

11. Базы данных для видео

11.1. Представление видеоданных

11.2. Сжатие видео

11.3. Организация поиска по видеоданным

Представление видеоданных

- Видео – упорядоченная по времени последовательность связанных изображений (называемых фреймами или кадрами)
- Непосредственное влияние телевизионных технологий:
 - SECAM (Россия, Франция)
 - PAL (большинство европейских стран, Китай)
 - NTSC (США, Япония)
- 25-30 фреймов/сек
- Чересстрочная развертка (четные/нечетные строки) для уменьшения мерцания
- Размеры фреймов: 352 x 240, 768 x 576 (PAL), 720 x 576 (CCIR 601), 720 x 480 (NTSC), 1440 x 1152, 1920 x 1080 (HDTV)
- Формат фрейма (ширина к высоте): 4:3, 16:9 (широкоэкранный)
- Цветное видео: разложение на яркостную и две цветовые компоненты (модель YUV)
- Типичная частота дискретизации: 720 значений яркости на строку и 360+360 значений цветности на строку; YUV 4:2:2 (Y:Cb:Cr - яркость:голубая компонента цвета:красная комп.) – два пикселя хранятся в виде Y1:Y2:Cb:Cr

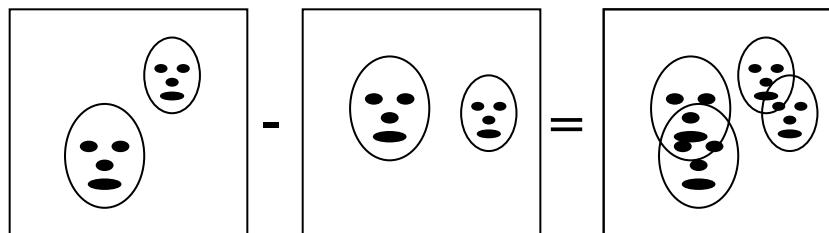
Сжатие видео

- «Очевидный» способ сжатия – сжимать последовательность фреймов (каждый из которых – обычное изображение) один за другим (□ Motion-JPEG; используется в веб-камерах)
- Недостаток: не учитывается, что соседние фреймы очень похожи
- Дефекты (искажения), заметные человеческому глазу, в обычных (неподвижных) изображениях отличаются от дефектов в видео
- Различные методы в зависимости от области применения
- Важные параметры:
 - Скорость упаковки/распаковки
 - Ошибкоустойчивость
- Большинство стандартов основано на дискретном косинус-преобразовании (DCT)
- Типичные коэффициенты сжатия: от 50:1 до 100:1; сжатое видео почти неотличимо от оригинала

Сжатие видео

Компенсация движения [1]:

- Если соседние фреймы мало отличаются друг от друга имеет смысл сжимать попиксельную разность двух последовательных фреймов (большая часть площади такой разницы близка к нулю и, значит, хорошо сожмется)
- Попиксельной разности недостаточно для видеоряда с подвижными объектами
- Рассматривается разность текущего фрейма и скомпенсированного фрейма (скомпенсированный - соседний, уже сжатый фрейм, в котором часть объектов (в идеале - все) перемещены и/или трансформированы так, чтобы этот измененный фрейм был как можно более близок в некоторой метрике к текущему кодируемому фрейму)
- Скомпенсированный фрейм = соседний фрейм (как правило предыдущий) плюс информация о движении
- Иллюстративный пример попиксельной разности последовательных фреймов без компенсации движения:



Сжатие видео

- Текущий кадр разбивается на непересекающиеся блоки одного размера $B(x,y)$ (например, 8 на 8 пикселей)
- Для каждого блока $B(x,y)$ в небольшой окрестности ищется наиболее «похожий» на него блок $BPrev(x+u,y+v)$ в предыдущем кадре; «похожесть» определяется выбранной метрикой
- Вектор $d=(u,v)^T$, на котором достигается минимум выбранной функции ошибки, считается вектором смещения (движения) для данного блока



Рисунок из [1]

Сжатие видео

Области применения:

а) Видеоконференции и видеотелефоны:

- Кодер и декодер должны быть одинаково быстрыми («симметричная» компрессия)
- Стандарт ITU-T H.621 [2]; уже устарел – текущий H.624; большинство систем видеоконференций используют проприетарные форматы
- Для увеличения скорости:
 - Во время компенсации движения расстояние между предсказываемым и блоком из предыдущего фрейма не более 15 пикселей (по вертикали и по горизонтали)
 - Единица обработки: макроблок - 16 на 16 значений яркости плюс два массива по 8 на 8 значений цветности
- Двухмерный пространственный фильтр: для сглаживания острых углов в предсказываемом блоке
- Преобразование: применение DCT к 8x8 блокам с разностью между предсказанным и предыдущим блоками
- Преобразованные блоки квантуются, кодируются и отправляются получателю

Сжатие видео

- Отправитель также должен декодировать фрейм и временно его сохранить (для декодирования следующего фрейма)
- Затруднение: высокая активность (много движения) в видео генерирует много данных, низкая активность много меньшее количество
- Решение (в H.621): 32 различных квантизатора – грубое квантование для высокоактивных блоков, детальное для низкоактивных блоков
- Получатель уведомляется об используемых квантизаторах: для каждого макроблока или, на практике, для группы макроблоков, в пределах которых квантизатор может быть заменен для проблематичных блоков
- Контроль скорости передачи:
Буфер передачи не должен быть ни пустым ни полным;
Скорость передачи можно менять изменением квантизаторов

Сжатие видео

б) Асимметричные приложения:

- Мультимедийные архивы, CD- и DVD-диски
- Однократное сжатие и многократная распаковка после
- Алгоритмы кодирования могут быть очень сложными
- Распаковка в режиме реального времени

MPEG (Moving Picture Experts Group)

- Набор стандартов для сжатия аудио и видео
- С 1988 года; основные шаги:
 - MPEG-1: первоначальный стандарт для сжатия аудио- и видеоданных
 - MPEG-2: цифровое телевидение и DVD-форматы
 - MPEG-3: планировался для телевидения высокой четкости; работы над ним были прекращены, т.к. MPEG-2 с незначительными модификациями может достигать тех же результатов
 - MPEG-4: с 1998 года; объектно-базированное кодирование; поддержка аудио/видео объектов, различных видов интерактивности

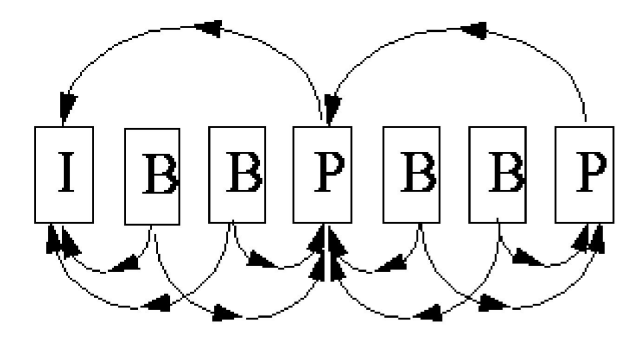
Сжатие видео

Возможность произвольного доступа:

- Цифровое вещание и доступ к видео из видеохранилищ:
Должна быть возможность начать просмотр с (почти) любого фрейма
- Решение в MPEG – три типа фреймов:
 - I-фреймы (intra): закодированные как неподвижные изображения, без ссылок на последующие или предыдущие фреймы; дают наименьшую нагрузку при сжатии; используются как стартовые
 - P-фреймы (predicted): предсказанные (компенсация движения) на основе последнего (ближайшего) I- или P-фрейма
 - B-фреймы (bi-directional): предсказанные из двух ближайших I- или P-фреймов, одного предыдущего и другого – последующего; дают наибольшую нагрузку при сжатии; ошибки не распространяются
- GOP (Group of Pictures): наименьшая единица доступа (независимо декодируемая); может состоять только из I-фреймов, I-фреймов и P-фреймов, или содержать все три типа; обычно GOP включает 12 или 15 фреймов и начинается с I-фрейма, например IBBPBVPBVPBVPBV

Сжатие видео

Пример:



- Два порядка фреймов:
 - Порядок показа
 - Порядок в потоке (P-фрейм должен быть раньше B-фрейма)
- Для конвертации из порядка в потоке в искомый порядок показа необходима буферизация
- Присутствует небольшая задержка
- Замечание: предсказываемый фрейм и фрейм для предсказания не обязательно соседние
Расстояние между ними влияет на размер области поиска похожих блоков при компенсации движения

Сжатие видео

Квантование:

- Как в H.261 могут использоваться различные квантизаторы

Контроль скорости передачи:

- К В-фреймам может применяться более грубое квантование (В-фреймы не используются в предсказании других фреймов, ошибка не распространяется)
- Для отдельных фреймов:
 - Увеличить/уменьшить шаг квантизатора
 - Добавить/убрать DCT-коэффициенты

MPEG-1:

- Сравнимое с VHS (видеокассеты) качество для видеопоследовательностей со «средним» или «медленным» движением
- Худшее в сравнении с VHS качество для видеопоследовательностей с «быстрым» движением

MPEG-2:

- Универсальный, независимый от приложений стандарт
- Большое количество опций

Сжатие видео

- Профили MPEG-2:
 - Простой (SP): нет В-фреймов
 - Основной (SP)
 - Масштабируемый (SNR): по отношению сигнал-шум
 - Пространственно-масштабируемый (SP)
 - Высший (HP)

В последних трех профилях может быть более одного потока видеоданных:

- Основной поток: низкоскоростное кодирование
 - Другие потоки: улучшение качества
- Уровни MPEG-2 (в зависимости от размера фрейма): low, main, high 1440, high
 - Поддержка чересстрочного видео
 - Два режима предсказания (макроблок, половина макроблока)

Сжатие видео

в) Пакетная передача видео:

- Передача сжатого видео по сетям передачи данных
- Данные рассылаются в пакетах:
 - Различные пакеты могут следовать по различным маршрутам
 - Пакеты могут иметь приоритеты
 - Пакеты могут теряться
- В случае перегрузок сети схема сжатия должна обладать возможностью менять скорость передачи данных (также учитывать возможные искажения при передаче)
- Несколькоуровневая обработка (аналогично постепенному отображению графики):
 - Низкоскоростной, высокоприоритетный уровень
 - Низкоприоритетные уровни для повышения качества
- Предлагаемая техника:
 - Многополосное кодирование (с помощью временных и пространственных фильтров)
 - DCT-базируемое кодирование; например, разбиение наборов коэффициентов на уровни (низко- ... высокочастотные); аналогично профилям MPEG-2

Организация поиска по видеоданным

Вопросы:

- Какие аспекты видео представляют наибольший интерес?
- Как лучше всего организовать и хранить эти аспекты?
- Какие языки запросов лучше использовать?
- Извлечение содержимого осуществляется вручную или в автоматическом режиме?

Аспекты, представляющие интерес:

- Одушевлённые объекты (люди и т.д.)
- Неодушевлённые объекты (дома, машины и т.д.)
- Действия (активности) и события (гуляющий человек, движение машины и т.д.)

Свойства объектов:

- Фреймо-зависимые: действительны для какого-то подмножества фреймов
- Фреймо-независимые: действительны для всего видео в целом

Организация поиска по видеоданным

Свойства действий:

- Множество вовлеченных объектов
- Роли каждого из объекта в действии
- Последовательность фреймов в которых происходит данное действие

Виды запросов:

- 1) По названию извлечь видео полностью
- 2) Найти сегменты (последовательности фреймов), содержащие определенные объекты или действия
- 3) Найти все видео/сегменты, содержащие объекты/действия с определенными свойствами
- 4) В заданном сегменте найти все объекты (определенного типа), находящиеся в нескольких или во всех фреймах данного сегмента
- 5) В заданном сегменте найти все действия (определенного типа), происходящие в данном сегменте

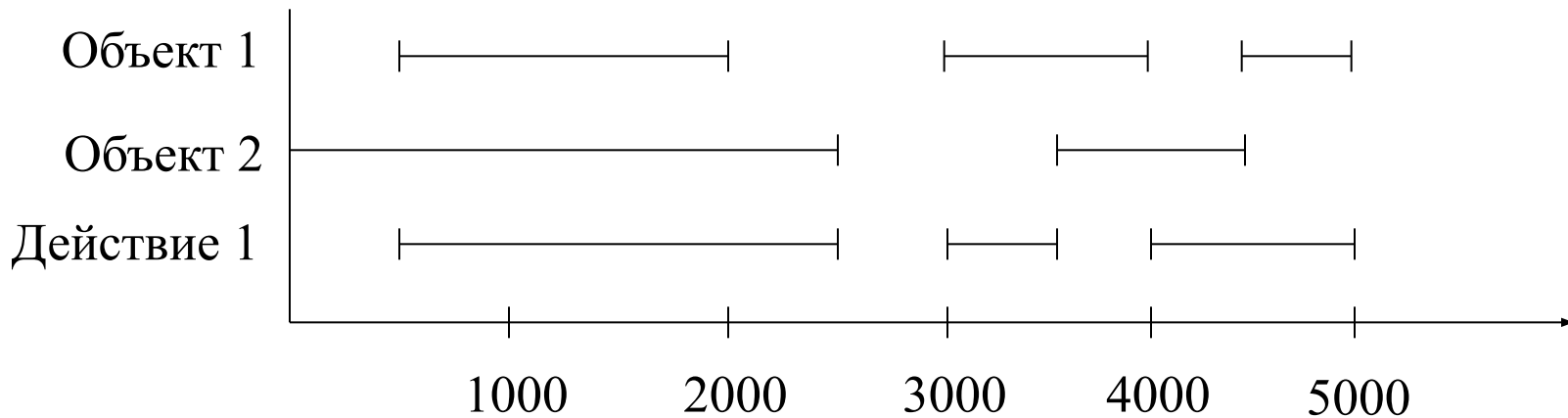
Организация поиска по видеоданным

Индексация содержимого видео:

- Вследствие большого количества фреймов описание содержимого строится как правило не для каждого фрейма
- Необходимо компактное представление
- Понятия:
 - Последовательность фреймов:
Непрерывное подмножество фреймов (смежных)
 - Упорядоченное множество последовательностей фреймов:
Упорядоченных по времени, без перекрытий
 - Сплошное (solid) множество последовательностей фреймов:
Упорядоченно и непустые промежутки между последовательностями (т.е. не смежные последовательности)
 - Ассоциативная карта сегментов:
Для каждого объекта и действия известно соответствующее сплошное множество последовательностей фреймов (т.е. известны все фреймы в которых присутствует тот или иной объект или действие)

Организация поиска по видеоданным

Иллюстрация:

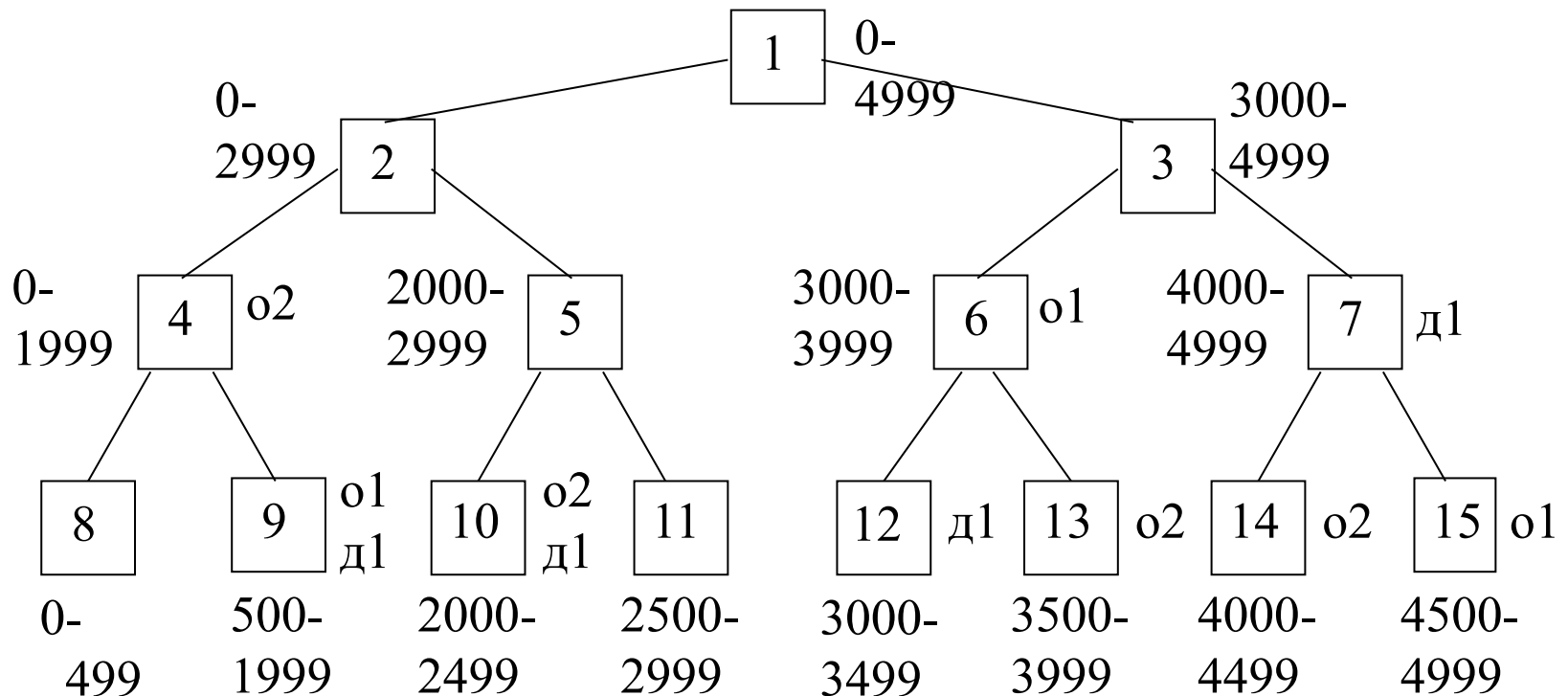


Дерево сегментов (frame segment tree):

- Двоичное дерево
- Специальный (одномерный) случай пространственного представления пространственных объектов на основе отсечения (см. тему 9)
- Листья задают основные интервалы в видеопоследовательности:
 - Листья упорядочены (по времени), и покрывают все видео
 - Внутренний узел описывает сегмент видео, полученный конкатенацией интервалов, представляемых потомками этого узла
 - Корень дерева соответствует всему видео

Организация поиска по видеоданным

Дерево сегментов для примера с предыдущего слайда:



Индексация:

Объект 1 → 6, 9, 15

Объект 2 → 4, 10, 13, 14

Действие 1 → 7, 9, 10, 12

Организация поиска по видеоданным

Индексация:

- Список указателей на узлы дерева сегментов для каждого объекта и действия
- Сами объекты и действия могут индексироваться традиционными способами
- Каждый узел дерева сегментов указывает на список указателей на объекты и действия, которые находятся/происходят в сегменте, описываемым данным узлом
- Например, в предыдущем примере:
 - Узел 4 → Объект 2
 - Узел 6 → Объект 1
 - Узел 7 → Действие 1
 - Узел 9 → Объект 1, Действие 1
 - Узел 10 → Объект 2, Действие 1
 - Узел 12 → Действие 1
 - Узел 13 → Объект 2
 - Узел 14 → Объект 2
 - Узел 15 → Объект 1
- Может быть обобщено для нескольких видео (общее дерево, объединенный набор объектов/действий)

Организация поиска по видеоданным

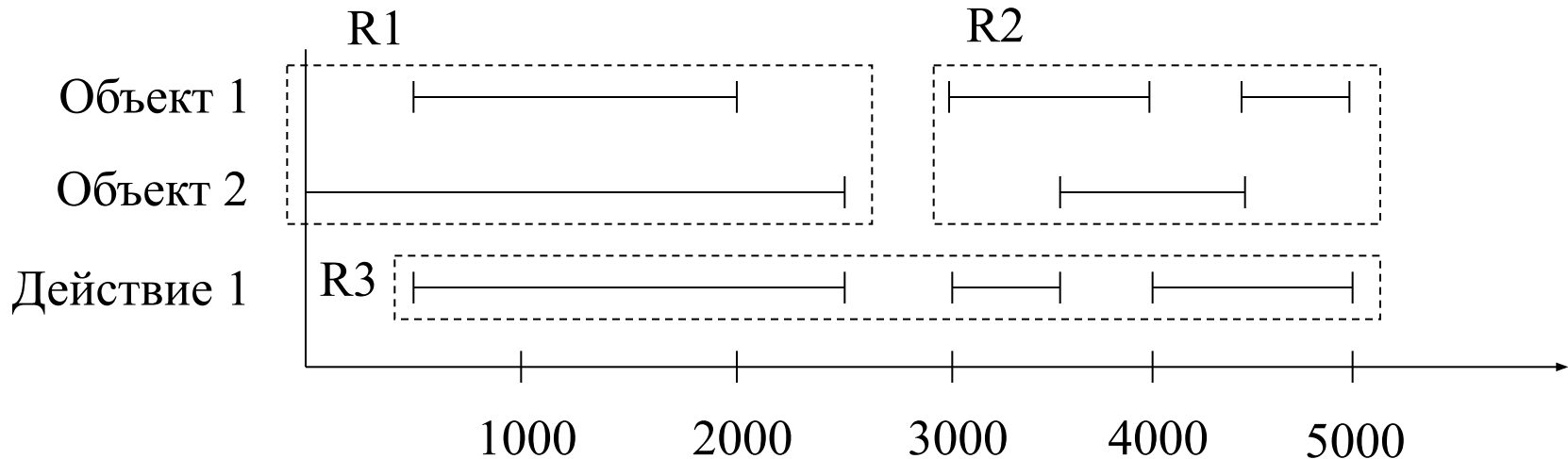
Выполнение запросов:

- 1) Найти сегменты, в которых есть заданный объект или происходит заданное действие:
Тривиально – следовать указателям
- 2) Найти объекты (действия), находящиеся между фреймами s и e :
Упорядоченно обходить дерево (обозначим: I – интервал текущего узла):
 - Если $I \cap [s, e) = \emptyset$, не просматривать поддереву, образуемое данным узлом
 - Если $I \subseteq [s, e)$, просмотреть все поддереву и вернуть все содержащиеся в нем объекты
 - Иначе, вернуть все объекты текущего узла и продолжить поиск в двух поддеревьях
- 3) Найти объекты/действия, находящиеся/произошедшие вместе с объектом x :
Просмотреть сегменты (соответствующие узлы) с x и вернуть объекты/действия, имеющие место в этих сегментах и в их потомках

Организация поиска по видеоданным

RS-дерево (R-segment tree):

- Специальный случай R-дерева
- Две возможных реализации:
 - Одномерное пространство (время)
 - Двухмерное пространство, где второе измерение – перенумерация объектов/действий (т.е. искусственное пространственное измерение)



Ссылки на литературу

- [1] Д.Кубасов и Д.Ватолин. Обзор методов компенсации движения. Графика и Мультимедиа, 2005 (<http://cgm.graphicon.ru/content/view/76/64/>)
- [2] ITU-T Recommendation H.261. Video Codec for Audiovisual Services at px64 kbit/s. ITU-T, 1993