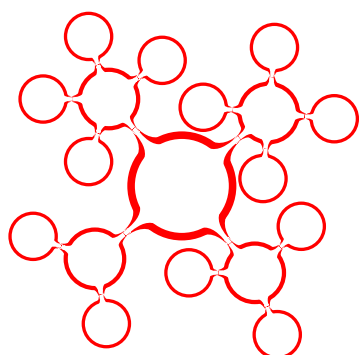


И. К. НИКИТИН w@w-495.ru

# **Обзор методов комплексного ассоциативного поиска видео**



## **An Overview Of Complex Content-Based Video Retrieval**

РАСПРОСТРАНЯЕТСЯ: на правах рукописи.

ДАТА: 31 августа 2014

МОСКВА 2014 Г.

УДК 004.932.4

И. К. Никитин  
Московский Авиационный Институт  
(Национальный Исследовательский Университет),  
Волоколамское шоссе, д. 4,  
г. Москва, 125993, Россия  
Email: w@w-495.ru

## Обзор методов комплексного ассоциативного поиска видео

В статье предлагается обзор различных существующих методов ассоциативного поиска по видео. В течение прошлого десятилетия наблюдался стремительный рост количества видео размещаемых в Интернете, что создало острую необходимость в появлении поиска по видео. Видео имеет сложную структуру. Одна и та же информация может быть выражена различными способами. Это серьёзно усложняет задачу видео-поиска. Заголовки и описания видео не могут дать полного представления о самом видео, что влечет за собой необходимость использования ассоциативного поиска по видео. Существует семантический разрыв между низкоуровневыми характеристиками видео и восприятием пользователей. Комплексный ассоциативный видео-поиск может рассматриваться как связующее звено между обычным поиском и смысловым поиском по видео.

**Ключевые слова:** анализ видео; аннотирование видео; видео-поиск; кадры; классификация видео; нечеткие дубликаты видео; ранжирование видео; сцены; съёмки.

Ilya Nikitin  
postgraduate student  
of Moscow Aviation Institute  
(National Research University)  
Moscow, Russia  
E-mail: w@w-495.ru

## An Overview Of Complex Content-Based Video Retrieval

The paper focuses on an overview of the different existing methods in content-based video retrieval. During the last decade there was a rapid growth of video posted on the Internet. This imposes urgent demands on video retrieval. Video has a complex structure and can express the same idea in different ways. This makes the task of searching for video more complicated. Video titles and text descriptions cannot give the hole information about objects and events in the video. This creates a need for content-based video retrieval. There is a semantic gap between low-level video features, that can be extracted, and the users' perception. Complex content-based video retrieval can be regarded as the bridge between traditional retrieval and semantic-based video retrieval.

**Keywords:** frames; near-duplicates video; scenes; shots; video annotation; video classification; video mining; video reranking; video retrieval.

# СОДЕРЖАНИЕ

<b>1. Введение</b>	<b>5</b>
1.1. Конкретные применения	5
1.2. Исследования	6
1.3. Комплексность	7
1.4. Обзоры	8
1.5. Общая схема	8
<b>2. Деление видео</b>	<b>11</b>
2.1. Обнаружение границ съёмок	12
2.1.1 Извлечение признаков	13
2.1.2 Измерение сходства	13
2.1.3 Поиск границы	14
2.2. Извлечение ключевых кадров	17
2.2.1 Последовательное сравнение	18
2.2.2 Глобальное сравнение	19
2.2.3 Ссылочные кадры	21
2.2.4 Упрощение кривых	21
2.2.5 Кластеризация	22
2.2.6 Объекты и события	22
2.3. Сегментация сцен	23
2.3.1 Деление по ключевым кадрам	24
2.3.2 Деление по аудиовизуальной информации	24
2.3.3 Деление по фону	25
2.4. Сегментация звука	26

<b>3. Выделение признаков</b>	<b>27</b>
3.1. Характеристики ключевых кадров	27
3.1.1 Цвета	28
3.1.2 Текстуры	28
3.1.3 Контуры	29
3.1.4 Края	29
3.2. Характеристики объектов	29
3.3. Характеристики движения	30
3.4. Звуковые характеристики	30
<b>4. Анализ видео</b>	<b>30</b>
<b>5. Классификация видео</b>	<b>31</b>
5.1. Жанры	32
5.2. События	34
5.3. Объекты	34
<b>6. Аннотирование видео</b>	<b>35</b>
<b>7. Обработка запроса</b>	<b>36</b>
7.1. Типы запросов	36
7.1.1 Семантические запросы	36
7.1.2 Не семантические запросы	37
7.2. Оценка сходства	37
7.3. Сравнение характеристик	38
7.4. Сравнение текста	38
7.5. Сравнение онтологий	38
7.6. Оценка релевантности	38
<b>8. Заключение</b>	<b>39</b>
<b>Список использованных источников</b>	<b>41</b>
<b>Предметный указатель</b>	<b>58</b>

# 1. ВВЕДЕНИЕ

Видео обладает тремя отличиями от других типов медийных ресурсов [63].

- 1) Видео более насыщено информацией, чем, например изображения.
- 2) Видео содержит много необработанных данных.
- 3) Видео имеет слабо выраженную структуру.

Эти особенности делают поиск и индексацию видео весьма сложными. Ранее, видео базы были крошечными. Поиск и индексация в таких базах были основаны на ручной аннотации по ключевым словам. Недавно, размеры таких баз сильно возросли и возникла необходимость автоматического анализа видео с минимальным привлечением человеческого труда.

В длинном видео сложно автоматизировано найти интересующий отрывок. А, размечать и искать видео вручную весьма трудоемко. Смысловой разрыв между низкоуровневой информацией и потребностями пользователя заставляет работать с видео на более высоком уровне. Тем не менее, большинство методов поиска следуют парадигме прямого отображения низкоуровневых характеристик видео на смысловые понятия. Этот подход требует предварительной обработки данных. А результаты такого отображения не будут устойчивы. Без учета конкретной предметной области задача кажется неразрешимой.

Анализ большого объема видео-данных для выделения нужной информации является сложной задачей. Для ее решения применяют ассоциативный поиск. В англоязычной литературе ассоциативный видео-поиск называют «content based video retrieval» (CBVR) — поиск по содержанию.

## 1.1. КОНКРЕТНЫЕ ПРИМЕНЕНИЯ

Существует много разных применений видео-поиска. Например:

- исследование электронной коммерции и анализ тенденций выбора и упорядочивания материалов;

- быстрый поиск видео в массиве данных;
- анализ новостных событий [68];
- интеллектуальное администрирование и контроль интернет-видео — поиск нужных видео и отсеивание нежелательных.

Последнее время стало появляться много клипов с очень схожим содержанием (нечеткие дубликаты видео). Задача эффективной идентификации нечетких дубликатов играет ключевую роль в задачах поиска, защите авторских прав, и многих других.

Ассоциативный поиск используется для автоматического реферирования видео, анализа новостных событий, видеонаблюдения, и в образовательных целях [4].

Эти и многие другие прикладные задачи мотивируют ученых заниматься проблемами индексации и поиском по содержанию.

## 1.2. ИССЛЕДОВАНИЯ

Существует два вида исследований: научные и промышленные.

Примером крупных научных исследований может служить регулярная конференция TREC Vid (Text REtrieval Conference: Video), которую с 2003 года проводит Американский Национальный институт стандартов и технологий (NIST). Многие участники присылают свои алгоритмы для выпуска коллективной монографии, например [101], [5] и [93].

В промышленных исследованиях важную роль играют экспертные группы стандарта TV-Anytime и MPEG (Moving Picture Experts Group) [69]. Существуют работы, в которых видео-поиск рассматривается в рамках задачи сжатия видео [8].

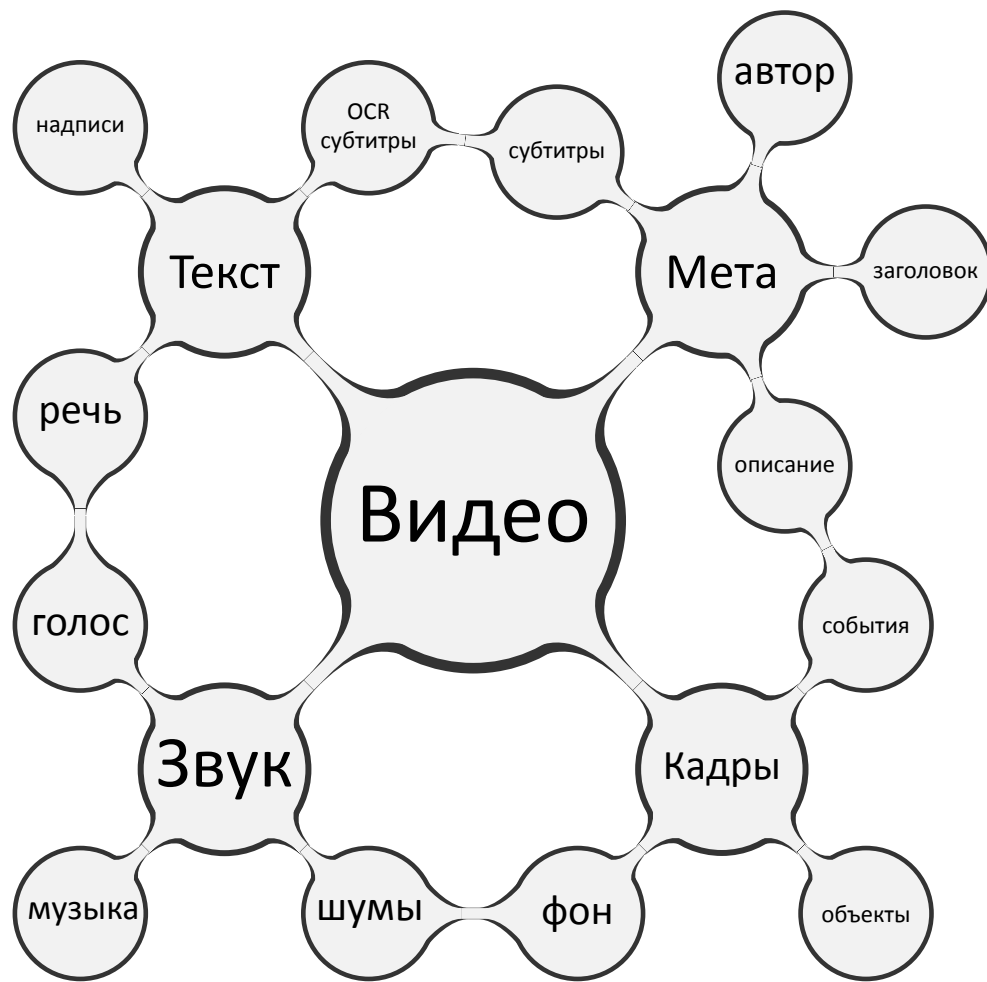


Рис. 1. Структура видео

### 1.3. КОМПЛЕКСНОСТЬ

Видео содержит в себе несколько видов данных. Авторы [70] и [92] выделяют четыре вида.

- 1) Метаданные — заголовок, автор и описание.
- 2) Звуковая дорожка.
- 3) Тексты полученные при помощи технологии оптического распознавания символов (OCR) или при помощи технологий распознавания речи.
- 4) Визуальная информация кадров видео.

Важно, что источники данных разной природы не являются независимыми. Они тесно связаны и дополняют друг друга, позволяя воссоздать полную картину происходящего.



Таким образом, видео обладает комплексностью. Комплексность (системность, мультимодальность) — способность взаимодействовать с пользователем по различным каналам информации и извлекать и передавать смысл автоматически [65]. Комплексность видео состоит в возможности автора выражать мысли, используя по крайней мере два информационных канала. Каналы могут быть визуальными, звуковыми или текстовыми.

## 1.4. ОБЗОРЫ

Поиску и индексации видео было посвящено много обзоров. Обычно обзоры фокусируются на какой-то подзадаче поиска. Например, в работе Смитона [93] дается хороший обзор методов обнаружения границ съёмок за семь лет работы TRECVID. В обзоре Сноека [94] приведен полный анализ смыслового поиска по видео. Основной акцент сделан на оценке алгоритмов, использующих базы данных TRECVID, поиске видео по семантическим понятиям. Шеффман с соавторами [114] сделали обзор по интерфейсам систем видео просмотра и их применению. В работе [96] приведен анализ методов выделения пространственно-временной смысловой информации из видео. В работе [39] дают хороший обзор аннотации видео. В работе [9] описываются свежие исследования методов ранжирования видео.

## 1.5. ОБЩАЯ СХЕМА

Ассоциативный поиск видео состоит из следующих шагов.

- 1) Анализ временной структуры видео — деление видео на фрагменты, которое включает обнаружение границ съёмок.
- 2) Определение характеристик фрагментов: параметров ключевых кадров, характеристик движения и объектов.
- 3) Извлечение информации из характеристик.
- 4) Аннотация видео, построение семантического индекса.

- 5) Обработка пользовательского запроса и выдача результата.
- 6) Обратная связь и переранжирование результатов для улучшения поиска характеристик.

Таблица 1. Некоторые научные работы по поиску видео от 2008 года

Год	Работа	Тема
2008	[85]	Сегментация видео
2009	[17]	
2010	[93] [83]	
2011	[51] [111] [99]	
2013	[27]	
2010	[61]	Реферирование видео
2012	[31]	
2012	[20], [47], [117]	Индексация видео
2013	[112]	
2014	[63]	
2012	[23]	Комплексный ассоциативный поиск
2013	[102]	
2011	[52], [21]	Представление видео
2009	[94]	Смысловой ассоциативный поиск
2012	[32], [109], [55]	
2012	[39], [109]	Аннотирование видео
2012	[22]	Видео-поиск по движению
2011	[9]	Ранжирование видео
2012	[39]	
2010	[67]	Классификация видео
2011	[82]	
2012	[2], [110]	

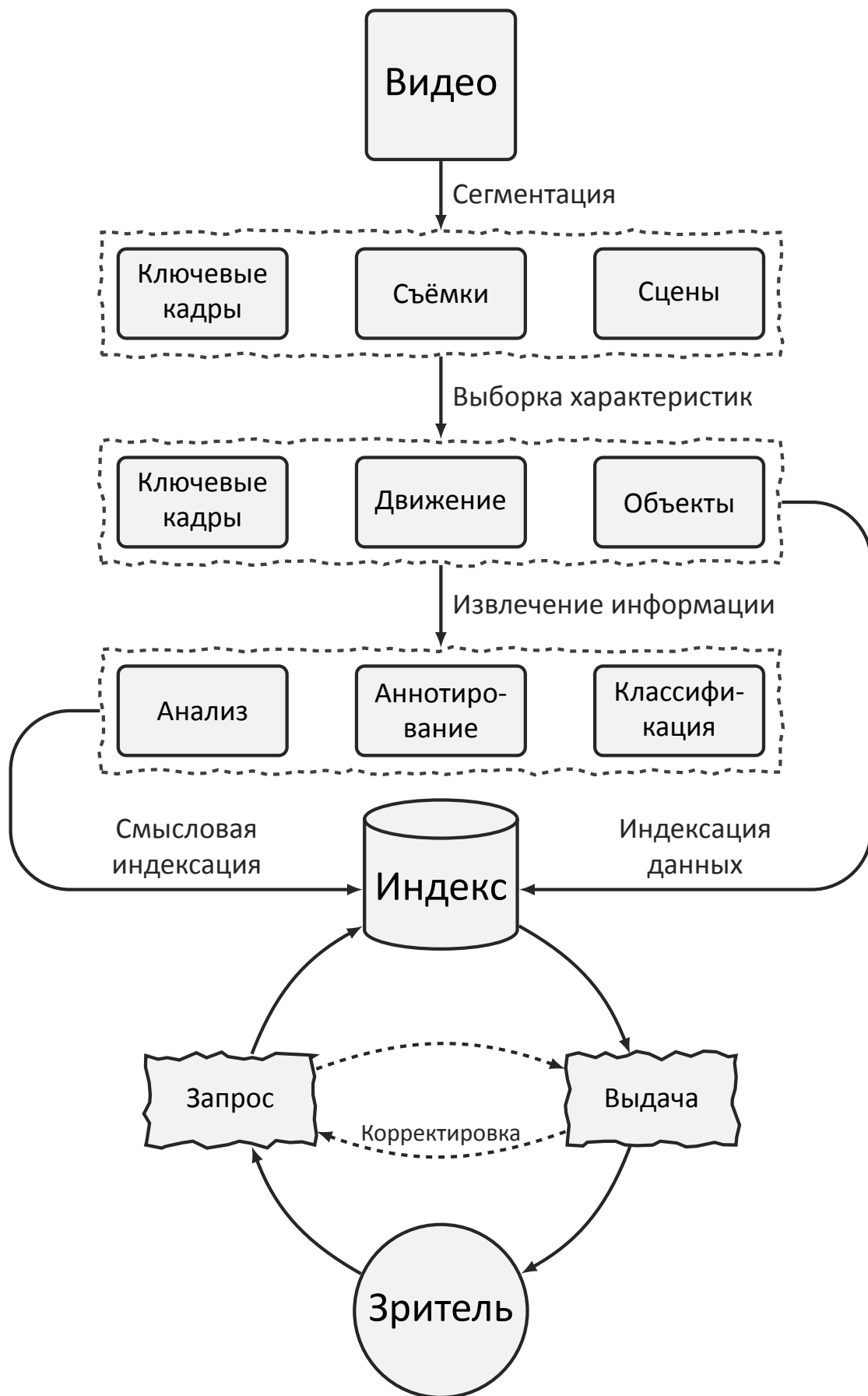


Рис. 2. Схема поиска по видео

## 2. ДЕЛЕНИЕ ВИДЕО

Деление видео включает в себя обнаружение границ съёмок, извлечение ключевых кадров, сегментацию сцен и аудио.

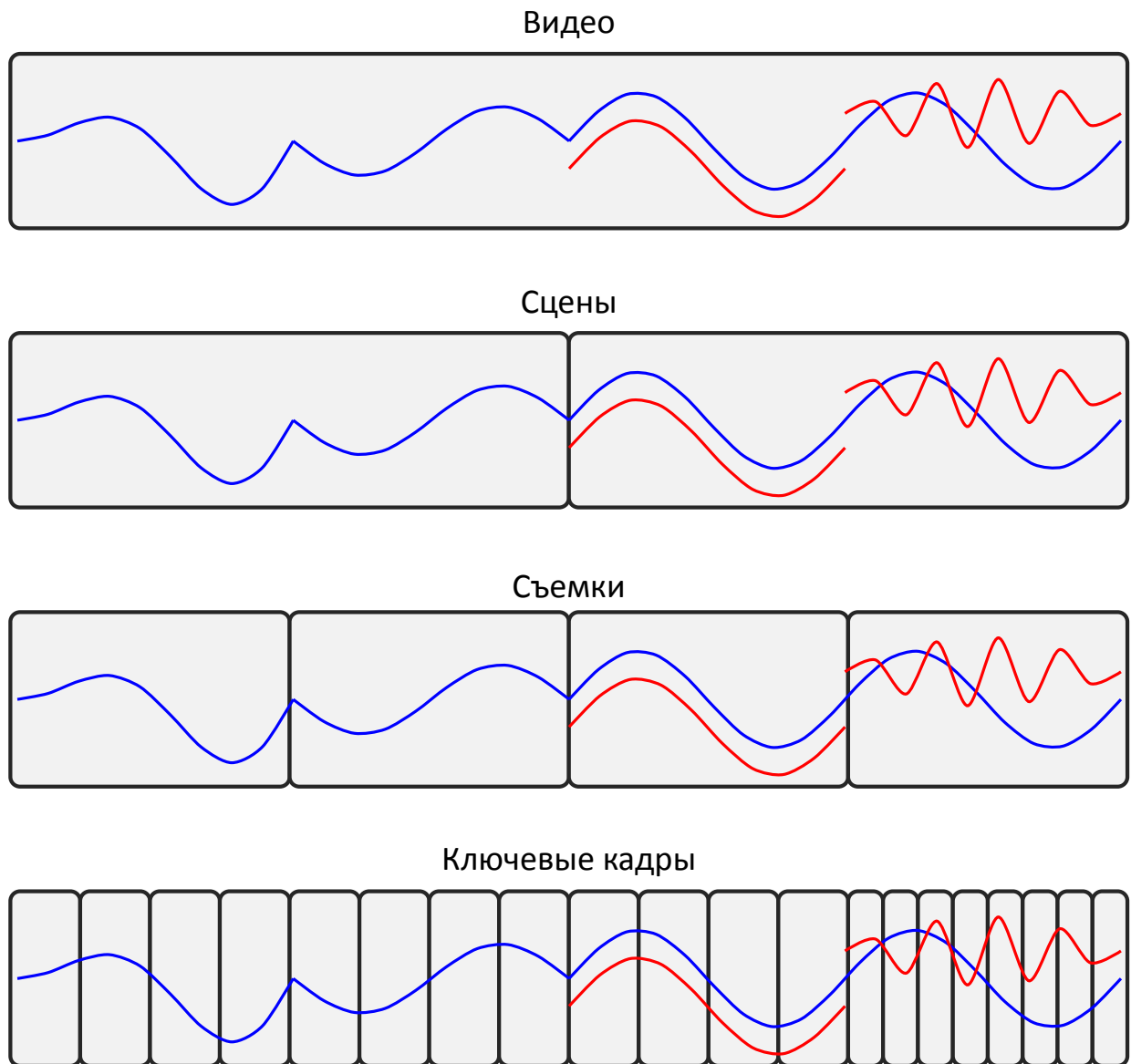


Рис. 3. Схема визуальной сегментации видео на примере движения точек. Временная ось отложена по горизонтали.

Это может показаться не естественным, но сегментацию сцен и выделение ключевых кадров проще проводить, если уже известны границы съёмок.

## 2.1. ОБНАРУЖЕНИЕ ГРАНИЦ СЪЁМОК

Видео делят на фрагменты по времени. В качестве таких фрагментов могут выступать съёмки. Съёмка (кинематографический кадр, монтажный план) — отрезок киноплёнки, на котором запечатлено непрерывное действие между пуском и остановкой камеры, или между двумя монтажными склейками.

С точки зрения семантики, самым мелким элементом видео является кадр (фотографический кадр, кадрик). Съёмка является более крупным делением. Из съёмок складываются сцены, а из сцен видео целиком.

Кадры съёмок связаны между собой по смыслу. В задачах анализа и поиска съёмки считают основной структурной единицей видео.

Границы съёмок бывают трех типов:

- линейная склейка — съёмка внезапно прерывается и начинается другая, обычно такой переход занимает 1—2 кадра;
- постепенное проявление, исчезновение, проявление или затухание (в монохромном кадре);
- вытеснение — исчезновения одной съёмки, и появления другой (растворение, вытеснение шторкой).

В литературе приводят и иную классификацию границ съёмок: линейную склейку противопоставляют плавному переходу. При этом, затухание, проявление и вытеснение рассматривают как частные случаи плавного перехода [43].

Плавную смену съёмок обнаружить сложнее. Основные подходы обнаружения границ съёмок вместе с их описаниями, преимуществами и недостатками всесторонне представлены в работах [93], [65] и [34].

Для обнаружения границ съёмок, как правило, сначала извлекают визуальные характеристики каждого кадра. Затем, на основе выделенных признаков, оценивают сходство между кадрами. Границы съёмок определяют по смене неоднородных кадров.

Выделение съёмок состоит из следующих этапов:

- извлечение признаков [16];
- измерение сходства;
- и поиска границы.

### 2.1.1. ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ

Для определения смен съёмок используют

- векторы движения;
- цветовые гистограммы [44];
- изменения отношения края [119] [122] и [73];
- преобразование характеристик инвариантное к масштабу (SIFT) [107];
- карты салиентности [85];
- угловые точки [37] и так далее.

Для цветовых гистограмм малые движения камеры оказываются очень ощутимыми. При изменениях освещения и движении характеристики краев более устойчивы. А влияние смещений объекта и камеры могут быть эффективно обработаны характеристиками движения. Но как правило, простые цветовые гистограммы не уступают характеристикам краев и движения и более сложными методам [34].

### 2.1.2. ИЗМЕРЕНИЕ СХОДСТВА

Согласно работам [87] [86] [25] для оценки подобия между векторами характеристик кадров используют:

- расстояние Евклида;
- косинусное расстояние;

- сходство по хи-квадрат;
- и пересечение гистограмм.

В работе [44] предлагают использовать транспортное расстояние (по метрике Канторовича — Рубинштейна — Вассерштейна — Орнштейна). В англоязычной литературе его называют «землеройным расстоянием» — Earth Mover's Distance (EMD).

### 2.1.3. ПОИСК ГРАНИЦЫ

Для поиска границ съёмок применяют как попарное сравнение кадров, так и сравнение участков кадров заданного размера [25]. При сравнении участок постоянно смещается. Такие участки называют окнами, а сам подход методом скользящего окна.

Оконные методы используют контекстную информацию кадров, что позволяет сократить влияние локальных шумов и помех. Однако они требуют дополнительных вычислительных затрат.

На данный момент существует два типа методов обнаружения границ съёмок: пороговые и статистические.

#### ПОРОГОВЫЕ МЕТОДЫ

Пороговые методы попарно сравнивают подобию кадров с заданным порогом [113] [13]. Если сходство между кадрами или «окнами» оказывается ниже порога, то граница съёмки считается найденной. Порог может быть глобальным, адаптивным или комбинированным.

Глобальный порог подбирается экспериментально для всего видео целиком [13]. Локальные особенности при этом не учитываются, что отрицательно влияет на точность обнаружения.

Алгоритмы адаптивного порога, основанные на скользящем окне, вычисляют порог локально [44] [85]. Адаптивный порог дает более высокую точность нахождения границ [42], но требует больше вычислительных ресурсов

чем глобальный. Кроме того, для применения адаптивного порога необходимы априорные знания о самом видео, например, для выбора размера окна.

Комбинация адаптивных и глобальных порогов позволяет подстраивать локальные пороги при наличии значения общих параметров. Например, в работе [75] введена пороговая функция. Значение функции изменяется локально в пределах двух глобальных порогов. Глобальные пороги настраиваются в зависимости от требований точности и полноты. Однако, отношения локальных адаптивных порогов и двух глобальных порогов не могут быть легко определены.

## СТАТИСТИЧЕСКИЕ МЕТОДЫ

Статистические методы находят границы сцен на основе характеристик кадров. Определение границ съёмок может рассматриваться как задача классификации на основе статистического обучения. Кадры классифицируются как принадлежащие съёмке, и не принадлежащие. Используются классификаторы с учителем — контролируемые и без учителя.

### КЛАССИФИКАЦИЯ С УЧИТЕЛЕМ

Преимущество методов классификации с учителем в том — что не нужно искусственно устанавливать пороги определения границ. Вместе с тем, для повышения точности могут быть использованы самые разные виды характеристик. Минус такой классификации — необходимость аккуратного выбора примеров для обучения, как положительных, так и отрицательных.

К контролируемым классификаторам относят AdaBoost (Adaptive Boosting — адаптивное улучшение) и метод опорных векторов (MOB, SVM —support vector machine).

Применяют и некоторые другие алгоритмы. В работе [25] применяют классификатор kNN на основе поиска ближайших соседей, в котором сходства кадров на определенном временном интервале используются как входные данные. В работе [10] различные виды перехода съёмок находят с помощью скрытых моделей Маркова (НММ).



## МЕТОД ОПОРНЫХ ВЕКТОРОВ

Метод опорных векторов [64] [86] [80] используется для того, чтобы отличить склейку от ее отсутствия. Для обработки структур высокой размерности используется, чтобы уменьшить влияние изменений при освещенности и быстром перемещении объекта используется функция ядра. В работе Чжао [7] для скользящего окна применяются два МОВ-классификатора, чтобы отличить плавную смену съёмов от линейных склеек. В работе [38] МОВ использовали для классификации кадров в трех категориях: постепенный переходы, склейки и прочие кадры.

В работах [34] и [49] МОВ комбинируют с пороговыми методами. Предварительно границы-кандидаты ищут с помощью порога. После этого применяют МОВ-классификатор для более строго отбора.

## АДАПТИВНОЕ УЛУЧШЕНИЕ

Суть адаптивного улучшения AdaBoost состоит в усилении простых классификаторов, путём их комбинирования [35]. В контексте поиска границ съёмов AdaBoost хорош тем, что позволяет использовать большое число различных характеристик. Классификация с помощью адаптивного улучшения используется в работах [122] и [100].

Чжао и Цай [122] применяют алгоритм к обнаружению границ съёмов для сжатого видео. Сначала приближенно классифицируются характеристики цвета и движения. Затем каждый кадр классифицируется как линейная склейка, плавный переход между съёмками или отсутствие перехода.

## КЛАССИФИКАЦИЯ БЕЗ УЧИТЕЛЯ

Алгоритмы неконтролируемого обучения делятся на два типа:

- использующие сходство кадров;
- использующие сами кадры.

Методы на основе сходства кадров сравнивают меры подобия пар кадров, и разбивают их на две группы — подобия кадров одной съёмки, и подобия кадров, относящихся к разным съёмкам. Для классификации применяются методы К-средних и нечетких К-средних.

Методы на основе кадров классифицируют сами кадры по визуальному сходству. Группы похожих кадров относят к общей съёмке [107]. В работе [87] применяется метод К-средних. В работе [26] применяют спектральную кластеризацию.

Преимущество таких методов в том, что какая-либо обучающая выборка не требуется. Однако методы не учитывают временную информацию и поэтому, плавные смены съёмки не могут быть обнаружены.

## СЖАТОЕ ВИДЕО

Классификация без учителя применяется для определения границ съёмок в сжатом видео. Для выявления съёмок в таком видео могут быть использованы, например коэффициенты дискретного косинус-преобразования (ДКТ) и типы макроблоков [122] [123] [88]. Макроблоками называют участки кадра размером 16 на 16 пикселей. Макроблоки содержат информацию о типе временного предсказания и векторы компенсации движения [54]

Работа со сжатым видео позволяет сэкономить время, затраченное на распаковку, но точность таких методов ниже.

## 2.2. ИЗВЛЕЧЕНИЕ КЛЮЧЕВЫХ КАДРОВ

Среди кадров одной съёмки есть избыточность. Для ее уменьшения выделяют кадры, которые наиболее полно отражают содержание съёмки. Ключевые кадры являются самым простым способом описания длинной последовательности видео. При извлечении ключевых кадров используют различные характеристики:

- цветовые гистограммы;
- края;

- очертания;
- оптические потоки.

Самый простой способ выбирать ключевые кадры — это использовать каждый  $k$ -тый кадр в качестве ключевого. Однако такой способ не позволяет учитывать изменения визуальной динамики съёмки.

Способы извлечения подразделяются на шесть категорий:

- последовательное сравнение;
- глобальное сравнение;
- на основе ссылочных кадров;
- на основе кластеризации;
- и на основе объектов или событий [105].

Из-за субъективности определения ключевого кадра, нет общего метода оценки поиска.

Обычно, в качестве меры для оценки результата извлечения ключевых кадров, используются частота ошибок и степень сжатия видео. Предпочтены методы, дающие меньше число ошибок и большие степени сжатия.

### 2.2.1. ПОСЛЕДОВАТЕЛЬНОЕ СРАВНЕНИЕ

При последовательном сравнении ключевой кадр сравнивают с другими кадрами до тех пор пока не будет найден «сильно отличный». Для сопоставления кадров применяют цветовые гистограммы [29]. В работе [29] предложена накопительная энергетическая функция в качестве расстояния между парой кадров.

Последовательное сравнение — интуитивно понятный способ поиска ключевых кадров. К преимуществам подхода строит отнести простоту, точность и не высокую вычислительную сложность. У подхода есть существенные недостатки.

- 1) Ключевые кадры описывают только локальные свойства фрагмента. Глобальные свойства съёмки никак не отражены.
- 2) Количество ключевых кадров никак не контролируется. Подход нельзя применять, если требуется заданное число ключевых кадров.
- 3) При повторении визуального содержимого одни и те же ключевые кадры будут выделены многократно.

### **2.2.2. ГЛОБАЛЬНОЕ СРАВНЕНИЕ**

Методы глобального сравнения используют различия между кадрами в съёмке и распределяют ключевые кадры, минимизируя предопределённую целевую функцию [105].

Ключевые кадры найденные такими методами отражают глобальные характеристики съёмки. К преимуществам относится и возможность задавать заранее количество ключевых кадров. По сравнению с последовательным, глобальное сравнение даёт менее избыточную последовательность ключевых кадров, но требует больше вычислительных затрат.

### **РАВНАЯ ИЗМЕНЧИВОСТЬ КАДРОВ**

Ключевые кадры выбирают так, чтобы сегменты съёмки представленные ключевыми кадрами менялись с одинаковой скоростью — т. е. обладали равной изменчивости. В сегменте, изменение кадров может быть оценено по содержанию последовательных кадров [29] или как разница между первым и последним кадрами сегмента. Целевая функция в подходе выбирается как сумма разностей изменчивости сегментов.

Например, в работе [28] ключевые кадры получают путем деления съёмки на сегменты с равным количеством перемещений в кадрах. Количество перемещений измеряют с помощью дескриптора движения MPEG-7. Ключевым кадром считают медианный кадр каждого сегмента.

## МАКСИМАЛЬНОЕ ПОКРЫТИЕ

В методе максимального покрытия ищут кадры, которые могут выступать представителями для максимального количества других кадров [105]. Например, в работе [14] покрытие ключевого кадра, определено как количество кадров, которые визуальны похожи на данный.

Если число ключевых кадров не задано, то минимизируют число результирующих кадров, удовлетворяющих предопределенному критерию доверия. В противном случае, — если число ключевых фиксировано, то максимизируют число кадров, которые ключевые кадры могут представлять [79] [24] Для поиска соответствий кадров используются жадные алгоритмы.

## МИНИМАЛЬНАЯ КОРРЕЛЯЦИЯ

Выбирают наиболее непохожие кадры. В методе минимизируют сумму корреляций между ключевыми кадрами, Особенно важно, чтобы некоррелированными были последовательные ключевые кадры.

В работе [72] кадры съёмки и их отношения представляют с помощью взвешенного ориентированного графа. Вершины кратчайшего пути в графе соответствуют кадрам с наименьшей корреляцией. Такие кадры назначают ключевыми.

## МИНИМАЛЬНАЯ ОШИБКА ВОССТАНОВЛЕНИЯ

Ключевые кадры извлекают, используя минимизацию сумму различий между предсказанным «восстановленным» кадром и кадром соответствующим ему реальным. Восстановленный кадр получают аппроксимацией набора известных ключевых кадров. Такие методы применимы в первую очередь к анимации [99].

В работе [56] для выделения заданного числа ключевых кадров внутри съёмки применяют итерационную процедуру, уменьшающую ошибку восстановления.

В работе предлагают [89] аппроксимировать «восстановленные» кадры на основе инерции ключевых кадров, в которых изображено движение.

### 2.2.3. ССЫЛОЧНЫЕ КАДРЫ

Выделение ключевых кадров в таких методах состоит из двух шагов.

- 1) Ссылочные кадры создаются алгоритмически.
- 2) Сравнивают все кадры съёмки со ссылочными.

Важно заметить, что не обязательно пытаться воспроизвести искусственный ссылочный кадр целиком. Достаточно сконструировать определенную характеристику кадра.

Например, в работе [33] строится альфа-усеченная средняя цветовая гистограмма съёмки. Альфа-усеченное усреднение (alpha trimmed mean) представляет из себя смесь усредняющего и медианного фильтров. «Альфа» ( $\alpha$ ) в данном случае, не имеет никакого отношения к каналу прозрачности в модели (RGBA). В данном случае «альфа» — это коэффициент усечения (усреднения), который позволяет подавить нежелательные шумы. После построения альфа-усеченной средней цветовой гистограммы Вычисляют расстояния между этой гистограммой и цветовыми гистограммами кадров съёмки. Далее на основе этого расстояния кадры собирают в кластеры и для каждого кластера выбирают представителя, который и будет являться ключевым кадром. Авторы предлагают брать центроиды кластеров.

В работе [97] конструируют средний кадр съёмки, который отражает пространственное и временное распределение пикселей фрагмента. На основе похожести кадров съёмки и среднего кадру строят кривую расстояний. Пики этой кривой выбирают в качестве ключевых кадров. К плюсам методов со ссылочными кадрами строит отнести понятность и простоту реализации. К минусам — зависимость от построения ссылочного кадра, потенциально есть риск потери информации [63].

### 2.2.4. УПРОЩЕНИЕ КРИВЫХ

В работе [12] предложен алгоритм реального времени, который выделяет ключевые кадры на основе статистики макроблоков из потока сжатого видео. Авторы работают с MPEG-2.

В предложенном методе авторы строят кривую разницы кадров. При этом учитываются различные типы макроблоков и варианты смены контекста видео на определенном макроблоке. Построенная кривая разницы кадров содержит в себе шум. С помощью фильтра Гаусса и дискретного выделения контура (Discrete Contour Evolution) кривая «спрямляется» — выделяются сегменты с наибольшей разницей.

Авторы анализируют углы наклона сегментов кривой и на основе этого выделяют ключевые кадры. Аналогичным методом в работе выделяются и съёмки. К минусам подхода стоит отнести рост вычислительной сложности и ростом качества спрямления кривой разницы кадров.

### 2.2.5. КЛАСТЕРИЗАЦИЯ

Алгоритмы выделения ключевых кадров, основанные на кластеризации, группируют похожие кадры в кластеры. Те кадры, которые оказываются максимально близко к центрам кластеров считают ключевыми. Преимущества этих алгоритмов — в том, что они могут использовать общие методы кластеризации, полученные таким образом кадры будут отражать глобальные особенности видео. Однако, в предложенных методах никак не учитывается последовательная структура видео. Кроме того, на больших объемах данных сложно построить семантически значимые группы кадров, а на практике соседние кадры могут попасть в разные группы.

В работе [40] для иерархической кластеризации кадров применяют метод полной связи (complete-linkage clustering).

Авторы [59] предлагают использовать модели гауссовых смесей (МГС) в собственном пространстве кадров. Число компонентов МГС определяет обязательное количество кластеров.

В работе [116] используют нечеткий метод К-средних.

### 2.2.6. ОБЪЕКТЫ И СОБЫТИЯ

Методы сегментации, основанные на поиске объектов и событий, выбирают ключевые кадры, содержащие объекты или информацию о событии. Кадры, полученные таким образом, будут иметь семантическое значение.

Ограничение этих алгоритмов заключается в том, что обнаружение предмета или события сильно зависит от эвристических правил, определенных в соответствии с решаемой задачей. Эти алгоритмы, эффективны только тогда, когда экспериментальные параметры были подобраны правильно.

В работе [48] предложено считать ключевыми кадры, в которых происходит взаимодействие объектов (области объектов соприкасаются). В работе [53] для извлечения ключевых кадров используют функции формы, которые описывают действия и жесты человека.

Авторы [58] предварительно выделяют ключевые кадры на основе цветowych гистограмм. Потом к этим кадрам применяют оценку модели гауссовых смесей (МГС). для выделения объектов. Обученный МГС и выделенные объекты используются для уточнения предварительных ключевых кадров.

В работе [95] предлагают извлекать ключевые кадры совместно с выделением объектов. Для этого строят общее пространство признаков для обоих процессов. Извлечение ключевого кадра сформулировано как процесс выбора характеристик для сегментации объекта в контексте моделирования видео на основе модели гауссовых смесей (МГС).

Авторы [59] считают движение наиболее важной характеристикой в представлении действий или событий в видео. Таким образом, должна быть функция для определения ключевых кадров. Для моделирования закономерности движения в видео предлагается треугольная модель воспринимаемой энергии движения (perceived motion energy, PME). Предложенная модель аппроксимирует энергетический график движения треугольниками [121]. Вершины треугольников выбираются в качестве ключевых кадров. Таким образом, ключевые кадры отвечают моментам смены ускорения и торможения движения

Очень похожая идея обсуждается в работе [41]: ключевыми кадрами считают наибольшие изгибы траектории движения камеры. Эти изгибы образуют временные особые точки, по аналогии с особыми точками внутри кадра.

## 2.3. СЕГМЕНТАЦИЯ СЦЕН

Сегментация сцен также известна как деление сюжета на блоки.



Сцена представляет собой группу смежных съёмок. Эти съёмки связаны между собой конкретной темой или предметом. Сцены обладают семантикой более высокого уровня чем съёмки.

Существует три способа сегментации сцен:

- деление по ключевым кадрам;
- деление на основе объединения визуальной и звуковой информации;
- деление по фону.

### **2.3.1. ДЕЛЕНИЕ ПО КЛЮЧЕВЫМ КАДРАМ**

При делении сцен по ключевым кадрам каждая съёмка представляется набором ключевых кадров. Для кадров выявляют их характеристики. Близкие по времени кадры с близкими характеристиками группируют в сцены [106].

Например, в работе [42] сходство съёмок оценивают используя сравнение блоков ключевых кадров. Похожие съёмки склеиваются и образуют сцены.

В работе [60] анализируют траектории движения во временных срезах последовательностей кадров. Таким образом, для компактного представления съёмки используется стратегия выделения ключевого кадра на основе движения. Смены сцены определяют по схожести ключевых кадров в соседних съёмках.

Ограничение деления по ключевому кадру заключается в том, что кадры не могут эффективно представить динамическое содержание съёмки. Съёмки в пределах сцены, как правило, связаны динамическим развитием сюжета, а не сходством ключевых кадров.

### **2.3.2. ДЕЛЕНИЕ ПО АУДИОВИЗУАЛЬНОЙ ИНФОРМАЦИИ**

При одновременном анализе звуковой и визуальной информации сменой сцен считают границу съёмки, где содержимое обоих каналов изменяется одновременно. Для определения соответствия между этими двумя наборами сцен используют алгоритм поиска ближайшего соседа с ограничением по времени [98]. К минусам подхода следует отнести сложность определения связи между аудио сегментами и визуальными съёмками.

### 2.3.3. ДЕЛЕНИЕ ПО ФОНУ

Деление сцен по фону основано на гипотезе, что съёмки, принадлежащие к одной сцене часто имеют один и тот же фон. Такое предположение является и главной проблемой этого подхода — в видео иногда съёмки одной сцены имеют разный фон.

Для восстановления фона каждого кадра в работе [19] используют так называемый «мозаичный подход» — объединение близких по цвету пикселей в одноцветные прямоугольные области. Сходство съёмок определяют с помощью оценки распределения цвета и текстуры всех фоновых изображений в кадре. Для управления процессом группировки съёмок применяют кинематографические правила.

### СЛИЯНИЕ, ВОСХОДЯЩИЙ ПОДХОД

Подобные съёмки последовательно объединяют. Таким образом формируется сцена.

В работе [76] предлагается двупроходный алгоритм выделения сцен. При первом проходе производятся анализ подобия цветов. На его основе выделяют потенциальные границы сцен. Авторы вводят понятие обратной связи съёмок (Backward Shot Coherence, BSC) — мера по подобия данной съёмки относительно ранее предыдущих съёмок. Обратной связью оценивается как максимальное пересечение цветовых гистограмм кадров данной съёмки с кадрами  $N$  предыдущих съёмок. Минимумы обратной связи съёмок помечают как потенциальные границы сцен. Во втором проходе оценивается динамика видео вблизи потенциальных границ сцен. Потенциальные границы сцен, по обе стороны от которых динамика движения превышает заданный порог — удаляются. Оставшиеся потенциальные границы объявляют границами сцен.

В работе [115] используется вероятностный алгоритм кластеризации (best-first model merging [66]). В качестве вероятностной модели выступает скрытая марковская модель (НММ). Аналогичная кластеризация применяется в работе [71] для группировки изображений в альбомы. В данном случае,

скрытая переменная — принадлежность сцене, а наблюдаемая переменная — визуальная информация съёмки. Для представления съёмки вычисляют усредненную цветовую гистограмму ключевых кадров.

## РАЗДЕЛЕНИЕ, НИСХОДЯЩИЙ ПОДХОД

Авторы [77] строят граф различий для съёмок видео. Далее, используя нормализованный разрез, выделяют похожие съёмки, которые образуют сцены.

Согласно работе [104] сцена это

## СТАТИСТИЧЕСКАЯ МОДЕЛЬ

[120]

[103]

[30]

## КЛАССИФИКАЦИЯ ГРАНИЦ

[81]

## 2.4. СЕГМЕНТАЦИЯ ЗВУКА

Звуковая дорожка — богатый источник информации о содержании для всех жанров видео.

Как показано в лингвистической литературе границы «высказываний» выделяются интонационно. На существенные изменения темы обычно указывают:

- длинные паузы;
- изменения тона;
- изменением амплитуды колебаний.

Для автоматического деления речи на темы применяется вероятностная модель связи интонационных и лексических сигналов. Сначала извлекают большое количество интонационных характеристик. И, таким образом, получают два главных типа речевой просодии: продолжительность и тон.

На основе дерева принятия решений выбирают типичную интонационную функцию. После чего, лексическая информация извлекается с помощью Скрытых Моделей Маркова (НММ) и статистических моделей языка.

Аудио является перспективным источником информации для анализа лекционных видео. Обычно такие видео длятся 60 — 90 минут. Сложно искать интересующий отрывок по всему такому видео [78]. Для решения проблемы используют технологии распознавания речи. Сначала текст извлекают из аудио, а потом производят индексацию стенограммы для поиска по ней [51]. Например, система распознавания речи Sphinx-4 при поиске по видео достигает полноты 72% и средней точности 84%.

## 3. ВЫДЕЛЕНИЕ ПРИЗНАКОВ

Из полученных частей видео выделяют признаки. К признакам относят:

- характеристики ключевых кадров;
- объекты;
- движение в кадре;
- характеристики аудио и текста.

### 3.1. ХАРАКТЕРИСТИКИ КЛЮЧЕВЫХ КАДРОВ

Выделяют цветовые, текстурные, формовые, краевые характеристики.



Рис. 4. Классификация признаков для выделения при ассоциативном видео-поиске

### 3.1.1. ЦВЕТА

Цветовые характеристики включают в себя цветовые гистограммы, цветовые моменты, цветовые коррелограммы, смесь Гауссовых моделей. При выделении локальной цветовой информации изображения разбивают на блоки  $5 \times 5$  [118].

### 3.1.2. ТЕКСТУРЫ

Текстурными характеристиками называют визуальные особенности поверхности некоторого объекта. Они не зависят от тона или насыщенности цвета

объекта. Текстурные характеристики отражают однородные явления в изображениях. Для выделения текстурной информации из видео применяют фильтры Габора [36].

### **3.1.3. КОНТУРЫ**

Контурные или формовые характеристики, описывают формы объектов в изображениях. Они могут быть извлечены из контуров или областей объектов.

### **3.1.4. КРАЯ**

На конференции TRECVID-2005 для получения пространственного распределения краев в задаче поиска по видео был предложен дескриптор гистограммы границ (EHD) [46].

## **3.2. ХАРАКТЕРИСТИКИ ОБЪЕКТОВ**

Такие характеристики включают параметры областей изображения, которые соответствуют объектам:

- основной цвет;
- текстуру;
- размер и т. д.

В работе [90] предложена система поиска лиц. По видео-запросу с конкретным человеком система способна выдать ранжированный список съёмок с этим человеком. Текстовая индексация и поиск приводят к расширению семантики запроса и делают возможным использования Glimpse-метода (agrep), для поиска нечеткого соответствия [57].

### 3.3. ХАРАКТЕРИСТИКИ ДВИЖЕНИЯ

Характеристики движения ближе к смысловым понятиям, чем характеристики ключевых статических кадров и объектов. Движение в видео может быть вызвано движением камеры и движением предметов в кадре.

Движения камеры такие как «приближение или удаление», «панорамирование влево или вправо» и «смещение вверх или вниз» используются для индексации видео. Движения объектов на данный момент являются предметом исследований.

### 3.4. ЗВУКОВЫЕ ХАРАКТЕРИСТИКИ

Преимущество аудио-подходов состоит в том, что они обычно требуют меньше вычислительных ресурсов, чем визуальные методы. Кроме того, аудио-записи могут быть очень короткими.

Многие звуковые характеристики выбраны на основе человеческого восприятие звука. Характеристики аудио можно разделить на три уровня [19]:

- низкоуровневая акустика, такая как средняя частота для кадра;
- средний уровень, как признак объекта, например звук скачущего мяча;
- высокоуровневые, такие как речь и фоновая музыка; играющая в определенных типах видео.

В работе [108] используют блочные характеристики аудио. Аудио-поток при этом разделяется на отрезки в 2048 отсчетов. Для выделения таких характеристик применяют функцию Ханна и логарифмическую шкалу.

## 4. АНАЛИЗ ВИДЕО

Интеллектуальный анализ данных в больших базах видео стал доступен недавно. Задачи анализа видеoinформации можно сформулировать как выявление:

- структурных закономерностей видео;
- закономерностей поведения движущихся объектов;
- характеристик сцены;
- шаблонов событий и их связей;
- и других смысловых атрибутов в видео.

В работах применяют извлечение объектов — группировку различных экземпляров того же объекта, который появляется в различных частях видео.

Для классификации пространственных характеристик кадров применяют метод поиска ближайших соседей [3].

Обнаружение специальных шаблонов применяется к действиям и событиям, для которых есть априорные модели, такие как действия человека, спортивные мероприятия, дорожные ситуации или образцы преступлений [74].

Поиск моделей — автоматическое извлечение неизвестных закономерностей в видео. Для поиска моделей используют экспертные системы с безнадзорным или полуконтролируемым обучением.

Поиск неизвестных моделей полезен для изучения новых данных в наборе видео. Неизвестные образцы обычно находят благодаря кластеризации различных векторов характеристик.

Для выявления закономерностей поведения движущихся объектов используют n-граммы и суффиксные деревья. При этом анализируют последовательности событий по многократным временным масштабам.

## 5. КЛАССИФИКАЦИЯ ВИДЕО

Задача классификации состоит в том, чтобы отнести видео к predetermined категории. Для этого используют характеристики видео или результаты интеллектуального анализа данных.



Классификация видео — хороший способ увеличить эффективность видео-поиска. Семантический разрыв между низкоуровневыми данными и интерпретацией наблюдателя, делает ассоциативную классификацию очень трудной задачей.

Смысловая классификация видео может быть выполнена на трех уровнях [102]:

- жанры:
  - например, «фильмы», «новости», «спортивные соревнования», «мультфильмы», «реклама» и т. д.
- события видео;
- и объекты в видео.

## 5.1. ЖАНРЫ

Жанровая классификация разделяет видео на подмножество соответствующее жанру и несоответствующее [117].

В работе [50] предложена классификация большого числа видео только по заголовку видео. Для этого использован поэтапный метод опорных векторов.

Видео классифицируют также на основе статистических моделей различных жанров. Для этого анализируют структурные свойства: статистику цвета, съёмки, движение камеры и объектов. Свойства используются, чтобы получить более абстрактные атрибуты стиля. К абстрактным атрибутам стиля можно отнести: панорамирование камеры и изменение масштаба, речь и музыке. Строят отображение этих атрибутов на жанры видео.

В работе [110] для классификации жанров используется комбинация из четырех дескрипторов:

- блоковый аудио дескриптор:

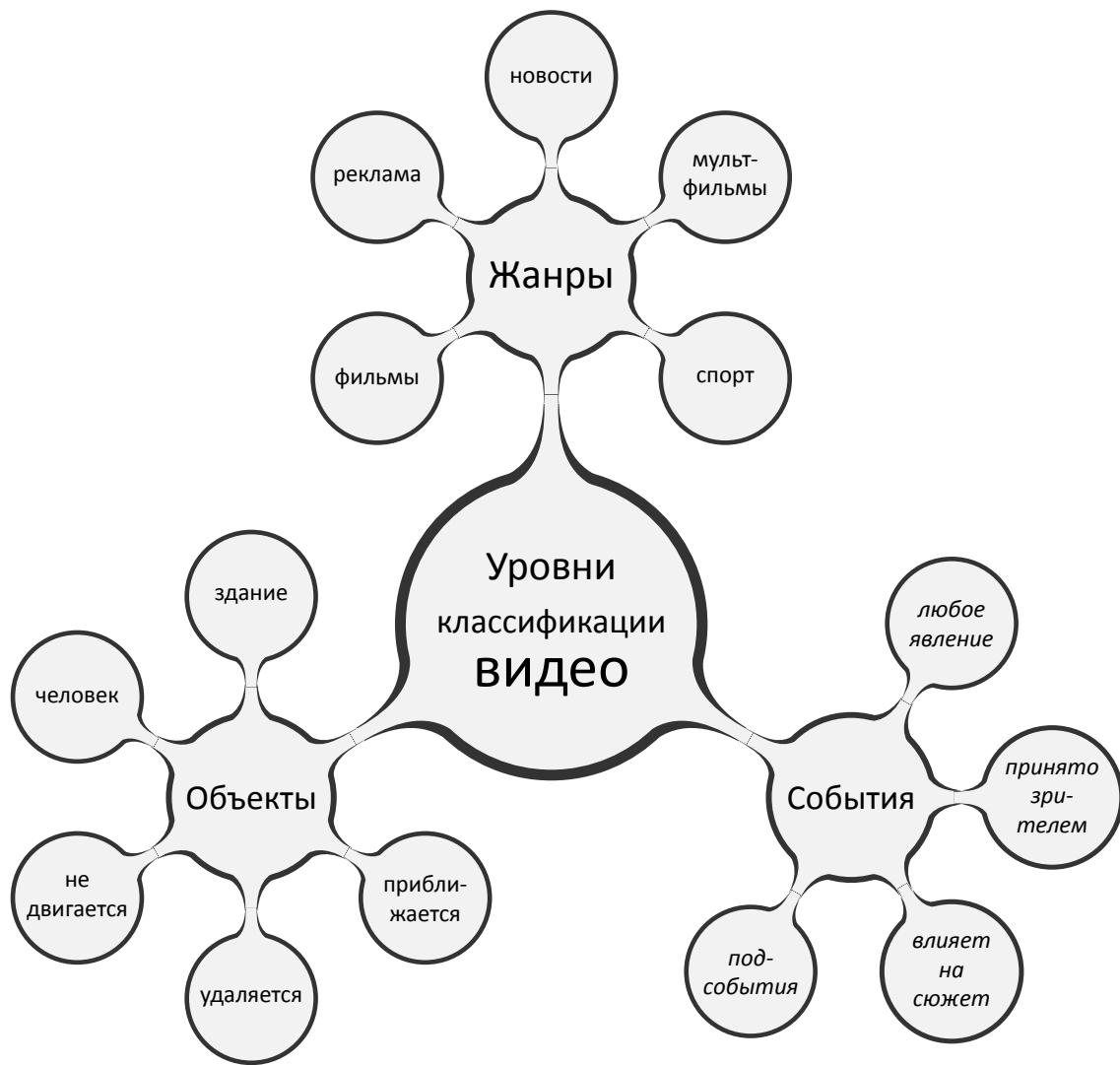


Рис. 5. Уровни классификации видео

- захватывает локальную временную информацию;
- дескриптор визуальной временной структуры:
  - использует информацию о смене съёмки,
  - оценивает количество съёмок за определенный интервал времени («ритм» видео),
  - описывает «активные» и «не активные» смены съёмок;
- дескриптор цвета:
  - использует статистику распределения цвета, элементарных оттенков, цветовых свойств, и отношений между цветами;

- статистика фигур контуров.

Были проведены эксперименты на видеоматериалах общей продолжительностью 91 час видео. Классификация проводилась на семи жанрах видео: мультфильмы, реклама, документальные фильмы, художественные фильмы, музыкальные клипы, спортивные соревнования и новости. Комплексный дескриптор позволил авторам достичь точности 87% — 100% и полноты 77% — 100%.

## 5.2. СОБЫТИЯ

Событие может быть определено как любое явление в видео, которое

- может быть воспринято зрителем;
- играет роль для представления содержимого.

Каждое видео может состоять из многих событий, и каждое событие может состоять из многих подсобытий. Таким образом складывается иерархическая модель [15].

## 5.3. ОБЪЕКТЫ

Объектная классификация является самым низкоуровневым типом классификации. Съёмки классифицируют тоже на основе объектов. Объекты в съёмках представлены с помощью параметров цвета, текстуры и траектории. В работе [45] для кластеризации связанных съёмок используется нейронная сеть. Каждый кластер отображен на одну из 12 категорий. Объекты разделяются по положению в кадре и характеру движения.

## 6. АННОТИРОВАНИЕ ВИДЕО

Процесс присваивания переопределенных смысловых понятий фрагментам видео называют аннотированием. Примеры смысловых понятий: человек, автомобиль, небо и гуляющие люди.

Аннотирование видео подобно классификации, за исключением двух различий.

- 1) Для классификаций важны жанры, а для аннотирования понятия. Жанры и понятия имеют различную природу, несмотря на то, что некоторые методы могут быть использованы в обоих задачах.
- 2) Классификация видео применяется к полным видео, в то время как аннотируют обычно фрагменты [62].

Аннотирование, основанное на обучении, необходимо для анализа и понимания видео. Было предложено много различных способов автоматизации процесса.

Например, в работе [39] было разработано «быстрое полуконтролируемое графовое обучение на нескольких экземплярах» (Fast Graphbased Semi-Supervised Multiple Instance Learning — FGSSMIL). Алгоритм работает в рамках общей платформы для разных типов видео одновременно (спортивные передачи, новости, художественные фильмы). Для обучения модели используется небольшое число видео, размеченных вручную, и значительный объем не размеченного материала.

В работе [84] предлагается создавать частичную ручную аннотацию видео как часть практической профессиональной подготовки. Авторы рассматривают лабораторные занятия студентов-медиков. Во время занятия идет запись видео. Кроме того одновременно происходит запись изменения состояния тренировочного манекена (виртуального пациента). Таким образом, к записанному видео добавляется семантическая разметка на основе показаний датчиков манекена. После происходит разбор занятия и анализ допущенных ошибок, В результате к видео добавляется разметка, созданная самими студентами.



Рис. 6. Типы запросов при ассоциативном видео-поиске

## 7. ОБРАБОТКА ЗАПРОСА

После построения поискового индекса может быть выполнен ассоциативный поиск. Поисковая выдача оптимизируется на основе связи между запросами.

### 7.1. ТИПЫ ЗАПРОСОВ

Существует две категории запросов: семантические и не семантические.

#### 7.1.1. СЕМАНТИЧЕСКИЕ ЗАПРОСЫ

К семантическим запросам относят наборы ключевых слов и поисковые фразы. Ключевые слова — наиболее очевидный и простой вид запроса. При таких запросах частично учитывается семантика видео. Поисковые фразы или запросы на естественном языке — самый естественный и удобный способ взаимодействия человека с поисковой системой. Для выбора и ранжирования видео используется смысловая близость слов [6].

## 7.1.2. НЕ СЕМАНТИЧЕСКИЕ ЗАПРОСЫ

Не семантические запросы используются для поиска по образцу, эскизам, объектам и т. д... Запросом может быть изображение или видео.

### ПОИСК ПО ОБРАЗЦУ

При таком поиске из запроса выделяют низкоуровневые характеристики и сравнивают их с данными в базе с помощью меры сходства.

### ПОИСК ПО ЭСКИЗУ

Пользователи могут изобразить нужное видео с помощью эскиза. Далее для эскиза применяется поиск по образцу.

### ПОИСК ОБЪЕКТА

В качестве запроса выступает изображение объекта. Система находит и возвращает все вхождения объекта в материалах из базы [91]. В отличие от предыдущих видов запросов, в данном случае, привязка происходит не к видео, а именно к изображенному объекту.

## 7.2. ОЦЕНКА СХОДСТВА

Критерии близости видео является важным фактором при поиске. Выделяют несколько способов сравнения видео:

- сравнение характеристик;
- сравнение текста;
- сравнение онтологий.

Применяют также комбинации методов. Выбор конкретного метода зависит от типа запроса.

## 7.3. СРАВНЕНИЕ ХАРАКТЕРИСТИК

При сравнении характеристик видео оценивают среднее расстояние между особенностями соответствующих кадров [11].

## 7.4. СРАВНЕНИЕ ТЕКСТА

Для сравнения запроса и описания видео применяют текстовое сопоставление. Описание и запрос нормализуют, а затем вычисляют их смысловое сходство, используя пространственные векторные модели [1].

## 7.5. СРАВНЕНИЕ ОНТОЛОГИЙ

При сравнении онтологии оценивают смысловое сходство отношений между ключевыми словами запроса и описанием аннотированного видео [6].

Для усиления влияния смысловых понятий автоматически подбирают комбинации методов. Для этого исследуют различные стратегии на учебном наборе видео.

## 7.6. ОЦЕНКА РЕЛЕВАНТНОСТИ

Видео из поисковой выдачи оцениваются или пользователем или автоматически. Эту оценку используют для уточнения дальнейших поисков. Обратная связь релевантности устраняет разрыв между смысловым понятием адекватности поискового ответа и низкоуровневым представлением видео.

Явная обратная связь предлагает пользователю выбрать релевантные видеоролики из ранее полученных ответов. На основе мнений пользователей системы меняют коэффициенты мер подобия [18].

Неявная обратная связь уточняет результаты поиска на основе кликов и переходов пользователя.

Псевдообратная связь выделяет положительные и отрицательные выборки из предыдущих результатов поиска без участия человека.

Рассматривая текстовую и визуальную информацию с вероятностной точки зрения, визуальное ранжирование можно сформулировать как задачу байесовской оптимизации. Такое прием называют байесовским визуальным ранжированием.

## 8. ЗАКЛЮЧЕНИЕ

Многие вопросы остаются открытыми и требуют дальнейшего исследования, особенно в следующих областях.

Большинство современных подходов индексации видео сильно зависят от предварительных знаний о предметной области.

Это ограничивает их расширяемость для новых областей.

Устранение зависимости от предварительных знаний — важная задача будущих исследований.

Индексация и поиск видео в среде «облачных» вычислений сформировали новое направление исследований видео-поиска.

Важной особенностью «облачных» вычислений является то, что искомые видео и сама база данных меняются динамически.

Современные подходы к смысловому поиску видео, как правило, используют набор текстов для описания визуального содержания видео.

В этой области пока осталось много неразрешенных вопросов.

Например, отдельной темой для исследования может быть эмоциональная семантика видео [102]. Эмоциональная семантика описывает человеческие психологические ощущения, такие как радость, гнев, страх, печаль, и пр.

Эмоциональный видео-поиск — поиск материалов, которые вызывают конкретные чувства у зрителя. Для имитации человеческого восприятия могут быть использованы новые подходы к видео-поиску.

Темой для дальнейшего изучения является мультимедийный человеко-машинный интерфейс, в частности:



- расположение мультимедийной информации;
- удобство интерфейса для решения задач пользователя;
- пригодность интерфейса для оценки и обратной связи пользователей;
- и способность интерфейса адаптироваться к привычкам пользователей и отражать их индивидуальность.

Организация и визуализация результатов поиска — также интересная тема исследования. На данный момент проблема сочетания множественных информационных моделей на различных уровнях абстракции остается нерешенной.

Эффективное использование информации о движении имеет большое значение для поиска видео. Важными задачами направления являются:

- способность различать движения фона и переднего плана;
- умение обнаруживать движущиеся объекты и определять события в кадре;
- объединение статических характеристик и характеристик движения;
- построение индекса движения.

Интересными вопросами для исследования остаются:

- быстрый видео-поиск с помощью иерархических индексов;
- адаптивное обновление иерархической индексной модели;
- обработка временных характеристик видео во время создания и обновления индекса;
- динамические меры сходства видео на основе выбора статистических функций.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Adding Semantics to Detectors for Video Retrieval / Cees G. M. Snoek, Bouke Huurnink, Laura Hollink et al. // IEEE Transactions on Multimedia. — 2007. — August. — Vol. 9, no. 5. — P. 975–986. [38](#)
2. Adult movie classification system based on multimodal approach with visual and auditory features / JaeDeok Lim, ByeongCheol Choi, SeungWan Han, ChoelHoon Lee // Information Science and Digital Content Technology (ICIDT), 2012 8th International Conference on. — Vol. 3. — 2012. — June. — P. 745–748. [9](#)
3. Anjulan A, Canagarajah CN. A unified framework for object retrieval and mining // IEEE Transactions on Circuits and Systems for Video Technology. — 2009. — Vol. 19 (1). — P. 63 – 76. — Publisher: IEEE. [31](#)
4. Applications of Video-Content Analysis and Retrieval / Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray et al. // IEEE MultiMedia. — 2002. — jul. — Vol. 9, no. 3. — P. 42–55. — URL: <http://dx.doi.org/10.1109/MMUL.2002.1022858>. [6](#)
5. Awad George, Over Paul, Kraaij Wessel. Content-Based Video Copy Detection Benchmarking at TRECVID // ACM Trans. Inf. Syst. — 2014. — jul. — Vol. 32, no. 3. — P. 14:1–14:40. — URL: <http://doi.acm.org/10.1145/2629531>. [6](#)
6. Aytar Yusuf, Shah Mubarak, Luo Jiebo. Utilizing semantic word similarity measures for video retrieval // 2013 IEEE Conference on Computer Vision and Pattern Recognition. — 2008. — Vol. 0. — P. 1–8. [36](#), [38](#)
7. BUPT at TRECVID 2007: Shot Boundary Detection. / Zhi-Cheng Zhao, Xing Zeng, Tao Liu, An-Ni Cai // TRECVID / Ed. by Paul Over, George Awad, Wessel Kraaij, Alan F. Smeaton. — National Institute of Standards and Technology (NIST), 2007. — URL: <http://dblp.uni-trier.de/db/conf/trecvid/trecvid2007.html>. [16](#)

8. Babu R. Venkatesh, Ramakrishnan K. R. Compressed domain video retrieval using object and global motion descriptors. // *Multimedia Tools Appl.* — 2007. — Vol. 32, no. 1. — P. 93–113. — URL: <http://dblp.uni-trier.de/db/journals/mta/mta32.html>. 6
9. Bayesian Visual Reranking / Xinmie Tian, Linjun Yang, Jingdong Wang et al. // *Trans. Multi.* — 2011. — aug. — Vol. 13, no. 4. — P. 639–652. — URL: <http://dx.doi.org/10.1109/TMM.2011.2111363>. 8, 9
10. Boreczky J.S., Wilcox L.D. A hidden Markov model framework for video segmentation using audio and image features // *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on.* — Vol. 6. — 1998. — May. — P. 3741–3744 vol.6. 15
11. Browne Paul, Smeaton Alan F. Video retrieval using dialogue, keyframe similarity and video objects. // *ICIP (3).* — IEEE, 2005. — P. 1208–1211. — URL: <http://dblp.uni-trier.de/db/conf/icip/icip2005-3.html>. 38
12. Calic J., Izuierdo E. Efficient key-frame extraction and video analysis // *Information Technology: Coding and Computing, 2002. Proceedings. International Conference on.* — 2002. — April. — P. 28–33. 21
13. Cernekova Zuzana, Pitas Ioannis, Nikou Christophoros. Information theory-based shot cut/fade detection and video summarization. // *IEEE Trans. Circuits Syst. Video Techn.* — 2006. — Vol. 16, no. 1. — P. 82–91. — URL: <http://dblp.uni-trier.de/db/journals/tcsv/tcsv16.html>. 14
14. Chang Hyun Sung, Sull Sanghoon, Lee Sang Uk. Efficient video indexing scheme for content-based retrieval // *Circuits and Systems for Video Technology, IEEE Transactions on.* — 1999. — Dec. — Vol. 9, no. 8. — P. 1269–1279. 20
15. Chang Peng, Han Mei, Gong Yihong. Extract highlights from baseball game video with hidden Markov models // *Image Processing. 2002. Proceedings. 2002 International Conference on.* — Vol. 1. — 2002. — P. I–609–I–612 vol.1. 34

16. Chantamunee Siripinyo, Gotoh Yoshihiko. University of Sheffield at TRECVID 2007: Shot Boundary Detection and Rushes Summarisation. // TRECVID / Ed. by Paul Over, George Awad, Wessel Kraaij, Alan F. Smeaton.— National Institute of Standards and Technology (NIST), 2007.— URL: <http://dblp.uni-trier.de/db/conf/trecvid/trecvid2007.html>. 13
17. Chasanis Vasileios, Kalogeratos Argyris, Likas Aristidis. Movie Segmentation into Scenes and Chapters Using Locally Weighted Bag of Visual Words // Proceedings of the ACM International Conference on Image and Video Retrieval.— CIVR '09.— New York, NY, USA : ACM, 2009.— P. 35:1–35:7.— URL: <http://doi.acm.org/10.1145/1646396.1646439>. 9
18. Chen Liang-Hua, Chin Kuo-Hao, Liao Hong-Yuan. An Integrated Approach to Video Retrieval // Nineteenth Australasian Database Conference (ADC 2008) / Ed. by Alan Fekete, Xuemin Lin.— Vol. 75 of CRPIT.— Wollongong, NSW, Australia : ACS, 2008.— P. 49–55. 38
19. Chen Liang-Hua, Lai Yu-Chun, Mark Liao Hong-Yuan. Movie Scene Segmentation Using Background Information // Pattern Recogn.— 2008.— mar.— Vol. 41, no. 3.— P. 1056–1065.— URL: <http://dx.doi.org/10.1016/j.patcog.2007.07.024>. 25, 30
20. Chen Xu, Hero AO., Savarese S. Multimodal Video Indexing and Retrieval Using Directed Information // Multimedia, IEEE Transactions on.— 2012.— Feb.— Vol. 14, no. 1.— P. 3–16. 9
21. Cheng Xiangang, Chia Liang-Tien. Stratification-Based Keyframe Cliques for Effective and Efficient Video Representation // IEEE Transactions on Multimedia.— 2011.— Vol. 13, no. 6.— P. 1333–1342. 9
22. Chu Wei-Ta, Tsai Shang-Yin. Rhythm of Motion Extraction and Rhythm-Based Cross-Media Alignment for Dance Videos // Multimedia, IEEE Transactions on.— 2012.— Feb.— Vol. 14, no. 1.— P. 129–141. 9
23. Content-Based Analysis Improves Audiovisual Archive Retrieval / B. Huurnink, C. G M Snoek, M. de Rijke, A W M Smeulders // Multimedia,

- IEEE Transactions on. — 2012. — Aug. — Vol. 14, no. 4. — P. 1166–1178. 9
24. Cooper M., Foote J. Discriminative techniques for keyframe selection // Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. — 2005. — July. — P. 4 pp.—. 20
25. Cooper Matthew, Liu Ting, Rieffel Eleanor G. Video Segmentation via Temporal Pattern Classification. // IEEE Transactions on Multimedia. — 2007. — Vol. 9, no. 3. — P. 610–618. — URL: <http://dblp.uni-trier.de/db/journals/tmm/tmm9.html>. 13, 14, 15
26. Damnjanovic U., Izquierdo E., Grzegorzec M. Shot Boundary Detection Using Spectral Clustering // 15th European Signal Processing Conference / Ed. by M. Domanski, R. Stasinski, M. Bartkowiak. — Poznan, Poland : PTETiS, Poznan, 2007. — September. — P. 1779–1783. 17
27. Del Fabro Manfred, Böszörményi Laszlo. State-of-the-art and future challenges in video scene detection: a survey. — 2013. — Vol. 19, no. 5. — P. 427–454. — URL: <http://dx.doi.org/10.1007/s00530-013-0306-4>. 9
28. Divakaran A., Radhakrishnan R., Peker K.A. Motion activity-based extraction of key-frames from video shots // Image Processing. 2002. Proceedings. 2002 International Conference on. — Vol. 1. — 2002. — P. I–932–I–935 vol.1. 19
29. Dynamic Selection and Effective Compression of Key Frames for Video Abstraction / Xu-Dong Zhang, Tie-Yan Liu, Kwok-Tung Lo, Jian Feng // Pattern Recogn. Lett. — 2003. — jun. — Vol. 24, no. 9-10. — P. 1523–1532. — URL: [http://dx.doi.org/10.1016/S0167-8655\(02\)00391-4](http://dx.doi.org/10.1016/S0167-8655(02)00391-4). 18, 19
30. EMS: Energy Minimization Based Video Scene Segmentation / Zhiwei Gu, Tao Mei, Xian-Sheng Hua et al. // Multimedia and Expo, 2007 IEEE International Conference on. — 2007. — July. — P. 520–523. 26

31. Event Driven Web Video Summarization by Tag Localization and Key-Shot Identification / Meng Wang, R. Hong, Guangda Li et al. // Multimedia, IEEE Transactions on. — 2012. — Aug. — Vol. 14, no. 4. — P. 975–985. [9](#)
32. Fast Semantic Diffusion for Large-Scale Context-Based Image and Video Annotation / Yu-Gang Jiang, Qi Dai, Jun Wang et al. // Image Processing, IEEE Transactions on. — 2012. — June. — Vol. 21, no. 6. — P. 3080–3091. [9](#)
33. Ferman AM., Tekalp AM. Two-stage hierarchical video summary extraction to match low-level user browsing preferences // Multimedia, IEEE Transactions on. — 2003. — June. — Vol. 5, no. 2. — P. 244–256. [21](#)
34. A Formal Study of Shot Boundary Detection. / Jinhui Yuan, Huiyi Wang, Lan Xiao et al. // IEEE Trans. Circuits Syst. Video Techn. — 2007. — Vol. 17, no. 2. — P. 168–186. — URL: <http://dblp.uni-trier.de/db/journals/tcsv/tcsv17.html>. [12](#), [13](#), [16](#)
35. Freund Yoav. Boosting a Weak Learning Algorithm by Majority // Inf. Comput. — 1995. — sep. — Vol. 121, no. 2. — P. 256–285. — URL: <http://dx.doi.org/10.1006/inco.1995.1136>. [16](#)
36. Fxpal experiments for trecvid 2004 / John Adcock, Andreas Girgensohn, Matthew Cooper et al. // Proceedings of the TREC Video Retrieval Evaluation (TRECVID). — 2004. — P. 70–81. [29](#)
37. Gao Xinbo, Li Jie, Shi Yang // RSKT / Ed. by Guoyin Wang, James F. Peters, Andrzej Skowron, Yiyu Yao. — Lecture Notes in Computer Science. — Springer, 2006. — P. 651–658. [13](#)
38. A General Method for Shot Boundary Detection. / Ling Xue, Li Chao, Li Huan, Xiong Zhang // MUE. — IEEE Computer Society, 2008. — P. 394–397. — URL: <http://dblp.uni-trier.de/db/conf/mue/mue2008.html>. [16](#)
39. A Generic Framework for Video Annotation via Semi-Supervised Learning / Tianzhu Zhang, Changsheng Xu, Guangyu Zhu et al. // IEEE Transactions on Multimedia. — 2012. — P. 1206–1219. [8](#), [9](#), [35](#)

40. Girgensohn Andreas, Boreczky John. Time-Constrained Keyframe Selection Technique // Multimedia Tools Appl.— 2000.— aug.— Vol. 11, no. 3.— P. 347–358.— URL: <http://dx.doi.org/10.1023/A:1009630817712>. 22
41. Han Seung-Hoon, Kweon In-So. Scalable temporal interest points for abstraction and classification of video events // Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on.— 2005.— July.— P. 4 pp.—. 23
42. Hanjalic Alan. Shot-boundary detection: unraveled and resolved? // IEEE Trans. Circuits Syst. Video Techn.— 2002.— Vol. 12, no. 2.— P. 90–105.— URL: <http://dblp.uni-trier.de/db/journals/tcsv/tcsv12.html>. 14, 24
43. Hattarge A.M., Thakre K.S. Analysis And Review Of Formal Approaches To Automatic Video Shot Boundary Detection // International Journal of Advanced Research in Computer and Communication Engineering.— 2014.— Vol. 3.— P. 5266–5271. 12
44. Hoi Steven C.H., Wong Lawson, Lyu Albert. Chinese University of Hong Kong at TRECVID 2006: Shot Boundary Detection and Video Search // Int. TREC Video Retrieval workshop (TRECVID'06).— 2006. 13, 14
45. Hong G.Y., Fong B., Fong A.C.M. An intelligent video categorization engine // Kybernetes.— 2005.— Vol. 34, no. 6.— P. 784–802.— <http://www.emeraldinsight.com/doi/pdf/10.1108/03684920510595490>. 34
46. Informedia at TRECVID 2003: Analyzing and Searching Broadcast News Video / A. G. Hauptmann, R. V. Baron, M. Y. Chen et al. // Proceedings of the TRECVID Workshop.— 2003.— URL: <http://www.science.uva.nl/research/publications/2003/HauptmannPTRECVID2003>. 29
47. Interactive Video Indexing With Statistical Active Learning / Zheng-Jun Zha, Meng Wang, Yan-Tao Zheng et al. // Multimedia, IEEE Transactions on.— 2012.— Feb.— Vol. 14, no. 1.— P. 17–27. 9



48. J. Calic, B. Thomas. Spatial analysis in key-frame extraction using video segmentation // Proc. Workshop Image Anal. Multimedia Interactive Services. — Portugal, 2004. 23
49. Liu Chunxi, Liu Huiying, Jiang Shuqiang et al. JDL at Trecvid 2006 Shot Boundary Detection. — 2006. 16
50. Jiang Yu-Gang, Ngo Chong-Wah, Yang Jun. Towards Optimal Bag-of-features for Object Categorization and Semantic Video Retrieval // Proceedings of the 6th ACM International Conference on Image and Video Retrieval. — CIVR '07. — New York, NY, USA : ACM, 2007. — P. 494–501. — URL: <http://doi.acm.org/10.1145/1282280.1282352>. 32
51. Kamabathula Vijaya Kumar, Iyer Sridhar. Automated Tagging to Enable Fine-Grained Browsing of Lecture Videos // 2012 IEEE Fourth International Conference on Technology for Education. — 2011. — Vol. 0. — P. 96–102. 9, 27
52. Karpenko A, Aarabi P. Tiny Videos: A Large Data Set for Nonparametric Video Retrieval and Frame Classification // Pattern Analysis and Machine Intelligence, IEEE Transactions on. — 2011. — March. — Vol. 33, no. 3. — P. 618–630. 9
53. Kim Changick, Hwang Jenq-Neng. Object-based video abstraction using cluster analysis // Image Processing, 2001. Proceedings. 2001 International Conference on. — Vol. 2. — 2001. — Oct. — P. 657–660 vol.2. 23
54. Le Gall D.J. The MPEG video compression standard // Compcon Spring '91. Digest of Papers. — 1991. — Feb. — P. 334–335. 17
55. Learning Semantic and Visual Similarity for Endomicroscopy Video Retrieval / B. Andre, T. Vercauteren, AM. Buchner et al. // Medical Imaging, IEEE Transactions. — 2012. — June. — Vol. 31, no. 6. — P. 1276–1288. 9
56. Lee Hun-Cheol, Kim Seong-Dae. Iterative key frame selection in the rate-constraint environment // Signal Processing: Image Communication. — 2003. — Vol. 18, no. 1. — P. 1 – 15. — URL:



<http://www.sciencedirect.com/science/article/pii/S0923596502000899>. 20

57. Li Huiping, Doermann D. Video indexing and retrieval based on recognized text // Multimedia Signal Processing, 2002 IEEE Workshop on. — 2002. — Dec. — P. 245–248. 29
58. Liu Lijie, Fan Guoliang. Combined key-frame extraction and object-based video segmentation // Circuits and Systems for Video Technology, IEEE Transactions on. — 2005. — July. — Vol. 15, no. 7. — P. 869–884. 23
59. Liu Tianming, Zhang Hong-Jiang, Qi Feihu. A novel video key-frame-extraction algorithm based on perceived motion energy model // Circuits and Systems for Video Technology, IEEE Transactions on. — 2003. — Oct. — Vol. 13, no. 10. — P. 1006–1013. 22, 23
60. Motion-based video representation for scene change detection / Chong-Wah Ngo, Ting-Chuen Pong, Hong-Jiang Zhang, R.T. Chin // Pattern Recognition, 2000. Proceedings. 15th International Conference on. — Vol. 1. — 2000. — P. 827–830 vol.1. 24
61. Multi-View Video Summarization / Yanwei Fu, Yanwen Guo, Yanshu Zhu et al. // Multimedia, IEEE Transactions on. — 2010. — Nov. — Vol. 12, no. 7. — P. 717–729. 9
62. Multi-modality Web Video Categorization / Linjun Yang, Jiemin Liu, Xiaokang Yang, Xian-Sheng Hua // Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval. — MIR '07. — New York, NY, USA : ACM, 2007. — P. 265–274. — URL: <http://doi.acm.org/10.1145/1290082.1290119>. 35
63. Nabeel Muhammad Asghar, Hussain Fiaz, Manton Rob. Video Indexing: A Survey // International Journal of Computer and Information Technology. — 2014. — January. — Vol. 03. — P. 148–169. 5, 9, 21
64. Ngo Chong-Wah. A Robust Dissolve Detector by Support Vector Machine // Proceedings of the Eleventh ACM International Conference on Multimedia. — MULTIMEDIA '03. — New York, NY, USA : ACM,

2003. — P. 283–286. — URL: <http://doi.acm.org/10.1145/957013.957072>. 16
65. Nigay Laurence, Coutaz Joelle. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion // Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. — CHI '93. — New York, NY, USA : ACM, 1993. — P. 172–178. — URL: <http://doi.acm.org/10.1145/169059.169143>. 8, 12
66. Omohundro Stephen. Best-First Model Merging for Dynamic Learning and Recognition // Advances in Neural Information Processing Systems 4. — Morgan Kaufmann, 1992. — P. 958–965. 25
67. Optimizing support vector machine based classification and retrieval of semantic video events with genetic algorithms / Bashar Tahayna., Mohammed Belkhatir., .M. Saadat Alhashmi, Thomas O'Daniel // Image Processing (ICIP), 2010 17th IEEE International Conference on. — 2010. — Sept. — P. 1485–1488. 9
68. Peng Y.X., Ngo C.W. Hot Event Detection and Summarization by Graph Modeling and Matching // CIVR05. — 2005. — P. 257–266. 6
69. Pereira F., Vetro A., Sikora. Multimedia retrieval and delivery: essential metadata challenges and standards // Proceedings of the IEEE. — 2008. — April. — Vol. 96, no. 4. — P. 721–744. 6
70. Performance Analysis of Using Wavelet Transform in Content Based Video Retrieval System / Yuk Ying Chung, Wai Kwok Jess Chin, Xiaoming Chen et al. // Proceedings of the 2007 Annual Conference on International Conference on Computer Engineering and Applications. — CEA'07. — Stevens Point, Wisconsin, USA : World Scientific and Engineering Academy and Society (WSEAS), 2007. — P. 277–282. — URL: <http://dl.acm.org/citation.cfm?id=1348258.1348307>. 7
71. Platt John C. AutoAlbum: Clustering Digital Photographs using Probabilistic Model Merging. — Institute of Electrical and Electronics Engineers, Inc., 2000. — June. — URL: <http://research.microsoft.com/apps/pubs/default.aspx?id=68971>. 26

72. Porter S.V., Mirmehdi M., Thomas B.T. A shortest path representation for video summarisation // Image Analysis and Processing, 2003.Proceedings. 12th International Conference on. — 2003. — Sept. — P. 460–465. [20](#)
73. Porter Sarah Victoria. Video Segmentation and Indexing Using Motion Estimation : Ph.D. dissertation / Sarah Victoria Porter ; University of Bristol. — 2004. [13](#)
74. Quack T., Ferrari V., Gool L. Video mining with frequent item set configurations // Int. Conf. Image Video Retrieval. — 2006. — P. 360–369. [31](#)
75. Quénot Georges, Moraru Daniel, Besacier Laurent. CLIPS at TRECvid: Shot Boundary Detection and Feature Detection // TRECVID 2003 Workshop Notebook Papers. — 2003. — P. 35–40. [15](#)
76. Rasheed Z., Shah M. Scene detection in Hollywood movies and TV shows // Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. — Vol. 2. — 2003. — June. — P. II–343–8 vol.2. [25](#)
77. Rasheed Z., Shah M. Detection and Representation of Scenes in Videos // Trans. Multi. — 2005. — dec. — Vol. 7, no. 6. — P. 1097–1105. — URL: <http://dx.doi.org/10.1109/TMM.2005.858392>. [26](#)
78. Repp Stephan, Grob Andreas, Meinel Christoph. Browsing within Lecture Videos Based on the Chain Index of Speech Transcription // IEEE Transactions on Learning Technologies. — 2008. — Vol. 1, no. 3. — P. 145–156. [27](#)
79. Rong Jiawei, Jin Wanjun, Wu Lide. Key frame extraction using inter-shot information // Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on. — Vol. 1. — 2004. — June. — P. 571–574 Vol.1. [20](#)
80. SVM-Based Shot Boundary Detection with a Novel Feature. / Kazunori Matsumoto, Masaki Naito, Keiichiro Hoashi, Fumiaki Sugaya //

- ICME. — IEEE, 2006. — P. 1837–1840. — URL: <http://dblp.uni-trier.de/db/conf/icmcs/icme2006.html>. 16
81. An SVM Framework for Genre-Independent Scene Change Detection / N. Goela, Kevin Wilson, Feng Niu et al. // Multimedia and Expo, 2007 IEEE International Conference on. — 2007. — July. — P. 532–535. 26
82. Sargin Mehmet Emre, Aradhye Hrishikesh. Boosting video classification using cross-video signals // Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference. — 2011. — May. — P. 1805–1808. 9
83. Scene Segmentation in Artistic Archive Documentaries / Dalibor Mitrović, Stefan Hartlieb, Matthias Zeppelzauer, Maia Zaharieva // Proceedings of the 6th International Conference on HCI in Work and Learning, Life and Leisure: Workgroup Human-computer Interaction and Usability Engineering. — USAB'10. — Berlin, Heidelberg : Springer-Verlag, 2010. — P. 400–410. — URL: <http://dl.acm.org/citation.cfm?id=1947789.1947824>. 9
84. Semantic Annotation of Ubiquitous Learning Environments / M. J. Weal, D. T. Michaelides, K. Page et al. // IEEE Transactions on Learning Technologies. — 2012. — Vol. 5, no. 2. — P. 143–156. 35
85. Shot Boundary Detection: An Information Saliency Approach / X. Wu, P.C. Yuen, C. Liu, J. Huang // Image and Signal Processing, 2008. CISP '08. Congress on. — Vol. 2. — 2008. — May. — P. 808–812. 9, 13, 14
86. Shot Boundary Detection by a Hierarchical Supervised Approach / G. Camara-Chavez, F. Precioso, M. Cord et al. // Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on. — 2007. — June. — P. 197–200. 13, 16
87. Shot boundary detection using unsupervised clustering and hypothesis testing / Hong Lu, Yap-Peng Tan, Xiangyang Xue, Lide Wu // Communications, Circuits and Systems, 2004. ICCAS 2004. 2004

- International Conference on. — Vol. 2. — 2004. — June. — P. 932–936 Vol.2.  
[13](#), [17](#)
88. Shot boundary detection without threshold parameters. / Harilaos Koumaras, Georgios Gardikis, Georgios Xilouris et al. // J. Electronic Imaging. — 2006. — Vol. 15, no. 2. — P. 020503. — URL: <http://dblp.uni-trier.de/db/journals/jei/jei15.html>. [17](#)
89. Shot reconstruction degree: a novel criterion for key frame selection. / Tie-Yan Liu, Xu-Dong Zhang, Jian Feng, Kwok-Tung Lo // Pattern Recognition Letters. — 2004. — Vol. 25, no. 12. — P. 1451–1457. — URL: <http://dblp.uni-trier.de/db/journals/prl/prl25.html>. [20](#)
90. Sivic J., Everingham M., Zisserman A. Person Spotting: Video Shot Retrieval for Face Sets // ACM International Conference on Image and Video Retrieval. — 2005. [29](#)
91. Sivic Josef, Zisserman Andrew. Video Google: Efficient Visual Search of Videos // Toward Category-Level Object Recognition. — 2006. — P. 127–144. [37](#)
92. Smeaton A. F. Techniques used and open challenges to the analysis, indexing and retrieval of digital video // Information Systems. — 2006. — Vol. 32, no. 4. — P. 545–559. [7](#)
93. Smeaton Alan F., Over Paul, Doherty Aiden R. Video Shot Boundary Detection: Seven Years of TRECVID Activity // Comput. Vis. Image Underst. — 2010. — apr. — Vol. 114, no. 4. — P. 411–418. — URL: <http://dx.doi.org/10.1016/j.cviu.2009.03.011>. [6](#), [8](#), [9](#), [12](#)
94. Snoek Cees G. M., Worring Marcel. Concept-Based Video Retrieval // Found. Trends Inf. Retr. — 2009. — apr. — Vol. 2, no. 4. — P. 215–322. — URL: <http://dx.doi.org/10.1561/1500000014>. [8](#), [9](#)
95. Song Xiaomu, Fan Guoliang. Joint Key-Frame Extraction and Object Segmentation for Content-Based Video Analysis // IEEE Trans. Cir. and Sys. for Video Technol. — 2006. — jul. — Vol. 16, no. 7. — P. 904–914. — URL: <http://dx.doi.org/10.1109/TCSVT.2006.877419>. [23](#)

96. State-of-the-art on Spatio-temporal Information-based Video Retrieval / W. Ren, S. Singh, M. Singh, Y. S. Zhu // Pattern Recogn. — 2009. — feb. — Vol. 42, no. 2. — P. 267–282. — URL: <http://dx.doi.org/10.1016/j.patcog.2008.08.033>. 8
97. Sun Zhonghua, Jia Kebin, Chen Hexin. Video Key Frame Extraction Based on Spatial-Temporal Color Distribution // Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP '08 International Conference on. — 2008. — Aug. — P. 196–199. 21
98. Sundaram H., Chang Shih-Fu. Video scene segmentation using video and audio features // Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on. — Vol. 2. — 2000. — P. 1145–1148 vol.2. 24
99. A Survey on Visual Content-Based Video Indexing and Retrieval / Weiming Hu, Nianhua Xie, Li Li et al. // Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. — 2011. — Nov. — Vol. 41, no. 6. — P. 797–819. 9, 20
100. TRECVID 2007 by the Brno Group. / Adam Herout, Vítězslav Beran, Michal Hradis et al. // TRECVID / Ed. by Paul Over, George Awad, Wessel Kraaij, Alan F. Smeaton. — National Institute of Standards and Technology (NIST), 2007. — URL: <http://dblp.uni-trier.de/db/conf/trecvid/trecvid2007.html>. 16
101. TRECVID 2011-an overview of the goals, tasks, data, evaluation mechanisms and metrics / Paul Over, George Awad, Jon Fiscus et al. // TRECVID 2011-TREC Video Retrieval Evaluation Online / NIST, USA. — 2011. 6
102. Tamizharasan C, S Chandrakala. A Survey On Multimodal Content Based Video Retrieval // International Journal of Emerging Technology and Advanced Engineering. — Vol. 3. — Chennai, INDIA : Sri Sai Ram Engineering College. An ISO 9001:2008 Certified and NBA Accredited Engineering Institute, 2013. — January. 9, 32, 39
103. Tan Yap-Peng, Lu Hong. Model-based clustering and analysis of

- video scenes // Image Processing. 2002. Proceedings. 2002 International Conference on. — Vol. 1. — 2002. — P. I–617–I–620 vol.1. [26](#)
104. Tavanapong W., Zhou Junyu. Shot clustering techniques for story browsing // Multimedia, IEEE Transactions on. — 2004. — Aug. — Vol. 6, no. 4. — P. 517–527. [26](#)
105. Truong Ba Tu, Venkatesh Svetha. Video Abstraction: A Systematic Review and Classification // ACM Trans. Multimedia Comput. Commun. Appl. — 2007. — feb. — Vol. 3, no. 1. — URL: <http://doi.acm.org/10.1145/1198302.1198305>. [18](#), [19](#), [20](#)
106. Truong Ba Tu, Venkatesh S., Dorai C. Scene Extraction in Motion Pictures // IEEE Trans. Cir. and Sys. for Video Technol. — 2003. — jan. — Vol. 13, no. 1. — P. 5–15. — URL: <http://dx.doi.org/10.1109/TCSVT.2002.808084>. [24](#)
107. Unsupervised Video Shot Detection Using Clustering Ensemble with a Color Global Scale-Invariant Feature Transform Descriptor. / Yuchou Chang, Dah-Jye Lee, Yi Hong, James K. Archibald // EURASIP J. Image and Video Processing. — 2008. — Vol. 2008. — URL: <http://dblp.uni-trier.de/db/journals/ejivp/ejivp2008.html>. [13](#), [17](#)
108. Seyerlehner Klaus, Schedl Markus, Pohle Tim, Knees Peter. Using Block-Level Features For Genre Classification, Tag Classification And Music Similarity Estimation. — 2010. [30](#)
109. Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning / Hong Qing Yu, C. Pedrinaci, S. Dietze, J. Domingue // Learning Technologies, IEEE Transactions on. — 2012. — April. — Vol. 5, no. 2. — P. 130–142. [9](#)
110. Video Genre Categorization and Representation using Audio-Visual Information / Bogdan Ionescu, Klaus Seyerlehner, Christoph Rasche et al. // SPIE – Journal of Electronic Imaging. — 2012. — Vol. 21(2). [9](#), [32](#)
111. Video Scene Segmentation with a Semantic Similarity. / Niraj Kumar, Piyush Rai, Chandrika Pulla, C. V. Jawahar // IICAI / Ed. by Bhanu Prasad,



- Pawan Lingras, Ramakant Nevatia. — ICAI, 2011. — P. 970–981. — URL: <http://dblp.uni-trier.de/db/conf/iicai/iicai2011.html>. 9
112. Video Search and Indexing with Reinforcement Agent for Interactive Multimedia Services / Anand Paul, Bo-Wei Chen, Karunanithi Bharanitharan, Jhing-Fa Wang // ACM Trans. Embed. Comput. Syst. — 2013. — feb. — Vol. 12, no. 2. — P. 25:1–25:16. — URL: <http://doi.acm.org/10.1145/2423636.2423643>. 9
113. Video Shot Boundary Detection Algorithm. / Kyong-Cheol Ko, Young Min Cheon, Gye-Young Kim et al. // ICVGIP / Ed. by Prem Kumar Kalra, Shmuel Peleg. — Vol. 4338 of Lecture Notes in Computer Science. — Springer, 2006. — P. 388–396. — URL: <http://dblp.uni-trier.de/db/conf/icvgip/icvgip2006.html>. 14
114. Video browsing interfaces and applications: a review / Klaus Schoeffmann, Frank Hopfgartner, Oge Marques et al. // SPIE Reviews. — 2010. — mar. — Vol. 1, no. 1. — P. 018004. — URL: <http://link.aip.org/link/?SV2/1/018004/1>. 8
115. Video shot grouping using best-first model merging / L. Zhao, W. Qi, Y.-J. Wang et al. // Proc. Storage Retrieval Media Database. — 2001. — P. 262–269. 25
116. Wang Ting, Wu Yu, Chen Long. An Approach to Video Key-frame Extraction Based on Rough Set // Multimedia and Ubiquitous Engineering, 2007. MUE '07. International Conference on. — 2007. — April. — P. 590–596. 22
117. Wu Jun, Worring Marcel. Efficient Genre-Specific Semantic Video Indexing // IEEE Transactions on Multimedia. — 2012. — April. — Vol. 14, no. 2. — P. 291–302. 9, 32
118. Yan Rong, Hauptmann Alexander G. A review of text and image retrieval approaches for broadcast news video // Inf. Retr. — 2007. — Vol. 10, no. 4-5. — P. 445–484. 28



119. Yoo Hun-Woo, Ryoo Han-Jin, Jang Dong-Sik. Gradual shot boundary detection using localized edge blocks. // Multimedia Tools Appl. — 2006. — Vol. 28, no. 3. — P. 283–300. — URL: <http://dblp.uni-trier.de/db/journals/mta/mta28.html>. 13
120. Zhai Y., Shah M. Video scene segmentation using Markov chain Monte Carlo // Multimedia, IEEE Transactions on. — 2006. — Aug. — Vol. 8, no. 4. — P. 686–697. 26
121. Zhang H.J., Liu T. Method and system for extracting key frames from video using a triangle model of motion based on perceived motion energy. — 2008. — apr. — US Patent 7,359,440. URL: <http://www.google.com/patents/US7359440>. 23
122. Zhao Zhi-Cheng, Cai An-Ni. Shot Boundary Detection Algorithm in Compressed Domain Based on Adaboost and Fuzzy Theory. // ICNC (2) / Ed. by Licheng Jiao, Lipo Wang, Xinbo Gao et al. — Vol. 4222 of Lecture Notes in Computer Science. — Springer, 2006. — P. 617–626. — URL: <http://dblp.uni-trier.de/db/conf/icnc/icnc2006-2.html>. 13, 16, 17
123. A compressed-domain approach for shot boundary detection on H.264/AVC bit streams. / Sarah De Bruyne, Davy Van Deursen, Jan De Cock et al. // Sig. Proc.: Image Comm. — 2008. — Vol. 23, no. 7. — P. 473–489. — URL: <http://dblp.uni-trier.de/db/journals/spic/spic23.html>. 17

## ОБ АВТОРЕ

Никитин Илья Константинович, заочный аспирант Московского Авиационного Института.

На данный момент, автор — разработчик в компании «Яндекс» (yandex.ru) и консультант-исследователь в компании «ТиВиЗавр» (tvzavr.ru).

Адрес: 123098, Москва, ул. Живописная, 38—1—26

Электронная почта: w@w-495.ru

Телефон для связи: +7 916 536 05 41

## ABOUT THE AUTHOR

Ilya Konstantinovich Nikitin, post-graduate student of Moscow Aviation Institute.

At the moment, the author is a developer in Yandex company (yandex.com) and a consultant-researcher in TVzavr internet-cinema (tvzavr.com).

Address: 123098, Russia, Moscow, Zhivopisnaya, st., 38—1—26

E-mail: w@w-495.ru

Phone: +7 916 536 05 41

# ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

## A

AdaBoost, 14  
Agrep, 28

## B

Best-first model merging, 24  
BSC, 24

## C

CBVR, 4  
Complete-linkage clustering, 21

## D

DCE, 21  
DCT, 16

## E

EHD, 28  
EMD, 13

## F

FGSSMIL, 34

## G

Glimpse, 28

## H

HMM, 14, 24, 26

## K

Kernel function, 15  
kNN, 14

## M

MPEG, 5  
MPEG-2, 20  
MPEG-7, 18

## N

N-грамма, 30  
NIST, 5  
NNS, 30

## O

OCR, 6

## P

PME, 22

## S

SIFT, 12  
Sphinx-4, 26  
SVM, 14

## T

TRECvid, 5  
TV-Anytime, 5

## A

Адаптивное улучшение, 14

## Б

Байесовская оптимизация, 38

## В

Вассерштейна метрика, 13  
Вытеснение, 11

## Г

Гаусс, 21  
Гауссовы модели, 27  
Граф различий, 25

## Д

ДКТ, 16

## З

Землеройное расстояние, 13

## И

Иерархическая кластеризации, 21  
Интонация, 25

## К

К-means, 16

К-средние, 16

Кадрик, 11

Канторовича метрика, 13

Кинематографический кадр, 11

Ключевой кадр, 16

## М

МГС, 22

Метод опорных векторов, 14, 31

Метод полной связи, 21

Модели гауссовых смесей, 22

Монтажный план, 11

## Н

Нечеткие дубликаты, 5

Нормализованный разрез, 25

## О

Обратная связь

Неявная, 37

Псевдообратная, 37

Явная, 37

Обучение

Без учителя, 30

Полуконтролируемое, 30, 34

Оптический поток, 17

Оптическое распознавание текста, 6

Орнштейна метрика, 13

## П

Поиск ближайшего соседа, 30

Просодия, 26

## Р

Распознавание

речи, 6

текста, 6

Распознавание речи, 26

Рубинштейна метрика, 13

## С

Склейка, 11

Скользящее окно, 13

Скрытые Модели Маркова, 14, 24, 26

Спектральная кластеризация, 16

Ссылочный кадр, 17

Суффиксное дерево, 30

Съемка, 11

## Т

Транспортное расстояние, 13

Треугольная модель, 22

## Ф

Фильтр Гаусса, 21

Фильтры Габора, 28

Фотографический кадр, 11

Функция Ханна, 29

## Ц

Цветовая гистограмма, 16

Цветовая коррелограмма, 27

Цветовой момент, 27

## Я

Ядро, 15