат діаграм Вороного використаний з метою просторової сегментації відеокадрів, методи оцінки кольору, текстури, форми областей застосовані для аналізу змісту відео, метод Харріса використаний для знаходження опорних точок у відеокадрах, алгоритм кластеризації k -середніх застосований для уточнення місця розташування опорних точок, крім того використані основні положення теорії множин і математичної статистики.

Наукова новизна отриманих результатів. Наукова новизна дисертаційної роботи полягає в наступному:

– вперше запропонована модель представлення змісту відеокадрів на основі діаграм Вороного, яка, враховуючи усі переваги оперування з опорними точками на зображеннях, дозволяє вирішувати задачі порівняння в процесі аналізу відеокадрів;

– подальший розвиток отримав метод пошуку ключових кадрів, який на відміну від існуючих дозволяє комплексно враховувати зміну кольорових та текстурних ознак на розбиттях відеокадрів, що в свою чергу забезпечує покращення значень мір повноти й точності в процесі реферування відео;

– удосконалені методи побудови розбиттів зображень на базі опорних точок, які на відміну від існуючих за рахунок кластеризації k -середніх і управління порядком діаграм Вороного підвищують надійність та швидкодію метричного пошуку.

Практичне значення отриманих результатів. Запропонований метод дозволяє виконати інтерпретацію змісту відеокадрів за допомогою діаграм Вороного різного порядку. Порівняння розбиттів Вороного та пошук ключових кадрів можуть бути використані в інформаційно-пошукових системах з метою реферування відеоматеріалів. Метод може бути корисним для організацій з інтегрованою системою відеоспостереження, що дозволить прискорити перегляд відзнятого протягом багатьох годин матеріалу, у системах обробки зображень і відео для просторової сегментації та порівняння матеріалу, пошуку повторів та для розв'язання інших подібних задач. Застосування діаграм Вороного дозволило підвищити сталість розбиттів, а в результаті й надійність відбору значимих кадрів завдяки використанню не окремих точок, а областей навколо них. Самі по собі опорні точки дають менше інформації, аніж області, окрім цього, їх зміна від кадру до кадру значно більша, до того ж складно визначити, яка саме точка була зміщена, а яка залишилась на попередньому місці. Діаграми більше підходять для пошуку ключових кадрів і з точки зору відсутності необхідності сегментації на об'єкти, тому що дотепер є певні труднощі у виділенні об'єктів на відео різного жанру з динамічно змінюваним фоном.

Запропонована модель і метод успішно протестовані на відеоколекціях відкритого доступу, зокрема Internet Archive, Open Video Project, Movie Content Analysis Project, що підтвердило універсальність їх застосування. Крім того, метод пошуку ключових кадрів впроваджений на підприємстві АТ «Міжрегіональна електроенергетична асоціація «ЕЛТА» в процесі вдосконалення та додання нових функцій до створеної раніше та використовуваної спеціалізованої підсистеми візуального аналізу даних (акт про впровадження від 19.05.2014 р.), що дозволило

прискорити обробку ендоскопічних матеріалів експертами й забезпечило можливість підкріплення поточних звітів зображеннями на основі вилучених ключових кадрів. Наукові результати дисертаційних досліджень використані в навчальному процесі Харківського національного університету радіоелектроніки (акт про впровадження від 05.05.2014 р.). Модель представлення змісту відеоданих, запропонована в дисертаційній роботі, реалізована програмно, один із програмних компонентів сертифікований УкрСЕПРО в рамках системи імітаційного моделювання.

Особистий внесок здобувача. Усі наукові результати, описані в дисертації, отримані здобувачем особисто. У роботах, написаних у співавторстві, здобувачеві належить: аналіз особливостей обробки мультимедіа даних [1], вдосконалення методу Харріса для розміщення опорних точок у відеокадрах [2], оцінка результатів пошуку ключових кадрів за допомогою мір повноти й точності та аналіз можливостей оптимізації пошуку ключових кадрів [3], аналіз підходів до пошуку границь сцени [5], метод пошуку ключових кадрів за допомогою діаграм Вороного різних порядків [6], порівняння областей за їх формою за допомогою метрики на основі розбиттів [7], аналіз специфіки обробки відеоданих [8], вивчення обробки відеоданих та сегментації [9], аналіз застосування теоретико-множинного підходу для контекстного пошуку та деталізації відеоданих [10], визначення способів грануляції як інструментарію деталізації даних [13], модель представлення змісту відео на основі діаграм Вороного [14], аналіз існуючих методів пошуку ключових кадрів [15].

Апробація результатів дисертації. Основні результати дисертаційної роботи доповідались та обговорювались в рамках наступних наукових конференцій: IEEE International Conference on Advanced Video and Signal Based Surveillance (Сідней, Австралія, 22-24 листопада, 2006), XI Міжнародного молодіжного форуму «Радіоелектроніка й молодь в XXI столітті» (Харків, Україна, 10-12 квітня 2007), International Conference on Applied and Theoretical Information Systems Research (Тайбей, Тайвань, 10-12 лютого, 2012), International scientific and technical Internet conference “Computer graphics and image recognition” (Вінниця, Україна, 13-20 березня, 2012), Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (Євпаторія, Україна, 27-30 травня, 2012), 19-th Zittau East West Fuzzy Colloquium (Циттау, Німеччина, 5-7 вересня, 2012), 20-th Zittau East West Fuzzy Colloquium (Циттау, Німеччина, 25-27 вересня, 2013), Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» (смт. Залізний Порт, Україна, 28-31 травня, 2014), 12-th International Conference on Pattern Recognition and Information Processing (Мінськ, Білорусія, 28-31 травня, 2014).

Публікації. Основні результати дисертації опубліковані в 15 наукових працях, а саме: 4 статті у виданнях, що входять до переліку фахових видань України з технічних наук, та 2 статті в закордонних журналах (усього 3 індексуються у міжнародних наукометричних базах даних); 9 публікацій у матеріалах міжнародних наукових конференцій (з них 1 індексується у міжнародних наукометричних базах даних).

Структура дисертації. Дисертаційна робота складається із вступу, чотирьох

розділів, висновків, списку використаних джерел і додатка. Загальний обсяг дисертації становить 147 сторінок, з яких: основний текст дисертації займає 120 сторінок, 44 рисунка, 10 таблиць (з них 2 таблиці на 1 окремій сторінці), список використаних джерел з 155 пунктів на 17 сторінках, 1 додаток на 3 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дисертації, сформульовано мету, задачі, об'єкт, предмет і методи дослідження, визначено наукове та практичне значення отриманих результатів, наведено відомості про публікації (в тому числі відомості про особистий внесок здобувача до публікацій, написаних у співавторстві) та апробацію результатів дисертаційної роботи, надано інформацію стосовно впровадження результатів та участі здобувача у науково-дослідних роботах, що пов'язані з темою дисертаційних досліджень.

У **першому розділі** досліджено моделі представлення змісту відео за опорними точками, окремими об'єктами або областями та цілим кадром. Проведено аналіз сучасних методів статичного та динамічного реферування відео, наведено приклади областей застосування цих методів. Особлива увага приділяється підходам до пошуку ключових кадрів, котрі дозволяють відобразити сюжет відеоматеріалу у вигляді статичних зображень. Незважаючи на значну кількість існуючих методів штучного інтелекту та ознак (див. рис. 1), що використовуються на теперішній час для розв'язання цієї задачі, а також для аналізу та представлення змісту відео, пошуку повторів та супутніх задач, до основних проблем, що виникають, відносять складність сегментації на об'єкти при їх накладанні один на другий та нестационарному фоні,

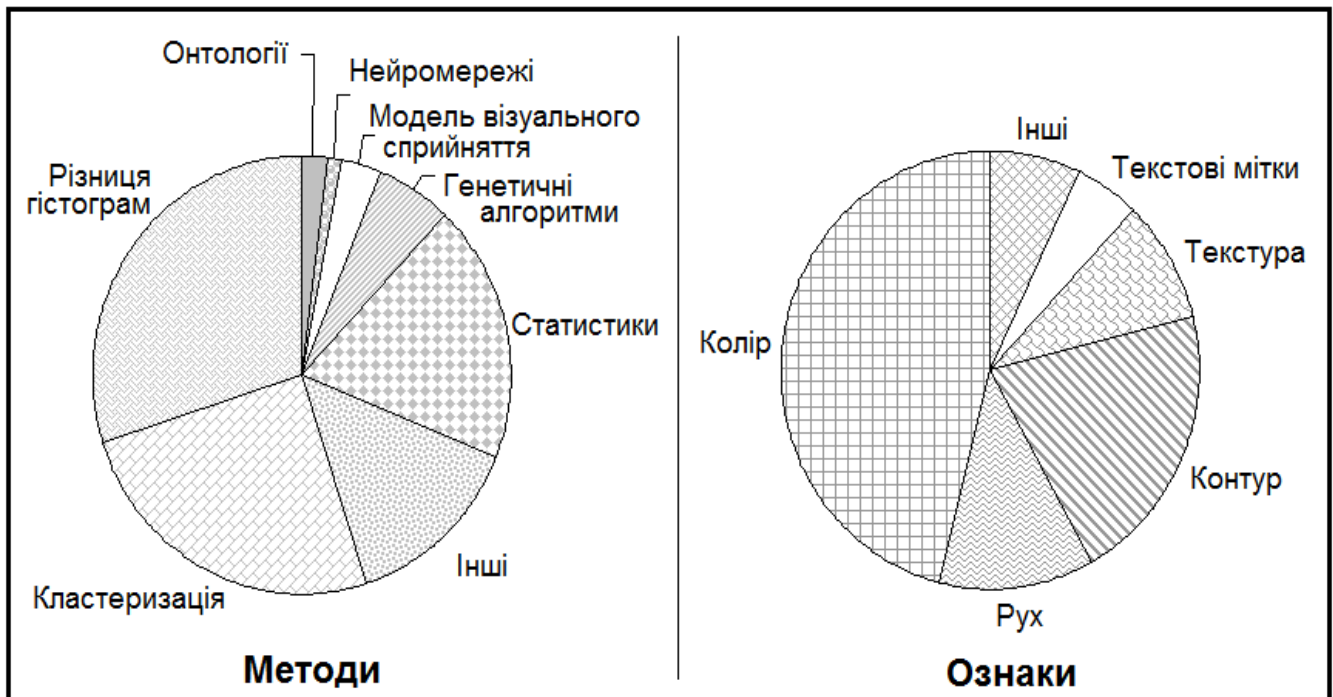


Рисунок 1 – Найпоширеніші методи штучного інтелекту та ознаки, що використовуються для пошуку ключових кадрів (побудовано на основі даних про більш ніж 50 методів, створених протягом останніх 10 років)

конфлікт низькорівневого та високорівневого представлення відеокадрів, суб'єктивізм оцінки результатів. Розробка конкурентоспроможної моделі та методу статичного реферування відео вимагає знаходження принципово нових рішень, що дозволить не тільки вилучати значимий зміст із відеоданих, але й дасть можливість усунути інформаційну перевантаженість, забезпечить зручність використання та доступу до відеоданих, що зберігаються.

Окрім цього, проаналізовано основні етапи, які традиційно використовуються в рамках різних методів пошуку ключових відеокадрів. Виявлення границь сцен часто виконується на попередньому етапі обробки даних, перед вилученням ключових кадрів. Незважаючи на важливість цього етапу з погляду машинного представлення змісту відео, його результати накладають обмеження на кількість вилучених з однієї сцени кадрів. Метод, що застосовується на основному етапі пошуку ключових кадрів, має особливий інтерес з погляду штучного інтелекту, тому що саме він повинен гарантувати точність результатів. Якщо кадри зі схожим змістом можуть бути усунуті на останньому етапі відбору ключових кадрів, то кількість пропущених кадрів зі значимим змістом і кадрів, невірно відібраних у якості ключових, може бути зведена до мінімуму тільки завдяки основному етапу та раціонально підібраному набору ознак.

Проаналізовано варіанти просторово-часової сегментації відеоряду Φ з урахуванням його динамічних властивостей, які відрізняють методи обробки зображень від відеоданих. Існуючі проблеми, а також перспективи машинного представлення й аналізу змісту відео визначають основний напрямок досліджень роботи як створення моделі та методу пошуку ключових кадрів $\{B_l^*(z)\}$ з мінімальною відстанню між кадрами в рамках однієї сцени $S_l(B_i, B_j)$, $l = 1, 2, \dots$, $1 \leq i \leq j \leq K$, $\forall l S_l(i, j) \neq \emptyset$, $\Phi = \bigcup_{l \in L} S_l(i, j)$, $\forall l', l'' S_{l'}(i, j) \cap S_{l''}(i, j) = \emptyset$, де B_i – перший кадр сцени S_l , а B_j – останній, K – загальна кількість кадрів відеоматеріалу, що підлягає реферуванню. Тобто відбір ключового кадру $B_r^*(z) \in S_l(i, j)$ зі сцени l здійснюється за наступним правилом: $r = \arg \min_{r \in L_l} (\sum_{t \in L_l, r \neq t} \rho(B_r(z), B_t(z)))$. Виходячи з цього, в розділі сформульовано задачу дослідження.

У другому розділі надано відомості стосовно запропонованої моделі представлення змісту відеоданих на основі діаграм Вороного, які дозволяють розбити простір зображень (відеокадрів) на сегменти, пов'язані з опорними точками (в літературі також можна зустріти синонімічні терміни: значима точка, гранична точка, точка інтересу, ключова точка, характеристична точка, характерна точка, сайт, атом), а потім відстежити зміни форми, місцезнаходження, площі, кольору та текстури цих сегментів від кадру до кадру, на основі чого зробити висновок про зміну змісту.

Для визначення кількості та місцезнаходження опорних точок у відеокадрах в рамках даної роботи пропонується використовувати метод Харріса з подальшим вдосконаленням розміщення точок за допомогою алгоритму кластеризації k -середніх з урахуванням кольорово-текстурних ознак, щоб знайдені точки більшою мірою відповідали змісту зображення в цілому, а не локальним різким змінам

інтенсивності, до визначення яких схильний метод Харріса. Алгоритм k -середніх обраний тому, що належність точки тому або іншому кластеру визначається відповідно до діаграми Вороного, згенерованої по середнім значенням, а метод Харріса вважається одним із найбільш швидких та якісних детекторів опорних точок.

На рис. 2 видно покращення сегментації кадру з урахуванням змісту: вихідна діаграма Вороного, побудована по опорним точкам Харріса, показана ліворуч; отримана за допомогою k -середніх діаграма Вороного з новими опорними точками (центроїдами кластерів) праворуч. Взагалі, гілки дерев на світлому фоні неба, текстова інформація у кадрі або будь-яка інша картата область створюють несприятливі умови роботи методу Харріса. У результаті опорні точки в занадто високій концентрації зосереджені в одній області кадру.



Рисунок 2 – Покращення сегментації кадру з урахуванням змісту

Застосування алгоритму k -середніх збільшує час обробки, однак ця міра повністю виправдана якістю сегментації. Крім того, новизна запропонованої процедури полягає в тому, що результат після застосування алгоритму k -середніх не залежить від вихідного розміщення опорних точок. Іншими словами, буде відрізнятися тільки кількість ітерацій, що призводять до одного результату, тобто отримана інваріантність щодо використовуваного методу пошуку опорних точок.

Уточнення (деталізація) в окремих областях кадру може виконуватись шляхом збільшення кількості вихідних опорних точок або шляхом побудови діаграм Вороного більш високих порядків (їх також називають узагальненими діаграмами Вороного). Експериментально підтверджено, що кластеризація великої кількості точок, особливо на зображеннях великого розміру з високою роздільною здатністю, виконується занадто довго. Таким чином, для того самого варіанта розміщення опорних точок можна побудувати різні діаграми (з різною кількістю областей), що заощадить час приблизно в півтора рази у порівнянні з додаванням нових опорних точок і перебудуванням діаграми першого порядку.

Діаграма Вороного k -го порядку, $V^{(k)}$, побудована на основі n опорних точок у двомірному просторі, являє собою розбиття площини на опуклі багатокутники, таке,

що точки z кожної області Вороного $v(p_i)^{(k)}$ мають однакову кількість найближчих опорних точок p_i , що дорівнює k . Зі збільшенням порядку k , розбиттів стає усе більше, тобто деталізація збільшується. Зменшення деталізації відбувається не тільки при максимально допустимому значенні порядку k (коли $k = n - 1$), але набагато раніше. Важливо зазначити, що для різної кількості опорних точок одержується різне граничне значення зменшення деталізації. У будь-якому разі зміна деталізації відбувається по параболі. Отримані області Вороного будь-якого порядку не призначені для навіть грубого обриса границь об'єкта, тому знаходження об'єкта в тій або іншій області лише свідчить про те, що розподіл пікселів кадру з погляду кольору та текстури є найбільш стабільним у даному поданні за умови вихідного місця розташування опорних точок та їх кількості, зумовлених методом Харріса.

Для того, щоб формально представити область Вороного, позначимо множину опорних точок як $\{p_1, p_2, \dots, p_n\}$. Діаграма Вороного являє собою розбиття поля зору $D = [a, b] \times [c, d]$, де $a, b, c, d = \text{const}$, на опуклі багатокутники $V = \{v(p_1) \cap D, v(p_2) \cap D, \dots, v(p_n) \cap D\}$ таким чином, що для кожної області Вороного виконується нерівність $v(p_i) = \{z \in R^2 : d(z, p_i) \leq d(z, p_j) \forall i \neq j\}$, де $d(\circ, \circ)$ – Евклідова метрика. Області Вороного будуються за серединними перпендикулярами між суміжними опорними точками p_i і p_λ . Області Вороного, пов'язані з суміжними опорними точками, мають невідроджену границю (ребро області Вороного, тобто більше однієї точки). Запишемо область Вороного $v(p_i)$ як перетин напівплощин, що проходять через суміжні опорні точки:

$$v(p_i) = \bigcap_{\lambda \in [1; \psi]} H(p_i, p_\lambda), \quad (1)$$

де ψ – кількість суміжних з p_i опорних точок p_λ .

Для порівняння розбиттів Вороного двох кадрів $V'(z)$ і $V''(z)$ з набором опорних точок $\{p'_1, p'_2, \dots, p'_n\}$ та $\{p''_1, p''_2, \dots, p''_m\}$ відповідно, скористаймося метрикою $\rho_1(V', V'')$:

$$\rho_1(V', V'') = \sum_{i=1}^n \sum_{j=1}^m \text{card}(v(p'_i) \Delta v(p''_j)) \text{card}(v(p'_i) \cap v(p''_j)), \quad (2)$$

де $v(p'_i) \Delta v(p''_j) = (v(p'_i) \setminus v(p''_j)) \cup (v(p''_j) \setminus v(p'_i))$ – симетрична різниця, що відображає кількість елементів (пікселів), по яким області $v(p'_i)$ і $v(p''_j)$ відрізняються одна від іншої.

Наступне рівняння дозволяє представити подібність кадрів через опорні точки, для чого достатньо записати рівняння прямої, яка містить серединний перпендикуляр між суміжними опорними точками з координатами $p_i(x_i; y_i)$ та $p_\lambda(x_\lambda; y_\lambda)$, тобто прямої, яка містить ребро області Вороного:

$$y = \frac{x_\lambda - x_i}{y_i - y_\lambda} \left(x - \frac{x_\lambda - x_i}{2} \right) + \frac{y_\lambda - y_i}{2}. \quad (3)$$

Враховуючи формули (1) і (3) можна переписати формулу (2) в термінах координат опорних точок:

$$\begin{aligned} \rho_1(V', V'') = & \sum_{i=1}^n \sum_{j=1}^m \text{card} \left(H \left(\bigcap_{\lambda \in [1; \psi]} \frac{x'_\lambda - x'_i}{y'_i - y'_\lambda} \left(x - \frac{x'_\lambda - x'_i}{2} \right) + \frac{y'_\lambda - y'_i}{2} \right) \Delta \right. \\ & \left. \Delta H \left(\bigcap_{\lambda \in [1; \psi]} \frac{x''_\lambda - x''_j}{y''_j - y''_\lambda} \left(x - \frac{x''_\lambda - x''_j}{2} \right) + \frac{y''_\lambda - y''_j}{2} \right) \right) \times \\ & \times \text{card} \left(H \left(\bigcap_{\lambda \in [1; \psi]} \frac{x'_\lambda - x'_i}{y'_i - y'_\lambda} \left(x - \frac{x'_\lambda - x'_i}{2} \right) + \frac{y'_\lambda - y'_i}{2} \right) \cap \right. \\ & \left. \cap H \left(\bigcap_{\lambda \in [1; \psi]} \frac{x''_\lambda - x''_j}{y''_j - y''_\lambda} \left(x - \frac{x''_\lambda - x''_j}{2} \right) + \frac{y''_\lambda - y''_j}{2} \right) \right). \end{aligned} \quad (4)$$

Вищенаведена метрика подібності вказує на те, наскільки дві діаграми схожі з точки зору форми областей. Для того, щоб прийняти до уваги кольорові та текстурні ознаки, визначимо ще дві метрики, $\rho_2(B'(z), B''(z))$ і $\rho_3(B'(z), B''(z))$ відповідно, задані на спільних частинах розбиттів:

$$\rho_2(B'(z), B''(z)) = \sum_{i=1}^n \sum_{j=1}^m \sum_{x_q} \sum_{y_u} (x_q, y_u) \in (v(p'_i) \cap v(p''_j)) (B'(x_q, y_u) - B''(x_q, y_u))^2, \quad (5)$$

де $B'(x_q, y_u)$ – значення інтенсивності пікселя з області $(v(p'_i) \cap v(p''_j))$,

та

$$\rho_3(B'(z), B''(z)) = \sum_{i=1}^n \sum_{j=1}^m (v(p'_i), v(p''_j)) \supseteq (v(p'_i) \cap v(p''_j)) \left| E(v(p'_i)) - E(v(p''_j)) \right|, \quad (6)$$

де $E(v(p'_i))$ – значення ентропії для області $v(p'_i)$.

Метрики оцінки подібності за кольором і текстурою можна представити аналогічно формулі (4) у термінах координат опорних точок, а не областей Вороного. Таким чином, отримано три ненормалізовані оцінки. Далі пропонується виконати нормалізацію для формул (2), (5) і (6), щоб одержати значення в діапазоні від 0 до 1. Перетворення вищевказаних метрик до обмеженої форми припускає використання функції компандеру діапазону:

$$\rho'(B'(z), B''(z)) = \frac{\rho(B'(z), B''(z))}{1 + \rho(B'(z), B''(z))}. \quad (7)$$

Комбінація цієї функції з метрикою в результаті як і раніше призведе до отримання метрики, що задовольняє вимогам рефлексивності, симетричності та

нерівності трикутника. Завдяки тому, що ненегативна лінійна комбінація метрик призводить до одержання метрики, можна використати наступну результуючу метрику:

$$\hat{\rho}(B'(z), B''(z)) = \alpha_1 \rho'_1 + \alpha_2 \rho'_2 + \alpha_3 \rho'_3, \quad \sum_{\gamma=1}^{\gamma=3} \alpha_\gamma = 1, \quad \alpha_\gamma \geq 0, \quad (8)$$

де $\hat{\rho}(B'(z), B''(z))$ показує подібність кадрів;

α_γ – важливість кожної використовуваної ознаки.

Незважаючи на різноманітність моделей представлення й аналізу змісту відео, використання діаграм Вороного має ряд переваг. Перш за все, для реалізації розбиття за допомогою діаграм Вороного необхідно менше обчислювальних ресурсів, аніж для аналізу руху сегментованих об'єктів. Діаграми Вороного різного порядку більш сталі до динамічних змін відео також у порівнянні з опорними точками. Останні дають менше інформації, аніж області, їх зміна від кадру до кадру більш значна та складніша в ідентифікації. Окрім того, діаграми Вороного раніше не використовувались з метою представлення змісту та реферування відео, що має безумовний інтерес для досліджень з точки зору конкурентоспроможності застосування нової моделі для пошуку ключових кадрів.

У першому розділі наголошувалось на тім, що будь-який метод відбору ключових кадрів, як правило, передбачає визначення границь сцен на попередньому етапі аналізу відео. У даному випадку пропонується не тільки виконати часову сегментацію на сцени, але й проаналізувати однотипність змісту відеоданих перед їх обробкою. Це дає можливість визначити поріг відбору дійсно значимих для даної відеопослідовності кадрів, а врахування границь сцен на останньому етапі пошуку ключових кадрів, замість першого, не накладає обмежень на кількість відібраних кадрів у порівнянні з багатьма методами. Далі наведено опис всієї процедури порівняння розбиттів Вороного для відбору ключових кадрів.

Етап 1. Визначення однотипності відеокadrів. Враховуючи текстуру (дисперсію ентропії) в усій відеопослідовності, виконаємо установку граничного значення відповідно до наступного правила:

$$Thresh = \begin{cases} \frac{1}{4}, \frac{1}{K-1} \sum_{k=1}^{k=K} \left(E(B_k(z)) - \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)) \right)^2 \rightarrow \infty, \\ \frac{1}{2}, \frac{1}{K-1} \sum_{k=1}^{k=K} \left(E(B_k(z)) - \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)) \right)^2 \rightarrow \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)), \\ \frac{3}{4}, \frac{1}{K-1} \sum_{k=1}^{k=K} \left(E(B_k(z)) - \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)) \right)^2 \rightarrow 0, \end{cases} \quad (9)$$

де K – загальне число кадрів відеопослідовності;

$E(B_k(z))$ – значення ентропії для k -ого відеокадру.

Через те, що близькість за змістом є відносною мірою подібності відеокадрів, доводиться використовувати граничне значення для завдання прийняттого ступеня їх подібності. В даному випадку граничне значення встановлюється з урахуванням змісту відео. Для відео з гетерогенним змістом та розмаїтістю сцен, це значення повинне бути менше $\frac{1}{4}$, щоб не вилучати дуже велику кількість кадрів. І навпаки, для відео з гомогенним змістом і невеликою кількістю сцен (або навіть із одною сценою), це значення повинно бути збільшене до $\frac{3}{4}$, щоб вилучати трохи більше ключових кадрів.

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

Етап 3. Порівняння діаграм Вороного. Відповідно до формули (8), розраховуємо значення $\hat{\rho}(B_k(z), B_{k+1}(z))$ для двох кадрів. Якщо значення $\hat{\rho}(B_k(z), B_{k+1}(z))$ виявилось меншим за встановлене граничне значення, у такому випадку обидва кадри $B_k(z)$ й $B_{k+1}(z)$ вважаються ключовими та вилучаються як $B_r^*(z)$ і $B_{r+1}^*(z)$, після чого переходимо до наступного етапу 4, у протилежному випадку вилучається тільки один кадр $B_k(z)$ як ключовий $B_r^*(z)$ і переходимо до етапу 5.

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

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

Етап 6. Повторюємо етап 3 поки $B_{k+1}(z) \leq K$.

Етап 7. Відбір ключових кадрів з урахуванням границь сцен. Останнім етапом буде видалення схожих кадрів, якщо такі були відібрані в рамках однієї сцени. Границі сцен можна визначити за допомогою одного з існуючих методів різниці кольорових гістограм або аналізу часових рядів. Порівнюємо ключові кадри попарно усередині кожної сцени з використанням процедури, описаної на етапі 3. Знайдений другий ідентичний ключовий кадр видаляємо з послідовності ключових кадрів. Результуюча послідовність ключових кадрів буде виглядати таким чином: $\{\hat{B}_1^*(z), \dots, \hat{B}_l^*(z)\}$.

Застосування спеціально розроблених метрик подібності, орієнтованих безпосередньо на аналізовані групи ознак, дозволяє підвищити якість відбору кадрів, адже, як відомо, на результативність будь-якої процедури впливає не тільки характер аналізованих даних і застосовуваних методів, але й значною мірою обрані й використовувані метрики подібності. Процедура може бути розширена використанням зважених оцінок ознак, тому що в даному випадку була задіяна їх лінійна комбінація з однаковими ваговими коефіцієнтами, крім того, можуть бути використані додаткові ознаки, наприклад, рух областей Вороного (хоча це значно збільшить час обробки відеоданих). Визначення важливості кожної використовуваної ознаки є пріоритетним напрямком подальших досліджень.

У третьому розділі надано інформацію стосовно заходів модифікації запропонованого методу пошуку ключових кадрів для покращення ефективності та прискорення обробки відеоданих. На рис. 3 наведено діаграму, яка наочно демонструє середній час, необхідний для обробки різних за жанрами відеоматеріалів тривалістю одну хвилину за допомогою запропонованого методу. Пошук ключових кадрів у документальних фрагментах та рекламних роликах виконано швидше завдяки лише тому, що деякі ролики цих категорій були чорно-білі. Чим менше деталей містить кадр та менше опорних точок виявлено на початковому етапі, тим скоріше виявляється обробка відеоданих.

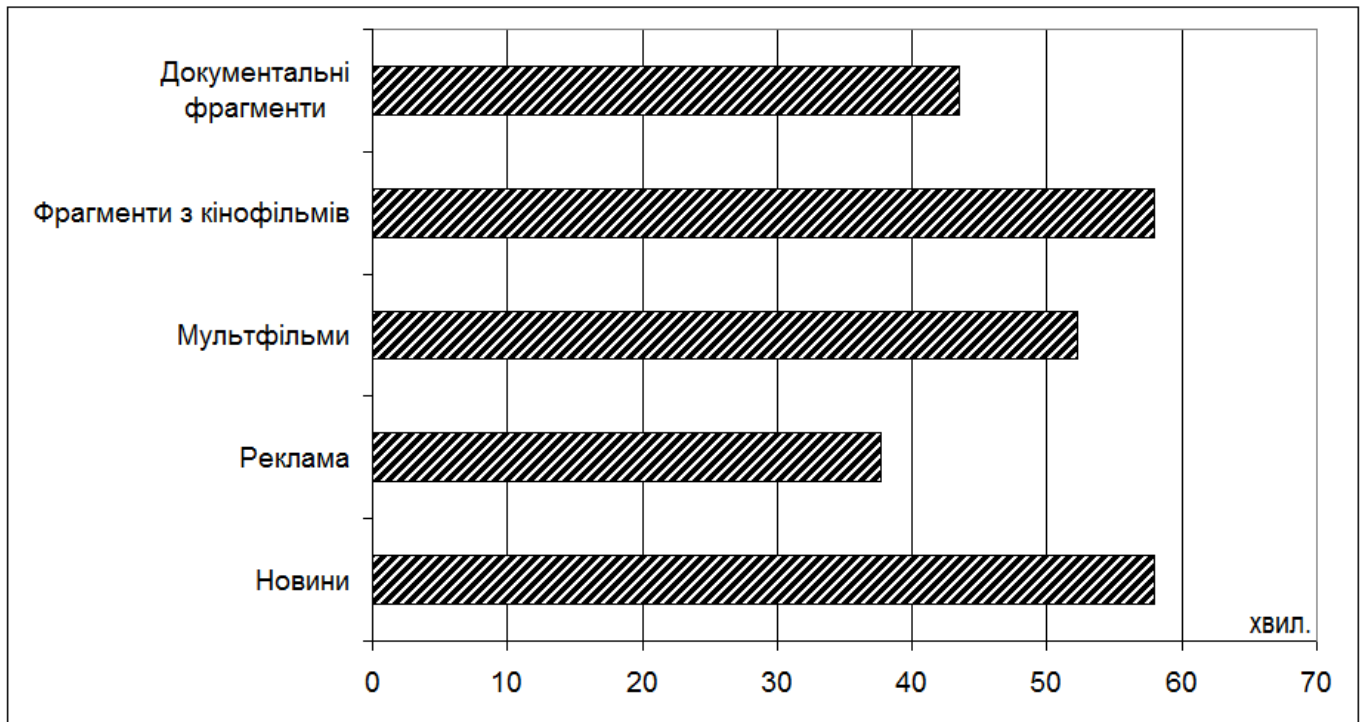


Рисунок 3 – Середній час (у хвилинах), необхідний для обробки різних за жанрами відеоматеріалів тривалістю одну хвилину

Тривалий час обробки зумовлено кількома причинами. По-перше, кожен з роликів має 29 кадрів на секунду. По-друге, усі дані оброблено на комп'ютері з низькою прожерною частотою 1,6 ГГц Dual CPU та усього 1 Гбайт оперативної пам'яті. Окрім численних деталей, які було присутні у відеороликах та зменшили швидкість обробки, 5 опорних точок буде встановлено на переміщено за одну секунду, тоді як 15 опорних точок для того самого кадру будуть обрані вже за 2 секунди. Це справедливо для кадру розміром 320*240 пікселів, однак збільшення розміру кадрів призведе до збільшення часу обробки даних. Проаналізовано вплив розміру та роздільної здатності на основні характеристики відеокadrів, а також на швидкість пошуку ключових кадрів. Встановлено, що при наявності можливості вибору розміру кадрів, краще використовувати кадри біля 800*600 пікселів, не уповільнюючи обробку значною кількістю пікселів.

Завдяки тому, що найбільш стійкими до зміни змісту є діаграми Вороного найвищого порядку, вони потенційно можуть використовуватись не тільки для порівняння відеокadrів між собою, але й для визначення границь сцен у

відеопослідовності. Діаграма найвищого порядку для кадрів, взятих з однієї сцени, може лишатись майже незмінною. Однак для знаходження границь сцен існує безліч інших методів. Визначення найкращого методу аналізу границь сцен не ставилося за мету досліджень, хоча переваги та недоліки таких методів було проаналізовано.

Запропоновано заходи по скороченню часу розміщення опорних точок, тому що в результаті експериментів встановлено, що саме цей етап займає найбільше часу в усій процедурі пошуку ключових кадрів. Проаналізовано варіанти встановленні кількості опорних точок та їх вихідного розміщення, адже метод Харріса вимагає ресурсномістке обчислення похідних. Також проаналізовано такі показники кластеризації як кількість ітерацій та сума відстаней під час зміни місця розташування опорних точок за допомогою алгоритму k -середніх для відеопослідовностей різного типу, досліджено швидкість процедури пошуку ключових кадрів при застосуванні інших алгоритмів кластеризації на основі центроїд (відмінних від k -середніх), а саме Simmulated Annealing, Genetic Algorithms, Tabu Search, комбінації першого та останнього методів, яка виявилась найбільш продуктивною.

У четвертому розділі проведено оцінку повноти та точності запропонованого методу на тестових відеопослідовностях різного жанру. Результат експертних оцінок методу дозволяє судити про його конкурентоспроможність у порівнянні з існуючими підходами, деякі з котрих було випробувано на тих саме тестових колекціях. На рис. 4 наведена лінійна діаграма з середніми значеннями коефіцієнтів Дайса, котрі отримано при вилученні ключових кадрів з різних за жанром відеопослідовностей. Причина низького середнього значення коефіцієнту Дайса для

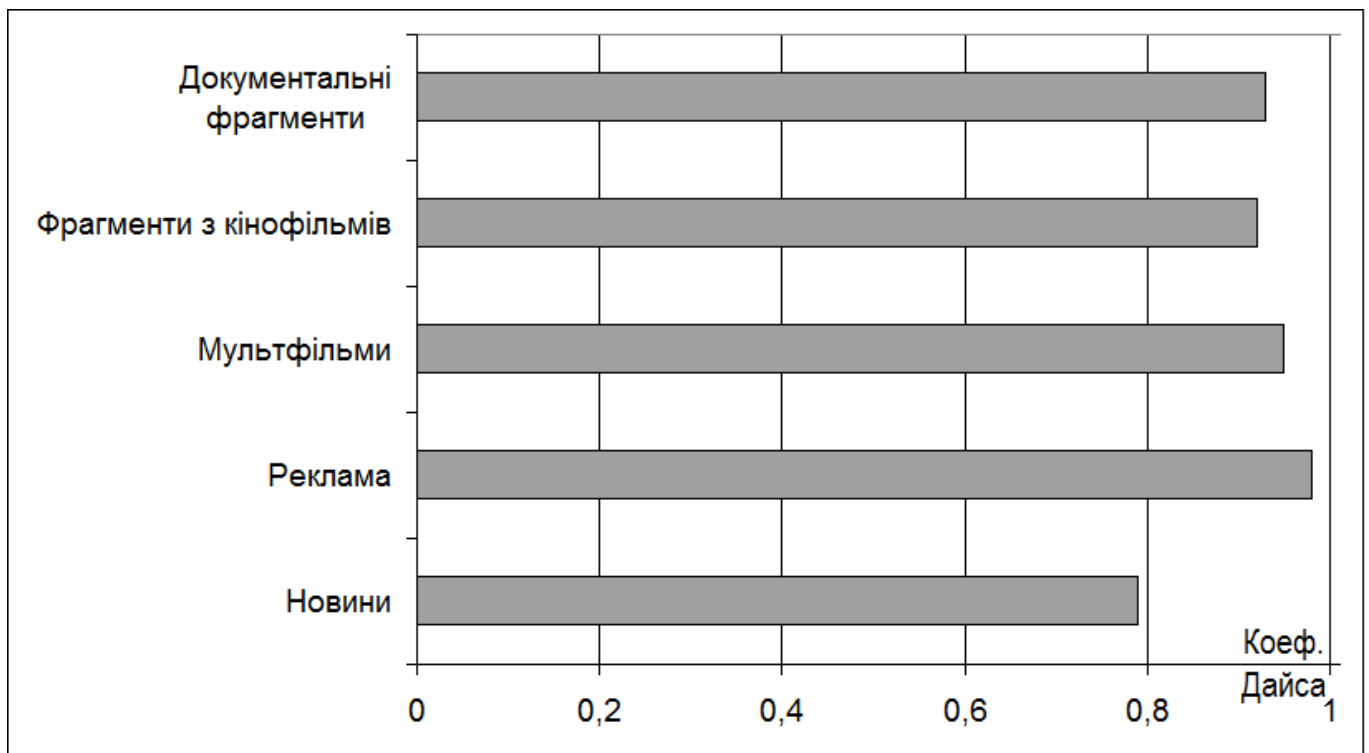


Рисунок 4 – Значення коефіцієнтів Дайса, котрі отримано при вилученні ключових кадрів з різних за жанром відеопослідовностей

відеоданих з новинами полягає у наявності в них текстової інформації, розпізнавання якої не ставилось за мету досліджень.

Порівняння результатів пошуку ключових відеокадрів існуючими методами з кадрами, вилученими за допомогою методу на основі діаграм Вороного, дозволяє дійти висновку, що запропонований метод дає більше можливостей врахувати значимі зміни змісту, при цьому немає необхідності задавати кількість ключових кадрів апріорно, як це робиться в деяких методах. Метод дозволяє вилучати кадри як при однорідному змісті сцен без істотного руху, так і при значних змінах експозиції, що особливо важливо з погляду універсальності.

Додаток до дисертації містить копії актів про впровадження результатів дисертаційної роботи.

ВИСНОВКИ

У дисертаційній роботі запропоновано нове розв'язання актуальної науково-практичної задачі розробки моделі представлення та аналізу змісту відео, а також створення методу пошуку ключових кадрів на основі цієї моделі для реферування відеоданих різного жанру. У рамках проведених досліджень досягнута мета та розв'язано усі поставлені задачі, отримано наступні результати та висновки.

1. Досліджено методи штучного інтелекту, які можуть використовуватись з метою реферування відеоданих, а також низькорівневі та високорівневі ознаки, що характеризують зміст відеокадрів. Виявлено їх переваги, недоліки та шляхи вирішення існуючих проблем.

2. В узагальненому вигляді запропонована модель представлення змісту відеокадрів на основі математичного апарата діаграм Вороного з урахуванням кольорово-текстурних ознак та ознак форми областей. Застосовуючи запропоновані метрики аналізу низькорівневих ознак і порівняння діаграм Вороного, побудованих для кожного кадру, отримана можливість машинного представлення змісту відео, що дозволило вилучати ключові кадри.

3. На основі запропонованої моделі та синтезованих метрик подібності розбиттів Вороного розроблено новий метод пошуку ключових кадрів з урахуванням змісту відео та границь сцен. Метод орієнтований не тільки на вилучення значимого змісту, але й на усунення повторів, якщо вони з'являються в результуючій послідовності ключових кадрів.

4. Установлено, що використання діаграм Вороного вищих порядків для деталізації представлення змісту відеокадрів є більш доцільним, чим збільшення або зменшення числа опорних точок з наступним перебудуванням діаграми Вороного першого порядку.

5. Метод Харріса адаптований під процедуру пошуку ключових кадрів з використанням діаграм Вороного та розширений за допомогою алгоритму кластеризації k -середніх, що забезпечує більшу орієнтацію на зміст кадрів при розміщенні опорних точок та усуває невірні розміщенні точки в областях з викидами інтенсивності.

6. Запропонована модель і метод пройшли перевірку на тестових колекціях відкритого доступу та самостійно відзнятому відеоматеріалі, крім того, проведено

порівняльний аналіз запропонованого методу з деякими з існуючих. Також для оцінки методу використовувались загальноприйняті міри повноти й точності. В процесі перекомпонування кластерів проведена перевірка їх валідності після переміщення опорних точок для формування більш стабільних областей Вороного. Досліджено можливості оптимізації запропонованого методу.

7. Результати теоретичних і практичних досліджень реалізовані програмно у вигляді модуля пошуку ключових кадрів і випробувані в ході роботи електроенергетичного підприємства АТ «Міжрегіональна електроенергетична асоціація «ЕЛТА», що дало можливість автоматизувати та прискорити обробку відеоданих експертами цього підприємства. Наукові результати дисертаційних досліджень застосовано у навчальному процесі Харківського національного університету радіоелектроніки при підготовці бакалаврів та магістрів за напрямком та спеціальністю «Інформатика», а також при проведенні лекційних занять та лабораторного практикуму з дисциплін «Стиснення даних» та «Теорія розпізнавання образів». Один з розроблених програмних компонентів сертифікований УкрСЕПРО.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Михнова, А. В. Особенности визуализации данных в системах диагностики / А. В. Михнова, Е. Д. Михнова // Системи обробки інформації. – Харків : ХУПС, 2012. – № 2 (100). – С. 67-70.
2. 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. (Входить до міжнародної наукометричної бази INSPEC.)
3. Mashtalir, S. V. Stabilization of key frame descriptions with higher order Voronoi diagram / S. V. Mashtalir, O. D. Mikhnova // Бионика интеллекта. – 2013. – № 1. – Р. 68-72.
4. Михнова, Е. Д. Анализ видеоданных на основе диаграмм Вороного различного порядка / Е. Д. Михнова // Збірник наукових праць ХУПС. – Харків : ХУПС, 2014. – № 1 (38). – С. 142-145.
5. On-line video segmentation using methods of fault detection in multidimensional time sequences / Y. Bodyanskiy, D. Kinoshenko, S. Mashtalir, O. Mikhnova // International Journal of Electronic Commerce Studies. – 2012. – Vol. 3, No. 1. – P. 1-20. (Входить до міжнародних наукометричних баз Australian Research Council Journal List, Ulrich's periodicals directory, Index Copernicus, EIJ, DOAJ.)
6. Mikhnova, O. Key frame partition matching for video summarization / O. Mikhnova, N. Vlasenko // International Journal of Information Models and Analyses. – 2013. – Vol. 2, No. 2. – P. 145-152. (Входить до міжнародних наукометричних баз ITHEA Bibliographic Database, Google Scholar, Cite Seer, DBLP.)
7. A novel metric on partitions for image segmentation / V. Mashtalir, O. Mikhnova, V. Shlyakhov, E. Yegorova // Video and Signal Based Surveillance : proc. of intern. conf.,

Sidney, 22-24 Nov. 2006. – Washington : IEEE CS, 2006. – P. 18. 1-6. (Входить до міжнародних наукометричних баз IEEE CS DL, ACM DL, Google Scholar.)

8. Михнова, А. В. Особенности графической визуализации результатов испытаний автоматических выключателей / А. В. Михнова, Е. Д. Михнова // Радиоэлектроника и молодежь в XXI веке : матер. XI междунар. молодежн. форума, Харьков, 10-12 апр. 2007 г. – Харьков : ХНУРЭ, 2007. – Ч. 1. – С. 411.

9. Adaptive video segmentation via non-stationary multidimensional time series analysis / Y. Bodyanskiy, D. Kinoshenko, S. Mashtalir, O. Mikhnova // Applied and Theoretical Information Systems Research : proc. of intern. conf., Taipei, 10-12 Feb. 2012. – New Taipei City : Academy of Taiwan Information Systems Research, 2012. – P. 1-14.

10. Mashtalir, V. Multigroups in granular computing paradigm. Part 1. Formalistic premises / V. Mashtalir, O. Mikhnova, V. Shlyakhov // East West Fuzzy Colloquium 2012 : proc. of the 19th Zittau fuzzy colloquium, Zittau, 5-7 Sep. 2012. – Zittau/Görlitz : Hochschule Zittau/Görlitz, 2012. – P. 226-233.

11. Mikhnova, O. A template-based approach to key frame extraction from video / O. Mikhnova // Computer graphics and image recognition : proc. of intern. scient. and techn. Internet conf., Vinnytsia, 13-20 Mar. 2012. – Vinnytsia : VNTU, 2012. – P. 120-127.

12. Михнова, Е. Д. Извлечение ключевых кадров из эндоскопических и рентгеноскопических материалов / Е. Д. Михнова // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта : матер. междунар. науч. конф., Евпатория, 27-30 мая 2012 г. – Херсон : ХНТУ, 2012. – С. 384-386.

13. Mashtalir, V. Conditions of multiring production by two ternary relations / V. Mashtalir, O. Mikhnova, V. Shlyakhov // East West Fuzzy Colloquium 2013 : proc. of the 20th Zittau fuzzy colloquium, Zittau, 25-27 Sep. 2013. – Zittau/Görlitz : Hochschule Zittau/Görlitz, 2013. – P. 104-111.

14. Машталир, С. В. Модель представления и анализа содержания видеок кадров на основе диаграмм Вороного / С. В. Машталир, Е. Д. Михнова // Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта : матер. междунар. науч. конф., Железный Порт, 28-31 мая 2014 г. – Херсон : ХНТУ, 2014. – С. 304-306.

15. 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.

АНОТАЦІЯ

Міхнова О.Д. Модель та метод пошуку ключових кадрів для реферування відеоданих. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки, Харків, 2014.

Дисертаційна робота присвячена розвитку актуального напрямку обробки відеоданих – пошуку ключових кадрів у різних за жанром відеопослідовностях із

залученням методів просторово-часової сегментації. З точки зору сегментації відео у часі досліджено переваги та недоліки існуючих методів визначення границь сцен. З точки зору просторової сегментації розбиття поля зору кожного кадру реалізовано за допомогою діаграм Вороного з урахуванням кольору, текстури та форми областей. Вдосконалено метод визначення місцезнаходження опорних точок, за якими будуються діаграми Вороного. Проаналізовано можливості та особливості побудови діаграм Вороного різних порядків для представлення змісту відео. Розроблено спеціалізовані метрики, які дозволили порівняти діаграми Вороного та визначити значні зміни змісту в усьому відео.

На основні запропонованої математичної моделі представлення та аналізу змісту відео побудовано метод пошуку ключових кадрів, який дозволяє не тільки видобувати кадри зі значимим змістом, але й усувати близькі за змістом кадри. Проведені експериментальні дослідження підтвердили конкурентоспроможність нової моделі та методу, які було успішно впроваджено на промисловому підприємстві.

Ключові слова: реферування відео, ключовий кадр, просторово-часова сегментація, діаграма Вороного, опорна точка.

АННОТАЦІЯ

Михнова Е.Д. Модель и метод поиска ключевых кадров для реферирования видеоданных. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. – Харьковский национальный университет радиоэлектроники, Харьков, 2014.

Диссертационная работа посвящена развитию актуального направления обработки видеоданных – поиску ключевых кадров в разных по жанру видеопоследовательностях с привлечением методов пространственно-временной сегментации. С точки зрения сегментации видео во времени исследованы достоинства и недостатки существующих методов определения границ сцен. С точки зрения пространственной сегментации разбиение поля зрения каждого кадра реализовано с помощью диаграмм Вороного с учетом цвета, текстуры и формы областей. Усовершенствован метод определения местоположения опорных точек, по которым строятся диаграммы Вороного. Проанализированы возможности и особенности построения диаграмм Вороного различных порядков для представления содержания видео. Разработаны специализированные метрики, которые позволили сравнить диаграммы Вороного и определить значительные изменения содержания во всем видео.

На основе предложенной математической модели представления и анализа содержания видео разработан метод поиска ключевых кадров, который позволяет не только извлекать кадры со значительным изменением содержания, но и устранять близкие по содержанию кадры. Проведенные экспериментальные исследования подтвердили конкурентоспособность новой модели и метода, которые были успешно внедрены на промышленном предприятии.

Ключевые слова: реферирование видео, ключевой кадр, пространственно-временная сегментация, диаграмма Вороного, опорная точка.

ABSTRACT

Mikhnova O.D. Model and method of key frame extraction for video summarization. – Manuscript.

Thesis for the candidate degree of technical sciences in the speciality 05.13.23 – Artificial Intelligence Systems and Methods. – Kharkiv National University of Radio Electronics, Kharkiv, 2014.

Integration of video endoscopes and cameras into industrial, medical, security and other tracking systems has caused great interest to intellectual video processing. As a rule, monitoring of any object assumes huge amount of time and human resources needed not only for decision making, that influences on the state of objects under control, but also for long lasting tracking and identification of interesting or non-ordinary conditions. Video summarization is an excellent tool for redundancy elimination in video, which is achieved by key frame extraction. These static images with significant content provide a brief overview on what was going on in video for hours. Key frames can also facilitate indexing, archiving, searching and cataloging video information.

Despite of variety of existing key frame extraction methods and content presentation models, the main problem they face is the gap between information retrieved from video at a low level and semantic interpretation at a high level required for efficient summarization. Another challenge consists in different lighting conditions and camera characteristics, with which a video is shot. That is why video summarization attracts more and more research and development efforts.

To overcome the aforementioned problems a novel model and method have been developed. The main mathematics which lies on its basis is Voronoi diagrams applied for spatial segmentation of frames. The diagram corresponds to decomposition of a plane (frames, in our case) into Voronoi tessellations or regions. Each tessellation of the first order Voronoi diagram is connected with its own salient point and built under the following rule: the distance from any point located at the same tessellation to its salient point is less or equal to its distance to any other salient point. Voronoi diagrams of higher order (or generalized Voronoi diagrams) are built according to the following rule: the distance between the farthest point of one Voronoi tessellation to its corresponding generator points is closer or equals to the distance to any nearest generator point of another tessellation. Arbitrary Voronoi tessellation of order k may contain from 0 to k generator points. A simple Voronoi diagram of the first order is a particular case of higher order Voronoi diagrams.

Application of Voronoi diagrams has several advantages over traditional segmentation into objects. First of all, Voronoi diagram takes much less computational resources to build and, unlike object tracing, it is strictly defined. Higher order Voronoi diagrams are more stable to dynamic changes in video content compared with salient points as well. The latter gives less information and their changes from frame to frame are more significant and hard to identify. In addition, Voronoi diagrams have never been used

for content presentation and key frame extraction, which offers great interest from the point of competitiveness with other methods and models of this kind.

Throughout the thesis, the constructed model is defined in terms of several metrics for comparing Voronoi diagrams which present content in each frame from different points of view, namely color, texture and shape. Other frame characteristics, such as motion and area of segments, have not been taken into consideration because of some reasons which have been provided. The resulting metric has also been proposed, uniting frame characteristics for machine level understanding. Each of the metrics has been presented using Voronoi tessellations and salient points only. The model aims at providing full presentation and description of frame content. A new method based on the model has been described in details.

Various improvement measures have been designed for the method. The first one consists in enhancement of Harris method (for salient point selection) via k -means clustering that brings to invariance of initial placement of salient points as a result (only from the point of quality, but not time needed for a number of clustering iterations). The second one lies in stabilization of key frame descriptions with higher order Voronoi diagrams. The third one implies integration of shot boundary detection method into the proposed key frame extraction procedure. Voronoi diagrams of highest order turned out to be the most stable for changes in content. Highest order Voronoi diagrams constructed for frames from the same shot may look almost the same and can also be used for the purpose of shot boundary detection along with some other methods being analyzed. Frame size and resolution optimization and centroid-based clustering techniques have been studied to select the most reasonable solution for key frame extraction.

The obtained results have been verified on test samples taken from open source TRECVID collection, Internet Archive, Movie Content Analysis Project, Open Video Project. Among the test samples are several commercials and self-made high definition videos (shot at the city centre). Different estimators have been applied to check validity, performance and quality of the proposed model and method. The latter has also been compared with some of the existing methods used for key frame extraction. Scientific results have been adopted in an industrial enterprise ELTA, which decreased experts' time needed for endoscopic material processing and enabled to add key frames as images for the reports. Scientific results have been also testified by UkrSEPRO and adopted in academic activity at Kharkiv National University of Radio Electronics.

Keywords: video summarization, key frame, spatio-temporal segmentation, Voronoi diagram, salient point.

Формат 60x84/16. Ум. друк. арк. 0.9. Тир. 100 прим. Зам. 264-14.
Підписано до друку 05.09.14. Папір офсетний.

Надруковано з макету замовника у ФОП Бровін О.В.
61022, м. Харків, вул. Трінклера, 2, корп.1, к.19. Т. (057) 758-01-08, (066) 822-71-30
Свідоцтво про внесення суб'єкта до Державного реєстру
видавців та виготовників видавничої продукції серія ДК 3587 від 23.09.09 р.

СТИЛЬ ™
ИЗДАТ 
ТИПОГРАФИЯ
www.stil-izdat.com