

ПОИСК ОБЪЕКТОВ В ХРАНИЛИЩАХ ВИДЕОДАНЫХ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ПАРАЛЛЕЛЬНОЙ СУБД И СОПРОЦЕССОРОВ INTEL XEON PHI

М.С. Миниахметова, М.Л. Цымблер

Южно-Уральский государственный университет (Челябинск)

Аннотация

В настоящее время аналитическая обработка видеоданных используется для решения многочисленных научных и практических задач. Объемы накапливаемых видеоданных в современном обществе возрастают экспоненциально, поэтому решение задач интеллектуального анализа видео требует разработки новых методов и алгоритмов хранения и обработки видео, эффективно использующих современные многопроцессорные и многоядерные вычислительные системы. В данной статье предлагается подход к организации видеохранилища, поддерживающего функцию поиска объектов на видео. В рамках данного подхода используются параллельная СУБД и многоядерные сопроцессоры. Параллельная СУБД хранит метаданные видео в форме реляционных таблиц, горизонтальные фрагменты которых распределяются по узлам кластерной системы и обрабатываются параллельно. Многоядерные сопроцессоры используются для извлечения метаданных из видео, загружаемого в хранилище. Для реализации запросов на поиск объектов в видео предлагается расширение SQL новыми типами данных Video и VideoObject и соответствующими конструкциями.

Введение

В настоящее время аналитическая обработка видеоданных используется для решения многочисленных научных и практических задач из широкого спектра предметных областей: безопасность, киноиндустрия, новостные службы, медицина, образование и др. Поиск объектов на видео является одной из актуальных задач интеллектуального анализа видеоданных [1]. Под объектом в данном случае понимается сущность, с которой ассоциируется набор изображений (например, с объектом «автомашина» могут быть связаны набор изображений городских автомобилей, гоночных болидов и детских игрушек-автомобильчиков). Целью поиска является нахождение видеок кадров, в которых фигурирует объект, *похожий* на множество ассоциированных изображений. В условиях современных видеохранилищ, включающих в себя сотни тысяч видеосюжетов с общим объемом в сотни терабайт, поиск объектов на видео представляет собой вычислительно сложную задачу. В соответствии с этим решение задачи поиска объектов в хранилищах видеоданных требует разработки новых методов и алгоритмов хранения и обработки видео, эффективно использующих современные многопроцессорные и многоядерные вычислительные системы.

В данной статье предлагается подход к организации видеохранилища, поддерживающего функцию поиска объектов на видео на основе использования параллельной СУБД и многоядерных сопроцессоров Intel Xeon Phi. Текст статьи организован следующим образом. В статье представлен обзор существующих подходов к организации хранилищ видеоданных и поиска объектов на видео, далее приведено описание алгоритмов, используемых для решения задачи поиска объектов, после чего представлено описание подхода, предлагаемого авторами данного исследования. В заключении резюмируются полученные результаты и рассматриваются направления дальнейших исследований.

Обзор подходов к организации хранилищ видеоданных и поиску объектов на видео

Одной из наиболее сложных задач обработки графической информации на сегодняшний день считается поиск объектов на изображении [2]. Для ее решения используются методы извлечения дескрипторов особых точек и алгоритмы классификации. Видео при этом может рассматриваться, как множество последовательно сменяющихся изображений. Таким образом, при стандартной частоте смены кадров 24 кадра в секунду, поиск объектов на видео может занимать продолжительное время. Если же рассматривать эту задачу применительно к массивам видеоданных, хранящимся в мультимедийной базе данных, то становится очевидно, что ее решение требует специальных архитектурных решений и новых подходов.

На сегодняшний день имеются сведения о множестве исследований, касающихся тех или иных аспектов баз данных мультимедиа, а также семантического поиска по видеоданным.

Одними из первых вопросы хранения и семантического поиска по видеоданным затронули авторы статьи [3], в которой предлагают подход, основанный на возможности фрагментации видеофайлов с выделением ключевых кадров, под которыми авторы понимают кадры, позволяющие определить некоторое действие без необходимости просмотра промежуточных кадров. Видео, содержание которого удовлетворяет этим требованиям, авторы называют "фрагментируемым" (quantizable video). В качестве примера приводится следующая ситуация: баскетбольный игрок Новитски передает нижний пас игроку Кидд. Здесь начальный и

конечный кадры будут ключевыми, поскольку содержат достаточно информации для определения совершенного действия — передачи паса одного игрока другому. На основе использования такого подхода предлагается расширение языка SQL, предназначенного для выполнения семантических пространственно-временных запросов к мультимедийной базе данных. Как указывают сами авторы, их система, основанная на реляционной модели данных и названная ими Grammar-Based Semantic Modelling and Retrieval Tool (G-SMART), будет наиболее эффективна для видео спортивного содержания.

В статье [4] рассматривается разработка системы управления базами данных видео (VDBMS) на основе реляционной модели данных. Для эффективного контекстного поиска авторы статьи вводят новые типы данных: video data type (VDT) и stream data type (SDT). В работе предлагается представление видео файла в виде иерархии сцен, фрагментов и ключевых кадров. Система предусматривает SQL-подобный язык запросов. Запросы к данным типа VDT предусматривают использование проиндексированной мета-информации, которая извлекается в ходе препроцессинга при добавлении видео в базу данных. К мета-информации авторы относят цветные гистограммы, текстуры и ориентацию границ каждого кадра. Эти данные затем индексируются и полученные индексы используются для выполнения запросов на поиск видео по кадру и поиск ближайших соседей. Поскольку такие запросы являются очень ресурсоемкими в виду большого объема обрабатываемых данных, авторы вводят оператор StopAfter, предусматривающий прерывание выполнения запроса после нахождения заданного числа значений, удовлетворяющих условиям запроса. Тип данных SDT применяется для задач трекинга объектов на видео, а также в тех случаях, когда объем свободного места в хранилище небольшой, и авторы считают пространственно невыгодным хранение кортежей с метаданными каждого кадра.

Исследование авторов статьи [5] посвящено разработке подхода к созданию базы данных видео с использованием механизмов и типов данных, предоставляемых пакетом расширения Oracle interMedia для СУБД Oracle. Данный пакет использует объектную модель данных для хранения и обработки видео. Центральное место авторы статьи отводят разработке механизма потоковой загрузки видео при воспроизведении файла из базы данных, запрашиваемого пользователем. Поиск по базе данных в организован с использованием ключевых слов и тегов, которые вводятся вручную.

В статье [6] предлагается подход к эффективному хранению и обработке видео в реляционной базе данных, основанный на фрагментации видео на отрезки, сцены и ключевые кадры. Для повышения точности фрагментации используется также анализ звуковой дорожки. Для организации эффективного поиска используются новые типы данных и алгоритмы кластеризации и классификации. Кроме того, разработанная авторами система реализует функцию поиска человеческих лиц на видео.

Кроме того, существует ряд исследований, направленных на разработку систем хранения и обработки видеоданных с использованием облачных сервисов и распределенных хранилищ. Тем не менее, ни в одном из рассмотренных исследований нами не было встречено упоминания об использовании параллельной реляционной СУБД, параллельных алгоритмов обработки видеоданных и использовании многоядерных ускорителей.

Алгоритмы анализа изображений и поиска объектов

Поиск объекта на изображении начинается с вычисления каких-либо характеристик изображения, которые называют *дескриптором* данного изображения. Из наиболее широко известных видов дескрипторов и алгоритмов их вычисления можно отметить:

- Алгоритмы, основанные на признаках Хаара. Среди них наибольшее распространение получил метод Виолы-Джонса [7]. Изначально метод предназначался для поиска на изображении человеческих лиц, но дальнейшие исследования показали его эффективность для распознавания различных классов объектов.
- Алгоритмы, основанные на гистограммах. Наиболее известным алгоритмом из данного класса в настоящее время является алгоритм вычисления гистограмм направленных градиентов (Histogram of Oriented Gradients, HOG). Метод был создан и впервые обнародован исследователями INRIA Навниитом Далалом и Биллом Триггсом в 2005 г. [2] В этой работе авторы использовали его для поиска пешеходов, но впоследствии область его применения также была расширена, и в настоящее время данный метод успешно используется и для других классов объектов.

Независимо от того, какими характеристиками оперирует алгоритм обработки изображения, для его применения к задаче распознавания образов необходимо использование алгоритмов классификации. Таким образом, первым шагом решения задачи поиска объектов является создание обучающей выборки изображений, часть из которых содержит искомый объект, а другая часть содержит элементы фона и шум. Затем выполняется вычисление дескрипторов всех изображений, составляющих обучающую выборку, а результаты вычислений используются для обучения алгоритма классификации. На последнем шаге решения задачи распознавания объектов последовательно вычисляются дескрипторы отдельных областей интересующего изображения и проверяются полученным классификатором на предмет соответствия характеристикам искомого класса.

Использование параллельной СУБД и сопроцессоров Xeon Phi для поиска объектов на видео

Для организации эффективного расширяемого хранилища сверхбольших объемов видеоданных нами было решено использовать параллельную реляционную СУБД PargreSQL, созданную в ходе научного

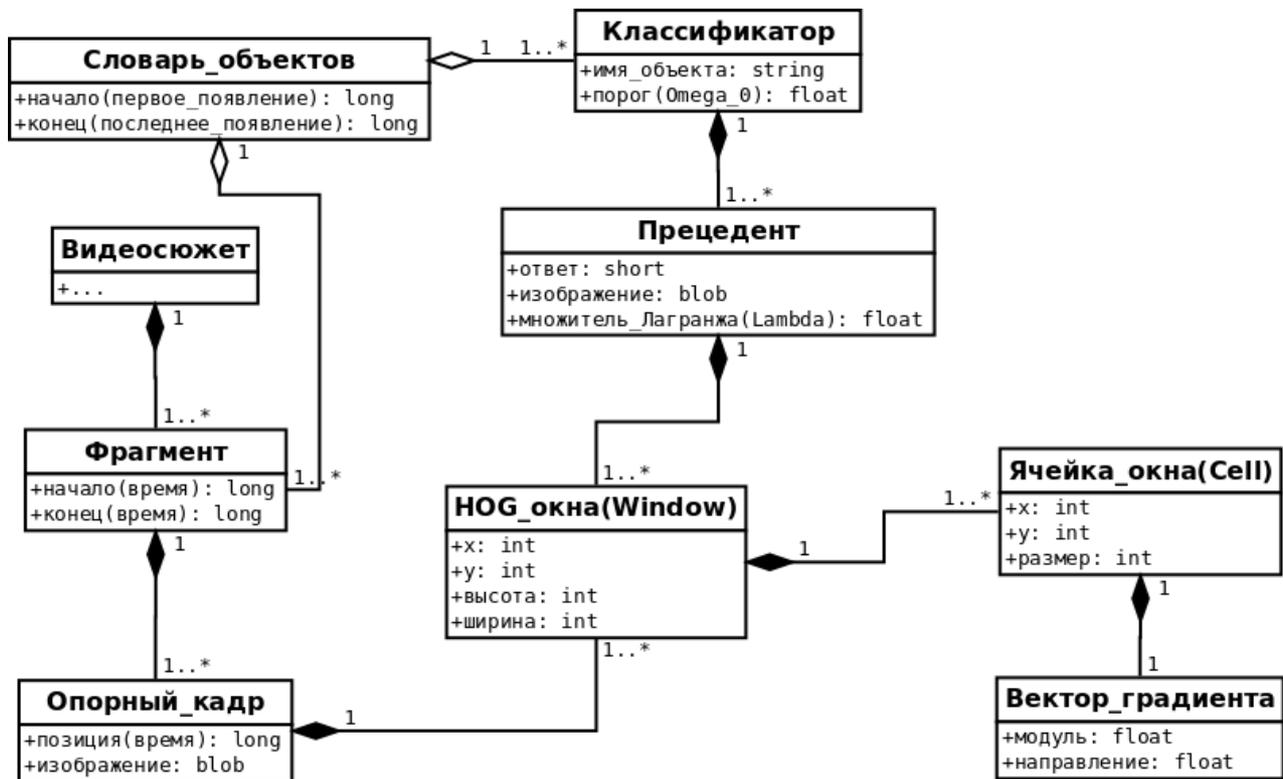


Рис. 3. Диаграмма классов системы

На рис. 3 представлена модель разрабатываемой системы.

Для организации обработки данных в спроектированной базе данных, а также интеграции в нее поддержки запросов на поиск объектов на хранимом в базе данных видео предлагается создать новый тип данных Video, инкапсулирующий данные и методы для обработки видео. Кроме того, необходимо расширить стандартный синтаксис языка запросов SQL, дополнив структуру запросов DDL и DML новыми ключевыми словами и функциями для написания запросов к типу данных Video. В качестве примера такого запроса можно служить следующий запрос на поиск временных интервалов видео, на которых появляется объект «Автомобиль» (Car), во всем видео, хранящимся в базе данных:

```
SELECT Name, TimeRange FROM Videos WHERE Occures(Car)
```

При выполнении подобных запросов требуется вычислять и анализировать дескрипторы ключевых кадров фрагментов видео. При большом размере и разнородности даже одного видеофайла это эта операция может занимать значительное количество времени, поэтому для повышения скорости обработки дескрипторы ключевых кадров должны быть вычислены в ходе выбора ключевых кадров во время препроцессинга. В данном исследовании будут использованы дескрипторы HOG, предложенные Далалом и Триггсом. При этом для ускорения обработки каждого видеофайла алгоритм вычислений гистограмм направлений градиентов будет адаптирован для параллельного выполнения на сопроцессорах Intel Xeon Phi с применением концепции параллелизма по данным.

Для определения принадлежности дескриптора анализируемого кадра классу искомого объекта будет использован алгоритм классификации SVM (Support Vector Machine, метод опорных векторов). Для минимизации числа промахов классификации для данного алгоритма требуется большой объем обучающей выборки, состоящий из порядка тысячи положительных прецедентов и в несколько раз большего количества негативных прецедентов. Очевидно, что обучение алгоритма классификации является ресурсоемкой и длительной операцией, поэтому для ее выполнения также необходимо использовать параллельный алгоритм вычисления дескрипторов изображений, предложенный выше. Для хранения и использования полученных в результате обучения параметров алгоритма классификации в синтаксис SQL будет добавлен тип данных VideoObject.

Так как ключевой кадр репрезентативно отражает содержимое фрагмента, то ожидается, что для нахождения объекта на данном фрагменте достаточно определить факт присутствия этого объекта на его ключевых кадрах. Тем не менее, остается ненулевая вероятность, что интересующий объект не попал ни на один из ключевых кадров. В этом случае проектируемое расширение SQL предполагает возможность использования в запросе ключевого слова DEEP для выполнения «глубокого» поиска, т.е. поиска объектов напрямую в видеофайле без использования служебных таблиц. В этом случае представленный выше запрос примет следующий вид:

SELECT DEEP Name, TimeRange FROM Videos WHERE Occures(Car)

Поскольку обработка видеофайла является дорогостоящей и длительной операцией, для ее существенного ускорения в данном исследовании предлагается заранее выполнять поиск всех имеющихся объектов на видео, хранящихся в базе данных, и хранить результаты вычислений в индексируемых служебных таблицах, называемых *словарем объектов*. Словарь объектов будет содержать информацию о наличии объекта на фрагментах видео, а также временных интервалах, на которых данный объект присутствует. Предполагается несколько способов заполнения словаря объектов:

- по команде администратора базы данных;
- при выполнении «глубокого» поиска;
- в моменты простоя вычислительного узла, содержащего реплику видеофайла, не имеющего индекса для данного объекта.

Таким образом, выполнение запроса к базе данных на поиск видео по представленному на нем объекту обрабатывается в несколько этапов, каждый из которых выполняется только в том случае, если предыдущий шаг не дал результатов:

1. Поиск информации об искомом объекте для данного видео в словаре объектов. При наличии такой информации, соответствующие данные из словаря мгновенно добавляются к ответу на запрос.
2. Поиск объекта на ключевых кадрах. При этом используются предварительно вычисленные дескрипторы изображений соответствующих кадров и параметры алгоритма классификации, соответствующие искомому объекту. То есть, поиск объекта в этом случае сводится к выполнению запросов на выборку данных из соответствующих служебных таблиц базы данных.
3. «Глубокий» поиск по видеофайлу. В этом случае используются алгоритм вычисления дескрипторов для каждого кадра видео и алгоритм классификации, адаптированные для параллельного выполнения на сопроцессорах Intel Xeon Phi с использованием концепции параллелизма по данным.

Заключение

В данной статье описан подход к организации видеохранилища, поддерживающего функцию поиска объектов на видео. В рамках данного подхода используются параллельная СУБД и многоядерные сопроцессоры. Параллельная СУБД хранит метаданные видео в форме реляционных таблиц, горизонтальные фрагменты которых распределяются по узлам кластерной системы и обрабатываются параллельно. Многоядерные сопроцессоры используются для извлечения метаданных из видео, загружаемого в хранилище. Для реализации запросов на поиск объектов в виде предлагается расширение SQL новым типом данных VideoObject и соответствующими конструкциями.

Дальнейшие исследования будут направлены на реализацию описанного подхода на основе параллельной СУБД PargreSQL [8, 9].

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект № 12-07-00443-а) и Фонда содействия развитию малых форм предприятий в научно-технической сфере (проект № 0000856ГУ1/2013).

ЛИТЕРАТУРА:

1. V. Vijayakumar, R. Nedunchezian A study on video data mining // International Journal of Multimedia Information Retrieval. Vol. 1, Issue 3, 2012, Springer-Verlag. P. 153-172.
2. N. Dalal, B. Triggs Histograms of Oriented Gradients for Human Detection // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 25 June, 2005. IEEE Computer Society, 2005. P. 886-893.
3. V. Jain, R. Aygun Spatio-Temporal Querying of Video Content Using SQL for Quantizable Video Databases // Journal of multimedia. Vol. 4, Issue 4, 2009. Academy Publisher. P. 215-227.
4. D.S. Kushwaha, A. Kushwaha Application of Video Database Management System (Vdbms) For Video Query Processing // International Journal Of Engineering And Computer Science. Vol. 2, Issue 11, 2013. P. 3278-3281.
5. S. Greca, D. Xhaja, E. Gani Multimedia Database Systems and Oracle Video Streaming // 1st International Symposium on Computing in Informatics and Mathematics (ISCIM 2011) Proceseefings. Tirana, Albania, 2-4 June, 2011. P. 158-166.
6. S. Zhu, Y. Liu Video scene segmentation and semantic representation using a novel scheme // Multimediz Tools snd Applications. Vol. 42, April 2009. Kluwer Academic Publishers. P. 183-205.
7. P. Viola, M.J. Jones Robust Real-time Object Detection // Cambridge Research Laboratory: Technical Reports Series. Vol. 1, 2001. P. 1-24.
8. Пан К.С., Соколинский Л.Б., Цымблер М.Л. Интеграция параллелизма в СУБД с открытым кодом // Открытые системы. СУБД. 2013. № 9. С. 56-58.
9. Pan C.S., Zymbler M.L. Taming Elephants, or How to Embed Parallelism into PostgreSQL // Database and Expert Systems Applications - 24th International Conference, DEXA 2013, Prague, Czech Republic, August 26-29, 2013. Proceedings, Part I. Springer, 2013. Lecture Notes in Computer Science. Vol. 8055. P. 153-164.