

Министерство образования и науки Украины
Харьковский национальный университет радиозлектроники

На правах рукописи

МАШТАЛИР СЕРГЕЙ ВЛАДИМИРОВИЧ

УДК 004.932.2:004.93'14

МОДЕЛИ И МЕТОДЫ ТЕМПОРАЛЬНОЙ ОБРАБОТКИ ВИДЕО
ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА

05.13.06 – информационные технологии

Диссертация на соискание ученой степени
доктора технических наук

Научный консультант
Бодянский Евгений Владимирович
доктор технических наук, профессор

Цей примірник дисертації ідентичний
за змістом з іншими, що подані до
спеціалізованої вченої ради Д 64.052.08

Учений секретар спеціалізованої вченої ради Д 64.052.08

І.П. Плісс

Харьков – 2016

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
РАЗДЕЛ 1 АНАЛИЗ СОСТОЯНИЯ И ТЕНДЕНЦИЙ РАЗВИТИЯ МОДЕЛЕЙ И МЕТОДОВ ОБРАБОТКИ ВИДЕОПОТОКОВ. ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЙ.....	15
1.1 Ретроспективный анализ трендов развития поиска мультимедийной информации по контексту.....	15
1.2 Изучение методов выделения ключевых кадров и поиска дескрипторов изображений видеопоследовательностей	21
1.3 Анализ задач и методов темпоральной обработки видео.....	35
1.4 Постановка основных задач исследований.....	44
Выводы по первому разделу.....	48
РАЗДЕЛ 2 МОДЕЛИ СЕГМЕНТАЦИИ ВИДЕО И МЕТРИЧЕСКОГО ПОИСКА СЕГМЕНТОВ.....	50
2.1 Модели структуризации видеопоследовательностей для задач информационного поиска	51
2.2 Формализация процедур поиска сегментов с запросами ‘ad exemplum’.....	57
2.3 Построение областей элиминации с одной опорной точкой....	62
2.4 Анализ процедур поиска при двух опорных точках	68
2.5 Исследование областей элиминации первого порядка для конфигураций с m опорными точками	76
2.6 Общий случай построения областей элиминации.....	81
Выводы по второму разделу.....	91
РАЗДЕЛ 3 МЕТОДЫ КЛАСТЕРИЗАЦИИ И СРАВНЕНИЯ ДАННЫХ В ВИДЕОПОТОКЕ ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА ...	92
3.1 Фрагментные методы кластеризация визуальной информации.....	92
3.2 Свойства преобразований сегментаций изображений.....	105
3.3 Анализ метрических свойств множества сегментаций	114

	3
3.4 Построение характеристических векторов видеокадров.....	120
Выводы по третьему разделу.....	134
РАЗДЕЛ 4 МЕТОДЫ И МОДЕЛИ СЕГМЕНТАЦИИ ВИДЕО НА	
 ОСНОВЕ ИНДУЦИРОВАННЫХ ВРЕМЕННЫХ РЯДОВ.....	136
4.1 Метод темпоральной сегментации на основе настраиваемой	
модели.....	137
4.2 Метод обнаружения изменений свойств многомерных	
временных рядов на основе экспоненциального сглаживания..	149
4.3 Метод обнаружения изменений свойств многомерных	
временных рядов на основе анализа главных компонент.....	157
4.5 Методы темпоральной сегментации на базе адаптивного	
прогнозирования многомерных временных рядов.....	163
Выводы по четвертому разделу.....	179
РАЗДЕЛ 5 МЕТОДЫ АНАЛИЗА ОПОРНЫХ ДАННЫХ СЕГМЕНТОВ В	
 ЗАДАЧАХ ИНФОРМАЦИОННОГО ПОИСКА.....	181
5.1 Анализ моделей формирования ключевой информации	181
5.2 Сравнение кадров, идентифицируемых диаграммами	
Вороного.....	197
5.3 Процедуры поиска ключевых кадров	209
Выводы по пятому разделу.....	217
РАЗДЕЛ 6 МЕТОДЫ И МОДЕЛИ КОМПРЕССИИ ИЗОБРАЖЕНИЙ	
 ДЛЯ ТЕМПОРАЛЬНОЙ ОБРАБОТКИ ВИДЕО.....	218
6.1 Методы компрессии видеокадров на базе многослойных	
нейронных сетей.....	218
6.2 Метод сжатия визуальной информации с помощью 4-х	
слойной автоассоциативной нейронной сети.....	226
6.3 Анализ эффективности компрессии изображений для	
темпоральной обработки	236
6.4 Использование самоорганизующихся карт Кохонена при	
обработке кадров видеопоследовательностей	242

	4
Выводы по шестому разделу	253
РАЗДЕЛ 7 ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ТЕМПОРАЛЬНОЙ ОБРАБОТКИ ВИДЕО ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА.....	254
7.1 Особенности экспериментального анализа временных рядов, индуцированных видеопотоками.....	254
7.2 Имитационное моделирование признаковых пространств для задач поиска с запросами «по образцу».....	271
7.3 Исследование специфики решения поисковых задач структурированного видео.....	279
7.4 Особенности внедрения результатов диссертационных исследований.....	290
Выводы по седьмому разделу	293
ВЫВОДЫ.....	295
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	298
ПРИЛОЖЕНИЕ Акты о внедрении результатов диссертации.....	334

ВВЕДЕНИЕ

Актуальность темы. Диверсификация использования мультимедиа и гипермедиа акцентирует внимание на развитии информационных технологий, комплекса методов, способов и средств хранения, обработки, передачи и отображения информации, ориентированных на плохоструктурированные или совсем неструктурированные данные больших объемов. Видеопотоки являются яркими представителями таких неструктурированных данных. Минимизация их избыточности, обеспечение адекватности, достоверности, полноты, объективности, доступности и актуальности становятся ключевыми вопросами поиска компромисса между информационной потребностью и объемами и свойствами накапливаемых видеоданных. Возможность в условиях неопределенности извлекать необходимые сведения из видео – основная проблема релевантности визуальных информационных ресурсов, связанных с нелинейным анализом, многоаспектным автоматическим индексированием, аннотированием и реферированием, поиском с различными формами запросов, в частности, ‘ad exemplum’ – «по образцу». Семантический разрыв между низкоуровневыми признаками визуальной информации и требованиями высокоуровневой интерпретации требует проведения исследований, потенциальной базой которых служит учет пространственных (между объектами сцен) связей и связей между видеокадрами во времени. Полиструктурность видеопотоков, многоуровневое и многозначное содержательное значение их элементов ставят в ряд важнейших задачу поиска совокупности устойчивых связей, обеспечивающих целостность анализа и тождественность представления видео в целом. Структуризация группирует элементы видеоданных, управляет степенью их огрубления или детализации, определяет логические связи между выделенными группами, стратифицирует видеоданные. Структурирование видеопоследовательностей обеспечивает создание «устойчивого каркаса», на основе которого решаются задачи синтеза информационных технологий обработки, анализа, хранения и интерпретации видеопотоков. В

силу природы формирования и хранения видео особое значение приобретают и возможность, и необходимость on line обработки. Анализ видеопоследовательностей и/или индуцированных ими многомерных временных рядов в режиме наращивания данных с учетом предыстории различной глубины и краткосрочного прогнозирования, пожалуй, единственный естественный путь развития многофакторной междисциплинарной интеграции моделей и методов управления (в широком смысле) видео. В этом плане определенную перспективу имеет совершенствование инструментов управления коллекциями видео, и в первую очередь, системами семантического поиска CBVIR (Content Based Video Information Retrieval). Применительно к CBVIR основное влияние на развитие моделей и методов обработки и распознавания статической и динамической визуальной информации оказали украинские и зарубежные ученые С.Г. Антощук, Р.А. Воробель, Г.Н. Жолткевич, В.Н. Крылов, Д.Д. Пелешко, Е.П. Путятин, Т.Е. Рак, М.И. Шлезингер, Н. Burkhardt, R.M. Haralick, B.K.P. Horn, G. Liu, M. Petkovic, M. Sonka, P. Zezula и др. Несмотря на многочисленные исследования в данном направлении, научно-техническая проблема создания систем валидного семантического поиска динамической визуальной информации остается актуальной. Основной парадигмой развития информационного поиска в коллекциях плоскоструктурированных данных является такая его организация, когда первоначально исключаются из рассмотрения заведомо ненужные объекты, а детальный поиск осуществляется лишь среди оставшихся.

Непрерывный рост в различных проблемно-ориентированных областях объемов видеоданных, обладающих различными характеристиками приводит к необходимости создания нового инструментария темпоральной обработки видеопоследовательностей – средств on line структуризации видеопотоков на основе пространственной (внутрикадровой) и временной (межкадровой) сегментации видеоданных. Таким образом, разработка моделей и методов темпоральной обработки видео для информационного поиска является важной в концептуальном и прикладном аспектах научно-технической проблемой.

Связь работы с научными программами, планами, темами. Диссертационная работа выполнена на кафедре информатики Харьковского национального университета радиозлектроники в рамках государственных научно-исследовательских работ: «Синтез методов обработки информации в условиях неопределенности на основе самообучения и мягких вычислений» (№ ГР0107U003028, исполнитель), «Эволюционные гибридные системы вычислительного интеллекта с переменной структурой для интеллектуального анализа данных» (№ ГР0110U000458, исполнитель), «Модели и методы грануляции и интерпретации многомерных данных» (№ ГР0110U002636, ответственный исполнитель), «Модели и методы анализа и интерпретации динамических процессов» (№ ГР0113U000357, ответственный исполнитель раздела), «Нейро-фаззи системы для потоковой кластеризации и классификации последовательностей данных в условиях искажений отсутствующими и аномальными наблюдениями» (№ ГР0113U000361, исполнитель), «Разработка гибридных систем и методов вычислительного интеллекта для обработки потоков нечеткой информации в условиях нестационарности и неопределенности» (№ ГР0116U002539, ответственный исполнитель раздела), которые выполнялись в соответствии с приказами Министерства образования и науки Украины по результатам конкурсного отбора научно-исследовательских работ. В рамках указанных работ соискателем получены модели и методы темпоральной обработки видеопотоков, направленные на метрический поиск в больших коллекциях видео с запросами 'ad exemplum'.

Цель и задачи исследований. Целью диссертационной работы является решение научно-технической проблемы разработки моделей и методов темпоральной обработки видеопотоков для интеллектуальных информационных технологий поиска плохоструктурированных данных.

Для достижения поставленной цели потребовалось решение следующих задач:

– анализ тенденций развития методов обработки видеопотоков в аспекте их структурирования (выделения однородных сегментов), поиска валид-

ных представителей сегментов, дескрипторов видеокадров для выделения перспективных средств, влияющих на эффективность многоаспектного индексирования в библиотеках динамической визуальной информации и метрического поиска в CBVIR;

– синтез моделей методов сегментации видео, обеспечивающих создание процедур информационного поиска с запросами ‘ad exemplum’ на основе построения областей элиминации для произвольных конфигураций объектов в поисковом пространстве;

– разработка методов кластеризации и сравнения разбиений конечных множеств для повышения валидности семантического поиска в больших библиотеках видеоданных;

– построение моделей и методов on line сегментации многомерных временных рядов, индуцируемых видеопотоками, путем обнаружения изменения их свойств;

– создание методов поиска ключевых кадров, представляющих сегменты видеоданных в задачах поиска видео на базе механизмов поиска изображений в больших библиотеках статической визуальной информации;

– разработка методов и моделей компрессии видеокадров и видео для темпоральной обработки видеопотоков в задачах многоаспектной индексации;

– исследование специфики использования разработанных моделей и методов темпоральной обработки видеопоследовательностей, создание и внедрение исследовательских и специализированных программных средств.

Объект исследования – процессы обработки и анализа видеопотоков в мультимедиа информационных технологиях.

Предмет исследования – методы пространственной и временной сегментации видеопоследовательностей для метрической индексации и поиска в видеокolleкциях.

Методы исследования – при разработке и исследовании методов и моделей сегментации для индексации видеопотоков использованы основные

положения математического аппарата распознавания образов и обработки изображений, анализа временных рядов, искусственных нейронных сетей и идентификации систем, а также элементы математической статистики при проведении и анализе результатов экспериментальных исследований.

Научная новизна. Научная новизна диссертации состоит в постановке и решении научно-технической проблемы пространственной и временной сегментации видеоданных для индексации и информационного поиска в больших библиотеках видеопоследовательностей:

- впервые предложена пространственно-временная модель последовательной сегментации видео на основе отношений сходства / неразличимости видеок кадров, что позволяет определять границы сегментов с учетом семантики визуальной информации и решать задачи темпоральной обработки в online режиме;

- впервые получены обобщенные критерии элиминации объектов при метрическом поиске, позволяющие сократить время поиска с запросом «по образцу» за счет исключения трудоемких вычислений расстояний от запроса до анализируемых объектов;

- впервые получены методы фрагментной кластеризации изображений, которые обеспечивают локальную в поле зрения темпоральную сегментацию видеок кадров, что позволяет осуществлять анализ видеопоследовательностей в окрестностях пространственных областей интереса в реальном времени;

- усовершенствованы методы обнаружения изменения свойств многомерных временных рядов, индуцированных видеопотоками, которые на основе настраиваемых моделей, экспоненциального сглаживания и анализа главных компонент в отличие от известных методов дают возможность обнаруживать сегменты при последовательном анализе видеоданных в различных признаковых пространствах;

- получили дальнейшее развитие методы обнаружения изменения свойств многомерных временных рядов на основе адаптивного прогнозирования видеопоследовательностей, что в отличие от известных методов дает

возможность выделения сегментов видеопотоков с учетом различной предыстории входных данных;

– получили дальнейшее развитие методы формирования ключевых кадров видеосегментов, которые в отличие от известных используют разбиения видеокадров на базе диаграмм Вороного, что позволяет учитывать «пространственное содержание» видеокадров, повышая валидность процедур поиска;

– усовершенствованы интеллектуальные методы компрессии видеокадров, что в отличие от известных методов, обеспечивает темпоральную обработку видеопоследовательностей в плане их структуризации для информационного поиска.

Практическая значимость полученных результатов. Практическая значимость состоит в разработке методов и моделей сегментации видеопотоков, что создает предпосылки для многоаспектной индексации в задачах информационного поиска и обеспечивает повышение быстродействия поиска в системах СВВІR с запросами «по образцу».

Результаты диссертационных исследований использованы в Харьковской областной государственной администрации для системы управления архивами селекторных конференций с использованием видеосвязи (акт от 09.10.2014 г.), в ГУМВД Украины в Харьковской области (акт от 28.01.2015 г.) с целью детектирования значимых фрагментов видео, в Харьковском техническом центре рентгеновского сервиса «Спектр» (акт от 14.12.2014 г.) и на Заводе рентгеновского оборудования «Квант» (акт от 18.12.2014 г.) для анализа и выделения фрагментов в потоках рентгеноскопической видеoinформации.

Теоретические и практические результаты работы нашли применение в учебном процессе Харьковского национального университета радиоэлектроники при подготовке специалистов и магистров по специальности «Информатика» (акт от 18.05.2016 г.).

Результаты теоретико-экспериментальных исследований реализованы в

виде прикладных и исследовательских программных комплексов, один из которых «Программный комплекс пространственно-временной сегментации видеоданных в задачах поиска с запросом «по образцу» сертифицирован УкрСЕПРО.

Личный вклад соискателя. Все положения, выносимые на защиту, получены соискателем лично. В работах, написанных в соавторстве, диссертанту принадлежат постановка задач, выбор и обоснование их решения, а именно: [2] – разработаны модели и методы структуризации и сегментации видеопотоков; [10] – разработан метод кластеризации многомерных данных в условиях большой размерности; [11] – разработан метод сегментации видеопоследовательностей при помощи анализа текстур отдельных кадров; [12] – разработан матричный метод кластеризации изображений в больших базах данных; [15] – модифицирован метод J -средних для фрагментной кластеризации изображений в больших базах данных; [17] – предложена модификация метода X -средних для фрагментной сегментации изображений; [24] – предложен метод поиска изменений в потоке видеоданных; [28] – предложен нейросетевой алгоритм сжатия изображений; [29] – разработан метод сжатия изображений при помощи главных и независимых компонент; [32] – предложен метод описания видеоданных за счет критерия стабильности; [37] – предложен метод детектирования изменений в видеоданных на базе экспоненциального сглаживания; [47] – предложен подход к выделению характерных признаков изображений; [74] – выделен вектор характеристик, который позволяет сравнивать результаты информационного поиска; [79] – разработана нейросеть для прогнозирования многомерных временных рядов; [80] – разработан матричный метод адаптивного прогнозирования многомерных временных рядов; [81] – предложена матричная прогнозирующая модель, которую можно использовать для анализа видеоданных; [83] – предложен метод фрагментного сжатия изображений; [84] – разработан метод построения характеристического вектора признаков для изображений; [86] – предложен метод сжатия изображений, на основе информационного крите-

рия; [89] – предложен метод поиска изменений в видеоряде на базе VAR модели; [91] – предложен метод построения областей достаточности и необходимости, которые можно использовать при сравнении кадров видеопоследовательностей при поиске «по образцу»; [99] – предложен метод сравнения разбиений, который можно использовать при сравнении кадров видеопоследовательностей при поиске с запросом «по образцу»; [132] – предложен алгоритм самообучения карты Кохонена для сжатия изображений; [183] – разработан метод нормализации изображений на основе поиска геометрических свойств их отдельных элементов; [218] – предложен метод сравнения изображений при помощи анализа их сегментаций; [229] – разработан метод поиска инвариантных характеристик скелетонов бинарных изображений; [230] – предложен метод темпоральной сегментации видеопоследовательностей; [251] – предложен метод адаптивного прогнозирования, который можно использовать для темпоральной сегментации видеоданных; [256] – предложен метод детектирования ключевых кадров; [258] – модифицировано описание ключевых кадров при помощи диаграмм Вороного высших порядков; [259] – разработана модель и метод выделения ключевой информации в видеопоследовательностях; [267] – разработан и проанализирован нейросетевой метод анализа многомерных временных рядов; [270] – разработан on line метод темпоральной сегментации видеопоследовательностей; в [13] рассмотрены модели и методы структуризации видеопотоков; в [14, 16] исследованы фрагментные методы кластеризации при темпоральной обработке видеоданных; в [61, 69, 254, 255, 262] конкретизированы и изучены признаковые пространства и методы поиска характерных признаков; в [85, 87, 88, 200, 257] исследованы нейросетевые методы сжатия видеокадров для темпоральной обработки; в [90, 157] рассмотрены методы обнаружения изменений свойств многомерных временных рядов; в [94] исследован фрагментный метод прогнозирования видеорядов; в [217] исследованы модели и методы поиска с запросом «по образцу».

Апробация результатов диссертации. Основные результаты работы докладывались, обсуждались и были одобрены на научно-технических конференциях и семинарах: 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); Международной научной конференции «Глобальные информационные системы. Проблемы и тенденции развития» (Туапсе, 3-6 октября 2006); 5-th International Conference 'I.Tech, Information Research & Applications' (Bulgaria, Varna, 26-30 June, 2007); Международных научных конференциях «Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта» (ISDMCI'2008, 2009, 2014, 2015, 2016) (Евпатория, 19-23 мая 2008 г., Евпатория, 18-22 мая 2009 г., Железный Порт, 28-31 мая 2014 г., Железный Порт, 25-28 мая 2015 г., Железный Порт, 24-28 Мая 2016 г.); International Workshop 'Computer Graphics, Vision and Mathematics (GraVisMa 2009)' (Czech Republic, Plzen, 2-5 September, 2009); 54-th Internationales Wissenschaftliches Kolloquium (IWK) 'Information Technology and 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 'Applied and Theoretical Information Systems Research (ATISR)' (Taiwan, Taipei, 10-12 February, 2012); 12-th International Conference 'Pattern Recognition and Information Processing' (Belarus, Minsk, 28-31 May, 2014); I Международном научно-практическом форуме «Наука и бизнес» (Днепропетровск, 29-30 июня 2015 г.); Международной научно-технической конференции «Электротехнические и компьютерные системы: теория и практика» (Одесса, 20-24 июля 2015 г.).

П у б л и к а ц и и . Основные положения диссертации отражены в 52 работах, среди которых: 1 монография; 24 статьи в специализированных изданиях Украины по техническим наукам; 8 статей в зарубежных журналах (всего 18 публикаций в международных наукометрических базах); 19 публикаций в материалах международных научно-технических конференций и семинаров (3 в международных наукометрических базах).

Автор выражает глубокую признательность научному консультанту доктору технических наук, профессору Бодянскому Евгению Владимировичу за многочисленные рекомендации и обсуждения представляемых результатов.

РАЗДЕЛ 1

АНАЛИЗ СОСТОЯНИЯ И ТЕНДЕНЦИЙ РАЗВИТИЯ МОДЕЛЕЙ И МЕТОДОВ ОБРАБОТКИ ВИДЕОПОТОКОВ. ПОСТАНОВКА ЗАДАЧ ИССЛЕДОВАНИЙ

Одним из важных направлений развития моделей и методов, используемых при обработке мультимедийной информации, является построение систем, позволяющих осуществлять поиск по контексту медиаданных. Необходимость синтеза новых методов контекстного поиска плохоструктурированной информации обусловлена непрерывным ростом объемов и диверсификацией видеоданных, вызвана развивающимися возможностями систем регистрации, хранения, идентификации и анализа видео, стремлением получать автоматическую интерпретацию визуальной информации. В большинстве случаев в новых условиях существующие методы не в полной мере удовлетворяют предъявляемым требованиям по релевантности, надежности и быстродействию.

Раздел посвящен анализу современного состояния и трендов развития моделей и методов решения мультидисциплинарных задач проблемы информационного поиска видеоданных. Проанализированы существующие подходы и методы сегментации видеопотока. На основе ретроспективного анализа сформулированы основные задачи диссертационных исследований.

1.1 Ретроспективный анализ трендов развития поиска мультимедийной информации по контексту

Поиск и интерпретация информации разного рода всегда является одной из важнейших задач [36, 46, 147, 158, 192, 209, 278]. С развитием же информационных технологий исследователей все меньше стала удовлетворять возможность поиска информации только по текстовым запросам, базирующимся на корреляции поискового запроса и отбираемых данных. Соответ-

ственно стали развиваться подходы, основанные на контекстном поиске информации. В то же время с начала 90-х годов все большую важность стали приобретать мультимедийные данные, на первом этапе это были естественно изображения.

Направление Content-Based Image Retrieval (CBIR) в научной литературе появилось в 1992 году в работах Т. Като по автоматическому извлечению растровых изображений из базы данных на основе цветовых и текстурных характеристик. Становление CBIR как самостоятельной научной проблематики совпало с ростом количества задач искусственного интеллекта [5, 20, 104, 133, 140, 184, 228, 250, 252, 253, 312], компьютерного зрения [35, 36, 43, 46, 121, 122, 142, 147, 152, 260, 303], распознавания образов [38, 39, 50, 113, 116, 123, 124, 127, 131, 141, 150, 272] и других смежных областей [41, 49, 51, 56, 65, 66, 82, 112, 117, 125, 126, 249], что позволило использовать наработки в этих областях для создания различных алгоритмов, исходя из математического аппарата, методы которого модифицировались под особенности представления мультимедийных данных. В качестве наиболее известного примера следует указать разработку методов сжатия изображений и видео с потерями таких, как Jpeg, Mpeg.

В то же время эти подходы, хоть и учитывали специфику представления мультимедийных данных, все же не относились напрямую к CBIR, т.к. нельзя утверждать, что они основывались на контексте данных. К подходам решения задачи информационного поиска, учитывающим подобия изображений следует отнести разработанный в середине 90-х QBIC (Query By Image Content), разработанный IBM [285]. Следует отметить, что зачастую понятия CBIR и QBIC являются синонимами с точки зрения задачи информационного поиска.

В дальнейшем, в первую очередь в связи с развитием сети Internet, поиск в базах данных изображений приобрел исключительную важность. В последнее время стали выделять и отдельный класс задач, связанных с анализом сверхбольших баз данных (Big Data). В тоже время улучшение техноло-

гий по хранению и скорости доступа к данным позволило исследователям перейти от анализа исключительно изображений к анализу визуальной информации в целом, что привело к возникновению соответствующих направлений Content-Based Video Retrieval (CBVR), Content-Based Visual Information Retrieval (CBVIR).

Нельзя не отметить и тот факт, что возникло еще одно направление информационного поиска мультимедийной информации, которое основывалось на индексировании/аннотировании информации. Другими словами, происходила замена мультимедиа данных их текстовым описанием. Это направление получило название Concept-Based Image Indexing (CBII) также называемое 'Description-Based' или 'Text-Based' Image Indexing/Retrieval. Одними из первых работ в этом направлении являются работы Расмуссена [177, 286].

В области анализа видеоданных в середине 90-х наибольшее внимание было уделено поиску устойчивого метода обнаружения границ наблюдаемого объекта. Широкую популярность приобрел метод вычисления расстояния между гистограммами соседних кадров, который используется и сейчас (часто как составная часть более сложных подходов) [237, 243, 246].

В сравнении с работами раннего периода, отличительной чертой работ последнего десятилетия является многократно возросшее количество моделей представления мультимедийной информации.

В общем виде формализация описания контента мультимедийной информации сводится к обработке изображения и вычислению характеристик так называемых дескрипторов (как правило, численных характеристик). Однако это не означает переход к СВII, т.к. нет «словесного» описания, а используется только некоторый набор формализованных характеристик. В этом контексте информационный поиск описан как спецификация минимального набора инвариантных условий, которые моделируют намерения Пользователя и направлены на уменьшение влияния семантического разрыва, обусловленного лишь случайными искажениями, помехами т.д. [219, 294, 308].

При этом возникает еще одна проблема, связанная с выбором меры/метрики для осуществления информационного поиска. Широко распространенное евклидово расстояние дает далеко не лучший результат, и выбор метрики остается весьма непростой задачей, на решение которой направлены многочисленные исследования [82, 98, 260, 293].

В [219] предметная область поиска изображений была классифицирована по типу обрабатываемого контента на узкую и широкую. Следует отметить, что такая классификация нередко применяется и сегодня, оставаясь своеобразным разделителем подходов к созданию систем, реализующих поиск изображения по содержанию.

Изображения конкретной предметной области, как правило, имеют сравнительно небольшое разнообразие объектов различения и хорошо детектируемые визуальные характеристики, что упрощает постановку задачи CBIR. Изображения же широкой предметной области усложняют возможность генерализации критериев поиска. Принято считать, что проблема широких и узких предметных областей может быть решена с помощью нахождения новых метрик вычисления подобия [98, 260, 293].

В целом современные работы по CBIR, связанные с поиском изображений, можно разделить: на поиск по ассоциации (поиск подобного изображения, согласно некоторому критерию с порогом подобия/различия); целевой поиск (поиск конкретного изображения); категориальный поиск (когда отдельное изображение отвечает за целый класс искомых данных)[1, 4, 7, 27, 31, 60, 71, 95, 96, 102, 151, 187, 211, 220, 275].

С момента появления CBVR, т.е. с развитием роли видеоинформации, сместились и акценты в анализе изображений, т.к. кроме пространственного анализа, характерного для анализа статических мультимедиа данных добавился и временной компонент, что в свою очередь добавило целый ряд задач и возможных методов их решения. В частности, среди них следует отметить:

– выделение и распознавание движущихся объектов, в условиях действия различного рода помех и возмущений [18, 53, 55, 105, 113, 302];

- отслеживание траектории движения найденных объектов [33, 108, 114, 115, 119];
- детектирование движения на сцене [42, 48, 63, 134, 137, 316];
- обнаружение изменений сцены [23, 30, 40, 68, 77, 100, 118, 196, 299, 301, 305, 309].

Для решения этих задач разработано большое количество специализированных систем, которые осуществляют выделение изображений движущихся объектов на сложном фоне, фильтрацию помех, скоростную фильтрацию, отделение объектов от фона, оценку скорости каждого объекта, его идентификацию и сопровождение. Эти системы построены с применением различных методов и подходов, обычно работают с данными, поступающими в реальном времени.

Процесс проектирования таких систем представляет собой сложную технологическую последовательность, включающую в себя регистрацию цифрового изображения, его обработку с целью выделения значимой информации и анализ этой информации для решения определенной задачи. В настоящее время создается большое количество подобных систем, имеющих различное назначение, однако последовательность обработки видеосигнала в них практически одна и та же, а в основе лежат почти одни и те же модули. Условно обработку видео/изображений в таких системах можно разделить на следующие этапы [36, 234, 236, 260, 278, 300]:

- выделение областей переднего и заднего планов;
- выделение и классификация движущихся объектов;
- определение траектории найденных движущихся объектов;
- распознавание и описание трансформаций объектов, представляющих интерес.

Процесс выделения переднего плана заключается в отделении подвижных фрагментов изображения от неподвижных, которые называют фоновыми или принадлежащими заднему плану. Сложность этой задачи обуславливается большим количеством различных факторов таких, как движения камеры

(объектов), собственные шумы камеры, изменение метеорологической дальности видимости, смена освещения сцены, локальные искажения (падающие тени, движение веток деревьев на ветру и т.д.) и т.п.

На втором этапе выделения и классификации сначала осуществляется сегментация изображения переднего плана, то есть находятся компактные участки, движущихся с одинаковой скоростью объектов, которые считаются подвижными. Далее они соотносятся с обычно заранее определенными классами: автомобили, люди, животные и т.д.

Для осуществления определения траектории движущегося объекта (трекинга) необходимо установить взаимно-однозначное соответствие между найденными объектами на последовательных кадрах. На заключительном этапе осуществляется распознавание и описание действий объектов, которые были выделены.

После распознавания и классификации объектов становится возможным поиск информации в хранилищах данных, когда найденный объект используется в качестве критерия поиска. Такой подход является актуальным при слежении за объектами, монтаже телевизионных программ, когда необходимо в существующем архиве найти видеоданные, контекст которых соответствует заданному образцу (т.е. ключевые моменты того или иного видеосюжета содержат распознанный объект).

Большинство СВIR-систем осуществляют выделение и вычисление характеристик на этапе предобработки данных. Именно они выступают в качестве исходных данных для последующих задач анализа мультимедийной информации таких, как оценка сходства, аннотирование, распознавание образов.

Говоря об извлечении локальных характеристик (дескрипторов не всего изображения, а лишь некоторой его области – сегмента), приходим к важности процесса сегментации изображений. Без надежной сегментации невозможно должным образом найти локальные дескрипторы, а значит, при формализации описания изображения возникает так называемый семантический

конфликт [6, 34, 156, 195, 203, 232, 235, 242]. Безусловно, предпринимались попытки снизить зависимость от валидности сегментации. Например, включать в результирующий набор признаков все полученные сегменты и использовать так называемые гибкие метрики или же блочные двумерные модели Маркова для описания пространственного распределения цвета и текстуры [70, 156]. В качестве еще одной возможной стратегии можно указать использование перцептивных принципов группировки для иерархического выделения структуры изображения [203, 232]. Однако проблема ликвидации семантического разрыва между низкоуровневыми признаками и высокоуровневой интерпретацией все еще остается открытой.

Таким образом, можно сделать важный промежуточный вывод. Перспективным направлением решения мультидисциплинарной проблемы создания эффективных систем CBVIR является разработка теоретического инструментария темпоральной *on line* обработки видеопоследовательностей, обеспечивающей структуризацию данных для технологии информационного поиска и включающей в себя модели и методы пространственной и временной сегментации видеорядов. В силу прелиминарной неструктурированности, широкой вариативности данных и практически полной априорной неопределенности необходимо иметь целые наборы моделей и методов, которые можно использовать или параллельно, или адаптировать к конкретным предметно-ориентированным областям.

1.2 Изучение методов выделения ключевых кадров и поиска дескрипторов изображений видеопоследовательностей

Первым этапом контекстного анализа видео является разбиение видеопоследовательности на сегменты, т.е. последовательности изображений, неразличимых с точки зрения заданного критерия. Сегменты рассматриваются как наименьшая единица обработки или индексации видео, когда не воспринимаются никакие изменения в содержании сцены, а концепции интер-

претации более высокого уровня строятся путем анализа межсегментных отношений [175, 204, 219, 236, 278, 281, 287, 291, 301, 306, 309, 312].

Ключевые кадры – статические изображения, извлеченные из оригинальных видеоданных и наилучшим образом представляющие содержание сегментов [174, 194, 224, 226, 244, 268, 318]. Ключевые кадры часто используются в качестве дополнения описания видео и являются некоторым визуальным описанием содержания видео для быстрого интерактивного просмотра и (или) поиска в видео. Но в основном они используются для представления видеопотоков в поисковой системе, и именно на визуальных особенностях ключевых кадров может быть реализована поисковая система видеоданных с использованием инструментов CBIR [168, 171, 179, 192, 201].

Разработанные на сегодняшний день методы анализа ключевых кадров видео можно условно объединить в группы, приведенные в таблице 1.1.

Таблица 1.1 – Классификация методов поиска ключевых кадров

Название группы методов	Авторы (год публикации)
Извлечение граничных элементов	A. Nagasaka, Y. Tanaka (1991)
Факторизация матриц	Y. Gong, X. Liu (2000), M. Cooper, J. Foote (2002)
Упрощение кривой	S. Lim, D. Thalmann (2001), K. Matsuda, K. Kondo (2004), E. Bulut, T. Capin (2007)
Кластеризация	L. Li, X. Zhang, Y. Wang, W. Hu, P. Zhu (2008), Z. Qu, L. Lin, T. Gao, Y. Wang (2013)
Модель зрительного внимания	J. Peng, Q. Xiaolin (2010), L.J. Lai, Y. Yi (2012), Q.-G. Ji, Z.-H. Xie, Z.-D. Fang, Z.-M. Lu (2013)
Генетические алгоритмы	X. Yang, Z. Wei (2011)
Разность гистограмм	B. Liang, W. Xiao, X. Liu (2012), G.I. Rathod, D.A. Nikam (2013)
Статистики	J. Almeida, N.J. Leite, R.S. Torres (2012), S.S. Kanade, P.M. Patil (2013)
Нейросети	D.P. Papadopoulos, V.S. Kalogeiton, S.A. Chatzichristofis, N. Papamarkos (2013)
Онтологии	K. Khurana, M.B. Chandak (2013)

Для первых методов поиска ключевых кадров, разработанных еще в 90 годах прошлого века, было характерно осуществлять поиск по граничным элементам сегментов. Так, в [265] предложено извлекать первый кадр каждого сегмента. Несмотря на простоту такого подхода, число ключевых кадров всегда зависит от числа сегментов, к тому же кадры извлекаются независимо от важности их содержания. Аналогичные подходы фиксируют заданное число кадров, распределяя их равномерно по всей длине видеопоследовательности.

Позднее стали появляться методы, учитывающие движение в качестве пространственно-временного признака [178, 204, 230, 241, 248, 281, 297, 315]. Методы оптического потока, как правило, используются для поиска локального минимума движения. Такое решение является очень сложным с вычислительной точки зрения, к тому же, хорошие результаты достигаются только при наличии значительного движения в видео.

При поиске ключевых кадров нередко объединяется движение объектов с визуальными признаками. Например, выполняется поиск границ сегмента путем анализа изменений характеристик цвета, а ключевым кадром считается тот, в котором наблюдается явное ускорение или замедление движения. Иногда число ключевых кадров устанавливается в зависимости от шаблона движения, используемого в анализируемом видеоматериале. Такой подход не привязан к установке пороговых значений и является достаточно быстрым, но может применяться только для ограниченного набора типов видеопоследовательностей [297, 301, 311].

Еще одной группой методов, которая стала применяться сравнительно давно для поиска ключевых кадров, является факторизация матриц. Так, в [207] предложено кластеризовать видеопоток, а разложение по сингулярным числам матрицы дает переупорядоченное пространство признаков, облегчающее классификацию. Этот метод позволяет устранить повторы в результирующей последовательности кадров, так как каждый ключевой кадр отбирается из кластера с аналогичными изображениями. Вместо границ сцен вни-

мание смещено на содержание, хотя выбор первого ключевого кадра должен осуществить Пользователь из набора кадров, предложенных ему без учета содержания. В качестве признаков здесь используются только цветные гистограммы, что может привести к неверному отбору кадров (когда кадры с близким распределением цветов и совершенно разным содержанием будут считаться подобными).

Методы кластеризации стали использоваться для поиска ключевых кадров на рубеже тысячелетия и до сих пор занимают по популярности лидирующие позиции [50, 97, 205, 212, 311, 313]. Основной проблемой методов кластеризации традиционно является то, что число кластеров, как правило, должно задаваться априорно.

Особо следует отметить метод [226], который предполагает расчет матрицы сходства, после чего определяются кластеры с последующим поиском и извлечением ключевых кадров. Уникальность этого метода состоит в том, что преодолено ограничение по выбору одного кадра на кластер. Количество кадров, отбираемых из одного кластера, зависит от продолжительности и сложности содержания сцены. Кроме того, данный метод является простым с вычислительной точки зрения по сравнению с другими методами поиска ключевых кадров на основе кластеризации. Что касается недостатков, вывод о значимости кадров делается исходя из предположения, что на значимых сценах дольше фокусируется камера. Когда ключевые кадры отбираются из длинных последовательностей кадров в кластере, средний кадр каждой последовательности считается значимым, что напоминает самые ранние подходы к поиску ключевых кадров [162, 174, 193, 194, 202, 224, 226, 292, 311, 318, 319].

Другая группа методов основывается на анализе некоторых временных функций. Такие методы позволяют представить видеопоследовательность в виде кривой из многомерного признакового пространства. Значимость каждого кадра может определяться с учетом порогового значения в локальных максимумах и минимумах кривой [170, 206]. В последнее время стали попу-

лярными и некоторые другие простые методы, например, анализ разности цветовых гистограмм между соседними кадрами. При выборе подобных методов, однако, надо учитывать, что кадры с подобными цветовыми гистограммами могут иметь разное содержание [237, 287]. Кроме разности кадров, для определения кандидата в ключевой кадр может определяться средняя разность внутри группы кадров, входящих в состав одного сегмента. Для этих целей часто используются методы классической или нечеткой кластеризации [50, 59, 97, 191, 205, 212, 298, 311, 313].

Для повышения степени учета семантики нередко делается акцент на добавление для объектов текстовых меток (аннотаций). В этом случае Пользователи вовлечены в процесс разметки, принимая или отрицая соответствие ключевых слов объектам, что существенно сужает область применения. К тому же, определяя высокоуровневые метки, невозможно предусмотреть все изменения содержания, которые могут иметь место в видеопоследовательности. Несмотря на то, что в семантической концепции поиска ключевых кадров уже достигнуты определенные успехи, корректная машинная интерпретация видеоданных все еще остается очень сложной задачей из-за необходимости вовлечения Пользователей [201, 227, 242, 275, 294].

Сравнительно недавно для поиска ключевых кадров были предложены методы на основе модели зрительного внимания. Они предполагают генерацию «карты важности» или так называемой «кривой внимания» по данным об изменениях фона и объектов переднего плана. Для извлечения ключевых кадров стали применяться генетические алгоритмы и нейронные сети и, однако авторы этих методов в своих работах предупреждают о необходимости многочисленных модификаций, которые нужны для повышения качества отбора значимых кадров [20, 153, 176, 180, 197, 210, 214, 215, 233, 266, 296]. На рис. 1.1 приведены основные подходы к поиску ключевых кадров и признаков, используемых при классификации.

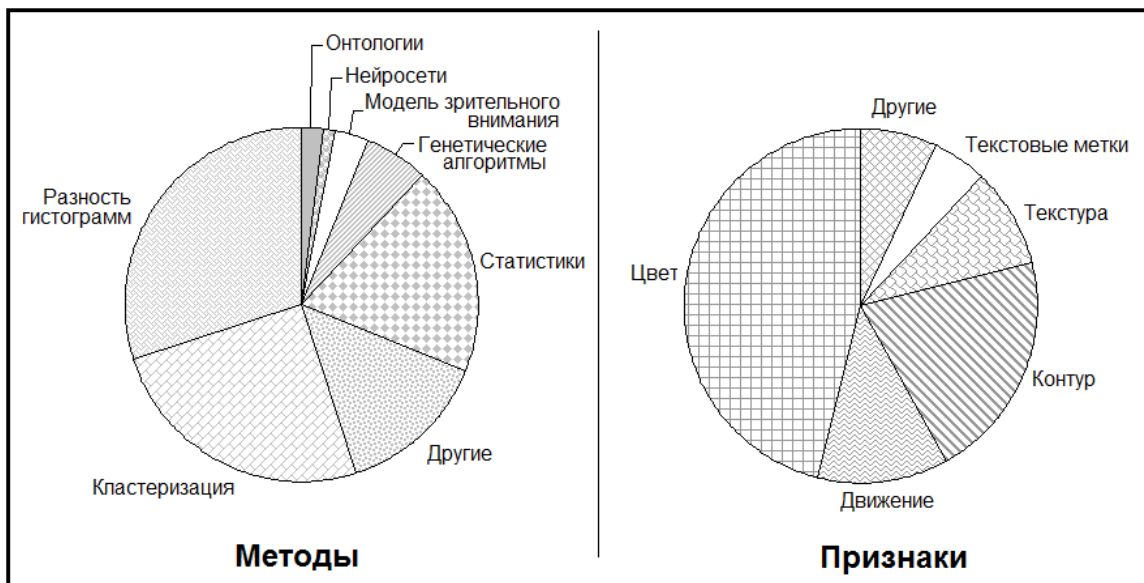


Рисунок 1.1 – Наиболее распространенные методы и признаки, используемые для поиска ключевых кадров

Несмотря на разнообразие существующих решений, по-прежнему сложно найти хорошее статическое представление видеопоследовательности. Это связано с различными источниками видео, размерами, качеством, используемыми камерами, углами съемки, наличием шума, с самим содержанием и целью отбора ключевых кадров. Ни один из существующих методов не справляется со всеми задачами одновременно. Решение одной проблемы непременно ведет к появлению другой. Кроме того, практически во всех этих методах, число ключевых кадров задается априорно, а кадры выбираются из предварительно сегментированного материала, что может привести к высокой корреляции ключевых кадров или пропуску действительно значимых кадров.

Рассмотрим подробнее широко используемые на практике методы, основанные на разности гистограмм, которые обобщенно можно представить следующим образом. Отметим, что они, как правило, основаны на использовании глобальных характеристик изображения, поэтому являются достаточно устойчивыми к движениям камеры и объектов. Кроме того, гистограммы инвариантны к незначительному вращению и изменению масштаба. Однако совершенно разные изображения могут иметь практически идентичные гисто-

граммы, поскольку они не учитывают информацию о расположении пикселей.

Простейшие гистограммные методы вычисляют разницу между гистограммами двух кадров. Более сложные методы используют при этом весовые коэффициенты. В наиболее развитых методах определяется оператор пересечения между гистограммами, или используют меры сходства типа EMD (Earth Mover's Distance) [165, 171, 192, 243, 278, 300].

Самый простой метод основан на вычислении следующего выражения:

$$d(t, t+1) = \frac{V}{a} \sum_{u=1}^V |H(I_{t+1}, u) - H(I_t, u)|,$$

где H – гистограмма интенсивностей изображения,

V – количество уровней квантования гистограммы.

Для цветных изображений расстояние между гистограммами вычисляется аналогично, или как максимум суммы разностей интенсивностей.

Достаточно большое распространение получил алгоритм двойного сравнения. Основу метода составляют операции сравнения с двумя порогами: большим d_h и меньшим d_r . На первом этапе используется d_h для определения резких переходов, на втором – d_r для определения возможного начала плавного перехода. Как только $d(t, t+1)$ превышает d_r , начинается вычисление суммы разностей d по всем следующим кадрам до тех пор, пока разница для двух последовательных кадров не будет меньше d_r .

Для повышения эффективности извлечения сегментов можно использовать один из уже традиционных подходов, а именно – фрагментные методы с последующим (возможно пирамидальным) анализом блоков. Данные методы в отличие гистограммных учитывают пространственную неоднородность изображений, а также могут применяться для распараллеливания с целью сокращения времени вычислений. После того, как входное изображение разби-

то на блоки, к каждому из них можно применить методы, приведенные на рис. 1.1. Основным преимуществом блочных методов является локальность анализа потенциальных областей интереса, их относительная нечувствительность к шумам и движениям камеры и объектов [95,108, 152, 123].

Обычно блоки, имеющие одинаковое расположение в последовательных кадрах, сравниваются попарно между собой. Мера сходства основана на использовании среднего значения и дисперсии, полученные на всем изображении. Степень сходства для блока b из двух кадров в моменты времени t и $t+1$ нередко определяется как

$$d(I_t, I_{t+1}, b) = \frac{\frac{1}{\sigma} (s_{t+1,b}^2 + s_{t,b}^2) + ((m_{t+1,b}^2 + m_{t,b}^2)/2)^2 \frac{1}{\sigma}}{s_{t+1,b}^2 s_{t,b}^2}$$

где $m_{t,b}$ и $s_{t,b}$ – соответственно среднее значение и дисперсия значений пикселей в блоке b в момент времени t .

После этого величину степени сходства сравнивают с пороговым значением

$$d_D(I_t, I_{t+1}, b) = \begin{cases} 1, & d(I_t, I_{t+1}, b) > d_D \\ 0, & d(I_t, I_{t+1}, b) \leq d_D \end{cases}$$

а граница сегмента фиксируется в том случае, если

$$\sum_{b=1}^B d_D(I_t, I_{t+1}, b) c_b > d$$

где B – общее количество блоков, а коэффициенты c_b используются для предоставления разного веса отдельным блокам.

Для увеличения надежности детектирования границ сегментов возможно применение комбинированных подходов. В этом случае одновремен-

но используется несколько методов различного типа и более сложные дескрипторы [38, 45, 62, 180, 245, 247, 292].

На рис. 1.1 приведены основные категории дескрипторов, которые в самом общем случае отражают визуальные свойства изображения, либо глобальные – для всего кадра, либо локальные – для небольшой группы пикселей изображения (цвет, форма, особые точки). Преимуществом глобального дескриптора является высокая скорость обработки как подсчета характеристик, так и вычисления меры подобия. Однако глобальные дескрипторы используются реже локальных, поскольку они часто огрубляют признаковое описание и чувствительны к геометрическим трансформациям объекта в поле зрения (по сути, за счет изменения фона переднего и заднего планов).

При использовании локальных дескрипторов набор характеристик вычисляется для каждого пикселя, используя его окрестность (например, среднее значение цвета в этой окрестности), или некоторой области – в частности блока.

В середине 90-х годов большое внимание уделялось алгоритмам обнаружения границ, основанных на анализе гистограмм: узкая гистограмма указывает на низкий контраст, а бимодальная предполагает наличие областей разной яркости. Обычно учитывались: среднее, дисперсия, коэффициент асимметрии, коэффициент эксцесса, энергия, энтропия. Применение этих характеристик для текстурного анализа рассматривается в [211, 244].

Одним из самых ранних использований цветowych гистограмм для индексирования изображений был метод Свона-Баларда (1991). Использование подходов, основанных на вычислении дескрипторов, в QBIC, Pictoseek, VisualSEEK и т.п. является значительным шагом в развитии CBIR и говорит о становлении основной методологии [285].

Инновационный подход к анализу цветовой информации был описан Финлэйсоном в 1996 г. Он предложил учитывать тень и зеркальное отражение объектов, а также особенности формы для того, чтобы различать изменения объектов в одном цветовом диапазоне.

В 1999 г. Ханг предложил использовать цветные кореллограммы вместо гистограмм, что позволяет учитывать пространственное распределение цвета. Вейвлет-преобразование Добеши показало хорошие результаты при извлечении цветных характеристик фона (этот подход нашел применение в системе WBIS), и акцентировало внимание исследователей на возможном решении проблемы семантического конфликта [232, 306].

Говоря о колориметрии, следует упомянуть о такой характеристике, как формула цветного отличия (Color difference). Распространенные определения цветного различия обычно используют формулу вычисления расстояния в евклидовом пространстве, однако стоит заметить, что не любое цветовое пространство является евклидовым. В [237] предлагается метод поиска по контексту с использованием гистограмм цветного отличия. Особенность заключается в том, что в качестве значений гистограммы используется не просто количество пикселей одного цвета, а количество перцептивно-однородных цветных отличий между двумя точками. Метод устойчив к фоновым различиям и работает в цветовом пространстве *Lab*.

Другим большим блоком возможных дескрипторов являются те, которые используют информацию о текстуре изображений. Хотя текстуры присутствуют в изображениях повсеместно, единого подхода к их описанию и строгого определения не существует. Как следствие, методы различения текстур, как правило, разрабатываются отдельно для каждого конкретного случая. В целом дается следующее определение текстуры: «пространственная организация элементов в пределах некоторого участка поверхности». Эта организация обусловлена определенным статистическим распределением интенсивности тонов. Иными словами, участок может считаться текстурным, если количество отмечаемых на нем перепадов интенсивности или изменений цвета достаточно велико.

Перед распознаванием текстур определяется размер скользящего окна, поскольку от этого зависит, какие свойства фоновых объектов будут содержаться в полученных дескрипторах. Известно [53, 121–123, 142, 303], что в

окнах размером 3×3 или 5×5 пикселей статистические текстурные меры больше действуют как обнаружители перепадов яркости, а не как измерители текстуры, хотя при этом сокращается время вычислений. Слишком большие размеры окон могут исказить результаты из-за влияния краев структур и границ изображений. Однако большое окно позволяет достигнуть высокой статистической достоверности. Как оказалось, окна 20×20 пикселей наиболее эффективны для текстурной обработки аэрокосмических фотографий сельскохозяйственных угодий, пастбищ, лесных массивов и других сходных объектов. При изменении размеров окна от 80×80 до 20×20 пикселей численные значения текстурных признаков изменялись на 5...10 %. Дальнейшее изменение размера окна приводит к значительному искажению текстурных признаков.

Текстурные признаки можно разделить на группы [142, 147, 152, 303]:

- признаки, основанные на измерении пространственных частот;
- признаки, основанные на статистических характеристиках уровней интенсивности элементов разложения;
- признаки, основанные на описании структурных элементов.

Интересные исследования об использовании текстурных характеристик в вейвлет-преобразованиях описаны в [4, 59, 60]. Джафари-Хоузани и Солтаниан-Задех в 2005 г. предложили использовать для описания дескриптор текстуры, вычисляемый на основе преобразования Радона.

Кроме цвета и текстуры одним из направлений исследований связанные с анализом геометрических характеристик, основанных на понятии формы и контура. Форма – ключевой атрибут сегментированного участка изображения, так как без ее устойчивого представления качественное распознавание невозможно. Говоря о распознавании форм, подразумевают способы и методы сопоставления дескрипторов. Для дескрипторов контура возникает проблема получения характеристик границ, для чего может быть использована полигональная аппроксимация, что ведет к упрощению контура и элиминации шума [72, 73, 196, 204, 287, 303].

Дескрипторы области не зависят от контура, однако содержат данные обо всех пикселях, содержащихся в области, что может быть избыточным, исказить значимую информацию. Возможными характеристиками, вычисляемыми по области, могут быть: центр тяжести, мера инерции, энергия, эксцентриситет, коэффициент кривизны отдельных сегментов изображения, эллиптические дисперсии, число Эйлера и т.д. Себастьян и Бартолини в 2005 г. предложили способ описания формы с помощью графа столкновений и преобразования Фурье соответственно.

Следует отметить, что по-прежнему большее внимание уделяется поиску дескрипторов, инвариантных к аффинным преобразованиям. Так, в 1998 г. Линдерерг предложил «шкалу уровней интереса», которая основана на определении максимума функции Лапласиана-Гауссиана. Позднее, в 2004 г., Миколайчук и Шмидт показали, что эта функция дает хорошие результаты в области автоматического выбора структуры.

Для формального представления изображений формируются сигнатуры изображения, как правило, представляемые вектором. Наиболее часто используемые сигнатуры – гистограммные и области – могут быть рассмотрены как набор взвешенных векторов. Распределения, полученные на основе коллекции векторов локальных характеристик, могут быть представлены в виде непрерывной функции плотности, или даже пространственной стохастической модели [225].

Несколько дескрипторов, основанных на цветовых, текстурных характеристиках и характеристиках формы, описаны в [240, 245]. Следует отметить, что они использовались в стандарте MPEG-7 и достаточно адекватны для описания видео и естественных изображений. При этом следует более подробно остановиться на этом стандарте, т.к. он подразумевает анализ контекста видеоданных.

MPEG-7 формально называется «Мультимедиа-интерфейс для описания содержимого» (Multimedia Content Description Interface) и имеет целью стандартизовать описание мультимедийного материала, поддерживающего

некоторый уровень интерпретации «содержания» информации. Стандарт не ориентирован на какое-то конкретное приложение, он формализует элементы, рассчитанные на поддержку широкого круга приложений [240].

MPEG-7 является стандартом ISO/IEC, разработанным MPEG (Moving Picture Experts Group), комитетом, который разработал также стандарты MPEG-1 (1992), MPEG-2 (1995), и MPEG-4 (версия 1 в 1998 и версия 2 в 1999). Стандарты MPEG-1 и MPEG-2 позволили производить широко распространенные продукты, такие как интерактивные CD, DVD, цифровое широкополосное аудио (DAB), цифровое телевидение, и многие другие коммерческие услуги. MPEG-4 является первым реальным мультимедийным стандартом для представления данных, позволяющим интерактивно работать с комбинациями натурального и синтетического материала, закодированного в виде объектов (он моделирует аудио-визуальные данные, как комбинацию таких объектов). MPEG-4 предоставляет стандартизованные технологические элементы, допускающие интеграцию производства, распределения и доступа к мультимедийному материалу. Это относится к интерактивному и мобильному мультимедиа, интерактивной графике и цифровому телевидению.

MPEG-7 предлагает полный набор аудиовизуальных средств описания, которые образуют базис для приложений, делая возможным высококачественный доступ к мультимедийному материалу, что предполагает хорошие решения для записи, идентификации материала, обеспечения прав собственности, и быстрой, эргономичной, точной целевой фильтрации, поиска.

MPEG-7 сконструирован так, чтобы учесть все подходы, учитывающие требования основных стандартов, таких как, SMPTE Metadata Dictionary, Dublin Core, EBU P/Meta, и TV Anytime. Эти стандарты ориентированы на специфические приложения и области применения, в то время как MPEG-7 достаточно универсален. MPEG-7 использует также схему XML в качестве языка выбора текстуального представления описания материала. Главными элементами стандарта MPEG-7 являются:

– Дескрипторы (D). Представление характеристик, которые определяют

синтаксис и семантику представления каждой из характеристик.

– Схемы описания DS (Description Scheme), которые специфицируют структуру и семантику взаимодействия между компонентами. Эти компоненты могут быть дескрипторами и схемами описания.

– Язык описания определений DDL (Description Definition Language), позволяющий создавать новые схемы описания и, возможно, дескрипторы.

Системные средства служат для поддержки мультиплексирования описаний, синхронизации описаний и материала, механизмов передачи, кодовых представлений (как текстовых, так и двоичных форматов) для эффективной записи и передачи, управления и защиты интеллектуальной собственности в описаниях MPEG-7.

Подводя некоторый промежуточный итог, резюмируем: признаки, подразделяющиеся на структурные, статистические и дескрипторы, основанные на моделях или преобразованиях, являются базой для формирования многомерных пространств описаний ключевых кадров, видеосегментов и видеопотоков в целом.. К структурным признакам относятся характеристики текстур, обычно «интегрирующих» анализ текселей (микротекстур), дескрипторы границ, линий, углов, точек и областей. Статистические признаки базируются на свойствах изображений в контексте одномерных (на границах, линиях, в заданном направлении) и двумерных (в поле зрения, на фрагментах, областях) распределений, гистограммных описаниях. Отдельно выделяются моменты изображений, обеспечивающих обработку в условиях геометрических преобразований (изменений масштаба, поворотов, смещений, действия группы евклидовых подобий). Признаки, основанные на моделях (фрактальных, Марковских случайных полях, других стохастических моделях) позволяют создавать синтетические признаковые пространства, которые могут эффективно использоваться в задачах классификации. Вейвлеты, преобразования Фурье, Зернике, Меллина, Уолша, Адамара, фильтры Габора и целый ряд других относятся к базовым трансформациям, также обеспечивающим формирование признаковых пространств [4, 8, 27, 38, 43, 71, 123, 142, 150, 152,

211, 260, 300, 303].

Таким образом, можно утверждать, что применительно к СВВIR определенную перспективу для сокращения семантического разрыва между низкоуровневыми описаниями изображений и стремлением к высокоуровневой интерпретации имеют модели и методы, базирующиеся на некоторых «совокупных» представлениях видеок кадров. Среди подобных характеристик следует выделить сегментацию (кластеризацию) изображений [36, 43, 142, 158, 175, 205], когда разбиения поля зрения целостно характеризуют «пространственное» содержание. Такой подход приемлем и для выделения ключевых кадров – характеристик видеосегментов, получаемых в результате темпоральной сегментации, и при продуцировании многомерных временных рядов на основе различных признаков.

1.3 Анализ задач и методов темпоральной обработки видео.

Исследования в области поиска информации на основе видеоконтента сегодня являются одними из активно развивающихся направлений интеллектуальных информационных технологий [66, 158, 161, 165, 179, 208, 219, 234, 278, 312, 314]. Представление характеристик, извлеченных при индексации кадров или видеосегментов, хранится в базе данных и используется для поиска видео [192, 210, 236, 265], а само множество исходно неструктурированных видео хранится в некоторых коллекциях.

При поиске видеоданных [66, 158, 161, 179, 208, 234, 278] в качестве запросов могут использоваться эскизы, признаки всего изображения или отдельных объектов, ключевые слова, любая комбинация перечисленного. Однако наиболее общий (и в то же время наиболее сложный) вид запроса – ‘*ad exemplum*’, т.е. запрос «по образцу». Естественно, что аппарат низкоуровневых признаков, непосредственно извлекаемых из видео [278, 290], к настоящему времени разработан наиболее полно [138, 202, 211, 231, 235, 317, 321], но семантический конфликт и трудности формирования и восприятия запро-

са, контроля адекватности манипуляций и требуют развития информационных технологий поиска с запросами ‘ad exemplum’ [171, 187, 192, 208, 219, 278, 312, 314], тем самым фаза поиска использует видеофрагменты в качестве запросов [106, 175, 295, 314]. Иначе говоря, структуризация и данных в коллекциях видеопотоков, и собственно запроса связана с темпоральной обработкой непосредственно видеорядов или многомерных временных рядов, индуцированных признаками видеок кадров [34, 42, 158, 175, 178, 191, 234, 295].

В таблице 1.2 представлены некоторые классы алгоритмов и их ориентация на особенности видеоматериала.

Таблица 1.2 – Сравнение методов анализа видеопотоков

Задача	Методы решения	Используемые характеристики	Тип приложений
Цветовая/текстурная классификация	Сегментирование кадров	Цвет, текстура	Визуальное восприятие
Методы мультимодального контекстного поиска	Поисковые системы	Выделение ключевых кадров, анализ формы	Поиск объектов
Выделение текстов в видео	Выделение объектов интереса в видео	Сегментация кадров	Поиск объектов
Семантическая суммаризация	Анализ баз видео для получения оптимальных вариантов просмотра	Матрицы смежности	Спортивное видео
Семантический анализ аудио	Автоматическая аудио категоризация	Аудио	Клипы, видео конференции
Автоматическое выделение сцен	Сравнение изображений	Вейвлеты, цветовые дескрипторы	3D сцены в фильмах

Для решения задачи обнаружения изменения свойств временных рядов сложился целый ряд относительно независимых подходов, среди которых выделяются статистический (корреляционный, спектральный, регрессионный анализы), адаптивный (экспоненциальное сглаживание, анализ следящего сигнала) и подход, основанный на методах классификации (кластеризация, сегментация, прогнозирование). Если для одномерных рядов имеется целое множество эффективных методов и алгоритмов [3, 26, 52, 67, 77, 107, 149, 209, 222, 225, 263, 271], то результаты для многомерного случая, особенно в on line режиме, все еще далеки от своего окончательного решения.

Вообще говоря, когда объектом анализа является многомерный временной ряд $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$, где $x_i \in \mathbb{R}^P$, $1 \leq k \leq N$ – дискретное время, в моменты которого производятся наблюдения, требуется найти сегменты $S(a, b) = \{a \leq k \leq b\}$, представляющие собой статистически однородные последовательности $x(a), x(a+1), \dots, x(b)$, а задача сегментации состоит в нахождении c непересекающихся сегментов $S^c = \{S_i(a_i, b_i), 1 \leq i \leq c\}$ таких, что $a_1 = 1, b_c = N, a_i = b_{i-1} + 1$.

Анализ многомерных временных рядов осуществляется на основе контроля некоторых обобщенных характеристик (спектров, корреляционных функций, средних, дисперсий, главных компонент и т.п.), для чего существует хорошо отработанный статистический инструмент. Но когда объем выборки N не фиксирован, количество сегментов c априори неизвестно, а данные $x(k)$ поступают на обработку последовательно в реальном времени – именно такова задача темпоральной обработки видеоданных – задача существенно усложняется, и требуются новые решения.

Для решения задачи информационного поиска представим многомерный временной ряд F , являющийся, по сути, видеопоследовательностью, в виде

$$F = \{B_i\}_{i=1}^N.$$

где $B(i)$ – i -й кадр видеопоследовательности,

$B(i) = (b_1(i), b_2(i), \dots, b_n(i))^T$, $b_j(i)$ – некоторые характеристики кадров или же компоненты многомерного временного ряда.

Тогда можно перейти к анализу временных последовательностей.

Вполне объяснимо, что наиболее компактной формой описания многомерного временного ряда является его математическая модель. Наиболее широкое распространение получили: модели авторегрессии-скользящего среднего (AR – Auto-Regressive, MA – Moving-Average, ARMA – Auto-Regressive Moving-Average, ARIMA – Auto-Regressive Integrated Moving-Average). Известен ряд их модификаций типа CARMAX (Control Auto-Regressive Moving-Average Model with auXiliary inputs, а также CARMA, CARIMA, CARIMAX), SSM (State-Space Models), SET (Spectral Expansion Technique) [26, 155, 173, 182, 209, 222, 225, 232, 310]. Проанализируем математические модели многомерных нестационарных временных рядов, ориентированные на обнаружение моментов изменения основных характеристик.

Для описания многомерных последовательностей достаточно часто используются модели [21, 107, 221, 225] в двух основных формах: структурной

$$\sum_{l=0}^N K_l B(i-l) + Dz(i) = h(i) \quad (1.1)$$

(здесь K_l – матричные коэффициенты при внутрисистемных (эндогенных) переменных, K_0 – невырожденная матрица при эндогенных переменных текущего времени, D – матрица коэффициентов при внесистемных (экзогенных) переменных, $z(i)$ – вектор экзогенных переменных, включающий и их запаздывающие значения, $h(i)$ – векторный возмущающий сигнал с нулевым математическим ожиданием и ограниченными вторыми моментами) и приведенной

$$B(i) = -K_0^{-1} \sum_{l=1}^N K_l B(i-l) + Dz(i) - h(i) \quad (1.2)$$

или

$$B(i) = CZ(i) + x(i), \quad (1.3)$$

где вектор $Z(i)$ включает в себя как экзогенные переменные, так и запаздывающие значения эндогенных переменных

$$x(i) = K_0^{-1} h(i). \quad (1.4)$$

Важной проблемой исследований является идентификация параметров форм (1.1 – 1.3). Для этого достаточно часто используются косвенный, двухшаговый или трехшаговый методы наименьших квадратов, которые предназначены для работы только с фиксированной и заранее заданной выборкой [221]. Алгоритмы, предназначенные для работы в последовательном режиме, такие, как [107] релаксационный, рекурсивный или метод неподвижной точки характеризуются недостаточно высокой скоростью сходимости для того, чтобы обеспечивать эффективную обработку сигнала в режиме реального времени.

В [169] рассматривается модель многомерного временного ряда вида

$$B(i) = \sum_{l=1}^N K_l B(i-l) + \sum_{p=1}^q D_p z(i-p) + Fy(i-1) + x(i) \quad (1.5)$$

или

$$K(z^{-1})B(i) = D(z^{-1})z(i-1) + Fy(i-1) + G(z^{-1})h(i), \quad (1.6)$$

Неизвестные коэффициенты, описывающие поведение контролирующей последовательности, входят либо в матрицы B_l , D_p , F соответствующих раз-

мерностей, либо в матричные полиномы $B(z^{-1})$, $D(z^{-1})$, $G(z^{-1})$ от оператора сдвига назад z^{-1} , $y(i)$ – некоторая детерминированная функция, описывающая тренд, присутствующий в сигнале $B(i)$.

При этом авторами [169] отмечается, что один и тот же временной ряд F может быть описан бесконечным множеством многомерных уравнений (1.5) или (1.6).

Существует три основных подхода к задаче оценивания параметров этих уравнений: метод максимального правдоподобия, Байесов подход и метод ограниченной информации. Если первые два из них реализуются в пакетной форме, то третий – в форме рекуррентного метода наименьших квадратов может обрабатывать и последовательно поступающие наблюдения. К сожалению, стандартный рекуррентный метод наименьших квадратов, будучи, по сути, алгоритмом идентификации с бесконечной памятью, принципиально не пригоден для работы с нестационарными объектами, сигналы которых могут резко изменить свои свойства.

Для решения задачи обнаружения изменения свойств многомерных рядов удобным, компактным и эффективным описанием являются так называемые векторные авторегрессионные модели (VAR– модели) [40, 159], введенные и исследованные в [22, 222, 283].

В общем случае VAR-модель связывает прошлые и текущие наблюдения исходного сигнала $B(i)$ в форме

$$B(i) = K_0 + \sum_{l=1}^N K_l B(i-l) + x(i), \quad (1.7)$$

где $K_0 = \{k_{0i}\}$ – $(n \times 1)$ вектор средних значений,

$K_l = \{k_{lij}\}$ – $(n \times n)$ матрицы параметров,

N – порядок модели.

Кроме формулы (1.7) VAR-модель может быть компактно описана в пространстве состояний

$$\begin{aligned} \dot{B}(i) &= PB(i-1) + P_0 + E(i), \\ \dot{Y}(i) &= CB(i), \end{aligned}$$

$$\text{где } B(i) = \begin{pmatrix} B(i) \\ B(i-1) \\ \vdots \\ B(i-N+1) \end{pmatrix}, P_0 = \begin{pmatrix} K_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, P = \begin{pmatrix} K_1 & L & K_{N-1} & K_N \\ I_n & L & 0 & 0 \\ M & O & M & M \\ 0 & L & I_n & 0 \end{pmatrix},$$

$$E(i) = \begin{pmatrix} E(i) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, C = (I_n, 0, K, 0), B(i) - (np' - 1) - \text{вектор состояний,}$$

P - $(np' \times np)$ переходная матрица, 0 и 0 - $(n' - 1)$ и $(n' \times n)$ нулевые вектор и матрица соответственно.

Описание (1.7) позволяет использовать для анализа многомерных сигналов мощный математический аппарат калмановской фильтрации.

Обнаружение изменения свойств многомерного временного ряда F при анализе его кадров $B(i)$ связано с контролем каждой из его компонент $b_j(i)$, $j = 1, 2, \dots, n$, при этом возможно возникновение следующих трех ситуаций:

а) изменение средних значений (уравнений) $l \in N$ компонент

$$k_{0j}(i) = \begin{cases} k_{0j}, & \text{если } i < i_a, \\ k_{0j}^a, & \text{если } i \geq i_a, \end{cases} \quad (1.8)$$

б) изменение характеристик (дисперсий) $l \in N$ возмущений $x_i(s_i^2)$

$$b_j(i) = \begin{cases} k_{0j} + \sum_{l=1}^N \sum_{m=1}^n k_{ljm} x_m(i-l) + x_j(i), & \text{если } i < i_a, \\ k_{0j} + \sum_{l=1}^N \sum_{m=1}^n k_{ljm} x_m(i-l) + x_j^a(i), & \text{если } i \geq i_a, \end{cases} \quad (1.9)$$

в) изменение коэффициентов k_{ljm} , приводящее к изменению автокорреляционных свойств нестационарного временного ряда

$$b_j(i) = \begin{cases} k_{0j} + \sum_{l=1}^N \sum_{m=1}^n k_{ljm} x_m(i-l) + x_j(i), & \text{если } i < i_a, \\ k_{0i} + \sum_{l=1}^N \sum_{m=1}^n k_{ljm}^0 x_m(i-l) + x_j(i), & \text{если } i \geq i_a, \end{cases} \quad (1.10)$$

где i_a – момент времени, в который происходит изменение свойств.

В [40] предложена группа методов обнаружения изменения свойств процессов, описываемых VAR-моделями, в апостериорном режиме, т.е. в ситуации, когда задана вся реализация на интервале от 1 до N со всеми произошедшими в ней изменениями. Переход к априорной диагностике в режиме реального времени может быть реализован на основе настраиваемых VAR-моделей, обучаемых с помощью тех или иных процедур адаптивной идентификации с конечной памятью.

Одним из эффективных подходов к темпоральной обработке видеорядов может служить применение методов прогнозирования: по значимому различию наблюдения и прогноза делается вывод об изменении свойств видеопотока, т.е. происходит смена видеосегментов. Наличие темпоральной зависимости (вообще говоря, неизвестной) $y = f(T, K, S, x, e)$, сформированной в результате воздействия различных факторов, создает предпосылки для определения оценок $\hat{y}(t_{n+k})$ для неизвестных значений $y(t_{n+k})$ по предыдущим значениям $y(t_1), y(t_2), \dots, y(t_n)$, где k – горизонт прогноза. Здесь полага-

ется: T – тренд, которому соответствует медленное изменение в некотором направлении, сохраняющееся в течение значительного промежутка времени; K – циклические колебания, т.е. квазипериодическая динамика, в которой есть фаза возрастания и фаза убывания; S – сезонные колебания, т.е. изменения, регулярно происходящие в течение некоторого периода; x – хаотические компоненты; e – случайные флуктуации, т.е. несистематический или случайный эффект, индуцируемый влиянием на анализируемый процесс разнородных событий. Существует значимое множество подходов: регрессионные и авторегрессионные, нейросетевые модели, модели экспоненциального сглаживания, модели на базе цепей Маркова, адаптивные и классификационные модели и др. [19, 22-24, 52, 54, 68, 77, 111, 120, 145, 146, 228, 250]. Среди сопоставимых задач с обработкой видеопотоков можно указать системы, связанные с чрезвычайными ситуациями [57, 58], в которых можно выделить прогнозирование загрязнений окружающей среды [163, 166, 188, 199, 239, 276, 277, 282, 288], а также и подходы, направленные на выявление различных дефектов [110, 120, 238].

Необходимо отметить, что особое место в задачах темпоральной обработки видеопоследовательностей занимают методы компрессии [43, 104, 177, 184, 213, 260, 280]. По-прежнему перспективу имеют фрагментные on line методы на основе специализированных искусственных нейронных сетей, которые работают по принципу самообучения.

На основе проведенного анализа можно заключить, что на фоне перманентной аккумуляции плохоструктурированной динамической визуальной информации необходимость развития методов темпоральной обработки в признаковых пространствах и пространствах изображений приобретает все большую актуальность. При этом одной из ключевых задач развития информационных технологий CBVIR является создание комплекса методов и моделей обнаружения в последовательном режиме изменений свойств многомерных временных рядов, индуцированных видеопоследовательностями. Кроме того обработка видеопотоков должна быть обеспечена и возможностью сег-

ментации (кластеризации) во времени и пространстве произвольных множеств фрагментов видеокадров.

1.4 Постановка основных задач исследований

Нередко наше время представляется «эпохой информационного взрыва», когда развитие методов и средств регистрации, хранения, обработки, анализа и интерпретации информации отстает от возрастающих запросов индивидуальных и корпоративных пользователей, а главное – от высочайших темпов роста объемов и доступности фиксируемой и производимой информации. Релевантность информационных ресурсов, разумный компромисс между накоплением данных и потребностями их рационального использования являются одной из ключевых проблем, на которых акцентируется внимание исследователей.

Адекватность – соответствие трактовки заложенной семантике, достоверность – степень отражения объективной реальности, полнота – достаточность для принятия решения, доступность – интеграция создания, хранения и виртуализации, актуальность – соответствие моменту времени и предметной области, наряду с учетом субъективности, избыточности или, напротив, дефицита при интерпретации – вот перечень основных свойств информации, предопределяющий механизмы создания той или иной информационной технологии. Типичным примером активно развивающейся формы представления информации является гипермедиа, когда одновременно используется графика, звук, видео, текст и ссылки для создания нелинейной среды восприятия. Следует отметить, что слабоструктурированные либо совсем неструктурированные данные (прежде всего видеоданные) играют все большую роль в ИТ, что обусловлено высочайшей информационной емкостью видео и возможностями бурного развития программных и технических средств.

Управление видеоданными во многом базируется на достаточно хорошо развитом инструментарии систем CBIR, методов обработки и распозна-

вания изображений, но имеет и ярко выраженную специфику. Она в основном связана с необходимостью on line обработки и анализа очень больших объемов информации в условиях существенной неопределенности при главенствовании не пространственных (в поле зрения) корреляционных связей, а сложных отношений между видеокадрами во времени. В самом общем виде требующее развития управление видеоданными можно свести к решению задач анализа, многоаспектного автоматического индексирования, аннотирования и реферирования, нелинейного просмотра и поиска с различными формами запросов. Решение этих задач во многом определяется результатами структуризации видеопоследовательностей. На рис. 1.2 приведена обобщенная схема поисковой системы динамической визуальной информации с выделением on line и off line задач.

Структуризация видео – это поиск в условиях различной степени детализации совокупности устойчивых связей между отдельными компонентами,



Рисунок 1.2 – Схема поиска в коллекциях видео

которые обеспечивают как формальный синтез видеоряда и его частей, так и анализ семантики (содержательного значения), по крайней мере, при заданных критериях, в конкретных предметно-ориентированных областях. Иерархическая структура видео связана с представлением данных как множества видеорядов (коллекции – плохоструктурированного аналога базы данных), отдельного видео, его элементов в форме эпизодов, в свою очередь, состоящих из возможно несвязанных во времени сегментов (множеств последовательных во времени изображений однородных в смысле заданного критерия сходства), отдельных видеокадров. Потокам видео свойственна полиструктурность, обусловленная возникновением необходимости анализа «событий», «сюжетов», «сценариев» и других семантических единиц, динамически изменяющих стратификацию. На рис. 1.3 приведена схема, детализирующая задачи, которые необходимо решить для создания эффективной информаци-

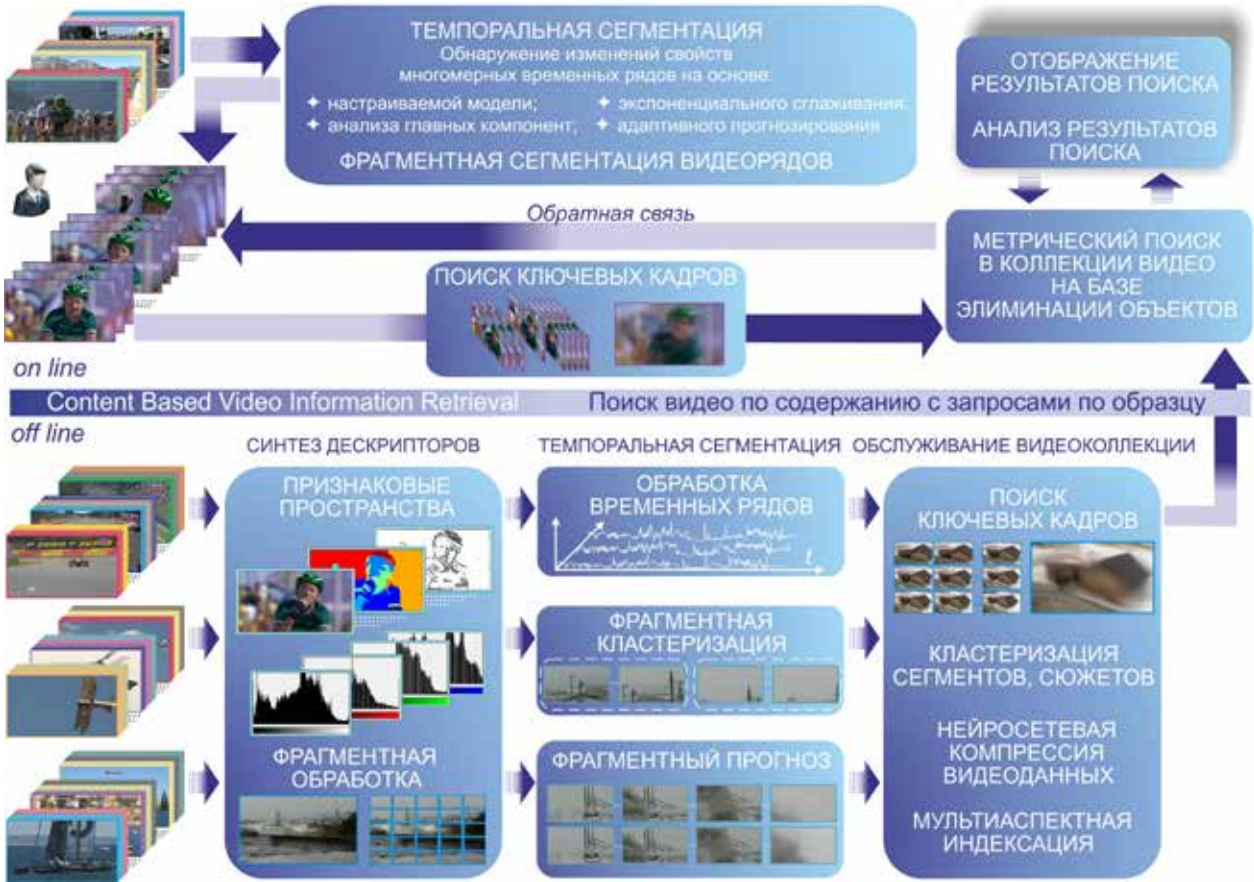


Рисунок 1.3 – On line и off line задачи в системах CBVIR

нной технологии поиска видео с запросами «по образцу».

Для создания моделей и методов, представляющих собой инструментарий информационных технологий, оперирующих с большими потоками видеоданных в плане их структуризации и метрического поиска, потребовалось решение следующих задач:

- анализ тенденций развития методов обработки видеопотоков в аспекте их структурирования (выделения однородных сегментов), поиска валидных представителей сегментов, дескрипторов видеок кадров для выделения перспективных средств, влияющих на эффективность многоаспектного индексирования в библиотеках динамической визуальной информации и метрического поиска в CBVIR;

- синтез моделей методов сегментации видео, обеспечивающих создание процедур информационного поиска с запросами ‘ad exemplum’ на основе построения областей элиминации для произвольных конфигураций объектов в поисковом пространстве;

- разработка методов кластеризации и сравнения разбиений конечных множеств для повышения валидности семантического поиска в больших библиотеках видеоданных;

- построение моделей и методов on line сегментации многомерных временных рядов, индуцируемых видеопотоками, путем обнаружения изменения их свойств;

- создание методов поиска ключевых кадров, представляющих сегменты видеоданных в задачах поиска видео на базе механизмов поиска изображений в больших библиотеках статической визуальной информации;

- разработка методов и моделей компрессии видеок кадров и видео для темпоральной обработки видеопотоков в задачах многоаспектной индексации;

- исследование специфики использования разработанных моделей и методов темпоральной обработки видеопоследовательностей, создание и внедрение исследовательских и специализированных программных средств.

Диссертация структурирована следующим образом. Второй раздел посвящен синтезу моделей сегментации видео и метрического поиска сегментов. В разделе получены обобщенные условия элиминации объектов при метрическом поиске без вычисления расстояний от них до запроса. В третьем разделе предложены фрагментные методы кластеризация, которые могут использоваться для пространственной и темпоральной сегментации визуальной информации. Изучены вопросы метрического сравнения классов эквивалентностей, являющихся результатом сегментации видеок кадров. Исследованы вопросы признаковых представлений изображений для построения многомерных временных рядов, соответствующих видеопоследовательностям. Результаты четвертого раздела представляют собой комплекс средств, предназначенных для обнаружения в on line режиме изменений свойств различных многомерных временных рядов. Пятый раздел посвящен моделям формирования ключевой информации, характеризующей отдельные части видеопотоков, разработке процедур поиска ключевых кадров. В шестом разделе рассмотрены методы и модели компрессии изображений для темпоральной обработки видео. Седьмой раздел посвящен экспериментальному исследованию методов и моделей темпоральной обработки видео для информационного поиска.

Выводы по первому разделу

1. Глобальная аккумуляция плохоструктурированных видеоданных требует развития информационных технологий в плане решения научно-технической проблемы, связанной со структуризацией видеопотоков, установлением семантического соответствия поискового запроса и поискового образа в условиях априорной неопределенности и ограничений на время принятия решений.

2. Для создания перспективных CBVIR требуется междисциплинарная многофакторная интеграция методов обработки многомерных сигналов, рас-

познавания образов, компьютерного зрения, управления базами данных, анализа многомерных временных рядов, которая в результате продуцирует инструментарий информационных технологий для хранения, обработки, передачи, отображения и интерпретации плохоструктурированных или совсем неструктурированных данных больших объемов.

3. Развитие средств on line структуризации видеопотоков на основе пространственной (внутрикадровой) и временной (межкадровой) сегментации видеоданных представляет определенную перспективу для сокращения семантического разрыва между признаковыми представлениями видео и целями предметно-ориентированной интерпретации, что в результате создает предпосылки валидности информационных технологий мультимедиа и гипермедиа.

РАЗДЕЛ 2

МОДЕЛИ СЕГМЕНТАЦИИ ВИДЕО И МЕТРИЧЕСКОГО ПОИСКА СЕГМЕНТОВ

Методы анализа мультимедийной (прежде всего визуальной) информации получают все большее развитие. Это связано с заметным совершенствованием программно-аппаратных средств регистрации, обработки и хранения растровой информации, возникновением сверхбольших библиотек, облачных хранилищ данных неограниченных объемов, интенсификацией и диверсификацией использования различных видеопоследовательностей. Спецификой работы с коллекциями видеопотоков и даже с отдельными видеорядами является их практически полная первичная неструктурированность, что, естественно, существенно затрудняет решение задач оперирования ими (хранения, поиска, анализа и интерпретации). Достаточно очевидной «единицей» структуризации является видеосегмент – последовательность кадров, «неразличимых» с точки заданного критерия. Более сложные, иерархические, конструкции, обеспечивающие индексацию видеоданных и создающие предпосылки для реферирования, аннотирования и интерпретации видеопоследовательностей, базируются именно на построении разбиений или покрытий видеопоследовательностей.

Раздел посвящен синтезу моделей сегментации видеопоследовательностей, которые предназначены для обеспечения возможностей поиска, в первую очередь, с запросами ‘ad exemplum’. Изучены подходы к поиску, когда на первом этапе выполняется элиминация информации, не удовлетворяющей требованиям сходства с запросом. При этом за счет учета конфигурационных свойств библиотек видео исключается существенное количество трудоемких операций сравнения видеосегментов или ключевых кадров, их характеризующих. Детально рассмотрены ситуации с различным числом опорных точек и различных «пакетах» запросов.

Раздел основан на работах автора [13,61,91,92].

2.1 Модели структуризации видеопоследовательностей для задач информационного поиска

Видеопоток представим в виде упорядоченного во времени набора изображений (видеокадров)

$$F = \{B_1, B_2, \dots, B_i, \dots, B_N\}, \quad (2.1)$$

где i – дискретное время,

$B_i(z) \in D \in \mathbb{R}^2, z \in \mathbb{R}^2$ – изображение в поле зрения D в момент времени i .

Видеопоток (2.1) индуцирует временной ряд $X = \{x_1, x_2, \dots, x_N\}$, в котором, не снижая общности рассмотрения, можно полагать, что $x_i \in \mathbb{R}^P$. Следует отметить, что временная модель видеопоследовательности может быть не только абсолютной (перечислением элементов), но и относительной – в виде отношений между видеокадрами. В качестве наиболее часто используемого варианта можно указать разностную модель, которая, как правило, представляется симметрической $N \times N$ матрицей различий между видеокадрами

$$D = \begin{pmatrix} 0 & d_{12} & d_{13} & \dots & d_{1N} \\ d_{21} & 0 & d_{23} & \dots & d_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ d_{(N-1)1} & d_{(N-1)2} & \dots & \dots & d_{(N-1)N} \\ d_{N1} & d_{N2} & \dots & d_{(N-1)N} & 0 \end{pmatrix} \quad (2.2)$$

где $d: F \times F \rightarrow \mathbb{R}_0^+$ – некоторый неотрицательный функционал, удовлетворяющий аксиомам рефлексивности ($d(B', B'') = 0 \iff B' = B''$) и симметрич-

ности ($d(B', B'') = d(B'', B')$).

Если для любого триплета (i, j, k) , $i, j, k \in \{1, \dots, N\}$ справедливо неравенство треугольника $d_{ij} + d_{jk} \geq d_{ik}$, то d – метрика. Отметим одно важное свойство матрицы D в этом случае: если $d_{ij} = \epsilon$, то для любого k справедливо неравенство $|d_{ik} - d_{jk}| \leq \epsilon$. Действительно, из того, что для любого k выполняются неравенства $d_{ik} + d_{ij} \geq d_{kj}$ и $d_{jk} + d_{ji} \geq d_{ki}$, с учетом свойства симметрии непосредственно следует $d_{ik} + \epsilon \geq d_{kj}$, $d_{jk} + \epsilon \geq d_{ki}$, что равносильно $d_{ik} + \epsilon \geq d_{kj} \geq d_{ik} - \epsilon$ или $|d_{ik} - d_{jk}| \leq \epsilon$. Необходимо подчеркнуть, если $\epsilon = 0$, то $d_{ik} = d_{jk}$. Данное свойство является ключевым, поскольку позволяет организовывать информационный поиск путем исключения (элиминации) объектов, явно не удовлетворяющих степени сходства с запросом.

В первом случае мы оперируем функционалом сходства, который связан с некоторыми отношениями толерантности и продуцирует покрытие видеоряда, во втором – имеем функционал неразличимости, индуцированный эквивалентностями, и разбиение видеопоследовательности.

Обработку видеоряда можно выполнять в двух режимах: пакетном и online. Первый предоставляет большие возможности по обработке и анализу, но вычислительные модели весьма сложны. Второй является более естественным, т.к. анализ видеoinформации реализуется по мере ее поступления. Ясно, что оперировать можно только предысторией, что приводит к необходимости создания специфического инструментария.

Пусть на декартовом квадрате $F \times F$ множества изображений или на декартовом квадрате множества признаков $X \times X$ (будем полагать их равносильными) задан некоторый набор бинарных отношений $T = \{t_g\}$, $g \in G$ (G – индексное множество), определяющих сходство (толерантность) или неразличимость (эквивалентность) пар $x_i, x_j \in X$. Если эквивалентность порождает естественную структуризацию видеопоследовательности в форме разбиения, то толерантность продуцирует покрытие видеоряда, т.е. требуется до-

полнительный анализ, позволяющий – либо сформировать из общих частей новый класс, либо перераспределить видеокadres по уже имеющимся классам. Рис. 2.1 демонстрирует примеры возникновения «проблемных» ви-



Рисунок 2.1 – Примеры толерантных видеокadres (посередине)

деокадров за счет эффекта «проявления».

Основной единицей структуризации видео является (видео)сегмент (отрезок) длиной $b - a + 1$ – упорядоченный во времени набор изображений $S_{ab} = \{V_a, V_{a+1}, \dots, V_{b-1}, V_b\}$. Изображения $V_i, V_j \in F$ являются связными относительно отношения t_g , если $\langle V_i, V_j \rangle \hat{=} t_g$ или $V_i t_g V_j$. Сегмент S_{ab} является связным, если $\forall i, j \in \{a, \dots, b\} \exists V_i t_g V_j$ и

$$\forall a > 1 \exists \langle V_{a-1}, V_a \rangle \hat{=} t_g, \forall b < N \exists \langle V_b, V_{b+1} \rangle \hat{=} t_g. \quad (2.3)$$

Безусловно, могут существовать разделенные во времени связные с заданным видеокadром изображения, но они образуют другие отрезки, которые должны учитываться на ином уровне стратификации. Связный сегмент будем обозначать $S_{ab}^* = [V_a] t_g$, подразумевая, что $[V_a] t_g$ – класс толерантности или эквивалентности. Ясно, что вместо V_a можно использовать любое изображение, но с практической точки зрения нередко следует применять ключе-

вой кадр, наиболее «похожий» на все остальные изображения отрезка, поскольку проверка отношений $T = \{t_g\}$, $g \in G$, как правило, выполняется с заданной точностью. Ключевым кадром B_{ab}^* связного сегмента $S_{ab}^* = [B_a]_t$ является

$$B_{ab}^* = \arg \min_{B^* \in [B_a]_t} \sum_{i=a}^b d(B^*, B_i). \quad (2.4)$$

Отметим, что ключевые кадры могут быть синтезированными. Так, на рис. 2.2 приведены примеры представления ключевых кадров «средними» изображениями сегмента.



Рисунок 2.2 – Примеры ключевых видеокadres связных сегментов

Будем называть сегмент сильно-связным ($^+S_{ab}$), если $B_i, B_j \in S_{ab}$, $g \in G \Rightarrow \langle B_i, B_j \rangle \in t_g$; если $B_i, B_j \in S_{ab}$, $\{g\}, g \in G \Rightarrow B_i t_g B_j$, сегмент будет слабо-связным ($^-S_{ab}$). Если набор $\{g\}$ одинаков для всех видеокadres сегмента, имеем однородную слабо-связность. В относительных координатах связный сегмент S_{ab}^* представляется матрицей

$$D_i = \begin{pmatrix} d_{a(i)(a(i)+1)} & d_{a(i)(a(i)+2)} & \dots & d_{a(i)b(i)} & 0 \\ 0 & d_{(a(i)+1)(a(i)+2)} & \dots & d_{(a(i)+1)b(i)} & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & d_{(b(i)-1)b(i)} & 0 \\ 0 & \dots & \dots & 0 & 0 \end{pmatrix} \quad (2.5)$$

В связи с тем, что обычно обработка видео производится в режиме *on line*, т.е. выполняется обработка растущего ряда, введем следующую конструкцию: k -аппроксимацией отрезка $[V_{a(i)}, V_{b(i)}]$ в точке V_m называется сегмент $[V_{a(i)}, V_m]_k$, для которого справедливо

$$x_m \mathbf{t} V_{x_{m-1}}, x_{m-1} \mathbf{t} V_{x_{m-2}}, \dots, V_{m-k+1} \mathbf{t} V_{m-k}.$$

В этом случае матрицы (2.2), (2.5) становятся наддиагональными. Если предыстория учитывается более детально, т.е. в каждой точке учитывается l предыдущих значений $V_j \mathbf{t} V_{j-1}, V_j \mathbf{t} V_{j-2}, \dots, V_j \mathbf{t} V_{j-l}, j \in \{a(i)+l, \dots, b(i)\}$, получаем наддиагональную ленточную матрицу (вплоть до верхнетреугольной). При этом, если \mathbf{t} – толерантность, следует рассматривать замыкание.

Проведенный анализ структуры видеоряда позволяет сформулировать задачу в следующей форме: необходимо найти представление видеопоследовательности F в виде связанных сегментов $S_{a(k)b(k)}^* \mathbf{t} \mathbf{A} \in \mathbf{A}$, $k = \overline{1, K}$ таких, что

$$F = \bigcup_{k=1}^K S_{a(k)b(k)}^*, \quad \forall i \neq j \in \{1, 2, \dots, K\} \quad S_{a(i)b(i)}^* \mathbf{t} S_{a(j)b(j)}^* = \mathbf{A}. \quad (2.6)$$

В относительных координатах это соответствует приведению матрицы (2.1) к блочно-диагональному виду

$$D^* = \begin{matrix} \emptyset & D_1 & & \emptyset \\ \circ & & & \circ \\ \circ & D_2 & & 0 \\ \circ & & \circ & \\ \circ & & & \\ \circ & & & \\ \circ & & D_i & \\ \circ & 0 & & \circ \\ \circ & & & \\ \circ & & & D_K \end{matrix} \begin{matrix} \circ \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \\ \div \end{matrix} \quad (2.7)$$

Поиск сюжетов в видеопоследовательности, т.е. объединение «не граничащих» сегментов в новый класс эквивалентности, связан с иерархией сегментаций на основе иных – семантических – отношений. Исследование иерархий выходит за пределы данного исследования.

Рассмотрим теперь некоторую формальную конструкцию. Пусть на $P \uparrow \downarrow$ задана функция $f(x)$, для которой при заданном отношении сходства t определим «следящую» функцию

$$W(x_0) = \max\{x > x_0 : \exists x', x'' \in [x_0, x] \text{ с } f(x') \neq f(x'')\}. \quad (2.8)$$

Если ее сопоставить некоторому разбиению $P = \bigcup_{k=1, K} [a(k), b(k)]$, то очевидно, что при $x_0 \in (a(k), b(k))$ имеем $W(x_0) = 0$, при $x_0 \in [a(k), b(k))$ имеем $W(x_0) = a(k)$. При $x'_0, x''_0 \in [a(k), b(k)]$ и $x'_0 < x''_0$ получаем $W(x'_0) > W(x''_0)$, а в целом $W(x_0)$ – кусочно линейна и имеет разрывы первого рода на границах сегментов разбиения. Адаптируя следящую функцию к on line анализу, получаем

$$W_d(x_0) = \max\{x > x_0 : \exists x' \in [x_0, x], \exists x'' \in [x_0, x'] \text{ с } f(x') \neq f(x'')\}$$

или с памятью на d видеокадров

$$W_d(k_0) = \max\{k > k_0 : \exists k' \in [k_0, k] \exists k'' = k' - l, l \in \{1, \dots, \min\{d, k' - k_0\}\} \text{ с } B(k') \neq B(k'')\}. \quad (2.9)$$

2.2 Формализация процедур поиска сегментов с запросами ‘ad exemplum’

Рассмотрим в метрическом пространстве F сегментов видео некоторую конечную относительно (известны лишь расстояния между всеми точками) конфигурацию $K \hat{=} F$, $K = \bigcup_{l \in L} \bigcup_{i \in I} S_{l,a(i)b(i)}^*$ (L, I – индексные множества в коллекции и видеопотоке соответственно). Задача заключается в поиске сегмента(ов), наиболее похожего(их) на запрос, который, в свою очередь, также задается видеосегментом/видеокадром. Необходимо особо подчеркнуть, что стремление к ликвидации семантического дисбаланса между признаками низкого уровня и высокоуровневой интерпретацией содержания динамической (и даже статической) визуальной информации приводит к введению сложных комбинированных метрик. При этом их вычислительная сложность существенно возрастает, а при больших L и I комбинаторная емкость задачи поиска становится такой, что говорить о приемлемом времени ее решения становится невозможным. В рамках традиционных подходов к метрическому поиску, которые, как правило, основываются на обратном неравенстве треугольника, т.е. за счет возможности предварительного вычисления расстояний между объектами поиска исключаются из рассмотрения элементы без вычисления расстояний от них до запроса, проанализируем с самых общих позиций возможности элиминации сегментов. Отметим, что такая парадигма поиска, вообще говоря, свойственна и человеку: сначала (при ослабленных критериях) отбрасываются явно неподходящие объекты, а затем на базе более детального анализа принимается решение.

Поясним дальнейшую (упрощенную) нотацию. Сегменты из видеокolleкции будем обозначать точками $x_1, \dots, x_n \hat{=} K \hat{=} F$, запрос ‘ad exemplum’ точкой $y \hat{=} F$, опорные точки, т.е. объекты, до которых в режиме on line вычисляются расстояния от запроса $y \hat{=} F$, будем представлять как $x_1^*, x_2^*, \dots, x_m^* \hat{=} K$. Анализируемую конфигурацию $K' \hat{=} K \hat{=} F$ представим

матрицей расстояний $P = (r(x_i, x_j))_{i,j=1}^n$, ($r(x_i, x_j)$ – расстояние между точками x_i, x_j), которая вычисляется в режиме off line. Поясним это подробнее. На отдельных этапах поиска конфигурация K' может представлять собой некоторую часть видеокolleкции, например, один из кластеров. Тогда матрица P является лишь отдельным блоком общей блочно диагональной матрицы, полученной в результате предварительной кластеризации. В связи с этим будем полагать, что на вид матрицы P не накладывается никаких ограничений, и, не нарушая общности рассмотрения, будем считать $K' = K$. Для запроса $y \in F$ расстояние до некоторой опорной точки $x^* \in K$ обозначим $d = r(y, x^*)$.

Введем, наконец, порог сходства объектов D , который характеризует меру сходства (уровень неразличимости) точек метрического пространства F с текущей точкой y . Если $r(y, x_k) > D$, то точка x_k настолько отличается от y , что она автоматически исключается из дальнейшего анализа.

Таким образом, результатом метрического поиска $p \in n$ сегментов в конфигурации $K \in F$ с запросом $y \in F$, ограничениями $r(y, x_l) \leq D$, $l \in \{i_1, i_2, \dots, i_p\}$ при известных расстояниях до некоторого множества опорных точек $\{x_k^*\}_{k \in K}$ является упорядоченное по возрастанию множество $K^s = \{x_{i_1}, x_{i_2}, \dots, x_{i_s}\} \in K$, для которого при $0 < s \in p \in n$ справедливы соотношения

$$\begin{aligned} & "x_{i_j} \in K^s, "x \in K \setminus K^s, "y \in F \quad \rho(y, x_{i_j}) \leq D, D^3 0, \\ & \rho(y, x_{i_j}) \leq \rho(y, x), \rho(y, x_{i_j}) \leq \rho(y, x_{i_{j+1}}), j = \overline{1, s-1}. \end{aligned} \quad (2.10)$$

Отметим, что параметр D – необязательный: полагая $D = \infty$, переходим к анализу всей конфигурации $K \in F$. На практике этот параметр выбирается из ряда эвристических соображений, связанных с собственно пространствен-

ной конфигурацией сегментов, свойствами используемой метрики и предполагаемой степенью релевантности поиска. Параметр p , точнее индуцируемый им параметр s , также является варьируемым: в граничных случаях при $p = 1$ находим ближайший к запросу сегмент, а при $p = n$ и $D = \forall$ ранжируем по расстоянию до запроса все точки конфигурации. Задачу (2.10) назовем (D, p) -поиском.

Поиск сегмента, наиболее похожего на запрос, можно представлять итерационной процедурой исключения подмножеств сегментов, для которых явно не выполняются соотношения (2.10). Варьируя опорные точки $x_1^*, x_2^*, \dots, x_m^*$, имеем возможность приходить к сокращению мощности анализируемой конфигурации K_i ($K_0 = K, K_{i+1} = K_i / \{x_{k1}, x_{k2}, \dots, x_{kr}\}$). Формализуем операцию элиминации (исключения) элементов.

Допустим, имеется конфигурация $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$. Кроме расстояний между n анализируемыми и m опорными точками известны расстояния $r(y, x_i^*) = d_i, i = \overline{1, m}$ от запроса y до каждой из опорных точек. Вводя обозначения $r(x_i, x_j) = a_{ij}, r(x_i^*, x_j^*) = b_{ij}, r(x_i, x_j^*) = g_{ij}$, можно представить матрицу расстояний для данной конфигурации в виде

$$P = \begin{array}{cccc|cccc}
 \textcircled{0} & a_{12} & K & a_{1n} & g_{11} & K & K & g_{1m} & \textcircled{0} \\
 \textcircled{0} & 0 & K & K & g_{21} & K & K & g_{2m} & \div \\
 & & O & K & K & K & K & K & \div \\
 & & & a_{n-1n} & K & K & K & K & \div \\
 & & & 0 & g_{n1} & K & K & g_{nm} & \div \\
 & & & & \text{---} & \text{---} & \text{---} & \text{---} & \div \\
 & & 0 & & 0 & b_{12} & K & b_{1m} & \div \\
 & & & & & 0 & K & K & \div \\
 & & & & & & O & b_{m-1m} & \div \\
 & & & & & & & 0 & \textcircled{0}
 \end{array} \tag{2.11}$$

Рис. 2.3 дает геометрическую интерпретацию off line и on line расстояний. Ясно, что конфигурация $\mathbb{K}=\{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$ порождает некоторое

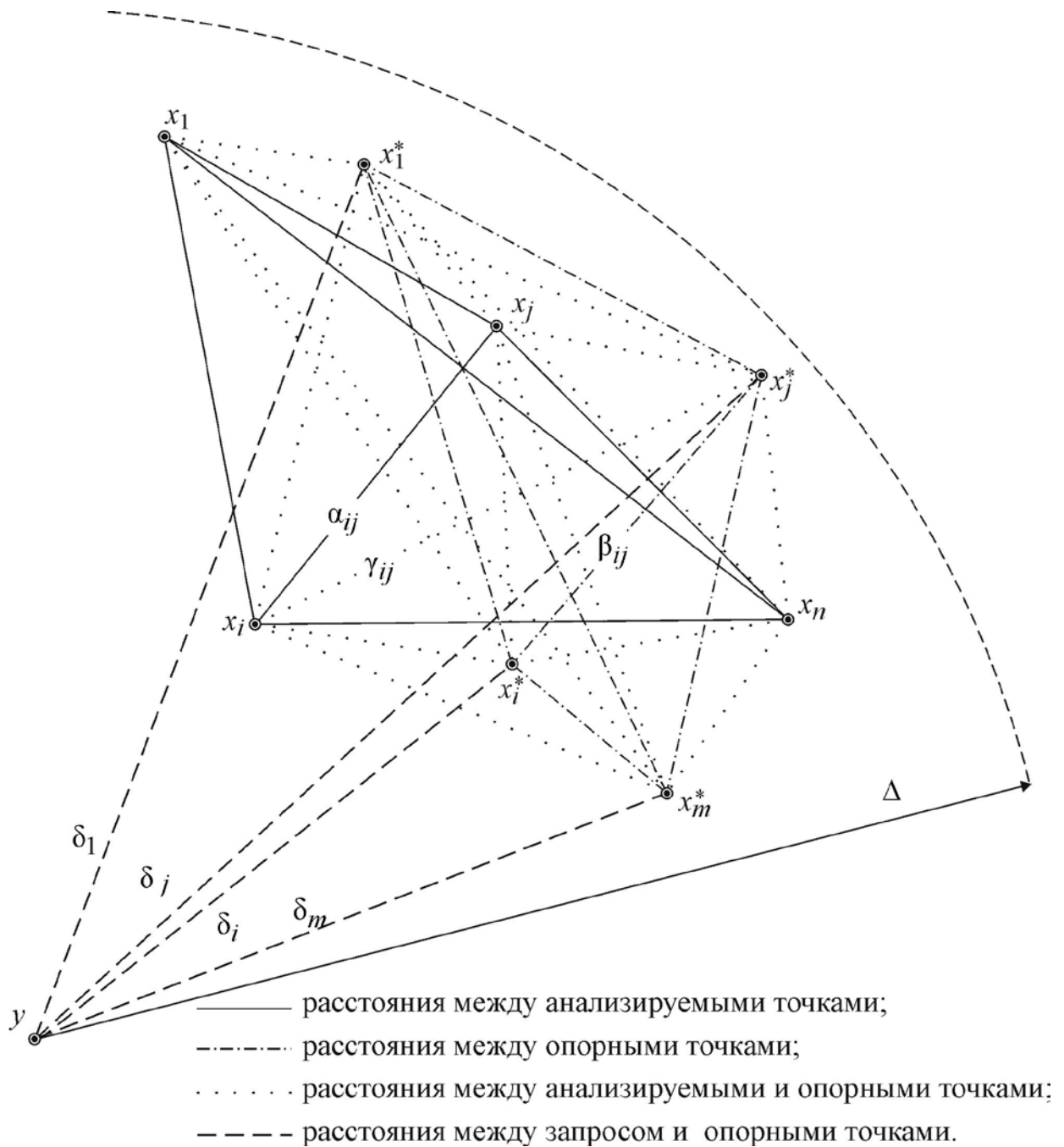


Рисунок 2.3 – Геометрическая интерпретация конфигурации $\mathbb{K}=(x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*)$

арифметическое пространство \mathbb{R}^q , в котором и следует формулировать условия поиска. Поскольку используются off line вычисленные расстояния (между всеми точками, включая опорные) и on line расстояния от запроса до

опорных точек, его размерность равна $C_{n+m}^2 + m$.

Для конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$ область D_r назовем областью элиминации r -го порядка, если

$$\hat{D}_r = \left\{ (a_{ij})_{\substack{i,j=1,n \\ i < j}}, (b_{ij})_{\substack{i,j=1,m \\ i < j}}, (g_j)_{j=1,n}, (d_i)_{i=1,m} \right\} \subset C_{n+m}^2,$$

то среди точек x_1, \dots, x_n можно указать r точек, среди которых отсутствует ближайшая к $y \in F$.

Определенный интерес представляют граничные ситуации. Нетрудно заметить, что если порядок элиминации равен $n - 1$, то фактически D_r уже определяет искомый объект. Если $r = 1$, то поиск области элиминации – это построение процедур выбора ближайшего среди двух объектов без вычисления расстояний до них. Очевидно, что в этом случае умение выбирать ближайшую из любой пары точек позволяет легко находить ближайшую из множества x_1, \dots, x_n , в частности:

- выбираем ближайшую к запросу y точку из x' и x'' ;
- пусть это точка x' , тогда исключаем x'' и переходим к следующему шагу;
- выбираем следующую (если перебор не завершен) точку x''' , затем – ближайшую из x' и x''' и возвращаемся к предыдущему шагу.

В результате находим точку, ближайшую к запросу на всем множестве x_1, \dots, x_n . Количество операций выбора при этом равно $n - 1$.

Построение областей элиминации будем проводить в следующем порядке. Сначала рассмотрим ситуацию выбора точки ближайшей к запросу при одной, двух и m опорных точках. Далее, наращивая число точек конфигурации, представим общую схему построения области D_r .

2.3 Построение областей элиминации с одной опорной точкой

Начнем с рассмотрения ситуации $n = 2, m = 1$, т.е. конфигурация имеет вид $K = \{x_1, x_2, x^*\}$ (рис. 2.4), и известно расстояние до опорной точки: $\rho(y, x^*) = \delta \geq 0$. Полагаем, что в рассмотрении участвуют точки, для которых выполняется пороговое ограничение $\rho(y, x^*) < D$ (в противном случае имеем вырожденный случай). Действительно, если хотя бы одна из точек конфигурации $\{x_1, x_2, x^*\} \in K$ не попадает в окрестность $U_D(y)$ радиуса D запроса $y \in F$, то возможны два варианта. Допустим, $x^* \notin U_D(y)$ – это означает, что в качестве опорной выбрана точка, не удовлетворяющая условиям поиска, и целесообразно выбрать другую. Если одна из точек x_1 или x_2 не принадлежит окрестности $U_D(y)$, например x_1 , то ближайшей является x_2 . Если же обе не принадлежат $U_D(y)$, то нет смысла искать среди них ближайшую. Од-

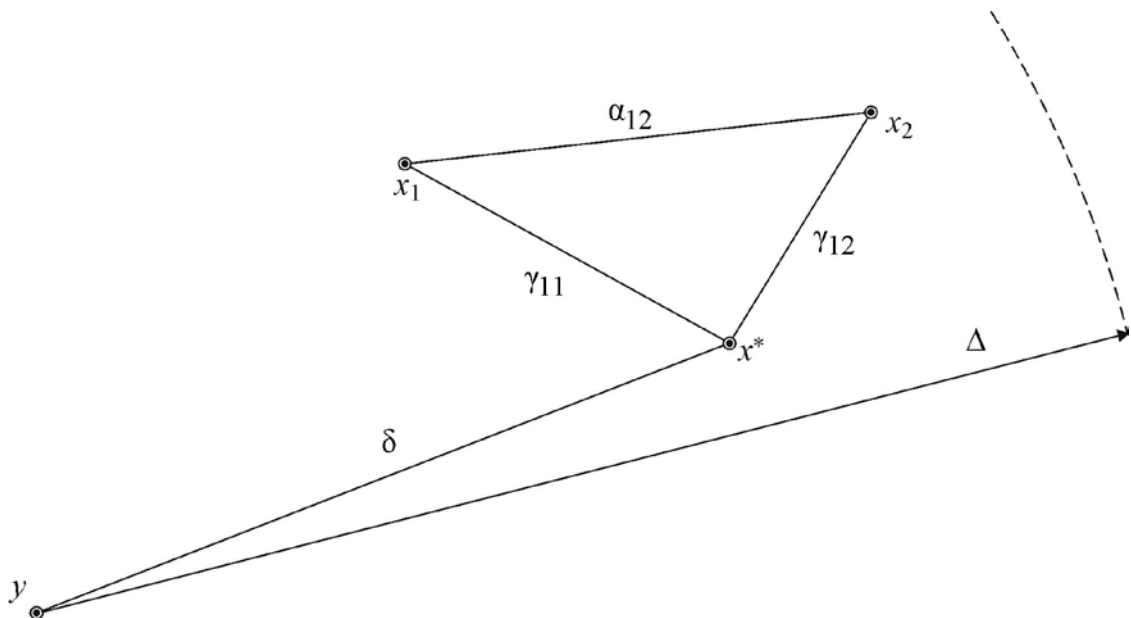


Рисунок 2.4 – Геометрическая интерпретация конфигурации $K = \{x_1, x_2, x^*\}$

нако расстояния $\rho(y, x_1)$ и $\rho(y, x_2)$ неизвестны, и если не предполагать, что точки $x_1, x_2 \in U_D(y)$, то необходимо учитывать дополнительную информа-

цию, например, о предварительной кластеризации в пространстве F .

Для конфигурации $K=\{x_1, x_2, x^*\}$ представленной на рис. 2.4, матрица расстояний (2.11) в этом случае примет вид

$$P = \begin{pmatrix} \alpha_{12} & \gamma_{11} \\ 0 & \gamma_{12} \\ 0 & 0 \end{pmatrix} \quad (2.12)$$

Область элиминации в i^3 определяется следующим образом. Если при известных $\rho(y, x^*) = \delta \geq 0$ и $D > 0$ элементы матрицы расстояний P удовлетворяют системе условий

$$\begin{cases} \alpha_{12}, \gamma_{11}, \gamma_{12} > 0; \\ |\delta - \gamma_{11}| \leq \delta + \gamma_{12}, \\ |\delta - \gamma_{12}| \leq \delta + \gamma_{11}; \\ \gamma_{11} + \gamma_{12} \leq \alpha_{12}, \\ \gamma_{11} + \alpha_{12} \leq \gamma_{12}, \\ \alpha_{12} + \gamma_{12} \leq \gamma_{11}; \\ \delta \leq D, \\ \delta + \gamma_{11} \leq D, \\ \delta + \gamma_{12} \leq D, \end{cases} \quad (2.13)$$

то можно без вычисления расстояний $\rho(y, x_1)$ и $\rho(y, x_2)$ сделать выбор ближайшей к запросу y точки x_1 или x_2 конфигурации $K=\{x_1, x_2, x^*\}$.

Рассмотрим элементы (2.13) подробнее. Ограничения $\alpha_{12}, \gamma_{11}, \gamma_{12} \geq 0$ естественны в силу неотрицательности метрики. В соответствии с рис. 2.4 из неравенства треугольника, учитывая запрос y , имеем

$$\begin{cases} |\delta - \gamma_{11}| \leq \rho(y, x_1) \leq \delta + \gamma_{11}, \\ |\delta - \gamma_{12}| \leq \rho(y, x_2) \leq \delta + \gamma_{12}, \end{cases} \quad (2.14)$$

а принимая во внимание $D(x_1 x_2 x^*)$ (здесь и далее будем таким образом обозначать треугольники с соответствующими вершинами) – $\gamma_{11} + \gamma_{12} \geq \alpha_{12}$, $\gamma_{11} + \alpha_{12} \geq \gamma_{12}$, $\alpha_{12} + \gamma_{12} \geq \gamma_{11}$.

Нетрудно понять, что в любом случае областью элиминации будет являться та область, в которой хотя бы одна из нижних границ неравенств (2.14) превосходит верхнюю границу тех же неравенств. С одной стороны, это означает $|\delta - \gamma_{11}| \geq \delta + \gamma_{12}$ или, что равносильно, $\rho(y, x_1) \geq \rho(y, x_2)$. С другой стороны, $|\delta - \gamma_{12}| \geq \delta + \gamma_{11}$, откуда следует, что точка x_1 ближе к запросу y , чем x_2 . Наконец, три последние неравенства системы (2.13) обеспечивают принадлежность конфигурации $K \hat{U}_D(y)$ допустимому уровню сходства. Действительно, $\delta \in D$ свидетельствует о том, что $x^* \hat{U}_D(y)$, а если $\delta + \gamma_{11} \in D$, то $\rho(y, x_1) \in \delta + \gamma_{11} \in D$ или $x_1 \hat{U}_D(y)$. Аналогично из $\delta + \gamma_{12} \in D$ следует $x_2 \hat{U}_D(y)$.

Поскольку анализ происходит в \mathbb{R}^3 , можно дать геометрическую интерпретацию системы (2.13) – в явном виде указать область элиминации как объект в трехмерном пространстве, ограниченный плоскостями.

Начнем с части системы вида

$$\begin{cases} |\delta - \gamma_{11}| \geq \delta + \gamma_{12}, \\ |\delta - \gamma_{12}| \geq \delta + \gamma_{11}; \end{cases} \quad (2.15)$$

Этот набор условий является наиболее существенным, поскольку именно он позволяет сделать выбор ближайшей к запросу точки. Построим проекцию области элиминации на плоскость переменных $(\gamma_{11}, \gamma_{12})$ или плоскость $\alpha_{12} = 0$. Система (2.15) может быть переписана в виде

$$\begin{cases}
 \delta - \gamma_{11} \leq \delta + \gamma_{12}, \\
 \delta - \gamma_{11} \leq \delta - \gamma_{12}, \\
 \delta - \gamma_{12} \leq \delta + \gamma_{11}, \\
 \delta - \gamma_{12} \leq \delta - \gamma_{11}.
 \end{cases}
 \quad (2.16)$$

Рассмотрим первую часть системы (2.16). Нетрудно заметить, что неравенство $\delta - \gamma_{11} \leq \delta + \gamma_{12}$ приводит к условию $\gamma_{11} + \gamma_{12} \leq 0$, а с учетом $\gamma_{11}, \gamma_{12} > 0$ является пустым множеством. Значит, остается только второе условие $\delta - \gamma_{11} \leq \delta - \gamma_{12}$, которое эквивалентно неравенству $\gamma_{11} - \gamma_{12} \leq 2\delta$, определяющее область первого квадранта, ограниченную прямыми $\gamma_{12} = 0$ и $\gamma_{12} = \gamma_{11} - 2\delta$. Рассуждая совершенно аналогично, из второй пары неравенств получаем область, ограниченную прямыми $\gamma_{11} = 0$ и $\gamma_{12} = \gamma_{11} + 2\delta$. Полученные области иллюстрируются рис. 2.5 в i^2 и рис. 2.6 в i^3 , где область, соответствующая объединению ограничена снизу плоскостью $\alpha_{12} = 0$ и плоскостями L_1, L_2, L_3, L_4 , уравнения которых имеют вид

$$\begin{aligned}
 L_1 : \gamma_{12} &= 0, \\
 L_2 : \gamma_{11} - \gamma_{12} - 2\delta &= 0, \\
 L_3 : \gamma_{11} - \gamma_{12} + 2\delta &= 0, \\
 L_4 : \gamma_{11} &= 0.
 \end{aligned}$$

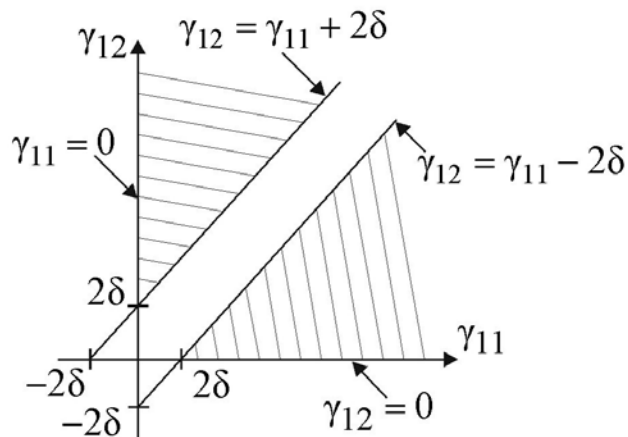


Рисунок 2.5 – Геометрическая интерпретация ограничений в i^2

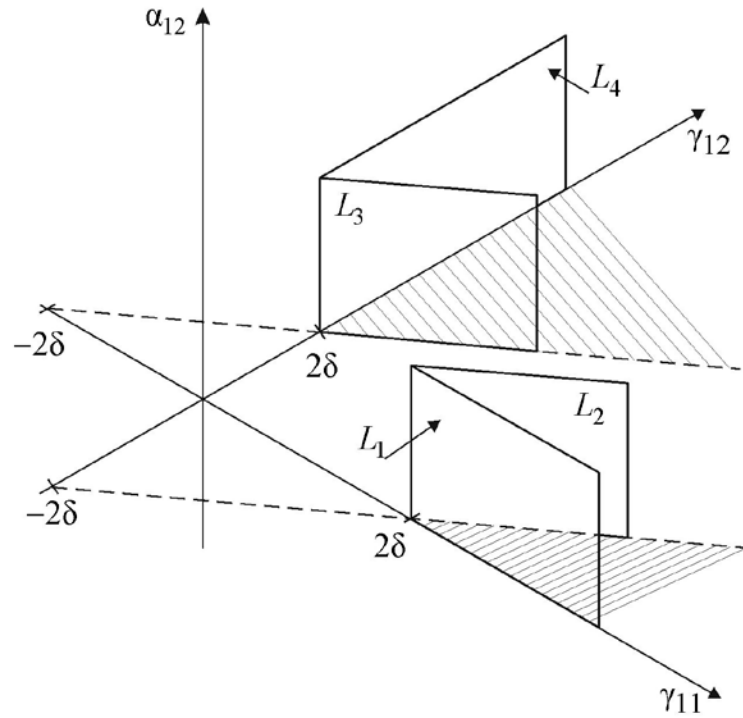


Рисунок 2.6 – Геометрическая интерпретация ограничений в i^3

Теперь остановимся на части системы (2.13), представляющей собой систему неравенств

$$\begin{cases} \dot{\gamma}_{11} + \gamma_{12} \leq \alpha_{12}, \\ \dot{\gamma}_{11} + \alpha_{12} \leq \gamma_{12}, \\ \dot{\alpha}_{12} + \gamma_{12} \leq \gamma_{11}. \end{cases} \quad (2.17)$$

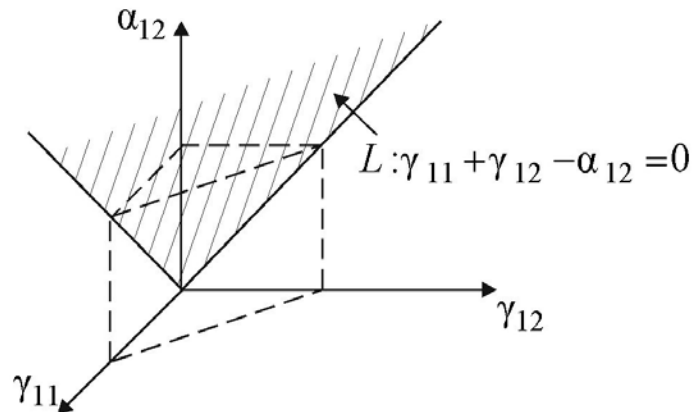


Рисунок 2.7 – Геометрическая интерпретация первого ограничения из (2.17)

Рассмотрим первое из них: его граница – это плоскость $\gamma_{11} + \gamma_{12} - \alpha_{12} = 0$, а область, удовлетворяющая условию $\gamma_{11} + \gamma_{12} \geq \alpha_{12}$, находится под плоско-

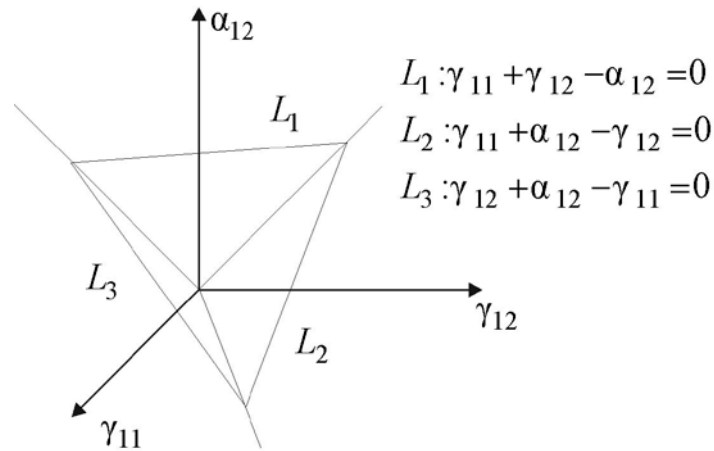


Рисунок 2.8 – Геометрическая интерпретация ограничений (2.17)

стью L (рис. 2.7). В силу симметрии ограничений относительно других осей область системы неравенств (2.17) ограничивается бесконечной правильной треугольной пирамидой (рис. 2.8)

Наконец, рассмотрим последние три неравенства системы (2.13)

$$\begin{aligned} \delta &\in D, \\ \delta + \gamma_{11} &\in D, \\ \delta + \gamma_{12} &\in D. \end{aligned} \quad (2.18)$$

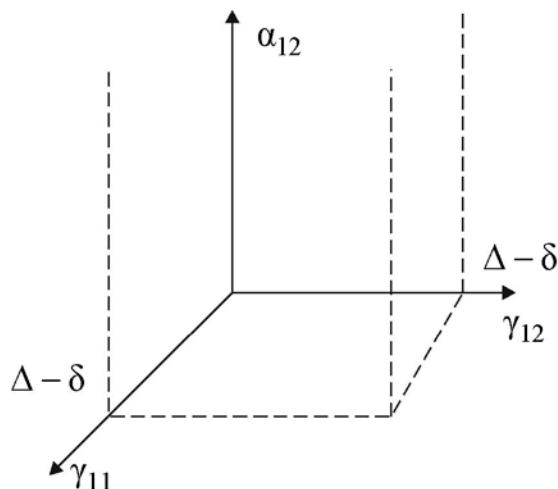


Рисунок 2.9 – Геометрическая интерпретация ограничений (2.18)

Строго говоря, первое из них это фактически условие, которым должны удовлетворять параметры системы (2.13) (действительно, $\gamma_{11}, \gamma_{12}, \alpha_{12}$ – переменные, относительно которых строится область элиминации, а δ и D – параметры системы), чтобы требуемая область была не пустой. Оставшиеся два неравенства задают внутреннюю часть бесконечного параллелепипеда (рис. 2.9), ограниченного плоскостями $\gamma_{11} = 0, \gamma_{12} = 0, \alpha_{12} = 0, \gamma_{11} = D - \delta, \gamma_{12} = D - \delta$.

В результате получаем, что область элиминации – это пересечение областей, представленных на рис. 2.6, 2.8, 2.9. Как видно из рис. 2.6, 2.9, чтобы область элиминации была непустой, необходимо выполнение достаточно очевидного неравенства $D - \delta > 2\delta$ что означает: параметры системы (2.18) δ и D должны удовлетворять более жесткому, чем $\delta \leq D$ требованию, а именно – $\delta < \frac{D}{3}$.

2.4 Анализ процедур поиска при двух опорных точках

В случаях, когда проведена прелиминарная кластеризация пространства F , нередко целесообразно использовать несколько опорных точек, например, центров кластеров. Рассмотрим сначала конфигурацию $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$, состоящую из двух сравниваемых точек x_1 и x_2 и двух опорных x_1^* и x_2^* , что схематично представлено рис. 2.10.

Матрица расстояний (2.11) конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$ имеет вид

$$P = \begin{matrix} \text{e} & \alpha_{12} & \gamma_{11} & \gamma_{12} & \text{ö} \\ \zeta & 0 & \gamma_{21} & \gamma_{22} & \div \\ \zeta & & 0 & \beta_{12} & \div \\ \zeta & & & & \div \\ \text{e} & & & & 0 \end{matrix} \quad (2.19)$$

что означает: область элиминации должна быть найдена в шестимерном арифметическом пространстве.

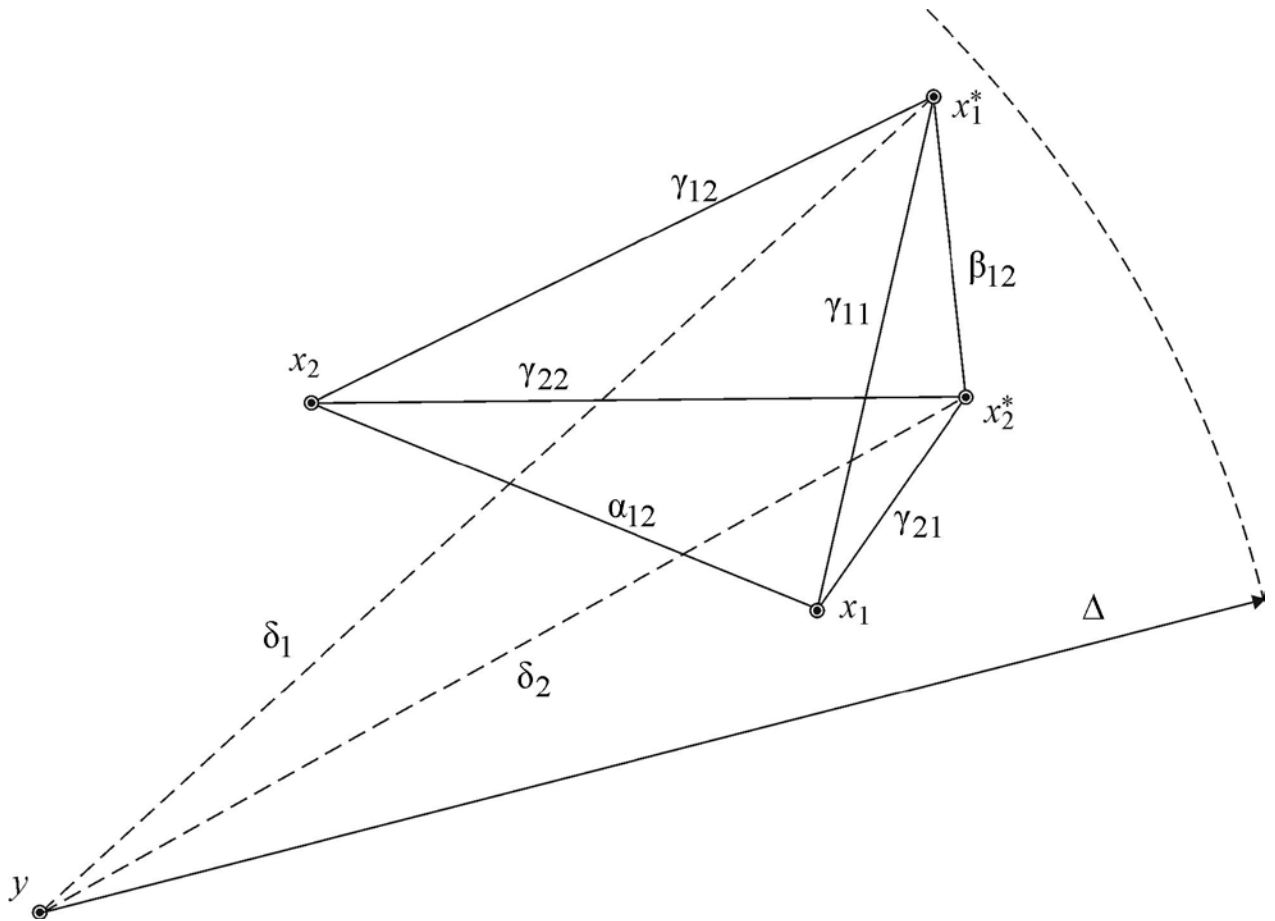


Рисунок 2.10 – Пример конфигурации $\mathbb{K}=\{x_1, x_2, x_1^*, x_2^*\}$

Первый тип ограничений связан с неотрицательностью расстояний. Второй тип ограничений базируется на неравенстве треугольника и, наконец, третий – внешнее условие, касающееся сходства или различия запроса y со сравниваемыми и опорными точками. Наиболее существенен второй тип ограничений, из которых вытекают условия, определяющие область элиминации.

Следует сказать, что ограничения, связанные с неравенством треугольника можно разделить на две группы, а именно

$$\begin{cases}
 \text{i} & |\delta_1 - \gamma_{11}| \leq \rho(y, x_1) \leq \delta_1 + \gamma_{11}, \\
 \text{ii} & |\delta_1 - \gamma_{12}| \leq \rho(y, x_1) \leq \delta_2 + \gamma_{12}, \\
 \text{iii} & |\delta_1 - \gamma_{21}| \leq \rho(y, x_2) \leq \delta_1 + \gamma_{21}, \\
 \text{iv} & |\delta_1 - \gamma_{22}| \leq \rho(y, x_2) \leq \delta_2 + \gamma_{22},
 \end{cases} \quad (2.20)$$

то есть на систему, где непосредственно участвуют сравниваемые неизвест-

ные расстояния $\rho(y, x_1)$ и $\rho(y, x_2)$, и набор систем, касающихся элементов матрицы расстояний (2.19). Например, для треугольника $D(x_1, x_2, x_1^*)$ (см. рис 2.10) имеем

$$\begin{array}{l} \dot{\gamma}_{11} + \gamma_{21} \stackrel{3}{=} \alpha_{12}, \\ \dot{\gamma}_{11} + \alpha_{12} \stackrel{3}{=} \gamma_{21}, \\ \ddot{\gamma}_{21} + \alpha_{12} \stackrel{3}{=} \gamma_{11}. \end{array} \quad (2.21)$$

На рис. 2.11 представлен еще ряд треугольников, для которых следуют ограничения:

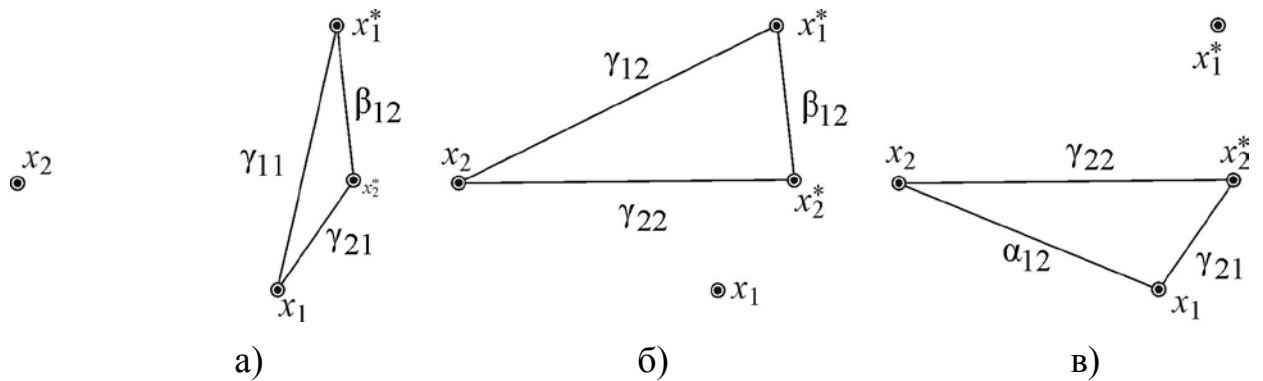


Рисунок 2.11 – Треугольники конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$ (рис. 2.8)

$D(x_1, x_1^*, x_2^*)$ – рис. 2.11, а)

$$\begin{array}{l} \dot{\gamma}_{11} + \gamma_{12} \stackrel{3}{=} \beta_{12}, \\ \dot{\gamma}_{11} + \beta_{12} \stackrel{3}{=} \gamma_{12}, \\ \ddot{\gamma}_{12} + \beta_{12} \stackrel{3}{=} \gamma_{11}, \end{array} \quad (2.22)$$

$D(x_2, x_1^*, x_2^*)$ – рис. 2.11, б)

$$\begin{array}{l} \dot{\gamma}_{21} + \gamma_{22} \stackrel{3}{=} \beta_{12}, \\ \dot{\gamma}_{21} + \beta_{12} \stackrel{3}{=} \gamma_{22}, \\ \ddot{\gamma}_{22} + \beta_{12} \stackrel{3}{=} \gamma_{21}, \end{array} \quad (2.23)$$

$D(x_1, x_2, x_2^*)$ – рис. 2.11, в)

$$\begin{cases} \delta_1 + \gamma_{12} \leq \alpha_{12}, \\ \delta_1 + \alpha_{12} \leq \gamma_{22}, \\ \delta_2 + \alpha_{12} \leq \gamma_{12}. \end{cases} \quad (2.24)$$

Всего же имеем $C_5^3 = 10$ треугольников, т.е. остается еще два треугольника, содержащие точку y , а именно: $D(y, x_1^*, x_2^*)$ и связанные с ним ограничения

$$\begin{cases} \delta_1 + \delta_2 \leq \beta_{12}, \\ \delta_1 + \beta_{12} \leq \delta_2, \\ \delta_2 + \beta_{12} \leq \delta_1. \end{cases}$$

но в них входит один параметр β_{12} , и существенно только ограничение на него

$$|\delta_1 - \delta_2| \leq \beta_{12} \leq \delta_1 + \delta_2. \quad (2.25)$$

Что касается треугольника $D(y, x_1, x_2)$, то он не дает ограничений на параметр α_{12} , поскольку расстояния $\rho(y, x_1)$ и $\rho(y, x_2)$ неизвестны.

Теперь вернемся к системе неравенств (2.20), которая является основной в построении области элиминации. Ясно, что для нахождения ближайшей точки среди точек x_1 и x_2 к запросу y достаточно, чтобы выполнялись следующие системы неравенств

$$\begin{cases} |\delta_1 - \gamma_{11}| \leq \delta_1 + \gamma_{21}, \\ |\delta_1 - \gamma_{11}| \leq \delta_2 + \gamma_{22}, \\ |\delta_1 - \gamma_{12}| \leq \delta_1 + \gamma_{21}, \\ |\delta_1 - \gamma_{12}| \leq \delta_2 + \gamma_{22}. \end{cases} \quad (2.26)$$

В этом случае $\rho(y, x_1) \geq \rho(y, x_2)$ и соответственно $\rho(y, x_1) \leq \rho(y, x_2)$, когда

$$\begin{cases} \hat{\rho}_1 - \gamma_{21} \geq \delta_1 + \gamma_{11}, \\ \hat{\rho}_1 - \gamma_{21} \leq \delta_2 + \gamma_{12}, \\ \hat{\rho}_2 - \gamma_{22} \geq \delta_1 + \gamma_{11}, \\ \hat{\rho}_2 - \gamma_{22} \leq \delta_2 + \gamma_{12}. \end{cases} \quad (2.27)$$

Действительно, есть по два ограничения снизу и по два ограничения сверху $\rho(y, x_1)$ и $\rho(y, x_2)$ и система (2.26) означает, что хотя бы одно из ограничений снизу расстояния $\rho(y, x_1)$ превосходит верхнюю границу расстояния $\rho(y, x_2)$. Отсюда естественно вытекает, что $\rho(y, x_1) \geq \rho(y, x_2)$. Соответственно, выполнение условий (2.27) достаточно для того, чтобы $\rho(y, x_1) \leq \rho(y, x_2)$.

Таким образом, установлено, если элементы матрицы расстояний P вида (2.19) удовлетворяют системам условий (2.21) – (2.27), лежат в «первом квадранте» (все положительны) и в D -окрестности запроса y , т.е. удовлетворяют ограничениям

$$\begin{cases} \hat{\rho}_1 + \gamma_{11} < D, \\ \hat{\rho}_2 + \gamma_{12} < D, \\ \hat{\rho}_1 + \gamma_{21} < D, \\ \hat{\rho}_2 + \gamma_{22} < D. \end{cases} \quad (2.28)$$

где $0 < \delta_1, \delta_2 < D$, то этого достаточно для выбора ближайшей точки конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$ к запросу y .

Для понимания этого факта остается прояснить природу системы (2.28). Смысл ее заключается в том, что непревышение D – порога меры сходства запроса y с опорными точками x_1^*, x_2^* и сравниваемыми точками x_1 и x_2 – должно удовлетворяться всегда (в противном случае точки сразу ис-

ключаются из рассмотрения). Иначе говоря, $x_1, x_2, x_1^*, x_2^* \in U_D(y)$, но если опорные точки попадают в нужную окрестность из условия $0 < \delta_1, \delta_2 < D$, то сравниваемые попадают, когда выполняется система (2.28) (напомним, что $\rho(y, x_1)$ и $\rho(y, x_2)$ неизвестны). Действительно, из неравенства треугольника следует, что

$$\begin{cases} \rho(y, x_1) \leq \delta_1 + \gamma_{11}, \\ \rho(y, x_1) \leq \delta_2 + \gamma_{12} \end{cases} \quad (2.29)$$

и

$$\begin{cases} \rho(y, x_2) \leq \delta_1 + \gamma_{21}, \\ \rho(y, x_2) \leq \delta_2 + \gamma_{22}. \end{cases} \quad (2.30)$$

Тогда ясно, что если имеет место (2.28) то из (2.29) и (2.30) следует, что $\rho(y, x_1) < D$ и $\rho(y, x_2) < D$, т.е. $x_1, x_2 \in U_D(y)$.

Итак, завершено построение области достаточности в случае двух опорных точек. Поскольку, это область в \mathbb{R}^6 , то с геометрической точки зрения, можно дать интерпретацию лишь ее одномерным, двумерным или трехмерным проекциям на определенные координатные подпространства.

Остановимся на этом подробнее. Следует отметить, что наибольший интерес представляют двумерные и трехмерные проекции. Конечно же, прежде всего имеет смысл указать проекции, соответствующие набору неравенств, описывающих область элиминации – это неравенства (2.20) – (2.24). Непосредственная проверка показывает, что неравенства (2.21) – (2.24) ничем не отличаются от неравенств, рассмотренных для конфигурации $K = \{x_1, x_2, x^*\}$, т.е. трехмерные проекции, им соответствующие, останутся теми же. А именно, в координатном подпространстве, например, переменных $\{\gamma_{11}, \gamma_{21}, \alpha_{12}\}$ будем иметь представление в виде рис. 2.8. Та же ситуация соответствует трехмерным проекциям в координатных подпространствах для троек параметров $\{\gamma_{11}, \gamma_{12}, \beta_{12}\}$, $\{\gamma_{21}, \gamma_{22}, \beta_{12}\}$ и $\{\gamma_{12}, \gamma_{22}, \alpha_{12}\}$.

Теперь рассмотрим двумерные проекции, индуцируемые неравенствами систем (2.26) и (2.27) и являющиеся основными при построении области элиминации. Рассмотрим (2.26). Нетрудно заметить, что в координатном подпространстве $\{\gamma_{11}, \gamma_{21}\}$ приходим к ситуации, описанной выше: проекция области элиминации будет иметь вид, как на рис. 2.5. Аналогичный вид проекции области элиминации будет и в подпространстве $\{\gamma_{12}, \gamma_{22}\}$.

Но в случае двух опорных точек возникают еще две однотипные пары координатных двумерных проекций $\{\gamma_{11}, \gamma_{22}\}$, $\{\gamma_{12}, \gamma_{21}\}$, которые не присутствовали в конфигурации $K = \{x_1, x_2, x^*\}$. Им соответствуют системы неравенств:

– для $\{\gamma_{11}, \gamma_{22}\}$

$$\begin{cases} \delta_1 - \gamma_{11} \geq \delta_2 + \gamma_{22}, \\ \delta_2 - \gamma_{22} \geq \delta_1 + \gamma_{11}, \end{cases} \quad (2.31)$$

– для $\{\gamma_{12}, \gamma_{21}\}$

$$\begin{cases} \delta_1 - \gamma_{12} \geq \delta_2 + \gamma_{21}, \\ \delta_2 - \gamma_{21} \geq \delta_1 + \gamma_{12}. \end{cases} \quad (2.32)$$

Рассмотрим первое неравенство системы (2.31). Ясно, что оно эквивалентно системе

$$\begin{cases} \delta_1 - \gamma_{11} \geq \delta_2 + \gamma_{22}, \\ \delta_1 - \gamma_{11} \leq \delta_2 + \gamma_{22}. \end{cases}$$

Отсюда следует

$$\begin{cases} \delta_1 - \delta_2 \geq \gamma_{11} + \gamma_{22}, \\ \delta_1 + \delta_2 \leq \gamma_{11} - \gamma_{22}. \end{cases}$$

В данном случае имеются два варианта. Если $\delta_1 < \delta_2$, то неравенство $\delta_1 - \delta_2 \geq \gamma_{11} + \gamma_{22}$, с учетом того, что $\gamma_{11}, \gamma_{22} \geq 0$ представляет собой пустое множество, а проекция области элиминации в координатном пространстве

$\{\gamma_{11}, \gamma_{22}\}$ будет задаваться неравенством $\delta_1 + \delta_2 \leq \gamma_{11} - \gamma_{22}$ и иметь вид, представленный на рис. 2.12, а). Если $\delta_1 > \delta_2$, то с учетом неравенства $\delta_1 - \delta_2 \leq \gamma_{11} + \gamma_{22}$ двумерная проекция области элиминации в координатном подпространстве $\{\gamma_{11}, \gamma_{22}\}$ будет представлена в виде рис. 2.12, б).

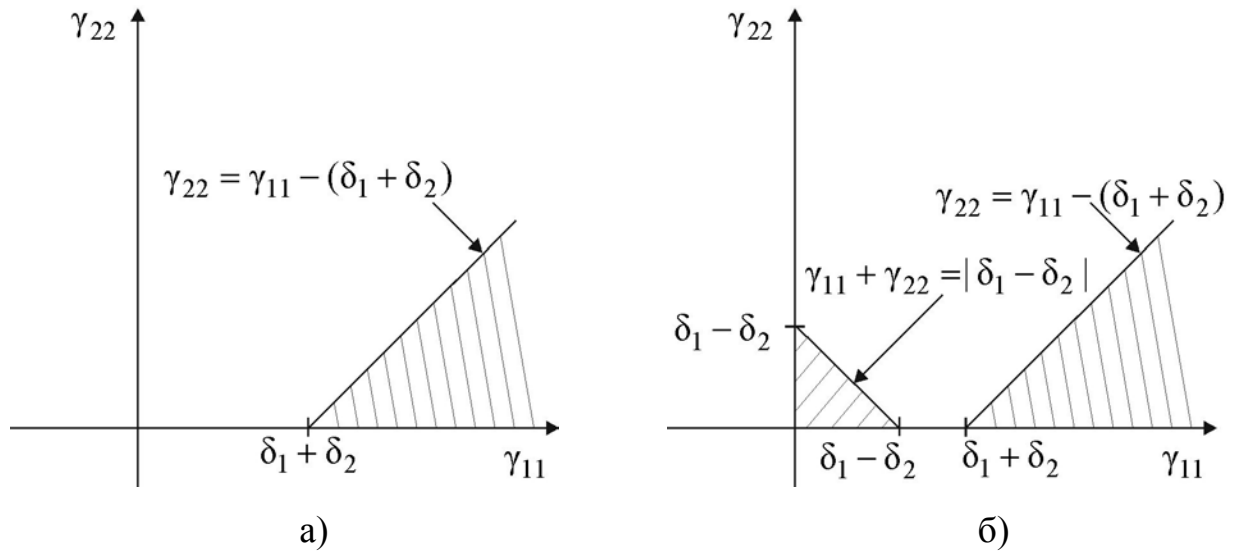


Рисунок 2.12 – Геометрическая интерпретация проекций ограничений конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$

Исследуя аналогично второе неравенство системы (2.31), получим, при $\delta_1 > \delta_2$ область, изображенную на рис. 2.13, а), а при $\delta_1 < \delta_2$ – на рис. 2.13, б),

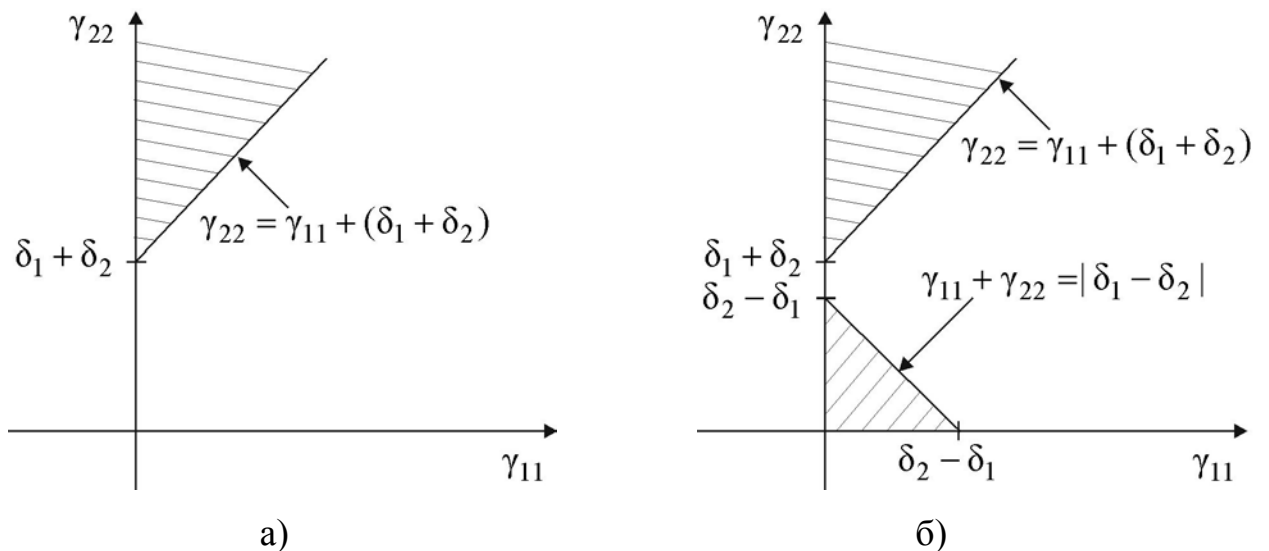


Рисунок 2.13 – Геометрическая интерпретация проекций ограничений конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*\}$

Таким образом, системе (2.31) соответствует область, представленная на рис. 2.14. Наконец, системе (2.32) соответствует двумерная проекция области элиминации в координатном подпространстве $\{\gamma_{12}, \gamma_{21}\}$, полностью аналогичная показанной на рис. 2.14.

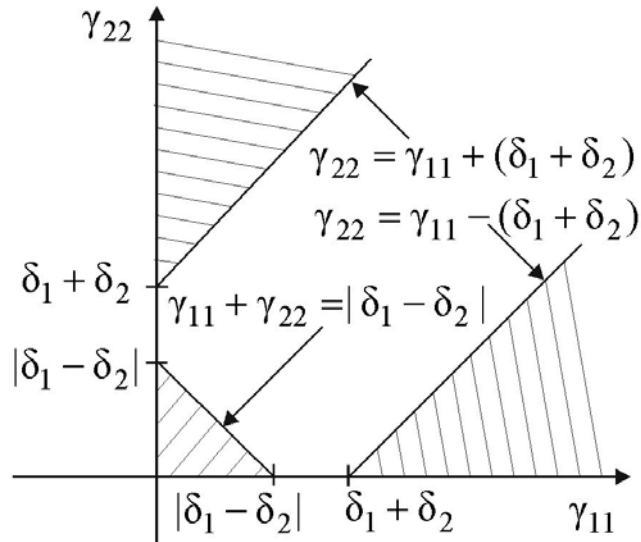


Рисунок 2.14 – Геометрическая интерпретация проекций ограничений (2.31)

2.5 Исследование областей элиминации первого порядка для конфигураций с m опорными точками

После анализа областей элиминации для одной и двух опорных точек естественным продолжением является изучение общего случая – построения областей элиминации для конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*, \dots, x_m^*\}$ ($m > 2$).

В принятой нотации для конфигурации, показанной на рис 2.15, матрица расстояний (2.11) P примет вид

$$P = \begin{matrix} \circ & a_{12} & g_{11} & K & K & g_{1m} & \circ \\ \circ & 0 & g_{21} & K & K & g_{2m} & \div \\ \circ & & 0 & b_{12} & K & b_{1m} & \div \\ \circ & & & 0 & K & K & \div \\ \circ & & & & \circ & b_{m-1m} & \div \\ \circ & & & & & 0 & \emptyset \end{matrix} \quad (2.33)$$

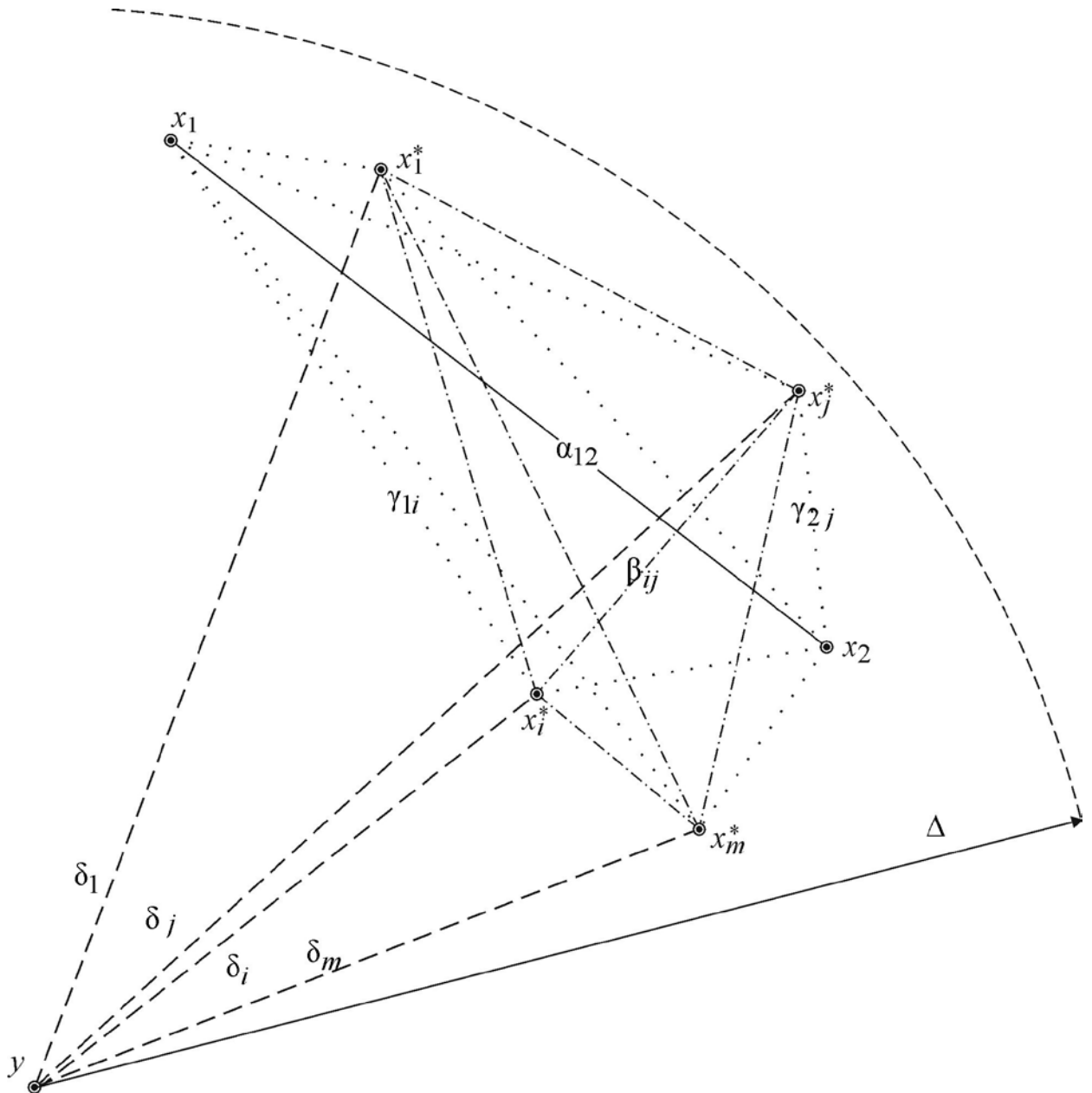


Рисунок 2.15 – Пример конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, x_2^*, \dots, x_m^*\}$

Также как и ранее, основная часть ограничений, обеспечивающая построение области элиминации, связана с неизвестными расстояниями $\rho(y, x_1)$ и $\rho(y, x_2)$. Если в случае одной опорной точки было 2 ограничения, в случае двух – 4, то при m опорных точках их будет $2m$, и они будут иметь вид

$$\begin{array}{l}
|\delta_1 - \gamma_{11}| \leq \rho(y, x_1) \leq \delta_1 + \gamma_{11}, \\
|\delta_1 - \gamma_{21}| \leq \rho(y, x_2) \leq \delta_1 + \gamma_{21}, \\
\vdots \\
|\delta_i - \gamma_{1i}| \leq \rho(y, x_1) \leq \delta_i + \gamma_{1i}, \\
|\delta_i - \gamma_{2i}| \leq \rho(y, x_2) \leq \delta_i + \gamma_{2i}, \\
\vdots \\
|\delta_m - \gamma_{1m}| \leq \rho(y, x_1) \leq \delta_m + \gamma_{1m}, \\
|\delta_m - \gamma_{2m}| \leq \rho(y, x_2) \leq \delta_m + \gamma_{2m}.
\end{array}$$

Если сгруппировать ограничения по точкам x_1 и x_2 , а именно

$$\begin{array}{l}
\{|\delta_i - \gamma_{1i}| \leq \rho(y, x_1) \leq \delta_i + \gamma_{1i}, i = \overline{1, m}, \\
|\delta_i - \gamma_{2i}| \leq \rho(y, x_2) \leq \delta_i + \gamma_{2i}, i = \overline{1, m},
\end{array}$$

то можно сделать выбор ближайшей точки к запросу y , если хотя бы одна из нижних границ первой системы неравенств превосходит хотя бы одну из верхних границ второй системы неравенств и наоборот. Точнее говоря, это означает, что область элиминации имеет проекцию в координатном подпространстве параметров $\{\gamma_{11}, \dots, \gamma_{1m}, \gamma_{21}, \dots, \gamma_{2m}\}$ общего пространства $i^{(m+2)(m+1)/2}$ (пока без учета остальных ограничений), которая задается объединением двух систем неравенств: первая определяет как ближайшую точку x_1 и имеет вид

$$[\delta_i + \gamma_{1i} \leq |\delta_j - \gamma_{2j}|,$$

где объединение ведется по всем парам с повторениями из набора индексов $\{1, 2, \dots, m\}$, т.е. $i, j = \overline{1, m}$. Количество этих пар равно n^2 .

Аналогичная система объединений вида

$$[\delta_j + \gamma_{2j} \leq |\delta_i - \gamma_{1i}|$$

обеспечивает выбор x_2 как ближайшей точки. В результате окончательно проекция области элиминации без учета ограничений, которые будут указаны ниже в координатном пространстве $\{\gamma_{11}, \dots, \gamma_{1m}, \gamma_{21}, \dots, \gamma_{2m}\}$, является объединением $2m^2$ неравенств, имеющих вид

$$\begin{aligned} \hat{e} \delta_i + \gamma_{1i} &\leq |\delta_j - \gamma_{2j}|, \\ \hat{e} \delta_j + \gamma_{2j} &\leq |\delta_i - \gamma_{1i}|, \end{aligned} \quad (2.34)$$

где $i, j = \overline{1, m}$.

Учтем теперь другие ограничения, относящиеся к тем треугольникам, которые описывают одну или две опорных точек и одну или две из сравниваемых, т.е. дано множество опорных точек $\{x_1^*, \dots, x_m^*\}$ и множество сравниваемых $\{x_1, x_2\}$. Ясно, что количество треугольников с вершинами из этих двух множеств можно находить следующим образом: если участвует одна опорная точка, то их m , а именно $D(x_i^*, x_1, x_2)$, где $i = \overline{1, m}$; если участвует две опорные точки, то их $2C_m^2$, т.к. из множества $\{x_1, x_2\}$ выбор одной точки – два способа выбора, а из множества $\{x_1^*, \dots, x_m^*\}$ выбор двух точек C_m^2 способа выбора. Таким образом, это треугольники вида $D(x_i^*, x_j^*, x_k)$, где $i \neq j$, $i, j = \overline{1, m}$ и $k = 1, 2$.

Из неравенства треугольника первому типу треугольников $D(x_i^*, x_1, x_2)$ соответствует m ограничений вида

$$\begin{aligned} \gamma_{1i} + \gamma_{2i} &\leq \alpha_{12}, \\ \gamma_{1i} + \alpha_{12} &\leq \gamma_{2i}, \\ \alpha_{12} + \gamma_{2i} &\leq \gamma_{1i}, \end{aligned} \quad (2.35)$$

где $i = \overline{1, m}$, второму типу треугольников $D(x_i^*, x_j^*, x_k)$ соответствует C_m^2 ограничений вида

$$\begin{aligned}
 & \dot{\gamma}_{1i} + \gamma_{1j} \geq \beta_{ij}, \\
 & \dot{\gamma}_{1i} + \beta_{ij} \geq \gamma_{1j}, \\
 & \dot{\beta}_{ij} + \gamma_{1j} \geq \gamma_{1i},
 \end{aligned} \tag{2.36}$$

когда $k=1$, т.е. $D(x_i^*, x_j^*, x_1)$ и когда $k=2$, т.е. $D(x_i^*, x_j^*, x_2)$ ограничения будут выглядеть следующим образом

$$\begin{aligned}
 & \dot{\gamma}_{1i} + \gamma_{2j} \geq \beta_{ij}, \\
 & \dot{\gamma}_{1i} + \beta_{ij} \geq \gamma_{2j}, \\
 & \dot{\beta}_{ij} + \gamma_{2j} \geq \gamma_{1i}.
 \end{aligned} \tag{2.37}$$

Отметим, что для систем (2.36) и (2.37) $i, j = \overline{1, m}$ и $i \neq j$

Следует сказать, что описанный тип ограничений аналогичен ситуации двух опорных точки. Однако, начиная с трех опорных точек, возникают дополнительные ограничения, которые связаны с треугольниками $D(x_i^*, x_j^*, x_l^*)$, содержащими в качестве вершин только опорные точки. Всего таких треугольников C_m^3 и им соответствуют системы неравенств

$$\begin{aligned}
 & \dot{\beta}_{ij} + \beta_{il} \geq \beta_{jl}, \\
 & \dot{\beta}_{ij} + \beta_{jl} \geq \beta_{il}, \\
 & \dot{\beta}_{jl} + \beta_{il} \geq \beta_{ij},
 \end{aligned} \tag{2.38}$$

где $i, j, l = \overline{1, m}$ и $i \neq j, i \neq l, l \neq j$.

Наконец, остается последний тип ограничений, связанный с неравенством треугольника: одной вершиной треугольника является запрос y , а две другие – опорные точки x_i, x_j , где $i \neq j$ и $i, j = \overline{1, m}$. Треугольникам вида $D(y, x_i^*, x_j^*)$ (их всего C_m^2) соответствует система неравенств

$$\begin{cases} \delta_i + \delta_j \leq \beta_{ij}, \\ \delta_i + \beta_{ij} \leq \delta_j, \\ \beta_{ij} + \delta_j \leq \delta_i, \end{cases} \quad (2.39)$$

где $i \neq j$ и $i, j = \overline{1, m}$.

Но из матрицы (2.33) в наборе неравенств системы (2.39) участвует один параметр β_{ij} , поэтому она трансформируется к виду

$$\{|\delta_i - \delta_j| \leq \beta_{ij} \leq \delta_i + \delta_j, \quad (2.40)$$

где $i \neq j$ и $i, j = \overline{1, m}$.

Подводя некоторый промежуточный итог, можно заключить, что в случае m опорных точек, если элементы матрицы расстояний P вида (2.33) удовлетворяют системам условий (2.34) – (2.38), (2.40), являются строго положительными и удовлетворяют системе неравенств

$$\begin{cases} \delta_i + \gamma_{1i} < D, \\ \delta_i + \gamma_{2i} < D, \end{cases}$$

обеспечивающей попадание всех точек конфигурации $\mathbb{K} = \{x_1, x_2, x_1^*, \dots, x_m^*\}$ в D -окрестность запроса y , то этого достаточно для выбора ближайшей точки без вычисления расстояний.

Особо следует подчеркнуть, что полученный результат обеспечивает создания простых итерационных процедур направленного поиска путем попарного сравнения, что дает существенное повышение быстродействия в условиях предварительной кластеризации множества сегментов видео.

2.6 Общий случай построения областей элиминации

До сих пор рассматривались только области элиминации первого по-

рядка. Проанализируем теперь область D_r для конфигурации $\mathbb{K}=\{x_1, \dots, x_n, x^*\}$, пример которой приведен на рис. 2.16.

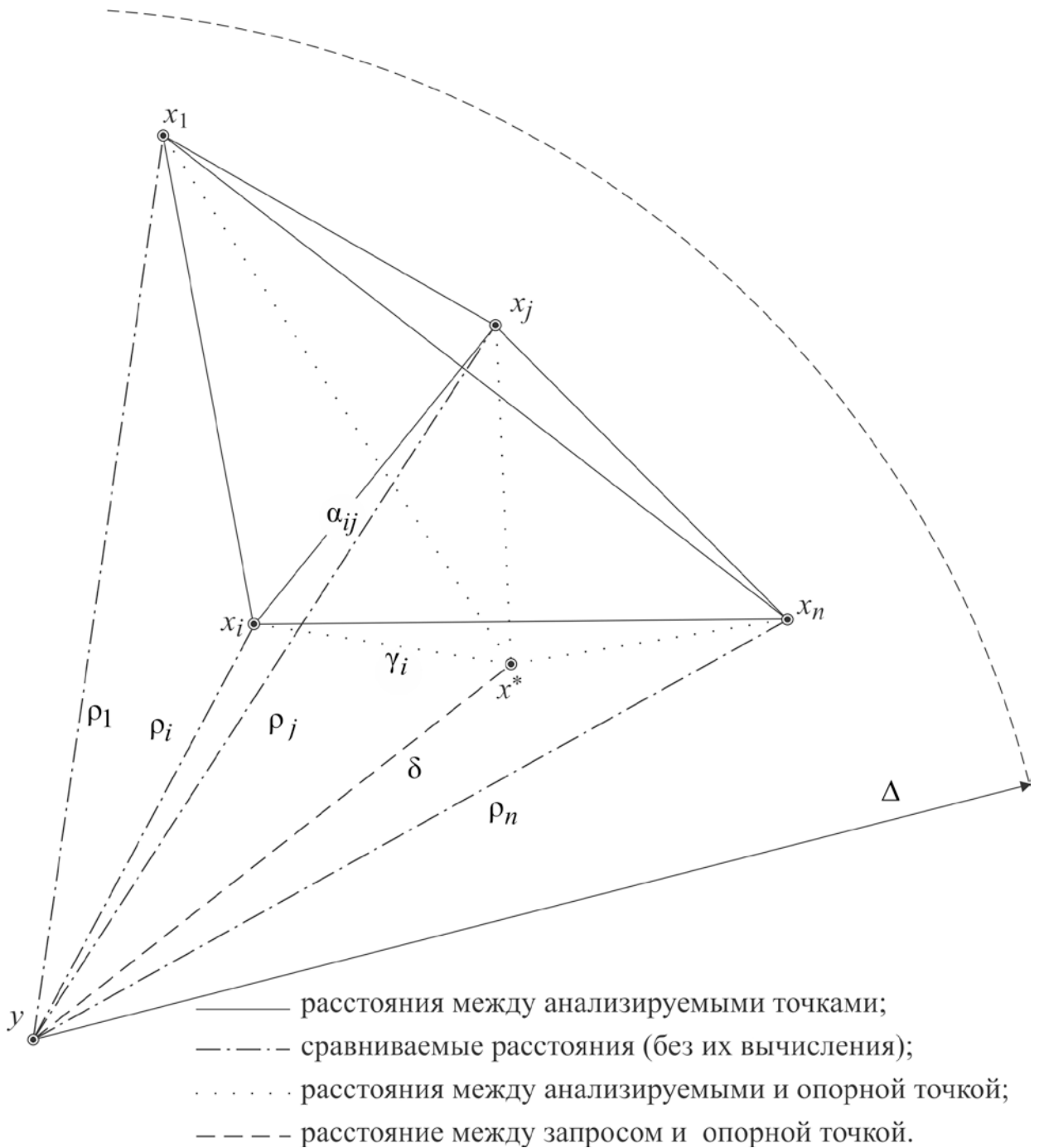


Рисунок 2.16 – Пример конфигурации $\mathbb{K}=\{x_1, \dots, x_n, x^*\}$

Обозначим неизвестные расстояния $r(y, x_i)$, $i = \overline{1, n}$ через r_i , а известные расстояния $r(x^*, x_i)$, $i = \overline{1, n}$ через g_{li} . Матрица расстояний (2.11) примет

ВИД

$$P = \begin{matrix} \text{a} \\ \text{c} \\ \text{c} \\ \text{c} \\ \text{c} \\ \text{c} \\ \text{e} \\ \text{e} \\ \text{e} \\ \text{e} \\ \text{e} \\ \text{e} \end{matrix} \begin{matrix} a_{12} & K & a_{1n} & g_{11} \ddot{o} \\ 0 & K & a_{2n} & g_{12} \ddot{\div} \\ & \circ & K & K \ddot{\div} \\ 0 & & a_{n-1n} & K \ddot{\div} \\ & & 0 & g_{1n} \ddot{\emptyset} \end{matrix}$$

Нетрудно заметить, что в этом случае возникает n треугольников $D(y, x_i, x^*)$, где $r(y, x_i) = r_i$, $r(x_i, x^*) = \gamma_{1i}$, $r(y, x^*) = \delta$. Для каждого из этих треугольников ($i = \overline{1, n}$) справедливы соотношения

$$|\delta - \gamma_{1i}| \leq r_i \leq \delta + \gamma_{1i}. \tag{2.41}$$

Обозначая $\mu_i = |\delta - \gamma_{1i}|$, $M_i = \delta + \gamma_{1i}$ левую и правую части (2.41) соответственно, подчеркнем, что если $i \neq j \in \{1, 2, \dots, n\}$

$$M_i \leq \mu_j, \tag{2.42}$$

то $r_i \leq r_j$, т.е. сегмент x_i ближе к запросу y , чем сегмент x_j .

Рассмотрим сначала некоторое формальное построение. Для заданной конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x^*\}$ сформируем два набора $\{\mu_i\}_{i=1}^n$ и $\{M_i\}_{i=1}^n$. Сделаем на сегментах x_1, \dots, x_n выборку объемом r , т.е. зафиксируем точки x_{s_1}, \dots, x_{s_r} – им соответствует набор чисел m_{s_1}, \dots, m_{s_r} . Обозначим

$$h(x_{s_1}, \dots, x_{s_r}) = \min\{m_{s_1}, \dots, m_{s_r}\}.$$

Исключим выбранные точки из рассмотрения

$$\{x_{p_1}, \dots, x_{p_{n-r}}\} = \{x_1, \dots, x_n\} \setminus \{x_{s_1}, \dots, x_{s_r}\}.$$

Точкам $x_{p_1}, \dots, x_{p_{n-r}}$ поставим в соответствие набор чисел $M_{p_1}, \dots, M_{p_{n-r}}$ и обозначим

$$H(x_{s_1}, \dots, x_{s_r}) = \min\{M_{p_1}, \dots, M_{p_{n-r}}\}.$$

Можно утверждать, что если

$$H(x_{s_1}, \dots, x_{s_r}) \leq h(x_{s_1}, \dots, x_{s_r}), \quad (2.43)$$

то исключение из конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x^*\}$ точек x_{s_1}, \dots, x_{s_r} оставляет в ней точку, ближайшую к запросу y .

Это действительно так. Неравенство (2.43) означает, что минимальное значение из числового набора m_{s_1}, \dots, m_{s_r} превосходит минимальное значение из числового набора $M_{p_1}, \dots, M_{p_{n-r}}$ (эти минимумы, безусловно, достигаются, т.к. рассматриваемые наборы конечны). Без ограничения общности положим: в первом наборе это m_{s_k} , во втором – M_{p_l} , где k и l – фиксированные числа из множеств $\{1, \dots, r\}$ и $\{1, \dots, n-r\}$ соответственно. Отсюда следует

$$\begin{array}{l} M_{p_l} \leq m_{s_1}, \\ M_{p_l} \leq m_{s_2}, \\ \dots \\ M_{p_l} \leq m_{s_r}. \end{array} \quad (2.44)$$

Но из этих неравенств с учетом (2.42) непосредственно вытекает, что для расстояний между запросом y и точками конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x^*\}$ справедливы неравенства $r_{p_l} \leq r_{s_1}, r_{p_l} \leq r_{s_2}, \dots, r_{p_l} \leq r_{s_k}$. Иначе говоря, точка x_{p_l} ближе к запросу, чем точки зафиксированной выборки x_{s_1}, \dots, x_{s_r} . Тем самым, ближайший к запросу сегмент находится среди точек $x_{p_1}, \dots, x_{p_{n-r}}$, что и требовалось показать.

Поясним эти рассуждения более детально для случая $n = 3$. Иначе говоря, изучим конфигурацию $\mathbb{K} = \{x_1, x_2, x_3, x^*\}$, учитывая все возможные ограничения. Из рис. 2.16 имеем треугольники $D(y, x_1, x^*)$, $D(y, x_2, x^*)$, $D(y, x_3, x^*)$, ограничивающие расстояния, которые подлежат оценке: $r(y, x_i), i = 1, 2, 3$. Кроме того, имеем 4 набора неравенств, ограничивающих известные расстояния:

– для $D(x^*, x_1, x_2)$

$$\begin{cases} \dot{\gamma}_{11} + \gamma_{12} \leq \alpha_{12}, \\ \dot{\gamma}_{11} + \alpha_{12} \leq \gamma_{12}, \\ \ddot{\gamma}_{12} + \alpha_{12} \leq \gamma_{11}; \end{cases} \quad (2.45)$$

– для $D(x^*, x_1, x_3)$

$$\begin{cases} \dot{\gamma}_{11} + \gamma_{13} \leq \alpha_{13}, \\ \dot{\gamma}_{11} + \alpha_{13} \leq \gamma_{13}, \\ \ddot{\gamma}_{13} + \alpha_{13} \leq \gamma_{11}; \end{cases} \quad (2.46)$$

– для $D(x^*, x_2, x_3)$

$$\begin{cases} \dot{\gamma}_{12} + \gamma_{13} \leq \alpha_{23}, \\ \dot{\gamma}_{12} + \alpha_{23} \leq \gamma_{13}, \\ \ddot{\gamma}_{13} + \alpha_{23} \leq \gamma_{12}; \end{cases} \quad (2.47)$$

– для $D(x_1, x_2, x_3)$

$$\begin{cases} \dot{\alpha}_{12} + \alpha_{13} \leq \alpha_{23}, \\ \dot{\alpha}_{12} + \alpha_{23} \leq \alpha_{13}, \\ \ddot{\alpha}_{13} + \alpha_{23} \leq \alpha_{12}. \end{cases} \quad (2.48)$$

Наконец, имеем ограничения, вытекающие из критерия сходства

$$\begin{cases} \dot{\rho}(y, x^*) = \delta \in D, \\ \dot{\rho}(y, x_1) \in D, \\ \dot{\rho}(y, x_2) \in D, \\ \dot{\rho}(y, x_3) \in D. \end{cases} \quad (2.49)$$

Далее для произвольного числа опорных точек (здесь $k = 1$) ограничения первого типа будем обозначать через $V_j(x_k^*)$, второго – $C(x_k^*)$, третьего – $L(x_k^*)$.

Найдем явный вид областей $V_1(x^*)$, $V_2(x^*)$, $V_3(x^*)$. Учитывая обозначения $\mu_i = |\delta - \gamma_{1i}|$, $M_i = \delta + \gamma_{1i}$, $i \in \{1, 2, 3\}$, получаем

$$\begin{aligned} \mu_1 &\leq \rho(y, x_1) \leq M_1, \\ \mu_2 &\leq \rho(y, x_2) \leq M_2, \\ \mu_3 &\leq \rho(y, x_3) \leq M_3. \end{aligned}$$

Ясно, что если, например, $\mu_3 \leq M_2$, то $\rho(y, x_2) \leq \rho(y, x_3)$, т.е. точка x_2 ближе к запросу y , чем x_3 . Если дополнительно $\mu_2 \leq M_1$, $\rho(y, x_2) \leq \rho(y, x_1)$, то x_1 ближе, чем x_2 . Таким образом, точка x_1 из конфигурации $\mathbb{K} = \{x_1, x_2, x_3, x^*\}$ ближайшая к запросу, если $|\delta - \gamma_{12}| \leq \delta + \gamma_{13}$ и $|\delta - \gamma_{13}| \leq \delta + \gamma_{11}$ или же (рассуждая по аналогии для $\mu_2 \leq M_3$ и $\mu_3 \leq M_1$) $|\delta - \gamma_{13}| \leq \delta + \gamma_{12}$ и $|\delta - \gamma_{12}| \leq \delta + \gamma_{11}$.

В явном виде требуемые области определяются в виде

$$V_1: \begin{cases} |\delta - \gamma_{12}| \leq \delta + \gamma_{13}, \\ |\delta - \gamma_{13}| \leq \delta + \gamma_{11}; \\ |\delta - \gamma_{13}| \leq \delta + \gamma_{12}, \\ |\delta - \gamma_{12}| \leq \delta + \gamma_{11}; \end{cases} \quad V_2: \begin{cases} |\delta - \gamma_{11}| \leq \delta + \gamma_{13}, \\ |\delta - \gamma_{13}| \leq \delta + \gamma_{12}; \\ |\delta - \gamma_{13}| \leq \delta + \gamma_{11}, \\ |\delta - \gamma_{11}| \leq \delta + \gamma_{12}; \end{cases} \quad V_3: \begin{cases} |\delta - \gamma_{11}| \leq \delta + \gamma_{12}, \\ |\delta - \gamma_{12}| \leq \delta + \gamma_{13}; \\ |\delta - \gamma_{12}| \leq \delta + \gamma_{11}, \\ |\delta - \gamma_{11}| \leq \delta + \gamma_{13}. \end{cases}$$

Отметим, что области V_1, V_2, V_3 лежат в четырехмерном пространстве, их пересечение возможно только на границах, что соответствует равноудаленности точек. Иначе говоря, если мы попали во внутреннюю часть области V_i , то точка x_i является ближайшей к запросу y . Общие ограничения этого типа представляются объединением областей: $V = V_1 \cup V_2 \cup V_3$. Окончатель-

но: область $D_2 \hat{I}_i^7$ для конфигурации $\mathbb{K}=\{x_1, x_2, x_3, x^*\}$ определяется областью V при условии верности ограничений C , задаваемых (2.45) – (2.48) и L – (2.49). Здесь для условий сходства L необходимо отметить следующее.

При проверке аналога (2.49) для выбора точки, ближайшей к запросу из двух анализируемых, нарушение условий идентифицировалось как вырожденный случай. Для (2.49) возможны следующие варианты: $x^* \notin U_D(y)$ – вырожденный случай (выбор опорной точки неудачен); только одна из точек x_1, x_2, x_3 попадает в D - окрестность точки y (именно она ближайшая, вырожденный случай); $x_1, x_2, x_3 \in U_D(y)$ – вырожденный случай (все точки отбрасываются); любые две точки из x_1, x_2, x_3 попадают в D - окрестность точки y (невырожденный случай, но он рассмотрен выше). Таким образом, в общем случае условия сходства L должны при анализе конфигурации $\mathbb{K}=\{x_1, x_2, x_3, x^*\}$ всегда выполняться.

Вернемся теперь к конфигурации $\mathbb{K}=\{x_1, \dots, x_n, x^*\}$. Неравенства (2.44) в более явном виде могут быть представлены следующим образом

$$\begin{array}{l} |\delta - \gamma_{1p_l} \mathbf{E} | \delta - \gamma_{1s_1} |, \\ \vdots \\ |\delta - \gamma_{1p_l} \mathbf{E} | \delta - \gamma_{1s_2} |, \\ \vdots \\ \dots\dots\dots \\ |\delta - \gamma_{1p_l} \mathbf{E}_{s_r} | \delta - \gamma_{1s_r} |. \end{array}$$

Эта система неравенств в арифметическом пространстве параметров (см. определение 2.1) $(a_{ij})_{i,j=1,\dots,n}, (g_{li})_{i=1,\dots,n}, d_{li} \in D_r \hat{I}_i^{C_{n+1}^{2+1}}$ задают область элиминации, попадание в которую позволяет отбросить из исходной конфигурации конкретный набор точек объема r . Это ограничения типа $V(x_{s_1}, \dots, x_{s_r}, x^*)$ (ограничения типа C и L от выбора точек x_{s_1}, \dots, x_{s_r} не зависят). Рассмотрев другую выборку, получаем аналогичные ограничения, а их

проверка обеспечивает исключение одной или другой из этих выборок. Иначе говоря, объединение по всем выборкам объема r дает построение области

$$V_r = \bigcup_{\substack{(s_1, \dots, s_r) \in \{1, \dots, n\} \\ s_i \neq s_j, i, j=1, r}} V(x_{s_1}, \dots, x_{s_r}, x^*). \tag{2.50}$$

Таким образом, окончательно для конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x^*\}$ получаем область элиминации

$$D_r = V_r \mid C \mid L, \tag{2.51}$$

где

$$C = \bigcup_{\substack{i, j, k=0, n \\ i \neq j \neq k}} C_{ijk}, \quad C : \begin{cases} \delta + \alpha_{ij} + \alpha_{ik} \leq \alpha_{jk}, \\ \delta + \alpha_{jk} + \alpha_{ik} \leq \alpha_{ij}, \\ \delta + \alpha_{ik} + \alpha_{ij} \leq \alpha_{jk}; \end{cases} \tag{2.52}$$

$$L : \begin{cases} \delta \in D, \\ \delta + \alpha_{01} \in D, \\ \dots \\ \delta + \alpha_{0n} \in D; \end{cases} \tag{2.53}$$

(здесь и выше: $x_0 = x^*$, т.е. " $i \in \{1, \dots, n\} \Rightarrow a_{0i} = a_{i0} = g_{li}$).

Полученные результаты дают возможность построения области элиминации в общем случае для конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$ (см. рис. 2.3), представляемой матрицей P (2.11).

Рассмотрим выборку x_{s_1}, \dots, x_{s_r} из точек x_1, \dots, x_n . Обозначим через $D_r(x_i^*)$ область элиминации (2.51) для i -ой опорной точки. Можно утверждать, что

$$D_r = \bigcup_{i=1}^m D_r(x_i^*). \tag{2.54}$$

Действительно, по определению D_r – это такая область в $\mathbb{R}^{C_{n+1}^2+1}$, по-

падение в которую кортежа $\overline{(a_{ij})_{i,j=1,n}}, \overline{(b_{ij})_{i,j=1,m}}, \overline{(g_{ij})_{i=1,n, j=1,m}}, \overline{(d_i)_{i=1,m}}$

означает существование выборки x_{s_1}, \dots, x_{s_r} , среди элементов которой нет точки, ближайшей к запросу y . С другой стороны, ясно, что хотя бы для одной опорной точки x_i^* происходит попадание кортежа параметров в $D_r(x_i^*)$ (иначе не может быть по определению), что и предопределяет (2.54).

Из (2.51) непосредственно следует

$$D_r(x_i^*) = V_r(x_i^*) \cap C(x_i^*) \cap L(x_i^*).$$

Нетрудно заметить, что ограничения неравенств треугольника $C(x_i^*)$ и ограничения сходства $L(x_i^*)$ строятся по аналогии с (2.52) и (2.53) соответственно, а построение $V_r(x_i^*)$ имеет специфику.

Рассмотрим фиксированную выборку x_{s_1}, \dots, x_{s_r} из точек конфигурации x_1, \dots, x_n . Как и ранее, сформируем два набора чисел, соответствующих этой выборке: $m_{s_1}(x_i^*), \dots, m_{s_r}(x_i^*)$, $M_{p_1}(x_i^*), \dots, M_{p_{n-r}}(x_i^*)$, $i = \overline{1, m}$. Обозначим

$$g(x_{s_1}, \dots, x_{s_r}) = \min_{i=1, m} \min_{j=1, r} \{m_{s_j}(x_i^*)\}, \quad (2.55)$$

$$G(x_{s_1}, \dots, x_{s_r}) = \min_{i=1, m} \min_{j=1, n-r} \{M_{p_j}(x_i^*)\}. \quad (2.56)$$

Можно утверждать: если имеет место

$$G(x_{s_1}, \dots, x_{s_r}) \leq g(x_{s_1}, \dots, x_{s_r}), \quad (2.57)$$

то, отбросив из конфигурации $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$ точки x_{s_1}, \dots, x_{s_r} , мы не потеряем точку, ближайшую к y .

Действительно, поскольку мы имеем дело с конечными наборами, то $G(x_{s_1}, \dots, x_{s_r})$ достигается для какой-то опорной точки x_k^* . Это означает, что

$$G(x_{s_1}, \dots, x_{s_r}) = \min \{M_{p_1}(x_k^*), \dots, M_{p_{n-r}}(x_k^*)\}.$$

С другой стороны, рассмотрим соответствующий этой опорной точке набор чисел $m_{s_1}(x_k^*), \dots, m_{s_r}(x_k^*)$. Из (2.55) следует, что

$$g(x_{s_1}, \dots, x_{s_r}) \leq \min \{m_{s_1}(x_k^*), \dots, m_{s_r}(x_k^*)\}. \quad (2.58)$$

Но тогда из (2.57) и (2.58) следует

$$\min \{M_{p_1}(x_k^*), \dots, M_{p_{n-r}}(x_k^*)\} \leq \min \{m_{s_1}(x_k^*), \dots, m_{s_r}(x_k^*)\},$$

что фактически соответствует доказанному выше соотношению (2.43) для опорной точки x_k^* , т.е. соотношение (2.57) верно.

Таким образом, изучена конфигурация $\mathbb{K} = \{x_1, \dots, x_n, x_1^*, x_2^*, \dots, x_m^*\}$, а именно: показано, что в самом общем виде при анализе близости точек x_1, \dots, x_n к некоторому запросу y можно исключать из рассмотрения (без трудоемкого вычисления расстояний до запроса) r точек, используя предварительно вычисленные расстояния между точками и *on line* вычисленные расстояния до опорных точек $x_1^*, x_2^*, \dots, x_m^*$. Построение областей элиминации позволяет придерживаться различных стратегий поиска, например, сразу находить ближайшую, анализируя область D_{n-1} , итерационно понижать размерность конфигурации, что особенно важно при стратифицированном поиске кластеризованных данных. Необходимо особо подчеркнуть, что границы областей элиминации любого порядка линейны. Количество ограничений велико, поскольку весьма значительна размерность анализируемого арифмети-

ческого пространства, но они весьма просты. Кроме того, основные вычисления, обеспечивающие построение областей элиминации по своей природе прелиминарны.

Выводы по второму разделу

1. Эффективный доступ к видеоданным по их содержанию базируется на разработке методов on line представления видеопотоков в виде однородных сегментов, устраняющих семантический конфликт между низкоуровневыми моделями видео и высокоуровневыми целями предметно-ориентированной трактовки.

2. В больших видеокolleкциях перспективным направлением является организация семантического поиска с запросами по образцу с парадигмой элиминации объектов без вычисления расстояний до них, что создает предпосылки для получения рациональных по быстродействию вычислительных моделей.

3. Построение в общем виде областей элиминации в произвольном метрическом пространстве обеспечивает для произвольных конфигураций при произвольных наборах опорных данных (центрах категоризации, кластеризации и т.п.) реализовывать различные стратегии поиска, характерным свойством которых является исключение (непосредственно при поиске) необходимости трудоемких вычислений расстояний от запроса до элементов библиотек видео.

РАЗДЕЛ 3

МЕТОДЫ КЛАСТЕРИЗАЦИИ И СРАВНЕНИЯ ДАННЫХ В ВИДЕОПОТОКАХ В ЗАДАЧАХ ИНФОРМАЦИОННОГО ПОИСКА

Раздел посвящен описанию моделей и методов кластеризации больших объемов данных, что применимо к задаче информационного поиска в видеоданных. Предложены подходы, основанные на плотностях, которые отличаются вычислительной простотой, что является важным критерием при обработке видеопоследовательностей. В основу модификаций положено распространение методов традиционной векторной кластеризации на фрагментную обработку визуальной информации. Под фрагментами понимаются как «прямоугольные окна» при кластеризации-сегментации изображений, так и собственно видеокadres при кластеризации коллекций изображений.

Предложены и обоснованы инструменты сравнения изображений-запросов с кадрами видеопоследовательностей для определения достаточно корректных результатов информационного поиска. С этой целью введена метрика на разбиениях, позволяющая сравнивать «множества множеств», что удобно не только при сравнении результатов кластеризации, но и при использовании сегментаций изображения-запроса и кадров видео. Рассмотрены вопросы синтеза характеристик кадров на основе оценок разбиений поля зрения с целью построения векторов признаков для решения задачи сегментации видеопоследовательностей.

Раздел основан на работах автора [10-12, 14-17, 47, 84, 99, 183, 217, 218, 229, 254, 255, 261, 262].

3.1 Фрагментные методы кластеризация визуальной информации

Особую группу методов кластеризации образуют алгоритмы, предназначенные для обработки информации, хранящейся в сверхбольших базах данных (VLDB), где на первый план выходят быстрдействие и простота

численной реализации. В данной ситуации в качестве достаточно эффективных показали себя методы кластеризации, основанные на плотности распределения данных, при этом применяемое здесь понятие плотности по смыслу близко к плотности распределения, используемой в теории вероятностей и математической статистике. Именно методы, основанные на плотности, позволяют формировать кластеры произвольной формы в условиях, когда обрабатываемые данные искажены возмущениями, а само число кластеров p заранее неизвестно.

Наиболее распространенным методом из этого класса является DBSCAN (Density-Based Spatial Clustering of Applications with Noise), отличающийся вычислительной простотой и устойчивостью к возмущениям. В основе метода лежит ряд понятий и определений, основными из которых являются внутренние и граничные точки, D-досыгаемость (Density reachability) и D-связность (D-connectedness), порог ($\epsilon = Eps$) и минимальное количество наблюдений в кластере (MinPts). При этом полагается, что произвольная точка $x(t)$ непосредственно досыгаема из любой точки $x(q)$, если она удалена в смысле принятой метрики (традиционно евклидовой) на расстояние, не превышающее заданного априори порога ϵ . Именно $\epsilon = Eps$ является основным исходным параметром алгоритма, задаваемым пользователем, при этом подразумевается, что этот пользователь является квалифицированным специалистом в конкретной предметной области, в рассматриваемом нами случае – специалистом в области видеообработки и компьютерных наук. На основании выбранного порога формируется ϵ -окрестность точки $x(q)$, которую образуют все точки x , удовлетворяющие неравенству

$$\|x - x(q)\| < \epsilon.$$

Что же касается минимального количества наблюдений в кластере MinPts, то это также параметр, выбираемый экспериментально, обычно

$N > \text{MinPts}^3 N + 1$, при этом утверждается, что если в ϵ -окрестности точки $x(q)$ содержится не менее MinPts точек, то $x(k)$ и $x(q)$ относятся к одному кластеру.

Если рассмотреть понятие D-достижимости, то точка $x(k)$ является D-достижимой из $x(q)$, если можно сформировать «цепочку» наблюдений $x(q), \dots, x(r), \dots, x(k)$ такую, что каждый ее элемент $x(r)$ непосредственно достижим своими соседями.

Важным фактом является то, что понятие D-достижимости не является симметричным. Если $x(k)$ лежит на границе кластера, то симметрия нарушается, т.е. эта точка может содержать меньше, чем MinPts точек в своей окрестности. Именно нахождением таких граничных точек и завершается формирование кластеров. Понятно, что в этом случае априорно предполагается, что формируемые кластеры не пересекаются. Все наблюдения, принадлежащие конкретному кластеру и имеющие не менее MinPts наблюдений в своей ϵ -окрестности, называются внутренними точками кластера. Описанная асимметрия порождает понятие D-связности, при этом точки $x(k)$ и $x(q)$ называются D-связными, если они обе достижимы из $x(r)$, при этом очевидно, что понятие D-связности – симметрично.

Опираясь на введенные понятия, можно определить кластер (сегмент) как множество D-связных точек, причем, что важно, такую формулировку можно распространить и на другие подходы к задаче кластеризации, где используется понятие метрики. Сам же процесс кластеризации может быть сведен к последовательности элементарных действий, которая, стартуя из произвольной точки, находит множество D-связных с ней данных. После того как все такие наблюдения найдены, процедура запускается вновь из произвольной прежде непроанализированной точки и находит все относящиеся к ней D-связные данные. Так происходит до исчерпания всех N наблюдений анализируемой группы объектов-образов. Множество всех объектов, не включенных ни в один кластер и содержащих менее, чем MinPts наблюдений

в своей ϵ -окрестности, в рамках стандартного подхода рассматриваются как шумы, хотя может оказаться, что именно в этих точках содержится уникальная информация, которая должна быть тщательно проанализирована вне рамок DBSCAN.

Следует отметить, что метод DBSCAN в силу своей простоты и наглядности получил широкое распространение во множестве прикладных задач анализа данных, в том числе и для сегментации различного рода изображений, где в соответствие каждому пикселю ставится многомерный набор признаков, задаваемый в векторной форме. Понятно, что количество таких векторов в выборке может быть очень велико. Конечно, в рассмотрение могут быть введены некоторые дополнительные характеристики анализируемого изображения, однако, во всех случаях для успешного решения задачи квалификация пользователя должна быть достаточно высока. Именно это обстоятельство, а также низкий уровень формализации этого метода и чувствительность к выбору параметров алгоритма породили целый ряд модификаций, лишенных некоторых из недостатков прототипа.

На сегодня известен целый ряд модификаций, причем каждая новая из них стремилась минимизировать влияние субъективного фактора, связанного с каждым конкретным пользователем и дополнительно формализовать базовую процедуру.

Одной из таких модификаций является DBCLASD (Distribution-Based Clustering of LArge Spatial Databases), с помощью которой также можно формировать кластеры произвольной формы из «зашумленных» данных. Основным достоинством DBCLASD является возможность обработки данных в последовательном (on line) режиме, при этом каждый вновь поступающий на обработку образ может быть отнесен к тому или иному кластеру на основе анализа распределений расстояний от анализируемого образа до каждого из кластеров основанного на c^2 -тесте. Данный метод обладает пониженной чувствительностью к выбору параметров Eps и $MinPts$, однако, в его основе лежит предположение, что данные в каждом кластере подчинены равномер-

ному закону распределения, что далеко не всегда бывает в реальных задачах, особенно связанных с обработкой изображений.

Развитием DBSCAN также является алгоритм OPTICS (Ordering Points to Identify the Clustering Structure), позволяющий решать задачи кластеризации в условиях, когда кластеры имеют не только различную форму, но и разную плотность распределения данных в каждом классе. OPTICS кроме основных понятий и определений, используемых в DBSCAN, вводит дополнительные характеристики для каждого наблюдения такие, как внутреннее расстояние (core distance) и расстояние досягаемости (reachability distance). OPTICS структурно эквивалентен DBSCAN, обладает расширенными функциональными возможностями, однако с вычислительной точки зрения значительно сложнее и медленнее прототипа, что усложняет его использование в задачах, связанных с VLDB.

Интересным гибридом DBSCAN и популярного метода k -средних является Bridge, с помощью которого исходный массив данных сначала обрабатывается с помощью стандартного метода k -средних, а затем к каждой сформированной группе данных применяется DBSCAN, подавляющий шумы и восстанавливающий плотность распределения данных в каждом кластере. Понятно, что Bridge с вычислительной точки зрения сложнее, чем DBSCAN, однако в настоящее время он используется для решения ряда задач, связанных с VLDB.

Наиболее формализованным и математически обоснованным алгоритмом, основанным на плотности, является DENCLUE (DENsity-based CLUstEring), созданный для обработки больших массивов мультимедийных данных, формируя кластеры произвольной формы при высоком уровне шумов. Данный метод основан на ряде предположений:

- влияние каждого вектора-образа на соседние наблюдения формально может быть описано с помощью некоторой функции, обычно ядерной, называемой функцией влияния, которая описывает взаимосвязь всех наблюдений в некоторой окрестности данного образа;

– общая плотность распределения данных в n -мерном пространстве признаков формально описывается как сумма функций влияния каждого наблюдения;

– кластеры определяются как окрестности аттракторов плотности (D-аттракторов), являющихся, по сути, локальными максимумами общей функции плотности распределения данных.

Для некоторой произвольной точки в пространстве признаков y ее влияние на образ x может быть описано с помощью функции влияния

$$f^y(x) = f(x, y),$$

при этом наиболее часто в качестве таких функций используется либо прямоугольная конструкция

$$f(x, y) = \begin{cases} 0, & \text{если } D(x, y) > \sigma; \\ 1, & \text{в противном случае,} \end{cases} \quad (3.1)$$

либо гауссиан

$$f(x, y) = \exp\left\{-\frac{D^2(x, y)}{2\sigma^2}\right\}, \quad (3.2)$$

где σ – параметр ширины ядерной функции,

$D(x, y)$ – расстояние, обычно евклидово, между точками x и y .

Тогда для множества наблюдений $X = \{x(1), x(2), \dots, x(k), \dots, x(N)\}$ общая функция плотности может быть представлена в виде

$$f^x(x) = \prod_{k=1}^N f(x, x(k)). \quad (3.3)$$

Функция (3.3), является суммой ядерных функций, характеризуется наличием множества локальных экстремумов-максимумов, именуемых D-аттракторами, каждый из которых представляет отдельный кластер и может быть определен с помощью тех или иных оптимизационных процедур. Здесь же заметим, что использование функции влияния (3.1) превращает DENCLUE в стандартный DBSCAN, а если $f(x, y)$ непрерывна и дифференцируема, как например (3.2), для нахождения локальных максимумов может быть использована стандартная градиентная оптимизация. При этом произвольная точка x притягивается к D-аттрактору x^* , если последовательность итераций

$$x^i = x^{i-1} + \eta \frac{\tilde{N}f^x(x^{i-1})}{\|\tilde{N}f^x(x^{i-1})\|}; i=1, 2, \dots; x^0 = x \quad (3.4)$$

сходится к x^* .

Если в качестве $f^x(x)$ используются соотношения (3.2), (3.3), то

$\tilde{N}f^x(x) = \mathring{\mathbf{a}} \sum_{k=1}^N (x(k)-x)f(x, x(k))$, а процедура (3.4) приобретает вид

$$x^i = x^{i-1} + \eta \frac{\mathring{\mathbf{a}} \sum_{k=1}^N (x(k) - x^{i-1})f^{x^i}, x(k)}{\left\| \mathring{\mathbf{a}} \sum_{k=1}^N (x(k) - x^{i-1})f^{x^i}, x(k) \right\|}, \quad (3.5)$$

где η – параметр шага поиска.

Каждый из D-аттракторов характеризуется собственной функцией плотности

$$f^{x^*}(x) = \mathring{\mathbf{a}} \sum_{x(k) \uparrow \text{near } x^*} f(x, x(k)), \quad (3.6)$$

где $\text{near } x^* = \{x(k) : D(x^*, x(k)) \leq \sigma_{\text{near}}\}$,

а ее экстремум определяет координаты центроида кластера.

Конечно, с вычислительной точки зрения DENCLUE сложнее любого из описанных выше алгоритмов, однако к его преимуществам следует отнести высокий уровень формализации, а также то, что он обобщает рассмотренные выше процедуры кластеризации, основанные на плотности.

При решении задач кластеризации всегда предполагается, что каждое многомерное наблюдение-образ описывается n -мерным вектором $x(k)$, а весь процесс решения связан именно с векторными операциями. В ситуации, когда имеется большая коллекция изображений, подлежащих кластеризации, каждое двумерное изображение сначала должно быть подвергнуто векторизации, далее решается задача кластеризации, а ее результат подвергается де-векторизации переводящей векторное описание в матричную форму. Существенно упростить процесс кластеризации массивов можно, не переводя их в векторную форму, а непосредственно, оперируя с матрицами. Таким образом, набором исходных образов является набор матриц $x(k) = \{x_{i_1 i_2}(k)\}$, $x_1 = 1, 2, \dots, m$; $x_2 = 1, 2, \dots, n$; $k = 1, 2, \dots, N$, $x(k) \hat{=} m \times n$. Далее, вводя вместо стандартной векторной евклидовой нормы ее сферический матричный аналог

$$D_S^2(x, y) = Sp(x - y)(x - y)^T,$$

можно ввести матричную функцию влияния

$$f_S(x, y) = \exp\left\{-\frac{D_S^2(x, y)}{2\sigma^2}\right\} = \exp\left\{-\frac{Sp(x - y)(x - y)^T}{2\sigma^2}\right\} \quad (3.7)$$

и матричную функцию плотности $f_S^X(x) = \prod_{k=1}^N f_S(x, x(k))$.

При этом произвольная $(m \times n)$ матрица-образ x притягивается к мат-

ричному D-аттрактору x^* , если последовательность итераций типа (3.4)

$$x^i = x^{i-1} + \eta \frac{\begin{pmatrix} \nabla f_S^x(x^{i-1}) \\ \nabla x_{i1} \\ \nabla x_{i2} \end{pmatrix}}{\left(Sp \begin{pmatrix} \nabla f_S^x(x^{i-1}) \\ \nabla x_{i1} \\ \nabla x_{i2} \end{pmatrix} \begin{pmatrix} \nabla f_S^x(x^{i-1}) \\ \nabla x_{i1} \\ \nabla x_{i2} \end{pmatrix}^T \right)^{\frac{1}{2}}}, \quad i=1, 2, \dots; \quad x^0 = x \quad (3.8)$$

сходится к x^* . Здесь $\begin{pmatrix} \nabla f_S^x(x) \\ \nabla x_{i1} \\ \nabla x_{i2} \end{pmatrix}$ – $(m \times n)$ матрица, образованная производными $f_S^x(x)$ по компонентам матрицы x .

Если в качестве матричной функции используется выражение (3.7), алгоритм оптимизации (3.8), может быть переписан в простой форме

$$x^i = x^{i-1} + \eta \frac{\sum_{k=1}^N r(k, i-1)}{\left(Sp \left(\sum_{k=1}^N r(k, i-1) \left(\sum_{k=1}^N r(k, i-1) \right)^T \right) \right)^{\frac{1}{2}}},$$

где $r(k, i-1) = (x(k) - x^{i-1}) f_S(x^{i-1}, x(k))$.

Отметим, что, по сути, это расширение (3.5) на матричный случай.

Использование вместо векторного описания его матричного аналога позволяет существенно повысить быстродействие процесса обработки информации и избежать ряда проблем, возникающих в задаче кластеризации данных, описываемых векторами высокой размерности. Особо следует подчеркнуть, что алгоритмы такого типа используют фрагменты

$$p_{ij}(m, n) = \{i, i+1, \dots, m-1\} \times \{j, j+1, \dots, n-1\}$$

поля зрения $P = \{1, 2, \dots, M\} \cup \{1, 2, \dots, N\}$, которые задают его разбиение

$$\cup p_{ij}(m,n) = P, \quad p_{ij}(m,n) \cap p_{i'j'}(m,n) = \emptyset \quad i \neq i', j \neq j'$$

или покрытие. При анализе видеопоследовательности $\mathcal{P} = \{P_1, P_2, \dots, P_K\}$ имеем $p_{ij}^{(l)}(m,n) \in P_l, p_{ij}^{(l+1)}(m,n) \in P_{l+1}, \dots, p_{ij}^{(k)}(m,n) \in P_k, l \in \{1, 2, \dots, K\}$. Отсюда следует, что фрагментные методы равнозначно могут применяться к статической и динамической визуальной информации, обеспечивая пространственно-временную сегментацию видеоданных.

Кроме предложенного подхода также можно использовать матричные модификации других методов кластеризации, в частности, X -средних. Исходной информацией для решения задачи сегментации-кластеризации, как и ранее, является массив матричных наблюдений $X = \{x(1), x(2), \dots, x(N)\}$, $x(k) = \{x_{i_1 i_2}(k)\} \in R^{m \times n}; i_1 = 1, 2, \dots, m; i_2 = 1, 2, \dots, n; k = 1, 2, \dots, N$, который должен быть разделен в процессе самообучения на p сегментов-кластеров, описываемых центроидами $C = \{C(1), C(2), \dots, C(p)\}$. При этом число кластеров p априорно не задается, а должно быть определено непосредственно в процессе обработки информации в диапазоне $2 \leq p_{min} \leq p \leq p_{max} \leq N - 1$.

Основная идея метода X -средних состоит в многократном, применении к исходному массиву данных X алгоритма k -средних с разными значениями p и оценивании получаемых результатов с помощью того или иного критерия, основанного на байесовском оценивании.

Среди критериев качества кластеризации, не опирающихся на статистические предположения и пригодных для нахождения количества кластеров в массиве данных, в качестве весьма эффективного показал себя критерий Цалиньского-Харабаша, который может быть адаптирован на случай матричных объектов. Для оценки качества разбиения критерий

Цалинского-Харабаша может быть записан в виде

$$CH(p) = \frac{\frac{1}{p-1} SpS_B(p)}{\frac{1}{N-p} SpS_w(p)},$$

где $S_B(p) = \sum_{l=1}^p N_l (C(l) - \bar{C})(C(l) - \bar{C})^T$ – матрица межкластерного

рассеивания,

$S_w(p) = \sum_{l=1}^p \sum_{k=1}^N m(x(k), C(l))(x(k) - C(l))(x(k) - C(l))^T$ – матрица внутри-

кластерного рассеивания,

$$C(l) = \frac{1}{N_l} \sum_{x(k) \in Cl_l} x(k) = \frac{\sum_{k=1}^N m(x(k), C(l))x(k)}{\sum_{k=1}^N m(x(k), C(l))} \text{ – центр оид (центр тяжести)}$$

кластера Cl_l ,

$$\bar{C} = \frac{1}{N} \sum_{l=1}^p N_l C(l) \text{ – матричный центр тяжести массива } X.$$

Таким образом, предлагаемая модификация метода X -средних для обработки изображений на основе K -средних и критерия Цалинского-Харабаша может быть реализована в виде последовательности этапов:

- задание $p = p_{min}$, решение задачи с помощью k -средних и расчет $CH(p_{min})$;
- расщепление любого из кластеров, решение задачи с помощью k -средних и расчет $CH(p_{min} + 1)$;
- при $CH(p + 1) \leq CH(p)$ полагается, что p есть наилучшая оценка количества кластеров в массиве X .

Итак, предложен модифицированный метод кластеризации X -средних для решения задачи сегментации изображений. Особенностью предложенной

модификации является возможность обработки матричных сигналов в отсутствии информации о статистических характеристиках этих сигналов и числа кластеров, которое автоматически определяется в процессе анализа исходного массива данных.

Кроме метода X -средних также применим и матричный вариант J -средних для задачи сегментации отдельных кадров. Суть J -средних состоит в том, что при нахождении некоторого центроида $C(l)$, в котором процедура кластеризации останавливается, организуется процесс перемещения этого центроида в незанятые точки, находящиеся в его окрестности. Такие скачки «центроид – незанятая точка» производится до тех пор, пока не будет найден более глубокий экстремум, соответствующий меньшему значению целевой функции E . С формальной точки зрения суть J -средних состоит в следующем: пусть получено некоторое текущее разбиение $P_p = \{Cl_1, Cl_2, \dots, Cl_p\}$ с соответствующим набором центроидов $C = \{C(1), C(2), \dots, C(p)\}$. В рассмотрение вводится набор областей соседства $H_l(P_p)$, где $l \in \{1, 2, \dots, p\}$, а $H_l(P_p)$ определяет все возможные перемещения центроида $C(l)$ в точки $x(s)$, принадлежащие его окрестности, где $s \in \{1, 2, \dots, \overset{\circ}{N}\}$. Понятно, что при таких перемещениях постоянно корректируется и набор центроидов C_s , приобретая вид

$$C_s = \{C(1), \dots, C(l-1), x(s), C(l+1), \dots, C(p)\}.$$

Сам алгоритм в этом случае можно описать в виде следующей последовательности шагов:

1. Задание начального, достаточно произвольного, разбиения $P_p = \{Cl_1, Cl_2, \dots, Cl_p\}$, определение его центров $C(1), C(2), \dots, C(p)$ и расчет целевой функции $E(x(k), C(l))$, соответствующей этому разбиению.

2. Определение всех незанятых точек, т.е. наблюдений, не

совпадающих с центроидами или не находящихся в их непосредственной близости.

3. Скачки в окрестностях центроидов, т.е. их замена незанятыми точками $x(s)$ с расчетом целевых функций $E(x(k), C(q), x(s))$ вида

$$E(x(k), C(q), x(s)) = \sum_{k=1}^{\dot{N}} \left(\sum_{q=1}^p m(x(k), C(q)) Sp(AA^T) - m(x(k), x(s)) Sp(BB^T) \right)$$

4. Принятие решений об остановке или дальнейшем поиске. Если

$$E(x(k), C(q), x(s)) > E(x(k), C(q)),$$

то полагается, что оптимальное решение было найдено на предыдущих итерациях, и процесс кластеризации останавливается. В противном случае выбирается решение, соответствующее наименьшему значению целевой функции $E(x(k), C(q), x(s))$ для всех реализованных $x(s)$. Полученное разбиение полагается начальным и производится возврат к шагу 2.

Главная особенность J -средних, отличающая его от всех методов кластеризации с использованием центроидов, – это наличие шага 3 (скачки в окрестностях), удачная реализация которого определяет эффективность полученного решения в целом. А соответственно имеет смысл представить этот шаг в виде совокупности действий:

– Формирование нового центроида $C(p+1)$ в незанятой точке $x(s)$ и нахождение «наилучшего» центроида $C(l)$, дающего наибольшее значение целевой функции $E(x(k), C(q), x(s))$.

– Удаление центроида $C(l)$ и замена его на $x(s)$.

– Формирование нового разбиения P_p^{new} , в котором $C^{new}(l) \circ x(s)$.

Наибольшие проблемы, связанные с практическим использованием метода J -средних, возникают при задании размеров (диаметра) окрестности

скачка вокруг каждого центроида. Слишком малая окрестность может не обнаружить незанятые точки с меньшими значениями целевой функции, а слишком большие окрестности резко увеличивают вычислительную сложность алгоритма, особенно в случаях, когда эти окрестности перекрываются. Основной же проблемой данного алгоритма является необходимость многократного полного перебора всех сочетаний наблюдений и рассчитанных центроидов. Понятно, что при больших \dot{N} и p подход, основанный на J -средних оказывается неэффективным.

Необходимо особо подчеркнуть, что при использовании традиционных методов кластеризации для сегментации изображений, как правило, ослабляется одно из требований – связность областей в поле зрения. В связи с этим для повышения валидности сегментации необходима постобработка. Удобным инструментом оказываются алгоритмы бинарной морфологии, поскольку каждую область можно трактовать как двухградационное изображение. Кроме того, различные характеристики (прежде всего геометрические) таких областей создают предпосылки для формирования векторов признакововых описаний изображения в целом. Перейдем к анализу метрических свойств разбиений, а затем конкретизируем возможности синтеза признакововых описаний, сравнение которых может осуществляться на базе традиционных метрик.

3.2. Свойства преобразований сегментаций изображений

Рассмотрим свойства изменений результатов сегментации изображений, на основе которых построим функционал для метрического сравнения различных разбиений в задачах информационного поиска.

Рассмотрим произвольное изображение, например, ключевой кадр $\mathbb{B}_{ab} \hat{=} S_{ab}^*$ некоторого сегмента. Будем ассоциировать изображение с множеством (носителем) B и его разбиением на собственные подмножества P с

конечным числом элементов, т.е.

$$\begin{cases} P = \{p_1, \dots, p_n\}, \\ p_k \subseteq B, k = \overline{1, n}, \\ \bigcup_{k=1}^n p_k = B, \\ p_i \cap p_j = \emptyset, i, j = \overline{1, n}, i \neq j. \end{cases} \quad (3.9)$$

Будем считать, что на $B \subseteq \mathbb{R}^2$ задана мера $\mu(B)$ в виде количество точек, попадающих в область; она конечна, откуда следует и конечность мер для элементов произвольных разбиений.

На множестве произвольных разбиений $P = \{p_1, \dots, p_n\}$, $G = \{g_1, \dots, g_m\}$ множества B введем функционал

$$r(P, G) = \sum_{k=1}^n \sum_{l=1}^m |p_k \setminus g_l| + |p_k \cap g_l|,$$

где « \setminus » – операция симметрической разности множеств.

Подчеркнем, что изучение свойств разбиений для оценки сходства изображений по результатам сегментации сводится к проверке выполнения аксиом рефлексивности, симметричности.

Нетрудно заметить, что из вида функционала непосредственно следует, что он удовлетворяет аксиоме симметричности. Докажем рефлексивность. Будем считать, что разбиение P состоит из n элементов, тогда

$$\begin{aligned} r(P, P) &= \sum_{k=1}^n \sum_{l=1}^n |p_k \setminus p_l| + |p_k \cap p_l| = \\ &= \sum_{k=1}^n |p_k \setminus p_k| + |p_k \cap p_k| + 2 \sum_{\substack{k, l=1 \\ k > l}}^n |p_k \setminus p_l| + |p_k \cap p_l|, \end{aligned}$$

т.е. $\rho(P, P)$ состоит из n слагаемых с совпадающими номерами, а значит, и с

совпадающими элементами разбиения, и из $n^2 - n$ слагаемых с различными элементами разбиения. Первая группа слагаемых состоит из нулей, т.к. $|p_k \vee p_k| = 0$, для $k = \overline{1, n}$, а вторая группа дает нулевые слагаемые, поскольку $|p_k \wedge p_l| = 0$ для " $k, l = \overline{1, n}$ и $k \neq l$ ". Таким образом, в одну сторону рефлексивность доказана.

Покажем рефлексивность в обратную сторону, т.е. пусть для двух, в общем случае, различных разбиений P и G имеет место $r(P, G) = 0$. Поскольку все слагаемые функционала $r(P, G)$ положительны, то он равен 0 только тогда, когда каждое из них равно 0. Выберем некоторый элемент разбиения $p \in P$. В исходном функционале он входит в набор «нулевых» слагаемых $|p \vee g_l| |p \wedge g_l| = 0$, где $l = \overline{1, m}$, $G = \{g_1, \dots, g_m\}$. Если предположить, что p не является элементом разбиения G , то для всех $g_l \in G$ выполняется $|p \vee g_l| \neq 0$. Но тогда для всех номеров l должно выполняться $|p \wedge g_l| = 0$. Однако это возможно только тогда, когда $p \notin \mathcal{A}$, т.к. $p \in P$, а весь набор $G = \{g_1, \dots, g_m\}$ покрывает множество B . В то же время p – элемент покрытия P этого же изображения B и, естественно, $p \in \mathcal{A}$. Тогда найдутся элементы разбиения G , покрывающие p и имеющие с ним непустые пересечения (рис.3.1). Но тогда для этих элементов $|p \vee g_l| \neq 0$, т.е. получаем противоречие.

В итоге можно утверждать, что произвольный элемент $p \in P$ является элементом G или $p \in G$. Ясно, что в силу симметрии аналогично можно показать $G \subseteq P$, т.е. $P = G$. Таким образом, рефлексивность доказана.

Для конечного множества $B = \{1, \dots, N\}$ и множества всех его подмножеств \mathcal{Y}_B (разбиений на конечный набор подмножеств) зафиксируем элемент $P \in \mathcal{Y}_B$, т.е. $P = \{p_1, \dots, p_n\}$, где $p_i \subseteq B$, $\bigcup_{i=1}^n p_i = B$ и $p_i \cap p_j = \emptyset$, когда $i \neq j$. Тогда, изучим функционал сходства $r(P, G)$

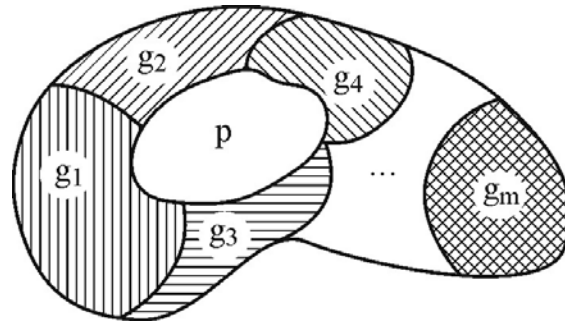


Рисунок 3.1 – Пересечение покрытий P и G множества B

$$r(P, G) = \sum_{i=1}^n \sum_{j=1}^m |p_i \vee g_j| |p_i \wedge g_j| \quad (3.10)$$

в плане анализа его изменений при вариациях разбиений.

Нетрудно заметить, что если к множеству B добавить $(N + 1)$ -ый элемент, то разбиение $P \hat{=} Y_B$ может трансформироваться двумя способами: с одной стороны, этот элемент может добавляться как отдельно взятый элемент разбиения

$$P \hat{=} \{p_1, \dots, p_n, p_{n+1}\}, \quad (3.11)$$

где $P \hat{=} Y_{B \hat{=} B \hat{=} \{1, \dots, N + 1\}}$.

С другой стороны, он может «вливаться» в один из элементов разбиения P , тогда новое разбиение множества $B \hat{=}$ будет выглядеть

$$P \hat{=} \{p_1, \dots, \{p_k, N + 1\}, \dots, p_n\}, \quad (3.12)$$

где $k = \overline{1, n}$.

При $k = \overline{1, n}$ трансформацию вида (3.11) обозначим $F_0(P)$, а (3.12) – через $F_k(P)$.

Рассмотрим теперь, как трансформируется функционал (3.10) при добавлении одного элемента.

Рассмотрим сначала несколько вспомогательных формальных положений. Нетрудно заметить, что если над элементами $G, P \hat{Y}_B$ выполнить трансформацию F_i , то функционал (3.10) изменится: область определения станет $Y_{B\phi} B\phi = \{1, 2, \dots, N+1\}$. При этом новый функционал $r(P, G)$, естественно, связан с первоначальным, а именно, вид функционала не меняется: он остается в нотации (3.10).

Покажем, что для любых двух разбиений P, G множества B имеет место

$$r(P, G) = r(F_0(P), F_0(G)), \quad (3.13)$$

где $F_0(P), F_0(G) \hat{Y}_{B\phi}$
 $B\phi = \{1, 2, \dots, N+1\}$.

Действительно, с учётом того, что

$$F_0(P) = P\phi = \{p_1, \dots, p_n, (N+1)\}, \quad F_0(G) = G\phi = \{g_1, \dots, g_n, (N+1)\},$$

выражение (3.10) может быть представлено в виде

$$\begin{aligned} r(P, G) = & \sum_{i=1}^n \sum_{j=1}^m |p_i \vee g_j| |p_i \wedge g_j| + \sum_{i=1}^n [|p_i \vee (N+1)| |p_i \wedge (N+1)| + \\ & + \sum_{j=1}^m |g_j \vee (N+1)| |g_j \wedge (N+1)| + |(N+1) \vee (N+1)| |(N+1) \wedge (N+1)| \end{aligned} \quad (3.14)$$

Ясно, что в равенстве (3.14) первое слагаемое представляет собой $\rho(P, G)$, два последующих равны 0, т.к. $|p_i \wedge (N+1)| = |g_j \wedge (N+1)| = 0$ при " $i = \overline{1, n}, j = \overline{1, m}$ ", а в силу того, что $|(N+1) \vee (N+1)| = 0$, обнуляется последнее слагаемое. Таким образом, получаем равенство (3.13), что и требовалось.

Теперь покажем, что для произвольных разбиений P, G множества B

и любого номера $s \in \{1, \dots, m\}$ имеет место

$$r \in F_0(\Pi), F_s(G) = r(\Pi, G) + 2|g_s|, \quad (3.15)$$

где $F_0(\Pi), F_s(G) \in Y_{B'}$,

$$B' = \{1, 2, \dots, N+1\}.$$

Исходя из трансформации (3.12), имеем

$$\begin{aligned} F_0(\Pi) &= \Pi \in \{p_1, \dots, p_n, (N+1)\}, \\ F_s(G) &= G \in \{g_1, \dots, (g_s, N+1), \dots, g_m\}. \end{aligned}$$

Тогда

$$\begin{aligned} r \in \Pi \in G \in &= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq s}}^m |p_i \vee g_j| |p_i \wedge g_j| + \sum_{i=1}^n [|p_i \vee (g_s, N+1)| \\ &+ |p_i \wedge (g_s, N+1)|] + \sum_{\substack{j=1 \\ j \neq s}}^m |(N+1) \vee g_j| |(N+1) \wedge g_j| + \\ &+ |(N+1) \vee (g_s, N+1)| |(N+1) \wedge (g_s, N+1)|. \end{aligned}$$

Нетрудно заметить, что при любых $i = \overline{1, n}, s = \overline{1, m}$

$$\begin{aligned} |p_i \vee (g_s, N+1)| &= |p_i \vee g_s| + 1, \\ |p_i \wedge (g_s, N+1)| &= |p_i \wedge g_s|. \end{aligned} \quad (3.16)$$

Далее, для $j = \overline{1, m}$

$$|(N+1) \wedge g_j| = 0 \quad (3.17)$$

и, наконец,

$$\begin{aligned} |(N+1)\vee(g_s, N+1)| &= |g_s|, \\ |(N+1) \uparrow (g_s, N+1)| &= 1. \end{aligned} \quad (3.18)$$

Учитывая равенства (3.16) – (3.18), получаем

$$\begin{aligned} r \llbracket \Pi \llbracket G \rrbracket \rrbracket &= \mathring{a} \prod_{i=1}^n \mathring{a} \prod_{\substack{j=1 \\ j \neq s}}^m |p_i \vee g_j| |p_i \uparrow g_j| + \mathring{a} \prod_{i=1}^n (|p_i \vee g_s| + 1) |p_i \uparrow g_s| + |g_s| = \\ &= \left[\mathring{a} \prod_{i=1}^n \mathring{a} \prod_{j=1}^m |p_i \vee g_j| |p_i \uparrow g_j| + \mathring{a} \prod_{i=1}^n |p_i \vee g_s| |p_i \uparrow g_s| \right] + \mathring{a} \prod_{i=1}^n |p_i \uparrow g_s| + |g_s|. \end{aligned} \quad (3.19)$$

Ясно, что в правой части равенства (3.19) слагаемое в квадратных скобках представляет собой ни что иное, как $r(\Pi, G)$, а с учетом того, что $\Pi = \{p_1, \dots, p_n\}$ – разбиение множества B , и при этом $g_s \uparrow B$, то $\mathring{a} \prod_{i=1}^n |p_i \uparrow g_s| = |g_s|$. Следовательно, окончательно $r \llbracket \Pi \llbracket G \rrbracket \rrbracket = r(\Pi, G) + 2|g_s|$, т.е. получаем равенство (3.15), что и требовалось доказать.

Нетрудно понять, что в силу симметрии для любых разбиений P, G множества B и произвольного номера $k \in \{1, \dots, n\}$ будет иметь место аналог равенства (3.15)

$$r \llbracket F_k(\Pi), F_0(G) \rrbracket = r(\Pi, G) + 2|p_k|,$$

где $F_k(\Pi), F_0(G) \uparrow Y_{B'}$,

$$B \llbracket = \{1, 2, \dots, N+1\}.$$

На основе проведенных рассуждений покажем, что если зафиксировать два произвольных номера $k \in \{1, \dots, n\}$ и $s \in \{1, \dots, m\}$, и рассмотреть два произвольных разбиения P, G множества B , то будет выполняться равенство

$$r \Phi F_k(\Pi), F_s(G) = r(\Pi, G) + 2|p_k \vee g_s|,$$

где $F_k(\Pi), F_s(G) \in Y_{B'}$,

$$B \Phi = \{1, 2, \dots, N+1\}.$$

Поскольку

$$\begin{aligned} F_k(\Pi) &= \{p_1, \dots, (p_k, N+1), \dots, p_n\}, \\ F_s(G) &= \{g_1, \dots, (g_s, N+1), \dots, g_m\}, \end{aligned} \quad (3.20)$$

то для функционала $r \Phi$ и равенств (3.20) находим

$$\begin{aligned} r \Phi F_k(\Pi), F_s(G) &= \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^n \sum_{s=1}^m |p_i \vee g_j| |p_i \wedge g_j| + \sum_{i=1}^n \sum_{k=1}^n |p_i \vee (g_s, N+1)| |p_i \wedge (g_s, N+1)| + \\ &+ \sum_{j=1}^m \sum_{s=1}^m |(p_k, N+1) \vee g_j| |(p_k, N+1) \wedge g_j| + \\ &+ |(p_k, N+1) \vee (g_s, N+1)| |(p_k, N+1) \wedge (g_s, N+1)|. \end{aligned} \quad (3.21)$$

Учитывая (3.18), а именно

$$|p_i \vee (g_s, N+1)| = |p_i \vee g_s| + 1,$$

$$|p_i \wedge (g_s, N+1)| = |p_i \wedge g_s|,$$

можно представить (3.21) в виде

$$r \Phi F_k(\Pi), F_s(G) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^n \sum_{s=1}^m |p_i \vee g_j| |p_i \wedge g_j| +$$

$$+ \mathring{a}_{i=1}^n (|p_i \vee g_s| + 1) |p_i \mid g_s| + \mathring{a}_{j=1}^m (|p_k \vee g_j| + 1) |p_k \mid g_j| + |p_k \vee g_s| (|p_k \mid g_s| + 1).$$

При этом имелось в виду, что

$$\begin{aligned} |(p_k, N+1) \vee (g_s, N+1)| &= |p_k \vee g_s|, \\ |(p_k, N+1) \mid (g_s, N+1)| &= (|p_k \mid g_s| + 1). \end{aligned}$$

Таким образом, имеем

$$\begin{aligned} r \mathring{P} \mathring{G} \mathring{G} \mathring{G} &= \left[\mathring{a}_{i=1}^n \mathring{a}_{j=1}^m |p_i \vee g_j| |p_i \mid g_j| + \mathring{a}_{i=1}^n |p_i \vee g_s| \right. \\ &\quad \left. |p_i \mid g_s| + \mathring{a}_{j=1}^m |p_k \vee g_j| |p_k \mid g_j| + \mathring{a}_{i=1}^n |p_i \mid g_s| + \right. \\ &\quad \left. + \mathring{a}_{j=1}^m |p_k \mid g_i| + |p_k \vee g_s| |p_k \mid g_s| + |p_k \vee g_s|. \right. \end{aligned} \quad (3.22)$$

С учетом того, что слагаемые, стоящие в квадратных скобках равенства (3.22) вместе со слагаемым этого же равенства вида $|p_k \vee g_s| |p_k \mid g_s|$, дают $r(\mathring{P}, \mathring{G})$, можно далее представить $r \mathring{P} \mathring{G} \mathring{G} \mathring{G}$ в виде

$$\begin{aligned} r \mathring{P} \mathring{G} \mathring{G} \mathring{G} &= r(\mathring{P}, \mathring{G}) + \left(\mathring{a}_{i=1}^n |p_i \mid g_s| - |p_k \mid g_s| \right) + \\ &\quad + \left(\mathring{a}_{j=1}^m |p_k \mid g_j| - |p_k \mid g_s| \right) + |p_k \vee g_s|. \end{aligned} \quad (3.23)$$

Здесь в каждой из круглых скобок мы фактически добавили и вычли по одному слагаемому $|p_k \mid g_s|$. Но суммы вида $\mathring{a}_{i=1}^n |p_i \mid g_s|$ и $\mathring{a}_{j=1}^m |p_k \mid g_j|$ рас-

смаатривались выше, т.е. каждая из них равна

$$\mathring{a}_{j=1}^u |p_k \mid g_s| = |g_s|, \quad \mathring{a}_{j=1}^m |p_k \mid g_j| = |p_k| \quad (3.24)$$

Тогда из (3.23) и (3.24) имеем

$$r \mathring{P} \mathring{G} \emptyset = r(P, G) + (|p_k| + |g_s| - 2|p_k \mid g_s|) + |p_k \vee g_s|.$$

Осталось заметить, что поскольку для любых конечных множеств D и E $|D \vee E| = |D| + |E| - 2|D \mid E|$, то $r \mathring{P} \mathring{G} \emptyset = r(P, G) + 2|p_k \vee g_s|$.

Подводя некоторый промежуточный итог, следует подчеркнуть, что изменение функционала (3.10) на двух произвольных конечных разбиениях $P = \{p_1, \dots, p_n\}$ и $G = \{g_1, \dots, g_m\}$, заданных на множестве $B = \{1, \dots, N\}$, подчиняется на $Y_B \times Y_B$ следующим соотношениям (и только им)

$$\begin{aligned} r'(P', G') &= r(P, G), & \text{если } P' &= F_0(P), G' = F_0(G); \\ r'(P', G') &= r(P, G) + 2|g_s|, & \text{если } P' &= F_0(P), G' = F_s(G); \\ r'(P', G') &= r(P, G) + 2|p_k|, & \text{если } P' &= F_k(P), G' = F_0(G); \\ r'(P', G') &= r(P, G) + 2|p_k \vee g_s|, & \text{если } P' &= F_k(P), G' = F_s(G), \end{aligned}$$

Установленные свойства означают, что, во первых, функционал (3.10) может использоваться в качестве оценки сходства разбиений-сегментаций (проверены аксиомы рефлексивности и симметричности), а во-вторых, эти свойства обеспечивают предпосылки для установления факта степени неразличимости разбиений. Иначе говоря, необходимо проверить выполнение неравенства треугольника.

3.3 Анализ метрических свойств множества сегментаций

Проверку того, что функционал (3.10), заданный на $Y_B \times Y_B$, где B –

конечное множество, а Y_B – множество разбиений B на конечный набор подмножеств, является метрикой, проведем доказательством по индукции.

Очевидно, что базой индукции является случай $|B|=3$. Вводя обозначение $B = \{a, b, g\}$, получаем множество всех разбиений в виде

$$Y_B = \{ \{\{a\}, \{b\}, \{g\}\}, \{\{a\}, \{b, g\}\}, \{\{b\}, \{a, g\}\}, \{\{g\}, \{a, b\}\}, \{a, b, g\} \}. \quad (3.25)$$

Такое представление правомочно, поскольку выбор трех точек множества Y_B может быть осуществлен $C_3^3 = 10$ способами:

$$\begin{array}{l} \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \\ \text{i} \end{array} \begin{array}{l} 1) \{\{a\}, \{b\}, \{g\}\}; \{\{a\}, \{b, g\}\}; \{\{b\}, \{a, g\}\}; \\ 2) \{\{a\}, \{b\}, \{g\}\}; \{\{a\}, \{b, g\}\}; \{\{g\}, \{a, b\}\}; \\ 3) \{\{a\}, \{b\}, \{g\}\}; \{\{b\}, \{a, g\}\}; \{\{g\}, \{a, b\}\}; \\ 4) \{\{a\}, \{b\}, \{g\}\}; \{\{a\}, \{b, g\}\}; \{a, b, g\}; \\ 5) \{\{a\}, \{b\}, \{g\}\}; \{\{b\}, \{a, g\}\}; \{a, b, g\}; \\ 6) \{\{a\}, \{b\}, \{g\}\}; \{\{g\}, \{a, b\}\}; \{a, b, g\}; \\ 7) \{\{a\}, \{b, g\}\}; \{\{b\}, \{a, g\}\}; \{\{g\}, \{a, b\}\}; \\ 8) \{\{a\}, \{b, g\}\}; \{\{b\}, \{a, g\}\}; \{a, b, g\}; \\ 9) \{\{a\}, \{b, g\}\}; \{\{g\}, \{a, b\}\}; \{a, b, g\}; \\ 10) \{\{b\}, \{a, g\}\}; \{\{g\}, \{b, a\}\}; \{\{a, b, g\}\}, \end{array} \quad (3.26)$$

но, как легко заметить, варианты 1) – 3), 4) – 6), 8) – 10) равносильны с точностью до обозначений. Тем самым, достаточно проверить существование четырех треугольников (с учетом уникальности случая 7) в (3.26)), а именно из (3.25) имеем: $DG_1G_2G_3$, $DG_1G_2G_5$, $DG_2G_3G_5$, $DG_1G_2G_4$, где $G_1 = \{\{a\}, \{b\}, \{g\}\}$, $G_2 = \{\{a\}, \{b, g\}\}$, $G_3 = \{\{b\}, \{a, g\}\}$, $G_4 = \{\{g\}, \{a, b\}\}$, $G_5 = \{a, b, g\}$.

Теперь необходимо найти расстояния между всеми разбиениями. Два элемента множества Y_B можно выбрать $C_3^2 = 10$ способами. В силу симметрии введенной нотации укажем, что

$$\begin{aligned} r(G_1, G_2) &= r(G_1, G_3) = r(G_1, G_4), \\ r(G_2, G_3) &= r(G_2, G_4) = r(G_3, G_4), \\ r(G_5, G_2) &= r(M_5, G_3) = r(G_5, G_4), \end{aligned}$$

а значение $r(G_1, G_5)$ должно быть вычислено отдельно. Непосредственные вычисления дают: $r(G_1, G_2) = 2$, $r(G_2, G_3) = 4$, $r(G_2, G_5) = 4$, $r(G_1, G_5) = 6$. Полученные треугольники схематично иллюстрируются рис. 3.2 и рис. 3.3.

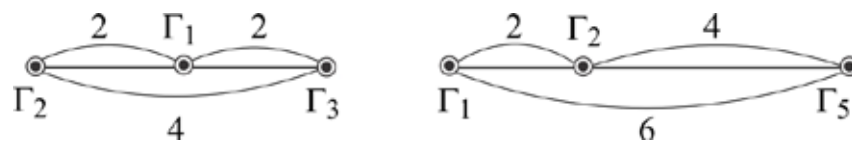


Рисунок 3.2 – Треугольники: $DG_1G_2G_3$, $DM_1M_2M_5$

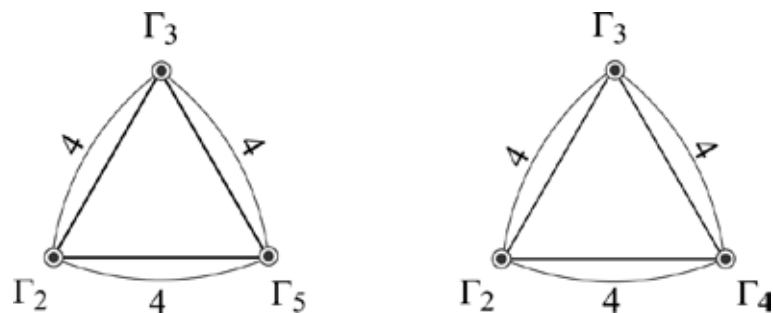


Рисунок 3.3 – Треугольники: $DG_2G_3G_5$, $DG_2G_3G_4$

Таким образом, для всех случаев (3.26) база для проведения слабой математической индукции доказана.

Рассмотрим индуктивный переход. Допустим, неравенство треугольника имеет место для множества B мощности N . Рассмотрим множество мощности $N+1$ и его произвольное конечное разбиение $P \hat{=} Y_{B'}$, где $B' = \{1, 2, \dots, N+1\}$.

Отметим, что для элемента $N+1$ возможны, как отмечалось выше, два варианта: он продуцирует новый сегмент или попадает в уже имеющийся.

При этом $B = B' \setminus \{(N+1)\}$ множество мощности N , а само разбиение, как было установлено выше, получено одной из трансформаций $P' = F_0(P)$ или $P' = F_k(P)$, где $P \hat{=} Y_B$. Отметим, что на B' для любых трех разбиений P', G', H' вида (3.9) их попарные расстояния имеют вид, представленный на рис. 3.4, где для функционалов $r(P, G), r(P, H), r(G, H)$ неравенство треугольников выполняется по сделанному предположению индукции, поскольку $P, G, H \hat{=} Y_B$ – конечные разбиения на множестве B мощности N . Тем самым, если неравенства треугольников на множестве $B' \setminus B'$ справедливы для приращений $\Delta r(P', G')$, $\Delta r(P', H')$, $\Delta r(G', H')$, то индуктивный переход будет верифицирован.

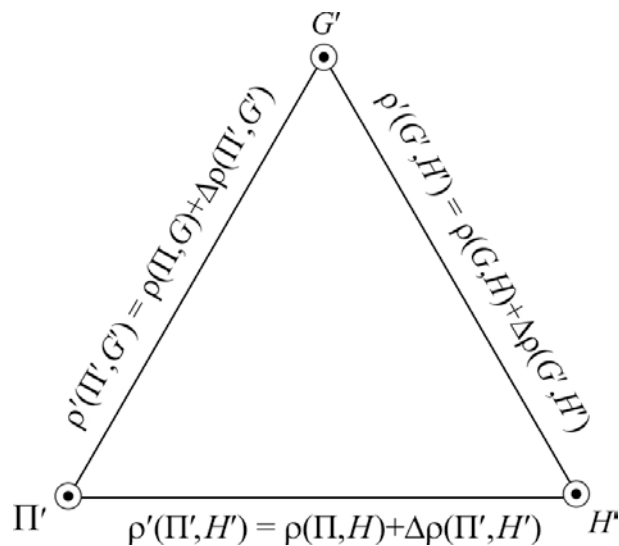


Рисунок 3.4 – Попарные расстояния для разбиений P', G', H'

Как нетрудно заметить, может возникнуть всего восемь возможных ситуаций, которые представлены на рис. 3.5. При этом для случаев 1), 2), 3), 5) выполнение неравенства треугольника очевидно. Ситуация 8) также очевидна, поскольку симметрическая разность – метрика. В силу симметрии варианты 4), 6), 7) сводятся к одному случаю, показанному на рис. 3.6. Для проверки выполнения неравенства треугольника достаточно показать

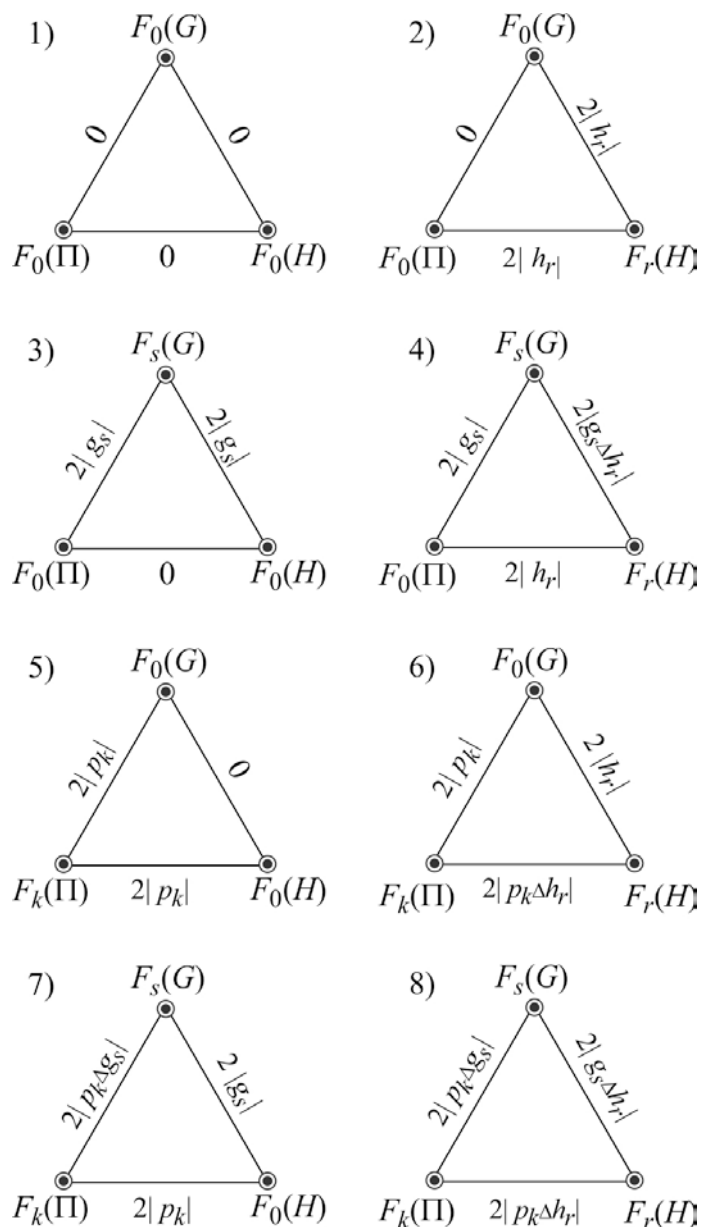


Рисунок 3.5 – Изменения расстояний при индуктивном переходе

$$|p^*| + |g^*| - |p^* \vee g^*| \geq 0,$$

$$|p^* \vee g^*| + |p^*| - |g^*| \geq 0.$$

Принимая во внимание, что $|p^* \vee g^*| = |p^*| + |g^*| - 2|p^* \wedge g^*|$, получаем

$$|p^*| + |g^*| - |p^* \vee g^*| = 2|p^* \wedge g^*| \geq 0,$$

откуда следует

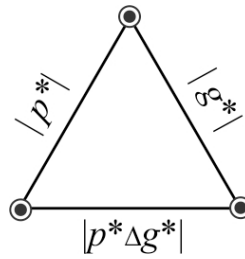


Рисунок 3.6 – Обобщение вариантов 4), 6), 7) рис. 3.5

$$\begin{aligned}
 |p^* \vee g^*| + |p^*| - |g^*| &= |p^*| + |g^*| - 2|p^* \setminus g^*| + |p^*| - |g^*| = \\
 &= 2(|p^*| + |p^* \setminus g^*|) = 2|p^* \setminus g^*| \geq 0,
 \end{aligned}$$

что и требовалось.

Таким образом, функционал (3.10) является метрикой. Отметим, что общий случай – для разбиений любых измеримых множеств этот результат доказан позже в [231]. Достоинством такой метрики является то, что она в некотором интегральном виде оценивает «пространственное» содержание видеокadra. Более того, она мультипликативно учитывает степень различия (в виде симметрической разности элементов разбиений) и степень сходства (в виде пересечения этих областей), что в результате повышает ее валидность. Несомненно, эта метрика является ключевой в декартовых произведениях метрик, рассмотренных во втором разделе. Основное ее назначение – использование при решении задачи информационного поиска видеоданных на этапе сравнения изображения-запроса с множеством ключевых кадров, представляющих видеопоследовательности. Однако при сравнении результатов кластеризации-сегментации для выбора наиболее адекватного разбиения данных эта метрика также весьма полезна.

Примеры ключевых кадров видеопоследовательностей различной тематики, а также их сегментации приведены на рис. 3.7. Другими словами, мы можем построить сегментацию для изображения-запроса сделанного Пользователем и, используя функционал (3.10), провести метрическое сравнение с сегментациями кадров видеопоследовательностей, на основании чего сделать вывод о мере сходства данных и получить результат поиска.

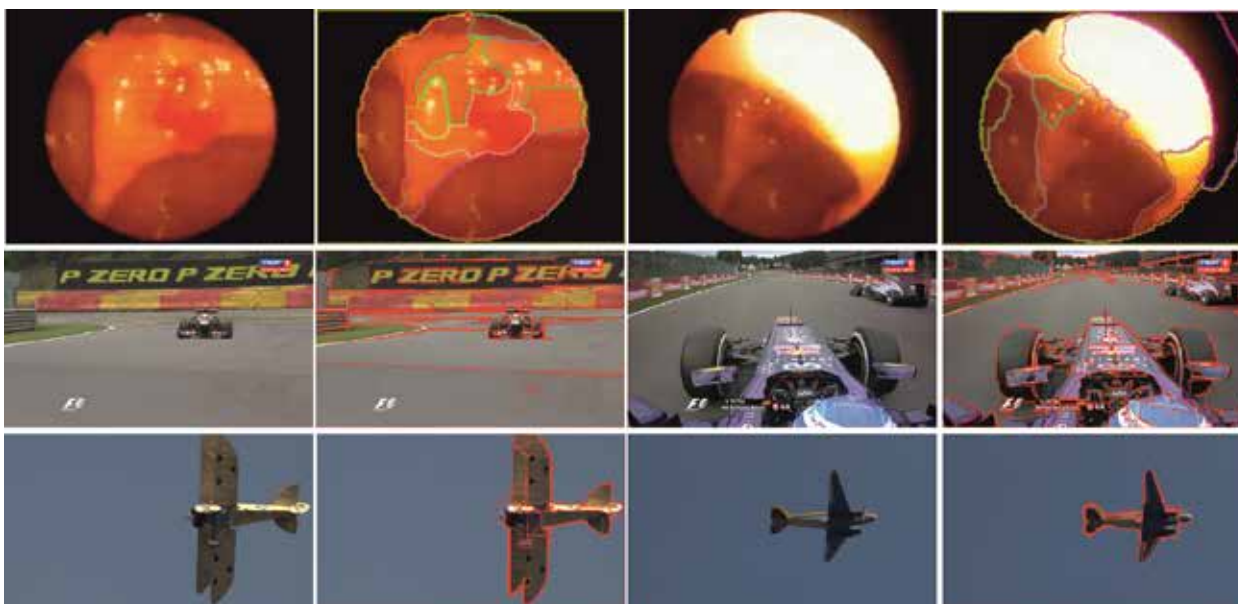


Рисунок 3.7 – Примеры кадров видеопоследовательностей и их сегментации

Ясно, что введенная метрика не всегда достаточно адекватно отражает «содержание» видеокadra: носители различных объектов могут быть похожи. Поэтому имеет смысл использовать не только метрическое сравнение результатов сегментаций, но и учитывать другие характеристики изображений. При этом одним из вполне естественных подходов видится анализ контуров сегментов (отдельных областей разбиений), входящих в ключевые/поисковые кадры.

3.4 Построение характеристических векторов видеокadров

На сегодняшний день существует большое количество подходов к выделению и представлению контуров. Их можно разделить, в зависимости от способов обработки, на одномерные функции представления фигуры (One-dimensional function), аппроксимацию многоугольников (Polygonal approximation), взаимосвязь пространственных признаков (Spatial interrelation feature), моменты (Moments) и многие другие. Классификация подходов по анализу формы представлена в табл. 3.1.

Таблица 3.1 – Классификация подходов к анализу формы

Подходы к анализу формы областей	Методы и характеристики
Одномерные функции	Различные функции, которые строятся на основе точек контура, в частности, функция комплексных координат
Полигональная аппроксимация	Методы слияния и расщепления
Функции взаимосвязи областей	Выпуклая оболочка, цепные коды, контекст области, шок графы (скелетоны), матрица области (представление в полярных или декартовых координатах) и т.д.
Моменты	Моменты границ и областей такие, как Зерника, инвариантные, алгебраические, радиальные (Чебышева), гомоцентричные, ортогональные и т.д.
Пространственный масштаб	Точки пересечения областей, кривизна областей.
Преобразование областей	Гармонические функции, преобразования Фурье, волновые преобразования (Wavelet), радиальные преобразования, R -функции и т.д.

Применительно к СВВIR не так важна точная идентификация объекта, а скорее важны такие простые с вычислительной точки зрения характеристики, по совокупности приближений которых можно провести валидное объективное сравнение. Кроме вполне естественных геометрических характеристик сегментов таких как, например, площадь, периметр, определенный интерес также представляют методы упрощенного представления изображений, а также методы инвариантных к преобразованиям характеристик. Одним из актуальных подходов является метод полигональной аппроксимации, кото-

рый позволяет представить каждый сегмент в приближенном виде, оставляя только значимые для данного класса детали. Но более важно то, что разбиения Вороного (как будет показано в разделе 5), описываемые обычно прямоугольниками, являются достаточно адекватными характеристиками ключевых видеок кадров.

Учитывая то, что каждая сегментированная область может рассматриваться как некоторое бинарное изображение, весьма эффективным подходом может быть использование операций математической морфологии. Другими словами, задача сравнения двух изображений может быть сформулирована как задача нахождения и сравнения векторов характеристик объектов на изображении. Следует отметить, что выбор и самих компонент этого вектора, и их количества является довольно сложной и нетривиальной задачей.

Одной из характеристик, основанной на центре тяжести сегментированной области, являются комплексные координаты. Для кадра B_i и его сегментации R_i каждой j -ой области сопоставляется множество координат точек контура $P = \{x_n, y_n\}, n \hat{=} \overline{1, N}$ (N – количество точек контура)

$$R_i(j) = P_n(j), n \hat{=} \overline{1, N}, \quad (3.27)$$

тогда комплексная координата имеет следующий вид:

$$z_n = x_n + iy_n.$$

Полученные комплексные числа можно использовать в качестве дескриптора сегмента. Данная характеристика очень проста в реализации, но имеет ряд недостатков: получаемый результат является неинвариантным к смещению, масштабированию и вращению. С другой стороны, если отслеживать движение в видеоданных, изменение этой характеристики может быть крайне информативным. Чтобы комплексные координаты были инвариантными к смещению, z_n можно вычислять с учетом центра тяжести (центроида):

$$z_n = [x_n - g_x] + i[y_n - g_y],$$

где $g = (g_x, g_y)$ – центр тяжести соответствующего сегмента, который представлен ниже

$$\begin{aligned} g_x &= \frac{1}{6S} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i), \\ g_y &= \frac{1}{6S} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i), \end{aligned} \quad (3.28)$$

где S – площадь сегмента

$$S = \frac{1}{2} \left| \sum_{i=0}^{N-1} (x_i y_{i+1} - x_{i+1} y_i) \right|.$$

В качестве второй характеристики вектора признаков вычислим функцию расстояния $r(n)$, которую найдем как расстояние от каждой точки контура сегмента до некоторой неподвижной точки, в качестве которой также может выступать центр тяжести. Тогда для (3.27) функция расстояния принимает вид

$$r(n) = [(x(n) - g_x)^2 + (y(n) - g_y)^2]^{1/2}.$$

Функция расстояния имеет те же преимущества и недостатки, что и комплексные координаты. Получаемый результат является инвариантным к смещению, масштабированию и вращению. Так как в расчетах функции расстояния было включено значения центра тяжести фигуры (3.28).

В качестве следующей характеристики рассмотрим угол касательной в каждой точке контура

$$q(n) = q_n = \arctan \frac{y(n) - y(n-w)}{x(n) - x(n-w)},$$

где w – окно небольшого размера для более точных расчетов $q(n)$.

Несмотря на простоту реализации, характеристика имеет два больших недостатка: чувствительность к шуму и дискретности, что может привести к ошибочным вычислениям. Для уменьшения влияния шума до расчета касательной функции угла, контур необходимо применять фильтр нижних частот с соответствующей пропускной способностью. Другая проблема – дискретность, возникающая из-за того, что угол касательной функции принимает значения в диапазоне длины 2π , как правило, в интервале $[-\rho, \rho]$ или $[0, 2\rho]$. Поэтому q_n в целом содержит разрывы размером 2π . Для преодоления проблемы разрывов используется кумулятивная угловая функция

$$j(n) = |q(n) - q(0)|,$$

где $q(0)$ – угол касательной к случайно выбранной точке на контуре.

В частности, если предположить, что

$$t = 2\rho n / N, j(n) = j(tN / 2\rho),$$

то кумулятивная угловая функция определяется как

$$y(t) = j\left(\frac{N}{2\rho}t\right) - t, t \in [0, 2\rho].$$

Вариант такой функции из примера контура одного из сегментов кадра из рис 3.7 представлен на рис. 3.8.

Еще одной вполне естественной характеристикой является функция

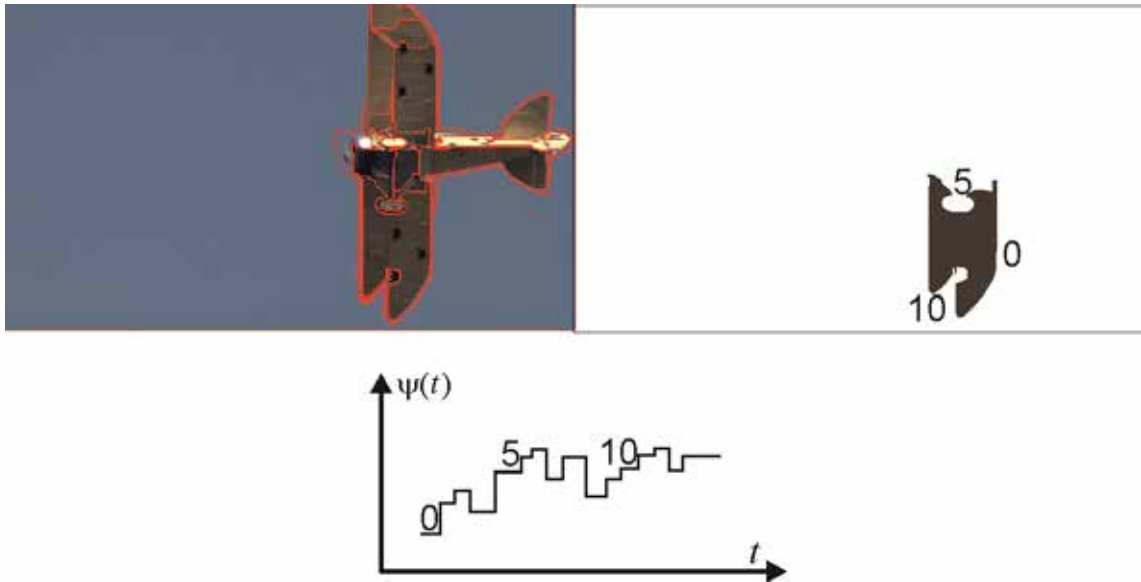


Рисунок 3.8 – Ступенчатая функция угла касательной к контуру

площади сегментов. При изменении граничных точек вдоль контура объекта, площадь треугольника, образованная двумя последовательными точками границы и центром тяжести (3.26) также изменяется. Это формирует функцию площади, которая может быть использована для представления формы объекта (рис. 3.9).

Функция площади линейна относительно аффинных преобразований. При необходимости функцию площади изображения можно получить, используя функцию сечения для симметрических фигур и функцию радиус-вектора для выпуклых областей, соответствующие формулы представлены ниже:

$$S(x) = 2 \int_{x_0}^{x_1} q_x(x) dx, S(x) = \frac{1}{2} \int_0^{2\pi} r_x^2(j) dj$$

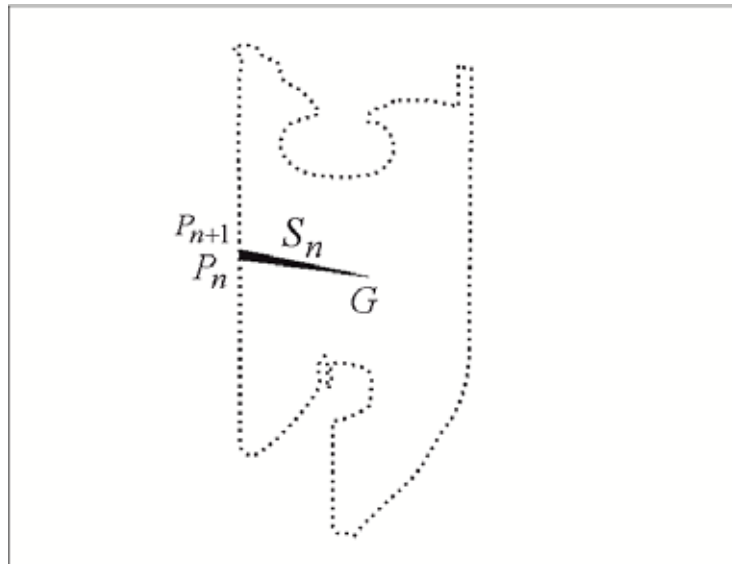


Рисунок 3.9 – Функция площади P_n

Функция представления фигуры с помощью треугольников вычисляется с помощью площади треугольников, образованных точками контура объекта. Кривизна в точке контура (x_n, y_n) измеряется следующим образом. Для каждой трех точек

$$P_{n-t_s}(x_{n-t_s}, y_{n-t_s}), P_n(x_n, y_n), P_{n+t_s}(x_{n+t_s}, y_{n+t_s}),$$

где $n \in [1, N], t_s \in [1, N/2 - 1]$ и N предполагается четным.

Площадь треугольника, образованного данными точками, определяется по формуле:

$$TAR(n, t_s) = \frac{1}{2} \begin{vmatrix} x_{n-t_s} & y_{n-t_s} & 1 \\ x_n & y_n & 1 \\ x_{n+t_s} & y_{n+t_s} & 1 \end{vmatrix}.$$

При обходе контура против часовой стрелки, положительные, отрицательные и нулевые значения TAR означают выпуклые, вогнутые и прямолинейные участки соответственно.

Большинство объектов в предметно-ориентированной области можно

классифицировать, причем объекты внутри класса различаются мелкими деталями структуры, а объекты разных классов имеют большие различия. Таким образом, нередко, чтобы классифицировать объекты, необходимо найти их упрощенное представление, состоящее из одних крупных деталей (рис. 3.10).

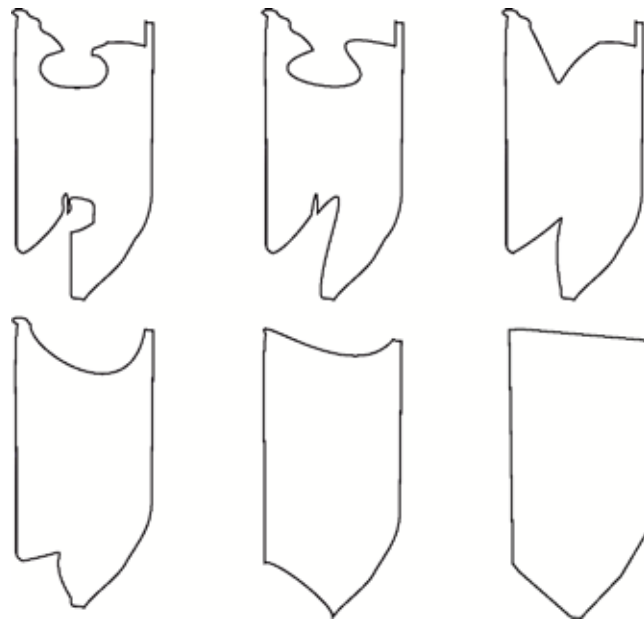


Рисунок 3.10 – Полигональная аппроксимация

Один из способов упрощенного представления плоского объекта со сложным контуром заключается в использовании особых точек контура и связей между этими точками. Данный подход используется в методе полигональной аппроксимации, которая обеспечивает сглаживание сложных контуров с целью упрощения формы, что уменьшает эффект дискретной пикселизации контура.

Основная идея полигональной аппроксимации заключается в следующем: на каждом шаге, пару последовательных отрезков S_1, S_2 (отрезком линия между двумя соседними вершинами) заменяют одним сегментом линией, соединяющей их конечные точки. Ключевой особенностью данного подхода является порядок замещения. Замена происходит в соответствии со степенью актуальности K

$$K(S_1, S_2) = \frac{b(S_1, S_2)l(S_1)l(S_2)}{l(S_1) + l(S_2)},$$

где $b(s_1, s_2)$ – угол поворота на общую вершину S_1, S_2 и

$l(a)$ – длина a , нормированная относительно общей длины кривой.

Метод полигональной аппроксимации решает задачу упрощения формы, то есть в процессе построения определяется значение вершин контура на основе степени значимости. Поскольку любую цифровую кривую можно рассматривать как многоугольник без существенной потери информации, то для извлечения необходимых признаков достаточно изучить эволюцию полигональной формы.

Еще одной интересной характеристикой является выпуклая оболочка сегмента. Ясно, что любая сегментация кадра может быть представлена покрытием, состоящим только из выпуклых оболочек сегментов. Они могут быть найдены следующим образом: выпуклость произвольной формы определяется, как отношение периметров выпуклой оболочки $O_{Convexhull}$ к периметру исходного контура O :

$$Convexity = \frac{O_{convexhull}}{O}.$$

Область в \mathbb{R}^2 является выпуклой тогда и только тогда, когда для любых двух точек $P_1, P_2 \in \mathbb{R}^2$, отрезок P_1P_2 принадлежит области. На рис. 3.11 представлена выпуклая оболочка сегмента.

Выпуклая оболочка H области изображения является наименьшей выпуклой оболочкой области. Другими словами, для области, выпуклая оболочка $Convexity$ определена как наименьший выпуклый набор в \mathbb{R}^2 , содержащий область изображения. Чтобы уменьшить эффект шума, необходимо сначала сгладить границы объекта.

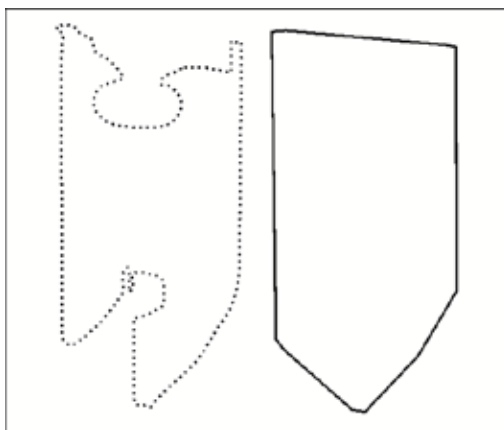


Рисунок 3.11 – Выпуклая оболочка области, соответствующая рис. 3.8,3.9

Представление формы может быть получено рекурсивным процессом, который приводит к дереву вогнутости. Каждая вогнутость может быть описана функцией площади данной области, длиной хорд, функцией максимального искривления, функцией расстояния от максимального искривления до хорды.

Представление фигуры с помощью выпуклой оболочки является инвариантным к вращению, масштабированию и перемещению, а также устойчивым к зашумленным границам формы (после фильтрации).

В работе [254] рассматривается вариант использования выпуклых оболочек для нормализации изображений. При этом рассмотрен вариант использования не всей выпуклой оболочки, а только некоторых характерных точек и метод нахождения соответствий между изображением и эталоном, который можно применять для нахождения соответствий между сегментацией изображения-запроса и сегментациями ключевых кадров видеопоследовательностей.

В этой же работе рассмотрен еще один подход представления изображений, а именно скелетизация. Метод анализа формы плоского объекта на основе построения его скелета широко применяется в обработке изображений. Непрерывный скелет определяется для областей, заданных «непрерывно», т.е. таких, чья граница описывается в виде непрерывных (обычно ломаных) линий. Наиболее быстрым в реализации построения скелета изображе-

ния является способ, основанный на МАТ (Medial Axis Transform), что является достаточно важным, с точки зрения обработки большого количества сегментов и кадров видеопоследовательностей.

Для установления соответствий необязательно использовать весь скелет изображения, а достаточно использовать некоторые множества характерных точек скелета. На рис. 3.12 показан пример (I – крайние, V – внутренние) таких характерных точек.

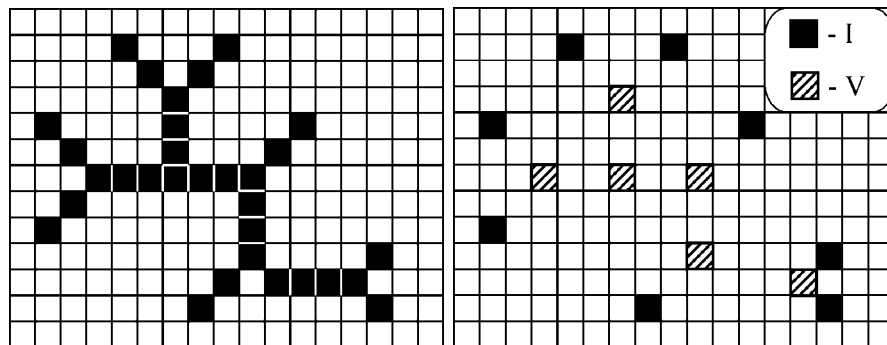


Рисунок 3.12 – Скелет изображения и множество характерных точек

Кроме приведенных характеристик существует еще целый ряд полезных параметров, таких как:

– центр оид или центр тяжести: расположение этого элемента неизменно на изображении и инвариантно относительно аффинных преобразований

$$G(x, y): \begin{cases} \ddot{g}(x) = \frac{1}{N} \sum_{i=1}^N x_i, \\ \ddot{g}(y) = \frac{1}{N} \sum_{i=1}^N y_i; \end{cases}$$

– эксцентриситет (Elo): отношение длины к ширине описанного прямоугольника вокруг исходного объекта

$$Elo = 1 - W / L,$$

где W - высота,

L – ширина;

– дисперсия эллипса: определяет ошибку того, насколько объект соответствует эллипсу при условии одинаковой матрицы ковариации (C_{ell}).

$$E_{va} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N-1} (\sqrt{V_i^T C_{ell}^{-1} V_i} - \frac{1}{N} \sum_{i=1}^N \sqrt{V_i^T C_{ell}^{-1} V_i})^2}}{\frac{1}{N} \sum_{i=1}^N \sqrt{V_i^T C_{ell}^{-1} V_i}},$$

где $V_i = \begin{pmatrix} x_i - g(x) \\ y_i - g(y) \end{pmatrix}$ – моменты осей.

– функция расстояния центраоида: функция, которая строится на базе расстояний от центра тяжести до границы объекта

$$r_n = \sqrt{(x_n - g_x)^2 + (y_n - g_y)^2};$$

– радиусы максимальной вписанной R_{in} и минимальной описанной R_{out} окружностей вокруг объекта;

– длины полуосей осей a_e, b_e эллипса с площадью S и периметром P , которые находятся по формуле

$$a_e = a(X) + \sqrt{a(X)^2 - S(X)/\rho}, b_e = S(X)/(\rho a_e(X)),$$

где $a(X) = \frac{1}{3}(\sqrt{S(X)/\rho} + P(X)/\rho)$;

– параметры вытянутости (их целый ряд и они позволяют оценить вытянутость объекта, и являются во многом обратной величиной к отношению округлости);

– диаметр Мартина $MD(X)$ (расстояние между левым верхним углом

объекта X и его правым нижним углом); различают также диаметры Мартина по осям ($MD_x(X)$, $MD_y(X)$), соответственно, это максимальное расстояние по осям;

– «звездность» объекта (показывает видимую область границы объекта из произвольной точки – используется при выборе точки отсчета);

– параметр Вейделла (во многом похож на параметр Хортон и также служит для выявления компактности объекта);

– угловой параметр Фишера

$$\sum_i a_i(X) / 360^0, \sum_j b_j(X) / \sum_k g_k(X);$$

– проекция Пентланда вычисляется по формуле

$$4S(X) / (pF_{\max}(X));$$

– параметр округлости Пирарда вычисляется по формуле

$$1 / \sqrt{\sum_{i=1}^n \frac{1 + R_{in}(X) / r_i(X)}{r_i(X)} / (n - 1)},$$

где r_i – радиус максимально вписанной окружности, пересекающей i -ю точку контура X ;

– параметр округлости Венворта

$$4r(X) / (F_{\max}(X) + B(X)),$$

где $r(x)$ – радиус максимально выпуклой части контура X ,

$B(X)$ – ширина X , ортогональная к $F_{\max}(X)$;

– отношение Ферре (позволяет оценить разброс исходных данных по

осям).

С помощью моментов можно получить представление об ориентации объекта в пространстве – угол минимальных моментов:

$$Q(X) = \tan^{-1} \frac{M_{xx}(X) - M_{yy}(X) + \sqrt{(M_{xx}(X) - M_{yy}(X))^2 + 4(M_{xy}(X))^2}}{2M_{xy}(X)}$$

$$\text{где } M_x = \frac{1}{S_{obj}} \sum_{i=1}^N x_i^2, \quad M_y = \frac{1}{S_{obj}} \sum_{i=1}^N y_i^2, \quad M_{xy} = \frac{1}{S_{obj}} \sum_{i=1}^N x_i y_i.$$

Также существует целый ряд отношений, основанных на вышеописанных характеристиках, в частности:

- отношение радиусов R_{in} / R_{out} ;
- отношение эллипсов $a_e(X) / b_e(X)$;
- отношение диаметров $MD_x(X) / F_x(X)$ и $MD_y(X) / F_y(X)$;
- отношение протяженности $S(X) / S(F_x(X) \cdot F_y(X))$;

где $F_x(X) \cdot F_y(X)$ описанный прямоугольник вокруг X , получаемый вращением фигуры до тех пор, пока не будет выполнено условие $F_x(X) = F_{min}(X)$.

Анализ представленных выше характеристик позволяет утверждать, что комплексные координаты и функция расстояния, при расчете с учетом центра тяжести объекта являются инвариантными относительно смещения, но в тоже время они не инвариантны к масштабированию и вращению. Следовательно, можно сделать вывод, что в векторе характеристик объекта можно использовать одну из функций, а именно ту, которая будет давать наилучший результат во взаимодействии с другими характеристиками.

Линейной относительно аффинных преобразований является функция площади, поэтому можно сделать вывод, что вероятность ошибки значения

при расчете данной характеристики будет минимальной. Функция угла прямой касательной чувствительна к шуму и дискретности, поэтому при использовании данной характеристики необходимо, прежде всего, произвести фильтрацию изображения.

В случае, когда необходимо отнести объект к какому либо классу, и мелкие детали данного объекта не столь важны, можно использовать полигональную аппроксимацию. Она сглаживает контур, что «нормализует» форму объекта и упрощает классификацию. Целесообразно использовать и функцию выпуклости оболочки, которая отражает отношение периметра фигуры к его выпуклой области, то есть помогает определить, насколько объект близок к заданному классу.

В тоже время, если вести речь о задаче анализа видеопотоков и поиска с запросом «по образцу», вектор признаков не должен содержать все описанные параметры. Это связано с тем, что кроме повышения сложности вычислительной модели, возможно возникновение ситуации, аналогичной «проклятию размерности», когда некоторые характеристики путем компенсации значений друг относительно друга будут «скрывать» особенности сегментов, т.е. приводить к ошибочным идентификациям. Таким образом, можно и необходимо говорить о совместном использовании метрики на разбиениях с характеристиками типа площади, периметра, скелета, угла наклона к касательной.

Выводы по третьему разделу

1. Фрагментные модификации традиционных алгоритмов кластеризации создают предпосылки для on line сегментации видеопоследовательностей и видеок кадров. В первом случае используется множество фрагментов, соответствующих изменению во времени некоторой области интереса, во втором – ее разбиение или покрытие на одном изображении. Результат (в том числе и на полных изображениях) – пространственно-временная сегментация

обеспечивает структуризацию данных для задач информационного поиска.

2. Метрика на сегментациях изображений, интегрально учитывающая и сходство и различия областей носителей визуальной информации, может использоваться как достаточно адекватный инструмент сравнения «пространственного» содержания видеок кадров. Ее применение при структуризации видеопоследовательностей позволяет алгоритмизировать не только поиск границ видеосегментов, но оперировать ключевыми кадрами для организации метрического поиска.

3. Расширение анализа общего «пространственного» содержания на одновременный учет локальных характеристик областей объектов, элементов переднего или заднего плана позволяет управлять детализацией признаков представлений, что в итоге дает возможность варьировать сложность вычислительных моделей, обеспечивая с тем или иным приближением требуемое время вычислений.

РАЗДЕЛ 4

МЕТОДЫ И МОДЕЛИ СЕГМЕНТАЦИИ ВИДЕО НА ОСНОВЕ ИНДУЦИРОВАННЫХ ВРЕМЕННЫХ РЯДОВ

При решении задачи информационного поиска в видеоданных по их содержимому одним из вполне естественных подходов видится возможность разбиения исходных данных на однородные с точки зрения заданных критериев сегменты для дальнейшего ответа на запросы Пользователей о контексте видео. Такое разбиение представляет собой темпоральную сегментацию видеоданных. Вообще говоря, сегментация может осуществляться либо темпоральной обработкой непосредственно видеоряда, либо путем анализа многомерных временных рядов, индуцированных признаками видеокадров. В этих рядах существуют различные виды межкадровой корреляции, в частности, сходство соседних кадров видеопоследовательности, что позволяет применить аппарат анализа временных рядов с целью обнаружения разладок. Разладки и будут означать смену однородными сегментами, т.е. будут определять границы этих самых сегментов.

Поскольку многомерные временные ряды, соответствующие видеопоследовательностям, как правило, содержат избыточную информацию, их анализ осуществляется на основе контроля некоторых обобщенных характеристик таких, как спектры, корреляционные функции, тренды средних значений, дисперсии, главные компоненты и т.п.

Задача радикально усложняется в случаях, когда объем выборки не фиксирован, что присуще последовательному, покадровому, анализу видеоданных в реальном или приведенном масштабе времени. В подобных растущих выборках количество сегментов неизвестно, а в силу величайшей вариативности содержания видео и условий регистрации никаких априорных предположений о свойствах видеоряда обычно сделать нельзя. Таким образом, на первый план выступает уже не просто сегментация, а последовательное обнаружение моментов изменения свойств видеоданных.

В разделе рассмотрены модели многомерных временных рядов применительно к задаче информационного поиска в видеоданных, а также методы обнаружения разладок временных рядов для выделения однородных фрагментов видеоданных.

Раздел основан на работах автора [2, 24, 37, 79-81, 89, 90, 93, 157, 230, 251, 267, 270].

4.1. Метод темпоральной сегментации на основе настраиваемой модели

В подразделе рассматривается задача последовательного (в реальном или приведенном масштабе времени) обнаружения свойств многомерных временных рядов в процессе адаптивной идентификации модели VAR процесса (1.7). При этом исходной информацией для решения задачи идентификации и обнаружения изменений является только сам n -мерный временной ряд F , значения которого в реальном времени поступают на адаптивный идентификатор.

Для упрощения дальнейших выкладок введем в рассмотрение составные матрицу $K = (K_0 \parallel K_1 \parallel \dots \parallel K_N)$ и вектор $B\phi(i) = (1, B^T(i-1), \dots, B^T(i-p))^T$ размерности $((Nn+1) \times 1)$ соответственно, после чего перепишем уравнение (1.7) в виде

$$B(i) = KB\phi(i) + x(i), \quad (4.1)$$

где матрица априори неизвестных параметров K содержит практически всю необходимую информацию о свойствах контролируемого сигнала.

Задача идентификации состоит в том, что в соответствие реальному сигналу (4.1) ставится настраиваемая модель

$$\hat{B}(i) = K(i-1)B\phi(i), \quad (4.2)$$

матрица параметров $B\Phi(i)$ которой, уточняется на каждом такте времени i путем минимизации принятого критерия идентификации, являющегося некоторой функцией различия рассчитанных $\hat{B}(i)$ и наблюдаемых данных $B(i)$. При этом синтезированная модель (4.2) обязана быть работоспособной и в режиме прогнозирования, а нарушение прогнозирующих свойств может быть признаком возникновения рекуррентных процедур, которые могут быть представлены в обобщенном виде [76]

$$\begin{cases} \dot{K}(i) = K(i-1) + g(i)e(i)B\Phi^T(i), \\ \dot{e}(i) = B(i) - \hat{B}(i) = B(i) - K(i-1)B\Phi(i), \end{cases} \quad (4.3)$$

где $g(i)$ – скалярный или матричный коэффициент усиления алгоритма, определяющий его свойства и зависящий от принятого критерия идентификации,

$e(i)$ – векторная ошибка идентификации.

На практике наибольшее распространение получили алгоритмы, связанные с критерием минимума суммы квадратов ошибок идентификации

$$I(i) = \underset{u=1}{\overset{i}{\mathop{\text{a}}}} b(u) \|e(u)\|^2 = \underset{u=1}{\overset{i}{\mathop{\text{a}}}} \underset{j=1}{\overset{n}{\mathop{\text{a}}}} b(u) e_j(u)^2 \quad (4.4)$$

и его модификации, определяемые принятой системой весов $b(u)$. Наиболее популярным является метод наименьших квадратов, в котором все веса имеют одинаковый вес, т.е.

$$I(i) = \underset{u=1}{\overset{i}{\mathop{\text{a}}}} \|e(u)\|^2 \quad (4.5)$$

Алгоритм, соответствующий (4.3), (4.5), имеет вид

$$\begin{aligned} K(i) &= K(i-1) + \frac{e(i)B\Phi^T(i)P(i-1)}{1 + B\Phi^T(i)P(i)B\Phi}, \\ P(i) &= P(i-1) - \frac{P(i-1)B\Phi(i)B\Phi^T(i)P(i-1)}{1 + B\Phi^T(i)P(i-1)B\Phi(i)} \end{aligned} \quad (4.6)$$

и при неизменных параметрах (4.1) обеспечивает монотонную сходимость оценок $K(i)$ к истинным значениям параметров K .

Альтернативой (4.5), (4.6) является одношаговая процедура

$$K(i) = K(i-1) + \frac{e(i)B\Phi^T(i)}{B\Phi^T(i)B\Phi(i)}, \quad (4.7)$$

порождаемая одношаговым же критерием идентификации $I(k) = \|e(k)\|^2$ и являющаяся обобщением алгоритма Качмажа [223] на векторно-матричную модель (4.2). Обладая высоким быстродействием, процедура (4.7) не имеет фильтрующих свойств, а потому не способна различить изменения в сигнале и влияние стохастического компонента $x(i)$.

В связи с этим представляется целесообразным использование алгоритмов с конечной памятью, обладающих как сглаживающими, так и следящими свойствами, компромисс между которыми и задается величиной памяти. Признаком возникающих изменений может служить потеря прогнозирующих свойств модели (4.2) и необходимость перестройки памяти алгоритма.

Возвращаясь к критерию (4.4), заметим, что в классе порождаемых им алгоритмов наибольшее распространение получил экспоненциально взвешенный метод наименьших квадратов с критерием

$$I(i) = \sum_{u=1}^i b^{i-u} \|e(u)\|^2$$

и рекуррентной процедурой настройки

$$\begin{aligned} \dot{\ddot{}} K(i) &= K(i) + \frac{e(k)B\Phi^T(i)P(i-1)}{b + B\Phi^T(i)P(i-1)B\Phi(i)}, \\ \dot{\ddot{}} P(i) &= \frac{1}{b} \left(P(i-1) - \frac{P(i-1)B\Phi(i)B\Phi^T(i)P(i-1)}{b + B\Phi^T(i)P(i-1)B\Phi(i)} \right), \end{aligned} \quad (4.8)$$

где $0 \leq b \leq 1$ – параметр сглаживания.

Здесь следует заметить, что идентификатор с экспоненциальным сглаживанием в общем случае является неустойчивым, что ведет к «взрыву параметров» ковариационной матрицы, который особенно часто возникает при высоких размерностях обрабатываемого сигнала $B(i)$. Таким образом, использование традиционного экспоненциально взвешенного рекуррентного метода наименьших квадратов усложняется плохой обусловленностью информационной матрицы

$$\sum_{u=1}^i b^{k-u} B\Phi(u)B\Phi^T(u),$$

порождаемой высоким уровнем корреляции между компонентами $x_j(i)$.

В [19] эта проблема решается использованием вместо операции обращения взвешенной информационной матрицы, основанной на формуле Шермана-Моррисона, операции псевдообращения с использованием теоремы Гревилля. Полученный при этом алгоритм весьма громоздок с вычислительной точки зрения, что особенно сказывается при больших n .

В связи с этим предлагается ввести в рассмотрение многомерную модификацию алгоритма экспоненциально взвешенной стохастической аппроксимации [25] в виде

$$\begin{aligned} K(i) &= K(i-1) + \frac{e(i)B\Phi^T(i)}{br(i-1) + \|B\Phi(i)\|^2}, \\ r(i) &= br(i-1) + \|B\Phi(i)\|^2, \end{aligned} \quad (4.9)$$

являющуюся своеобразным компромиссом между процедурами (4.7) и (4.8) и обладающую необходимыми сглаживающими и следящими свойствами.

Механизм адаптации алгоритмов (4.8), (4.9) основан на «подавлении» устаревшей информации, при этом динамические свойства алгоритмов полностью определяются памятью, на основе которой происходит уточнение матрицы текущих оценок $K(i)$.

Как уже отмечалось, наилучшими фильтрующими свойствами обладают алгоритмы с неограниченной памятью, такие, как рекуррентный метод наименьших квадратов, в то же время эти процедуры имеют плохие следящие свойства в случае, если характеристики контролируемого сигнала изменяются во времени. В этом случае преимущество имеют алгоритмы с малой памятью типа алгоритма Качмажа, однако эти процедуры плохо работают в условиях помех.

Таким образом, при идентификации систем, в которых могут происходить изменения, объем памяти алгоритма следует выбирать на основе компромисса между его сглаживающими и следящими свойствами. К сожалению, в реальных ситуациях характеристики возмущений, дрейфов и скачков априори неизвестны и могут изменяться в процессе функционирования контролируемой системы. В этих условиях тяжело отдать предпочтение одному алгоритму с фиксированной памятью, а поэтому целесообразно использовать алгоритмы с переменной памятью, величину которой можно оперативно изменять в зависимости от наличия или отсутствия изменений.

В [148] предложен метод регулирования параметра сглаживания b , основанный на контроле статистики, характеризующей ошибку прогнозирова-

ния одномерного сигнала. Предполагается, что настройка параметров модели производится с помощью экспоненциально взвешенного алгоритма Калмана-Мейна, который для j -й компоненты $b_j(i)$ может быть записан в виде

$$\begin{aligned} \dot{k}_j(i) &= k_j(i-1) + \frac{(e(i)B\Phi^T(i)P_j(i-1))}{bs_i^2 + B\Phi^T(i)P_j(i-1)B\Phi(i)}, \\ P_j(i) &= \frac{1}{b} \frac{\Phi P_j(i-1) \Phi^T - \frac{P_j(i-1)B\Phi(i)B\Phi^T(i)P_j(i-1)}{bs_i^2 + B\Phi^T(i)P_j(i-1)B\Phi(i)}}{\Phi} \end{aligned} \quad (4.10)$$

где $k_j(i)$ – j -я строка матрицы $K(i)$,

$(0)_j$ – j -я строка соответствующего произведения матриц.

Если же дисперсии отдельных компонент $x_j(i)$ вектора возмущений неизвестны, то в (4.10) можно использовать оценку в виде

$$\begin{aligned} \dot{s}_j^2(i) &= s_j^2(i-1) + p_j(i-1)(s_j^2(i-1) - l_j^2(i)), \\ p_j(i) &= \frac{1}{b} \frac{\Phi p_j(i-1) \Phi^T - \frac{p_j^2(i-1)}{b + p_j(i-1)}}{\Phi} \end{aligned} \quad (4.11)$$

Также предполагается, что модель достаточно точно описывает контролируемый сигнал на временных интервалах t наблюдений, а параметры могут изменяться скачками в произвольные моменты времени i_a . Переменное значение $b(i)$ регулируется с помощью статистики

$$T_j(i) = \frac{\sum_{u=i-t}^i l_j^2(u)}{b(i-1)s_j^2 + B\Phi^T(u)p_j(u-1)B\Phi(u)}, \quad (4.12)$$

имеющей c^2 распределение с t степенями свободы, при этом $b(0) = 1$. Регулирование $b(i)$ производится в дискретные моменты времени $j t$ по такому правилу

$$b(i) = \begin{cases} 1 & \text{при } i < t, i = j t \text{ и } T_j(i) \leq c_g^2, \\ b(i-1) - \Delta b & \text{при } i = j t \text{ и } T_j(i) > c_g^2, \\ b(i-1) & \text{при } j t < i < (j+1)t, j = 1, 2, \dots \end{cases} \quad (4.13)$$

Здесь c_g^2 – квантиль закона c^2 , соответствующий уровню значимости g , Δb – шаг регулирования.

Правило (4.13) предусматривает изменение $b(i)$ при значениях i кратных t , при промежуточных i величина $b(i)$ остается неизменной, а факт изменений фиксируется в момент реализации второго соотношения (4.13).

Громоздкость и инерционность этой процедуры вынуждает искать иные более быстрые и эффективные приемы обнаружения изменений. Так, в [181] предложен метод регулирования параметра сглаживания b на основе критерия Манна-Уитни. При этом контролируемой характеристикой является значение

$$\sum_{u=i-t+1}^i \text{sign}(b_j(u) - \hat{b}_j(u))^3 g \quad (4.14)$$

где g – некоторое пороговое значение,

t – величина скользящего контрольного окна,

$$\text{sign}(b_j(u) - \hat{b}_j(u)) = \begin{cases} 0 & \text{при } b_j(u) = \hat{b}_j(u), \\ +1 & \text{при } b_j(u) > \hat{b}_j(u), \\ -1 & \text{при } b_j(u) < \hat{b}_j(u). \end{cases}$$

Процесс контроля начинается со значения $b(1) = 0$, что соответствует максимальному быстродействию алгоритма (4.19). Экспоненциально взвешенный рекуррентный метод наименьших квадратов в этой ситуации, естественно, принципиально неработоспособен. В процессе идентификации возможно возникновение следующих ситуаций:

$$\mathring{a} \sum_{u=i-t+1}^t \text{sign}(b_j(u) - \hat{b}_j(u))^3 d,$$

что означает доминирование стохастической компоненты x_i сигнала b_j над «дрейфовой». В этом случае необходимо улучшить сглаживающие свойства алгоритма, а именно, увеличить память по правилу

$$b(i) = b(i - 1) + Db.$$

В ситуации (4.14) преобладает дрейфовая компонента сигнала, и алгоритм не успевает отслеживать возникшие изменения. В этой ситуации необходимо уменьшить память

$$b(i) = b(i - 1) - Db$$

и зафиксировать факт изменений.

Для контроля за изменениями многомерного временного ряда предлагается использовать модификацию (4.14) в виде

$$\max_j \left(\mathring{a} \sum_{u=i-t+1}^t \text{sign}(b_j(u) - \hat{b}_j(u))^3 \right) g, \quad (4.15)$$

т.е. одновременно контролировать все компоненты и фиксировать факт из-

менения, если разладка произойдет в хотя бы одной компоненте $b_j(i)$.

На практике более широкое распространение получили не статистические, а эвристические процедуры, такие, как методы Чоу, Робертса-Рида, Брауна, Тригга-Лича, Шоуна и т.п. [145, 181, 263], имеющие в своей основе некоторые эвристики, а поэтому несущие элемент субъективизма. В условиях дефицита априорной и текущей информации о характеристиках и свойствах контролируемого сигнала предпочтение следует отдать этим методам.

Основа большинства эвристических методов следующая: задается набор значений параметра сглаживания \mathbf{b} и набор характеристик, определяющих качество идентификации. Как правило, это:

- текущая ошибка оценивания i -й компоненты

$$e_j(i, \mathbf{b}) = b_j(i) - \hat{b}_j(i, \mathbf{b});$$

- кумулятивная сумма ошибок

$$S_j(i, \mathbf{b}) = e_j(i, \mathbf{b}) + S_j(i - 1, \mathbf{b});$$

- среднее абсолютное значение ошибки

$$d_j(i, \mathbf{b}) = (1 - \mathbf{b})|e_j(i, \mathbf{b})| + \mathbf{b}d_j(i - 1, \mathbf{b});$$

- средняя ошибка

$$\bar{e}_j(i, \mathbf{b}) = (1 - \mathbf{b})e_j(i, \mathbf{b}) + \mathbf{b}\bar{e}_j(i - 1, \mathbf{b});$$

- относительная средняя ошибка

$$\bar{\delta}_j(i, \mathbf{b}) = (1 - \mathbf{b})\frac{e_j(i, \mathbf{b})}{b_j(i)} + \mathbf{b}\bar{\delta}_j(i - 1, \mathbf{b});$$

- средняя квадратичная ошибка

$$\bar{e}_j^2(i, b) = (1 - b)e_j^2(i, b) + b\bar{e}_j^2(i - 1, b).$$

Если значение выбранной контролируемой характеристики превышает некоторый порог g , принимается решение о том, что произошли изменения и необходимо соответствующим образом откорректировать параметр сглаживания b .

Стационарный стохастический сигнал обычно сталкивается с b , лежащим в интервале $0,7 \leq b \leq 0,99$ [263]. Простейшая форма контроля состоит в том, что когда значение $\bar{e}_j^2(k, b)$ при некотором b превышает порог 0.05, параметр сглаживания уменьшается по правилу $b(i) = b(i - 1) - Db$ и процесс идентификации продолжается с новым $b(i)$. Если значение b превосходит порог $0.7 (b(i) \leq 0.7)$, принимается решение о возникновении изменения в сигнале.

Более эффективным, но и более сложным является метод Чоу. Здесь одновременно настраиваются три модели со значениями параметров сглаживания b , $b + Db$ и $b - Db$. Если в текущий момент времени k лучший результат получен, например, с $b(i) = b + Db$, то на следующем такте времени используется новая тройка параметров сглаживания b , $b + Db$ и $b + 2Db$. Если же лучшая модель получена при $b(i) = b - Db$, то формируется тройка b , $b - Db$ и $b - 2Db$, и, наконец, если лучший результат достигнут при $b(i) = b$, то сохраняется набор b , $b + Db$ и $b - Db$.

Более простыми и оперативными являются методы, использующие контрольный следящий сигнал (трекинг-сигнал), являющийся индикатором изменений контролируемых сигналов [145, 263]. Существует ряд форм следящего сигнала, при этом его выход за определенные границы свидетельствует о возникших изменениях.

Р. Брауном в качестве следящего сигнала предложено использовать выражение [173]

$$T_j^K(i) = \frac{\sum_{u=1}^i e_j(u)}{\sqrt{s_j^2(i)}}, \quad (4.16)$$

физический смысл которого состоит в том, что в случае, если настраиваемая модель адекватна контролируемому сигналу, сумма ошибок, в силу случайности которых, варьирует около 0, не превышая при этом некоторых границ, которые устанавливаются априорно для данного уровня вероятности при определенной дисперсии суммы ошибок прогноза, которая стремится к значению

$$\lim_{i \rightarrow \infty} s_j^2(i) = \frac{1}{1 - (1 - b)^{2(Nn+1)}} s_j^2.$$

Для практических расчетов Р. Браун вместо среднеквадратических ошибок s_j предложил использовать среднее абсолютное отклонение

$$d = \int_{-\infty}^{\infty} |e - M\{e\}| p(e) de \approx \sum_{u=0}^n |e(u) - M\{e\}| P_j.$$

Это отклонение пропорционально среднеквадратическому значению ошибки, поскольку

$$D = (e - M\{e\}) / s_j,$$

$$d_j = s_j \int_{-\infty}^{\infty} |D| p(D) dD,$$

при этом для широкого класса распределений коэффициент пропорциональности изменяется незначительно [145] (для нормального распределения

$$\frac{d_j}{s_j} = \sqrt{2p} = 0,7979).$$

Если величина d_j рассчитана согласно выражению

$$d_j(i) = (1 - b)|e_j(i)| + b d_j(i - 1),$$

а следящий сигнал по формуле

$$T_j^b(i) = \frac{\sum_{u=1}^i e_j(u)}{d_j(i)},$$

то $(1 - P_j)\%$ -ные границы можно записать в виде

$$\pm t_{j,1-P_j} = \pm \frac{s_j}{2} \sqrt{\frac{p(1-b)}{1 - (1-b)^{2(Nn+1)}}}.$$

Д. Тригг и А. Лич предложили в качестве следящего сигнала использовать отношение [307]

$$T_j^{TL}(i) = \frac{T_j^b(i)}{d_j(i)}, \quad (4.17)$$

где

$$T_j^b(i) = (1 - b)e_j(i) + b T_j^b(i - 1) \quad (4.18)$$

не общая сумма отклонений, а сглаженная ошибка; при этом должно соблюдаться неравенство $b \leq b$.

При $b = b$ следящий сигнал будет варьироваться между -1 и $+1$. Для введения автоматической обратной связи Д. Тригг и А. Лич предложили рас-

считывать параметр сглаживания согласно соотношению $b(i) = 1 - |T^{TL}(i)|$, а разладки в сигнале фиксировать при существенных изменениях $b(i)$.

Известна также модификация Шоуна [263] с $b(i) = 1 - |T^{TL}(i - 1)|$.

Возрастание следящего сигнала свидетельствует об увеличении расхождения между моделью и контролируемой последовательностью, для компенсации которого необходима более быстрая реакция идентификатора, которая обеспечивается более низким значением параметра сглаживания. Таким образом, обеспечивается отрицательная обратная связь.

В [145] подчеркивается, что при идентификации процессов с достаточно плавным дрейфом, преимущества имеет метод Брауна. Резкие скачки лучше определяются с помощью следящего сигнала Тригга-Лича. Именно поэтому эта форма может быть полезна при анализе изменений временного ряда, описываемых соотношениями (1.8) – (1.10).

4.2 Метод обнаружения изменений свойств многомерных временных рядов на основе экспоненциального сглаживания

Основным недостатком подхода к обнаружению изменений, рассмотренного выше, является значительное число параметров настраиваемой модели (4.12), составляющее $n(Nn + 1)$, что может вызывать определенные трудности при больших n и высоких частотах поступления информации на обработку.

Достаточно удобным математическим аппаратом для решения задачи обнаружения изменения свойств одномерных стохастических последовательностей является экспоненциальное сглаживание [145, 173, 263], суть которого можно пояснить следующим примером.

Введем в рассмотрение элементарную модель вида

$$b_j(i) = k_j + x_j(i) \quad (4.19)$$

и предположим, что коэффициент k_j время от времени может измениться скачком так, как это показано на рис. 4.1.

При этом необходимо отметить, что величина и момент времени изменения коэффициента k_j априори неизвестны, а интервал времени $d_j - a_j$, в течение которого значение коэффициента k_j остается неизменным, существенно превышает шаг квантования сигнала (время между двумя последовательными наблюдениями).

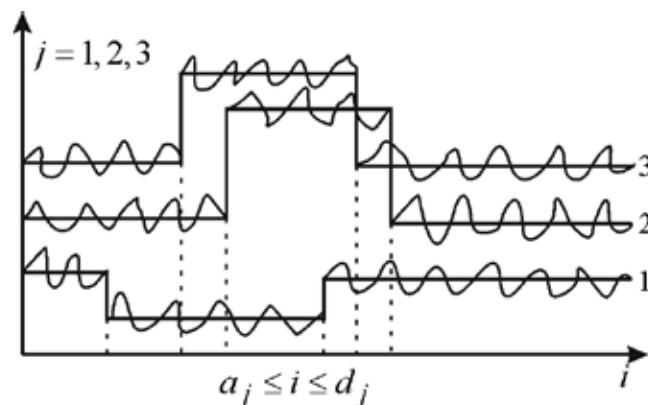


Рисунок 4.1 – Скачкообразные изменения среднего уровня временного ряда

Поскольку экспоненциальное сглаживание разработано для решения задачи прогнозирования, рассмотрим сначала задачу построения прогноза $\hat{b}_j(i-1)$ в момент времени i . Предположим также, что параметры модели (4.19) в момент $i+1$ совпадают с теми, которые получены к моменту времени i . В этом случае задача сводится к оценке текущего значения $k_j(i)$ по i предыдущим наблюдениям. Поскольку значение k_j с течением времени изменяется, то для получения этой оценки наблюдения $b_j(i), b_j(i-1), \dots$ следует брать с большим весом, или наблюдения, полученные значительно раньше. В принципе эта задача может быть успешно решена с помощью простейшего метода скользящего окна, при этом, если задан размер этого окна t , то сглаженная оценка имеет вид

$$\bar{b}_j(i) = \frac{1}{t} \sum_{u=i-t+1}^i b_j(u). \quad (4.20)$$

Скорость реакции идентификатора, использующего аппарат скользящего окна, на изменение в сигнале зависит от величины S усредняемых последних наблюдений. Эта скорость, увеличивается с уменьшением t в отсутствие изменений уменьшает точность получаемой оценки $k_j(i) = \bar{b}_j(i)$, поскольку ее дисперсия представляется в виде

$$s_{k_j}^2 = \frac{s_j^2}{t}.$$

Отсюда непосредственно следует, что точность оценки $k_j(i)$ и скорость реакции идентификатора являются противоречащими друг другу требованиями.

С учетом очевидного соотношения

$$\bar{b}_j(i) = \bar{b}_j(i-1) + \frac{b_j(i) - b_j(i-t)}{t}$$

несложно переписать (4.20) в рекуррентной форме

$$\bar{b}_j(i) = \frac{1}{t} b_j(i) + \left(1 - \frac{1}{t}\right) \bar{b}_j(i-1)$$

или

$$t_j(i) = a b_j(i) + (1-a) t_j(i-1) = a b_j(i) + b t_j(i-1). \quad (4.21)$$

Процедура (4.21) и является экспоненциальным сглаживанием.

В выражении (4.21) для отличия экспоненциального сглаживания от

скользящего среднего введено обозначение $t_j(i)$ вместо $\bar{b}_j(i)$. Величина a , являющаяся аналогом $\frac{1}{t}$ в скользящем среднем, определяет веса наблюдений, присутствующих в сглаженной оценке.

Из выражения (4.21) следует, что текущее значение сглаженной величины $t_j(i)$ равно предыдущему ее значению плюс некоторая доля разницы между текущим наблюдением и предыдущим значением сглаженной величины. Поскольку операция (4.21) реализуется для всех наблюдений временного ряда, ее можно переписать с учетом всех предыдущих наблюдений в виде

$$\begin{aligned} t_j(i) &= ab_j(i) + (1-a)(ab_j(i-1) + (1-a)t_j(i-2)) = \\ &= ab_j(i) + (1-a)(ab_j(i-1) + (1-a)(ab_j(i-2) + (1-a)t_j(i-3))) = \\ &= \dots = a \overset{i-1}{\underset{u=0}{\overset{\circ}{a}}} (1-a)^u b_j(i-u) + (1-a)^i b_j(0). \end{aligned}$$

Таким образом, величина $t_j(i)$ является некоторой комбинацией всех предыдущих наблюдений, вес которых со временем убывает по геометрической прогрессии. Текущее наблюдение имеет вес a , значение которого лежит в интервале $[0,1]$. Предельное значение $a=0$ соответствует случаю $t=\infty$ в скользящем среднем. При этом $t_j(i) = t_j(i-1)$, т.е. значение t_j не зависит от новой информации. Предельное значение $a=1$ означает, что предыстория вообще не влияет на текущую оценку, т.е. $t_j(i) = b_j(i)$, а $s_{k_j}^2 = s_j^2$.

Тем самым, точность и скорость реакции идентификатора на скорость изменения в сигнале полностью зависит от принятого значения a . Малая величина a обеспечивает большую точность оценки b_i при стационарном сигнале, но медленную реакцию на изменения, в то время как увеличение a будет способствовать увеличению скорости этой реакции.

Обычно [173], величина a лежит в пределах от 0.01 до 0.3, а число t –

в пределах от 6 до 200. Поскольку этот диапазон достаточно велик, в каждой конкретной задаче этот параметр нужно выбирать специальным образом. Необходимость использования больших значений a (малых значений t) указывает на несоответствие выбранной и реальной моделей контролируемого сигнала, т.е. может служить признаком изменения его свойств.

Выше отмечалось, что при экспоненциальном сглаживании вес текущего наблюдения имеет величину a , а веса предыдущих наблюдений убывают в обратном времени. В скользящем среднем средний вес последних t наблюдений принимается одинаковым и равным $\frac{1}{t}$. Веса же всех ранних наблюдений равны 0.

В [145] введено понятие среднего «возраста» в скользящем среднем в виде

$$r = \frac{1}{t}(0 + 1 + 2 + \dots + i - 1) = \frac{t(t - 1)}{2t} = \frac{1}{2}(t - 1). \quad (4.22)$$

Как следует из (4.22), средний «возраст» наблюдений есть среднее «возрастов» всех отдельных наблюдений, взятых с весами, равными весам этих отдельных наблюдений. При экспоненциальном сглаживании вес наблюдения, сделанного в момент времени $i - l$, будет ab^l , поэтому средний «возраст» наблюдений может быть записан в виде

$$r = (0ab^0 + 1ab^1 + \dots + lab^l + \dots) = a \sum_{u=0}^{\infty} ub^l.$$

Условие равенства среднего «возраста» наблюдений в скользящем среднем и при экспоненциальном сглаживании позволяет найти зависимость между параметром a и окном t в виде

$$\frac{b}{a} = \frac{1 - a}{a} = \frac{t - 1}{2}$$

или

$$\begin{cases} \dot{a} = \frac{t-1}{2}, \\ \dot{b} = \frac{t-1}{t+1}. \end{cases}$$

Особый интерес представляет реакция идентификатора с экспоненциальным сглаживанием на скачки, возникающие в контролируемом сигнале. Допустим, что в момент времени i_a происходит единичный скачок, т.е. $k_j^a = k_j + 1$ при $i \geq i_a$. Воспользовавшись стандартным дискретным z -преобразованием [221], с учетом того, что $z[1] = \frac{z}{z-1}$, можно записать

$$z[k_j(i)] = \frac{az}{z-b^{T_0}} \frac{z}{z-1}, \quad \text{или} \quad \frac{az}{z-b^{T_0}} \frac{z}{z-1} = \frac{b^{T_0}}{b^{T_0}-1} \frac{az}{z-b^{T_0}} + \frac{a}{1-b^{T_0}} \frac{z}{z-1}, \quad \text{где}$$

T_0 – период квантования контролируемого сигнала.

Переходя обратно во временную область, получаем

$$b_j(i) = 1 \frac{a}{1-b^{T_0}} - a \frac{b^{T_0}}{1-b^{T_0}} b^i,$$

откуда следует, что идентификатор с экспоненциальным сглаживанием приходит к новому установившемуся состоянию $k_j + 1$ или быстрее, чем меньше b (больше a).

Общая процедура экспоненциального сглаживания (4.21) предназначена для обработки одномерных сигналов $b_j(i)$. В рамках рассматриваемой нами проблемы целесообразно ввести экспоненциальное сглаживание многомерных последовательностей в форме

$$t(i) = AB\phi(i) + (I - A)t(i-1), \quad (4.23)$$

где $t(i) = (t_1(i), t_2(i), \dots, t_n(i))^T$,

$A = \text{diag}(a_1, a_2, \dots, a_n)$ - $n \times n$ диагональная матрица,

I - $n \times n$ единичная матрица.

Если для контроля изменений использовать следящий сигнал Триггера Лича

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

по аналогии с (4.23) несложно ввести его векторный аналог

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

при этом, естественно, контролируется каждая компонента ($n \times 1$) – вектора $T^{TL}(i)$.

Возникновение изменений в процессе обнаруживается путем проверки неравенства

$$\max_i (T^{TL}(i) - T^{TL}(i-1)) \geq g$$

по типу (4.15).

Несколько сложнее обстоит дело с обнаружением изменений характеристик (дисперсий) возмущений $x_j(s_j^2)$ (1.9) и внутренней структуры мно-

гомерного временного ряда (1.10).

Для контроля дисперсий введем вектор квадратов ошибок (4.16)

$S(i) = (e_1^2(i), e_2^2(i), \dots, e_n^2(i))^T$ и экспоненциально сглаженный вектор дисперсий

$$t_{s2}(i) = AS(i) + (I - A)t_{s2}(i - 1). \quad (4.24)$$

Понятно, что для стационарного сигнала и $a_j = 1/i$ выражение (4.24) описывает дисперсии $s_{k_j}^2$ оценок компонент k_j , однако в нестационарной ситуации при $0 \leq a_j \leq 1$, рост компонент вектора (4.24) свидетельствует о возникновении изменений. Поскольку в данной ситуации использование следящего сигнала невозможно, необходимо контролировать условие

$$\max_i (t_{s2}(i) - t_{s2}(i - 1)) \leq g_{s2}.$$

Анализ внутренней структуры стационарной многомерной временной последовательности может быть проведен с помощью ее корреляционной матрицы вида

$$R(i, t) = \frac{1}{i} \sum_{u=1}^i (B(u) - \bar{B})(B(u - t) - \bar{B})^T, \quad t = 0, 1, 2, \dots, t_{max},$$

которая содержит информацию об автокорреляционных и взаимно корреляционных функциях всех компонент $b_j(i)$.

Для обнаружения изменений свойств можно ввести экспоненциально сглаженную корреляционную матрицу

$$t_R(i,t) = a(B(i) - t(i))(B(i-t) - t(i))^T,$$

а контроль вести с помощью неравенства

$$Sp(t_R(i,t)t_R^T(i,t)) - Sp(t_R(i-1,t)t_R^T(i-1,t)) \geq \theta_R,$$

где $Sp(0)$ – символ следа матрицы,

$Sp(t_R t_R^T)$ – квадрат сферической нормы матрицы t_R .

Таким образом, на основе методологии экспоненциального сглаживания можно в реальном времени обеспечить контроль за изменениями всех характеристик многомерных временных рядов.

4.3 Метод обнаружения изменения свойств многомерных временных рядов на основе анализа главных компонент

Важной проблемой при анализе больших массивов (как по объему, так и по размерности) наблюдений, заданных в форме временных рядов, является задача их сжатия с целью выделения латентных факторов, определяющих внутреннюю структуру контролируемого сигнала, что в конечном итоге преследует цель сделать исходный временной ряд более просто интерпретируемым с точки зрения обнаружения изменений свойств.

Одним из наиболее эффективных подходов к решению этой задачи является аппарат факторного анализа, в рамках которого наиболее широкое распространение получил метод главных компонент особенно в задачах распознавания образов, обработки изображений, спектрального анализа и т.п. и известного еще как преобразование Карунена-Лоева.

Исходной информацией для анализа является $(i \times n)$ матрица наблюдений

$$B = \begin{pmatrix} b_1(1) & b_2(1) & \dots & b_n(1) \\ b_1(2) & b_2(2) & & b_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(u) & b_2(u) & \dots & b_n(u) \\ \vdots & \vdots & \ddots & \vdots \\ b_1(i) & b_2(i) & \dots & b_n(i) \end{pmatrix}$$

образованная массивом из i n -мерных векторов наблюдений $B(u) = (b_1(u), b_2(u), \dots, b_n(u))^T$, что для задачи информационного поиска в видеоданных является представлением кадра при помощи некоторого вектора характеристик и ее корреляционная ($n \times n$) матрица вида

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

где $B^c(u) = B(u) - \bar{B}$ центрированные относительно среднего исходные данные.

Метод главных компонент состоит в проецировании наблюдаемых исходных данных из исходного n -мерного пространства в m -мерное ($n > m \geq 1$) выходное и сводится к нахождению системы w^1, w^2, \dots, w^m ортогональных собственных векторов матрицы $R(i)$ таких, что $w^1 = (w_1^1, w_2^1, \dots, w_n^1)^T$ соответствует наибольшему собственному значению λ_1 матрицы $R(i)$, $w^2 = (w_1^2, w_2^2, \dots, w_n^2)^T$ – второму по величине собственному значению λ_2 и т.д. Иначе говоря, речь идет об отыскании решения матричного уравнения

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

такого, что $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ и $\|w^l\|^2 = 1$.

Таким образом, в алгебраических терминах решение этой задачи тесно связано с проблемой собственных значений и нахождения ранга корреляционной матрицы; в геометрическом смысле – это задача перехода в пространство более низкой размерности с минимальной потерей информации; в статистическом смысле – это задача нахождения множества ортонормальных векторов в пространстве входов, принимающих на себя максимальную вариацию данных, и, наконец, в алгоритмическом смысле – это задача последовательного определения (выделения) множества собственных векторов w^1, w^2, \dots, w^m путем оптимизации каждого из локальных функционалов, образующих глобальный критерий

$$I_w(i) = \frac{1}{i} \sum_{l=1}^m \sum_{u=1}^i (B^c T(u) w^l)^2$$

при ограничениях

$$\begin{cases} w^{lT} w^N = 0, & \text{при } l \neq N, \\ w^{lT} w^N = 1, & \text{иначе.} \end{cases}$$

Первая главная компонента, несущая максимум информации о контролируемом сигнале, может быть найдена путем максимизации локального критерия

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

с помощью стандартного метода неопределенных множителей Лагранжа.

Далее из каждого вектора $B^c(u)$ вычитается его проекция на первую главную компоненту и вычисляется первая главная компонента остатков, яв-

ляющаяся второй главной компонентой исходных данных и ортонормальная к первой.

Третья главная компонента вычисляется путем проектирования каждого исходного вектора на первую и вторую главные компоненты, вычитание этой проекции из каждого $B^c(u)$ и нахождения первой главной компоненты полученных остатков, являющейся третьей компонентой исходных данных. Остальные главные компоненты вычисляются рекурсивно согласно описанной стратегии.

К настоящему времени разработано достаточно развитое математическое и программное обеспечение для реализации преобразования Карунена-Лоева, объединение одним общим недостатков: необходимостью априорного задания матрицы X фиксированной размерности. Если же данные поступают последовательно в реальном времени, стандартные процедуры факторного анализа становятся неработоспособными.

Тем самым, следует использовать рекуррентные процедуры реального времени для нахождения собственных векторов матрицы $R(i)$ путем последовательной обработки наблюдений многомерного временного ряда $B(1), B(2), \dots, B(i), B(i+1) \dots$ без вычисления самой корреляционной матрицы. Введем искусственный нейрон на основе адаптивного линейного ассоциатора для вычисления первой главной компоненты в реальном времени. На рис. 4.2 произведена схема этого нейрона, модифицированная для решения задачи обнаружения изменения свойств в многомерном сигнале на основе анализа главной компоненты.

Для предварительно центрированных данных алгоритм обучения имеет вид

$$\begin{cases} \dot{w}^1(i+1) = w^1(i) + h(i+1)(B(i+1) - y(i)w^1(i))y(i+1), \\ \dot{y}(i+1) = B^T(i+1)w^1(i), \quad w^1(0) = 0, \quad y(1) = B^T(1)w^1(0) \end{cases} \quad (4.25)$$

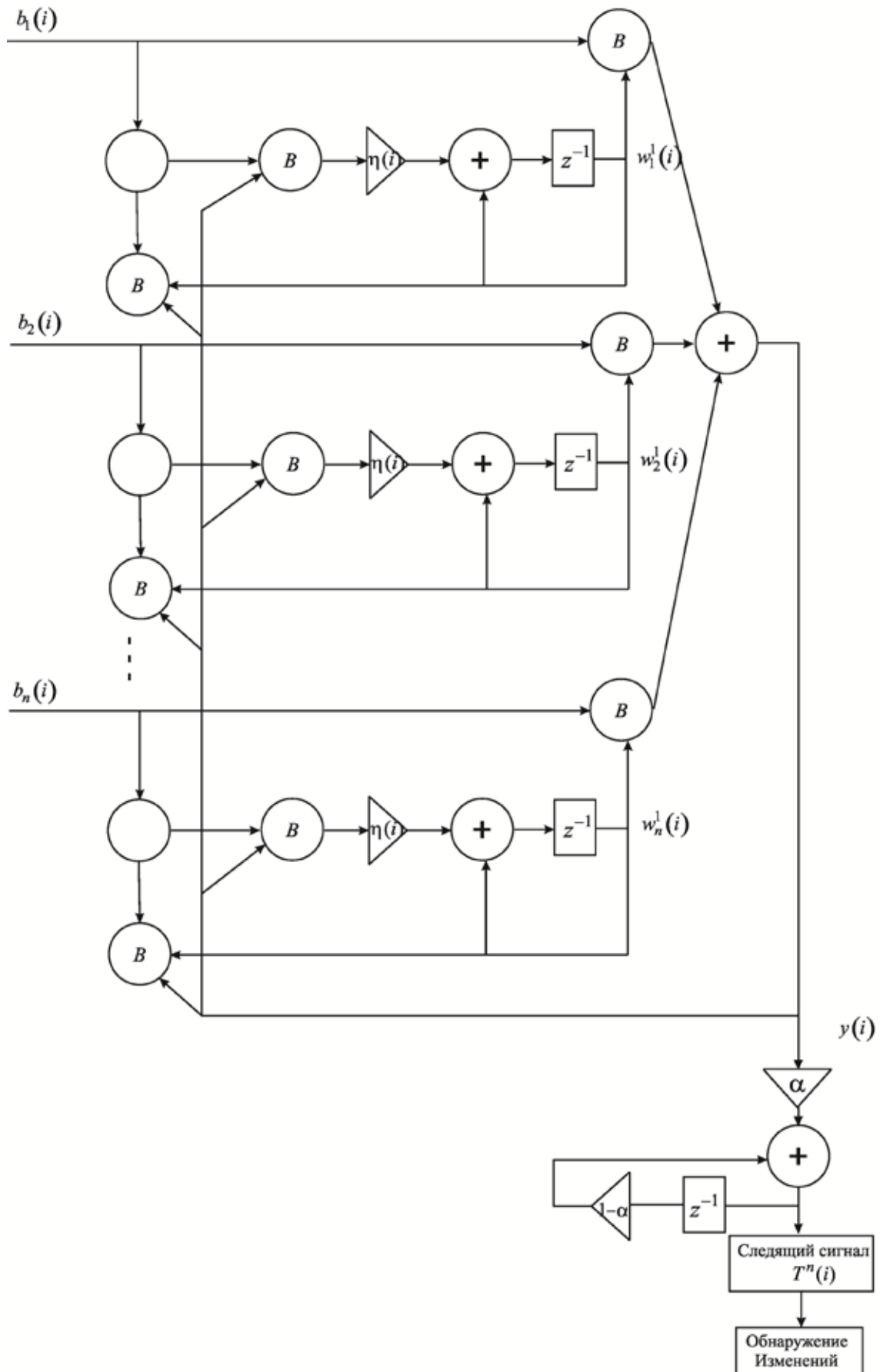


Рисунок – 4.2 Модифицированный нейрон для обнаружения изменения свойств главной компоненты многомерного временного ряда

где $h(i+1)$ – параметр шага настройки выбираемый достаточно малым.

Алгоритм (4.25) обеспечивает нормирование вектора $w^1(i)$

$$\|w^1(i)\|^2 = 1,$$

а сам вектор $w^1(i)$ является собственным вектором матрицы $R(i)$, соответствующим максимальному собственному значению, а выходной сигнал $y(i)$ характеризуется максимально возможной дисперсией, т.е. содержит максимум информации о многомерном входном сигнале $b_n(i)$.

Далее выходной сигнал $y(i)$ подвергается экспоненциальному сглаживанию, фильтрующему шумовые компоненты $x(i)$, а обнаружение изменения свойств происходит с помощью одномерного следящего сигнала $T^{TL}(i)$ (4.17), (4.18).

В результате исследований на практических видеорядах получен следующий пример выявления разладок многомерного временного ряда при помощи нейронной сети (рис. 4.3). Как видно из рисунка, колебания значений кадров происходят практически постоянно, однако тренд меняется только в

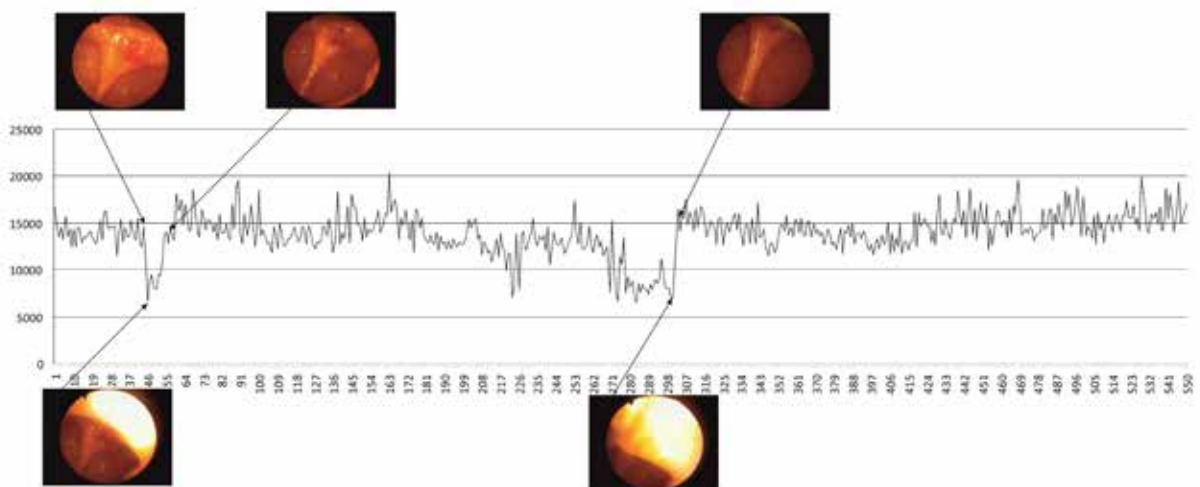


Рисунок 4.3 – Детектирование изменений в видеоряде

случае серьезных колебаний таких, как показано на приведенных кадрах, произошедших на двух сегментах видео.

4.4 Методы темпоральной сегментации на базе адаптивного прогнозирования многомерных временных рядов

Еще одним подходом по выделению сегментов видео на базе анализа многомерных временных рядов может являться их адаптивное прогнозирование, когда заметный разрыв между прогнозом и наблюдаемыми данными служит индикатором.

В основе традиционных математических методов прогнозирования (статистических, адаптивных, нейросетевых и т.п.) лежат математические модели того или иного вида, получаемые в результате решения задачи структурной и параметрической идентификации. Именно на основе математических моделей решается задача временной экстраполяции, при этом в модели явно или неявно присутствует аргумент дискретного времени. Если же данных для построения математической модели недостаточно, синтез прогнозирующей модели просто невозможен. В этом случае вместо временной экстраполяции может быть использовано пространственное прогнозирование (экстраполяция), сводящееся к оценке значений векторного поля по отдельным наблюдениям. Среди методов пространственной экстраполяции в качестве одного из наиболее перспективных следует отметить многомерную линейную экстраполяцию [130], подтвердившую свою эффективность при решении ряда реальных задач проектирования и управления сложными многомерными нелинейными объектами.

Сначала рассмотрим метод многомерной линейной экстраполяции применительно к задаче одношагового прогнозирования n -мерной нелинейной нестационарной временной последовательности применительно к опи-

санной нами модели. Полагая, как и ранее $\bigcup_{i=1}^N B(i)$, тогда анализируемый ряд

(i -я его компонента) может быть описана нелинейной зависимостью (NARX-модель) вида

$$\begin{aligned} \hat{B}_i(k) &= f_i(B_i(k-1), \dots, B_i(k-n_{A,i}), B\Phi(k-1), \dots, B\Phi(k-n_C), \\ &B\Phi(k-1), \dots, B\Phi(k-n_C), \dots, B\Phi(k-l), \dots, B\Phi(k-n_C)) = \\ &= f_i(R_{i1}(k), \dots, R_{i,n_{A,i}}(k), \dots, R_{i,n_{A,i}+n_C,q}(k)), \end{aligned} \quad (4.26)$$

где $f_i(0)$ – априори неизвестная нелинейная зависимость, подлежащая восстановлению на основании имеющихся наблюдений;

$\hat{B}_i(k)$ – оценка (прогноз) контролируемой последовательности $B_i(k)$ по данным, имеющимся к $(k-1)$ -му моменту времени;

$i=1, \dots, n$; $n_{A,i}$ – глубина учитываемой предыстории по контролируемой последовательности;

$B\Phi(k-l)$ – p -я компонента многомерного экзогенного сигнала, влияющего на $B_i(k)$; $l=1, \dots, n_C$; $p=1, \dots, q$.

Выражение (4.26) может быть представлено и в векторно-матричной форме

$$\hat{B}(k) = F(B(k-1), \dots, B(k-n_{A,i}), B\Phi(k-1), \dots, B\Phi(k-n_C)) = F(R(k)). \quad (4.27)$$

При наличии репрезентативной обучающей выборки нелинейные преобразования $f_i(0)$ и $F(0)$ могут быть восстановлены в процессе обучения той или иной искусственной нейронной сети, однако если эта выборка мала, нейросетевой подход оказывается неработоспособным, в то время как многомерная линейная экстраполяция позволит получить вполне приемлемые по точности результаты.

Задачу многомерной линейной экстраполяции применительно к проблеме прогнозирования многомерных временных рядов можно описать, используя обозначения (4.26), (4.27), следующим образом [128-130]. Пусть за-

дана матрица прецедентов (предыстории) в виде

$$R = \begin{matrix} \begin{matrix} \text{æ} \\ \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ø} \end{matrix} R^T(1), & B^T(1) \\ \begin{matrix} \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ø} \end{matrix} R^T(2), & B^T(2) \\ \begin{matrix} \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ø} \end{matrix} \mathbb{M} & \mathbb{M} \\ \begin{matrix} \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ç} \\ \text{ø} \end{matrix} R^T(N), & B^T(N) \end{matrix} \begin{matrix} \text{ö} \\ \div \\ \div \\ \div \\ \div \end{matrix}$$

размерности N' ($nn_A + qn_C + n$). Сама экстраполяция сводится к нахождению в момент времени N оценки

$$B(N+1) = y(R(N+1), R). \quad (4.28)$$

Здесь $y(0)$ – собственно алгоритм экстраполяции, который должен удовлетворять ряду требований, основным из которых является то, что в результате его использования должны точно восстанавливаться все следствия матрицы прецедентов, т.е.

$$\hat{B}(k) = B(k) = y(R(k), R), \quad k = 1, 2, \dots, N \quad (4.29)$$

Важно отметить, что методы прогнозирования, основанные на тех или иных математических моделях, практически никогда не обеспечивают выполнение условия (4.29).

Из других требований можно отметить то, что алгоритм $y(0)$ должен быть векторным, т.е.

$$y = (y_1, \dots, y_n), \quad B_i = y_i(Z, R), \quad i = 1, \dots, n; \quad (4.30)$$

трудоемкость реализации должна расти по n и N не быстрее чем линейно, алгоритм должен быть работоспособным при любых N (даже при $N=1$).

Заметим также, что при малых N , не говоря уже о $N=1$, достаточно точная математическая модель в принципе не может быть построена, хотя в случае анализа видеоданных данный недостаток не является столь критичным.

Следуя [128], представим алгоритм многомерной линейной экстраполяции с евклидовой метрикой в виде следующей последовательности шагов:

1. Формирование предыстории прогнозируемого процесса в виде матриц

$R\Phi = (R(1), R(2), \dots, R(N))$ – матрица размерности $(nn_A + qn_C) \times N$

$B = (B(1), B(2), \dots, B(N))$ – матрица размерности $n \times N$.

2. Нахождение вектора весовых коэффициентов $l = (l_1, l_2, \dots, l_N)^T$, доставляющих минимум норме

$$\|R(N+1) - \sum_{k=1}^N l_k R(k)\|^2 = \|R\Phi(N+1) - R\Phi l\|^2.$$

3. Формирование оптимального прогноза в виде линейной комбинации

$$\hat{B}(N+1) = \sum_{k=1}^N l_k B(k) = \bar{B}l. \quad (4.31)$$

Минимизация выражения (4.31) может быть проведена с помощью стандартного метода наименьших квадратов, в результате чего получаем известное решение

$$l = (R\Phi^T R\Phi)^{-1} R\Phi^T R(N+1), \quad (4.32)$$

которое существует только тогда, когда матрица $R\Phi^T R\Phi$ невырождена. Поскольку при малых N ($N < nn_A + qn_C$) это не так, вместо операции обращения матриц предлагается использовать операцию псевдообращения, в результате чего приходим к выражению

$$l = R\phi R(N+1), \quad (4.33)$$

откуда следует, что фактически задача сводится к нахождению ортогональной проекции вектора $R(N+1)$ на линейную оболочку, образованную векторами предыстории $R\phi$

С вычислительной точки зрения никаких затруднений при реализации данной процедуры не возникает, однако решение задачи усложняется, если данные на обработку поступают последовательно в реальном времени $k=1,2,\dots$. При этом все предыдущие соотношения можно переписать в следующей форме:

$$\begin{aligned} R\phi &= (R(1), R(2), \dots, R(k)), \\ \bar{B}_k &= (B(1), B(2), \dots, B(k)), \\ \|R(k+1) - \hat{a}_{l=1}^k l_l R(l)\|^2 &= \|R(k+1) - R\phi L_k\|^2, \end{aligned} \quad (4.34)$$

$$L_k = (l_1, l_2, \dots, l_k)^T,$$

$$\bar{B}(k+1) = \hat{a}_{l=1}^k l_l B(l) = \bar{B}_k L_k, \quad (4.35)$$

$$L_k = R\phi^+ R(k+1). \quad (4.36)$$

В [130] для расчета матрицы $R\phi_{k+1}^+$ по имеющейся $R\phi$ и поступившим на обработку $R(k+1)$ и $B(k+1)$ предлагается использовать формулу Гревилля, хотя гораздо проще применить регуляризованную версию (4.32) в виде

$$L_k = (R\phi^T R\phi + gI_k)^{-1} R\phi R(k+1), \quad (4.37)$$

где g – параметр регуляризации,

$I - (k \times k)$ – единичная матрица.

Для обработки нестационарных временных рядов, характеристики которых непредсказуемо меняются во времени, вместо обработки всей имеющейся выборки, целесообразно решать задачу на «скользящем окне», состоящей из c последних наблюдений. При этом соотношения (4.34) – (4.36) могут быть представлены следующим образом:

$$\begin{aligned}
 R_{k,c} &= (R(k - c + 1), R(k - c + 2), \dots, R(k)), \\
 \bar{B}_{k,c} &= (B(k - c + 1), B(k - c + 2), \dots, B(k)), \\
 \|R(k + 1) - \hat{a}_{l=k-c+1}^k \|_1 R(l)\|^2 &= \|R(k + 1) - R_{k,c} L_{k,c}\|^2, \\
 L_{k,c} &= (l_{k-c+1}, \dots, l_k)^T, \\
 \bar{B}(k + 1) &= \hat{a}_{l=k-c+1}^k \|_1 B(l) = \bar{B}_{k,c} L_{k,c}, \\
 L_{k,c} &= R_{k,c}^+ R(k + 1).
 \end{aligned} \tag{4.38}$$

Для реализации этой процедуры в реальном времени в [129] был предложен рекуррентный алгоритм псевдообращения на «скользящем окне», отличающийся громоздкостью и вычислительной сложностью.

В связи с этим целесообразно вместо операции рекуррентного псевдообращения использовать модификацию (4.37) в «оконном» варианте, при этом его пакетная форма может быть записана в виде

$$L_{k,c} = (R_{k,c}^T R_{k,c} + gI_c)^{-1} R_{k,c}^T R(k + 1), \tag{4.39}$$

а рекуррентная [20] следующим образом

$$\begin{aligned}
 \dot{L}_{k+1,c} &= L_{k,c} + \frac{R(k)(R(k+1) - L_{k,c}^T R(k))}{1 + R^T(k)R(k)R(k)} R(k), \\
 \dot{R}(k-1) &= R(k-1) + \frac{R(k-1)R(k-c)R^T(k-c)R(k-1)}{1 - R^T(k-c)R(k-1)R(k-c)}, \\
 \dot{R}(k) &= \dot{R}(k-1) - \frac{\dot{R}(k-1)R(k)R^T(k)\dot{R}(k-1)}{1 + R^T(k)\dot{R}(k-1)R(k)}, \\
 \dot{L}_{0,c} &= 0, R(k) = g^{-1}I_c.
 \end{aligned}$$

Собственно же прогноз вычисляется согласно соотношению (4.38).

Понятно, что применение для прогнозирования соотношений (4.38), (4.39) резко упрощает использование метода, однако остается вопрос обоснованного выбора величины окна c , которое, как правило, задается из сугубо субъективных соображений, что снижает эффективность подхода в целом.

Предлагаемый ниже метод многомерной экстраполяции основан на использовании некоторой близости (расстояния) последнего вектора предыстории $R(N+1)$ ко всем предыдущим данным $R(1), \dots, R(N)$ и формировании прогноза $\hat{B}(N+1)$ с помощью этой же функции.

Реализация метода состоит из последовательности следующих шагов:

1. Расчет расстояния между вектором $R(N+1)$ и всеми предыдущими $R(k)$ на основе некоторой функции близости $d(N+1, k)$ – в простейшем случае евклидовой метрики

$$d(N+1, k) = \|B(N+1) - B(k)\| \quad k.$$

2. Упорядочение (ранжирование) этих расстояний в порядке возрастания так, что

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

3. Отбор первых c векторов, для которых выполняется условие

$$d^c(N+1, k_c) \leq \epsilon,$$

где ϵ – некоторый порог.

4. Формирование набора весовых коэффициентов l_l в виде

$$l_l = \frac{(d^l)^{-1}}{\sum_{l=1}^c (d^l)^{-1}}, 1 \leq l \leq c,$$

отвечающих условию $\sum_{l=1}^c l_l = 1$.

5. Расчет прогноза

$$\hat{B}(N+1) = \sum_{l=1}^c l_l R(l).$$

При поступлении нового наблюдения процесса $B(N+1)$ все итерации повторяются.

Таким образом, на каждом шаге в формировании прогноза также участвуют c наблюдений, однако это значение может изменяться, при этом ясно, что чем меньше c , тем более нестационарен сигнал. Несложно также заметить, что если $B(k) = const$, то $c = N$, $l_l = 1/N$.

Данный метод не требует больших объемов априорной информации (малая обучающая выборка), нет необходимости решения дополнительных задач оптимизации или псевдообращения, крайне прост в вычислительной реализации.

Однако для успешного использования всех традиционных методов прогнозирования, исходная выборка наблюдений – временной ряд – должна

быть достаточно репрезентативной. При этом, чем больше параметров содержит прогнозирующая модель, тем больше по объему должна быть исходная информация. Вместе с тем в реальных задачах достаточно часто возникает ситуация, когда эта выборка либо мала по объему, либо прогнозируемый процесс нестационарен – содержит как нерегулярные тренды, так и внезапные скачки, поэтому его предыстория не может быть использована для нахождения параметров модели.

Кроме предложенного подхода, предлагается использовать процесс, характеризующийся двумерными полями. При этом необходимо использовать какую-либо гипотезу либо о механизме их генерации, либо о наиболее приемлемой математической модели. Например, можно использовать операция линейного преобразования видеоряда b в матрицу z той же размерности вида

$$z = A * b = \left(\overset{v}{\underset{j=1}{\mathring{a}}} A_{1j} b_j \parallel \overset{v}{\underset{j=1}{\mathring{a}}} A_{2j} b_j \parallel \dots \parallel \overset{v}{\underset{j=1}{\mathring{a}}} A_{vj} b_j \right), \quad (4.40)$$

где b_j – j -й столбец $(n' v)$ -матрицы b ,

A_{ij} – $(n' n)$ -матрицы преобразований, подлежащие оцениванию.

При этом преобразование (4.40) содержит $(nv)^2$ неизвестных параметров.

Вводя далее операции циклической перестановки $[b]^{(j)} = (b_j \parallel b_{j+1} \parallel \dots \parallel b_v \parallel b_1 \parallel \dots \parallel b_{j-2} \parallel b_{j-1})$ и трансплантации $\{b\}_{(j)} = (\overset{!}{0} \parallel \overset{!}{0} \parallel \dots \parallel \overset{!}{0} \parallel b_j \parallel \overset{!}{0} \parallel \dots \parallel \overset{!}{0})$, преобразование (4.40) можно переписать в виде матричной свертки

$$z = A * b = \overset{n}{\underset{i=1}{\mathring{a}}} \overset{v}{\underset{j=1}{\mathring{a}}} \left\{ A_{ij} [b]^{(j)} \right\}_{(i)}. \quad (4.41)$$

Использование выражения (4.41) для синтеза прогнозирующей модели неудобно в силу его громоздкости и большого числа оцениваемых параметров, что, в свою очередь, требует больших объемов обучающих выборок для решения задачи параметрической идентификации. Можно, конечно, воспользоваться операциями векторизации и девекторизации, при этом для $(n \times v)$ -матрицы y в результате столбцовой векторизации получаем $(nv \times 1)$ -вектор-столбец $b\Phi = (b_1^T, b_2^T, \dots, b_j^T, \dots, b_v^T)^T$, а в результате строчной векторизации – $(1 \times nv)$ -вектор-строку $b\Phi = (b_1, b_2, \dots, b_i, \dots, b_n)$, где b_j , b_i – j -ый столбец и i -ая строка матрицы b соответственно.

Вводя далее матрицу преобразования в виде:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1v} \\ A_{21} & A_{22} & \dots & A_{2v} \\ \dots & \dots & \dots & \dots \\ A_{v1} & A_{v2} & \dots & A_{vv} \end{pmatrix}$$

(здесь A_{ij} – $(n \times n)$ -матрицы) и произведение

$$Ab\Phi = \left(\left(\sum_{j=1}^v A_{vj} b_j \right)^T, \left(\sum_{j=1}^v A_{vj} b_j \right)^T, \dots, \left(\sum_{j=1}^v A_{vj} b_j \right)^T \right)^T,$$

в результате столбцовой девекторизации получаем:

$$z = (Ab) = \left(\sum_{j=1}^v A_{1j} b_j \quad \sum_{j=1}^v A_{2j} b_j \quad \dots \quad \sum_{j=1}^v A_{vj} b_j \right).$$

Аналогичным образом преобразование типа (4.41) может быть получено и с помощью строчных векторизации-девекторизации.

При анализе видео достаточно часто контролируемые процессы и явле-

ния являются случайными, в связи с чем целесообразно ввести матричные аналоги векторных случайных процессов и последовательностей. Так можно использовать матричную марковскую последовательность

$$b(i) = A * b(i - 1) + x(i), \quad (4.42)$$

где $x(i) = \{x_{ij}(i)\}$ – матричный белый шум.

В принципе, используя операции векторизации-девекторизации, можно переписать (4.42) в форме $b(i) = (Ab(i - 1)) + x(i)$ и для оценки неизвестных параметров матрицы A использовать стандартный метод наименьших квадратов, однако высокая размерность $((nv)^2)$ существенно усложняет его использование. В связи с этим, было предложено использовать конструкцию

$$b(i) = A * b(i - 1) + x(i) = \left(\mathop{\mathbb{A}}_{j=1}^v c_{j1} Ab_j(i - 1) \mathbb{A} \mathop{\mathbb{A}}_{j=1}^v c_{j1} Ab_j(i - 1) \mathbb{A} \dots \right. \\ \left. \mathbb{A} \mathop{\mathbb{A}}_{j=1}^v c_{j1} Ab_j(i - 1) \right) + x(i),$$

соответствующую выражению

$$b(i) = Ab(i - 1)C + x(i), \quad (4.43)$$

где A и C – $(n' n)$ и $(v' v)$ -матрицы преобразования, подлежащие определению.

При этом описание содержит всего $n^2 + v^2 < (nv)^2$ параметров.

В развитие описания (4.43) предлагается ввести матричную модель авторегрессии вида

$$b(i) = \mathop{\mathbb{A}}_{h=1}^{n_A} A^h b(i - h) + x(i) = \mathop{\mathbb{A}}_{h=1}^{n_A} A^h b(i - h) C^h + x(i) = \\ = \mathbb{A}(b(i - 1) \mathbb{A} b(i - 2) \mathbb{A} \dots \mathbb{A} b(i - n_A)) \mathbb{C} + x(i) = \mathbb{A} b(i - 1) \mathbb{C} + x(i),$$

содержащую $(n^2 + \nu^2)n_A$ параметров. Здесь $\overset{\circ}{A} = (A^1 \text{M} A^2 \text{M} \dots \text{M} A^{n_A})$,

$$\overset{\circ}{C} = \begin{pmatrix} \text{C}^1 & \text{L} & \text{D} \\ \text{C}^2 & \text{L} & \text{D} \\ \dots & \dots & \dots \\ \text{C}^{n_A} & \text{L} & \text{D} \end{pmatrix}$$

$$b(i-1) = \begin{pmatrix} b(i-1) & 0 & 0 \\ 0 & b(i-2) & 0 \\ \dots & \dots & \dots \\ 0 & 0 & b(i-n_A) \end{pmatrix}$$

Ясно, что на значение контролируемых показателей в задачах прогнозирования существенным образом влияет множество экзогенных факторов. В связи с этим введем в рассмотрение прогнозирующую матричную модель с экзогенными входами вида:

$$b(i) = \sum_{h=1}^{n_A} \overset{\circ}{a}^h A^h b(i-h) C^h + \sum_{l=1}^{n_K} \overset{\circ}{a}^l K^l x(i-l) D^l + x(i) \quad (4.44)$$

где $x(i-l)$ – $(n' \nu)$ -матрица экзогенных факторов в дискретный момент времени $i-l$, $l=1,2,\dots,n_K$, содержащая $(n^2 + \nu^2)' (n_A + n_K)$ параметров, подлежащих определению.

Вводя далее матрицу

$$\overset{\circ}{A} = (A^1 \text{M} A^2 \text{M} \dots \text{M} A^{n_A} \text{M} K^1 \text{M} \dots \text{M} K^{n_K})$$

размерности $n' n(n_A + n_K)$, матрицу

$$C = \begin{pmatrix} C^1 & \emptyset \\ C & M \\ C & C^{n_A} \\ C & D^1 \\ C & M \\ C & D^{n_K} \\ \emptyset & \emptyset \end{pmatrix}$$

размерности $v(n_A + n_K) \times v$ и

$$b(i) = \begin{pmatrix} b(i-1) \\ b(i-1) \\ b(i-1) \\ b(i-1) \\ b(i-1) \\ b(i-1) \\ b(i-1) \end{pmatrix} = \begin{pmatrix} A(i-1) & \emptyset \\ \emptyset & M \\ \emptyset & C^{n_A} \\ \emptyset & D^1 \\ \emptyset & M \\ \emptyset & D^{n_K} \\ \emptyset & \emptyset \end{pmatrix} \begin{pmatrix} b(i-1) \\ b(i-n_A) \\ x(i-1) \\ 0 \\ x(i-n_K) \end{pmatrix} + \begin{pmatrix} \emptyset \\ \emptyset \\ \emptyset \\ \emptyset \\ \emptyset \\ \emptyset \\ \emptyset \end{pmatrix}$$

размерности $n(n_A + n_K) \times v(n_A + n_K)$, можно переписать (4.44) в компактной форме

$$b(i) = A(i)b(i-1)C + x(i), \tag{4.45}$$

после чего поставить в соответствие описанию (4.45) настраиваемую прогнозирующую модель

$$b(i+1) = A(i)b(i)C(i), \tag{4.46}$$

где $A(i), C(i)$ – матрицы уточняемых оценок, полученных по данным i наблюдений.

Понятно, что в i -й момент времени настраивается не модель (4.46), а ее «сдвинутый» назад аналог

$$b(i) = A(i-1)b(i-1)C(i-1). \tag{4.47}$$

Поскольку в задачах экологического мониторинга контролируемые показатели описываются нестационарными стохастическими или хаотическими

последовательностями, для оценивания параметров модели (4.47) следует использовать адаптивные процедуры идентификации, однако известные алгоритмы не ориентированы на матричные модели, поэтому требуется их серьезная модификация.

Предложен оптимальный градиентный алгоритм идентификации, минимизирующий в on line режиме критерии оценивания

$$\begin{aligned} J^A(i) &= Sp(b(i) - \hat{A}(i-1)b\hat{\Phi}(i-1)\hat{C}(i-1))(b(i) - \hat{A}(i-1)b\hat{\Phi}(i-1)\hat{C}(i-1))^T = \\ &= SpV^A(i)V^{AT}(i) = \|V^A(i)\|^2, \\ J^C(i) &= Sp(b(i) - \hat{A}(i)b\hat{\Phi}(i-1)\hat{C}(i-1))(b(i) - \hat{A}(i)b\hat{\Phi}(i-1)\hat{C}(i-1))^T = \\ &= SpV^C(i)V^{CT}(i) = \|V^C(i)\|^2 \end{aligned}$$

(здесь, $\| \cdot \|$ – евклидова матричная норма) и имеющий вид

$$\begin{aligned} \hat{A}(i) &= \hat{A}(i-1) + h_A(i)V^A(i)\hat{C}^T(i-1)b\hat{\Phi}(i-1), \\ \hat{C}(i) &= \hat{C}(i-1) + h_C(i)b\hat{\Phi}^T(i-1)\hat{A}(i)V^C(i), \end{aligned} \quad (4.48)$$

где $h_A(i)$, $h_C(i)$ – скалярные параметры шага обучения.

На основе процедуры (4.48) можно ввести матричную модификацию адаптивного аддитивно-мультипликативного алгоритма Качмажа вида:

$$\begin{aligned} \hat{A}(i) &= \hat{A}(i-1) + g(SpV^A(i)\hat{C}^T(i-1)b\hat{\Phi}^T(i-1)) \cdot \\ &\cdot b\hat{\Phi}(i-1)\hat{C}(i-1)V^{AT}(i)(\beta + SpV^A(i)\hat{C}^T(i-1)) \cdot \\ &\cdot b\hat{\Phi}^T(i-1)b\hat{\Phi}(i-1)\hat{C}(i-1)\hat{C}^T(i-1)b\hat{\Phi}^T(i-1) \cdot \\ &\cdot b\hat{\Phi}(i-1)\hat{C}(i-1)V^{AT}(i)^{-1}\hat{C}^T(i-1)b\hat{\Phi}^T(i-1), \\ \hat{C}(i) &= \hat{C}(i-1) + g(SpV^C(i)\hat{A}^T(i-1)b\hat{\Phi}(i-1)) \cdot \\ &\cdot b\hat{\Phi}^T(i-1)\hat{A}^T(i)V^C(i)(\beta + Sp\hat{A}(i)b\hat{\Phi}(i-1)) \cdot \\ &\cdot b\hat{\Phi}^T(i-1)\hat{A}^T(i)V^C(i)V^{CT}(i)\hat{A}(i) \cdot \\ &\cdot b\hat{\Phi}(i-1)b\hat{\Phi}^T(i-1)\hat{A}^T(i)^{-1}b\hat{\Phi}^T(i-1)\hat{A}^T(i)V^C(i), \end{aligned} \quad (4.49)$$

где $0 < g < 2$, $b^3 0$ – свободные параметры, выбираемые из эмпирических соображений.

Несложно также заметить, что, если $b(i)$ и $b(i-1)$ – скаляр и вектор соответственно, $g=1$, $b=0$, то (3.20) автоматически превращается в обычный одношаговый алгоритм Качмажа, являющийся одним из самых популярных в теории и практике адаптивной идентификации.

Поскольку в реальных задачах видеообработки исходные данные обычно «зашумлены» интенсивными возмущениями, а сами контролируемые последовательности существенно нестационарны, одношаговые алгоритмы типа (4.49) могут не обеспечивать требуемую точность прогнозирования, поэтому имеет смысл использовать «оконные» процедуры, обладающие сглаживающими свойствами. Поэтому был предложен оптимальный градиентный алгоритм идентификации, минимизирующий в on line режиме критерий оценивания

Введем в рассмотрение матричную модель

$$b(i) = A(i)x(i)$$

где $b(i)$ – исходный видеопоток,

$b(i), x(i) = b(i-1)C, A$ – $(n \times v)$, $n(n_A + n_K) \times v$, $n \times n(n_A + n_K)$ -матрицы, и критерий идентификации

$$J_A(i) = \sum_{l=1}^i Sp(b(l) - A(i)x(l))(b(l) - A(i)x(l))^T,$$

где l – глубина прогноза,

минимизация которого ведет к оценке

$$\hat{A}(i) = p_A(i)R_A^{-1}(i), \quad (4.50)$$

$$\text{где } p_A(i) = \sum_{l=1}^i b(l)x(l)^T,$$

$$R_A(i) = \sum_{l=1}^i x(l)x(l)^T.$$

Несложно организовать пересчет оценок (4.50) на скользящем окне, содержащем s последних наблюдений:

$$\begin{cases} p_A^s(i) = p_A^s(i-1) + b(i)\mathbb{W}^T(i) - b(i-s)\mathbb{W}^T(i-s), \\ R_A^s(i) = R_A^s(i-1) + \mathbb{W}(i)\mathbb{W}^T(i) - \mathbb{W}(i-s)\mathbb{W}^T(i-s), \\ A^s(i) = p_A^s(i)(R_A^s(i))^{-1}. \end{cases} \quad (4.51)$$

Аналогично предыдущему введем также модель

$$b(i) = \mathbb{W}(i)\mathbb{C} + x(i),$$

где $\mathbb{W}(i) = \mathbb{W}(i-1)$, $\mathbb{C} - n \times n(n_A + n_K)$, $n(n_A + n_K) \times n$ -матрицы, и критерий идентификации

$$J^{\mathbb{C}}(i) = \sum_{l=1}^i \text{Sp}(b(l) - \mathbb{W}(l)\mathbb{C}(i))(b(l) - \mathbb{W}(l)\mathbb{C}(i))^T,$$

минимизация которого ведет к оценке

$$\mathbb{C}(i) = R_{\mathbb{C}}^{-1}(i)p_{\mathbb{C}}(i),$$

$$\text{где } R_{\mathbb{C}}(i) = \sum_{l=1}^i \mathbb{W}^T(l)\mathbb{W}(l),$$

$$p_{\mathbb{C}}(i) = \sum_{l=1}^i y^T(l)\mathbb{W}(l).$$

Вводя далее процедуру оценивания на скользящем окне, получаем оценки

$$\begin{cases}
 \dot{P}_{\mathcal{C}}^s(i) = P_{\mathcal{C}}^s(i-1) + b^T(i) \mathcal{X}(i) - b^T(i-s) \mathcal{X}(i-s), \\
 \dot{R}_{\mathcal{C}}^s(i) = R_{\mathcal{C}}^s(i-1) + \mathcal{X}^T(i) \mathcal{X}(i) - \mathcal{X}^T(i-s) \mathcal{X}(i-s), \\
 \dot{C}_{\mathcal{C}}^s(i) = (R_{\mathcal{C}}^s(i))^{-1} P_{\mathcal{C}}^s(i).
 \end{cases} \quad (4.52)$$

Таким образом, алгоритмы оценивания (4.49), (4.51), (4.52) являются обобщением на матричную модель популярных процедур адаптивной идентификации.

Тем самым, можно заключить: рассмотрена задача синтеза матричных прогнозирующих моделей двумерных полей наблюдений что позволяет проводить исследования по выявлению разладок в многомерных временных рядах, иными словами, проводить темпоральную сегментацию видеопоследовательностей. Введены матричные процедуры оценивания параметров этих моделей на основе существующих одномерных оценок, что также позволяет облегчить обработку видеопоследовательностей, т.к. не требует их перевода в одномерное пространство, в результате чего могут появляться существенные потери данных.

Таким образом, в разделе предложены методы временной сегментации видеопоследовательностей на основе анализа многомерных временных рядов, которые позволяют в зависимости от вида входных данных и поставленной задачи выявлять изменения в рядах с разной степенью чувствительности, что в целом определяет инструментарий темпоральной обработки видеопотоков в аспекте их структуризации для CBVIR.

Выводы по четвертому разделу

1. Структуризация видеоданных может осуществляться на основе адаптивного обнаружения изменения свойств многомерных временных рядов в условиях априорной неопределенности на основе настраиваемой модели. Введенная на базе многомерной авторегрессионной модели, параметры кото-

рой настраиваются с помощью специализированных адаптивных алгоритмов с конечной памятью, модификация экспоненциально взвешенной стохастической аппроксимации, обладает фильтрующими и следящими свойствами с автоматическим регулированием параметра сглаживания.

2. Предварительная компрессия индуцированных видеоданными многомерных временных рядов с помощью адаптивной модификации анализа главных компонент позволяет сужать рассмотрение вплоть до одномерных сигналов, что создает предпосылки для использования хорошо апробированных процедур обнаружения сегментов видеопоследовательностей.

3. Модифицированные методы адаптивного прогнозирования на основе многомерной линейной экстраполяции и полиномиальной нейронной сети позволяют решать задачу структурирования видеопоследовательностей в режиме on line на базе учета различия прогноза и реального наблюдения.

4. Вычислительные модели предложенных процедур адаптивного обнаружения изменения свойств многомерных временных рядов отличаются простотой, обладают достаточно высоким быстродействием и позволяют в режиме роста рядов («просмотра» видео) решать задачу темпоральной сегментации на основе их динамических признаков описаний.

РАЗДЕЛ 5

МЕТОДЫ АНАЛИЗА ОПОРНЫХ ДАННЫХ СЕГМЕНТОВ В ЗАДАЧАХ ИНФОРМАЦИОННОГО ПОИСКА

В результате методов сегментации видеоданных, предложенных в разделе 4, получено множество связанных сегментов видеоданных. Одной из возможностей организации поиска сегментов является использование достаточно хорошо разработанного инструментария поиска изображений. С этой целью для решения задачи информационного поиска необходимо выделить опорные данные V_{ab} связанного сегмента, ограниченного кадрами V_a и V_b . Кроме того использование этой парадигмы существенно сокращает время, необходимое для сравнения входных запросов с хранимыми видеоданными для поиска соответствий «по образцу». Вполне логичным выглядит тот факт, что в качестве опорных данных могут быть использованы ключевые кадры. Тогда представление каждого сегмента возможно при помощи одного ключевого кадра или некоторого множества ключевых кадров, автоматически извлекаемых из видео. Одним из вариантов формирования ключевых кадров могут быть разбиения Вороного, представляющие собой разумный компромисс между сложностью обработки и валидностью представления информации в аспекте ликвидации семантического разрыва CBVIR, способы анализа которых и рассматривается в данном разделе.

Рассмотрены процедуры поиска ключевых кадров на основе разбиений Вороного, изучена специфика получения генераторов диаграмм Вороного.

Раздел основан на работах автора [87, 229, 254, 256-259].

5.1 Анализ моделей формирования ключевой информации

В первом разделе были рассмотрены различные подходы по выделению ключевой информации в видео. При этом был сделан вывод о том, что

на текущий момент не существует некоего универсального подхода, что связано со спецификой собственно видеоинформации и способов ее получения.

В связи с этим предлагается сравнение диаграмм Вороного с помощью метрик в рамках исследуемого признакового пространства.

Чтобы формально описать область Вороного, обозначим поле зрения как $D=[a,b] \times [c,d]$, где $a,b,c,d = const$. Пусть $\{p_1, p_2, \dots, p_n\}$ – множество опорных точек. Диаграмма Вороного представляет собой разбиение поля зрения D на выпуклые многоугольники $V = \{v(p_1) \mid D, \dots, v(p_n) \mid D\}$ таким образом, что для каждой области Вороного выполняется следующее неравенство:

$$v(p_i) = \{z \in R^2 : d(z, p_i) \leq d(z, p_j) \forall j \neq i\}$$

где $d(0,0)$ – обычно евклидова метрика.

Иными словами, область Вороного $v(p_i)$, связанная с опорной точкой p_i , представляет собой набор точек Z , расстояние каждой из которых до связанной опорной точки меньше или равно расстоянию до любой другой опорной точки.

Исходя из определения, диаграмма Вороного, лежащая в основе исследуемой процедуры поиска ключевых кадров, строится по опорным точкам. Для определения количества и размещения опорных точек в поле зрения видеокадров предлагается использовать метод Харриса, который принадлежит группе так называемых граничных методов поиска опорных точек, характеризуется относительной простотой и высокой производительностью по сравнению с другими методами и позволяет получить наиболее точный результат, сравнимый с тем, что можно получить при использовании вейвлет-анализа [293].

Метод Харриса инвариантен относительно вращения, изменения освещения и наличия шума на изображениях. Кроме того, качество диаграмм Во-

роного (и как следствие найденных опорных точек) предполагается улучшать путем перекомпоновки методом нечетких J -средних (или других методов кластеризации), что в результате приведет к инвариантности относительно используемого метода поиска опорных точек.

По сути метод Харриса нужен для исходного размещения и определения количества опорных точек, после чего они будут переупорядочены. Изменение интенсивности играет важную роль в поиске опорных точек. Метод Харриса позволяет определить значимость точки в небольшой окрестности в рамках выбранного окна. Как правило, используется окно Гаусса

$w(n) = e^{-\frac{1}{2} \frac{(n - N + 1)^2}{s^2}}$, где N – число измерений, а s – стандартное отклонение распределения Гаусса. В основе метода Харриса лежит автокорреляционная функция, которая позволяет найти точки в кадре со значительным изменением интенсивности по двум направлениям. Наличие двух больших собственных значений свидетельствует о том, что в центре окна находится искомая опорная точка [300].

Предположим, что двумерное изображение, представляющее собой кадр видеопоследовательности, задано функцией интенсивности от его координат $B(x, y)$. Пусть задано окно Гаусса для двух измерений $W = e^{-\frac{x^2 + y^2}{2s^2}}$ со сдвигом на Dx, Dy . Значения каждого пикселя в окне будут средне взвешенными для окрестности. Наибольший вес принимает центральный пиксель (x, y) в окне, веса пикселей окрестности распределяются в зависимости от расстояния до них. Запишем автокорреляционную функцию S в виде суммы квадратов разности для значений текущего окна и окна со сдвигом:

$$S_W(Dx, Dy) = \sum_{x \in W} \sum_{y \in W} (B(x, y) - B(x + Dx, y + Dy))^2. \quad (5.1)$$

Воспользуемся рядом Тейлора первого порядка для аппроксимации значения в окне со сдвигом:

$$B(x + Dx, y + Dy) \approx B(x, y) + \frac{\partial B(x, y)}{\partial x} Dx + \frac{\partial B(x, y)}{\partial y} Dy, \quad (5.2)$$

где $\frac{\partial B(x, y)}{\partial x}$ и $\frac{\partial B(x, y)}{\partial y}$ – частные производные по x и y соответственно [300].

Подстановка (5.2) в (5.1) даст следующее равенство

$$\begin{aligned} S_W(Dx, Dy) &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} (B(x, y) - B(x, y) - \frac{\partial B(x, y)}{\partial x} Dx - \frac{\partial B(x, y)}{\partial y} Dy)^2 dx dy = \\ &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \frac{\partial^2 B(x, y)}{\partial x^2} Dx^2 + \frac{\partial^2 B(x, y)}{\partial y^2} Dy^2 + 2 \frac{\partial^2 B(x, y)}{\partial x \partial y} Dx Dy dx dy. \end{aligned} \quad (5.3)$$

Учитывая, что $u^2 = u^T u$ получим

$$\begin{aligned} S_W(Dx, Dy) &= [Dx, Dy] \begin{bmatrix} \frac{\partial^2 B(x, y)}{\partial x^2} & \frac{\partial^2 B(x, y)}{\partial x \partial y} \\ \frac{\partial^2 B(x, y)}{\partial x \partial y} & \frac{\partial^2 B(x, y)}{\partial y^2} \end{bmatrix} \begin{bmatrix} Dx \\ Dy \end{bmatrix} = \\ &= [Dx, Dy] \begin{bmatrix} \frac{\partial^2 B(x, y)}{\partial x^2} & \frac{\partial^2 B(x, y)}{\partial x \partial y} \\ \frac{\partial^2 B(x, y)}{\partial x \partial y} & \frac{\partial^2 B(x, y)}{\partial y^2} \end{bmatrix} \begin{bmatrix} Dx \\ Dy \end{bmatrix}. \end{aligned} \quad (5.4)$$

Матрицей Харриса $A_W(x, y)$ называется вторая производная S в точке

(x, y) [103]

$$A_W(x, y) = \begin{pmatrix} \frac{\partial^2 B(x, y)}{\partial x^2} & \frac{\partial B(x, y)}{\partial x} \frac{\partial B(x, y)}{\partial y} \\ \frac{\partial B(x, y)}{\partial x} \frac{\partial B(x, y)}{\partial y} & \frac{\partial^2 B(x, y)}{\partial y^2} \end{pmatrix} \quad (5.5)$$

Эта матрица дает представление об интенсивности в локальной окрестности одного пикселя. Основные изменения интенсивности пикселя можно отследить по собственным значениям матрицы Харриса. Возможно 3 варианта (см. рис. 5.1):

– оба собственных значения невелики, тогда изменения интенсивности в локальной окрестности исследуемого пикселя отсутствуют;

– одно собственное значение значительно больше другого, тогда локальная окрестность исследуемого пикселя находится на границе изменения интенсивности, значительные изменения интенсивности изображения наблюдаются при небольшом смещении в перпендикулярном к границе направлении;

– оба собственных значения достаточно большие, тогда небольшое смещение в любом направлении приведет к значительным изменениям интенсивности в локальной окрестности исследуемого пикселя, иными словами – найдена опорная точка [300].

На рис. 5.1 показаны примеры применения метода Харриса для поиска опорных точек в кадрах с разным содержанием.

При этом количество точек было ограничено 50. Несмотря на то, что в общем случае количество опорных точек определяется автоматически, следует дополнительно ограничивать их максимальное значение, так как слишком большое число точек ведет к сложностям в анализе большого количества, полученных по ним областей диаграмм Вороного, и усложняет процедуру расчетов, не улучшая при этом качество результатов.



Рисунок 5.1 – Примеры применения метода Харриса для поиска опорных точек на изображениях с разным содержанием

Эксперименты, проведенные на более, чем 20 образцах тестовых коллекций открытого доступа с низким разрешением и самостоятельно отснятом видеоматериале с более высоким разрешением, показали, что построение диаграмм Вороного по опорным точкам, число которых больше 15, не является целесообразным. Пространственная сегментация видеок кадров разного содержания с учетом до 10 и 20 опорных точек приведена в качестве примера в конце этого подраздела. В любом случае, число объектов в кадре не является столь большим, а отслеживание изменений границ и сопоставление боль-

шего числа опорных точек и связанных с ними областей ведет к появлению ошибок [101].

На рис. 5.1 видно равномерное распределение точек с учетом элементов содержания. Однако один из недостатков метода состоит в том, что когда встречается текст поверх изображения, точки концентрируются в большей степени в области текстовых элементов. Это объясняется максимальным изменением интенсивности именно в таких областях. Использование различных модификаций метода Харриса позволяет устранить этот недостаток, однако это не существенно в данном случае, так как перед поиском ключевых кадров найденные характерные точки предполагается перераспределить (с соответствующей перекомпоновкой диаграмм Вороного) с учетом кластерного анализа цвето-текстурных признаков для всех видеок кадров, чтобы они в большей степени соответствовали содержанию видеок кадров в целом, а не локальным выбросам интенсивности.

Принято считать, что кластеризация может быть применена для поиска оптимального разбиения [50], поэтому для улучшения размещения опорных точек можно использовать методы кластеризации, описанные в разделе 2.1. Кроме того, предлагается использовать метод нечетких J -средних.

Реализация FJM метода производится в два этапа: нахождение локальных оптимумов с помощью стандартного FCM и нахождение более «глубоких» минимумов с помощью FJM-эвристики.

На первом этапе реализуется последовательность шагов.

1. Задание начального достаточно произвольного разбиения $P_p = \{Cl_1, Cl_2, \dots, Cl_p\}$ с центроидами $C(1), C(2), \dots, C(p)$; фаззификатора b и порогового значения $e > 0$, определяющего условия останова алгоритма.

2. Расчет уровней принадлежности $m(x(k), C(l))$ с центроидами, полученными на предыдущем шаге.

3. Перерасчет центроидов $C(1), C(2), \dots, C(p)$ с уровнями принадлежности, полученными на предыдущем шаге.

4. Оценка сферической нормы разности между ранее полученными

центроидами и центроидами, рассчитанными на третьем шаге.

5. Проверка условий останова. Если полученная норма меньше ϵ – окончание работы алгоритма; если же полученная норма больше ϵ – возврат к шагу 2 с центроидами, полученными на шаге 3.

Итерации продолжаются до выполнения условия останова, а полученное решение есть координаты локального оптимума задачи нелинейного программирования.

Второй этап – это фаза скачков, когда из полученного локального минимума совершаются случайные движения в его окрестности с целью отыскать более глубокий экстремум.

Было показано, что общая задача оптимизации может быть приведена к задаче безусловной оптимизации специального вида целевой функции, которая в матричном случае может быть записана в виде

$$E = E(x(k), C(l)) = \mathring{\mathbf{a}} \left(\mathring{\mathbf{a}} \left(\sum_{k=1}^{\mathring{N}} \sum_{l=1}^p (Sp(x(k) - C(l))(x(k) - C(l))^T)^{1-b} \right)^{1-b} \right), \quad (5.6)$$

для произвольных значений фаззификатора b .

Далее из любого из полученных центроидов $C(l)$, $l = 1, 2, \dots, p$ совершаются скачки, состоящие в том, что выбранный центроид заменяется любым наблюдением $x(r)$, после чего рассчитывается значение целевой функции (5.6) в форме

$$E = E(x(k), x(r)) = \mathring{\mathbf{a}} \left(\mathring{\mathbf{a}} \left(\sum_{k=1}^{\mathring{N}} \sum_{l=1}^p (Sp(x(k) - x(r))(x(k) - x(r))^T)^{1-b} \right)^{1-b} \right) \quad (5.7)$$

Если окажется, что для некоторого $x(r)$ значения (5.7) окажутся меньшими чем (5.6), принимается решение о том, что найден новый улучшенный центроид кластера Cl_l , после чего производится перерасчет всех уровней.

Такие скачки совершаются в окрестности каждого из центроидов, полученных на первом этапе. Если окажется, что скачки в окрестностях всех центроидов не привели к улучшению значения целевой функции (5.7), то либо принимается решение об окончании процесса оптимизации, либо производятся скачки в окрестности увеличенного радиуса. В принципе такой процесс может продолжаться до исчерпания всех незанятых точек. Найденный последним локальный экстремум полагается глобальным.

На рис. 5.2 показана предложенная схема переупорядочения опорных точек, где изображение под номером 1 является исходным видеокадром; 2 – выбор опорных точек методом Харриса; 3 – первоначальное разбиение на кластеры по опорным точкам, отобранным методом Харриса; 4 – назначение

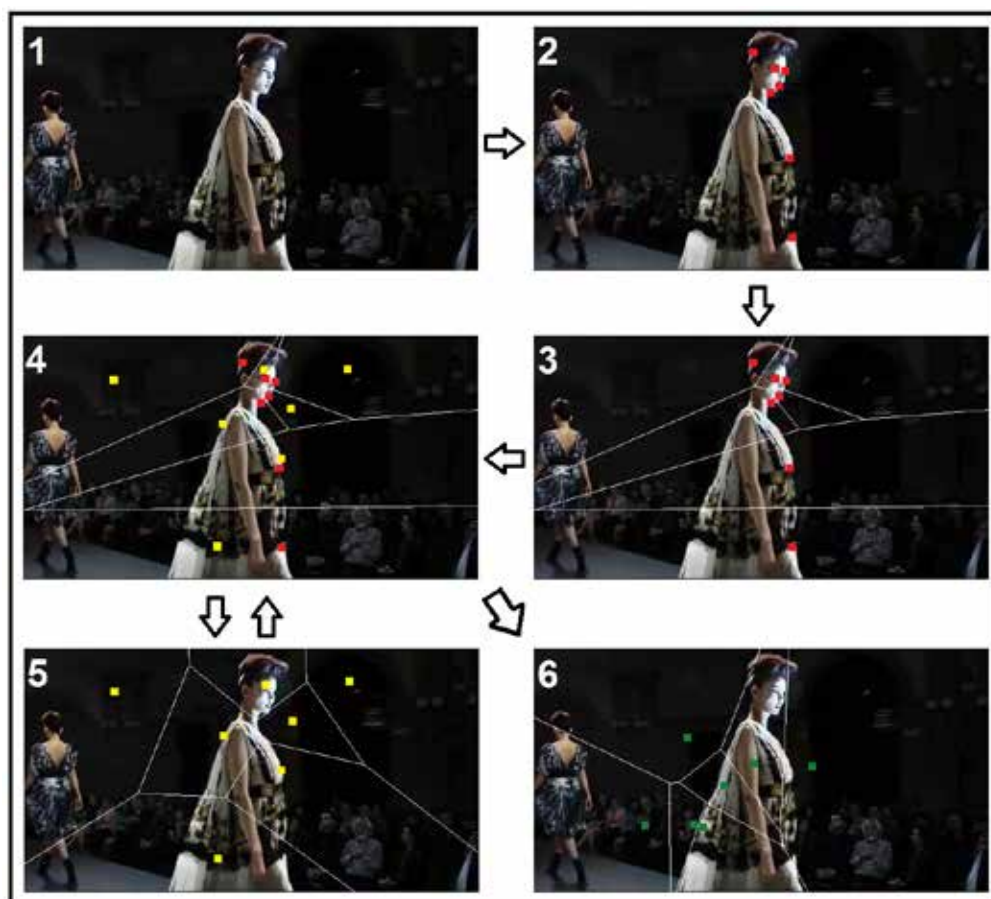


Рисунок 5.2 – Предложенная схема переупорядочения опорных точек

новых центроид; 5 – переупорядочение кластеров и возврат к этапу 4 до тех пор, пока центроиды не перестанут смещаться относительно опорных точек и

не будет получено результирующее разбиение 6. Красным маркером обозначены исходные опорные точки, размещенные с помощью метода Харриса, желтым – центроиды и опорные точки, полученные на промежуточных этапах кластеризации FJM с учетом цвета и текстуры в областях, зеленым – результирующие опорные точки, на базе которых формируется разбиение видеокадра.

В качестве вектора признаков для кластерного анализа использовалась интенсивность как цветовой признак и энтропия как текстурный признак. В качестве меры близости использовался квадрат Евклидова расстояния, чтобы придать больший вес удаленным друг от друга объектам.

Применение метода Харриса позволяет достичь большей концентрации областей в зоне объекта, тогда как после применения алгоритма FJM области диаграммы Вороного становятся более равномерно распределенными по всему кадру. Однако это не означает, что после переупорядочения опорных точек детали будут утеряны.

На рис. 5.3 показана сегментация кадров разного содержания с учетом до 10 и 20 опорных точек, переупорядоченных с помощью алгоритма FJM. Очевидна слишком высокая концентрация областей, которые нужно анализировать при обнаружении до 20 опорных точек, что отмечалось ранее. При этом контуры областей, занимающих основную площадь кадра, и их размещение сравнимы при разном количестве опорных точек.

На рис. 5.4 показаны диаграммы Вороного для нескольких кадров с похожим содержанием. Очевидно улучшение качества (с точки зрения соответствия значимым элементам содержания) диаграмм справа после перекомпоновки методом FJM. Слева показаны диаграммы Вороного, полученные при размещении опорных точек методом Харриса.

Для кадров с похожим содержанием диаграммы Вороного практически не отличаются друг от друга (см. рис. 5.4). На основании этого сходства и выполняется процедура отбора ключевых кадров, обеспечивая извлечение уникального содержания и отсутствие повторов.



Рисунок 5.3 – Сегментация кадров разного содержания,
выполненная с учетом до 10 и 20 опорных точек

Применение алгоритма FJM увеличивает время обработки, однако данная мера полностью оправдана качеством сегментации. Кроме того, новизна предложенной процедуры состоит в том, что результат после применения алгоритма FJM не зависит от выбора исходных опорных точек. Иными словами, задача их оптимального выбора не стоит. Независимо от расположения

исходных точек, кластеризация приведет к одному результату размещения для данного изображения. Будет отличаться только количество итераций, приводящих к одному результату.

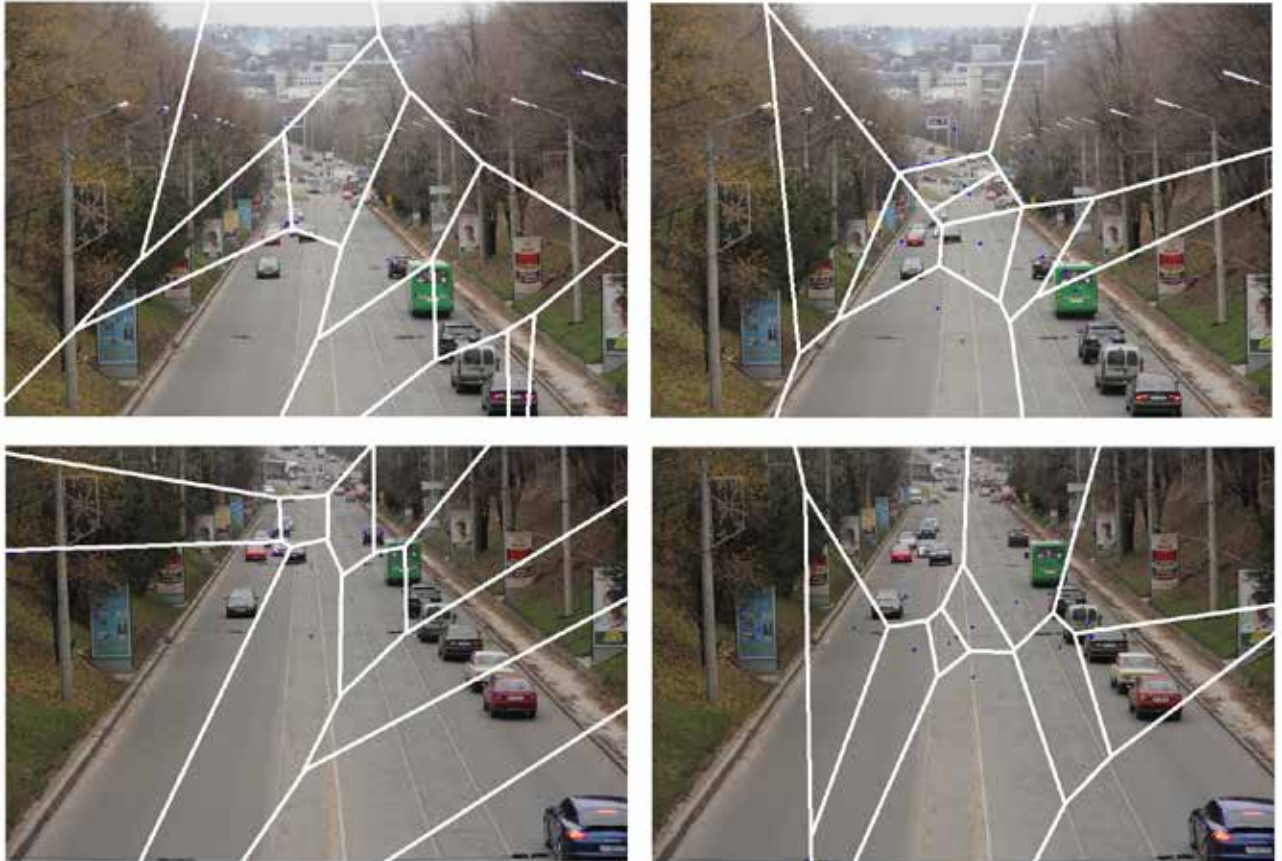


Рисунок 5.4 – Диаграммы Вороного для двух кадров с похожим содержанием

Диаграммы Вороного более высоких порядков изучались многими исследователями, начиная с 1970 годов. Среди них наиболее известны работы Майлса (Miles), Шамоса (Shamos), Ауэрхаммера (Aurenhammer), Агарвала (Agarwal) и др. Термин «порядок» для диаграмм Вороного означает количество точек, которые формируют область Вороного, а «более высокий» или порядок k предполагает наличие более одной точки (в отличие от обычных диаграмм Вороного, где одна точка формирует единственную область Вороного). Здесь высокий порядок не имеет отношения к размерности пространства. Поиск ключевых кадров осуществляется только в двумерном пространстве, так как все кадры расположены на плоскости XU .

Уточнение (детализация) в отдельных областях кадра может выполняться путем увеличения количества исходных опорных точек или путем построения диаграмм Вороного более высоких порядков (или, как их еще называют, обобщенных диаграмм Вороного). Экспериментально подтверждено, что кластеризация большого числа точек особенно на изображениях большого размера с высоким разрешением выполняется слишком долго. Таким образом, для одного и того же варианта размещения опорных точек можно построить разные диаграммы (с разным количеством областей), как показано на рис. 5.5, что сэкономит время примерно в полтора раза по сравнению с добавлением новых опорных точек и перестроением диаграммы первого порядка [101].

На рис. 5.5 показано три видеокадра с движением автомобилей по проезжей части, при этом каждый из трех кадров сегментирован с помощью диаграмм Вороного первого, второго и восьмого порядка (слева направо). Исходная обработка этих кадров методом Харриса позволила обнаружить 9 опорных точек, поэтому восьмой порядок является максимальным в данном случае.

Увеличение числа сегментов при использовании диаграмм второго порядка очевидно, равно как и сокращение числа сегментов при использовании диаграмм восьмого порядка. Из рис. 5.5 видно, что наиболее устойчивыми к изменению содержания являются диаграммы Вороного наивысшего порядка. Приведенные на рисунке кадры взяты из одной сцены, где диаграмма наивысшего порядка остается практически неизменной. Это легко объяснить исходя из определения диаграмм Вороного высших порядков.

Диаграмма Вороного k -го порядка, $V^{(k)}$, построенная на базе n опорных точек в двумерном пространстве, представляет собой разбиение плоскости на выпуклые многоугольники, такое, что точки z каждой области Воро-

ного $v(p_i)^{(k)}$ имеют одинаковое число ближайших опорных точек p_i , которое равно k . Ранее рассмотренные обычные диаграммы Вороного являются частным случаем диаграмм Вороного k -го порядка при $k = 1$.

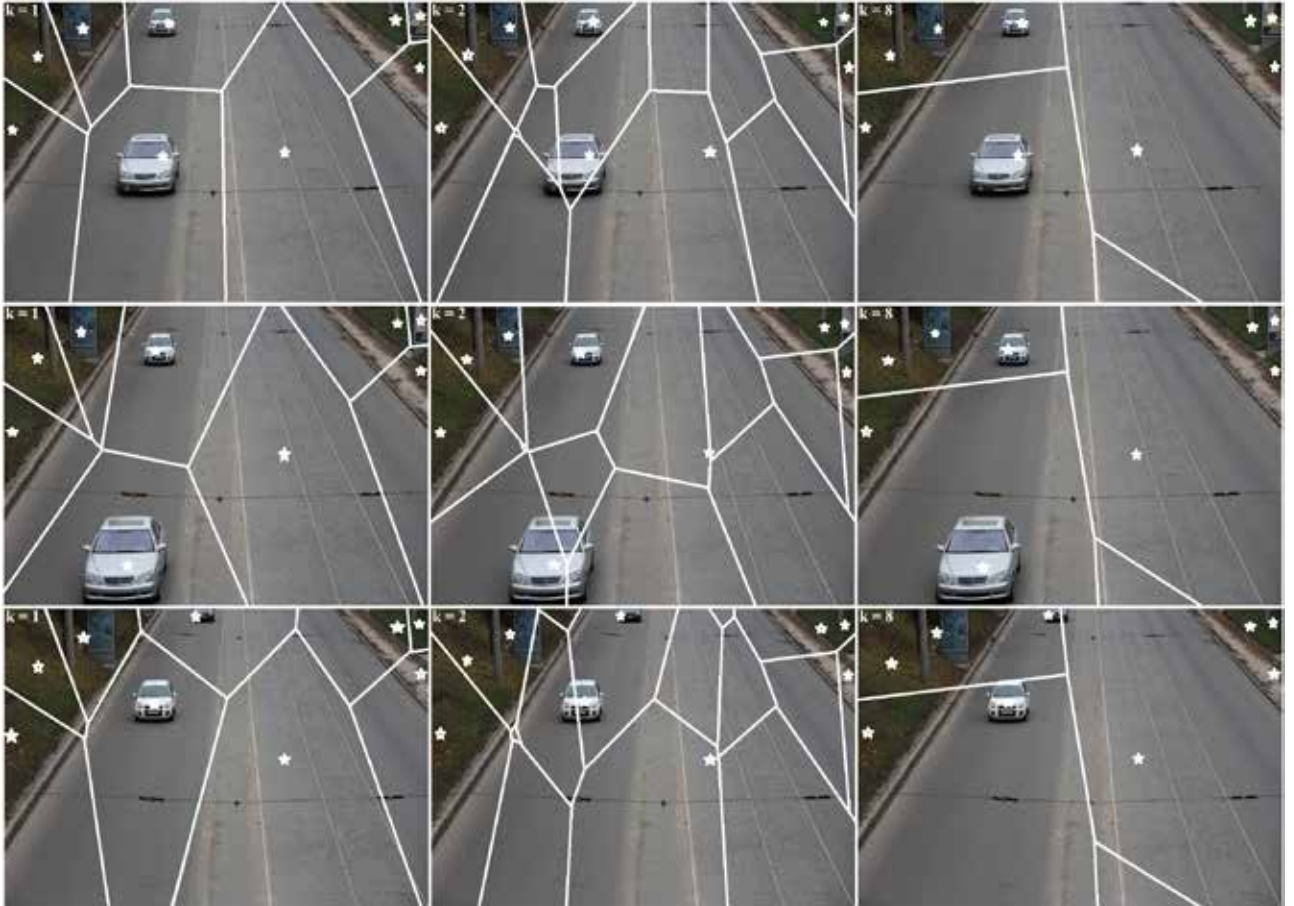


Рисунок 5.5 – Использование диаграмм Вороного разных порядков в пространственной сегментации видеокадров

Для формального представления областей Вороного k -го порядка предположим, что $\{p_1, p_2, \dots, p_m\}$ – множество опорных точек, а $\{\{p_{1,1}, \dots, p_{1,k}\}, \dots, \{p_{l,1}, \dots, p_{l,k}\}\}$ – множество, состоящее из подмножеств с k ближайшими опорными точками, тогда выпуклый многоугольник Вороного $v(p_i)^{(k)}$ порядка k , формируемый набором опорных точек $\{p_{i,1}, \dots, p_{i,k}\}$, может быть описан следующим образом:

$$v(p_i)^{(k)} = \{z \in \mathbb{R}^2 : \max\{d(z, p_{i,h}), p_{i,h} \in v(p_i)^{(k)}\} \leq \min\{d(z, p_{i,j}), p_{i,j} \in V^{(k)} \setminus v(p_i)^{(k)}\}\}. \quad (5.8)$$

Иными словами, расстояние максимально удаленной точки одной области до связанных с ней опорных точек меньше или равно расстоянию до ближайшей опорной точки любой другой области.

Экспериментально было подтверждено, что на некотором этапе, при постоянном увеличении порядка, наблюдается обратное явление – детализация уменьшается. Причем уменьшение детализации происходит не только при максимально допустимом значении k -го порядка (когда $k = n - 1$), но гораздо раньше.

При $k = n - 1$, то есть когда порядок максимально допустим и равен общему количеству опорных точек за исключением одной, диаграмма разделяет наиболее удаленные точки (furthest-site Voronoi diagram). Важно отметить, что для разного количества опорных точек получается разное пороговое значение уменьшения детализации. В таблице 5.1 приведены пороговые значения с шагом в 5 точек в интервале от 5 до 35 опорных точек [258]. Хотя, как упоминалось ранее, построение диаграмм Вороного по опорным точкам, число которых больше 15, не целесообразно.

На рис. 5.6 показан усредненный график зависимости количества областей Вороного от порядка диаграммы для 15 опорных точек. Хотя количество областей Вороного зависит от расположения опорных точек на плоскости, эксперименты на тестовой коллекции TRECVID доказали, что количество областей Вороного отличается в пределах 20-30 % для диаграмм одного порядка с одинаковым числом опорных точек [258].

Следует отметить, что уменьшение детализации разбиений зачастую происходит раньше порогового значения, однако при достижении порогового значения, точек становится меньше не только по сравнению с предыдущим порядком, но по сравнению с диаграммами любого порядка, в том числе и

Таблица 5.1 – Пороговые значения уменьшения детализации разбиений Вороного

Количество опорных точек	Порядок k , при котором достигается пороговое значение уменьшения детализации
5	4
10	9
15	13
20	18
25	22
30	28
35	32

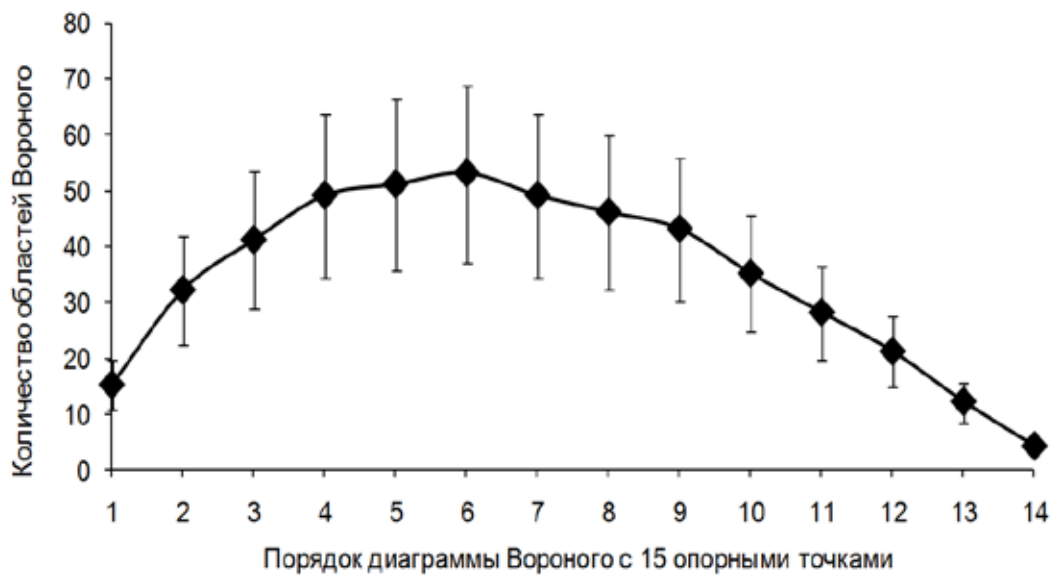


Рисунок 5.6 – Зависимость количества областей Вороного от порядка диаграммы для 15 опорных точек

первого. В любом случае изменение детализации происходит по параболе – сначала детализация увеличивается, после чего плавно уменьшается, достигая своего порогового значения ближе к последним порядкам. Таким образом, диаграмма Вороного k -го порядка является хорошим инструментом для

детализации содержания изображений, которые предполагается сравнивать между собой в дальнейшем [258].

Полученные области Вороного любого порядка не предназначены для огрубленного очертания границ объекта, поэтому нахождение объекта в той или иной области лишь свидетельствует о том, что распределение пикселей кадра с точки зрения цвета и текстуры является наиболее стабильным в данном представлении при условии исходного местоположения опорных точек и их количества, определяемых методом Харриса [101].

5.2 Сравнение кадров, сегментированных с помощью диаграмм Вороного

Для сравнения видеок кадров, их отдельных сегментов или наборов признаков принято использовать различные метрики или расстояния, показывающие близость объектов исследования рассмотренную в разделе 3. С точки зрения терминологии, “расстояние” и “мера близости” или “сходства” являются обратными друг к другу понятиями. Можно говорить, что расстояние между двумя множествами – это различие между ними [261, 262]. Напомним, что метрикой является любая неотрицательная функция, удовлетворяющая требованиям рефлексивности, симметричности и неравенства треугольника

$$\begin{aligned}
 1) \ r(B'(z), B(z)) = 0 \ \hat{=} \ B'(z) = B(z), \\
 2) \ r(B'(z), B(z)) = r(B(z), B'(z)), \\
 3) \ r(B'(z), B(z)) \leq r(B(z), B(z)) + r(B(z), B(z)),
 \end{aligned}
 \tag{5.9}$$

Как правило, сравнение кадров в известных поисковых системах выполняется с помощью широко известных Манхэттенского расстояния, расстояния Хаусдорфа, Махаланобиса, Евклидова расстояния, иногда различных мер близости или специализированных метрик, разработанных авторами

методов поиска ключевых кадров. Применяются линейные комбинации (иногда называемые агрегацией) метрик или мер, а иногда полученным с их помощью значениям назначаются весовые коэффициенты. Использование статистических мер не дает высококачественных результатов, однако они могут использоваться для обработки данных в реальном времени или для решения узкоспециализированных задач [8].

Чтобы получить возможность сравнения содержания кадров, представленного с помощью диаграмм Вороного, необходимо разработать принципиально новую метрику, так как ранее учеными не ставилась задача сравнения разбиений Вороного. Как упоминалось ранее, это позволит не только максимально отразить содержание видеопоследовательности, но и стабилизировать представление кадров.

Попытку сравнения разбиений Вороного, но не с помощью метрик сходства, предпринял японский ученый Yukio Sadahiro. Он предложил использовать различные методы визуального и количественного анализа, включая критерий c^2 , индекс Каппа и их дополнения, площадь и периметр разбиений Вороного, дисперсию и стандартное отклонение, центры масс и др. Вклад Yukio Sadahiro заключался в том, что для сравнения различных систем разбиений на административные области Японии он предложил использовать меру плотности детализации и иерархические отношения (наложение, частичное наложение и включение). Однако, площадные методы иногда имеют двойственное значение для приложений обработки видео, так как объекты могут быть отсняты с разным масштабом. Разные объекты видеок кадров могут быть одного размера. Таким образом, нельзя отслеживать объекты видео основываясь только на свойствах их площади. В нашем случае нужно учитывать различные признаки: пространственные, текстурные, цветовые, которые являются основными атрибутами, используемыми в системах СВIR и СВVIR.

Для того, чтобы формально описать процедуру сравнения видеок кадров, содержание которых представлено с помощью разбиений Вороного, опреде-

лим k -ый ключевой кадр видеопоследовательности F как $B_k(z), z = (x, y) \in D$ (здесь и далее $k = 1, 2, \dots, K$ – дискретное время). Учитывая, что $1 \leq i < j \leq K$ и $B_i(z), B_j(z) \in F$, обозначим сегмент как $S_l(i, j) = [B_i(z), B_j(z)], l = 1, 2, \dots, i, j \in L_l, \# L_l = K$. Сцена представляет собой набор последовательно идущих кадров, полученных после временного разбиения на значимые сегменты таким образом, что выполняется следующее: $\bigcup_{l \in L} S_l(i, j) \in \mathcal{A}, F = \bigcup_{l \in L} S_l(i, j), \forall l', l'' \in L, S_{l'}(i, j) \cap S_{l''}(i, j) = \emptyset$. Для фиксированного l обозначим ключевой кадр как изображение $B_r^*(z) \in S_l(i, j)$, для которого характерно следующее свойство:

$$r = \arg \min_{r \in L_l} \int_{L_l} \rho(B_r(z), B_t(z)). \quad (5.10)$$

После извлечения всех ключевых кадров будет получен набор кадров $\{B_l^*(z)\}$ для видеопотока F . Иными словами, для каждой сцены будет извлечен один кадр (или несколько), при этом каждый ключевой кадр является репрезентативным для своей сцены (или ее части).

Чтобы сравнить разбиения Вороного двух кадров $B'(z)$ и $B''(z)$ с набором опорных точек $\{p'_1, p'_2, \dots, p'_n\}$ и $\{p''_1, p''_2, \dots, p''_m\}$ соответственно, воспользуемся метрикой на разбиениях, рассмотренной в разделе 3, которая в интерпретации к данной задаче будет иметь вид $\rho_1(V', V'')$:

$$\rho_1(V', V'') = \sum_{i=1}^n \sum_{j=1}^m |v(p'_i) \Delta v(p''_j)| |v(p'_i)| |v(p''_j)| \quad (5.11)$$

где $v(p'_i) \Delta v(p''_j) = (v(p'_i) \setminus v(p''_j)) \cup (v(p''_j) \setminus v(p'_i))$.

Значение, полученное из симметрической разности, не зависит от размера областей, то есть две большие области с тремя точками несоответствия (отличия между ними), окажутся точно такими же по расстоянию, как и две

небольшие области, также с тремя отличающимися точками. Умножение симметрической разности на пересечение областей позволяет получить большой разброс значений для разных и для похожих кадров. Использование только симметрической разности не дает такого большого разброса значений, что видно из примера следующих двух графиков. На рис. 5.7 показан график значений симметрической разности областей Вороного для последовательно идущих кадров фрагмента новостного видео “factories_512kb.mp4” из тестовой коллекции открытого доступа Internet Archive.

На рис. 5.8 показан график для этого же фрагмента видео, но для его построения использовались значения симметрической разности умноженной на пересечение областей Вороного.

Диапазон значений, полученных в результате умножения симметрической разности на пересечение областей Вороного для рассмотренного новостного видео, равен $2,3 \cdot 10^9$, тогда как для симметрической разности значения варьируют от $148 \cdot 10^3$ до $289 \cdot 10^3$, что примерно в 10^5 раз меньше. Кроме того, изменения на графике, полученные

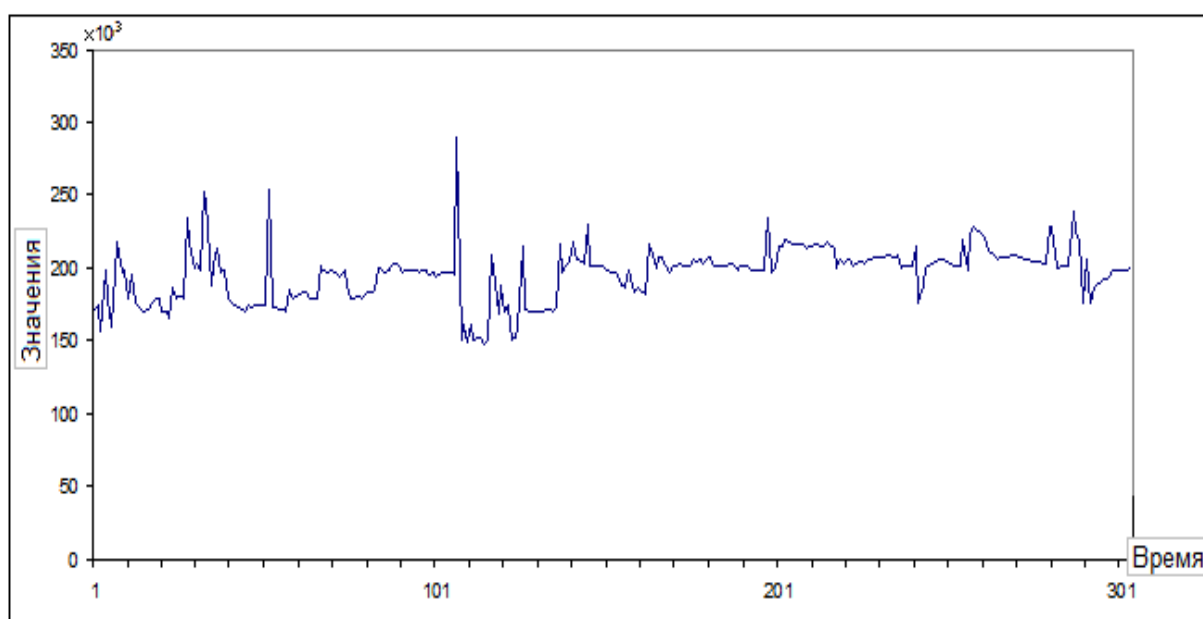


Рисунок 5.7 – Сравнение формы областей Вороного с помощью симметрической разности

при расчете симметрической разности, меньше соответствуют изменениям сюжета видеоматериалов, чем расчеты, выполненные в соответствии с предложенной формулой 5.11. Поэтому только по симметрической разности

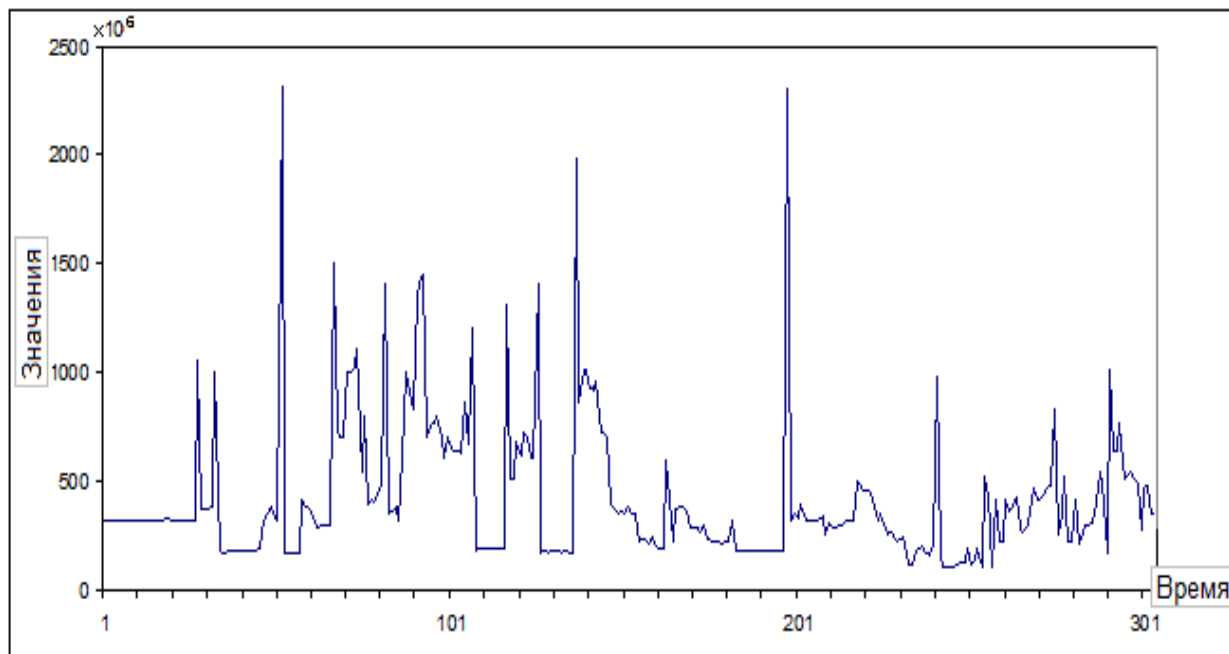


Рисунок 5.8 – Сравнение формы с помощью симметрической разности умноженной на пересечение областей Вороного

сложно судить об изменениях формы областей при значительном изменении сюжета.

При использовании пересечения областей, для разных по форме областей значение симметрической разности значительно выше, в то время как для одинаковых разбиений за счет меньшего количества пересечений разных областей большинство множителей принимают значение, равное 0, что обуславливает большой разброс значений [87].

Для дальнейших исследований важно определить следующие положения. Исходя из того, что области Вороного строятся по серединным перпендикулярам между смежными опорными точками p_i и p_l [147], запишем область Вороного $v(p_i)$ как пересечение полуплоскостей, проходящих через смежные опорные точки:

$$v(p_i) = \frac{1}{|\hat{\Gamma}[1; y]|} H(p_i, p_l), \quad (5.12)$$

где y – количество смежных с p_i опорных точек p_l .

Для смежных опорных точек p_i и p_l выполняется следующее:

1) $\$ \mathfrak{g}_y, d(p_i, \mathfrak{g}_y) = d(p_l, \mathfrak{g}_y)$, где \mathfrak{g}_y – вершина многоугольника Вороного;

2) $(\frac{x_l - x_i}{2}; \frac{y_l - y_i}{2}) \hat{\Gamma} v(p_i), v(p_l)$, где $p_i(x_i; y_i), p_l(x_l; y_l)$;

3) $v(p_i) | v(p_l) \in \mathfrak{AE}$, где $v(p_i) | v(p_l)$ – ребро многоугольника Вороного или его вершина.

Однако опорные точки являются смежными только в том случае, когда соответствующие им области имеют невырожденную границу (ребро области Вороного, то есть более одной точки). Наглядный пример, иллюстрирующий, когда в результате пересечения прямых, содержащих серединный перпендикуляр, получается точка и ребро можно увидеть на рис. 5.9.

Чтобы представить сходство кадров непосредственно через опорные точки достаточно записать уравнения для серединных перпендикуляров между смежными опорными точками p_i и p_l одного кадра $B'(z)$ и смежными опорными точками p_i и p_l другого кадра $B''(z)$ [87].

Уравнение прямой, проходящей через две точки с координатами $p_i(x_i; y_i)$ и $p_l(x_l; y_l)$, выглядит следующим образом:

$$\frac{y - y_i}{y_l - y_i} = \frac{x - x_i}{x_l - x_i} \text{ или } (y_l - y_i)x + (x_l - x_i)y + (x_i y_l - x_l y_i) = 0. \quad (5.13)$$

Из формулы (5.13) можно определить угол наклона α для прямой, проходящей через смежные опорные точки, $\text{tg}\alpha = -\frac{y_l - y_i}{x_l - x_i}$. Исходя из рис. 5.10,

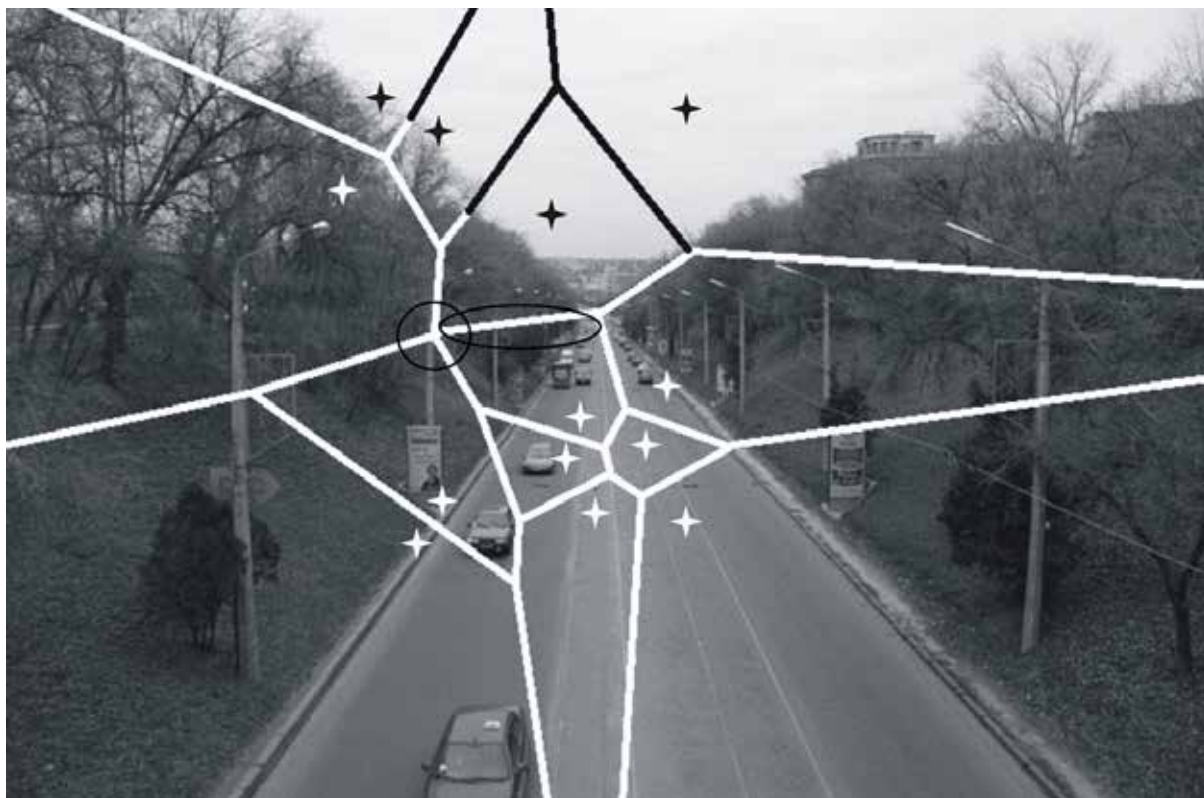


Рисунок 5.9 – Точка и ребро, образуемые в результате пересечения прямых, содержащих серединный перпендикуляр между смежными опорными точками

$\operatorname{tg} j = \operatorname{tg}(q + \frac{p}{2})$, поэтому угол наклона j для серединного перпендикуляра

можно выразить как $\operatorname{tg} j = -\operatorname{ctg} q$. Таким образом, $\operatorname{tg} j = \frac{x_l - x_i}{y_i - y_l}$.

Зная угол наклона j и координаты точки $(\frac{x_l - x_i}{2}; \frac{y_l - y_i}{2})$, которая лежит на прямой, образующей полуплоскость, несложно определить уравнение прямой, содержащей серединный перпендикуляр между смежными опорными точками. С учетом того, что координаты точки должны удовлетворять

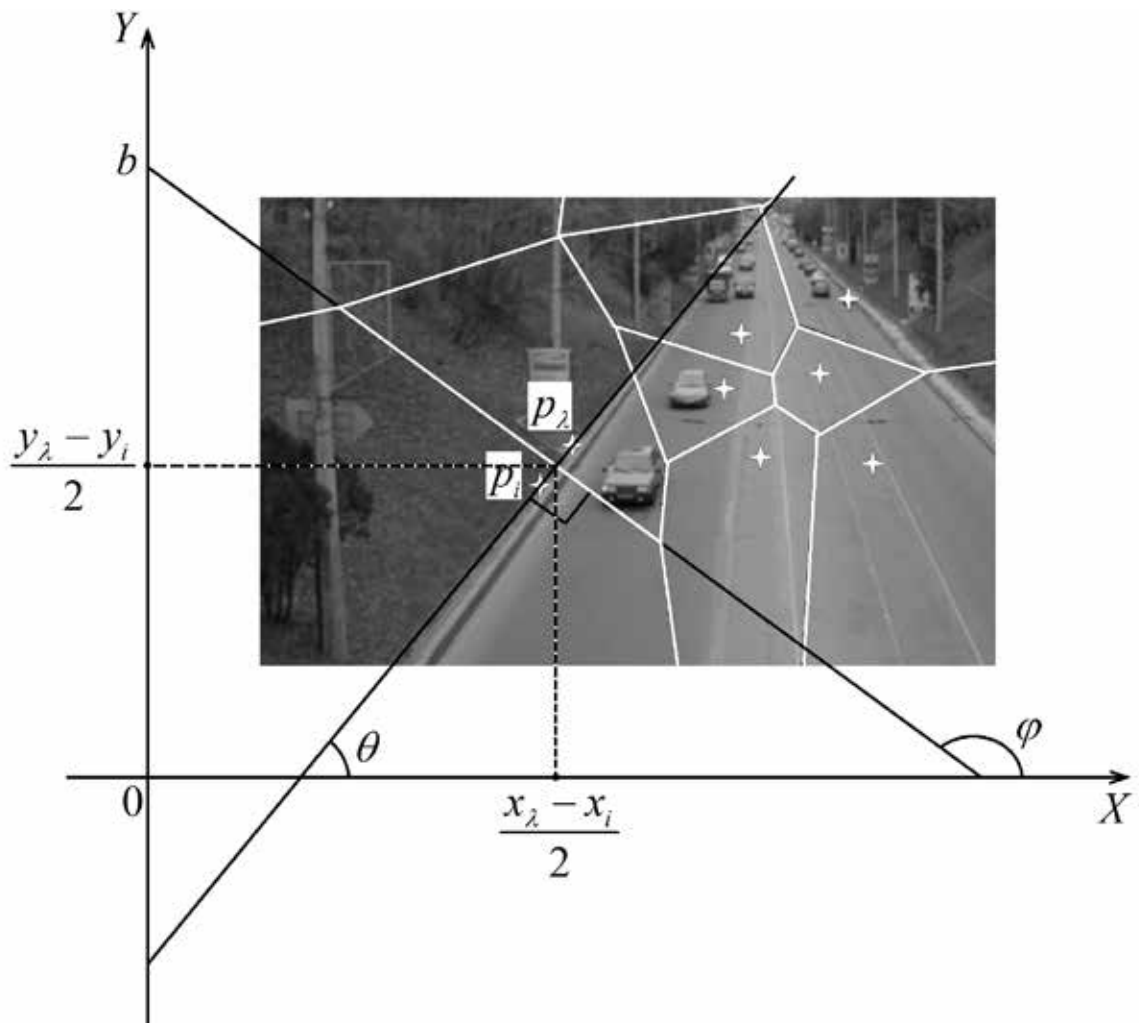


Рисунок 5.10 – Угол наклона прямой, которая содержит серединный перпендикуляр между смежными опорными точками, на примере разбиения Вороного для одного из видеок кадров

уравнению прямой с угловым коэффициентом ($y = kx + b$, $k = \text{tg}j$), определим точку пересечения серединного перпендикуляра с осью ординат:

$$b = \frac{y_l - y_i}{2} - \text{tg}j \cdot \frac{x_l - x_i}{2}, \text{ т.е. } b = \frac{y_l - y_i}{2} - \frac{x_l - x_i}{y_l - y_i} \cdot \frac{x_l - x_i}{2}.$$

Запишем уравнение прямой, содержащей серединный перпендикуляр между смежными опорными точками

$$y = \frac{x_1 - x_i}{y_i - y_1} x + \frac{y_1 - y_i}{2} - \frac{x_1 - x_i}{y_i - y_1} \cdot \frac{x_1 - x_i}{2}$$

или

$$y = \frac{x_1 - x_i}{y_i - y_1} \left(x - \frac{x_1 - x_i}{2} \right) + \frac{y_1 - y_i}{2}. \quad (5.14)$$

Пересечение прямых, содержащих серединные перпендикуляры между смежными опорными точками, образуют области Вороного в виде многоугольников. Чтобы найти координаты вершин многоугольников достаточно определить координаты точек пересечения таких прямых

$$\left\{ \begin{array}{l} \frac{x_{i+1} - x_i}{y_{i+1} - y_i} \left(x - \frac{x_{i+1} - x_i}{2} \right) + \frac{y_i - y_{i+1}}{2} \\ \frac{x_j - x_k}{y_j - y_k} \left(x - \frac{x_j - x_k}{2} \right) + \frac{y_k - y_j}{2} \end{array} \right.$$

Зная координаты опорных точек $(x_i; y_i)$ и координаты вершин многоугольников $(x_{g_y}; y_{g_y})$ можно с помощью формулы Герона определить площадь каждого многоугольника, разбив их на треугольники с вершиной в опорной точке и основанием в виде ребер многоугольников (сторона между двумя вершинами с известными координатами) или отрезка поля зрения, ограничивающего кадр, координаты которого также известны. При этом, если опорная точка находится внутри треугольника, основанием которого является отрезок поля зрения, то площадь такой области рассчитывается с учетом вершины и отрезка в основании, без учета самой опорной точки (см. рис. 5.11) [87]. Таким образом, можно рассчитать площадь каждой области одного кадра $A(v(p'_i))$ и другого $A(v(p''_i))$, чтобы сравнить кадры впоследствии.

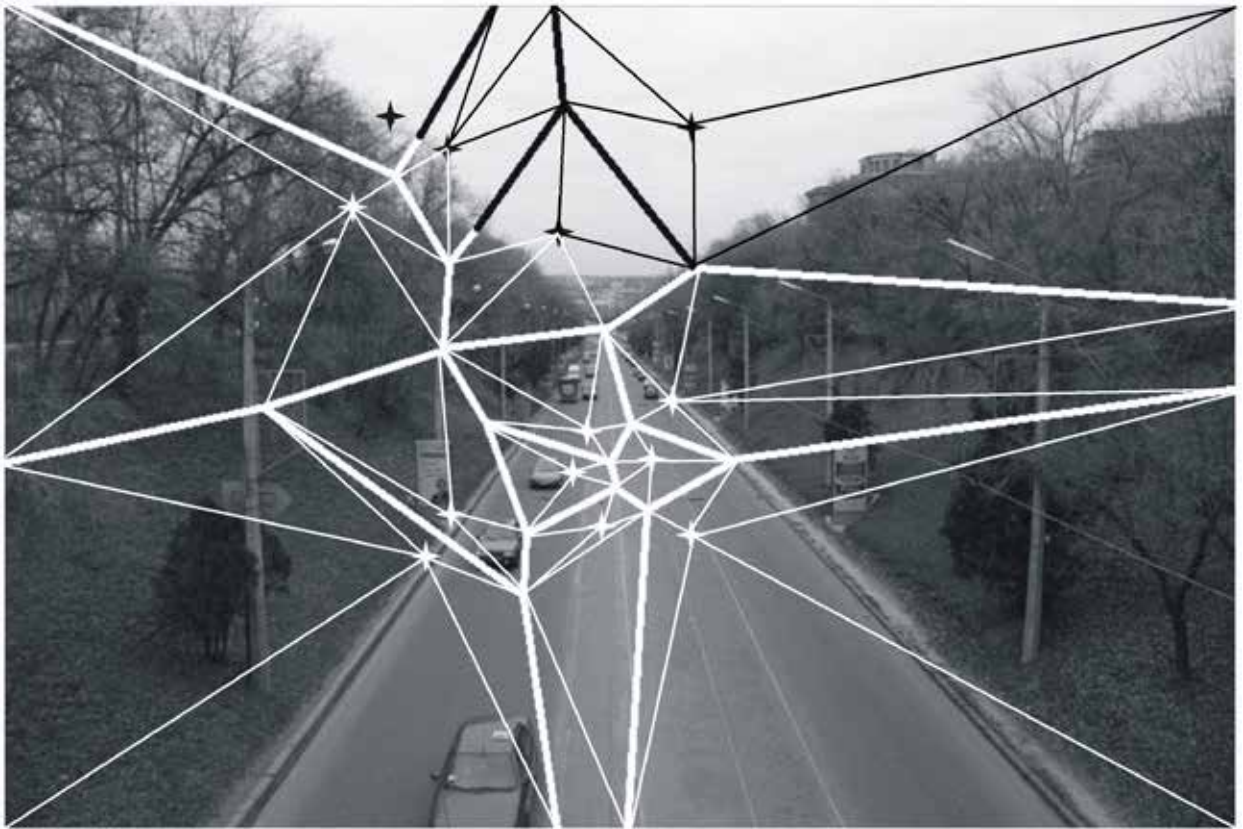


Рисунок 5.11 – Разбиение многоугольников Вороного на треугольники для определения их площади

При этом гранулярность разбиений кадра $B'(z)$ вычисляется как отношение множества точек каждой области ко всем точкам кадра:

$$G(v(p'_i)) = \frac{\sum_i \{A(v'(p_i))\}^2}{\{A(B'(z))\}^2}. \quad (5.15)$$

Общая же площадь всех видеокадров одинаковая $A(B'(z)) = A(B''(z))$ и определяется полем зрения. Диапазон значений гранулярности разбиений (или плотности областей Вороного) варьирует в пределах $0 < G(B'(z)) \leq 1$. Данный показатель площади областей тем больше, чем меньше количество областей и чем они больше по площади. Площадную близость разбиений двух кадров можно вычислить по формуле 5.16:

$$\rho_G(B'(z), B''(z)) = |G(v'(p_i)) - G(v''(p_j))|. \quad (5.16)$$

Площадная близость кадров не учитывает пространственное размещение разбиений, поэтому по-разному упорядоченные, но в то же время одинаковые по площади, области будут иметь общее значение гранулярности. Кроме того, площадная близость не учитывает размещения объектов [87].

Рассмотрим возможные варианты размещения объекта на диаграммах Вороного в двух последовательных видеокдрах. Объект (в данном случае это автомобили) может находиться:

- в одной области обоих кадров $B'(z)$ и $B''(z)$
- в одной области одного кадра $B'(z)$ и в разных областях другого кадра $B''(z)$;
- в нескольких областях одного кадра $B'(z)$ и в одной области другого кадра $B''(z)$;
- в разных областях обоих кадров $B'(z)$ и $B''(z)$.

Поэтому целесообразным представляется использование цветотекстурных признаков наряду с признаком, характеризующим форму областей Вороного, хотя другие признаки также могут быть задействованы при условии назначения им меньшего веса.

С учетом (5.12) и (5.14) можно переписать (5.11) в терминах координат опорных точек

$$\rho_1(V', V'') = \sum_{i=1}^n \sum_{j=1}^m \text{card} \left(\begin{array}{c} \mathfrak{e} \\ \mathfrak{e} \end{array} \cap \begin{array}{c} \mathfrak{e} \\ \mathfrak{e} \end{array} \right) \left| \frac{x_i - x_j}{y_i - y_j} \left(x - \frac{x_i - x_j}{2} \right) + \frac{y_i - y_j}{2} \right| D$$

$$D \left| \frac{x_i - x_j}{y_i - y_j} \left(x - \frac{x_i - x_j}{2} \right) + \frac{y_i - y_j}{2} \right|$$

$$\begin{aligned} \text{card} \left(\begin{array}{c} \mathbb{R} \times \mathbb{R} \\ \mathbb{C} \times \mathbb{C} \\ \mathbb{E} \times \mathbb{E} \end{array} \right) \hat{=} \int_{[1; y]} \left| \frac{x_{\mathbb{C}} - x_{\mathbb{C}}}{y_{\mathbb{C}} - y_{\mathbb{C}}} \left(x - \frac{x_{\mathbb{C}} - x_{\mathbb{C}}}{2} \right) + \frac{y_{\mathbb{C}} - y_{\mathbb{C}}}{2} \right| \\ \left| \begin{array}{c} \mathbb{R} \\ \mathbb{C} \\ \mathbb{E} \end{array} \right| \hat{=} \int_{[1; y]} \left| \frac{x_{\mathbb{C}} - x_{\mathbb{C}}}{y_{\mathbb{C}} - y_{\mathbb{C}}} \left(x - \frac{x_{\mathbb{C}} - x_{\mathbb{C}}}{2} \right) + \frac{y_{\mathbb{C}} - y_{\mathbb{C}}}{2} \right| \frac{\ddot{\cdot}}{\emptyset} \end{aligned} \quad (5.17)$$

Вышеприведенная метрика сходства показывает, насколько две диаграммы похожи с точки зрения формы областей. Чтобы принять во внимание цветовые и текстурные признаки, определим еще две метрики, $\rho_2(B'(z), B''(z))$ и $\rho_3(B'(z), B''(z))$ соответственно, заданные на общих частях разбиений:

$$\rho_2(B'(z), B''(z)) = \sum_{i=1}^n \sum_{j=1}^m \hat{=} (x_q, y_u) \hat{=} (v(p'_i) | v(p''_j)) (B'(x_q, y_u) - B''(x_q, y_u))^2, \quad (5.18)$$

где $B'(x_q, y_u)$ – значение интенсивности пикселя из области $(v(p'_i) | v(p''_j))$,

$$\rho_3(B'(z), B''(z)) = \sum_{i=1}^n \sum_{j=1}^m \hat{=} v(p'_i, v(p''_j)) \hat{=} (v(p'_i) | v(p''_j)) | E(v(p'_i)) - E(v(p''_j)) |, \quad (5.19)$$

где $E(v(p'_i))$ – значение энтропии для области $v(p'_i)$.

Измерения сходства цвета и текстуры, позволяют учесть соответствующие изменения в анализируемых областях Вороного двух кадров. Квадрат евклидова расстояния использовался для того, чтобы придать большие веса удаленным по цвету объектам. Манхэттенское расстояние выбрано для оценки сходства текстур потому, что расчет значений энтропии выполняется для каждой области в целом и каждое значение представлено одним числом с плавающей точкой (получается одно значение для каждой области Вороно-

го), тогда как сходство по цвету оценивается в каждом пикселе, присутствующем в обоих кадрах.

Метрики оценки сходства по цвету и текстуре можно представить аналогично формуле (5.17) в терминах координат опорных точек, а не областей Вороного [87]. Таким образом, получены три ненормализованные оценки. Далее предлагается выполнить нормализацию для формул 5.11, 5.18 и 5.19, чтобы получить значения в диапазоне от 0 до 1. Преобразование вышеуказанных метрик к ограниченной форме предполагает использование функции, называемой компандером диапазона:

$$\rho(B'(z), B''(z)) = \frac{\rho(B'(z), B''(z))}{1 + \rho(B'(z), B''(z))} \quad (5.20)$$

Комбинация этой функции с метрикой в результате по-прежнему приведет к метрике, удовлетворяющей требованиям рефлексивности, симметричности и неравенству треугольника.

Благодаря тому, что неотрицательная линейная комбинация метрик по-прежнему приводит к получению метрики, можно использовать следующую результирующую метрику:

$$\hat{\rho}(B'(z), B''(z)) = a_1\rho_1 + a_2\rho_2 + a_3\rho_3, \quad \sum_{g=1}^{g=3} a_g = 1, \quad a_g \geq 0 \quad (5.21)$$

где $\hat{\rho}(B'(z), B''(z))$ показывает сходство кадров;

a_g – важность каждого используемого признака.

5.3 Процедуры поиска ключевых кадров

В первом разделе упоминалось о том, что любой метод отбора ключевых кадров, как правило, предполагает определение границ сцен в качестве

предварительного этапа анализа видеоданных. В данном случае предлагается не только выполнить временную сегментацию на сцены, но и проанализировать однотипность содержания видеоматериала перед его обработкой. Это даст возможность определить порог извлечения действительно значимых для данной видеопоследовательности кадров. А учет границ сцен на последнем этапе отбора ключевых кадров, вместо первого, не будет накладывать ограничения на количество извлекаемых кадров, как это делается во многих общепринятых методах.

Для анализа однотипности содержания видеоматериала можно использовать значения энтропии, характеризующие каждый кадр. Высокий показатель энтропии свидетельствует о большом разбросе значений пикселей, тогда как низкий показатель энтропии говорит об однотипности пикселей, а следовательно, и деталей. Таким образом, значение энтропии показывает, насколько много деталей содержится в локальной области, для которой выполняется расчет энтропии.

Видеокадр без объектов, не содержащий текстуры (полностью черный, белый или серый), будет иметь значение энтропии, равное 0. Интересно, что любой другой цвет заливки (от светло-желтого до темно-синего) дает значение энтропии 1,58, однако появление каких-либо деталей на белом фоне приведет к появлению лишь десятых долей значения энтропии. Однако, для реальных видеоматериалов какая-то текстура все-таки присутствует в видеокдрах и, как правило, значение энтропии варьирует в пределах от 3,25 до 7,75. Значение энтропии менее 5 показывает наличие незначительной текстуры и небольшого количества деталей. Больше всего значений, как правило, приходится на диапазон от 7 до 7,75.

Для последовательно идущих кадров отличие энтропии на одну десятую, вероятнее всего означает, что объект наблюдения не изменился, а лишь переместился в пространстве. Хотя вероятность начала новой сцены также присутствует, потому что одно и то же значение энтропии может быть определено для абсолютно разного содержания видеокдра. По этим причинам

только значения энтропии не могут служить для корректного определения границ сцен, но для оценки однотипности содержания всего видеоматериала это является простым в расчете и достаточно верным показателем. В подтверждение вышесказанного, в таблице 5.2 показаны миниатюры кадров четырех видеопоследовательностей и соответствующие им значения энтропии.

Номера кадров в левом верхнем углу обозначают видеоролик, из которого они были взяты. Так, диапазон значений энтропии первого видеоролика варьирует от 3,25 до 7,74. Этот видеоматериал характеризуется наибольшим изменением сцен, среди анализируемых видео. Диапазон значений энтропии для второго видео составляет 4,12 – 6,59. Здесь также много изменений сцен, но съемка в основном выполнена в ночное время и темный фон несколько сужает диапазон видимых деталей. Третий видеоматериал характеризуется практически полным отсутствием изменений сцены, перемещение камеры обуславливает съемку разных объектов, однако все они отсняты при одинаковых условиях и очень близки по содержанию, диапазон изменения энтропии от 7,1 до 7,53. Для четвертого видеоролика характерно наличие нескольких сцен, диапазон значений энтропии здесь достаточно широк – от 5,85 до 7,75, просто кадры не из всего диапазона значений были внесены в таблицу.

Далее приводится описание всей процедуры сопоставления разбиений Вороного для отбора ключевых кадров:

1. Определение однотипности видеокадров. С учетом разброса значений текстуры (дисперсии энтропии) во всей видеопоследовательности выполним установку порогового значения согласно следующему правилу

$$Threshold = \begin{cases} \frac{1}{4}, & S_B \in \mathbb{R}^+ \\ \frac{1}{2}, & S_B \in \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)), \\ \frac{3}{4}, & S_B \in \mathbb{R}^0 \end{cases}, \quad (5.22)$$

$$\text{где } S_B = \frac{1}{K-1} \sum_{k=1}^{k=K} E(B_k(z)) - \frac{1}{K} \sum_{k=1}^{k=K} E(B_k(z)) \frac{2}{\sigma^2},$$

K – общее число кадров видеопоследовательности;

$E(B_k(z))$ – значение энтропии для k -ого видеокadra.

Таблица 5.2 – Миниатюры видеокadров и соответствующие им значения энтропии

Диапазон значений энтропии	Миниатюры видеокadров
3,25 – 4,99	
5 – 5,99	
6 – 6,99	
7 – 7,25	
7,26 – 7,5	
7,51 – 7,75	

Из-за того, что близость содержания является относительной мерой подобия видеокadров, приходится использовать пороговое значение для зада-

ния приемлемой степени их сходства. В данном случае пороговое значение устанавливается с учетом содержания видео. Для видео с гетерогенным содержанием и большим разнообразием сцен, это значение должно быть меньше $\frac{1}{4}$, чтобы не извлекать слишком большого числа кадров. И наоборот, для видео с гомогенным содержанием и небольшим числом сцен (или даже с одной сценой), это значение должно быть увеличено до $\frac{3}{4}$, чтобы извлечь не-много больше ключевых кадров.

Энтропия была выбрана в целях определения однотипности потому, как проведенные исследования [162] выявили показательность изменения этих значений для характеристики содержания видео. Кроме того, энтропия считается хорошей мерой разнообразия для набора векторов признаков. Признаки, которые уменьшают неопределенность, являются более информативными, то есть уменьшение энтропии ведет к более рациональному выбору признаков [141].

2. Берем первый ($B_k(z)$) и второй ($B_{k+1}(z)$) кадр для сравнения. Устанавливаем $k = 1$.

3. Покадровое сравнение диаграмм Вороного. Согласно формуле (5.21), рассчитываем значение $\hat{\rho}(B_k(z), B_{k+1}(z))$ для двух кадров. Если значение $\hat{\rho}(B_k(z), B_{k+1}(z))$ оказалось меньшим, чем установленное пороговое значение, в таком случае оба кадра $B_k(z)$ и $B_{k+1}(z)$ считаются ключевыми и извлекаются как $B_r^*(z)$ и $B_{r+1}^*(z)$, после чего переходим к следующему этапу 4, в противном случае извлекается только один кадр $B_k(z)$ как ключевой кадр $B_r^*(z)$ и переходим к этапу 5.

4. Переопределяем значения $B_k(z) = B_{k+1}(z)$, $B_{k+1}(z) = B_{k+2}(z)$ и переходим к этапу 6.

5. Оставляем $B_k(z) = B_k(z)$ и устанавливаем $B_{k+1}(z) = B_{k+2}(z)$.

6. Повторяем этап 3 до тех пор, пока $B_{k+1}(z) \notin K$.

7. Отбор ключевых кадров с учетом границ сцен. Последним этапом будет удаление похожих кадров, если такие были отобраны в рамках одной сцены. Границы сцен можно определить с помощью одного методов анализа временных рядов, рассмотренных в разделе 4. Сравниваем ключевые кадры попарно внутри каждой сцены с использованием процедуры, описанной на этапе 3. Найденный второй идентичный ключевой кадр удаляем из последовательности ключевых кадров. Таким образом, результирующая последовательность ключевых кадров будет выглядеть следующим образом $\{\hat{B}_1^*(z), \dots, \hat{B}_l^*(z)\}$.

Применение специально разработанных метрик сходства, ориентированных непосредственно на анализируемые группы признаков, позволяет повысить качество отбора кадров, ведь, как известно, на результативность любой процедуры влияет не только характер анализируемых данных и применяемых методов, но и в значительной мере выбранные и используемые метрики сходства.

Процедура может быть расширена использованием взвешенных оценок признаков, так как в данном случае была задействована их линейная комбинация с одинаковыми весовыми коэффициентами, кроме того, могут быть использованы дополнительные признаки, например, движение областей Вороного (хотя это значительно увеличит время обработки видеоданных). Определение важности каждого используемого признака является приоритетным направлением дальнейших исследований.

Далее необходимо провести анализ предложенной процедуры поиска ключевых кадров. Прежде всего, были проведены исследования с помощью наиболее общепринятых оценок, какими являются меры полноты и точности. В целом, задача стоит в достижении определенной полноты поиска при удовлетворительном уровне ложно позитивных ключевых кадров. Показатель, позволяющий найти баланс между точностью и полнотой поиска, называется F-мерой (F-measure):

$$F = \frac{1}{a \frac{1}{P} + (1-a) \frac{1}{R}} = \frac{(b^2 + 1)PR}{b^2P + R} \quad (5.23)$$

где $a \in [0,1]$;

$$b^2 = \frac{1-a}{a}, \quad b^2 \in [0, \infty).$$

При $b < 1$ отдается предпочтение точности, а при $b > 1$ – полноте поиска ключевых кадров. Если точности и полноте назначить одинаковые веса ($a = 1/2$ или $b = 1$), получится сбалансированная F-мера, которую принято обозначать F_1 ($F_1 \in [0,1]$) – сокращение от $F_{b=1}$. При $b = 1$ формула (5.23) упрощается и становится равной коэффициенту Дайса (Dice coefficient) [78] для множеств найденных и релевантных кадров:

$$F_{b=1} = \frac{2PR}{P+R} = \frac{2|P \cap R|}{|P \cup R|}.$$

Для кадров, отобранных в качестве ключевых с помощью диаграмм Вороного из тестовой коллекции TRECVID, коэффициент Дайса в среднем составил 0,92, что свидетельствует о неплохих результатах по сравнению с оценкой метода оптического потока, которая в среднем составила 0,65; оценка метода кластеризации k -средних для поиска ключевых кадров составила в среднем 0,78; тогда как кадры отобранные методом упрощения кривой были оценены в среднем в 0,83.

Помимо мер полноты и точности, для оценки качества кластеров, полученных во время перекомпоновки опорных точек методом k -средних, использован показатель достоверности кластеров (cluster validity), расчет которого выполняется согласно формуле

$$r(n) = \frac{1}{n} \sum_{i=1}^n \max_{j \in \{1, \dots, n\}} \frac{\sum_{j \in C_i} P_{ij}}{P_{ij}}, \quad n \geq 2,$$

где n – общее количество кластеров;

s_i – дисперсия кластера с индексом i ;

P_{ij} – расстояние Минковского между кластерами i и j .

Чем лучше n кластеров отличаются друг от друга, тем меньшим будет значение показателя достоверности кластеров и тем больше вероятность того, что количество кластеров, равное n , является оптимальным для данного видеокадра. На рис. 5.12 показан график изменения показателя достоверности кластеров для видеопоследовательности с высоким разрешением и однородным содержанием. Анализ изменения показателя достоверности кластеров для различных видеопоследовательностей показал корректность разбиения на классы.

С точки зрения быстродействия предложенного подхода следует отметить следующее, предложенный метод поиска ключевых кадров состоит из ряда этапов обработки данных и включает в себя различные алгоритмы, соответственно говорить о высоком быстродействии сложно. Таким образом, для повышения быстродействия можно, например, использовать видео меньшего разрешения.

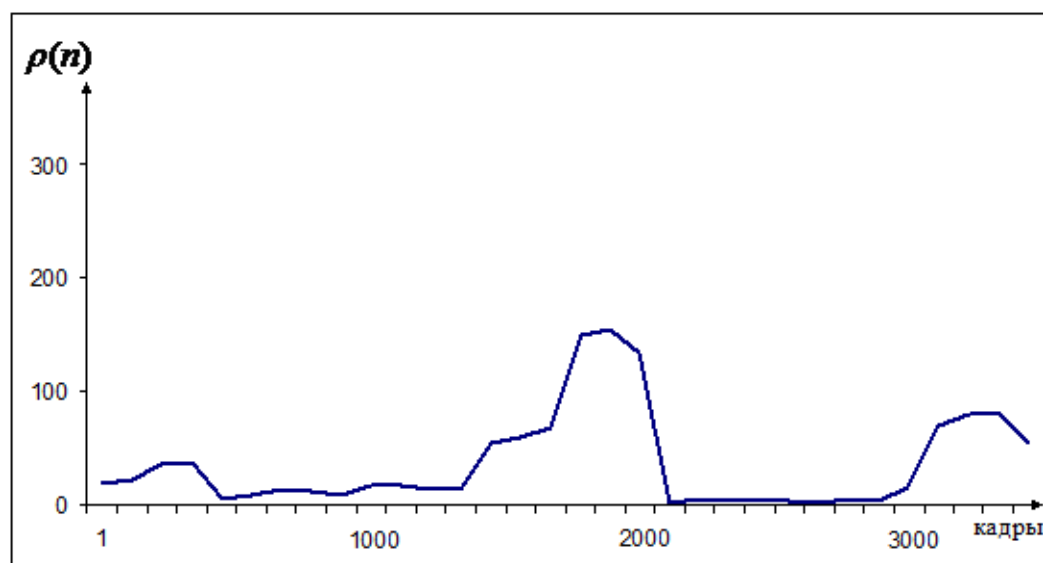


Рисунок 5.12 – Изменение показателя достоверности кластеров для видеопоследовательности с высоким разрешением и однородным содержанием

Кроме размера и разрешения видео, которые влияют на скорость поиска ключевых кадров, среди всех этапов в рамках предложенного метода наиболее ресурсоемким является определение опорных точек методом Харриса за счет вычисления производных. Следовательно, возможно следует использовать другие методы поиска опорных точек, которые могут быть использованы вместо метода Харриса (например, методы SURF, SIFT или FAST) для приближения процедуры поиска ключевых кадров к реальному времени. Если говорить о качестве поиска, переупорядочение опорных точек с помощью алгоритмов кластеризации делает инвариантным результирующее размещение опорных точек относительно выбранного метода их исходного размещения, хотя и изменяет число итераций.

Выводы по пятому разделу

1. Методы выделения и анализа ключевых кадров обеспечивают сведение механизмов CBVIR к схемам CBIR, что позволяет использовать для видеопотоков накопленный арсенал средств поиска статических изображений по их содержанию.

2. Представление «содержания» видеокадров на основе диаграмм Вороного различных порядков с учетом цвето-текстурных признаков и признаков формы областей обеспечивает возможности управления числом опорных точек (степенью детализации видеокадров) и позволяет осуществлять достаточно надежное сравнение видеокадров на основе анализа индуцированных разбиений поля зрения.

3. Вариации в наборах ключевых кадров дают возможность реферирования видео с различными характеристиками уточнения и/или огрубления моделей сегментов, что создает предпосылки для стратифицированного аннотирования видеопотока в целом.

РАЗДЕЛ 6

МЕТОДЫ И МОДЕЛИ КОМПРЕССИИ ИЗОБРАЖЕНИЙ ДЛЯ ТЕМПО-
РАЛЬНОЙ ОБРАБОТКИ ВИДЕО

Раздел направлен на описание методов, позволяющих сократить объемы хранимых данных, что весьма актуально с точки зрения анализа видеопотоков, т.к. решение задачи информационного поиска подразумевает хранение не только самих видео, но и дополнительной информации, как минимум, в виде границ сегментов и ключевых кадров каждого из сегментов. Информационные технологии сокращения статистической, структурной, психофизиологической и содержательной избыточности визуальной информации достаточно хорошо развиты и непрерывно совершенствуются в различных аспектах повышения эффективности их использования. Применительно к СВВIR перспективным направлением развития информационных технологий является разработка интеллектуальных, в первую очередь, нейросетевых методов компрессии данных.

Раздел основан на работах автора [28, 32, 83, 85, 86, 88, 132, 200].

6.1 Методы сжатия изображений, связанные с понижением размерности исходных данных на базе многослойных нейронных сетей

В задачах анализа и обработки сигналов высокой размерности, возникающих, прежде всего, при распознавании и обработке изображений, космических снимков, данных экологического мониторинга, биометрической информации и т.п., на предварительных этапах исследования достаточно часто появляется необходимость понижения размерности исходного пространства признаков, для чего традиционно используется стандартный анализ главных компонент. С теоретической точки зрения анализ главных компонент сводится к хорошо изученной алгебраической проблеме собственных значений, а ее решение в задачах, связанных с обработкой изображений, известно как пре-

образование Карунена-Лоэва. При этом предполагается, что исходная информация задана в виде фиксированного массива данных, образованного N n -мерными векторами $x(1), x(2), \dots, x(k), \dots, x(N)$, где $k = 1, 2, \dots, N$ – номера наблюдения в исходном массиве, а результатом решения является набор доминантных собственных значений $\lambda_1 > \dots > \lambda_j > \dots > \lambda_m$ и соответствующих собственных значений векторов $w_1, w_2, \dots, w_j, \dots, w_m$, $w_j = (w_{j1}, w_{j2}, \dots, w_{jn})^T$ корреляционной $(n \times n)$ – матрицы исходных данных

$$\begin{aligned} R(N) &= \frac{1}{N} \sum_{k=1}^N (x(k) - \bar{x}(N))(x(k) - \bar{x}(N))^T, \\ \bar{x}(N) &= \frac{1}{N} \sum_{k=1}^N x(k). \end{aligned} \quad (6.1)$$

Собственно компрессия (понижение размерности) исходного пространства производится с помощью отображения

$$y(k) = Wx(k),$$

где $y(k) = (y_1(k), y_2(k), \dots, y_m(k))^T$,

$W = (w_1, w_2, \dots, w_{m-1}, w_m)^T$ – $(m \times n)$ – матрица проецирования, образованная доминантными собственными векторами корреляционной матрицы $R(N)$.

Существенные проблемы при решении этих задач возникают, когда исходная размерность n достаточно велика, k имеет смысл текущего дискретного времени, сами данные в виде векторов $x(k)$ поступают на обработку последовательно в реальном времени. В этой ситуации на первый план выходит подход, основанный на применении искусственных нейронных сетей [20, 135, 184, 200, 213, 272].

К настоящему времени известно достаточно много нейросетевых архи-

текстур и алгоритмов обучения, решающих задачи выделения главных компонент, которые можно условно разделить на две большие подгруппы: последовательные и параллельные системы.

В последовательных системах [20, 64, 143], как следует из названия, главные компоненты выделяются последовательно одна за одной, при этом каждая следующая компонента вычисляется с помощью уже вычисленной предыдущей. Преимуществом таких систем является возможность адаптивно изменять количество нейронов в сети (а, следовательно, и число необходимых главных компонент) по ходу процесса обучения, а основным недостатком – низкая скорость настройки, ограничивающая использование этих нейросетей в системах видеобработки реального времени.

В параллельных системах [20] все нейроны сети работают одновременно, благодаря чему достигается увеличение быстродействия, однако результатом вычисления являются не собственно главные компоненты, а, так называемые, главные подпространства, натянутые на доминантные собственные векторы. В таких системах, хотя и обеспечивается эффективное сжатие данных, физическая интерпретация результатов становится невозможной.

В основе настройки большинства из указанных сетей лежит минимизация критерия, основанного на квадрате ошибки обучения, в силу чего эти процедуры, так или иначе, являются ближайшими родственниками алгоритмов обучения аппроксимирующих нейронных сетей. Несколько в стороне стоит сеть Мяо-Хуа, обучение которой основано на использовании не квадратичного (в смысле ошибки) информационного критерия (NIC), благодаря чему достигается некоторое увеличение быстродействия по сравнению с другими указанными выше нейросетями. К сожалению, нейросеть Мяо-Хуа обеспечивает лишь анализ главных подпространств вместо желаемого анализа главных компонент.

В связи с вышеизложенным, рассмотрим оптимальный алгоритм самообучения на основе взвешенного информационного критерия. В основе предлагаемого алгоритма лежат подходы, предложенные Оя с соавторами для од-

нослойной нейронной сети, образованной адаптивными линейными ассоциаторами и реализующей отображение

$$y(k) = W(k - 1)x(k),$$

предложили в качестве критерия самообучения (энергетической функции) использовать конструкцию

$$\min_W \{E^{\text{ooW}}(W) = -\frac{1}{2} Sp(WR(k)W^T)\} \quad (6.2)$$

при ограничениях

$$WW^T = A, \quad (6.3)$$

где $A = \text{diag}\{a_i\}$ – произвольная $(m \times m)$ – диагональная матрица с элементами $a_1 > a_2 > \dots > a_m$.

Было показано, что решение задачи оптимизации (6.2), (6.3) приводит к получению набора из m доминантных собственных векторов w_j корреляционной матрицы $R(N)$, однако эффективного алгоритма решения предложено не было.

Мяо и Хуа ввели так называемый новый информационный критерий (NIC), имеющий вид

$$\min_W \{E^{\text{NIC}}(W) = -\frac{1}{2} Sp(\log(WR(N)W^T) - Sp(WW^T))\} \quad (6.4)$$

и градиентную процедуру его минимизации, в результате чего могут быть найдены главные подпространства, натянутые на доминантные собственные

векторы матрицы $R(N)$. К сожалению, минимизация критерия (6.4) не приводит к нахождению главных компонент.

На основе (6.2), (6.3) и (6.4) был введен взвешенный информационный критерий (WINC)

$$\min_W \{E^{\text{WINC}}(W) = -\frac{1}{2} \text{Sp}(\log(WRW^T A) - \text{Sp}(WW^T))\} \quad (6.5)$$

и алгоритм обучения нейронной сети на его основе вида

$$W(k) = W(k-1) - hG(k) = W(k-1) + h(A^{-1}W(k-1) - R(k)W^T(k-1)A^{-1}W(k-1)R(k) - W(k-1)), \quad (6.6)$$

где h – постоянный параметр шага поиска,

$$G = -((A^{-1}WRW^T A)^{-1}WR - W) = \left\{ \frac{\partial E^{\text{WINC}}}{\partial w_{ji}} \right\} - (m \times n) \text{ матрица, обра-}$$

зованная частными производными (6.5) по настраиваемым параметрам w_{ji} ,

$H(k)$ – корреляционная матрица, вычисляемая по k наблюдениям.

В случае, если обработка данных проводится в реальном времени, для вычисления этой матрицы вместо (6.1) могут быть использованы рекуррентные соотношения

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

для стационарного случая и

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

для нестационарного. Здесь $0 < \mathbf{a} < 1$ – эмпирически выбираемый параметр забывания устаревшей информации. Авторами алгоритма (6.6) доказана его сходимость к доминантным собственным векторам. При этом интересно заметить, что при единичной матрице A ($A = I_m$) алгоритм (6.6) минимизирует критерий НИС (6.4), т.е. с помощью одного алгоритма можно находить и главные компоненты, и главные подпространства.

Скорость сходимости алгоритма (6.6), как и всех градиентных алгоритмов оптимизации, существенным образом зависит от скалярного параметра шага поиска h , который в данном случае полагается постоянным. Ясно, что оптимальный выбор этого параметра позволит повысить эффективность решения задачи в целом.

Введя в рассмотрение апостериорную ошибку восстановления

$$\mathcal{E}(k) = e(k) + G(k)y(k), \quad (6.7)$$

где $G(k) = -((A^{-1}W(k-1)H(k)W^T(k-1)A)^{-1}W(k-1)^T H(k) - W(k-1))$, можно ввести вспомогательный критерий оптимизации

$$\min_h \{E(h) = \|\mathcal{E}(k)\|^2\}, \quad (6.8)$$

минимизация которого обеспечивает максимально возможную скорость сходимости алгоритма (6.6).

Записав квадрат нормы вектора ошибок (6.7)

$$\|\mathcal{E}(k)\|^2 = \|e(k)\|^2 - 2he^T(k)G(k)y(k) + h^2 \|(G(k)y(k))\|^2$$

и решив уравнение

$$\frac{\partial \|e(k)\|^2}{\partial h} = -e^T(k)G(k)y(k) + h\|(G(k)y(k))\|^2 = 0,$$

получаем оптимальное значение параметра шага поиска в виде

$$h(k) = \frac{e^T(k)G(k)y(k)}{\|(G(k)y(k))\|^2}.$$

Окончательно оптимальный алгоритм самообучения на основе взвешенного информационного критерия может быть представлен в виде

$$W(k) = W(k-1) + \frac{((x(k) - W^T(k-1)y(k))^T G(k)y(k))}{\|G(k)y(k)\|^2} G(k). \quad (6.9)$$

Архитектура нейронной сети, решающая рассматриваемую задачу, представлена на рис. 6.1. Сеть имеет два слоя, образованных m (в первом скрытом слое) и n (в выходном слое) адаптивными линейными ассоциаторами. В первом скрытом слое, синаптические веса которого образуют $(m \times n)$ – матрицу $W = \{w_{ji}\}$, производится сжатие информации, при этом на его выходе вычисляются значения главных компонент y_1, y_2, \dots, y_m . Выходной слой служит для восстановления входного сигнала, с помощью $(n \times m)$ матрицы синаптических весов $W^T = \{w_{ij}\}$, в связи с чем на его выходе появляются значения $\hat{x}_1(k), \hat{x}_2(k), \dots, \hat{x}_n(k)$, являющиеся оценками компонент входного сигнала:

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

Понятно, что такое восстановление возможно при $m = n$, однако при $m < n$ обеспечивается восстановление с максимально возможной точностью в смысле критерия (6.8)

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

что выгодно отличает рассматриваемые здесь конструкции от традиционных подходов.

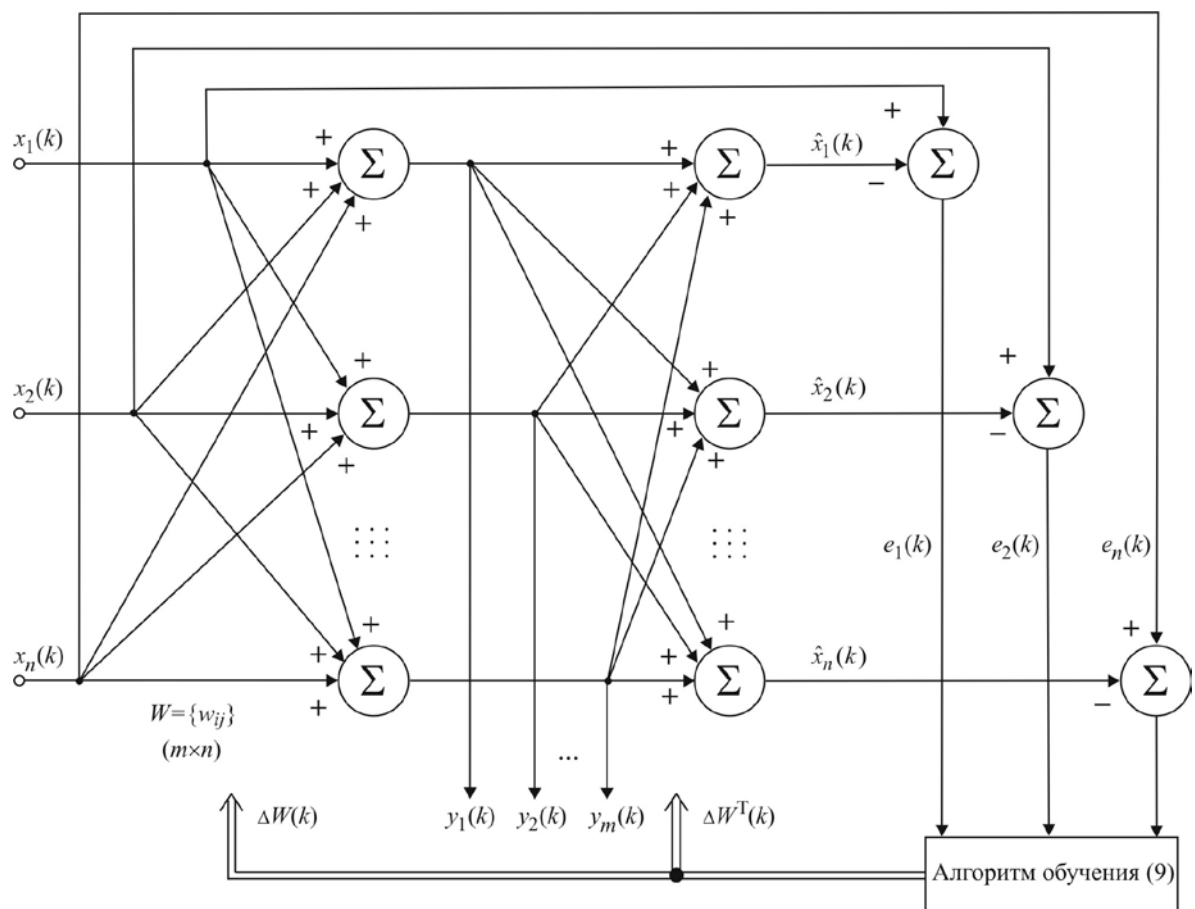


Рисунок 6.1 – Параллельная искусственная нейронная сеть для нахождения главных компонент

Однако данный подход не лишен недостатков, в связи с чем целесообразным оптимизировать процесс обучения четырехслойной гетерогенной автосоциативной нейронной сети с прямой передачей информации, предна-

значенной для понижения размерности обрабатываемых векторов-образов, с учетом влияния помех и возможным нарушением выпуклости целевых функций скрытых слоев.

6.2 Метод сжатия с помощью 4-х слойной автоассоциативной нейронной сети

Архитектура рассматриваемой автоассоциативной нейронной сети, предназначенной для понижения размерности пространства исходных образов, приведена на рис. 6.2 и содержит четыре последовательно соединенных слоев нейронов.

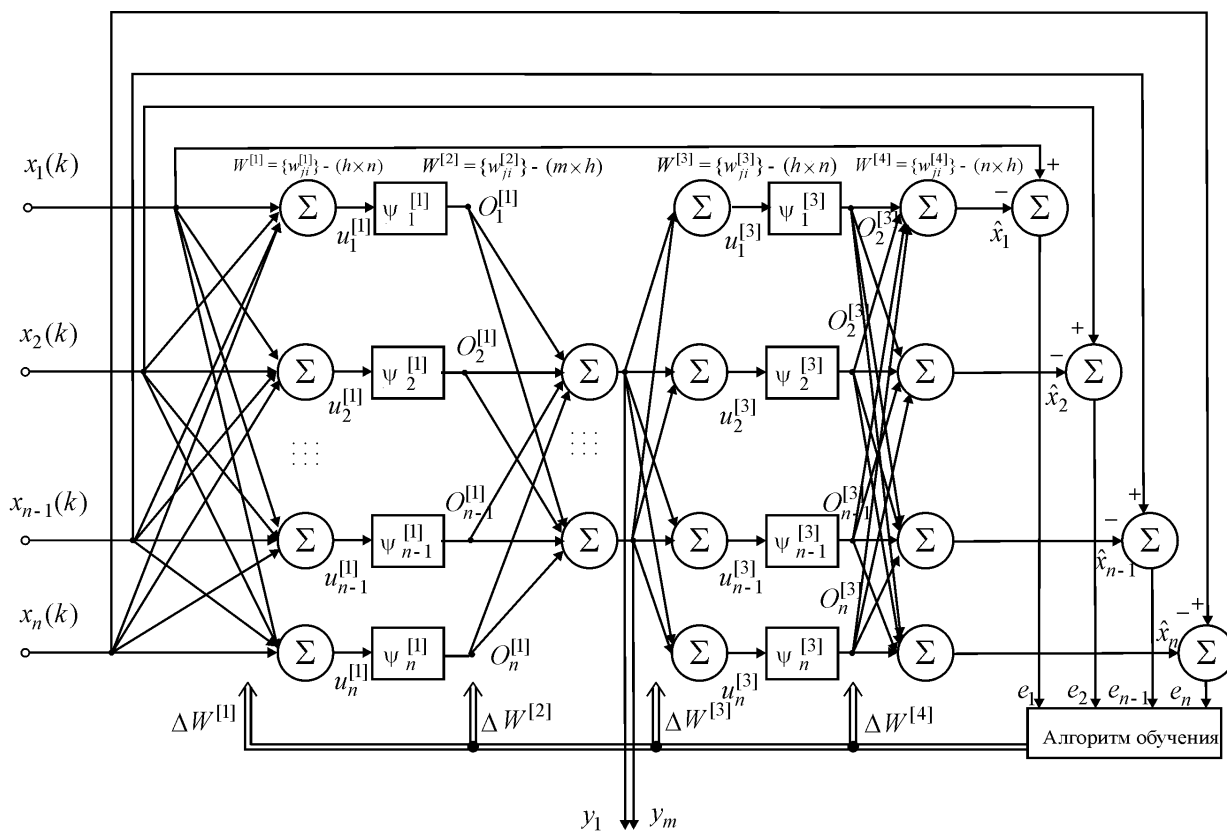


Рисунок 6.2 – Автоассоциативная многослойная нейронная сеть

На вход (рецепторный нулевой слой) подается последовательность векторов образов $x(1), x(2), \dots, x(k), \dots, x(N)$, аналогично предыдущему случаю, $x(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T \hat{1} \mid N$, предварительно пронормированных на

гиперкуб так, что $x_i(k) \in [-1, 1]$, $i = 1, 2, \dots, n$.

Первый скрытый слой содержит $h^3 n$ нейронов – элементарных персептронов Розенблатта с сигмоидальными активационными функциями. В случае отсутствия сигнала смещения, первый скрытый слой содержит hn настраиваемых синаптических весов и описываемых соотношениями вида:

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

$$O^{[1]}(k) = Y^{[1]}(W^{[1]}x(k)),$$

где $O_j^{[1]}(k)$ – выходной сигнал j -го нейрона первого скрытого слоя $x(k)$,

$y_j^{[1]}(0)$ – активационная функция j -го нейрона первого скрытого слоя,

$u_j^{[1]}(k)$ – сигнал внутренней активации j -го нейрона первого скрытого слоя,

$w_{ji}^{[1]}$ – синаптический вес i -го входа j -го нейрона первого скрытого слоя,

$$O^{[1]}(k) = (O_1^{[1]}(k), \dots, O_j^{[1]}(k), \dots, O_h^{[1]}(k))^T,$$

$Y^{[1]} = \text{diag}(Y_1^{[1]}, \dots, Y_j^{[1]}, \dots, Y_h^{[1]})$ – $(h \times h)$ -диагональная матрица активационных функций,

$$W^{[1]} = \{w_{ji}^{[1]}\} - (h \times n) - \text{матрица синаптических весов.}$$

Если нейроны первого скрытого слоя включают сигнал смещения, то приведенные выше соотношения принимают вид:

$$O_j^{[1]}(k) = y_j^{[1]}(u_j^{[1]}(k)) = y_j^{[1]}(\sum_{i=1}^n w_{ji}^{[1]} x_i(k) + Q_j^{[1]}) = y_j^{[1]}(\sum_{i=0}^n w_{ji}^{[1]} x_i(k)),$$

$$O^{[1]}(k) = Y^{[1]}(W^{[1]}x(k)),$$

где $Q_j^{[1]}$ – сигнал смещения j -го нейрона первого скрытого слоя,
 $w_{j0}^{[1]} = Q_j^{[1]}, x_0(k) = 1, W^{[1]}$ – $(h' n + 1)$ -матрица синаптических весов,
 $x(k) = (1, x_1(k), \dots, x_n(k))^T$.

В этом случае слой содержит $h(n + 1)$ настраиваемых параметров.

В качестве активационных функций используются либо логистическая функция

$$y_j^{[1]}(u_j^{[1]}) = \frac{1}{1 + e^{-g_j u_j^{[1]}}},$$

либо гиперболический тангенс

$$y_j^{[1]}(u_j^{[1]}) = \tanh(g_j u_j^{[1]}) = \frac{1 - e^{-2g_j u_j^{[1]}}}{1 + e^{-2g_j u_j^{[1]}}},$$

где g_j – положительный параметр (возможно настраиваемый), задающий «крутизну» функции.

Полезно также ввести в рассмотрение производные логистической функции

$$\frac{\nabla y_j^{[1]}(u_j^{[1]})}{\nabla u_j^{[1]}} = g_j y_j^{[1]}(u_j^{[1]}) (1 - y_j^{[1]}(u_j^{[1]}))$$

и гиперболического тангенса

$$\frac{\Psi(y_j^{[1]}(u_j^{[1]}))}{\Psi(u_j^{[1]})} = g_j(1 - (y_j^{[1]}(u_j^{[1]}))^2)$$

соответственно.

Второй скрытый слой содержит $m < n$ нейронов – адаптивных линейных ассоциаторов, а его выходной сигнал $y = (y_1, \dots, y_m)^T$ является выходом нейронной сети в целом и представляет собой «сжатый» входной образ x .

В отсутствие смещения второй скрытый слой описывается соотношениями

$$y_j(k) = \mathop{\text{â}}\limits_{i=1}^h w_{ji}^{[2]} o_i^{[1]}(k), \quad j = 1, 2, \dots, m$$

$$y(k) = W^{[2]} O^{[1]}(k),$$

где $W^{[2]} = \{w_{ji}^{[2]}\}$ – $(m \times h)$ – матрица синаптических весов.

Если же нейроны второго скрытого слоя включают сигнал смещения, то приведенные выше соотношения принимают вид

$$y_j(k) = \mathop{\text{â}}\limits_{i=1}^h w_{ji}^{[2]} o_i^{[1]}(k) + Q_j^{[2]} = \mathop{\text{â}}\limits_{i=0}^h w_{ji}^{[2]} o_i^{[1]}(k)$$

$$y(k) = W^{[2]} O^{[1]}(k),$$

где $W^{[2]} = \{w_{ji}^{[2]}\}$ – $(m \times (h + 1))$ – матрица синаптических весов,

$$O^{[1]}(k) = (1, O_1^{[1]}(k), \dots, O_h^{[1]}(k))^T.$$

Третий скрытый слой подобен первому и также содержит h элементарных персептронов Розенблатта. В отсутствие смещения этот слой описывается соотношениями

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

$$O^{[3]}(k) = Y^{[3]}(W^{[3]}y(k)),$$

где $W^{[3]}$ – $(h \times m)$ -матрица синаптических весов.

Если же нейроны третьего скрытого слоя включают сигнал смещения, то приведенные выше соотношения имеют вид

$$O_j^{[3]}(k) = y_j^{[3]}(u_j^{[3]}(k)) = y_j^{[3]}(\mathring{\mathbf{a}} \sum_{i=1}^m w_{ji}^{[3]} y_i(k) + Q_j^{[3]}), j = 1, 2, \dots, h,$$

$$O^{[3]}(k) = Y^{[3]}(W^{[3]}y(k)),$$

где $W^{[3]}$ – $(h \times (m + 1))$ -матрица синаптических весов.

Четвертый выходной слой подобен второму скрытому и содержит n нейронов – адаптивных линейных ассоциаторов. В отсутствии смещения четвертый слой описывается соотношениями

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

где $\mathring{\mathbf{x}}(k)$ – $(n \times 1)$ -вектор, являющийся оценкой входного сигнала $x(k)$, восстановленного после сжатия,

$W^{[4]}$ – $(n \times h)$ -матрица синаптических весов.

Если адаптивные линейные ассоциаторы четвертого слоя включают сигнал смещения, то приведенные выше соотношения имеют вид

$$\mathring{x}_j(k) = \mathring{\mathbf{a}} \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k) + Q_j^{[4]} = \mathring{\mathbf{a}} \sum_{i=1}^h w_{ji}^{[4]} o_i^{[3]}(k), o_0^{[3]}(k) = 1,$$

$$\hat{x}(k) = W^{[4]}O^{[3]}(k),$$

где $W^{[4]}$ – $(n \times (h+1))$ -матрица синаптических весов.

Таким образом, отображение, реализуемое четырехслойной автоассоциативной нейронной сетью, приведенной на рис.6.2, имеет вид

$$\hat{x}(k) = W^{[4]}(Y^{[3]}(W^{[3]}W^{[2]}Y^{[1]}(W^{[1]}x(k)))).$$

В качестве критерия обучения будем использовать стандартную квадратичную функцию

$$\begin{aligned} E(k) &= \sum_{j=1}^n E_j(k) = \frac{1}{2} \sum_{j=1}^n (x_j(k) - \hat{x}_j(k))^2 = \\ &= \frac{1}{2} \sum_{j=1}^n e_j^2(k) = \frac{1}{2} \|e(k)\|^2 = \frac{1}{2} \|x(k) - \hat{x}(k)\|^2, \end{aligned}$$

где $e(k) = (e_1(k), \dots, e_n(k))^T$ – вектор ошибок восстановления (обучения) входного сигнала по всем координатам.

Сам же процесс обучения будем проводить на основе градиентной процедуры вида

$$w_{ji}^{[s]}(k+1) = w_{ji}^{[s]}(k) - h^{[s]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[s]}}, \quad s = 1, 2, 3, 4,$$

или

$$Dw_{ji}^{[s]} = w_{ji}^{[s]}(k+1) - w_{ji}^{[s]}(k) = - h^{[s]}(k) \frac{\partial E(k)}{\partial w_{ji}^{[s]}},$$

где $h^{[s]}(k)$ – параметр шага обучения s -го слоя, определяющий скорость процесса настройки.

Следуя концепции обратного распространения ошибок, обучение сети начинается с четвертого (выходного) слоя при этом

$$Dw_{ji}^{[4]} = -h^{[4]}(k) \frac{\nabla E(k)}{\nabla w_{ji}^{[4]}} = -h^{[4]}(k) \frac{\nabla E_j(k)}{\nabla \hat{x}_j} \frac{\nabla \hat{x}_j}{\nabla w_{ji}^{[4]}} = h^{[4]}(k) e_j(k) O_i^{[3]}(k).$$

Поскольку выходной сигнал слоя $\hat{x}(k)$ линейно зависит от настраиваемых параметров, для обучения можно использовать либо оптимальный по быстродействию алгоритм Качмажа-Уидроу-Хоффа [223]

$$Dw_{ji}^{[4]} = \frac{e_j(k) O_i^{[3]}(k)}{\|O^{[3]}(k)\|^2},$$

в имеющий вид векторной форме

$$Dw_j^{[4]} = \frac{e_j(k) O^{[3]}(k)}{\|O^{[3]}(k)\|^2}, \quad j = 1, 2, \dots, n,$$

либо, если требуется дополнительная фильтрация возмущений и сглаживание, процедуру вида

$$\begin{cases} Dw_j^{[4]} = (p^{[4]}(k))^{-1} e_j(k) O^{[3]}(k), \\ p^{[4]}(k) = \alpha p^{[4]}(k-1) + \|O^{[3]}(k)\|^2, \end{cases} \quad (6.10)$$

где $0 \leq \alpha \leq 1$ – параметр сглаживания.

Обучение параметров третьего скрытого слоя производится с помощью алгоритма

$$Dw_{ji}^{[3]} = -h^{[3]}(k) \frac{\nabla E(k)}{\nabla w_{ji}^{[3]}} = -h^{[3]}(k) \frac{\nabla E(k)}{\nabla u_j^{[3]}} \times \frac{\nabla u_j^{[3]}}{\nabla w_{ji}^{[3]}} = h^{[3]}(k) d_j^{[3]}(k) y_i(k), \quad (6.11)$$

где $d_j^{[3]}(k) = \frac{\nabla E(k)}{\nabla u_j^{[3]}}$, $j=1,2,\dots,h$ так называемая начальная ошибка (d-

ошибка) третьего скрытого слоя, которая также может быть представлена в форме

$$d_j^{[3]}(k) = - \frac{\nabla E(k)}{\nabla O_j^{[3]}(k)} \times \frac{\nabla O_j^{[3]}(k)}{\nabla u_j^{[3]}}$$

а с учетом того, что

$$\nabla O_j^{[3]}(k) = y_j^{[3]}(u_j^{[3]}(k)),$$

ЛОГИЧНО ЗАПИСАТЬ

$$d_j^{[3]}(k) = - \frac{\nabla E(k)}{\nabla O_j^{[3]}(k)} \times \frac{y_j^{[3]}(u_j^{[3]}(k))}{\nabla u_j^{[3]}}$$

представив далее $-\frac{\nabla E(k)}{\nabla O_j^{[3]}(k)}$ в виде

$$\begin{aligned}
-\frac{\|E(k)\|}{\|O_j^{[3]}(k)\|} &= -\frac{\sum_{i=1}^n \|E(k)\| \|\hat{x}_i(k)\|}{\sum_{i=1}^n \|\hat{x}_i(k)\| \|O_j^{[3]}\|} = \frac{\sum_{i=1}^n \|E(k)\|}{\sum_{i=1}^n \|\hat{x}_i(k)\|} \times \frac{\|O_j^{[3]}\|}{\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k)} = \\
&= \frac{\sum_{i=1}^n e_i(k)}{\|O_j^{[3]}\|} \times \frac{\|O_j^{[3]}\|}{\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k)} = \frac{\sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k)}{\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k)},
\end{aligned}$$

можно получить окончательное выражение

$$d_j^{[3]}(k) = \frac{\|y_j^{[3]}(u_j^{[3]}(k))\|}{\|u_j^{[3]}\|} \frac{\sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k)}{\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k)}.$$

Тогда процедура (6.11) приобретает форму

$$Dw_{ji}^{[3]} = h^{[3]}(k) d_j^{[3]}(k) y_i(k) = h^{[3]}(k) y_i(k) \frac{\|y_j^{[3]}(u_j^{[3]}(k))\|}{\|u_j^{[3]}\|} \frac{\sum_{i=1}^n e_i(k) w_{ij}^{[4]}(k)}{\sum_{p=1}^h w_{ip}^{[4]}(k) O_p^{[3]}(k)},$$

или в векторном виде

$$Dw_j^{[3]} = h^{[3]}(k) d_j^{[3]}(k) y(k). \quad (6.12)$$

Свойства общей процедуры (6.12) определяются выбором параметра $h^{[3]}(k)$, от обоснованного выбора которого зависит время обучения в целом.

Введем в рассмотрение модификацию регуляризованного алгоритма обучения [221] вида

$$Dw_j^{[3]} = \frac{h^{[3]}(k) d_j^{[3]}(k) y(k) + r^{[3]} Dw_j^{[3]}(k-1)}{\|y(k)\|^2}, \quad (6.13)$$

где $h^{[3]} = \text{const} > 0$,

$0 \leq r < 1$ – параметр регуляризации.

Несложно заметить, что при $r^{[3]} = 0$, процедура приобретает форму нелинейной модификации мультипликативного алгоритма Качмажа, характеризующегося высоким быстродействием. Если же требуется дополнительная фильтрация можно от процедуры (6.13) перейти к процедуре общего вида

$$\begin{cases} \dot{D}w_j^{[3]} = (p^{[3]}(k))^{-1} (h^{[3]}(k) d_j^{[3]}(k) y(k) + W_3, \\ \dot{p}^{[3]}(k) = a p^{[3]}(k-1) + \|y(k)\|^2, \end{cases} \quad (6.14)$$

где $W_3 = r^{[3]} D_w^{[3]}(k-1)$, объединяющей в себе быстродействие алгоритма Качмажа-Уидроу-Хоффа, сглаживающие свойства алгоритма, и возможность проходить «плоские» участки целевой функции присущие регуляризованным алгоритмам с $r^{[3]} > 0$.

Аналогичным образом можно записать алгоритм обучения второго скрытого слоя в виде

$$Dw_{ji}^{[2]} = h^{[2]}(k) d_j^{[2]}(k) O_i^{[1]}(k),$$

где $d_j^{[2]}(k) = \sum_{i=1}^h d_i^{[3]}(k) w_{ij}^{[3]}(k)$.

Окончательно же

$$\begin{cases} \dot{D}w_j^{[2]} = (p^{[2]}(k))^{-1} (h^{[2]}(k) d_j^{[2]}(k) O^{[1]}(k) + W_2, \\ \dot{p}^{[2]}(k) = a p^{[2]}(k-1) + \|O^{[1]}(k)\|^2. \end{cases} \quad (6.15)$$

где $W_2 = r^{[2]} D_w^{[2]}(k-1)$.

И, наконец, для первого скрытого слоя справедливы соотношения

$$Dw_{ji}^{[1]} = h^{[1]}(k)d_j^{[1]}(k)x_i(k),$$

$$d_j^{[1]}(k) = \frac{\prod_{i=1}^h y_j^{[1]}(u_j^{[1]}(k))}{\prod_{i=1}^h u_j^{[1]}}$$

$$\begin{cases} Dw_j^{[1]} = (p^{[1]}(k))^{-1} (h^{[1]}(k)d_j^{[1]}(k)x(k) + W_1, \\ p^{[1]}(k) = ap^{[1]}(k-1) + \|x(k)\|^2. \end{cases} \quad (6.16)$$

где $W_1 = r^{[1]}Dw^{[1]}(k-1)$

Таким образом, процесс обучения нейронной сети, показанной на рис. 6.2, сводится к последовательному использованию процедур (6.10), (6.14), (6.15) и (6.16) для каждого предъявляемого вектора-образа $x(k)$. Для достаточно длинной обучающей выборки процесс настройки может быть реализован в реальном времени. Если же объем исходных данных ограничен, обучение может быть организовано по эпохам путем многократного «прогона» данных через нейронную сеть.

6.3 Анализ эффективности компрессии изображений для темпоральной обработки

Нам необходимо оценить, насколько предложенные нами нейросетевые методы сжатия изображений эффективны по сравнению с существующими методами сжатия и позволяют работать с видеопотоками. Для этого существует ряд формальных критериев, по которым можно сказать насколько тот или иной метод применим для того или иного класса изображений. В таком случае к основным количественным характеристикам по оценке качества изображений в целом относятся критерии описанные выражениями (6.17) – (6.29):

– средняя разность (AD)

$$AD = \frac{1}{H \times W} \sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y)); \quad (6.17)$$

– нормированная корреляция (normalized cross-correlation)

$$NK = \frac{\sum_{x=1}^H \sum_{y=1}^W f(x, y) \times \bar{f}(x, y)}{\sqrt{\sum_{x=1}^H \sum_{y=1}^W f(x, y)^2}}; \quad (6.18)$$

– качество корреляции (correlation quality)

$$CQ = \frac{\sum_{x=1}^H \sum_{y=1}^W f(x, y) \bar{f}(x, y)}{\sum_{x=1}^H \sum_{y=1}^W f(x, y)}; \quad (6.19)$$

– максимальная разность (maximum difference)

$$MD = \max \left(\left| f(x, y) - \bar{f}(x, y) \right| \right); \quad (6.20)$$

– верность изображения (image fidelity)

$$IF = 1 - \frac{\sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2}{\sum_{x=1}^H \sum_{y=1}^W f(x, y)^2}; \quad (6.21)$$

– среднеквадратичная лапласианова погрешность (laplasian mean square error)

$$LMSE = \frac{\sum_{x=2}^{H-1} \sum_{y=2}^{W-1} (O(f(x, y)) - O(\bar{f}(x, y)))^2}{\sum_{x=2}^{H-1} \sum_{y=2}^{W-1} O(f(x, y))^2}, \quad (6.22)$$

где $O(f(x, y)) = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y)$;

– среднеквадратичная погрешность (mean square error) или среднеквадратическое отклонение (СКО)

$$MSE = \frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2; \quad (6.23)$$

– максимальная среднеквадратичная погрешность (peak mean square error)

$$PMSE = \frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2 / (\max(f(x, y)))^2; \quad (6.24)$$

– нормированная абсолютная погрешность (normalized absolute error)

$$NAE = \frac{\sum_{x=1}^H \sum_{y=1}^W |f(x, y) - \bar{f}(x, y)|}{\sum_{x=1}^H \sum_{y=1}^W |f(x, y)|}; \quad (6.25)$$

– нормированная среднеквадратичная погрешность

$$NMSE = 1 - IF = \frac{\sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2}{\sum_{x=1}^H \sum_{y=1}^W f(x, y)^2}; \quad (6.26)$$

– норма L_p (Минковского)

$$L_p = \left(\frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^p \right)^{1/p}, \text{ для } p = 1, 2, 3, \dots; \quad (6.27)$$

– отношение сигнал/шум

$$SNR = 10 \log_{10} \left(\frac{\sum_{x=1}^H \sum_{y=1}^W f(x, y)^2}{\sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2} \right); \quad (6.28)$$

– максимум отношения сигнала/шум

$$PSNR = 10 \log_{10} \left(\frac{\sum_{x=1}^H \sum_{y=1}^W \max(f(x, y))^2}{\sum_{x=1}^H \sum_{y=1}^W (f(x, y) - \bar{f}(x, y))^2} \right). \quad (6.29)$$

Оценки качества, представленные выражениями (6.17) – (6.29) удобны в использовании, тем не менее, большинство из них не всегда позволяют объективно оценить качество изображения, в особенности это касается ситуации связанной с потерями при наличии шумов на исходных данных. Т.к. некоторые из представленных критериев в подобной ситуации выдают серьезные выбросы значений, говорящие о существенных изменениях на изображении, а в то же время для человеческого глаза они практически не различимы. В связи с этим следует отметить, что в большинстве работ на практике для оценки качества новых алгоритмов сжатия используются среднеквадратическое отклонение (MSE) (6.23) и пиковое отношение сигнал/шум (PSNR) (6.29). Учитывая же специфику анализа видеопотоков, PSNR является менее актуальным, т.к. при анализе видео влияние шумов незначительно. В связи с этим для анализа качества компрессии изображений, получаемого при помощи построенной нейронной сети был взят критерий, основанный на среднеквадратичной ошибке значений пикселей (mean square error – MSE).

Он подразумевает сравнение исходного изображения и восстановленного после сжатия. Поскольку предложенный метод сжатия следует отнести к сжатию с потерями, то естественно желание выяснить степень потерь. Однако, следует отметить, что данный критерий дает правильные результаты при отсутствии шума на изображении. При наличии же шума получим серьез-

езные ошибки для данного критерия. В такой ситуации лучше использовать критерий отношения пикового значения сигнала к шуму (PSNR — peak signal-to-noise ratio), который является более устойчивым к подобным воздействиям.

В результате сравнения исходных изображений и после применения предлагаемого метода сжатия, использующего представленные двухслойную и четырехслойную нейронные сети на основе критерия (6.23) получены результаты представленные на рис. 6.3.

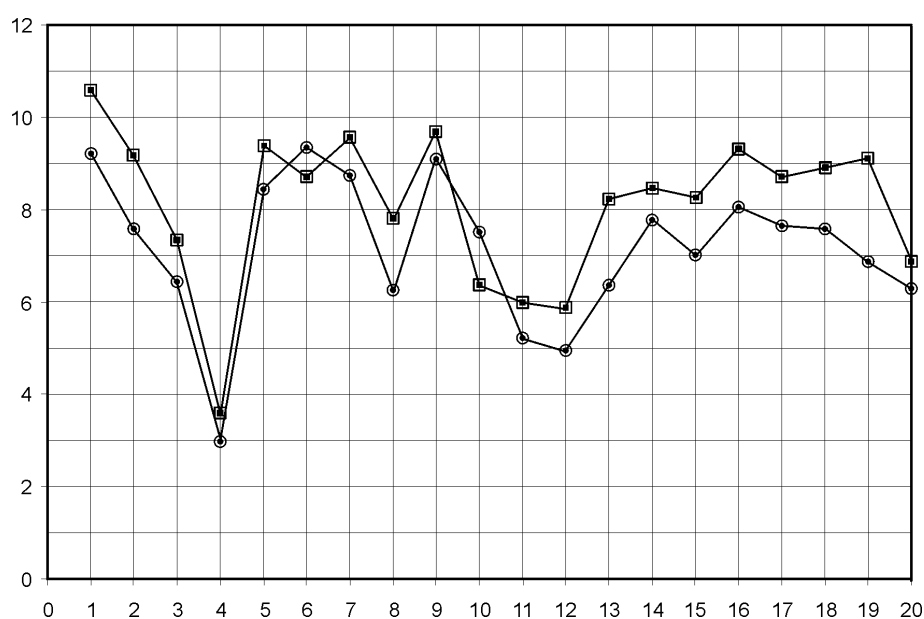


Рис. 6.3 – Коэффициенты сжатия для множества изображений и предложенных нейронных сетей

Кроме того, необходимо было провести анализ предложенных подходов на кадрах видеопоследовательностей. Результаты приведены в таблице 6.1. Для этого использовались 4 тестовых видеопоследовательности: съемка авиашоу, гонка формулы 1, велогонка и парусная регата. В таблице в первой колонке приведены по 2 кадра из каждой последовательности, во второй и третьей колонках степени сжатия для данных кадров при помощи. Под степенью компрессии понимаем отношение размеров исходного и сжатого изоб-

Таблица 6.1 – Сжатие кадров видеопоследовательностей

Исходные изображения	Адаптивное нейросетевое сжатие	Автоассоциативная многослойная нейронная сеть
	11.19	15.47
	9.54	12.40
	8.47	11.9
	5.48	9.14
	4.88	8.32
	6.35	10.03
	7.08	11.91
	5.05	8.89

ражений. Как видно из полученных результатов, 4-х слойная нейронная сеть дает более высокий коэффициент сжатия в то же время можно сделать вывод о том, что наличие большого количества разных объектов негативно влияет на степень сжатия обоих методов.

В тоже время надо отметить, что время, необходимое на сжатие при помощи автоассоциативной многослойной нейронной сети, значительно превышает аналогичный параметр для адаптивного нейросетевого сжатия, что, учитывая специфику обработки видео, заставляет применять оба метода в зависимости от поставленной задачи. С одной стороны, если требуется максимальное сохранение дискового пространства, но нет ограничений на время обработки (или они незначительно), то вполне естественно использование автоассоциативной многослойной нейронной сети. В условиях обработки видеопотоков это может быть связано с задачами, не требующими on line обработки, когда существует некоторое множество данных, которое просто необходимо обработать вне зависимости от требуемых ресурсов. В то же время вполне естественно, что сжимать все видеопоследовательности данными методами нет смысла, т.к. эти подходы значительно медленнее существующих алгоритмов компрессии видео и обладают меньшей степенью сжатия по сравнению с методами сжатия видео, т.к. они не учитывают корреляцию между кадрами.

Кроме рассмотренных подходов ключевые кадры сегментов исходных мультимедиа данных можно также подвергать обработке при помощи самоорганизующейся карты Кохонена для устранения возможных помех и улучшения качества этих изображений.

6.4 Использование самоорганизующихся карт Кохонена при обработке кадров видеопоследовательностей

Обработка изображений с помощью самоорганизующейся карты обычно производится следующим образом: входное изображение размером

$M \times N$ пикселей разбивается на блоки размером $m \times n$, каждый из которых преобразуется в векторную форму размерности $mn \times 1$. Затем с помощью обучающей выборки, содержащей $MN(mn)^{-1}$ векторов образов $x(k)$, $k = 1, 2, \dots, MN(mn)^{-1}$, пронормированных в принятой метрике, нейронная сеть настраивается с помощью принятого правила самообучения. При этом предполагается, что архитектура в слое Кохонена содержит p нейронов с синаптическими весами, заданными в форме $mn \times 1$ – векторов $w_q = (w_{q,1}, w_{q,2}, \dots, w_{q,mn})^T$, $q = 1, 2, \dots, p$. После замены каждого из предъявляемых векторов $x(k)$ на номер нейрона-победителя j , соответствующего этому образу, запоминаем номер j , соответствующий этому образу, и соответствующий ему вектор синаптических весов w_j^* . Для восстановления исходного изображения каждый номер нейрона-победителя j заменяется на w_j^* , девекторизованный к матричной форме. Понятно, что чем меньше отношение $p \cdot mn / MN$, тем выше эффективность сети.

В основе процедуры самоорганизации карты Кохонена лежат принципы конкурентного самообучения, а ее работа, как и любого другого алгоритма обучения, начинается с инициализации синаптических весов сети, которая обычно выбирается случайным образом. При этом весьма желательно, чтобы для каждого из нейронов выполнялось условие $\|w_q(0)\| = 1$. Сама же процедура самоорганизации выполняется в три основных этапа [230]: конкуренции, кооперации, синаптической адаптации.

Процесс конкуренции начинается с анализа образа $x(k)$, поступающего с рецепторного слоя на все нейроны слоя Кохонена. Для каждого из нейронов вычисляется его выходной сигнал

$$y_q(k) = w_q^T(k-1)x(k), q = 1, 2, \dots, p$$

и расстояние (как правило, в евклидовой метрике)

$$D(x(k), w_q(k-1)) = \|x(k) - w_q(k-1)\|,$$

причем если входы и синаптические веса предварительно пронормированы так, что $\|x(k)\| = \|w_q(k-1)\| = 1$, а в качестве расстояния используется традиционная евклидова метрика, то мерой близости между векторами $x(k)$ и $w_q(k-1)$ может служить также скалярное произведение

$$D(x(k), w_q(k-1)) = w_q^T(k-1)x(k) = \cos(w_q(k-1), x(k)) = y_q(k).$$

Далее определяется нейрон-победитель, ближайший ко входному образу такой, что

$$D(x(k), w_j^*(k-1)) = \min D(x(k), w_q(k-1)),$$

т.е.

$$\|x(k) - w_j^*(k-1)\| < \|x(k) - w_q(k-1)\|, \quad q \neq j,$$

после чего в простейшем случае, «перепрыгивая» через этап кооперации, можно подстроить синаптические веса сети с помощью правила конкурентного самообучения [230]

$$w_j(k) = \begin{cases} w_j(k-1) + h(k)(x(k) - w_j(k-1)), & \text{если } j\text{-нейрон победил;} \\ w_j(k-1) & \text{в противном случае.} \end{cases} \quad (6.26)$$

Процедура (6.26) реализует принцип «победитель получает все» (WTA), при этом вектор синаптических весов нейрона-победителя подтягивается ко входному образу на расстояние, определяемое шагом поиска $h(k)$.

Регулирование величины шага поиска обычно производится, исходя из эмпирических соображений [20], а общая рекомендация сводится к тому, что он должен монотонно уменьшаться в процессе самообучения. С этой рекомендацией хорошо согласуется процедура

$$h(k) = p^{-1}(k), \quad p(k) = ap(k-1) + \|x(k)\|^2, \quad 0 \leq a \leq 1,$$

которая для нормированных входов изображений простой вид

$$p(k) = ap(k-1) + 1,$$

а при $a = 1$ автоматически дает оценку

$$h(k) = \frac{1}{k},$$

удовлетворяющую условиям стохастической аппроксимации А. Дворецкого.

Можно заметить, что правило настройки (6.26) является обобщением на многомерный случай простейшего линейного абсолютно-оптимального алгоритма обучения [230], минимизирующего критерий

$$E^k = \frac{1}{k} \sum_{s=1}^k (x(s) - w)^2,$$

где $x(s)$ и w – скаляры.

С позиций же теории оптимизации, (6.26) есть не что иное, как гради-

ентная процедура стохастической аппроксимации, минимизирующая критерий

$$E_j^k = \frac{1}{k} \sum_{s=1}^k \|x(s) - w\|^2,$$

и ведущая, при $h(k) = \frac{1}{k}$, к векторной оценке среднего арифметического

$$w_j(k) = \frac{1}{k} \sum_{s=1}^k x(s),$$

при этом естественно предполагается, что j -й нейрон k раз оказывается победителем.

Этап же кооперации предполагает, что нейрон-победитель определяет так называемую локальную область топологического соседства, в которой обучается не только он сам, но и его ближайшее окружение, при этом более близкие к победителю нейроны активируются сильнее, чем удаленные. Эта топологическая область определяется функцией соседства $h_{jq}(k)$, зависящей от расстояния $D(w_j^*(k), w_q(k))$ между победителем $w_j^*(k)$ и любым из нейронов слоя Кохонена $w_q(k)$ и некоторого скалярного параметра \mathbf{s} , задающего ее «ширину». Как правило, $h_{jq}(k)$ – это потенциальная функция, симметричная относительно максимума в точке $w_j^*(k)$ ($D(w_j^*(k), w_j^*(k)) = 0$), принимающая в ней единичное значение и монотонно стремящаяся к нулю с ростом расстояния. Часто в качестве функции соседства используются гауссиан, треугольная функция, перевернутая парабола, «мексиканская шляпа» и целый ряд других [20].

Использование функции соседства приводит к правилу самообучения вида

$$w_q(k) = w_q(k-1) + h(k)h_{jq}(k-1)(x(k) - w_q(k-1)) \quad (6.27)$$

реализующему принцип «победитель получает больше» (WTA) вместо традиционного WTA. Естественно, что при $h_{jq}(k-1) = d_{jq} = 1$ при $j = q$ и 0 в остальных случаях, вновь приходим к стандартному алгоритму (6.26), обеспечивающему на каждом шаге настройку единственного нейрона победителя w_j^* . Использование же в качестве $h_{jq}(k)$ потенциальных функций ведет к тому, что все нейроны сети в большей или меньшей мере подтягивают векторы своих синаптических весов $w_q(k-1)$ к предъявляемому образу $x(k)$.

Можно показать, что алгоритм (6.27) минимизирует критерий

$$E_j^k = \mathop{\text{a}}\limits_{s=1}^k h_{jq}(s) \|x(s) - w_j\|^2$$

и сходится к оценке взвешенного среднего арифметического

$$w_j(k) = \frac{\mathop{\text{a}}\limits_{s=1}^k h_{jq}(s)x(s)}{\mathop{\text{a}}\limits_{s=1}^k h_{jq}(s)}.$$

Известно, что статистические оценки, связанные с квадратичными критериями и средними арифметическими, являются эффективными лишь в случае, когда наблюдения, принадлежащие конкретной группе (кластеру), подчинены нормальному закону распределения. Если же распределение данных в кластере отличается от гауссовского, а сами данные загрязнены аномальными выбросами (что достаточно часто бывает в задачах обработки изображений), эффективность оценок среднего арифметического быстро падает с «утяжелением» хвостов распределений. Ясно, что в этом случае алгоритмы (6.26) и (6.27) также становятся неэффективными.

В случае, когда в данных присутствуют наблюдения, резко выделяю-

щиеся на фоне остальных, целесообразно использовать критерии, менее реагирующие на выбросы. Таковым, например, является достаточно популярный критерий наименьших модулей

$$E_j^k = \mathop{\text{arg}}_{s=1}^k \mathop{\text{arg}}_{i=1}^{mn} |x_i(s) - w_{ji}|, \quad (6.28)$$

порождаемый манхэттенской метрикой

$$D(x(k), w_q(k-1)) = \mathop{\text{arg}}_{i=1}^{mn} |x_i(k) - w_{qi}(k-1)|.$$

Алгоритм обучения самоорганизующейся карты в этом случае приобретает вид [20]

$$w_j(k) = \begin{cases} w_j(k-1) + \mathbf{h}(k) \text{sign}(x(k) - w_j(k-1)), & \text{если } j\text{-нейрон победил} \\ w_j(k-1) & \text{в противном случае.} \end{cases} \quad (6.29)$$

где $\text{sign } x(k) = (\text{sign } x_1(k), \text{sign } x_2(k), \dots, \text{sign } x_{mn}(k))^T$.

Минимизация критерия (6.28) приводит к оценке, являющейся выборочной медианой, которая по сравнению с оценками, основанными на вычислении среднего, менее чувствительна к различного рода выбросам. Фактически с помощью алгоритма (6.29) производится не только кластеризация данных, но и их медианная фильтрация.

Известно, что медиана является эффективной оценкой математического ожидания случайной величины, когда данные распределены по закону Лапласа с функцией плотности распределения

$$f(x_i, w_i) = \frac{1}{2s_i} \exp \left\{ -\frac{|x_i - w_i|}{s_i} \right\},$$

где s_i – параметр ширины (масштаба) распределения по i -й координате, в общем случае неизвестный,

w_i – математическое ожидание, подлежащее оцениванию.

Оптимальная оценка математического ожидания в этом случае связана с минимизацией критерия

$$E_j^k = \sum_{s=1}^k \sum_{i=1}^{mn} \frac{1}{s_i} |x_i(s) - w_{ji}|$$

и может быть получена путем обобщения на многомерный случай абсолютно оптимального релейного алгоритма оценивания и его адаптации к форме (6.29), т.е.

$$w_j(k) = \begin{cases} w_j(k-1) + \frac{S}{k} \in \text{sign}(x(k) - w_j(k-1)), & \text{если } j\text{-нейрон победил;} \\ w_j(k-1) & \text{в противном случае.} \end{cases} \quad (6.30)$$

где $S = (s_1, s_2, \dots, s_{mn})^T$,

\in – символ прямого произведения.

Поскольку вектор S в (6.30) неизвестен, был ранее введен одномерный алгоритм настройки параметра масштаба, который применительно к рассматриваемому случаю приводит к обобщению процедуры (6.30) в форме

$$w_j(k) = \begin{cases} w_j(k-1) + \frac{S(k)}{k} \in \text{sign}(x(k) - w_j(k-1)), \\ S(k) = S(k-1) - \frac{1}{k} (S(k-1) - |x(k) - w_j(k-1)|), & \text{если } j\text{-нейрон победил,} \\ w_j(k-1) & \text{в противном случае.} \end{cases} \quad (6.31)$$

Переходя к этапу кооперации, следует заметить, что поскольку медиана является ранговой оценкой, для построения функции соседства целесообразно использовать также ранговые конструкции, к каковым, например, относится так называемый «нейронный газ».

В рамках этой концепции все нейроны карты Кохонена ранжируются в порядке возрастания расстояний так, что $D(x(k), w_j^*(k-1)) < D(x(k), w^1(k-1)) < \dots < D(x(k), w^{mn-1}(k-1))$, где верхний индекс обозначает не номер нейрона в сети, а его ранг в смысле расстояния от предъявленного образа $x(k)$, т.е.

$$R(w^*(k-1)) = 0 < R(w^1(k-1)) = 1 < \\ < R(w^2(k-1)) = 2 < \dots < R(w^{mn-1}(k-1)) = mn - 1.$$

Далее для каждого из нейронов определяются значение функции соседства

$$h_q(k-1) = \exp\left\{-\frac{R(w_q(k-1))}{s(k)}\right\} \quad (6.32)$$

(здесь $s(k)$ – параметр ширины) и производится уточнение синаптических весов сети согласно правилу

$$w_q(k) = w_q(k-1) + h(k)h_q(k-1)sign(x(k) - w_q(k-1)) \quad q = 1, 2, \dots, p,$$

являющемуся робастной модификацией алгоритма (6.27).

Несложно видеть, что параметры нейрона с нулевым рангом $w^*(k)$ (фактически нейрона-победителя) при этом уточняются согласно первому соотношению алгоритма (6.29).

Аналогично параметру шага $h(k)$ параметр ширины функции соседства (6.32) также должен монотонно уменьшаться в процессе самообучения, для чего ранее была предложена простая процедура, которая для рассматриваемого нами случая может быть записана в виде

$$\begin{aligned} h(k) &= h(0) \left(\frac{h(\min)}{h(\max)} \right)^{\frac{kmn}{MN}}, \\ s(k) &= s(0) \left(\frac{s(\min)}{s(\max)} \right)^{\frac{kmn}{MN}}, \end{aligned} \quad (6.33)$$

где $h(\min)$, $s(\min)$ и $h(\max)$, $s(\max)$ – минимально и максимально возможные значения соответствующих параметров.

Заметим, что соотношения (6.33) также удовлетворяют условиям А. Дворецкого.

Следует заметить, что в «чистом» виде ни гауссовское, ни лапласово распределение на практике не встречается, поэтому Я.З. Цыпкиным было рассмотрено приблизительно нормальное распределение, отличающееся от классического гауссовского наличием «тяжелых» хвостов. Функция плотности этого распределения имеет вид

$$f(x_i, w_i) = \frac{1}{2s_i} \exp\left(-\frac{x_i - w_i}{s_i}\right) \quad (6.34)$$

связанная с ним метрика –

$$D(x(k), w_q(k-1)) = \sum_{i=1}^{mn} s_i \ln \cosh \frac{x_i(k) - w_{qi}(k-1)}{s_i}, \quad (6.35)$$

а критерий обучения –

$$E_j^k = \sum_{s=1}^k \sum_{i=1}^{mn} s_i \ln \cosh \frac{x_i(k) - w_{ji}(k-1)}{s_i}.$$

Далее несложно записать алгоритм самообучения карты Кохонена на основе WTA-принципа в виде

$$w_j(k) = \begin{cases} w_j(k-1) + h(k) \tanh \frac{x(k) - w_j(k-1)}{s}, \\ \text{если } j\text{-нейрон победил,} \\ w_j(k-1) \text{ в противном случае,} \end{cases} \quad (6.36)$$

$$\text{где } \tanh \frac{x(k) - w_j(k-1)}{s} = \left(\tanh \frac{x_1(k) - w_{j1}(k-1)}{s_1}, \dots, \tanh \frac{x_{mn}(k) - w_{jmn}(k-1)}{s_{mn}} \right)^T,$$

а параметры s_i в данном случае определяют крутизну функции гиперболического тангенса, при этом, чем меньше значение s_i , тем ближе алгоритм (6.36) по свойствам к медианной процедуре (6.29). С ростом же s_i (6.36) приближается по своим свойствам к стандартному правилу (6.26).

Что же касается выбора функции соседства, резонно предположить, что если для стандартной процедуры (6.27) в качестве таковой наиболее часто используется гауссиан, то в данном случае с таким же успехом может быть применена функция (6.34) в виде

$$h_q(k-1) = \frac{1}{2s_i} \operatorname{sech}^2 D(w_j^*(k-1), w_q(k-1)),$$

где $D(\times)$ определяется выражением (6.35).

Тогда комбинированный робастный алгоритм самообучения карты Кохонена на основе WTM-принципа окончательно может быть записан в виде

$$w_q(k) = w_q(k-1) + h(k) h_q(k-1) \tanh \frac{x(k) - w_q(k-1)}{s}, \quad q = 1, 2, \dots, p. \quad (6.37)$$

Алгоритмы (6.36) и (6.37) в зависимости от принятого вектора параметров s могут обладать свойствами как традиционных правил Кохонена (6.26), (6.27) так и их медианных модификаций (6.29), (6.30), обеспечивая

устойчивую обработку информации при наличии различного рода возмущений и помех, что может быть достаточно удобным инструментом повышения эффективности обработки кадров видеопоследовательностей. Особенно это важно в тех случаях, когда получение видеоданных затруднено произвольными сложными условиями съемки (например, наличие метеорологических помех), или ограничено низким разрешением съемки (например, web-камеры), или повреждением исходных данных при передаче по каналам связи, вызвавших появление помех. Наличие подобных искажений может существенно влиять как на методы выделения однородных сегментов видео, так и на поиск ключевой информации в них, а, следовательно, привести к ухудшению валидности информационного поиска при запросах «по образцу».

Выводы шестому по разделу

1. Фрагментное сжатие изображений в последовательном режиме обработки информации с учетом ресурсоемкости процессов кодирования-декодирования, степени сжатия, потери качества и т.п. может быть получено на основе использования нейросетевых технологий, реализующих парадигму самообучения.

2. Адаптивная нейросетевая компрессия видеопотоков на основе взвешенного информационного критерия путем поиска главных компонент и доминантных собственных векторов корреляционных матриц сохраняет валидность сегментации, что дает возможность выполнять темпоральную обработку видео с существенной экономией памяти.

3. Процедура самообучения автоассоциативной нейронной сети позволяет строить симметричные подпространства в режиме реального времени со снижением размерности синаптических весов нейронной сети за счет фрагментной темпоральной обработки.

РАЗДЕЛ 7

ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ МЕТОДОВ И МОДЕЛЕЙ ТЕМПОРАЛЬНОЙ ОБРАБОТКИ ВИДЕО ДЛЯ ИНФОРМАЦИОННОГО ПОИСКА

Полученные теоретические результаты представляют собой фундамент информационной технологии контекстного поиска видеоданных с запросом «по образцу». Основным моментом является структуризации данных, обеспечиваемая темпоральной обработкой видеорядов в последовательном (по мере «наблюдения») режиме. Сама парадигма поиска в видеокolleкциях порождает разбиение темпоральной обработки на две фазы – on line и off line. Хотя в обоих случаях обработка и анализ видеопоследовательностей выполняются последовательно кадр за кадром, off line фаза не имеет принципиальных временных ограничений и дополнительно требует решения целого ряда задач по обслуживанию видеокolleкций. Перечень задач и подходов к их решению иллюстрируются рис. 1.2 и рис. 1.3.

Информационные технологии поиска, реализуя весьма сложный теоретический инструментарий, достаточно прозрачны концептуально. On line фаза заключается в структуризации (сегментации) видеопотока-запроса. Затем выделяются ключевые кадры сегментов, выбирается конфигурация из видеокolleкций, опорные точки – ключевые кадры. Далее после выбора критерия поиска на основе областей элиминации с использованием результатов off line фазы выполняется быстрый метрический поиск. В off line фазе структуризации видеопотоков проводится для пополнения видеокolleкций, а вместо процедур поиска выполняется подготовка конфигураций и индексация видеопоследовательностей (ключевых кадров). Практическое использование информационной технологии поиска требует экспериментального анализа специфики темпоральной обработки, чему и посвящен настоящий раздел.

Раздел основан на работах автора [14, 16, 61, 69, 74, 85, 87, 88, 90, 93, 94, 157, 200, 217, 254, 255, 257, 261, 262].

7.1 Особенности экспериментального анализа временных рядов, индуцированных видеопотоками

В целом любой временной ряд должен обладать двумя неотъемлемыми характеристиками: временем, на котором он рассматривается, с некоторыми точками отсчета и значениями ряда. В нашем случае под отсчетом времени следует понимать кадр видеопоследовательности, а под значением (уровнем ряда) – те характеристики кадра, которые удалось выделить предложенными методами анализа видеопоследовательностей. Учитывая такую постановку, можно утверждать, что ряд является моментным, т.е. уровень ряда характеризует изучаемое явление в конкретный момент времени. Также ряд является полным, т.к. интервалы подсчета стабильны и неизменны (кадр). Следует отметить, что существующая проблема анализа рядов, связанная с сопоставимостью значений актуальна для анализа результатов сегментаций видео только в плане кросс-корреляций между различными предлагаемыми подходами, т.к. они могут давать разную динамику изменения значений уровня, в тоже время когда рассматривается один метод данная проблема не стоит, т.к. для выделения характеристик каждого кадра используется одинаковый подход.

Исходя из логики задачи сегментации видеопоследовательностей, можно предполагать, что в результате каждый сегмент последовательности должен обладать неким трендом, под которым можно подразумевать и некий ключевой кадр сегмента. При этом можно вести речь только о трендах в сегментах, т.к. переход от сегмента к сегменту никоим образом не предсказуем с точки зрения выделения тренда всей видеопоследовательности. Таким образом, имеем некоторое множество трендов, мощность которого напрямую зависит от количества сцен в исходных данных. Ту же логику можно применить и при определении стационарности видеопоследовательности. Опять же, внутри сегмента ряд должен быть стационарен, в то время как между сегментами говорить об одинаковом математическом ожидании значений ха-

рактических, конечно же, не приходится. В целом можно говорить о периодизации динамики, где период ограничивается рамками сегмента.

Для проверки стационарности временных рядов используют тест Дики-Фуллера (DF). Поскольку временной ряд имеет единичный корень, или порядок интеграции один, если его первые разности образуют стационарный ряд, то при помощи этого теста проверяют значение коэффициента a в авторегрессионном уравнении первого порядка AR(1)

$$x_t = ax_{t-1} + \epsilon_t, \quad (7.1)$$

где x_t – временной ряд,

ϵ – ошибка.

Если в результате $a = 1$, то в этом случае ряд x_t не стационарен. В случае, когда $|a| < 1$, то ряд является стационарным. Сущность теста сводится к следующему: можно преобразовать уравнение (7.1) к виду:

$$Dx_t = bx_{t-1} + \epsilon_t,$$

где $b = a - 1$,

$Dx_t = x_t - x_{t-1}$ – оператор разности.

В зависимости от модели различают три теста:

1. Без константы и тренда: $Dx_t = bx_{t-1} + \epsilon_t$.
2. С константой, но без тренда: $Dx_t = b_0 + bx_{t-1} + \epsilon_t$.
3. С константой и линейным трендом: $Dx_t = b_0 + b_1t + bx_{t-1} + \epsilon_t$.

И для каждого из тестов существуют свои критические значения DF-статистики, которые берутся из специальной таблицы Дики-Фуллера (Мак-Киннона). Если значение статистики лежит левее критического значения (критические значения – отрицательные) при данном уровне значимости, то нулевая гипотеза о единичном корне отклоняется и процесс признается ста-

ционарным (в смысле данного теста). В противном случае гипотеза не отвергается и процесс может содержать единичные корни, то есть быть нестационарным (интегрированным) временным рядом.

В случае, когда процесс может быть авторегрессией более высокого порядка, для проверки стационарности используют так называемый расширенный тест Дики-Фуллера (ADF). В этом случае все сводится к уравнению вида

$$Dx_t = (a_1 + a_2 - 1)x_{t-1} - a_2 Dx_{t-1} + e_t$$

Для приведения ряда к стационарному виду в классических подходах используют алгоритмы сглаживания. Это обосновывается тем, что данные подходы позволяют выделить и удалить так называемую сезонную составляющую временного ряда, что соответственно может помочь привести ряд к стационарному виду.

Для экспериментального анализа предложенных подходов по сегментации видео использовались видеотрекеры спортивного содержания длительностью 1500 кадров или иными словами длительностью 1 минута каждый, которые включали в себя:

1. Фрагменты этапа велогонки Тур де Франс. (видео «Тур де Франс»).
2. Видеотрекер полетов самолетов на авиашоу. (видео «Авиашоу»).
3. Фрагменты гонки этапа Формулы 1. (видео «Формула 1»).
4. Фрагменты соревнований по парусному спорту. (видео «Парусный спорт»).

На рисунке 7.1 представлены 11-е кадры каждой из этих видеопоследовательностей и их пространственные сегментации для иллюстрации содержимого многомерных рядов. Результаты DF теста для этих видеотрекеров после получения временных рядов при помощи сегментации видео, описанной в разделе 4 представлены в таблице 7.1.



Рис. 7.1 – Кадры тестовых видеопоследовательностей

Таблица 7.1 – Тест Дики-Фулера

Тип видео	Var модель	Нейронная сеть	Настраиваемая модель
1. Тур де Франс	-4.92	-4.58	-3.97
2. Авиашоу	-5.23	-4.88	-4.25
3. Формула 1	-4,71	-4,64	-4,05
4. Парусный спорт	-4,18	-4,35	-4,18

Из таблицы 7.1 видно, что гипотеза о стационарности всего видеофрагмента не соблюдается ни для одного из исходных видеоданных. Хотя для модели со сглаживанием результаты были близки критическим, но все же превысили данные значения. Таким образом, в результате можно сделать вывод о том, что для обработки видеопоследовательностей нужно отталкиваться от того факта, что временные ряды, полученные на их основе будут носить нестационарный характер.

При этом надо понимать, что вид получаемых одномерных рядов колеблется в зависимости от выбранной модели. На рисунках 7.2 – 7.4 приведены примеры анализируемых одномерных рядов одной и той же видеопоследовательности длительностью 550 кадров, на основании единой пространственной сегментации и при аналогичном выборе вектора характеристик. Так на рисунке 7.2 приведен пример, полученный на основании VAR- модели. На следующем графике (рис. 7.3) приведен результат работы нейронной сети.

Как можно увидеть при сравнении, в этом случае модель более чувствительна к изменениям, что выражается в более скачкообразных изменениях ряда. Хотя общий тренд сохраняется в обоих случаях. Вполне логичным выглядит то, что на более коротких выборках предпочтительным будет выглядеть вто-

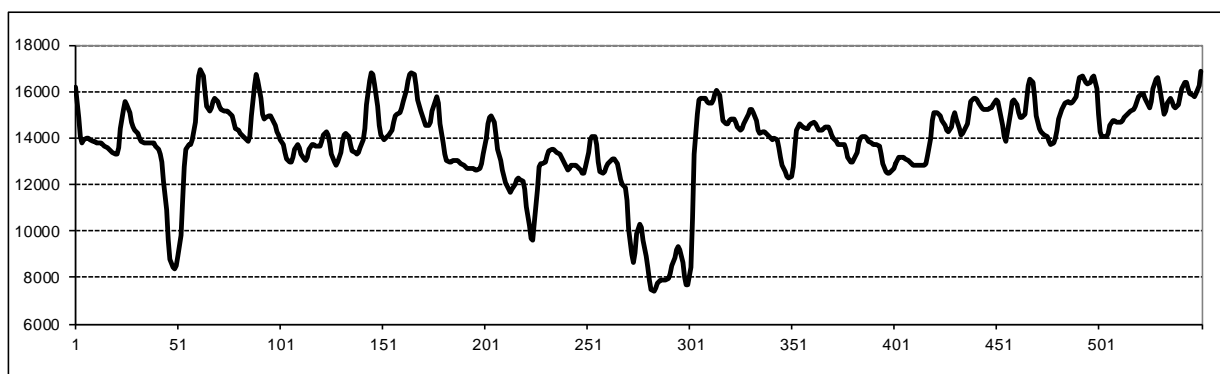


Рисунок 7.2 – Пример ряда, полученный с помощью VAR-модели

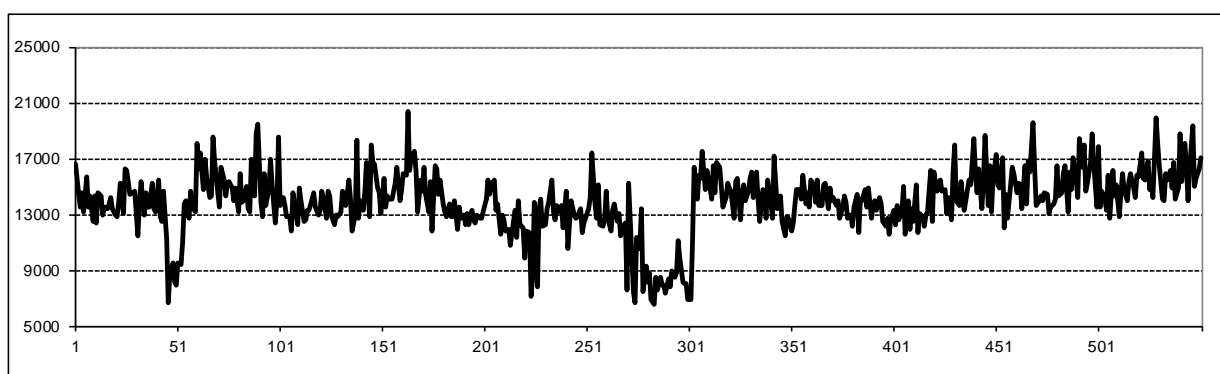


Рисунок 7.3 – Пример ряда, полученный с помощью нейронной сети

рой вариант, т.к. при этом снижается вероятность пропустить изменения ряда. Иначе говоря, данный подход применим для более «тонкого» анализа.

Если же нас интересуют изменения, происходящие на больших интервалах, то первая модель предпочтительней, т.к. в этом случае отражаются в основном «грубые» изменения без учета возможных выбросов значений на малых интервалах времени. На рисунке 7.4 рассмотрена настраиваемая модель для той же видеопоследовательности. В этом случае, если изменения от-

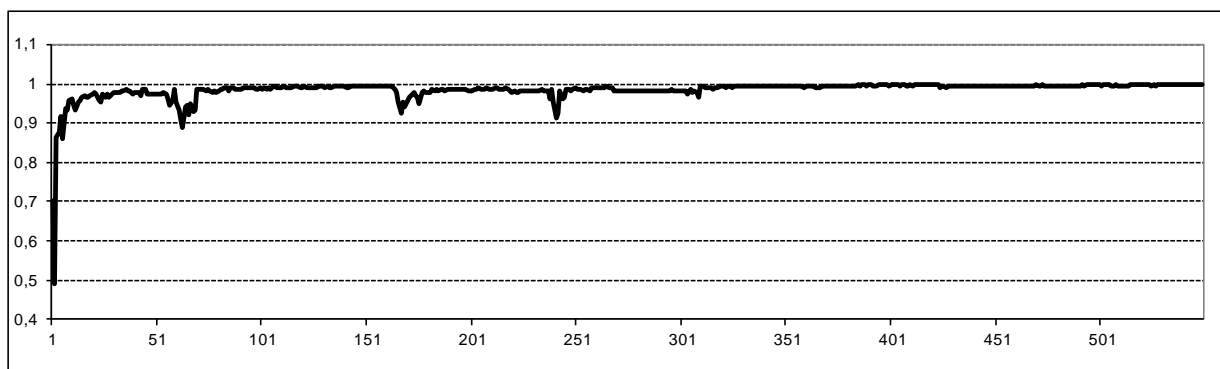


Рисунок 7.4 – Пример ряда, полученный с помощью настраиваемой модели

сутствуют, то линия графика стремится к 1. При возникновении разладок во временном ряду происходят аномальные выбросы значений. При этом на приведенной иллюстрации четко наблюдается один из недостатков данного подхода: изменения регистрируются с запаздыванием в несколько кадров, которое необходимо для перенастройки модели. Этот момент четко наблюдается в первых нескольких кадрах видеопоследовательности, когда выборка данных минимальна и настройка модели затруднена. Если исключить из рассмотрения всего 2 начальных значения, то информативность представления результата значительно улучшается, как показано на рис. 7.5.

Если сравнить все три подхода, то видно, что в каждом случае регистрируются разладки в видео на определенных интервалах. Во-первых, это интервал от 40 по 70 кадры, а также неоднократные колебания значений характеристик на интервале от 150 до 300 кадра. Зарегистрировано как максимальное, так и минимальное значение параметров в первых двух подходах, и есть четкое стабильное отклонение в случае настраиваемой модели.

Исходя из приведенных экспериментальных результатов работы методов сегментации видеопоследовательностей на одинаковых входных данных, можно утверждать, что каждый из этих подходов регистрирует происходящие изменения в видеопоследовательности, что в свою очередь позволяет выделить границы, а соответственно решить задачу сегментации, т.е. найти разбиения всей последовательности данных на однородные сегменты.

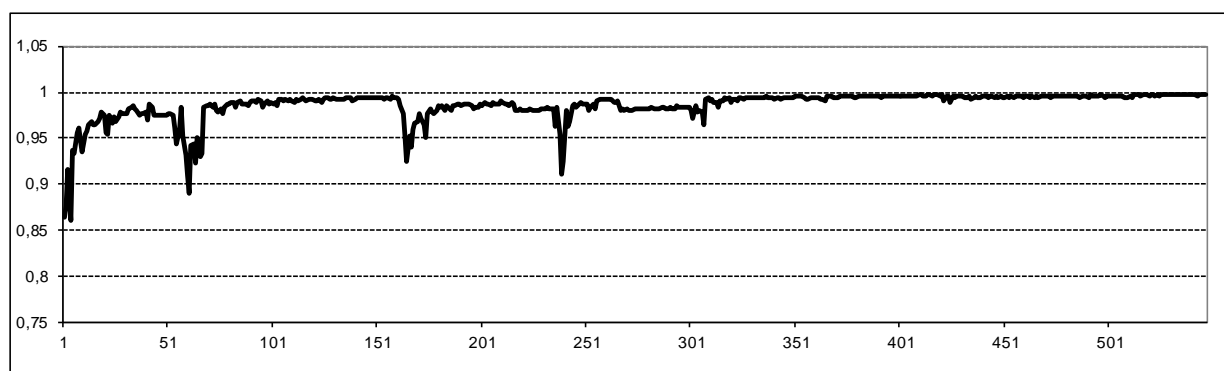


Рисунок 7.5 – Пример ряда, полученный с помощью настраиваемой модели

Однако остается открытым вопрос выбора вектора характеристик из описанных в разделе 3 для наиболее качественного разбиения исходных видеопоследовательностей. На рисунке 7.6 приведены 3 различных варианта выбора вектора параметров. При этом они были нормированы с тем, чтобы значения были сравнимы. Как видно из результатов первый и второй варианты векторов характеристик дают неоднозначные результаты, по которым сложно определить наличие разладок в последовательности видеоданных. Третий же вариант дает сходные результаты с тем вектором, который был выбран для проведения экспериментов. Следует отметить, что в первом и втором случаях, в отличие от третьего варианта, в качестве одного из исключяющихся из рассмотрения признаков была выбрана площадь сегментов пространственной сегментации кадров. Исходя из этого, можно утверждать, что данная характеристика является одной из необходимых для вычислений при построении характеристического вектора кадров видеопоследовательности.

Учитывая, что обработка видеопоследовательностей является достаточно ресурсоемкой процедурой, требующей значительных временных затрат, вполне логичными выглядят попытки сократить ресурсные затраты. Одним из интуитивно понятных подходов является прореживание, т.е. рассмотрение не каждого кадра видеопоследовательности, а через определенные промежутки. Таким образом, с одной стороны можно значительно уменьшить время необходимое на обработку данных, в частности, один из самых

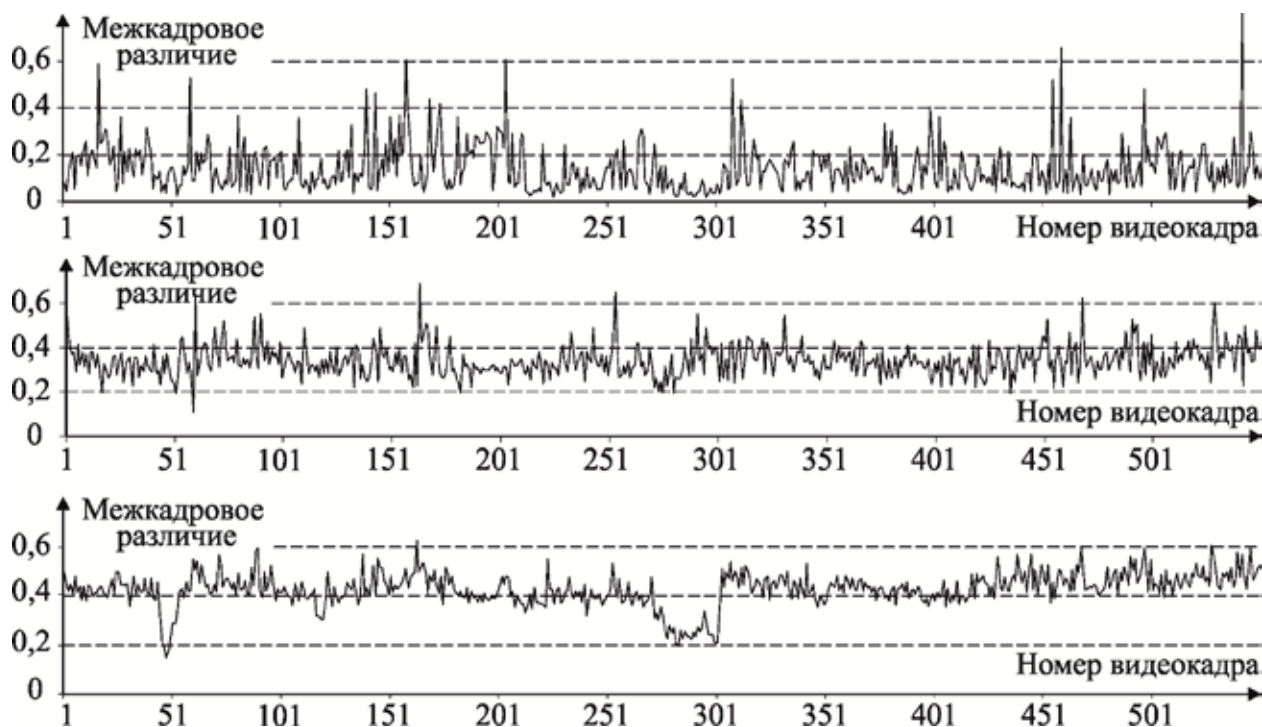


Рисунок 7.6 – Зависимость результатов от выбора вектора характеристик

затратных по времени этапов – пространственную сегментацию кадров. С другой стороны, в результате такого прореживания видеопоследовательности могут произойти потери информации об изменениях структуры временного ряда. На рис. 7.7 приведен пример использования прореживания на том же видеофрагменте. При этом наблюдаются как позитивные моменты, связанные с меньшим случайным разбросом данных в соседних кадрах, так и негативные, в той ситуации, когда выбранный изначально являлся аномальным выбросом. Например, изменения, происходившие около 300 кадра, стали менее явными по сравнению с анализом всей последовательности.

Рассмотрим результаты сегментации тестовых видеопоследовательностей. При визуальном рассмотрении каждого исходных данных люди стабильно выделяют границы сегментов. На рисунке 7.8 приведены примеры перехода от одного сегмента к другому каждой из тестовых видеопоследовательностей. При этом иногда возникают ситуации и с неявными переходами, которые могут приводить к проблемам при детектировании изменений параметров временного ряда. Такая ситуация может возникать при наличии раз-

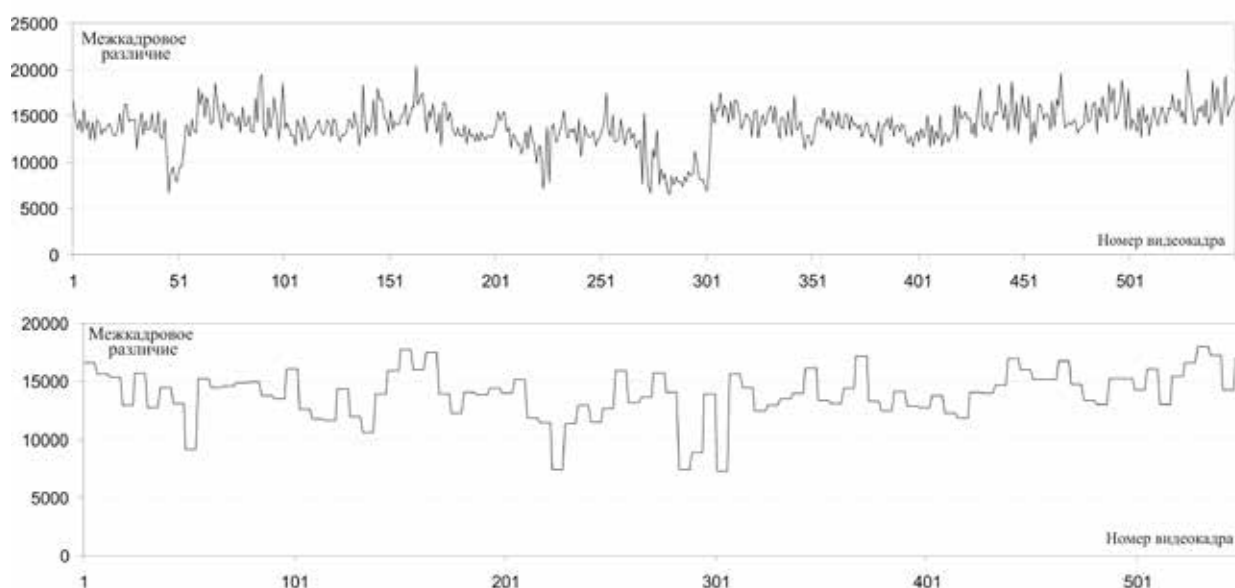


Рисунок 7.7 – Использование всей последовательности и каждого 5-го кадра

личных эффектов в видеоданных, которые призваны сгладить визуальные переходы для зрителя, но являются скорее негативным фактором при математическом анализе подобных последовательностей. Пример подобного перехода, соответствующий толерантности, показан на примере кадров из видеопоследовательности отражающей фрагменты этапа гонок «Формулы 1».

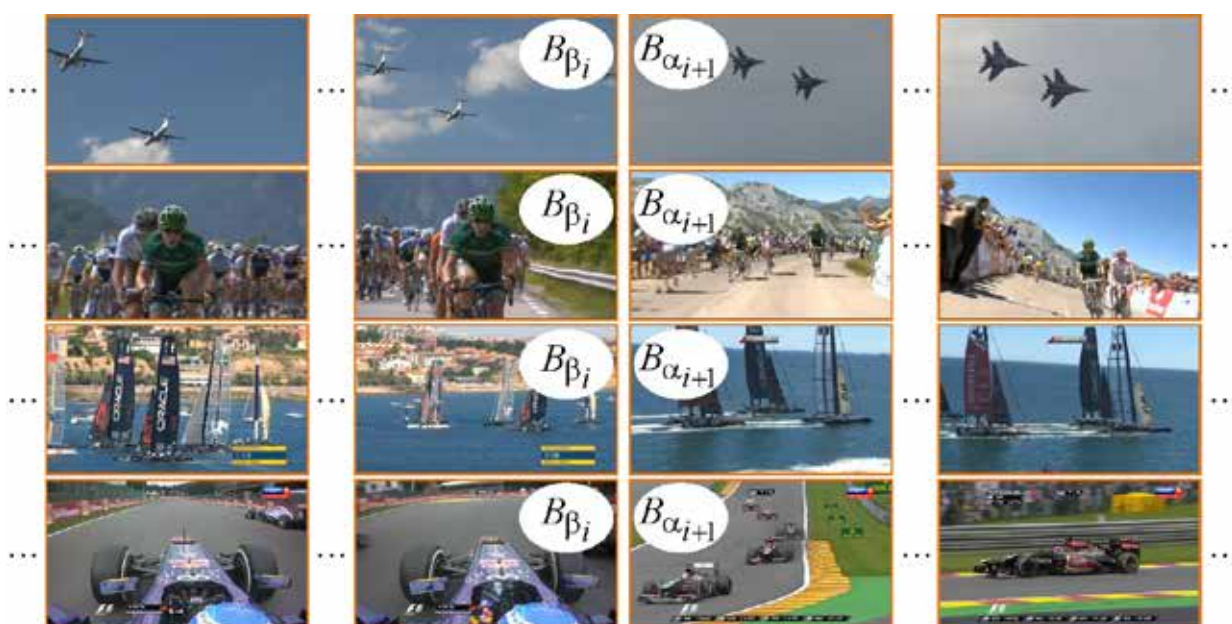


Рисунок 7.8 – Примеры границ между сегментами в исходных видеорядах

На рис. 7.9 показаны 3 кадра из последовательности перехода от одного сегмента к другому, где на каждом из кадров присутствуют элементы 2 различных сегментов видеоданных. Вполне логично, что выявление подобных переходов усложняется именно «благодаря» наличию характеристик, принадлежащих разным группам данных.



Рисунок 7.9 – Пример кадров с элементами разных сегментов

Следует отметить тот факт, что в зависимости от сложности фона можно получать более явные показатели переходов между сегментами. Так, для съемок авиашоу, где в качестве фона выступает небо, переход от одного сегмента видео к другому выделяется более отчетливо (рис. 7.10), чем для последовательности, отражающей перипетии этапа велогонки, на которой зачастую существует ситуация в которой скорее создается впечатление движения не объекта, а фона, что достигается ввиду того, что камера не является стационарной и основное ее внимание направлено на спортсменов.

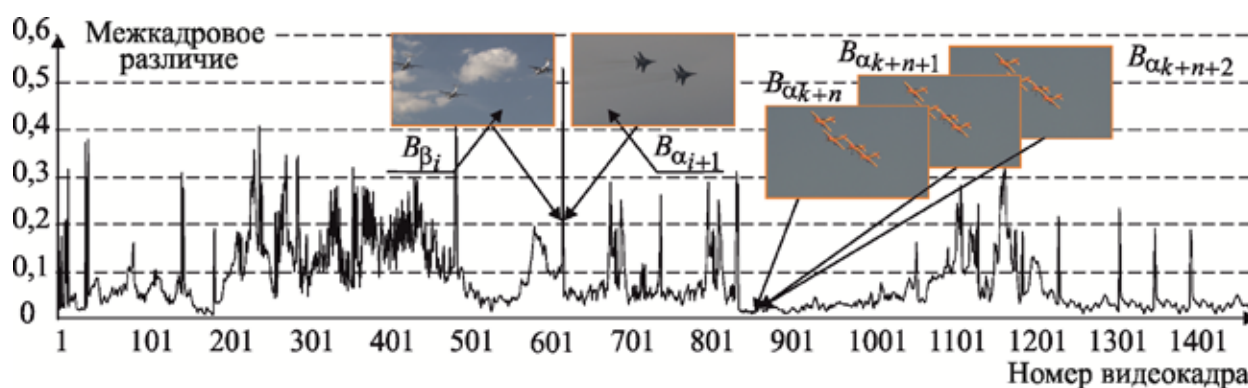


Рисунок 7.10 – Сегментация видеопоследовательности «Авиашоу»

При этом фон является далеко не однородным, на котором могут при-

существовать свои движущиеся элементы, например, болельщики, что также влияет на детектирование границ сегментов. Сегментация такой тестовой видеопоследовательности представлена на рисунке 7.11. Из приведенного временного ряда создается впечатление, что изменения структуры происходят практически постоянно, но на самом деле некоторые из всплесков являются ложными, поэтому для подобного рода последовательностей скорее подойдет подход, менее ярко реагирующий на малейшие изменения структуры временного ряда. В частности, вместо нейронной сети можно было бы использовать VAR модель подобно тому, как это было сделано на рис. 7.2.

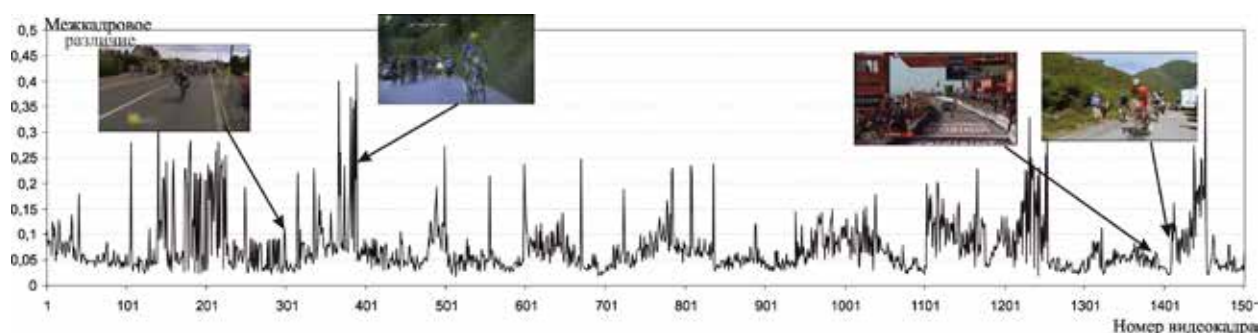


Рисунок 7.11 –Сегментация видеопоследовательности «Тур де Франс»

Следует отметить, что в рассматриваемых сюжетах «Авиашоу» и «Тур де Франс» в интерактивном режиме устойчиво выделяются 19 и 16 сегментов соответственно. Для видеопотока «Авиашоу» имеем следующие выделенные сегменты, совпадающие с сегментами, указанными интерактивно: [1, 19], [20, 80], [81, 142], [143, 199], [200, 280], [281, 311], [312, 354], [355, 389], [390, 433], [434, 520], [521, 603], [604, 651], [652, 725], [726, 807], [808, 875], [876, 951], [952, 1237], [1238, 1414], [1415, 1500]. Для видеопотока «Тур де Франс» – [1, 316], [317, 457], [458, 538], [539, 576], [577, 624], [625, 679], [680, 734], [735, 823], [824, 869], [870, 924], [925, 967], [968, 1007], [1008, 1063], [1064, 1114], [1115, 1148], [1149, 1500].

До этих пор мы строили сегментации видеопоследовательностей на основе одной и той же пространственной сегментации кадров. При этом был

выбран алгоритм текстурной сегментации JSeg. Однако, как и у большинства других методов автоматической сегментации изображений, он может выполняться с различными параметрами, которые будут влиять на количество и характеристики сегментов, на которые будет производиться разбиение отдельных кадров.

Для того чтобы выяснить насколько влияет изменение параметров сегментации изображений на результат сегментации видеопоследовательностей для тестовых видеопоследовательностей были построены временные ряды на основании нейронной сети, в которую подавались в качестве исходных данных многомерные характеристические векторы, полученные на основании различных сегментаций кадров. В результате на рисунке 7.12 представлено 4 варианта разложения видеопоследовательности «Авиашоу» с различными параметрами сегментации алгоритмом JSeg.

Как видно из полученных последовательностей, различные параметры сегментации методом JSeg влияют только на общие значения характеристик кадра, т.е. числовые значения рядов различны. Но в тоже время влияние на общий тренд ряда незначительно, было незначительным.

Таким образом, в результате анализа видео можно сделать предварительный вывод о целесообразности использования моделей анализа многомерных временных рядов, описанных в разделе 4, с целью пространственно-временной сегментации видео.

К особенностям «спортивного» видео стоит отнести тот факт, что съемки подобных роликов обычно проводятся множеством видеокамер в онлайн режиме с постоянной сменой ракурса телетрансляции для выделения наиболее интересной для зрителя информации. В результате в рамках одного видеоролика возможны абсолютно различные по съемки сюжетов, что может быть связано, как с различным местонахождением камер относительно центра событий, что приводит к изображениям одних и тех же действий в разном ракурсе и масштабе, например, видео с бортовой камеры болида в «Формуле 1» и камеры, размещенной за пределами трека; так и возможным

нестационарным размещением видеокамеры, например, камера – на мотоцикле во время велогонок.

Абсолютно иными особенностями обладают художественные фильмы, в которых смена сцены диктуется сюжетом и никоим образом не связана с изменениями, которые могут происходить в реальном времени. Соответственно в таких случаях возможен как факт движения объекта в поле зрения с фиксированным фоном, так и изменение фона, например, удаление или приближение с фактически неподвижным объектом.

Для экспериментальной проверки сегментации данных подобного типа был использован ролик из художественного фильма «Истребители», т.к. по своему контекстному содержанию (визуальному восприятию) его можно сравнивать с роликом «Авиашоу». Длительность этого видео также составляла 1 минуту (1500 кадров). Пример темпоральной зависимости от различных пространственных сегментаций для «Авиашоу» представлен на рисунке 7.12, а для ролика из художественного фильма «Истребители» на рисунке 7.13.

Пространственные сегментации видеокадров из этого ролика иллюстрируются рисунком 7.14.

Как видно из полученных результатов, наблюдаются аналогичные предыдущим случаям ситуации, т.е. детектирование изменений в видеопотоке выполняется достаточно валидно, что также позволяет разбить его на множество однородных сегментов, а зависимость от параметров сегментации кадров хоть и существует, однако не носит критического характера.

В тоже время интересной является возможность проверить зависимость от параметров пространственной на коротких выборках. Для этого исследуем последовательность в 50 кадров. Результат этой пространственно-временной сегментации представлен на рис. 7.15.

В данном случае зависимость уже выше, например, для одного из графиков существует всего один всплеск, в отличие от двух других, хотя общий вид графиков все же остается схожим и всплеск происходит на тех же кадрах последовательности.

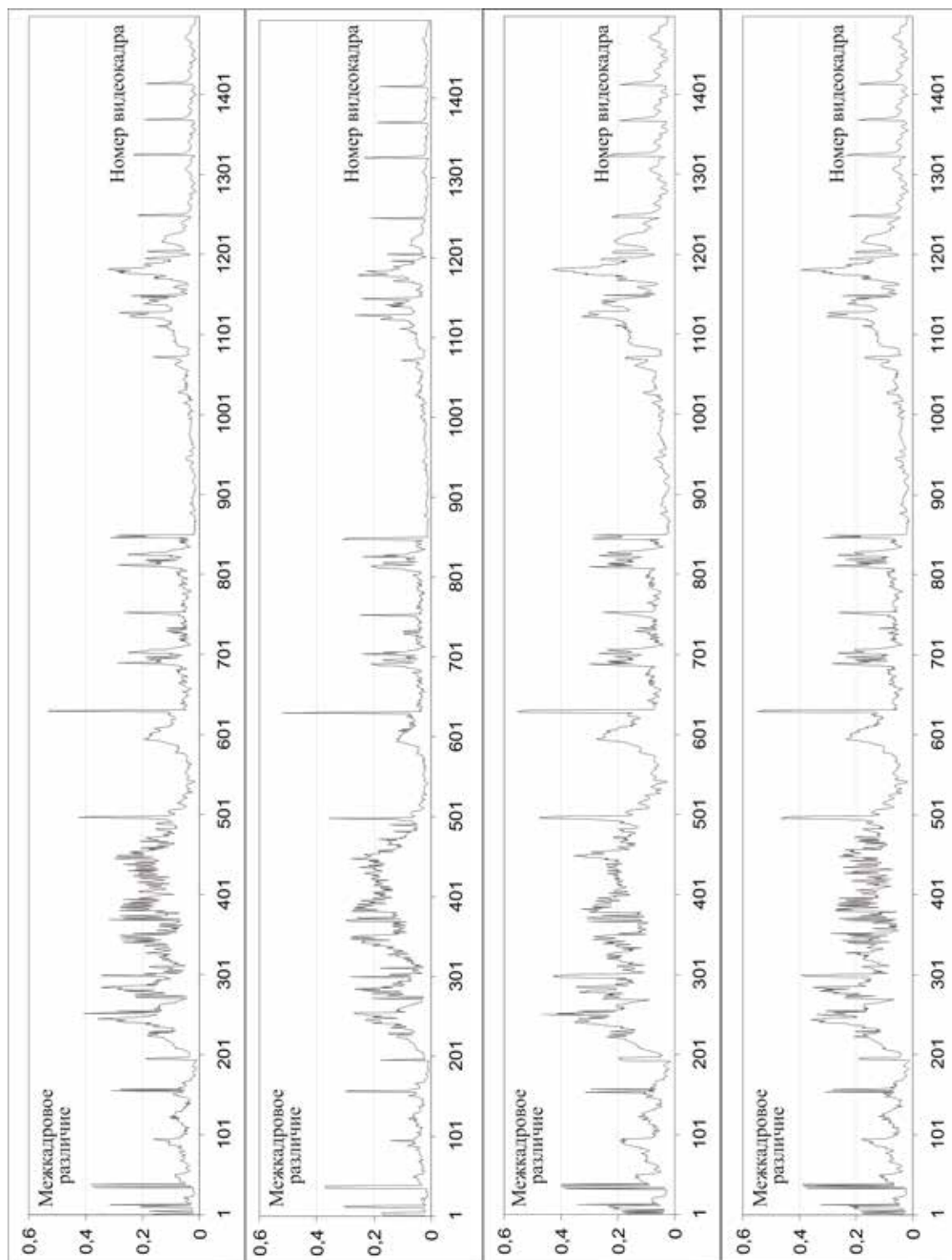


Рисунок 7.12 – Зависимость сегментации видео от параметров сегментации отдельных кадров в ролике «Авиашоу»

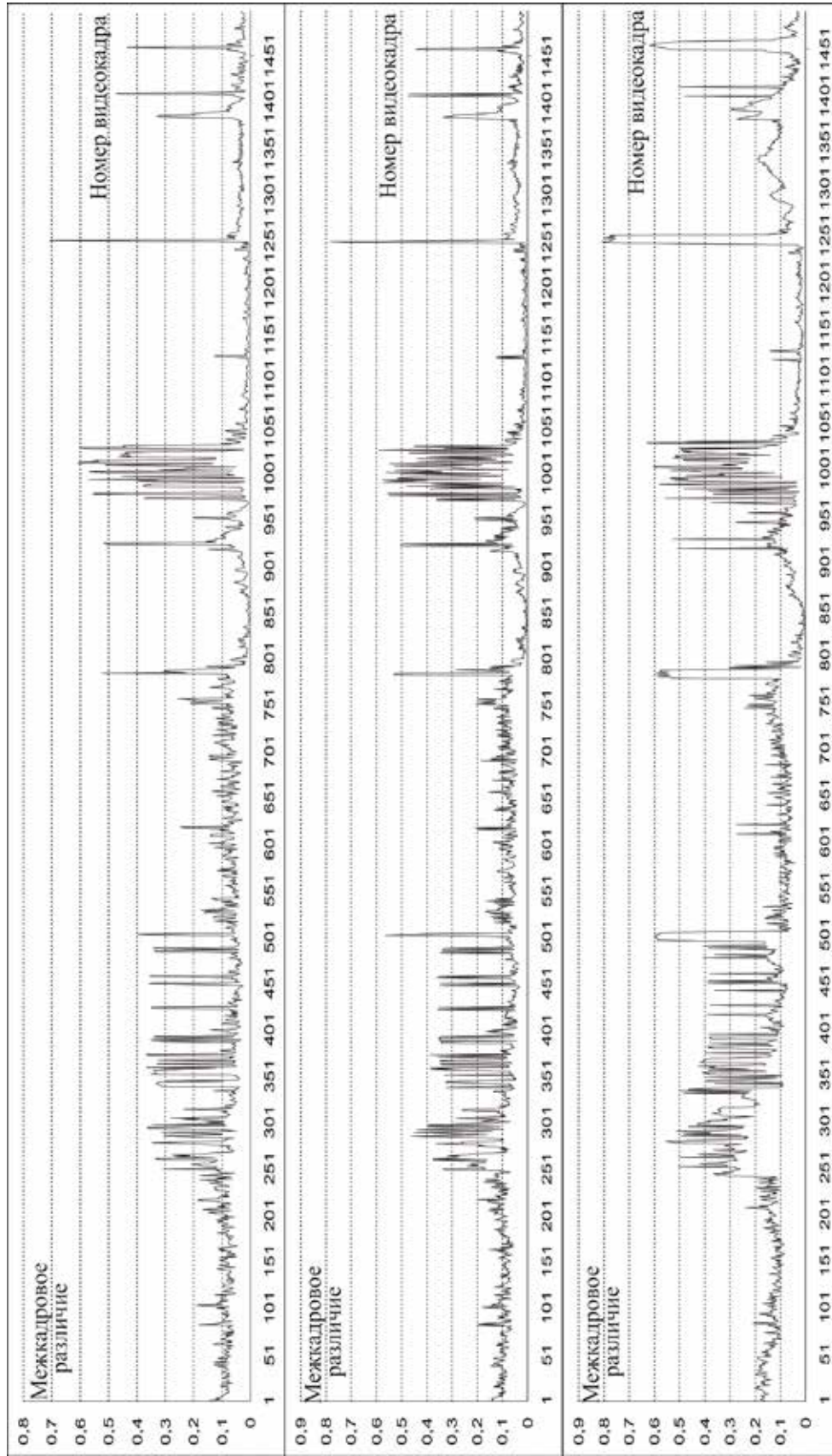


Рисунок 7.13 – Зависимость сегментации видео от параметров сегментации отдельных кадров в ролике «Истребители»

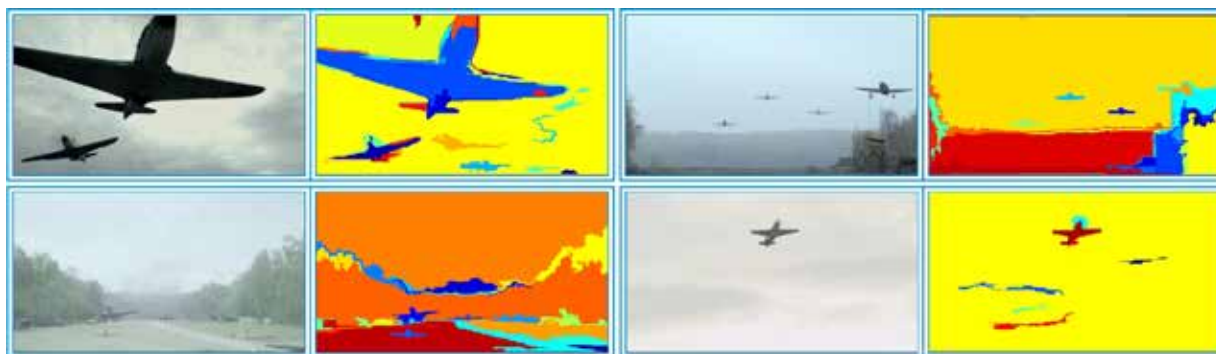


Рисунок 7.14 – Кадры кинофильма «Истребители» и их сегментация

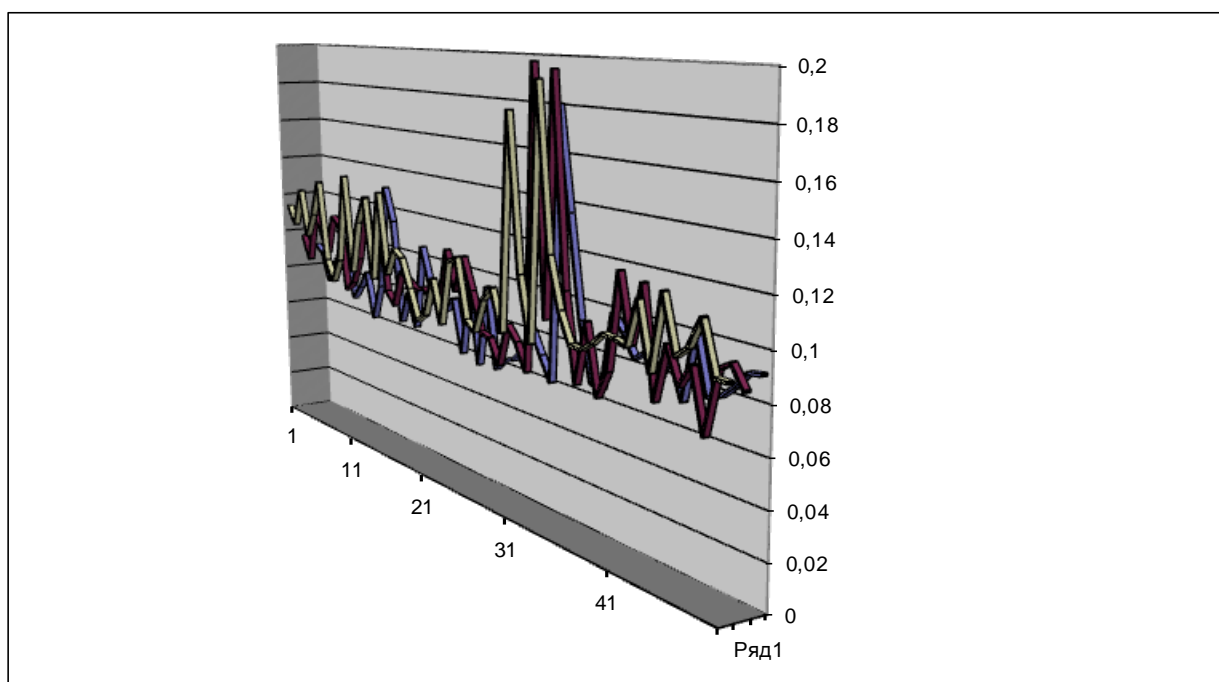


Рисунок 7.15 – Зависимость сегментации во времени от параметров пространственной сегментации в ролике «Истребители» на короткой выборке

Аналогичные эксперименты были проведены для различного разрешения видео. Оказалось, что вариации в разрешении также не оказывают принципиального влияния на структуризацию видеоданных. Однако при разрешении менее 400 строк колебания различий между видеокадрами выравниваются с тенденцией уменьшения, а уже при 200 строках задача сегментации существенно усложняется. При увеличении степени детализации видео колебания различий между видеокадрами также выравниваются, но с тенденцией

увеличения размаха, что уже при 800 строках может оказать влияние на выделение ложных сегментов, что предопределяет чрезмерную темпоральную сегментацию.

Перейдем к экспериментальному анализу выделения ключевой информации для полученных результатов темпоральной сегментации.

7.2 Имитационное моделирование признаков пространств для задач поиска с запросами «по образцу»

Основными источниками тестовых видеороликов считаются CERN Document Server (база лаборатории физики элементарных частиц), Open Video project, Movie Content Analysis project и Internet Archive. Также можно использовать данные из библиотек цифровых видеороликов (DVL – Digital Video Libraries): Informedia DVL (<http://www.informedia.cs.cmu.edu>), Consumer DVL (<http://www.cdvl.org/>), Hemispheric Institute DVL (<http://hidvl.nyu.edu/>), Harvard-Smithsonian Center for Astrophysics DVL (<http://hsdvl.org/>) и др. Доступ к материалам некоторых библиотек закрыт, однако его можно получить по запросу для исследовательских целей.

Все видеоматериалы открытого доступа, как правило, предлагаются небольшого размера с низким разрешением. Однако при низком разрешении мелкие детали сливаются с фоном, а с увеличением разрешения повышается уровень детализации информации кадра, поэтому тестирование методов с различным разрешением и размером имеет ключевое значение. Для тестирования разработанных методов на видео с высоким разрешением приходится использовать видеоролики собственного производства. Пользовательские видео также применяются для тестирования под специфические предметные области, такие как наблюдение за офисом банка или номерными знаками автомобилей на проезжей части. Наиболее часто используемыми стандартами видео принято считать PAL (720 ´ 576), HD (1280 ´ 720) и Full HD (1920 ´ 1080).

Некоторые исследователи полагают, что 3-4 видеоролика с разным типом содержания (или даже одинаковым) достаточно для тестирования, другие выполняют проверку разработанных методов на 20 роликах продолжительностью более часа, тестовые коллекции третьих насчитывают около сотни видеороликов. Жанр видео оказывает влияние на задействованные визуальные признаки, поэтому для получения правдивых результатов необходимо проводить тестирование на видеоданных разного типа: новостных, спортивных, мультипликационных роликах, кинофильмах, ток-шоу, если только предлагаемый метод не ориентирован на решение узкого класса задач с однородным содержанием.

Кроме того, тестовые видео должны включать движение камеры, изменение масштаба и условий освещения, потому как некоторые методы, особенно ориентированные на наблюдение, не смогут корректно обрабатывать такого рода изменения. Также метод должен быть протестирован на видеоданных разного размера и разрешения, большой и незначительной межкадровой разностью. Часто методы, которые показывают неплохой результат на большей части видеопоследовательностей, плохо справляются с обработкой нескольких типов жанров, особенно это касается распознавания лиц и наложения текстовой информации на видео. Кроме тестовой коллекции TRECVID с низким разрешением и небольшим размером кадров (320 × 240) из открытого доступа Интернет-архива, разработанный метод был опробован на нескольких рекламных роликах с разрешением, ориентированным для просмотра на телеэкране. В табл. 7.2 приведена информация об использованных тестовых коллекциях свободного доступа.

На рис. 7.16 показаны примеры исходных видеопоследовательностей, использованных для тестирования предложенного метода поиска. Для того, чтобы уместить кадры видеороликов с разным содержанием на одном рисунке, они были извлечены с интервалом в 250 кадров (для 1 и 3 видеороликов), а для 2 видеоролика показаны кадры только небольшого фрагмента. Для поиска ключевых кадров использовались все кадры исходного материала.

Таблица 7.2 – Описание использованной тестовой коллекции из свободного доступа

Категория	Имя файла	Продол- жи- тельность (чч:мм:сс)	К-во сцен	Источ- ник
News (новости)	clusterpiece2-3_256kb.mp4	00:04:33	42	Internet Archive*
	factories_512kb.mp4	00:06:38	54	
Commercial (реклама)	LuckyStr1948_2.mpeg	00:00:57	10	OV Project**
	amc_jeep_512kb.mp4	00:01:24	28	Internet Archive*
Cartoon (мульти- фильм)	Flintstone_full.mpg	00:01:02	60	MoCA Project** *
	TheKangarooKid.mp4	00:07:42	47	Internet Archive*
Movie (ки- но-фильм)	VideoAbstract.mpg	00:03:25	31	MoCA Project***
	What_Happened_1901_512kb.mp4	00:01:17	1	Internet Archive*
Documen- tary (доку- ментальный фильм)	anni005.mpg	00:06:19	63	OV Project**
	apo16005.mpg	00:00:23	5	
	20030322-bbs-sheetz1_64kb.mp4	00:59:58	1	Internet Archive*

* Internet Archive (<http://archive.org>)

** Open Video Project (<http://www.open-video.org>)

*** Movie Content Analysis Project (<http://pi4.informatik.uni-mannheim.de/pi4.data/content/projects/moca/index.html>)

Для проверки работоспособности предложенного метода на видео Full HD и HD (1920 ´ 1080 и 1280 ´ 720) и сжатием MPEG 4 Video использовался самостоятельно отснятый материал продолжительностью около 2 минут каждый (25 кадров в секунду). Непрофессиональной камерой с высоким разрешением отснято движение автомобилей под мостом, людей на мосту, и открывающийся взору пейзаж. Эти видео характеризуется отсутствием смены сцен, лишь наличием движения объектов и камеры.

В демонстрационных целях был использован видеоматериал продол-



Рисунок 7.16 – Примеры исходных видеопоследовательностей

жительностью 30 секунд (25 кадров в секунду) с размером кадров 576 ´ 432 и сжатием Xvid – это китайский телевизионный рекламный ролик автомобилей компании Mercedes Benz C-Class, выпущенный телекомпанией Stink Berlin (3-й пример на рис. 7.16). Материал интересен тем, что он отснят в сумерках с изменением освещения, кроме того, он идеально подходит для тестирования функций поиска ключевых кадров, так как характеризуется очень частой сменой сцен. На рис. 7.17 показаны ключевые кадры, отобранные из этого видеоролика с использованием предложенного в диссертационной работе метода.



Рисунок 7.17 – Результат отбора ключевых кадров для рекламного видеоролика

Последние кадры видеоролика содержат пример наложения текста на видео. Разработанный метод не предназначен для поиска ключевых кадров при наложении текста, поэтому эти кадры не были включены в результирующую последовательность отобранных кадров.

Для того, чтобы наглядно представить отличия существующих методов и сравнить результаты поиска ключевых кадров, полученные с использова-

нием разрабатываемого метода, предлагается смоделировать системы на основе анализа оптического потока движения, на основе кластерного анализа и анализа кривой, построенной по данным выделенных признаков.

Первый метод поиска ключевых кадров, результаты которого будут проанализированы, – это анализ движения с помощью оптического потока. Расчет оптического потока выполнялся вариационным методом Хорна-Шунка, который анализирует изменение энергии полей движения. Данный метод был выбран, как один из тех, которые сочетают в себе относительно хорошее качество и высокую скорость расчетов согласно общему рейтингу алгоритмов вычисления оптического потока из базы Middlebury.

На рис. 7.18 показаны кадры рекламного ролика автомобилей Mercedes (3-й пример на рис. 7.16), отобранные при появлении максимального количества движения. Из-за того, что видео предварительно не разбивалось на сцены, получилось, что кадры с незначительным движением не были приняты во внимание, тогда как из сцен со значительным количеством движения было извлечено слишком много кадров (в данном случае это обусловлено еще и изменением освещения, см. рис. 7.18). Однако, из каждой сцены хотя бы один кадр попал в результирующую последовательность, так как скачкообразное движение объектов наблюдается именно на границах сцен (то есть первые кадры каждой сцены были по большей части отобраны как ключевые). К сожалению, первые кадры каждой сцены далеко не всегда заключают в себе наиболее значимое содержание.

Как видно из рис. 7.18, в результате было отобрано слишком много похожих кадров со значительным количеством движения, что обусловлено изменением освещения в видеоролике. Следует отметить, что максимальное число движущихся объектов наблюдается либо на границах сцен, либо в середине сцены при сложной текстуре анализируемых кадров. Сложность извлечения ключевых кадров состоит в том, что необходимо вручную определять порог чувствительности (для каждого типа видеоданных нужно уста-



Рисунок 7.18 – Ключевые кадры рекламного ролика автомобилей Mercedes с мелкими штрихами, обозначающими оптический поток движения

навливать условие поиска максимума или минимума).

Как упоминалось ранее, данная группа методов позволяет выделить кадры только со значительным движением. Кроме того, методам на основе оптического потока присуща низкая вычислительная мощность. Данный метод может оказаться совершенно непригодным для анализа типового движения, аналогичного второй видеопоследовательности на рис. 7.16 с однородной сценой и экспозицией. Несмотря на наличие движения объектов, эта группа методов не позволяет определить, какое движение считать более значимым, а какое менее значимым. Но в целом результаты, полученные с использованием этого метода, достаточно неплохие, исходя из рис. 7.18. Про-

блема в вычислительной скорости и значительно худших результатах при иных начальных условиях (исходной видеопоследовательности).

Следующая группа методов не требует предварительного определения границ сцен, но число извлекаемых кадров всегда ограничено количеством кластеров. На рис. 7.19 показаны кадры рекламного ролика автомобилей Mercedes, отобранные с помощью кластеризации методом J -средних.

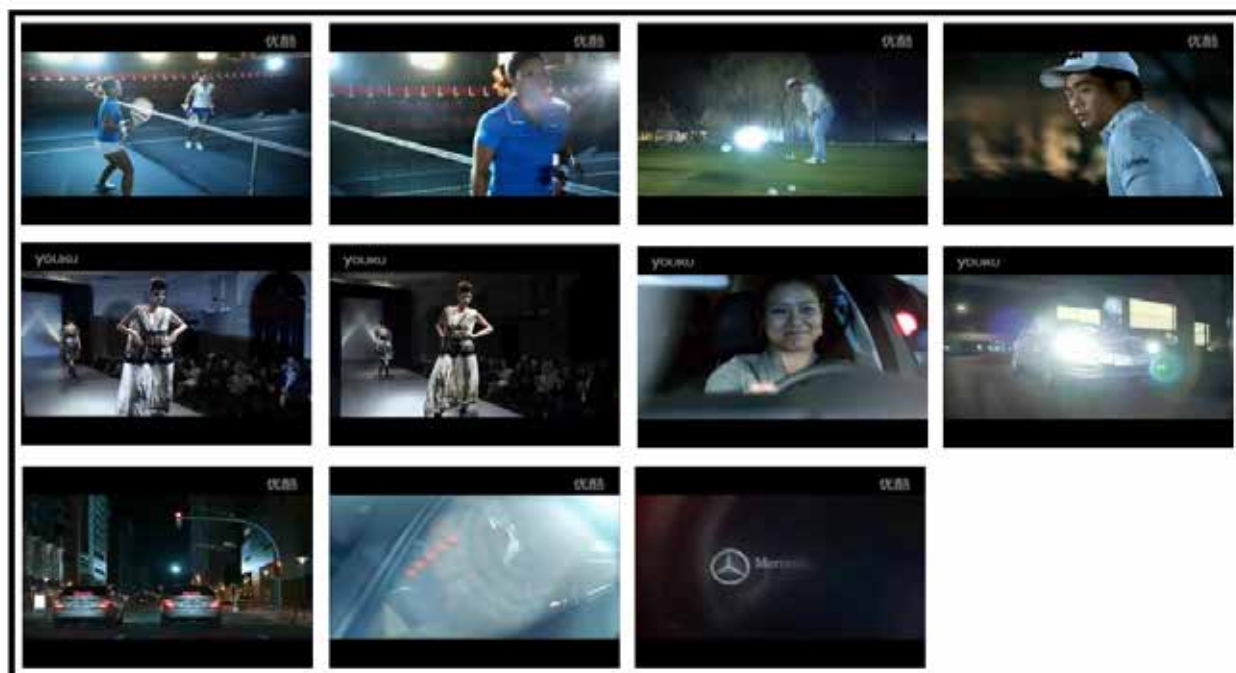


Рисунок 7.19 – Ключевые кадры рекламного ролика автомобилей Mercedes, отобранные методом J -средних

Еще один подход, который достаточно часто встречается в литературе по поиску ключевых кадров, – это методы, связанные с упрощением кривой, которая строится на основании выбранных признаков. На рис. 7.20 показаны кадры рекламного ролика автомобилей Mercedes, отобранные путем упрощения кривой, которая построена по цвето-текстурным признакам.

Эксперименты показали, что набор упрощенных признаков формы с достаточной валидностью позволяет находить ключевые кадры, что в полной мере подтверждает правомочность использования средств CBIR для CBVIR.



Рисунок 7.20 – Кадры рекламного ролика автомобилей Mercedes, отобранные с помощью метода упрощения кривой

7.3 Исследование специфики решения поисковых задач структурированного видео

Для оценки качества работы поисковой системы применяются два различных типа оценок, основанные на анализе поиска и результатов работы системы. При поиске информации соответствующей некоторому запросу в поисковой системе или при классификации информации в рубрицирующей системе (при вынесения решения о соответствии или несоответствии некоторому классу) часто используют подход, основанный на весовых коэффициентах. На втором же этапе оценки результата поиска используют метрики.

На первом этапе мы использовали стандартный подход – весовые коэффициенты. В качестве весовых коэффициентов использовались следующие:

a_1 – коэффициент Луна $w_{i,j} = tf_{i,j}$ (tf – поиск данных),

где $tf_{i,j}$ – количество раз, когда данные i соответствуют данным j .

a_2 – TFIDF $w_{i,j} = tf_{i,j} \cdot \ln \frac{J}{df_i}$ (TFIDF – поиск данных в видеопотоке с

учетом их встречаемости в видео),

где J – общее количество видео в коллекции,

df_i – количество видео, в которых встречаются i данные.

$$a_3 - \text{нормализованный TFIDF } w_{i,j} = \frac{tf_{i,j}}{ndl_j} \ln \frac{J}{df_i},$$

где $ndl_j = avdl / dl_j$,

$avdl$ – средняя длина видео,

dl_j – длина j обрабатываемого видео.

a_4 – DTW_{LM} (также носит название LM). Весовой коэффициент, получивший наибольшее применение в статистической модели и выражающийся

формулой: $w_{i,j} = \frac{tf_{ij}}{dl_j}$.

a_5 – INQUERY(учет запроса): система INQUERY предложенная Калланом в 1992 и получившая развитие в 1998 базируется на сходстве сетевой информации с данными, по которым производится индексация и содержанием запроса (данными, содержащимися в нем) Таким образом, вероятность соответствия вычисляется по формуле

$$w_{i,j} = 0.4 + 0.6 \times \frac{tf_{ij}}{tf_{ij} + 0.5 + 1.5 \times \frac{dl_j}{avdl}} \times \frac{\ln \frac{J+0.5}{df_i}}{\ln(J+1)}.$$

a_6 – обратное соответствие (IDF – подсчет количества видео в коллекциях, в которых встречаются искомые данные) заключается в том, что данные, которые встречаются в очень малом количестве видео, приобретают дополнительный вес (предложен Спарксом):

$$w_{i,j} = idf_{i,j} = \ln \frac{J}{df_i}.$$

a_7 – BM25, выражающийся в виде

$$w_{i,j} = \frac{(k_1 + 1) \times tf_{i,j}}{K + tf_{i,j}} \times \ln \frac{J - df_i + 0.5}{df_i + 0.5} \times (k_2 + 1) \times \frac{qtf_i}{k_2 + qtf_i},$$

где qtf_i – количество данных i в запросах; k_1, k_2 – const соответственно равные 2, 5; $K = 2 \times ((1 - b) + b \times \frac{dl_j}{avdl})$, $b = 0.75$.

a_8 – T-INQUERY, вариация a_{16} , выражающаяся в виде

$$w_{i,j} = \frac{df_i}{df_i + 200 \times ((1 - b) + \frac{dl_j}{avdl})}.$$

a_9 – Окарі коэффициент, выражающийся формулой

$$w_{i,j} = \frac{(k_1 + 1) \times tf_{i,j}}{\frac{dl_j}{avdl} + tf_{i,j}}.$$

Кроме использованных весовых коэффициентов также могут использоваться ltn, lnc, dtc, ltc, dtu, Lnu, atn, prn, ntc, VM26 и т.д. Эти коэффициенты выражаются формулами:

$$\begin{aligned} \text{ltn } w_{ij} &= (\ln(tf_{ij}) + 1) \cdot idf_j \\ \text{atn } w_{ij} &= idf_j \cdot [0.5 + 0.5 \cdot tf_{ij} / \max tf_i] \\ \text{lnc } w_{ij} &= \frac{\ln(tf_{ij}) + 1}{\sqrt{\mathring{a}_{k=1}^t (\ln(tf_{ik}) + 1)^2}} \\ \text{ntc } w_{ij} &= \frac{tf_{ij} \cdot idf_j}{\sqrt{\mathring{a}_{k=1}^t (tf_{ik} \cdot idf_k)^2}} \\ \text{dtc } w_{ij} &= \frac{(\ln(\ln(tf_{ij}) + 1) + 1) \cdot idf_j}{\sqrt{\mathring{a}_{k=1}^t ((\ln(\ln(tf_{ik}) + 1) + 1) \cdot idf_k)^2}} \end{aligned}$$

$$\begin{aligned} \text{ltc } w_{i,j} &= \frac{(\ln(tf_{i,j}) + 1) \cdot idf_j}{\sqrt{\mathbf{a}_{k=1}^t ((\ln(tf_{i,k}) + 1) \cdot idf_k)^2}} \\ \text{dtu } w_{i,j} &= \frac{(1 + \ln(1 + \ln(tf_{i,j}))) \cdot idf_j}{(1 - slope) \times pivot + slope \times nt_i} \\ \text{Lnu } w_{i,j} &= \frac{(\ln(tf_{i,j}) + 1 / \ln(1 / nt_i) + 1)}{(1 - slope) \times pivot + slope \times nt_i} \\ \text{npr } w_{i,j} &= tf_{i,j} \cdot \ln\left[\frac{(n - df_j)}{df_j}\right] \\ \text{BM26 } w_{i,j} &= \frac{(k_1 + 1) \times tf_{i,j}}{K + tf_{i,j}} \times \ln \frac{J - df_i + 0.5}{df_i + 0.5} \times (k_2 + 1) \end{aligned}$$

Рассмотрим некоторую специфику использования коэффициентов. Предположим необходимо выделить все данные, которые встречаются в малом количестве видеоданных; тогда необходимо рассматривать a_2 и a_7 . a_2 должен неограниченно расти $a_2 \geq 50$, а a_7 должен быть положительным и иметь значение $20 \leq a_7 \leq 50$. Если же мы введем еще одно ограничение, например, общий размер должен быть небольшим, то в этом случае дополнительно используются a_3, a_4 и a_6 . a_3 должен быть минимальным (в диапазоне от 0 до a_2), a_4 увеличивается и имеет значение $5 \times 10^{-3} \leq a_4 \leq 5 \times 10^{-2}$, а a_6 имеет значение больше 1.

Рассмотрим теперь следующий вариант. Необходимо выделить данные, которые часто встречаются в большом количестве видеопотоков, т.е. данные по которым сложно идентифицировать отдельную видеопоследовательность, но в тоже время это данные, характерны для этого класса видеопотоков. В этом случае мы можем использовать a_7 . Если $a_7 < 0$, то сегмент встречается минимум в 50% видео коллекции.

При необходимости использования оценки запроса пользователя нужно брать коэффициенты a_5, a_7, a_8 , которые включают встречаемость запрашиваемых данных в запросах других пользователей. При одинаковой встречаемости в видео больший вес в a_5 и меньший a_8 будет иметь смысл – чаще

встречающийся в запросах.

Следует отметить, что во многих коэффициентах (a_3 , a_5 , a_7 , a_9) используется отношение длины текущего видео и средней длины видео в коллекции. Таким образом, можно учитывать в каком по объему видео встречаются запрашиваемые данные.

После проведения поиска и получения результата запроса необходимо оценить качество проведенной работы. Естественно, что «идеальным» алгоритмом считается тот, для которого выводы, сделанные поисковой системой, согласуются с мнением экспертов оценщиков. В качестве метрик оценки качества работы систем на практике используются следующие:

При оценке поиска по запросу:

1. Полнота (recall).
2. Точность (precision).
3. Средняя точность (average precision).
4. Точность на уровне 5 элементов (precision(5)).
5. Точность на уровне 10 элементов (precision(10)).
6. R-точность (R-precision).

А при классификации:

1. Полнота (recall).
2. Точность (precision).
3. Тщательность (accuracy).
4. Ошибка (error).
5. F-мера (F-measure).

Опишем каждую из этих метрик. Сначала дадим описание оценки качества работы системы в применении к одному запросу, а затем опишем методики усреднения метрик для получения интегральных показателей качества поиска/классификации. Для многих метрик спорным вопросом является случай, когда для данного запроса нет релевантных откликов. Например, значение полноты в этом случае — неопределенность типа 0/0. В TRECVideo такие запросы не учитываются при вычислении метрик.

1. Метрики на множествах объектов

Большинство метрик, применяемых в современной оценке поиска, основываются на отношении релевантности (принадлежности) результата запросу. Метрики для неупорядоченного множества основаны, как правило, на бинарной классификации объектов «релевантен/нерелевантен» по отношению к выбранному запросу.

Данные меры основываются на матрице классификации (Табл. 7.3).

Таблица 7.3 Основные категории ответа системы

	Релевантны	Не релевантны
Найдено системой	<i>a</i>	<i>B</i>
Не найдено системой	<i>c</i>	<i>D</i>

В таблице: *a* — количество видео, найденных системой и релевантных с точки зрения экспертов; *b* — количество видео, найденных системой, но не релевантных с точки зрения экспертов; *c* — количество релевантных видео, не найденных системой; *d* — количество нерелевантных видео, не найденных системой.

Полнота (*recall*) вычисляется как отношение найденных релевантных видео к общему количеству релевантных видео:

$$recall = \frac{a}{a + c}.$$

Полнота характеризует способность системы находить нужные пользователю видео, но не учитывает количество нерелевантных видео, выдаваемых пользователю. Например, если полнота равна 50%, то это значит, что половина релевантных видео системой не найдена.

Точность (*precision*) вычисляется как отношение найденных релевантных видео к общему количеству найденных видео:

$$precision = \frac{a}{a+b}$$

Точность характеризует способность системы выдавать в списке результатов только релевантные видео. Например, если точность равна 50%, то это значит, что среди найденных объектов половина релевантных и половина – нерелевантных.

Тщательность (accuracy) вычисляется как отношение правильно принятых системой решений к общему числу решений. Формально:

$$Accuracy = \frac{(a+d)}{(a+b+c+d)},$$

поскольку предполагается, что система принимает решение о принадлежности к данной категории для каждого видео в коллекции. Таким образом, знаменатель не зависит от рассматриваемой категории. При вычислении оценки в качестве знаменателя использовалось общее число видео оценивавшихся хотя бы для одной категории (что сильно меньше числа объектов, но это сказывается лишь на масштабе).

Отметим, что микроусреднение и макроусреднение для этой оценки дают одинаковый результат.

Ошибка (error) вычисляется как отношение неправильно принятых системой решений к общему числу решений. Формально:

$$Error = \frac{(b+c)}{(a+b+c+d)}.$$

F-мера часто используется как единая метрика, объединяющая метрики полноты и точности в одну метрику. F-мера для данного запроса (рубрики) вычисляется по формуле

$$F = \frac{2}{\frac{1}{p} + \frac{1}{r}} \quad (7.2)$$

Отметим основные свойства метрики F : $0 \leq F \leq 1$, если $r=0$ или $p=0$, то $F=0$, если $r=p$, то $F=r=p$, $\min(r, p) \leq F \leq \frac{r+p}{2}$.

Одним из важных вопросов построения метрик при информационном поиске (особенно в случае классификации и фильтрации) является метод усреднения результатов, что зачастую может оказывать значительное влияние его на результаты оценки.

В случае построения усредненной по множеству заданий той или иной множественной метрики можно рассмотреть две последовательности действий:

- сначала вычислить метрики по каждому запросу отдельно и затем их усреднить;
- найти общее количество данных, относящихся к категориям и уже на их основе вычислить искомую метрику.

Первый способ вычисления принято называть микроусреднением, второй – макроусреднением. Первый способ характерен для оценки задач поиска, в которых важен результат в среднем по запросу, независимо от мощности ответа на этот запрос. Второй же способ нашел большее применение в оценке классификации и фильтрации, где необходимо учитывать «размеры» запросов.

С точки зрения анализа видеопоследовательностей имеет смысл рассмотреть меры на последовательностях данных.

Метрики для списка информации, отсортированного по релевантности, учитывают не только факт наличия информации в общем списке найденных данных, но и его положение в этом списке. Такой вид метрик имеет смысл только в оценке систем поиска.

Точность на уровне n элементов определяется как количество реле-

вантных данных среди первых n выданных данных, делённое на n . Если система выдала более данных, то эта величина равна точности системы на первых n результатов запроса. Если система выдала менее n элементов, то точность на уровне n элементов будет заведомо не выше точности системы. Точность на уровне n элементов характеризует способность системы выдавать релевантные элементы в начале списка результатов. Например, если система выдает не более 10 элементов на первой странице, то $\text{precision}(10)$ отражает качество результатов системы, получаемых на первой странице. Эта метрика имеет ряд недостатков. В частности, для различных запросов метрики $\text{precision}(n)$ могут быть несравнимы. Несмотря на известные недостатки, точность на уровне является незаменимой метрикой современных систем поиска так как, в частности, позволяет оценить полезность первой страницы ответа системы для Пользователя.

R-точность равна точности на уровне n элементов для n равного количеству релевантных элементов для данного запроса. Данная метрика призвана заменить точность на уровне в тех случаях, когда необходимо учесть большую разницу в количестве релевантных элементов разных запросов.

Средняя точность для данного запроса определяется следующим образом: пусть для данного запроса имеется k релевантных элементов. Точность на уровне i -го релевантного элемента $\text{prec_rel}(i)$ равна $\text{precision}(\text{pos}(i))$, если i -й релевантный элемент находится в результатах запроса на позиции $\text{pos}(i)$. Если i -й релевантный элемент не найден, то $\text{prec_rel}(i)=0$. Средняя точность для данного запроса равна среднему значению величины $\text{prec_rel}(i)$ по всем k релевантным элементам:

$$\text{AvgPrec} = \frac{1}{k} \sum_{i=1}^k \text{prec_rel}(i).$$

Отметим основные свойства метрики «средняя точность»:

– $\text{AvgPrec} \leq \text{recall}$;

– если релевантные элементы находятся только в начале списка результатов, то $AvgPrec = recall$;

– если релевантные элементы равномерно распределены по списку результатов, то $AvgPrec \approx precision \times recall$;

– количество элементов, ранжированных ниже последнего релевантного, не влияет на значение $AvgPrec$ (отсекается «хвост»).

Средняя точность позволяет оценивать качество работы системы, учитывая приоритет высоко ранжированных элементов перед элементами, находящимися в конце списка. В отличие от метрик $precision(n)$ и R - $precision$, средняя точность учитывает все найденные элементы. Необходимо отметить, что метрика $AvgPrec$ обладает высокой устойчивостью относительно вариаций оценки экспертов при вычислении средней оценки по множеству запросов.

Примеры кадров видеопоследовательностей, которые использовались в качестве тестовой информации для выявления работоспособности системы информационного поиска представлены на рисунке 7.21. В качестве базы для поиска были выбраны документальные видео катастроф различного рода, произошедших как в результате действий человека, так и действий природного характера. Также стоит подчеркнуть, что отдельным множеством выступали техногенные катастрофы.

Частным случаем формулы (7.2) является так называемый коэффициент Дайса. На рис. 7.22 приведена линейная диаграмма, показывающая средние значения коэффициентов Дайса, полученных после проведения сегментации при извлечении ключевых кадров из разных по жанру видеопоследовательностей.

Для кадров, отобранных в качестве ключевых с помощью диаграмм Вороного из тестовой видеокolleкции, коэффициент Дайса в среднем составил 0,89, что свидетельствует о неплохих результатах по сравнению с оценкой метода оптического потока, которая в среднем составила 0,61.



Рисунок 7.21 – Примеры кадров входных видеопоследовательностей

Точность на уровне 10 элементов по результатам запросов к тестовой коллекции по соответствию типам видео в среднем составила 62%, что с одной стороны не является прекрасным результатом, но все же предоставляет Пользователю достаточно много релевантной к его запросу информации.

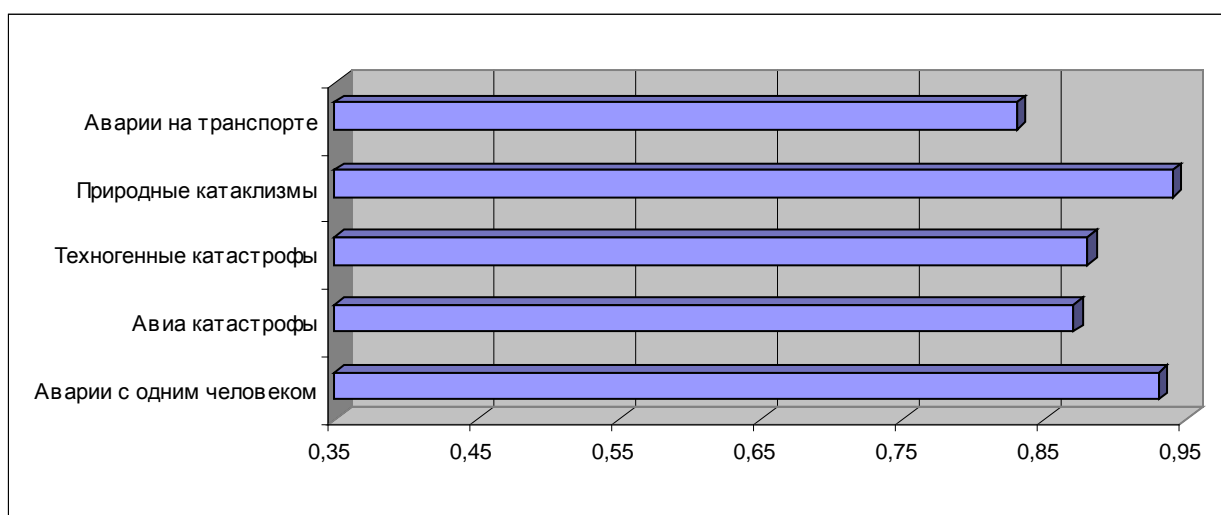


Рисунок 7.22 – Значения коэффициентов Дайса, полученные при извлечении ключевых кадров из разных по жанру видеопоследовательностей

7.4 Особенности внедрения результатов диссертационных исследований

Результаты диссертационных исследований использованы в Харьковской областной администрации для системы селекторных конференций видеосвязи (акт от 09.10.2014), в ГУМВД Украины в Харьковской области (акт от 28.01.15) с целью упрощения детектирования фрагментов видео, которые содержат наиболее важную для розыска информацию. В частности, для улучшения отдельных сегментов видеопотоков, снятых с помощью видеокамер с небольшим разрешением, для выявления личностей преступников. На рисунке 7.23 и 7.24 представлены примеры выявленной опорной информации видеопоследовательностей, содержащие информацию о личности, необходимую для ее идентификации в ходе розыска.

Также результаты внедрены на ТОО Харьковского технического центра рентгеновского сервиса «Спектр» (акт от 14.12.2014) для поиска информации в потоках видеоданных рентгеновского спектра и в ТОО Завода рентгеновского оборудования «Квант» (акт от 18.12.2014) для анализа и выделения необходимых фрагментов в потоках рентгеноскопической видеоинформации.



Рисунок 7.23 – Пример обнаружения в видеопоследовательности
«областей интереса»

мации. Пример поиска ключевых кадров при рентгеноскопическом кардиологическом исследовании представлен на рисунке 7.25.

Кроме того результаты внедрены в Харьковском национальном университете радиоэлектроники (акт от 18.05.2016) при проведении лекционных и лабораторных работ.

Результаты теоретико-экспериментальных исследований были реализованы в виде прикладных и исследовательских программных комплексов, один из которых «Программный комплекс пространственно-временной сегментации видеоданных в задачах поиска по запросу «по образцу»» сертифицирован УкрСЕПРО.

В результате исследований установлено, что предложенные модели и методы анализа видеопоследовательностей позволяют проводить сегментирование исходных данных и выделение опорной информации как всей последовательности так и отдельных ее сегментов, что позволяет значительно упростить процедуру поиска «по образцу». Полученная опорная информация может храниться в больших коллекциях визуальной информации, что позволяет при поиске корреляций между запросом Пользователя и хранящейся информацией использовать динамические визуальные образы, а не только

текстовое и признаковое описание или статическое изображение, что уже в настоящее время позволяет значительно улучшать поиск видеoinформации, особенно в больших коллекциях.

Результаты практического использования результатов диссертационных исследований акцентируют внимание на вопросах кластеризации больших коллекций видео (центры кластеров выступают в качестве опорных



Рисунок 7.24 – Пример обнаружения в видеопоследовательности «областей интереса»

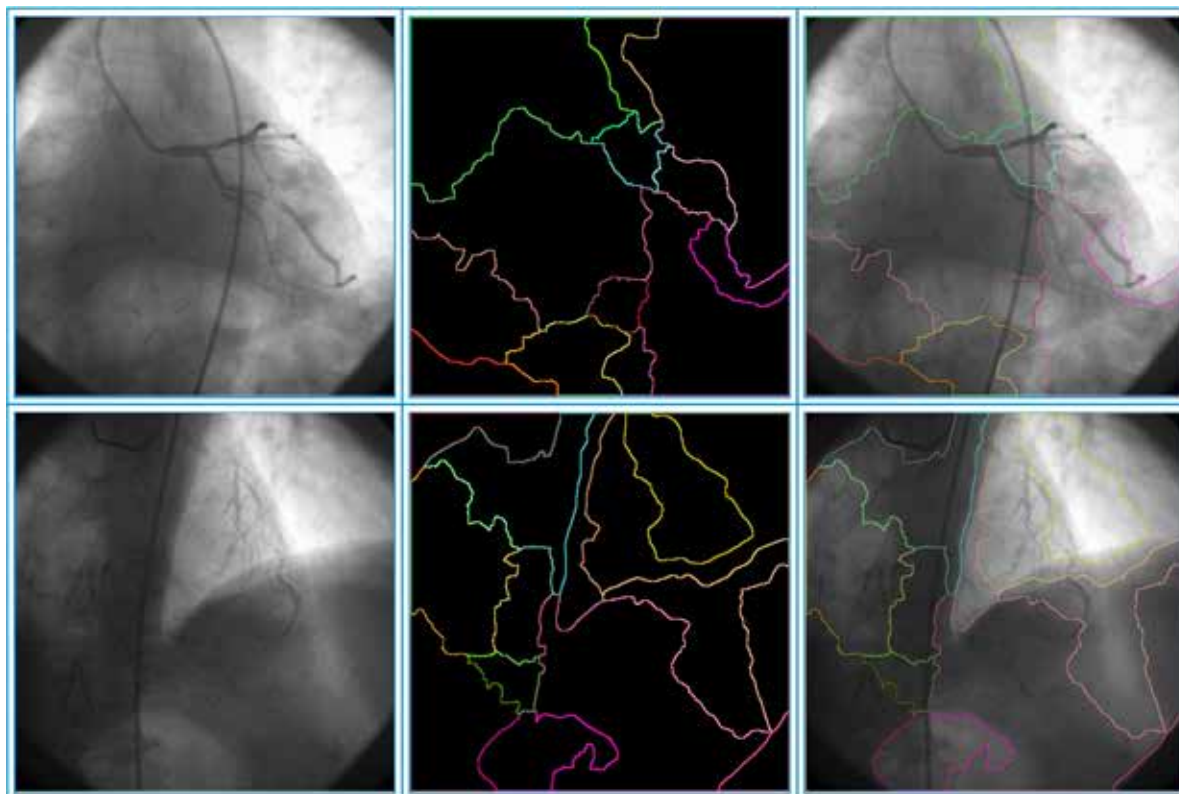


Рисунок – 7.25 Пример выделения ключевых кадров
из рентгеноскопического видео

точек), причем методы кластеризации должны быть аддитивными в том плане, чтобы обеспечивать пополнение коллекций. Кроме того кластеризация найденных сегментов должна быть стратифицированной, что в конечном итоге обеспечит семантический анализ видео на уровнях «событий», «сцен», «сценариев».

Выводы по седьмому разделу

1. При темпоральной сегментации многомерных временных рядов, индуцированных видеопоследовательностями, в силу широкой вариативности свойств этих рядов следует комплексно использовать предложенные on line методы обнаружения разладок, адаптивно выбирая разбиения или покрытия видео.

2. При предварительной обработке видеокolleкций для дальнейшего

on line построения областей элиминации целесообразно использовать различные варианты метрик и их комбинаций, что в итоге приведет к снижению семантического разрыва при поиске сегментов (ключевых кадров) с запросами «по образцу».

3. Рациональный синтез признаков пространств применительно к темпоральной обработке видео обеспечивает разумный компромисс между избыточностью и дефицитом информации на всех этапах принятия решений.

4. Фрагментные методы темпоральной обработки видеопоследовательностей (кластеризации и прогноза) создают предпосылки локализации областей интереса, что может заметно повышать точность и надежность анализа видео при тех же вычислительных ресурсах.

ВЫВОДЫ

Развитие моделей и методов темпоральной обработки, анализа и интерпретации плохоструктурированных потоков информации (прежде всего, видеопотоков) предопределяет развитие интеллектуальных информационных технологий мультимедиа и гипермедиа. Релевантность информационных ресурсов, разумный компромисс между накоплением данных и потребностями их рационального использования остаются одной из ключевых проблем. Тенденции перманентного расширения номенклатуры решаемых задач, интенсификации обработки видеопотоков, 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 устранение семантического конфликта между низкоуровневыми моделями видео и высокоуровневыми целями их предметно-ориентированной трактовки остается ключевой проблемой. Специфика решения этих задач в СВВIR определяется необходимостью *on line* обработки и анализа очень больших объемов информации в условиях существенной неопределенности при главенствовании не пространственных (в поле зрения) корреляционных связей, а сложных отношений между видеокадрами во времени в условиях ограничений на время принятия решений и вычислительных ресурсов.

При проведении теоретико-экспериментальных диссертационных исследований получены следующие основные результаты:

1. Установлено, что для больших библиотек видео перспективным направлением исследований является разработка моделей и методов темпоральной обработки видеопотоков, обеспечивающих организацию семантического поиска с запросами по образцу. Содержание видеоданных индуцируется двойной сегментацией: во времени (*on line* поиск изменений событий или

сюжетов) и пространстве (поиск областей, коррелирующих с объектами сцен).

2. Предложены, обоснованы и исследованы модели структуризации видеопоследовательностей для задач информационного поиска с запросами 'ad exemplum' на основе построения областей элиминации для произвольных конфигураций объектов в произвольном метрическом пространстве.

3. Разработаны фрагментные методы обработки визуальной информации, которые равносильно могут использоваться для темпоральной и пространственной сегментации видеоданных с целью повышения валидности семантического поиска в больших библиотеках видеоданных.

4. Изучены методы валидного сравнения конечномерных разбиений видеокладов, проанализированы множества дескрипторов элементов разбиений и изображений, которые позволяют повысить адекватность представления видеорядов в признаковых пространствах.

5. Синтезирован комплекс моделей и методов on line сегментации видеопоследовательностей и многомерных временных рядов, индуцируемых видеопотоками, путем обнаружения изменения их свойств, что обеспечивает валидную структуризацию видео.

6. 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.

7. Разработаны нейросетевые методы компрессии видеокладов и видео для темпоральной обработки видеопотоков в задачах многоаспектной индексации, исследованы вопросы эффективности компрессии изображений.

8. Путем экспериментальных исследований установлена валидность разработанных моделей и методов темпоральной обработки видеопоследовательностей, определены условия их рационального использования в задачах интеллектуального информационного поиска.

9. На основе предложенных методов и моделей темпоральной обработ-

ки видеопоследовательностей разработан ряд прикладных и исследовательских программных комплексов, используемых при поиске в больших видеоархивах с запросами 'ad exemplum'.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Абрамов, С.К. Мера содержания фона на основе энтропии для поиска и сортировки изображений в базах данных / С.К. Абрамов, В.В. Лукин, Н.Н. Пономаренко // Радиоелектронні і комп'ютерні системи. – 2007. – №2 (21). – С. 24-28.
2. Аналіз та обробка даних засобами обчислювального інтелекту / Є.В. Бодянський, Д.Д. Пелешко, О.А. Винокурова, С.В. Машталір, Ю.С. Іванов. – Львів: Видавництво Львівської політехніки, 2016. – 236 с.
3. Андерсен, Т. Статистический анализ временных рядов / Т. Андерсен; пер с англ. под ред. И.Г. Беляева. – М.: Мир, 1976. – 755 с.
4. Антощук, С.Г. Анализ базисных функций вейвлет-преобразования при мультимасштабном контурном представлении изображений / С.Г. Антощук, А.А. Николенко, Е. В. Ткаченко // Електромашинобудування та електрообладнання. – 2009. – Вип. 72. – С. 15-19.
5. Антощук, С.Г. Информационная технология локализации текстовых областей на сложном фоне / С.Г. Антощук, А.А. Николенко, Т.К. Тьен // Електротехнічні та комп'ютерні системи. – 2014. – №16(92). – С. 114-121.
6. Антощук, С.Г. Оценка эффективности моделей событий в задачах семантического анализа видеопотока / Н.А. Годовиченко, С.Г. Антощук // Електротехнічні та комп'ютерні системи. – 2014. – №13(89). – С. 144-151.
7. Ахметшин, А.М. Нечеткая сегментация слабоконтрастных радиологических изображений / А.М. Ахметшин, Л.Г. Ахметшина // Клиническая информатика и телемедицина. – 2010. – Т. 6, Вып. 7. – С. 37-42.
8. Березький, О.М. Статистичне оброблення цитологічних зображень / О.М. Березький, К.М. Березька, С.Ю. Попіна // Вісник Хмельницького національного університету. – 2012. – №5. – С. 161-164.
9. Богословский, А.В. Методы обработки видеопоследовательностей на основе анализа энергетических параметров видеосигнала изображения / А.В. Богословский, И.В. Жигулина, И.С. Маслов // Радиотехника. – 2015. – № 4.

С. 112-119.

10. 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.

11. Богучарский, С. Анализ текстур в последовательности изображений на основе векторного квантования / С. Богучарский, С. Машталир // Радиоэлектроника, информатика, управление. –2014. – № 2(31) – С. 94-99. (Входит в международные наукометрические базы Ulrich's Periodicals Directory, BASE, ADAT, Index Copernicus, DOAJ, EBSCO, WorldCat и др.)

12. Богучарский, С.И. Иерархическая агломеративная кластеризация изображений в больших базах данных / С.И. Богучарский, А.Г. Каграманян, С.В. Машталир // Системи обробки інформації. – 2014. – Вип. 8. – С. 93-97. (Входит в международные наукометрические базы BASE, Index Copernicus, SIS.)

13. 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.

14. 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: сб. научн. трудов междунар. научн. конф., Железный Порт, 25-28 мая 2015 г. – Херсон: ХНТУ, 2015. – С. 251-253.

15. Богучарский, С.И. Модификация метода J -средних в задачах фрагментной сегментации изображений / С.И. Богучарский, А.Г. Каграманян, С.В. Машталир // Бионика интеллекта. – 2015. – №1(84) – С. 19-22.

16. 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 в

международные наукометрические базы Ulrich's Periodicals, Index Copernicus.)

17. Богучарский, С.И. Модифицированный метод кластеризации X -средних в задачах сегментации изображений / С.И. Богучарский, С.В. Машталир // Электротехнические и компьютерные системы. – 2015. – № 20 (96). – С. 106-110. (Входит в международные наукометрические базы Ulrich's Periodicals, Index Copernicus.)

18. Богуш, Р.П. Анализ алгоритмов обработки динамических изображений в приложении к задаче сегментации движения на сложном фоне / Р.П. Богуш, Г.А. Самощенко // Актуальные вопросы современной информатики: матер. междунар. заочн. научн.-практ. конф., Коломна, 1-15 апр. 2011 г. – Коломна: МГОСГИ, 2011. – Т. 2. – С. 13-16.

19. Бодянский, Е.В. Адаптивное обобщенное прогнозирование многомерных случайных последовательностей / Е.В. Бодянский, И.П. Плисс, Т.В. Соловьева // Доклады АН УССР. – 1989. – Сер. А. – № 9. – С. 73-75.

20. Бодянский, Е.В. Искусственные нейронные сети: архитектуры, обучение, применения / Е.В. Бодянский, О.Г. Руденко. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с.

21. Бодянский, Е.В. О решении задачи управления матричным объектом в условиях неопределенности / Е.В. Бодянский, И.П. Плисс // Автоматика и телемеханика. – 1990. – №2. – С 175-178.

22. Бодянский, Е.В. Ортосинапс, ортонейроны и нейропредиктор на их основе. / Е.В. Бодянский, Е.А. Викторов, А.Н. Слипченко // Системи обробки інформації. – 2007. – Вип.4(62). – С.139-143.

23. Бодянский, Е.В. Рекуррентная нейронная сеть для обнаружения изменений свойств нелинейных стохастических последовательностей / Е.В. Бодянский, С.А. Воробьев // Автоматика и телемеханика. – 2000. – №7. – С. 55-67.

24. Бодяньський, Є.В. Виявлення змін у потоці відеоданих на основі аналізу багатовимірних часових рядів / Бодяньський Є.В., Машталір С.В.//

Доповіді Національної академії наук України. –2012. –№11. – С.30-33.

25. Бодянський, Є.В. Синтез квазіпрямих адаптивних регуляторів / Є.В. Бодянський, І.П. Плісс, Г.В. Соловійова // Доповіді АН УРСР. – Сер. А. – 1987. – №1. – С. 59-61.

26. Бокс, Дж. Анализ временных рядов / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974. – 406 с.

27. Введение в контурный анализ; приложения к обработке изображений и сигналов / Я.А. Фурман, А.В. Кревецкий, А.К. Передреев и др; под ред. Я.А. Фурмана. – 2-е изд., испр. – М.: Физматлит, 2003. – 592 с.

28. 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.

29. 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 в международные наукометрические базы BASE, Index Copernicus, SIS.)

30. Воробейчиков, С. Обнаружение изменения параметров авторегрессионных процессов с неизвестным распределением помех / С. Воробейчиков, В. Конев // Статистические проблемы управления. – 1988. – Вып. 83. – С. 175-179.

31. Використання логифмічних перетворень для сегментації зображень зварних швів / Р.А. Воробель, І.Б. Івасенко, Т.С. Мандзій, В.В. Боцян // Відбір і обробка інформації. – 2011. – Вип. 35 (111). – С. 109-115.

32. 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 в международные наукометрические базы

Inspec, EBSCO, BASE, World Cat.)

33. Вишняков, Б.В. Использование модифицированного метода оптических потоков в задаче обнаружения и межкадрового прослеживания движущихся объектов / Б.В. Вишняков, Ю.В. Визильтер, А.В. Лагутенков // Вестник компьютерных и информационных технологий. – 2007. – № 5. – С. 2-8.

34. Годовиченко, Н.А. Моделирование событий для семантического анализа видеопотока в условиях неопределенности входных данных / Н.А. Годовиченко, Антощук С.Г. // Сучасні інформаційні та електронні технології: праці 15-ої міжн.-практ. конфю, Одеса 26-30 травня 2014 р. – Одесса: «Политтехперіодика». – 2014. – С. 50-51.

35. Головастов, А. Машинное зрение и цифровая обработка изображений / А. Головастов // Современные технологии автоматизации. – 2010. – № 4. – С. 8-18.

36. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс; пер с англ. под ред. П. А. Чочиа. – М.: Техносфера, 2005. – 1072 с.

37. Гончаренко, М.О. Детектирование изменений сцены в потоке видеоданных / М.О. Гончаренко, С.В. Машталир // Электротехнические и компьютерные системы. – 2012.– № 7 (83) – С. 143-147. (Входит в международные наукометрические базы Ulrich's Periodicals, Index Copernicus.)

38. Гороховатский, В.А. Структурное распознавание изображений на основе моделей голосования признаков характерных точек / В.А. Гороховатский, Е.П. Путятин // Реєстрація, зберігання і обробка даних.– 2008.– Т.10. – № 4. – С.75-85.

39. Гороховатский, В.А. Применение кластеризации к элементам структурного описания в задаче классификации изображений / В.А. Гороховатский // Системи управління, навігації та зв'язку: зб. наук. пр.– К.: ЦНДІНУ. – 2011. – Вип. 1(17). – С. 64-67.

40. Гребенюк, Е.А. Методы анализа нестационарных временных рядов с неявными изменениями свойств / Е.А. Гребенюк // Автоматика и

телемеханика. – 2005. – №12. – С. 3-30.

41. Гришин, В.Н. Модели, алгоритмы и устройства идентификации сложных систем. / В.Н. Гришин, В.А. Дятлов, Л.Г. Милов // Л.: Энергоатомиздат, 1985. – 104 с.

42. Гришин, С.В. Метод повышения эффективности временной сегментации видео на сценах с медленным движением / С.В. Гришин, Д. С. Ватолин // ГрафиКон'2007: сб. трудов конф., Москва, 23-27 июня 2007 г. – М.: МАКС Пресс, 2007. – С. 253-256.

43. Грузман, И.С. Цифровая обработка изображений в информационных системах / И.С. Грузман. – Новосибирск: НГТУ, 2002. – 352 с.

44. Дискретное моделирование динамических систем симптоматики возникновения хронической сердечной недостаточности у подростков / Носов К.В., Жолткевич Г.Н., Беспалов Ю. Г., Высоцкая Е.В., Печерская А.И., Рак Л.И. // Научно-методический электронный журнал «Концепт». – 2013. – № 11. – С. 101-105.

45. Дружинин, Д.В. Комбинированный алгоритм сжатия ключевых кадров экранного видео / Д.В. Дружинин // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2011. – №3. – С.67-77.

46. Дуда, Р. Распознавание образов и анализ сцен / Р. Дуда, П. Харт; пер с англ. Г.Г. Вайнштейн, А.М. Васьковский; под ред. В.Л. Стефанюка. – М.: Мир, 1976. – 512 с.

47. Егоров, А. Сравнительный анализ методов морфологической нормализации / Егоров А., Машталир С. // Радиоэлектроника и информатика. – 2005. – №4. – С. 90-95. (Входит до міжнародних наукометричних баз BASE, World Cat.)

48. Жигулина, И.В. Алгоритмы обработки видеопоследовательностей при обнаружении движения по краям изображения / И.В. Жигулина // Радиотехника. – 2013. – №9. – С.83-86.

49. Жолткевич, Г.Н. Многовариантный анализ интегрированной

компонентно-ориентированной информационной технологии создания автоматизированной системы управления / Г.Н. Жолткевич, Л.Е. Сергеев // Радіоелектронні і комп'ютерні системи. – 2003. – №4. – С. 189–192.

50. Журавлев, Ю.И. Распознавание. Математические методы. Программная система. Практические применения / Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько. – М.: Фазис, 2006. – 176 с.

51. Загоруйко, Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: Наука, 1999. – 264 с.

52. Загоруйко, Н.Г. Эмпирическое предсказание / Н.Г. Загоруйко. – Новосибирск: Наука, 1979. – 120 с.

53. Иванов, Ю.П. Комплексная фильтрация и классификация сигналов / Ю.П. Иванов. – Л.: Изд-во Ленинградского университета, 1988. – 212с.

54. Ивахненко, А.Г. Системы эвристической самоорганизации в технической кибернетике / А.Г. Ивахненко. – К.: Техника, 1971. – 372 с.

55. Ивахненко, А.Г. Помехоустойчивость моделирования / А.Г. Ивахненко, В.С. Степашко. – К.: Наук. Думка, 1985. – 216 с.

56. Ивахненко, А.Г. Самоорганизация прогнозирующих моделей / А.Г. Ивахненко, Й.А. Мюллер – К.: Техніка, 1985. – 223 с.

57. Інформаційна структура базових знань для прийняття рішень в надзвичайних ситуаціях як основа розробки тестів / Т.Є. Рак, Р.Л. Ткачук, Л.С. Сікора, М.П. Сорочич, В.І. Кунченко-Харченко // Моделювання та інформаційні технології. Зб. наук. пр. ІПМЕ НАН України. – 2013.– Вип.67. – С. 146-159.

58. Інформаційні і системні, технології структуризації ієрархічних систем для забезпечення підтримки рішень при ліквідації надзвичайних ситуацій / Т.Є. Рак, Р.Л. Ткачук, Л.С. Сікора, Б.Л. Якимчук, В.І. Кунченко-Харченко // Моделювання та інформаційні технології, 2013. – №68. – С.151-170.

59. Исследование адаптивного субградиентного метода кластеризации в пространстве вейвлет-преобразования / Г.Ю. Щербакова, В.Н. Крылов, О.В.

Логвинов, Р.А. Писаренко // Радиоэлектронні і комп'ютерні системи. – 2012. – №7. – С.142-146.

60. Ищенко, А.В. Разработка преобразования с обобщенными гребенчатыми масштабными и вейвлет функциями для сегментации изображений / А.В. Ищенко, В.Н. Крылов, М.В. Полякова // Восточно-Европейский журнал передовых технологий. – 2014. – Т.5, №2(71). – С.33-37.

61. 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. – 2010. – Ужгород: УжНУ. – С. 109-110.

62. Кадомский, К.К. О задаче описания ситуации на основе прототипов / К.К. Кадомский, А.А. Каргин // Системні дослідження та інформаційні технології. – 2013. – № 1. – С. 107-117.

63. Казаков, Б.Б. Эффективные методы выделения движения объектов в последовательности изображений / Б.Б. Казаков, С.А. Ивановский // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Сер. Информатика. Телекоммуникации. Управление. – 2009. – Т. 5, № 86. – С. 55-60.

64. Каминскас, В. Последовательное обнаружение изменений параметров динамических объектов в процессе адаптивного управления / В. Каминскас, Д. Виткуте // Статистические проблемы управления. – 1988. – Вып. 83. – С. 187-191.

65. Каргин, А.А. Введение в интеллектуальные машины. Книга 1. Интеллектуальные регуляторы / А.А. Каргин // – Донецк: Норд-Пресс, ДонНУ, 2010. – 526 с.

66. Каргин, А.А. Модели обучения системы мотивированного контекстного ситуационного управления / А.А. Каргин, Н.В. Крачковский // Вісник ХНТУ. – Херсон: ХНТУ. – 2012. – №1(44). – С.257–260.

67. Кендалл, М.Дж. Многомерный статистический анализ и временные

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

68. Клигенс, Н. Методы обнаружения моментов изменения свойств случайных процессов / Н. Клигенс, Л. Тельскис // Автоматика и телемеханика. – 1983. – №10. – С.5-56.

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

70. Коваленко, Н.В. Автоматическое аннотирование видео на основе онтологий и байесовской сети / Н.В. Коваленко // Электротехнические и компьютерные системы. – 2015. – №18 (94). – С. 81-85.

71. Крылов, В.Н. Повышение быстродействия определения границ сегмента упорядоченной текстуры на изображении с однородным фоном с помощью одноканального обнаружения пачки импульсов / В.Н. Крылов, М.В. Полякова // Оптико-електронні інформаційно-енергетичні технології. – 2013. –Т. 15, № 1 . – С. 73-80.

72. Кунцевич, В.М. О решении задачи двумерной дискретной фильтрации. / В.М. Кунцевич // Автоматика и телемеханика. – 1987. – №6. – С. 68–78.

73. Кунцевич, В.М. Синтез оптимальных и адаптивных систем управления. Игровой подход. / В.М. Кунцевич, М.М. Лычак. – К.: Наукова думка, 1985. – 248 с.

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

75. Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов / Ю.П. Лукашин. – М.: Финансы и статистика, 2003 – 414 с.

76. Льюнг, Л. Идентификация систем. Теория для пользователя / Л.

Льюнг. – М.: Наука, 1991. – 432 с.

77. Маликаускас, В. Оценка моментов изменения свойств многомерных случайных последовательностей / В. Маликаускас // Статистические проблемы управления. – 1988. – Вып. 83. – С. 199-204.

78. Маннинг, К.Д. Введение в информационный поиск / К.Д. Маннинг, П. Рагхаван, Х. Шютце; пер с англ. Д. А. Ключин. – М.: Вильямс, 2011. – 528 с.

79. 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 в международные наукометрические базы Ulrich's Periodicals Directory, DRIVER, BASE, Index Copernicus, WorldCat, DOAJ, EBSCO и др.)

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

81. 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 в международные наукометрические базы Ulrich's Periodicals, Index Copernicus.)

82. Машталир, В.П. Точечно-множественные методы в задачах обработки информации / В.П. Машталир. – Харьков: Бизнес Информ, 2001.– 199 с.

83. 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.

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

85. 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.

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

87. 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.

88. 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.

89. 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. – С. 65-69.

90. 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. – Том 2. – С. 431-435.

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

92. 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.

93. 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.

94. 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.

95. Мельник, Р. Экстракция признаков изображений при фрагментации интенсивности / Р. Мельник, Ю. Каличак // Вісник Національного університету “Львівська політехніка”. Комп'ютерні науки та інформаційні технології. – 2011. – №710. – С. 248-255.

96. Мельник, Р.А. Поиск изображений у базах даних за структурными коэффициентами на основе алгоритму триступеневої кластеризації / Р.А. Мельник, Р.Б. Тушницький // Вісник Національного університету “Львівська політехніка”: Комп'ютерні системи проектування. Теорія і практика. – 2009. – № 651. – С. 190–196.

97. Мельник, Р.А. Сегментування зображень алгоритмом ієрархічної кластеризації / Р.А. Мельник, Р.Б. Тушницький // Відбір і обробка інформації. – 2008. – Вип. 28 (104). – С. 97-102.

98. Меры подобия изображений для поисковых систем / Н.Н. Пономаренко, В.В. Лукин, С.К. Абрамов, А.А. Зеленский // Радиоэлектронные и компьютерные системы. – 2007. – №2 (21). – С. 49-58.

99. 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 в

международные наукометрические базы BASE, World Cat.)

100. Миляков, Е. Алгоритм последовательного обнаружения разладок процессов авторегрессии с использованием автокорреляций остатков / Е. Миляков, И. Габриелян, Э. Шпилевский // Статистические проблемы управления. – 1989. – Вып.85. – С. 29-47.

101. 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.

102. Мотыко, А.А. Синтез панорамных изображений с учетом оценки достоверности характерных точек / А.А. Мотыко, Н.А. Обухова // Научно-техн. конф. проф.-препод. состава, научн. сотр. и аспирантов: матер. 63-ей научн.-техн. конф., Санкт-Петербург, 21-25 февраля 2011 г. – СПб.: СПб ГУТ, 2011. – С. 254-256.

103. Мухіна, М.П. Алгоритм розпізнавання рухомого об'єкта за шаблоном з використанням детектора Харріса / М.П. Мухіна, Н.Ю. Баграновська, Ю.І. Почапська // Електроніка та системи управління. – 2010. – №2 (24). – С. 70-75.

104. Нейросетевые методы сжатия изображений / О.Г. Руденко, А.А. Бессонов, М.С. Сныткин, С.В. Мирошниченко, С.О. Руденко // Научные технологии в инфокоммуникациях: обработка и защита информации / под ред. В.М. Безрука, В.В. Баранника. – Харьков: СМИТ, 2013. – 398 с.

105. Нижибицкий, Е.А. Композиции признаков для видеотрекинга при помощи фильтра частиц / Е.А. Нижибицкий // Машинное обучение и анализ данных. – 2015. – Т.1, №11. – С. 1515-1528.

106. Никитин, И.К. Методы поиска нечетких дубликатов видео / И.К. Никитин // Нейрокомпьютеры: разработка, применение.– 2015. – №9. – С.60-66.

107. Никифоров, И.В. Последовательное обнаружение изменения свойств временных рядов / И.В. Никифоров. – М.: Наука, 1983. – 198 с.

108. Обухова, Н.А. Обнаружение и сопровождение движущихся

объектов методом сопоставления блоков / Н.А. Обухова // Информационно-управляющие системы. – 2004 – №1 (8). – С. 30-37.

109. Одейчук, А.Н. Информационная технология прогнозирования с оценкой риска нестационарных временных рядов / А. Н. Одейчук // Вісник Національного технічного університету «ХПІ». – Харків: НТУ «ХПІ» – 2011. – №23. – С. 95-107.

110. Одейчук, А.Н. Метод синтеза моделей прогнозирования временных рядов в информационных управляющих системах / А.Н. Одейчук // Східно-Європейський журнал передових технологій. – 2011. – № 3/9 (51). – С. 42-46.

111. Орлов, Ю.Н. Нестационарные временные ряды. Методы прогнозирования с примерами анализа финансовых и сырьевых рынков / Ю.Н. Орлов, К. П. Осминин. – М. Либроком, 2011. – 384 с.

112. Парамонов, А.И. Формирование модели мира автономного робота на основе интерпретации разнородной сенсорной информации / А.И. Парамонов, А.А. Каргин // Штучний інтелект. – 2014. – №3. – С.121-128.

113. Пелешко, Д.Д. Концепція усунення спотворень зображення, викликаних рухом окремих об'єктів / Д.Д. Пелешко, Н.О. Кустра, А.В. Ключак // Вісник Хмельницького національного університету: Технічні науки. – 2011. – №4. – С.243-246.

114. Пелешко, Д.Д. Розробка програмної системи аналізу кількості і оцінки режиму руху автотранспортних засобів на контрольно-перепускних пунктах / Д.Д. Пелешко, Ю.С. Іванов // Моделювання та інформаційні технології. – 2013. – Вип. 68. – С. 56-61.

115. Пелешко, Д.Д. Супровід рухомих об'єктів у локальних системах відеоспостереження міської інфраструктури / Д.Д. Пелешко, Ю.С. Іванов // Вісник Львівського державного університету безпеки життєдіяльності: збірник наукових праць. – Львів: ЛДУБЖ. – 2013 – № 8. – С. 37-41.

116. Пелешко, Д.Д. Частотна фільтрація наборів зображень / Д.Д. Пелешко // Наук. вісник НЛТУ України: Зб. наук.-техн. праць. – 2008. – Вип. 186. – С. 252-258.

117. Перельман, И.И. Оперативная идентификация объектов управления. / И.И. Перельман. – М.: Энергоатомиздат, 1982. – 272 с.

118. Полякова, М.В. Методология выбора подчеркивающего преобразования при сегментации изображений иерархических объектов и анализе сцен / М. В. Полякова, В. Н. Крылов, Н. П. Волкова // Бионика интеллекта. – 2011. – № 1. – С. 56-65.

119. Пошук траєкторій рухомих об'єктів для автоматизованих систем відеоспостереження / Д.Д. Пелешко, Ю.С. Іванов, А.М. Ковальчук, М.З. Пелешко // Вісник НУ «Львівська політехніка»: Комп'ютерні науки та інформаційні технології. – Львів: Видавництво Львівської політехніки, 2012. – № 744. – С. 239-242.

120. Прогнозирование вероятности скрытых дефектов критического ПО с заданной точностью / Б.М. Конорев, В.В. Сергиенко, В.С. Харченко, Г.М. Жолткевич // Радиоэлектронні і комп'ютерні системи. – 2014. – №5(69). – С.50-54.

121. Прэтт, У. Цифровая обработка изображений: в 2 кн. Кн. 1 / У. Прэтт; пер. с англ. под ред. Д. С. Лебедева. – М.: Мир, 1982. – 312 с.

122. Прэтт, У. Цифровая обработка изображений: в 2 кн. Кн. 2 / У. Прэтт; пер. с англ. под ред. Д. С. Лебедева. – М.: Мир, 1982. – 480 с.

123. Путятин, Е.П. Обработка изображений в робототехнике / Е.П. Путятин, С.И. Аверин. – М.: Машиностроение, 1990. – 320 с.

124. Пытьев, Ю. П. Методы морфологического анализа изображений / Ю. П. Пытьев, А. И. Чуликов. – М.: Физматлит, 2010. – 336 с.

125. Райбман, Н.С., Построение моделей процессов производства / Н.С. Райбман, В.М. Чадеев. – М.: Энергия, 1975. – 376 с.

126. Рак, Т. Є. Інформаційні технології оцінки професійного рівня і інтелекту операторів для роботи в умовах ризику / Т.Є. Рак // Зб. наук. пр. ІПМЕ НАН України. – Вип. 62. – 2012. – С. 202-207.

127. Рак Т.Є. Лінійні форми з елементами алгоритму RSA і додаткове зашумлення у захисті півтонових зображень /Т.Є. Рак, Ю.О. Борзов //

Науковий вісник НЛТУ України, – 2015. – №3. – С. 377-382.

128. Растригин, Л.А. Системы экстремального управления / Л.А. Растригин. – М.: Наука, 1974. – 632 с.

129. Растригин, Л.А. Статистические методы поиска / Л.А. Растригин. – М.: Наука, 1968. – 376 с.

130. Растригин, Л.А. Экстраполяционные методы проектирования и управления / Л.А. Растригин, Ю.П. Пономарев. – М.: Машиностроение, 1986. – 120 с.

131. Рашкевич Ю.М. Зміна роздільної здатності зображень з використанням власних векторів деяких квадратних матриць/ Ю.М. Рашкевич, А.М. Ковальчук, Д.Д. Пелешко// Зб. наук. праць інституту проблем моделювання в енергетиці ім. Г. Є. Пухова НАН України: «Моделювання та інформаційні технології». – 2008. – Вип. 49. – С. 145 – 153.

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

133. Роджерс, Д. Алгоритмические основы машинной графики / Д. Роджерс; пер. с англ. С. А. Вичеса, Г. В. Олохтоновой, П. А. Монахова; под ред. Ю. М. Баяковского, В. А. Галактионова. – М.: Мир, 1989. – 512 с.

134. Розробка програмної системи детектування руху у відеопотоці/ Д.Д. Пелешко, Ю.Є. Иванов, А.В. Ключак, Р.С. Марцишин // Технічні вісті. – Львів, 2009. – №1(29), 2(30). – С.78-80.

135. Руденко, О.Г. Сжатие изображений с помощью нейронных сетей на основе теории адаптивного резонанса / О.Г. Руденко, М.С. Сныткин // Вестник Херсонского государственного технического университета. – 2007. – №4(27). – С. 25-28.

136. Семесенко, М.П. Методы обработки и анализа измерений в научных исследованиях / М.П. Семесенко. – К.: Выща школа, 1983. – 240 с.

137. Симонян, К. А. Адаптивный метод оценки движения в видео / К. А. Симонян, С. В. Гришин, Д. С. Ватолин // Сборник статей молодых

ученых факультета ВМиК МГУ. – 2008. – № 5. – С. 112-119.

138. Система видеонаблюдения за поведением лабораторных животных с автоматической сегментацией на поведенческие акты / А. С. Конушин, Д. П. Ветров, П. А. Воронин и др. // ГрафиКон'2008: сб. трудов конф., Москва, 23-27 июня 2008 г. – М.: МАКС Пресс, 2008. – С. 199-205.

139. Сосулин, Ю.Г. Оптимальное обнаружение сигналов со случайным моментом появления / Ю.Г. Сосулин, М.М. Фишман // Известия АН СССР. Техническая Кибернетика. – 1977. – №3. – С. 149-155.

140. Субоптимальное управление стохастическими процессами. / Е.В. Бодянский, С.Г. Удовенко, А.Е. Ачкасов, Г.К. Вороновский // Харьков: Основа, 1997. – 140 с.

141. Ту, Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес; пер. с англ. И.Б. Гуревич; под ред. Ю.И. Журавлева. – М.: Мир, 1978. – 411 с.

142. Форсайт, Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс; пер. с англ. А.В. Назаренко, И.Ю. Дорошенко; под ред. А.В. Назаренко. – М.: Вильямс, 2004. – 928 с.

143. Фу, К. Последовательные методы в распознавании образов и обучении машин / К. Фу; пер. с англ. Э.Ф. Зайцева; под ред. Л.А. Мееровича, Я.З. Цыпкина. – М.: Наука, 1971. – 256 с.

144. Чебышев, П.Л. Об интерполировании. Избранные труды / П.Л. Чебышев. – М.: Физматгиз, 1965. – 661 с.

145. Чуев, Ю.В. Прогнозирование количественных характеристик процессов / Ю.В. Чуев, Ю.Б. Михайлов, И.В. Кузьмин. – М.: Сов. радио, 1975. – 400 с.

146. Шамша, Б.В. Информационная технология построения моделей прогнозирования нелинейных временных рядов в условиях гетероскедастичности / Б.В. Шамша, А.Ю. Гуд, А.Н. Одейчук // Східно-Європейський журнал передових технологій. – 2010. – № 1/4 (43). – С. 58-61.

147. Шапиро, Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман; пер. с англ. А.А. Богуславского; под ред. С. М. Соколова. – М.: Бином.

Лаборатория знаний, 2006. – 752 с.

148. Шильман, С.В. Итеративное линейное оценивание с регулируемым объемом предистории / С.В. Шильман // Автоматика и телемеханика. – 1983. – №5. – С. 93-98.

149. Ширлев, А.Н. Статистический последовательный анализ / А.Н. Ширлев. – М.: Наука, 1976. – 272 с.

150. Шлезингер, М.И. Десять лекций по статистическому и структурному распознаванию / М.И. Шлезингер, В. Главач. – К.: Наук. думка, 2004. - 546 с.

151. Щербакова Г.Ю. Исследование автоматизированной классификации с использованием вейвлет-преобразования / Г.Ю. Щербакова, В.Н. Крылов// Системи обробки інформації. – 2015. – № 6. – С.153-156.

152. Яне, Б. Цифровая обработка изображений / Б. Яне; пер. с англ. А. М. Измайловой. – М.: Техносфера, 2007. – 584 с.

153. A database and evaluation methodology for optical flow / S. Baker, D. Scharstein, J. P. Lewis et al. // International Journal of Computer Vision. – 2011. – Vol. 92, Iss. 1. – P. 1–31.

154. Adaptive appearance model tracking for still-to-video face recognition / M. Ali Akber Dewan, E. Granger, G.-L. Marcialis, R. Sabourin, F.Roli // Pattern Recognition. – 2016. – Vol. 49, Iss. 1. – P. 129–151.

155. A novel Taylor series based approach for control computation in NN–ANARX structure based control of nonlinear systems / J. Belikov, K. Vassiljeva, E. Petlenkov, S. Nomm // 27th Chinese Control Conference: Proc., China, Kunming, 16-18 July, 2008 – Kunming: Beihang University Press, 2008. – P. 494-498.

156. A survey of content-based image retrieval with high-level semantics / Y. Liu, D. Zhanga, G. Lua, W.-Y. Ma // Pattern Recognition. – 2007. – Vol. 40, Iss. 1. – P. 262-282.

157. 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.

158. Advances in image and video segmentation / Zhang Y.-J. (ed.). – Hershey- London-Melbourne-Singapore: IRM Press, 2006. – 457 p.

159. Aktan, M. The neural network modeling of suspended particulate matter with autoregressive structure / M. Aktan, H. Bayraktar // Ecology. – 2010. – Vol. 19, Iss. 74. – P. 32-37.

160. Almeida, J. VISON: video summarization for online applications / J. Almeida, N.J. Leite, R.S. Torres // Pattern Recognition Letters. – 2012. – Vol. 33, Iss. 4. – P. 397-409.

161. An efficient approach to content-based object retrieval in videos / C. Hong, N. Li, M. Song, et al. // Neurocomputing. – 2011. – Vol. 74, Iss. 17. – P. 3565-3575.

162. An improved keyframe extraction method based on HSV color space / Z. Qu, L. Lin, T. Gao, Y. Wang // Journal of Software. – 2013. – Vol. 8, Iss. 7. – P. 1751-1758.

163. An online air pollution forecasting system using neural networks / A. Kurt, B. Gulbagci, F. Karaca, O. Alagha // Environmental International. – 2008. – Vol. 34, Iss. 5. – P. 592-598.

164. Antani, S. A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video / S. Antani, R. Kasturi, R. Jain // Pattern Recognition. – 2002. – Vol. 35, Iss. 4. – P. 945-965.

165. Asghar, M.N. Video indexing: a survey / M.N. Asghar, F. Hussain, R. Manton // International Journal of Computer and Information Technology. – 2014. – Vol. 3, Iss. 1. – P. 148-169.

166. Athanasiadis, I.N. Classification techniques for air quality forecasting / I.N. Athanasiadis, K.D. Karatzas, P.A. Mitkas // 17th European Conference on Artificial Intelligence, proc., Italy, Riva del Garda, 28 August – 1 September, 2006. – Amsterdam: IOS Press, 2006, Vol. 4. – P. 4-1–4-7.

167. Automatic initial segmentation of speech signal based on symmetric

matrix of distances / D. Peleshko, Y. Pelekh, M. Rashkevych, Y. Ivanov, I. Verbenko // *International Journal of Computer Science and Technology*. – 2014. – Vol. 13, Iss. 9. – P. 4783-4790.

168. Automatic summarization and annotation of videos with lack of metadata information / D.P. Papadopoulos, V.S. Kalogeiton, S.A. Chatzichristofis, N. Papamarkos // *Expert Systems with Applications*. – 2013. – Vol. 40, Iss. 14. – P. 5765-5778.

169. Badavas, P.C. Real-time statistical process control / P.C. Badavas. – N.Y.: Prentice-Hall Inc., 1993. – 232 p.

170. Billings, A. Extended model set, global data and threshold model identification of severely non-linear systems / A. Billings, S. Chen // *International Journal of Control*. – 1989. – Vol. 50, Iss. 5. – P. 1897-1923.

171. Bohm, C. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases / C. Bohm, S. Berchtold, D.A. Keim // *ACM Computing Surveys*. – 2001. – Vol. 33, Iss. 3. – P. 322-373.

172. Brown, M. Neural networks for modeling and control / M. Brown, C.J. Harris // *Advances in Intelligent Control* / C.J. Harris (ed.). – 1994. – London: Taylor and Francis. – P. 85-112.

173. Brown, R.G. Smoothing, forecasting, and prediction of discrete time series / R.G. Brown. – N.Y.: Prentice Hall Inc., 1963. – 468 p.

174. Bulut, E. Key frame extraction from motion capture data by curve saliency / E. Bulut, T. Capin // *Computer Animation and Social Agents: proc. of the 20th annual conf., Belgium, Hasselt, 11-13 June, 2007*. – Hasselt: UHasselt, 2007. – P. 63-67.

175. Casual video segmentation using supersedes and graph matching / V.N. Gangapure, S. Nanda, A.S. Chowghury, X. Jiang // *Graph-Based Representations in Pattern Recognition: proceedings of 10th IAPR-TC-15 intern. workshop, China, Beijing, 13-15 May, 2015*. – *Lecture Notes in Computer Science*. – Berlin-Heidelberg-New York: Springer, 2015. – Vol. 9069. – P. 282-291.

176. Chan, L.W. An adaptive learning algorithm for back propagation

networks / L.W. Chan, F. Fallside // *Computer Speech & Language*. – 1987. – Vol. 2, Iss. 3-4. – P. 205-218.

177. Chen, H.-L. Intellectual access to images / Chen, H.-L., Rasmussen, E.M. // *Library Trends*. – 1999. – Vol.48, Iss. 2. – P. 291-302.

178. Chen, L. Video sequence matching based on temporal ordinal measurement / L. Chen, F.W.M. Stentiford // *Pattern Recognition Letters*. – 2008. – Vol. 29, Iss. 13. – P. 1824-1831.

179. Cheng, P.J. A new content-based access method for video databases / P.J. Cheng, W.P. Yang // *Information Sciences*. – 2011. – Vol. 118, Iss. 1-4. – P. 37-73.

180. Cheng, S.-C. Visual object retrieval via block-based visual-pattern matching / S.-C. Cheng, C.-T. Kuo, H.-J. Chen // *Pattern Recognition*. – 2007. – Vol. 40, Iss. 6. – P. 1695-1710.

181. Chow, E.Y. Issues in the development of a general design algorithm for reliable failure detection / E.Y. Chow, A.S. Willsky // *19-th IEEE Conference on Decision and Control: proc., USA, Albuquerque: 10-12 December, 1980*. – N.Y.: IEEE, 1980, Vol. 2. – P. 1006-1012.

182. Chowdhury, F.N. Input-output modeling of nonlinear systems with time-varying linear models / F.N. Chowdhury // *IEEE Trans. on Automatic Control*. – 2000. – Vol. 45, Iss. 7. – P. 1355-1358.

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

184. Cichocki, A. Neural networks for optimization and signal processing / A. Cichocki, R. Unbehauen. – Stuttgart: Teubner, 1993. – 526 p.

185. Classifier based graph construction for video segmentation / A. Khoreva, F. Galasso, M. Hein, B. Schiele // *Computer Vision and Pattern Recognition (CVPR 2015): proc. of the IEEE conference, USA, Boston, 7-12 June,*

2015. – Washington: IEEE CS. – P. 951-960.

186. Concept of critical software independent verification based on invariant-oriented model-checking approach / G.N. Zholtkevych, et al. // *Радіоелектронні і комп'ютерні системи*. – 2012. – 5 (57). – С. 184-190.

187. Content-based image retrieval by feature adaptation and relevance feedback / A. Grigorova, F.G.B. De Natale, C. Dagli, T.S. Huang // *IEEE Transactions on Multimedia*. – 2007. – Vol. 9, Iss. 6. – P. 1183-1192.

188. Corani, G. Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. / Corani G. // *Ecological Modeling*. – 2005. – Vol. 185, Iss. 2-4. – P. 513-529.

189. Cooper, M. Summarizing video using non-negative similarity matrix factorization / M. Cooper, J. Foote // *Multimedia Signal Processing: proc. of IEEE 5th workshop, USA, St. Thomas, 9-11 December, 2002*. – Washington: IEEE CS, 2002. – P. 25-28.

190. Cybenko, G. Approximation by superposition of a sigmoidal function / G. Cybenko // *Mathematics of Control, Signals and Systems*. – 1989. – Vol. 2, Iss. 4. – P. 303-314.

191. Dadashi, R. AVCD-FRA: a novel solution to automatic video cut detection using fuzzy-rule-based approach / R. Dadashi, H.R. Kanan // *Computer Vision and Image Understanding*. – 2013. – Vol. 117, Iss. 7. – P. 807-817.

192. Divakaran, A. *Multimedia content analysis: theory and applications* / A. Divakaran (ed.). – N.Y.: Springer Science+Business Media. – 2009. – 390 p.

193. Ejaz, N. Efficient visual attention based framework for extracting key frames from videos / N. Ejaz, I. Mehmood, S.W. Baik // *Signal Processing: Image Communication*. – 2013. – Vol. 28, Iss. 1. – P. 34-44.

194. Ejaz, N. Adaptive key frame extraction for video summarization using an aggregation mechanism / N. Ejaz, T.B. Tariq, S.W. Baik // *Journal of Visual Communication and Image Representation*. – 2012. – Vol. 23, Iss. 7. – P. 1031-1040.

195. Exploiting semantic features for image retrieval at CLEF 2005 /

J. L. Martínez-Fernández, J. Villena, A. García-Serrano et al. // Cross Language Evaluation Forum (CLEF) / ImageCLEF workshop on Image and Video Retrieval Evaluation, 2005: working notes for the CLEF 2005, Austria, Vienna, 21-23 September, 2005. – Vienna: CEUR, 2005. – P. 1-9.

196. Embedding spatial information into image content description for scene retrieval / N.V. Hoàng, V. Gouet-Brunet, M. Rukoz, M. Manouvrier // Pattern Recognition. – 2010. – Vol. 43, Iss. 9. – P. 3013-3024.

197. Esau, I. On application of artificial neural network methods in large-eddy simulations with unresolved urban surfaces / I. Esau // Modern Applied Science. – 2010. – Vol. 4, Iss. 8. – P. 3-11.

198. Estimating of critical software latent faults presence with required trustworthiness / G. Zholtkevych, B. Konorev, V. Sergiienko, V. Kharchenko // Reliability: Theory & Applications. – 2014. – Vol. 9, Iss. 2/33. – P. 27-35.

199. Evolving the neural network model for forecasting air pollution time series / H. Niska, T. Hiltunen, A. Karppinen, et al. // Engineering Applications of Artificial Intelligence. – 2004. – Vol. 17, Iss. 2. – P. 159-167.

200. 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. (Входит в международные наукометрические базы Scopus, DBLP, ACM Digital Library, MS Academic Search и др.)

201. Fayka, M.B. Particle swarm optimization based video abstraction / M. B. Fayka, H. A. El Nemr, M.M. Moussa // Journal of Advanced Research. – 2010. – Vol. 1, Iss. 2. – P. 163-167.

202. Feature extraction and clustering for dynamic video summarization / H. Zhou, A. H. Sadka, M. R. Swash et al. // Neurocomputing. – 2010. – Vol. 73, Iss. 10-12. – P. 1718-1729.

203. Feng, L. Semantic concept co-occurrence patterns for image annotation

and retrieval / L. Feng, B. Bhanu // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2016. – Vol. 38, Iss. 4. – P. 785-799.

204. Finding temporally consistent occlusion boundaries in videos using geometric context / S.H. Raza, A. Humayun, I. Essa, et al. // IEEE Winter Conference Applications of Computer Vision (WACV): proc., USA, Waikoloa Beach, 5-9 January, 2015. – Washington: IEEE CS. – P. 1022-1029.

205. Fuzzy models and algorithms for pattern recognition and image processing / J. C. Bezdek, J. Keller, R. Krisnapuram, N. R. Pal. – N.Y.: Springer, 2005. – 776 p.

206. Gokilavani, C. Segmentation of pedestrian video using thresholding algorithm and its parameter analysis / C. Gokilavani, M. Gowtham. // International Journal of Applied Research. – 2015. – Vol. 4, Iss 1. – P.43-46.

207. Gong, Y. Video summarization using singular value decomposition / Y. Gong, X. Liu. // Computer Vision and Pattern Recognition: proc. of IEEE conf., USA, Hilton Head Island, 13-15 June, 2000. – Washington: IEEE CS, 2000. – Vol. 2. – P. 174-180.

208. Geetha, P. A survey of content-based video retrieval / P.A. Geetha, V. Narayanan // Journal of Computer Science. – 2008. – Vol. 4, Iss. 6. – P. 474-486.

209. Handbook of time series analysis recent theoretical developments and applications / B. Schelter, M. Winterhalder, J. Timmer (eds.). – Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA, 2006. – 496 p.

210. Hammoud, R. A probabilistic framework of selecting effective key frames from video browsing and indexing / R. Hammoud, R. Mohr // Real-Time Image Sequence Analysis : proc. of intern. workshop, Finland, Oulu, 31 August – 1 September, 2000. – Oulu, 2000. – P. 79-88.

211. Haralick, R. M. Textural features for image classification / R. M. Haralick, K. Shanmugam, I. Dinstein // IEEE Transactions on Systems, Man, and Cybernetics. – 1973. – Vol. 3, Iss. 6. – P. 610–621.

212. Harris, C.J. Intelligent control. Aspects of fuzzy logic and neural nets / C.J. Harris, C.G. Moore, M. Brown. – Singapore: World Scientific, 1993. – 380 p.

213. Haykin, S. Neural networks: a comprehensive foundation / S. Haykin. – Upper Saddle River, N.J.: Prentice-Hall, Inc., 1999. – 842 p.
214. Hornik, K. Approximation capabilities of multilayer feedforward networks / K. Hornik // *Neural Networks*. – 1991. – Vol. 4, Iss. 2. – P. 251-257.
215. Hornik, K. Multilayer feedforward networks are universal approximators / K. Hornik, M. Stinchcombe, H. White // *Neural Networks*. – 1989. – Vol. 2, Iss. 5. – P. 359-366.
216. Hua, X.-S. AVE: automated home video editing / X.-S. Hua, L. Lu, H.-J. Zhang // *Multimedia'03: proc. of the 11th ACM intern. conf., USA, New York, 2-8 November, 2003*. – N.Y.: ACM, 2003. – P. 490-497.
217. Image partition transforms for faithful segmentation search / D. Kinoshenko, S. Mashtalir, K. Shcherbinin, E. Yegorova // *Information Research & Applications: proc. of fifth intern. conf. I.Tech, Bulgaria, Varna, 26-30 June, 2007*. – Sofia: ITHEA, 2007. – Vol. 2. – P. 385-389.
218. Image quotient set transforms in segmentation problems / D. Kinoshenko, S. Mashtalir, K. Shcherbinin, E. Yegorova // *International Journal «Information Technologies & Knowledge»*. – 2008. – Vol. 2. – Iss. 4. – P. 372-378. (Входит в международные наукометрические базы BulDML, OpenAIRE, BASE и др.)
219. Image retrieval: Ideas, influences, and trends of the new age / Datta R. et al. // *ACM Computing Surveys (CSUR)*. – 2008. – Vol. 40. – Iss. 2. – P. 51-60.
220. Interactive digital photomontage / A. Agarwala, M. Dontcheva, M. Agrawala et al. // *ACM Transactions on Graphics*. – 2004. – Vol. 23, Iss. 4. – P. 294–301.
221. Isermann, R. Process fault detection based modeling and estimating methods – a survey. / R. Isermann // *Automatica*. – 1984. – Vol. 20, Iss. 4. – P. 387-404.
222. Juselis, K. The cointegrated VAR-model / K. Juselis. – N.Y.: Oxford University Press, 1994. – 799 p.
223. Kaczmarz, S. Approximate solution of system of linear equations / S.

Kaczmarz // International Journal of Control. – 1993. – Vol. 57, Iss. 6. – P. 1269-1271.

224. Kanade, S.S. Dominant color based extraction of key frames for sports video summarization / S. S. Kanade, P. M. Patil // International Journal of Advances in Engineering & Technology. – 2013. – Vol. 6, Iss. 1. – P. 504-512.

225. Kashyap, R.L. Dynamic stochastic models from empirical data / R.L. Kashyap, A.R. Rao. – N.Y., San Francisco, London: Academic Press, 1976. – 334 p.

226. Key-frame extraction using dominant-set clustering / X. Zeng, W. Hu, W. Liy et al. // Multimedia and Expo : proc. of intern. conf., Hannover, 23-26 June, 2008. – Washington: IEEE CS, 2008. – P. 1285-1288.

227. Khurana, K. Video annotation methodology based on ontology for transportation domain / K. Khurana, M. B. Chandak // International Journal of Advanced Research in Computer Science and Software Engineering. – 2013. – Vol. 3, Iss. 6. – P. 540-548.

228. Kingdom, S. Intelligent systems and financial forecasting / S. Kingdom – Berlin: Springer-Verlag, 1997. – 227 p.

229. 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.

230. Kinoshenko, D. Temporal video segmentation via spatial image segmentation / D. Kinoshenko, S. Mashtalir, V. Shlyakhov // International Journal «Information Technologies & Knowledge». – 2013. – Vol. 7. – Iss 3. – P. 212-219. (Входит в международные наукометрические базы BulDML, OpenAIRE, BASE и др.)

231. Kinoshenko, D. A partition metric for clustering features analysis / D. Kinoshenko, V. Mashtalir, V. Shlyakhov // International Journal “Information Theories and Applications”. – 2007. – Vol. 14, Iss. 3. – P. 230-236.

232. Kotta, Ü. Two approaches for state space realization of NARMA models: bridging the gap / Ü. Kotta, N. Sadegh // Mathematical and Computer

Modeling of Dynamical Systems. – 2002. – Vol. 8, Iss. 1. – P. 21-32.

233. Kruschke, J.K. Benefits of gain: speeded learning and minimal hidden layers in back propagation networks / J.K. Kruschke, J.R. Movellan // *IEEE Transactions on Systems, Man, and Cybernetics*. – 1991. – Vol. 21, Iss. 1. – P. 67-70.

234. Kundu A. A survey on video segmentation the future roadmap / A. Kundu, N. Janwe // *International Journal of Modern Trends in Engineering and Research*. – 2015, Vol. 2, Iss. 3. – P.527-534.

235. Learning mid-level features for recognition / Y.-L. Boureau, F. Bach, Y. LeCun, J. Ponce // *Computer Vision and Pattern Recognition: proc. of IEEE Computer Society conf.*, USA, San Francisco, 13-18 June, 2010. – Washington: IEEE CS. – 2010. – P. 2559-2566.

236. Learning to segment moving objects in videos / K. Fragkiadaki, P. Arbeláez, P. Felsen, J. Malik // *Computer Vision and Pattern Recognition (CVPR 2015): proc. of the IEEE conf.*, USA, Boston, 7-12 June, 2015. – Washington: IEEE CS. – 2015. – P. 4083-4090.

237. Lee, W.-T. Histogram-based interest point detectors / W.-T. Lee, H.-T. Chen // *Computer Vision and Pattern Recognition (CVPR 2009): proc. of the IEEE conf.*, USA, Miami, 20-25 June, 2009. – N.Y.: Curran Associates, Inc. – 2009. – Vol. 3. – P. 1590-1596.

238. Lewis, D.C. Industrial and business forecasting methods / D.C. Lewis. – London: Butterworths Scientific, 1982. – 145 p.

239. Lira, T.S. Air quality prediction in Uberlandia, Brasil, using linear models and neural networks / T.S. Lira, M.A.S. Barrozo, A.J. Assis // *Computer Aided Process Engineering: proc. of 17th European Symp. Romania, Bucharest 27-30 May 2007*. – Amsterdam: Elsevier. – 2007.– Vol. 24. – P. 51-56.

240. Liang, B. Design of video retrieval system using MPEG-7 descriptors / B. Liang, W. Xiao, X. Liu // *Procedia Engineering*. – 2012. – Vol. 29. – P. 2578-2582.

241. Lin, G.-S. Detection of frame duplication forgery in videos based on spatial and temporal analysis / G.-S. Lin, J.-F. Chang // *International Journal of*

Pattern Recognition and Artificial Intelligence. – Vol. 26, Iss. 7. – 2013. – P. 1-18.

242. Liu, B. Multiclass semantic video segmentation with object-level active inference / B. Liu, X. He // Computer Vision and Pattern Recognition (CVPR 2015): proc. of the IEEE conf., USA, Boston, 7-12 June, 2015. – Washington: IEEE CS. – 2015. – P. 4286-4294.

243. Liu G.H. Content-based image retrieval using color difference histogram / G.H. Liu, J.Y. Yang // Pattern Recognition. – 2013. – Vol. 46, Iss 1. – P. 188-198.

244. Liu, T. A novel video key-frame-extraction algorithm based on perceived motion energy model / T. Liu, H.-J. Zhang, F. Qi // IEEE Transactions on Circuits and Systems for Video Technology. – 2003. – Vol. 13, Iss. 10. – P. 1006-1013.

245. Liu, G. Key frame extraction from MPEG video stream / G. Liu, J. Zhao // Computer Science and Computational Technology (ISCST '09): proceedings of the second international symposium, China, Huangshan, 26-28 December, 2009. – Oulu: Academy Publisher, 2009. –P. 7-11.

246. Liu G.H. Image retrieval based on multi-texton histogram / Liu G. H., L. Zhangb, Y.-K. Houd, Z.-Y. Lic // Pattern Recognition. – 2010. – Vol. 43, Iss. 7. – C. 2380-2389.

247. Liu, L. Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition / L. Liu, L. Shao, P. Rockett // Pattern Recognition. – 2013. – Vol. 46, Iss. 7. – P. 1810-1818.

248. Liwicki S. Online kernel slow feature analysis for temporal video segmentation and tracking / S. Liwicki, S.P. Zafeiriou, M. Pantic // IEEE Transactions on Image Processing. – 2015. – Vol. 24, Iss. 10. – P. 2955-2970.

249. Ljung, L. System identification: Theory for the User. / L. Ljung. – Englewood Cliffs: PTR Prentice Hall, 1987. – 512 p.

250. Mandic, D.P. Recurrent neural networks for prediction / D.P. Mandic, J.A. Chambers. – Chichester: John Wiley & Sons, Ltd., 2001. – 285 p.

251. 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 в международные наукометрические базы ANED, Index Copernicus, Jour Informatics)

252. Masters, T. Neural, novel & hybrid algorithms for time series prediction / T. Masters. – N.Y.: John Wiley & Sons, Inc., 1995. – 514 p.

253. Masters, T. Practical neural networks recipes in C++ / T. Masters. – San Diego: Academic Press, Inc., 1993. – 493 p.

254. 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. (Входит в международные наукометрические базы Scopus, DBLP, BASE, MS Academic Search.)

255. 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.

256. 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 в международные наукометрические базы INSPEC, OpenAIRE. ADAT, Inspec.)

257. 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., Belarus, Minsk, 28-31 May, 2014. – Minsk: UIIP NASB, 2014. – P. 181-186.

258. Mashtalir, S.V. Stabilization of key frame descriptions with higher order Voronoi diagram / S.V. Mashtalir, O.D. Mikhnova // Бионика интеллекта. –

2013. – № 1(80). – P. 68-72.

259. 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. – Iss. 2. – P. 68-79. (Входит в международные наукометрические базы ACM Digital Library, Bacon's Media Directory, Index Copernicus, Cabell's Directories, DBLP и др.)

260. Machine learning in computer vision / Sebe, N., Cohen, I., Garg, A., Huang, Th.S. – Dordrecht: Springer, 2005. – 242 p.

261. Metrical properties of nested partitions / 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. – P. 57-58.

262. 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.

263. Montgomery, D.C. Forecasting and time series analysis / D.C. Montgomery, I.A. Johnson, J.S. Yardiner. – N.Y.: McGraw-Hill, 1990. – 394 p.

264. Nagaraja, N.S. Video segmentation with just a few strokes / N.S. Nagaraja, F.R. Schmidt, T. Brox // *Computer Vision (ICCV): proc IEEE intern. conf.*, Chile, Santiago, 7-13 December, 2015. – IEEE CS, 2015. – P. 3235-3243.

265. Nagasaka, A. Automatic video indexing and full-video search for object appearances / A. Nagasaka, Y. Tanaka // *Visual Database Systems: proc. of the IFIP second working conf.*, Budapest, 30 September – 3 October, 1991. – Amsterdam: North-Holland Publishing Co., 1991. – Vol. A-7. – P. 113-127.

266. Nelles, O. Nonlinear systems identification: from classical approaches

to neural networks and fuzzy models / O. Nelles. – Berlin-Heidelberg-N.Y.: Springer-Verlag, 2001. – 785 p.

267. 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, Iss. 3. – P. 232-242.

268. Nonparametric motion feature for key frame extraction in sports video / L. Li, X. Zhang, Y. Wang et al. // Pattern Recognition: proc. of Chinese conf., Beijing, 22-24 October, 2008. – Piscataway: IEEE, 2008. – P. 1-5.

269. NUS-PRO: A new visual tracking challenge / A. Li, M. Lin, Y. Wu et al. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2016. – Vol. 38, Iss. 2. – P. 335-349.

270. 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 в международные наукометрические базы Engineering Village (EI)-GEOBASE, ABI/INFORM (ProQuest), Australian Research Council Journal List, Ulrich's periodicals directory, Index Copernicus, EIJ, DOAJ.)

271. Palit, A.K. Computational intelligence in time series forecasting: theory and engineering applications / A.K. Palit, D. Popovic. – London: Springer-Verlag, 2005. – 372 p.

272. Pao, Y.H. Adaptive pattern recognition and neural networks / Y.H. Pao. – Addison-Wesley: Reading, 1989. – 320 p.

273. Partially-supervised learning from facial trajectories for face recognition in video surveillance / M. De la Torre Gomera, E. Granger, R. Sabourin, D. Gorodnichy // Information Fusion. – 2014. – Vol. 24. – P. 31-53.

274. Patra, J.C. Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks / J.C. Patra, A.C. Kot // IEEE Transactions on Systems, Man, and Cybernetics. – 2002. – Vol. 32, Iss. 4. – P. 505-511.

275. Perceptual feature selection for semantic image classification /

D. Depalov, T. N. Pappas, D. Li, B. Gandhi // Image Processing (ICIP 2006): proc. of 13th intern. conf., Atlanta, 8-11 October, 2006. – Washington: IEEE CS, Vol. 2. – 2006. – P. 2921-2924.

276. Perez, P. An integrated neural network model for PM10 forecasting / P. Perez, J. Reyes // Atmospheric Environment. – 2006. – Vol. 40, Iss. 16. – P. 2845-2857.

277. Perez, P. Prediction of PM2.5 concentration several hours in advance using neural networks in Santiago, Chile. / P. Perez, A. Trier, J. Reyes // Atmospheric Environmental. – 2000. – Vol. 34, Iss. 8. – P. 1189-1196.

278. Petkovic, M. Content-based video retrieval: a database perspective (Multimedia systems and applications) / M. Petkovic, W. Jonker. – Boston-Dordrecht-London: Kluwer Academic Publishers, 2004. – 168 p.

279. Pham, D.T. Modeling and prediction using GMDH networks of adalines with nonlinear preprocessors / D.T. Pham, X. Liu // International Journal of Systems Science. – 1994. – Vol. 25, Iss. 11. – P. 1743-1759.

280. Pham, D.T. Neural networks for identification, prediction and control / D.T. Pham, X. Liu. – London: Springer-Verlag, 1995. – 238 p.

281. Porter, S. Temporal video segmentation and classification of edit effects / S. Porter, M. Mirmehdi, B. Thomas // Image and Vision Computing. – 2003. – Vol. 21, Is. 13-14. – P. 1097-1106.

282. Pouliot, G. Emission processing for an air quality forecasting model/ G. Pouliot, T. Pierce // Emission Inventories: proc. of 12th int. conf. – USA, San Diego, 28 April – 1 May, 2003. – San Diego: EIIP. – 2003. – P. 1-9.

283. Pouliezos, A.D. Real time fault monitoring of clustering processes / A.D. Pouliezos, Y.S. Stavrakalis. – Dordrecht: Kluwer Academic Publishers, 1994. – 542 p.

284. Prest, A. Explicit modeling of human-object interactions in realistic videos / A. Prest, V. Ferrari, and C. Schmid // IEEE Transactions on Pattern Analysis and Machine Intelligence, –2013. – Vol. 35, Iss. 4. – P. 835-848.

285. Query by image and video content: the QBIC system / M. Flickner, H.

- Sawhney, W. Niblack, et al. // IEEE Computer. Vol. 28, Iss. 9. – 1995. – P. 23-32.
286. Rasmussen, E.M. Indexing images / Rasmussen, E.M. // Annual Review of Information Science and Technology. – 1997. – Vol. 32, Iss. 2. – P. 169-196.
287. Rathod, G.I. An algorithm for shot boundary detection and key frame extraction using histogram difference / G.I. Rathod, D.A. Nikam // International Journal of Emerging Technology and Advanced Engineering. – 2013. – Vol. 3, Iss. 8. – P. 155-163.
288. Reich, S.L. Artificial neural network for the identification of unknown air pollution sources / S.L. Reich, D.R. Gomez, L.E. Dawidowski // Atmosphere Environment. – 1999. – Vol. 33, Iss. 18. – P. 3045-3052.
289. Ren, J. Activity-driven content adaptation for effective video summarization / J. Ren, J. Jiang, Y. Feng // Journal of Visual Communication and Image Representation. – 2010. – Vol. 21, Iss. 8. – P. 930-938.
290. Roth, V. Content-based retrieval from digital video / V. Roth // Image and Vision Computing. – 1999. – Vol. 17, Iss. 7. – P. 531-540.
291. Sakarya, U. Video scene detection using graph-based representations / U. Sakarya, Z. Telatar // Signal Processing: Image Communication. – 2010. – Vol. 25, Iss. 10. – P. 774-783.
292. Scalable feature matching by dual cascaded scalar quantization for image retrieval / W. Zhou, M. Yang, X. Wang et al. // IEEE Transactions on Pattern Analysis and Machine Intelligence. – Vol. 38, Iss. 1. – 2016. – P. 159-171.
293. Sebe, N. Comparing salient point detectors / N. Sebe, M.S. Lew // Pattern Recognition Letters. – 2003. – Vol. 24, Iss. 1. – P. 89-96.
294. Semantic video segmentation: exploring inference efficiency / S. Tripathi, S. Belongie, Y. Hwang, T. Nguyen // SoC Design (ISOCC 2015): proc. of 12th int. conf., South Korea, Gyeongju, 2-5 November, 2015. Piscataway: Curran Associates, Inc. – 2015. – P. 157-158.
295. Shanmugam, T.N. An enhanced content-based video retrieval system based on query-clip / T.N. Shanmugam, P. Rajendran // International Journal of

Research and Reviews in Applied Sciences. – 2009. – Vol. 1, Iss. 3. – P. 236-253.

296. Sharkeya, A.J.C. On combining artificial neural nets / A.J.C. Sharkeya // *Connection Science*. – 1996. – Vol. 8, Iss. 3. – P. 299-314.

297. Shot boundary detection and label propagation for spatio-temporal video segmentation / S. Piramanayagam; E. Saber; N.D. Cahill; D. Messinger // *Image Processing: Machine Vision Applications VIII: proc. of SPIE-IS&T Electronic Imaging, USA, San Francisco, 8 February, 2015*. – Springfield: SPIE, 2015, Vol. 9405. – P. 9405-12 (7 p.).

298. Shot retrieval based on fuzzy evolutionary aiNet and hybrid features / X.-H. Li, Y.-Z. Zhan, J. Ke, H.-W. Zheng // *Computers in Human Behavior*. – 2011. – Vol. 27, Iss. 5. – P. 1571-1578.

299. Smeaton, A.F. Video shot boundary detection: seven years of TRECVID activity / A.F. Smeaton, P. Over, A.R. Doherty // *Computer Vision and Image Understanding*. – 2010. – Vol. 114, Iss. 4. – P. 411-418.

300. Sonka, M. Image processing, analysis, and machine vision / M. Sonka, V. Hlavac, R. Boyle. – 3rd ed. – Toronto: Thomson, 2007. – 850 p.

301. Spatio-temporal video segmentation of static scenes and its applications / Jiang H., G. Zhang, H. Wang, H. Bao // *IEEE Transactions on Multimedia*. – 2015. – Vol. 17, Iss. 1. – P. 3-15.

302. Sun, J. Streaming analysis of track data from video / J. Sun, D.W. Messinger // *Geospatial Informatics, Fusion, and Motion Video Analytics V: proc. of SPIE, USA, Baltimore, 20-21 April, 2015*. Washington: SPIE, 2015, Vol. 9473. – P. 9473-1 (15 p.).

303. Szeliski, R. Computer vision. Algorithms and applications / R. Szeliski. – London: Springer, 2011. – 813 p.

304. Tamane, S. Content based image retrieval using high level semantic features / S. Tamane // *Computing for Nation Development. INDIACOM-2008: proc. of the 2nd national conf., India, New Delhi, 8-9 February, 2008*. – New Delhi: BVICAM, 2008. – P. 204-207.

305. Thakare, S. Intelligent processing and analysis of image for shot

boundary detection / S. Thakare // International Journal of Emerging Technology and Advanced Engineering. – 2012. – Vol. 2, Iss. 2. – P. 208-212.

306. Tousch, A.-M. Semantic hierarchies for image annotation: A survey / A.-M. Tousch, S. Herbin, J.-Y. Audibert // Pattern Recognition. – 2012. – Vol. 45, Iss. 1. – P. 333-345.

307. Trigg, D.W. Exponential smoothing with an adaptive response rate / D.W. Trigg, A.G. Leach // Operational Research. – 1967. – Vol. 18, Iss 1. – P. 53-59.

308. Tsai, Y.-H. Salient points reduction for content-based image retrieval / Y.-H. Tsai // International Journal of Computer, Electrical, Automation, Control and Information Engineering. – 2009. – Vol. 3, Iss. 1. – P. 87-90.

309. Tziakos, I. Video event segmentation and visualisation in non-linear subspace / I. Tziakos, A. Cavallaro, L.-Q. Xu // Pattern Recognition Letters. – 2009. – Vol. 30, Iss. 2. – P. 123-131.

310. Vassiljeva, K. State-space control of nonlinear systems identified by ANARX and neural network based SANARX models / K. Vassiljeva, E. Petrenkov, J. Belikov // Computational Intelligence: proc. of IEEE World Congress (WCCI 2010), Spain, Barcelona, 18-23 July, 2010. – Piscataway: IEEE CS. – 2010. – P. 3816-3823.

311. Vázquez-Martín, R. Spatio-temporal feature-based keyframe detection from video shots using spectral clustering / R. Vázquez-Martín, A. Bandera // Pattern Recognition Letters. – 2013. – Vol. 34, Iss. 7. – P. 770-779.

312. Veltkamp, R.C. State-of-the-art in content-based image and video retrieval / R.C. Veltkamp, H. Burkhardt, H.-P. Kriegel // Computational Imaging and Vision. – Netherlands: Kluwer Academic Publishers. – 2001. – 343 p.

313. Video abstraction based on the visual attention model and online clustering / Q.-G. Ji, Z.-D. Fang, Z.-H. Xie, Z.-M. Lu // Signal Processing: Image Communication. – 2013. – Vol. 28, Iss. 3. – P. 241-253.

314. Video search and mining / D. Schonfeld, C. Shan, D. Tao, L. Wang (eds.) // Studies in Computational Intelligence. – Berlin Heidelberg: Springer-Verlag. – 2010. – Vol. 287. – 388 p.

315. Video summarization from spatio-temporal features / R. Laganière, R. Bacco, A. Hocevar et al. // TRECVID Video Summarization: proc. of the 2nd ACM TVS workshop, Canada, Vancouver, 31 October, 2008. – N.Y.: ACM, 2008. – P. 144-148.

316. Wang, L. Video action detection with relational dynamic-poselets / L. Wang, Y. Qiao, and X. Tang // Computer Vision – ECCV2014: proc. of 13th European conf. Part V, Switzerland, Zurich, 6-12 September, 2014. – Lecture Notes in Computer Science. – Cham-Heidelberg-New York-Dordrecht-London: Springer, 2014. Vol. 8693. – P. 565-580.

317. Wang, W. Saliency-aware geodesic video object segmentation / W. Wang, J. Shen, F. Porikli // Computer Vision and Pattern Recognition (CVPR 2015): proc. of the IEEE conf., USA, Boston, 7-12 June, 2015. – Washington: IEEE CS. – 2015. – P.3395-3402.

318. Wolf, W. Key frame selection by motion analysis / W. Wolf // Acoustics, Speech and Signal Processing: proc. of intern. conf., USA, Atlanta, 7-10 May 1996. – Washington: IEEE CS, 1996. – Vol. 2. – P. 1228-1231.

319. Yang, X. Genetic keyframe extraction for soccer video / X. Yang, Z. Wei // Procedia Engineering. – 2011. – Vol. 23. – P. 713-717.

320. Zholtkevych, G. Hierarchic decomposition of pre-machines as models of software system components / G Zholtkevych, I Perepelytsya // Системи управління, навігації і зв'язку. – 2011. – 4(20). – P. 233–238.

321. Zhou, F. Generalized canonical time warping / F. Zhou and F. De la Torre // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2016. – Vol. 38, Iss. 2. – P. 279-294.

ПРИЛОЖЕНИЕ

Акты о внедрении результатов диссертации

«Затверджую»
 проректор з наукової роботи
 Харківського національного
 університету радіоелектроніки
 докт. фіз.-мат. наук, професор
 М.І. Сліпченко
 «30» 01 2015 г.

«Затверджую»
 Т.В.О. начальника УКР
 ГУМВС України в Харківській області
 полковник міліції О.І. Лобунец
 «28» 01 2015 г.

АКТ

про впровадження результатів дисертаційної роботи
 «*Моделі та методи темпоральної обробки відео для інформаційного пошуку*»
 доцента кафедри інформатики
 Харківського національного університету радіоелектроніки
 канд. техн. наук, доцента Машталіра Сергія Володимировича

В період від 27.01.2015 г. до 28.01.2015 г. комісія у складі:

Голови – нач. від. УКР Птушка Є.О.
 та членів комісії – зас. нач. від. УКР Кириченко О.М.
 – о/у УКР Була Д.В.
 – о/у УКР Побежимов В.М.

розглянула результати використання матеріалів докторської дисертаційної роботи «*Моделі та методи темпоральної обробки відео для інформаційного пошуку*» доцента кафедри інформатики Харківського національного університету радіоелектроніки Машталіра Сергія Володимировича в рамках розв'язання задачі аналізу відео потоків для покращення діяльності карного розшуку.

Запропоновані підходи до сегментації відео потоків дозволяють спростити виявлення фрагментів відео, які містять найбільш корисну для розшуку інформацію, що дозволяє в подальшому звертати увагу тільки на ці сегменти з всього потоку відео даних, знятого здебільшого з камер спостереження з невеликою роздільною здатністю для їх подальшого покращення при розв'язанні задачі ідентифікації злочинців.

Комісія підтверджує доцільність впровадження даного підходу для підвищення ефективності використання відеоматеріалів в розшуковій діяльності за рахунок видалення інформації, яка є несуттєвою в даному випадку та можливості сфокусуватися на аналізі більш значущих фрагментів із великих в загальному випадку об'ємів відеоданих.

Даний акт не дає підстав для будь-яких фінансових розрахунків.

Голова комісії


 Птушка

Члени комісії


 Кириченко


 Була


 Побежимов

<p>«Затверджую» проректор з наукової роботи Харківського національного університету радіоелектроніки докт. фіз.-мат. наук, професор</p>  <p>М.І. Сліпченко</p>  <p>« 14 » 10 2014 р.</p>	<p>«Затверджую» Заступник голови – керівник апарату Харківської обласної державної адміністрації</p>  <p>В.В. Хома</p>  <p>« 9 » 10 2014 р.</p>
--	--

АКТ

про впровадження результатів дисертаційної роботи
«Моделі та методи темпоральної обробки відео для інформаційного пошуку»
 доцента кафедри інформатики
 Харківського національного університету радіоелектроніки
 канд. техн. наук, доцента Машталіра Сергія Володимировича

В період від 01.10.2014 г. до 03.10.2014 г. комісія у складі:

- Голови – Хоми Василя Васильовича
 та членів комісії – Светличного Володимира Вікторовича
 – Ситника Артема Юрійовича

розглянула результати щодо використання матеріалів докторської дисертаційної роботи *«Моделі та методи темпоральної обробки відео для інформаційного пошуку»* доцента кафедри інформатики Харківського національного університету радіоелектроніки Машталіра Сергія Володимировича при здійсненні пошуку та аналізу відеоматеріалів, які були накопичені в ході проведення селекторних нарад обласної системи відеоконференцв'язку. Дисертаційне дослідження спрямоване на підвищення швидкодії інформаційного пошуку відеоданих із запитом «за зразком», результати якого забезпечують необхідні показники точності і повноти в умовах нестационарності часових рядів, індукованих відео, високого рівня апіорної та поточної невизначеності, нелінійності процесів, що аналізуються, зазвичай, відповідно зростаючій вибірці відеорядів, тобто безпосередньо при реєстрації динамічної візуальної інформації або при її безпосередньому «перегляді».

Розглянувши матеріали, пов'язані з виконанням та впровадженням указаних робіт, комісія відзначає, що програмні засоби зберігання, обробки та багатоаспектного пошуку сюжетів, які базуються на методах та моделях, запропонованих у дисертації, дозволяють створювати достатньо ефективні багатокористувацькі інтерактивні системи, направлені в цілому на аналіз та інтерпретацію великих об'ємів неструктурованих відеоданих.

Розроблені та всебічно досліджені в дисертації моделі та методи темпоральної обробки відео, зокрема, сегментації відеорядів на базі різноманітних критеріїв та підходів (слідкуючих функцій, прогнозування рядів) в ознакових просторах та просторах зображення, створюють достатні передумови для побудови

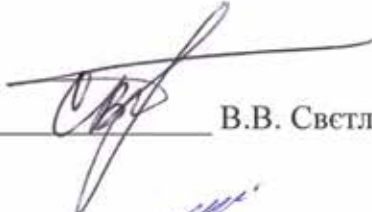
стратифікованої семантичної індексації відеоданих. Запропоновані моделі розширюють представлення про можливості автоматичного розбиття та покриття візуальної інформації як в просторі, так і її трансформацій у часі.

Методи, які запропоновані в дисертації, є достатньо простими з точки зору програмної реалізації обчислювальних моделей та, в той же час, забезпечують достатньо високі показники точності та надійності.

Комісія вважає необхідним відзначити, що в цілому розроблені моделі та методи можливо використовувати як базу для створення уніфікованих програмних засобів щодо розв'язання різноманітних задач аналізу, перетворення та тематичної інтерпретації відеоданих. Цінність отриманих результатів є очевидною для реалізації концепції «електронний уряд» в межах органів місцевого державного управління.

Даний акт не дає підстав для будь-яких фінансових розрахунків.

Заступник керівника апарату
Харківської обласної державної
адміністрації


В.В. Светличний

Начальник інформаційно-комп'ютерного
відділу апарату Харківської обласної
державної адміністрації


А.Ю. Ситник

<p>«Затверджую» проректор по науковій роботі Харківського національного університету радіоелектроніки докт. фіз.-мат. наук, професор</p> <p> М. Сліпченко</p> <p>«26» _____ 2014 г.</p>	<p>«Затверджую»</p> <p>Директор ТОВ Заводу рентгенівського обладнання «Квант»</p> <p> О.М. Коляда</p> <p>«18» _____ 12 2014 г.</p>
---	--

АКТ

про впровадження результатів дисертаційної роботи
«Моделі та методи темпоральної обробки відео для інформаційного пошуку»
 доцента кафедри інформатики
 Харківського національного університету радіоелектроніки
 канд. техн. наук, доцента Машталіра Сергія Володимировича

В період від 01.12.2014 г. до 17.12.2014 г. комісія у складі:

- Голови – Афанасьєв С.В. – керівник виробництва
 та членів комісії – Седак О.В. – заступник керівника відділу технічного контролю
 – Непіпенко В.І. – інженер-розробник

провела аналіз використання запропонованих в дисертації методів та склала даний акт про те, що на заводі рентгенівського обладнання «Квант» був впроваджений нейромережевий метод сегментації відео потоку.

Результати впровадження довели, що розроблені Машталіром С.В. методи дозволяють провести аналіз рентгеноскопічної інформації та виявити однорідні за просторово-часовим складом фрагменти, що в свою чергу дозволяє використовувати методи покращення відеоданих на найбільш вагомих фрагментах, а не на всіх вихідних даних з метою зменшення часу необхідного на відповідну обробку.



Комісія підтверджує доцільність впровадження даного підходу, який дозволяє більш вдало покращувати динамічну інформацію отримувану в рамках процедури рентгеноскопії за рахунок вибору найбільш вагомих з точки зору аналізу сегментів відео потоку. Що в свою чергу дозволяє спростити процес діагностики захворювань фахівцем.

Даний акт складений для пред'явлення до спеціалізованої вченої ради і не дає підстав для будь-яких фінансових розрахунків.

Голова комісії

 /Афанасьєв С.В./

Члени комісії

 /Седак О.В./
 /Непіпенко В.І./

<p>«Затверджую» проректор по науковій роботі Харківського національного університету радіоелектроніки докт. фіз.-мат. наук, професор</p> <p> М.І. Сліпченко</p> <p>« 26 » 12 2014 г.</p> 	<p>«Затверджую» Директор ТОВ Харківського технічного центру рентгенівського сервісу «Спектр»</p> <p> С.О. Федотова</p> <p>« 17 » 12 2014 г.</p> 
--	---

АКТ

про впровадження результатів дисертаційної роботи
 «Моделі та методи темпоральної обробки відео для інформаційного пошуку»
 доцента кафедри інформатики

Харківського національного університету радіоелектроніки
 канд. техн. наук, доцента Машталіра Сергія Володимировича

В період від 24.11.2014 г. до 12.12.2014 г. комісія у складі:

- Голови – Мельник С.О. – начальник сервісної служби
 та членів комісії – Оніщенко Ю.І. – інженер-налогоджувальник рентгенівської
 техніки
 – Сватов Н.В. – рентгеномеханік


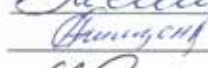

склала даний акт про те, що у ХТЦ РС «Спектр» у відділі сервісної служби був застосований метод аналізу потокової мультимедійної інформації з метою пошуку еталонів в вихідних даних при розв'язанні задачі ідентифікації.

Результати впровадження довели, що розроблені Машталіром С.В. методи дозволяють ефективно виділяти фрагменти в потоці мультимедійних рентгенівських даних з інформацією, час пошуку в рамках якої потрібної інформації значно зменшується в порівнянні з аналізом всього вихідного потоку даних. Зокрема при наявності множини зображень еталонів з якими потрібно порівняти вихідну інформацію запропонований підхід дозволяє щонайменше вдвічі скоротити необхідний час обробки за рахунок видалення фрагментів інформації яка гарантовано не має перетину з еталонною множиною.

Комісія підтверджує доцільність використання запропонованого підходу для підвищення якості та швидкодії розв'язання задачі ідентифікації/класифікації в потоці рентгенівських даних.

Даний акт складений для пред'явлення до спеціалізованої вченої ради і не дає підстав для будь-яких фінансових розрахунків.

Голова комісії
 Члени комісії

 Мельник С.О./
 Оніщенко Ю.І./
 Сватов Н.В./

«Затверджую»
Перший проректор
Харківського національного
університету радіоелектроніки



І.І. Ключник

2016 р.

про використання в навчальному процесі результатів дисертаційної роботи
«Моделі та методи темпоральної обробки відео для інформаційного пошуку»
доцента кафедри інформатики
Харківського національного університету радіоелектроніки
Машталіра Сергія Володимировича

Комісія у складі голови – декана факультету ПММ, проф. Дорошенка В.О., та членів: зав. каф. інформатики, д.т.н., проф. Путятіна Є.П.; проф. каф. інформатики, д.т.н. Гороховатського В.О. провела роботу по встановленню використання результатів дисертаційної роботи «Моделі та методи темпоральної обробки відео для інформаційного пошуку» доцента кафедри інформатики ХНУРЕ Машталіра С.В. в навчальному процесі при підготовці бакалаврів, спеціалістів та магістрів за напрямом «Інформатика».

Розглянувши матеріали роботи та зміст навчального процесу на кафедрі інформатики, комісія відзначає, що при проведенні лекційних занять та лабораторного практикуму за курсами «Методи та системи розпізнавання образів», «Алгоритми та програмне забезпечення синтезу зображень», «Стиснення даних» використовуються наступні результати дисертаційних досліджень:

- методи сегментації та пошуку змін властивостей в відеопослідовностях за допомогою обробки багатовимірних часових рядів;
- методи виділення опорних даних сегментованих відеопослідовностей з використанням аналізу окремих кадрів за допомогою побудови векторів характеристик;
- моделі та методи компресії зображень для темпоральної обробки відео на базі використання нейронних мереж.

Комісія відзначає, що використання нових результатів дисертаційних досліджень дає можливість покращити якість вивчення інформаційних технологій у контексті розширення форм сприйняття обробки та інтерпретації відеоданих.

Голова комісії
декан факультету ПММ, проф.

 В.О. Дорошенко

Члени комісії
зав. каф. інформатики, д.т.н., проф.

 Є.П. Путятін

проф. каф. інформатики, д.т.н.

 В.О. Гороховатський



ОРГАН З СЕРТИФІКАЦІЇ ХНУРЕ
зареєстрований Державним комітетом стандартизації, метрології та сертифікації України
реєстраційний № ОДС–054 від 03.06.2002 р.

СЕРТИФІКАТ ВІДПОВІДНОСТІ

Серія ПЗ

Зареєстровано в Реєстрі ОДС ХНУРЕ за № UA.ОДС–054. 068-14
Зареєстровано в Реєстрі ОДС ХНУРЕ

Термін дії з «02» жовтня 2014 р. до «01» жовтня 2017 р.
Срок действия

код УКТ ЗЕД, ТН ЗЕД

72.30.2

код ДКП, ОКП

Продукція
Продукция

Програмний засіб

**Програмний комплекс просторо-часової сегментації відеоданих
в задачах пошуку із запитом “за зразком”**

Відповідає вимогам
Соответствует требованиям

ТЗ 804-02071197-00048-01 90, ДСТУ 2850-94 Програмні засоби
ЕОМ. Показники і методи оцінювання якості

Виробник продукції
Изготовитель продукции

Харківський національний університет радіоелектроніки
м. Харків, пр. Леніна, 14, код ЄДРПОУ 02071197

Сертифікат видано
Сертификат выдан

Харківський національний університет радіоелектроніки
м. Харків, пр. Леніна, 14, код ЄДРПОУ 02071197

Додаткова інформація
Дополнительная информация

Одинична продукція
Автор Машталір Сергій Володимирович

Сертифікат видано органом з сертифікації
Сертификат выдан органом по сертификации

ОДС ХНУРЕ, 61166, м. Харків, пр. Леніна, 14
тел.: (057) 702–1331, 702–0557

На підставі
На основании

Рішення ОДС ХНУРЕ № 96-14 від 01.10.14

Керівник органу з сертифікації



М.П.

І.В. Руженцев



До уваги постачальників, покупців та контролюючих органів. Сертифікат має юридичну силу на всій території України.
Копія не дійсна без оригіналу підпису і печатки органу, що його видав.
Чинність сертифіката відповідності можна перевірити в Реєстрі ОДС ХНУРЕ за тел. (057) 702–0557

00118