УДК 004.932.4

И. К. Никитин

Национальный исследовательский университет Московский авиационный институт
Волоколамское шоссе, 4, Москва, 125993, Россия

w@w-495.ru

ОБЗОР МЕТОДОВ КОМПЛЕКСНОГО АССОЦИАТИВНОГО ПОИСКА ВИДЕО

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 серьезно усложняет задачу видеопоиска. 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 видеопоиск может рассматриваться как связующее звено между обычным поиском и смысловым поиском по видео.

Ключевые слова: анализ видео, аннотирование видео, видеопоиск, кадры, классификация видео, нечеткие дубликаты видео, ранжирование видео, сцены, съемки.

Введение

В связи с увеличением пропускной способности сетей многие пользователи получили доступ к видео в Интернете. Для примера, каждую минуту на сайт YouTube загружается более 48 часов новых видео. Более 14 миллиардов клипов были просмотрены в мае 2010.

В длинном видео сложно автоматизированно найти интересующий отрывок, а размечать и искать видео вручную очень 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. 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. Необходимость анализа большого объема данных для выделения нужной информации является серьезной проблемой. Для ее решения применяют ассоциативный поиск. В англоязычной литературе ассоциативный видеопоиск называют «content based video retrieval» (CBVR) – поиск по содержимому.

Ассоциативный поиск используется для автоматического реферирования видео, анализа новостных событий, видеонаблюдения, а также в образовательных целях [1].

Видео содержит в себе несколько типов данных. Авторы [2; 3] выделяют следующие:

1) метаданные – заголовок, автор и описание;
2) звуковая дорожка;
3) тексты, полученные при помощи технологии оптического распознавания символов (OCR);
4) визуальная информация кадров видео.

Никитин И. К. Обзор методов комплексного ассоциативного поиска видео // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, вып. 4. С. 71–82.

Некоторые научные работы по ассоциативному поиску видео

Год	Работа	Тема
2008 2011	[5] [6]	Сегментация видео
2010 2012	[7] [8]	Автоматическое реферирование видео
2012 2014	[9; 10; 11] [12]	Индексация видео
2012 2013	[13] [14]	Комплексный ассоциативный поиск
2011	[15; 16]	Представление видео
2012	[17–19]	Смысловой ассоциативный поиск
2012	[20; 18]	Аннотирование видео
2012	[21]	Видеопоиск по движению
2011 2012	[22] [20]	Ранжирование видео
2010 2011 2012	[23] [24] [25; 26]	Классификация видео

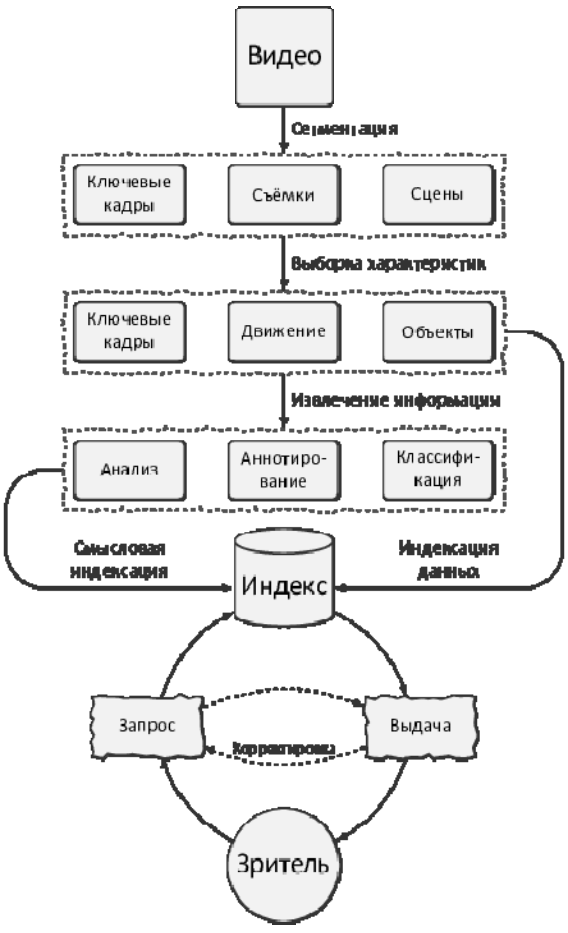

Схема поиска по видео

Таким образом, видео обладает комплексностью. Комплексность (системность, мультимодальность) – способность взаимодействовать с пользователем по различным каналам информации и извлекать и передавать смысл автоматически [4].

Комплексность видео состоит в возможности автора выражать мысли, используя по крайней мере два информационных канала. Каналы могут быть визуальными, звуковыми или текстовыми.

В работе [20] дают хороший обзор аннотации видео. В работе [22] описываются свежие исследования методов ранжирования видео.

Ассоциативный поиск видео состоит из следующих шагов (см. рисунок).

1. Анализ временной структуры видео – деление видео на фрагменты, которое включает обнаружение границ съемок.
2. Определение характеристик фрагментов.
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.

Деление видео

Деление видео включает обнаружение границ съемок, извлечение ключевых кадров, сегментацию сцен и аудио.

Обнаружение границ съемок. Видео делят на фрагменты по времени. В качестве таких фрагментов могут выступать съемки. Съемка (кинематографический кадр, монтажный план) – отрезок кинопленки, 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). Съемка является более крупным делением. Из съемок складываются сцены, а из сцен – видео целиком.

Границы съемок бывают трех типов:

- линейная склейка – съемка внезапно прерывается и начинается другая;
- постепенное исчезновение или проявление (в монохромном кадре);
- вытеснение – исчезновения одной съемки и появления другой (растворение, вытеснение шторкой).

Для обнаружения границ съемок, как правило, сначала извлекают визуальные характеристики каждого кадра. Затем, на основе выделенных признаков, оценивают сходство между кадрами. Границы съемок определяют по смене неоднородных кадров. В работе [4] описаны параметры смены кадров и ошибки выделения на основе глобальных и локальных характеристик для обнаружения съемок и классификации.

Существует два типа методов обнаружения съемок:

1) Пороговые – попарно сравнивают подобия кадров с заданным порогом;
2) Статистические – обнаруживают границы сцен на основе характеристик кадров.

Извлечение ключевых кадров. Среди кадров одной съемки есть избыточность. Для ее уменьшения выделяют кадры, которые наиболее полно отражают содержание съемки.

При извлечении ключевых кадров используют различные характеристики:

- цветовые гистограммы;
- края;
- очертания;
- оптические потоки.

Способы извлечения подразделяются на шесть категорий:

- последовательное сравнение;
- глобальное сравнение;
- на основе ссылочных кадров;
- на основе кластеризации;
- на основе упрощения кривых;
- и на основе объектов или событий [27].

При последовательном сравнении ключевой кадр сравнивают с другими кадрами до тех пор, пока не будет найден «сильно отличный». Для сопоставления кадров используются цветовые гистограммы [28].

Методы глобального сравнения используют различия между кадрами в съемке и распределяют ключевые кадры, минимизируя предопределенную целевую функцию. Методы на основе ссылочных кадров генерируют систему отсчета кадров и затем сравнивают ключевые кадры съемки со ссылочными.

В работе [29] описано создание средней гистограммы без канала прозрачности. С помощью такой гистограммы описывается цветовое распределение кадров в съемке.

Сегментация сцен, *или деление сюжета на блоки*. Сцена представляет собой группу смежных съемок. Эти съемки связаны между собой конкретной темой или предметом. Сцены обладают семантикой более высокого уровня, чем съемки.

Существует три способа сегментации сцен:

- по ключевому кадру;
- на основе объединения визуальной и звуковой информации;
- по фону.

При делении сцен по ключевому кадру каждая съемка представляется набором ключевых кадров. Для кадров выявляют их характеристики. Близкие по времени кадры с близкими характеристиками группируют в сцены [30]. Далее, используя сравнение блоков ключевых кадров, вычисляют сходство между съемками. Ограничение деления по ключевому кадру заключается в том, что кадры не могут эффективно представить динамическое содержание съемки.

Съемки в пределах сцены, как правило, связаны динамическим развитием сюжета в пределах сцены, а не сходством ключевых кадров.

При одновременном анализе звуковой и визуальной информации сменой сцен считают границу съемки, где содержимое обоих каналов изменяется одновременно. Для определения соответствия между этими двумя наборами сцен используют алгоритм поиска ближайшего соседа с ограничением по времени [31]. К минусам подхода следует отнести сложность определения связи между аудиосегментами и визуальными съемками.

Деление сцен по фону основано на гипотезе, что съемки, принадлежащие к одной сцене, часто имеют один и тот же фон. Для восстановления фона каждого кадра используют объединение близких по цвету пикселей в одноцветные прямоугольные области. Сходство съемок определяют с помощью оценки распределения цвета и текстуры всех фоновых изображений в кадре. Для управления процессом группировки съемок применяют кинематографические правила [32].

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;
- изменение тона;
- изменение 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 (HMM) 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 видео [33]. Для решения проблемы используют технологии распознавания речи. Сначала текст извлекают из аудио, а потом производят индексацию стенограммы для поиска по ней [6]. Например, система распознавания речи Sphinx-4 при поиске по видео достигает полноты 72 % и средней точности 84 %.

Выделение признаков

Из полученных частей видео выделяют следующие признаки:

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

Характеристики ключевых кадров. Выделяют цветовые, текстурные, формовые, краевые характеристики.

Цвета. Цветовые характеристики включают цветовые гистограммы, цветовые моменты, цветовые коррелограммы, смесь Гауссовых моделей. При выделении локальной цветовой информации изображения разбивают на блоки 5×5 [34].

Текстуры. Текстурными характеристиками называют визуальные особенности поверхности некоторого объекта. Они не зависят от тона или насыщенности цвета объекта, а отражают однородные явления в изображениях. Для выделения текстурной информации из видео применяют фильтры Габора [35].

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

Края. На конференции TRECVid-2005 для получения пространственного распределение краев в задаче поиска по видео был предложен дескриптор гистограммы границ (EHD) [36].

Характеристики объектов. Такие характеристики включают параметры областей изображения, которые соответствуют объектам:

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

В работе [37] предложена система поиска лиц. По видеозапросу с конкретным человеком система способна выдать ранжированный список съемок с этим человеком. Текстовая индексация и поиск приводят к расширению семантики запроса и делают возможным использования Glimpse-метода (agrep) для поиска нечеткого соответствия [38].

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.

Звуковые характеристики. Преимущество аудиоподходов состоит в том, что они обычно требуют меньше вычислительных ресурсов, чем визуальные методы. Кроме того, аудиозаписи могут быть очень короткими.

Многие звуковые характеристики выбраны на основе человеческого восприятие звука. Характеристики аудио можно разделить на три уровня [32]:

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

В работе [39] используют блочные характеристики аудио. Аудиопоток при этом разделяется на отрезки в 2 048 отсчетов. Для выделения таких характеристик применяют функцию Ханна и логарифмическую шкалу.

Представление видео

В работе [5] сформулирована проблема машинного представления видео. В работе [40] разработаны многослойные, графические аннотации видео – мультимедийные потоки. Они представляют собой визуальный язык как способ представления видеоданных. Особое внимание уделено проблеме создания глобального архива видео, допускающего повторное использование. Нисходящие поисковые системы используют высокоуровневые знания определенной предметной области, чтобы генерировать надлежащие представления.

Но, как сказано выше, это не самый удобный подход. Представление, управляемое данными, – стандартный способ извлечь низкоуровневые характеристики и получить соответствующие представления без любых предварительных знаний о предметной области.

Представления на основе данных могут быть сведены к двум основным классам.

1. Сигнальные признаки, которые характеризуют низкоуровневое аудиовизуальное содержание. Это цветовые гистограммы, формы, текстуры,

2. Описательное представление с помощью текста, атрибутов или ключевых слов. Авторы работы [16] предлагают для описания видео использовать послойные графовые клики ключевых кадров (SKCs), которые более компактны и информативны, чем последовательность изображений или ключевые кадры.

Анализ видео

Интеллектуальный анализ данных в больших базах видео стал доступен недавно. Задачи анализа видеоинформации 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 [41]. Обнаружение специальных шаблонов применяется к действиям и событиям, для которых есть априорные модели, такие как действия человека, спортивные мероприятия, дорожные ситуации или образцы преступлений [42].

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.

Классификация видео

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

Классификация видео – хороший способ увеличить эффективность видеопоиска. Семантический разрыв между низкоуровневыми данными и интерпретацией наблюдателя делает ассоциативную классификацию очень трудной задачей.

Смысловая классификация видео может быть выполнена на трех уровнях [14]:

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

Жанры. Жанровая классификация разделяет видео на подмножество, соответствующее жанру и несоответствующее [11]. В работе [43] 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, съемки, 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 музыку. Строят отображение этих атрибутов на жанры видео.

В работе [26] для классификации жанров используется комбинация из четырех дескрипторов.

- Блоковый аудиодескриптор:
  - захватывает локальную временную информацию.
- Дескриптор визуальной временной структуры:
  - использует информацию о смене съемок;
  - оценивает количество съемок за определенный интервал времени («ритм» видео);
  - описывает «активные» и «неактиные» смены съемок.
- 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 %.

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 [44].

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

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 обеих задачах.

2. Классификация видео применяется к полным видео, в то время как аннотируют обычно фрагменты [46].

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

Например, в работе [20] 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 неразмеченного материала.

В работе [47] 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.

Обработка запроса

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

Существует две категории запросов: семантические и несемантические.

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 естест-

венный и удобный способ взаимодействия человека с поисковой системой. Для выбора и ранжирования видео используется смысловая близость слов [48].

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

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

Поиск по эскизу позволяет пользователям изобразить нужное видео с помощью эскиза. Далее для эскиза применяется поиск по образцу.

В качестве запроса при поиске объекта выступает изображение объекта. Система находит и возвращает все вхождения объекта в материалах из базы [49]. В отличие от предыдущих видов запросов в данном случае привязка происходит не к видео, а именно по изображенному объекту.

Оценка сходства. Критерии близости видео являются важным фактором при поиске. Выделяют несколько способов сравнения видео. Это характеристики, текст или онтологии. Применяют также комбинации методов. Выбор конкретного метода зависит от типа запроса.

При сравнении характеристик видео оценивают среднее расстояние между особенностями соответствующих кадров [50].

Для сравнения запроса и описания видео применяют текстовое сопоставление. Описание и запрос нормализуют, а затем вычисляют их смысловое сходство, используя пространственные векторные модели [51].

При сравнении онтологии оценивают смысловое сходство отношений между ключевыми словами запроса и описанием аннотированного видео [48]. Чтобы усилить влияние смысловых понятий, автоматически подбирают комбинации методов. Для этого исследуют различные стратегии на учебном наборе видео.

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 в системе меняют коэффициенты мер подобия [52].

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. Такой прием называют байесовским визуальным ранжированием.

Заключение

Большинство современных подходов индексации видео сильно зависит 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 видеопоиска. 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 [14]. Эмоциональная семантика описывает человеческие психологические ощущения, такие как радость, гнев, страх, печаль, и проч.

Эмоциональный видеопоиск – поиск материалов, которые вызывают конкретные чувства у зрителя. Для имитации человеческого восприятия могут быть использованы новые подходы к видеопоиску.

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 запроса 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 неразрешенной.

Эффективное использование информации о движении имеет большое значение для поиска видео. Важные задачи направления:

- способность различать движения фона и переднего плана;
- обнаружение движущихся объектов и определение события в кадре;
- объединение статических характеристик и характеристик движения;
- построение индекса движения.

Интересными вопросами для исследования остаются:

- быстрый видеопоиск 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 литературы

1. *Nevenka Dimitrova*, *Hong-Jiang Zhang*, *Behzad Shahraray*, *Ibrahim Sezan*, *Thomas Huang*, *and Avideh Zakhor*. Applications of video-content analysis and retrieval // IEEE MultiMedia. 2002. Vol. 9 (3). P. 42–55.

2. *Yuk Ying Chung*, *Wai Kwok Jess Chin*, *Xiaoming Chen*, *David Yu Shi*, *Eric Choi*, *and Fang Chen*. Performance analysis of using wavelet transform in content based video retrieval system // Proceedings of the 2007 Annual Conference on International Conference on Computer Engineering and Applications, CEA'07. Stevens Point, Wisconsin, USA, 2007. P. 277–282.

3. *Smeaton A. F*. Techniques used and open challenges to the analysis, indexing and retrieval of digital video // Information Systems. 2006. Vol. 32 (4). P. 545–559.

4. *Laurence Nigay, Joëlle Coutaz*. A design space for multimodal systems: Concurrent processing and data fusion // In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York, 1993. P. 172–178.

5. *Haase B.*, *Marc Eliot Davis*, *Davis M*. Media streams: Representing video for retrieval and repurposing. Technical report. 1995.

6. *Vijaya Kumar Kamabathula and Sridhar Iyer*. Automated tagging to enable fine-grained browsing of lecture videos // 2012 IEEE Fourth International Conference on Technology for Education. 2011. P. 96–102.

7. *Yanwei Fu*, *Yanwen Guo*, *Yanshu Zhu*, *Feng Liu*, *Chuanming Song*, *and ZhiHua Zhou*. Multi-view video summarization // Multimedia, IEEE Transactions on. 2010. Vol. 12 (7). P. 717–729.

8. *Meng Wang*, *R. Hong*, *Guangda Li*, *Zheng-Jun Zha*, *Shuicheng Yan*, *and TatSeng Chua*. Event driven web video summarization by tag localization and key-shot identification // Multimedia, IEEE Transactions on. 2012. Vol. 14 (4). P. 975–985.

9. *Xu Chen*, *AO. Hero*, *and S. Savarese*. Multimodal video indexing and retrieval using directed information // Multimedia, IEEE Transactions on. 2012. Vol. 14 (1). P. 3–16.

10. *Zheng-Jun Zha*, *Meng Wang*, *Yan-Tao Zheng*, *Yi Yang*, *Richang Hong*, *Chua T.-S*. Interactive video indexing with statistical active learning. Multimedia, IEEE Transactions on. 2012. № 14 (1). P. 17–27.

11. *Jun Wu, Marcel Worring*. Efficient genre-specific semantic video indexing // IEEE Transactions on Multimedia. 2012. № 14 (2). P. 291–302.

12. *Muhammad Nabeel Asghar*, *Fiaz Hussain*, *Rob Manton*. Video indexing: A survey // International Journal of Computer and Information Technology. 2014. Vol. 3. P. 148–169.

13. *Huurnink B., Snoek C. G. M., Rijke M. de, Smeulders A. W. M.* Contentbased analysis improves audiovisual archive retrieval // Multimedia. IEEE Transactions on. 2012. Vol. 14 (4). P. 1166–1178.

14. *Tamizharasan C., Chandrakala S.* A survey on multimodal content based video retrieval // International Journal of Emerging Technology and Advanced Engineering. Chennai, INDIA, 2013. Vol. 3.

15. *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. Vol. 33 (3). P. 618–630.

16. *Xiangang Cheng, Liang-Tien Chia*. Stratification-based keyframe cliques for effective and efficient video representation // IEEE Transactions on Multimedia. 2011. Vol. 13 (6). P. 1333–1342.

17. *Yu-Gang Jiang*, *Qi Dai*, *Jun Wang*, *Chong-Wah Ngo*, *Xiangyang Xue*, *Shih-Fu Chang*. Fast semantic diffusion for large-scale context-based image and video annotation // Image Processing, IEEE Transactions on. 2012. Vol. 21 (6). P. 3080–3091.

18. *Hong Qing Yu*, *Pedrinaci C.*, *Dietze S.*, *Domingue J.* Using linked data to annotate and search educational video resources for supporting distance learning // Learning Technologies, IEEE Transactions on. 2012. Vol. 5 (2). P. 130–142.

19. *Andre B.*, *Vercauteren T.*, *Buchner A. M.*, *Wallace M. B.*, *Ayache N*. Learning semantic and visual similarity for endomicroscopy video retrieval // Medical Imaging, IEEE Transactions. 2012. Vol. 31 (6). P. 1276–1288.

20. *Tianzhu Zhang, Changsheng Xu, Guangyu Zhu, Si Liu, Hanqing Lu.* A generic framework for video annotation via semi-supervised learning // IEEE Transactions on Multimedia. 2012. P. 1206–1219.

21. *Wei-Ta Chu, Shang-Yin Tsai*. Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos // Multimedia, IEEE Transactions on. 2012. Vol. 14 (1). P. 129–141.

22. *Xinmie Tian*, *Linjun Yang*, *Jingdong Wang*, *Xiuqing Wu*, *Xian-Sheng Hua*. Bayesian visual reranking // Trans. Multi. 2011. Vol. 13 (4). P. 639–652.

23. *Bashar Tahayna*, *Mohammed Belkhatir*, *M. Saadat Alhashmi*, *O'Daniel Th.* Optimizing support vector machine based classification and retrieval of semantic video events with genetic algorithms // Image Processing (ICIP). 2010 17th IEEE International Conference on. 2010. P. 1485–1488.

24. *Mehmet Emre Sargin, Hrishikesh Aradhye*. Boosting video classification using cross-video signals // Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference. 2011. P. 1805–1808.

25. *JaeDeok Lim*, *ByeongCheol Choi*, *SeungWan Han*, *ChoelHoon Lee*. Adult movie classification system based on multimodal approach with visual and auditory features // In Information Science and Digital Content Technology (ICIDT), 2012 8th International Conference on. 2012. Vol. 3. P. 745–748.

26. *Ionescu B.*, *Seyerlehner K.*, *Rasche Ch.*, *Vertan C.*, *Lambert P*. Video Genre Categorization and Representation using Audio-Visual Information. 2012. Vol. 21 (2).

27. *Ba Tu Truong, Svetha Venkatesh*. Video abstraction, A systematic review and classification // ACM Trans. Multimedia Comput. Commun. 2007. Vol. 3 (1).

28. Xu-*Dong Zhang*, *Tie-Yan Liu*, *Kwok-Tung Lo*, *Jian Feng*. Dynamic selection and effective compression of key frames for video abstraction // Pattern Recogn. Lett. 2003. Vol. 24 (9–10). P. 1523–1532.

29. *Kazunori Matsumoto*, *Masaki Naito*, *Keiichiro Hoashi*, *Fumiaki Sugaya*. Svm-based shot boundary detection with a novel feature // IEEE International Conference on Multimedia and Expo. 2006. P. 1837–1840.

30. *Ba Tu Truong*, *S. Venkatesh*, *C. Dorai*. Scene extraction in motion pictures // IEEE Trans. Cir. and Sys. for Video Technol. 2003. Vol. 13 (1). P. 5–15.

31. *H. Sundaram and Shih-Fu Chang*. Video scene segmentation using video and audio features // Multimedia and Expo, 2000. IEEE International Conference on. 2000. Vol. 2. P. 1145–1148.

32. *Liang-Hua Chen*, *Yu-Chun Lai*, *Hong-Yuan Mark Liao*. Movie scene segmentation using background information // Pattern Recogn. 2008. Vol. 41 (3). P. 1056–1065.

33. *Stephan Repp*, *Andreas Grob*, *Christoph Meinel*. Browsing within lecture videos based on the chain index of speech transcription // IEEE Transactions on Learning Technologies. 2008. Vol. 1 (3). P. 145–156.

34. *Rong Yan, Alexander G. Hauptmann*. A review of text and image retrieval approaches for broadcast news video // Inf. Retr. 2007. Vol. 10 (4–5). P. 445–484.

35. *John Adcock*, *Andreas Girgensohn*, *Matthew Cooper*, *Ting Liu*, *Lynn Wilcox*, *Eleanor Rieffel*. Fxpal experiments for trecvid 2004 // Proceedings of the TREC Video Retrieval Evaluation (TRECVID). 2004. P. 70–81.

36. *Hauptmann A. G.*, *Baron R. V.*, *Chen M. Y.*, *Christel M.*, *Duygulu P.*, *Huang C.*, *Jin R.*, *Lin W. H.*, *Ng D.*, *Moraveji N.*, *Papernick N.*, *Snoek C. G. M.*, *Tzanetakis G.*, *Yang J.*, *Yan R.*, *Wactlar H. D*. Informedia at trecvid 2003: Analyzing and searching broadcast news video // Proceedings of the TRECVID Workshop. 2003.

37. *Sivic J.*, *Everingham M.*, *Zisserman A.*. Person spotting: Video shot retrieval for face sets // In ACM International Conference on Image and Video Retrieval. 2005.

38. *Huiping Li, Doermann D*. Video indexing and retrieval based on recognized text // Multimedia Signal Processing, 2002 IEEE Workshop on. 2002. P. 245–248.

39. *Seyerlehner K.*, *Schedl M.*, *Pohle T.*, *Knees P*. Using blocklevel features for genre classification, tag classification and music similarity estimation. 2010.

40. *Chih-Wen Su*, *H.-Y.M. Liao*, *Hsiao-Rong Tyan*, *Chia-Wen Lin*, *Duan-Yu Chen*, *Kuo-Chin Fan*. Motion flow-based video retrieval // Multimedia, IEEE Transactions on. 2007. Vol. 9 (6). P. 1193–1201.

41. *Anjulan A., Canagarajah C. N*. A unified framework for object retrieval and mining // IEEE Transactions on Circuits and Systems for Video Technology. 2009. Vol. 19 (1). P. 63–76.

42. *Quack T.*, *Ferrari V.*, *Gool L.* Video mining with frequent item set configurations // Int. Conf. Image Video Retrieval. 2006. P. 360–369.

43. *Yu-Gang Jiang*, *Chong-Wah Ngo*, *Jun Yang*. Towards optimal bag-offeatures for object categorization and semantic video retrieval // In Proceedings of the 6th ACM International Conference on Image and Video Retrieval, CIVR '07. New York, NY, USA, 2007. P. 494–501.

44. *Peng Chang*, *Mei Han*, *Yihong Gong*. Extract highlights from baseball game video with hidden markov models. 2002. P. 609–612.

45. *Hong G. Y.*, *Fong B.*, *Fong A. C. M*. An intelligent video categorization engine // Kybernetes. 2005. Vol. 34 (6). P. 784–802.

46. *Linjun Yang*, *Jiemin Liu*, *Xiaokang Yang*, *Xian-Sheng Hua*. Multimodality web video categorization // Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval, MIR '07. New York, NY, USA, 2007. P. 265–274.

47. *Weal M. J.*, *Michaelides D. T.*, *Page K.*, *D. Roure C. De*, *Monger E.*, *Gobbi M.* Semantic annotation of ubiquitous learning environments // IEEE Transactions on Learning Technologies. 2012. Vol. 5 (2). P. 143–156.

48. *Yusuf Aytar*, *Mubarak Shah*, *Jiebo Luo*. Utilizing semantic word similarity measures for video retrieval // 2013 IEEE Conference on Computer Vision and Pattern Recognition. 2008. P. 1–8.

49. *Sivic J., Zisserman A*. Video google. Efficient visual search of videos // Toward Category-Level Object Recognition. 2006. P. 127–144.

50. *Browne P., Smeaton A. F*. Video retrieval using dialogue, keyframe similarity and video objects // ICIP. 2005. Vol. (3). P. 1208–1211.

51. *Cees G. M. Snoek*, *Bouke Huurnink*, *Laura Hollink*, *Maarten de Rijke*, *Guus Schreiber*, *Marcel Worring*. Adding semantics to detectors for video retrieval // IEEE Transactions on Multimedia. 2007. Vol. 9 (5). P. 975–986.

52. *Liang-Hua Chen*, *Kuo-Hao Chin*, *Hong-Yuan Liao*. An integrated approach to video retrieval // In Alan Fekete and Xuemin Lin, editors, Nineteenth Australasian Database Conference (ADC 2008). Wollongong, NSW, Australia, 2008. Vol. 75. P. 49–55.

53. *Kulesh V.*, *Petrushin V. A.*, *Sethi I. K*. The perseus project: Creating personalized multimedia news portal. O. R. Zaiane, S. J. Simoff (eds.). MDM/KDD. University of Alberta, 2001. P. 31–37.

54. *Kexue Dai*, *Jun Zhang*, *Guohui Li*. Video mining: concepts, approaches and applications // Multi-Media Modelling Conference Proceedings, 2006 12th International. 2006.

Материал поступил в редколлегию 02.09.2014

I. K. Nikitin

Moscow Aviation Institute
4 Volokolamskoe shosse, *4*, *Moscow*, *125993*, *Russian Federation*

w@w-495.ru

AN OVERVIEW OF COMPLEX CONTENT-BASED VIDEO RETRIEVAL METHODS

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.