МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ «ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ
по научно-исследовательской работе
ТЕМА: ИССЛЕДОВАНИЕ МЕТОДОВ СЕМАНТИЧЕСКОЙ РАЗМЕТКИ В ВИДЕОПОТОКЕ

Студент гр. 3304 _____ Бенитез Х.

Руководитель _____ Кринкин К.В

Санкт-Петербург
2018

ЗАДАНИЕ
НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Студент Бенитез Х.

Группа 3304

Тема НИР: Исследование методов семантической разметки в видеопотоке

Задание на НИР: Анализ алгоритмов: его описание, проблемы, связанные с анализом распределенных данных. Обзор соответствующей работы по применению распределенных данных.

Сроки выполнения НИР: 25.10.2018 – 20.12.2018
Дата сдачи отчета: 20.12.2018
Дата защиты отчета: 20.12.2018

Студент(ка) _____ Бенитез Х.

Руководитель _____ Кринкин К.В

АННОТАЦИЯ

Анализировать существующие алгоритмы и подходы к решению проблемы. Описание алгоритмов и проблем, возникающих при работе с поиском видео. Приведите список решений, основанных на алгоритмах группировки. Результаты проведенной работы дают краткий анализ того, что было достигнуто в результате осуществления исследований и разработок.

SUMMARY

Analyze existing algorithms and approaches to solve the problem. Description of algorithms and problems that arise when working with video searches. Bring a list of solutions based on grouping algorithms. The findings of the work carried out provide a brief analysis of what has been achieved as a result of the implementation of research and development.

СОДЕРЖАНИЕ

Оглавление

ВВЕДЕНИЕ

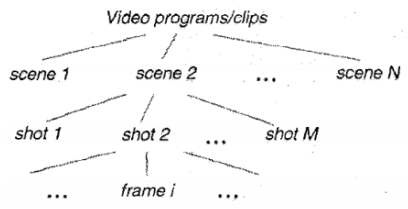
Индексирование и поиск видео на основе контента (CBVIR), в приложении проблемы поиска изображений, то есть проблемы поиска цифрового видео в больших базах данных. «На основе контента» означает, что поиск будет анализировать фактическое содержание видео. Термин «контент» в этом контексте может относиться к цветам, формам, текстурам. Без возможности просмотра видеоконтента поиск должен основываться на изображениях, предоставленных пользователем.

Хотя термин «поисковая система» часто используется без разбора для описания поисковых систем на основе искателя, каталогов, созданных человеком, и всего, что между ними, они не все одинаковы. Каждый тип «поисковой системы» собирает и ранжирует списки совершенно разными способами.

Поисковые системы, такие как Google, автоматически формируют свои списки. Они «ползают» или «паукают» в Интернете, и люди ищут по их спискам. Эти списки составляют индекс или каталог поисковой системы. Индекс можно представить как массивный электронный шкаф для хранения документов, содержащий копию каждой веб-страницы, найденной пауком. Поскольку пауки регулярно ищут информацию в Интернете, любые изменения, внесенные в веб-сайт, могут повлиять на рейтинг в поисковых системах. Также важно помнить, что добавление страницы-указателя в индекс может занять некоторое время. Пока это не произойдет, он не будет доступен для тех, кто ищет в поисковой системе [8].

Каталоги, такие как Open Directory, зависят от редакторов-людей для составления своих списков. Веб-мастера отправляют адрес, название и краткое описание своего сайта, а затем редакторы просматривают материалы. Однако гибридные поисковые системы обычно предпочитают один тип листинга другому.

Сегментация видео является первым шагом к поиску контента на основе контента с целью сегментирования движущихся объектов в видеопоследовательностях. Сегментация видео первоначально сегментирует первый кадр изображения как кадр изображения на несколько движущихся объектов, а затем отслеживает эволюцию движущихся объектов в последующих кадрах изображения. После сегментирования объектов в каждом кадре изображения эти сегментированные объекты имеют множество приложений, таких как наблюдение, манипулирование объектами, композиция сцены и извлечение видео [10]. Видео создается с помощью набора снимков и их составления с использованием указанных операторов композиции. Извлечение структурных примитивов является задачей сегментации видео, которая включает в себя обнаружение временных границ между сценами и между кадрами, как показано на рисунке 1.



Первым этапом анализа видеоконтента, просмотра и поиска видео на основе контента является разбиение видеопоследовательности на кадры. Снимок определяется как последовательность изображений, представляющая непрерывное действие, которое снимается с одной операции одной камеры. Снимки объединяются на этапе редактирования видеопродукции для формирования полной последовательности. Снимки можно эффективно рассматривать как наименьшую единицу индексации, при которой не может быть воспринято никаких изменений в содержании сцены, а концепции более высокого уровня часто создаются путем объединения и анализа взаимосвязей между кадрами и между кадрами.

Ключевые кадры - это неподвижные изображения, извлеченные из исходных видеоданных, которые наилучшим образом представляют содержание снимков абстрактно. Ключевые кадры часто использовались в качестве дополнения к тексту видео-журнала, хотя в прошлом они выбирались вручную. Ключевые кадры, при правильном извлечении, представляют собой очень эффективный визуальный конспект содержимого видео и очень полезны для быстрого просмотра видео. , Сводка видео, такая как предварительный просмотр фильма, представляет собой набор выбранных сегментов из длинной видеопрограммы, которые выделяют видеоконтент, и она лучше всего подходит для последовательного просмотра длинных видеопрограмм. Помимо просмотра, ключевые кадры также могут быть использованы при представлении видео в поисковом видео. Индекс может быть построен на основе визуальных особенностей ключевых кадров, и запросы могут быть направлены на ключевые кадры с использованием запросов с помощью алгоритмов поиска.

Как только ключевые кадры извлечены, следующим шагом является извлечение функций. Функции обычно извлекаются в автономном режиме, поэтому эффективные вычисления не являются существенной проблемой, но для больших коллекций все еще требуется больше времени для вычисления функций. Функции видеоконтента можно классифицировать на функции низкого и высокого уровня.

Низкоуровневые функции, такие как движение объекта, цвет, форма, текстура, громкость, спектр мощности, ширина полосы и высота тона, извлекаются непосредственно из видео в базе данных. Функции на этом уровне объективно получены из средств массовой информации, а не ссылаются на какую-либо внешнюю семантику. Функции, извлеченные на этом уровне, могут отвечать на запросы, такие как «поиск изображений с более чем 20% -ным распределением синего и зеленого цвета», что может привести к получению нескольких изображений с голубым небом и зеленой травой. Многие эффективные подходы к низкоуровневому извлечению признаков были разработаны или для различных целей.

Функции высокого уровня также называют семантическими функциями. Такие особенности, как тембр, ритм, инструменты и события, включают в себя различные степени семантики, содержащиеся в медиа. Предполагается, что высокоуровневые функции имеют дело с семантическими запросами (например, «поиск картины воды» или «поиск улыбки Моны Лизы»). Последний запрос содержит семантику более высокой степени, чем первый. Поскольку вода на изображениях отображает однородную текстуру, представленную в низкоуровневых объектах, такой запрос легче обрабатывать. Чтобы

получить последний запрос, поисковая система требует предварительных знаний, которые могут определить, что Мона Лиза - женщина, которая является конкретным персонажем, а не какой-либо другой женщиной в картине.

Трудность обработки запросов высокого уровня возникает из-за внешних знаний с описанием функций низкого уровня, известных как семантический разрыв. Процесс поиска требует механизма перевода, который может преобразовать запрос «Улыбка Моны Лизы» в низкоуровневые функции. Было предложено два возможных решения для минимизации семантического разрыва. Первый - это автоматическая генерация метаданных на носитель. Автоматическая аннотация по-прежнему включает семантическую концепцию и требует различных схем для различных сред. Второй использует обратную связь по релевантности, чтобы позволить поисковой системе изучать и понимать семантический контекст операции запроса. Соответствие обратной связи обсуждается в разделе-III.

Многие мультимедийные базы данных содержат большое количество функций, которые используются для анализа и запроса базы данных. Такой набор векторов признаков рассматривается как высокая размерность. Например, Tieu&Viola использовала более 10000 функций изображений, каждое из которых описывает локальный шаблон. Высокая размерность вызывает проблему «проклятия измерения», когда сложность и вычислительные затраты запроса экспоненциально возрастают с увеличением количества измерений. Уменьшение размеров - это популярный метод решения этой проблемы и поддержки эффективного поиска в крупных базах данных. Тем не менее, существует компромисс между эффективностью, полученной посредством измерения сокращения, и полнотой, полученной посредством извлеченной информации. Если каждая информация представлена меньшим числом измерений, скорость поиска увеличивается. Тем не менее, некоторая информация может быть потеряна. Одним из наиболее широко используемых методов поиска мультимедиа является анализ основных компонентов (PCA). PCA используется для преобразования исходных данных высокой размерности в новую систему координат с низкой размерностью путем поиска данных с высокой способностью различать. Новая система координат удаляет избыточные данные, и новый набор данных может лучше представлять важную информацию.

Поисковая система обычно содержит два механизма: измерение сходства и многомерное индексирование. Измерение сходства используется для поиска наиболее похожих объектов. Многомерная индексация используется для ускорения выполнения запросов в процессе поиска.

Чтобы измерить сходство, общий подход состоит в том, чтобы представить элементы данных в виде многомерных точек, а затем вычислить расстояния между соответствующими многомерными точками. Выбор метрик оказывает непосредственное влияние на производительность поисковой системы. Евклидово расстояние является наиболее распространенной метрикой, используемой для измерения расстояния между двумя точками в многомерном пространстве. Однако для некоторых применений евклидово расстояние не совместимо с воспринимаемым человеком сходством. Ряд методов был предложен для конкретных целей и представлен в разделе IV. Остальная часть статьи организована следующим образом: в разделе 2 рассматриваются различные

используемые алгоритмы и подходы, в разделе 3 мы объясняем различные функции, которые можно использовать для индексации видео. В разделе 4 рассматриваются различные подходы к измерению сходства, в разделе 5 мы представляем проблемы исследования с системами поиска видео на основе контента, и, наконец, мы завершаем в разделе 6.

2. ИСТОРИЯ

Несмотря на многие исследовательские усилия, существующие низкоуровневые функции все еще недостаточно мощны для представления содержимого фрейма индекса. Некоторые функции могут достигать относительно хорошей производительности, но их размеры обычно слишком велики, или реализация алгоритма затруднена [3]. Извлечение функций является очень важным шагом в поисковой системе для описания видео с минимальным количеством дескрипторов. Основные визуальные особенности рамки указателя включают цвет и текстуру [4]. В настоящее время исследования в области поиска видео на основе контента представляют собой живую дисциплину, расширяющуюся по ширине [5]. Представительные функции, извлеченные из индексных кадров, хранятся в базе данных объектов и используются для поиска видео на основе объектов [6]. Текстура является еще одним важным свойством индексных кадров. Различные представления текстуры были исследованы в распознавании образов и компьютерном зрении. Методы представления текстуры можно разделить на две категории: структурные и статистические. Структурные методы, включая морфологический оператор и граф смежности, описывают текстуру, идентифицируя структурные примитивы и правила их размещения. Они имеют тенденцию быть наиболее эффективными, когда применяются к текстурам, которые очень регулярны. Статистические методы, включая спектры мощности Фурье, матрицы совпадений, анализ главных компонент (SPCA), не зависящий от сдвига, особенность Тамуры, разложение по Вольду, случайное поле Маркова, фрактальная модель и методы фильтрации с высоким разрешением, такие как вейвлет-преобразование Габора и Хаара, охарактеризовать текстуру по статистическому распределению интенсивности изображения [1].

Поиск видео на основе текстовых запросов [30] представил подход, который позволяет осуществлять поиск на основе текстовой информации, представленной в видео. Области текстовой информации выделены в рамках видео. Затем видео снабжается текстовым контентом, представленным на изображениях. Автоматический поиск на основе контента и семантическая классификация видеоконтента [31] представили учебную среду, в которой построение видео высокого уровня визуализируется путем синтеза его набора элементарных функций. Это делается с помощью опорных векторных машин (SVM). Машины опорных векторов связывают каждый набор точек данных в многомерном пространстве признаков с одним из классов во время обучения. В основанном на контенте телевизионном спортивном извлечении видео на основе аудиовизуальных функций и текста [32] авторы предлагают основанное на контенте извлечение видео, которое является своего рода поиском по его семантическому содержимому. Поскольку видео состоит из мультимодальных информационных потоков, таких как визуальные, слуховые и текстовые потоки, авторы описывают стратегию использования мультимодального анализа для автоматического анализа спортивного видео. Вначале в документе определяется базовая структура системы базы данных спортивного видео, а затем

вводится новый подход, который объединяет анализ визуальных потоков, распознавание речи, обработку речевого сигнала и извлечение текста для осуществления поиска видео. Экспериментальные результаты телевизионного спортивного видео о футбольных играх показывают, что мультимодальный анализ эффективен для поиска видео благодаря быстрому просмотру древовидных видеоклипов или вводу ключевых слов в предварительно определенном домене.

Видео-поиск почти дубликатов с использованием k-NearestNeighborRetrieval пространственно-временных дескрипторов [33] описывает новую методологию для реализации функций поиска видео, таких как поиск почти дублированных видео и распознавание действий в видео наблюдения. Видео делятся на полсекундные клипы, чьи уложенные кадры производят объемные объемные пиксели в 3D. Пиксельные области с постоянными свойствами цвета и движения извлекаются из этих трехмерных объемов с помощью беспороговой иерархической пространственно-временной сегментации. Затем каждая область описывается многомерной точкой, компоненты которой представляют положение, ориентацию и, по возможности, цвет области. На этапе индексации для видеобазы данных этим точкам назначаются метки, которые указывают их исходный видеоклип. Все отмеченные точки для всех клипов сохраняются в одном двоичном дереве для эффективного поиска k-ближайшего соседа. Фаза поиска использует видео сегменты в качестве запросов. Работа, представленная в разделе «Быстрый поиск видео через статистику движения в интересующих регионах» [34], имеет дело с очень важной проблемой для быстрого извлечения семантической информации из обширной мультимедийной базы данных. В этой работе авторы предлагают алгоритм на основе статистики для извлечения видео, которые содержат запрошенный объект, из базы данных видео.

Чтобы ускорить алгоритм, авторы используют только локальное движение, встроенное в интересующую область, в качестве запроса для извлечения данных из потоков битов MPEG. В траектории на основе извлечения видео с использованием моделей процесса Dirichlet [35], введена основанная на траектории система извлечения видео с использованием моделей процесса Dirichlet. Основной вклад этой структуры в четыре раза. (1) Применяется к модели смеси процессов Дирихле (DPMM) для обучения траектории без надзора. DPMM - это бесконечно смешанная модель смеси, компоненты которой растут сами собой. (2) Использовать чувствительную ко времени модель смеси процесса Дирихле (tDPMM), чтобы узнать характеристики временных рядов траекторий. Кроме того, новый алгоритм оценки вероятности для tDPMM используется впервые. (3) основанная на tDPMM схема сопоставления вероятностных моделей, которая, как эмпирически показано, является более устойчивой к ошибкам и способной обеспечить более высокую точность извлечения, чем аналогичные методы в литературе. (4) Структура имеет хорошую масштабируемость и адаптивность в том смысле, что при представлении новых данных кластера платформа автоматически идентифицирует новую информацию кластера без необходимости повторного обучения. Теоретический анализ и экспериментальные оценки с использованием самых современных методов демонстрируют перспективность и эффективность системы. Аннотация к видео для поиска на основе контента с использованием анализа поведения человека и знания предметной области [39] представляет метод автоматической аннотации спортивного

видео для поиска на основе контента. Этот подход включает в себя анализ поведения человека и знание конкретной области с традиционными методами, чтобы разработать интегрированный модуль рассуждений для большей выразительности событий и надежного распознавания.

Как показано в Таблице I, нашей целью является не исчерпывающий обзор или полнота исследований в этой области, а качественное обсуждение подходов по мере их развития в этой области, чтобы представить наш взгляд на эволюцию этой области исследований.

ТАБЛИЦА I: Сравнение алгоритмов, используемых для извлечения признаков и их поиска приложения

Алгоритм	Подход	Используемые функции	Поисковое приложение
Классификация цветовой текстуры	Сегментируйте изображения по регионам	Цвет, Текстура Особенности	Важность восприятия
Мультимодальный контент на основе Просмотр и поиск методы	Peer2Peer поиск системы	Извлечение ключевого кадра, особенность формы	Основанный на объекте
Идентификация персонажа	Поиск объекта фильма	сегментация	Основанный на объекте
Семантически значимый Сводки	Разбор видео последовательности для выявления соответствующих видов камеры и отслеживания движений мяча	Основанные на сценариях, Со вхождения Матрицы	Спортивное видео
Семантическое видео поиск	Автоматическая аудио категоризация	аудио	музыка, речь и т. д.
Распознавание объекта в видеопоследовательностях	Методы последовательной классификации для BLOB-объектов	Фильтр Калмана применяется над BLOB-объектов	Объект на основе изображения
Автоматическое сопоставление сцен	Соответствие изображения	Вейвлеты, цветовые дескрипторы	3D сцена в фильме
Содержание на основе поиска лица	Соответствие изображения	самоорганизующиеся карты (SOM) и отзывы пользователей Визуальная кластеризация	Распознавание лица

3. ВИДЕО ИНДЕКСИРОВАНИЕ И ПОЛУЧЕНИЕ

Индексирование видео - это процесс маркировки видео и их эффективной организации для быстрого доступа и поиска. Автоматизация индексации может значительно снизить стоимость обработки при одновременном устранении утомительной работы [4]. Обычные функции, используемые в большинстве существующих систем поиска видео, включают такие функции, как цвет, текстура, форма, движение, объект, лицо, звук, жанр и т. Д. Очевидно, что чем больше функций используется для представления данных, тем лучше точность поиска. Однако, поскольку размер вектора объектов увеличивается с увеличением количества объектов, существует компромисс между точностью поиска и сложностью. Поэтому важно иметь минимальные возможности, представляющие видео, компактно. В этой статье мы обсуждаем индексирование ключевых кадров видео, текстуру, цвет, форму, аудио и т. Д., Используемые для индексации и поиска.

3.1. Особенности текстуры

Текстура может быть определена как визуальные образцы, которые имеют свойства однородности, которые не являются результатом присутствия только одного цвета или интенсивности.

Тамура и др. (1978) предложили метод извлечения и описания признаков текстуры, основанный на психологических исследованиях человеческого восприятия. Метод состоит из шести статистических характеристик, включая грубость, контрастность, направленность, сходство линий, регулярность и шероховатость, для описания различных свойств текстуры.

Серая матрица совпадений (GLC) - один из самых элементарных и важных методов извлечения и описания признаков текстуры. Его оригинальная идея впервые была предложена в Julesz (1975). В своих знаменитых экспериментах по визуальному восприятию текстуры человеком Юлез обнаружил, что для большого класса текстур нельзя различить пару текстур, если они согласуются в своей статистике второго порядка. Квантованный индексный кадр с матрицей GLC показан на фиг.2.

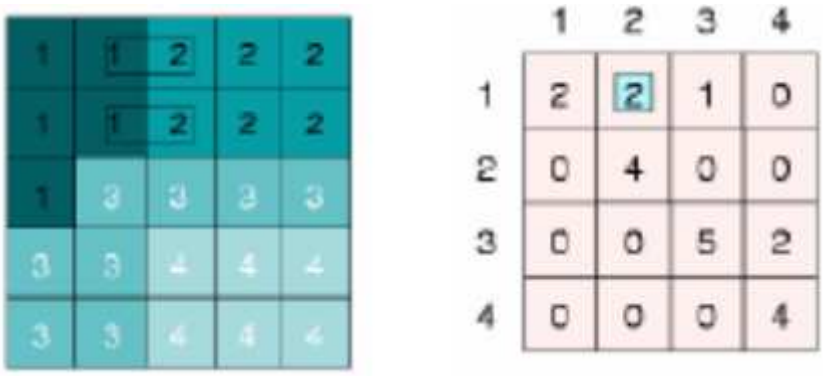


Рисунок 2. Квантованная слава слева с GLC справа

Следующие текстуры кадра индекса извлекаются из матрицы GLC.

Энтропия $-\sum_i \sum_j C(i, j) \log C(i, j)$ =

Энергия $\sum_i \sum_j C(i,j)^2$ =

Однородность $-\sum_i \sum_j C(i,j) / (1 + |i+j|)$ =

Контрастность измеряет уровень серого q; q = 0, 1, ..., qmax, изменяются на изображении g и в какой степени их распределение смещено к черному или белому. Центральные моменты второго порядка и нормированные четвертого порядка гистограммы уровня серого (эмпирическое распределение вероятностей), то есть дисперсия σ2, и эксцесс, α4, используются для определения контраста

$$F_{con} = \sigma / \alpha_4^n$$ где,

$$\alpha_4 = \frac{\mu_4}{\sigma^4}; \quad \sigma^2 = \sum_{q=0}^{q_{max}} (q-m)^2 \Pr(q|\mathbf{g}); \quad \mu_4 = \sum_{q=0}^{q_{max}} (q-m)^4 \Pr(q|\mathbf{g})$$

3.2 Разложение на основе Wold и особенности текстуры Gabor

Если текстура моделируется как выборка из стационарного двумерного случайного поля, разложение Вольда также можно использовать для поиска на основе подобия. В модели Уолда пространственно однородное случайное поле разлагается на три взаимно ортогональных компонента, которые приблизительно представляют периодичность, направленность и чисто случайную часть поля.

Двумерная функция Габора γ (x, y) и ее преобразование Фурье Г (u, v) имеют следующий вид

$$\gamma(x,y) = \frac{1}{2\pi\delta x \delta y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\delta x^2} + \frac{y^2}{\delta xy^2}\right) + 2\pi\sqrt{-1}Wx\right]$$

$$\tau(u,v) = \exp\left[-\frac{1}{2}\left(\frac{(u-W)^2}{\delta u^2}\right) + \frac{v^2}{\delta v^2}\right]$$

где σu = 1 / 2πσx и σv = 1 / 2πσy. Функция Габора является продуктом эллиптической волны Гаусса и волны комплексной плоскости и минимизирует совместную двумерную неопределенность как в пространственной, так и в частотной области. Соответствующие расширения и повороты этой функции дают класс самоподобных фильтров Габора для ориентации и масштабируемой настройки краев и линий. Фильтры образуют полный, но неортогональный базис для расширения изображения и получения его локализованного

описания пространственной частоты. Общее количество фильтров Габора равно произведению чисел шкал и ориентаций.

3.3. Особенности цвета

Цвет является одной из наиболее широко используемых визуальных функций в мультимедийном контексте и, в частности, для поиска изображений и видео. Для поддержки связи через Интернет данные должны хорошо сжиматься и подходить для гетерогенной среды с различными пользовательскими платформами и устройствами просмотра, большим разбросом мощности компьютера пользователя и изменяющимися условиями просмотра. Системы CBIR обычно не знают о разнице в исходных, кодированных и воспринимаемых цветах, например, о различиях между колориметрическими данными и данными о цвете устройства.

3.3.1. Цветовые дескрипторы

Цветовые дескрипторы изображений и видео могут быть глобальными и локальными. Глобальные дескрипторы определяют общее цветовое содержание изображения, но без информации о пространственном распределении этих цветов. Локальные дескрипторы относятся к конкретным областям изображения и, в сочетании с геометрическими свойствами этих последних, описывают также пространственное расположение цветов. В частности, дескрипторы цвета MPEG-7 состоят из ряда дескрипторов гистограммы, дескриптора доминирующего цвета и дескриптора цветовой компоновки (CLD).

3.3.2. Цветные гистограммы

Цветовая гистограмма h (изображение) = (hk (изображение) k = 1, ..., K) представляет собой K-мерный вектор, так что каждый компонент hk (изображение) представляет относительное количество пикселей цвета Ck в изображении, то есть доля пикселей, которые наиболее похожи на соответствующий цвет. Для построения цветовой гистограммы цвета изображения должны быть преобразованы в соответствующее цветовое пространство и количественно определены в соответствии с конкретной кодовой книгой размера K.

3.3.3. Цветовая коррелограмма

Цветовая коррелограмма изображения - это таблица, индексированная по парам цветов, где k-я запись для (i, j) указывает вероятность нахождения пикселя цвета на расстоянии от пикселя цвета в изображении. Такая особенность изображения оказывается устойчивой к большим изменениям внешнего вида одной и той же сцены, вызванным изменениями в положениях просмотра, изменениями в фоновой сцене, частичными окклюзиями, масштабированием камеры, вызывающим радикальные изменения в форме, и т. Д. пространственная корреляция цветов, эффективная и недорогая для контентного поиска изображений. Коррелограмма устойчиво переносит большие изменения внешнего вида и формы, вызванные изменениями положения просмотра, увеличения камеры и т. Д. Цветовая коррелограмма не является ни методом разделения изображений, ни методом уточнения гистограммы. В отличие от чисто локальных свойств, таких как положение пикселей, направление градиента, или чисто глобальных свойств, таких как распределение

цвета, коррелограммы учитывают локальную цветовую пространственную корреляцию, а также глобальное распределение этой пространственной корреляции.

3.4. Семантические особенности высокого уровня

Семантический разрыв относится к разнице между ограниченной описательной силой низкоуровневых индексных фреймов и богатством пользовательской семантики. Для поддержки запросов высокоуровневыми концепциями система должна обеспечивать полную поддержку в преодолении этого «семантического разрыва» между характеристиками фрейма числового индекса и богатством человеческой семантики. В этом обзоре мы рассмотрели методы, используемые для сокращения семантического разрыва на пять категорий, которые наиболее широко используются:

3.4.1. Использование онтологий объектов для определения понятий высокого уровня:

Для баз данных со специально собранными изображениями простая семантика, полученная на основе объектной онтологии, может работать нормально, но с большой коллекцией изображений требуются более мощные инструменты для изучения семантики.

3.4.2. Использование инструментов машинного обучения для связи низкоуровневых функций с концепциями запросов:

Упоминаются следующие методы: машинное обучение, байесовская классификация, нейронные сети и т. Д. Недостатками этих методов является то, что они требуют большой коллекции базы данных изображений для изучения данных.

3.4.3. Введение обратной связи по релевантности (RF) в цикл поиска для непрерывного изучения намерений пользователей:

Большинство современных RF-систем используют только функции ключевых кадров низкого уровня для оценки идеальных параметров запроса и не учитывают «семантическое» содержание индексного кадра.

3.4.4. Создание семантического шаблона (ST) для поддержки поиска изображений высокого уровня:

Этот метод повышает точность поиска по сравнению с традиционными методами с использованием цветовой гистограммы и особенностей текстуры.

3.4.5. Использование как визуального содержимого ключевых кадров, так и текстовой информации, полученной из Интернета для поиска изображений в Интернете (Интернет):

Преимущество заключается в том, что некоторая дополнительная информация поможет в семантическом поиске изображений. Следующая таблица 2 суммирует различные методы, связанные с семантическими признаками.

No.	Техника	Описание	Преимущество	Недостаток
1	Онтологии объектов для определения понятий высокого уровня	Семантика может быть получена с использованием низкоуровневых функций изображения и используется для запросов высокого уровня. Различные методы, такие как присвоение имен цветам, используются для описания изображений.	Проще классифицировать изображения по их характеристикам, таким как цвет и т. Д.	Для баз данных со специально собранными изображениями простая семантика, полученная на основе объектной онтологии, может работать нормально, но с большой коллекцией изображений требуются более мощные инструменты для изучения семантики.
2	Инструменты машинного обучения для связи низкоуровневых функций с концепциями запросов	Используя двоичный байесовский классификатор, концепции высокого уровня природных сцен извлекаются из классифицированных на общие типы как внутренние / наружные, а наружные изображения далее классифицируются по городским / ландшафтным и т. Д.	Результат здесь основан на наборе входных показателей и с использованием методов искусственного интеллекта, они могут быть более точными и помочь уменьшить семантический разрыв	Они требуют большой коллекции базы данных фреймов iindex для изучения данных. Обычные алгоритмы обучения страдают от двух проблем: (1) требуется большое количество помеченных обучающих образцов, и предоставление таких данных очень утомительно и подвержено ошибкам; (2) обучающий набор фиксируется на этапах обучения и применения.
3	релевантная обратная связь (RF)	Система предоставляет	RF это онлайн-обработка	Большинство современных

		начальные результаты поиска посредством запроса по примеру, эскиза и т. Д. Пользователь оценивает вышеуказанные результаты относительно того, являются ли они и в какой степени они релевантными (положительные примеры) или не относящимися (отрицательные примеры) к запросу. Алгоритм машинного обучения применяется для изучения		RF-систем используют только функции низкоуровневых индексных фреймов для оценки идеальных параметров запроса и не учитывают «семантическое» содержание индексных фреймов.
4	Генерация семантического шаблона (ST) для поддержки поиска видео высокого уровня	Пользователь сначала определяет шаблон для конкретной концепции, указав объекты и их пространственные и временные ограничения, веса, назначенные каждому элементу каждого объекта. Этот начальный сценарий запроса предоставляется системе. Благодаря взаимодействию с пользователями система в конечном итоге сводится к небольшому набору примерных запросов, которые «наилучшим образом» соответствуют (максимизируют отзыв) концепции в уме пользователя.	Этот метод повышает точность поиска по сравнению с традиционными методами с использованием цветовой гистограммы и особенностей текстуры. Это наиболее перспективный для поиска видео.	Не используется широко
5	Использование как визуального содержимого	Данные, относящиеся к индексным фреймам,	Некоторая дополнительная информация в	Точность получения недостаточна,

	индексных фреймов, так и текстовой информации, полученной из Интернета для поиска в Интернете (в Интернете).	используются для их описания. URL-адрес файла фрейма индекса часто имеет четкую иерархическую структуру, включая некоторую информацию о фрейме индекса, такую как категория фрейма индекса.	Интернете доступна для облегчения семантического поиска фрейма индекса.	поскольку они не могут подтвердить, действительно ли в полученных индексных фреймах содержатся концепции запроса. В результате пользователи должны пройти весь список, чтобы найти нужное видео. Это трудоемкий процесс, поскольку возвращаемые результаты всегда содержат несколько тем, которые смешаны вместе.

3.5. Аудио Особенности

Следующие аудио функции также используются для индексации видео и поиска видео.

3.5.1. Кратковременная энергия:

Основное использование этой функции - для отделения речи от неречевых сегментов в аудиосигнале. Это очень полезно в шумной обстановке, потому что шумовые сигналы имеют среднюю кратковременную энергию ниже, чем обычная речь. Средняя кратковременная энергия m образцов может быть выражена с использованием следующего члена:

$$E_m = \frac{1}{N} \sum_{n=0}^{N-1} (x(n)w(m-n))^2$$

3.5.2. Подача:

Высота тона (основная частота - F0) является очень важной функцией для анализа звука, особенно для обнаружения подчеркнутой человеческой речи. Он представляет ведущую

частоту сложного аудиосигнала. В речи высота тона повышается в результате возбуждения динамика. Мы решили использовать автокорреляционную функцию в нашем подходе к оценке основного тона. Автокорреляционная функция для случайного сигнала определяется как

$$A(k) = \frac{1}{2N+1} \sum_{n=-N}^{N} x(n)x(n+k)$$

Из приведенного выше уравнения мы можем вычислить пиковые значения для разных значений k. Шаг для конкретного окна определяется как местоположение наибольшего пикового значения (max (A (k))) в выбранном окне, если функция автокорреляции выше определенного порога (0,3 в нашей работе). Если последнее ограничение выполнено, мы можем рассчитать шаг как:

$$P = \frac{f}{k_{\max}} \quad [Hz]$$

3.5.3.Мелкочастотные кепстральные коэффициенты:

Это тип функций фонемного уровня для характеристики аудиосигналов. Он также основан на разделении поддиапазонов всего частотного спектра. MFCC основаны на шкале Мел. Шкала Mel постепенно искажает линейный спектр с более грубым разрешением на более высоком и более точным разрешением на низких частотах. Он направлен на предоставление компактного представления огибающей спектра аудиосигнала. Функции, подобные MFCC, для MPEG-1 и AAC на основе преобразования вычисляются на длинных симметричных окнах для оценки долгосрочной статистики, тогда как для 8xMDCT на основе преобразования они рассчитываются на покадровой основе, где вектор функции вычисляется для каждого рамка из 8192 образцов.

3.5.4. Скорость паузы:

Функция паузы предназначена для определения количества речи в аудиоклипе. Частота паузы может использоваться как показатель подчеркнутой человеческой речи. Это можно легко рассчитать, посчитав количество тихих звуковых кадров в аудиоклипе. Если мы обозначим количество тихих сегментов в аудиоклипе u как u S, мы можем написать:

$$S_u = \sum_{j=0}^{k} \begin{cases} 1 & \text{if silent audio frame} \\ 0 & \text{otherwise} \end{cases}$$

Затем мы можем усреднить это для каждого аудиоклипа с количеством аудиокадров в клипе (| u |) и получить частоту пауз (Pr):

$$P_r = \frac{S_u}{|u|}$$

3.5.5 Обнаружение начала:

Функции обнаружения начала представляют собой представления среднего уровня, которые направлены на локализацию переходных процессов в аудиосигнале. Используется эталонная функция обнаружения начала, основанная на окне анализа Хеннинга с длиной выборки 2048, которое дает временное разрешение 23,2 мс при 44,1 кГц. Функция обнаружения начала для MPEG-1 и AAC одинакова, а временное разрешение составляет 576 (13 мс) и 1024 (23,2 мс) выборок при 44,1 кГц. Эта функция проще 8xMDCT и дает разрешение по времени относительно опорной функции с 512 образцами.

3.5.6. Chromagram:

Профиль хроматограммы или класса основного тона (PCP) обычно определяется как 12-мерный вектор, где каждое измерение соответствует интенсивности класса полутона (цветности). Опорная хроматограмма основана на постоянном Q-преобразовании. Алгоритм, использованный автором для вычисления вычислений на основе преобразований, одинаков для трех кодеков. Входом в алгоритм является лучшее разрешение по частоте кодеков. Очевидно, что результаты для кодека 8xMDCT лучше, чем у двух других кодеков, поскольку они имеют самое высокое разрешение по частоте.

3.5.7. Латентное восприятие

Весь аудиоклип представлен в виде одного вектора в скрытом пространстве восприятия (LPS). Это делает вычислительно интенсивную меру подобия на основе сигналов управляемой. Этот метод также выявляет скрытую структуру восприятия аудиоклипов и на основе этого измеряет сходство. Аудио база данных разделена на 20 взаимно разных категорий (таких как самолет, промышленность, толпа, строительство). Категории выводятся с использованием доступных текстовых подписей.

Из аудиоклипа можно выделить векторный вектор. Затем он характеризуется вычислением количества векторов признаков, которые квантуются в каждый из эталонных кластеров признаков сигналов. Существует разреженная матрица, в которой каждая строка представляет количественную характеристику полного клипа в терминах эталонных кластеров. Эталонные кластеры получают путем неконтролируемой кластеризации всей коллекции объектов, извлеченных из клипов в базе данных. Посредством разложения по сингулярным числам (SVD) это разреженное представление отображается на точки в LPS. Таким образом, каждый аудиоклип представлен как один вектор в пространстве восприятия.

3.6. Другие особенности и представление

3.6.1. Особенности формы

Статистический подход распознавания образов был распространен в течение многих лет для распознавания формы. Набор измерений, которые независимо характеризуют некоторый аспект формы. Большая коллекция примеров, характеризующих форму статистически. Предположим, например, что миссия состоит в том, чтобы различать акул и жало лучей. Измерения могут включать в себя свойства области, такие как площадь, периметр, соотношение сторон, собственные значения, выпуклое расхождение и различные центральные моменты. Хотя вычисление таких особенностей может быть сложным, мы не обсуждаем фактический вычислительный процесс здесь, а скорее отсылаем читателя к текстам по вычислительной геометрии. Простой алгоритм выращивания областей состоит в том, чтобы сегментировать черно-белое изображение по областям.

3.6.2. Представление Key-Object

Ключевые кадры обеспечивают подходящую абстракцию и структуру для просмотра видео. Ключевой объект определяется как меньшие единицы в ключевом кадре. «ключевые объекты» используются для представления ключевых областей, которые участвуют в различных действиях в кадре. Это чрезвычайно сложная задача анализа, потому что ключевые объекты не обязательно соответствуют семантическим объектам, мы можем избежать проблемы анализа семантических объектов путем поиска областей когерентного движения. Когерентность движения может охватывать некоторые аспекты объектов, желательных для поиска. К ключевым объектам можно прикрепить несколько атрибутов, включая цвет, текстуру, форму, движение и жизненный цикл. Атрибуты цвета и текстуры вычисляются с помощью алгоритмов.

Каждый снимок представлен одним или несколькими ключевыми кадрами, которые далее разлагаются на ключевые объекты. В дескрипторе движения движения предоставляет информацию о вероятных действиях в кадре. Он также фиксирует общие движения в кадре, такие как глобальные движения, возникающие при панорамировании или масштабировании камеры. Например, дескриптор движения может использоваться для различения «шатких» последовательностей, снятых ручной камерой, от профессионально снятых последовательностей.

Кроме того, есть некоторые преимущества в разложении движения ключевого объекта на глобальный компонент и локальный / объектный компонент. При разложении движение ключевого объекта может быть более легко использовано для отражения движения относительно фона и других ключевых объектов в сцене. Без этого различия движение ключевого объекта вместо этого представляло бы движение относительно рамки изображения. Таким образом, эта декомпозиция обеспечивает более содержательное и эффективное описание для поиска.

3.6.3. Функция масштабного инвариантного преобразования объектов (SIFT)

Извлечение видео также можно выполнить с помощью функции SIFT. Видео делится на кадры, а кадры делятся на изображения. Объект отделен от изображения сегментацией изображения. Сегментированный объект является частью изображения. Функция извлекается из сегментированного изображения. В этом методе объекты извлекаются с помощью преобразования масштабируемого инвариантного объекта (SIFT) и используются для поиска ключевых точек на изображениях, поскольку они инвариантны к изображению.

В этом методе сначала видео конвертируется в изображения. Эти изображения сегментируются с использованием алгоритма сегментации, чтобы получить изображение объекта. Элементы извлекаются из изображения объекта с использованием алгоритма SIFT. Сопоставление объектов затем выполняется на объектах базы данных с помощью Mahalanobis Distance для извлечения видео из базы данных.

Видеокадр, называемый статическим изображением, является основной единицей видеоданных. Последовательность кадров определяется как набор интервалов кадров, где интервал кадров [i, j] представляет собой последовательность видеокадров от кадра i до j. Отдельные кадры разделены линиями кадров. Первым шагом является сегментирование видео на элементарные кадры, каждый из которых содержит непрерывные по времени и пространству. Видеопоток состоит из кадров, кадров, сцен и эпизодов. Физически связанные последовательности кадров генерируют видеокадры. Затем объедините связанные сцены в последовательности.

Сегментация - это совокупность методов, позволяющих интерпретировать пространственно близкие части изображения как объекты. Используется для поиска объектов и границ на изображении. Сегментация текстуры, используемая в качестве описания областей на сегменты, два ее типа основаны на областях и границах. Разделы на основе областей или группы областей в соответствии с общими свойствами изображения, такими как значения интенсивности из исходных изображений, текстуры или шаблоны, являются уникальными для каждого типа области, спектральные профили, которые предоставляют данные многомерного изображения. Методы на основе границ, используемые для поиска явных или неявных границ между регионами, соответствующими различным типам проблем.

Функции SIFT (масштабного инвариантного преобразования объектов), используемые при распознавании объектов, имеют очень большой размер. Они инвариантны к изменениям масштаба, 2D-трансляции и преобразованиям вращения. Большие вычислительные затраты, связанные с сопоставлением всех функций SIFT для задач распознавания, ограничивают его применение проблемами распознавания объектов.

4. ИЗМЕРЕНИЕ ПОДОБИЯ

Измерение сходства играет важную роль в поиске. Фрейм запроса передается системе, которая извлекает похожие видео из базы данных. Метрика расстояния может быть названа мерой сходства, которая является ключевым компонентом в извлечении видео на основе контента. При обычном поиске евклидовы расстояния между базой данных и

запросом вычисляются и используются для ранжирования. Фрейм запроса больше похож на фрейм базы данных, если расстояние меньше. Если x и y - двухмерные векторы объектов фрейма индекса базы данных и фрейма запроса соответственно. Следующая таблица 3 обобщает популярные методы измерения расстояний.

факторы название	**Формула**	**Рамка изображения / указателя**	**Время вычислений**	**семья**
евклиды расстояние	$D_1 = \sqrt{\sum_i (x_i - y_i)^2}$	Цветная, Черно-белая	Меньше	Minkowski family
в квадрате Аккорд	$d_{sc}(x,y) = \sum_{i=1}^{d} (\sqrt{x_i} - \sqrt{y}$	цвет	Средний	Squaredchord family
Хи-квадрат	$D_7 = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$	Образ текстура	Сокращение времени вычислений, но высокая стоимость вычислений	Squared L2 family or $\chi 2$ family
Манхеттен	$D_2 = \sum_i \lvert x_i - y_i \rvert$	Черный цвет и белый	Требуется меньше вычислений	Minkowski family
дивергенция	$\delta_{rs} = \sum_{i=1}^{d} \frac{(x_{ri} - x_{si})}{(\lvert x_{ri} \rvert + \lvert x$	Медицинские изображения	Больше	Squared L2 family or $\chi 2$ family
Волна хеджирования	$\delta_{rs} = \sum_{i=1}^{d} \frac{(x_{ri} - x_{si}}{(\lvert x_{ri} \rvert + \lvert x$	Цветная, Черно-белая	Больше	Intersection family

5. ИССЛЕДОВАТЕЛЬСКИЕ ПРОБЛЕМЫ

Ниже приведены меры исследования проблемы, связанные с индексацией и поиском видео.

5.1. Query Language Design:

Язык запросов основан на понятии «семантические показатели», а синтаксис фиксирует основные закономерности восприятия человеком семантических категорий. По сравнению с другими методами сокращения «семантического разрыва» язык запросов относительно плохо понят и заслуживает большего внимания.

5.2. Многомерная индексация индексных фреймов:

Поскольку размер базы данных индексных фреймов быстро увеличивается, важным фактором, который следует учитывать, является относительная скорость. Необходимы автономные данные многомерного индексного фрейма.

5.3. Стандартная СУБД, расширенная для поиска видео:

Создание видео-поиска в качестве подключаемого модуля в существующей СУБД также обеспечит естественную интеграцию с функциями, полученными из других источников. Интегрированная система CBIR потребует интеграции основанного на контенте сходства, взаимодействия с пользователями, визуализации базы данных видео, управления базами данных для поиска релевантных видео и т. Д.

5.4. Стандартный испытательный стенд и модель оценки производительности:

Стандартная база данных индексного фрейма с набором запросов и соответствующей моделью измерения производительности крайне необходима для объективной оценки производительности систем CBVR.

6. ВЫВОДЫ

Несмотря на значительный прогресс в академических исследованиях в области поиска видео, результаты поиска видео по поиску контента на коммерческих приложениях были относительно небольшими.

некоторые исключения ниши, такие как сегментация видео. Выбор функций, отражающих реальные человеческие интересы, остается открытым вопросом. Одним из многообещающих подходов является использование мета-обучения для автоматического выбора или объединения соответствующих функций. Другая возможность заключается в разработке интерактивного пользовательского интерфейса, основанного на визуальной интерпретации данных с использованием выбранной меры, чтобы помочь процессу выбора. Обширные эксперименты, сравнивающие результаты функций с фактическим интересом человека, могут быть использованы в качестве другого метода анализа. Поскольку взаимодействия с пользователем необходимы для определения характеристик, желательно разработать новые теории, методы и инструменты для облегчения вовлечения пользователя.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. G. Utz Westermann and Ramesh Jain,(2007)" Towards a Common Event Model for Multimedia Applications", in IEEE Multimedia.

2. M. Lew, N. Sebe, C Djerba, and R. Jain (2006), "Content-based Multimedia Information Retrieval: State of the Art and Challenges", ACM TOMCAPP vol.2, No. 1, pp. 1-19.

3. Milind Naphade , John R. Smith , Jelena Tesic , Shih-Fu Chang , Winston Hsu , Lyndon Kennedy , Alexander Hauptmann , Jon Curtis (2006), "Large-Scale Concept Ontology for Multimedia," IEEE Multimedia, April 2006.

4. Keiji Yanai, Kobus Barnard (2006), "Finding Visual Concepts by Web Image Mining", in proc. Of WWW 2006, Edinburgh, Scotland.

5. Michael S. Lew, N. Sebe, C. Djeraba, R. Jain (2006), "Content-Based Multimedia Information Retrieval: State of the Art and Challenges", ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No. 1, February 2006, Pp. 1–19.

6. Tuomas Aura, Thomas A. Kuhn, Micheal Roe (2006), "Scanning electronic documents for personally identifiable information", WEPS, USA, ACM.

7. R. Varadarajan, V. hristidis (2006), "A system for query-specific document summarization", CIKM, USA, ACM, June 2006.Adam Jatowt, Mitsuru Ishizuka, "Temporal multi-page summarization", Web Intelligence and Agent System, Volume 4 Issue 2, IOS Press.

8. B. V. Patel, B. B. Meshram(2006), "Mining and clustering images to improve image search engines for geo-informatics database", In the proceedings of National Conference on Geoinformatics, VPM Polytechnic, Mumbai, Dec-2006.

9. Ryutarou Ohbuchi, Jun Kobayashi (2006)," Unsupervised learning from a Corpus for Shape-Based 3D Model Retrieval", MIR'06, October 26–27, 2006, Santa Barbara, California, USA.

10. Michael S. L., Nicu Sebe, Chabane Djeraba, Ramesh Jain(2006), "Content-Based Multimedia Information Retrieval: State of the Art and Challenges", A

11. R. Dufour, Y. Estève, P. Deléglise, and F. Béchet (2009), "Local and global models for spontaneous speech segment detection andcharacterization," in ASRU 2009, Merano, Italy.

12. Tristan Glatard, Johan Montagnat, Isabelle E. Magnin (2004)," Texture Based Medical Image Indexing and Retrieval: Application to Cardiac Imaging", MIR'04, October 15–16, 2004, New York, New York.

13. Egon L. van den Broek, Peter M. F. Kisters, and Louis G. Vuurpijl (2004)," Design Guidelines for a Content-Based Image Retrieval Color-Selection Interface"ACM Dutch Directions in HCI, Amsterdam.

14. Tristan Glatard, Johan Montagnat, Isabelle E. Magnin (2004), "Texture Based Medical Image Indexing and Retrieval: Application to Cardiac Imaging", ACM Proc. Of MIR'04, October 15–16, 2004, New York, New York, USA.