

081-3
A86

В. М. Артюшенко

О. И. Шелухин

М. Ю. Афонин

ЦИФРОВОЕ СЖАТИЕ ВИДЕОИНФОРМАЦИИ И ЗВУКА

MPEG JPEG JBIG



087.5
А86
Издательско-торговая корпорация «Дашков и К^о»

В. М. Артюшенко, О. И. Шелухин,
М. Ю. Афонин

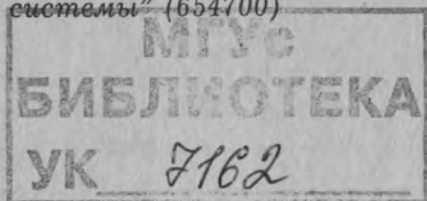
Цифровое сжатие видеоинформации и звука

Учебное пособие

Под редакцией доктора технических наук,
профессора В. М. Артюшенко

Допущено

*Учебно-методическим объединением вузов
по университетскому политехническому образованию
в качестве учебного пособия для студентов высших
учебных заведений, обучающихся по специальности
“Информационные системы и технологии” (071900)
направления подготовки дипломированных специалистов
“Информационные системы” (654700)*



Москва, 2003

УДК 681.3.053
ББК 32.81
А86

Рецензенты:

- Б. И. Шахтарин* — лауреат Государственной премии, доктор технических наук, профессор МГТУ им. Баумана;
В. П. Заболоцкий — лауреат премии Ленинского комсомола, кандидат технических наук, старший научный сотрудник, начальник теоретического отдела Центра исследований авиационно-космических материалов РАН.

А86 **Артюшенко В. М., Шелухин О. И., Афонин М. Ю. Цифровое сжатие видеoinформации и звука: Учебное пособие /** Под ред. В. М. Артюшенко. — М.: Издательско-торговая корпорация «Дашков и К^о», 2003. — 426 с.

ISBN 5-74978-258-7

В учебном пособии рассмотрены методы кодирования и сжатия изображений. Приводится описание и методы реализации основных стандартов сжатия аудиосигналов и видеоизображений MPEG-1, MPEG-2, MPEG-4, MPEG-7, фрактальных и вейвлет-преобразований.

Учебное пособие может быть рекомендовано специалистам в области разработки и эксплуатации цифровых систем сжатия, передачи, обработки, хранения и воспроизведения видеoinформации, а также студентам и аспирантам в области радиотехнических, аудиовизуальных и информационных специальностей.

ISBN 5-74978-258-7

© В. М. Артюшенко, О. И. Шелухин,
М. Ю. Афонин, 2003

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	7
Глава 1. МЕТОДЫ КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ	10
1.1. Основные алгоритмы сжатия изображений	10
1.2. Алгоритм Хаффмана	
1.3. Алгоритм JBIG	38
1.4. Алгоритм JPEG	39
1.5. Фрактальный алгоритм сжатия изображений	46
<i>Литература</i>	62
Глава 2. СЖАТИЕ АУДИОДАНЫХ АЛГОРИТМОМ MPEG-1	64
2.1. Краткий обзор алгоритмов сжатия MPEG Audio	64
2.1.1. Кодирование	64
2.1.2. Декодирование	65
2.2. Описание кодированного аудиопотока	65
2.2.1. Заголовок	66
2.2.2. error_chek (проверка ошибок)	67
2.2.3. audio_data (аудио данные)	67
2.2.4. ancillary_data (вспомогательные данные)	67
2.3. Детальное описание кодирование	68
2.3.1. Банк фильтров	68
2.3.2. Распределение бит	77
2.3.3. Форматирование	83
2.4. Процесс декодирования аудиосигнала	83
2.4.1. Обобщение	83

2.4.2. Декодирование Слоя 2 MPEG-1 Audio	83
2.5. Алгоритмы компрессии цифровых аудиоданных	88
<i>Литература</i>	106
Глава 3. MPEG-2	108
3.1. Кодирование динамических изображений и звукового сопровождения	108
3.2. Принципы кодирования изображений в MPEG-2	118
3.2.1. Черезстрочная развертка	119
3.2.2. Временное и пространственное сжатие	121
3.3. Профили и уровни	125
3.4. Структура кодированного потока MPEG-2 Видео	127
3.4.1. Общие положения	127
3.4.2. Последовательность видео	127
3.4.3. Группа изображений (Group of Pictures, GOP).....	128
3.4.4. Изображение (кадр)	131
3.5. Декодирование видео	135
3.5.1. Восстановление блоков размером 8x8	135
3.5.2. Декодирование кодов с переменной длиной	135
3.5.3. Инверсное сканирование	136
3.5.4. Обратное квантование	137
3.5.5. Обратное дискретное косинусное преобразование	137
3.5.6. Компенсация движения	137
3.5.7. Пропущенные макроблоки	138
3.6. Режимы предсказания	138
3.7. Единый поток данных	138
3.7.1. Пакетный элементарный пакет	138
3.7.2. Программный поток	141
3.7.3. Транспортный поток	142
3.7.4. Транспортный пакет	147
3.7.5. Заголовок транспортного пакета	148
3.7.6. Временная модель	151

3.7.7. Условный доступ	151
3.7.8. Multi-plex-wide operations	151
3.7.9. Transport Stream System Target Decoder	151
3.8. Операции с отдельными потоками (уровень пакетов PES)	152
3.8.1. Демультимплексирование	152
3.8.2. Синхронизация	152
3.8.3. Связь с уровнем сжатия	153
3.8.4. Пакеты PES	153
3.9. Программно-зависимая информация	155
3.9.1. Типы общих полей PSI	156
3.9.2. Таблица объединения программ (PAT)	157
3.9.3. Таблица состава программы (Program Map Table, PMT)	159
3.9.4. Таблица условного доступа (Conditional Access Table)	162
3.9.5. Таблица сетевой информации (Network Information Table)	163
3.9.6. Поле адаптации (Adaptation Field)	163
3.10. Кодирование аудиоданных в стандарте MPEG-2 ISO/IEC 13818-3	164
<i>Литература</i>	169
Глава 4. СТАНДАРТ MPEG-4	171
4.1. Цели и возможности стандарта MPEG-4	171
4.2. Детальное техническое описание стандарта MPEG-4	178
4.3. Список основных функций, обеспечиваемых MPEG-4	213
4.4. Профили MPEG-4	218
4.5. Кодирование видеоданных в MPEG-4	225
4.6. Кодирование аудиоданных в MPEG-4 ISO/IEC 14496-3	231
<i>Литература</i>	234

Глава 5. MPEG-7	235
5.1. Область применения стандарта MPEG-7	235
5.2. Главные функции MPEG-7	246
5.3. Техническое описание стандарта MPEG-7	249
5.4. Описание контента	289
<i>Литература</i>	295
Глава 6. СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ	297
6.1. Отличительные характеристики видеопоследовательностей	297
6.2. Моделирование источника	298
6.3. Определение смены сцен	303
6.4. Обоснование моделей	311
6.5. Нестационарность MPEG2-видеографика	314
6.6. Анализ и модели VBR-видеоданных	322
6.7. Аналитические методы оценки долговременной зависимости видеопоследовательности	329
<i>Литература</i>	334
ПРИЛОЖЕНИЕ 1. Стандартизация в области компьютерной видеоконференцсвязи (КВКС)	335
ПРИЛОЖЕНИЕ 2. Рекомендации H.323. КВКС в сетях с негарантированным уровнем качества обслуживания	353
ПРИЛОЖЕНИЕ 3. Рекомендации H.324. КВКС в низкоскоростных сетях	373
ПРИЛОЖЕНИЕ 4. Методы субъективной оценки качества видео для мультимедийных приложений	394

ПРЕДИСЛОВИЕ

Области применения методов кодирования и сжатия видеоинформации весьма разнообразны: от передачи и хранения изображений до спутниковых цифровых телекоммуникационных систем. Внимание к сжатию видеоинформации особенно возросло в последнее десятилетие в связи с разработкой принципиально новых цифровых телекоммуникационных систем. Создание новейших цифровых устройств обработки, передачи и хранения видеоизображений связано с радикальным изменением технологических возможностей новейших процессорных систем, создаваемых ведущими мировыми фирмами, специализирующимися в области совершенствования аппаратных и программных компьютерных средств. Использование новейших процессоров с производительностью несколько миллиардов операций в секунду обеспечивает реализацию самых сложных и вычислительно емких алгоритмов сжатия, что невозможно было осуществить ранее.

В учебном пособии излагаются вопросы методов кодирования и сжатия видеоинформации, приводится описание и методы реализации основных стандартов сжатия видеоизображений и аудиосигналов MPEG-1, MPEG-2, MPEG-4, MPEG-7, фрактальных и вейвлет-преобразований. Пособие базируется на результатах научных исследований, выполненных при непосредственном участии авторов.

Учебное пособие состоит из шести глав, предисловия, заключения и четырех приложений.

В первой главе рассмотрены основные методы и алгоритмы сжатия изображений, такие как: алгоритм Хаффмана, JBIG, JPEG, вейвлет-преобразования фрактальный алгоритм сжатия видеоинформации. Вторая глава посвящена сжатию аудиоданных алгоритмом MPEG-1. В третьей главе рассмотрены принципы кодирования изображений в MPEG-2. В

четвертой главе приведено детальное техническое описание, цели и возможности стандарта MPEG-4. Пятая глава посвящена основным характеристикам и области применения стандарта MPEG-7. В шестой главе рассмотрены статистические характеристики видеопоследовательностей, исследованы их отличительные характеристики, осуществлен анализ VBR-видеоданных и аналитических методов оценки долговременной зависимости, произведено моделирование видеоисточника.

В Приложении 1 рассмотрены вопросы стандартизации в области компьютерной видеоконференцсвязи (КВКС). В Приложении 2 приведены рекомендации H.323 КВКС в сетях с негарантированным уровнем качества обслуживания. Приложение 3 посвящено рекомендациям H.324 КВКС в низкоскоростных сетях. В Приложении 4 рассмотрены методы субъективной оценки качества видео для мультимедийных приложений.

Предлагаемое учебное пособие основано на курсах лекций "Обработка и передача видеоинформации в компьютерных сетях", "Системы цифровой обработки информационных сигналов", в течение ряда лет читаемых авторами в МГУС. Подготовка специалистов в области информационных систем и компьютерных технологий требует обобщения накопленного опыта в области разработки новых принципов цифрового сжатия видеоинформации. Этой цели и посвящено учебное пособие.

Рукопись отличается от существующих изданий, посвященных вопросам сжатия видеоинформации, тем, что она наиболее полно систематизирует материал новейших исследований проводимых в этой области.

Изложение материала в учебном пособии дается на основе современного математического аппарата, на достаточно высоком научном уровне. В пособии подробно освещены вопросы практической реализации рассматриваемых систем, приводятся оценки основных рабочих характеристик, структурные схемы устройств кодирующих и декодирующих видеоданные.

Рассмотренные вопросы представляют интерес для специалистов в области разработки и эксплуатации цифровых систем сжатия, передачи, обработки, хранения и воспроизведения видеоинформации, а также студентов и аспирантов в области радиотехнических, аудиовизуальных и информационных специальностей.

Учебному пособию присвоен гриф Учебно-методического объединения вузов по университетскому политехническому образованию для студентов высших учебных заведений, обучающихся по специальности "Информационные системы и технологии" (071900) направления подготовки дипломированных специалистов "Информационные системы" (654700), однако оно будет полезно и при подготовке специалистов по другим специальностям.

Глава 1

МЕТОДЫ КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ

1.1. Основные алгоритмы сжатия изображений

Главным направлением развития различных систем связи, телевидения и информационно-измерительных систем различного назначения является разработка и внедрение цифровых методов и устройств обработки, преобразования и передачи сигналов.

Однако использование цифровых методов передачи приводит к увеличению полосы занимаемых частот и соответственно уменьшению скорости передачи изображений. Эта проблема может быть решена путем разработки эффективных методов цифрового кодирования (сжатия), компьютерных и телевизионных изображений [1—4].

Существуют потребности и в других услугах: видеоконференцсвязь, видеотелефония, обмен накопленными образовательными или развлекательными базами данных, передача видеоизображений в реальном и нереальном масштабе времени, обмен программами между студиями и компьютерами и т. п.

С возникновением нового научно-технического направления, такого как мультимедиа (multimedia), проблема сжатия различных сигналов встала на первое место. Под термином “мультимедиа” понимается объединение и решение на базе компьютеров многообразных задач синтеза, анализа и передачи (обмена) речи, звука, видеоизображений, дизайна, факса, графики и текста, алгоритмов и микропрограмм и многих других.

Можно сказать, что интенсивное развитие методов и устройств сжатия различных изображений стимулируется в настоящее время развитием цифровых сетей интегрального обслуживания и технологии “мультимедиа” [12—16].

Каждое из этих направлений обычно характеризуется своими особыми требованиями, и в большинстве случаев невозможно использование разработанного кодека для другой области применения.

Следует отметить, что методы и устройства сжатия различных сообщений (в том числе и телевидения) интенсивно развивались последние 20—30 лет, но только на базе новых технологий достигнуты выдающиеся результаты.

Возможности сокращенного описания изображений обуславливаются в основном двумя факторами.

Первый связан с ограничениями зрительного восприятия получателя, благодаря которым исходное изображение можно без ущерба для субъективного качества аппроксимировать, более простым, экономно описываемым изображением (первичное сжатие).

Второй — опирается на избыточность цифрового представления изображений (даже после “первичного сжатия”). Последовательность дискретных величин, поступающих с выхода устройства первичного сжатия изображений, как правило, содержит остаточную избыточность, которую можно уменьшить методами неравномерного (статистического, энтропийного) кодирования источников.

Первый класс методов сжатия основан на свойствах изображения и динамики его изменения от кадра к кадру. Его применение ведет к неизбежным потерям информации и ухудшению качества изображения. Этот способ может быть реализован как в виде внутрикадрового кодирования, так и в виде межкадрового кодирования: обычно используется сочетание этих процедур. Энтропийное (статистическое) кодирование обеспечивает сжатие за счет использования статистических свойств сигнала и в принципе не ведет к потерям информации.

В настоящее время известно и разрабатывается значительное число методов сжатия телевизионных и других видеосигналов с учетом указанных факторов.

К первой группе методов следует отнести:

- ◆ кодирование с предсказанием (например, видео — АДИКМ);
- ◆ линейное кодирование с преобразованием (обычно с использованием ортогональных рядов Фурье, Карунена—Лозва, Адамара, Хартли, Гильберта, косинусного и др.) [4, 9, 20];
- ◆ межкадровое кодирование, основанное на том, что большинство изображений незначительно изменяется от кадра к кадру;
- ◆ прореживание отсчетов (и кадров) с последующим их восстановлением путем интерполяции и экстраполяции;
- ◆ кодирование с расчленением изображения на фрагменты;
- ◆ гибридное (комбинированное) внутри- и межкадровое кодирование, представляющее собой различные сочетания перечисленных методов кодирования с преобразованием и некоторые другие.

Ко второй группе методов статистического сжатия следует отнести:

- ◆ использование статистических кодов, например, кода Хаффена и его модификаций [18, 19], кода Джелинека [20], арифметических кодов [21];

- ◆ векторного квантования и некоторые другие [22—24].

В современных устройствах сжатия видеоизображений обычно также используются различные комбинации из первой и второй групп.

Некоторым особняком от этих двух названных групп методов изображения стоит так называемый фрактальный метод сжатия, которому в настоящее время уделяется большое внимание [17, 7].

Далее коротко рассмотрим основные методы и устройства сжатия видеоизображений, на которые в настоящее время или уже имеются стандарты или они находятся на конечной стадии разработки.

В настоящее время имеется несколько стандартов (рекомендаций) международных и европейских организаций и групп экспертов по различным видам услуг в области кодирования

следующих видеосигналов: видеоконференцсвязи, видеотелефонии, передачи подвижных и неподвижных изображений и т. д.

Основными требованиями, предъявляемыми к кодекам неподвижных изображений (как и речевым) являются: возможно более высокое сжатие без существенной потери качества изображения, удовлетворение требований международных стандартов и рекомендаций, относительно невысокая сложность и стоимость.

Первоначально перед разработчиками алгоритмов сжатия видеоизображения ставилась задача передачи цветного изображения с разрешением 720×576 элементов и скоростью 64 кбит/с, исходя из потребностей фотовидеотекста и возможного применения в цифровых сетях с интегральным обслуживанием.

Разработанный алгоритм обеспечивал хорошее качество, но время передачи составило 6,5 с. Для ускорения процесса передачи был реализован режим "прогрессивного кодирования", удобный для быстрого (2 с) просмотра [5, 6].

В ходе исследований стало ясно, что для большого числа разновидностей фотоизображений необходима разработка универсального алгоритма, способного адаптироваться к разнообразию разрешающих способностей, цветопередач и полюс пропускания.

В настоящее время алгоритм сжатия неподвижных кадров (видеоизображений) основывается на стандарте JPEG (Joint Photographic Experts Group) [6—11].

Указанный стандарт обеспечивает очень хорошее или отличное качество изображения для различных классов неподвижных видеоизображений.

Основой стандарта является алгоритм адаптивного дискретного косинусного преобразования (ADCT — Adaptive Discrete Cosine Transform), состоящей из трех основных частей: базовой системы, ряда расширителей возможностей и статистического кодирования.

Базовая система обеспечивает совместимость различных вариантов кодирования. В ней используется метод дискретного косинусного преобразования 8×8 , равномерное квантование и код Хаффмана. Расширители возможностей могут применяться порознь или в сочетаниях.

Структурная схема видеокодека включает в себя дискретный косинусный преобразователь и статистический кодер Хаффмана.

На приемной стороне (в декодере) производятся обратные преобразования. Алгоритм является достаточно сложным.

Кодирование изображения по алгоритму JPEG обычно начинается с преобразования цветового пространства в сигнал яркости Y и два цветоразностных сигнала U , V . Хотя это и необязательно, однако такой подход позволяет повысить эффективность сжатия. При этом степень сжатия компоненты яркости будет меньше, чем цветоразностных компонент, так как люди в гораздо меньшей степени замечают изменения в цвете.

После преобразования цветового пространства обычно, но необязательно, производится прореживание (U , V) данных цветности. При прореживании отбрасываются цветоразностные компоненты строк или столбцов пикселей с определенными номерами (например, каждой второй строки и каждого второго столбца). Следующий этап процедуры сжатия данных заключается в преобразовании небольших блоков изображения при помощи двумерного дискретного косинусного преобразования. Обработка ведется блоками 8×8 пикселей, т. е. сразу обрабатывается 64 пикселя. Выбор такого размера блока объясняется несколькими причинами: во-первых, такой блок с большой вероятностью содержит пиксели близкого цвета; во-вторых, является достаточно большим, превышающим интервал корреляции видеосигнала. Дискретное косинусное преобразование во многом напоминает дискретное преобразование Фурье.

При выполнении этой операции 64 пикселя преобразуются в матрицу из 64 коэффициентов. Важнейшей особенностью

этой матрицы является то, что основную энергию несут первые ее коэффициенты, а энергия последующих быстро убывает (максимальные значения коэффициентов концентрируются в левом верхнем углу матрицы, а минимальные — в нижнем правом углу). Далее преобразованная матрица проходит операцию квантования, которая применяется для сокращения разрядности коэффициентов. Фактически квантование означает деление матрицы 8×8 , коэффициентов дискретного косинусного преобразования на матрицу квантования также с размерностью 8×8 . При этом результат деления округляется до ближайшего целого. После квантования значения чисел в левом верхнем углу оказываются значительно меньше чем ранее, а в правом углу получаются преимущественно нули. Именно здесь происходит основная, необратимая, потеря информации. Причем цветоразностные компоненты могут квантоваться более грубо. После этого матрица, представляющая результат квантования, вытягивается в строку данных так, что все последовательности нулей правого нижнего угла оказываются в конце строки.

В некоторых версиях информация о яркости и цвете затем кодируется так, что сохраняются только отличия между соседними блоками.

После этого производится статистическое кодирование по методу Хаффмана метод сжатия без потерь, который широко используется в универсальных утилитах файлов. Сначала анализируется вся последовательность символов. Часто повторяющимся сериям бит присваиваются короткие элементы (маркеры). В частности, последние нули в конце строки могут быть заменены одним символом конца блока. Поскольку все блоки имеют одинаковую длину, всегда известно сколько нулей было опущено.

При восстановлении (декодировании) данных перечисленные выше операции выполняются в обратном порядке.

Возможная степень сжатия зависит как от вида изображений, так и от применяемых типов матриц квантования. Гладкие полутоновые изображения без мелких деталей

сжимаются лучше, чем с мелкими деталями. Например, изображение человеческого лица можно сжимать с большим коэффициентом, чем картинку с текстом.

Увеличение размера матрицы квантования приводит к повышению коэффициента сжатия. Для различных компонентов Y, U, V могут использоваться свои матрицы квантования.

Следует отметить, что в настоящее время имеется несколько различных пакетов программ по алгоритму JPEG. В некоторых программах требуется дополнительное дисковое пространство в дополнение к основной памяти. Некоторые предусматривают несколько фиксированных режимов сжатия, другие дают возможность выбирать их. Важным фактором является скорость сжатия, которая может изменяться от нескольких десятков секунд до нескольких минут. Некоторые алгоритмы создают самовосстанавливающиеся изображения. Многие системы сжатия по стандарту JPEG требуют очень большого объема памяти. Следует отметить, что JPEG регламентирует лишь только метод сжатия, а не формат файлов. Спецификацией для хранения файлов, сжатых по алгоритму JPEG, является JFIF (JPEG File Interchange Format).

Заметим, что пользователи алгоритма JPEG могут устанавливать различную степень сжатия, идя на компромисс между качеством изображения и размером файла (временем вычислений). Разные программы могут обладать весьма различными функциональными и скоростными возможностями.

Следует также отметить, что для оценки качества восстанавливаемых JPEG-файлов важно иметь специальный монитор. Стандартные VGA и Super VGA-дисплеи (не говоря уже о CGA) не способны воспроизводить, например, 24-битовый TIFF-файл изображения.

Для цветного изображения при использовании полного алгоритма в зависимости от качества изображения обеспечиваются следующие затраты бит на растровый элемент (пиксел);

-0,25...0,5 бит/эрл — хорошее качество;

-0,5...0,75 бит/эрл — очень хорошее качество;

-0,75...1,5 бит/эрл — отличное качество для большинства применений.

Размеры сжатого файла составляют от 1,2 Мбайт до 130 Кбайт, время сжатия на компьютере IBM PC — от нескольких десятков секунд до нескольких минут, а на ввод-вывод затрачивается несколько минут.

В [5] приводятся результаты испытаний ряда программ сжатия неподвижного цветного изображения размером 203×254 мм с разрешением 105 точек/см (файл размером порядка 14 Мбайт) в средах DOS и Windows по стандарту JPEG.

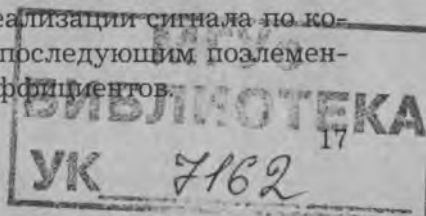
Заметим, что обычно при передаче телевизионного изображения затрачивается 8 бит/эрл. Таким образом, в соответствии со стандартом обеспечивается сжатие “подвижных” изображений примерно в 8—80 раз.

Следует отметить, что обычно стандарты определяют процедуры сжатия (кодирования) и декодирования, а не сами схемные и алгоритмические решения, этим самым они дают возможность их совершенствования. В настоящее время уже разработаны и ведется дальнейшее совершенствование СБИС и высокоэффективных сигнальных процессоров для видеокодеков различного назначения.

В настоящее время имеется несколько пакетов сжатия по стандартам JPEG, MPEG и H.261. Эти пакеты существенно отличаются по стоимости, совместимости файлов, скорости, качеству изображений и требуемым аппаратным ресурсам. Почти все они могут работать с изображениями объемом 2—3 Мбайта.

Основой алгоритмов сжатия неподвижных изображений являются дискретные косинусные преобразования, векторное квантование и статистическое кодирование. Рассмотрим коротко эти преобразования.

Основная идея дискретного косинусного преобразования состоит в разложении кодируемой реализации сигнала по косинусному ортогональному базису с последующим поэлементным квантованием полученных коэффициентов.



Прямое и обратное (инверсное) двумерное дискретное косинусное преобразование исходного массива данных $s(x, y)$, где $x = 0, 1, \dots, 7$ и $y = 0, 1, \dots, 7$, определяется соответственно выражениями

$$F(u, v) = 0.25C(u)C(v)[S(x, y)\cos\{(2x + 1)u\pi/16\}\cos\{(2y + 1)v\pi/16\}], \quad (1.1)$$

$$S(x, y) = 0.25[C(u)C(v)F(u, v)\cos\{(2x + 1)u\pi/16\}\cos\{(2y + 1)v\pi/16\}], \quad (1.2)$$

где $C(u), C(v) = 1/2$ для $u = 0$ и $v = 0$; $C(u), C(v) = 1$ для других значений u и v .

$S(x, y)$ обозначает матрицу исходных данных в общем случае размером $N1 \times N2$ (в большинстве случаев 8×8);

$F(u, v)$ — матрица коэффициентов дискретного косинусного преобразования (чаще всего от -1023 до 1023);

$S(x, y)$ — матрица исходных данных (чаще всего от -127 до 127).

Двумерные преобразования вычисляются в результате $N1 \times N2$ — кратного применения одномерного косинусного преобразования.

$$F(u, v) = (1/4)C(u)C(v) \sum_{x=0}^7 \sum_{y=0}^7 p(x, y) \left[\cos \frac{(2x+1)u\pi}{16} \right] \left[\cos \frac{(2y+1)v\pi}{16} \right], \quad (1.3)$$

где v — горизонтальная координата графического блока, u — вертикальная, x — вертикальная координата внутри блока, а y — горизонтальная координата внутри блока, $C(u), C(v) = \sqrt{2}$ для $u, v = 0$ и $C(u), C(v) = 1$ в противном случае.

Два члена в квадратных скобках являются ядрами преобразования, показанными ниже на рис. 1.1, а $p(x, y)$ представляет собой пиксельные данные блока реального рисунка. Начало координат в обоих случаях в верхнем левом углу. Процесс кодирования сводится к разбиению изображения на блоки 8×8 пикселей и выполнению процедуры двумерного ДСТ для каждого из этих блоков. Полученные коэффициенты преобразования дискретизируются. 64 числа, характеризующие уровень сигнала, превращаются в 64 коэффициента преобразования (амплитуды пространственных частот), которые хо-

рошо поддаются процедуре сжатия. Дискретизатор округляет коэффициенты, эта процедура вносит некоторые ошибки, но обратное преобразование на принимающей стороне за счет усреднения частично устраняет вносимые искажения. На практике дискретизатор реализует несколько более сложный алгоритм.

Интуитивно метод DCT базируется на выявлении того, насколько вышестоящий блок отличается от нижестоящего. Для реального представления (сжатия) коэффициентов преобразования здесь также используются коды Хаффмана.

DCT обеспечивает сжатие на уровне 0,5—1,0 бит/пиксель при хорошем качестве изображения. Сжатие требует времени, а максимально приемлемым временем задержки при пересылке изображения является 5 секунд. На рис. 1.2. приведена качественная оценка четкости и соответствия оригиналу изоб-

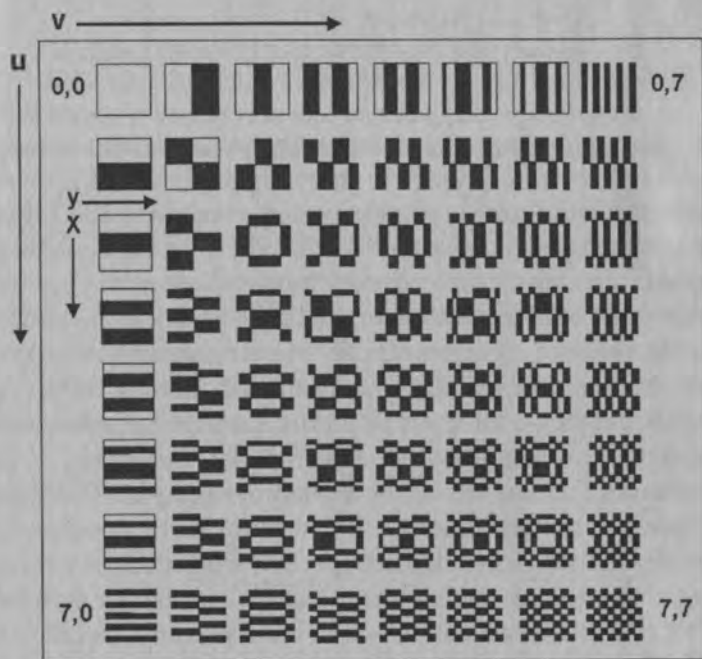


Рис. 1.1. Графическое представление двумерного преобразования



Рис. 1.2. Качество DCT-изображения для различных значений сжатия информации (картинка имеет разрешение 512×512 пикселей; заполненные квадратики соответствуют цветному изображению, а незаполненные — черно-белому)

ражения в зависимости от величины сжатия (DCT). Если использовать скорость обмена 64 кбит/с, то степени сжатия 0,01 бита на пиксель будет соответствовать время передачи изображения 0,04 секунды, а сжатию 10 — время передачи 40 сек.

Отображение графического образа может выполняться последовательно (примерно так, как мы читаем текст: слева-направо и сверху-вниз) или с использованием прогрессивного кодирования (сначала передается вся картинка с низким разрешением, затем последовательно четкость изображения доводится до максимальной). Последний метод весьма удобен для систем WWW, где, просмотрев изображение низкого разрешения, можно отменить передачу данных улучшающих четкость и тем самым сэкономить время. Хорошо распознаваемое изображение получается при сжатии порядка 0,1 бита на пиксель.

Энергетические спектры сигналов изображений достаточно быстро убывают с возрастанием номера спектрального коэффициента в базисе дискретного косинусного преобразования.

Известно также, что искажение модулей коэффициентов преобразования сигнала для этого базиса до известных пределов мало сказывается на качестве восстановления изображения. Отсюда вытекает возможность дополнительного сжатия, использующего избыточность в амплитудной составляющей для сокращения объема цифрового описания.

При этом часть коэффициентов, дающих наименьший вклад в суммарную энергию сигнала, отбрасывается. Экономия в полученном цифровом представлении сигнала достигается за счет этого отбрасывания.

Использование дискретного косинусного преобразования объясняется еще и тем, что оно для большого класса изображений хорошо аппроксимирует оптимальное дискретное представление Карунена—Лоэва, имеет алгоритм быстрого преобразования, просто в реализации и обеспечивает при этом существенное сжатие.

Сканирование коэффициентов дискретного косинусного преобразования производится по зигзагу, результирующая последовательность будет содержать длинные цепочки нулей, особенно в концевой части последовательности. Одна из задач предложенного алгоритма — “собрать” все нули по возможности вместе, что облегчает дальнейшую процедуру эффективного статистического кодирования. После преобразования сигнал обрабатывается путем равномерного квантования. Шаг квантования может регулироваться в соответствии со скоростью передачи, характеризуемой степенью занятости буферного ЗУ. Если скорость передачи достигает предельной, шаг квантования увеличивается, чем уменьшается объем подлежащей кодированию информации. Если скорость передачи позволяет, то, наоборот, шаг квантования может быть уменьшен для получения более высокого качества изображения.

Считается весьма перспективным и частично используется в JPEG векторное квантование.

Векторное квантование эффективно, когда требуемое число битов на элемент изображения должно быть меньше одной двоичной единицы. Обычно векторный квантователь состоит из множества, называемого кодовой книгой, содержащей L кодовых векторов размерности K . Векторы формируются путем деления исходного изображения на смежные неперекрывающиеся блоки изображений. Если кодовая книга создана и она имеется на передающей и приемной стороне, то при получении номера индекса вектора приемник выбирает из своей кодовой книги соответствующий вектор и заполняет им необходимое место на изображении.

Статистическое (энтропийное) кодирование используется для уменьшения избыточности (сжатия) сообщений, обусловленной неравной вероятностью элементов.

Код, уменьшающий избыточность, можно построить в разных вариантах. Однако во всех случаях часто встречающиеся высоковероятные элементы или буквы должны кодироваться короткими кодовыми комбинациями, а для реже встречающихся можно применять более длинные кодовые комбинации. Необходимо также, чтобы короткие кодовые комбинации, соответствующие высоковероятным элементам, не совпадали с началами более длинных. В противном случае невозможно будет правильно декодировать закодированное сообщение.

На возможность такого кодирования, учитывающего статистические свойства сообщений, указал Шеннон, а удобное правило (алгоритм) этого кодирования предложил Фэно. Поэтому коды такого рода называются статистическими или эффективными, или энтропийными, а также кодами Фэно—Шеннона. Алгоритм статистического кодирования Фэно—Шеннона заключается в следующем:

1. Подлежащие кодированию элементы сообщений располагаются в первом столбце таблицы в порядке убывания их вероятностей.
2. Элементы сообщений разбиваются на две группы с примерно равными суммарными вероятностями. Элементом пер-

вой группы в качестве первого знака кодовой комбинации присписывается 0, а элементам второй группы 1.

3. Элементы, входящие в каждую из групп, вновь разбиваются на две группы с примерно равными суммарными вероятностями. Элементом вновь полученных первых групп в качестве второго знака кодовой комбинации присписывается 0, а элементам вторых групп 1.

4. Этот процесс продолжается, пока в каждой из групп не останется по одному элементу.

Проблема сжатия и передачи движущегося изображения еще сложнее. Алгоритм кодирования такого изображения описан в рекомендациях CCITT H.261 и предполагает, что скорость передачи лежит в интервале 40 кбит/с — 2 Мбит/с. Следует иметь в виду, что видеотелефония и видеоконференции требуют синхронной передачи звука и изображения. Например, в стандарте H.221, это 46,4 кбит/с для видео и 16 кбит/с для звука). Нормальный формат телевидения имеет 625 и 525 строк развертки и частоту кадров 25—30 в секунду. Цветное телевидение использует сигналы R (red), G (green) и B (blue), причем яркость луча (y) определяется соотношением: $Y = 0.30R + 0.59G + 0.11B$ (при отображении белого цвета). Информация о цветах определяется формулами: $C_B = B - Y$ и $C_R = R - Y$. Зная величины y , C_B и C_R , можно восстановить значения R, G и B. При сжатии цветного изображения учитывается тот факт, что человеческий глаз извлекает большую часть информации из контуров предметов, а не из цветных деталей. Например, в рекомендации CCIR 601 предлагается использовать полосу 135 МГц для кодирования Y и только по 675 МГц для C_B и C_R . Такая схема требует 216 Мбит/с, что в 3375 раза превышает возможности стандартного 64кбит/с В-канала ISDN. Приемлемыми решениями могут быть:

а) снижение числа строк до 288 (формат 625 строк) для отображения яркости;

б) использование максимально возможного сжатия графических данных;

с) повышение пропускной способности канала. Для разрешение по горизонтали вполне достаточно 3 МГц. Рекомендация 601 требует 720 пикселей для яркости и 360 для каждой из составляющих цветов. В настоящее время используется стандарт CIF (Common Intermediate Format). Для некоторых приложений рекомендовано вдвое более низкое разрешение по каждой из осей (quarter CIF). РСМ-кодирование CIF с 8 битами на пиксель требует $352 \times 288 \times (1 + 1/4 + 1/4) \times 29.97 \times 8 = 365$ Мбит/с.

Проблема сжатия информации была, есть и всегда будет актуальной. При известных современных методах, чем больше эффективность сжатия — больше задержка (наилучший результат можно получить, используя сжатие всего фильма, чем кадра или тем более строки). В каждом конкретном случае выбирается то или иное компромиссное решение. При работе в реальном масштабе времени, где в процессе обмена участвует человек, задержки более секунды вызывают раздражение, и приходится ограничиваться сравнительно скромными коэффициентами сжатия.

При пересылке движущегося изображения производится сравнение текущего кадра с предшествующим. Если кадры идентичны, никакого информационного обмена не происходит. Если кадры отличаются лишь смещением какого-то объекта, выявляются границы этого объекта, направление и величина вектора его перемещения. Так как использование индивидуальных векторов перемещения для каждого пикселя слишком расточительно, используется общий вектор для блока пикселей 16×16 по яркости и для соответствующего блока 8×8 по цвету. Точность задания вектора перемещения обычно лежит в пределах $1/2$ пикселя (стандарт MPEG-2). Только эта информация и передается по каналу связи. Выявление движущихся объектов осуществляется путем вычитания изображения двух последовательных кадров. Если бы передавалась всегда только разница кадров, происходило бы накопление ошибок. Кроме того, как кодер, так и декодер содержат прямой и обратный DCT-преобразователь. Если комбинация прямого и обратного DCT-преобразования не при-

водит к получению исходного объекта, то такого рода эффекты могут заметно усилиться. Для исключения этого времени от времени производится передача непосредственно видеосигнала. Практически преобразователь изображения представляет чудо современной технологии, которое даст работу еще не одному поколению математиков и инженеров.

Нисколько не проще система передачи и мультиплексирования потока видео данных, который содержит помимо обычной информации описания формы движущихся объектов, векторы перемещения, коэффициенты дискретизации и многое другое. Схема передачи графической информации имеет 4-уровневую, иерархическую структуру. Передача каждого кадра изображения начинается с 20-битного кода PSC (Picture Start Code, эта сигнатура позволяет выделить начало кадра изображения в общем потоке), далее следует 5-битовый код TR (Temporal Reference, временная метка, которая позволяет поместить соответствующую часть изображения в правильную точку экрана). Изображение пересылается частями, имеется 4 уровня: кадр, группа блоков GoB (Group of Blocks), макроблоки (MB) и просто блоки.

Ядро всей структуры составляет процедура передачи кадра (внутренний слой, существуют еще слои GoB, MB и блока, см. рис. 1.3, 1.4, 1.5).

Поле Ptype содержит 6 бит, которые характеризуют формат изображения (используется ли формат CIF или QCIF). Однобитное поле PEI указывает на то, следует ли далее 8-битное поле PSpare (предназначено на будущее). Если



Рис. 1.3. Схема передачи кадра изображения

PEI = 0, начинается цикл передачи GoB. Группа блоков составляет одну двенадцатую картинки CIF или одну треть QCIF. GoB описывает Y (яркость), 176 пикселей для каждой из 48 строк и соответствующие 88×24 элементов для C_B и C_R .

GBSC — (Group of Blocks Start Code) представляет собой 16-разрядное слово, за которым следует 4 бита номера GoB (GN — GoB number). GN указывает, какой части изображения соответствует данный GoB. Поле gquant имеет 5 бит и указывает на номер преобразователя (одного из 31 дискретизаторов), который используется данным GoB. Смысл GEI идентичен PEI. GEI и GSpere позволяют сформировать структуру данных, идентичную той, что используется на уровне кадра.

Формат пересылки mb сложнее. Каждый GoB делится на 33 макроблока (MB), каждый из которых соответствует 16 строкам по 16 пикселей Y (четыре блока 8×8) и C_B и C_R . Каждый макроблок начинается с его адреса MBA (MacroBlock Address), имеющего переменную длину и определяющего положение макроблока в GoB.



Рис. 1.4. Блок-схема кодирования и передачи изображения

Макроблоки не передаются, если данная часть изображения не изменилась. За MBA следует код переменной длины Mtype, характеризующий формат макроблока (применен ли метод подвижного вектора MVD и т. д.) и последующую информацию. СВР (Coded Block Pattern) представляет собой кодовое слово переменной длины, которое несет в себе информацию о том, какой из шести блоков преобразования (8×8)

содержит коэффициенты (слой блоков). СВР нужно не для всех типов макроблоков. Каждый блок завершается флагом EOB (End of Block).

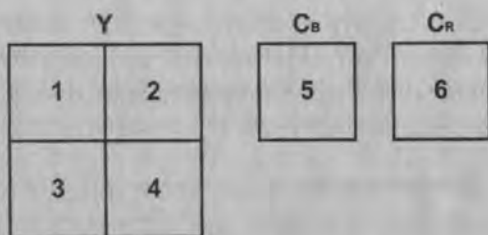


Рис. 1.5. Размещение блоков в макроблоках

Сама природа алгоритма кодирования и передачи графических данных такова, что число бит передаваемых в единицу времени зависит от характера изображения. Чем динамичнее изменяется картинка, тем больше поток данных. Для выравнивания потока данных широко используется буферизация. Буферизация в свою очередь порождает дополнительные задержки, которые в случае видеоконференций или видеотелефонии не должны превышать нескольких сотен миллисекунд.

Так как при передаче изображения широко используются коды переменной длины, она крайне уязвима для любых искажений. В случае ошибки будет испорчена вся информация вплоть до следующего стартового кода GoV. Из-за рекурсивности алгоритма формирования картинки, искажения будут оставаться на экране довольно долго. Использование векторов перемещения может привести к дрейфу искажений по экрану и расширению их области. Для того чтобы уменьшить последствия искажений, в передаваемый информационный поток включаются коды коррекции ошибок BCH (511, 493; Forward Error Correction Code), которые позволяют исправить любые две ошибки или кластер, содержащий до 6 ошибок в блоке из 511 бит (см. рис. 1.6). Алгоритм работает в широком диапазоне скоростей передачи информации. Для реализации коррекции ошибок в поток двоичных данных включается 8 пакетов, каждый из которых включает в себя 1 кадровый бит, 1 бит индикатор заполнения, 492 бита кодированных данных и 18 бит четности. Поле Fi (индикатор заполне-

ния) может равняться нулю, тогда последующие 492 бита не являются графической информацией и могут игнорироваться. Алгоритм предназначен для работы в динамическом диапазоне частот 40:1.

Порядок пересылки \longrightarrow (S1 S2 S3 S4 S5 S6 S7 S8) = (01010011)

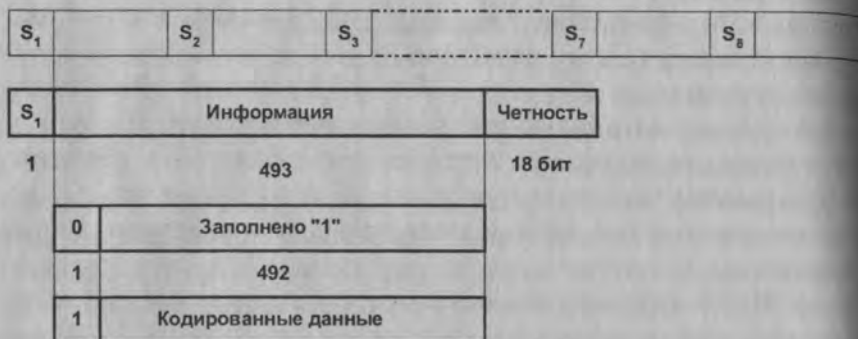


Рис. 1.6. Схема передачи данных с коррекцией ошибок

Во время переговоров или в ходе видеоконференции может возникнуть необходимость отобразить текст, выделить на экране какой-то объект, послать факс и т. д. Для решения таких задач можно использовать D-канал, но это не оптимально, так как он имеет свои специфические функции. Поэтому более привлекательным представляется создание специального протокола, работающего в рамках В-канала (H.221). Для этих целей используется младший бит каждого из октетов, что позволяет создать канал с пропускной способностью 8 кбит/с. Этот сервисный канал использует кадры по 80 бит. Первые 8 бит служат для целей синхронизации (FAS — Frame Alignment Signal) и выполняют следующие функции:

- ◆ выделение начала кадра (исключение имитации этого в информационном потоке);
- ◆ выделение начала блока кадров (опционно до 16 кадров);
- ◆ выполнение функций счетчика в многокадровых блоках (по модулю 16), может использоваться в многоточечных соединениях;

- ◆ нумерация соединений;
- ◆ CRC-контроль (опционно);
- ◆ “А-бит” для определения кадр/мультикадр/синхронизация при пересылке в противоположном направлении ($A = 0$ — передача).

При работе с каналами на 384, 1536 и 1920 Кбит/с сервисный канал использует тайм-слот 1. Следующие 8 бит имеют название BAS (Bit Allocation Signal) и выполняют следующие функции:

- ◆ код, характеризующий возможности канала (узко/широко полосная передача звука, различные видеопараметры, тип шифрования и т. д.);
- ◆ коды команд, определяющие значения передаваемых кадров;
- ◆ ESC-последовательности.

1.2. Алгоритм Хаффмана Классический алгоритм Хаффмана

Один из классических алгоритмов, известных с 60-х гг. Использует только частоту появления одинаковых байт в изображении. Сопоставляет символам входного потока, которые встречаются большее число раз, цепочку бит меньшей длины, и наоборот, встречающимся редко — цепочку большей длины. Для сбора статистики требует двух проходов по изображению.

Для начала введем несколько определений.

Определение. Пусть задан алфавит $\Psi = \{a_1, \dots, a_r\}$, состоящий из конечного числа букв. Конечную последовательность символов из Ψ

$$A = a_{i_1} a_{i_2} \dots a_{i_n} \quad (1.4)$$

будем называть *словом* в алфавите Ψ , а число n — *длиной слова* A . Длина слова обозначается как $l(A)$.

Пусть задан алфавит Ω , $\Omega = \{b_1, \dots, b_q\}$. Через B обозначим слово в алфавите Ω и через $S(\Omega)$ — множество всех непустых слов в алфавите Ω .

Пусть $S = S(\Psi)$ — множество всех непустых слов в алфавите Ψ , и S' — некоторое подмножество множества S . Пусть также задано отображение F , которое каждому слову $A, A \in S(\Psi)$, ставит в соответствие слово

$$B = F(A), B \in S(\Omega). \quad (1.5)$$

Слово B будем называть *кодом сообщения* A , а переход от слова A к его коду — *кодированием*.

Определение. Рассмотрим соответствие между буквами алфавита Ψ и некоторыми словами алфавита Ω :

$$\begin{aligned} a_1 &— B_1, \\ a_2 &— B_2, \\ &\dots \\ a_r &— B_r. \end{aligned} \quad (1.6)$$

Это соответствие называют *схемой* и обозначают через Σ . Оно определяет кодирование следующим образом: каждому слову $A = a_i a_j \dots a_n$ из $S'(\Psi) = S(\Psi)$ ставится в соответствие слово $B = B_i B_j \dots B_n$, называемое *кодом слова* A . Слова $B_1 \dots B_r$ называются *элементарными кодами*. Данный вид кодирования называют *алфавитным кодированием*.

Определение. Пусть слово B имеет вид

$$B = B' B'' \quad (1.7)$$

Тогда слово B' называется *началом* или *префиксом* слова B , а B'' — *концом* слова B . При этом пустое слово L и само слово B считаются началами и концами слова B .

Определение. Схема S обладает свойством *префикса*, если для любых i и j ($1 \leq i, j \leq r, i \neq j$) слово B_i не является префиксом слова B_j .

Теорема 1. Если схема Σ обладает свойством *префикса*, то алфавитное кодирование будет взаимно однозначным.

Предположим, что задан алфавит $\Psi = \{a_1, \dots, a_r\}$ ($r > 1$) и набор вероятностей p_1, \dots, p_r ($\sum_{i=1}^r p_i = 1$) появления символов a_1, \dots, a_r . Пусть, далее, задан алфавит W , $W = \{b_1, \dots, b_q\}$ ($q > 1$). Тогда можно построить целый ряд схем S алфавитного кодирования

$$\begin{aligned} a_1 & \text{ — } B_1, \\ & \dots \\ a_r & \text{ — } B_r \end{aligned} \quad (1.8)$$

обладающих свойством взаимной однозначности.

Для каждой схемы можно ввести среднюю длину l_{cp} , определяемую как математическое ожидание длины элементарного кода:

$$l_{cp} = \sum_{i=1}^r p_i l_i, \quad l_i = l(B_i) \text{ — длины слов.} \quad (1.9)$$

Длина l_{cp} показывает, во сколько раз увеличивается средняя длина слова при кодировании со схемой Σ .

Можно показать, что l_{cp} достигает величины своего минимума l_* на некоторой Σ и определена как

$$l_* = \min_{\Sigma} l_{cp}^{\Sigma}. \quad (1.10)$$

Определение. Коды, определяемые схемой Σ с $l_{cp} = l_*$, называются кодами с минимальной избыточностью, или кодами Хаффмана.

Коды с минимальной избыточностью дают в среднем минимальное увеличение длин слов при соответствующем кодировании.

В нашем случае алфавит $\Psi = \{a_1, \dots, a_r\}$ задает символы входного потока, а алфавит $\Omega = \{0, 1\}$, т. е. состоит всего из нуля и единицы.

Алгоритм построения схемы Σ можно представить следующим образом:

Шаг 1. Упорядочиваем все буквы входного алфавита в порядке убывания вероятности. Считаем все соответствующие слова B_{ir} из алфавита $\Omega = \{0, 1\}$ пустыми.

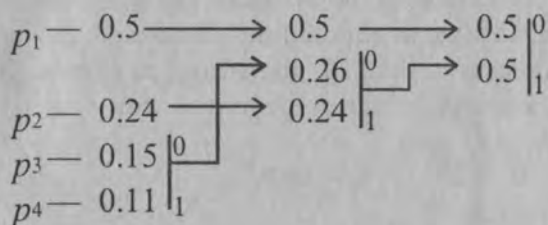
Шаг 2. Объединяем два символа a_{ir-1} и a_{ir} с наименьшими вероятностями p_{ir-1} и p_{ir} в псевдосимвол $a'\{a_{ir-1}a_{ir}\}$ с вероятностью $p_{ir-1} + p_{ir}$. Дописываем 0 в начало слова B_{ir-1} ($B_{ir-1} = 0B_{ir-1}$) и 1 в начало слова и B_{ir} ($B_{ir} = 1B_{ir}$).

Шаг 3. Удаляем из списка упорядоченных символов a_{ir-1} и a_{ir} , заносим туда псевдосимвол $a'\{a_{ir-1}a_{ir}\}$. Проводим шаг 2, добавляя при необходимости 1 или ноль для всех слов B_{ir} соответствующих псевдосимволам, до тех пор, пока в списке не останется 1 псевдосимвол.

Пример: Пусть у нас есть 4 буквы в алфавите $\Psi = \{a_1, \dots,$

$a_4\}$ ($r = 4$), $p_1 = 0.5$, $p_2 = 0.24$, $p_3 = 0.15$, $p_4 = 0.11$ $\left(\sum_{i=1}^4 p_i = 1 \right)$.

Тогда процесс построения схемы можно представить так:



Производя действия, соответствующие второму шагу, мы получаем псевдосимвол с вероятностью 0.26 (и приписываем 0 и 1 соответствующим словам). Повторяя же эти действия для измененного списка, мы получаем псевдосимвол с вероятностью 0.5. И наконец, на последнем этапе мы получаем суммарную вероятность 1.

Для того чтобы восстановить кодирующие слова, нам надо пройти по стрелкам от начальных символов к концу получившегося бинарного дерева. Так, для символа с вероятностью

стью p_4 получим $B_4 = 101$, для p_3 — получим $B_3 = 100$, для p_2 — получим $B_2 = 11$, для p_1 — получим $B_1 = 0$. Что означает схему:

$$\begin{aligned}a_1 &— 0, \\a_2 &— 11 \\a_3 &— 100 \\a_4 &— 101\end{aligned}$$

Эта схема представляет собой префиксный код, являющийся кодом Хаффмана. Самый часто встречающийся в потоке символ a_1 мы будем кодировать самым коротким словом 0, а самый редко встречающийся a_4 длинным словом 101.

Для последовательности из 100 символов, в которой символ a_1 встретится 50 раз, символ a_2 — 24 раза, символ a_3 — 15 раз, а символ a_4 — 11 раз, данный код позволит получить последовательность из 176 бит ($100 \cdot l_{cp} = \sum_{i=1}^4 p_i l_i$). То есть в среднем мы потратим 1.76 бита на символ потока.

Доказательства теоремы, а также того, что построенная схема действительно задает код Хаффмана, смотри в [10].

Как стало понятно из изложенного выше, классический алгоритм Хаффмана требует записи в файл таблицы соответствия кодируемых символов и кодирующих цепочек.

На практике используются его разновидности. Так, в некоторых случаях резонно либо использовать постоянную таблицу, либо строить ее “адаптивно”, т. е. в процессе архивации/разархивации. Эти приемы избавляют нас от двух проходов по изображению и необходимости хранения таблицы вместе с файлом. Кодирование с фиксированной таблицей применяется в качестве последнего этапа архивации в JPEG и в рассмотренном ниже алгоритме CCITT Group 3.

Характеристики классического алгоритма Хаффмана:

Коэффициенты компрессии. 8, 1, 5, 1 (лучший, средний, худший коэффициенты).

Класс изображений. Практически не применяется к изображениям в чистом виде. Обычно используется как один из этапов компрессии в более сложных схемах.

Симметричность. 2 (за счет того, что требует двух проходов по массиву сжимаемых данных).

Характерные особенности. Единственный алгоритм, который не увеличивает размера исходных данных в худшем случае (если не считать необходимости хранить таблицу перекодировки вместе с файлом).

Алгоритм Хаффмана с фиксированной таблицей CCITT Group 3

Близкая модификация алгоритма используется при сжатии черно-белых изображений (один бит на пиксел). Полное название данного алгоритма CCITT Group 3. Это означает, что данный алгоритм был предложен третьей группой по стандартизации Международного консультационного комитета по телеграфии и телефону (Consultative Committee International Telegraph and Telephone). Последовательности подряд идущих черных и белых точек в нем заменяются числом, равным их количеству. А этот ряд уже, в свою очередь, сжимается по Хаффману с фиксированной таблицей.

Определение. Набор идущих подряд точек изображения одного цвета называется *серией*. Длина этого набора точек называется *длиной серии*.

В таблице приведенной ниже заданы два вида кодов:

- ◆ *Коды завершения серий* — заданы с 0 до 63 с шагом 1.
- ◆ *Составные (дополнительные) коды* — заданы с 64 до 2560 с шагом 64.

Каждая строка изображения сжимается независимо. Мы считаем, что в нашем изображении существенно преобладает белый цвет, и все строки изображения начинаются с белой точки. Если строка начинается с черной точки, то мы считаем, что строка начинается белой серией длины 0. Например, последовательность длин серий 0, 3, 556, 10,... означает, что в этой строке изображения идут сначала 3 черных точки, затем 556 белых, затем 10 черных и т. д.

На практике в тех случаях, когда в изображении преобладает черный цвет, мы инвертируем изображение перед компрессией и записываем информацию об этом в заголовок файла.

Алгоритм компрессии выглядит так:

```
for(по всем строкам изображения) {
Преобразуем строку в набор длин серий;
for(по всем сериям) {
  if(серия белая) {
    L= длина серии;
    while(L > 2623) { // 2623=2560+63
      L=L-2560;
      ЗаписатьБелыйКодДля(2560);
    }
    if(L > 63) {
      L2=МаксимальныйСостКодМеньшеL(L);
      L=L-L2;
      ЗаписатьБелыйКодДля(L2);
    }
    ЗаписатьБелыйКодДля(L);
    //Это всегда код завершения
  }
  else {
    [Код аналогичный белой серии,
с той разницей, что записываются
черные коды]
  }
}
// Окончание строки изображения
}
```

Поскольку черные и белые серии чередуются, то реально код для белой и код для черной серии будут работать попеременно.

В терминах регулярных выражений мы получим для каждой строки нашего изображения (достаточно длинной, начинающейся с белой точки) выходной битовый поток вида:

$$((\langle B-2560 \rangle)^* \langle B-сст. \rangle \langle B-зв. \rangle (\langle C-2560 \rangle)^* \langle C-сст. \rangle \langle C-зв. \rangle)^+ \\ [(\langle B-2560 \rangle)^* \langle B-сст. \rangle \langle B-зв. \rangle],$$

где $()^*$ повтор 0 или более раз, $()^+$ — повтор 1 или более раз, $[]$ — включение 1 или 0 раз.

Для приведенного ранее примера: 0, 3, 556, 10... алгоритм сформирует следующий код: <Б-0><Ч-3><Б-512><Б-44><Ч-10>, или, согласно таблице, 001101011001100101001011010000100 (разные коды в потоке выделены для удобства). Этот код обладает свойством префиксных кодов и легко может быть свернут обратно в последовательность длин серий. Легко подсчитать, что для приведенной строки в 569 бит мы получили код, длиной в 33 бита, т. е. коэффициент сжатия составляет примерно 17 раз.

Заметим, что единственное “сложное” выражение в алгоритме: $L2 = \text{МаксимальныйДопКодМеньше}L(L)$ — на практике работает очень просто: $L2 = (L \gg 6) * 64$, где \gg — побитовый сдвиг L влево на 6 битов (можно сделать то же самое за одну побитовую операцию $\&$ — логическое И).

Приведенные ниже таблицы построены с помощью классического алгоритма Хаффмана (отдельно для черных и белых длин серий). Значения вероятностей появления конкретных длин серий были получены путем анализа большого количества факсимильных изображений.

Если в одном столбце встретятся два числа с одинаковым префиксом, то это опечатка.

Этот алгоритм реализован в формате TIFF.

Характеристики алгоритма CCITT Group 3

Коэффициенты компрессии. Лучший коэффициент стремится в пределе к 213 (3), средний 2, в худшем случае увеличивает файл в 5 раз.

Класс изображений. Двухцветные черно-белые изображения, в которых преобладают большие пространства, заполненные белым цветом.

Симметричность. Близка к 1.

Характерные особенности. Данный алгоритм чрезвычайно прост в реализации, быстр и может быть легко реализован аппаратно.

Таблица кодов завершения:

Длина серии	Код белой подстроки	Код черной подстроки
0	00110101	0000110111
1	00111	010
2	0111	11
3	1000	10
4	1011	011
5	1100	0011
6	1110	0010
7	1111	00011
8	10011	000101
9	10100	000100
10	00111	0000100
11	01000	0000101
12	001000	0000111
13	000011	00000100
14	110100	00000111
15	110101	000011000
16	101010	0000010111
17	101011	0000011000
18	0100111	0000001000
19	0001100	00001100111
20	0001000	00001101000
21	0010111	00001101100
22	0000011	00000110111
23	0000100	00000101000
24	0101000	00000010111
25	0101011	00000011000
26	0010011	000011001010
27	0100100	000011001011
28	0011000	000011001100
29	00000010	000011001101
30	00000011	000001101000
31	00011010	000001101001

Длина серии	Код белой подстроки	Код черной подстроки
32	00011011	000001101010
33	00010010	000001101011
34	00010011	000011010010
35	00010100	000011010011
36	00010101	000011010100
37	00010110	000011010101
38	00010111	000011010110
39	00101000	000011010111
40	00101001	000001101100
41	00101010	000001101101
42	00101011	000011011010
43	00101100	000011011011
44	00101101	000001010100
45	00000100	000001010101
46	00000101	000001010110
47	00001010	000001010111
48	00001011	000001100100
49	01010010	000001100101
50	01010011	000001010010
51	01010100	000001010011
52	01010101	000000100100
53	00100100	000000110111
54	00100101	000000111000
55	01011000	000000100111
56	01011001	000000101000
57	01011010	000000101100
58	01011011	000000101101
59	01001010	000000101011
60	01001011	000000101100
61	00110010	000000101101
62	00110011	000001100110
63	00110100	000001100111

Таблица составных кодов:

Длина серии	Код белой подстроки	Код черной подстроки
64	11011	0000001111
128	10010	000011001000
192	01011	000011001001
256	0110111	000001011011
320	00110110	000000110011
384	00110111	000000110100
448	01100100	000000110101
512	01100101	0000001101100
576	01101000	0000001101101
640	01100111	0000001001010
704	011001100	0000001001011
768	011001101	0000001001100
832	011010010	0000001001101
896	011010011	000000110010
960	011010100	0000001110011
1024	011010101	0000001110100
1088	011010110	0000001110101
1152	011010111	0000001110110
1216	011011000	0000001110111
1280	011011001	0000001010010

Длина серии	Код белой подстроки	Код черной подстроки
1344	011011010	0000001010011
1408	011011011	0000001010100
1472	010011000	0000001010101
1536	010011001	0000001011010
1600	010011010	0000001011011
1664	011000	0000001100100
1728	010011011	0000001100101
1792	00000001000	совп. с белой
1856	00000001100	— // —
1920	00000001101	— // —
1984	000000010010	— // —
2048	000000010011	— // —
2112	000000010100	— // —
2176	000000010101	— // —
2240	000000010110	— // —
2304	000000010111	— // —
2368	000000011100	— // —
2432	000000011101	— // —
2496	000000011110	— // —
2560	000000011111	— // —

1.3. Алгоритм JBIG

Алгоритм разработан группой экспертов ISO (Joint Bi-level Experts Group) специально для сжатия однобитных черно-белых изображений [5]. Например, факсов или отсканированных документов. В принципе может применяться и к 2- и к 4-битовым картинкам. При этом алгоритм разбивает их на отдельные битовые плоскости. JBIG позволяет управлять такими параметрами, как порядок разбиения изображения на битовые плоскости, ширина полос в изображении, уровни масштабирования. Последняя возможность позволяет легко ориентироваться в базе больших по размерам изображений, просматривая сначала их уменьшенные копии. Настраивая эти параметры, можно использовать описанный выше эффект "огрубленного изображения" при получении изображения по сети или по любому другому каналу, пропускная способность

которого мала по сравнению с возможностями процессора. Распаковываться изображение на экране будет постепенно, как бы медленно “проявляясь”. При этом человек начинает анализировать картинку задолго до конца процесса разархивации.

Алгоритм построен на базе Q-кодировщика [6], патентом на который владеет IBM. Q-кодер так же, как и алгоритм Хаффмана, использует для чаще появляющихся символов короткие цепочки, а для реже появляющихся — длинные. Однако, в отличие от него, в алгоритме используются и последовательности символов.

1.4. Алгоритм JPEG

JPEG — один из самых новых и достаточно мощных алгоритмов. Практически он является стандартом де-факто для полноцветных изображений [1]. Оперирует алгоритм областями 8×8 , на которых яркость и цвет меняются сравнительно плавно. Вследствие этого при разложении матрицы такой области в двойной ряд по косинусам (см. формулы ниже) значимыми оказываются только первые коэффициенты. Таким образом, сжатие в JPEG осуществляется за счет плавности изменения цветов в изображении.

Алгоритм разработан группой экспертов в области фотографии специально для сжатия 24-битных изображений. JPEG — Joint Photographic Expert Group — подразделение в рамках ISO — Международной организации по стандартизации. Название алгоритма читается [‘jei’peg]. В целом алгоритм основан на дискретном косинусоидальном преобразовании (в дальнейшем ДКП), применяемом к матрице изображения для получения некоторой новой матрицы коэффициентов. Для получения исходного изображения применяется обратное преобразование.

ДКП раскладывает изображение по амплитудам некоторых частот, таким образом, при преобразовании мы получа-

ем матрицу, в которой многие коэффициенты либо близки, либо равны нулю. Кроме того, система цветового восприятия человека слабо распознает определенные частоты. Поэтому можно аппроксимировать некоторые коэффициенты более грубо без заметной потери качества изображения. Для этого используется квантование коэффициентов (quantization). В самом простом случае — это арифметический побитовый сдвиг вправо. При этом преобразовании теряется часть информации, но могут достигаться большие коэффициенты сжатия.

Принцип работы алгоритма

Пусть сжимается 24-битное изображение.

Шаг 1. Переводим изображение из цветового пространства RGB, с компонентами, отвечающими за красную (Red), зеленую (Green) и синюю (Blue) составляющие цвета точки, в цветовое пространство YCrCb (иногда называют YUV).

В нем Y — яркостная составляющая, а Cr, Cb — компоненты, отвечающие за цвет (хроматический красный и хроматический синий). За счет того, что человеческий глаз менее чувствителен к цвету, чем к яркости, появляется возможность архивировать массивы для Cr и Cb компонент с большими потерями и, соответственно, большими коэффициентами сжатия. Подобное преобразование уже давно используется в телевидении. На сигналы, отвечающие за цвет, там выделяется более узкая полоса частот.

Упрощенно перевод из цветового пространства RGB в цветовое пространство YCrCb можно представить так:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.5 & -0.4187 & -0.0813 \\ 0.1687 & -0.3313 & 0.5 \end{pmatrix} \times \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}. \quad (1.11)$$

Обратное преобразование осуществляется умножением вектора YUV на обратную матрицу.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1.402 \\ 1 & -0.34414 & -0.71414 \\ 1 & 1.772 & 0 \end{pmatrix} \times \begin{pmatrix} Y & 0 \\ Cb & -128 \\ Cr & 128 \end{pmatrix}. \quad (1.12)$$

Шаг 2. Разбиваем исходное изображение на матрицы 8×8 . Формируем из каждой три рабочие матрицы ДКП — по 8 бит отдельно для каждой компоненты. При больших коэффициентах сжатия, этот шаг выполняется чуть сложнее. Изображение делится по компоненте Y — как и в первом случае, а для компонент Cr и Cb матрицы набираются через строчку и через столбец. То есть из исходной матрицы размером 16×16 получается только одна рабочая матрица ДКП. При этом, как нетрудно заметить, теряется $3/4$ полезной информации о цветовых составляющих изображения и получаем сразу сжатие в два раза. Так можно поступать благодаря работе в пространстве $YCrCb$. На результирующем RGB изображении, как показала практика, это сказывается не сильно.

Шаг 3. Применяя ДКП к каждой рабочей матрице получаем матрицу, в которой коэффициенты в левом верхнем углу соответствуют низкочастотной составляющей изображения, а в правом нижнем — высокочастотной.

В упрощенном виде это преобразование можно представить так:

$$Y[u, v] = \frac{1}{4} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} C(i, u) \times C(j, v) \times y[i, j], \quad (1.13)$$

где
$$C(i, u) = A(u) \times \cos\left(\frac{(2xi+1) \times i \times \pi}{2 \cdot n}\right), \quad (1.14)$$

$$A(u) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{for } u \equiv 0 \\ 1, & \text{for } u \neq 0 \end{cases}. \quad (1.15)$$

Шаг 4. Производим квантование. В принципе это просто деление рабочей матрицы на матрицу квантования поэлемент-

но. Для каждой компоненты (Y, U и V) в общем случае задается своя матрица квантования $q[u, v]$ (далее — МК).

$$Yq[u, v] = \text{IntegerRound} \left(\frac{Y[u, v]}{q[u, v]} \right). \quad (1.16)$$

На этом шаге осуществляется управление степенью сжатия, и происходят самые большие потери. Понятно, что, задавая МК с большими коэффициентами, получаем больше нулей и, следовательно, большую степень сжатия.

В стандарт JPEG включены рекомендованные МК, построенные опытным путем. Матрицы для большего или меньшего коэффициентов сжатия получают путем умножения исходной матрицы на некоторое число γ .

С квантованием связаны и специфические эффекты алгоритма. При больших значениях коэффициента γ потери в низких частотах могут быть настолько велики, что изображение распадется на квадраты 8×8 . Потери в высоких частотах могут проявиться в так называемом “эффекте Гиббса”, когда вокруг контуров с резким переходом цвета образуется своеобразный “нимб”.

Шаг 5. Переводим матрицу 8×8 в 64-элементный вектор при помощи “зигзаг”-сканирования, т. е. берем элементы с индексами (0, 0), (0, 1), (1, 0), (2, 0)... Это показано на рис. 1.7.

Таким образом, в начале вектора получаем коэффициенты матрицы, соответствующие низким частотам, а в конце — высоким.

$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$	$a_{0,4}$	$a_{0,5}$	$a_{0,6}$	$a_{0,7}$
$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{1,5}$	$a_{1,6}$	$a_{1,7}$
$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,0}$			
$a_{3,0}$	$a_{3,0}$	$a_{3,0}$	$a_{3,0}$				
$a_{4,0}$	$a_{4,1}$	$a_{4,2}$					
$a_{5,0}$	$a_{5,1}$						
$a_{6,0}$	$a_{6,1}$						
$a_{7,0}$	$a_{7,1}$						

Рис. 1.7. Алгоритм “зигзаг”-сканирования

Шаг 6. Свертывая вектор с помощью алгоритма группового кодирования получаем пары типа (пропустить, число), где “пропустить” является счетчиком пропускаемых нулей, а “число” — значение, которое необходимо поставить в следующую ячейку. Так, вектор 42 3 0 0 0 -2 0 0 0 0 1... будет свернут в пары (0, 42) (0, 3) (3, -2) (4, 1)...

Шаг 7. Свертываем получившиеся пары кодированием по Хаффману с фиксированной таблицей.

Процесс восстановления изображения в этом алгоритме полностью симметричен. Метод позволяет сжимать некоторые изображения в 10—15 раз без серьезных потерь.

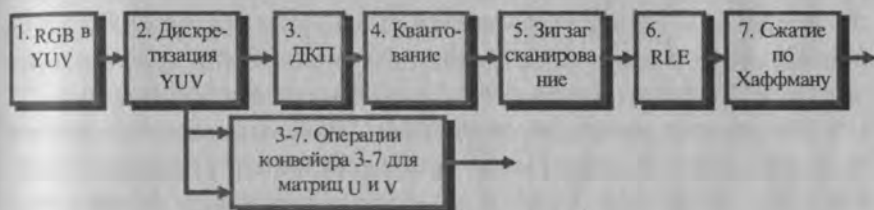


Рис. 1.8. Конвейер операций, используемый в алгоритме JPEG

Существенными положительными сторонами алгоритма является то, что задается степень сжатия; выходное цветное изображение может иметь 24 бита на точку.

Отрицательными сторонами алгоритма является то, что при повышении степени сжатия изображение распадается на отдельные квадраты (8×8). Это связано с тем, что происходят большие потери в низких частотах при квантовании, и восстановить исходные данные становится невозможно.

Проявляется эффект Гиббса — ореолы по границам резких переходов цветов.

Как уже говорилось, стандартизован JPEG относительно недавно — в 1991 году. Но уже тогда существовали алгоритмы, сжимающие сильнее при меньших потерях качества. Дело в том, что действия разработчиков стандарта были ог-

раничены мощностью существовавшей на тот момент техники. То есть даже на персональном компьютере алгоритм должен был работать меньше минуты на среднем изображении, а его аппаратная реализация должна быть относительно простой и дешевой. Алгоритм должен был быть симметричным (время разархивации примерно равно времени архивации).

Последнее требование сделало возможным появление таких бытовых изделий, как цифровые фотоаппараты, — устройства размером с небольшую видеокамеру, снимающие 24-битовые фотографии на 10—20 Мбитную флэш-карту с интерфейсом PCMCIA. Потом эта карта вставляется в разъем на портативном компьютере и соответствующая программа позволяет считать изображения. Не правда ли, если бы алгоритм был несимметричен, было бы неприятно долго ждать, пока аппарат “перезарядится” — сожмет изображение.

Не очень приятным свойством JPEG является также то, что нередко горизонтальные и вертикальные полосы на дисплее абсолютно не видны и могут проявиться только при печати в виде муарового узора. Он возникает при наложении наклонного раstra печати на горизонтальные и вертикальные полосы изображения. Из-за этих сюрпризов JPEG не рекомендуется активно использовать в полиграфии, задавая высокие коэффициенты. Однако при архивации изображений, предназначенных для просмотра человеком, он на данный момент незаменим.

Широкое применение JPEG долгое время сдерживалось, пожалуй, лишь тем, что он оперирует 24-битными изображениями. Поэтому для того, чтобы с приемлемым качеством посмотреть картинку на обычном мониторе в 256-цветной палитре, требовалось применение соответствующих алгоритмов и, следовательно, определенное время. В приложениях, ориентированных на придирчивого пользователя, таких, например, как игры, подобные задержки неприемлемы. Кроме того, если имеющиеся у вас изображения, допустим, в 8-битном формате GIF перевести в 24-битный JPEG, а потом обратно

в GIF для просмотра, то потеря качества произойдет дважды при обоих преобразованиях. Тем не менее выигрыш в размерах архивов зачастую настолько велик (в 3—20 раз!), а потери качества настолько малы, что хранение изображений в JPEG оказывается очень эффективным.

Несколько слов необходимо сказать о модификациях этого алгоритма. Хотя JPEG и является стандартом ISO, формат его файлов не был зафиксирован. Пользуясь этим, производители используют свои несовместимые между собой форматы и, следовательно, могут изменить алгоритм. Так, внутренние таблицы алгоритма, рекомендованные ISO, заменяются ими на свои собственные. Кроме того, легкая неразбериха присутствует при задании степени потерь. Например, при тестировании выясняется, что “отличное” качество, “100%” и “10 баллов” дают существенно различающиеся картинки. При этом, кстати, “100%” качества не означает сжатие без потерь. Встречаются также варианты JPEG для специфических приложений.

Как стандарт ISO JPEG начинает все шире использоваться при обмене изображениями в компьютерных сетях. Поддерживается алгоритм JPEG в форматах Quick Time, PostScript Level 2, Tiff 6.0 и, на данный момент занимает видное место в системах мультимедиа.

Характеристики алгоритма JPEG:

Коэффициенты компрессии. 2—200 (задается пользователем).

Класс изображений. Полноцветные 24 битные изображения, или изображения в градациях серого без резких переходов цветов (фотографии).

Симметричность. 1.

Характерные особенности. В некоторых случаях алгоритм создает “ореол” вокруг резких горизонтальных и вертикальных границ в изображении (эффект Гиббса). Кроме того, при высокой степени сжатия изображение распадается на блоки 8×8 пикселей.

1.5. Фрактальный алгоритм сжатия изображений

История фрактального сжатия (25)

Рождение фрактальной геометрии обычно связывают с выходом в 1977 г. книги Б. Мандельброта "Фрактальная геометрия природы". Одна из основных идей книги заключалась в том, что средствами традиционной геометрии (то есть используя линии и поверхности), чрезвычайно сложно представить природные объекты. Фрактальная геометрия задает их очень просто. В 1981 г. Джон Хатчинсон опубликовал статью "Фракталы и самоподобие", в которой была представлена теория построения фракталов с помощью системы итерируемых функций (IPS, Iterated Function System). Четыре года спустя появилась статья Майкла Барнсли и Стефана Демко, в которой приводилась уже достаточно стройная теория IPS. В 1987 г. Барнсли основал Iterated Systems, компанию, основной деятельностью которой является создание новых алгоритмов и ПО с использованием фракталов. Всего через год, в 1988 г., он выпустил фундаментальный труд "Фракталы повсюду". Помимо описания IPS, в ней был получен результат, известный сейчас как Collage Theorem, который лежит в основе математического обоснования идеи фрактальной компрессии. Если построение изображений с помощью фрактальной математики можно назвать прямой задачей, то построение по изображению IPS — это обратная задача. Довольно долго она считалась неразрешимой, однако Барнсли, используя Collage Theorem, построил соответствующий алгоритм. (В 1990 и 1991 гг. эта идея была защищена патентами.) Если коэффициенты занимают меньше места, чем исходное изображение, то алгоритм является алгоритмом архивации.

Первая статья об успехах Барнсли в области компрессии появилась в журнале BYTE в январе 1988 г. В ней не описывалось решение обратной задачи, но приводилось несколько изображений, сжатых с коэффициентом 1:10 000, что было совершенно ошеломительно. Но практически сразу было отмечено, что, несмотря на броские названия ("Темный лес",

“Побережье Монтере”, “Поле подсолнухов”), изображения в действительности имели искусственную природу. Это вызвало массу скептических замечаний, подогреваемых еще и заявлением Барнсли о том, что “среднее изображение требует для сжатия порядка 100 часов работы на мощной двухпроцессорной рабочей станции, причем с участием человека”. Отношение к новому методу изменилось в 1992 г., когда Арнауд Джеквин, один из сотрудников Барнсли, при защите диссертации описал практический алгоритм и опубликовал его. Этот алгоритм был крайне медленным и не претендовал на компрессию в 10000 раз (полноцветное 24-разрядное изображение с его помощью могло быть сжато без существенных потерь с коэффициентом 1:8 — 1:50); но его несомненным достоинством было то, что вмешательство человека удалось полностью исключить. Сегодня все известные программы фрактальной компрессии базируются на алгоритме Джеквина. В 1993 г. вышел первый коммерческий продукт компании Iterated Systems. Ему было посвящено достаточно много публикаций, но о коммерческом успехе речь не шла, продукт был достаточно “сырой”, компания не предпринимала никаких рекламных шагов, и приобрести программу было тяжело. В 1994 г. Ювал Фишер был предоставлен во всеобщее пользование исходные тексты исследовательской программы, в которой использовалось разложение изображения в квадродерево и были реализованы алгоритмы оптимизации поиска. Позднее появилось еще несколько исследовательских проектов, которые в качестве начального варианта программы использовали программу Фишера. В июле 1995 г. в Тронхейме (Швеция) состоялась первая школа-конференция, посвященная фрактальной компрессии. Таким образом, многие важные события в области фрактальной компрессии произошли за последние три года: алгоритм только-только начинает развиваться.

Идея метода

Идея метода фрактальной архивации основана на том, что мы представляем изображение в более компактной фор-

ме — с помощью коэффициентов системы итерируемых функций (Iterated Function System — далее по тексту как IFS). Прежде чем рассматривать сам процесс архивации, разберем, как IFS строит изображение, т. е. процесс декомпрессии.

Строго говоря, IFS представляет собой набор трехмерных аффинных преобразований, в нашем случае переводящих одно изображение в другое. Преобразованию подвергаются точки в трехмерном пространстве (координата x , координата y , яркость).

Наиболее наглядно этот процесс продемонстрировал Барнсли в своей книге “Fractal Image Compression”. Там введено понятие Фотокопировальной Машины, состоящей из экрана, на котором изображена исходная картинка, и системы линз, проецирующих изображение на другой экран:

- ◆ Линзы могут проецировать часть изображения произвольной формы в любое другое место нового изображения.

- ◆ Области в которые проецируются изображения не пересекаются.

- ◆ Линза может менять яркость и уменьшать контрастность.

- ◆ Линза может зеркально отражать и поворачивать свой фрагмент изображения.

- ◆ Линза должна (масштабировать) уменьшать свой фрагмент изображения.

Расставляя линзы и меняя их характеристики, мы можем управлять получаемым изображением. Одна итерация работы Машины заключается в том, что по исходному изображению с помощью проектирования строится новое, после чего новое берется в качестве исходного. Утверждается, что в процессе итераций мы получим изображение, которое перестанет изменяться. Оно будет зависеть только от расположения и характеристик линз и не будет зависеть от исходной картинки. Это изображение называется “неподвижной точкой” или аттрактором данной IFS. Соответствующая теория гарантирует наличие ровно одной неподвижной точки для каждой IFS.

Поскольку отображение линз является сжимающим, каждая линза в явном виде задает самоподобные области в нашем изображении. Благодаря самоподобию получаем сложную структуру изображения при любом увеличении. Таким образом, интуитивно понятно, что система итерируемых функций задает *фрактал* (не строго — самоподобный математический объект).

Наиболее известны два изображения, полученных с помощью IFS: “треугольник Серпинского” и “папоротник Барнсли”. “Треугольник Серпинского” задается тремя, а “папоротник Барнсли” четырьмя аффинными преобразованиями (или, в нашей терминологии, “линзами”). Каждое преобразование кодируется буквально считанными байтами, в то время как изображение, построенное с их помощью, может занимать и несколько мегабайт.

Из вышесказанного становится понятно, как работает архиватор, и почему ему требуется так много времени. Фактически, фрактальная компрессия — это поиск самоподобных областей в изображении и определение для них параметров аффинных преобразований.

В худшем случае, если не будет применяться оптимизирующий алгоритм, потребуется перебор и сравнение всех возможных фрагментов изображения разного размера. Даже для небольших изображений при учете дискретности мы

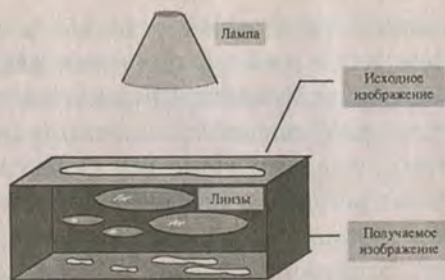


Рис. 1.9. Машина Барнсли



Рис. 1.10. Папоротник Барнсли

получим астрономическое число перебираемых вариантов. Причем даже резкое сужение классов преобразований, например, за счет масштабирования только в определенное количество раз, не дает заметного выигрыша во времени. Кроме того, при этом теряется качество изображения. Подавляющее большинство исследований в области фрактальной компрессии сейчас направлены на уменьшение времени архивации, необходимого для получения качественного изображения.

Приведем основные определения и теоремы, на которых базируется фрактальная компрессия. Этот материал более детально и с доказательствами рассматривается в [3] и [4].

Определение. Преобразование $w : R^2 \rightarrow R^2$, представленное в виде

$$w(\bar{x}) = w \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \bullet \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}, \quad (1.17)$$

где a, b, c, d, e, f действительные числа и $(x \ y) \in R^2$ называется *двумерным аффинным преобразованием*.

Определение. Преобразование $w : R^3 \rightarrow R^3$, представленное в виде

$$w(\bar{x}) = w \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a & b & t \\ c & d & u \\ r & s & p \end{pmatrix} \bullet \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} e \\ f \\ q \end{pmatrix}, \quad (1.18)$$

где $a, b, c, d, e, f, p, q, r, s, t, u$ действительные числа и $(x \ y \ z) \in R^3$ называется *трехмерным аффинным преобразованием*.

Определение. Пусть $f : X \rightarrow X$ — преобразование в пространстве (X, d) . Точка $x_f \in X$ такая, что называется *неподвижной точкой (аттрактором)* преобразования.

Определение. Преобразование $f : X \rightarrow X$ в метрическом пространстве (X, d) называется *сжимающим*, если существует число s : $0 \leq s < 1$, такое, что

$$d(f(x), f(y)) \leq s \cdot d(x, y) \quad \forall x, y \in X.$$

Замечание: Формально мы можем использовать любое сжимающее отображение при фрактальной компрессии, но реально используются лишь трехмерные аффинные преобразования с достаточно сильными ограничениями на коэффициенты.

Теорема. (О сжимающем преобразовании)

Пусть $f : X \rightarrow X$ в полном метрическом пространстве (X, d) . Тогда существует в точности одна неподвижная точка $x_f \in X$ этого преобразования и для любой точки $x \in X$, последовательность $\{f^n(x) : n=0, 1, 2, \dots\}$ сходится к x_f .

Более общая формулировка этой теоремы гарантирует нам сходимость.

Определение. Изображением называется функция S , определенная на единичном квадрате и принимающая значения от 0 до 1 или $S(x, y) \in [0..1] \forall x, y \in [0..1]$.

Пусть трехмерное аффинное преобразование $w : R^3 \rightarrow R^3$ записано в виде

$$w_i(\bar{x}) = w_i \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & p \end{pmatrix} \bullet \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} e \\ f \\ q \end{pmatrix} \quad (1.19)$$

и определено на компактном подмножестве R_i декартова квадрата $[0..1] \times [0..1]$. Тогда оно переведет часть поверхности S в область D_i , расположенную со сдвигом (e, f) и поворотом, заданным матрицей

$$\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (1.20)$$

При этом если интерпретировать значение S как яркость соответствующих точек, то она уменьшится в p раз (преобразование обязано быть сжимающим) и изменится на сдвиг q .

Определение. Конечная совокупность W сжимающих трехмерных аффинных преобразований, определенных на облас-

тях R_i , таких, что $w_i(R_i) = D_i$ и $D_i \cap D_j = \emptyset \quad \forall i \neq j$ называется *системой итерируемых функций (IFS)*.

Системе итерируемых функций однозначно сопоставляется неподвижная точка — изображение. Таким образом, процесс компрессии заключается в поиске коэффициентов системы, а процесс декомпрессии — в проведении итераций системы для получения неподвижной точки. Области R_i в дальнейшем будут именоваться ранговыми, а области D_i — доменными.

Построение алгоритма

Таким образом, основной задачей при компрессии фрактальным алгоритмом является нахождение соответствующих аффинных преобразований. В самом общем случае можно переводить любые по размеру и форме области изображения, однако в этом случае получается астрономическое число перебираемых вариантов разных фрагментов, которое невозможно обработать на текущий момент даже на суперкомпьютере.

В *учебном варианте алгоритма*, изложенном далее, сделаны следующие ограничения на области:

1. Все области являются квадратами со сторонами параллельными сторонам изображения. Это ограничение достаточно жесткое. Фактически мы собираемся приближать все многообразие геометрических фигур лишь квадратами.

2. При переводе ранговой области в доменную уменьшение размеров производится ровно в два раза. Это существенно упрощает как компрессор, так и декомпрессор, так как задача масштабирования небольших областей является нетривиальной.

3. Все доменные блоки — квадраты и имеют фиксированный размер. Изображение равномерной сеткой разбивается на набор доменных блоков.

4. Ранговые области берутся “через точку” и по X и по Y , что сразу уменьшает перебор в 4 раза.

5. При переводе ранговой области в доменную поворот куба возможен только на 0° , 90° , 180° или 270° . Также допускается зеркальное отражение. Общее число возможных преобразований (считая пустое) 8.

6. Масштабирование (сжатие) по вертикали (яркости) осуществляется в фиксированное число раз — в 0,75.

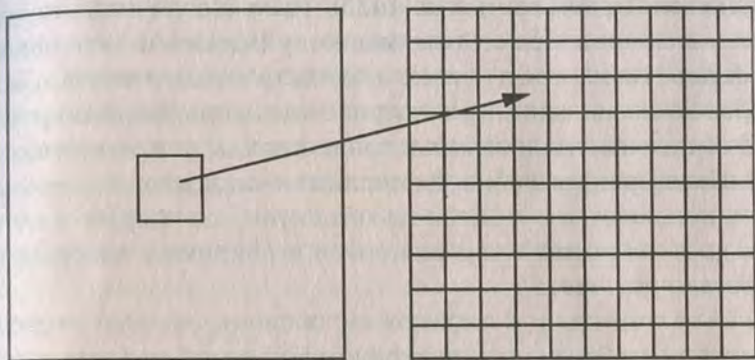


Рис. 1.11. Области сжатия

Эти ограничения позволяют:

1. Построить алгоритм, для которого требуется сравнительно малое число операций даже на достаточно больших изображениях.

2. Очень компактно представить данные для записи в файл. Нам требуется на каждое аффинное преобразование в IFS:

- ◆ два числа, для того чтобы задать смещение рангового блока. Если мы ограничим входные изображения размером 512×512 , то достаточно будет по 8 бит на каждое число.

- ◆ три бита, для того чтобы задать преобразование симметрии при переводе рангового блока в доменный.

- ◆ 7—9 бит, для того чтобы задать сдвиг по яркости при переводе.

Информацию о размере блоков можно хранить в заголовке файла. Таким образом мы затратили менее 4 байт на одно аффинное преобразование. В зависимости от того, каков размер блока, можно высчитать, сколько блоков будет в изображении. Таким образом, мы можем получить оценку степени компрессии.

Например, для файла в градациях серого 256 цветов 512×512 пикселей при размере блока 8 пикселей аффинных преобразований будет 4096 ($512/8 \cdot 512/8$). На каждое потребу-

ется 3.5 байта. Следовательно, если исходный файл занимал 262144 (512·512) байт (без учета заголовка), то файл с коэффициентами будет занимать 14336 байт. Коэффициент архивации — 18 раз. При этом мы не учитываем, что файл с коэффициентами тоже может обладать избыточностью и архивироваться методом архивации без потерь, например LZW.

Отрицательные стороны предложенных ограничений:

1. Поскольку все области являются квадратами невозможно воспользоваться подобием объектов, по форме далеких от квадратов (которые встречаются в реальных изображениях достаточно часто.)

2. Аналогично мы не сможем воспользоваться подобием объектов в изображении, коэффициент подобия между которыми сильно отличается от 2.

3. Алгоритм не сможет воспользоваться подобием объектов в изображении, угол между которыми не кратен 90° .

Такова плата за скорость компрессии и за простоту упаковки коэффициентов в файл.

Сам алгоритм упаковки сводится к перебору всех доменных блоков и подбору для каждого соответствующего ему рангового блока. Ниже приводится схема этого алгоритма.

```
for (all domain blocks) {
    min_distance = MaximumDistance;
    Dij = image->CopyBlock(i, j);
    for (all range blocks) { // С поворотами и отр.
        current=Координаты текущего преобразования;
        R=image->CopyBlock(current);
        current_distance = Dij.L2distance(R);
        if(current_distance < min_distance) {
            // Если коэффициенты best хуже:
            min_distance = current_distance;
            best = current;
        }
        Save_Coefficients_to_file(best);
    } //Next range
} //Next domain
```

Как видно из приведенного алгоритма, для каждого доменного блока производится его проверка со всеми возможными ранговыми блоками (в том числе с прошедшими преобразование симметрии), находим вариант с наименьшей мерой L_2 (наименьшим среднеквадратичным отклонением) и сохраняем коэффициенты этого преобразования в файл. Коэффициенты — это (1) координаты найденного блока, (2) число от 0 до 7, характеризующее преобразование симметрии (поворот, отражение блока), и (3) сдвиг по яркости для этой пары блоков. Сдвиг по яркости вычисляется как:

$$q = \left[\sum_{i=1}^n \sum_{j=1}^n d_{ij} - \sum_{i=1}^n \sum_{j=1}^n r_{ij} \right] / n^2, \quad (1.20)$$

где r_{ij} — значения пикселей рангового блока (R), а d_{ij} — значения пикселей доменного блока (D). При этом мера считается как:

$$d(R, D) = \sum_{i=1}^n \sum_{j=1}^n (0.75r_{ij} + q - d_{ij})^2. \quad (1.21)$$

Вычисление квадратного корня из L_2 меры и деления ее на n не производится, поскольку данные преобразования монотонны и не мешают поиску экстремума. Однако при этом можно выполнять на две операции меньше для каждого блока.

Посчитаем количество операций, необходимых для сжатия изображения в градациях серого 256 цветов 512×512 пикселей при размере блока 8 пикселей:

Часть программы	Число операций
for (all range blocks)	4096 (=512/8·512/8)
for (all range blocks) + symmetry transformation	492032 (= (512/2-8) * (512/2-8) * 8)
Вычисление q и $d(R, D)$	> 3*64 операций “+” > 2*64 операций “.”
Итого:	> 3* 128983236608 операций “+” > 2* 128983236608 операций “.”

Таким образом, нам удалось уменьшить число операций алгоритма компрессии до вполне вычисляемых (пусть и за несколько часов) величин.

Схема алгоритма декомпрессии изображений

Декомпрессия алгоритма фрактального сжатия чрезвычайно проста. Необходимо провести несколько итераций трехмерных аффинных преобразований, коэффициенты которых были получены на этапе компрессии.

В качестве начального может быть взято абсолютно любое изображение (например, абсолютно черное), поскольку соответствующий математический аппарат гарантирует нам сходимость последовательности изображений, получаемых в ходе итераций IFS к неподвижному изображению (близкому к исходному). Обычно для этого достаточно 16 итераций.

```
Прочитаем из файла коэффициенты всех блоков;  
Создадим черное изображение нужного размера;  
Until(изображение не станет неподвижным){  
  For(every domain (D)){  
    R=image->CopyBlock(R_coord_for_D);  
    For(every output pixel(i,j) in the domain){  
       $D_{ij} = R_{ij} + oD$ ;  
    } //Next pixel  
  } //Next domain  
} //Until end
```

Поскольку запись коэффициентов для блоков D_{ij} (которые, как мы оговорили, в нашем частном случае являются квадратами одинакового размера) производится последовательно, то получается, что *последовательно* заполняется изображение по квадратам сетки разбиения использованием аффинного преобразования.

Как можно легко подчитать, количество операций на один пиксел изображения при восстановлении необычайно мало, благодаря чему декомпрессия изображений для фрактального алгоритма проходит гораздо быстрее декомпрессии, например, для алгоритма JPEG.

Оценка потерь и способы их регулирования

При кратком изложении упрощенного варианта алгоритма были пропущены многие важные вопросы. Например, что делать, если алгоритм не может подобрать для какого-либо фрагмента изображения подобный ему? Достаточно очевидное решение — разбить этот фрагмент на более мелкие и попытаться поискать для них. В то же время понятно, что эту процедуру нельзя повторять до бесконечности, иначе количество необходимых преобразований станет так велико, что алгоритм перестанет быть алгоритмом компрессии. Следовательно, априорно допускаются потери в какой-то части изображения.

Для фрактального алгоритма компрессии, как и для других алгоритмов сжатия с потерями, очень важны механизмы, с помощью которых можно будет регулировать степень сжатия и степень потерь. К настоящему времени разработан достаточно большой набор таких методов. Во-первых, можно ограничить количество аффинных преобразований, заведомо обеспечив степень сжатия не ниже фиксированной величины. Во-вторых, можно потребовать, чтобы в ситуации, когда разница между обрабатываемым фрагментом и наилучшим его приближением будет выше определенного порогового значения, этот фрагмент дробился обязательно (для него обязательно заводятся несколько “линз”). В-третьих, можно запретить дробить фрагменты размером меньше, допустим, четырех точек. Изменяя пороговые значения и приоритет этих условий, мы будем очень гибко управлять коэффициентом компрессии изображения, в диапазоне от побитового соответствия, до любой степени сжатия. Причем эта гибкость будет гораздо выше, чем у ближайшего “конкурента” — алгоритма JPEG.

Рекурсивный (волновой) алгоритм

Английское название рекурсивного сжатия — wavelet. На русский язык оно также переводится как волновое сжатие и как сжатие с использованием всплесков. Этот вид ар-

хивации известен довольно давно и напрямую исходит из идеи использования когерентности областей. Ориентирован алгоритм на цветные и черно-белые изображения с плавными переходами и идеален для картинок типа рентгеновских снимков. Коэффициент сжатия задается и варьируется в пределах 5—100 раз. При попытке задать больший коэффициент, на резких границах, особенно проходящих по диагонали, проявляется “лестничный эффект” — ступеньки разной яркости, размером в несколько пикселей.

Идея алгоритма заключается в том, что мы сохраняем в файл разницу число между средними значениями соседних блоков в изображении, которая обычно принимает значения, близкие к 0.

Так два числа a_{2i} и a_{2i+1} всегда можно представить в виде $b_i^1 = (a_{2i} + a_{2i+1})/2$ и $b_i^2 = (a_{2i} - a_{2i+1})/2$. Аналогично последовательность a_i может быть попарно переведена в последовательность b_i^1, b_i^2 .

Разберем конкретный пример: пусть мы сжимаем строку из 8 значений яркости пикселей (a_i): (220, 211, 212, 218, 217, 214, 210, 202). Мы получим следующие последовательности b_i^1 и b_i^2 : (215.5, 215, 215.5, 206) и (4.5, -3, 0.5, 4). Заметим, что значения b_i^2 достаточно близки к 0. Повторим операцию, рассматривая b_i^1 как a_i . Данное действие выполняется как бы рекурсивно, откуда и название алгоритма. Мы получим из (215.5, 215, 215.5, 206): (215.25, 210.75) (0.25, 4.75). Полученные коэффициенты, округлив до целых и сжав, например, с помощью алгоритма Хаффмана с фиксированными таблицами, мы можем поместить в файл.

Заметим, что мы применяли наше преобразование к цепочке только два раза. Реально мы можем позволить себе применение wavelet — преобразования 4—6 раз. Более того, дополнительное сжатие можно получить, используя таблицы алгоритма Хаффмана с неравномерным шагом (т. е. придется сохранять код Хаффмана для ближайшего в таблице значения). Эти приемы позволяют достичь заметных коэффициентов сжатия.

Алгоритм для двумерных данных реализуется аналогично. Если имеется квадрат из 4 точек с яркостями $a_{2i, 2j}$, $a_{2i+1, 2j}$, $a_{2i, 2j+1}$, и $a_{2i+1, 2j+1}$, то

$$\begin{aligned} b_{i,j}^1 &= (a_{2i,2j} + a_{2i+1,2j} + a_{2i,2j+1} + a_{2i+1,2j+1})/4 \\ b_{i,j}^2 &= (a_{2i,2j} + a_{2i+1,2j} - a_{2i,2j+1} - a_{2i+1,2j+1})/4 \\ b_{i,j}^3 &= (a_{2i,2j} - a_{2i+1,2j} + a_{2i,2j+1} - a_{2i+1,2j+1})/4 \\ b_{i,j}^4 &= (a_{2i,2j} - a_{2i+1,2j} - a_{2i,2j+1} + a_{2i+1,2j+1})/4 \end{aligned} \quad (1.22)$$

Используя эти формулы, для изображения 512×512 пикселей получим после первого преобразования 4 матрицы размером 256×256 элементов.

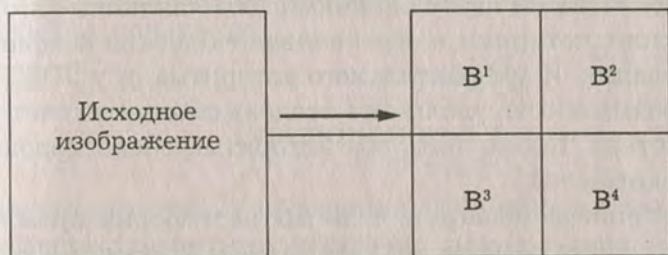


Рис. 1.12. Деление исходного изображения на 4 матрицы

В первой, как легко догадаться, будет храниться уменьшенная копия изображения. Во второй — усредненные разности пар значений пикселей по горизонтали. В третьей — усредненные разности пар значений пикселей по вертикали. В четвертой — усредненные разности значений пикселей по диагонали. По аналогии с двумерным случаем мы можем повторить наше преобразование и получить вместо первой матрицы 4 матрицы размером 128×128 . Повторив наше преобразование в третий раз, мы получим в итоге: 4 матрицы 64×64 , 3 матрицы 128×128 и 3 матрицы 256×256 . На практике, при

записи в файл, значениями, получаемыми в последней строке ($b_{i,j}^4$), обычно пренебрегают (сразу получая выигрыш примерно на треть размера файла — $1-1/4-1/16-1/64\dots$).

К достоинствам этого алгоритма можно отнести легкость реализации постепенного “проявления” изображения при передаче изображения по сети. Кроме того, поскольку в начале изображения мы фактически храним его уменьшенную копию, это упрощает показ “огрубленного” изображения по заголовку.

Сравнение с JPEG

Сегодня наиболее распространенным алгоритмом архивации графики является JPEG. Сравним его с фрактальной компрессией. Во-первых, заметим, что и тот, и другой алгоритм оперируют 8-битными (в градациях серого) и 24-битными полноцветными изображениями. Оба являются алгоритмами сжатия с потерями и обеспечивают близкие коэффициенты архивации. И у фрактального алгоритма, и у JPEG существует возможность увеличить степень сжатия за счет увеличения потерь. Кроме того, оба алгоритма очень хорошо распараллеливаются.

Различия начинаются, если мы рассмотрим время, необходимое алгоритмам для архивации/разархивации. Так, фрактальный алгоритм сжимает в сотни и даже в тысячи раз дольше, чем JPEG. Распаковка изображения, наоборот, произойдет в 5—10 раз быстрее. Поэтому если изображение будет сжато только один раз, а передано по сети и распаковано множество раз, то выгодней использовать фрактальный алгоритм.

JPEG использует разложение изображения по косинусоидальным функциям, поэтому потери в нем (даже при заданных минимальных потерях) проявляются в волнах и ореолах на границе резких переходов цветов. Именно за этот эффект его не любят использовать при сжатии изображений, которые готовят для качественной печати: там этот эффект может стать очень заметен. Фрактальный алгоритм избавлен от

этого недостатка. Более того, при печати изображения каждый раз приходится выполнять операцию масштабирования, поскольку растр (или линиятура) печатающего устройства не совпадает с растром изображения. При преобразовании также может возникнуть несколько неприятных эффектов, с которыми можно бороться либо масштабируя изображение программно (для дешевых устройств печати типа обычных лазерных и струйных принтеров), либо снабжая устройство печати своим процессором, винчестером и набором программ обработки изображений (для дорогих фотонаборных автоматов). Как можно догадаться, при использовании фрактального алгоритма таких проблем практически не возникает. Вытеснение JPEG фрактальным алгоритмом в повсеместном использовании произойдет еще не скоро (хотя бы в силу низкой скорости архивации последнего), однако в области приложений мультимедиа, в компьютерных играх его использование вполне оправдано.

Заключение

Сделаем некоторые обобщения. С одной стороны, приведенные выше алгоритмы достаточно универсальны и покрывают все типы изображений, с другой — у них, по современным меркам, слишком маленький коэффициент архивации. Используя один из алгоритмов сжатия без потерь, можно обеспечить коэффициент архивации изображения примерно в два раза. В то же время алгоритмы сжатия с потерями оперируют с коэффициентами 10—200 раз. Помимо возможности модификации изображения, одна из основных причин подобной разницы заключается в том, что традиционные алгоритмы ориентированы на работу с цепочкой. Они не учитывают так называемую «когерентность областей» в изображениях. Идея когерентности областей заключается в малом изменении цвета и структуры на небольшом участке изображения. Все алгоритмы, о которых речь пойдет ниже, были со-

зданы позднее специально для сжатия графики и используют эту идею.

Справедливости ради следует отметить, что и в классических алгоритмах можно использовать идею когерентности. Существуют алгоритмы обхода изображения по "фрактальной" кривой, при работе которых оно также вытягивается в цепочку; но за счет того, что кривая обегает области изображения по сложной траектории, участки близких цветов в получающейся цепочке удлиняются.

Литература

1. Методы передачи изображений, сокращение избыточности / Под ред. У. К. Прэтта; Пер. с англ.; Под ред. Л. С. Вилленкина. — М.: Радио и связь. 1983.
2. Харатишвили Н. Г. Цифровое кодирование с предсказанием непрерывных сигналов. — М.: Радио и связь, 1986.
3. Ахмед Н., Рао К. Р. Ортогональные преобразования при обработке цифровых сигналов / Пер. с англ.; Под ред. И. Б. Фоменко. — М.: Связь. 1980.
4. Синепол В. С., Цикин И. А. Системы компьютерной видеоконференцсвязи. — М.: Мобильные коммуникации, 1999.
5. Рауччи Р., Фразер Б. Средства сжатия изображений для работы со сканером // Мир ПК. 1992. № 4. С. 35—45.
6. Киволович П. Сжатие изображений по стандарту JPEG // Мир ПК. 1992. № 4. С. 46—51.
7. Ансон Л., Барнсли М. Фрактальное сжатие изображений // Мир ПК. 1992. № 4. С. 52—58
8. Сейттер И. Сжатие данных // Мир ПК. 1991. № 2. С. 46—59.
9. Ахмед Н., Рао К. Р. Ортогональные преобразования при обработке цифровых сигналов/Пер. с англ.; Под ред. И. Б. Фоменко. — М.: Связь, 1980.
10. Кричевский Р. Е. Сжатие и поиск информации. — М.: Радио и связь, 1989.

11. Wallace G. K. The JPEG still picture compression standard. Commun. ACM 34, 4, 1991. P. 30—45.
12. Le Gall D. L. MPEG-. A video compression standard for multimedia applications. Commun. ACM 34, 4, 1991. P. 46—58.
13. Liou M. L. Overview of the px64 kbit/s video coding standard. Commun. ACM 34, 4, 1991, pp. 60—63.
14. Lippman A. Feature sets for interactive images. Commun. ACM 34, 4, 1991. P. 92—102.
15. Hamey K., Keith M., Lavelle G., Kyan L. and Stark. D. The 1750 video processor: A total multimedia solution. Commun. ACM 34, 4, 1991. P. 64—78.
16. Hoffert E. M. and Gretsch G. The digital news system at EDUCOM: A convergence of interactive computing, newspapers, television and high-speed networks. Commun. ACM 34.4., 1991. P. 113—116.
17. Barnslev M. F. and Sloan A. D. Fractal image compression. In Proceedings of the Scientific Data Compression Workshop (NASA, 1988).
18. Lu M., Chen Ch. A Haffman-type Code Generator. IEEE Trans. on ASSP V. 38. N 9. P. 1619—1626.
- 19.
20. Ahmed N., Rao K. R., Discrete Cosine Transform IEEE Trans. Computers C-23, 1974, (I). 90—93.
21. Pennebaker, W. B., Mitchell, J. L., et. al. Arithmetic coding articles. IBM J. Res. Dev. 32, 6 Special Issue. Nov. 1988. P. 717—774.
22. Conway J., Sloane N. J. Fast Quantizing and algorithms for Lattice Quantizers and Codes. IEEE Trans. 1982, V. II—28, № 2. P. 227—232.
23. Special issue quantization. IEEE Trans 1982, V. II—28, № 2. P. 128—261.
24. Goldbery M., Bochen P. Image Compression using Adaptive Vector Quantization. EEE Trans. 1986, V. Com—34, № 2. P. 180—187.
25. Епаничников В. Ю., Цуккерман И. И. Подавление структурных помех при адаптивном групповом кодировании ТВ-изображений. Техника кино и телевидения. 1997. № 11. С. 17—22.

Глава 2

СЖАТИЕ АУДИОДАНЫХ АЛГОРИТМОМ MPEG-1

2.1. Краткий обзор алгоритмов сжатия MPEG Audio

Алгоритм MPEG-1 Audio (ISO 11172—3) ориентирован на кодирование высококачественного стерео звука. В настоящее время для доставки аудиоданных все известные DBS системы используют Уровень II (Layer II) MPEG-1 Аудио со скоростью потока 192 кбит/с.

MPEG-1 Аудио обеспечивает большое количество дополнительных опций для частот дискретизации, выходной скорости потока и имеет три уровня. Уровни MPEG-1 Аудио не представляют из себя уровни в том смысле, что один основан на другом. Точнее, смысл уровней в том, что декодер большего уровня может декодировать звук, кодированный декодером более низкого уровня.

2.1.1. Кодирование

До появления MPEG-1 Аудио сжатие звука сводилось к удалению статистической избыточности из электронного аналога акустических волн. MPEG-1 Аудио позволяет достичь дополнительного сжатия путем удаления ненужных элементов звука при использовании психоакустических фонем как во временной, так и в спектральной областях.

Основной принцип работы MPEG-1 аудио кодера состоит в следующем. Входные отсчеты аудио сигнала подаются на вход кодера. В DBS эти отсчеты следуют с частотой 48 кГц (для каждого из стерео каналов), поэтому на протяжении всей этой главы будет использоваться именно эта частота. Динамический диапазон каждого отсчета 16 бит. Упаковка

(mapping) создает фильтрационное и субполосное представление входного потока. В Слое 2 (Layer-2) 32 этих упакованных отсчета каждого канала называются отсчетами субполос (субполосными отсчетами, subband samples).

Параллельно психоакустическая модель (Psychoacoustic Model) осуществляет вычисления, контролирующие квантование и кодирование. Оценки границ (пороговых значений) маскирования используются для осуществления этого контроля квантования. Распределение бит в субполосах рассчитывается на основе соотношений сигнал/маска всех субполос. Максимальный уровень сигнала и минимальная граница маскирования выводятся из быстрого преобразования Фурье (FFT transform) дискретизированного входного сигнала.

2.1.2. Декодирование

В основе структуры декодера лежит процесс декодирования, обратный процессу кодирования. Сначала поток распаковывается и происходит разделение потоков основных данных и служебных (дополнительных) данных. Затем, если в процессе кодирования применялся флаг `error_check`, происходит проверка наличия ошибок. Поток данных затем распаковывается для восстановления различных информационных частей. Особенности параметров аудио потока MPEG-1 для DBS систем приведены в табл. 2.1.

Таблица 2.1

Параметр	Значение
Частота дискретизации	48 кГц для каждой пары стереосигналов
Скорость потока	192 кбит/сек
Режимы стерео	Joint stereo/intensity stereo
Психоакустическая модель	Психоакустическая модель 1

2.2. Описание кодированного аудиопотока

Основной составной блок в MPEG-1 Аудио стандарте — аудиокадр (Audio Frame). Аудиокадр показывает длительность

ность, эквивалентную количеству отсчетов разделенному на скорость отсчетов. Для DBS используется уровень 2 (Layer 2), имеющий 1152 отсчета, следующие с частотой 48 кГц. Период кадра в этом случае равен:

$$\frac{1152 \frac{\text{отсчетов}}{\text{кадр}}}{48000 \frac{\text{отсчетов}}{\text{секунду}}} = 24 \text{ миллисекунды на кадр.}$$

Аудиокадры объединяются для формирования аудиопоследовательности.

Аудиокадр — это пакет фиксированной длины в пакетированном потоке. В Уровне 2 он состоит из 1152 отсчетов для каждого из двух входящих стерео сигналов плюс перерыв в битах, содержащихся в пакете. Каждый пакет в общей сложности имеет 4608 бит. Он начинается со слова “синхронизация” и заканчивается байтом перед следующим словом синхронизации.

Каждый аудиокадр состоит из четырех независимых частей:

1. Все пакеты в пакетированном потоке содержат заголовок, обеспечивающий синхронизацию и перенос остальной информации, требующейся для декодирования потока.

2. **Error_check** (проверка ошибок) содержит 16 бит, которые могут обнаружить ошибки в процессе хранения данных или передачи.

3. **Audio_data** (аудиоданные) — полезная нагрузка каждого пакета. Содержит кодированные отсчеты и информацию, необходимую для их декодирования.

4. **Ancillary_data** (дополнительные данные) содержат биты, которые могут быть добавлены пользователем.

2.2.1. Заголовок

Информация, содержащаяся в первых 32 битах (4 байта) аудиокадра, является заголовком аудиокадра. Заголовок состоит из следующих составляющих:

Двенадцатибитная строка синхронизации **syncword** 0xFFF. 2-битный указатель уровня (для уровня 2 его значение равно "10").

Однобитный флаг **protection_bit** (защитный бит), показывающий, где могут быть добавлены биты коррекции ошибок.

4-битный параметр **bitrate_index** (индекс скорости потока). В случае DBS систем значение этого параметра всегда равно "1010" (192 Кбит/с) для стереосигнала.

Двубитный параметр **sampling_frequency** (частота отсчетов) показывает частоту следования отсчетов аудиосигнала. Для DBS систем его значение равно "01" (48 кГц) для стерео.

Параметр **mode** (режим) показывает режим работы канала. Для DBS систем его значение равно "01" (интенсивное стерео).

2.2.2. error_chek (проверка ошибок)

error_chek — это опциональная часть потока, которая содержит циклическую проверку на избыточность (Cyclical Redundancy Check, **crc_check**), оно представляет из себя 16-битное слово для проверки на четность. Оно обеспечивает обнаружение ошибок в кодированном потоке.

2.2.3. audio_data (аудио данные)

audio_data — это секция для полезной нагрузки в пакете данных. Она содержит кодированные отсчеты аудиосигнала и информацию о том, каким образом эти пакеты должны быть декодированы.

2.2.4. ancillary_data (вспомогательные данные)

Поток MPEG-1 Audio содержит элементы резервирования для доставляемых пользователю данных. Количество используемых вспомогательных бит (**no_of_ancillary_bits**) должно быть вычтено из общего количества бит в кадре. Так как длина кадра постоянна, то это вычитание возможно для кодирования отсчетов аудио и может иметь сильное влияние на качество аудио.

2.3. Детальное описание кодирования

MPEG-1 Audio алгоритм называют психоакустическим алгоритмом. В следующих частях описаны четыре основных части кодера.

2.3.1. Банк фильтров

Банки фильтров обеспечивает дискретизацию частотного спектра для каждого из 32 входящих отсчетов. Эти банки фильтров сильно дискретизированы. Это значит, что там столько отсчетов в анализируемой (частотной) области, сколько и во временной области. Банк фильтров для кодера называется “Банк анализирующих фильтров” (Analysis Filterbank). В декодере восстанавливающие фильтры называются “Банк синтезирующих фильтров” (Synthesis Filterbank).

В слое 2 (Layer 2) MPEG-1 Audio используется банк фильтров с 32 субполосами. В каждой субполосе 36 отсчетов, которые сгруппированы в три группы по 12 отсчетов каждая.

Начальный фильтр верхних частот

Алгоритмы кодирования обеспечивают частотный отклик вплоть до D. С. Рекомендуется ставить на вход кодера фильтр верхних частот для того, чтобы он не пропускал частоты в диапазоне 2—10 Гц. Основной задачей этого фильтра является удаление из процесса кодирования звуков, которые не могут быть услышаны человеческим ухом.

Банк анализирующих фильтров

Данный банк фильтров используется для разделения широкополосного сигнала с частотой дискретизации f_s на 32 субполосы с частотами дискретизации $f_s/32$. Алгоритм этого процесса с соответствующими формулами дан в следующем ниже его описании.

Алгоритм работы банка анализирующих фильтров

1. Перемещение 32 новых отсчетов аудио сигнала в регистр сдвига. Это показано на рис. 2.1.

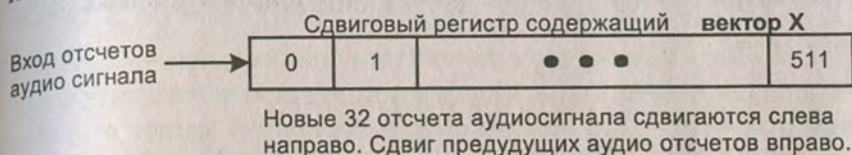


Рис. 2.1. Сдвиг аудио отсчетов

2. Умножение каждого значения вектора X на соответствующий коэффициент.

$$Z_i = C_i * X_i \quad i = 0, 1, \dots, 511 \quad (2.1)$$

3. Затем формируется 64-компонентный вектор Y .

$$Y_i = \sum_{j=0}^7 Z_{i+64*j}$$

$$Y_0 = Z_0 + Z_{64} + \dots + Z_{448}$$

$$Y_1 = Z_1 + Z_{65} + \dots + Z_{449} \quad (2.2)$$

...

$$Y_{63} = Z_{63} + Z_{127} + \dots + Z_{511}$$

4. Затем формируются выходные отклики субполос.

$$\begin{bmatrix} S_0 \\ S_1 \\ \cdot \\ \cdot \\ S_{31} \end{bmatrix} = \begin{bmatrix} M_{0,0} & M_{0,1} & \cdot & \cdot & \cdot & M_{0,63} \\ M_{1,0} & M_{1,1} & \cdot & \cdot & \cdot & M_{1,63} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ M_{31,0} & M_{31,1} & \cdot & \cdot & \cdot & M_{31,63} \end{bmatrix} * \begin{bmatrix} Y_0 \\ Y_1 \\ \cdot \\ \cdot \\ Y_{63} \end{bmatrix} \quad (2.3)$$

$$M_{ik} = \cos\{[(2*i + 1)*(k - 16)*\pi]/64\}, \text{ для } i = 0 \dots 31 \text{ и } k = 0 \dots 63. \quad (2.4)$$

На рис. 2.2 показаны значения D_i , используемые в синтезирующем субполосном фильтре (Synthesis Subband Filter).

Значения C_i , используемые в вычислении формулы (2.1), получены путем деления значений коэффициентов D на 32 для каждого индекса. В стандарте MPEG значения C и D коэффициентов представлены в соответствующих таблицах.

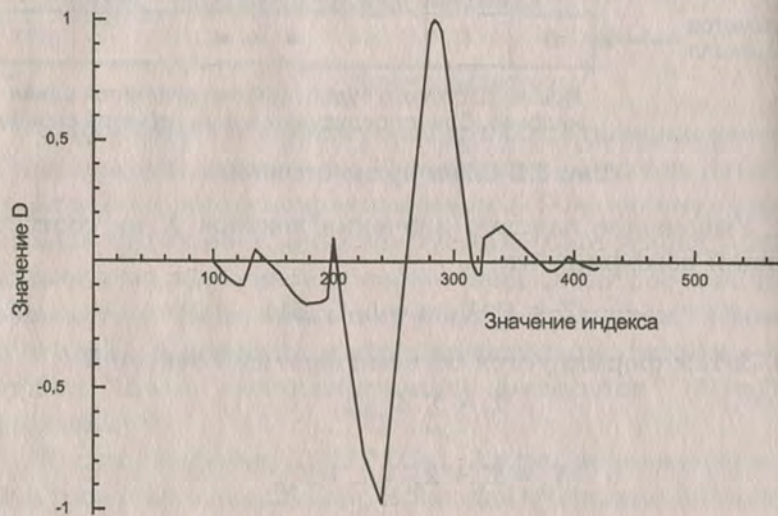


Рис. 2.2. Значения коэффициентов D

Психоакустическая модель

В психоакустической модели 1 FFT имеет окно из 1024 отсчетов. Несмотря на то что это число меньше, чем количество отсчетов в кадре, это не имеет сильного влияния на качество.

Конечным результатом работы психоакустической модели является следующее отношение сигнал—маска:

$$SMR_{sb}(n) = L_{sb}(n) - LT_{min}(n) \text{ dB} \quad (2.5)$$

Это отношение вычисляется для каждой субполосы n . $L_{sb}(n)$ — это уровень энергии сигнала (Signal Pressure Level, SPL) в каждой полосе. $LT_{min}(n)$ — это минимальный уровень, при котором происходит маскирование шума полезным сигналом в каждой субполосе n .

Вычисление уровня энергии сигнала

Формула (2.5) требует вычисления SPL $L_{sb}(n)$ в субполосе n . $L_{sb}(n)$ определяется по формуле:

$$L_{sb}(n) = \text{MAX}[X(k), \{20 \cdot \log[\text{scf}_{\text{max}}(n) \cdot 32768] - 10\}] \text{ dB} \quad (2.6)$$

$X(k)$ — это уровень энергии звука с k -ым индексом FFT, имеющим максимальную амплитуду в диапазоне частот, соответствующем субполосе n . Выражение $\text{scf}_{\text{max}}(n)$ — это максимум из трех масштабных коэффициентов субполосы n в пределах аудиокадра. Уровень энергии сигнала $L_{sb}(n)$ вычисляется для каждой субполосы n .

Следующим шагом в расчете выражения (2.5) является вычисление значения $LT_{\text{min}}(n)$, что является более сложным процессом. Минимальный уровень маскирования $LT_{\text{min}}(n)$ в субполосе n определяется следующим выражением:

$$LT_{\text{min}}(n) = \text{Min}[LT_g(i)] \text{ dB} . \quad (2.7)$$

$f(i)$ в субполосе n

Вычисление выражения (2.7) требует определения значения $LT_g(i)$ (глобальной пороговой величины маскирования для каждого i -го частотного отсчета).

$$LT_g(i) = 10 \log_{10} \left[10^{LT_g(i)/10} + \sum_{j=1}^m 10^{LT_{tm}[z(i), z(j)]/10} + \sum_{j=1}^n 10^{LT_{nm}[z(i), z(j)]/10} \right], \quad (2.8)$$

где $LT_q(i)$ — это абсолютная граничная величина, $LT_t_m[z(j), z(i)]$ — отдельная граница маскирования для тоновых маскировщиков, $LT_{nm}[z(j), z(i)]$ — отдельная граница маскирования для нетоновых маскировщиков.

Определение абсолютных границ (граничных значений)

Когда присутствует тон с определенной амплитудой и частотой, другие тоновые сигналы или шумы около этой частоты не могут быть услышаны человеческим ухом. Максимальный уровень низко амплитудного сигнала, который незаметен ухом, называется граничным значением (границей)

Параметры FFT (быстрого преобразования Фурье)

Параметры Layer 2	Значения
Длина преобразования	1024 отсчета
Размер окна при $f_s = 48$ кГц	21,3 мс
Частотное разрешение	<code>sampling_frequency/1024</code>
Окно Ханна (Hann window), $h(i)$:	
$h(i) = \sqrt{8/3} * 0.5 * \{1 - \cos[(2 * \pi * i) / N]\}$	$0 \leq i < N - 1$
Спектр плотности мощности, $X(k)$:	
$X(k) = 10 * \log_{10} \left[\frac{1}{N} \sum_{l=0}^{N-1} h(l) * s(l) * e^{-j * k * l * 2 * \pi / N} \right]^2$,	$k = 0 \dots N/2$
где $s(l)$ — входящий сигнал, $j = \sqrt{-1}$	

маскирования (Masking Threshold). В субполосе есть граничное значение, являющееся функцией частоты, лежащее ниже тона или шума, неслышимое и не зависящее от того, присутствует ли в субполосе тон или шум. Это граничное значение называется абсолютной границей (Absolute Threshold) и показано на рис. 2.3.

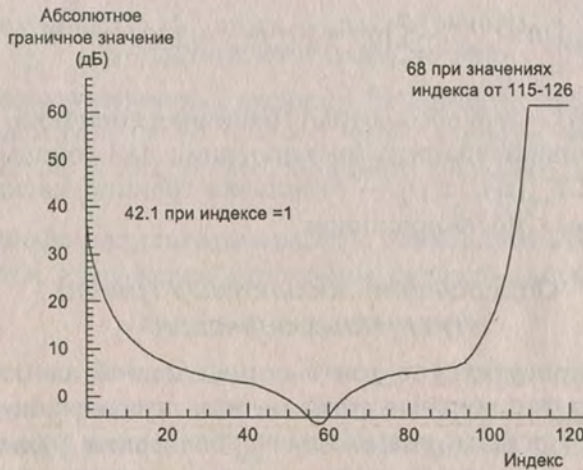


Рис 2.3. Абсолютные границы

Абсолютная граница обозначается как $LT_q(k)$. Ее часто называют бесшумной границей (threshold in quiet). В стандарте ISO Таблица D.1f дает набор значений индексов и частот для критических полос (Critical Bands), скоростей для критических полос (Critical Bands Rates, CBR) и абсолютных границ. Значения индексов лежат в диапазоне от 1 до 126. Для большинства приложений значения этих параметров могут быть определены при помощи кривых, показанных на рис. 2.3, 2.4 и 2.5.

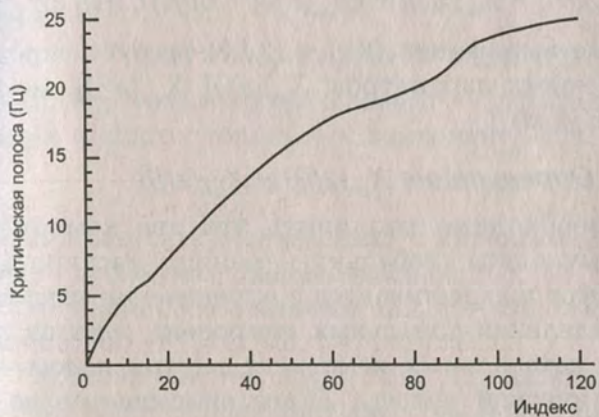


Рис. 2.4. Полоса критических частот

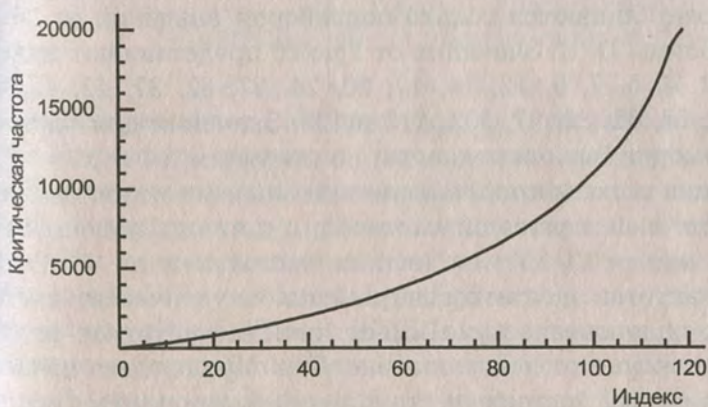


Рис. 2.5. Зависимость критической частоты от индекса

Определение отдельных границ маскирования

Тональные маскировщики близки к синусоидам и имеют спектр похожий на импульсы. С другой стороны, нетональные маскировщики имеют сложные характеристики. Отдельные границы маскировки для тональных и нетональных компонент определяются следующими выражениями:

$$LT_{tm}[z(j), z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j), z(i)] \text{ dB} \quad (2.9)$$

$$LT_{nm}[z(j), z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j), z(i)] \text{ dB} \quad (2.10)$$

Вычисление выражений (2.9) и (2.10) требует определения еще пяти новых параметров: $X_{tm}[z(j)]$, $X_{nm}[z(j)]$, $av_{tm}[z(j)]$, $av_{nm}[z(j)]$ и $vf[z(j), z(i)]$.

Определение $X_{tm}(z(j))$ и $X_{nm}(z(j))$

Сначала необходимо различить эти две компоненты, прежде чем вычислять глобальную границу маскировки из FFT спектра. Этот шаг начинается с определения локального максимума, выделения тональных компонент и вычисления интенсивности нетональных компонент внутри полосы пропускания критической полосы. Далее описывается то, как могут быть определены границы критической полосы.

В табл. D.2f Стандарта ISO даны эти значения, которые в свою очередь являются только поднабором значений от 26 до 126 из таблицы D. 1f. Значения от 1 до 26 представляют индексы 1, 2, 3, 4, 5, 7, 9, 12, 14, 17, 20, 24, 27, 32, 37, 42, 49, 53, 59, 65, 73, 77, 82, 89, 97, 103, 113 и 126. Значения для частоты и полосы могут быть определены по рисункам 2.4 и 2.5.

Ширина полосы пропускания критических полос изменяется вместе с центральной частотой, с ширины полосы пропускания около 0,1 кГц на низких частотах и до 4 кГц на высоких частотах. Это известно из психоакустических экспериментов, человеческое ухо имеет лучшее частотное разрешение в низкочастотном диапазоне. Для определения локального максимума, например, тональной компоненты, исследуется частотный диапазон (df) вокруг этого локального мак-

симула. Для частот до 3 кГц $df = 93.75$ Гц. Для частот от 3 кГц до 6 кГц $df = 140.63$ Гц. Для частот от 6 кГц до 12 кГц $df = 281.25$ Гц. Для частот от 12 кГц до 24 кГц $df = 562.5$ Гц.

Для формирования набора спектральных линий $X(k)$ для тональных и не тональных компонент выполняются следующие пять операций:

1. Селекция возможных вариантов локальных максимумов — спектральная линия (составляющая) $X(k)$, определяется как локальный максимум при выполнении следующего условия:

$$X(k) > X(k-1) \text{ и } X(k) > = X(k+1). \quad (2.11)$$

2. Выбор тональных компонент — локальный максимум, вносится в список тональных компонент, если

$$X(k) - X(k+j) > = 7 \text{ дБ}, \quad (2.12)$$

где j выбирается в соответствии с индексами, относящимися к данной субполосе. Для индексов от 2 до 63 $j = \pm 2$; для индексов больше 63, но меньше 127, $j = \pm 2, \pm 3$; для индексов от 127 до 255 $j = \pm 2, \pm 3, \pm 4, \pm 5, \pm 6$.

3. Уровень энергии звука SPL. Для каждого k , для которого $X(k)$ было определено как тональная компонента, вычисляется следующее выражение:

$$X_{tm}(k) = 10 * \log_{10} \left[10 \frac{X(k-1)}{10} + 10 \frac{X(k)}{10} + 10 \frac{X(k+1)}{10} \right]. \quad (2.13)$$

Затем значения всех спектральных составляющих внутри проверенного диапазона устанавливается равным ∞ дБ.

4. Нетональные компоненты. Нетональные компоненты (шум) вычисляются по остальным спектральным составляющим. Для вычисления этих компонент из спектральных составляющих $X(k)$ с помощью рис. 2.5 определяются критические полосы $z(k)$. В Слое 2 MPEG-1 при частоте дискретизации используется 26 критических полос.

5. Вычисление энергии. Внутри каждой критической полосы энергия спектральных составляющих (после того как

тональные компоненты были обнулены) суммируется для определения SPL новых нетональных компонент $X_{nm}(k)$ соответствующих данной критической полосе.

Выемка тональных и нетональных маскирующих компонент

Выемка — это процедура, которая используется для уменьшения количества маскировщиков, что весьма обосновано для вычисления глобальной границы маскирования (Global Masking Threshold). В следующем ниже описании Bark — это единица критической полосы.

$X_{tm}(k)$ и $X_{nm}(k)$ учитываются при вычислении границы маскировки только в том случае, если они больше чем $LT_q(k)$ для каждого индекса k .

Выемка двух или более тональных компонент происходит в диапазоне менее чем 0.5 Bark: Сохраняется $X_{tm}[z(j)]$ с наибольшей энергией и удаляется наименьшая компонента (или компоненты) из перечня тональных компонент. Для этой операции в области критической полосы используется скользящее окно (sliding window) шириной 0.5 Bark.

В этой операции используется элемент av , называемый индексом маскировки, и его разница для тональной и нетональной компонент (av_{tm} и av_{nm}). Индексы маскировки av_{tm} и av_{nm} определены эмпирическим путем. Для тональных маскировщиков он определяется следующим образом:

$$av_{tm} = -1.525 - 0.275 \cdot z(j) - 4.5 \text{ дБ} \quad (2.14)$$

Для нетональных компонент:

$$av_{nm} = -1.525 - 0.175 \cdot z(j) - 0.5 \text{ дБ} \quad (2.15)$$

Элемент vf , называемый функцией маскировки компоненты маскировки $X_{tm}[z(j)]$, характеризуемая различными углами наклонами кривой (низким и высоким), которые зависят от интервала в Bark-ах, $dz = z(i) - z(j)$. В этом выражении i — это индекс спектральной линии для которой вычисляется функция маскировки, а j — индекс маски. CBR $z(j)$ и $z(i)$ могут быть найдены по графику на рис. 2.6.

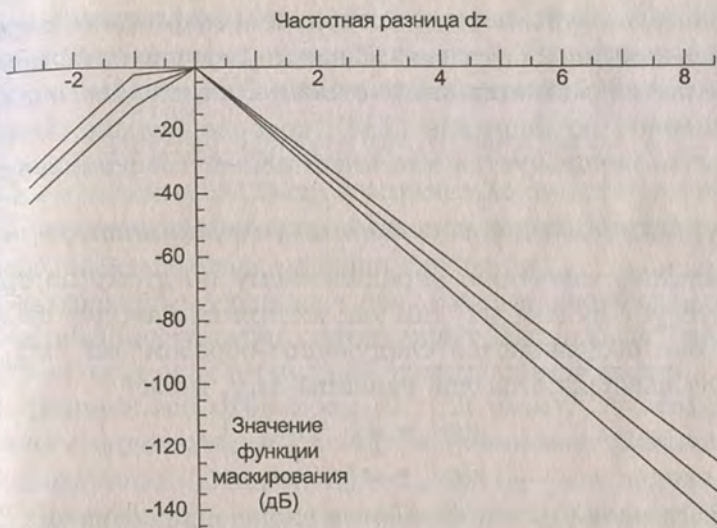


Рис. 2.6. График функции маскирования для $X[z(j)] = -10, 0, 10$

2.3.2. Распределение бит

Процесс распределения бит между кодированными параметрами должен максимально удовлетворять требования для скорости потока и требования к маскированию. Если имеется недостаточное количество бит, то они должны использоваться в порядке из расположения. В слое 2 MPEG-1 этот метод называется процессом распределения бит (т. е. назначение количества бит, предназначенных для каждого отсчета или группы отсчетов в каждой субполосе).

Вычисление масштабного коэффициента.

Масштабные коэффициенты могут храниться в таблице (см. табл. В. 1 стандарта ISO) или вычисляться. Для вычисления значения масштабного коэффициента для каждого индекса n используется число 1.2599210498949:

$$\frac{2.00000000000000}{1.2599210498949^n}, \quad (2.16)$$

где n может принимать значения от 0 до 62.

В каждой субполосе масштабный коэффициент вычисляется для каждых 12 отсчетов. В каждой группе из 12 отсчетов определяется максимальный отсчет. Наименьшее значение, вычисленное по формуле (2.16), которое больше, чем этот максимум, используется как масштабный коэффициент.

Кодирование масштабных коэффициентов

Каждому значению определенному по этому алгоритму соответствует индекс scf , так как их три на каждую субполосу, то они обозначаются следующим образом: scf_1 , scf_2 , scf_3 . Сначала вычисляются две разницы $dscf_1$ и $dscf_2$:

$$dscf_1 = scf_1 - scf_2 \quad (2.17)$$

$$dscf_2 = scf_2 - scf_3 \quad (2.18)$$

Класс разницы определяется следующим образом:

- ◆ 1 класс: $dscf \leq -3$;
- ◆ 2 класс: $3 < dscf < 0$;
- ◆ 3 класс: $dscf = 0$;
- ◆ 4 класс: $0 < dscf < 3$;
- ◆ 5 класс: $dscf \geq 3$.

Так как в данном случае присутствует два $dscf$, то им соответствует два набора классов.

Кодирование информации отбора масштабных коэффициентов

Для субполос с ненулевым распределением бит информация отбора масштабных коэффициентов ($scfsi$) кодируется 2-битным словом.

Распределение бит

До начала процесса распределения бит должно быть определено число бит (adb), которое имеется в распоряжении для кодирования полезных данных. Это число может быть получено путем вычитания числа бит необходимых для заголовка ($bhdr$, 32 бита), слова проверки CRC ($bcrc$, 16 бит), числа

бит для передачи информации о распределении бит ($bbal$) и числа бит необходимых для дополнительных данных ($banc$) из общего количества имеющихся в распоряжении бит (cb):

$$adb = cb - (bhdr + bcrc + bbal + banc) \quad (2.19)$$

Из выражения (2.19) определяется количество имеющих-ся в распоряжении для кодирования полезной нагрузки в виде отсчетов и масштабных коэффициентов бит.

Квантование в субполосе: эти данные (значения порогов квантования, *quantization value*) могут быть представлены в табличном виде или вычислены следующим образом:

1. Первые три субполосы (0, 1,2) имеют 15 возможных значений с индексами от 1 до 15. Значения могут быть вычислены следующим образом: $2^{n+1}-1$, где i — это индекс.

2. Следующие 8 субполос (3—10) также имеют 15 возможных значений. Индексам 1, 2, 3 и 4 соответствуют значения 3, 5, 7 и 9. Значения, соответствующие индексам 5—14, вычисляются по формуле 2^{i-1} . Индексу 15 соответствует значение 65535.

3. Субполосы 11—22 имеют только 7 индексов. Индексам 1—6 соответствуют значения 3, 5, 7, 9, 15 и 31 соответственно. Индексу 7 соответствует значение 65535.

4. Субполосы 23—26 имеют только три индекса. Им соответствуют значения 3, 5 и 65535.

5. Субполосы не имеют порогов квантования.

Параметр $nbal$ показывает количество бит требующееся для определения количества занятых бит. Соответственно, $nbal = 4$ для первых 10 субполос, 3 для субполос 11—22 и 2 для субполос 22—26. Количество бит, требующееся для представления этих квантованных отсчетов, может быть вычислено по алгоритму для “Классов квантования Слоя 2” (*Layer 2 Classes of Quantization*). Распределение является итерационной процедурой. Основное правило — это определение субполосы, которая выигрывает, и увеличение числа бит, распределенных в этой субполосе.

Для каждой субполосы отношение маска—шум (mask-to-noise ratio, MNR) вычисляется путем вычитания отношения сигнал-маска (signal-to-mask ratio, SMR) из отношения сигнал-шум (signal-to-noise ratio, SNR). Минимальное соотношение MNR вычисляется для каждой субполосы.

$$MNR = SNR - SMR . \quad (2.20)$$

Так как эти величины измеряются в децибелах, то выражение (2.20) эквивалентно делению SNR на SMR:

$$\frac{\frac{S}{N}}{\frac{S}{M}} = \frac{M}{N} = MNR . \quad (2.21)$$

Соотношение сигнал—шум (SNR) может быть найдено по графикам на рис. 2.7. и 2.8. Соотношение SMR “вытекает” из психоакустической модели, так как длина шага (step size) растет экспоненциально, то график на рис. 2.7 имеет логарифмическую шкалу и не отражает как следует значений SNR при низких значениях длины шага. График на рис. 2.8 отражает только низкие значения длины шага.

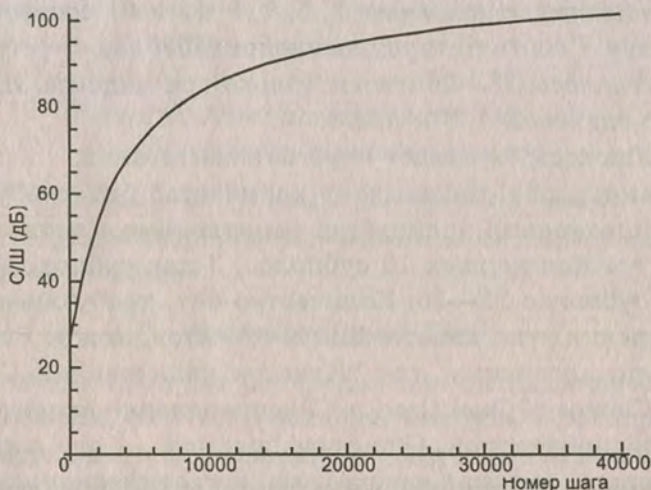


Рис. 2.7. Зависимость соотношения С/Ш от номера шага

Количество бит для отсчетов "bspl" (1), масштабные коэффициенты "bscf", и (3) информация отбора масштабных коэффициентов "bsel" сбрасываются в ноль. Минимальное соотношение MNR увеличивается путем использования следующего наибольшего индекса в предыдущей секции "Possible Quantization per Subband" и затем вычисляется новое значение MNR для данной субполосы.

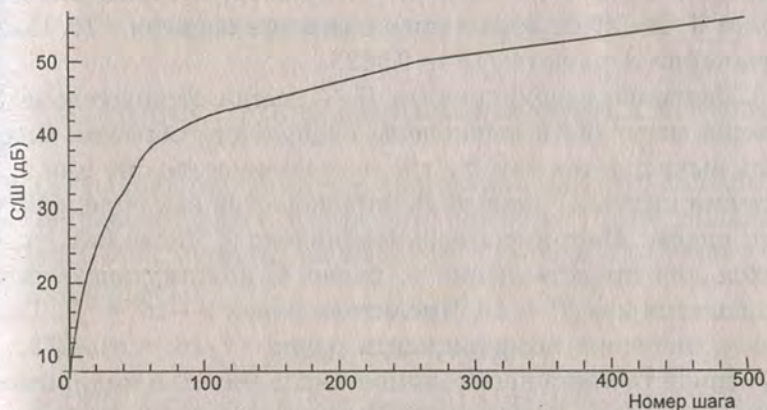


Рис. 2.8. Зависимость соотношения C/Ш от номера шага

Количество бит для отсчетов bspl обновляется в соответствии с дополнительным количеством требующихся бит. Если с первого раза (при первой итерации) для субполосы требуется ненулевое количество бит, то обновляется и bsel и bscf. Затем вычисляется значение abd:

$$abd = cb - (bhdr + bcrc + bbal + bsel + bscf + bspl + banc) \quad (2.21)$$

Итерации продолжаются до тех пор, пока значение abdnexth становится не меньше, чем любое возможное значение bspl, bsel и bscf в одной петле цикла.

Квантование и кодирование отсчетов субполосы

Квантованный отсчет $Q(i)$ вычисляется следующим образом:

$$Q(i) = A \cdot X(i) + B, \quad (2.22)$$

где $X(i)$ — отсчеты субполосы.

А и В определяются при помощи следующего алгоритма:

1. Всего 17 индексов. Первые 14 — те же самые, что и описанные в разделе "Possible Quantization". Значение индексов 15—17 определяется как $2^{i-1}-1$.

2. Значение коэффициента А — это значение индекса деленное на минимальное из 2 значений энергии, которые больше, чем значение индекса. Например, возьмем значение индекса 9. Следующее большее значение энергии — 16. Поэтому значение А равно $9/16 = 0.5625$.

3. Значение коэффициента В — всегда отрицательно. Его значения могут быть вычислены следующим образом: Знаменатель вычисляется как 2^n , где n — количество бит для представления индекса. Числитель определяется как определитель минус индекс. Например, возьмем индекс 9. Число бит, требующихся для представления 9, равно 4, поэтому знаменатель определяется как $2^4 = 16$. Числитель равен $9-16 = -7$. Таким образом, значение коэффициента равно $-7/16 = -0.4375$.

N представляет необходимое число бит для кодирования количества шагов. Инверсия MSB производится в порядке уклонения от появления кода "1", который используется для слова синхронизации.

Алгоритм "Possible Quantization" показывает, какая из групп используется. Если группировка не требуется, то группы из трех отсчетов кодируются отдельными кодовыми словами.

Если требуется группировка, то три следующих друг за другом отсчета кодируются одним кодовым словом. Для этой тройки передается только одно значение v_m (первое значение MSB). Эмпирическое правило для определения v_m ($m = 3, 5, 9$) по трем последовательным отсчетам субполосы x, y и z :

$$v_3 = 9z + 3y + x \quad (v_3 = 0...26) \quad (2.23)$$

$$v_5 = 25z + 5y + x \quad (v_5 = 0...124) \quad (2.24)$$

$$v_9 = 81z + 9y + x \quad (v_9 = 0...728) \quad (2.25)$$

2.3.3. Форматирование

Форматирование пакетов слоя 2 MPEG-1 показано на рис. 2.9.

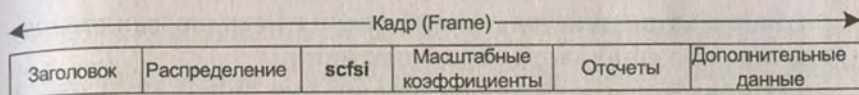


Рис. 2.9. Формат пакета слоя 2 MPEG-1

2.4. Процесс декодирования аудиосигнала

Для начала процесса декодирования декодер должен синхронизировать себя с входящим потоком. После включения декодера этот процесс начинается с поиска 12-битного кода синхронизации.

2.4.1. Обобщение

В DBS количество бит в заголовке уже известно декодеру и поэтому может рассматриваться как расширение кодового слова синхронизации, что позволяет реализовать более надежную синхронизацию. Для Слоя 2 количество бит в пакете определяется как:

$$N = 144 * \frac{\text{скорость потока}}{\text{частота дискретизации}} \text{ (байт)}. \quad (2.26)$$

Для DBS (Direct Broadcast Satellite)

$$N = 144 * \frac{192\,000}{48\,000} = 576 \text{ байт} = 4,608 \text{ бит}. \quad (2.27)$$

2.4.2. Декодирование Слоя 2 MPEG-1 Audio

Аудиодекодер сначала распаковывает поток и посылает дополнительные (служебные) данные на свой выход. Оставшиеся данные используются для функции восстановления и затем

происходит их инверсная расстановка для создания на выходе декодера отсчетов для левого и правого стереоканалов.

Декодирование аудиопакетов

Анализ аудиопакетов производится с использованием следующего трехшагового приближения.

1. Первый шаг состоит из чтения бит "nbal" (2, 3 или 4) для одной субполосы из потока. Эти биты интерпретируются как целое число без знака.

2. Параметр nbal и номер субполосы используются как индексы для вычисления параметра nlevels, используемого для квантования отсчетов в субполосе.

3. Количество бит, используемых для кодирования квантованных отсчетов, коэффициенты деквантования (восстановления) и все коды отсчетов для трех следующих друг за другом субполос, сгруппированные в один код, могут быть определены таким образом.

Идентификатор sblimt показывает номер ближайшей субполосы, которая не имеет распределенных в ней бит.

Декодирование информации отбора масштабных коэффициентов

Для каждой из 12 субполос есть три тождественных части (0, 1, 2). Каждая из этих частей может иметь свой масштабный коэффициент. Количество масштабных коэффициентов, которые должны быть считаны из потока, определяется в scfi[sb]. Информация отбора декодируется в порядке, обратном кодированию, показанном ранее в данной главе.

Декодирование масштабного коэффициента

Параметр scfi[sb] определяет количество кодированных масштабных коэффициентов и часть отсчетов субполосы, которые относятся к ним. Шесть бит закодированного масштабного коэффициента должны быть определены как целый индекс, не имеющий знака. В разделе "Possible Quantization of Subbands" (см. выше) показаны масштабные коэффициенты,

на которые должны быть умножены отсчеты субполос после восстановления.

Восстановление отсчетов субполос

Кодированные отсчеты появляются триплетами (по три штуки). Из раздела "Possible Quantization" можно узнать, сколько бит должны быть считаны для каждого триплета для каждой субполосы. Как там показано, известно, что весь код состоит либо из трех следующих друг за другом независимых друг от друга кодов каждого отсчета или одного кода объединенных вместе как один код (группировка). Если при кодировании использовалась группировка, то при декодировании должна быть произведена разгруппировка этих кодов. Если группировка при кодировании не производилась, то считанное из потока кодовое слово рассматривается как целое число не имеющее знака и обозначаемое c . Три следующих друг за другом кода $s[0]$, $s[1]$ и $s[2]$ вычисляются следующим образом:

$$\begin{aligned} s[0] &= c \text{ по модулю } [nlevels]^*; \\ c[1] &= c \text{ DIV } nlevels; \\ s[1] &= c[1] \text{ по модулю } [nlevels]; \\ c[2] &= c[1] \text{ DIV } nlevels; \\ s[2] &= c[2] \text{ по модулю } [nlevels]. \end{aligned} \tag{2.28}$$

Для изменения выполненной в процессе кодирования инверсии MSB первый бит каждого из трех кодов должен быть инвертирован. Восстановленные (деквантованные) значения вычисляются следующим образом:

Для вычисления коэффициента C (не надо путать с коэффициентом C_r , используемом в банке фильтров для анализа (Analysis Filterbank)) берется n , равное числу шагов. Затем вычисляется C :

$$C = \frac{2^j}{n},$$

где j — наименьшее число бит, которыми можно представить n .

Например, если $n = 9$, то $j = 4$, так как для представления числа 9 требуется 4 бита. Поэтому:

$$C = \frac{2^4}{9} = \frac{16}{9} = 1,77777.$$

Параметр D (не надо путать с D_i используем в банке фильтров для синтеза (Synthesis Filterbank)) можно вычислить по формуле:

$$D = \frac{1}{2^{j-1}},$$

где j — число бит, требующихся для представления n . Например, если $n = 31$, то $j = 5$ и

$$D = \frac{1}{2^4} = 0,0625.$$

Если число шагов 5 или 9, то $D = 0,5$. Если число шагов 3, 5 или 9, то это значит, что в процессе кодирования была произведена группировка кодов и в одном кодовом слове закодировано три отсчета, или же, если группировки не было, количество отсчетов в одном кодовом слове равно 1. Количество бит в одном кодовом слове равно j , за исключением тех случаев, где количество шагов 3, 5 или 9, где число бит в одном кодовом слове равно 5, 7 или 10 соответственно.

Параметры C и D используются в вычислении выражения (2.28) Масштаб восстановленных значений должны быть изменен. Коэффициенты умножения могут быть найдены при описанном выше вычислении масштабных коэффициентов. Значение коэффициента для изменения масштаба s'' вычисляется следующим образом:

$$s'' = \text{factor} * s'' . \quad (2.30)$$

Субполосный фильтр синтеза

Если в субполосе нет распределенных бит, то значения всех отсчетов данной субполосы равны нулю. Каждый раз, когда отсчеты для всех 32 субполос одного канала уже вы-

числены, они могут быть применены в субполосном фильтре синтеза и затем могут быть вычислены 32 следующих друг за другом отсчета. Детальный алгоритм этого процесса показан ниже.

Коэффициенты N_{ik} для операции матрицирования можно найти из следующего выражения:

$$N_{ik} = \cos \left[(16+i) * (2k+1) * \frac{\pi}{64} \right], 0 \leq i \leq 63, 0 \leq k \leq 31. \quad (2.31)$$

Коэффициенты D_i для оконных функций могут быть найдены по графику на рис. 2.2. Один кадр содержит $36 * 32 = 1152$ отсчетов субполос, из которых получается (после фильтрации) 1152 отсчета аудиосигнала.

Алгоритм работы банка фильтров синтеза

1. На вход фильтра поступают 32 отсчета субполос.
2. Осуществляется операция умножения матриц с использованием этих отсчетов:

$$\begin{bmatrix} V_0 \\ V_1 \\ \cdot \\ \cdot \\ \cdot \\ V_{63} \end{bmatrix} = \begin{bmatrix} N_{0,0} & N_{0,1} & \cdot & \cdot & \cdot & N_{0,31} \\ N_{1,0} & N_{1,1} & \cdot & \cdot & \cdot & N_{1,31} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ N_{63,0} & N_{63,1} & \cdot & \cdot & \cdot & N_{63,31} \end{bmatrix} * \begin{bmatrix} F_0 \\ F_1 \\ \cdot \\ \cdot \\ \cdot \\ F_{31} \end{bmatrix}$$

3. Шестнадцать этих векторов V перемещаются в таблицу FIFO показанную ниже. Осуществляется отображение V в U .

0	64	128	192	256	320	384	448	512	578	640	704	768	832	896	960
X		X		X		X		X		X		X		X	
31	95	159	223	287	351	415	479	543	607	671	735	799	863	927	991
32	96	160	224	288	352	416	480	544	608	672	736	800	864	928	992
	X		X		X		X		X		X		X		X
63	127	191	255	319	383	447	511	577	639	703	767	831	895	959	1023

X показывает, где V FIFO отображается в U .

4. Вектор U преобразуется в вектор W :

$$W_i = D_i * U_i \quad i = 0, 1, 2, \dots, 511 \quad (2.32)$$

5. Вычисляются отсчеты аудио сигнала:

$$S_j = \sum_{i=0}^{15} W_{j+32*i}$$

2.5. Алгоритмы компрессии цифровых аудиоданных

Представление высококачественных звуковых сигналов в цифровой форме широко используется в системах радиовещания и телевидения, в системах записи, хранения и воспроизведения информации, а также при обработке, монтаже и редактировании фонограмм, при создании новых необычных видов звучаний, в мультимедийных технологиях и т. д.

Аналоговый звуковой сигнал в студийном тракте преобразуется в цифровую форму в аналого-цифровом преобразователе (A/D , рис. 2.10). При этом так называемом первичном кодировании используется равномерное квантование отсчетов звукового сигнала (ЗС) с разрешением $\Delta A = 16 \dots 24$ бит/отсчет при частоте дискретизации $f_d = 44, 1 \dots 96$ кГц. В каналах студийного качества часто $\Delta A = 16$ бит/отсчет, $f_d = 48$ кГц, полоса частот кодируемого звукового сигнала $\Delta F = 20 \dots 20000$ Гц. Динамический диапазон такого цифрового канала составляет не менее 54 дБ. Если $f_d = 48$ кГц и $\Delta A = 16$ бит/отсчет, то скорость цифрового потока при передаче одного такого сигнала составит $v = 48 * 16 = 768$ кбит/с, что требует суммарной пропускной способности канала связи при передаче сигнала формата 3/2 + канал СНЧ более 3,840 Мбит/с. Но человек способен сознательно обрабатывать лишь около 100 бит/с информации. Поэтому можно говорить о присущей первичным цифровым звуковым сигналам избыточности. Компрессия цифровых аудиоданных обычно выполняется в кодере источника (см. рис. 2.10). На приемной стороне цифровой системы передачи

перед цифроаналоговым преобразованием (D/A) компрессированный сигнал восстанавливается в декодере источника.

Новейшие исследования, выполненные в области комп-



Рис. 2.10. Упрощенная схема передающей и приемной частей цифровой системы передачи звуковых сигналов

рессии цифровых аудиоданных в течение последних 20—25 лет, привели к появлению новых методов кодирования MUSICAM (Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing); разработанных совместно Институтом техники радиовещания (Германия); фирмой Philips (Голландия) и ССЕТТ (Франция); OCF (Optimum Coding in the Frequency Domain), PXXM (Perceptual Transform Coder), AT&T Hybrid, CNET и далее ASPEC (Adaptive Spectral Perceptual Entropy Coding, разработан совместно фирмами AT&T Bell Laboratories (США), Deutsche Thomson Brandt (Франция), университетами Eriangen и Hannover и фирмой Fraunhofer Gesellschaft (Германия); ATRAC (Adaptive Transform Acoustic Coding, США). На их основе были разработаны европейские стандарты MPEG-1 ISO/IEC 11172—3 [1], MPEG-2 ISO/IEC 13818—3, 13818—7 [2, 3], MPEG-4 ISO/IEC 14496—3 [4], а также стандарт ATSC Dolby AC—3 [5], США.

Наиболее важные методы кодирования высококачественных ЗС представлены в табл. 2.3.

Названия методов кодирования в табл. 2.3 даны так, как они предложены разработчиками и приведены в публикациях. Здесь даны также основные характеристики и области применения алгоритмов компрессии. Заметим лишь, что методы кодирования звуковых сигналов, изложенные в стандартах группы MPEG (Moving Pictures Expert Group) и ATSC

**Классификация основных методов кодирования
высококачественных звуковых сигналов
с компрессией цифровых аудиоданных**

Наименование алгоритма	Метод компрессии	Частота дискретизации, кГц	Кол-во каналов передачи	Скорость передачи кбит/с на канал	Величина компрессии	Области применения
ASPEC	Кодирование с преобразованием	32, 44.1, 48	1,2	64...192	1 : 6	ISDN
ATRAC	Субполосное кодирование с преобразованием	44.1	2	256	1 : 5	MiniDisk
MASCAM	Субполосное кодирование	48	1,2	128...256	1 : 4	DAB (Digital Audio Broadcasting)
MUSICAM	Субполосное кодирование	48	1,2	128...256	1 : 4	DAB
MPEG-1, Layer-1 и -2	Субполосное кодирование (MUSICAM)	32, 44.1, 48, Layer-1 только 48	1,2	32...448 (Layer-1) 32...384 (Layer-2)	1 : 4 (Layer-1) 1 : 6 (Layer-2)	DAB (Layer-2, 128...256 кбит/с), DBS (Direct Broadcast Satellite, Layer-2, 224 кбит/с), DCC (Digital Compact Cassette, Layer-1, 384 кбит/с)
MPEG-1, Layer-3	Субполосное кодирование с преобразованием	32, 44.1, 48	1,2	32...320	1 : 9	Internet-вещание
MPEG-2	Субполосное кодирование Субполосное кодирование с преобразованием	16, 22, ..., 48	1-5,1	32...384	< 1 : 9	Многоканальное стереофоническое вещание
MPEG-2 AAC	Субполосное кодирование с преобразованием	16, 22, ..., 48	5,1	16...384	1 : 15	Многоканальное стереофоническое вещание
MPEG-4	Субполосное кодирование с преобразованием Параметрическое кодирование	16, 22, ..., 48	1-5,1	2...64		Мультимедиа приложения
Dolby AC-3	Кодирование с преобразованием	44.1	1-5,1	32...384	1 : 13	Кинематограф, HDTV, спутниковое вещание
Гибридное кодирование	Субполосное кодирование с преобразованием Параметрическое кодирование	44.1	1,2	32...64	1 : 15..20	Радиовещание, хранение информации

(Advanced Television System Committee), нашли применение в системах цифрового звукового вещания DAB (Digital Audio Broadcasting) [10], Европа [6, 8]; телевидения (Dolby AC-3) [7], США; Internet;

ISDN. В системе DAB предпочтение отдано скорости передачи 128 кбит/с (Layer-2), а в системе Dolby AC-3 при формате 3/2 плюс канал СЧ — скорости 384 кбит/с. Однако в новейших гибридных методах кодирования высококачественных ЗС [9], сочетающих в себе идеи параметрического кодирования и кодирования с преобразованием, удалось уменьшить скорость цифрового потока при передаче сигнала стереопары до 32... 64 кбит/с, т. е. примерно в 25 раз по сравнению с ее исходной величиной 1, 536 Мбит/с.

Кодер с компрессией цифровых аудиоданных

Обобщенная структурная схема такого устройства представлена на рис. 2.11. В блоке время частотного анализа и сегментации исходного звукового сигнала $s(n)$ разделяется на субполосные составляющие и сегментируется по времени. Длина кодируемой выборки зависит от формы временной функции ЗС. При отсутствии резких выбросов по амплитуде используется так называемая длинная выборка, обеспечивающая высокое разрешение по частоте. В случае же резких изменений амплитуды сигнала длина кодируемой выборки резко уменьшается, что дает более высокое разрешение по времени. Решение об изменении длины кодируемой выборки принимает блок психоакустического анализа, вычисляя значение психоакустической энтропии сигнала. После сегментации субполосные сигналы квантуются и затем кодируются. В наиболее эффективных алгоритмах компрессии аудиоданных кодированию подвергаются не сами отсчеты ЗС, а соответствующие им коэффициенты модифицированного дискретного косинусного преобразования (МДКП). Обычно при компрессии цифровых аудиоданных используется так называемое энтропийное кодирование, при котором одновременно учитываются как свойства слуха человека, так и статистические

характеристики ЗС. Однако основную роль при этом играют процедуры устранения психоакустической избыточности, основанные на учете свойств слуха. Если известно, какие доли (части) звукового сигнала ухо воспринимает, а какие нет вследствие маскировки, то можно вычлениить и затем передать по каналу связи только те части сигнала, которые ухо способно воспринять, а неслышимые доли (части, составляющие исходного сигнала) можно отбросить (не передавать по каналу связи). Кроме того, отсчеты ЗС или соответствующие им коэффициенты МДКП можно квантовать с возможно меньшим разрешением по уровню так, чтобы искажения квантования, изменяясь по величине с изменением уровня самого сигнала, еще оставались бы неслышимыми, т. е. маскировались исходным сигналом. Учет закономерностей слухового восприятия ЗС выполняется в психоакустической модели или, иначе говоря, в блоке психоакустического анализа. Здесь по специальной процедуре для каждого субполосного сигнала рассчитывается допустимый уровень искажений (шумов) квантования, при котором они еще маскируются полезным сигналом данной субполосы. Именно этот блок в конечном счете определяет эффективность и качество кодера с компрессией цифровых данных.

Блок динамического распределения бит в соответствии с



Рис. 2.11. Обобщенная структурная схема кодера источника с компрессией цифровых аудиоданных

требованиями психоакустической модели для каждой субполосы кодирования выделяет такое минимально возможное

их количество, чтобы уровень искажений, вызванных квантованием, не превышал порог их слышимости.

Для обеспечения правильного декодирования компрессированных сигналов кроме кодовых слов отсчетов ЗС или соответствующих им коэффициентов МДКП (основные аудиоданные) к декодеру передается также и определенная дополнительная информация (данные дополнительной информации). После кодирования цифровые потоки основной и дополнительной информации форматируются. При этом наиболее важная ее часть подвергается помехоустойчивому кодированию (CRC-код).

Рассмотрим подробнее особенности алгоритмов компрессии цифровых аудиоданных, реализованных в стандартах группы MPEG.

Кодеры стандарта MPEG-1 ISO/IEC 11172—3 (8).

Первичный цифровой сигнал ($f_d = 48$ кГц, $\Delta A = 16$ бит/отсчет, $v = 768$ кбит/с) разделяется банком цифровых фильтров на 32 субполосные составляющие (рис. 2.12 и 2.13). Аудиофрейм образует выборка, состоящая из 384 (для Layer-1) или из 1152 (для Layer-2 и —3) отсчетов звукового сигнала. Время передачи аудиофрейма составляет 8 мс (Layer-1) и 24 мс (Layer-2 и Layer-3). Банк фильтров имеет относительно малое время задержки ЗС при одновременной компенсации искажений, возникающих за счет интерференции, в местах стыковки субполосных сигналов. Все 32 субполосы имеют постоянную ширину $\Delta F = f_d / (2n)$, где f_d — частота дискретизации ЗС, n — число субполос, равную 750 Гц. После фильтрации отсчеты ЗС в каждой субполосе собираются в блоки. Кодированию могут подвергаться либо сами отсчеты ЗС (Layer-1 и -2), либо коэффициенты МДКП (MDCT- Modified Discrete Cosine Transform, Layer-3). В первом случае (рис. 2.12) в каждой субполосе кодирования имеем по $384:32 = 12$ (Layer-1) или по $1152:32 = 36$ отсчетов ЗС (Layer-2), во втором случае (рис. 2.13) — по 18 коэффициентов МДКП. Заметим, что в Layer-1 и в Layer-2 используется банк полифазных квадра-

турных зеркальных фильтров PQMF (Polyphase Quadrature Mirror Filter); в Layer-3 применен так называемый гибридный банк фильтров (Hybrid PQMF Filter Bank, рис. 2.14), который дополнительно содержит блок МДКП (МОСТ) и частотный корректор ЧК. В Layer-3 при помощи банка PQMF-фильтров 1152 отсчета ЗС двумя блоками по 576 отсчетов каждый прежде всего разделяются на 32 субполосные составляющие. На выходе каждого такого PQMF-фильтра имеем соответственно 36 (18 + 18) отсчетов ЗС. При расфилтровке значение в каждом субполосном канале понижается в 32 раза. Далее, в каждом таком субканале для группы отсчетов ЗС выполняется МДКП. При этом длина N ортогонального преобразования зависит от формы временной функции выборок ЗС. При отсутствии всплесков в сигнале выборки МДКП выполняется для всей группы из 36 отсчетов субполосного сигнала, а при их наличии — отдельно для трех групп. Каждая группа содержит уже по 12 отсчетов ЗС, образуя три подблока. При длинном преобразовании ($N = 36$) общее число коэффициентов МДКП во всех субполосах (субканалах) равно $18 \times 32 = 576$, при коротком преобразовании ($N = 12$) имеем соответственно по $6 \times 32 = 192$ коэффициента МДКП в каждом коротком блоке. Далее эти три коротких блока объединяются в гранулу, так что в обоих случаях общее число коэффициентов МДКП в выборке, звукового сигнала равно $192 \times 3 = 576$.

Для изменения длины N ортогонального преобразования используется набор оконных функций: синусное окно типа "long" для длинного преобразования, синусное окно "short" для короткого преобразования и два окна перехода "start" и "stop", необходимые для уменьшения искажений, возникающих при переходе от длинных к коротким окнам и наоборот (рис. 2.15). Частотный корректор (ЧК) необходим для уменьшения искажений (см. рис. 2.14), возникающих вследствие добавления в банк PQMF фильтров (Layer-1 и Layer-2) процедуры ортогонального преобразования. Решение о выборе длины МДКП принимает психоакустическая модель кодера, вы-

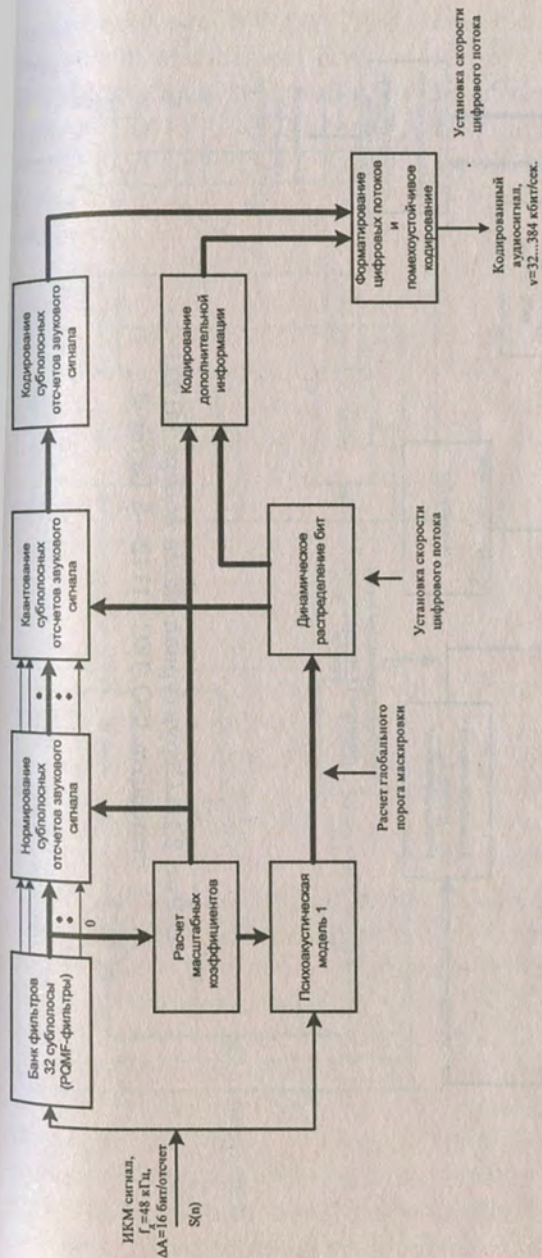


Рис. 2.12. Структурная схема кодера MPEG стандарта ISO/IEC 11172—3, Layer-1 и Layer-2

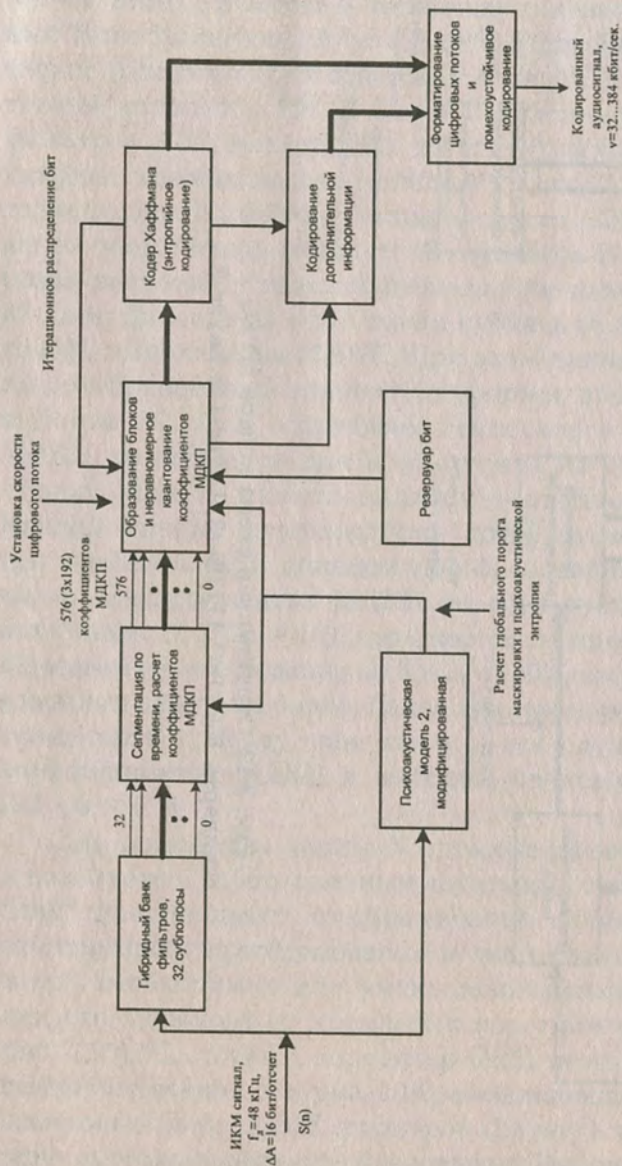


Рис. 2.13. Структурная схема кодера MPEG стандарта ISO/IEC 11172—3 Layer-3

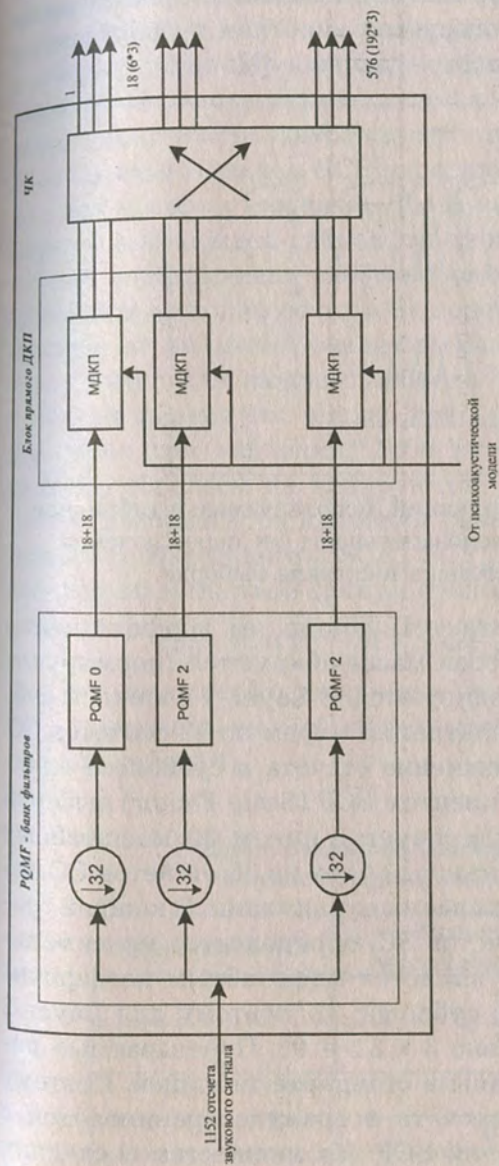


Рис. 2.14. Структурная схема гибридного банка фильтров (кодер MPEG стандартов ISO/IEC 11172—3 или 13818—3, Layer-3)

числя для сигнала каждого аудиофрейма величину психоакустической энтропии PE [10]. Если PE превышает некоторое пороговое значение, то используется короткое преобразование ($N = 12$), в противном случае — длинное ($N = 36$).

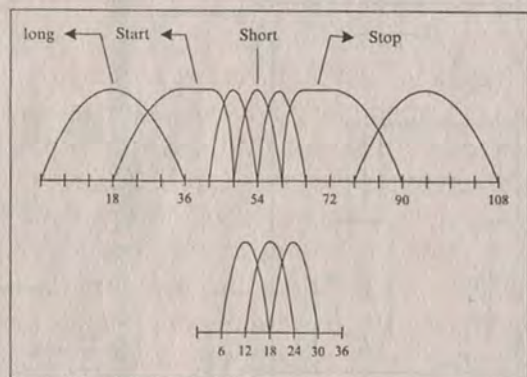


Рис. 2.15. Типы оконных функций, используемые в гибридном банке фильтров, и последовательность их переключения при наличии всплеска в сигнале выборке

Отсчеты ЗС или соответствующие им коэффициенты МДКП после банка фильтров масштабируются (нормируются), квантуются и затем кодируются. В Layer-1 в каждом субполосном канале после фильтрации имеем по 12 отсчетов ЗС. При этом максимальное значение отсчета в субполосе является масштабным коэффициентом SCF (Scale Factor) выборки субполосного ЗС. Всего для Layer-1 имеем 32 масштабных коэффициента. В Layer-2 каждый блок из 36 отсчетов ЗС делится на три подблока, называемые гранулами. В каждой грануле, включающей 12 отсчетов ЗС, определяется максимальный отсчет, его значение является масштабным коэффициентом SCF гранулы. Всего субполос 32, поэтому для Layer-2 общее количество SCF равно $3 \times 32 = 96$. Передаваемые декодеру значения SCF заданы в стандарте таблицей. Поэтому максимальное значение отсчета в грануле сравнивается с набором табличных значений SCF. Из множества последних выбирается ближайшее большее. Оно и принимается за SCF

гранулы. Каждому табличному значению SCF соответствует свой индекс $iscf$, указанный в соответствующей таблице стандарта ISO/IEC 11172—3. Кодировается не само значение SCF, а соответствующий ему индекс $iscf$, который представляет собой целое число, изменяющееся от 0 до 62. Значение индекса масштабного коэффициента $iscf$ (0...62) кодируется шестью битами. Индексы SCF передаются только для тех субполос, для которых выделены биты на кодирование отсчетов. В Layer-1 для каждой субполосы передается лишь один масштабный коэффициент (либо ни одного, если отсчеты данной субполосы не кодируются и не передаются). В Layer-2 в зависимости от различий значений масштабных коэффициентов трех гранул для каждого субполосного сигнала может передаваться разное их число: либо одно максимальное, либо два, либо три значения, либо не одного, если отсчеты ЗС данной субполосы не кодируются. Для этой цели предварительно вычисляются значения разностей $dscf_1$ и $dscf_2$ масштабных коэффициентов SCF_1, SCF_2, SCF_3 соответственно первой второй и третьей гранул субполосного сигнала:

$$dscf_1 = SCF_1 - SCF_2; \quad dscf_2 = SCF_2 - SCF_3.$$

Все множество возможных значений разностей $dscf_1$ и $dscf_2$ разделено на 5 областей, называемых классами различий (см. табл. 2.4).

Таблица 2.4

Класс различий. Значение разности пар масштабных коэффициентов

Класс различий	Значение разности пар масштабных коэффициентов
1	$dscf \leq -3,$
2	$-3 < dscf < 0,$
3	$dscf = 0,$
4	$0 < dscf < +3,$
5	$dscf \geq +3.$

Классы различий определяют количество и номера SCF для каждой субполосы, передаваемое декодером. Информация о выбранных масштабных коэффициентах (SCFSI- Scalenoct Selection Information) кодируется 2 бит для каждой субполосы. Расчет значений SCF и выбор их числа для передачи к декодеру выполняется в блоке расчета масштабных коэффициентов.

Таблица 2.5

Набор квантователей

Номер субполосы, N	Индекс квантователя (следующая строка) и соответствующее ему число ступеней квантования (остальные строки)														
	1	2	3	3	4	5	6	7	8	9	10	11	12	13	14
0, 1, 2	3	7	15	31	63	127	255	511	1023	2047	4095	8191	16383	32767	65535
3, 4, ..., 103	5	7	9	15	31	63	127	255	511	1023	2047	4095	8191	65535	
11, 12, ..., 22	3	5	7	9	31	65536									
23, 24, ..., 26	3	5	65535												

После того как значения масштабных коэффициентов SCF найдены, отсчеты субполосных сигналов нормируются. Нормирующим множителем для группы отсчетов является соответствующий им масштабный коэффициент SCF. Далее нормированные значения отсчетов ЗС квантуются и кодируются.

В Layer-1 и -2 используется равномерное квантование. Для каждой субполосы n число ступеней квантования или набор квантователей задается в стандарте для Layer-1 и -2 таблицей. Например, для Layer-2 и для скоростей передачи 56...384 кбит/с (табл. 2.5).

Для каждой скорости передачи (или группы скоростей) существует своя такая таблица. Значения f_d и скорость передачи цифровых данных v определяют также и максимальный номер субполосы n , отсчеты ЗС которой еще кодируются и передаются на приемную сторону системы. Например, при $f_d = 48$ кГц и $v = 56...192$ кбит/с биты не выделяются для кодирования отсчетов ЗС с 27-й по 31-ю субполосу; при $f_d = 32$ кГц и $v = 32$ кбит/с биты не выделяются для кодирования в субполосы с 8-й по 31-ю и т. п.

Требуемая величина шага квантования n отсчетов ЗС для каждой субполосы вычисляется из отношения сигнал—маска SMR , найденного с помощью психоакустической модели, где SMR — минимально возможное отношение энергии полезного сигнала к энергии шумов квантования в субполосе Δ_n , дБ, при котором этот шум еще маскируется полезным сигналом. Значение Δ_n меняется от одной субполосы кодирования к другой. Однако для всех нормированных отсчетов ЗС внутри субполосы n величина шага квантования Δ_n остается постоянной. Число ступеней квантования при кодировании отсчетов ЗС в субполосах составляет 2^m , где m — число разрядов в кодовом слове отсчета. Число ступеней квантования изменяется дискретно в пределах 3...65535, соответственно число разрядов m в кодовом слове отсчета должно составлять при этом 2...16. Однако биты можно расходовать более экономно, используя процедуру группирования. Она применяется при кодировании отсчетов ЗС, прошедших квантователи с тремя, пятью и семью ступенями квантования. При другом числе ступеней квантования процедура группирования отсчетов не применяется. При группировании каждые три следующих друг за другом квантованных субполосных отсчета ЗС кодируются одним кодовым словом. При этом требуемое на кодирование каждого отсчета в такой группе число бит составляет 1.6.7 вместо 2 (при трех ступенях квантования), 2.3.3 вместо 3 (при 5 ступенях квантования) и 3.3.3 вместо 4 (при 9 ступенях квантования), т. е. для всего набора используемых квантователей изменяется в пределах 1.6.7, 2, 3.3, 3, 3.3.3, 4, 5, 6, 7, ..., 16. Кодовые слова отсчетов, прошедших процедуру группирования, содержат соответственно 5, 7 и 10 бит.

Кодируется не число ступеней квантования, а соответствующий данному квантователю индекс, изменяющийся от 1 до 15. Если значение этого индекса принято равным 0, то это значит, что отсчеты ЗС данной субполосы не кодируются и не передаются. Подсчитаем число бит, необходимое для передачи информации о выбранном для каждой субполосы

числе ступеней квантования. Заметим, что это поле данных называется информацией о распределении бит (Bit Allocation) по субполосам кодирования. Например, для субполосы с номером $n = 26$ может быть использован один из трех возможных квантователей с числом ступеней квантования, равным или 3, или 5, или 65535, имеющих соответственно индексы 1, 2 и 3. Для кодирования этих трех чисел достаточно кодовое слово из 2 бит. Для 7 квантователей, используемых при кодировании отсчетов в субполосах с 11-й по 22-ю, для передачи соответствующих им индексов кодовое слово должно иметь уже 3 бита, а для наборов квантователей групп субполос 0, 1, 2 и 3, 4,...,10 — соответственно по 4 бита. В общей сложности для передачи этой информации для всех 26 субполос потребуется 88 бит, а вообще эта цифра может находиться в пределах от 26 (если кодируются сигналы 0, 1,...,7 субполос, $f_d = 48$ или 44, 1 кГц, $v = 32$ кбит/с) до 94 бит (когда кодируются сигналы 0, 1, 2,...,29 субполос, $f_d = 48, 44, 1$ или 48 кГц, $v = 96...192$ кбит/с).

Процедура квантования отсчетов выполняется по формуле

$$S_{\text{кв. sb}}(n) = A * S_{\text{sb}}(n) + B, \quad (2.33)$$

где A и B — коэффициенты квантования, выбираются по соответствующей таблице стандарта для каждого числа ступеней квантования;

$S_{\text{sb}}(n)$ — нормированное n -е значение субполосного отсчета;

$S_{\text{кв. sb}}$ — квантованное значение n -го субполосного отсчета.

После кодирования выполняется процедура инвентирования разрядов в каждом кодовом слове отсчета.

Далее о самой процедуре выделения бит при кодировании отсчетов ЗС в Layer-1 или в Layer-2. Скорость передачи цифровых данных и длина аудиофрейма определяют общее количество бит sb , имеющееся на кодирование всей информации выборки ЗС, Отсюда следует исключить биты преам-

булы ($bhdr$, 32 бита) и избыточные биты помехоустойчивого кодирования ($bcrc$, 16 бит CRC кода). Оставшееся после этого количество бит может быть использовано непосредственно для кодирования аудиоинформации, включающей поля данных: Bit Allocation, SCFSI, SCF и субполосных отсчетов ЗС (Subband Samples). Очевидно, что общее количество бит, доступное в каждом аудиофрейме для кодирования информации о SCFSI, SCF и Subband Samples, равно:

$$adb = cb - (bhdr + bcrc + bbal + banc), \quad (2.34)$$

где $bbal$ — число бит, требуемое для передачи информации о распределении бит по субполосам (Bit Allocation), $banc$ — биты дополнительных данных (Ancillary Data), не имеющие прямого отношения к информации кодируемой выборки отсчетов ЗС.

Процедура выделения бит при кодировании аудиоинформации из их доступного числа adb в Layer-1 и Layer-2 носит итерационный характер. Процесс выделения бит начинается с субполосы, в которой рассчитанное с помощью психоакустической модели требуемое значение SMR максимально. Для этой субполосы, например, с номером i сначала выделяется число бит, требуемое для передачи информации о распределении масштабных коэффициентов по субполосам (SCFSI, 2 бита) и для передачи значений этих коэффициентов (SCF; 6, 12 или 18 бит). После этого начинают выделяться биты для кодирования отсчетов ЗС данной субполосы i , начиная с минимально возможной длины кодового слова, определяемой наименьшим числом возможных ступеней квантования отсчетов ЗС в данной субполосе, с постепенным увеличением его длины, в соответствии с имеющимся в данной субполосе i набором квантователей. Процесс выделения бит в субполосу i продолжается до тех пор, пока фактическое значение отношения SMR в этой субполосе не станет меньше любого другого из множества этих значений, рассчитанных для остальных субполос психоакустической моделью. На этом процесс выде-

ления битов в субполосу i прерывается. Полученное для этой субполосы фактическое значение SMR_i заменяет требуемое значение SMR , рассчитанное для нее ранее с помощью психоакустической модели, и биты начинают выделяться в новую субполосу, например, полосу j , где теперь это отношение стало максимальным. Этот новый цикл опять начинается с выделения бит, необходимых для кодирования информации о распределении масштабных коэффициентов и значениях самих коэффициентов субполосы j . Затем начинают выделяться биты для кодирования субполосных отсчетов новой субполосы по процедуре, описанной ранее. Процесс продолжается до тех пор, пока полученное фактическое значение SMR_j для этой субполосы не станет меньше, чем в какой-либо другой из субполос, включая и субполосу i . Затем начинается процесс выделения бит в следующую субполосу. Этот итерационный процесс продолжается до тех пор, пока не будет исчерпано все имеющееся доступное количество бит.

В Layer-3 (см. рис. 2.13) используется неравномерное квантование, его аналитическая запись имеет следующий вид:

$$X'(k) = \text{nint}(|X(k)| / [(4\sqrt{2})^{q_{\text{quant}} + q_{\text{quantf}}}]^{3/4} - 0,0946), \quad (2.35)$$

где nint — означает, что берется ближайшее целое значение;

$X(k)$ — амплитудное значение коэффициента МДКП с номером k ;

$X''(k)$ — квантованное значение k -го коэффициента МДКП; q_{quant} — параметр кривой компрессии; q_{quantf} — шаг изменения параметра q_{quant} .

Процедура неравномерного квантования выполняется с помощью двух итерационных циклов: внешнего (Outer Iteration Loop) и внутреннего (Inner Iteration Loop). Во внутреннем цикле осуществляется собственно квантование коэффициентов МДКП в соответствии с требованиями к установленной скорости передачи, а внешний цикл приводит результаты

квантования коэффициентов МДКП в соответствии с требованиями психоакустической модели. Для кодирования квантованных значений коэффициентов МДКП используются таблицы кодов Хаффмана. Это так называемое энтропийное кодирование, учитывающее статистические особенности звукового сигнала. Перед кодированием все множество квантованных коэффициентов МДКП разделяется на подобласти. Прежде всего, начиная с верхней части спектра, подсчитывается количество пар коэффициентов МДКП, равных нулю, затем — число четверок коэффициентов МДКП, величины которых находятся в пределах $-1...+1$ (область малых значений коэффициентов МДКП). Оставшаяся часть больших значений коэффициентов МДКП разбивается на подобласти. Их число зависит от типа используемого окна (см. рис. 2.15). Для окна типа "long" число таких подобластей равно 3, в случае остальных типов окон число эти подобластей равно 2. Для кодирования малых значений коэффициентов МДКП используются две таблицы (A и B) кодов Хаффмана. При кодировании больших значений коэффициентов МДКП используются 32 таблицы. Каждая таблица кодов Хаффмана оптимизирована для различных статистических характеристик сигнала. При этом из имеющегося множества разных таблиц с кодами Хаффмана выбираются (путем перебора) те из них, которые обеспечивают наименьшее количество бит при кодировании коэффициентов МДКП.

Для правильного декодирования компрессированного сигнала необходимо кодирование и передача к декодеру дополнительной информации. В Layer-1 и Layer-2 к ней относятся: информация о распределении бит по субполосам (Bit Allocation) при кодировании отсчетов ЗС, информации о распределении (SCFSI) и величинах (SCF) передаваемых масштабных коэффициентов. В Layer-3 к дополнительной информации относятся: значения переменной, определяющей число бит, использованное для кодирования коэффициентов МДКП; число пар квантованных коэффициентов МДКП, зна-

чения которых не равны 0 и не находятся в пределах $-1...+1$, информация о шаге квантования в полосах кодирования; число бит, использованное для кодирования корректирующих множителей (масштабных коэффициентов); тип окна; номера, использованных при кодировании таблиц кодов Хаффмана. Далее цифровые данные субполосных отсчетов и дополнительной информации объединяются в единый цифровой поток (форматирование) и наиболее важная часть этой информации подвергается помехоустойчивому кодированию (CRC-код) для защиты от ошибок.

Литература

1. International Standard ISO/IEC 11172-3. Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s — Part 3: Audio, 1993-08-01.

2. International Standard ISO/IEC 13818-3. Information technology — Coding of moving pictures and associated audio information. — Part 3: Audio, 1995-05-15.

3. International Standard ISO/IEC 13818-7. Information technology — Coding of moving pictures and associated audio information. — Part 7: Advanced Audio Coding (AAC), 1997(E).

4. ISO/IEC FCD 14496-3 Subpart 1. Information Technology — Very Low Bitrate Audio-Visual Coding. Part 3: Audio (ISO/JTC 1/SC 29. — № 2203).

5. Digital Audio Compression Standard (AC-3). Doc.A/52

6. ATSC Digital Television Standard. Doc.A/53.

7. *Levine S.* Audio Representation for Data Compression and Compressed Domain Processing. A Dissertation Submitted to the Department of Electrical Engineering and the Committee on Graduate Studies of Stanford University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy. — Dezember. — 1998.

8. *Musman H.G., Werner O., Fuchs H.* Kompressionsalgorithmen fuer interactive Multimedia — Systeme // Informationstechnik und Technische Informatik. — 1993. — № 2. Seite 4—18.

9. *Kapust R.* Qualitaetsbeurteilung codierter Audiosignale mittels einer Bark-Transformation. Der Technischen Fakultaet der Univarsitaet Erlangen — Nuernberg, zur Eriangen ders Grades. Eriangen. — 1993. S. 166.

10. ETS I: Draft European Telecommunication Standard pr ETS 300 401 (1994) // Radio Broadcasting System; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers, ETSI, January. — 1994.

Глава 3 MPEG-2

3.1. Кодирование динамических изображений и звукового сопровождения

MPEG-2 — стандарт, состоящий из 9 частей [1—6]. Первые три части MPEG-2 достигли статуса международного стандарта, другие части находятся на различных стадиях разработки. Одна часть была исключена.

ISO/IEC DIS 13818—1 Информационная технология — Обобщенное кодирование динамических изображений и звукового сопровождения: Системная часть.

ISO/IEC DIS 13818—2 Информационная технология — Обобщенное кодирование динамических изображений и звукового сопровождения: Видеосигнал.

ISO/IEC 13818—3:1995 [1] Информационная технология — динамических изображений и звукового сопровождения — Часть 3: Звуковой сигнал.

ISO/IEC DIS 13818—4 [3] Информационная технология — Обобщенное кодирование динамических изображений и звукового сопровождения — Часть 4: Тесты на совместимость.

ISO/IEC DTR 13818—5 [4] Информационная технология — Обобщенное кодирование динамических изображений и звукового сопровождения — Часть 5: Программное моделирование (будущее TR).

ISO/IEC — 13818—6 [5] Информационная технология — Обобщенное кодирование динамических изображений и звукового сопровождения — Часть 6: Распространение DSM-CC — полное внедрение программного обеспечения.

ISO/IEC — 13818—9 [6] Информационная технология — Обобщенное кодирование динамических изображений и зву-

кового сопровождения — Часть 9: Распространение интерфейса реального времени для системных декодеров.

Часть 1 MPEG-2 посвящена объединению одного или более элементарных потоков видео и звуковых сигналов, а также других данных в одинарные или многоканальные потоки, удобные для хранения или передачи информации. Это определено в двух формах: Программный Поток и Транспортный Поток. Каждый из них оптимизирован для различного набора приложений. Модель приведена на рис. 3.1.

Программный Поток подобен мультиплексным системам MPEG-1. Это результат объединения одного или нескольких элементарных пакетных потоков (PES, Packetized Elementary Streams), имеющих одинаковую временную базу, в один поток. Программный Поток разработан для использования в относительно свободных от ошибок средах и подходит для приложений, в которых может использоваться программное обеспечение. Информационные пакеты могут иметь различную, в том числе и относительно большую длину.

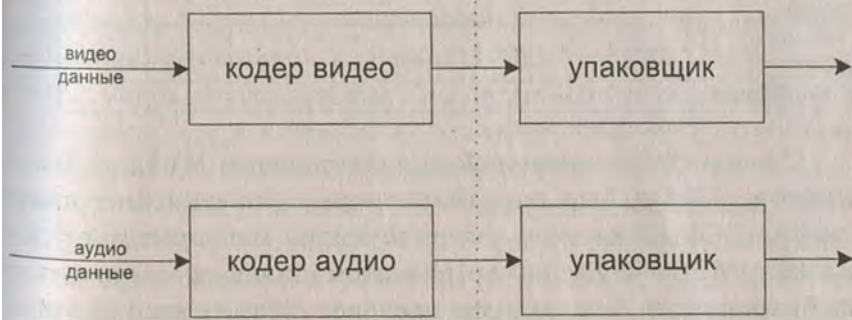


Рис. 3.1. Модель для систем MPEG-2

Транспортный поток объединяет один или более элементарных пакетных потоков (PES) с одним или более независимыми временными осями в один поток. Из элементарных потоков, разделяющих общую временную базу, формируются программы. Транспортный поток разработан для использования в такой среде, где вероятны такие ошибки, как потери

при хранении или передаче данных либо шумы. Пакеты транспортного потока имеют длину 188 байтов.

Часть 2 MPEG-2 основана на мощных возможностях сжатия видеосигнала стандарта MPEG-1 и предлагает большой набор инструментов кодирования. Они были сгруппированы по профилям, предлагающим различные функциональные возможности. Стандартом приняты только те комбинации, которые помечены "X".

Таблица 3.1

Видеопрофили MPEG-2

	Простой	Основной	С масштабируемым соотношением С/Ш	С пространственным масштабированием	Высокий	Multi-view	4:2:2
Высокий уровень		×			×		
Высокий-1440 уровень		×		×	×		
Основной уровень	×	×	×		×	×	×
Низкий уровень		×	×				

С момента окончательного утверждения MPEG-2 Видео в ноябре 1994 г. был разработан еще один дополнительный профиль. Он использует существующие инструменты кодирования MPEG-2 Видео, но при этом способен иметь дело с изображениями, имеющими цветное разрешение соответствующее формату 4:2:2 и требующих более высокую скорость передачи данных. Несмотря на то что MPEG-2 Видео был разработан без учета применения в студиях, ряд сравнительных тестов, проведенных MPEG, подтвердили, что MPEG-2 Видео имеет, как минимум, хорошие характеристики и во многих случаях даже лучшие, чем стандарты или спецификации, разработанные для высокой скорости передачи данных или применения в студиях.

Профиль 4:2:2 был окончательно одобрен в январе 1996 г., и в настоящее время является неотъемлемой частью MPEG-2 Видео.

Multiview профиль (MVP) — дополнительный профиль, разрабатываемый в настоящее время. Данный профиль должен дать возможность эффективно кодировать видео последовательности, исходящие из двух камер, снимающих одинаковый вид с маленьким углом между ними при использовании существующих в MPEG-2 инструментах кодирования видео. Этот профиль будет окончательно принят в июле 1996 г.

Часть 3 MPEG-2 является расширенным и совместимым вариантом стандарта MPEG-1 Аудио. Рисунок 3.2 показывает структуру блока данных MPEG-2 Аудио, иллюстрируя это свойство.

Части 4 и 5 MPEG-2 соответствуют частям 4 и 5 MPEG-1. Они были окончательно одобрены в марте 1996 г.

Часть 6 MPEG-2 — Digital Storage Media Command and Control (DSM-CC) — это описание набора протоколов, обеспечивающих функции управления и операции, характерные для управления потоками данных MPEG-1 и MPEG-2. Эти протоколы могут использоваться в различных типах сетей, как однородных, так и гибридных. В модели DSM-CC поток данных исходит от сервера и доставляется клиенту. Подразумевается, что и сервер, и клиент являются пользователями сети DSM-CC. DSM-CC определяет логическую сущность, называемую менеджером сеанса и ресурса (SRM, Session and Resource Manager), который обеспечивает (логически) централизованное управление сеансами DSM-CC и ресурсами (см. рис. 3.3).

Часть 6 была окончательно одобрена в качестве Международного стандарта в июле 1996 г.

Часть 7 MPEG-2 определяет алгоритм многоканального звукового кодирования, не совместимого с MPEG-1 Аудио.

Часть 8 MPEG-2 была первоначально запланирована как описание кодирования видео состоящего из 10 битовых отсчетов. Работа над этой частью была прекращена, когда ста-

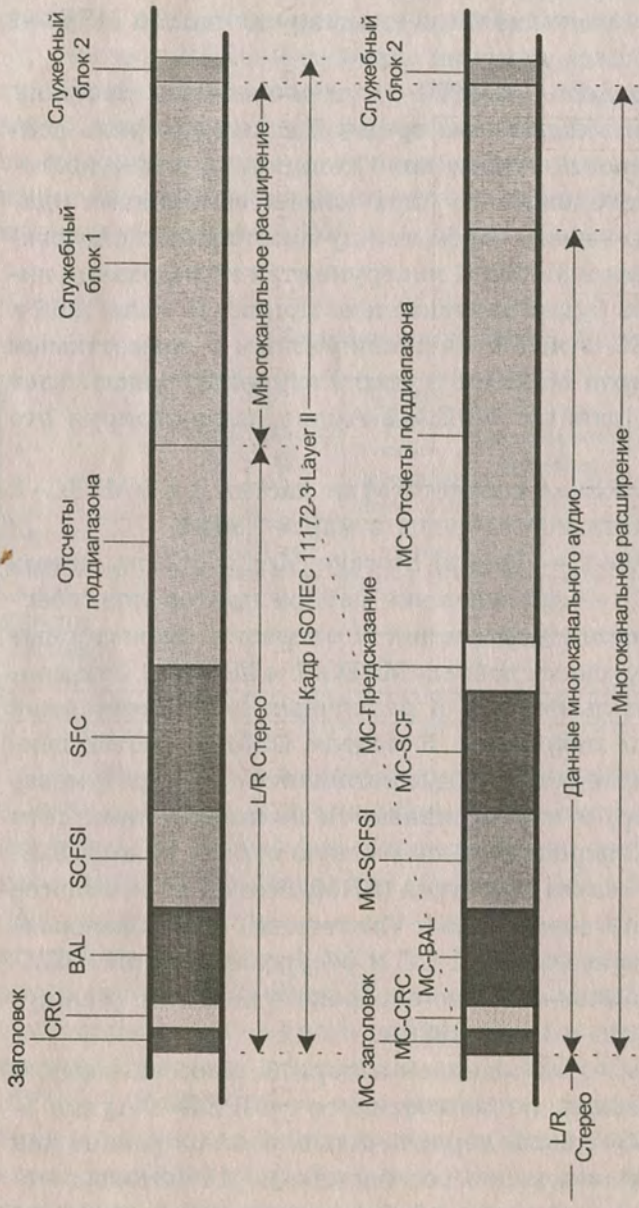


Рис. 3.2. Структура блока данных MPEG-2 Аудио

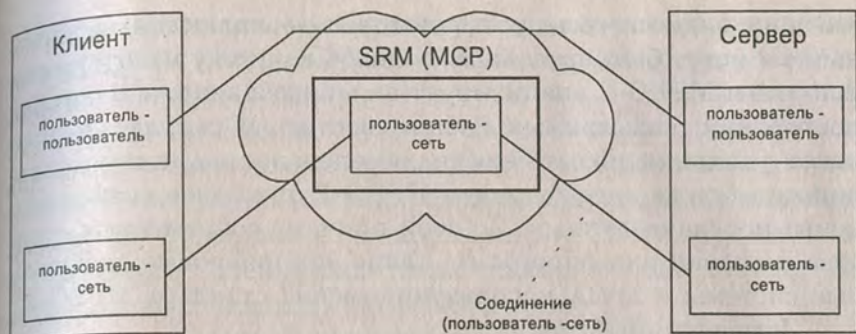


Рис. 3.3. Эталонная модель DSM-CC

до очевидно, что у производителей нет достаточного интереса к этому стандарту.

Часть 9 MPEG-2 — это определение оперативного интерфейса (RTI) для декодеров транспортного потока, которые могут использоваться для адаптации ко всем соответствующим сетям, по которым передаются транспортные потоки (см. рис. 3.4).

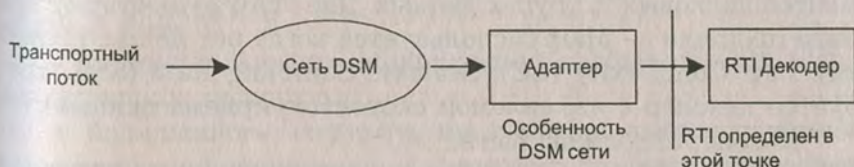


Рис. 3.4. Эталонная конфигурация оперативного интерфейса

Часть 9 была окончательно одобрена в качестве Международного стандарта в июле 1996 г.

Часть 10 — это часть DSM-CC.

MPEG-2 — upgrade для MPEG-1

Компрессия по стандарту MPEG-2 кардинально меняет положение вещей. Более 97% цифровых данных, представ-

ляющих видеосигнал дублируются, т. е. являются избыточными и могут быть сжаты без ущерба качеству изображения. Алгоритм MPEG-2 анализирует видеоизображение в поисках повторений, называемых избыточностью. В результате процесса удаления избыточности, обеспечивается превосходное видеоизображение в формате MPEG-2 при более низкой скорости передачи данных. По этой причине современные средства поставки видеопрограмм, такие как цифровые спутниковые системы и DVD, используют именно стандарт MPEG-2.

Изменения в Audio:

- ◆ появились новые виды частот 16, 22.05, 24 кГц.

- ◆ поддержка многоканальности — возможность иметь 5 полноценных каналов (left, center, right, left surround, right surround) + 1 низкочастотный (subwoofer).

- ◆ появился AAC (Advanced Audio Coding — прогрессивное кодирование звука) стандарт — обеспечивает очень высокое качество звука со скоростью 64 kbps per channel (килобит в сек. на канал), возможно использовать 48 основных каналов, 16 низкочастотных каналов для звуковых эффектов, 16 многоязыковых каналов и 16 каналов данных. До 16 программ может быть описано используя любое количество элементов звуковых и других данных. Для AAC существуют три вида профиля — Main (используется когда нет лишней памяти). Low Complexity (LC), Scalable Sampling Rate (SSR, требуется декодер с изменяемой скоростью приема данных).

Декодеры должны быть:

1. “*forwards-compatible*” (вперед совместимыми) — MPEG-2 Audio Decoder понимает любые MPEG-1 аудиоканалы.

2. “*backward compatible*” (обратно совместимыми) — MPEG-1 Audio Decoder должен понимать ядро MPEG-2 Audio (L-канал, R-канал).

3. “*matrixable*” (матрицируемыми) — MPEG1 Audio Decoder должен понимать 5-канальный MPEG-2 ($E = \text{left signal} + a * \text{center signal} + b * \text{left surround signal}$, $R = \text{right signal} + a * \text{center signal} + b * \text{right surround signal}$).

4. MPEG-1 Звуковой декодер не обязан понимать MPEG-2 AAC.

Вследствие этого совершенно спокойно можно использовать MPEG-1 Video + MPEG-2 Audio или, наоборот, MPEG-2 Audio + MPEG-1 Video.

Изменения в Видео:

◆ Требуется, чтобы разрешение по вертикали и горизонтали было кратно 16 в кодировщике кадров (frame-encoder) стандартах (покадровое кодирование), и 32 по вертикали в кодировщике полей (field-encoder, каждое поле состоит из двух кадров) стандартах (interlaced video).

◆ Возможность форматов 4:4:4, 4:2:2 (Next profile).

◆ Введены понятия Profile (форма, профиль) и Levels (уровни).

◆ Размер frame до 16383×16383 .

◆ Возможность кодировать interlaced video.

◆ Наличие режимов масштабирования (Scalable Modes)

◆ Pan&Scanning вектор (вектор панорамирования и масштабирования), который укажет декодеру как преобразовывать, например 16:9 в 4:3.

Изменения, связанные с алгоритмами кодирования:

◆ Точность частотных коэффициентов выбирается пользователем (8, 9, K), 11 бит на одно значение — в MPEG-1 только 8 бит).

◆ Нелинейный quantization процесс (разбиение непрерывных данных в дискретные).

◆ Возможность загрузить quantization matrix (матрица преобразований непрерывных данных в дискретные) перед каждым кадром.

◆ Новые режимы предсказания движения (16x8 MC, field MC, Dual Prime)

Scalable Modes (доступно только в Next и Main + Profile) делят MPEG-2 на три слоя (base, middle, high), для того чтобы организовать уровни приоритета в видеоданных (например, более приоритетный канал кодируется с большим количеством информации по коррекции ошибок, чем менее):

1. Spatial scalability (пространственное масштабирование) — основной слой кодируется с меньшим разрешением и

затем он используется как предсказание для более приоритетных.

2. Data Partitioning (дробление данных) — разбивает блок из 64 quantization коэффициентов в два потока, из которых более приоритетный переносит низкочастотные (наиболее критичные к качеству), а менее приоритетный — высокочастотные.

3. SNR (Signal to Noise Ratio) Scalability (масштабирование соотношения сигнал/шум) — каналы кодируются с одинаковой скоростью, но с разным качеством (менее приоритетный слой содержит плохую картинку — более дискретные шаги, а высокоприоритетный слой содержит довесок позволяющий построить качественную картинку)

4. Temporal Scalability (временное масштабирование) — менее приоритетный слой содержит канал с низкой скоростью передачи кадров, а высокоприоритетный содержит информацию позволяющую восстановить промежуточные кадры используя для предсказания менее приоритетные.

Таблица 3.2

Уровни

Уровень	Максимальное разрешение	Максимальная скорость	Примечание
Low (Низкий)	352 × 240 × 30	4 Мбит/сек	CIF, кассеты
Main (Основной)	720 × 480 × 30	15 Мбит/сек	CCIR 601, студийное ТВ
High 1440 (Высокий 1440)	1440 × 1152 × 30	60 Мбит/сек	4 × 601, бытовое ТВЧ (HDTV)
High (Высокий)	1920 × 1080 × 30	80 Мбит/сек	Продукция SMPTE 240M std

Таблица 3.3

Профили

Профиль	Комментарии
Simple (Простой)	Такой же как и Main, только без B-картинок. Используется в программах и КТВ (кабельное ТВ)
Main (Основной)	Стандартный MPEG-1, 95% пользователей, КТВ, спутники
Main+ (Основной+)	Main с пространственным масштабированием и масштабированием соотношения сигнал/шум
Next (Следующий)	Main+ с масштабом 4:2:2

Таблица 3.4

Допустимые комбинации профилей и уровней

Simple	Main	Main+	Next
High	Нет	Нет	4:2:2
High 1440	Нет	Main с пространственным масштабированием	4:2:2
Main	90% от всех	Main с масштабированием соотношения сигнал/шум	4:2:2
Low	Нет	Main с масштабированием соотношения сигнал/шум	Нет

Таблица 3.5

Наиболее популярные стандарты

Разрешение	Комментарии
352 × 480 × 24 (прогрессивная развертка)	VHS, хорош для фильмов
544 × 480 × 30 (черезстрочная развертка)	Laserdisc (LD), D-2, качество как у
704 × 480 × 30 (черезстрочная развертка)	Качество CCIR 601. Studio D-1



Рис. 3.5. Организация транспортного потока MPEG-2

Системный уровень MPEG-2 обеспечивает два уровня объединения данных:

1. Packetized Elementary (PES) — разбивает звук и видео на пакеты.

2. Второй уровень делится на:

◆ MPEG-2 Program Stream (совместим с MPEG-1 System) — для локальной передачи в среде с маленьким уровнем ошибок

◆ MPEG-2 Transport Stream — внешнее вещание в среде с высоким уровнем ошибок — передает транспортные пакеты (длиной 188 либо 188 + 16 бит) двух типов (сжатые данные — PES — и сигнальную таблицу Program Specific Information — PSI).

3.2. Принципы кодирования изображений в MPEG-2

Цифровое видео — это последовательность кадров, в которой каждый кадр рассматривается как набор отсчетов аналогового изображения. Отдельные отсчеты называются элементами изображения или пикселями.

Для создания цветного изображения, требуется три цветовых оси (составляющих). Большинство устройств отображения видео используют три основных цвета для получения цветных изображений: красный (red), зеленый (green) и синий (blue). При этом RGB составляющие коррелированы друг с другом. В соответствии с этим все алгоритмы сжатия выполняют шаг предварительной обработки, который состоит в преобразовании координат цвета в яркостную составляющую (Y) и две цветовые компоненты (Cr и Cb). Y компонента отображается так, как она есть в черно-белом видео. На выходе декодера идет шаг заключительной обработки, конвертирующий цветовое пространство обратно в RGB схему для дальнейшего воспроизведения.

Преобразование координат цвета состоит в перемножении трехкомпонентного вектора (RGB) с матрицей размером 3×3.

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.334 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

На выходе декодера, на шаге заключительной обработки происходит конвертация вектора $[Y, Cr, Cb]$ обратно в вектор $[R, G, B]$ путем следующей операции умножения матрицы на вектор.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0 & 1.404 \\ 1.000 & -0.3434 & -0.712 \\ 1.000 & 1.773 & 0 \end{bmatrix} \cdot \begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix}.$$

3.2.1. Черезстрочная развертка

В 1930 году, когда телевидение только создавалось, в качества устройства для отображения видео могли быть использованы только катодно-лучевые трубки (CRT, ЭЛТ). При скорости кадров 30 к/сек (это параметр для NTSC, для PAL и SECAM скорость кадров равна 25 к/сек и скорость полей 50 полей/сек), мерцание на ЭЛТ было недопустимым. Для решения этой проблемы разработчики ТВ-технологий придумали концепцию черезстрочной развертки.

Черезстрочные кадры разделены на два поля. Скорость полей при этом 60 полей/сек, т. е. удвоенная скорость кадров. При такой частоте обновления мерцания становится приемлемыми (вообще-то некоторые мерцания можно увидеть и при частоте 60 полей/сек, при частоте 72 поля/сек мерцания исчезают совсем).

В пространстве горизонтальные линии от двух полей перемежаются, как это показано на рис. 3.6. Часто заявляется, что черезстрочная развертка обеспечивает сжатие полосы пропускания 2:1. Это нонсенс, так как с каждым полем посылается только половина информации о кадре.

Дисплей, который представляет линии кадров последовательно, называют прогрессивным. В настоящее время современные технологии позволяют говорить о том, что прогрессивные дисплеи экономически пригодны и они позволяют значительно улучшить качество изображения. Однако примерно 2—3 миллиарда ТВ-приемников в мире работают с

черезстрочным изображением и поэтому DBS системы должны быть приспособлены для работы с черезстрочным изображением.

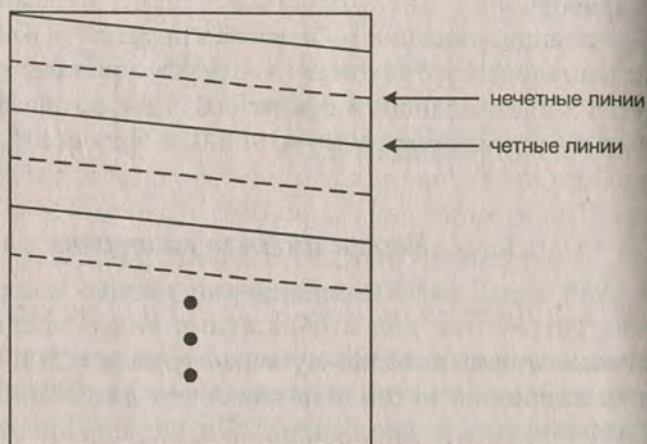


Рис. 3.6. Черезстрочный кадр

Преддискретизация цвета

Для достижения лучшего качества цифровое видео должно быть дискретизировано с сеткой с равным количеством отсчетов для каждого из трех цветовых планов. ТВ-инженеры открыли, что Human Vision System (HVS, человеческая зрительная система) менее чувствительна к разрешению по C_r и C_b осям (цветовые оси), чем по Y (яркость). Не удивительно, что меньшая чувствительность к C_r и C_b составляющим была взята как преимущество, имеющее огромную значимость для цифрового видео.

Стандарт CCIR 601—1 для записи цифрового видео использует цветовую схему "4:2:2". Это значит, что C_r и C_b цветовые планы уменьшены в двое по отношению к яркостной составляющей. Если цветовые планы в оригинале имеют размер 480×720 пикселей, то после уменьшения яркостный план все еще имеет размер 480×720 пикселей, а C_r и C_b уже имеют размер 480×320 пикселей. Еще один способ описания этого про-

цесса: количество бит на пиксель уменьшается с 24 (по 8 на каждый цветовой план) до 16, таким образом мы имеем компрессию 1:3.

Разработчики стандарта MPEG пошли другим путем. Поскольку уменьшение C_r и C_b составляющих в горизонтальном направлении асимметрично, они также используют уменьшение этих составляющих в вертикальном направлении. Получается дискретная структура 4:2:0, в которой C_r и C_b составляющие уменьшаются как по вертикали, так и по горизонтали. При этом когда Y план имеет размер 480×720 пикселей, C_r и C_b составляющие имеют размер 240×360 пикселей.

3.2.2. Временное и пространственное сжатие

Сжатие видео в MPEG-2 состоит из следующих шагов: пространственное сжатие и временное сжатие.

Пространственное сжатие

На рис. 3.7 показаны три основных шага пространственного сжатия. Сначала пиксели изображения преобразуются в блок 8×8 дискретным косинусным преобразованием (Discrete Cosine Transform, DCT). Затем коэффициенты преобразования кодируются при помощи кодирования с переменной длиной (Variable Length Coding, VLC). Затем идет процесс квантования.

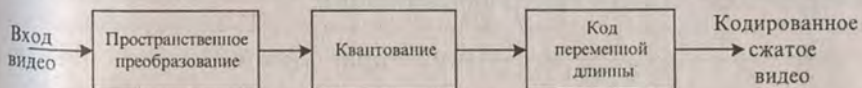


Рис. 3.7. Этапы пространственного сжатия

I-кадры. Пространственное преобразование

До начала детального разговора о DCT преобразовании важно понять роль этого преобразования. На вход DCT преобразователя поступают 8×8 массивы пикселей изображения, в котором значение интенсивности каждого пикселя может принимать значения от 0 до 255 для каждого цвета. На выходе мы имеем уже другой массив размером 8×8 — массив чисел. Пространственное преобразование преобразует блок изобра-

ражения размером 8×8 элементов в блок коэффициентов того же размера, который может быть закодирован с использованием значительно меньшего количества бит, чем оригинальный блок, который мы имели на первом этапе.

Самый первый коэффициент преобразования, имеющий индекс $(0, 0)$ особенный. Он представляет среднее значение 64 входящих пикселов. При движении коэффициента слева направо по горизонтали или вниз по вертикали преобразования говорят о росте пространственной частоты. DCT преобразование эффективно из-за того, что оно имеет тенденцию концентрировать энергию преобразования в коэффициентах преобразования, расположенных в верхнем левом углу матрицы, где наименьшая пространственная частота.

DCT преобразование: DCT блок размером $N \times N$ определяется как:

$$F(u, v) = \frac{2}{N} * C(u) * C(v) * \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) * \cos \left[\frac{(2 * x + 1) * u * \pi}{2 * N} \right] * \cos \left[\frac{(2 * y + 1) * v * \pi}{2 * N} \right], \quad (3.1)$$

где u, v, x, y принимают значения $0, 1, \dots, 7$, x и y — пространственные координаты в области выборки, u и v — координаты в области преобразования, и

$$C(u), C(v) = \begin{cases} 1/\sqrt{2} & \text{при } u, v = 0 \\ 1 & \text{при } u, v \neq 0 \end{cases} \quad (3.2)$$

Для того чтобы получить блоки размером 8×8 , значение $N = 8$.

Квантование

Второй шаг пространственного сжатия — квантование коэффициентов преобразования, которое уменьшает число бит для представления DCT коэффициентов. Квантование выполняется путем деления коэффициентов преобразования на целое число с последующим округлением до ближайшего целого числа.

Целый делитель каждого DCT коэффициента состоит из двух частей. Первая часть уникальна для каждого коэффициента в DCT матрице 8×8. Набор этих уникальных чисел также является матрицей и называется матрицей квантования. Вторая часть делителя (*quantizer_scale*) — это целое число, которое фиксировано для каждого следующего макроблока. В соответствии с этим $Dct[i][j]$ — это DCT матрица, при этом квантованная DCT матрица $Qdct[i][j]$ определяется как:

$$Qdct[i][j] = 8 * Dct[i][j] / ((quantizer_scale) * Q[i][j]).$$

Для не масштабируемых частей MPEG-2, которые представляют наибольший интерес для DBS систем, существуют две матрицы квантования: одна для Intra-кадров, другая для не Intra-кадров. Эти две матрицы показаны в табл. 3.6 и 3.7.

Таблица 3.6

Матрица квантования для I-кадров

	0	1	2	3	4	5	6	7
0	8	16	19	22	26	27	29	34
1	16	16	22	24	27	29	34	37
2	19	22	26	27	29	34	34	38
3	22	22	26	27	29	34	37	40
4	22	26	27	29	32	35	40	48
5	26	27	29	32	35	40	48	58
6	26	27	29	34	38	46	56	69
7	27	29	35	38	46	56	69	83

Таблица 3.7

Матрица квантования для не I-кадров

	0	1	2	3	4	5	6	7
0	16	16	16	16	16	16	16	16
1	16	16	16	16	16	16	16	16
2	16	16	16	16	16	16	16	16
3	16	16	16	16	16	16	16	16
4	16	16	16	16	16	16	16	16
5	16	16	16	16	16	16	16	16
6	16	16	16	16	16	16	16	16
7	16	16	16	16	16	16	16	16

Коды переменной длины

Следующим шагом в пространственной сжатии является размещение квантованных DCT коэффициентов в одномерный вектор, который затем будет являться частью таблицы кодов переменной длины. Процесс этого размещения называется зигзагообразным сканированием.

Кодовая книга (Codebook)

После того как квантованные коэффициенты DCT занесены в одномерный вектор, этот вектор становится частью книги кодов переменной длины. Назначение VLC состоит в минимизации среднего числа бит, требующихся для кодирования этого вектора.

VLC используемое в MPEG-2 Video известно как информационная теория кодов Хаффмана. Коды Хаффмана — это компактные коды, а это значит, что если основная статистика постоянна, то никакой другой код не может дать кодовое слово меньшей длины.

Кодовая книга Хаффмана для DCT коэффициентов основана на концепции неравномерных кодов. В этой технике количество следующих друг за другом нулей становится частью кодовой книги. Необычным здесь является то, что в данном

Таблица 3.8

Десять самых коротких кодов Хаффмана для DCT

Код переменной длины ¹	Расстояние	Уровень
10	конец блока	
10s ²	0	1
11s	0	1
011s	1	1
0100s	0	2
0101s	2	1
0010 1s	0	3
0011 1s	3	1
0011 0s	4	1
0001 10s	1	2

¹ "s" стоящее в конце каждого кода показывает знаковый разряд уровня.

² Эти коды используются для коэффициентов DCT преобразования.

случае в кодовую книгу вносится не только длина кода, но и амплитуда коэффициента, на котором заканчивается изменение длины кода. Для примера, в табл. 3.8 показаны несколько первых значений из таблицы В-14 (документ ISO 13818—2).

Р-кадры

Р-кадры образованы при помощи предсказания по предыдущим I- или Р-кадрам. Предыдущие I или Р кадры, используемые для предсказания Р кадров, называются *базовыми изображениями* (reference picture). Предсказание Р кадров состоит из двух независимых шагов: компенсация движения и кодирование остаточного изображения (Residual Image coding).

Кодирование остаточного изображения

Качество предсказания может очень сильно изменяться. Например, если сцена была вырезана между изображением, которое предсказано, и основным изображением, то тогда не будет соответствия с основным кадром и макроблок будет закодирован как Intra, несмотря на то что он относится к Р-изображению.

В этом случае основной макроблок (макроблок, на который идет ссылка при предсказании) будет вычтен из макроблока, имеющегося на входе для формирования остаточного изображения. Это остаточное изображение затем кодируется при помощи специальной схемы. В декодере сначала делается предсказание для макроблока с компенсацией движения, а затем к нему добавляется декодированное остаточное изображение для полного завершения процесса предсказания.

3.3. Профили и уровни

Все стандарты MPEG созданы как основные. Это значит, что они предназначены для обеспечения сжатия видеоизображений для широкого набора приложений.

Чрезвычайно широкий набор приложений требует соответствующего набора скоростей, разрешений и качества видео. Для соответствия этому широкому набору параметров в MPEG использована концепция Профилей и Уровней.

Профиль — это определенный поднабор полного синтаксиса потока данных. В рамках каждого профиля имеется широкий набор допустимых параметров. *Уровень* отражает этот набор. В восьмой части ISO 13818—2 описан набор таблиц определяющих параметры Профилей и Уровней.

Например, таблица максимальных скоростей показывает, что максимальная скорость данных для Основного профиля (Main Profile) может находиться в диапазоне от 15 Мбит/сек для Основного уровня (Main Level) до 80 Мбит/сек для Высокого уровня (High Level). Всего существует 11 комбинаций из пяти Профилей и четырех Уровней.

Профиль	Простой
	Уровень: Основной
Профиль:	Основной
	Уровень: Основной, Высокий 1440, Высокий
Профиль:	SNR
	Уровень: Низкий, Основной
Профиль:	Пространственный
	Уровень: Высокий 1440
Профиль:	Высокий
	Уровень: Основной, Высокий 1440, Высокий

Эти уровни покрывают чрезвычайно широкий диапазон видео параметров. Так, например, в DBS (Direct Broadband Satellite) системах в настоящее время используется Основной профиль (Main Profile) с Основным уровнем (Main Level), что обычно обозначается как MP@ML. В будущем в DBS системах, возможно, будет использован Высокий профиль (High Profile).

3.4. Структура кодированного потока MPEG-2 Видео

3.4.1. Общие положения

Поток видеоданных представляет собой иерархическую структуру со следующими типами элементов: видеопоследовательность, группа изображений, изображение, срез, макроблок, блок.

Видеопоследовательность — элемент потока видеоданных высшего уровня. Это серия последовательных кадров ТВ-изображения:

- ◆ построчная;
- ◆ чересстрочная.

Изображение в соответствии с используемыми методами дифференциального кодирования делится на типы:

◆ I (Intra-coded picture) — изображение кодируется с использованием только той информации, которая содержится в нем самом. В нем устраняется только пространственная избыточность;

◆ P (Predictive-coded picture) — изображение, при кодировании которого формируется разность между исходным изображением и предсказанием, полученным на основе предшествующего или последующего изображения типа I;

◆ B (Bidirectionally-predicted coded picture) — изображение, при кодировании которого используется предсказание, сформированное на основе предшествующего и последующего изображения типа I и P. При кодировании P и B изображений используется межкадровое кодирование. В них устраняется и пространственная и временная избыточность.

Начало каждого уровня в иерархии показывается уникальным 32-битным стартовым кодом. Первые шесть разрядов являются общими для всех кодов: 0x000001. Затем следует двухразрядный суффикс, который является уникальным для каждого уровня.

3.4.2. Последовательность видео

Видеопоследовательность — это структура самого высокого уровня в данной иерархии. Она начинается с заголовка

последовательности `sequence_header_code` (0x3B), за которыми может следовать заголовок группы картинок (*Group of Pictures, GOP*) и один или более кодированных кадров. Порядок передачи кадров — это порядок, в котором декодер обрабатывает их, но не обязательно корректный порядок для их последующего воспроизведения.

После кода `sequence_header_code` заголовка сам заголовок содержит ту информацию о видео, которая обычно не меняется часто. Она включает в себя информацию о размерах по вертикали и горизонтали (`horizontal_size_value` и `vertical_size_value`), которые указываются 12-битными целыми числами. Таким образом может быть описан размер изображения до 4096×4096 пикселей.

Расширение отображения последовательности (*Sequence Display Extension*) содержит информацию о том, какой используется видеоформат (NTSC (010), PAL (001), SECAM (011)). Данная информация отображается параметром `video_format`.

Видеопоследовательность заканчивается кодом конца последовательности `sequence_end_code`.

3.4.3. Группа изображений (*Group of Pictures, GOP*)

GOP — это набор изображений, который включает в себя кадры, не требующие информации от других кадров происходящих раньше или позже (определены как *Intra* или *I* кадры), кадры с однонаправленным предсказанием по предыдущим кадрам (*P* кадры) и кадры с двунаправленным предсказанием по предыдущим и следующим кадрам (*B* кадры). На рис. 3.8 показана типовая *GOP*.

Группа изображений — серия изображений, содержащих одно *I*-изображение (рис. 3.8), где стрелками показаны направления предсказания в пределах одной группы изображений.

С информационной точки зрения каждое изображение представляет собой три прямоугольных матрицы отсчетов изображений: яркостную *Y* и две матрицы цветности *C*. Соот-

ношение между количеством отсчетов яркости и цветности определяется форматом дискретизации:

◆ 4:2:0 — размеры матриц C_B и C_R в два раза меньше, чем Y , и в горизонтальном, и в вертикальном направлении (рис. 3.9);

◆ 4:2:2 — все три матрицы имеют одинаковые размеры по вертикали, но в горизонтальном направлении матрицы цветности имеют в два раза меньшее количество элементов (рис. 3.10a);

◆ 4:4:4 — все матрицы одинаковы (рис. 3.10б).

Возникает справедливый вопрос: для чего нужны различные типы кадров? На рис. 3.11 показан типовой кадр (это может быть I , P или B кадр). Для достижения максимального сжатия, избыточность должна быть удалена в трех направлениях: двух пространственных и одном временном. P и B кадры используются для достижения временного сжатия.

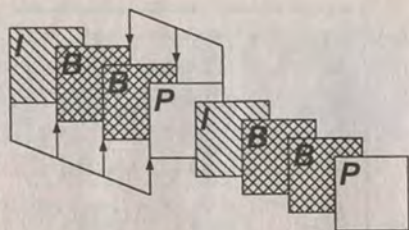


Рис. 3.8. Типовая GOP

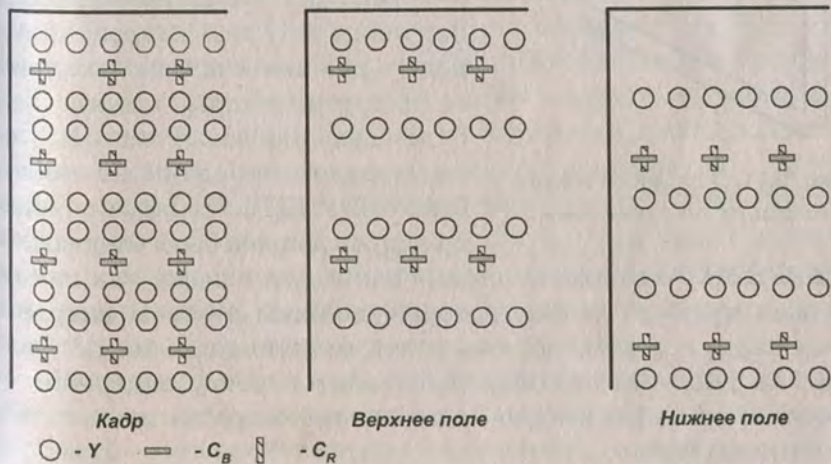


Рис. 3.9. Формат дискретизации 4:2:0

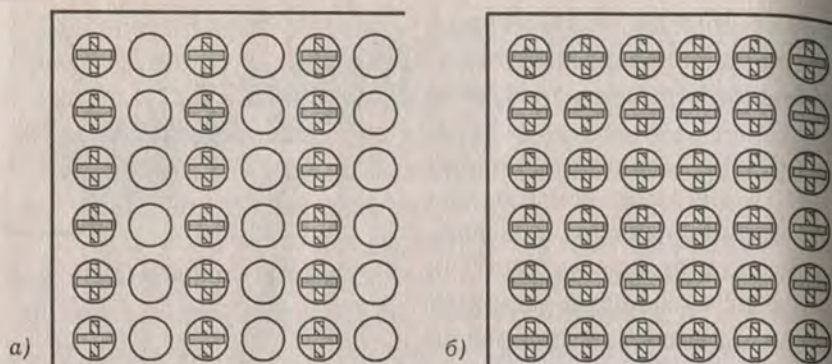


Рис. 3.10. Форматы дискретизации 4:2:2 и 4:4:4

Опыты показывают, что P кадрам требуется только 40%, а B кадрам 10% от количества бит, требуемых для I кадров.

Структуру, показанную на рис. 3.11, обычно называют гибридным кодированием. В основном из-за того, что пространственное сжатие достигается техникой преобразования, а временное сжатие достигается компенсацией движения.

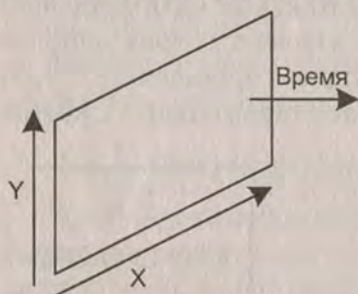


Рис. 3.11. Удаление избыточности по трем осям

Так как B кадры должны быть получены из I и P кадров, то декодер должен сначала получить те и другие кадры, прежде чем будет декодирован B кадр. Поэтому порядок передачи не может быть $IBBVBVP...$ Порядок передачи кадров должен быть следующий

$IPBVBVP...$ Это вносит ограничения для кодера, так как он должен хранить кадры, которые должны стать B кадрами, пока кадр, который должен стать следующим I или P кадром, не поступит в кодер. Это ведет к росту задержки — время от момента начала кодирования кадра до момента его воспроизведения.

Когда поток включает в себя заголовок группы изображений, то в него включены следующие параметры. Стартовый

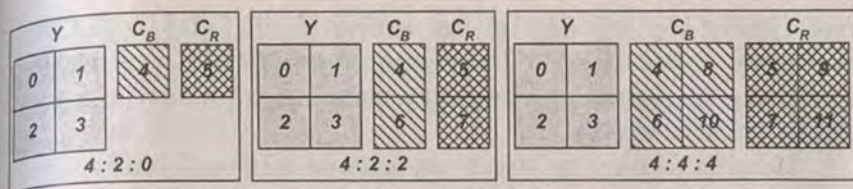


Рис. 3.12. Форматы дискретизации

код группы `group_start_code` = 0xB8. 2.5-битная двоичная строка показывающая время кодирования `time_code` в часах (5 бит), минутах (6 бит) и секундах (6 бит). Так как здесь может идти до 60 изображений в секунду, в заголовок включен 6-битный параметр `time_code_pictures`, который отсчитывает изображения. Строка `time_code` начинается с 1-битного флага `drop_frame_flag`, значение которого равно 1 только тогда, когда скорость кадров равна 29, 97 кадров в секунду. Бит маркировки `marker_bit`, значение которого всегда равно 1, расположен между параметрами `time_code_minutes` и `time_code_seconds`.

3.4.4. Изображение (кадр)

Изображения — это основная единица которая принимается как входящая и выходящая для отображения. Как уже было отмечено ранее, кодированные изображения могут быть I, P или B типа. Изображения также могут быть полевыми изображениями (Field Picture) и кадровыми изображениями (Frame Picture). Полевые изображения появляются в парах: одно верхнее (нечетное) поле и одно нижнее (четное) поле, которые вместе составляют кадр. Когда происходит кодирование черезстрочных изображений с использованием кадровых изображений, то два поля перемежаются друг с другом и кодируются как однокадровое изображение.

Заголовок изображения `picture_header` начинается со стартового кода изображения `picture_start_code` 0x00. Следующий параметр — это временная ссылка `temporal_reference` — 10 битное целое число, которое считает изображения и сбрасывается до 0 после достижения значения 1023. Следующий параметр

показывает тип кодированного изображения: `intra_coded` (001), `predicive_coded` (010) и `bidirectional_predictive_coding` (011). 16-битное целое число `vbv_delay` несет информацию о занятости буфера декодера.

Изображения делятся на срезы (Slices), макроблоки (Macroblocks) и блоки (Blocks), каждый из которых имеет свой заголовок.

Срез

Макроблок состоит из 16×16 пикселей. Срез — это ряд или неполный ряд макроблоков. Деление изображений на срезы является одной из новаций в алгоритмах сжатия MPEG. Если приходят искаженные данные, то они могут быть замкнуты в отдельный срез, который может быть пропущен декодером, так как срез не является целым изображением. Заголовок среза также содержит масштабный коэффициент квантователя, который позволяет декодеру правильно произвести восстановление изображения.

Каждое изображение делится на срезы, которые состоят из макроблоков (рис. 3.13). Макроблок складывается из блоков размером 8×8 элементов изображения (пикселей). Каждый макроблок содержит группу из четырех блоков с отсчетами яркости с размерами 16×16 пикселей и группу блоков с

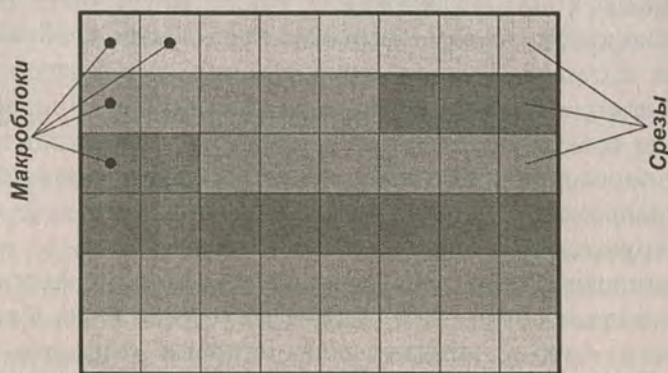


Рис. 3.13. Деление изображения на срезы и макроблоки

отсчетами цветности, взятых из той же области изображения, что отсчеты блоков яркости.

Число блоков с отсчетами цветности для различных форматов (рис. 3.12):

- ◆ 4:2:0 — по одному блоку C_B и C_R ;
- ◆ 4:2:2 — по два блока C_B и C_R ;
- ◆ 4:4:4 — по четыре блока C_B и C_R .

В изображениях типа “кадр”, в которых может использоваться и кадровое, и полевое кодирование, возможны два варианта внутренней организации макроблока (рис. 3.14):

- ◆ в случае кадрового кодирования каждый блок яркости образуется из чередующихся строк двух полей (рис. 3.14а);
- ◆ при полевом кодировании каждый блок яркости образуется из строк только одного из двух полей (рис. 3.14б).

Сокращение пространственной избыточности выполняется в изображениях типа I и достигается на уровне блока. Набор операций такого кодирования:

- ◆ дискретное косинусное преобразование;
- ◆ взвешенное квантование, определяемое элементами матрицы квантования;
- ◆ энтропийное кодирование серии коэффициентов косинусного преобразования, полученной в результате диагонального сканирования матрицы коэффициентов.

Для повышения точности предсказания и сокращения объема необходимых данных для представления изображения используется компенсация движения.

Оценивается скорость перемещения движущихся

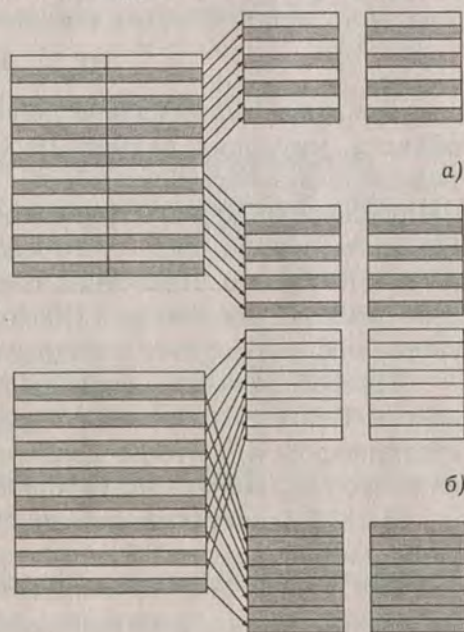


Рис. 3.14. Внутренняя организация макроблока

объектов от кадра к кадру и при определении предсказания производится соответствующая коррекция в положении опорного изображения, по отношению к которому находится ошибка предсказания.

Заголовок среза `slice_header` имеет стартовые коды от `0x01` до `0xAF`. Указатель позиции среза по вертикали `slice_vertical_position` — это последние восемь бит стартового кода среза `slice_start_code`. Для очень больших изображений, содержащих более чем 2800 линий, вводится параметр расширения позиции среза по вертикали `slice_vertical_position_extention`, который показывает расположение среза, однако, это не используется в DBS системах.

1-битный флаг `intra_slice_flag` устанавливается в значение 1 в том случае, если в потоке встречаются intra-срезы. Если макроблоки в срезе не относятся I-кадрам, то значение этого флага устанавливается в значение 0.

Макроблок

Как уже было сказано ранее, макроблок яркости — это область, имеющая размер 16×16 пикселей. Цветовая часть макроблока зависит от выборки отсчетов яркости. Структура макроблока `MP@ML` с форматом `4:2:0` показана на рис. 3.12. Из этого рисунка видно, что макоблок формата `4:2:0` состоит из шести блоков. Яркостная составляющая представляет из себя квадрат из четырех блоков размером 8×8 , а цветовые составляющие состоят из одного блока 8×8 каждая.

Макроблок — это основная единица для компенсации движения. Векторы движения определяются для яркостной составляющей, а векторы движения для цвета определяются из векторов движения для яркостной составляющей.

В DCT (дискретное косинусное преобразование) кодировании кадров каждый блок состоит из линий для обоих полей (четные и нечетные) в порядке их отображения на экране (нечетная, четная, нечетная...). Макроблок всегда делится на четыре блока. В DCT преобразовании полей каждый блок состоит из линий только четного или только нечетного поля.

У макроблоков нет заголовков. Они кодируются последовательно вместе со слайсами. Из рис. 3.13 видно, что адреса макроблоков не обязательно последовательны. Если разница адресов более, чем 33, то 11-битная двоичная строка 0000 0001 000, `macroblock_escape`, показывает, что надо прибавить 33 к значению, показываемому в поле приращения адреса макроблока `macroblock_address_increment`, представляющему из себя код длиной от 1 до 11 бит.

Следом в макроблоке идет 5-битный код шкалы квантования `quantizer_scale_code`. Наличие этого параметра в каждом макроблоке очень важно, так как оно позволяет изменять параметры квантования для каждого макроблока.

Блок

Блоки имеют размер 8×8 пикселей и являются наименьшим синтаксическим элементом MPEG-2 Видео. Блоки являются основными элементами для DCT кодирования.

3.5. Декодирование видео

3.5.1. Восстановление блоков размером 8×8

Процесс декодирования всегда представляет собой последовательность операций, обратных кодированию. Поэтому первым шагом в процессе декодирования является восстановление 64-элементного вектора кодированного при помощи VLC ($QFS[n]$). Обратное сканирование затем "укладывает" $QFS[n]$ в матрицу размером 8×8 ($QF[v][u]$). Процесс, обратный квантованию, создает массив $F[v][u]$. Последним шагом является обратное DCT преобразование этого массива с последующим формированием на выходе DCT преобразователя массива пикселей $f[x][y]$.

3.5.2. Декодирование кодов с переменной длиной

DC компонента Intra блока

DC компонента DCT преобразования для Intra блоков переработана по-другому, нежели другие DCT компоненты. Тех-

ника данного преобразования подобна импульсно-кодовой модуляции (PCM). В начале каждого слайса Intra кадра устанавливается предсказатель значения DC для каждого цветового плана.

В кодере VLC используется для кодирования параметра определяющего размер DC компоненты `dct_dc_size`. Если значение этого параметра 0, то DC коэффициент для этого блока является предиктором (предсказателем). Если его значение не равно 0, то оно указывает на длину идущего следом кода переменной длины (до 11 бит). Это дифференциальная величина, которая добавляется к предиктору для создания DC коэффициента.

Предикторы определяются как `dc_dct_pred[cc]`. В него включен 2-битный параметр, называемый `cc` (color component, цветовая компонента), который может принимать следующие значения: 00 для яркостного блока, 01 для C_b , 10 для C_r .

Другие DCT коэффициенты

Не Intra коэффициенты декодируются из других кодовых книг VLC. Возможны три случая:

1. *Конец блока.* В этом случае в блоке больше нет ненулевых коэффициентов и все значение остальных коэффициентов устанавливается равным 0.

2. *Обычный коэффициент.* Это комбинированное значение изменения длины и уровня следующее за отдельным битом, показывающим уровень.

3. *Исчезнувший код.* В этом случае используются специальные меры для кодирования изменения и уровня.

3.5.3. Инверсное сканирование

Декодированные VLC коды представляют одномерную версию квантованных DCT коэффициентов. В порядке, обратном квантованию, эти значения DCT коэффициентов подвергаются преобразованию в одномерный вектор. Этот одномерный вектор должен быть отображен в двумерную матрицу. Этот процесс называется инверсным сканированием.

3.5.4. Обратное квантование

Двумерный массив коэффициентов $QF[v][u]$ подвергается процессу обратному квантованию для восстановления DCT коэффициентов. Этот процесс, по существу, является умножением на шаг квантования. Этот шаг изменен двумя разными заголовками: весовая матрица ($W[w][v][u]$) модифицирует шаг квантования внутри блока, а скалярное значение `quant_scale` показывает, что шаг квантования может быть изменен на несколько бит. Результатом первого шага обратного квантования является матрица $F''[u][v]$, которая формируется следующим образом:

$$F''[v][u] = (2 * QF[v][u] + k) * (k * W[w][v][u] * (\text{quantizer_scale})) / 32,$$

где $k = 0$ для Intra блоков и $k = \text{sine}[QF[v][u]]$ для не Intra блоков.

Для MP@ML используются две весовые матрицы ($W[w][v][u]$), одна для Intra и одна для не Intra блоков. Параметр w определяется из параметра cs . Для формата 4:2:0 $cs = 0$ для Intra блоков и $cs = 1$ для не Intra блоков.

3.5.5 Обратное дискретное косинусное преобразование

И наконец, обратное дискретное косинусное преобразование создает заново пространственные значения для каждого блока, используя следующее выражение:

$$f(x, y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u, v) \cdot \cos[(2x+1)u\pi/2N] \cos[(2y+1)v\pi/2N], \quad (3.13)$$

где N равно 8 для блока размером 8×8 .

3.4.6. Компенсация движения

На рис. 3.15 показана примерная схема выхода устройства компенсации движения. Результат обратного дискретного косинусного преобразования $f[x][y]$ суммируется с результатом

предсказания движения $p[y][x]$. Эта сумма затем ограничивается диапазоном значений от 0 до 255 при помощи функции насыщения (ограничителя). На выходе ограничителя получаются декодированные пиксели, готовые для отображения.

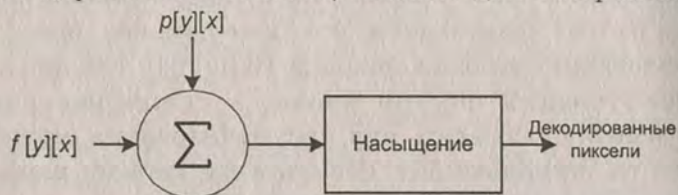


Рис. 3.15. Примерная схема процесса компенсации движения

3.5.7. Пропущенные макроблоки

При получении пропущенных макроблоках (у которых параметр `macroblock_address_increment` больше 1) декодер не получает ни информации о DCT коэффициентах, ни о векторе движения. В этом случае декодер должен сформировать предсказание путем использования конечных значений декодированных отсчетов.

3.6. Режимы предсказания

Существует два основных режима предсказания: предсказание поля и предсказание кадра.

В режиме предсказания полей предсказание делается независимо для каждого поля с использованием одного и более одного ранее предсказанных кадров. В режиме предсказания кадров предсказание формируется для кадра по одному или более ранее декодированному кадру.

3.7. Единый поток данных

3.7.1. Пакетный элементарный пакет

Видео, аудио и другие данные для MPEG-2 служб должны быть мультиплексированы в единый поток данных. Это является первой задачей раздела системы MPEG-2.

Сразу возникает следующий вопрос: когда мультиплексор получает сжатые параллельно видео- и аудиопотоки, то как они мультиплексируются так, чтобы декодер мог представить их синхронизировано (например, синхронизация движения губ и звука)? Вторая роль раздела системы MPEG-2 это обеспечения этой синхронизации.

Однородный поток данных MPEG представляет из себя непрерывный поток бит, которые нужно преобразовать в группы (пакеты) так, чтобы ошибки в этих битах не могли распространяться за границы единичного пакета. В основном чем больше длина пакета, тем больше он подвержен ошибкам. С другой стороны, группировка бит в пакеты создает дополнительные расходы ресурсов для передачи заголовков пакетов. Таким образом, идет выбор длины пакета между устойчивостью к ошибкам и эффективностью передачи. В любом случае формирование пакетов является третьей функцией уровня систем MPEG-2.

В большинстве случаев декодеры требуют наличия программно-зависимой информации (Program Specific Information, PSI) для декодирования входящего потока. Если суммировать все вышесказанное, то уровень систем MPEG выполняет следующее:

1. Мультиплексирование отдельных потоков в один поток данных.
2. Обеспечение способов синхронизации компонентов потоков, которые формируют аудио- и/или видеосервисы.
3. Пакетирование бит в группы.
4. Обеспечение PSI.

В системах MPEG-2 программа определяется как набор несущих полезную нагрузку элементарных потоков, таких как аудио и видео, которые имеют одну временную базу.

MPEG-2 регламентирует две возможных формы единого потока данных:

- ◆ программный поток;
- ◆ транспортный поток.

Первый шаг на пути получения единого потока — формирование пакетного элементарного потока PES (*Packetised Elementary Stream*), представляющего собой последовательность PES-пакетов (рис. 3.16).

Каждый пакет состоит из:

- ◆ заголовка
- ◆ данных пользователя или полезной нагрузки, которая представляет собой фрагменты исходного элементарного потока.

PES-пакеты могут быть переменной длины. В начале PES пакета идет 32-битный код старта, состоящий из стартового префикса и идентификатора потока (рис. 3.17). Идентификатор потока позволяет выделить PES-пакеты, принадлежащие одному элементарному потоку ТВ-программы. Флаги 1 и 2 — биты, указывающие на наличие или отсутствие в заголовке дополнительных полей, которые не являются обязательными. Эти поля служат для переноса дополнительной информации: авторские права, скремблирование, приоритет. Особую значимость имеют биты *P* и *D* флага 2, указывающие на наличие полей с метками времени представления *DTS* (*Decoding Time Stamps*), обеспечивающие синхронизацию потоков данных в декодере.

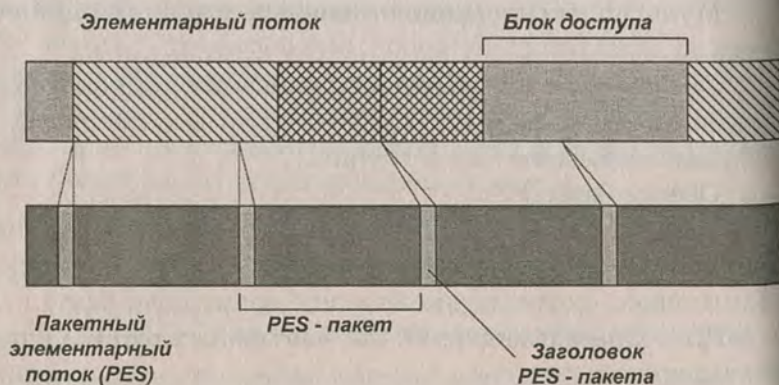


Рис. 3.16. Формирование PES

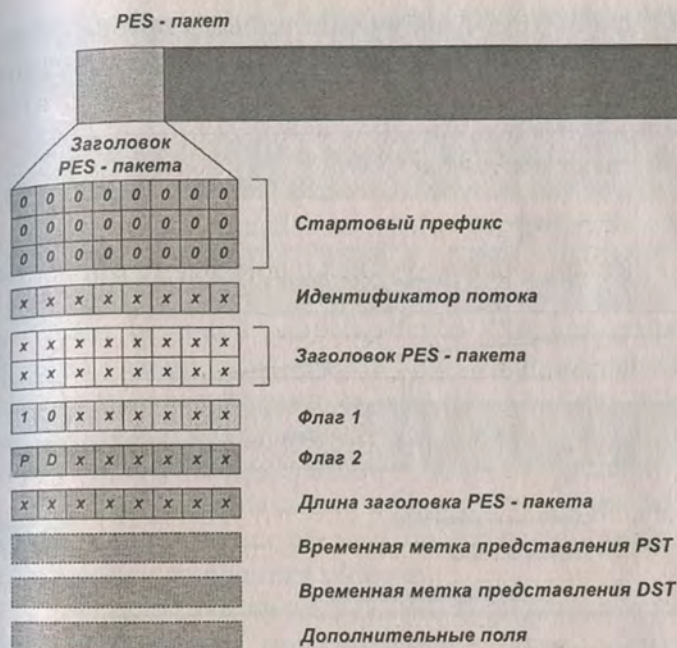


Рис. 3.17. Пакет PES

3.7.2. Программный поток

Программный поток объединяет элементарные потоки, образующие ТВ-программу. При формировании программного потока образуются блоки из PES-пакетов (рис. 3.18). Блок содержит:

- ◆ заголовок блока должен появляться не реже, чем через 0,7 с;
- ◆ системный заголовок — содержит информацию о характеристиках программного потока (скорость передачи данных, число видео и звуковых элементарных потоков). Декодер использует эту информацию, для того чтобы решить, может ли он декодировать это программный поток;
- ◆ определенное количество PES-пакетов.

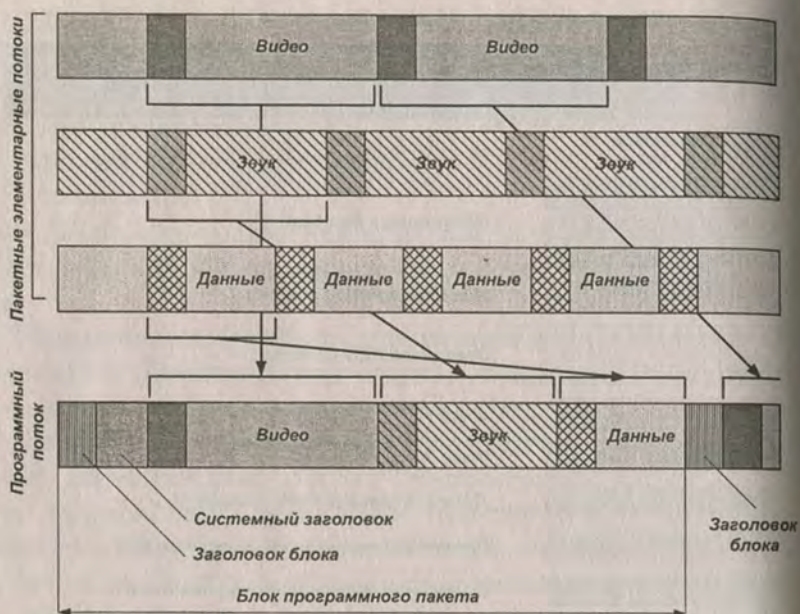


Рис. 3.18. Формирование программного потока

Программный поток объединяет элементарные потоки одной программы, имеющие общую временную базу.

3.7.3. Транспортный поток

Транспортные потоки (Transport Streams, TS) состоят из пакетов транспортных потоков (Transport Streams packets) с заголовками, содержащими информацию, которая определяет время, когда каждый байт должен поступить в декодер транспортного потока (*Transport Stream Decoder*) из канала.

Благодаря дополнительной операции мультиплексирования декодер может определить, какие ресурсы требуются для декодирования транспортного потока. Транспортный поток содержит информацию, позволяющую идентифицировать относящиеся к делу характеристики и взаимосвязь между элементарными потоками (*Elementary Streams, ES*), которые составляют каждую программу. Такая информация может вклю-

чать язык, на котором идет речь в аудиоканалах, а также взаимосвязь между видеопотоками, когда применяется многоуровневое кодирование.

На рис. 3.20 показано, что транспортный поток состоит из двух уровней: уровень систем (системный уровень) и уровень элементарного потока. Кодер транспортного потока (*Transport Stream encoder*) как бы складывает системный уровень около ES уровня.

Структура пакета транспортного потока дала толчок для нескольких решений разработчиков. С одной стороны, длина заголовка пакета должна быть как можно меньше для уменьшения расходов ресурсов на заголовок пакета, с другой стороны — возможности контроля должны быть как можно более полны, что подразумевает наличие длинного заголовка.

MPEG-2 системы имеют возможность достижения компромисса между этими обоими противоречащими друг другу требованиям следующим образом:

- ◆ Для доставки полезных данных используется четырехбайтовый заголовок фиксированной длины. Обычно 2 бита идентифицируют, какая информация в пакете является полезной нагрузкой или полем адаптации (*Adaptation field*).

- ◆ Если поле адаптации указано, то некоторое количество специальных параметров может быть включено в полезную нагрузку пакета.

Транспортный поток является конвертом для составляющих его элементарных потоков (*ES*). Однако элементарные потоки обычно независимы друг от друга. Например, аудио-и/или видеосервисы имеют свои аудио *ES* и видео *ES*. Эти два элементарных потока связаны и называются *программа*. Элементарные потоки, составляющие каждую программу, должны быть синхронизированы.

Данные элементарных потоков упаковываются в пакеты пакетированного элементарного потока (*Packetized Elementary Stream, PES*). Пакет *PES* состоит из заголовка следующим за данными. Пакеты *PES* вставляются в пакеты транспортного потока. Первый байт каждого заголовка пакета *PES* распола-

гается на первом доступном месте среди полезных данных пакета транспортного пакета.

В транспортном потоке переносятся и таблицы программно-зависимой информации (*Program Specific Information, PSI*). Существует четыре типа *PSI* таблиц:

- ◆ Таблица объединения программ (*Program Association Table, PAT*);
- ◆ Таблица состава программы (*Program Map Table, PMT*);
- ◆ Таблица условного доступа (*Conditional Access Table, CAT*);
- ◆ Таблица сетевой информации (*Network Information Table, NIT*).

Эти таблицы содержат информацию, необходимую для демультиплексирования и представления программ. Таблица *PMT* описывает остальную информацию о идентификаторах пакетов (*packet IDs, PIDs*) и благодаря этому информацию о том, какой элементарный поток к какой программе относится.

Так как основным направлением применения виделось использование цифровых накопителей медиа данных (*digital storage media, DSM*) с присущими им скоростями с очень малым числом ошибок, системы *MPEG-1* имеют очень большую длину пакета. Программные потоки (*Program Streams*) *MPEG-2* разработаны для тех же типов приложений. С другой стороны, одним из основных типов приложений для *MPEG-2* является широко вещание и использование в транспортных средах подверженных ошибкам. Таким образом, часть транспортных потоков систем *MPEG-2* предназначена для *DBS*.

Транспортный поток состоит из последовательности пакетов фиксированной длины (188 байт). Примечание: 188 байт не является очевидной длиной, пока не будут обоснованы несколько специфичных факторов. В укороченном блочном корректирующем коде Рида—Соломона 188 — это нормальное число информационных бит, чтобы увидеть это, вспомним, что t корректирующий код Рида—Соломона определяется как:

$$\text{Длина блока: } n = q - 1$$

$$\text{Число проверок на четность: } n - k = 2 \times t$$

$$\text{Минимальная длина (кодовое расстояние): } d_{\min} = 2 \times t + 1$$

Если мы зададим $q = 28 = 256$, тогда $n = 255$. Если мы затем зададим $t = 8$, то восемь байт ошибок могут быть исправлены, как результат мы получим (255, 239) код. Этот код является стандартным кодом NASA на протяжении нескольких лет. Мы можем сформировать код с $n'' = n - 1$ и $k'' = k - 1$, который известен как укороченный код Рида—Соломона. Задавая $l = 51$, мы получим (204, 188) код Рида—Соломона. (204, 188) код Рида—Соломона используется, например, в стандарте DVB. С его помощью возможна коррекция 8-байтовых ошибок в каждом пакете длиной 204 байта.

Второй причиной, из которой вытекло значение 188, — это длина пакетов ATM. Рис. 3.19 иллюстрирует структуру пакета ATM. Пакет состоит из 5-байтового заголовка, 1 байт уровня адаптации ATM (AAL) и 47 байт данных. Так как $4 \times 47 = 188$, то транспортный пакет MPEG-2 может быть введен в 4 ATM пакета.

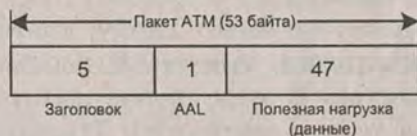


Рис. 3.19. Структура пакета ATM

Скорость транспортного потока

Транспортный поток определяет поток данных, который оптимизирован для связи или хранения одной или более программ кодированных данных MPEG-2 и других данных в средах подверженных ошибкам. Скорость транспортного потока определяется положением полей (меток) PCR (*Program Clock Reference*, ссылка на программные часы).

В основном для каждой программы есть своя метка PCR. Один транспортный поток может содержать несколько программ, каждая из которых имеет свою временную базу. Это необходимо для DBS, так как программные материалы, мультиплексированные в транспортный поток для индивидуального транспондера, достигают спутникового канала связи при помощи

различных средств. Каждый DBS транспондер имеет ограниченную полосу пропускания, поэтому скорость транспортного потока является постоянной для каждого транспондера.

Транспортный поток может объединять пакетные элементарные потоки, переносящие данные нескольких программ с независимыми временными базами. Он состоит из коротких пакетов фиксированной длины (188 байт). Элементарные потоки видео, звука и дополнительных данных (например, телетекст) разбиваются на фрагменты, равные по длине полезной нагрузке транспортного пакета (184 байта) и мультиплексируются в единый поток (рис. 3.20). Это процесс подчиняется ряду ограничений:

- ◆ первый байт каждого PES-пакета элементарного потока должен быть первым байтом полезной нагрузки транспортного пакета;
- ◆ каждый транспортный пакет может содержать данные лишь одного PES-пакета;
- ◆ если PES-пакет не имеет длину, кратную 184 байтам, то один из транспортных пакетов не заполняется данными PES-пакета полностью. В этом случае избыточное пространство заполняется полем адаптации. Транспортные пакеты, переносящие разные элементарные потоки, могут появляться в произвольном порядке, но пакеты, принадлежащие одному элементарному потоку, должны следовать в транспортном потоке в хронологическом порядке, т. е. в порядке их "нарезания" из PES-пакета.

Структура транспортного потока оптимизирована для условий передачи данных в каналах связи с шумами. Это проявляется, прежде всего, в небольшой длине пакетов. Типичные параметры защиты от ошибок данных транспортного потока дают системы цифрового телевизионного вещания. В системах DVB и ISDB к 188 байтам каждого пакета добавляются 16 проверочных байтов кода Рида—Соломона, что позволяет исправлять в каждом пакете до 8 пораженных шумами байтов. В ATSC к каждому пакету добавляется 20 проверочных байтов, что позволяет исправить до 10 байтовых ошибок в одном пакете.

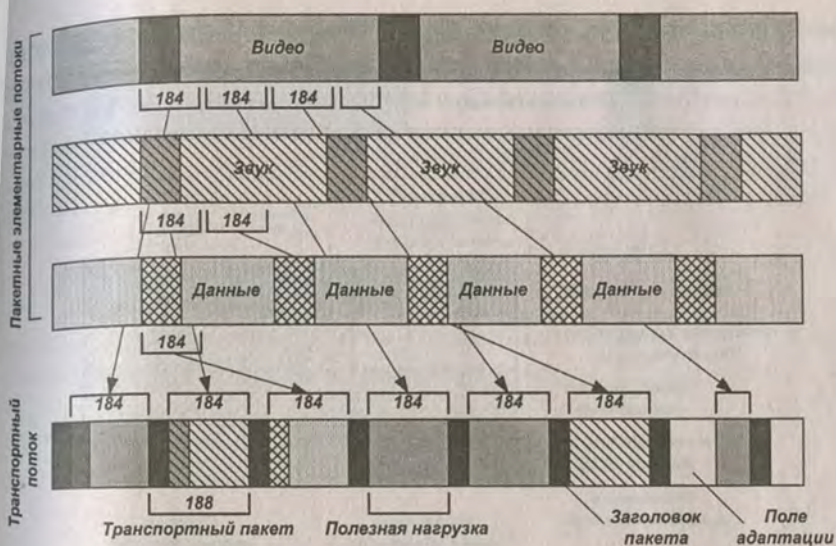


Рис. 3.20. Структура транспортного потока

3.7.4. Транспортный пакет

Транспортный пакет начинается с 4-байтного заголовка (рис. 3.21), первый байт которого — синхронизирующий (число 47 в шестнадцатеричном коде). Это значение не является уникальным и может появляться в других полях транспортного пакета. Определение начала пакета упрощает то, что заголовки всегда следуют с интервалом в 188 байт.

Транспортный поток может переносить несколько ТВ-программ, состоящих из набора элементарных потоков. Для опознавания пакетов, принадлежащих одному элементарному потоку, используется 13-битный идентификатор. Из 2^{13} возможных значений 17 зарезервировано для специальных полей, а остальные 8175 могут использоваться для присвоения в качестве номеров элементарным потокам. Таким образом, один транспортный пакет может переносить до 8175 элементарных потоков.

Важный компонент заголовка — счетчик непрерывности, который инкрементируется в последовательных транспорт-



Рис. 3.21. Транспортный пакет

ных пакета, принадлежащих одному и тому же элементарному потоку. Это позволяет декодеру обнаруживать потерю транспортного пакета и принимать меры к маскированию ошибок, которые могут возникнуть из-за потери.

Поле адаптации не является обязательным. Оно может использоваться не только для заполнения “пустот”. Это поле также переносит важную дополнительную информацию об использовании данных пакета, например, опорное время программы PCR (*Program Clock Reference*).

3.7.5. Заголовок транспортного пакета

Заголовок транспортного пакета содержит всю информацию, необходимую декодеру для декодирования различных структур пакетов. Рисунок 3.22 показывает структуру заголовка. Согласно этому рисунку, пакеты транспортного потока могут содержать три типа данных: пакетированные элементарные потоки (PES), программно-зависимую информацию (PSI) и поля адаптации. Только один из этих типов дан-

ных может быть передан в одном пакете. PSI пакеты состоят из четырех типов пакетов. Причем только один из них может присутствовать в одном пакете транспортного потока.



Рис. 3.22. Структура заголовка транспортного потока

Достаточно поучительно рассмотреть то, как декодер может произвести анализ заголовка транспортного пакета. Сначала проверяется значение метки `adaptation_field_control`, равно ли оно 2. Если это так, то пакет содержит поле адаптации и начинается последующий его анализ. Если значение метки `adaptation_field_control` равно 1 или 3, то декодер продолжает производить анализ заголовка для определения какие данные содержатся в пакете, PES или PSI. Если на этом этапе значение метки не равно 1 или 3, то пакет отбрасывается как не содержащий полезных данных.

Если же значение метки `adaptation_field_control` равно 1 или 3, то декодер определяет значение PID. И PES, и PSI данные могут содержать персональные данные. Если пакет несет поле адаптации, то флаг `transport_stream_data_flag` проверяется. Если значение метки равно "1", то пакет несет персональные данные, если "0", то пакет переносит другой тип поля адаптации.

Следующий ниже список показывает составляющие заголовка пакета транспортного потока.

Все пакеты содержат синхронизирующие байты, которые позволяют декодеру синхронизировать поток данных. Значение `sync_byte` равно "0100 0111" (0x47).

Когда поток содержит хотя бы одну ошибку, то флаг `transport_error_indicator` устанавливается равным "1".

`Payload_unit_start_indicator` — 1-й флаг, показывающий, то, какие данные содержатся в пакете (пакеты PES или PSI данные). Если значение флага "1", то пакет содержит первый байт пакета PES или первый байт PSI данных. Когда значение флага "0", то пакет не несет этого байта. Когда транспортный пакет содержит PSI данные и значение флага `packet_unit_start_indicator` равно "1", то первый байт после заголовка — `pointer_field`.

`Pointer_field` (только PSI данные) — это 8-байтовое поле, значение которого показывает число байт следующих непосредственно за `pointer_field` до первого байта первой секции полезных данных содержащихся в пакете.

Флаг `transport_priopity` со значением "1" показывает, что пакет имеет наибольший приоритет, чем другие пакеты с тем же PID, но со значением флага "0".

PID (идентификатор пакета) — это 13-байтовое поле, которое определяет полезную нагрузку пакета. Определены следующие значения PID: Таблица объединения программ (*Program Association Table, PAT*) и Таблица условного доступа (*Conditional Access Table, CAT*). Остальные значения PID зарезервированы.

`Transport_scrambling_control` — 2-битовое поле, показывающее режим кодирования (шифрования) данных, содержащихся в транспортном пакете. Заголовки пакетов транспортного потока включая поле адаптации не могут быть закодированы.

Пакеты с одним и тем же PID подсчитываются 4-битным счетчиком и представлены в заголовке транспортного потока как `contitunity_counter`. Этот флаг не увеличивает свое значе-

ние, когда в качестве полезной нагрузки пакет несет поле адаптации.

Когда флаг `discontinuity_counter` установлен в истинное значение в любом пакете транспортного потока, то флаг `contitunity_counter` в том же пакете может прерываться по отношению к предыдущему пакету транспортного потока с тем же PID.

3.7.6. Временная модель

MPEG-2 Системы, Видео и Аудио имеют временную модель, в которой сквозная задержка от входа в кодер до выхода декодера является величиной постоянной. Данная задержка вычисляется как сумма всех задержек от кодирования сигнала до его представления. Интервалы между кадрами и аудио отсчетами в кодере и декодере являются величинам одинаковыми.

3.7.7. Условный доступ

В коммерческих системах цифрового телевидения весьма велика роль условного доступа (Conditional Access, CA) для DBS систем, поэтому MPEG-2 должен поддерживать CA. Шифрование (кодирование) для условного доступа к программам закладывается в транспортном потоке системного уровня MPEG-2 (уровня систем).

3.7.8. Multiplex-wide operations

Данные операции включают в себя координацию восстановления данных в канале, установку часов, управление буферами (организация буферизации данных). В основном для целей широко вещания, декодеры элементарных потоков должны подчиняться синхронизации с данными поступающими из канала для избежания переполнения или опустошения буферов данных.

3.7.9. Transport Stream System Target Decoder

Это гипотетический декодер, известный как Transport Stream System Target Decoder (T-STD), используемый для оп-

ределения временных задержек. Время прибытия байтов и их декодирование не устанавливаются уровнем систем MPEG-2. Для получения более подробных сведений и этой модели см. разд. 2.4.2 ISO/IEC 13818—1.

3.8. Операции с отдельными потоками (уровень пакетов PES)

Демультимплексирование и синхронизация многочисленных элементарных потоков (ES) принципиально являются операциями проводимыми с потоками.

3.8.1. Демультимплексирование

Элементарные потоки мультиплексируются в транспортные потоки. Аудио- и видеопотоки являются основными элементарными потоками. В пакете PES может быть только один элементарный поток. Пакеты PES обычно имеют относительно большую длину. ID коды пакетов в транспортном потоке дают возможность произвести необходимое для реконструкции элементарных потоков декодирование и демультимплексирование.

3.8.2. Синхронизация

Временные метки позволяют производить синхронизацию между множеством элементарных потоков. Ссылка на системные часы (*System Clock Reference, SCR*) и ссылка на программные часы (*Program Clock Reference, PCR*) и такая опция, как ссылка на часы элементарного потока (*Elementary Stream Clock Reference, ESCR*), имеют величину с разрешением до 27 МГц. Так как представление временных меток используется для декодирования отдельных элементарных потоков, то они присущи уровню пакетов PES транспортных потоков.

Временная метка PCR позволяет синхронизировать систему декодирования с каналом. Метка PCR кодирует распределение во времени потока данных и получается из времен-

ной базы используемой для аудио и видео. Так как каждая программа может иметь свою временную базу, то для каждой программы существуют свои отдельные PCR поля в транспортном потоке, состоящем из множества программ. Каждая программа может иметь только одну свою PCR метку.

3.8.3. Связь с уровнем сжатия

Пока сжатые аудио- и видеоэлементарные потоки содержат только свою информацию для декодирования они являются полезной нагрузкой для пакетов PES. Пакеты PES не интересуют находящиеся в них полезные данные. Например, стартовые коды видео могут находиться в середине пакета PES. Однако временные метки, закодированные в заголовках PES пакетов (PES Packet Headers), используются для представления времени в процессах, происходящих на уровне сжатия.

3.8.4. Пакеты PES

Пакетированный элементарный поток (*Packetized Elementary Stream, PES*)

Транспортные потоки для мультимедиа контента логически состоят из пакетированных элементарных потоков (PES), как это показано на рис. 3.16. Поля ссылки на часы элементарного потока (ESCR) и поля скоростей элементарного потока (*Elementary Stream Rates, ES_Rate*) должны быть включены в пакеты PES. PES пакеты являются основной формой полезной нагрузки транспортных пакетов (*Transport Packet*). Структура этих пакетов рассмотрена ниже.

Структура пакетов PES

Пакет PES начинается с заголовка пакета (*Packet Header*). Заголовок пакета PES намного длиннее и сложнее, чем остальные заголовки в MPEG-2 Системах.

Как и в большинстве пакетов, первым параметром является синхронизация или `start_code`.

`packet_start_code_prefix` — это длинная строка — “0000 0000 0000 0000 0001” (0x000001). Обычно служит для того, чтобы не происходила эмуляция стартового кода.

stream_id определяет тип и количество элементарных потоков и может принимать значения в диапазоне от 0xBC до 0xFF. Значение 0xBC указывает на **Program_Stream_Map**. Все ID, которые начинаются с 0xC или 0xD, используются для аудиопотоков. ID (идентификационные номера) 0xE, используются для видеопотоков. Идентификационные номера 0xF0 и 0xF1 используются для контрольных сообщений (*Entitlement Control Message, ECM*) и (*Entitlement Management Message, EMM*).

PES_packet_length — это 16-битное поле, определяющее количество байт в пакете PES. Максимальная длина PES пакета — 65, 536 байт.

PES_scrambling_control — это 2-битное поле, показывающее режим скремблирования полезной нагрузки пакета PES.

Следующими 11 параметрами заголовка являются следующие флаги: **PES_priority_data_alignment_indicator**, **copyright**, **original_or_copy**, **PTS_DTS_flags**, **ESCR_flag**, **ES_rate_flag**, **DSM_trick_mode_flag**, **additional_copy_info_flag**, **PES_CRC_flag** и **PES_extention_flag**. Все эти параметры представляют из себя 1-битные флаги, показывающие наличие или отсутствие условий для последующей обработки данных, за исключением флагов **PTS_DTS_flags**, которые имеют размер 2 бита. Значения флагов **PTS_DTS_flags** показаны в табл. 3.8.

Если значение флага **PES_extention_flag** равно "1", то 8-битное поле **PES_header_data_length** описывает общее количество байт занятых каким-либо необязательным полем и/или какими-либо холостыми битами в данном заголовке PES пакета. **marker_bit** — это 1-битовое поле, которое всегда имеет значение "1".

Следующие три параметра представляют из себя различные временные метки (если они присутствуют): **PTS (presentation_time_stamp)**, **DTS (decoding_time_stamp)** и **ESCR (Elementary_Stream_Clock_Reference)**. Метки PTS и DTS показывают время, необходимое для представления и декодирования. Обе эти метки представляют из себя 33-битные числа, которые измеряются в числе периодов 27 МГц.

Возможные значения PTS_DTS_flags

PTS_DTS_flags	Значение
00	Не присутствует ни PTS, ни DTS поле
01	Запрещенное значение
10	Присутствует поле PTSD
11	Представлены оба поля (PTS и DTS)

часов, деленным на 300. **ESCR** — это 42-битное значение, составленное 33-битной базы (**ESCR_base**) и 9-битного расширения (**ESCR_extension**). База измеряется в периодах частоты 90-кГц генератора.

Поле **ES_rate** (**elementary_stream_rate**) представляет из себя положительное целое число, задающее скорость потока PES.

В MPEG-2 видео макроблок представляет из себя блок размером 16×16 пикселей. Если некоторые макроблоки пропущены, то значение флага **intra_slice_refresh** устанавливается равным "1". Это может быть полезно при скрытии ошибок.

2-битное поле **frequency_truncation** может показать, что ограниченный набор коэффициентов может быть использован для кодирования видеоданных.

7-битовое поле **additional_copy_info** содержит информацию о персональных данных, относящуюся к авторским правам.

Персональные данные в PES пакетах хранятся в 16-битных полях **PES_private_data**. Эти данные не должны эмулировать поле **packet_start_code_prefix** (0×000001).

Длина поля **packet_header_field** в байтах показывается 8-байтовым полем **packet_field_length**.

Редкие пакеты PES подсчитываются 7-битным флагом **program_packet_sequence_counter**.

3.9. Программно-зависимая информация

Программно-зависимая информация (*Program Specific Information, PSI*) включает в себя большую часть данных

MPEG-2 систем и персональных данных, позволяющих производить демультиплексирование программ декодером. Доступ к программам элементарных потоков или их частям может быть получен условно. Обычно PSI не может быть зашифрована.

В транспортных потоках PSI состоит из четырех таблиц. Таблицы, составляющие PSI-пакеты, имеют сходную структуру (см. табл. 3.9). Общие поля выделены жирным шрифтом.

Таблица 3.9

Таблицы PSI

Таблица объединения программ (PAT)	Таблица состава программ (PMT)	Таблица условного доступа (CAT)	Private Section
table_id	table_id	table_id	table_id
section_syntax_indicator	section_syntax_indicator	section_syntax_indicator	section_syntax_indicator
section_length	section_length	section_length	section_length
transport_stream_id	program_number		private_indicator
version_number	version_number	version_number	* version_number
current_next_indicator	current_next_indicator	current_next_indicator	* current_next_indicator
section_number	section_number	section_number	* section_number
last_section_number	last_section_number	last_section_number	* last_section_number
program_number	PCR_PID		* private_section_length
network_PID	program_info_length		* private_data_byte
program_map_PID	stream_type		* table_id_extention
	elementary_PID		
	ES_info_length		
CRC_32	CRC_32	CRC_32	CRC_32

3.9.1. Типы общих полей PSI

Так как некоторое количество полей в таблицах являются общими для четырех таблиц, то сначала будут рассмотрены именно они.

table_id — 8-битное поле идентифицирующее содержание PSI секции транспортного потока следующим образом: 0x00 определяет **program_association_table**; 0x01 указывает на **conditional_access_section**; значения от 0x03 до 0x0F яв-

ляются резервными; значения от 0x40 до 0xFE указывают на частные пользовательские данные (User Private); значение 0xFF является запрещенным значением.

section_syntax_indicator — 1-битное поле, которое всегда должно иметь значение “1”.

section_length — 12-битное поле, первые два бита которого “00”. Оно показывает количество байт раздела, начинающегося непосредственно за ним, включая CRC.

5-битовое поле **version_number** показывает номер версии таблицы объединения программ (*Program Association Table*, PAT). Всякий раз, когда таблица объединения программ меняется, то номер версии увеличивается на единицу. Когда достигается значение 31, то оно сбрасывается до 0. Флаг **current_next_indicator** определяет, к какой таблице PAT относится поле **version_number** (к той, которая применяется в данный момент времени или к следующей). Если значение **current_next_indicator** установлено равным “1”, то в данный момент должна применяться передаваемая PAT. Если значение указателя “0”, то это означает, что передаваемая таблица еще не должна применяться, но она будет следующей действительной таблицей.

8-битное поле **section_number** указывает номер передаваемой секции (если таблица разбита на несколько секций). **last_section_number** — определяет номер последней секции таблицы (секция с наибольшим номером) полной таблицы объединения программ.

3.9.2. Таблица объединения программ (PAT)

Все транспортные потоки содержат один или несколько пакетов транспортных потоков со значением PID 0x0000. Все вместе эти пакеты транспортных потоков содержат полный список всех программ внутри транспортного потока. Любые изменения в программах, переносимых транспортным потоком, описаны в обновленной версии таблицы объединения про-

грамм (*Program Association Table, PAT*), передаваемой пакетами транспортного потока со значением PID равным 0x0000. Эти секции используют значение указателя **table_id**, равное 0x0000. Секции только с этим значением **table_id**, могут находиться в пакетах транспортного потока со значением PID 0x0000.

Максимальное количество байт в секции MPEG-2 систем определено размером PSI-таблицы — 1 Кб (1 024 байт). Максимальное количество байт в поле **private_section** равно 4 Кб (4096 байт). Здесь нет ограничений на расположение стартовых кодов, байтов синхронизации и других подобных битов в PSI данных.

Таблица распределения программ обеспечивает соответствие между **program_number** (номер программы) и значением PID (идентификатор пакета) пакетов транспортного потока. **Program_number** — это числовой указатель, соответствующей программы. Номер программы 0x0000 зарезервирован для определения сетевого PID. Он указывает на пакеты транспортного потока, которые переносят таблицу сетевой информации (*Network Information Table*).

Определение уникальных полей в разделе PAT

Транспортный поток может быть выделен из любого мультиплексного потока в сети с помощью 16-битного поля **transport_stream_id**.

Программа, к которой относится **program_map_PID**, устанавливается с помощью 16-битного поля номера программы **program_number**. Если значение этого поля 0x0000, то тогда следующая PID ссылка должна быть на сетевой PID. Номер программы может использоваться, например, как указатель вещательного канала.

Поле **network_PID** является 13-битным полем, задающим таблицу сетевой информации (*Network Information Table*).

Пакеты транспортного потока, которые должны содержать **program_map_section**, применяются для программ, задаваемых 13-битным полем **program_map_PID**.

3.9.3. Таблица состава программы (Program Map Table, PMT)

Все транспортные потоки должны содержать один или более транспортный пакет со значением идентификатора пакета (PID) отведенным для таблицы состава программы, так как пакеты транспортного потока содержат секции карты (состава) программы (Program Map). Каждая программа, перечисленная в таблице PAT, описана как отдельная секция карты программы транспортного потока (Transport Stream Program Map).

Последняя переданная версия **TS_program_map_section** с установленным в "1" значением индикатора **current_next_indicator** всегда используется для текущих данных в транспортном потоке. Любые изменения значений в любых программах переносимых в транспортном потоке описываются в обновленной версии передаваемой секции таблицы состава программы переносимой в пакетах транспортного потока со значением идентификатора пакета (PID) определенного как **program_map_id** (идентификатор состава (карты) программы) для определенных программ. В течение непрерывного наличия программы идентификатор **program_map_id** не может быть изменен.

Секции со значением PID 0x02 содержат информацию таблицы состава программы. Эти секции могут быть переданы в пакетах транспортного потока с различными значениями идентификаторов пакетов (PID).

Характеристики программы (program definition) отображаются между номером программы и элементарными потоками составляющими ее. Таблица состава программы обеспечивает полный набор всех характеристик программы для транспортного потока. Таблица может быть разбита на несколько секций до момента ее включения в пакеты транспортного потока.

Описание уникальных полей в секции состава программы транспортного потока

Program_number — см. описание таблицы объединения программ (PAT).

PCR поля, действительные для программ заданных идентификатором **program_number**, должны быть определены 13-битным идентификатором **PCR PCR_PID**.

Число байт дескрипторов показывается 12-битным полем **program_info_length**, первые два бита которого должны иметь значение 00.

Тип элементарного потока или данных, составляющих полезную нагрузку, задается идентификатором **elementary_PID**, указываемому в 8-битном поле **stream_type** (тип потока).

Идентификатор пакетов транспортного потока (PID), которые переносят соответствующие элементарные потоки или полезные данные, указывается 13-битным идентификатором **elementary_PID**.

Дескрипторы элементарного потока

Дескрипторы элементарного потока (*Elementary Stream Descriptors*) дают дополнительную информацию об элементарных потоках. Все дескрипторы потока имеют формат, начинающийся со значения 8-битного тага (указателя). За значением тага следуют 8-битный дескриптор и поле данных.

Дескрипторы потока

8-битные поля — **descriptor_tag**, которое идентифицирует каждый дескриптор и длина дескриптора **descriptor_length**, которое описывает число байтов следующего за ним дескриптора — общие для всех дескрипторов потока.

8-битная структура поля **descriptor_tag** позволяет задать 256 дескрипторов, однако используется только 13. Значения тага 2 (**video_stream_descriptor**), 7 (**target_background_grid_descriptor**) и 8 (**video_window_descriptor**) являются атрибутами потоков видео.

Дескрипторы видеопотока

Дескриптор видеопотока **video_stream_descriptor** определяет ключевые параметры элементарных потоков видео. В

частности, он определяет код скорости кадров. Профиль и уровень также показаны вместе с остальными параметрами.

Дескриптор сетки заднего фона **target_background_grid_descriptor** описывает то, как расположено изображение тогда, когда зона изображения больше, чем размер полученной декодером видеокартинки.

Дескриптор окна видео **video_window_descriptor** используется для описания характеристик окна соответствующего элементарного потока. Его значения ссылаются на дескриптор **target_background_grid_descriptor** того же элементарного потока.

Дескриптор аудиопотока (значение тага 3)

Дескриптор аудиопотока **audio_stream_descriptor** дает основную информацию, которая определяет версию кодирования элементарного потока аудиоданных.

Дескриптор регистрации (значение тага 5)

Дескриптор регистрации **registration_descriptor** обеспечивает возможность однозначной идентификации форматов персональных данных. Он включает в себя 32-битное значение, полученное из авторских прав путем, разработанным ISO.

Дескриптор условного доступа (значение тага 9)

Если в транспортном потоке присутствует информация об управлении условным доступом (*Conditional Access*) в системе, то на соответствующей секции потока должен находиться дескриптор условного доступа (**CA_descriptor**). Этот дескриптор используется для определения большей части информации об управлении условным доступом.

ISO 369 Language Descriptor (значение тага 10)

Часть 2 документа ISO369 определяет 3 характеристических языковых кода. **ISO_369_language_descriptor** может быть использован для определения языка соответствующего элементарного потока.

Дескриптор системных часов (System Clock Descriptor, значение тага 11)

Этот дескриптор передает информацию о системных часах, которые используются для генерации временных меток. Точность частоты часов (генератора)

$\text{clock_frequency_integer} * 10^{-\text{clock_accuracy_exponent}}$ ppm,

где $\text{clock_accuracy_exponent}$ — это 3-битное число, а $\text{clock_frequency_integer}$ — это 6-битное число.

Если используется ссылка на внешние часы, то применяется индикатор ссылки на внешние часы $\text{external_clock_reference_indicator}$. Если системные часы имеют точность более чем требуемая 20-ppm, то точность часов может быть передана путем кодирования полей clock_accuracy .

Если значение переменной $\text{clock_frequency_integer} = 0$, то точность системных часов будет равна 20-ppm. Если используются обе части дескриптора, то точность часов соответствует точности внешних часов.

Дескриптор авторских прав (Copyright Descriptor, значение тага 13)

Дескриптор $\text{copyright_descriptor}$ содержит идентификатор авторских прав $\text{copyright_identifier}$, 32-битное значение которого получается путем регистрации авторских прав.

Дескриптор максимальной скорости (Maximum Bitrate Descriptor, значение тага 14)

Дескриптор $\text{maximum_bitrate_descriptor}$ кодируется как 22-битное положительное целое число в поле, отражающем значение максимальной скорости потока в единицах, соответствующих 50 байтам в секунду.

3.9.4. Таблица условного доступа (Conditional Access Table)

Пакеты транспортного потока должны переносить соответствующие дескрипторы условного доступа CA_descriptor

всякий раз, когда в транспортном потоке кодируются один или более элементарных потоков. Пакеты транспортного потока со значениями PID 0x0001 содержат эти дескрипторы. Полная версия таблицы условного доступа должна содержаться в декодере. Когда значение индикатора **urgent_next_indicator** равно 1, эта таблица требуется для зашифрованных (закодированных) файлов в транспортных потоках. Любые изменения в кодировании (шифровании) делают существующую таблицу ISO неверной или неполной и описываются в обновленной таблице условного доступа. Все эти секции используют идентификатор таблицы **table_id** со значением 0x01. В пакетах транспортного потока с PID 0x0001 возможно применение секций только с этим значением идентификатора таблицы.

Таблица условного доступа обеспечивает соответствие между одной или более СА систем, их ЕММ потоками и другими соответствующими им специальными параметрами. Поля условного доступа используют только общие поля.

3.9.5. Таблица сетевой информации (Network Information Table)

Таблица сетевой информации является необязательной таблицей. Если она присутствует, то она переносится в пакетах транспортного потока, имеющих одно и то же значение идентификатора пакета (PID) — **network_PID**. Значение идентификатора **network_PID** определяется пользователем и используется в таблице объединения программ.

3.9.6. Поле адаптации (Adaptation Field)

Поле адаптации указывается 2-битным флагом в заголовке транспортного пакета. Посылая нечастые изменения параметров, можно передать ценную вспомогательную информацию без существенного уменьшения всей эффективности заголовка (без сильного увеличения его размера).

3.10. Кодирование аудиоданных в стандарте MPEG-2 ISO/IEC 13818-3

Он дополняет стандарт MPEG-1, расширяя его на всю совокупность конфигураций звуковых форматов, предусмотренных документом 10/63E ITU-R. В MPEG-2 ISO/IEC 13818-3 [1], может быть организована передача звуковых сигналов следующих форматов:

- ◆ 3/2 (L, R, C — соответственно, левый, правый и центральный фронтальные сигналы; LS, RS — левый и правый пространственные сигналы) плюс канал сверхнизких частот СНЧ;

- ◆ 3/0 (L, R, C) плюс 2/0 ($L2, R2$ — дополнительные второй левый и второй правый сигналы);

- ◆ 3/1 (L, C, R, S , где S — сигнал “окружения” — Surround);

- ◆ 2/0 (L, R) плюс 2/0 ($L2, R2$);

- ◆ 3/0 (L, C, R);

- ◆ 2/1 (L, R, S);

- ◆ 2/0 (L, R — обычное стерео, ISO/IEC 11172-3);

- ◆ 1/0 (моно, ISO/IEC 11172-3).

При этом пространственные сигналы LS и RS имеют полосу частот 100...7000 Гц, как это принято в системах Dolby-Surround и Dolby-Pro-Logic. Полоса частот сигнала C ограничена сверху значением 9000 Гц и лишь сигналы L и R имеют полную полосу частот 20...20000 Гц.

В стандарте MPEG-2 ISO/IEC 13818-3 расширен также и состав применяемых кодеков и используемых частот дискретизации: дополнительно к ранее предусмотренным значениям частот дискретизации 32, 44, 1 и 48 кГц (MPEG-1) добавлены значения 16, 22, 05 и 24 кГц. Он позволяет передавать многоканальные стереофонические сигналы форматов 3/2 + СНЧ, Dolby-Surround, Dolby-Pro-Logic, а также одновременно: сигналы обычной стереофонии и плюс речь на двух языках; моносигнал и дополнительно до семи комментаторских каналов

с более низким качеством, используя частоты дискретизации 16, 22, 05 и 24 кГц.

Структура многоканальной системы передачи стандарта MPEG-2 ISO/IEC 3818-3 представлена на рис. 3.24. В соответствии с документом 10/63E ITU-R исходным на студийной стороне является пятиканальный сигнал вида L, C, R, LS, RS — соответственно, левый, центральный, правый фронтальные и левый и правый пространственные сигналы. На передающей стороне с помощью матрицы $M1$ они преобразуются в новую совокупность сигналов, два из которых ($T1, T2$) являются сигналами обычной двухканальной стереофонии (для них выполняется требование совместимости: $T1 + T1 = M$, где M — полноценный монофонический сигнал), а три сигнала $T3, T4$ и $T5$ представляют собой так называемое многоканальное расширение. Уравнения матрицирования имеют вид:

$$T1 = \alpha(L + \beta C + \gamma LS),$$

$$T2 = \alpha(R + \beta C + \gamma RS);$$

$$T3 = \alpha\beta C, T4 = \alpha\gamma LS, T5 = \alpha\beta RS,$$

где α, β, γ — коэффициенты матрицирования, их значения зависят от режима работы системы передачи. Коэффициент α необходим для устранения возможной перегрузки при одновременной передаче сигналов $T1, \dots, T5$. Всего возможных режимов работы четыре:

$$0 - \alpha = 1/(1+\sqrt{2}), \beta = 1/\sqrt{2}, \gamma = 1/\sqrt{2};$$

$$1 - \alpha = 1/(1,5+0,5\sqrt{2}), \beta = 1/\sqrt{2}, \gamma = 0,5;$$

$$2 - \alpha = 1/(1+\sqrt{2}), \beta = 1/\sqrt{2}, \gamma = 1/\sqrt{2};$$

$$3 - \alpha = \beta = \gamma = 1.$$

Режим 2 соответствует передаче сигналов системы Dolby-Surround. Поэтому перед кодированием должен быть сформирован сигнал окружения $jS = 0,5(jLS + jRS)$, который в

противофазе далее должен быть добавлен к сигналам L и R , как это обычно и выполняется в кодере системы Dolby.

После матрицирования полученные сигналы T_1, T_2, \dots, T_5 кодируются (см. рис. 3.23). Для кодирования сигналов T_1 и T_2 используется кодер MPEG-1 стандарта ISO/IEC 11172-3 со всеми имеющимися в нем уровнями компрессии Layer-1, Layer-2 и Layer-3. Дополнительные сигналы T_3, T_4 и T_5 кодируются отдельно в так называемом MPEG-2 кодере многоканального расширения. Далее оба цифровых потока форматируются, образуя аудиофрейм стандарта ISO/IEC 13818-3. Структура цифровых данных сигналов T_1 и T_2 соответствует формату данных аудиофрейма MPEG-1. Информационные данные сигналов многоканального расширения T_3, T_4, T_5 располагаются в части аудиофрейма, относящейся к передаче дополнительных данных (Ancillary Data). Требуемая скорость для передачи битов сигналов T_3, T_4 и T_5 может варьироваться от фрейма к фрейму в зависимости от характеристик ЗС. Остаток этой части аудиофрейма может использоваться для передачи других дополнительных данных, например, речевых сигналов. Частота дискретизации при передаче основных сигналов T_1 и T_2 составляет 32, 44, 1 и 48 кГц, дополнительных сигналов многоканального расширения T_3, T_4, T_5 — 16, 22, 05 и 24 кГц.

В стандарте MPEG-2 ISO/IEC 13818-3 для получения максимально возможной компрессии цифровых аудиоданных



Рис. 3.23. Структурная схема кодера MPEG стандарта ISO/IEC 13818—3 (многоканальная конфигурация)

при кодировании сигналов многоканального расширения T3, T4, T5 учитывается также и степень корреляции между основными и дополнительными сигналами. Если корреляция сильная, то от независимого кодирования каждого из дополнительных сигналов можно отказаться и передавать лишь их различия по отношению к основным сигналам, используя адаптивное предсказание в MPEG-2 кодере. Кроме того, для еще большей компрессии данных здесь используется также процедура динамического изменения переходного затухания между дополнительными сигналами. Весьма часто встречаются ситуации, когда эти сигналы отличаются только уровнями, поэтому достаточно для ряда из них передать только масштабные коэффициенты, а сами значения отсчетов восстанавливаемого сигнала могут быть взяты из сигнала другого канала. Все это позволяет дополнительно экономить биты при их передаче сигналов многоканального расширения.

После кодирования оба цифровых потока объединяются в мультимплексе (MUX). Наиболее важная часть информации также защищается с помощью помехоустойчивого кодирования (CRC-код).

На приемной стороне системы передачи общий цифровой поток разделяется демультимплексором (DEMUX) на два: один поток поступает на MPEG-1 декодер, а второй — на декодер MPEG-2 многоканального расширения. После декодирования сигналы поступают на инверсную матрицу M2 для их окончательного восстановления.

Психоакустическая часть кодеров стандартов ISO/IEC 11172-3 и ISO/IEC 13818-3 одинакова.

Стандарт MPEG-2 ISO/IEC 13818-7 AAC [2]

Алгоритм компрессии AAC (Advanced Audio Coding) разработан в рамках MPEG-2. Он базируется на учете опыта, накопленного при разработке алгоритма компрессии Layer-3 стандартов ISO/IEC 11172-3 и 13818-3, поддерживает все известные звуковые форматы: моно, обычное стерео, разно-

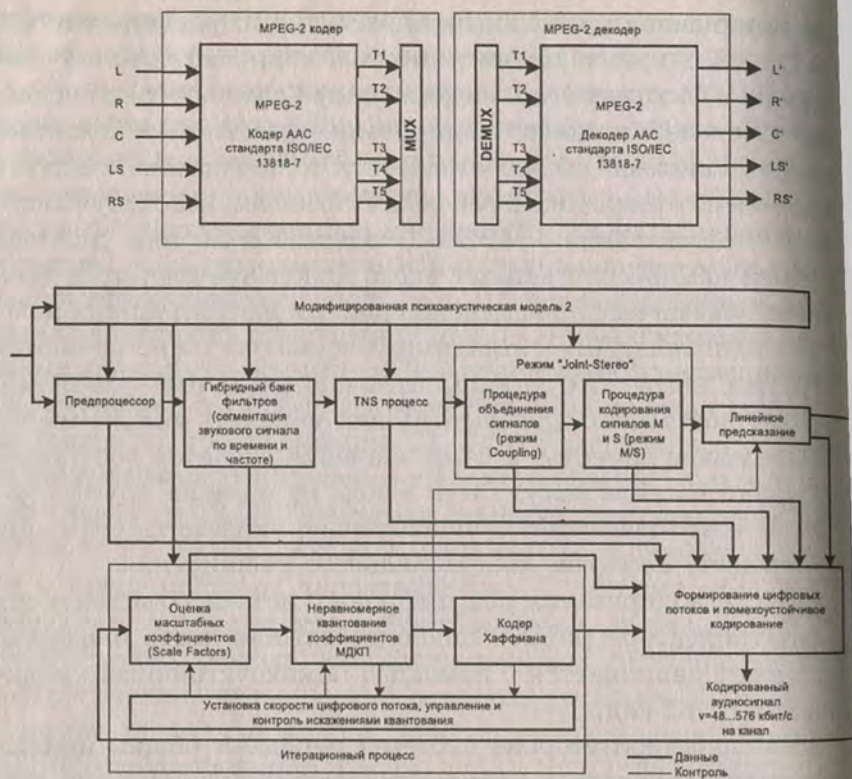


Рис. 3.24. Многоканальная конфигурация и структурная схема кодера MPEG-2 AAC стандарта ISO/IEC 13818-7

видности систем Dolby, пятиканальный звуковой формат 5.1. В отличие от Layer-3 здесь расширен набор возможных частот дискретизации: 8, 11, 0,25, 16, 22, 24, 32, 44, 1, 48, 64, 88, 2 и 96 кГц; изменены форма и длины оконных функций: используются окна Кайзера—Бесселя вместо синусных: длинное, включающее 2048 отсчетов ЗС, и короткое, включающее 256 отсчетов ЗС, что обеспечивает более высокое разрешение по частоте, при этом в обоих случаях используется 50%-ное перекрытие выборок отсчетов ЗС. Кодированию подвергаются коэффициенты МДКП, однако несколько изменена форма кривой компрессии при неравномерном квантова-

нии, применены иные книги кодов Хаффмана. Также как и в Layer-3, управление величиной искажений квантования выполняется с помощью двух итерационных циклов: внутреннего и внешнего. Для повышения качества алгоритма компрессии цифровых аудиоданных в алгоритме AAC используются специальные процедуры минимизации. Точнее говоря, управление микроструктурой искажений квантования внутри каждой из субполос (так называемая техника TNS-Temporal Noise Shaping) изменены. При объединении субполосных сигналов и кодировании (Coupling), предусмотрена, как и ранее, возможность работы кодера в режиме M/S кодирования, когда кодированию в отдельных субполосах подвергаются не сигналы L и R стереопары, а их сумма $M = (L + R)\sqrt{2}$ и разность $S = (L - R)\sqrt{2}$. При линейном предсказании учитывается не только корреляция между отсчетами многоканального сигнала, но и форма спектра шумов квантования и его изменение во времени. Введены уточнения и дополнительные процедуры при расчете глобального порога маскировки в психоакустической модели кодера AAC. Однако и здесь основой является модифицированная психоакустическая модель 2, как и в Layer-3.

Алгоритм MPEG-2 ISO/IEC 13818-7 AAC обеспечивает наиболее высокое качество кодирования звукового сигнала, например, при формате 5.1. искажения, вызванные компрессией, лежат ниже порогов их слуховой заметности уже при суммарной скорости цифрового потока 320...384 кбит/с.

Литература

1. International Standard ISO/IEC 13818-3. Information technology — Coding of moving pictures and associated audio information. — Part 3: Audio, 1995-05-15.
2. International Standard ISO/IEC 13818-7. Information technology — Coding of moving pictures and associated audio information. — Part 7: Advanced Audio Coding (AAC), 1997.

3. International Standard ISO/IEC 13818-4. Information technology — Coding of moving pictures and associated audio information. — Part 4: Conformance Testing, 1998.

4. International Standard ISO/IEC 13818-5. Information technology — Coding of moving pictures and associated audio information. — Part 5: Software Simulation, 1997.

5. International Standard ISO/IEC 13818-6. Information technology — Coding of moving pictures and associated audio information. — Part 6: Extensions for DSM-CC, 1998.

6. International Standard ISO/IEC 13818-9. Information technology — Coding of moving pictures and associated audio information. — Part 9: Extension for real time interface for systems decoders, 1996.

Глава 4

СТАНДАРТ MPEG-4

4.1. Цели и возможности стандарта MPEG-4

Международные нормы MPEG-1 и MPEG-2 считаются одними из самых успешных решений по стандартизации алгоритмов сжатия [1—3].

Формат MPEG предназначен для сохранения видео и звуковой информации на цифровых носителях. Часто речь идет о записи игровых и анимационных фильмов на жесткий диск или CD. Формат MPEG-2 идеально подошел для кодирования и передачи телевизионного сигнала. Именно телевизионная техника стала движущей силой развития форматов MPEG и в значительной мере повлияла на направление дальнейших усилий по стандартизации. Надо подчеркнуть, что механизм телевизионного вещания значительно отличается от простой кадровой анимации. На современных телестудиях передаваемый сигнал создается смещением большого количества различной информации, особенно в так называемых виртуальных студиях. Примером использования виртуальной студии является телепередача прогноза погоды, когда реальный ведущий двигается в специальном помещении на однородном фоне, где возможна цветовая фильтрация, в итоге дающая только изображение его фигуры. Далее оно комбинируется с фоном, на котором отображается метеорологическая информация. Итоговая картинка может быть дополнена текстовой информацией, титрами или другими объектами. В некоторых случаях к передаваемому изображению добавляется миниатюрное изображение сурдопереводчика. Все эти отдельные блоки информации комбинируются в единую последовательность телевизионных кадров, которая направляется на передатчик.

Основной предпосылкой для кодирования изображения (видео) по нормам MPEG-1 и MPEG-2 является тот факт, что передаваемый сигнал может быть дискретно структурирован на так называемые “кадры” или “полукадры”. Под кадрами понимается прямоугольное поле цветных точек, количество которых отвечает размерам изображения в данном телевизионном формате. У кадра нет внутренней структуры, он лишь представляет собой JPEG-картинку. Для повышения эффективности компрессии кадр перед “упаковкой” в видеоряд необходимо предварительно обработать, выделив дополнительно два последующих кадра и определив с их помощью так называемый “вектор движения”, посредством которого можно предсказать относительное перемещение отдельных областей в следующем кадре. В MPEG существует три разновидности кадров (обозначаемых I, B и P), которые после свертывания и кодирования организуются в линейную последовательность — видеоряд.

Члены группы MPEG определили, что передача последовательности изображений не является самым правильным путем для телевизионного сигнала, отдельные элементы (например, передний и задний планы, титры и т. д.) смешиваются с каждым кадром, хотя их самостоятельное кодирование было бы намного проще и экономнее. Более того, в будущем интерактивное телевидение будет требовать идентификации отдельных логических частей изображения, которая не может быть проведена в обратном направлении из передаваемых смешанных кадров. Наглядным примером такой возможности будет изменение языка титров или озвучивания DVD-фильма.

Поэтому группа MPEG создала новый стандарт MPEG-4, который был принят в конце 1999 г. под обозначением ISO/IEC 14496 (Information Technology — Coding of audio-visual objects) [4]. Номер 4 в рабочем названии MPEG при этом показывает, что после предыдущих двух норм произошел значительный логический скачок.

В отличие от принципа “покадрового” кодирования в стандартах MPEG-1 и MPEG-2, концепция MPEG-4 принципиально изменена. Изображение и звук не комбинируются перед отправкой, а передаются посредством нескольких параллельных потоков. Большой объем работы переведен на принимающую и воспроизводящую часть процесса. Если в предыдущих форматах единственной задачей клиентского декодера была обработка отдельных кадров и последовательное их воспроизведение, то теперь декодер должен провести полное воссоздание и микширование звука и изображения, то есть то, что до сих пор обычно делало оборудование в телевизионной студии.

Применение вычислительных мощностей на стороне пользователя фигурирует в информационных технологиях все чаще: вспомним, например, стандарты www-документов, при использовании которых к пользователю передается только интерпретируемая информация, из которой браузеры создают странички с надписями, абзацами, страницами, формулами и т. п.

Байтовые потоки, которые в терминологии MPEG-4 называются элементарными, могут переносить любую информацию, имеющую отношение к окончательному изображению. Каждый вид информации далее можно распределять по кадрам наиболее подходящим методом. Таким образом, отдельно кодируются текстовые данные, статические изображения, последовательность движущихся изображений, общие звуки, речь и т. п. Следующий самостоятельный поток содержит описание сцены, т. е. способ, которым складываются элементарные блоки информации при окончательном воспроизведении. Этот поток также организован соответствующим бинарным кодированием.

Стандарт MPEG-4 должен обеспечить набор технологий, удовлетворяющих требованиям авторов, провайдеров различных услуг и пользователей.

♦ Для авторов MPEG-4 дает возможность создания такого контента, который дает меньше возможностей для повтор-

ного использования (защита авторских прав) и имеет большую гибкость, что сегодня возможно с применением индивидуальных технологий, таких как цифровое телевидение, компьютерная графика, WWW и др.

♦ Для провайдеров услуг MPEG-4 предлагает прозрачность информационного трафика проходящего через их сети и обеспечить основные QoS параметры для различных медиа приложений. Передача данных соответствующая качеству QoS дает возможность оптимизации транспортной нагрузки в распределенных сетях.

♦ Для конечных пользователей MPEG-4 дает возможность воспользоваться многими функциями, которые потенциально могут быть аппаратно объединены в одном компактном терминале и дает возможность взаимодействия с контентом на высоком уровне с ограничениями устанавливаемыми только автором данного контента.

MPEG-4 достигает всего этого за счет обеспечения следующих стандартизованных направлений:

1) представление элементов акустического, визуального и аудиовизуального контентов, называемых аудио/видеообъектами (audio/video objects, AVO). Эти объекты могут быть как натуральными, так и синтезированными, что означает, что они могут быть как записанными при помощи камеры или микрофона, так и сгенерированными при помощи компьютера;

2) композиция этих элементарных объектов вместе для создания целостных аудиовизуальных объектов, что является процессом формирования аудиовизуальных сцен;

3) мультиплексирование и синхронизация данных соответствующих AVO, что дает возможность их транспортировки через сетевые каналы обеспечивающие QoS соответствующее различным AVO;

4) взаимодействие с аудиовизуальной сценой, обеспечиваемое на приемном конце (интерактивность).

Иллюстрация описанных выше функций MPEG-4 при помощи аудиовизуальной сцены показана на рис. 4.1.

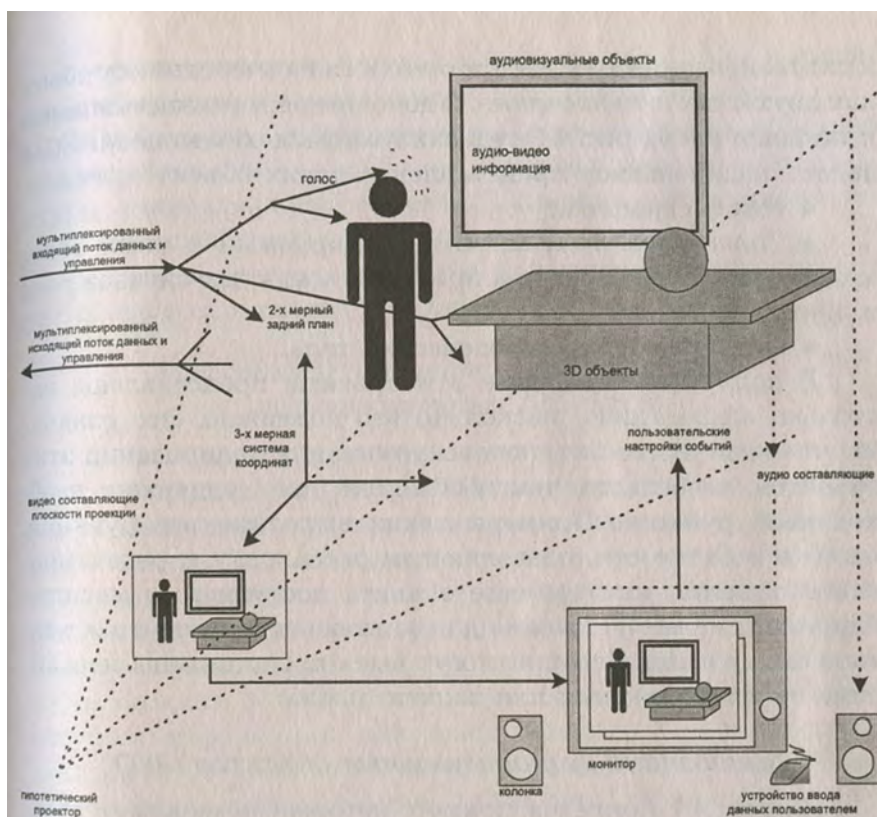


Рис. 4.1. Аудиовизуальная сцена

Представление примитивных AVO

Аудиовизуальные сцены построенные из нескольких AVO имеют иерархическую организацию. На ступенях этой иерархии можно наблюдать примитивные AVO, такие как:

- ◆ фиксированный 2-мерный задний план;
- ◆ картинка говорящего человека (без заднего плана);
- ◆ голос, соответствующий говорящему человеку;
- ◆ и др.

MPEG стандартизирует количество таких примитивных объектов, допускающее представление как натуральных, так

и синтезированных типов, которые в свою очередь могут быть как двух-, так и трехмерные. В дополнение к указанным выше и показанным на рис. 4.1 аудиовизуальным объектам MPEG-4 задает кодированное представление таких объектов, как:

- ◆ текст и графика;
- ◆ “говорящие головы” и ассоциируемый с ними текст может быть использован на приемном конце для синтеза речи и анимации “головы”;
- ◆ анимированные человеческие тела.

В кодированной форме эти объекты представлены настолько эффективно, насколько это возможно. Это означает, что количество бит, используемых для кодирования этих объектов, не больше, чем необходимо для поддержки необходимых функций. Примеры таких выполняемых функций: ошибка робастности, позволяющая распаковку и редактирование объекта, или наличие объекта доступного в масштабируемой форме. Это важно для запоминания того, что объекты в кодированной форме могут быть представлены независимо от их окружения или заднего плана.

Композиция аудиовизуальных объектов (AVO)

На рис. 4.1 приведен пример, который показывает путь, при котором аудиовизуальная сцена в MPEG-4 составлена из отдельных объектов. Образ, содержащий сочетание объектов, — это группа элементарных AVO, объединенных вместе. Например, визуальный объект соответствующий говорящему человеку и голос, объединенные вместе, формируют новый составной AVO, состоящий из акустического и визуального компонентов разговаривающего человека.

Такая группировка позволяет автору конструировать сложные сцены и дает возможность потребителю манипулировать набором объектов.

Более того, MPEG-4 обеспечивает возможность составления сцен, допускающих, например, следующее:

- ◆ расположение аудиовизуальных объектов в любом месте соответствующей координатной системы;

- ◆ расположение группы элементарных AVO порядке формирования составного аудиовизуального объекта;
- ◆ применять поток данных к аудиовизуальным объектам для модификации их атрибутов (например, динамической текстуры принадлежащей объекту; анимации двигающейся головы посылкой параметров анимации);
- ◆ интерактивное изменение пользовательского видения каких-либо параметров в любой сцене.

Мультиплексирование и синхронизация аудиовизуальных объектов

Данные аудио/видеообъектов передаются в одном или более элементарных потоках (Elementary Streams). Дескриптор объекта определяет все потоки, связанные с одним медиа объектом. Это позволяет управлять иерархически кодированными данными, а также соответствующими им мета данными о контенте (информация о содержании объекта) и соответствующими правами на интеллектуальную собственность. Каждый поток характеризуется набором дескрипторов, несущих информацию для конфигурации, т. е. определения декодером ресурсов, требующихся для декодирования потоков и точной синхронизации. Кроме того, дескрипторы могут нести информацию о требуемом качестве услуги (QoS, Quality of Service), необходимом для передачи данных (т. е. максимальная битовая скорость, частота повторения ошибок, приоритеты и т. д.).

Синхронизация элементарных потоков достигается путем установления временных меток индивидуальных “элементов доступа” внутри элементарных потоков. Уровень синхронизации управляет идентификацией каждого элемента доступа и расстановкой временных меток. Этот уровень, независимый от типа медиаданных, позволяет производить идентификацию типа элемента доступа (например, видео- или аудиокадры, команды описания сцены) в элементарных потоках, восстановление аудиовизуальных объектов и выполняет их временную синхронизацию.

Взаимодействие с AVO

В основном пользователь наблюдает сцену, которая составлена в соответствии с настройками автора этой сцены. Пользователь, как правило, зависящий от степени свободы, предоставленной автором, имеет возможность взаимодействия с наблюдаемой сценой. Пользователь может производить следующие операции:

- ◆ изменять точку обзора сцены, например, путем использования навигации в сцене;
- ◆ изменять местоположение объектов в сцене;
- ◆ инициировать каскад действий путем нажатия на специфические объекты, например, старт или остановка видео потока;
- ◆ выбрать желаемый язык в случае возможности использования нескольких языков;
- ◆ может быть инициировано больше комплексных режимов, например, виртуальные телефонные звонки, пользовательские ответы и установка коммуникационных соединений.

Идентификация и защита авторских прав на AVO

Достаточно важно иметь возможность хранения IPR информации (информации о защите прав на интеллектуальную собственность) об AVO MPEG-4. Для этого группа MPEG работала с представителями организаций держателей прав для определения синтаксиса и инструментальных средств для поддержки IPR идентификации и защиты авторских прав. Защита контента будет описана в MPEG-4 Version 2.

4.2. Детальное техническое описание стандарта MPEG-4

Как показано на рис. 4.2., потоки, приходящие из сети (или с устройства хранения данных) и обозначенные как TransMux потоки, демультиплексируются в FlexMux потоки

и направляются в соответствующие FlexMux демультимплексоры на выходе которых получаются элементарные потоки (Elementary Streams, ES). Элементарные потоки анализируются и перенаправляются на соответствующие им декодеры. Декодирование возвращает данные в форму аудио/видео (AV) объекта и выполняются операции, необходимые для реконструкции оригинального AV объекта, готового к передаче на соответствующее устройство. Аудио- и видеообъекты представлены в их кодированной форме, которая будет описана ниже. Реконструированные AV объекты сделаны доступными для составления и последующего использования в процессе рендеринга (воспроизведения) сцены. Декодированные AVO совместно с информацией описания сцены используются для составления сцены в соответствии с описанием автора. Процесс описания сцен будет описан ниже. Пользователь может в пределах, позволенных автором, взаимодействовать со сценой, которая уже составлена.

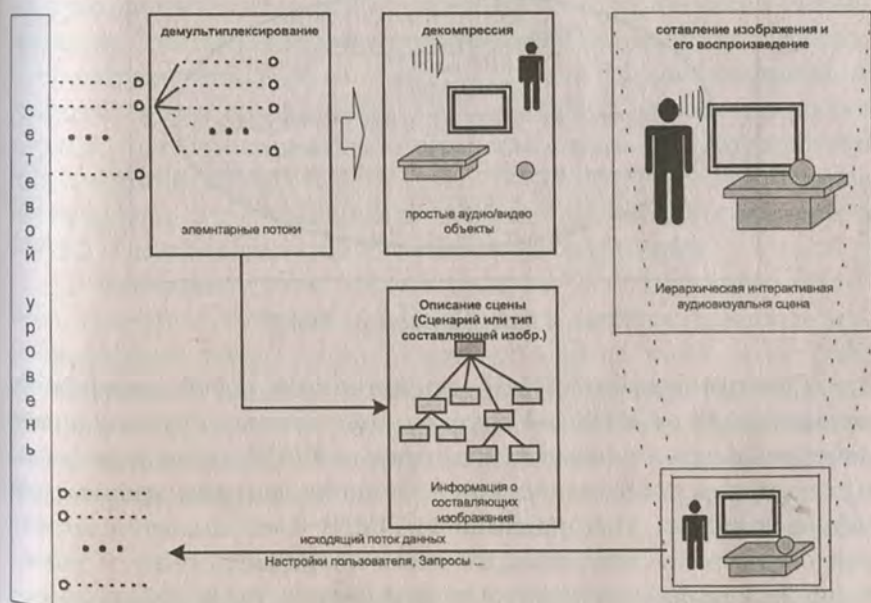


Рис. 4.2. Основные компоненты MPEG-4 терминала (приемная сторона)

DMIF

DMIF (Delivery Multimedia Integration Framework, мультимедийная интегрированная система доставки) предназначена для управления операциями над мультимедиаприложениями через интерактивные сети в широковещательных средах и на дисках. Архитектура DMIF предназначена для управления данными операциями в любых коммуникационных средах с любыми методами установления связи. Внедрение DMIF позволяет передавать сетям детали представляющие приложения с простым интерфейсом.

DMIF расположена между MPEG-4 приложением и транспортной сетью, как показано на рис. 4.3.



Рис. 4.3. Архитектура DMIF

Для приложений DMIF представляет собой интерфейс, независимый от MPEG-4 потоков, полученных путем взаимодействия с отделенным интерактивным DMIF узлом через сети и/или путем взаимодействия с вещаемыми или хранимыми медиа данными. Интерактивный DMIF узел является конечной системой в сети, которая может создавать сеанс с намеченной точкой. Намеченной точкой может быть набор широковещательных потоков, интерактивное приложение или набор хранимых MPEG-4 файлов.

MPEG-4 приложения могут устанавливаться через DMIF множественное соединение с приложениями в конечных узлах, каждый из которых представлен уникальным адресом.

MPEG-4 приложение может запрашивать DMIF на установку каналов со специфическими требованиями к QoS и шириной канала для каждого элементарного потока. DMIF гарантирует установление каналов со специфическими условиями пропускания в сетях между взаимодействующими узлами; позволяет каждому узлу по своему видеть сеть, что уменьшает количество стеков, поддерживаемых каждым терминалом. Контроль DMIF затрагивает оба FlexMux и TransMux уровня. В отношении FlexMux, DMIF обеспечивает контроль установления FlexMux каналов. В отношении TransMux DMIF использует открытый интерфейс, который совмещает существующие и будущие сети при помощи моделей, называемых дескрипторами ресурсов соединения. MPEG-4 предлагает прозрачный интерфейс с примитивной семантикой сигнализации. Эта семантика и интерфейс DMIF интерпретируется и транслируется для соответствия сигнальным сообщениям каждой сети. Что касается QoS, то MPEG-4 обеспечивает первый шаг, определяющий основной набор QoS параметров для медиа данных относительно DMIF интерфейса. Точное определение этих трансляций выходит за пределы стандарта MPEG-4 и производится сетевыми провайдерами.

В типовом случае конечный пользователь может получить доступ к AVO, распространяемым через удаленные интерактивные узлы, широкоэмитательные системы и устройства хранения данных. Если контент позволяет, то конечный пользователь может улучшить качество соединения добавлением расширенных AVO потоков через ресурсы соединения с гарантированным QoS.

Демультимплексирование, буферизация данных и синхронизация

Каждый элементарный поток может быть извлечен из приходящих данных при подключении к сети или устройству

хранения данных. Каждое подключение к сети определено TransMux каналом в системной модели MPEG-4. Демультимплексирование частично или полностью делается за пределами MPEG-4 и зависит от приложения. В целях интеграции MPEG-4 в системные среды создан интерфейс мультимплексирования потоков (Stream Multiplex Interface). В этом интерфейсе определены пакетизированные потоки адаптивного уровня (Adaptation Layer-packetized Streams, AL-packetized Streams). FlexMux уровень определяет опции инструментального набора FlexMux. Интерфейс TransMux определяет как каждый из AL-пакетизированных потоков (без использования уровня FlexMux) или FlexMux потоки восстанавливаются из TransMux уровня. Транспортные функции этого интерфейса в MPEG не определены. Информационный уровень интерфейсов обоснована, пока уровень контроля работает с DMIF. Это показано на рис. 4.4.

Точно так же, как MPEG-1 и MPEG-2 описывали режим работы идеализированного декодирующего устройства совместно с синтаксисом и семантикой потока данных, в MPEG-4 определена системная модель декодера (System Decoder Model, SDM). Это делает возможным идеальную точность терминальных операций без выполнения ненужных описаний деталей реализации. Это означает, что производителям дается свобода при разработке MPEG-4 терминалов и декодирующих устройств. Эти устройства, начиная от ТВ-приемников, которые не имеют возможности соединения с отправителем сигнала до компьютеров, полностью приспособлены для двунаправленных соединений. Некоторые устройства будут принимать потоки MPEG-4 через изохронные сети, пока остальные будут использовать неизохронные значения (например, Интернет) для обмена MPEG-4 информацией. Системная модель декодера дает общую модель, на базе которой могут быть созданы разные модели MPEG-4 терминалов.

Спецификации буфера и временных моделей являются неотъемлемыми частями кодирующих устройств, которые могут не знать о том, какое терминальное устройство будет

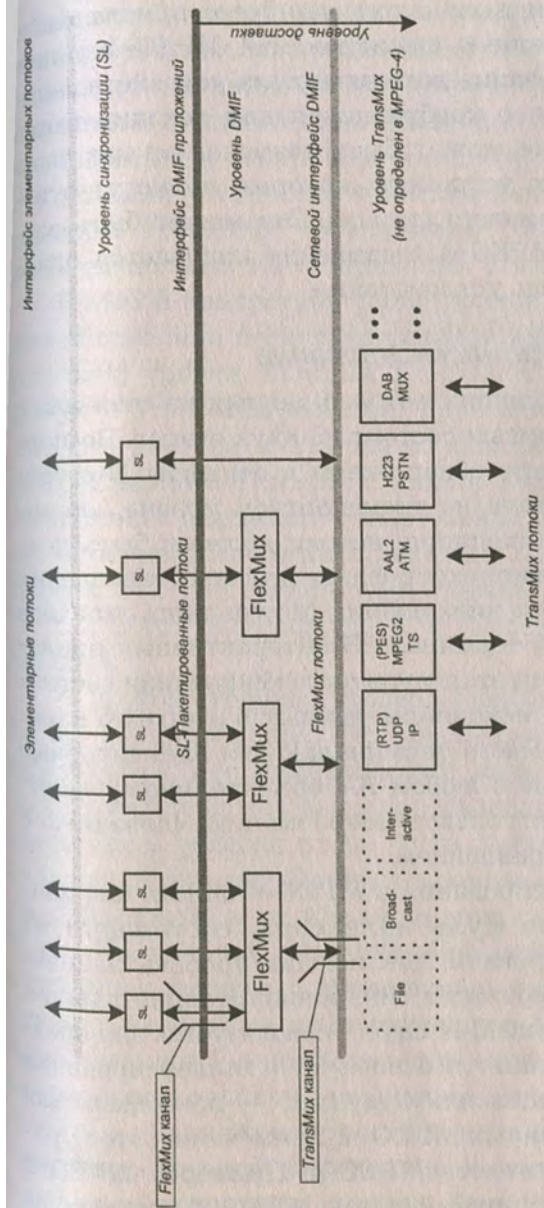


Рис. 4.4. Модель системного уровня стандарта MPEG-4

декодировать поток или даже о том, как будет принят кодированный поток. С помощью спецификаций MPEG-4 станет возможным информирование декодирующих устройств кодирующими устройствами о требующихся для декодирования ресурсах. Однако это не может быть возможным, как было сказано ранее, для тех устройств, которые не могут отвечать отправителю первичного сигнала. Это может быть возможным потому, что MPEG-4 соединение получается одновременно очень разными устройствами.

Демультимплексирование

Восстановление входящих потоков данных из сети или с устройства хранения данных состоит из двух этапов. Во-первых, каналы должны быть обнаружены и открыты. Это требует организации контроля на транспортном уровне, например, DMIF. Во-вторых, входящие потоки должны быть должным образом демультимплексированы для восстановления элементарных потоков из низходящих (downstream) каналов (входящих на приемный терминал). В интерактивных приложениях соответствующий этап мультиплексирования состоит в мультиплексировании исходящих данных в upstream каналы (исходящие из приемного терминала). Эти элементарные потоки переносят данные о любом AV объекте, информацию описания сцены или контрольную информацию, связанную с AVO или системным управлением.

Этап демультимплексирования в MPEG-4 описан как концепция двухуровневого мультиплексора, состоящего из TransMux и FlexMux уровней, так же как уровень элементов доступа, который переносит синхронизирующие данные.

Основной термин "TransMux Layer" используется для представления любых, лежащих в основе мультиплексирования функций, — существующих или будущих, — пригодных для транспортных потоков данных MPEG-4. Примечание: этот уровень не определен в контексте MPEG-4. Примеры: MPEG-2 Transport Stream (Транспортный поток MPEG-2), протоколы H.223, ATM AAL 2, IP/UDP (см. рис. 4.4). TransMux уровень

создан как структура защитного подуровня и подуровня мультиплексирования показывающего, что данный уровень отвечает за обеспечение соответствующего качества. Защитный подуровень функционально включает в себя средства защиты от ошибок и обнаружения ошибок соответствующие данной сети или устройству хранения данных. В некоторых реализациях TransMux уровня может быть включена возможность отдельной идентификации этих подуровней.

В любой конкретной сцене приложения может использоваться один или несколько реализаций TransMux (TransMux Instances) уровня. Каждый TransMux демультимплексор дает доступ к TransMux каналам. Требования к интерфейсу данных для доступа к TransMux каналам те же самые, что и для реализаций TransMux уровня. Они включают в себя необходимость надежного обнаружения ошибок, доставки, если это возможно, ошибочных данных с подходящей индикацией ошибок. Эти требования суммированы в информативной части TransMux интерфейса и в системной части стандарта MPEG-4.

С другой стороны, уровень FlexMux полностью описан в MPEG спецификации. Это обеспечивает гибкий, с низкими задержками, инструментальный набор для расслоения данных, который может быть использован как опция и особенно может быть полезен тогда, когда размер пакета данных имеет большой размер. FlexMux уровень сам по себе не является робастным к ошибкам и может быть использован в TransMux каналах с высоким QoS или для объединения элементарных потоков с приемлемым числом ошибок. FlexMux требует надежного детектирования ошибок и достаточного кадрирования FlexMux пакетов (для случайного доступа и исправления ошибок). Эти требования суммированы в интерфейсе мультиплексирования потоков (Stream Multiplex Interface), который определяет доступ данных к индивидуальным транспортным каналам. FlexMux демультимплексор извлекает AL-пакетизированные потоки из FlexMux потоков.

Уровень элементов доступа (Access Unit Layer) имеет минимальный набор средств для последовательного сравнения, отправления временной (синхронизирующей) информации и переноса элементов доступа с временными метками при помощи элементарных потоков. Каждый пакет состоит из одного элемента доступа (Access Unit, AU) или фрагмента AU. Форма этих AU с временными метками — это семантическая структура элементарных потоков видимая на данном уровне. AU уровень требует надежного обнаружения ошибок и кадрирование каждого пакета из предыдущего уровня, например использованием FlexMux. То, как данные могут доставлены на уровень сжатия, описано в информации по интерфейсу элементарных потоков (Elementary Streams Interface), которая может быть найдена в Системной части стандарта MPEG-4. AU уровень восстанавливает элементарные потоки из AL-пакетизированных потоков.

Для правильного определения связи ES с AVO в сцене используются дескрипторы объектов (Object Descriptors) и StreamMap таблицы. Дескрипторы объектов переносят информацию о количестве и свойствах элементарных потоков (ES) которые соответствуют определенным AV объектам. StreamMap таблицы связывают каждый поток с указателем канала (ChannelAssociationTag), который служит указателем канала, который несет данный поток.

Буферизация данных

Для прогнозирования того, как декодер будет работать во время декодирования разных элементарных потоков данных, созданных во время MPEG-4 сессии, системная модель декодера дает возможность кодеру определять и отслеживать минимальные ресурсы буфера, которые нужны для декодирования поступивших во время соединения данных. Данные о требуемых для буфера ресурсах передаются декодеру с дескрипторами объекта во время настройки MPEG-4 сессии, и поэтому декодер может сам принять решение о возможности обработки данной сессии.

В процессе управления ограниченным объемом буферного пространства модель позволяет отправителю, например, передавать обычные данные (не в реальном времени), если на приемной стороне имеется достаточно места для их хранения. Доступ к этим данным может быть получен тогда, когда это необходимо, позволяя в это время данным реального времени использовать больший объем канальной емкости, если это необходимо.

Синхронизация

Для операций, производимых в реальном времени, модель временной синхронизации предполагает, что конечная задержка сигнала от кодера и сигнала, входящего в декодер, является величиной постоянной. Более того, переданные потоки данных должны содержать скрытую или явную синхронизирующую информацию. Существует два типа таких данных. Первый используется для передачи скорости хода часов кодера к декодеру. Второй тип состоит из временных отметок привязанных к частям (порциям) кодированных AV данных, содержащих желательное время декодирования для элементов доступа (Access Units, AU) или время составления для элементов композиции (Composition Units, CU). Эта информация передается в AL-PDU заголовках, созданных в уровне элементов доступа (Access Unit Layer). С этой временной информацией интервал между изображениями и скорость аудио отсчетов может быть установлена в декодере для соответствия аналогичным параметрам кодера для выполнения операции синхронизации.

Разные AV объекты могут быть кодированы с различными масштабами по времени с незначительно отличающимися скоростями. Можно привести эти временные масштабы к оси времени приемного терминала.

Описание синтаксиса

В MPEG-4 определен синтаксический описательный язык (syntactic description language) для описания бинарного син-

таксиса представления потоков AV объектов так же, как информация описания сцены. Это отклонение от традиционного MPEG приближения, использующего язык псевдо-Си. Этот язык является расширением C++ и используется для описания синтаксического представления объектов, определения всего класса AV объектов и информации описания сцены путем их интеграции. Это дает возможность согласования и унификации способа описания синтаксиса в очень четкой форме, в то же время упрощая процесс тестирования потоков данных.

Кодирование аудиообъектов.

Кодирование аудиообъектов в MPEG-4 обеспечено набором инструментов для представления как натуральных звуков (речь, музыка), так и синтезированных. Это представление обеспечивает компрессию и другие функции, такие как масштабирование или проигрывание на разных скоростях. Представление синтезированного звука может быть выполнено текстовыми или инструментальными дескрипторами и кодированием параметров для обеспечения различных эффектов.

Естественный звук

MPEG-4 стандартизирует кодирование естественных звуков в диапазоне скоростей от 2 кбит/с до 64 кбит/с. Присутствие MPEG-2 AAC стандарта совместно с инструментарием MPEG-4 дает возможность для компрессии аудио на более высоких скоростях. Для скоростей от 2 кбит/с до 64 кбит/с MPEG-4 нормализует синтаксис потока данных и процесс декодирования в пределах набора инструментальных средств. В порядке достижения наивысшего качества звука во всем диапазоне скоростей и в то же время обеспечения дополнительной функциональности было создано три типа кодеров. Нижний диапазон скоростей 2—4 кбит/с для речи с частотой дискретизации 8 кГц и диапазон 4—16 кбит/с для аудио с частотой дискретизации 16 кГц покрыт техникой параметри-

ческого кодирования. Для кодирования речи в среднем диапазоне скоростей 6—24 кбит/с используется техника CELP кодирования. В этом диапазоне две частоты дискретизации (8 и 16 кГц) используются для поддержки большинства узкополосных и широкополосных голосовых приложений. Для скоростей от 16 кбит/сек используется T/F техника кодирования, известная как *TwinV/Q* и AAC кодеки. Аудиосигналы в этом диапазоне, как правило, имеют скорость отсчетов от 8 кГц.

Для обеспечения мягких переходов между скоростями и для обеспечения масштабирования скоростей и диапазонов определена структура скоростей. Это показано в табл. 4.1.

Таблица 4.1

Техника кодирования	Количество каналов	Скорость (кб/сек)
AAC	5	320
MPEG-2 Layer 2	5	640
AAC	2	128
AAC	2	96
MPEG-2 Layer 2	2	192
MPEG-2 Layer 3	2	128
AAC	1	6 (CELP) + 18 (AAC)
Масштабируемый: CELP совместно с AAC	1	
Масштабируемый: двойное VQ преобразование совместно с AAC	1	6 (Twin VQ) + 18 (AAC)
AAC	1	18
G.723	1	6.3
Широкополосное CELP кодирование	1	18.2
BSAC	2	96
BSAC	2	80
BSAC	2	64
AAC-LD (с 20 мс задержкой)	1	64
G.722	1	32
AAC-LD (с 30 мс задержкой)	1	32

Системный уровень MPEG-4 облегчает использование различных инструментов таким образом, что кодеки, соответствующие имеющимся стандартам, могут быть приспособ-

лены для кодирования. Из-за этого MPEG-4 позволяет использовать различные оптимизированные кодеры, стандартизированные ITU-T и которые разработаны для обеспечения специфических наборов требований. Каждый из этих кодеров разработан для работы в автономном режиме со своим синтаксисом. Дополнительные функции реализованы в конкретных кодерах посредством дополнительных инструментов.

Синтезированный звук

Декодеры также позволяют синтезировать звук, основанный на структурированных исходных данных. Введенный текст конвертируется в речь посредством Text-To-Speech (TTS) декодера, в то же время достаточно много звуков, включая музыку, может быть синтезировано. Синтезированная музыка может быть передана на очень низких скоростях с точным описанием звукового сигнала.

Text-To-Speech. TTS позволяет синтезировать речь из введенного текста. Данный процесс включает в себя выполнение следующих функций:

- ◆ Синтез речи с использованием оригинальной речи.
- ◆ Выполнение функций переключения режимов: пауза, окончание, перемотка вперед/назад.
- ◆ Поддержка многоязыковых режимов работы для текстов.
- ◆ Поддержка международных символов для фонем.
- ◆ Поддержка определения специфики возраста, пола, языка и диалекта оратора.

Структурированный аудиодекодер (Structured Audio Decoder, SAD) декодирует входящие данные и создает на своем выходе звук. Это декодирование управляется специальным языком синтеза SAOL (Structured Audio Orchestra Language, специальный структурированный оркестровый язык), стандартизированный как часть стандарта MPEG-4. Этот язык использует определяемые как "оркестровые" специальные "инструменты" (загруженные в потоке данных, не

фиксированные в терминале), с помощью которых создают и управляют данными. Инструменты представляют собой небольшую цепочку обработки простейших сигналов, которые могут имитировать некоторые специфические звуки, в том числе и звуки, издаваемые акустическими инструментами. Цепочка обработки сигналов может быть реализована и на аппаратном уровне, и в виде программного обеспечения.

MPEG-4 не стандартизирует метод синтеза звука, а лишь дает метод описания синтеза. Любой из уже существующих или создаваемых методов синтеза звука может быть описан в SAOL.

Контроль синтеза заканчивается загрузкой "scores" (партитур) и "scripts" в поток. Партитура — это упорядоченный во времени набор команд, которые позволяют имитировать разные инструментов в определенные моменты времени для создания общего качественного звука или генерации звуковых эффектов. Описания партитуры, загруженное в языке, который называется SASL (Structured Audio Score Language, структурированный язык партитур), могут быть использованы для создания новых звуков, а также содержат в себе дополнительную информацию, позволяющую модифицировать уже существующие звуки. Для процесса синтеза, который не требует сильного контроля, может быть использован MIDI протокол.

Тщательный контроль в соединении с настроенной разборчивостью инструментов позволяет генерировать звуки начиная от простейших аудиоэффектов (звук шагов, скрип двери) и закачивая сложнейшими натуральными звуками (звук дождя, музыка, проигрываемая набором разных инструментов, в том числе и синтетическими).

Все эти возможности полностью нормированы в MPEG-4 и все описанные выше звуки гарантировано одинаково будут воспроизводиться на каждом терминале. Для терминалов с меньшим набором функций и для приложений, которые не требуют столь изощренного синтеза, может быть использован

“wavetable bank format”. При использовании данного набора инструментов качество звука сохраняется, но становится меньше возможностей для обработки звука.

Эффекты

Язык SAOL используется при описании специальной обработки эффектов для использования в системном двоичном формате описания сцен MPEG-4 (MPEG-4 Systems Binary Format for Scene Description). Аудиосистема BIFS (Binary Format for Scenes) обрабатывает декодированные аудиоданные для создания такого выходящего потока аудиоданных, которым можно манипулировать для создания специальных эффектов с последующей совместимостью частоты следования отсчетов эффектов и звука. Эффекты, по существу, повторяют функции, выполняемые различными инструментами с входными сигналами (фильтрация, ограничители, микшеры и т. д.).

Общий звук

При кодировании звука используются экономичные методы записи из предыдущих стандартов MPEG. Более того, удобно записывать и синтезировать звук, созданный из образцов различных музыкальных инструментов и из нотной записи конкретного сочинения. Для этой цели при описании свойств инструментов и партитур используются особые языки. В принципе, с этого и началось развитие отдельного формата передачи данных.

Речь

При использовании синтезированной речи прежде всего нужно передать на сторону приемника библиотеку так называемых фонем (звуков произносимых букв, их двух- и трехбуквенных комбинаций и других акустических проявлений, таких как, например, смех или откашливание), а потом кодировать речь, состоящую только из данных фонем. Этот метод давно уже используют многие компьютерные игры.

Звуковые операторы

Отдельные потоки звуков и речи можно обрабатывать при помощи звуковых операторов, которые мы можем себе представить в виде виртуального пульта звукооператора. В нем к основным звуковым источникам добавляются различные эффекты, такие как запаздывание, изменение громкости, скорости и т. п. Звуковые операторы и источники звука организуются в форме иерархии событий.

Кодирование визуальных объектов

Визуальные (видео) объекты могут быть натуральными (естественными) и синтетическими.

Естественные текстуры, изображения и видео

Наборы инструментов для представления естественного видео в видеостандарте MPEG-4 направлены на создание стандартизированной основы технологий, позволяющих обеспечить эффективное хранение, передачу и манипуляцию с текстурами, изображениями и видеоданными в мультимедийных средах. Эти инструменты позволяют декодировать и представлять крошечные единицы изображений и видеоконтента, называемые “видеообъектами” (Video Objects, VOs). Для примера, видеообъектом может быть говорящая персона (без заднего плана), которая может быть совмещена с другими AVO (аудио/видеообъектами) для создания сцены.

В порядке достижения основной цели скорее как решения для узкого набора приложений набор выполняемых функций общий для различных приложений разбит на части. Вследствие этого “видеочасть” стандарта MPEG-4 обеспечивает решения в форме наборов инструментов и алгоритмов для:

- ◆ эффективное сжатие изображений и видео;
- ◆ эффективное сжатие текстур для их отображения в 2- и 3-мерных сетках;
- ◆ эффективное сжатие скрытых 2-мерных сеток;

- ◆ эффективное сжатие изменяющихся во времени геометрических фигур;
- ◆ эффективный доступ к различным видео объектам;
- ◆ эффективное воздействие на функции, выполняемые с изображениями и видеопоследовательностями;
- ◆ пространственное и временное масштабирование.

Синтетические объекты

Форма синтетических объектов — это набор класса компьютерной графики и в первом приближении синтезированные объекты могут быть описаны следующим образом:

- ◆ параметрическое описание:
 - а) параметрическое описание человеческого лица и тела;
 - б) анимационные потоки лица и тела;
- ◆ статическое и динамическое сеточное кодирование с отображением текстур;
- ◆ кодирование текстур для View Depend приложений.

Анимация лица

Лицо — это объект позволяющий производить лицевую геометрическую подготовку к рендерингу и анимации.

Подобно тому, как при передаче речи объем данных снижается разложением голоса на фонемы и их обратным сложением в поток речи, можно, например, создать графический алфавит лица говорящего человека. Вместо непрерывной передачи изображения диктора можно сначала перенести только несколько основных видов лица (фас, профиль), создать пространственную модель его головы и затем менять изображение на основе нескольких десятков так называемых визем (аналог фонемы в визуальной области). Каждая визема определяет движения определенных частей лица, например, открытие рта при произнесении звука "о", движение уголков рта при улыбке, поднятие бровей при удивлении и т. п. Сам формат уже содержит 14 основных визем для типичных фонем, таких как гласные звуки.

Человеческое лицо при этом характеризуется множеством определяющих точек. Трансформациями этих точек лицо при-

водится в движение так, что его вид отвечает синхронизированной с ним синтетической речи. Очевидно, что этот подход требует сложных расчетов как при создании модели головы отдельного человека, так и при последующей анимации этой трехмерной модели, комбинируемой с текстурами.

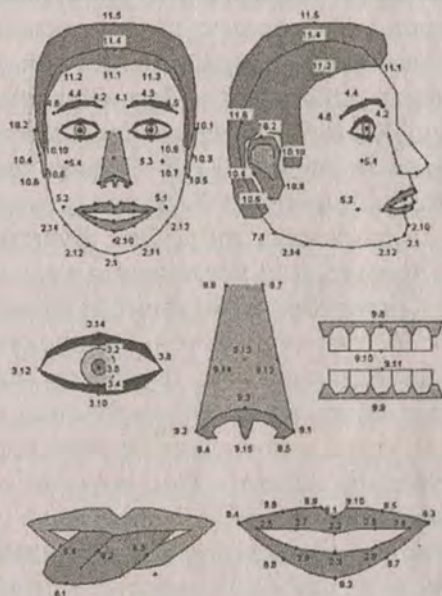


Рис. 4.5. Группа точек для управления анимацией частей лица

Форма, текстура и выражение лица в основном контролируется потоком, содержащим ссылки на FDP (Facial Definition Parameter, параметры определяющие лицо) и FAP (Facial Animation Parameter, параметры анимации лица) наборы. Изначально объект лица состоит из основного лица с нейтральным выражением. Оно может быть непосредственно управляться принимаемыми из потока FAP параметрами, которые обеспечивают создание анимации лица: изменение выражения, речь и т. д. Если принимаются FDP параметры, то они используются для преобразования основной модели лица в другое определенное лицо определяемое своей формой и

текстурой. Есть возможность (в виде опции) загрузки полной модели лица посредством набора FDP параметров.

Объект лица также может принимать локальные средства управления, которые могут быть использованы для модификации облика или мимику лица локально программированием или пользователем. Существует три возможности для локального контроля. Во-первых, путем посылки набора FDP параметров для изменения формы и/или текстуры лица. Во-вторых, может быть определен набор коэффициентов усиления (Amplification Factors), каждый коэффициент переносит параметр анимации в наборе FAP. Объект лица может использовать эти коэффициенты в FAP для усиления или уменьшения выбранных движений лица. Эта функция может быть использована, например, для увеличения визуального эффекта произношения речи для облегчения понимания движения губ. Третий способ обеспечивается определением функции фильтрации (Filter Function). Эта функция, если она определена, способствует обновлению объекта лица перед каждым рендерингом. Объект лица посылает оригинальный набор FAP параметров в функцию фильтрации, которая производит модификацию (изменение) в них и возвращает для последующего использования в рендеринге. Эта функция может включать в себя взаимодействие с пользователем (интерактивность). Можно также использовать функцию фильтрации как источник анимации лица, если нет потока, который контролирует лицо, например, в случае, если лицо управляется TTS системой, которая начинает управлять тогда, когда начинает приходить текст из потока данных.

Анимация тела

Тело представляет из себя объект, совместимый с производством виртуальных моделей тела и анимацией в форме набора 3D полигонов (сеток) для рендеринга. Для тела определены два набора параметров: набор параметров, определяющих тело (Body Definition Parameters, BDP), и параметров анимации тела (Body Animation Parameters, BAP). BDP опре-

деляет набор параметров для преобразования начального (стоящего по умолчанию) тела в измененное тело со своими собственным внешним видом, размерами и (в виде опции) текстурой. BAP параметры, если их интерпретировать корректно, обеспечивают разумные по качеству параметры анимации тела без необходимости инициализации или калибровки модели.

Изначально, до конструирования, объект тела представляет собой основное человеческое тело со стоящей по умолчанию позой. Это тело можно сразу отображать (без изменений). Эта модель может принимать из потока данных BAP параметры, обеспечивающие анимацию тела. Если принимаются BDP параметры, то они используются для преобразования основной модели тела в другую модель, определенную данными параметрами. Стоящая по умолчанию поза — это поза стоящего человека.

Модель человеческого тела должна быть совместима с поддержкой различных приложений, от реалистичной симуляции человеческих движений до сетевых игр, использующих простые человекоподобные модели.

2-мерные анимированные сетки

Если изображение представлено полем цветных точек (растром), то с ним нельзя проводить некоторые преобразования без потери или искажения информации. Растровое изображение можно, но с большим трудом, поворачивать на малый угол. Так что лучше сохранить описание изображения как набор геометрических элементов, манипуляции с которыми производить гораздо проще. Окончательное изображение возникнет путем проецирования геометрических элементов на плоскость, причем ее размер может быть совершенно произвольным. По такой схеме, например, работает известный редактор CorelDRAW. Стандарт MPEG-4 позволяет описать изображение при помощи многих различных объектов, среди которых наиболее часто используются ломаная линия, состоящая из отрезков, кривая Безье, окружность, прямоу-

гольник и правильные многоугольники. Им можно придать цвет или заполнить их текстурой. Отдельные объекты можно собирать в группы, произвольно их трансформировать и определять взаимное перекрытие по принципу слоев.

Двухмерная сетка представляет из себя мозаику из плоских 2-мерных зон (полигональных участков). Эти зоны пересекаются и точки пересечения называются узлами сетки. В MPEG-4 рассматриваются только треугольные сетки, в которых участки являются треугольниками. Двумерная динамическая сетка основывается на геометрии двумерной сетки и информации о движении всех узловых точек. Треугольные сетки уже достаточно давно используются для моделирования форм трехмерных объектов и рендеринга в компьютерной графике. Моделирование двумерных сеток может быть представлено как проецирование трехмерной треугольной сетки на план изображения. Пример двумерной сетки дан на рис. 4.6.

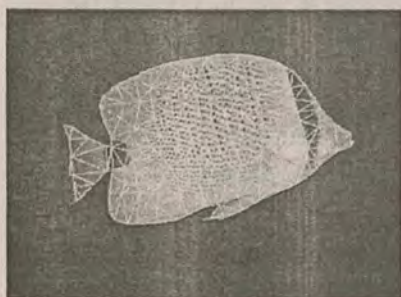


Рис. 4.6. Пример двумерной сетки

Динамическая сетка — это измененная сетка, в которой все узловые точки начальной сетки изменены во времени соответствующими векторами движения. Изначальная сетка может быть постоянной (стандартной) или подогнанной к содержанию изображения. Второй случай называется контентно-наполненной сеткой (content-based mesh), или основанной на контенте.

В двумерной сетке наполненной текстурой треугольные участки в данном кадре деформированы движением узловых точек сетки в опорном кадре и текстура внутри каждого участка в опорном кадре искажена для данного кадра путем использования параметрического отображения, определенного как функция векторов движения узловых точек сетки. Для треугольных сеток аффинное отображение является общим выбором. Эта линейная форма отображения подразумевает

отображение текстур с низкой вычислительной сложностью. Аффинное отображение может создавать преобразование, вращение, масштабирование и т. д.

Привлекательность моделирования двумерных сеток состоит в том, что двумерные сетки могут быть получены из единичной оценки объекта и не требуют ряда данных. В принципе, представление объектов двумерными сетками позволяет выполнять следующие функции:

Управление видеообъектами

◆ Расширенная реальность: Слияние виртуальных изображений с реальным видео для создания усовершенствованного изображения. Виртуальные изображения должны быть в полной совместимости с движущимися реальными объектами (отсюда требование подстройки).

◆ Анимация синтезированных объектов: перенос натурального видеообъекта в видеоклип внутрь другого видеообъекта. Перемещенный видеообъект может быть извлечен из другого натурального видеоклипа или может быть трансформирован из неподвижного объекта изображения использованием информации о движении объекта (отсюда необходимость долгосрочного временного предсказания).

◆ Пространственно-временная интерполяция.

Сжатие видео объектов

◆ моделирование двумерными сетками может быть использован для сжатия, если выбирать передачу изображения текстур только выбранных кадров и анимировать эти изображения текстур для вспомогательных кадров.

Индексация контентно-наполненного видео

◆ сеточное представление обеспечивает возможность уточнять информацию о траектории движения объекта, что может быть использовано для восстановления визуальных объектов со специфическим движением.

◆ Сеточное представление обеспечивает вершинное представление формы объектов, которое более эффективно, чем растровое отображение.

Основные трехмерные сетки

Учитывая, что сцена в телестудии по-своему “трехмерна”, нет ничего удивительного в том, что стандарт MPEG-4 позволяет кодировать и передавать данные, описывающие трехмерные объекты. Для решения данной задачи группа MPEG использовала основу формата VRML для описания виртуальной реальности и приняла ее за основное средство для моделирования пространственной сцены в MPEG-4. Первоначальное текстовое кодирование языка VRML здесь было обогащено экономичным двоичным кодированием, подходящим для передачи в рамках элементарного потока MPEG. Практическим примером использования трехмерной модели является глобус в виртуальной студии прогноза погоды. Шар, с нанесенным на его поверхность рисунком, разумеется, может вращаться.

Стандарт MPEG-4 поддерживает основные сетки для представления синтезированных 3D объектов. Эти сетки обеспечивают поддержку следующих свойств: цвет, нормирование затенение и координация текстур для отображения натуральных текстур, изображений и видео в сетках. Эта инструментальная панель обеспечивает алгоритмы для:

◆ *эффективного сжатия основных сеток;*

◆ *(Уровень детализации, Level Of Detail, LOD) масштабирование трехмерных сеток* — позволяет декодеру декодировать выборку из всего потока данных для реконструкции упрощенной версии сетки содержащей меньше углов чем оригинальная сетка. Такая упрощенная реконструкция полезна для уменьшения времени рендеринга объектов, которые удалены от наблюдателя (LOD управление), а также требует менее мощных устройств для рендеринга объектов с меньшим качеством.

♦ *Пространственное масштабирование* — позволяет декодеру декодировать выборку из полного потока данных сгенерированного кодером для реконструкции сетки с меньшим пространственным расширением. Это свойство наиболее полезно, когда совмещены LOD с масштабированием.

Масштабирование, зависящее от точки обзора

Данный тип масштабирования дает возможность отображения потока текстур, которые используются в реалистичных виртуальных средах. Принцип такого масштабирования заключается в том, что считаются позиции обзора в трехмерном виртуальном мире в порядке передачи наиболее важной информации. Как только посылается часть информации, зависящей от геометрии объекта, то сразу изменяется точка обзора. Эта аппроксимация позволяет сильно уменьшить количество передаваемой информации между удаленной базой данных и пользователем.

Структура инструментов для представления натурального видео

Алгоритмы MPEG-4 для кодирования видео и изображений дают эффективное представление визуальных объектов с достижением поддержки так называемых функций, основанных на контенте. Это дает возможность поддержки большинства тех функций, которые уже включены в MPEG-1 и MPEG-2, включая обеспечение эффективного сжатия последовательностей прямоугольных изображений с варьированием форматом входящих данных, скоростей кадров, скоростей передачи данных и с разными уровнями пространственного и временного масштабирования, с разными уровнями качества.

Основная классификация скоростей передачи данных и выполняемых функций в настоящее время обеспечена визуальным стандартом MPEG-4 (MPEG-4 Visual standard) для натуральных изображений и видео изображена на рис. 4.8,

вместе с попыткой классификации уровней скоростей передачи данных в сравнении с наборами выполняемых функций.

На другом конце находится “VLBV ядро” (Very Low Bit-rate Video, низкоскоростное видео), которое обеспечивает

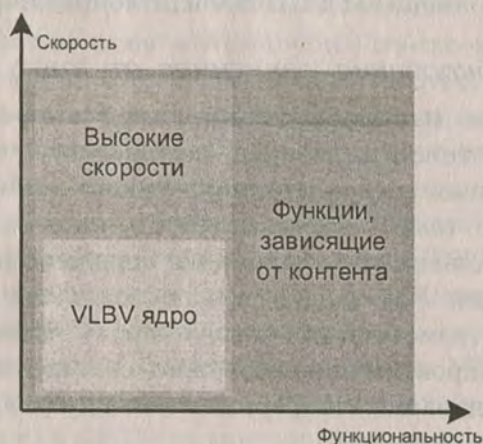


Рис. 4.7. Классификация инструментов кодирования MPEG-4

алгоритмы и инструменты для приложений, работающих со скоростями в диапазоне 5...64 кбит/с, поддерживая последовательности изображений с низким пространственным разрешением (не выше CIF разрешения) и с низкими скоростями кадров (обычно до 15 Гц). Основные приложения, которые поддерживаются VLBV ядром:

- а) VLBV кодирование последовательностей прямоугольных изображений с высокой эффективностью и низкой сложностью для мультимедийных приложений реального времени;
- б) обеспечение “произвольного доступа” и операций быстрой “прокрутки вперед” и “прокрутки назад” для устройств хранения VLB мультимедиаданных и доступа к приложениям.

Некоторые основные функции, описанные выше, также поддерживаются на высоких скоростях с набором пространственных и временных параметров увеличенным вплоть до рекомендаций ITU-R rec. 601, при этом используются

алгоритмы идентичные алгоритмам VLBV ядра. Диапазон скоростей при этом от 64 кбит/сек до 4 Мбит/с, что соответствует приложениям уровня вещания с организацией обратного канала и качеству цифрового ТВ.

Функции, основанные на контенте, поддерживают раздельное кодирование и декодирование контента. Это свойство MPEG-4 обеспечивает простой механизм интерактивности, гибкого представления и манипулирования контентом (изображениями и видео в сжатой области) без необходимости дальнейшей сегментации и транскодирования в приемнике.

Расширенные алгоритмы и инструменты MPEG-4 для функций, основанных на контенте, можно рассматривать как супернабор из VLBV ядра и инструментов для высоких скоростей, т. е. это инструменты VLBV и HBV ядер с дополнительными функциями.

Поддержка для стандартных и основанных на контенте функций

MPEG-4 Video стандарт поддерживает декодирование обычных прямоугольных изображений и видео так же, как и декодирование изображений, и видео с произвольной формой.

Кодирование обычных изображений и видео производится подобно стандартному процессу кодирования MPEG-1/2 и включает в себя алгоритмы предсказания/компенсации движения следующие за кодированием текстур. Для функций, основанных на контенте, где в последовательность изображений входят изображения с произвольной формой и положением, данное приближение расширено кодированием формы и информацией о структуре объекта. Форма может быть представлена 8 битными структурными компонентами, которые позволяют описать структуру объекта.

Расширенную аппроксимацию, основанную на контенте в MPEG-4, можно рассматривать как логическое расширение стандартного MPEG-4 VLBV ядра.

Схема кодирования видео в MPEG-4

Рисунок 4.8 позволяет увидеть в первом приближении MPEG-4 алгоритмы кодирования последовательностей как прямоугольных изображений, так и изображений с произвольной формой.

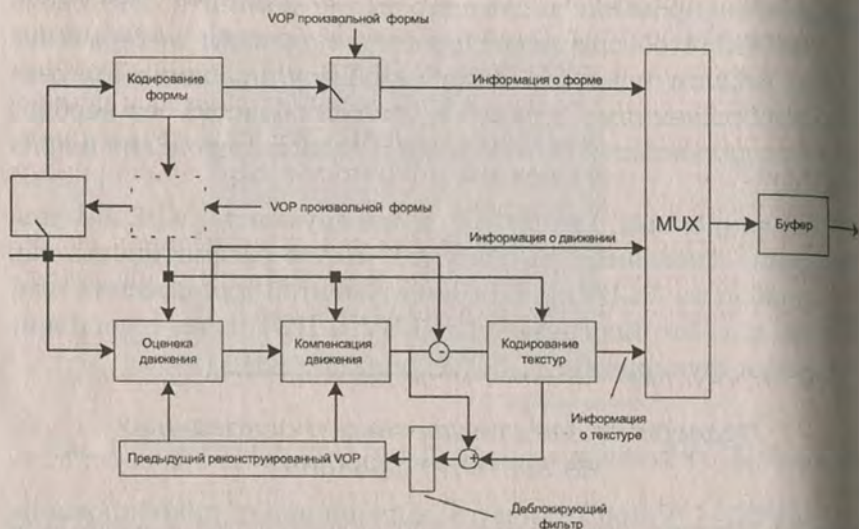


Рис. 4.8. Приближенная структура видеокодера MPEG-4

Основная структура процесса кодирования включает в себя кодирование формы (для видеообъектов с произвольной формой) и компенсацию движения как в DCT кодировании текстур (используется стандартное 8×8 DCT преобразование или адаптивное DCT).

Важное преимущество кодирования, основанного на контенте, — это то, что эффективность сжатия в этом случае может быть сильно улучшена, особенно для видеопоследовательностей с использованием инструментов предсказания движения отдельно для каждого объекта в сцене. Способы предсказания движения, которые могут быть использованы для эффективного кодирования и гибкого представления объектов:

- ♦ Стандартная, основанная на блоках размером 8×8 или 16×16 пикселей оценка и компенсация.

◆ Глобальная компенсация движения использующая 8 параметров движения и описываемое аффинным преобразованием.

◆ Глобальная компенсация движения, основанная на передаче статических “спрайтов”. Статический спрайт — это наибольшая часть неподвижного изображения, описывающая, например, панорамный задний план. Для каждого конкретного изображения в последовательности для реконструкции объекта кодируются только 8 глобальных параметров движения, описывающих движение камеры. Эти параметры представлены соответствующим аффинным преобразованием спрайта переданного в первом кадре.

◆ Глобальная компенсация движения, основанная на динамических спрайтах. Спрайты не передаются с первым кадром, но динамически генерируются в сцене.

Кодирование текстур и неподвижных изображений

Эффективное кодирование текстур и неподвижных изображений в MPEG-4 обеспечивается режимом визуальных текстур. Этот режим основан на алгоритме zerotree wavelet, который обеспечивает большую эффективность кодирования в достаточно большом диапазоне скоростей передачи данных. Вместе с большой эффективностью компрессии это так же обеспечивает пространственную стабильность и стабильное качество.

Масштабируемое кодирование видеообъектов

MPEG-4 поддерживает кодирование изображений и видеообъектов с пространственным и временным масштабированием. Масштабирование основано на возможности декодировать только часть потока данных и восстанавливать изображения или последовательности изображений с:

- ◆ уменьшением сложности декодера при уменьшении качества;
- ◆ уменьшенным пространственным разрешением;
- ◆ уменьшенным временным разрешением;

♦ с одинаковыми (эквивалентными) временным и пространственным разрешениями, но с меньшими качеством.

Это функционально желательно для прогрессивного кодирования изображений и видео в гетерогенных сетях для тех приложений, где приемник не способен отобразить изображения или видео последовательности с полным разрешением или высоким качеством. Это может быть в тех случаях, например, когда разрешение отображающего устройства ограничено.

Для декодирования неподвижных изображений стандарт MPEG-4 обеспечивает пространственное масштабирование с 11 уровнями градации и временное масштабирование до битового уровня. Для видеопоследовательностей поддерживается максимум 3 уровня градации.

MPEG-4 обеспечивает робастность ошибки для обеспечения доступа к изображениям или видеоинформации через широкий спектр носителей медиаданных и передающих средств. В особенности в условиях бурного роста мобильных коммуникаций важно обеспечить этот доступ к аудио и видеоинформации через беспроводные сети. Это подразумевает необходимость использования алгоритмов сжатия аудио и видео в средах с малым количеством ошибок на низких скоростях (т. е. менее чем 64 кбит/с).

Средства исправления ошибок, внедренные в MPEG-4, можно разделить на три основные категории. Эти категории включают в себя ресинхронизацию, восстановление данных и скрывание ошибок. Следует отметить, что эти категории не являются отдельными и самодостаточными.

Ресинхронизация

Инструменты ресинхронизации, как подразумевает их название, производят ресинхронизацию декодера и потока данных после обнаружения ошибок.

Приближение в ресинхронизации, принятое в MPEG-4, основано на пакетном приближении, так как структура группы блоков (Group of Blocks, GOB) используется в ITU-T стан-

дартах H.261 и H.263. В этих стандартах GOB определяется как один или более макроблоков. При поступлении новой GOB, информация, заложенная в ее заголовке, расположена в потоке данных. Эта информация в заголовке содержит в себе стартовый код GOB, который отличается от стартового кода изображения и позволяет декодеру обнаруживать эту группу. Более того, заголовок GOB содержит информацию, которая позволяет перезапустить процесс декодирования (например, ресинхронизировать декодер и поток и сбросить все предсказанные данные).

Аппроксимация GOB с ресинхронизацией основана на пространственной ресинхронизации. Это означает, что когда определенный макроблок включается в процесс кодирования, то в поток вставляется маркер ресинхронизации. При этом может возникнуть следующая проблема при аппроксимации: пока процесс кодирования идет с переменной скоростью, большинство маркеров могут вставляться в поток *unevenly*. Поэтому определенные части сцен, такие как зоны движения, будут *susceptible* к ошибкам.

Аппроксимация видео, принятая в MPEG-4, основана на обеспечении периодичности положения маркеров ресинхронизации в потоке данных. Другими словами, длина пакетов видео зависит не от количества макроблоков, а от числа бит в пакете. Если число бит содержащихся в данном пакете превышает определенный порог, то тогда формирование нового пакета начнется при поступлении нового макроблока.

Маркер ресинхронизации используется для различения момента старта нового видео пакета. Маркер отличается от всех возможных VLC кодов так же, как стартовый код VOP. Информация, содержащаяся в заголовке, также обновляется при старте пакета видео. Содержание этого заголовка — это информация, необходимая для рестарта процесса декодирования, включает в себя: номер макроблока, первого в этом пакете и параметры квантования, необходимые для декодирования этого первого макроблока. Номер макроблока необходим для пространственной ресинхронизации, в то вре-

мя как параметры квантования необходимы для ресинхронизации процесса декодирования.

Еще один заголовок, который включается в пакет, — это код расширения (Header Extension Code, HEC). HEC это единичный бит, который показывает наличие дополнительной информации о ресинхронизации. Эта дополнительная информация может понадобиться в том случае, если испорчен VOP заголовок.

Совместно с аппроксимацией пакетов видео для ресинхронизации в MPEG-4 может быть использован метод называемый синхронизацией с фиксированными интервалами. Этот метод требует, чтобы VOP стартовый код и маркеры ресинхронизации появлялись только в определенных фиксированных интервалах в потоке. Это помогает избегать проблем, связанных с эмуляцией стартовых кодов. Это означает, что если в потоке данных присутствуют ошибки, то они могут эмулировать VOP стартовый код. В этом случае когда используется фиксированный интервал синхронизации, то декодеру нужно искать VOP стартовый код только в начале каждого фиксированного интервала.

Восстановление данных

После того как синхронизация нарушена, инструменты восстановления данных начинают процесс восстановления потерянных данных. Эти инструменты не являются простыми кодами, корректирующими ошибки. Например, один из таких инструментов, который называется RVLC (Reversible Variable Length Code). В приближении это коды переменной длины, разработанные так, что они могут быть распознаны при чтении как в прямом, так и в обратном направлении.

Пример иллюстрирующий применение RVLC, дан на рис. 4.11. В ситуации, подобной приведенной на этом рисунке, когда несколько ошибок нарушили часть данных, все данные, расположенные между двумя точками синхронизации, будут потеряны. Как показано на рис. 4.9., RVLC позволяет

восстановить часть потерянных данных. Нужно отметить, что параметры, показанные на рисунке (QP и HES), представляют поля, зарезервированные в заголовке видеопакета для параметров квантования и кода расширения заголовка.

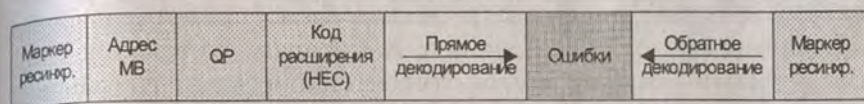


Рис. 4.9. Пример RVL

Скрытие (маскировка) ошибок

Скрытие ошибок является одним из наиболее важных компонентов в любом устойчивом к ошибкам кодере. Подобно инструментам восстановлению ошибок, описанным выше, эффективность стратегии скрытия ошибок сильно зависит от характеристик схемы ресинхронизации. По существу, если метод ресинхронизации может эффективно локализовать ошибку, то тогда проблема скрытия ошибки становится намного более легко решаемой. Для низких скоростей используемая в приложениях с низкой задержкой схема ресинхронизации обеспечивает достаточно приемлемый результат с простой стратегией скрытия, например, такой, как копирование блоков из предыдущих кадров.

В распознавании необходимость обеспечения расширенных возможностей скрытия заставила создать дополнительный режим устойчивый к ошибкам, что улучшает возможности декодера по локализации ошибок.

Данный режим использует разбиение данных разделением движения и текстур. Данное приближение требует, чтобы второй маркер ресинхронизации был вставлен между информацией о движении и о текстуре. Если информация о текстуре будет потеряна, то можно будет использовать информацию о движении для скрытия потери данных. Это значит, что, в то время как информация о движении исключается, данные о движении используются для компенсации движения декодированного ранее AVO.

Описание сцены

Поток описания сцены может включать в себя не только выбор титров на определенном языке или включение/выключение сопроводительных трехмерных предметов, но и процессы, которые до сих пор нельзя было представить себе применительно к теле вещанию. К методам, которые нас могут удивить сильнее всего, относится просмотр трехмерной сцены с различных ракурсов по желанию пользователя. Если информация передается с нескольких камер одновременно, не проблема выбрать любую из них. Если телестудия полностью виртуальная, т. е. находится в памяти компьютера, то можно вообще забыть о камере и без ограничений передвигаться по студии, естественно только виртуально, в памяти локальной пользовательской вычислительной системы.

В дополнение к обеспечению поддержки кодирования отдельных объектов MPEG-4 также обеспечивает удобства составления набора объектов внутри сцены. Необходимая для компоновки информация формирует описание сцены, которое кодируется и передается совместно с AV-объектами.

В порядке улучшения расширения возможностей для авторов инструменты манипуляции и взаимодействия и дескрипторы сцен кодируются независимо от потоков связанных с примитивными AV-объектами. Особое внимание посвящено идентификации параметров принадлежащих описанию сцен. Это производится дифференцированием параметров, которые используются для улучшения эффективности кодирования объекта и таких же параметров, которые используются для модификации объекта (например, изменения позиции объекта в сцене). Эти параметры расположены в описании сцены, а не в примитивных AV-объектах.

Как объекты собираются вместе. Сцена в MPEG-4 может быть представлена иерархической структурой в виде направленного нециклического графика. Каждый узел этого графа представлен AV-объектом, как показано на рис. 4.10. Структура этого дерева не является статичной, атрибуты узлов могут быть изменены (узлы могут быть добавлены, перемещены или удалены).

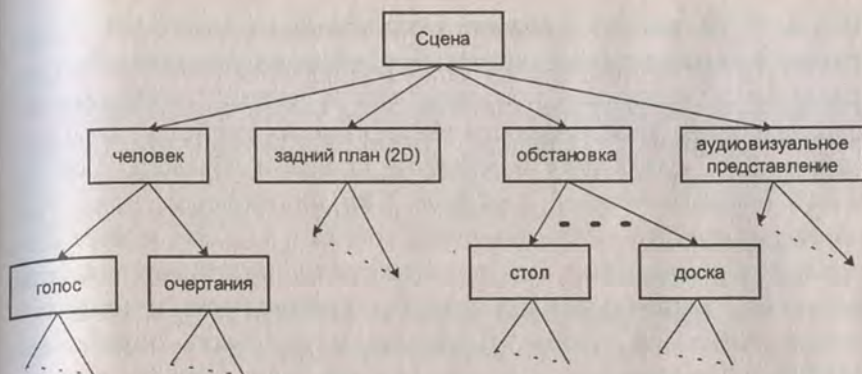


Рис. 4.10. Логическая структура сцены

Как объекты расположены в пространстве и времени. В MPEG-4 модели аудиовизуальные объекты имеют пространственные и временные параметры (размеры). Каждый AV-объект имеет свою локальную систему координат. *Локальная система координат* для объекта, в которой объект имеет фиксированное пространственно-временное положение и шкалу, одна. Локальная система координат служит как рукоятка управления расположением AV-объекта в пространстве и времени. AV-объект позиционируется в сцене путем точного определения преобразования координат из локальной координатной системы объекта в глобальную координатную систему определяемую одним или несколькими начальными узлами в дереве описания сцены.

Другие преобразования AVO. Структура описания сцены и семантика узлов сильно зависят от VRML. Это обеспечивает MPEG-4 большой набор операторов составления сцен, включая примитивную графику.

Взаимодействие с пользователем

MPEG-4 дает возможность пользователю взаимодействовать с контентом. Это взаимодействие может быть разделено на две категории: взаимодействие со стороны пользователя и взаимодействие со стороны сервера (client-side, server-side).

Взаимодействие со стороны пользователя включает в себя такую манипуляцию с контентом, которая производится локально на терминале пользователя и может принимать несколько форм. В особенности модификация свойств узлов описания сцен, например, изменение позиции объекта, возможность сделать объект видимым или невидимым, изменение размера шрифта синтетического текста и т. д. Это может быть сделан перевод действий пользователя (например, нажатие на кнопку мыши или ввод команд с клавиатуры) в корректировку описания сцены. Команды могут быть обработаны MPEG-4 терминалом в точно так же, как если бы они поступали из начального источника контента. Этот тип взаимодействия не требует стандартизации.

Другие формы "взаимодействия со стороны пользователя" требуют поддержки синтаксиса описания сцен и они должны быть стандартизированы. Использование VRML дает модель, с помощью которой создатели контента могут создавать полностью интерактивный контент.

Взаимодействие со стороны сервера (server-side interaction) включает в себя такие манипуляции с контентом, которые производятся на передающем конце и инициируются действиями пользователя. Это требует наличия обратного канала.

Связанная с контентом IPR идентификация и защита

MPEG-4 обеспечивает механизм для защиты интеллектуальной собственности (IPR). Это достигается пополнением кодированных AVO с опциональным набором IPR данных, несущих информацию о контенте (тип контента и держатель прав на контент). Набор данных, если он представлен, находится в заголовке потока, который взаимосвязан с объектом или описанием сцены. Число наборов таких данных гибко связано с каждым AVO, разные AVO могут разделять одни и те же или иметь обособленные наборы IPR данных. Возможности этих наборов позволяют осуществлять внедрение механизмов мониторинга, биллинга и защиты от копирования.

Информация о наполнении объектов (Object Content Information, OCI)

MPEG-4 позволяет добавлять к информации об объектах информацию о их наполнении. Пользователи стандарта могут использовать этот "OCI" поток для отправки текстовой информации вместе с MPEG-4 контентом.

4.3. Список основных функций, обеспечиваемых MPEG-4

DMIF

DMIF поддерживает выполнение следующих функций:

- ◆ прозрачность интерфейса MPEG-4 DMIF приложений независимо от типа медиаданных.
- ◆ Контроль за созданием FlexMux каналов
- ◆ Использование гетерогенных сетей: IP, ATM, мобильные сети, PTSN, узкополосные ISDN сетями.

Системы

◆ Описание сцен для создания составных AV объектах. Описание сцены обеспечивает большой набор узлов для двумерных и трехмерных операторов для композиции.

◆ Текст с поддержкой многоязыковости, выбора шрифтов и синхронизации.

◆ Интерактивность, включая: интерактивность на стороне пользователя и на стороне сервера; основная модель переключения событий и маршрутизации действий пользователей и т. д.

◆ Расслоение составных потоков в один поток (мультиплексирование).

◆ Независимость транспортного уровня.

◆ Инициализация и управление буферами приемного терминала.

◆ Временная идентификация, синхронизация и механизмы восстановления данных.

- ◆ Наборы данных позволяющие производить идентификацию прав на интеллектуальную собственность.

Аудио

Данный набор функций обеспечивает удобство широкого варьирования приложений от обычной речи до высококачественного многоканального звука и от натуральных звуков до синтезированных. MPEG-4 аудио поддерживает высокоэффективное представление аудио объектов состоящих из:

- ◆ *Речевые сигналы.* Кодирование речи может быть выполнено при скоростях от 2 кбит/с до 24 кбит/с с использованием инструментов кодирования речи. Для коммуникационных приложений можно обеспечить низкую задержку.

- ◆ *Синтезированная речь.* TTS алгоритм позволяет преобразовывать текст параметрами в четкую речь.

- ◆ Контроль за анимацией лица с образцами формы губ (Lip Shape Patterns) или с информацией о фонемах.

- ◆ Фонемная цепочка с информацией о длительности генерируемой для конвертера фонем в FAP.

- ◆ Функции режима поиска: пауза, продолжение, перемотка вперед/назад.

- ◆ Поддержка многоязыковых режимов для текста.

- ◆ Поддержка международных символов для фонем.

- ◆ *Узкополосные аудио сигналы.* Выполнение этого обеспечивает представление музыки и речи в каналах с ограниченной пропускной способностью (например, Интернет аудио). Это возможно при использовании T/F инструментов кодирования.

- ◆ *Широкополосные аудио сигналы.* Поддержка кодирования аудио с высоким качеством с использованием T/F декодеров. Это дает возможность поддержки качества звука широкополосного моно до широкополосного многоканального.

- ◆ *Синтезированное аудио.* Поддержка синтезированного звука обеспечена внедрением структурированного аудио декодера (Structured Audio Decoder).

◆ *Ограниченно сложное синтетическое аудио.* Это обеспечено внедрением структурированного аудио декодера (Structured Audio Decoder).

◆ *Масштабирование скоростей передачи данных* позволяет разбирать поток данных в поток с более низкой скоростью как комбинацию, которая может быть декодирована в сигнал несущий смысловую нагрузку. Разбор потока может быть сделан как во время передачи, так и в декодере.

◆ *Масштабирование полосы пропускания.*

◆ *Масштабирование сложности кодера.*

◆ *Масштабирование сложности декодера.*

◆ *Устойчивость к ошибкам.*

◆ *Создание аудио эффектов.*

Визуальная часть

MPEG-4 Visual стандарт обеспечивает возможность гибкого кодирования натуральных изображений и видео совместно с синтезированными (компьютерная графика) сценами. Это, например, позволяет обеспечить виртуальное присутствие участников видеоконференций. Видеостандарт обобщает инструменты и алгоритмы, поддерживающие кодирование натуральных неподвижных изображений и видеопоследовательностей и инструменты для сжатия синтезированной двухмерной и трехмерной графики.

Поддерживаемые стандарты

Поддерживаются следующие форматы и скорости:

◆ скорости: типовые между 5 кбит/сек и 4 Мбит/с;

◆ формата: прогрессивное и черезстрочное видео;

◆ разрешения: типовые от QCIF до ТВ.

Эффективность сжатия

◆ Эффективность сжатия видео поддерживается для всех скоростей передачи данных. Это включает в себя кодирование текстур с качеством регулируемым между "приемлемым" для высокоскоростного сжатия до "без потерь".

◆ Эффективность сжатия текстур для отображения текстур на двух- и трехмерных сетках.

◆ Случайный доступ к видео, позволяющий выполнять такие функции, как пауза, быстрая перемотка вперед и назад для сохраненного видео.

Функции, основанные на контенте

◆ Основанное на контенте кодирование изображений и видео позволяет отдельно декодировать и восстанавливать произвольные видео объекты.

◆ Случайный доступ контента в видео последовательностях для обеспечения выполнения таких функций, как пауза, быстрая перемотка вперед и назад для хранимых видео объектах.

◆ Расширенная манипуляция контентом в видео последовательностях для обеспечения выполнения таких функций, как деформирование синтетического и натурального текста, текстур, изображений и видео, основанное на реконструированном видеоконтенте. Например, отображение текста перед движущимся видео объектом, причем текст движется когерентно с объектом.

Масштабирование текстур, изображений и видео

◆ Масштабирование сложности в кодерах позволяет кодерам с разной сложностью генерировать несущие смысловую нагрузку потоки данных для конкретных текстур, изображений и видео.

◆ Масштабирование сложности в декодерах позволяет данным текстурам, изображениям и потокам видео быть декодированными декодерами с разными уровнями сложности. Качество восстановления в основном зависит от сложности используемого декодера. Это может означать, что менее мощный декодер декодирует только часть входящего потока.

◆ Пространственное масштабирование позволяет декодерам декодировать поднабор из полного потока данных созданного кодером для восстановления и отображения тек-

стур, изображений и видео объектов с меньшим пространственным разрешением. Для текстур и неподвижных изображений поддерживается 11 уровней масштабирования. Для видеопоследовательностей поддерживается 3 уровня.

◆ *Временное масштабирование* позволяет декодерам декодировать поднабор данных из полного потока для восстановления и отображения видео с меньшим временным масштабированием. Поддерживается 3 уровня масштабирования.

◆ *Масштабирование качества* позволяет анализировать поток данных как несколько потоков с разными скоростями (несколько уровней), такая комбинация из поднабора уровней затем может быть декодирована в полезный сигнал. Анализ потока может производиться как во время передачи, так и непосредственно в декодере. Качество восстановления при этом зависит от числа уровней используемых для декодирования и восстановления.

Кодирование формы

◆ *Кодирование формы* поддерживается для облегчения описания и компоновки стандартных изображений и видео.

Робастные и подверженные к ошибкам среды

◆ *Устойчивость к ошибкам* поддерживается для облегчения доступа изображений и видео к большому набору хранимых и передаваемых медиа данных. Это включает полезную работу алгоритмов сжатия изображений и видео в средах с большим количеством ошибок при низких скоростях (т. е. менее 64 кбит/с).

Анимация лица

Часть стандарта "Анимация лица" делает возможным посылку параметров, которые калибруют и анимируют синтезированные лица. Эти модели не стандартизированы в MPEG-4. Стандартизированы только параметры.

Функции и кодирование параметров анимации лица (независимая модель):

- ◆ характерные позиции и ориентация для анимации сеток определяющих лицо;

- ◆ движение губ, эквивалентное речевым фонемам.

Функции и кодирование параметров определяющих лица:

- ◆ характерные трехмерные позиции;

- ◆ трехмерная калибровка сеток для анимации;

- ◆ текстурная карта лица;

- ◆ персональные характеристики для каждого лица.

Кодирование текстур лица:

- ◆ кодирование двумерных сеток со скрытой структурой;

- ◆ предсказание, основанное на сетках и видеоизменение анимированных текстур;

- ◆ геометрическое сжатие.

4.4. Профили MPEG-4

Для DMIF и Систем в MPEG-4 имеются Профили: поднабор законченных инструментальных средств, которые дают возможность работы с большим набором приложений. Для Аудио и Видео частей в MPEG-4 также имеются Профили, но в этом случае существует различие между *профилем объекта (Object Profile)* и *профилем композиции (Composition Profile)*:

- ◆ профили объекта определяют синтаксис и инструменты для одиночных (Аудио или визуальных) объектов;

- ◆ профили композиции служат для определения того, какие объекты и как могут быть объединены в декодере (например, объект "текстура" может быть отображен на объекте "сетка"). Определение профилей композиции основывается на профилях объекта.

Для аудио- и видеообъектов определены свои профили композиции. Профили композиции не дают определения (описания) таким параметрам, как скорость потока, сложность, размер экрана, скорость отсчетов и т. д. Для определения

этих параметров вводятся уровни профиля композиции, но в первой версии стандарта MPEG-4 они не определены.

DMIF

Инструменты DMIF, содержащиеся в спецификациях DMIF профиле, охватывают следующие моменты:

- ◆ DMIF интерфейс приложений (DMIF-Applications Interface, DAI).

- ◆ Разработаны для прозрачности приложений для расположения и доступа к контенту MPEG.

- ◆ Для упаковки элементарных потоков MPEG-4 в транспортных сетях, основанной на QoS параметрах.

- ◆ DMIF-DMIF интерфейс (DDI) — для гарантии end-to-end QoS, основанного на предобслуживании пользовательского изображения на каждой DMIF конечной системе, участвующей в MPEG-4 соединении.

Эти инструменты подразделены на профили, соответствующие их применению для различных приложений.

Профиль 1: Вещательный DMIF

- ◆ DAI используется для прозрачного доступа контента MPEG-4 к различным вещательным медиаданным.

- ◆ DAI “подгоняет” различные пакеты с контентом к конкретным стандартам вещания.

Профиль 2: DMIF профиль вещания и хранения

DAI используется в дополнение к профилю 1 для:

- ◆ доступа к локально хранящимся файлам, использующим файловые форматы MPEG-4 на различных устройствах хранения медиаданных;

- ◆ запись на локальные устройства хранения с использованием файловых форматов MPEG-4.

DAI используется в дополнение к профилю 1 для:

- ◆ упаковка контента для соответствия отдельным форматам хранения мультимедиаданным.

Профиль 3: DMIF профиль вещания, хранения и удаленного взаимодействия

Уровень 1

DAI используется в дополнение к профилю 2 для:

- ◆ доступа для удаленного взаимодействия конечных DMIF конечных систем для улучшения базиса;
- ◆ доступа к файлам, хранимым на удаленных устройствах;
- ◆ установления доступа через различные транспортные медиа по единичным сетям.

DAI используется в дополнение к профилю 2 для:

- ◆ упаковки контента для соответствия стандартам (форматам) индивидуальных транспортных сетей.

Уровень 2

DAI используется в дополнение к профилю 2 для:

- ◆ доступа к удаленному взаимодействию оконечных DMIF систем с согласованием с конечным пользователем и определенным автором контента уровнем QoS;
- ◆ доступа к удаленным файлам на оконечных DMIF системах с согласованием с конечным пользователем и определенным автором контента уровнем QoS.

DAI используется в дополнение к профилю 2 для:

- ◆ упаковки контента для соответствия определенным стандартам транспортных сетей.

Основанные на DDI функции управления соединением и ресурсами сетей для управления сетевыми ресурсами и доступом к сети во время соединения.

Системы

Системные профили определены, чтобы ограничить производительность на участке разрешенных преобразований в наборщике. Эти преобразования передаются в BIFS узлах.

2D профиль

Приложения, поддерживающие только двухмерную графику, соответствуют "2D профилю". Этот профиль предназначен

начен для систем, которые работают с несложной двумерной графикой с двумерными преобразованиями.

3D профиль

Приложения, поддерживающие полный набор трехмерной графики, соответствуют “3D профилю”. Этот профиль предназначен для графических систем, которые работают с трехмерной графикой, которая требует более сложных операций, чем в “2D профиле”.

VRML профиль

Приложения, претендующие на соответствие “VRML профилю”, должны выполнять все узлы, и причем только описанные в спецификации стандарта MPEG-4, эти узлы также являются общими для спецификации DIS 14472-1 (VRML). Цель этого профиля в том, чтобы максимизировать возможность взаимодействия между MPEG-4 и VRML.

Полный профиль

Приложения, поддерживающие полный набор возможностей, описанных в Части 1 Systems Committee Draft, оформлены в “Полном профиле”.

Аудиофиль

Приложения, поддерживающие все связанное только с аудиоузлами соответствуют “аудио профилю систем”. Этот профиль предназначен только для аудиоприложений.

Аудио

Эта часть содержит профили объектов для натурального аудио, синтезированного аудио и преобразования текста в речь (TTS). Она также включает в себя секцию, обрисовывающую текущие согласования для профилей создания аудиоданных.

Таблица 4.2

Профиль объекта: натуральное аудио

Профиль	Иерархическая информация	Инструменты
Основной	Содержит простое масштабирование	<ul style="list-style-type: none"> • все инструменты для натурального аудио • все инструменты для обработки речи • синтаксис для масштабирования
Простой масштабируемый	Содержит натуральную речь, основанную на: CELP параметрических алгоритмах	<ul style="list-style-type: none"> • CELP инструменты • 13818-7 LC профиль • 13818-7 SSR профиль • AAC с PNS и LTP • масштабируемое CELP/AAC • другие T/F инструменты определяемые в тестах
Простое параметрическое аудио		<ul style="list-style-type: none"> • HILN • HVXC

Таблица 4.3

Профили объекта для синтезированного аудио

Профиль	Иерархическая информация	Инструменты
Основной	Содержит все остальные	• все структурированные аудио инструменты
Алгоритмический синтез	Содержит синтаксис основного MIDI профиля	<ul style="list-style-type: none"> • SAOL • SASL • MIDI
Wavetable синтез	Содержит синтаксис основного MIDI профиля	<ul style="list-style-type: none"> • SASBF • MIDI
Основной MIDI	нет	• MIDI

Таблица 4.4

TTS профили

Профиль	Иерархическая информация	Инструменты
TTS	нет	TTS интерфейсы

Профиль создания аудио

Профиль	Иерархическая информация	Инструменты
Основной	Содержит простой	<ul style="list-style-type: none"> • Основное натуральное аудио • Структурированное аудио • TTS интерфейс • Создание
Простой	Содержит речь	<ul style="list-style-type: none"> • Простое масштабируемое аудио • Алгоритмический синтез
Речь		<ul style="list-style-type: none"> • Параметрическое кодирование речи • Кодирование речи? основанное на CELP

Профили визуальных объектов

Первый профиль объекта называется простой профиль объекта. Второй называется базовый простой профиль. Эти два профиля иерархичны, базовый профиль обобщает простой. Так же существует профиль 12 битных видеообъектов, который представляет из себя простой профиль, к которому добавлены инструменты кодирования 4—12 битовых пикселей видео. Позже профиль объекта был определен для кодирования масштабируемых текстур.

Простой профиль видео объектов

Данный профиль использует следующий инструменты:

- ◆ Intra режим (I);
- ◆ режим с предсказанием (P);
- ◆ AC/DC предсказание;
- ◆ ресинхронизация слайсов;
- ◆ таблицы кодов с переменной длиной;
- ◆ разбиение данных;
- ◆ двоичная форма с единой прозрачностью.

Профиль 12 битных видеообъектов

Данный профиль аналогичен предыдущему, добавлен только один инструмент:

- ◆ кодирование видео с 4—12 битными пикселями.

Базовый профиль видеообъектов

Следующий профиль, имеющий рабочее название "базовый" (Core), включает в себя следующие инструменты:

- ◆ все инструменты для простого профиля;
 - ◆ режим двунаправленного предсказания (В-режим);
 - ◆ H.263/MPEG-2 таблицы квантования;
 - ◆ неограниченные вектора движения;
 - ◆ четыре вектора движения на макроблок;
 - ◆ статические спрайты (выборки);
 - ◆ временное масштабирование;
 - ◆ основанная на кадрах структура;
 - ◆ основанная на объектах структура;
 - ◆ пространственное масштабирование;
 - ◆ инструменты для кодирования черезстрочного видео.
- Профиль масштабируемой текстуры изображения:
- ◆ масштабируемое Wavelet преобразование;
 - ◆ двоичная форма.

SNHC

Объектовый профиль для анимации лица — Простой профиль

По существу, *Простой объектовый профиль для анимации лица (Simple Facial Animation Object Profile)* представляет из себя только набор требований к декодеру по анимации тела и лица (Facial & Body Animation, FBA) для правильного расположения или изменения положения для соответствующих характеристических 3D точек, для которых были получены FAP параметры (Facial Animation Parameters). Если приняты только FAP параметры (без FDP параметров), то характеристические точки будут расположены на локальной модели лица, которая известна декодеру. В этом профиле все FDP данные (характеристические точки, сетка, текстура) могут быть проигнорированы при декодировании лица.

Объектовый профиль для анимации лица — Продвинутый профиль

Данный профиль включает в себя все требования простого профиля. Кроме этого, в этом профиле декодер должен декодировать всю остальную принимаемую информацию.

4.5. Кодирование видеоданных в MPEG-4

MPEG-4 кодера представляет собой схему в виде соединения физических объектов с присущими им свойствами, такими как форма, движение и текстура. Следующие друг за другом VOP принадлежащие вышеупомянутым физическим объектам в сцене рассматриваются как VOP последовательность видеобъектов со случайной формой и позицией. Каждый кадр видеопоследовательности представляется как набор VOP, который должен быть закодирован и декодирован VM (Verification Model, контрольная модель кодера). Так же архитектура MPEG-4 кодера дает возможность достижения параллельного кодирования VOP, как это показано на рис. 4.11. Кодер выполняет процесс кодирования следующим образом: в каждый момент времени кодер, выбирает тот VOP, который имеет минимальные вес для кодирования, кодирует его и сохраняет закодированные данные на диске. Вдобавок происходит проверка на предмет того, не является ли данный VOP последним VOP в кадре, и если это так, то смешивает его и записывает смешанный кадр на диск, в то время как те же операции продолжаются со следующими VOP. Кодер прекращает работать тогда, когда завершается кодирование всех VOP. Смешивание осуществляется с помощью процедуры псевдо составления, которая составляет все VOP принадлежащие одному кадру в порядке его восстановления. Это делается для тестирования и не имеет сильного влияния на процедуру кодирования. Подлинное составление сцены происходит на системном уровне.

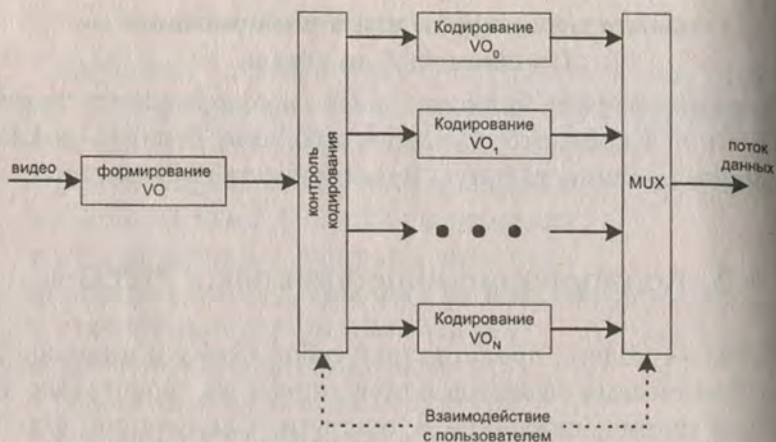


Рис. 4.11. Структура VM-кодера

Модель поставщик—потребитель

Обработка данных в реальном времени с большими частотами кадров и большими размерами изображений CCIR-601 является неотъемлемой частью реализации параллелизма в кодеке MPEG-4, и это дает возможность реализации процесса кодирования на многопроцессорных платформах. Принимая во внимание то, что кодер всегда выполняет одни и те же инструкции с различными данными, становится возможным параллельный запуск более одной ступени одной программы (кодера) для работы с различными наборами данных. Модель поставщик—потребитель реализует это. Различные этапы (ступени) программы (MPEG-4 кодера) могут выполняться с помощью разных процессоров (*потребители*), обрабатывающими данные, становящимися доступными в процессе их распределения (*поставщик*). В UNIX это выполняется путем использования многопоточного программирования (multithread programming, MT), где потоки (процессы) назначаются разным процессорам (в Solaris 2.0 для это имеется двухуровневая модель). Основные причины для выбора многопоточного внедрения:

1. Улучшение производительности при использовании мультипроцессорного оборудования: повторяющиеся подпро-

цессы (трэды) — это эффективный путь использования параллелизма оборудования.

2. Пропускная способность: использование таких подпроцессов делает реальным использование перерывов в работе CPU, например, подпроцессы, которые делают I/O запрос должны ждать, пока выполняется другой подпроцесс в процессе.

3. Системные ресурсы: внедрение мультипроцессорного метода в приложениях может создать сотни или даже тысячи подпроцессов, один для всех выполняет задачу синхронизации, что дает незначительное воздействие на системные ресурсы.

4. Единый источник для многозадачных платформ.

На рис. 4.12 показана версия Поставщик—Потребитель кодера VM 4.0. Она состоит из следующих частей: часть Поставщика, часть синхронизации Буферы/Подпроцессы, часть Кодер/Потребитель.

Часть Поставщика предназначена для выполнения задач чтения и распределения VO. Распределение VOP от поставщика к кодерам/потребителям происходит через часть синхронизирующих буферов. В этой модели существуют два типа буферов. Внешние буферы, куда поставщик посылает информация о VOP и внутренние буферы, из которых кодеры копируют данные о VOP для последующей обработки. Статус поставщика может быть активным или пассивным. Поставщик является пассивным тогда, когда внешние буферы наполнены VOP данными. Это значит, что поставщик посылает данных больше, чем потребитель может обработать. Подача данных блокируется до тех пор, пока хотя бы один из внешних буферов не опустеет. Контроль блокировки осуществляется за счет использования сигнализации, которая показывает емкость системных ресурсов (буферов).

Часть синхронизации буферов. Как сказано выше, каждому кодеру назначается внешний буфер в который поставщик записывает VOP данные и внутренний буфер (зона локальной памяти), из которого кодер копирует данные для пос-

ледующей обработки (векторы движения, текстура и т. д.). Когда данные переписываются из внешнего буфера во внутренний, тогда кодер готов для получения новых данных (внешний буфер пуст). Данные о VOP могут не копироваться во внутренний буфер, пока кодер/потребитель не закончит кодирование предыдущего VOP. Поставщик заполняет пустые буфера по следующему правилу: “Минимизировать время ожидания кодером/потребителем получения данных”. Это достигается за счет отказа от стандартного порядка управления буферами и отдельной проверкой статуса каждого буфера.

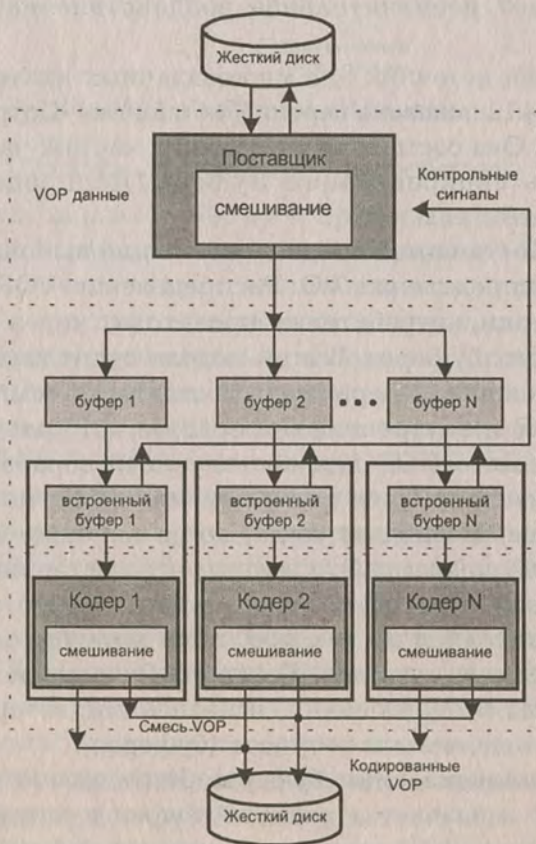


Рис. 4.12. Модель поставщик—потребитель

Часть пользователь/кодер. Эта часть осуществляет кодирование. Она предназначена для кодирования формы, движения (оценка и компенсация) и текстуры в соответствии с MPEG-4 алгоритмами кодирования VM 4.0. Более того, эта часть осуществляет запись потоков VOP бит на жесткие диски и смешивание для всех последующих кадров.

Анализ модели и ее производительность

Теоретически, если в модели поставщик—пользователь р используются процессоры (кодеры), то тогда общий рост скорости будет равен р. Это, конечно, не предполагает добавления служебных данных от переданных через архитектуру шины параллельной системы данных совместно с оптимизированным программированием на уровне процессора. Но если служебные данные для передачи VOP в процессоры нулевые, то скорость может и не достичь своего максимального значения по следующим причинам:

VOP зависимость. Стандарт MPEG-4, так же как и предыдущие (MPEG-1/2), использует временную избыточность последовательности видео кадров путем применения трех типов VOP: I-, P- и B-VOP. I-VOP может быть закодирован независимо, в то время как P и B-VOP кодируются с использованием предыдущих I, P-VOP. Итак, если n VOP существуют там, где $n/2$ I-VOP и $n/2$ P-VOP, то тогда в начале процесса кодирования не более, чем $n/2$ VOP, могут быть закодированы, даже если n процессоров свободны в данный момент.

Размер VOP. Время кодирования в некоторой степени связано с размером VOP. Размер VOP показывается в макроблоках. Макроблок является наименьшим элементом, который может быть обработан кодером. Он имеет размер 16×16 пикселей и состоит из четырех яркостных блоков размером 8×8 и двух цветных, также имеющих размер 8×8 . Если два VOP имеют один и тот же размер, то это совсем не означает, что они также имеют и одно время кодирования. Рисунок 4.13 иллюстрирует временные измерения для нескольких

VOP разных размеров из *coastgurd_cif.yuv* видеопоследовательности (cif, разрешение 352×288 пикселей). На этом графике явно видна 61 точка, каждая из которых представляет 61 размер VOP, и, как можно увидеть, время кодирования не всегда пропорционально размеру VOP.

Тип VOP. Время кодирования VOP зависит от его типа. Например, для кодирования P-VOP требуется времени больше

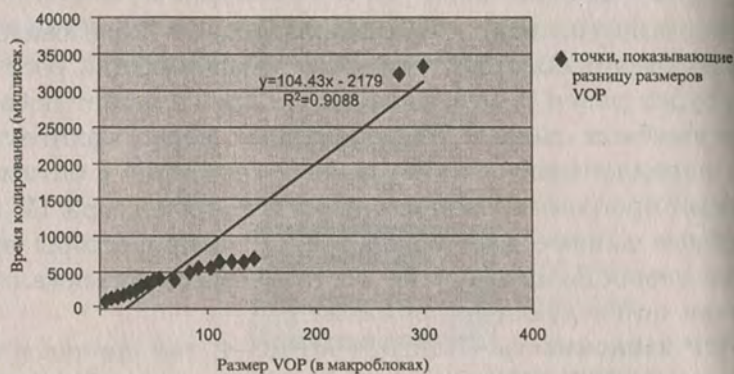


Рис. 4.13. Зависимость времени кодирования от размера VOP

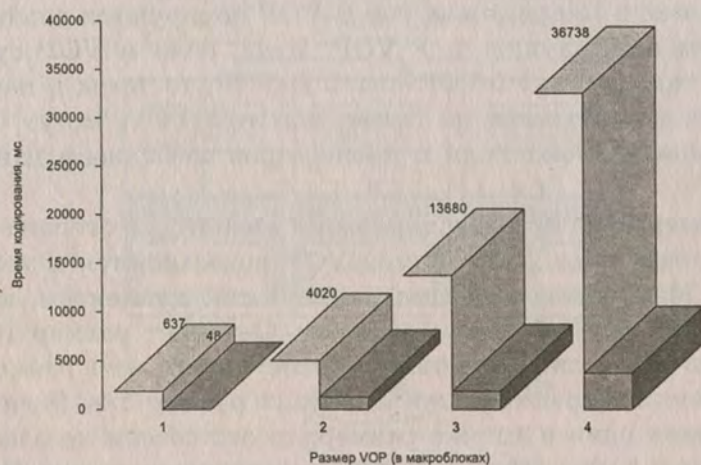


Рис. 4.14. Зависимость времени кодирования от размера VOP

ше, чем для кодирования I-VOP, в основном из-за оценки движения. На рис. 4.14 показано время кодирования I и P-VOP. Измерения были произведены для четырех размеров VOP (6, 64, 132 и 308 макроблоков), входящих в последовательность *coasguard_cif.yuv*. Серые колонки показывают время кодирования для I-VOP и белые/затененные — время кодирования для P-VOP.

4.6. Кодирование аудиоданных в MPEG-4 ISO/IEC 14496-3

Этот стандарт, разработанный для мультимедиаприложений, объединяет опыт, накопленный группой MPEG при разработке всех рассмотренных алгоритмов компрессии цифровых данных [4]. Здесь впервые при компрессии цифровых данных высококачественных звуковых сигналов предложено использовать параметрическое кодирование, при котором реальный звуковой сигнал представляется в виде модели, содержащей совокупность тональных и шумоподобных сигналов. В этом случае (рис. 4.15) исходный сигнал выборки $s(n)$ в блоке анализа-синтеза разделяется на тональные и шумоподобные составляющие, далее для базовой параметрической модели сигнала оцениваются значения текущих частот, фаз и амплитуд тональных сигналов и уровней энергии шумоподобных сигналов в определенных полосах частот. Значения перечисленных параметров квантуются и кодируются минимально возможным количеством бит, которое определяется с помощью психоакустической модели. На приемной стороне системы по этим параметрам синтезируется исходный сигнал. До последнего времени параметрическое моделирование использовалось только при компрессии цифровых данных речевых сигналов, более простых по своей структуре, чем музыкальный сигнал. Однако в последние годы благодаря успехам вычислительной техники математического моделирования, психофизики и электроники, параметрическое

представление все чаще начинает применяться и при кодировании высококачественных звуковых сигналов, обеспечивая более высокий уровень компрессии цифровых данных. Параметрическое кодирование, обладая очень сложными процедурами оценки параметров и требующее при реализации существенно больших вычислительных затрат, позволяет получить скорость цифрового потока 16...24 кбит/с при достаточно хорошем качестве.

Вторым альтернативным алгоритмом, обеспечивающим более высокое качество, однако при большей скорости цифрового потока является алгоритм AAC (стандарт ISO/IEC 13818—7), также вошедший в этот стандарт.



Рис. 4.15. Упрощенная схема кодера, использующего метод параметрического кодирования звукового сигнала

Гибридные методы кодирования звуковых сигналов.

Известны, как уже было сказано ранее, и гибридные методы кодирования (рис. 4.19), использующие параметрическое описание звукового сигнала на тех временных отрезках, где он более или менее монотонен, и алгоритм компрессии MPEG-2 AAC стандарта ISO/IEC 13818—7 на участках быстрого изменения его временной функции. Эти участки выделяются с помощью так называемого детектора выбросов

(атак). По данным [5], качество кодирования высококачественных звуковых сигналов оказывается уже достаточно высоким при скорости цифрового потока 32 кбит/с на канал.

Качество систем кодирования с компрессией цифровых аудиоданных

Результаты экспертных оценок качества систем кодирования с компрессией цифровых данных представлены [5]. Экспертизы выполнены в соответствии с Рек. 562—3 “Субъективная оценка качества звука” МККР. В качестве звукового материала использовались отрывки звучаний разных жанров, взятые с компакт-диска SQAM, рекомендуемого для проведения экспертиз исследовательской группой MPEG. Разница в звучании исходного отрывка и отрывка, прошедшего кодек, оценивалась по 5-балльной шкале.

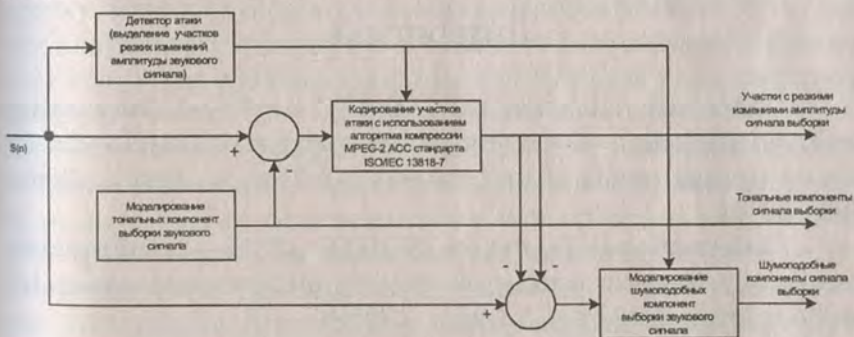


Рис. 4.16. Идея построения кодера, реализующего гибридный метод кодирования аудиоданных

Градации оценки ухудшения качества звучания (5-балльная шкала оценки):

- 0... незаметное;
- 1... заметное, но не раздражающее;
- 2... слегка раздражающее;
- 3... раздражающее;
- 4... очень раздражающее.

Из табл. 4.6. видно, что кодек MPEG-1 Layer-3 несколько уступает по скорости передачи при равном качестве только

MPEG-2 AAC на скорости передачи 128 кбит/с. Наиболее эффективными, допускающими наибольшее число степеней свободы для усовершенствования являются алгоритмы компрессии MPEG-1 ISO/IEC 11172—3 Layer-3 и MPEG-2 ISO/IEC 13818—7 AAC (Advanced Audio Coding) [3].

Таблица 4.6

Результаты экспертной оценки качества кодеков с компрессией цифровых аудиоданных

Наименование алгоритма	96 кбит/с	128 кбит/с	160 кбит/с	192 кбит/с
MPEG-2 AAC	-1,15	-0,17	—	—
MPEG Layer-3	—	-1,73	—	—
Dolby AC-3	—	-2,11	-1,04	-0,52
MPEG Layer-2	—	-2,14	-1,75	-1,18

Литература

1. International Standard ISO/IEC 11172—3. Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1, 5 Mbit/s- Part 3: Audio, 1993-08-01.

2. International Standard ISO/IEC 13818—3. Information technology-Generic coding of moving pictures and associated audio information. Part 3:Audio, 1995-05—15.

3. International Standard ISO/IEC 13818—7. Information technology-Generic coding of pictures and associated audio information. Part 7: Advanced Audio Coding (AAC), 1997 (E).

4. ISO/IEC FCD 14496—3 Subpart 1. Information Technology-Very Low Bitrate Audio-Visual Coding. Part 3: Audio, 1998-05—10 (ISO/JTC 1/SC 29. — N 2203).

5. *Levine S.* Audio Representation for Data Compression and Compressed Domain Processing. A Dissertation Submitted to the Department of Electrical Engineering and the Committee on Graduate Studies of Stanford University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy. — Dezember. — 1998.

Глава 5

MPEG-7

5.1. Область применения стандарта MPEG-7

Доступно все больше и больше аудиовизуального материала из самых разных источников. Информация может быть представлена в различных медийных формах, таких как статические изображения, графика, 3D модели, звук, голос, видео [1—5]. Аудиовизуальная информация играет важную роль в обществе, будучи записана на магнитную или фото пленку, или поступая в реальном масштабе времени от аудио или визуальных датчиков в аналоговой или цифровой форме. В то время как аудиовизуальная информация первоначально предназначалась для людей, в настоящее время все чаще такие данные генерируются и передаются и воспринимаются компьютерными системами. Это может быть, например, сопряжено с распознаванием голоса или изображения, медийным преобразованием (голос в текст, картинку в голос, голос в картинку, и т. д.). Другими сценариями являются извлечение информации (быстрый и эффективный поиск для различных типов мультимедийных документов, представляющих интерес для пользователя), фильтрация потоков описаний аудиовизуального материала (чтобы получить только те элементы мультимедиаданных, которые удовлетворяют предпочтениям пользователя). Например, программа во время телепередачи запускает соответствующим образом программируемый VCR, чтобы записать эту программу, или сенсор изображения выдает предупреждение, когда происходит определенное событие. Автоматическое транскодирование может быть выполнено для строки символов, преобразовав ее в аудиоданные, или можно провести поиск в потоке аудио или видеоданных. Во всех этих примерах, аудиовизуальная ин-

формация была приемлемым образом закодирована, что позволяет программе ЭВМ предпринять соответствующие действия.

Аудиовизуальные источники будут играть в перспективе все большую роль в нашей жизни, и будет расти необходимость обрабатывать такие данные. Это делает необходимым обработку видов аудиовизуальной информации, имеющей волновую форму, компрессированный формат (такой как MPEG-1 и MPEG-2) или даже объектно-ориентированный (такой как MPEG-4) формат. Необходимы формы презентации, которые позволяют некоторую степень интерпретации смысла информации. Эти формы могут быть переданы или доступны для прибора или программы ЭВМ. В примерах приведенных выше датчики изображения могут генерировать визуальные данные не в форме РСМ (значения пикселей), а в форме объектов с ассоциированными физическими величинами и временной информацией. Эти объекты могут быть запомнены и обработаны с целью проверки, выполняются ли в них определенные условия. Видеозаписывающий прибор может получить описания аудиовизуальной информации, ассоциированной с программой, которая при выполнении заданных условий выдаст команду на запись, например, только новости за исключением спорта или запись фильма с автоматическим вырезанием вставок рекламы (согласитесь, об этом сегодня можно только мечтать).

MPEG-7 будет стандартом для описания мультимедийных данных, которые поддерживают определенные операционные требования [5]. MPEG не стандартизует приложения. MPEG может, однако, использовать приложения для понимания требований и развития технологий. Должно быть ясно, что требования, сформулированные в данном документе, получены из анализа широкого диапазона потенциальных приложений, которые могут использовать описания MPEG-7. MPEG-7 не ориентирован на какое-то конкретное приложение, скорее, элементы, которые стандартизует MPEG-7, будут поддерживать максимально широкий диапазон приложений.

Цель MPEG-7

Аудиовизуальный материал MPEG-7 может включать в себя: статические изображения, графику, 3D модели, звук, голос, видео и композитную информацию о том, как эти элементы комбинируются при мультимедийной презентации. В особых случаях этих общих видов данных сюда может включаться выражения лица и частные характеристики личности.

Средства описаний MPEG-7, однако, не зависят от способа кодирования и записи материала. Можно сформировать описание MPEG-7 аналогового фильма или картинки, которая напечатана на бумаге, точно так же, как и цифрового материала.

MPEG-7, как и другие объекты семейства MPEG, предоставляют стандартное представление аудио-визуальных данных, удовлетворяющих определенным требованиям. Одной из функций стандарта MPEG-7 является обеспечение ссылок на определенные части мультимедийного материала. Например, дескриптор формы, используемый в MPEG-4, может оказаться полезным в контексте MPEG-7, точно также это может относиться к полям вектора перемещения, используемым в MPEG-1 и MPEG-2.

В своих описаниях MPEG-7 допускает различную гранулярность, предлагая возможность существования различных уровней дискриминации. Хотя описание MPEG-7 не зависит от кодового представления материала, он может использовать преимущества, предоставляемые кодированным материалом MPEG-4. Если материал кодирован с использованием MPEG-4, который предоставляет средства кодирования аудио-визуального материала в виде объектов, имеющих определенные связи во времени (синхронизация) и в пространстве (на сцене для видео или в комнате для аудио), будет возможно связать описания с элементами (объектами) в пределах сцены, такими как аудио- и видеообъекты.

Так как описательные характеристики должны иметь смысл в контексте приложения, они будут различными для разных приложений. Это подразумевает, что один и тот же

материал может быть описан различным образом в зависимости от конкретного приложения. Возьмем в качестве примера визуальный материал: нижним уровнем абстракции будет описание, например, формы, размера, текстуры, цвета, движения (траектории) и позиции ("где на сцене может размещаться объект"). А для аудио: ключ, тональность, темп, вариации темпа, положение в звуковом пространстве. Высшим уровнем представления будет семантическая информация: "Это сцена с лающей коричневой собакой слева и голубым мячом, падающим справа, с фоновым звуком проезжающих авто". Могут существовать промежуточные уровни абстракции.

Уровень абстракции относится к способу выделения определенных характеристик: многие характеристики нижнего уровня могут быть выделены полностью автоматически, в то время как характеристики высокого уровня требуют большего взаимодействия с человеком.

Кроме описания материала, требуется также включить другие виды информации о мультимедийных данных:

- ◆ **Форма.** Примером формы является используемая схема кодирования (например, JPEG, MPEG-2), или общий объем данных. Эта информация помогает определить, может ли материал быть воспринят пользователем.

- ◆ **Условия доступа к материалу.** Это включает учет ограничений на использование материала, учитывающих авторские права и права собственности, а также цену.

- ◆ **Классификация.** Это включает оценку происхождения материала и его классификацию по predetermined categories.

- ◆ **Связь с другим важным материалом.** Информация может помочь пользователю ускорить поиск.

- ◆ **Контекст.** В случае записанного документального материала очень важно знать обстоятельства записи (например Олимпийские игры 1996, финал 200-метрового забега для мужчин с барьерами).

Во многих случаях будет желательно использовать для описания текстовые данные. Необходимо позаботиться о том

чтобы полезность описаний была независима по возможности от языка. Хорошим примером текстуального описания являются указания авторов, названия фильма и пр.

Следовательно, средства MPEG-7 позволят формировать описания (т. е. наборы схем описания и соответствующих дескрипторов по желанию пользователя) материала, который может содержать:

- ◆ информацию, описывающую процессы создания и производства материала (директор, заголовок, короткометражный игровой фильм);

- ◆ информацию, относящуюся к использованию материала (указатели авторского права, история использования, описание вещания);

- ◆ информацию о характеристиках записи материала (формат записи, кодирование);

- ◆ структурную информацию о пространственных, временных или пространственно-временных компонентах материала (разрезы сцены, сегментация областей, отслеживание перемещения областей);

- ◆ информацию о характеристиках материала нижнего уровня (цвета, текстуры, тембры звука, описание мелодии);

- ◆ концептуальную информацию о реальном содержании материала (объекты и события, взаимодействие объектов);

- ◆ информацию о том, как эффективно просматривать материал (конспекты, вариации, пространственные и частотные субдиапазоны...);

- ◆ информацию о собрании объектов;

- ◆ информацию о взаимодействии пользователя с материалом (предпочтения пользователя, история использования).

Все эти описания являются, конечно, эффективно закодированными для поиска, отбора и т. д.

Чтобы удовлетворить этому многообразию дополнительных описаний материала, MPEG-7 осуществляет описание материала с нескольких точек зрения. Наборы средств описаний, разработанные с учетом этих точек зрения, представляются в виде отдельных объектов. Однако они взаимосвязаны и

могут комбинироваться множеством способов. В зависимости от приложения, некоторые будут присутствовать, а другие отсутствовать, могут присутствовать лишь частично.

Описание, сформированное с помощью средств MPEG-7, будет ассоциировано с самим материалом, чтобы позволить быстрый и эффективный поиск и фильтрацию материала, представляющего интерес для пользователя.

Данные MPEG-7 могут физически размещаться вместе с ассоциированным АВ-материалом, в том же информационном потоке или в той же системе памяти, но описания могут также размещаться на другом конце света. Когда материал и его описания размещены не совместно, необходим механизм для соединения АВ-материала и его описаний MPEG-7; эти связи должны работать в обоих направлениях.

Тип материала и запрос могут не совпадать; например, визуальный материал может быть запрошен, используя визуальное содержимое, музыка, голос, и т. д. Согласование данных запроса и описания MPEG-7 выполняется поисковыми системами и агентами фильтрации.

MPEG-7 относится ко многим различным приложениям в самых разных средах. Этот стандарт должен обеспечивать гибкую и масштабируемую схему описания аудиовизуальных данных. Следовательно, MPEG-7 не определяет монолитную систему описания материала, а предлагает набор методов и средств для различных подходов описания аудио-визуального материала. MPEG-7 сконструирован так, чтобы учесть все подходы, учитывающие требования основных стандартов, таких как SMPTE Metadata Dictionary, Dublin Силие, EBU P/Meta, и TV Anytime. Эти стандарты ориентированы на специфические приложения и области применения, в то время как MPEG-7 пытается быть как можно более универсальным. MPEG-7 использует также схему XML в качестве языка выбора текстуального представления описания материала. Главными элементами стандарта MPEG-7 являются:

◆ **Дескрипторы (D).** Представление характеристик, которые определяют синтаксис и семантику представления каждой из характеристик.

◆ Схемы описания DS (Description Scheme), которые специфицируют структуру и семантику взаимодействия между компонентами. Эти компоненты могут быть дескрипторами и схемами описания.

◆ Язык описания определений DDL (Description Definition Language), позволяющий создавать новые схемы описания и, возможно, дескрипторы и обеспечивающий расширение и модификацию существующих схем описания,

◆ Системные средства служат для поддержки мультиплексирования описаний, синхронизации описаний и материала, механизмов передачи, кодовых представлений (как текстуальных, так и двоичных форматов) для эффективной записи и передачи, управления и защиты интеллектуальной собственности в описаниях MPEG-7.

Область действия стандарта

MPEG-7 относится к приложениям, которые могут осуществлять запись (или реализовать поточную передачу, например, производить широковещательную пересылку в Интернет) и могут работать как в реальном времени так и off-line. "Среда реального времени" в данном контексте означает, что описание генерируется в процессе приема материала.

На рис. 5.1 показана блок-схема системы обработки данных MPEG-7. Чтобы полностью использовать возможности описаний MPEG-7, автоматическое извлечение характеристик (или "дескрипторов") может оказаться особенно заметным. Ясно также, что автоматическое извлечение не всегда возможно. Как было указано выше, чем выше уровень абстракции, тем труднее автоматическое извлечение характеристик и тем полезнее интерактивные средства.

Чтобы улучшить понимание терминологии, введенной выше (т. е. дескриптор, схема описания и DDL), рассмотрите рис. 5.2 и рис. 5.3.

На рис. 5.2 продемонстрирована масштабируемость рассмотренной концепции. Более того, там показано, что DDL

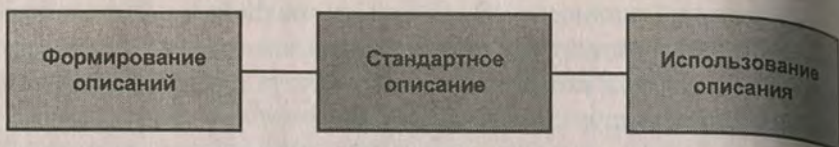


Рис. 5.1. Область MPEG-7.

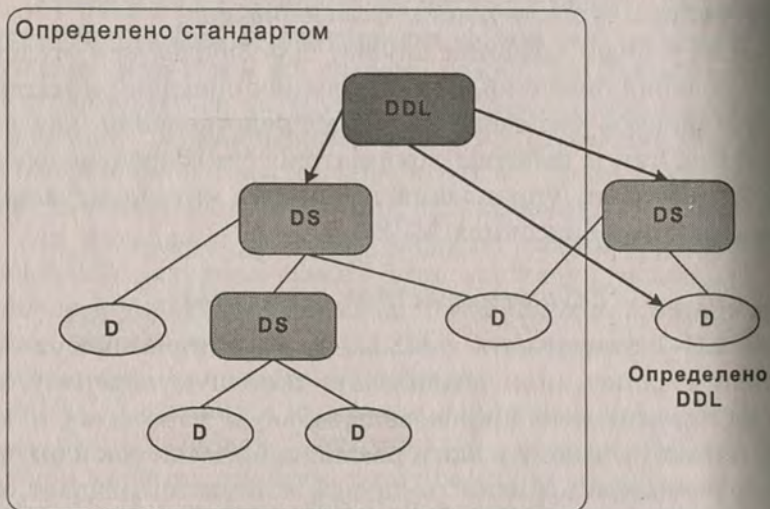


Рис. 5.2. Взаимодействие различных элементов MPEG-7

предоставляет механизм построения схемы описания, которая, в свою очередь, образует основу для формирования описания (см. также рис. 5.3).

Овалами обозначены средства, которые выполняют операции, такие как кодирование или декодирование, в то время как прямоугольниками отмечены статические элементы, такие как описания.

Главной задачей MPEG-7 будет предоставление новых решений для описания аудиовизуального материала. Таким образом, чисто текстовые документы не являются объектами MPEG-7. Однако аудиовизуальный материал может со-



Рис. 5.3. Абстрактное представление возможных приложений на основе MPEG-7

держат и сопряженный с ним текст. MPEG-7 будет, следовательно, рассматривать и поддерживать существующие решения, разработанные другими организациями стандартизации для текстовых документов.

Помимо самих дескрипторов на рабочие характеристики системы довольно сильно влияют DB-структуры. Чтобы быстро решить, представляет ли данный материал какой-то интерес, нужно структурировать индексную информацию, например, иерархическим или ассоциативным способом.

Область применения MPEG-7

Элементы, которые стандартизует MPEG-7, будут поддерживать широкий диапазон приложений (например, мультимедийные цифровые библиотеки, выбор широкоэвещательного медийного материала, мультимедийное редактирование, домашние устройства для развлечений и т. д.). MPEG-7 делает возможным мультимедийный поиск в WEB столь же простым, как и текстовый. Это станет применимо для огромных архивов, которые станут доступны для широкой публики, это придаст новый стимул для электронной торговли, так как

покупатели смогут искать нужный товар по видеообразцам. Информация, используемая для извлечения материала, может также применяться агентами для отбора и фильтрации широковещательного материала или целевой рекламы. Кроме того, описания MPEG-7 позволят быстрые и эффективные с точки зрения затрат полуавтоматические презентации и редактирование.

Все области применения, базирующиеся на мультимедиа, выиграют от использования MPEG-7. Ниже предлагается список возможных приложений MPEG-7, которые любой из читателей без труда сможет дополнить:

- ◆ архитектура, недвижимость и интерьерный дизайн (например, поиск идей);
- ◆ выбор широковещательного медийного канала (например, радио, TV);
- ◆ услуги в сфере культуры (исторические музеи, картинные галереи и т. д.);
- ◆ цифровые библиотеки (например, каталоги изображений, музыкальные словари, биомедицинские каталоги изображений, фильмы, видео- и радиоархивы);
- ◆ E-коммерция (например, целевая реклама, каталоги реального времени, каталоги электронных магазинов);
- ◆ образование (например, депозитарии мультимедийных курсов, мультимедийный поиск дополнительных материалов);
- ◆ домашние развлечения (например, системы управления личной мультимедийной коллекцией, включая манипуляцию содержимым, например, редактирование домашнего видео, поиск игр, караоке);
- ◆ исследовательские услуги (например, распознавание человеческих особенностей, экспертизы);
- ◆ журнализм (например, поиск речей определенного политика, используя его имя, его голос или его лицо);
- ◆ мультимедийные службы каталогов (например, Желтые страницы, туристская информация, географические информационные системы);

◆ мультимедийное редактирование (например, персональная электронная служба новостей, персональная медийная среда для творческой деятельности)

◆ удаленное опознавание (например, картография, экология, управление природными ресурсами);

◆ осуществление покупок (например, поиск одежды, которая вам нравится);

◆ надзор (например, управление движением, транспортом, неразрушающий контроль в агрессивной среде).

В принципе любой тип аудиовизуального материала может быть получен с помощью любой разновидности материала в запросе. Это означает, например, что видеоматериал может быть запрошен с помощью видео, музыки, голоса и т. д. Ниже приведены примеры запросов.

◆ Проиграйте несколько нот на клавиатуре и получите список музыкальных отрывков, сходных с проигранной мелодией, или изображений, соответствующим некоторым образом нотам, например, в эмоциональном плане.

◆ Нарисуйте несколько линий на экране и найдете набор изображений, содержащих похожие графические образы, логотипы, идеограммы...

◆ Определите объекты, включая цветовые пятна или текстуры, и получите образцы, среди которых вы выберете интересующие вас объекты.

◆ Опишите действия и получите список сценариев, содержащих эти действия.

◆ Используя фрагмент голоса Паваротти, получите список его записей, видеоклипов, где Паваротти поет, и графический материал, имеющий отношение к этому певцу.

Части MPEG-7

Стандарт MPEG-7 состоит из следующих частей:

1. Системы MPEG-7. Средства, которые необходимы при подготовке описаний MPEG-7 для эффективной передачи и записи, и для обеспечения синхронизации между материалом и описаниями. Эти средства имеют также отношение к охране интеллектуальной собственности.

2. Язык описания определений MPEG-7. Язык для определения новых схем описания и, возможно, новых дескрипторов.

3. MPEG-7 Audio — дескрипторы и схемы описания, имеющие отношение исключительно к описанию аудиоматериала.

4. MPEG-7 Visual — дескрипторы и схемы описания, имеющие отношение исключительно к описанию визуального материала

5. MPEG-7 Multimedia Description Schemes — дескрипторы и схемы описания, имеющие отношение к общим характеристикам описаний мультимедиа.

6. MPEG-7 Reference Software — программные реализации соответствующих частей стандарта MPEG-7.

7. MPEG-7 Conformance — базовые принципы и процедуры тестирования рабочих характеристик практических реализаций MPEG-7.

5.2. Главные функции MPEG-7

Системы MPEG-7

Системы MPEG-7 будут включать в себя средства, которые необходимы для подготовки описаний MPEG-7 для эффективной транспортировки и запоминания, а также позволяют синхронизировать мультимедийный материал и описания и средства, сопряженные с управлением и защитой интеллектуальной собственности. Стандарт определяет архитектуру терминала и нормативных интерфейсов.

Язык описания определений MPEG-7

Согласно определению в MPEG-7 язык описания определений DDL (Description Definition Language) представляет собой: "...язык, который позволяет формировать новые схемы описания и, возможно, дескрипторы. Он также позволяет расширение и модификацию существующих схем описания".

В качестве основы DDL был выбран язык XML. Как следствие, DDL может быть поделен на следующие логические нормативные компоненты:

- ◆ структурная схема языковых компонентов XML;
- ◆ компоненты типа данных схемы;
- ◆ специфические расширения MPEG-7.

Аудио MPEG-7

Окончательный проект аудио MPEG-7 представляет шесть технологий: система аудиоописаний (которая включает в себя дерево шкал и низкоуровневые дескрипторы), средства описания звуковых эффектов, средства описания тембра инструмента, описание голосового материала, сегмент молчания и дескрипторы мелодии, облегчающие обработку запросов.

Визуальный MPEG-7

Средства визуального описания MPEG-7, включенные в CD/ХМ состоят из базовых структур и дескрипторов, которые характеризуют следующие визуальные характеристики:

- ◆ цвет;
- ◆ текстура;
- ◆ форма;
- ◆ движение;
- ◆ локализация;
- ◆ прочие.

Каждая категория состоит из элементарных и сложных дескрипторов.

Основные объекты и схемы описания мультимедиа MPEG-7

Базисом схем описания мультимедиа MDS (Multimedia Description Schemes) является стандартизация набора средств описания (дескрипторы и схемы описания), имеющих дело с общими и мультимедийными объектами.

Общими объектами являются характеристики, которые используются в аудио, видео и текстовых описаниях и, следовательно, характеризуют все медийные типы материала. Такими характеристиками могут быть, например, вектор, время и т. д.

Помимо этого набора общих средств описания стандартизованы более сложные средства описания. Они используются, когда нужно описать более одного вида медийного материала (например, аудио и видео). Эти средства описания могут быть сгруппированы в 5 различных классов согласно их функциональному предназначению:

1. Описание материала: представление воспринимаемой информации.

2. Управление материалом: информация о характере медийного материала, формирование и использование АВ-материала.

3. Организация материала: представление анализа и классификации нескольких АВ-материалов.

4. Поиск и доступ: спецификация кратких характеристик и изменений АВ-материала.

5. Взаимодействие с пользователем: описание предпочтений пользователя и истории использования мультимедийного материала.

Эталонные программы MPEG-7: модель экспериментов (eXperimentation Model)

Программное обеспечение модели XM (eXperimentation Model) представляет собой систему моделирования для дескрипторов MPEG-7 (D), схем описания (DS), схем кодирования (CS), языка описания определений (DDL). Кроме нормативных компонентов, системе моделирования необходимы некоторые дополнительные элементы, существенные при исполнении некоторых процедурных программ. Структуры данных и процедурные программы образуют приложения. Приложения XM образуют две разновидности: приложения клиента и сервера.

5.3. Техническое описание стандарта MPEG-7

Системы MPEG-7 в настоящее время определяют архитектуру терминала и нормативных интерфейсов.

Архитектура терминала

Представление информации, специфицированное в стандарте MPEG-7, предоставляет средства описаний кодированного мультимедийного материала. Объект, который использует такое кодовое представление мультимедийного материала, называется терминалом. Этот терминал может соответствовать отдельно стоящему приложению или быть целой прикладной системой. Архитектура такого терминала изображена на рис. 5.4, а его работа описана ниже.

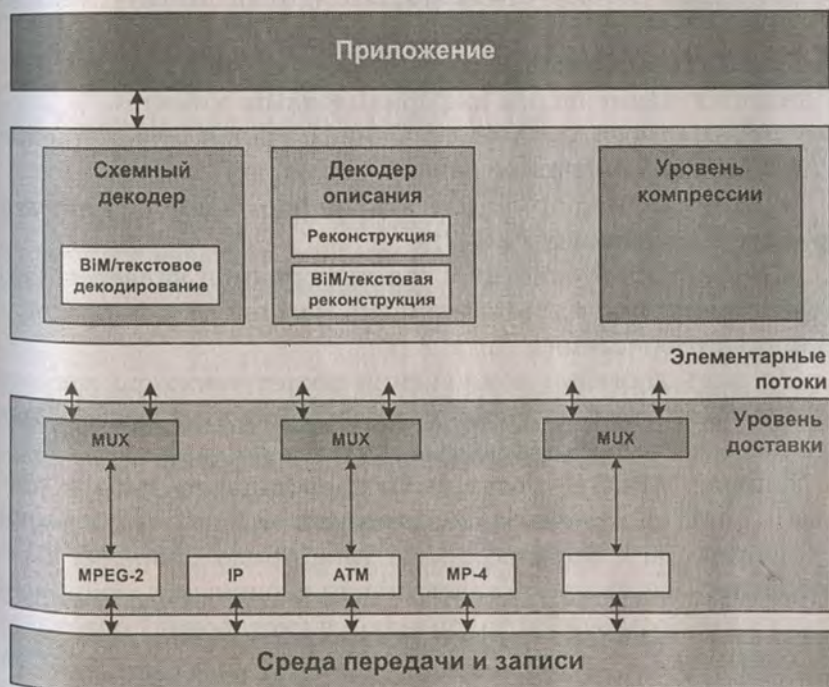


Рис. 5.4. Архитектура MPEG-7

В нижней части рис. 5.4 размещена система передачи/записи. Это относится к нижнему уровню инфраструктуры доставки (сетевой уровень и ниже). Эти уровни передают мультиплексированные потоки данных уровню доставки. Транспортная среда MPEG-7 базируется на многих системах доставки данных. Это включает, например, транспортные потоки MPEG-2, IP или MPEG-4 (MP4) файлы или потоки. Уровень доставки реализует механизмы, позволяющие выполнять синхронизацию, формирование кадров и мультиплексирование материала MPEG-7. Материал MPEG-7 может быть доставлен независимо или вместе с данными, которые он описывает. Архитектура MPEG-7 позволяет передавать данные (например, запросы) назад из терминала к отправителю или серверу.

Уровень доставки предоставляет уровню сжатия MPEG-7 элементарные потоки. Элементарные потоки MPEG-7 состоят из последовательности индивидуально доступных порций данных, называемых блоками доступа (Access Units). Блок доступа является наименьшим информационным объектом, к которому может относиться временная информация. Элементарные потоки MPEG-7 содержат данные различной природы:

- ◆ схемная информация: эта информация определяет структуру описания MPEG-7;
- ◆ информация описаний: эта информация является либо полным описанием мультимедийного материала или фрагментами такого описания.

Уровень доставки приложения может также по запросу доставлять мультимедийный материал. Для этих целей могут использоваться существующие средства доставки.

Данные MPEG-7 могут быть представлены либо в текстовом, либо в двоичном формате, или в виде комбинации этих форматов, в зависимости от типа приложения. MPEG-7 определяет однозначную связь между двоичным и текстовым форматами. Возможно установление двухсторонней однозначной связи между текстовым и двоичным представлениями. Следует заметить, что это не всегда доступно: некоторые приложения могут не захотеть передавать всю информацию.

содержащуюся в текстовом представлении, а могут предпочесть использовать более эффективную с точки зрения полосы двоичную кодировку с потерями.

Синтаксис текстуального формата определен в части 2 (DDL — Description Definition Language) стандарта. Синтаксис двоичного формата (BiM — двоичный формат для данных MPEG-7) определен в части 1 (системы) стандарта. Схемы определены в частях 3, 4 и 5 (визуальная, аудио и схемы описания мультимедиа) стандарта.

На уровне компрессии производится разборка потока блоков доступа (текстуальных или двоичных), а описания материала реконструируются. MPEG-7 не перепоручает реконструкцию текстуального представления в качестве промежуточного шага декодирующему процессу. Двоичный поток MPEG-7 может быть разобран с помощью BiM, передан в текстовом формате и затем в этом виде транспортирован для последующей реконструкционной обработки, или двоичный поток может быть разобран BiM и затем передан в подходящем формате для последующей обработки.

Блоки доступа MPEG-7 далее структурируются как команды, в которые инкапсулированы схемы описания. Команды придают материалу MPEG-7 динамический вид. Они позволяют пересылать описания одним куском или в виде небольших фрагментов. Команды делают возможными базовые операции с материалом MPEG-7, такие как актуализация дескриптора, удаление части описания или добавление новой структуры DDL. На реконструкционном этапе уровня компрессии выполняется актуализация описания и соответствующих схем посредством указанных команд.

Нормативные интерфейсы

MPEG-7 имеет два нормативных интерфейса, как это показано на рис. 5.5.

◆ **Материал:** это данные, которые должны быть представлены согласно формату, описанному в данной спецификации. Под материалом подразумеваются сами медийные данные либо их описание.



Рис. 5.5. Нормативные интерфейсы MPEG-7

♦ Двоичный/текстовый кодировщик MPEG-7: программа, осуществляющая преобразование материала к формату, который согласуется с данной спецификацией. Это может включать комплексное преобразование материала с целью извлечения деталей.

♦ Интерфейс текстового формата. Этот интерфейс описывает формат текстуальных блоков доступа. Текстовый декодер MPEG-7 воспринимает поток таких блоков доступа и реконструирует описание материала нормативным способом.

♦ Интерфейс двоичного формата. Этот интерфейс описывает формат двоичных блоков доступа. Двоичный декодер MPEG-7 воспринимает поток таких блоков доступа и реконструирует описание материала нормативным способом.

♦ Двоичный/текстовый декодер MPEG-7. Программа, осуществляющая преобразование материала к формату, который согласуется с данной спецификацией.

Верификация стандарта (рис. 5.6)

Кроме элементов нормативных интерфейсов, процесс валидации включает определение канонического представ-

ления описания материала. В каноническом пространстве, описания материала могут быть сравнены. Процесс валидации работает следующим образом:

1. Описание материала преобразуется в текстуальный и двоичный форматы без потерь, генерируя два разных представления одного и того же материала.
2. Два кодированных описания декодируются соответствующими двоичным и текстовым декодерами.
3. Из реконструированных описаний материала генерируются два канонических описания.
4. Два канонических описания должны быть эквивалентны.



Рис. 5.6. Процесс верификации

Описание канонической презентации XML-документа определено в Canonical XML [3].

Язык описания определений MPEG-7 (DDL)

Главными средствами, используемыми в описаниях MPEG-7, являются DDL (Description Definition Language), схемы описаний (DS) и дескрипторы (D). Дескрипторы связывают характеристики с набором их значений. Схемы описания являются моделями мультимедийных объектов и всего мно-

гообразия элементов, которые они представляют, например, модели данных описания. Они специфицируют типы дескрипторов, которые могут быть использованы в данном описании, и взаимоотношения между этими дескрипторами или между данными схемами описания.

DDL образует центральную часть стандарта MPEG-7. Она обеспечивает надежную описательную основу, с помощью которой пользователь может создать свои собственные схемы описания и дескрипторы. DDL определяет семантические правила выражения и комбинации схем описания и дескрипторов.

DDL не является языком моделирования, таким как UML (Unified Modeling Language), а языком схем для представления результатов моделирования аудиовизуальных данных, например DS и D.

DDL должен удовлетворять требованиям MPEG-7 DDL. Он должен быть способен выражать пространственные, временные, структурные и концептуальные взаимоотношения между элементами DS и между DS. Он должен предоставить универсальную модель для связей и ссылок между одним или более описаниями и данными, которые им описываются. Кроме того, язык не должен зависеть от платформы и приложения и быть читаемым как машиной, так и человеком. MPEG-7 должен базироваться на синтаксисе XML. Необходима также система разборки DDL (парсинга), которая должна быть способна проверять схемы описания (материал и структуру) и дескрипторы типа данных, как примитивные (целые, текст, дата, время), так и составные (гистограммы, нумерованные типы).

Разработка контекста

Так как схемный язык XML не был специально разработан для аудиовизуального материала, необходимы определенные расширения для того, чтобы удовлетворить всем требованиям MPEG-7 DDL.

Обзор схемы XML

Целью схемы является определение класса XML-документов путем использования конкретных конструкций, чтобы наложить определенные ограничения на их структуру: элементы и их содержимое, атрибуты и их значения, количество элементов и типы данных. Схемы можно рассматривать, как некоторые дополнительные ограничения на DTD.

Главной рекомендацией MPEG-7 АНГ было использование схемы, базирующейся на XML. В начале разработки имелось много решений, но ни одно из них не оказалось достаточно стабильным. В исходный момент группа DDL решила разработать свой собственный язык, следуя принципам, используемым группой W3C при подготовке схемы XML. В апреле 2000, рабочая группа W3C XML опубликовала последнюю версию спецификации схемы XML 1.0. Улучшенная стабильность схемного языка XML, его потенциально широкое поле применения, доступность средств и программ разборки, а также его способность удовлетворить большинству требований MPEG-7 привели к тому, что схема XML явилась основой DDL. Однако так как схема XML не была разработана специально для аудиовизуального материала, необходимы некоторые специфические расширения. DDL делится на следующие логические нормативные компоненты:

- ◆ схемные структурные компоненты XML;
- ◆ схемные компоненты типа данных XML;
- ◆ расширения для XML схемы MPEG-7.

Схема XML: Структуры

Схема XML: Структуры являются частью 2-частной спецификации XML-схемы. Она предоставляет средства для описания структуры и ограничений, налагаемых на материалы документов XML 1.0. Схема XML состоит из набора компонентов структурной схемы, которые могут быть разделены на три группы. Первичными компонентами являются:

- ◆ схема — внешний уровень определений и деклараций;
- ◆ определения простых типов;

- ◆ определения составных типов;
- ◆ декларации атрибутов;
- ◆ декларации элементов.

Вторичными компонентами являются:

- ◆ определения группы атрибутов;
- ◆ определения ограничений идентичности;
- ◆ определения группы;
- ◆ декларации нотации.

Третья группа образована компонентами “helper”, которые входят в другие компоненты и не могут существовать отдельно:

- ◆ аннотации;
- ◆ фрагменты (Particles);
- ◆ произвольные подстановки (Wildcards).

Определения типа задают внутренние компоненты схемы, которые могут использоваться в других компонентах, таких как элементы, атрибуты деклараций или другие определения типа. Схема XML предоставляет два вида компонентов определения типа:

- ◆ простые типы — являющиеся простыми типами данных (встроенными или вторичными), которые не могут иметь каких-либо дочерних элементов или атрибутов;
- ◆ составные типы — которые могут нести в себе атрибуты и иметь дочерние элементы или быть получены из других простых или составных типов.

Новые типы могут быть также определены на основе существующих типов (встроенных или вторичных) путем расширения базового типа. Детали использования этих компонентов можно найти в проекте DDL или в схеме XML: Спецификация структур.

Схема XML: Типы данных

XML Schema:Datatypes является второй частью 2-частной схемной спецификации XML. Она предлагает возможности определения типов данных, которые могут быть использованы для ограничения свойств типов данных элементов и ат-

рибутов в рамках схем XML. Она предлагает более высокую степень проверки типа, чем доступна для XML 1.0 DTD:

- ◆ набор встроенных примитивных типов данных;
- ◆ набор встроенных вторичных типов данных;
- ◆ механизмы, с помощью которых пользователи могут определить свой собственный вторичный тип данных.

Подробные детали встроенных типов данных и механизмы получения вторичных типов можно найти в окончательном проекте DDL или в спецификации XML Schema:Datatypes.

Расширения схемы XML MPEG-7

Для того чтобы удовлетворить специфическим требованиям MPEG-7, нужно добавить следующие характеристики к спецификации языка XML:

- ◆ массив и матрица типов — как фиксированного, так и параметризованного размеров;
- ◆ встроенные примитивные временные типы данных: `basicTimePoint` и `basicDuration`.

Программы разборки, специфические для MPEG-7, будут разработаны путем добавления валидации этих дополнительных конструкций к стандартным схемным разборщикам XML.

Аудио MPEG-7

Аудио MPEG-7 FCD включает в себя пять технологий: структура аудиоописания (которая включает в себя масштабируемые последовательности, дескрипторы нижнего уровня и равномерные сегменты тишины), средства описания тембра музыкального инструмента, средства распознавания звука, средства описания голосового материала и средства описания мелодии.

Описание системы аудио MPEG-7

Аудиоструктура содержит средства нижнего уровня, созданные для обеспечения основы для формирования аудио-

приложений высокого уровня. Предоставляя общую платформу структуры описаний, MPEG-7 Audio устанавливает базис для совместимости всех приложений, которые могут быть созданы в рамках данной системы.

Существует два способа описания аудиохарактеристик нижнего уровня. Один предполагает стробирование уровня сигнала на регулярной основе, другой может использовать сегменты (смотри описание MDS) для пометки сходных и отличных областей для заданного звукового отрывка. Обе эти возможности реализованы в двух типах дескрипторов нижнего уровня (один для скалярных величин, таких как мощность или частота, и один для векторов, таких как спектры), которые создают совместимый интерфейс. Любой дескриптор, воспринимающий эти типы, может быть проиллюстрирован примерами, описывающими сегмент одной результирующей величиной или последовательностью результатов стробирования, как этого требует приложение.

Величины, полученные в результате стробирования, сами могут подвергаться последующей обработке с привлечением другого унифицированного интерфейса: они могут образовать масштабируемые ряды (Scalable Series). Дерево шкал может также хранить различные сводные значения, такие как минимальное, максимальное значение дескриптора и его дисперсию.

Аудио дескрипторы нижнего уровня имеют особую важность при описании звука. Существует семнадцать временных и пространственных дескрипторов, которые могут использоваться в самых разных приложениях. Они могут быть грубо поделены на следующие группы:

- ◆ Базовая: мгновенные значения уровня волнового сигнала и мощности.
- ◆ Базовая спектральная: частотный спектр мощностей, спектральные характеристики, включая среднее значение, спектральная полоса и спектральная однородность.
- ◆ Параметры сигнала: фундаментальная частота квазипериодических сигналов и гармоничность сигналов.

- ◆ Временная группа по тембру: временной центроид.
- ◆ Спектральная группа по тембру: специфические спектральные характеристики в линейном пространстве частот, включая спектральный центроид и спектральные свойства, специфические для гармонических частей сигналов, включая спектральное смещение и спектральную ширину.
- ◆ Представления спектрального базиса: характеристики, используемые первично для распознавания звука.

Каждый из них может использоваться для описания сегмента с результирующим значением, которое применяется для всего сегмента или для последовательности результатов стробирования. Временная группа по тембру (Timbral Temporal) является исключением, так как ее значения приложимы только к сегменту, как целому.

В то время как аудиодескрипторы нижнего уровня вообще могут служить для многих возможных приложений, дескриптор однородности спектра поддерживает аппроксимацию сложных звуковых сигналов. Приложения включают в себя голосовую идентификацию.

Кроме того, очень простым, но полезным средством является дескриптор тишины. Он использует простую семантику "тишины" (т. е. отсутствие значимого звука) для аудиосегмента. Такой дескриптор может служить для целей дальнейшей сегментации аудиопотока.

Средства описания аудио верхнего уровня (D и DS)

Четыре набора средств описания аудио, которые приблизительно представляют области приложения, интегрированы в FCD: распознавание звука, тембр музыкального инструмента, разговорный материал и мелодическая линия.

Средства описания тембра музыкальных инструментов

Дескрипторы тембра служат для описания характеристик восприятия звуков. Тембр в настоящее время определен в литературе как характеристика восприятия, которая заставляет два звука, имеющих одну высоту и громкость, воспринимать-

ся по-разному. Целью средства описания тембра является представление этих характеристик восприятия сокращенным набором дескрипторов. Дескрипторы относятся к таким понятиям как "атака", "яркость" или "богатство" звука.

В рамках четырех возможных классов звуков музыкальных инструментов, два класса хорошо детализированы, и являются центральным объектом экспериментального исследования. В FCD представляются гармонические, когерентные непрерывные звуки и прерывистые, ударные звуки. Дескриптор тембра для непрерывных гармонических звуков объединяет спектральные дескрипторы тембра с временным дескриптором *log attack*. Дескриптор ударных инструментов комбинирует временные дескрипторы тембра с дескриптором спектрального центроида. Сравнение описаний, использующих один из наборов дескрипторов выполняется с привлечением метрики масштабируемого расстояния.

Средства распознавания звука

Схемы дескрипторов и описаний распознавания звука, представляют собой наборы средств для индексирования и категорирования звуков, с немедленным использованием для звуковых эффектов. Добавлена также поддержка автоматической идентификации звука и индексация. Это сделано для систематики звуковых классов и средств для спецификации онтологии устройств распознавания звука. Такие устройства могут использоваться для автоматической индексации сегментов звуковых треков.

Средства распознавания используют в качестве основы спектральные базисные дескрипторы низкого уровня. Эти базисные функции далее сегментируются и преобразуются в последовательность состояний, которые заключают в себя статистическую модель, такую как смешанная модель Маркова или Гаусса. Эта модель может зависеть от своего собственного представления, иметь метку, ассоциированную с семантикой исходного звука, и/или с другими моделями для того, чтобы категоризовать новые входные звуковые сигналы для системы распознавания.

Средства описания содержимого сказанного

Средства описания Spoken Content позволяют осуществить детальное описание произнесенных слов в пределах аудио-потoka. Учитывая тот факт, что сегодняшнее автоматическое распознавание речи ASR-технологий (Automatic Speech Recognition) имеет свои ограничения и что всегда можно столкнуться с высказыванием, которого нет в словаре, средства описания Spoken Content жертвует некоторой компактностью ради надежности поиска. Чтобы этого добиться, средства отображают выходной поток и то, что в норме может быть видно в качестве текущего результата автоматического распознавания речи ASR. Средства могут использоваться для двух широких классов сценария поиска: индексирование и выделение аудио потока, а также индексирование мультимедийных объектов аннотированных голосом.

Средства описания Spoken Content поделены на два широких функциональных блока: сетка, которая представляет декодирование, выполненное системой ASR, и заголовок, который содержит информацию об узнанных собеседниках и о самой системе распознавания. Сетка состоит из комбинаций слов голосовых записей для каждого собеседника в аудио потоке. Комбинируя эти сетки, можно облегчить проблему со словами, отсутствующими в словаре, и поиск может быть успешным, даже когда распознавание исходного слова невозможно.

Средства описания мелодии

DS мелодического очертания (Melody Contour) является компактным представлением информации о мелодии, которая позволяет эффективно и надежно контролировать мелодическую идентичность, например, в запросах с помощью наигрывания. DS мелодического очертания использует 5-ступенчатый контур (представляющий интервал между смежными нотами), в котором интервалы дискретизированы. DS мелодического очертания (Melody Contour DS) предоставляет также базовую информацию ритмики путем запоминания час-

тот, ближайших к каждой из нот, это может существенно увеличить точность проверки соответствия запросу.

Для приложений, требующих большей описательной точности или реконструкции заданной мелодии, DS мелодии поддерживает расширенный набор дескрипторов и высокую точность кодирования интервалов. Вместо привязки к одному из пяти уровней в точных измерителях используется существенно больше уровней между нотами (100 и более). Точная информация о ритмике получается путем кодирования логарифмического отношения разностей между началами нот способом аналогичным с используемым для кодирования уровней сигнала.

Визуальный MPEG-7

Средства визуального описания MPEG-7, включенные в CD/ХМ, состоят из базовых структур и дескрипторов, которые охватывают следующие основные визуальные характеристики:

- ◆ цвет;
- ◆ текстура;
- ◆ форма;
- ◆ движение;
- ◆ локализация;
- ◆ прочее.

Каждая категория состоит из элементарных и составных дескрипторов.

Базовые структуры

Существует пять визуально связанных базовых структур: сеточная выкладка, временные ряды (Time Series), многопрекционность (MultiView), пространственные 2D-координаты и временная интерполяция (TemporalInterpolation).

Сеточная выкладка

Сетка делит изображение на равные прямоугольные области, так что каждая область может быть описана отдельно.

Каждая область сетки описывается посредством других дескрипторов, таких как цвет или текстура. Более того, дескриптор позволяет ассоциировать субдескрипторы со всей прямоугольной областью или с произвольным набором прямоугольных областей.

Многовидовые 2D—3D

Дескриптор 2D/3D специфицирует структуру, которая комбинирует 2D дескрипторы, представляющие визуальные параметры 3D-объекта, видимые с различных точек. Дескриптор образует полное 3D-представление объекта на основе его проекций. Может использоваться любой визуальный 2D-дескриптор, такой как, например, форма контура, форма области, цвет или текстура. Дескриптор 2D/3D поддерживает интеграцию 2D-дескрипторов, используемых в плоскости изображения для описания характеристик 3D-объектов (реальный мир). Дескриптор позволяет осуществлять сравнение 3D-объектов путем сравнения их проекций.

Временные ряды

Этот дескриптор определяет в видеосегменте дескрипторы временных рядов и предоставляет возможность сравнения изображения с видеокadreом и видеокadreов друг с другом. Доступно два типа временных рядов (TimeSeries): `RegularTimeSeries` и `IrregularTimeSeries`. В первом дескрипторы размещаются регулярным образом (с постоянным шагом) в пределах заданного временного интервала. Это допускает простое представление для приложений, которые предполагают ограниченную сложность. Во втором дескрипторы размещаются нерегулярно (с переменными интервалами) в пределах заданного временного интервала. Это обеспечивает эффективное представление для приложений, которые требуют малой полосы пропускания или малой емкости памяти. Они полезны, в частности, для построения дескрипторов, которые содержат временные ряды дескрипторов.

Пространственные координаты 2D

Это описание определяет 2D пространственную координатную систему, которую следует использовать в других D/DS, где это важно. Оно поддерживает два вида координатных систем: “локальную” и “интегрированную” (рис. 5.7). В “локальной” координатной системе все изображения привязаны к одной точке. В “интегрированной” координатной системе каждое изображение (кадр) может быть привязано к разным областям. Интегрированная координатная система может использоваться для представления координат на мозаичном видео снимке.

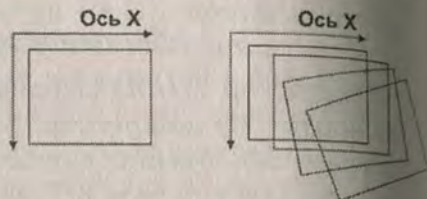


Рис. 5.7. “Локальная” и “интегрированная” координатная система

Временная интерполяция

Temporal Interpolation D описывает временную интерполяцию, использующую связанные многогранники. Это может использоваться для аппроксимации многомерных значений переменных, которые меняются со временем, такие как положение объекта в видео. Размер описания временной интерполяции обычно много меньше, чем описание всех величин. На рис. 5.8. 25 реальных величин представлены пятью линейными интерполяционными функциями и двумя квадратичными интерполяционными функциями. Начало временной интерполяции всегда привязывается ко времени 0.

Описатели цвета

Существует восемь дескрипторов цвета: цветового пространства, доминантных цветов, цветовой дискретизации, GoF/GoP цвета, цветовой структуры, цветового размещения и масштабируемой гистограммы цветов.



Рис. 5.8. Реальные данные и функции интерполяции

Цветовое пространство

Понятие цветового пространства используется в других описаниях, базирующихся на цвете. В текущем описании поддерживаются следующие цветовые пространства:

- ◆ R, G, B;
- ◆ Y, Cr, Cb;
- ◆ H, S, V;
- ◆ HMMD;
- ◆ матрица линейного преобразования с учетом R, G, B;
- ◆ монохромное.

Оцифровка цвета

Этот дескриптор определяет дискретизацию цветового пространства и поддерживает линейные и нелинейные преобразователи, а также lookup-таблицы. Число уровней квантования конфигурируемо так, чтобы обеспечить большую гибкость для широкого диапазона приложений. В случае нелинейного АЦП ширина канала преобразования может также конфигурироваться. Для разумных приложений в контексте MPEG-7 этот дескриптор должен комбинироваться с другими, например, чтобы характеризовать значения в цветовой гистограмме.

Доминантный цвет

Этот дескриптор цвета является наиболее удобным для представления локальных характеристик (области объекта или

изображения), где для предоставления цветовой информации достаточно малого числа цветов. Могут использоваться и полные изображения, например, картинки флагов или цветных торговых марок. Квантование цвета используется для получения малого числа характерных цветов в каждой области/изображении. Соответственно вычисляется процент каждого дискретизируемого цвета в области. Определяется также пространственная когерентность всего дескриптора.

Масштабируемый цвет

Дескриптор масштабируемого цвета (Scalable Color) является гистограммой цветов в цветном пространстве HSV, которая кодируется с помощью преобразования Хара. Ее двоичное представление является масштабируемым с точки зрения числа каналов и числа бит, характеризующих значение точности в широком диапазоне потоков данных. Дескриптор масштабируемого цвета полезен для сравнения изображений и поиска, базирующегося на цветовых характеристиках. Точность отображения возрастает с увеличением числа бит, используемых для описания.

Описатель структуры цвета

Дескриптор цветовая структура (Color Structure) является описателем цветовой характеристики, которая объединяет цветовое содержимое (аналогично цветовой гистограмме) и информацию о структуре материала. Его главная задача сравнение изображений главным образом для статических картинок. Метод выборки вводит данные о цветовой структуре в дескриптор, учитывая локально цвета окрестных пикселей, и не анализирует каждый пиксель отдельно. Дескриптор цветовая структура обеспечивает дополнительную функциональность и улучшенный поиск, базирующийся на подобию естественных изображений.

Выкладка цвета

Этот дескриптор специфицирует пространственное распределение цветов для быстрого поиска и просмотра. Его

целью является не только сравнение изображений и видеоклипов, но также поиск, базирующийся на раскладке цветов, такой как сравнение наброска с изображением, которое не поддерживается другими цветовыми дескрипторами. Этот дескриптор может использоваться для всего изображения или для любой его части. Данный дескриптор может также быть применен для областей произвольной формы.

Цвет GoF/GoP

Дескриптор цвета группа_кадров/группа_картинок расширяет возможности дескриптора масштабируемого цвета, который определен для статических изображений, чтобы выполнять цветовое описание видеосегментов или собрания статических изображений. Дополнительные два бита позволяют определить, была ли вычислена цветовая гистограмма, прежде чем было осуществлено преобразование Хара: для усреднения, медианы или пересечения. Усредненная гистограмма, которая соответствует усредненному значению счетчика для каждой ячейки всех кадров или изображений, эквивалентна вычислению совокупной цветовой гистограммы всех кадров или изображений с последующей нормализацией. Медианная гистограмма соответствует вычислению медианного значения счетчика для каждой ячейки совокупности кадров или изображений. Более надежно округлять ошибки и присутствие выбросов в распределении яркости изображения по сравнению с усредненной гистограммой. Гистограмма пересечения соответствует вычислению минимального значения счетчика для каждой ячейки совокупности кадров или изображений, чтобы получить цветовые характеристики "наименьшего общего" группы изображений. Заметим, что это отличается от гистограммы пересечения, которая является скалярной мерой. Аналогичные меры сходства/различия, которые используются для сравнения масштабируемых цветовых описаний, могут быть применены для сопоставления цветовых дескрипторов GoF/GoP.

Описатели текстуры

Существует три текстурных дескриптора: Edge Histogram, Homogeneous Texture и Texture Browsing.

Дескрипторы однородной текстуры

Однородная текстура представляет собой важный визуальный примитив для поиска и просмотра большой коллекции выглядящих сходно образов. Изображение может рассматриваться как мозаика однородных текстур, так что эти текстурные характеристики, соответствующие областям могут использоваться для индексации визуальных данных. Например, пользователь, просматривающий абстрактную базу данных изображений, может захотеть идентифицировать различные блоки в этой коллекции изображений. Блоки с автомашинами, запаркованными регулярным образом, являются хорошим примером однородного текстурного образца, рассматриваемого с большого расстояния, как это происходит при аэро съемке. Аналогично сельскохозяйственные области и участки растительности являются другим примером однородных текстур, встречающихся при аэро и спутниковых наблюдениях. Примеры запросов, которые могут поддерживаться в этом контексте, могут включать в себя "Поиск всех спутниковых изображений Санта-Барбары, которые имеют меньше чем 20% облачного покрытия" или "Найти растительный участок, который выглядит как эта область". Чтобы поддерживать такой поиск изображений, необходимо эффективное представление текстуры. Дескриптор однородной текстуры предоставляет количественное представление, используя 62 числа (по 8 бит каждое), которое удобно для поиска сходства. Получение данных осуществляется следующим образом; изображение сначала обрабатывается посредством набора фильтров Габора, настроенных на определенные ориентации и масштаб (смоделированные с помощью функций Габора). Дескриптор однородной текстуры предоставляет точное количественное описание текстуры, которое может ис-

пользоваться для поиска. Вычисление этого дескриптора базируется на фильтрации.

Просмотр текстуры

Дескриптор просмотра текстуры (Texture Browsing) полезен для представления однородной текстуры в приложениях, служащих для просмотра, и требует только 12 бит (максимум). Он предоставляет перцептуальную характеристику текстуры, аналогично человеческому описанию в терминах регулярности, шероховатости, ориентированности. Вычисление этого дескриптора осуществляется так же, как и дескриптора однородной текстуры. Сначала изображение фильтруется с помощью набора специально настроенных фильтров (смоделированных посредством функций Габо́ра); в отфильтрованном результате идентифицируются два доминантных ориентаций текстуры. Три бита используются для представления каждой из доминантных ориентаций. За этим следует анализ проекций отфильтрованного изображения вдоль доминантных направлений, чтобы определить регулярность (характеризуемую двумя битами) и закругленность (2 бита \times 2). Этот дескриптор совместно с дескриптором однородной текстуры предоставляет масштабируемое решение для представления областей изображения с однородной текстурой.

Краевая гистограмма

Дескриптор краевой гистограммы представляет пространственное распределение пяти типов краев, в частности, четырех ориентированных краев и одного неориентированного. Так как края играют важную роль для восприятия изображения, данный дескриптор помогает найти изображения со сходным семантическим значением. Таким образом, он изначально ориентирован на сравнение изображений (по образцам или наброскам), в особенности на естественные изображения с нерегулярными краями. В этом контексте свойства системы поиска изображения могут быть существенно улучшены, если дескриптор краевой гистограммы комбинируется с другими дескрипторами, такими как дескриптор цветовой гистограммы.

мы. Кроме того, наилучшие характеристики системы поиска изображения, учитывая только этот дескриптор, достигаются путем использования полуглобальных и глобальных гистограмм, получаемых непосредственно из дескриптора крайних гистограмм.

Описатели формы

Существует четыре типа дескрипторов формы: объектная форма, базирующаяся на областях, форма, базирующаяся на контурах, 3D-форма и 2D—3D множественные проекции.

Форма, базирующаяся на областях

Форма объекта может состоять из одной области или набора областей, а также некоторых отверстий в объектах, как это показано на рис. 5.9. Так как дескриптор формы, базирующийся на областях, использует все пиксели, определяющие форму в пределах кадра, он может описывать любую форму, т. е. не только простые формы с односвязными областями, как на рис. 5.9а и б), но также сложные формы, которые содержат отверстия или несколько несоединенных областей, как показано на рис. 5.9с, d и e), соответственно. Дескриптор формы, базирующийся на областях, может не только эффективно описать столь несхожие формы, но и минимизировать искажения на границах объекта.

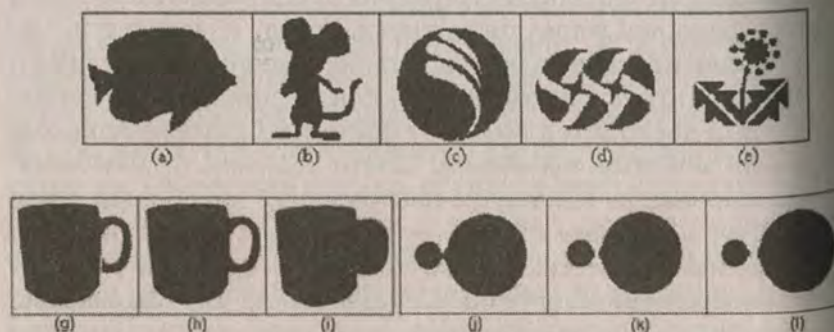


Рис. 5.9. Примеры различной формы

На рис. 5.9 *g*, *h* и *i* показаны очень схожие изображения чашки.

Различия имеются только в форме ручки. Форма (*g*) имеет трещину на нижней части ручки, в то время как на (*i*) ручка не имеет отверстия. Deskriptor формы, базирующейся на областях, рассматривает (*g*) и (*h*) подобными, но отличными от (*i*), так как там ручка не имеет отверстия. Аналогично на рис. 5.9 *j*, *k*, *l* показана часть видео последовательности, где два диска постепенно разделяются. С точки зрения deskriptора формы, базирующейся на областях, эти картинки схожи.

Заметим, что черный пиксель в пределах объекта соответствует 1 на изображении, в то время как пиксели белого фона соответствуют 0.

Deskriptor характеризуется малым размером и быстрым временем поиска. Размер данных для представления является фиксированным и равным 17.5 байт.

Форма, основанная на контуре

Deskriptor формы, базирующейся на контуре, получает параметры формы объекта или его контур, извлеченный из описания областей. Он использует так называемое Curvature Scale-Space представление, которое воспринимает значимые параметры формы.

Deskriptor формы, базирующейся на контуре объекта, использует Curvature Scale Space представление контура. Это представление имеет несколько важных особенностей, в частности:

- ◆ извлекает очень хорошие характеристики формы, делая возможным поиск, основанный на сходстве;
- ◆ отражает свойства восприятия визуальной системы человека и предлагает хорошее обобщение;
- ◆ устойчиво при плавном движении;
- ◆ устойчиво при частичном перекрытии формы;
- ◆ устойчиво по отношению преобразованиям перспективы, которые являются следствием изменения параметров видеокамеры, и представляются общими для изображений и видео;
- ◆ компактно.

Некоторые из вышеперечисленных свойств проиллюстрированы на рис. 5.10, каждый кадр содержит весьма сходные с точки зрения CSS изображения, основанные на результате действительного поиска в базе данных MPEG-7.

На рис. 5.10а продемонстрированы свойства обобщения формы (внешнее сходство различных форм), б — устойчивость по отношению к плавному движению (бегущий человек), в — устойчивость к частичному перекрытию (хвосты или ноги лошадей).

Рассматривая непрерывное развитие мультимедийных

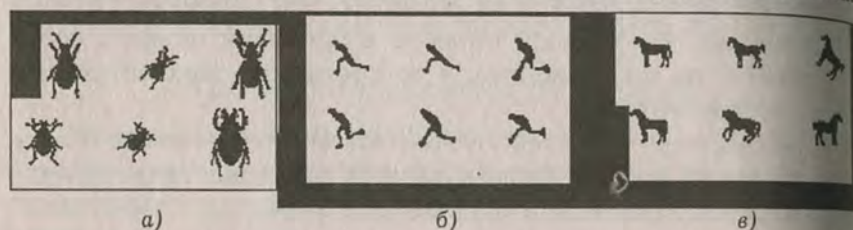


Рис. 5.10. Пример свойства обобщения формы 3D-форма

технологий, виртуальных миров, 3D-материал становится обычным для современных информационных систем. В большинстве случаев, 3D-информация представляется в виде сетки многоугольников. Группа MPEG-4, в рамках подгруппы SNHC, разрабатывала технологии для эффективного кодирования модели 3D-сеток. В стандарте MPEG-7 необходимы средства для интеллектуального доступа к 3D-информации. Главные приложения MPEG-7 имеют целью поиск, получение и просмотр баз 3D-данных.

Предлагаемый дескриптор 3D-формы имеет целью предоставление внутреннего описания формы сеточных 3D-моделей. Он использует некоторые локальные атрибуты 3D-поверхности.

Дескрипторы перемещения

Существует четыре дескриптора перемещения: перемещение камеры, траектория перемещение объекта, параметрическое движение объекта и двигательная активность.

Движение камеры

Этот дескриптор характеризует параметры перемещения 3D-камеры. Он базируется на информационных параметрах 3D-перемещения камеры, которые могут быть автоматически получены.

Дескриптор движения камеры поддерживает следующие стандартные операции с камерой (см. рис. 5.11): фиксированное положение, панорамное движение (горизонтальное вращение), слежение за движущимся объектом (горизонтальное поперечное перемещение), вертикальное вращение, вертикальное поперечное перемещение, изменение фокусного расстояния, наезд (трансфокация вдоль оптической оси) и вращение вокруг оптической оси. Отрывок, для которого все кадры характеризуются определенным типом перемещения камеры, относящееся к одному виду или нескольким, определяет базовые модули для дескриптора перемещения камеры. Каждый составляющий блок описывает начальный момент, длительность, скорость перемещения изображения и увеличение фокусного расстояния (FOE) (или сокращение фокусного расстояния — FOC).

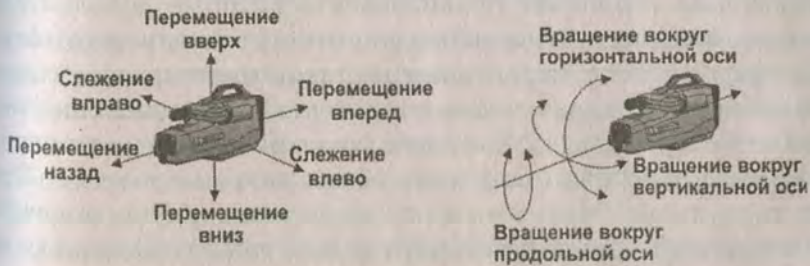


Рис. 5.11. Перемещения камеры

Дескриптор представляет объединение этих составляющих блоков, он имеет опцию описания смеси типов перемещения камеры. Смешанный режим воспринимает глобальную информацию о параметрах перемещения камеры, игнорируя детальные временные данные, путем совместного описания

нескольких типов движения, даже если эти типы перемещения осуществляются одновременно. С другой стороны, несмещанный режим воспринимает понятие чистых перемещений и их совмещения на протяжении определенного временного интервала. Ситуация, когда одновременно реализуется несколько типов перемещений, описывается как суперпозиция описаний чистых независимых типов перемещения. В этом режиме описания временное окно конкретного элементарного сегмента может перекрываться с временным окном другого элементарного сегмента.

Траектория движения

Траектория движения объекта является простой характеристикой высокого уровня, определяемая как позиция во времени и пространстве одной репрезентативной точки этого объекта.

Этот дескриптор полезен для поиска материала в объектно-ориентированных визуальных базах данных. Он также эффективен в большинстве специальных приложений. В данном контексте с предварительным знанием ряда параметров, траектория позволяет реализовать некоторые дополнительные возможности. При наблюдении могут выдаваться сигналы тревоги, если траектория воспринимается, как опасная (например, проходит через запретную зону, движение необычно быстро, и т. д.). В спорте могут распознаваться специфические действия (например, обмен ударами у сетки). Кроме того, такое описание позволяет также улучшить обработку данных: для полуавтоматического редактирования медиаданных, траектория может быть растянута, смещена и т. д., чтобы адаптировать перемещения объекта для любого контекста.

Дескриптор является списком ключевых точек (x, y, z, t) вместе с набором опционных интерполирующих функций, которые описывают путь объекта между ключевыми точками, в терминах ускорения. Скорость неявно известна с помощью спецификации ключевых точек. Ключевые точки специфициру-

ются путем задания моментов времени или их 2-D или 3-D декартовых координат, в зависимости от приложения. Интерполирующие функции определены для каждого компонента $x(t)$, $y(t)$ и $z(t)$ независимо. Некоторые свойства этого представления перечислены ниже:

- ◆ не зависит от пространственно-временного разрешения материала (например, 24 Hz, 30 Hz, 50 Hz, CIF, SIF, SD, HD, и т. д.), т. е. если материал существует во многих форматах одновременно, для описания траектории объекта необходим только один набор дескрипторов данного материала;
- ◆ компактно и масштабируемое. Вместо запоминания координаты объекта для каждого кадра, гранулярность дескриптора выбирается на основе ряда ключевых точек, используемых для каждого из временных интервалов;
- ◆ непосредственно допускает широкое разнообразие применений, типа поиска подобия, или категорирование по скорости (быстрые, медленные объекты), поведению (ускоряется, когда приближается к этой области) или по другим характеристикам движения высокого уровня.

Параметрическое движение

Модели параметрического движения были использованы в рамках различных схем анализа и обработки изображения, включая сегментацию перемещения, оценки глобального перемещения, и отслеживание объектов. Модели параметрического перемещения использовались уже в MPEG-4 для оценки перемещения и компенсации. В контексте MPEG-7 перемещение является крайне важной характеристикой, связанный с пространственно-временной структурой видео, относящейся к нескольким специфическим MPEG-7 приложениям, таким как запоминание и поиск в видео базах данных, и для целей анализа гиперсвязей. Движение является также критической характеристикой для некоторых специфических приложений, которые уже рассматривались в рамках MPEG-7.

Базовый принцип состоит из описаний движения объектов в видеопоследовательности, например, в параметричес-

кой 2D-модели. В частности, аффинные модели включают в себя трансляции, вращения, масштабирование и их комбинации, планарные модели перспективы делают возможным учет глобальных деформаций, сопряженных с перспективными проекциями, а квадратичные модели позволяют описать более сложные движения.

Параметрическая модель ассоциирована с произвольными фоновыми объектами или объектами переднего плана, определенными как области (группа пикселей) в изображении в пределах заданного интервала времени. Таким способом движение объекта записывается компактным образом в виде набора из нескольких параметров. Такой подход ведет к очень эффективному описанию нескольких типов перемещения, включая простые преобразования, вращения и изменения масштаба, или более сложные перемещения, такие как комбинации перечисленных выше элементарных перемещений.

Определение подобия характеристик моделей движения является обязательным для эффективного поиска объектов. Оно также необходимо для поддержки запросов нижнего уровня, полезно и в запросах верхнего уровня, таких как "поиск объектов приближающихся к камере", или для "объектов, описывающих вращательное движение", или "поиск объектов, перемещающихся влево" и т. д.

Двигательная активность

Просмотр человеком видео или анимационной последовательности воспринимается как медленная последовательность, быстро протекающий процесс, последовательность действий и т. д. Deskriptor активности воспринимает интуитивное понятие "интенсивность действия" или "темп действий" в видео сегменте. Примеры высокой "активности" включают такие сцены, как "ведение счета голов в футбольном матче", "автомобильные гонки" и т. д. С другой стороны сцены, типа "чтение новостей", "интервью", "снимок" и т. д. воспринимаются как кадры низкой активности. Видеоматериал охватывает диапазон от низкой до высокой активности, следовательно

но, нам нужен дескриптор, который позволяет нам точно выразить активность данной видеопоследовательности/снимка и всесторонне перекрывать упомянутый выше диапазон. Дескриптор активности полезен для приложений, таких как видеонаблюдение, быстрый просмотр, динамическое видеорежимирование, информационные запросы и т. д. Например, мы можем замедлить темп презентации кадров, если дескриптор активности указывает на высокую активность, так чтобы облегчить просмотр этой активности. Другим примером приложения является нахождение всех кадров высокой активности в новой видеопрограмме, которая может рассматриваться как просмотр, так и абстракцию.

Локализация. Локатор области

Этот дескриптор допускает локализацию областей внутри изображения или кадров путем спецификации их с помощью краткого и масштабируемого отображения боксов или многогранников.

Пространственно-временной локатор

Локатор описывает пространственно-временные области в видео последовательности, такой как области движущихся объектов, и обеспечивает функцию локализации. Главным его применением является гипермедиа, где выделенная точка находится внутри объекта. Другим ведущим применением является поиск объектов путем проверки, прошел ли объект

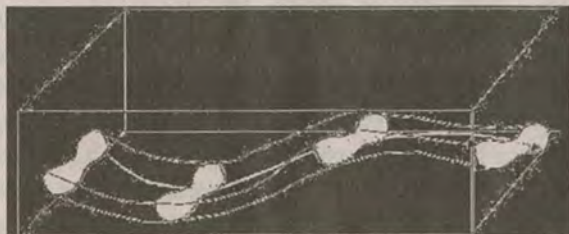


Рис. 5.12. Пространственно-временная область

определенные точки. Это может использоваться для наблюдения. Deskriptor SpatioTemporalLocator может описывать как связанные, так и несвязанные области.

Распознавание лица

Deskriptor FaceRecognition может использоваться для получения изображения лиц, которые соответствуют запросу. Deskriptor представляет проекцию вектора лица на набор базовых векторов, которые охватывают пространство возможных векторов лица. Набор параметров FaceRecognition получается из нормализованного изображения лица. Это нормализованное изображения лица содержит 56 строк с 46 значениями уровня в каждой строке. Центры двух глаз на каждом изображении лица размещаются на 24-ом ряду и 16-й и 31-й колонке для правого и левого глаз соответственно. Это нормализованное изображение затем используется для получения одномерного вектора лица, который состоит из значений яркости пикселей нормализованного изображения лица, которое получается в результате растрового сканирования, начинающегося в верхнем левом углу и завершающегося в нижнем правом углу изображения. Набор параметров FaceRecogniton вычисляется путем проектирования одномерного вектора лица на пространство, определяемое набором базисных векторов.

Схемы описания мультимедиа MPEG-7

Deskriptоры MPEG-7 сконструированы для описания следующих типов информации: низкоуровневые аудиовизуальные характеристики, такие как цвет, текстура, движение, уровень звука и т. д.; высокоуровневые семантические объекты, события и абстрактные принципы; процессы управления материалом; информация о системе памяти и т. д. Ожидается, что большинство deskriptоров, соответствующих низкоуровневым характеристикам, будут извлекаться автоматически, в то время как человеческое вмешательство будет необходимо для формирования высокоуровневых deskriptоров.

MPEG-7 DS преобразуются в дескрипторы путем комбинирования индивидуальных дескрипторов а также других DS в рамках более сложных структур и определения соотношения составляющих дескрипторов и DS. В MPEG-7 DS категоризируются в отношении к аудио или видеообластям или по отношению к описанию мультимедиа. Например, характерные DS соответствуют неизменным метаданным, связанным с формированием, производством, использованием и управлением мультимедиа, а также описанием материала. Обычно мультимедийные DS относятся ко всем типам мультимедиа, в частности, к аудио, видео и текстовым данным, в то время как специфичные для области дескрипторы, такие как цвет, текстура, форма, мелодия и т. д., относятся исключительно к аудио или видео областям. Как в случае дескрипторов, реализация DS может в некоторых вариантах базироваться на автоматических средствах, но часто требует вмешательства человека.

Средства организации MDS

На рис. 5.13 представлена схема организации мультимедийных DS MPEG-7 в следующих областях: базовые элементы, описание материала, управление материалом, организация материала, навигация и доступ, взаимодействие с пользователем.

Базовые элементы

Спецификация мультимедийных DS MPEG-7 определяет определенное число схемных средств, которые облегчают формирование и выкладку описаний MPEG-7. Схемные средства состоят из корневого элемента, элементов верхнего уровня и средств выкладки (Package Tools). Корневые элементы, которые являются начальными элементами описания.

MPEG-7 позволяют сформировать полные XML-документы и фрагменты описания MPEG-7. Элементы верхнего уровня, которые позволяют корневым элементам в описании MPEG-7 организовать DS для объектно ориентированных за-



Рис. 5.13. Обзор мультимедийных DS MPEG-7

дач описания, таких как описание изображения, видео, аудио или аудиовизуальный материал, собрания (коллекции), пользователи или семантики мира. Созданы пакетные средства для группирования или ассоциации связанных компонентов DS описаний в каталоги или пакеты. Пакеты полезны для организационных и передающих структур и типов описательной информации MPEG-7 для систем поиска и для помощи при просмотре пользователям, незнакомым с особенностями описаний MPEG-7.

Спецификация мультимедийных DS MPEG-7 определяет также некоторое число базовых элементов, которые используются повторно в качестве фундаментальной конструкции при определении MPEG-7 DS. Многие базовые элементы предоставляют специфические типы данных и математические структуры, такие как вектора и матрицы, которые важны для описания аудиовизуального материала. Они включаются также в качестве элементов для связи медиафайлов и локализации сегментов, областей и т. д. Многие базовые элементы предназначены для специальных нужд описания аудиовизуального материала, таких как описание времени, мест,

людей, индивидуальностей, групп, организаций, и других текстовых аннотаций. Из-за их важности для описания аудиовизуального материала давайте очертим подходы MPEG-7 к описанию временной информации и текстовых аннотаций.

◆ **Временная информация:** DS для описания времени базируется на стандарте ISO 8601, который был воспринят схемным языком XML. Временные DS предоставляют временную информацию в медиапоток и для реального мира. MPEG-7 расширяет спецификацию времени ISO 8601 для того, чтобы описать время в терминах стробирования аудиовизуального материала, например, путем подсчета периодов стробирования. Это позволяет поддерживать эффективное описание временной информации в больших массивах аудиовизуального материала.

◆ **Текстовая аннотация:** текстовая аннотация является также важным компонентом многих DS. MPEG-7 предоставляет некоторое число базовых конструкций для текстового аннотирования, включая свободный текст (слова, фразы), структурированный текст (текст плюс назначение слов) и зависимая структурированная аннотация (структурированный текст плюс взаимные связи), для того чтобы поддерживать широкий диапазон функций текстовых описаний.

Управление содержимым

MPEG-7 предоставляет также DS для управления материалом. Эти элементы описывают различные аспекты создания медиаматериала, медиакодирование, запись, форматы файлов и использование материала. Функциональность каждого из этих классов DS представлена ниже [5].

◆ **Создание информации:** описывает формирование аудиовизуального материала. Эта информация описывает создание и классификацию аудиовизуального материала и других данных, которые с ним связаны. Информация формирования выдает заголовок (который может быть текстовым или фрагментом аудиовизуального материала), текстовую аннотацию, а также данные о создателях, месте формирования и дате.

Классификационная информация описывает, как аудиовизуальный материал классифицируется в таких категориях, как жанр, тема, цель, язык и т. д. Она предоставляет также обзор и управляющую информацию, такую как классификация по возрасту, тематический обзор, рекомендации создателей и т. д. Наконец, информация, сопряженная с материалом, описывает, существует ли другой материал, который связан тематически с данным материалом.

◆ Использование информации: описывает информацию об использовании аудиовизуального материала, такую как права использования, доступность, записи об использовании и финансовая информация. Правовая информация не включается в описание MPEG-7, вместо этого предлагаются ссылки на владельцев прав и другие данные, относящиеся к защите авторских прав. Правовые DS предоставляют эти ссылки в форме уникальных идентификаторов, которые управляются извне. Базовая стратегия описаний MPEG-7 заключается в предоставлении доступа к текущей информации о владельце без возможности непосредственного обсуждения возможных условий доступа к самому материалу. DS доступности и DS записей об использовании предоставляют данные, относящиеся, соответственно к доступности и прошлому использованию материала, такому как широковещательная демонстрация, доставка по требованию, продажа CD и т. д. Наконец, финансовые DS предоставляют информацию, связанную со стоимостью производства и доходами, которые могут стать результатом использования материала. Информация использования является обычно динамической, меняющейся за время жизни аудиовизуального материала.

◆ Медиаописание: характеризует характер записи, например, сжатие данных, кодирование и формат записи аудиовизуального материала. DS медиаинформации идентифицирует источник материала. Образцы аудиовизуального материала называются медиапрофайлами, которые являются версиями исходного материала, полученными, возможно, посредством другого кодирования или записи в другом формате.

Каждый медиапрофайл описывается индивидуально в терминах параметров кодирования и положения.

Описание содержимого

MPEG-7 предоставляет также DS для описания материала. Эти элементы описывают структуру (области, видеокадры и аудиосегменты) и семантику (объекты, события, абстрактные понятия). Функциональность каждого из классов DS представлена ниже:

- ◆ Структурные аспекты. DS описывает аудиовизуальный материал с точки зрения его структуры. Структурные DS формируются на основе DS сегментов, которые представляют пространственную, временную или пространственно-временную структуру аудиовизуального материала. Для получения оглавления или индекса для поиска аудиовизуального материала DS сегменты могут быть организованы в иерархические структуры. Сегменты могут быть описаны на основе характеристик восприятия с помощью дескрипторов MPEG-7 для цвета, текстуры, формы, движения, аудиопараметров и т. д.

- ◆ Концептуальные аспекты. DS описывает аудиовизуальный материал с точки зрения семантики реального мира и концептуальных представлений. DS семантики включают в себя такие характеристики, как объекты, события, абстрактные концепции и отношения. DS структуры и DS семантики имеют отношение к набору связей, который позволяет описать аудиовизуальный материал на основе его структуры и семантики.

Навигация и доступ

MPEG-7 предоставляет также DS для облегчения просмотра и извлечения аудиовизуального материала путем определения резюме, разделов, составных частей и вариантов аудиовизуального материала.

- ◆ Резюме предоставляет компактное описание аудиовизуального материала, которое призвано облегчить поиск,

просмотр, визуализацию и прослушивание аудиовизуального материала. DS резюме содержат два типа режимов навигации: иерархический и последовательный. В иерархическом режиме информация организована в виде последовательности уровней, каждый из которых описывает аудиовизуальный материал с разной степенью детализации. Вообще, уровни более близкие к корневому предоставляют более общие резюме, периферийные же уровни повествуют о тонких деталях. Последовательные резюме предоставляют последовательность изображений или видеок кадров, возможно синхронизованных со звуком, которые могут служить для просмотра слайдов, или аудиовизуальный набросок.

◆ Разделы и декомпозиции описывают различные составляющие аудиовизуального сигнала в пространстве, времени и частоте. Разделы и декомпозиции могут использоваться для описания различных проекций аудиовизуальных данных, которые важны для доступа с разным разрешением.

◆ Вариации предоставляют информацию о различных вариантах аудиовизуальных программ, таких как резюме и аннотации; масштабируемые, сжатые версии и варианты с низким разрешением; а также версии на различных языках — звук, видео, изображение, текст и т. д. Одной из важных возможностей, обеспечиваемых DS вариации, является выбор наиболее удобной версии аудиовизуальной программы, которая может заменить оригинал, если необходимо, адаптироваться к различным возможностям терминального оборудования, сетевым условиям или предпочтениям пользователя.

Организация содержимого

MPEG-7 предоставляет также DS для организации и моделирования собрания аудиовизуального материала, а также его описания. DS собрания организует коллекцию аудиовизуального материала, сегментов, событий, и/или объектов. Это позволяет описать каждое собрание как целое на основе общих характеристик. В частности, для описания значений атрибутов собрания могут быть специфицированы различные модели и статистики.

Интеракция с пользователем

Наконец, последний набор DS MPEG-7 имеет отношение к взаимодействию с пользователем. DS взаимодействия с пользователем описывает предпочтения пользователя и историю использования мультимедийного материала. Это позволяет, например, найти соответствие между предпочтениями пользователя и описаниями аудиовизуального материала, для того чтобы облегчить индивидуальный доступ к аудиовизуальному материалу, презентации и пр.

Управление содержимым

Средства управления описанием материала позволяют охарактеризовать жизненный цикл материала.

Материал, охарактеризованный описаниями MPEG-7, может быть доступным в различных форматах и режимах, с разными схемами кодирования. Например, концерт может быть записан в двух разных режимах: звуковом и аудиовизуальном. Каждый из этих режимов может использовать различное кодирование. Это создает несколько медиапрофайлов. Наконец, могут быть получены несколько копий одного и того же материала. Эти принципы режимов и профайлов проиллюстрированы на рис 5.14.

◆ **Материал.** Реальное событие, такое как концерт, может быть представлено различными типами медиаматериала, например, звуковой материал, аудиовизуальный материал. Материал является объектом, который имеет специфическую структуру для отображения реальности.

◆ **Медиаинформация.** Физический формат материала описывается DS медиа информации. Одна копия описания DS будет ассоциирована с одним материалом.

◆ **Медиапрофайл.** Один объект может иметь один или более профайлов, которые соответствуют различным схемам кодирования. Один из профайлов является оригинальным, он называется мастерным профайлом, который соответствует первоначально созданному или записанному материалу. Дру-

гие будут получаться перекодированием из мастерного. Если материал закодирован тем же кодирующим средством, но с другими параметрами, формируется другой медиапрофайл.

◆ **Медиакопия.** Медиаобъект может быть поставлен в соответствие физическому объекту, называемому медиакопией. Медиакопия специфицируется идентификатором или локатором.



Рис. 5.14. Модель материала, профайла и копии

◆ **CreationInformation.** Информация о процессе формирования материала описывается DS CreationInformation. Одна копия описания DS будет ассоциирована с одним материалом.

◆ **UsageInformation.** Информация об использовании материала описывается DS UsageInformation. Одна копия описания DS будет ассоциирована с одним материалом.

Единственной частью описания, которая зависит от среды записи или формата кодирования является MediaInformation, описанная в этом разделе. Остальная часть описания MPEG-7 не зависит от профайлов или копий и, как след-

стве, может использоваться, чтобы описать все возможные копии материала.

Средства описания среды

Описание среды включает в себя один элемент верхнего уровня, DS MediaInformation. Оно состоит из опционального MediaIdentification D и одного или нескольких MediaProfile D

Идентификация среды (Media Identification) D содержит средства описания, которые являются специфическими по отношению к идентификации аудиовизуального материала вне зависимости от имеющихся различных копий.

Медиапрофайл D содержит различные средства описания, которые позволяют охарактеризовать один профайл аудиовизуального материала. Концепция профайла относится к различным вариациям, которые могут отклоняться от оригинала в зависимости от выбранного кодирования, формата записи и т. д. Профайл, соответствующий оригиналу или мастерной копии аудиовизуального материала, считается мастерным профайлом. Для каждого профайла может быть одна или более медиаконии мастерного медиапрофайла. MediaProfile D состоит из:

- ◆ MediaFormat D содержит средства описания, которые являются специфическими для формата кодирования медиапрофайла.

- ◆ MediaInstance D содержит средства описания, которые идентифицируют и локализуют различные копии медиапрофайлов.

- ◆ MediaTranscodingHints D содержит средства описания, которые специфицируют рекомендации по транскодированию для описываемого материала. Целью этого D (дескриптора) является улучшение качества и сокращение сложности транскодирующих приложений. Рекомендации по транскодированию могут использоваться в виде схем оценки кодирования с целью снижения вычислительной сложности.

- ◆ MediaQuality D предоставляет информацию об уровне качества аудио- или видеоматериала. Это может использо-

ваться для представления как субъективной, так и объективной оценки качества.

Создание и производство средств описания

Средства описания получения материала предоставляют авторские тексты, описания процесса формирования и/или производства аудиовизуального материала. Эта информация не может быть получена из самого материала. Эти данные связаны с материалом, но не описывают его буквально.

Описание формирования и производства материала содержит в качестве элемента верхнего уровня, DS Creation-Information, который состоит из одного Creation D, нуля или одного Classification D, и нуля или нескольких RelatedMaterial D.

Creation D содержит средства описания, имеющие отношение к формированию материала, включая место, дату, действия, материалы, персонал (технический и творческий) и организации, участвовавшие в процессе.

Classification D содержит средства описания, которые позволяют классифицировать аудиовизуальный материал. Classification D используется для описания классификации аудиовизуального материала. Это позволяет осуществлять поиск и отбор на основе предпочтений пользователя, ориентируясь на классификации пользователя (например, по языку, стилю, жанру и т. д.) и на классификации услуг (например, на цель, патентную защиту, сегментацию рынка, медиаревью и т. д.).

Related Material D содержит средства описания, имеющие отношение к дополнительной информации о аудиовизуальном материале, имеющемся в других материалах.

Средства описания использования содержимого

Средства описания информации об использовании материала предоставляют данные о процессе использования аудио-визуального материала.

Описание данных об использовании обеспечивается посредством DS UsageInformation, который может включать один Rights D, нуль или один Financial D и нуль или несколько Availability D и UsageRecord D.

Важно заметить, что описание DS UsageInformation предполагает добавление новых описаний, каждый раз, когда материал используется (например, DS UsageRecord, доход в Financial D), или когда имеются другие способы доступа к материалу (например, Availability D).

- ◆ Rights D предоставляет доступ к информации о правах владельцев и правах доступа.

- ◆ Financial D содержит информацию, относящуюся к издержкам и доходам от полученного аудиовизуального материала. Понятия частичных издержек и доходов позволяют классифицировать различные издержки и доходы, в зависимости от их типа. Итоговые издержки и доходы вычисляются приложением на основе указанных выше составляющих.

- ◆ Availability D содержит средства описания, относящиеся к доступности использования материала.

- ◆ DS UsageRecord содержит средства описания, относящиеся к прошлому использованию материала.

5.4. Описание контента

Описание структурных аспектов контента

Основным элементом этой части описания является DS сегмента. Она относится к описанию физического и логического аспектов аудиовизуального материала. DS сегмента может использоваться для формирования сегментных деревьев. MPEG-7 специфицирует также DS графа, который позволяет представлять сложные взаимоотношения между сегментами. Она используется для описания пространственно-временных соотношений, между сегментами, которые не описаны структурами дерева.

Сегмент представляет собой секцию аудиовизуального материала. DS сегмента является абстрактным классом (в

смысле объектно ориентированного программирования). Она имеет девять основных подклассов: DS мультимедийного сегмента, DS аудиовизуальной области, DS аудиовизуального сегмента, DS аудиосегмента, DS статической области, DS статической 3D-области, DS подвижной области, DS видеосегмента и DS электронной раскраски. Следовательно, она может иметь как пространственные, так и временные свойства. Временной сегмент может быть набором фрагментов аудиовизуальной последовательности, представленным DS аудиосегмента, набором кадров видеопоследовательности, представленным DS видеосегмента или комбинацией аудио и видеоинформации, охарактеризованной DS аудиовизуального сегмента. Пространственный сегмент может быть областью изображения или кадром в визуальной последовательности, представленным DS статической области для 2D-областей и DS статической области 3D для 3D-областей. Пространственно-временной сегмент может соответствовать подвижной области в видеопоследовательности, представленной DS подвижной области или более сложной комбинацией визуального и аудиоматериала, представленного, например, DS аудиовизуальной области. InkSegment DS описывает временной интервал или сегмент электронной раскраски, который соответствует набору чернильных капель, выбрасываемых из сопла. Наконец, наиболее общим сегментом является DS мультимедийного сегмента, который описывает составные сегменты, образующие мультимедийную презентацию. DS сегмента является абстрактным и не может быть отображен сам по себе: он используется для определения общих свойств его подклассов. Любой сегмент может быть описан с помощью информации формирования, использования медийных данных и текстовой аннотации. Более того, сегмент может быть поделен на субсегменты с помощью DS декомпозиции сегмента.

Сегмент не является обязательно связанным, он может быть составлен из нескольких несвязанных компонентов. Связность здесь относится как к пространственным, так и к времен-

ным доменам. Временной сегмент (видеосегмент, аудиосегмент или аудиовизуальный сегмент) считается связанными, если он является непрерывной последовательностью видеокадров или аудиофрагментов. Пространственный сегмент (статическая область или статическая 3D-область) считается связанными, если он является группой связанных пикселей. Пространственно-временной сегмент (подвижная область или аудиовизуальная область) считается связанным в пространстве и времени, если временной сегмент, где он размещен является связанным и если каждый кадр, в него входящий, является пространственно связанным (заметим, что это не является классической связностью в 3D-пространстве).

На рис. 5.16 проиллюстрированы несколько примеров временных или пространственных сегментов и их связности. Рис. 5.16а и 5.16б иллюстрируют временные и пространственные сегменты, содержащие один связный компонент. Рис. 5.16в и 5.16г иллюстрирует временной и пространственный сегменты, состоящие из трех связанных компонент. Заметим, что в последнем случае дескрипторы и DS, привязанные к сегменту, являются глобальными по отношению к объединению связанных компонент, образующих сегмент. На этом уровне невозможно индивидуально описать связанные компоненты сегмента. Если связанные компоненты должны быть описаны индивидуально, тогда сегмент разделяется покомпонентно.

DS Сегмента является рекурсивным, т. е., он может быть поделен на субсегменты и таким образом образовать древовидную структуру. Результирующее сегментное дерево используется для определения медиаисточника, временной и/или пространственной структуры аудиовизуального материала. Например, видеопрограмма может быть временно преобразована в ряд сцен различного уровня, снимков, и микро-сегментов; оглавление может, таким образом, генерироваться на основе этой структуры. Подобные стратегии могут использоваться для пространственных и пространственно-временных сегментов.



Рис. 5.15. Примеры разложения сегмента на компоненты: а и б — декомпозиции сегмента без зазоров и перекрытий; в и г — декомпозиции сегмента с зазорами и перекрытиями

Сегмент может также разделен на составные части по медиаисточникам, таким как различные звуковые дорожки или разные позиции видеокамер. Иерархическая декомпозиция полезна при формировании эффективных стратегий поиска (от глобального до локального). Она также позволяет описанию быть масштабируемым: сегмент может быть описан непосредственно с помощью его набора дескрипторов и DS, а может быть также описан набором дескрипторов и DS, которые относятся к его субсегментам. Заметим, что сегмент может быть разделен на субсегменты различного типа, например, видеосегмент может быть разложен на движущиеся области, которые, в свою очередь, разлагаются на статические области.

Так как это выполняется в пространственно-временном пространстве, декомпозиция должна описываться набором атрибутов, определяющих тип разложения: временное, пространственное или пространственно-временное. Более того, пространственная и временная подсекции могут располагаться с зазором или с перекрытием. Несколько примеров декомпозиций для временных сегментов описано на рис. 5.15. Рисунок 5.15а



Рис. 5.16. Примеры сегментов:

а и б — сегменты состоят из одного связного компонента;
в и г — сегменты состоят из трех связанных компонентов

и 5.15б описывают два примера декомпозиции без зазоров или перекрытий. В обоих случаях объединение дочерних объектов соответствует в точности временному продолжению родительского, даже если родитель сам не является связанным (см. пример на рис. 5.15б). Рисунок 5.15в демонстрирует пример декомпозиции с зазорами, но без перекрытий. Наконец, рис. 5.15г иллюстрирует более сложный случай, где родитель состоит из двух связанных компонентов и его декомпозиция создает три дочерних объекта: первый сам состоит из двух связанных компонентов, остальные два состоят из одного связанного компонента. Декомпозиция допускает зазоры и перекрытия. Заметим, что в любом случае декомпозиция означает, что объединение пространственно-временного пространства, определенное дочерними сегментами, включается в пространство, определенное его сегментом-предшественником (дочерние объекты содержатся в предшественниках).

Как упомянуто выше, любой сегмент может быть описан с помощью данных формирования, информации об использо-

вании медиаданных и текстовой аннотации. Однако специфические характеристики, зависящие от типа сегмента, также допускаются. Примеры специфических характеристик представлены в табл. 5.1. Большинство дескрипторов (D), соответствующих этим характеристикам, может быть получено автоматически из исходного материала. Для этой цели в литературе описано большое число различных средств.

Пример описания изображения представлен на рис. 5.17.

Исходные изображения описаны как стационарные области, SR1, которые описаны с помощью данных формирования (заголовки, создатель), информации использования (авторские права), медийной информации (формат файла), а также текстовой аннотации (обобщающей свойства изображения), гистограмм цвета и дескриптора текстуры. Исходная область может быть в дальнейшем разложена на составные области. Для каждого шага декомпозиции мы указываем, допустимы или нет зазоры и перекрытия. Дерево сегмента состоит из 8 стационарных областей (заметим, что SR8 является одиночным сегментом, составленным из двух связанных сегментов). Для каждой области на рис. 5.17 показан тип характеристики, которая реализована. Заметим, что в иерархическом дереве не нужно дублировать информацию формирования, использования и пр., так как предполагается, что дочерние сегменты наследуют эти характеристики.

Таблица 5.1

Примеры характеристик для описания сегмента

Характеристика	Видео-сегмент	Неподвижная область	Динамическая область	Видео-сегмент
Время	×	—	×	×
Форма	—	×	×	—
Цвет	×	×	×	—
Текстура	—	×	—	—
Движение	×	—	×	—
Движение камеры	×	—	—	—
Мозаика	×	—	—	—
Характеристики звука	—	—	×	×

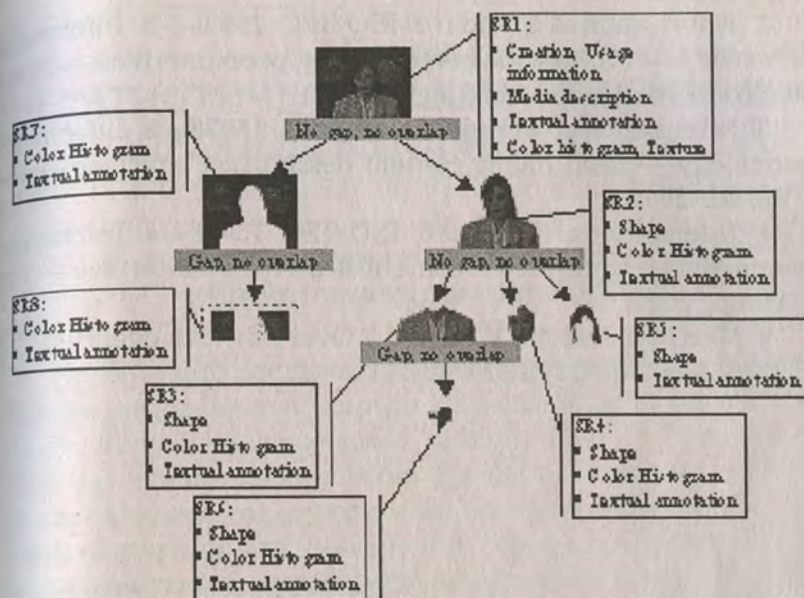


Рис. 5.17. Примеры описания изображения со стационарными областями

Описание структуры материала может выходить за рамки иерархического дерева. Хотя иерархические структуры, такие как деревья, удобны при организации доступа, поиска и масштабируемого описания, они подразумевают ограничения, которые делают их неприемлемыми для некоторых приложений. В таких случаях DS графа сегмента не используется. Структура графа определяется набором узлов, представляющих сегменты, и набора ребер, определяющих отношения между узлами.

Литература

1. International Standard ISO/IEC 15938—1. Information Technology — Multimedia content description interface — Part 1: Systems, 2002.

2. International Standard ISO/IEC 15938—2. Information Technology — Multimedia content description interface — Part 2: Description definition language, 2002.

3. International Standard ISO/IEC 15938—3. Information Technology — Multimedia content description interface — Part 3: Visual, 2002.

4. International Standard ISO/IEC 15938—4. Information Technology — Multimedia content description interface — Part 4: Audio, 2002.

5. *Martinez Jose M.* MPEG —7 Overview. Klagenfurt, July 2002 (ISO/IEC JTC1/SC29/WG11).

Глава 6

СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ

6.1. Отличительные характеристики видеопоследовательностей

Видео — это последовательность непрерывных пространственно неподвижных картин, называемых кадрами. Существует несколько физических причин, почему трассы видеоисточников являются особыми. Каждая неподвижная картина представляется кодирующим алгоритмом в цифровом виде и затем сжимается для уменьшения полосы пропускания. Обычно используемый способ уменьшения полосы пропускания — переслать начальный полный кадр, а затем пересылать разностные кадры. Это называется *межкадровым кодированием*. Так как расположенные рядом кадры не слишком отличаются друг от друга (поскольку движение является непрерывным), это приводит к существованию значительной корреляции кадров, находящихся рядом. Защититься от ошибок передачи можно, периодически передавая полный кадр. Кроме того, при изменении сцены кадры больше не зависят от прошлых кадров. В этом случае функциональная корреляция заканчивается, что может также положить конец и статистической корреляции в размерах кадра. Так как произошедшие изменения требуют, чтобы новый кадр был передан полностью, то продолжительности сцен отражаются на характере трассы. По этим и некоторым другим причинам видеотрафик отличается от широкополосного трафика данных. И как следствие, модели и выводы, полученные для видео, не могут быть применены к другим типам трафика [1].

Когда информация теряется в течение передачи или межвходные времена кадров большие или сильно меняются,

качество видео ухудшается. Качеством воспроизведения видео можно управлять, ограничивая размеры буфера; при этом кадры, которые поступают с опозданием, могут не учитываться вообще. Размер буфера часто характеризуют периодом времени до его опустошения (что является максимальной задержкой кадра, которая может возникнуть). Целью проектирования является не допустить задержек больше 100—200 мс. Так как от источника до получателя могут встретиться несколько буферов, а также существуют другие источники задержек (например, время распространения сигнала), то в некоторых исследованиях в качестве максимального размера буфера используется величина 10 мс.

Кадры передаются единицами постоянного размера, которые называют *ячейками*. Степень потери информации характеризуется степенью потери ячеек (CLR — cell lose rate). При моделировании часто интересует, где происходят потери ячеек из-за переполнения буфера. Типичной задачей обеспечения эффективности функционирования сети является сохранение CLR не более 10^{-k} , где k обычно берется в интервале 3...6.

6.2. Моделирование источника

Главной проблемой при моделировании источника является выбор того, как представить трассы данных статистическими моделями. Модель источника подбирается, чтобы рассматривать ее в качестве входного процесса к модели исследуемой системы. Очевидно, что модель источника является приемлемой, если она адекватно описывает трассу в построенной модели системы. Под адекватностью понимается следующее: когда модель источника используется в модели системы, то значения интересующих рабочих характеристик, полученных при моделировании, достаточно близки к значениям, которые дает реальная трасса. Определение достаточной близости может зависеть от того, использование какой модели системы положено в основу.

Видеоконференции

При видеоконференциях показываются только “разговаривающие головы”, и поэтому видеоконференции могут считаться самым простым типом видео для моделирования. Для описания развлекательного видео используются модели, произошедшие от них.

Приведем данные трех различных кодеров и четырех видеотелеконференций А, В, С и D примерно полчасовой длины. Каждые данные состоят из размеров неподвижных картин, т. е. кадров. Все телеконференции показывают сцену “голова и плечи” с небольшим движением и изменениями сцены и с небольшим приближением или удалением камеры (панорамированием). Все кодеры использовали версию H.261 стандарта видеокодирования.

Основные различия последовательностей в том, что, например, последовательность А была записана кодером, который не использует ни дискретно-косинусного преобразования (DCT), ни компенсации движения. Последовательность В была записана кодером, который использует и DCT и компенсацию движения. Последовательности С и D были записаны кодером, который использует DCT, но не использует компенсацию движения. Графики на рис. 6.1 показывают, что присутствие или отсутствие DCT или компенсации движения оказывает незначительное влияние на интересующие статистические характеристики. Сводка статистических характеристик этих последовательностей приведена в табл. 6.1.

Все наборы данных корректно описываются отрицательно-биномиальными маргинальными распределениями (диск-

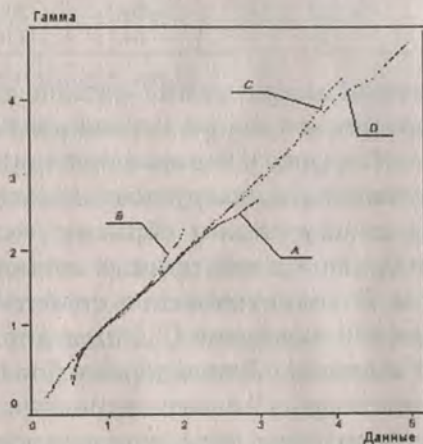


Рис. 6.1. QQ-графики для четырех последовательностей

Сводка статистических характеристик последовательностей

Последовательность	Биты на ячейку	Среднее, ячейки	Среднеквадратическое отклонение, ячейка	r
A	14	1506,4	512,7	0,981
B	48	104,9	29,7	0,984
C	64	130,3	74,4	0,985
D	64	170,6	107,6	0,970

ретный аналог гамма-распределения) и геометрически изменяющимися корреляционными функциями.

На рис. 6.2 показаны нормированные корреляционные функции анализируемых последовательностей. Ордината имеет лог-шкалу; таким образом, геометрические функции будут отображены как прямые линии. Для последовательностей B, C и D геометрическое свойство сохраняется, как минимум, для 100 задержек (2,5 с), а для последовательности A — для 50 задержек. Для задержек более чем 250 геометрическая функция недооценивает корреляционную функцию. Проведенные исследования последовательностей A, B и C позволяют сделать вывод о присутствии в них долговременной зависимости.

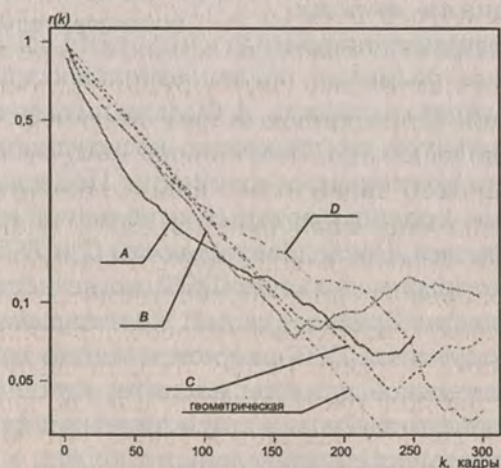


Рис. 6.2. Коэффициент корреляции последовательностей A, B, C и D

Поскольку отрицательно-биномиальное и гамма-распределения полностью определяются двумя параметрами, эти параметры могут быть легко оценены методом моментов с помощью математического ожидания и дисперсии количества ячеек на кадр. Для описания свойств трафика VBR-телеконференции требуются только эти два момента и коэффициент корреляции r , который может быть оценен из рис. 6.2.

Широковещательное видео

Обратимся к более динамичным последовательностям, таким как фильмы, новости, спорт и развлекательное телевидение. Так как основное назначение моделей — оказывать помощь в оценках сетевых характеристик, рассмотрим модели с позиции предсказания степени потери ячеек.

Широковещательное VBR-видео отличается по скорости потока от VBR-видео конференций. Последовательности видеоконференций состоят из картинок “голова и плечи” с небольшим панорамированием или без него, в то время как вещательное видео характеризуется постоянной сменой сцен. В связи с межкадровым кодированием очевидно, что при изменении сцены потребуется больше бит, чем для внутрисценных кадров. Таким образом, по крайней мере в этом отношении широковещательное видео будет отличаться от видеоконференций. Существуют также и другие различия. В частности, было показано, что количество бит на кадр для широковещательного видео имеет корреляционную функцию отличную от корреляционной функции видеоконференций или видеотелефонии. Для последних двух процессов корреляционные функции подобны друг другу и геометрически спадают к нулю. Для широковещательного видео корреляционная функция не затухает к нулю. Более того, первый кадр после изменения сцены содержит значительно больше бит, чем другие кадры в сцене. Отмечается также, что корреляционная функция при малых задержках спадает быстрее, чем при больших. В результате временные ряды могут быть описаны псевдомарковским процессом, который определяется бито-

выми интенсивностями различных типов сцен (и состоянием изменений сцен).

Моделирование широковещательного видео

Имеется несколько наборов данных, определяющих сумму битов на кадр, для последовательностей кодируемых внутриполеовой/межкадровой схемой ДИКМ кодирования без использования DST или компенсации движения. Стратегия моделирования заключается в развитии способа опознавания смены сцен и дальнейшего построения модели для длин сцен и суммы ячеек в кадре смены сцены.

Перед описанием статистических моделей приведем некоторые основные статистические характеристики используемых последовательностей. На рис. 6.3 показаны максимальные X_{\max} и средние X_{cp} битовые интенсивности и их отношения.

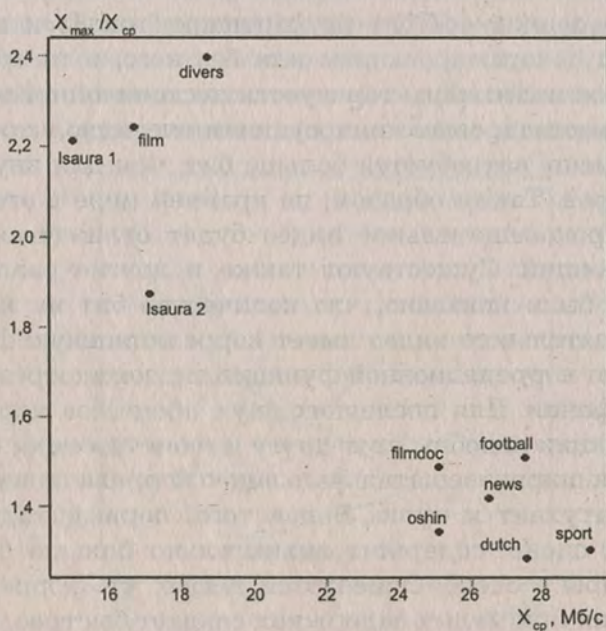


Рис. 6.3. Отношения максимума к среднему значению

Отношения максимума к среднему значению изменяются от 1,3 до 2,4. Для сравнения: отношение максимума к среднему для видеоконференции с таким же кодеком (последовательность А) равно 3,2. Отметим, что большие отношения максимума к среднему связаны со снижением средних битовых интенсивностей. Последовательности *divers*, *film*, *Isuara 1*, и *Isuara 2*, которые имеют пониженную среднюю интенсивность и высокие отношения максимума к среднему, были различными телепрограммами, записанными с сетевого кабельного телевидения (и помечены как нормальное качество широкоэмитательного видео). Последовательности с низкими отношениями максимума к среднему (такие как *football*, *sport*, *news* и др.) были взяты прямо из телевизионных студий (и обозначены как высококачественное широкоэмитательное видео).

6.3. Определение смены сцен

На рис. 6.4 показаны два участка трассы *film*. Визуально можно наблюдать, что в них существует несколько острых выступов, которые, возможно, появляются из-за смены сцен. Так как эти выступы могут быть основной причиной потери ячеек, необходимо смоделировать их длительность и амплитуду. Если просто фиксировать длительность, а не моделировать распределение этих длительностей, то мультиплексированные источники с неодинаковыми начальными условиями иногда не будут иметь совпадающие пики, что приведет к недооценке степени потери ячеек.

Будем полагать, что изменение сцены происходит, когда кадр содержит необычно большое количество ячеек сравнительно с его соседями. Для количественной оценки будем рассматривать X_i как количество ячеек в кадре i . При смене сцены разность второго порядка $\Delta X_i^2 = (X_{i+1} - X_i) - (X_i - X_{i-1})$ будет большой по величине и отрицательной по знаку. Чтобы определить, что понимается под "большим", разделим разность второго порядка на среднее нескольких прошлых кад-

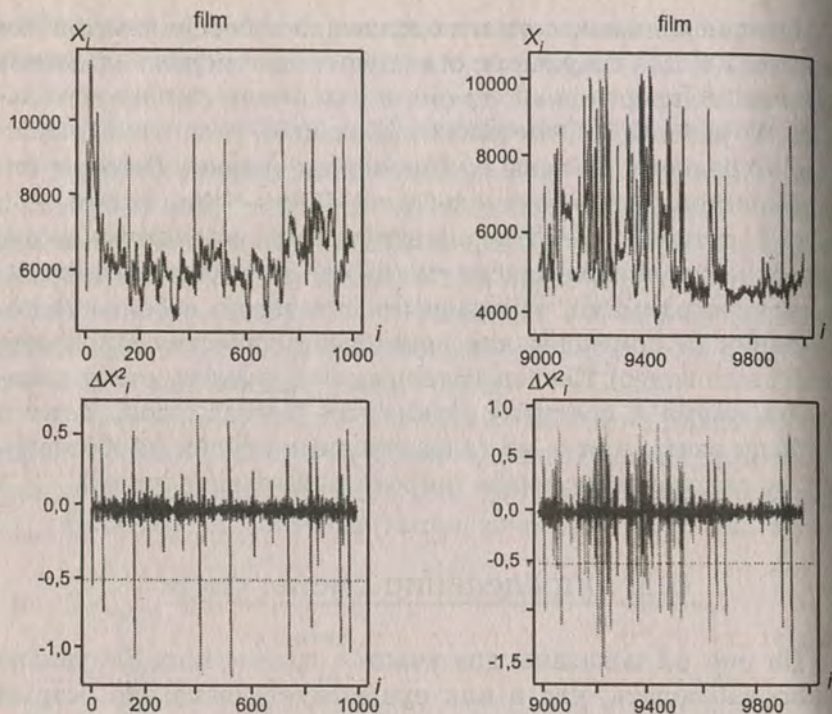


Рис. 6.4. Битовые интенсивности
и масштабные разности второго порядка

ров. В качестве критического значения выбиралась величина, равная $-0,5$. Поскольку выбор этого значения абсолютно субъективен, рассмотрим статистическую теорию определения резко выделяющегося значения для вывода о том, что предполагаемое событие в действительности имеет место.

Для определения того, точен ли критерий определения смены сцен, рассмотрим временной ряд X_i и соответствующие значения масштабных разностей второго порядка. ΔX_i^2 . Первые 1000 кадров показаны слева на рис. 6.4. Проверяемые значения выявляют смены сцен с безошибочными результатами. Видно, что выбора значения равного $-0,5$ в качестве критического недостаточно. Кадры с 9001 до 10000 показаны

справа. ΔX_i . Эта часть является более активной (в единицах измерения изменений битовой интенсивности) последовательностью, и изменение критического значения повлияет на количество кадров, определенных как смены сцен. Когда критическое значение равно $-0,5$, то присутствует 317 смен сцен (средняя длина сцены $6,5$ с); когда $-0,4$, то 374 и когда $-0,6$, то 283. ПРВ длин сцен, полученных с использованием этих критических значений, показаны на рис. 6.5. Заметим, что критическое значение не оказывает значительного влияния на ПРВ.

Длины сцен. Графики корреляционной функции показывают, что длины сцен не коррелированы, поэтому главная проблема при их моделировании состоит в описании распределения суммы кадров в сцене. Форма ПРВ (см. рис. 6.5) была

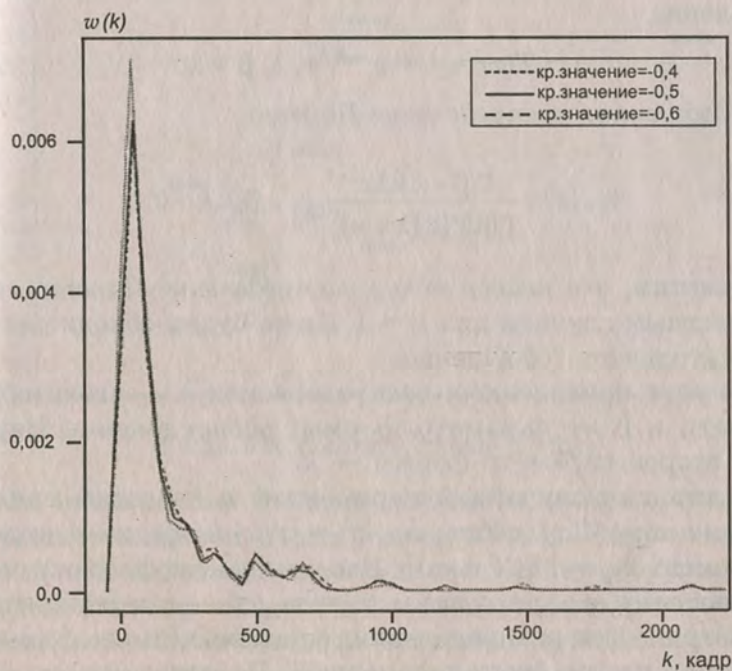


Рис. 6.5. ПРВ длин сцен

приблизительно одинаковой для всех последовательностей, за исключением news. (Это широковещательные новости, где присутствовали 134 смены сцен и 75 из них длиной в 3 кадра, причем большинство из них происходило последовательно.)

Полученный вид ПРВ выбирается в качестве опорного распределения для испытательных моделей. В качестве кандидатов на описание ПРВ длины сцены (и впоследствии количества ячеек на кадр) будем использовать следующие три распределения.

Гамма распределение, определяемое как

$$w_G(x) = \frac{\lambda(\lambda x)^\beta}{\Gamma(\beta+1)} e^{-\lambda x}, x > 0. \quad (6.1)$$

Распределение Вейбулла с дополнительной функцией распределения

$$1 - F(x) = 1 - e^{-\lambda x^\beta}, x, \lambda, \beta > 0. \quad (6.2)$$

Обобщенное распределение Парето

$$w_P(x) = \frac{\Gamma(\beta+k)\lambda^b x^{k-1}}{\Gamma(\beta)\Gamma(k)(\lambda+x)^{k+\beta}}, x, \beta, \lambda, k > 0. \quad (6.3)$$

Заметим, что классическое распределение Парето является частным случаем для $k = 1$. Ниже будем обходиться без прилагательного "обобщенное".

Во всех приведенных распределениях λ — масштабный параметр и β — параметр формы; распределение Парето имеет второй параметр формы — k .

Пусть для случайной переменной x известны среднее значение $m = M(x)$, дисперсия $\sigma^2 = \sigma^2(x)$ и третий центральный момент $m_3 = M[(X - m)^3]$. Назовем γ_1 коэффициентом изменчивости (т. е., $\gamma_1 = \sigma/m$), а $\gamma_3 = m_3/\sigma^3$ — коэффициентом асимметрии. Для распределений, описанных выше, γ_1 и γ_3 не зависят от масштабного параметра λ . Построение γ_3 на фоне γ_1 дает кривую для распределений Гамма и Вейбулла, и кри-

вую для каждого значения k распределения Парето. На рис. 6.6 показаны эти кривые, и точки выборочных значений для моментов длины сцены из исследуемых видеопоследовательностей. (Были включены данные с внутрикадровым кодированием фильма "Звездные войны".)

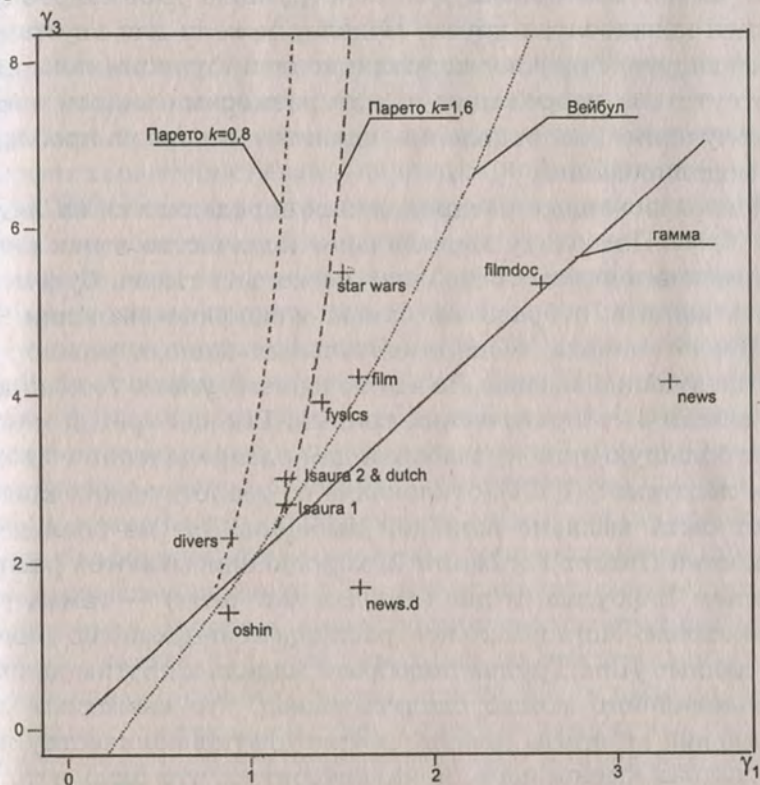


Рис. 6.6. Моменты длины сцены

За исключением последовательности news, моменты всех последовательностей расположены вблизи четырех кривых, полученных из распределений. В случае, когда выборки из гамма распределения сравниваются с гамма кривыми, расстояния от кривых находятся в пределах описанных отклонений. В итоге можно сделать вывод, что обычно (но не всегда) длины сцен соответствуют одновершинному распределению.

Кадры смены сцен

Смоделируем сумму ячеек в кадрах смены сцен. Очевидно, что кадры, которые начинают сцену, будут иметь больше ячеек, чем другие кадры, потому что большинство картинок имеют повторяющуюся конструкцию. Для кадров смены сцен понадобятся другие модели, нежели для внутрисценных кадров. Графики корреляционной функции показывают отсутствие корреляции между размерами кадров, поэтому получение распределения является основной проблемой при моделировании.

Использованные распределения определяются на интервале $(0; \infty)$. Поскольку минимальное количество ячеек в кадре изменения сцены составляет несколько тысяч, будем выбирать данные, отбрасывая самые маленькие значения.

После анализа экспериментальных данных можно сделать следующие выводы. За исключением *physics u football* данные далеки от теоретических кривых. Так как третий момент имеет большую ошибку выборки для распределений с "тяжелыми хвостами" (РТХ), отклонение от теоретических кривых может быть вызвано ошибкой выборки. Две из последовательностей (*Isaura 1* и *Isaura 2*) хорошо описываются распределением Вейбулла, и две (*dutch* и *star wars*) — гамма-распределением. Логнормальное распределение хорошо описывает данные *film*. Трудно подобрать модель к другим данным. Из изложенного можно сделать вывод, что известные распределения не всегда хорошо соответствуют количеству ячеек в кадрах смены сцен, и маловероятно, что одно и то же распределение будет хорошо подходить для всех последовательностей.

Внутрисценные кадры

Анализ графика битовой интенсивности (см. рис. 6.4) показывает, что воздействие смены сцен оказывается на два кадра, причем первый кадр после смены сцены также очень большой. Рассматривая *film*, *Isaura 1* и *Isaura 2*, находим,

что линейная регрессия хорошо моделирует количество ячеек в первом кадре после смены сцены. Будем считать Z_n количеством ячеек в n -м кадре смены сцены, и будем считать Y_n количеством ячеек в следующем $(n + 1)$ -м кадре. Тогда получаем

$$Y_{n+1} = a + bZ_n + \varepsilon_n, \quad (6.4)$$

где ε_n — независимые и одинаково распределенные нормальные случайные переменные с нулевым средним.

В остальных внутрисценивых кадрах важными статистическими характеристиками являются корреляционная функция и ПРВ. Исследования показали, что корреляционная функция сцен телевизионного сигнала не затухает геометрически к нулю, как это происходит для видеоконференций. Подобные функции получаются, когда все внутрисценивые кадры (исключая первые два кадра каждой сцены) рассматриваются как отдельный временной ряд. Рис. 6.7а, с иллюстрируют эту функцию для выборок *Isaura 1* и *film*. Заметим, однако, что этот метод оценки корреляционной функции игнорирует границы сцен. Например, последний кадр сцены и первый (по счету) кадр последующей сцены различаются всего на одну задержку при вычислении корреляционной функции. Для последовательности *film* это приводит к тому, что корреляционная функция имеет периодические пульсации, как показано на рис. 6.7с. Эти пульсации не представлены в корреляционных функциях первых шести сцен и вызваны искусственным объединением сцен. Более точная оценка $r(i)$ получается путем деления данных на сцены, вычисления опытной корреляционной функции для каждой сцены и усреднения по всем сценам. Вычислим $r_s(i)$, которая является корреляцией для i -й задержки в сцене s для всех i и всех s . Затем получаем корреляционную функцию $r(i)$ как

$$r(i) = \frac{1}{S} \sum_{s=1}^S R_s(i) 1_s(i), \quad i = 1, 2, \dots,$$

где S — это количество сцен в последовательности; функция $1_s(i)$ равна единице, если сцена s содержит по меньшей

мере $i + 3$ кадров (напомним, что первые два кадра рассматриваются отдельно, и для оценки корреляции для i -й задержки необходимо по меньшей мере $i + 1$ наблюдение). В противном случае $1_s(i) = 0$. Корреляционные функции, полученные таким образом, показаны на рис. 6.7b, d. Как видно, корреляционные функции геометрически спадают к нулю. Пунктирные линии на рис. 6.7b, d показывают геометрические функции, подобранные к вычисленным корреляционным функциям.

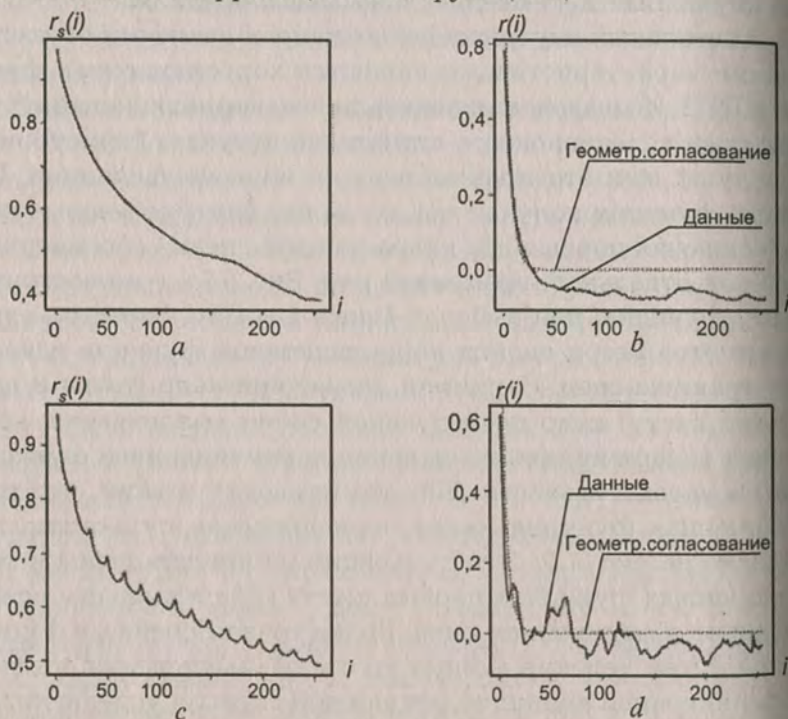


Рис. 6.7. Корреляции в пределах сцены:

a — внутрисценные кадры Isaura 1; b — корреляционная функция для Isaura 1; c — внутрисценные кадры film; d — корреляционная функция для film

Пользуясь QQ-графиками, приходим к выводу, что отображенные данные *Isaura 1*, *Isaura 2*, *divers* и *film* ведут себя в соответствии с распределением Парето; *dutch*, *sport* и

star wars соответствуют гамма-распределению; *football* и *filmdoc* могут генерироваться или гамма-, или вейбулла распределением, и *oshin*, *fysics* и *news* не соответствуют ни одному из стандартных распределений. (*News* выступает в качестве смеси 5—7 стандартных распределений, однако эту смесь трудно подобрать.)

Так как существует широкое многообразие возможных сцен, можно предположить, что распределение внутрисценивых кадров будет смесью одновершинных распределений.

6.4. Обоснование моделей

Из рис. 6.6 видно, что распределение Вейбулла является подходящим кандидатом для описания длины сцен *film*. Параметры этого распределения определялись с использованием метода моментов. В результате анализа также выяснилось, что число ячеек в кадрах изменения сцены соответствует распределению Вейбулла. Из (6.4) понятно, как вычислить количество ячеек в кадре, следующем за кадром изменения сцены. Количество ячеек на кадр в кадрах смены сцен не коррелировано между следующими друг за другом кадрами. Таким образом, нужно только построить модель для количества ячеек на кадр в пределах сцены. Используя рис. 6.8 для определения количества ячеек на кадр во внутрисценивых кадрах *film*, выбираем распределение Парето с параметром $k = 4$.

Рассмотрим модель на основе марковской цепи, которая создается следующим образом. Пусть X_n — количество ячеек в кадре n , и $[X_n]$ — целая часть $X_n/50$. (Выбор 50 является произвольным и связан с количеством состояний. При выборе 50 их количество составляет около 100.) Будем моделировать $[X_n]$; $n = 1, 2, \dots, N$ как марковскую цепь с матрицей переходов $P = (p_{ij})$. Оценки p_{ij} находим обычным способом:

$$p_{ij} = \frac{\text{количество переходов из } i \text{ в } j}{\text{количество переходов из } i},$$

где знаменатель больше нуля.

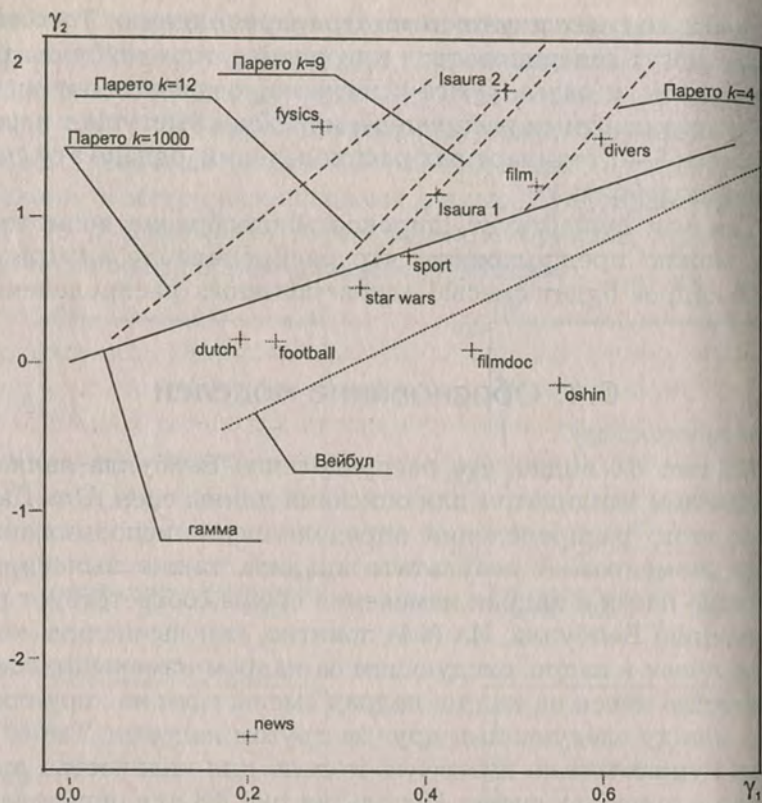


Рис. 6.8. Количество ячеек
во внутрисценивых кадрах

Isaura 1. Для последовательности *Isaura 1* модели генерируются подобным же образом. Распределение длин сцены для *Isaura 1* соответствует распределению Вейбулла. Распределение ячеек в кадрах в пределах сцены соответствует распределению Парето с параметром $k = 9$. Модели на основе марковской цепи (здесь моделируется $[X_n] = X_n/100$ вместо $X_n/50$) строятся таким же образом, как для последовательности *film*, за исключением того, что вместо Парето распределения используется распределение Вейбулла.

Результаты моделирования

На рис. 6.9 показаны результаты моделирования, оценивающие точность (по предсказанию степени потери ячеек) модели для последовательности *film* со средней скоростью 16,7 Мб/с и максимальной — 37, 31 Мб/с. Коэффициент использования канала $\rho = 0,849$. Моделирование было проведено для максимальных задержек в мультиплексоре, равных 0, 5; 1; 2; 3 и 4 мс. Максимальная интенсивность поступления ячеек от источника в мультиплексор была равной 43 Мб/с.

Было сгенерировано 10 выборок, которые использовались в 10 прогонах моделирования. На рис. 6.9 показаны вероятности потерь для 10 выборок при различных объемах буфера. Здесь изображены значения, соответствующие среднему плюс с. к. о ($m + \sigma$) и среднее минус с. к. о ($m - \sigma$). Для

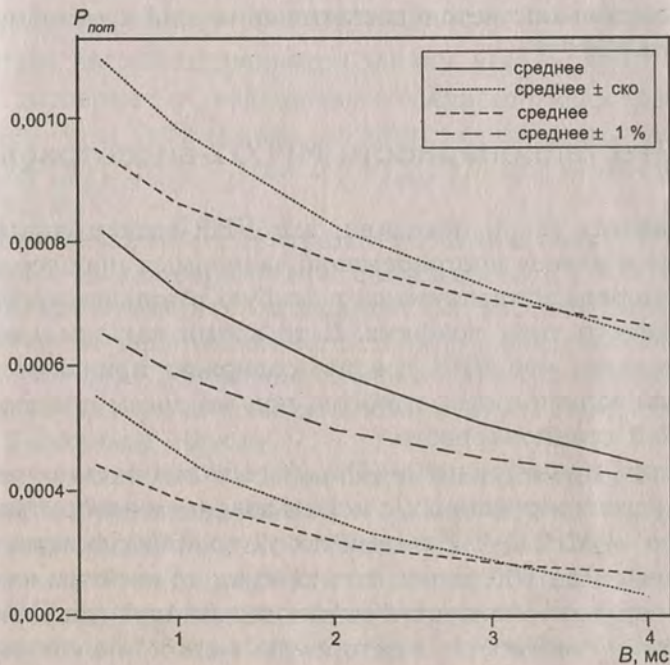


Рис. 6.9. Вероятности потери ячейки для *film*

реального трафика моделирование прогонялось, не только используя трассу реального трафика, но также с интенсивностями превышенными и заниженными на 1%, что также иллюстрируется на рисунке. Чтобы отразить диапазон значений, на фоне которых вероятности потерь, сгенерированные моделью, можно было сравнить, использовалось 1%-ное масштабирование. Так как интенсивности источника могут быть известны не очень точно, то разумно сравнивать потери для модели с реальными потерями с 1%-ной неточностью для средней интенсивности источника.

Как можно видеть из рисунка, для всех исследуемых размеров буфера в мультиплексе, средние интенсивности потерь 10 выборок попадали в область потерь полученных для реального трафика $\pm 1\%$. Реальная интенсивность потерь лежит в пределах средних потерь для модели \pm с. к. о. Следовательно, модель является достаточно точной для последовательности *film*.

6.5. Нестационарность MPEG2-видеотрафика

В работах [2; 3] показано, что VBR-видеотрафик принадлежит к классу долговременно зависимых процессов. Это, в свою очередь, подразумевает слабую стационарность рассматриваемого типа трафика. В то время как исследования подтверждают, что VBR-трафик содержит признаки долговременной зависимости, ставится под вопросом предположение слабой стационарности.

Анализ, проведенный в [4], основывался на трех потоках трафика, сгенерированных с использованием алгоритма сжатия видео — MPEG-2. Каждый поток трафика включал приблизительно 154 000 точек, где каждая точка отражала общую битовую интенсивность одного видеокadra (информацию о яркости и цветности, векторе движения и информацию заголовка). В видео, взятом из исходного фильма, оба поля каждого кадра идентичны. Для улучшения отношения сиг-

нал—помеха ОСП эти поля были усреднены, и получившиеся данные (720×288 пикселей, 25 кадров/с) закодированы программным кодером MPEG-2. Остановимся на самой простой схеме кодирования, где все кадры кодируются во внутриволевом режиме.

Для исследования был выбран фильм “Звездные войны”. Среди всех существующих жанров фильмов “экшэны” — самые требовательные к сетевому управлению, из-за того, что в них слишком быстро меняются сцены, быстро изменяются условия освещения и т. п.

В результате к настоящему времени стало очевидным, что видеотрафик переменной битовой интенсивности содержит признаки долговременной зависимости:

- ◆ корреляция r_k проявляет гиперболическое затухание для больших задержек k : $r_k \xrightarrow{k \rightarrow \infty} c_0 k^{-\beta}$;

- ◆ спектральная плотность мощности $S(\omega)$ на небольших значениях частот подчиняется закону $S(\omega) \xrightarrow{\omega \rightarrow 0} c_1 \omega^{\beta-1}$;

- ◆ дисперсия σ_n^2 выборочного среднего уменьшается медленнее, чем обратная величина размера выборки n : $\sigma_n^2 = \sigma^2(\bar{X}_n) \xrightarrow{n \rightarrow \infty} c_2 n^{-\beta}$, где $\bar{X}_n = \sum_{i=1}^n X_i / n$ для некоторых констант c_0, c_1, c_2 .

Постоянная $\beta \in [0; 2]$ отражает тип зависимости; $0 \leq \beta < 1$ указывает на долговременную зависимость; а $1 < \beta \leq 2$ говорит о кратковременной зависимости в данных. (Степень персистентности чаще выражается при помощи показателя Херста $H = 1 - \beta/2$.) Долговременная зависимость определяется в рамках структуры слабой стационарности [5], т. е. стационарности в широком смысле.

Стационарность и эргодичность позволяют из каждой отдельной выборки данных или в данном случае — отдельного временного ряда, получить статистические оценки, такие как среднее значение и дисперсия или другие параметры модели. Если предположения о стационарности и эргодичности не выполняются, то некоторые меры, такие как среднее значение и дисперсия, могут стать бессмысленными. В действительности среднее значение временного ряда VBR-видео

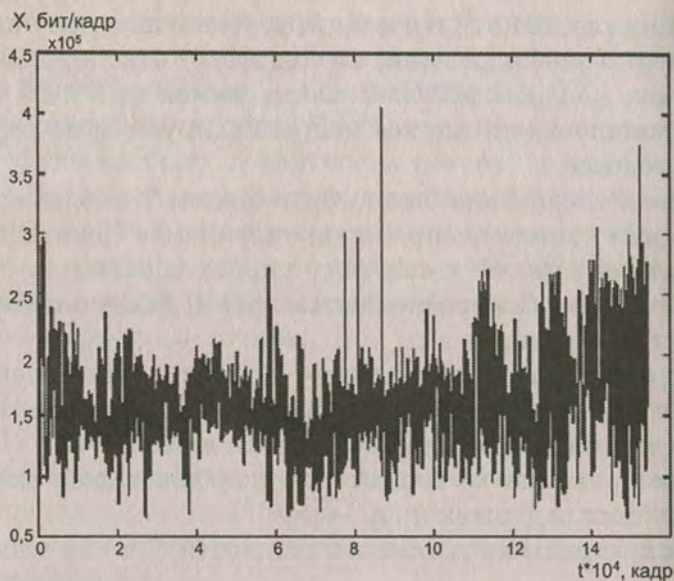


Рис. 6.10. Профиль битовой интенсивности для VBR-трафика

сходится очень медленно, что может быть вызвано нестационарностью и совсем не обязательно долговременной зависимостью. Из профиля битовой интенсивности для трафика (рис. 6.10) видно, что наиболее вероятным объяснением наблюдаемой долговременной зависимости является нестационарность.

Тестирование на стационарность

Пусть $X(n)$, $n = 0, 1, 2, \dots$, — стохастический процесс со спектральной плотностью мощности $S(\omega)$. Оценим периодограмму этого процесса в виде $\hat{I}_N(\omega) = \frac{1}{2\pi N} \left| \sum_{n=0}^{N-1} (X(n) - \bar{X}) e^{-j\omega n} \right|^2$, где \bar{X} — выборочное среднее, которое сходится к $\frac{1}{2} S(\omega) / \omega^2$ для $\omega \neq 0; \pm\pi; \pm 2\pi; \dots$ Это подразумевает, что оценка $\hat{I}_N(\omega)$ для больших N является несмещенной, но несостоятельной оценкой, так как $\lim_{N \rightarrow \infty} \sigma^2 \hat{I}_N(\omega) = S^2(\omega)$. Тем не менее справед-

ливо, что для двух фиксированных частот ω_1 и ω_2 ординаты периодограммы $\hat{I}_N(\omega_1)$ и $\hat{I}_N(\omega_2)$ приблизительно некоррелированы. Эти свойства справедливы также и для долговременно зависимых процессов. Применение *спектрального окна* $\Lambda(\omega)$ дает состоятельную оценку

$$\overline{I}_N(\omega) = \int_{-\pi}^{+\pi} \hat{I}_N(\omega) \Lambda(\Theta - \omega) d\omega, \quad (6.5)$$

Выбирая спектральное окно Бартлетта-Пристли (*Bartlett-Priestley*), получим для дисперсии $\sigma^2(\overline{I}_N(\omega)) \approx \frac{6N}{5N} S^2(\omega)$. Дисперсия все еще зависит от самой спектральной плотности мощности. Чтобы преодолеть эту функциональную зависимость, можно использовать логарифмическое *преобразование, стабилизирующее дисперсию*.

Для первого порядка точности получаем:

$$M\{\log(\overline{I}_N)\} \approx \log(I_N); \quad (6.6, \text{ а})$$

$$\sigma^2\{\log(\overline{I}_N)\} \approx \frac{2\pi}{N} \int_{-\pi}^{+\pi} \Lambda^2(\Theta) d\Theta, \quad (6.6, \text{ б})$$

где $\omega \neq 0; \pm\pi; \dots$

Таким образом, оценка $\log(\overline{I}_N)$ ближе к нормальности, чем нетрансформированная оценка. Чтобы подтвердить (или опровергнуть) предположение слабой стационарности, процесс X разбивается на I сегментов, каждый из которых центрирован по времени t_i и имеет длину N . Для каждого i -го сегмента вычисляется спектральная плотность мощности $\overline{I}_{N,i}(\omega)$ в соответствии с (6.5). Дискретизация сглаженной периодограммы (6.5) осуществляется по частотам $\omega_j = \pi j / N (j = j_0 + k\Delta j, k = 0; 1; \dots J)$, а логарифмирование дает двумерную случайную переменную $Y_{ij} = \log(\overline{I}_{N,i}(\omega_j))$. Если частоты ω_j , также как и времена t_i , имеют достаточно широкий разброс, то случайная величина Y_{ij} приблизительно нормально распределена и некоррелиро-

вана. Предположение приблизительной нормальности и некоррелированности Y_{ij} в обоих измерениях подразумевает приблизительную независимость Y_{ij} . Следовательно, чтобы определить структуру лежащего в основе случайного процесса, можно использовать методику анализа дисперсии:

$$Y_{ij} = \mu + a(t_i) + b(\omega_j) + c(t_i, \omega_j) + \eta_{ij}, \quad (6.7)$$

где η_{ij} являются независимо и одинаково нормально распределенными с нулевым средним значением и дисперсией σ^2 , определенной соотношением (6.6, δ).

Присутствие $c(t_j, \omega_j)$ и $a(t_i)$ проверяют, используя величины:

$$S_{I+R} = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - Y_{.j} - Y_{i.} + Y_{..})^2, \quad (6.8)$$

$$S_T = J \sum_{i=1}^I (Y_{i.} - Y_{..})^2, \quad (6.9)$$

где точка показывает среднее по индексу, который она заменяет, например, $Y_{.j} = \sum_{i=1}^I Y_{ij} / I$. Для стационарного процесса можно ожидать, что $c(t_j, \omega_j)$ и $a(t_i)$ исчезнут. В этом случае S_{I+R} / σ^2 и S_T / σ^2 являются χ^2 — распределенными с $(I-1)(J-1)$ и $(I-1)$ степенями свободы соответственно. Гипотеза стационарности отвергается, если один из статистических тестов превышает на 1% квантиль соответствующего χ^2 -распределения.

На рис. 6.11 и 6.12 показано, что шумовые слагаемые для VBR-трафика распределены нормально и некоррелированы. Следовательно, в присутствии долговременной зависимости этот тест не может использоваться, потому что шум не распределен нормально и коррелирован.

Типы нестационарности для VBR-видеотрафика

Тест стационарности предполагает, что модель сцены и видеотрафик имеют одинаковую структуру $M\{Y_{ij}\} = \mu + a(t_i) +$

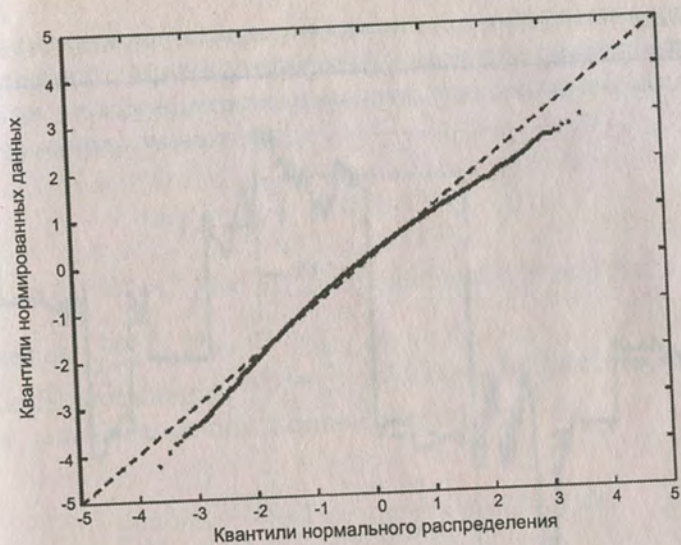


Рис. 6.11. QQ-график шумовых слагаемых η_{ij}

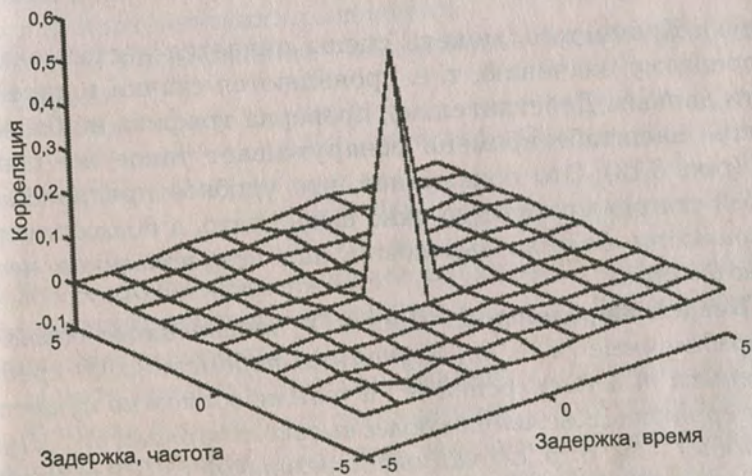


Рис. 6.12. Двухмерная корреляция шумовых слагаемых η_{ij}

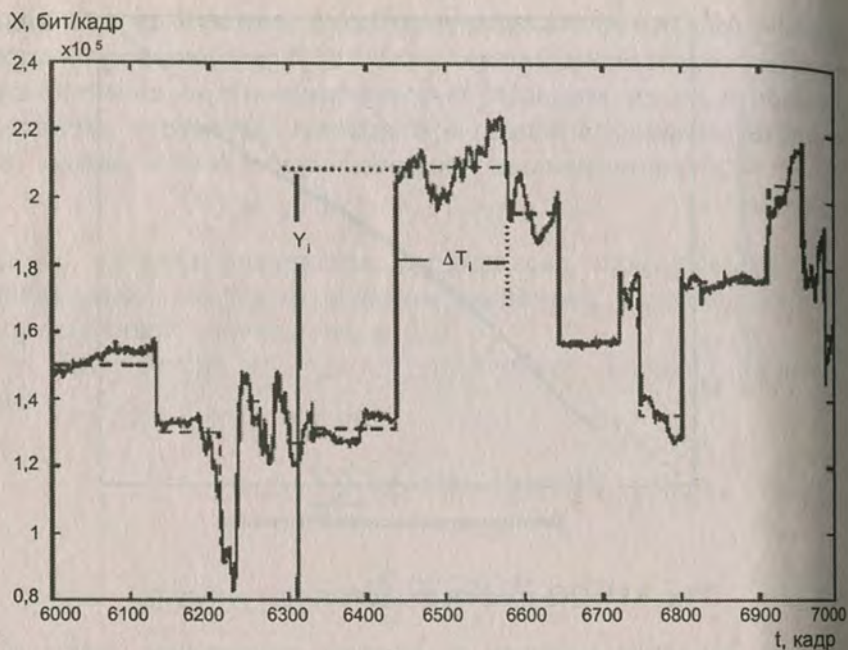


Рис. 6.13. Секция VBR-видеотрафика

+ $b(\omega)$). Кроме того, модель сцены является нестационарной по среднему значению, т. е. проявляются скачки в интенсивности данных. Действительно, проверка трафика на более коротком масштабе времени обнаруживает такое же поведение (рис. 6.13). Это показывает, что удобное предположение слабой стационарности должно быть снято, а долговременная зависимость должна выглядеть как искусственная нестационарность.

Введем определение. Пусть Y_i независимые одинаково распределенные (н. о. р.) случайные переменные со средним значением m и дисперсией σ_y^2 , а также возможно существование групповых моментов более высокого порядка $m_{y,r}^1$. Пусть (периоды) — н. о. р. случайные переменные с ПРВ и средним значением. Тогда стохастический процесс $X(t) = Y_i$ для $t_i \leq t < t_{i+1}$ ($i = 1, 2, \dots$) называется процессом сдвигающегося уровня.

Концепция нестационарности, на первый взгляд, не имеет существенной привлекательности и кажется слишком громоздкой. Тем не менее можно показать, что эмпирические групповые моменты имеют вид:

$$m_{X,r} = \frac{1}{T} \int_{t_1}^{t_{n+1}} X^r(t) dt = \sum_{i=1}^n Y_i^r \frac{\Delta T_i}{T}, \quad (6.10)$$

где $T = t_{n+1} - t_1$ и $r = 0, 1, \dots$, существуют с математическим ожиданием $\lim_{T \rightarrow \infty} M \{ m_{X,r}^n \} = m_{Y,r}^n$.

Даже корреляция $R(\tau) = \frac{1}{T} \int_{t_1}^{t_{n+1}} (X(t) - m_Y)(X(t+\tau) - m_Y) dt$ существует для бесконечно длинного процесса

$$\lim_{T \rightarrow \infty} M \left\{ \frac{\gamma(\tau)}{\sigma_Y^2} \right\} = 1 - \frac{\tau}{m_t} + \frac{1}{m_t} \int_k^\tau (\tau - t) w_t(t) dt \quad (6.11)$$

для положительных τ , где k — длина самого короткого периода.

Следовательно, эти процессы являются асимптотически слабо стационарными, что обеспечивает приемлемую основу для практических приложений.

Теперь предположим, что периоды имеют распределение Парето:

$$w_t(t) = \begin{cases} f(t) & \text{для } 0 < k \leq t < t_0, \\ \frac{\Theta t_1^\Theta}{t^{\Theta+1}} & \text{для } t \geq t_0, \end{cases} \quad (6.12)$$

где $f(t)$ — некоторая положительная функция, такая, что w_t — соответствующая плотность; а t_0, t_1 — некоторые положительные константы.

Можно показать, что справедлива следующая теорема.

Теорема 6.1. *Процесс сдвигающихся уровней имеет распределение периодов (6.12), если и только если соответствующая корреляционная функция проявляет долговременную зависимость с показателем Херста $H = (3 - \Theta)/2$.*

Показано, что спектральная плотность мощности вблизи нуля для процесса сдвигающихся уровней (ПСУ) с распределением (6.12) соответствует степенному закону. Это говорит о том, что нестационарность для среднего значения может послужить причиной долговременной зависимости.

В некоторых случаях на практике можно говорить только о локальной стационарности, и важно точно определить соответствующие масштабы времени стационарного фрактального поведения. Отметим, что иногда, помимо статистического доказательства фрактального поведения, физическое описание механизмов генерации трафика также может помочь при выборе фрактальных моделей.

6.6. Анализ и модели VBR-видеоданных

В случае, когда показатель Херста теряет свое первоначальное значение, о данных нельзя сказать, что они самоподобны. При этом аргументы о нестационарности, как правило, основываются на визуальном осмотре.

Можно, однако, проведя предварительно некоторую сортировку данных, ограничиться диапазоном, в котором процесс мог бы рассматриваться как стационарный. В качестве примера рассмотрим два VBR-видеофильма “Звездные войны”: первый — JPEG и второй — MPEG.

Выбирая две различные схемы кодирования одного и того же фильма, можно отделить статистические характеристики, вводимые схемой кодирования, от тех, что изначально принадлежат самому фильму. Оба набора данных содержат 171 000 целых чисел, которые отражают размер кадра в битах для фильма и при скорости просмотра 24 кадра в секунду представляют собой почти двухчасовую видеопоследовательность.

Выборка из первых 1200 кадров JPEG-видеоданных и соответствующая выборке гистограмма видеоданных представлены на рис. 6.14а, б.

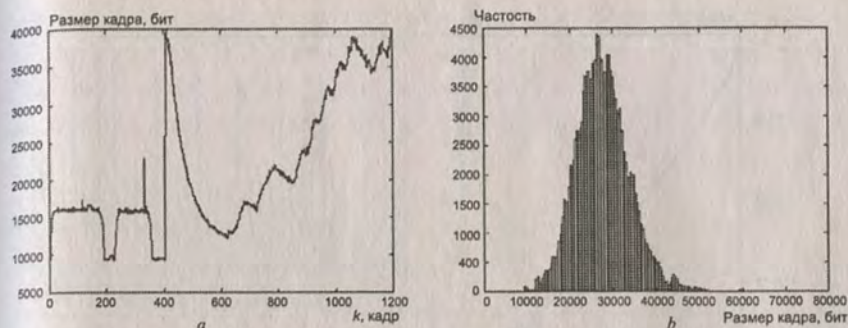


Рис. 6.14. Выборка из JPEG-видеоданных (а); гистограмма видеоданных (b)

Ниже приведены некоторые статистические характеристики трассы.

$$N = 171000; \sigma = 6254,2; \nabla = 8622;$$

$$m = 27791,2; \Delta = 78459,$$

где N — размер выборки; m — среднее значение; σ — среднее квадратичное отклонение; ∇ — минимальное значение выборки; Δ — максимальное значение выборки.

Корреляционная функция JPEG-данных приведена на рис. 6.15.

Аналогичная выборка из 1200 кадров и гистограмма MPEG-видеоданных приведены на рис. 6.16 а, б.

Ниже приведены статистические характеристики MPEG-видеоданных:

$$N = 171000; \sigma = 18164,95; \nabla = 476;$$

$$m = 15599,18; \Delta = 185267.$$

Из корреляционной функции на рис. 6.17а видно, что MPEG-кодирование вводит строгую периодичность, которой нет в JPEG-данных (рис. 6.15).

Для устранения периодичности можно сгруппировать MPEG-данные в блоки по 12 кадров, называемых группой изображений (GOP). Корреляционная функция GOP-данных приведена на рис. 6.1.7, б.

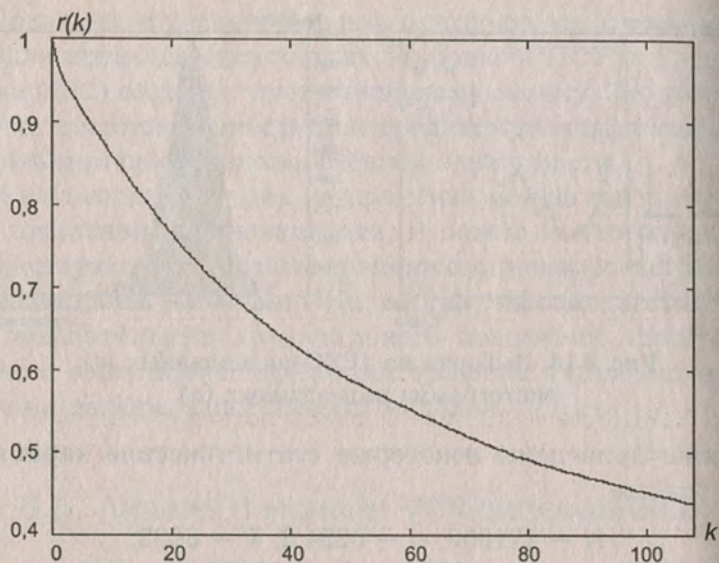


Рис. 6.15. Первые 108 точек корреляционной функции JPEG-данных

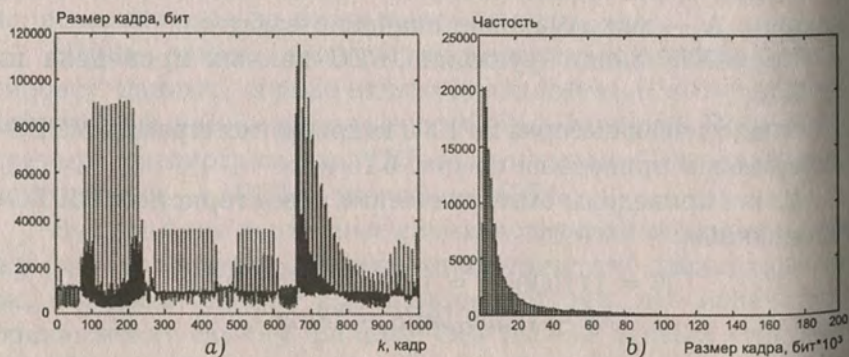


Рис. 6.16. Выборка MPEG-видеоданных (a); гистограмма видеоданных (b)

Основные статистические характеристики сгруппированных MPEG-данных имеют вид:

$$N = 14250; \sigma = 5514,3; \nabla = 6458; m = 14959,4; \Delta = 78720.$$

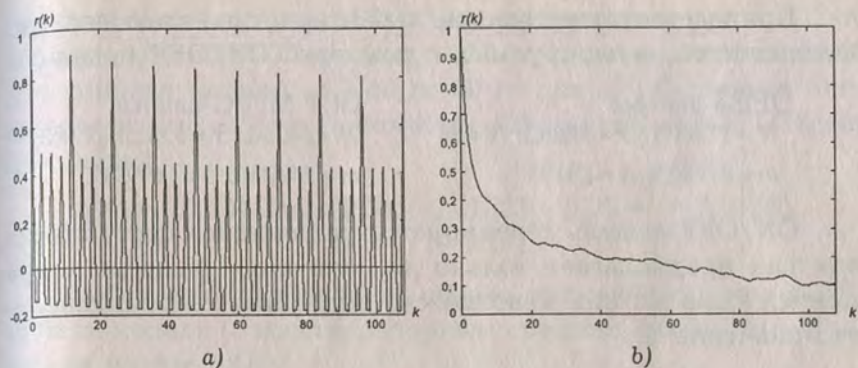


Рис. 6.17. Первые 108 точек корреляционной функции MPEG данных: *a* — до группировки; *b* — после группировки

Остановимся на построении моделей видеоданных, с помощью которых можно генерировать видеопоследовательности с различным характером корреляционных функций.

ON/OFF-модель видеопоследовательностей

Пусть ON/OFF-модель генерирует кадры $X(t)$, каждый размером p бит или 0 бит в соответствии с выходным значением генератора случайных чисел и пороговым значением a , т. е.

$$X[l] = \begin{cases} p & \text{если } RNG[l] \geq a \\ 0 & \text{если } RNG[l] < a. \end{cases} \quad (6.13)$$

Статистические характеристики кадров равны $M(X) = ap$ и $\sigma^2(X) = a(1 - a)p^2$.

Параметры модели a и p для JPEG, полученные в [41], равны $a = 0,952$ и $p = 29198$. Поскольку MPEG-данные являются слишком пульсирующими, чтобы описывать их ON/OFF-моделью, данная модель применялась лишь к GOP-данным. Вычисления показали, что параметры модели равны $a = 0,880$ и $p = 16999$.

Приведем статистические характеристики видеопоследовательностей, генерируемых с помощью ON/OFF-модели:

JPEG-данные

$N = 171000$; $\sigma = 6269,3$; $\nabla = 0$;

$m = 27783,3$; $\Delta = 29198$.

GOP MPEG-данные

$N = 14250$; $\sigma = 5582,3$; $\nabla = 0$;

$m = 14909,0$; $\Delta = 16999,0$.

ON/OFF-модель генерирует ограниченные поступления, так как предполагает только два значения размера кадра, причем одно из них невозможно встретить на практике, — это значение 0.

Авторегрессионные модели

Если видеоданные содержат ДВЗ, то они будут обладать и кратковременной, и долговременной корреляционными структурами. Следовательно, представляет интерес разработка модели с такими характеристиками, например, авторегрессионной интегральной модели скользящего среднего (ARIMA). В рассматриваемом случае можно упростить модель ARIMA, выбирая MA-порядок 0, оставляя только AR-часть. Причина в том, что оценка параметров ARMA-процесса привлекает решение системы из нелинейных уравнений. Используя AR-процесс, можно точно определить порядок модели и те значения AR-компонент, которые лучше всего подходят для данных.

В результате AR(p)-процесс может быть записан как

$$x[i] = - \sum_{k=1}^p a[k]x[i-k] + u[i],$$

где $\{u[i]\}$ — некоторый независимый источник гауссовского шума.

Чтобы оценить порядок модели, можно использовать информационный критерий (AIC), определяемый как $AIC(k) = N \ln \hat{p}_k + 2k$, где \hat{p}_k — оценка дисперсии шума для AR-модели k -го порядка; N — число подбираемых точек.

Оценка AIC показывает, что, например, когда объем видеоданных составляет 500 точек (среднее значение обнуля-

ется до подбора), для JPEG данных получаем модель 3-го порядка, а для MPEG-данных — 12-го порядка. Значения AIC для порядка модели от 1-го до 30-го для JPEG-данных приведены на рис. 4.83, а. Параметры AR-модели для JPEG-данных равны:

$$a[1] = -0,564635; a[2] = -0,319230; a[3] = -0,110245; \\ \sigma^2 = -1983281.$$

Выборка, полученная на основе применения данной AR-модели модели (с восстановленным средним значением), приведена на рис. 6.18б.

Для подбора ДВЗ можно заменить независимый и одинаково распределенный шум на фрактальный гауссовский шум (ФГШ) с единичной дисперсией и $H = 0,80$, создавая таким образом FARIMA-модель. В результате можно получить два набора данных — с $H = 0,5$ и с $H = 0,80$.

Статистические характеристики для случая для JPEG-данных

без ДВЗ

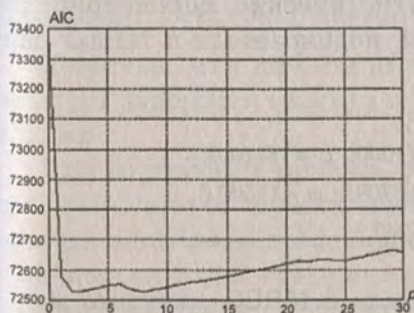
$$N = 171000; \sigma = 6324,6; \nabla = 6,26;$$

$$m = 27989,7; \Delta = 66965.$$

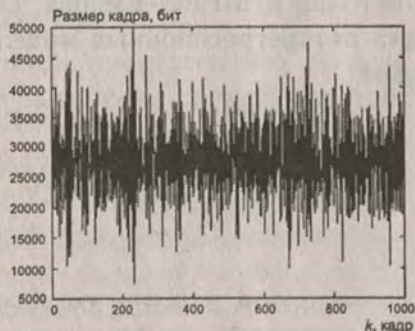
с ДВЗ

$$N = 171000; \sigma = 6522,7; \nabla = 2037,5;$$

$$m = 27789,8; \Delta = 48042,1.$$



а)



б)

Рис. 6.18. Оценка AIC [41]:

а — для модели от 1-го до 30-го порядков; б — выборка из JPEG-данных, сгенерированных при помощи AR-модели, в которой использовались н. о. р. $N(0; 1)$ инновации

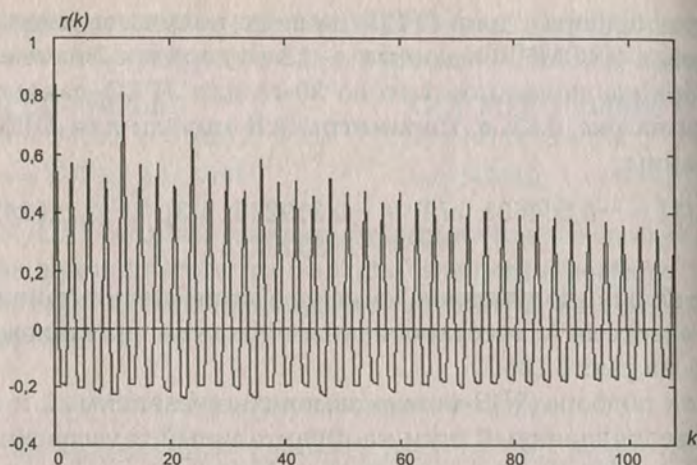


Рис. 6.19. Корреляционная функция для AR-модели, подобранной к MPEG-данным

Аналогичные модели также были построены и для MPEG-данных. Корреляционная функция AR-модели приведена на рис. 6.19. Большие компоненты, кратные 12 задержкам, все еще различимы, но уже не так сильно выделяются, как для настоящих MPEG-данных. Статистические характеристики для авторегрессионных моделей, подбираемых к MPEG-данным:

без ДВЗ	с ДВЗ
$N = 171000; \sigma = 18161,1;$	$N = 171000; \sigma = 18164,0;$
$\nabla = -69844,6; m = 16520,4;$	$\nabla = -66709,8; m = 15600,0;$
$\Delta = 97929,7.$	$\Delta = 95042,3.$

Заметим, что при согласовании с MPEG-данными моделью генерируются отрицательные значения трафика. Поэтому для получения реального потока трафика эти значения должны быть усечены.

Используя [1], можно сгенерировать дискретный поток поступлений (в этом случае бит/кадр) в соответствии с ал-

горитмом $A[i] = m + kX_H[i]$ где m — среднее значение, k — некоторая масштабная константа; $X_H[i]$, $i = 0, 1, 2, \dots$, — процесс ФГШ-типа. Чтобы оценить значение k , необходимо:

- 1) найти размер, среднее значение, дисперсию и показатель Херста для каждого набора данных VBR-видео;
- 2) сгенерировать ФГШ-трассу с показателем Херста и длиной в соответствии с этими JPEG-данными;
- 3) вычислить дисперсию ФГШ-трассы и использовать ее и дисперсию видеоданных.

В результате получены модели со следующими параметрами:

для JPEG-данных	для MPEG-данных
$N = 171000$; $m = 27791$;	$N = 171000$; $m = 18697,7$;
$H = 0,803$; $k = 6254$.	$H = 0,801$; $k = 18104,0$.

6.7. Аналитические методы оценки долговременной зависимости видеопоследовательности (6)

Одной из характеристик, используемых для описания самоподобного трафика, является нормированная функция корреляции $r(j) = M[(x(i) - \bar{x})(x(i+j) - \bar{x})] / \sigma_x^2$.

Известно, что для стационарных процессов она зависит только от задержки (шага) j и не зависит от времени дискретизации.

При построении кривых, отображающих степенную функцию корреляции, можно видеть, что зависимости для большинства видеопоследовательностей требуют разложения, которое нельзя выразить при помощи обычной экспоненциальной функции. На рис. 6.20а для примера показан график функции автокорреляции для GOP-последовательности фильма "Парк юрского периода". Из графика видно, что функция корреляции отличается от экспоненциальной и сильно уменьшается только после 1,4 мин (на рисунке этот участок соответствует $j > 20$). Для лучшего понимания влияния корреля-

ции видеопоследовательностей их функции корреляции могут быть представлены суммой геометрических прогрессий. Основная форма этого представления имеет следующий вид:

$$r(j) = \sum_{i=1}^M a_i r_i^j; \quad \sum_{i=1}^M a_i = 1, \quad (6.14)$$

где $a_i, i = 1, M$ — коэффициенты аппроксимации.

В частном случае аппроксимации двумя членами из (6.14) имеем:

$$r(j) = ar_1^j + (1-a)r_2^j. \quad (6.15)$$

Для аппроксимации функции корреляции последовательности “Парк юрского периода” (см. рис. 6.20а) коэффициенты разложения имеют следующие значения: $a = 0,716, r_1 = 0,854, r_2 = 0,997$.

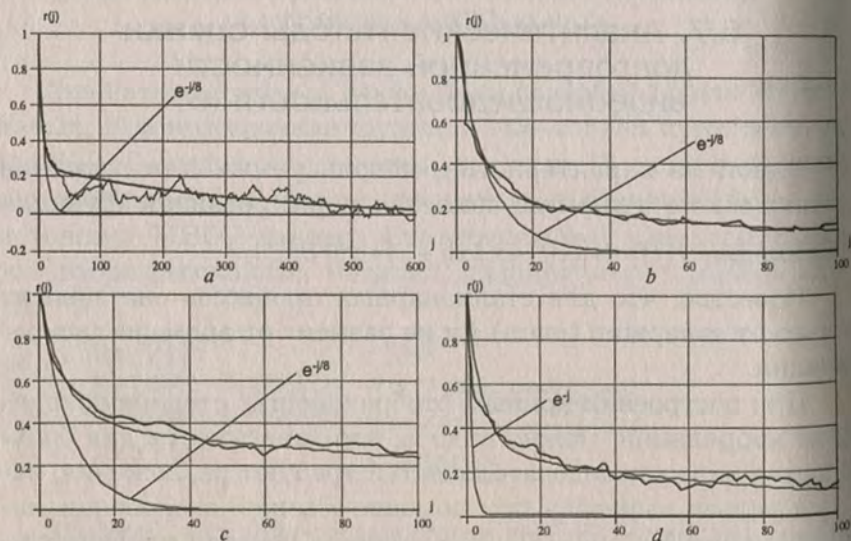


Рис. 6.20. График корреляционной функции для GOP-последовательности фильма “Парк юрского периода”:

a — $a = 0,716, r_1 = 0,854, r_2 = 0,997$; b — $a = 0,75, r_1 = 0,8, r_2 = 0,99$; c — $a = 0,6, r_1 = 0,86, r_2 = 0,995$; d — $a = 0,7, r_1 = 0,78, r_2 = 0,99$

Представление функции корреляции суммой (6.15) дает значительное преимущество при моделировании видеотрафика. Можно предположить, что если представлять коэффициент корреляции гиперболической функцией, то результирующая модель трафика будет соответствовать требованиям самоподобности. Примеры представления реальных корреляционных функций видеопоследовательностей представлены на рис. 6.20b, c, d.

Расчеты произведены по формуле $r(j) ar_1^j + (1 - a)r_2^j$.

Пусть x — случайная переменная, определенная на интервале $[0; \Gamma]$ с плотностью вероятности $w(x)$. Определим случайный процесс, представляющий собой последовательность отсчетов случайной переменной x , которые обозначим как $x_i (i = 0, 1, 2, 3, \dots)$. Значение i -го отсчета $x_i = \int_0^\infty x_i w(x_i) dx_i$, а корреляционная функция i -го и $(i + j)$ -го отсчетов определяется выражением $R(i, i + j) = \int_0^\infty \int_0^\infty x_i x_{i+j} w(x_i, x_{i+j}) dx_i dx_{i+j}$, где — совместная плотность распределения i -го и $(i + j)$ -го отсчетов.

Если предположить, что случайный процесс стационарен, то $x_i = x$ и $R(i, i) = x_i^2 = x^2$; тогда

$$\begin{aligned} R(i, j) &= (x_i - \bar{x} + \bar{x})(x_{i+j} - \bar{x} + \bar{x}) = \\ &= (x_i - \bar{x})(x_{i+j} - \bar{x}) + (\bar{x})^2 = r(j)\sigma_x^2 + \bar{x}^2, \end{aligned} \quad (6.16)$$

где $r(j)$ — коэффициент корреляции, определенный в общем случае выражением (6.14), или в случае двух членов соотношением (6.15), который дает разложение, имеющее гиперболическое затухание с ростом j .

Отметим, что $r(0) = 1 = a + (1 - a)$. Подставляя приближение в (6.16), получим

$$R(i, i + j) = ar_1^j \sigma_x^2 + (1 - a)r_2^j \sigma_x^2 + x^2 \quad (6.17)$$

Если x_j и x_{i+j} статистически независимы, то $x_i x_{i+j} = x_i x_{i+j} = x^2$, а $r(j) = \delta_{j0}$. Пусть среднее значение в k -й группе из N отсчетов равно $m_k = \frac{1}{N} \sum_{i \in J_k} x_i$, причем $J_k = \{kN, kN + 1, \dots, (k+1)N - 1\}$, тогда среднеквадратическое значение отсчета:

$$m^2 = \frac{1}{N^2} \sum_{i \in J_k} \sum_{l \in J_k} x_i x_l = \frac{1}{N^2} \sum_{i \in J_k} \sum_{l \in J_k} R(i, l) = \frac{1}{N^2} \sum_{i \in J_k} \sum_{l \in J_k} r(i-l) \sigma_x^2 + x^2. \quad (6.18)$$

Здесь были использованы представления коэффициента корреляции в зависимости от разности аргументов $r(i-l)$ и $l = i + j$. После преобразований (6.18) можно записать в виде

$$m_k^2 = \frac{2}{N^2} \sum_{j=0}^{N-1} (N-j) r(j) \sigma_x^2 - \frac{\sigma_x^2}{N} + x^2. \quad (6.19)$$

Выражение под знаком суммы в (6.19) является арифметико-геометрической прогрессией, для которой справедливо равенство:

$$\sum_{k=0}^{n-1} (a+kr) q^k = \frac{a - [a + (n-1)r] q^n}{1-q} + \frac{r q (1-q^{n-1})}{(1-q)^2}. \quad (6.20)$$

Воспользовавшись (6.20) и подставив значение $r(j) \sigma_x^2$ из (6.14), выполним непосредственное суммирование геометрической прогрессии в (6.19), получим

$$m_k^2 = \frac{1}{N} \sum_{i=1}^M \sigma_{x_i}^2 \left[\frac{1+r_i}{1-r_i} - \left(\frac{2r_i}{N} \right) \frac{1-r_i^N}{(1-r_i)^2} \right] + x^2,$$

где $\sigma_x^2 = a_i \sigma_{x_i}^2$, откуда

$$\frac{m_k^2}{\sigma_x^2} = \frac{1}{N} \sum_{i=1}^M a_i \left[\frac{1+r_i}{1-r_i} - \left(\frac{2r_i}{N} \right) \frac{1-r_i^N}{(1-r_i)^2} \right].$$

Для иллюстрации долговременной зависимости (продолжительной корреляции) в видеопоследовательностях построим график изменения дисперсии:

$$\frac{m_k^2}{\sigma_x^2} = \frac{a}{N} \left(\frac{1+r_1}{1-r_2} - \frac{2r_1}{N} \frac{1-r_1^N}{(1-r_2)^2} \right) + \frac{(1-a)}{N} \left(\frac{1+r_2}{1-r_2} - \frac{2r_2}{N} \frac{1-r_2^N}{(1-r_2)^2} \right), \quad (6.21)$$

где m_k^2 — дисперсия выборочного среднего всех мультиплексированных GOP-последовательностей. Переменные a_{ij} и r_i берутся из геометрической аппроксимации функции корреляции.

На рис. 6.21а нижняя кривая является графиком зависимости дисперсии от размера блока объединения. Верхняя кривая представляет модель, определяемую выражением (6.21) в случае $M = 2$.

Построенные таким образом графики изменения дисперсии для статистически независимого набора данных будут

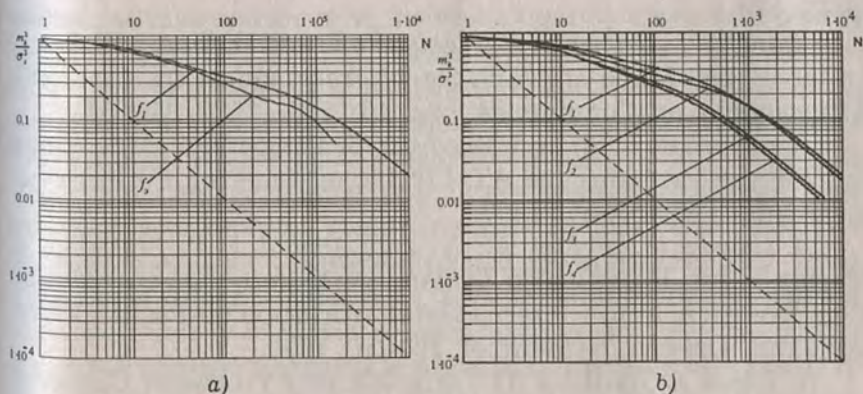


Рис. 6.21. Графики изменения дисперсии:

а — f_5 — экспериментально полученная кривая, f_1 — кривая полученная с помощью аппроксимации по формуле (6.21); б — f_1, f_2, f_3, f_4 кривые полученные с помощью аппроксимации по формуле (4.31), с использованием исходных значений для f_1 : $a = 0,6$; $r_1 = 0,86$ для f_2 : $a = 0,716$; $r_1 = 0,854$; $r_2 = 0,997$; для f_3 : $a = 0,7$; $r_1 = 0,78$; $r_2 = 0,99$; для f_4 : $a = 0,75$; $r_1 = 0,8$; $r_2 = 0,99$

иметь крутизну (наклон) -1 , в то время как долговременно зависимые наборы данных будут иметь наклон больше, чем -1 (см. рис. 6.21, а, б).

На основании представленных на рис. 6.21а графиков можно сделать вывод, что кривая дисперсии имеет линейную зависимость от среднего GOP после того, как этот размер достигает значения 1000. Значение корреляционной функции, соответствующее этой точке, равно 0,1. Для более детального анализа требуется изучение выборок, имеющих больший размер.

Как видно из графиков, данная модель достаточно хорошо представляет аппроксимирует экспериментальные данные.

Литература

1. Шелухин О. И., Тенякшев А. М., Осин А. В. Фрактальные процессы в телекоммуникациях. — М.: Радиотехника, 2003, 480 с.
2. Beran J., Sherman R., Taqqu R. R. and Willinger W. Long-range dependence in variable-bit-rate video traffic. IEEE Transactions on Communications, 43:1566—1579, 1995.
3. Krishnan K. R. and Meempat G. Long-range dependence in VBR video streams and ATM traffic engineering. Performance Evaluation, 30:45—56, 1997.
4. Grasse M., Frater M. R. and Arnold J. F. Statistics of Variable Bit Rate Video Coders with and without Motion Compensation. In 6th International Workshop on Packet Video, Portland, Oregon, September 26—27, 1994.
5. Beran J. Statistics for Long-Memory Processes. Chapman & Hall, New York, 1994.
6. Шелухин О. И., Симонян А. Г., Афонин М. Ю. Аналитические методы оценки долговременной зависимости видеопоследовательностей // Вестник МГУ. Серия: Радиоэлектроника и информатика. Тематический выпуск "Современные технологии в радио и телекоммуникациях". М., 2002. С. 6—11.

СТАНДАРТИЗАЦИЯ В ОБЛАСТИ КОМПЬЮТЕРНОЙ ВИДЕОКОНФЕРЕНЦСВЯЗИ (КВКС) (1)

Основную роль в работе по стандартизации в области КВКС играют Международный союз электросвязи (International Telecommunication Union, ITU), Международная организация по стандартизации (International Organization for Standardization, ISO) и Международный консорциум мультимедийных телеконференций (International Multimedia Teleconferencing Consortium, IMTC). Приведем краткую характеристику областей их деятельности.

Международный союз электросвязи (ITU) является межправительственной организацией, обеспечивающей сотрудничества государственных и частных организаций в процессе создания документов, регламентирующих ключевые характеристики телесервиса и телекоммуникационного оборудования. В марте 1993 г. в его составе были созданы ряд специализированных секторов, в частности, сектор ITU-T, который принял на себя функции, ранее принадлежавшие организации International Telegraph and Telephone Consultative Committee (СCITT). Основной целью деятельности ITU-T является разработка, принятие, распространение и внедрение рекомендаций в области телекоммуникаций. Рекомендации могут быть преобразованы в национальные или международные стандарты только после принятия решения соответствующими организациями, такими как ISO (см. ниже). Эти рекомендации разрабатываются усилиями более чем 4000 экспертов из разных стран, объединенных в 16 специализированных групп (Study Groups).

Результаты работы специализированных групп рассматриваются на Всемирной конференции ITU (World Telecom-

munication Standardization Conference, WTSC). Эта конференция собирается один раз в четыре года для того, чтобы определить основные направления деятельности секторов ITU и перечень вопросов, подлежащих рассмотрению в их специализированных группах. За эти годы были разработаны и приняты следующие важные рекомендации в области КВКС:

◆ Н.320 — видеотелефония в узкополосных цифровых сетях с коммутацией каналов (узкополосная ISDN, Switched 56 и т. п.);

◆ Н.321 — видеотелефония в широкополосных цифровых сетях с коммутацией каналов (широкополосная ISDN, АТМ технологии);

◆ Н.322 — видеотелефония в цифровых IP сетях с коммутацией пакетов и гарантированным качеством обслуживания (ЛВС);

◆ Н.323 — видеотелефония в цифровых IP сетях с коммутацией пакетов и негарантированным качеством обслуживания (ЛВС, Интернет);

◆ Н.324 — видеотелефония в аналоговых сетях с коммутацией каналов (аналоговые телефонные сети общего пользования);

◆ Т 120 — передача данных и организация многоточечных конференций.

Следует заметить, что в литературе по системам КВКС приведенные выше рекомендации часто называются также стандартами, хотя, строго говоря, они таковыми не являются. Как указывалось выше, их преобразование в национальные или международные стандарты возможно только после принятия решения соответствующими организациями типа ISO. Тем не менее подобная терминология широко используется.

Международная организация по стандартизации (ISO) объединяет национальные комитеты стандартизации более 100 стран мира. Каждая страна представлена в ISO одним национальным комитетом. Эта организация, созданная в 1947 г., является неправительственной по своему статусу. Решение о

ее учреждении была принято в 1946 году конференцией Международной электротехнической комиссии (International Electrotechnical Commission, IECJ), делегаты которой пришли к выводу о необходимости создания новой международной организации для координации разработок национальных стандартов в области техники. Основной миссией ISO является поддержка деятельности по разработке стандартов и связанных с этим процессом исследований во всех странах с целью создания лучших условий для международного обмена продуктами и услугами. Результатом деятельности ISO является подписание международных соглашений, которые становятся, по существу, международными стандартами в определенной области.

Текущая работа ISO имеет чрезвычайно распределенный характер и сосредоточена в 2700 технических комитетах подкомитетах и рабочих группах. Область деятельности этой федерации не ограничивается каким-либо определенным кругом проблем. Она включает в себя аспекты стандартизации во всех областях техники, кроме электротехники (последние регулируются Международной электротехнической комиссией). Деятельность по разработке стандартов в области информационных технологий ведется в рамках совместного комитета ISO/IEC, который курирует более 1300 стандартов. В их число входят и стандарты, определяющие функционирование аудио/видеосистем реального времени (алгоритмы кодирования аудио, графической, мультимедиа и гипермедиа информации, телекоммуникационный и информационный обмен между системами и т. д.). В поиске информации об этих стандартах и национальных дополнений к ним может оказать помощь специальная информационная сеть ISO (ISONET, <http://www.wssn.net/WSSN/>).

Международный консорциум мультимедийных телеконференций (IMTC) является некоммерческой межнациональной корпорацией, созданной в 1994 году на базе двух ранее существовавших международных объединений производителей

лей мультимедиа коммуникационного оборудования — Consortium for Audiographics Teleconferencing Standards, Inc. и Multimedia Communications Community of Interest, концентрировавших свое внимание на внедрении требований стандартов соответственно Т.120 и Н.320. Консорциум ИМТС объединил работу в этих направлениях, поскольку в современных системах КВКС функции, предусматриваемые указанными рекомендациями, являются одинаково важными.

Сами по себе рекомендации ITU-T и стандарты ISO еще не гарантируют полной совместимости продуктов, производимых различными фирмами, оставляя за фирмами право конкретных программно-аппаратных решений. Проблема совместимости различных систем КВКС является основным объектом внимания ИМТС.

В настоящее время в деятельности ИМТС участвуют более 140 компании и организаций, и его усилия направлены на:

- ◆ расширение использования мультимедиа телекоммуникационных систем, базирующихся на открытых стандартах, прежде всего на стандартах серии Н. xxx и Т.120, принятых ITU-T;
- ◆ распространение знаний в среде потенциальных потребителей о возможностях и преимуществах применения мультимедиа систем поддержки совместной работы, базирующихся на открытых стандартах. Эта работа проводится консорциумом посредством организации демонстраций, проведения тестирования совместимости оборудования различных производителей, публикаций соответствующих материалов;
- ◆ разработку методов реализации требований стандартов ITU-T, обеспечивающих совместимость продуктов разных производителей.

Деятельность ИМТС организована в рамках 6 рабочих групп, которые занимаются следующей тематикой:

- ◆ применение коммуникационных протоколов, предусмотренных рекомендациями Н.320, Н.323 и Т.120;
- ◆ совместимость программных средств, сетевого оборудования и сервисов в системах КВКС;

- ◆ взаимосвязь H.320-систем с системами КВКС, работающими в локальных сетях;
- ◆ разработка технических и организационных проблем проведения групповых многоточечных конференций;
- ◆ изучение и обобщение требований конечных пользователей систем КВКС;
- ◆ распространение информации о разработках ИМТС в прессе, в среде конечных пользователей и производителей мультимедиа коммуникационных систем.

Особенно полезными в практическом аспекте являются регулярно проводимые ИМТС тесты имеющихся на мировом рынке программно-аппаратных средств видеоконференцсвязи.

Перечисленные рекомендации серии H.xxx являются, по существу, “зонтичными” пакетами различных частных рекомендаций и сами по себе не определяют многие детали протоколов взаимодействия систем. Детальные требования к соответствующим системам и протоколам содержатся именно в указанных частных рекомендациях. При этом многие из этих частных рекомендаций фигурируют одновременно в различных “зонтичных пакетах”, являясь связующими звеньями, обеспечивающими решение проблемы совместимости различных систем КВКС.

Наибольшее распространение в настоящее время получили системы КВКС, работающие в соответствии с рекомендациями (стандартами) H.320, H.323, H.324 при широком использовании рекомендаций Т 120. Рассмотрим кратко содержание этих рекомендаций.

Рекомендации H.320

Эти рекомендации определяют набор требований по обеспечению совместимости по аудио- и видеоразличных систем видеоконференцсвязи в ISDN и других коммутируемых изохронных цифровых сетях, обеспечивающих скорости передачи до 1920 кбит/с. Указанные требования относятся к следующим службам:

- ◆ узкополосный видеотелефон ($p \times 64$ кбит/с);

- ◆ широкополосный видефон (применительно к широкополосной ISDN);
- ◆ узкополосные видеоконференции ($p \times 64$ кбит/с);
- ◆ широкополосные видеоконференции (применительно к широкополосной ISDN);
- ◆ аудиографические конференции;
- ◆ цифровая телефония;
- ◆ охранные видеосистемы.

Как указывалось выше, рекомендации H.320 являются зонтичным пакетом частных рекомендаций (рис. П. 1.1), каждая из которых включает детальные сведения о требованиях к соответствующим терминалам и протоколам. Как видно из рис. П. 1.1, рекомендации H.320 предполагают объединение и передачу трех типов информации: видео, аудио и данных. Соответствующее видео- и аудиооборудование включает видеокамеры и мониторы, микрофоны и громкоговорители. Передача данных предполагает реализацию режимов виртуальных рабочих досок, обмена файлами, окон обмена текстами, вводимыми в клавиатуры компьютера и т. д. В качестве главных составляющих рекомендации H.320 включают:

- ◆ рекомендации H.261, определяющие методы кодирования и декодирования видеоинформации;
- ◆ рекомендации G.711, G.722, G.728, определяющие методы кодирования и декодирования аудиоинформации;
- ◆ рекомендации серии T (T.120 и др.), определяющие взаимодействие программно-аппаратных средств при обмене данными. При этом серия T.120 определяет как часть стандартов H.320, связанную с аудиографикой, так и представляет собой самостоятельную группу стандартов;
- ◆ рекомендации H.221, определяющие методы мультиплексирования и демультиплексирования потоков аудио-, видеоданных и управления;
- ◆ рекомендации H.230, H.242 и H.243, определяющие сигнализацию и управление и контроль. В соответствии с рекомендациями серии 1.4.00 по D-каналу ISDN осуществляется сигнализация, как, например, вызов при наборе ISDN номера абонента и др.

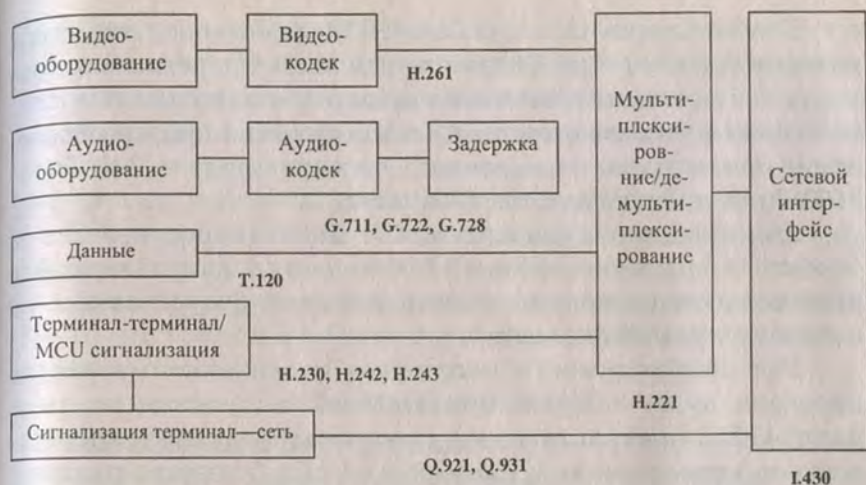


Рис. П. 1.1. Рекомендации ITU-T H.320

Рекомендации H.261 определяют вид кодирования и процедуру декодирования видеосигнала в реальном масштабе времени и механизм исправления ошибок при передаче видеoinформации. Принятым форматом видеосигнала является Common Intermediate Format (CIF), как некоторое компромиссное решение для стран, использующих различные видеостандарты (PAL, SECAM, NTSC и пр.). В дополнение к этому разработан стандарт с более низким разрешением — Quarter CIF (QCIF).

Рекомендации H.261 не определяют конкретную процедуру кодирования, оставляя производителю право выбора в рамках стандарта самого видеосигнале.

Определено два механизма обнаружения и исправления ошибок, каждый из которых основан на использовании дополнительных контрольных символов в кадре. В реальной системе возможен выбор любого из этих методов.

С 1997 г. набор возможных видеокодеков был расширен. Так, наряду с H.261-кодеком, рекомендации H.320 определили возможность использования кодеков стандартов H.262 и H.263 с повышенной в сравнении с H.261, эффективностью кодирования и расширенным набором форматов изображения.

Рекомендации G.711, G.722, G.728 определяют вид кодирования аудиосигнала. Обязательным является режим по стандарту G.711, обеспечивающий стандартное качество телефонной связи с жированием по А- и мю-законам при преобразовании аналогового телефонного сигнала (полоса 300 Гц — 3400 Гц) в цифровой поток 64 кбит/с.

Кодирование аудиосигнала в низкоскоростной поток 16 кбит/с по рекомендации G.728 является факультативным режимом, позволяющим отвести больший ресурс канала на передачу видеoinформации.

При необходимости получения повышенного качества передачи аудио возможно использование режима по стандарту G.722 (факультативно), предусматривающего преобразование аудиосигнала с полосой до 7 кГц (качество радиовещания) в цифровой поток до 64 кбит/с.

Сеанс связи всегда начинается в режиме использования G.711 с целью обеспечения возможности связи с любыми цифровыми телефонными аппаратами, после чего возможен переход на другие режимы, рассмотренные выше. При этом предусматривается дополнительная задержка аудиопотока относительно видеопотока для их синхронизации в связи с тем, что компрессия/декомпрессия видео приводит к большей задержке, чем компрессия/декомпрессия аудио.

Рекомендации H.221 определяют структуру кадра в диапазоне скоростей от 64 кбит/с до 1920 кбит/с. Внутри этого диапазона строго определены значения промежуточных скоростей:

- ♦ от 64 кбит/с до 384 кбит/с, кратные 64 кбит/с (соответствующие n В-каналам, где n от 1 до 6);

- ♦ от 384 кбит/с до 1920 кбит/с, кратные 384 кбит/с.

Рекомендации определяют механизм объединения каналов и позволяют устранить различие в задержках при передаче информации по различным В-каналам.

Рекомендации H.230 определяют процедуру синхронизации кадров во всем диапазоне используемых скоростей передачи.

Рекомендации H.242 определяют процедуру установления связи между передающим и принимающим терминалами,

позволяя передающему терминалу определить возможности принимающего. Это позволяет исключить применение режимов, которые не поддерживаются хотя бы одним из терминалов. При этом принимающий терминал может получать информацию о структуре графика и, как следствие, идентифицировать различные компоненты (аудио-, видеоданные).

Рекомендации H.231, 243 определяют процедуры взаимодействия между тремя и более терминалами и протоколы работы системы управления многоточечными соединениями MCU (Multi point Control Unif) для групповых видеоконференций.

Рекомендации H.323

Эти рекомендации определяют набор требований по обеспечению совместимости различных мультимедийным систем в сетях с пакетной коммутацией IP и IPX на базе Ethernet, Fast Ethernet и Token Ring, не обладающих возможностью гарантии требуемого уровня качества обслуживания QoS.

Рекомендации определяют интерфейс между локальной и иными (территориальными, Интернет) сетями. Ими также определяются процедуры взаимодействия терминалов как в двухточечном, так и многоточечном режимах видеоконференций.

Этот стандарт позволяет внедрять мультимедиа коммуникационные приложения без изменения существующей сетевой инфраструктуры, обладая механизмом компенсации высокой неизохронности потока IP-пакетов. При этом H.323 предусматривает возможность управления уровнем загруженности сети графиком конференц-приложений, что гарантирует сосуществование этих приложений с другими сетевыми сервисами.

Рекомендации H.323 тесно интегрированы с другими рекомендациями ITU серии H.32x и определяют способы взаимодействия с терминалами, соответствующими этим стандартам. В частности, важным свойством H.323-систем является их способность устанавливать соединения с H.320- и H.324-терминалами. В общую структуру H.323 входит также и семейство стандартов T.120.

Эти рекомендации поддержаны практически всеми ведущими производителями сетевых коммуникационных приложений, включая Intel, Netscape, PictureTel, Microsoft и т. д. В качестве главных составляющих рекомендации H.323 включают:

- ◆ рекомендации H.261 и H.263, определяющие методы кодирования и декодирования видеоинформации;

- ◆ рекомендации G.711, G.722, G.723, G.728 и G.729, определяющие методы кодирования и декодирования аудиоинформации;

- ◆ рекомендации H.225, определяющие методы мультиплексирования и демультимплексирования потоков аудио-, видеоданных и управления, а также контроль вызова, доступа, регистрации участников и определения их текущего состояния (модуль Registration/Admission/Status — RAS);

- ◆ рекомендации H.245, определяющие сигнализацию, управление и контроль, в том числе для многоточечных конференций с использованием модуля MCU;

- ◆ рекомендации серии T (T.120 и др.), определяющие взаимодействие программно-аппаратных средств при обмене данными.

Рекомендации H.324

Настоящие рекомендации распространяются на системы КВКС, функционирующие в обычных телефонных сетях общего пользования, H.324-терминалы могут обмениваться речевыми сообщениями, подвижным изображениями, данными в их любой комбинации в реальном времени как в симплексном, так и дуплексном режимах. Рекомендации определяют также возможность многоточечных конференций с помощью модуля управления групповыми конференциями (MCU). Требования к MCU и другим нетерминальным устройствам рекомендациями H.324 не определяются.

Терминал, соответствующий H.324, может быть интегрирован с РС или рабочей станцией, или реализован в виде отдельного устройства (видеофона).

Рекомендации H.324 определяют также возможности взаимодействия с другими терминалами семейства H.33x (в ISDN и IP-сетях).

В качестве главных составляющих рекомендации H.324 включают:

- ◆ рекомендации H.261 и H.263, определяющие методы кодирования и декодирования видеоинформации;
- ◆ рекомендации G.723, определяющие методы кодирования и декодирования аудиоинформации;
- ◆ рекомендации H.223, определяющие работу блока мультиплексирования отдельных медиапотоков (аудио-, видеоданные) в единый битовый поток для передачи его к другому терминалу и демультимплексирования принятого битового потока;
- ◆ рекомендации H.224, определяющие протокол управления устройствами в реальном времени (например, управление удаленной камерой в соответствии с рекомендацией H.281);
- ◆ рекомендации H.231 и H.245, определяющие блок протоколов сигнализации, управления и контроля, в том числе для многоточечных конференций с использованием модуля MCU;
- ◆ рекомендации серии T (T.120 и др.), определяющие взаимодействие программно-аппаратных средств при обмене данными (в частности, базисный набор протоколов доступа к аналоговым телефонным сетям общего пользования, определенный рекомендациями T.123);
- ◆ рекомендации, определяющие использование телефонного модема [стандарт V.34).

Рекомендации T.120

Рекомендации этого семейства охватывают все аспекты протоколов взаимодействия от транспортного до прикладного уровня и определяют правило построения основных приложений, позволяющих существенно расширить возможности КВКС. По существу, именно эти рекомендации и определяют основное отличие КВКС от собственно систем двухточечного аудио- и видеообщения. Важным при этом является обеспечение взаимодействия приложений, работающих в различных сетевых

средах, разделенных коммуникационными устройствами, включая аналоговые телефонные сети общего пользования, N-ISDN, В-ISDN, TCP/IP. При этом в одной многоточечной конференции одновременно могут участвовать приложения, работающие и на разных сетях, и на различных компьютерных платформах (PC, мощные рабочие станции). Стандарт T.120 включает следующие рекомендации (рис. П. 1.2):

- ◆ T.121 — обобщенный шаблон приложений;
- ◆ T.122/T.125 — служба многоточечных соединений;
- ◆ T.123 — стек протоколов транспортного уровня;
- ◆ T.124 — унифицированное управление телеконференциями;
- ◆ T.126/T. SI — прикладной протокол передачи неподвижных изображений;
- ◆ T.127/T. MBFT — протокол многоточечной передачи двоичных файлов.

Кроме того, дополняющим семейство T.120 является протокол T.AVC, реализующий ряд функций управления во время конференций.

Рекомендации T.121 обеспечивают совместное использование приложений, реализующих как стандартные, так и нестандартные протоколы. Эта задача решается путем зада-

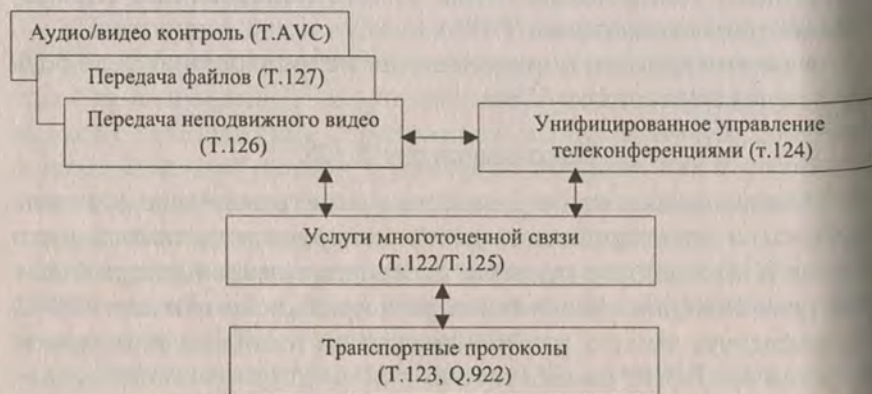


Рис. П. 1.2. Рекомендации семейства T.120

ния унифицированного шаблона приложения T.GAT (Generic Application Template), включающего модуль управления ресурсами ARM (Application Resource Manager), инвариантный к типу приложения. В то же время специфические функции приложения (например, пересылка файлов, обмен неподвижными изображениями и др.) обеспечиваются модулем сервиса приложений ASE (Application Service Element).

Рекомендации T.122/T.125 определяют службу многоточечной связи MCS (Multipoint "unications Service), позволяющую любому терминалу во время конференции устанавливать связь с другим терминалом. В рамках одного сеанса связи поддерживается одновременно до 64 тысяч соединений, выполняющих множество различных функций используемых приложений. При этом можно обеспечивать как общедоступные функции, так и функции с ограниченным для определенных терминалов доступом (фоновый обмен файлами, доступ к наборам данных и т. п.).

Рекомендации T.123 представляют собой стек протоколов транспортного уровня, обеспечивающих вышерасположенным уровням модели OSI инвариантность к типу сети передачи данных (телефонная сеть общего пользования, компьютерные сети, N-ISDN, B-1SDN). Сам стек протоколов изменяется в зависимости от типа сети. Так, для сетей ISDN синхронная передача пакетов по протоколу HDLC обеспечивается протоколом Q.922, а например, для аналоговой телефонной сети обеспечивается асинхронная передача данных по протоколу типа RS-232.

Рекомендации T.124 определяют модуль унифицированного управления конференциями GCC (Generic Conference Control), ответственный за установление и разрыв соединений через службу многоточечной связи MCS, управление модулем MCU, а также собственно менеджмент самой конференции (например, поддерживая функции ведущего конференции).

Рекомендации T.126 определяют протокол совместной работы пользователей с неподвижными изображениями и ан-

нотациями к ним T.SI (Still Image) для двух и более терминалов. Аннотации содержат набор функций, обеспечивающий совместное использование графической информации распределенные терминалами, включая такое важнейшее приложение, как виртуальная рабочая доска (White Board). При этом может быть обеспечен обмен графической информацией между приложениями, выполняющимися на разных компьютерных платформах (PC, MAC, UNIX).

Рекомендации T.127 определяют протокол как двухточечной, так и многоточечной передачи двоичных файлов T.MBF (Multipoint Binary File Protocol). Во время конференции любой файл может быть передан на любое число терминалов, участвующих в конференции. При этом предоставляется возможность выбора алгоритма сжатия пересылаемых файлов.

Протокол T.AVC (Audio Video Control) не входит непосредственно в рекомендации T 120. В его функции входит поддержка работы терминалов H.320 при удаленном управлении камерами, коммутации в блоке MCU и ряд других функций управления, связанных с передачей аудио- и видеoinформации.

Рекомендации T.120 обеспечивают все необходимые процедуры управления конференцией при одновременной передаче как голоса, так и изображений участников. При этом речевое общение участников мало отличается от режима обычных звуковых телеконференций, когда каждый участник может слышать все проходящие переговоры и в любой момент включаться в дискуссию. Предусмотрены также возможности управления дискуссией и при необходимости ограничения доступа к ней. В то же время технология использования видеоизображений участников отличается от технологии звукового общения. За редким исключением, примером которого является система Enhanced CU-SeeMe с одновременным изображением на экране большого числа участников, обычно на экран монитора для контроля выводится собственное изображение пользователя терминала (local video) и изображение только одного из участников конференции.

который в данный момент является “активным”, т. е. говорящим. При вступлении в разговор другого участника изображение немедленно переключается на него.

Таким образом, как показывает анализ современных программно-аппаратных средств, предназначенных для поддержки коллективной работы в компьютерных сетях, КВКС может рассматриваться как одно из таких средств с наиболее развитым уровнем мультимедийного сервиса, ориентированного на интерактивное общение в реальном времени. Значительная часть программно-аппаратных средств КВКС рассчитано на применение в сетях с пакетной коммутацией (локальные и корпоративные сети, Интернет) и соответствует рекомендациям серии Н.323 ITU-T. Однако в условиях высокой загрузки сетевых каналов, характерной именно для Интернет, сильно проявляется неизохронность пакетного трафика. Последнее, в свою очередь, приводит к снижению качества передачи мультимедиа данных, которые по своей природе очень чувствительны к задержкам. Поэтому реализуемое качество передачи звука и изображения часто не позволяет рекомендовать такие системы КВКС для профессионального использования. Естественным выходом в таких условиях может явиться, с одной стороны, снижение интенсивности генерации мультимедиа трафика и, с другой стороны, резервирование требуемых сетевых ресурсов для конкретных приложений и управления передачей потоков данных реального времени. Примером реализации первого пути решения указанной проблемы является групповая адресация и маршрутизация (IP-multicasting). Второй путь связан с разработкой соответствующих транспортных протоколов типа RTP, RTCP, RSVP, поддерживающих приложения реального времени в IP-сетях и учитывающих в том числе технологии IP-multicasting.

Применительно к ЛВС наряду с приведенными выше путями решения проблемы неизохронности пакетного трафика рассматривается также и возможность модификации самой

технологической основы традиционных ЛВС как, например, при использовании технологии isoEthernet (рекомендация Н.322 ITU-T). Тем не менее такое решение применимо лишь в ограниченных пределах сетей предприятий и до настоящего времени не получило широкого распространения.

Другой возможный путь, применимый как в ЛВС, так и в рамках Интернет, связан с использованием технологий АТМ на базе В-ISDN. Однако последнее решение связано с использованием чрезвычайно дорогостоящих каналов.

Применение выделенных цифровых каналов с высокой пропускной способностью для проведения разовых видеоконференций позволяет успешно решить проблему "представительских" конференций, однако широкое использование такой практики связано с высокими затратами на аренду каналов. Кроме того, в этом случае осуществляются лишь соединения типа "точка—точка" или многоточечные соединения с ограниченным количеством участников, и оказывается невозможным осуществить вызов произвольного абонента, как это делается в сетях с коммутацией каналов.

Использование сетей с коммутацией каналов привлекательно простотой обеспечения изохронности трафика мультимедиа, а также возможностью вызова любого абонента сети. Системы КВКС, предназначенные для использования в таких сетях, могут применяться даже в телефонных сетях общего пользования (системы, удовлетворяющие рекомендациям серии Н.324 ITU-T), однако ограниченность пропускной способности телефонных каналов не позволяет добиться качества, требуемого для профессионального использования КВКС. В то же время имеется большое число систем КВКС, которые предназначены для использования в сетях ISDN.

Использование ISDN для организации компьютерных видеоконференций как одна из форм предоставляемых услуг становится все более распространенным. Качество КВКС оказывается вполне приемлемым для профессиональной работы. Следует учесть, что при этом возможно соединение с любым

абонентом мировой сети ISDN аналогично соединениям в телефонных сетях общего пользования. При соединении "точка—точка" происходит конференция двух партнеров (или соединение двух студий с любым числом участников в каждой из них), а при использовании сервера MCU может осуществляться и многоточечная конференция. В режиме повышенного качества, применяя расширенный доступ ISDN PRA, можно использовать большие конференц-студии с проекцированием изображения с экрана компьютера на большой экран размером до 3—4 м по диагонали.

Одним из важнейших условий использования систем КВКС как средства коллективной работы удаленных пользователей является совместимость программно-аппаратных средств таких систем. Благодаря усилиям, прежде всего, таких международных организаций, как ITU, ISO и IMTC, проведена большая работа по стандартизации в сфере КВКС. Наибольшее распространение в настоящее время получили системы КВКС, работающие в соответствии с рекомендациями H.320, H.323, H.324 при широком использовании рекомендаций T.120. Важным результатом такой стандартизации является возможность использования аудио и видео кодеков, соответствующих единым и взаимосогласованным стандартам, что исключает необходимость перекодировки при обмене аудио- и видеoinформацией, заметно снижая время задержки при транспортировке данных между терминалами. Единый стандарт обмена данными, соответствующий рекомендациям T.120, охватывает все аспекты протоколов взаимодействия от транспортного до прикладного уровня и делает возможным применение мощных инструментов совместной работы, виртуальная рабочая доска, совместное использование приложений и пр.

Чрезвычайно важным результатом деятельности указанных организаций является обеспечение взаимодействия приложений КВКС, работающих в различных сетевых средах, разделенных коммуникационными устройствами, включая аналоговые телефонные сети общего пользования, N-ISDN,

V-ISDN, TCP/IP. При этом в одной многоточечной конференции одновременно могут участвовать приложения, работающие в разных сетях и на различных компьютерных платформах (PC, мощные рабочие станции). Таким образом, к настоящему времени можно считать принципиально решенным вопрос о совместимости систем КВКС, функционирующих в различных сетях.

Литература

1. Синепол В. С., Цикин И. А. Системы компьютерной видеоконференцсвязи. — М.: ООО "Мобильные коммуникации", 1999.

РЕКОМЕНДАЦИИ Н.323. КВКС В СЕТЯХ С НЕГАРАНТИРОВАННЫМ УРОВНЕМ КАЧЕСТВА ОБСЛУЖИВАНИЯ (1)

Общая характеристика рекомендаций Н.323

Рекомендации Н.323 являются интегральным документом, определяющим основные требования к оборудованию и программному обеспечению компьютерных систем, предоставляющих возможность обмена аудио- и видеоинформацией и данными в IP-сетях, включая и Интернет. Реализация положений этих рекомендаций обеспечивает совместимость продуктов разных производителей. Этот стандарт является ключевым для систем, предназначенных для работы в локальных сетях.

Н.323 определяет семейство рекомендаций, принятых ИТУ и устанавливающих основные требования к мультимедийным коммуникационным системам в сетях, не обладающих возможностью гарантии требуемого уровня качества обслуживания QoS. Такие сети сегодня являются преобладающими и включают в себя сети пакетной коммутации IP и IPX на базе Ethernet, Fast Ethernet и Token Ring. В число приложений, для которых Н.323 является фактором, обеспечивающим возможность взаимодействия, можно отнести следующие:

- ◆ компьютерная видеоконференцсвязь;
- ◆ Интернет-телефония и видеотелефония;
- ◆ системы для совместной работы распределенных групп;
- ◆ сетевые игры;
- ◆ системы дистанционного консультирования и технической поддержки.

Эти рекомендации актуальны как для специализированных систем, так и для систем, построенных на основе персо-

нального компьютера. Они определяют порядок взаимодействия коммуникационных систем в режиме “точка—точка” и в режиме многоточечной конференции. Рекомендациями также определяется интерфейс между локальной и территориальной сетями.

Рекомендации H.323 тесно интегрированы с другими рекомендациями ITU серии H.3xx и определяют способы взаимодействия с терминалами, соответствующими этим стандартам. Отметим характерные черты настоящих рекомендаций:

- ◆ H.323 ориентированы на системы, функционирующие в широко распространенных IP-сетях. Определяя механизмы компенсации высокой неизохронности потока IP-пакетов, этот стандарт позволяет внедрять мультимедиа коммуникационные приложения без изменения существующей сетевой инфраструктуры. Это тем более важно в условиях, когда ЛВС становятся все более производительными, освоена технология Fast Ethernet (100 Мбит/сек) и активно развивается технология гигабитного Ethernet;

- ◆ определяют механизм взаимодействия на уровнях “устройство — устройство” и “приложение — приложение”, рекомендации обеспечивают взаимодействие продуктов и приложений различных производителей;

- ◆ стандарт определяет способы взаимодействия мультимедиа терминалов территориальными сетями;

- ◆ H.323 предусматривает возможность управления уровнем загрузки сети графиком конференцприложений, что гарантирует сосуществование этих приложений с другими сетевыми сервисами;

- ◆ рекомендации H.323 не содержат элементов, специфических для определенного типа сети, и тем самым обеспечивают возможность совместной работы приложений, развернутых в разных сетях. Отсутствуют в стандарте и элементы, специфические для конкретных операционных систем или компьютерных платформ;

- ◆ H.323 определяет механизм многоточечных конференций без использования специализированного мультимедиа-

моста MCU. Стандарт предусматривает возможность реализации функций групповых конференций в различных компонентах H.323-совместимых систем. Кроме этого, в стандарте заложена рекомендация поддержки механизма групповой адресации (multicasting), что позволяет более экономично использовать пропускную способность сетевых коммуникационных каналов;

◆ рекомендации предусматривают возможность совместной работы систем с различными функциональными возможностями. Так, например, терминал, способный работать лишь в режиме аудиоконференции, может успешно взаимодействовать со станцией с полным набором функциональных возможностей ВКС. Равным образом в групповой конференции H.323-мультимедиа терминал может работать с терминалом, обладающим только возможностью документконференции, и одновременно успешно использовать свои аудио/видеовозможности для работы с другим терминалом;

◆ важным свойством H.323-систем является их способность устанавливать соединения с H.320- и H.324-терминалами;

◆ рекомендации H.323 поддержаны практически всеми ведущими производителями сетевых коммуникационных приложений, включая Intel, Netscape, PictureTel, Microsoft и т. д. Это дает надежду обеспечения действительной совместимости их продуктов и способствует широкому внедрению систем мультимедиа в компьютерных сетях.

Архитектура стандарта H.323

Рекомендации H.323 определяют технические требования для аудио- и видеокommunikационных служб в ЛВС с пакетной коммутацией. В общую структуру H.323 входит и семейство стандартов документконференций T.120. В сферу влияния рекомендаций H.323 не входит ЛВС как таковая, однако элементы системы ВКС, необходимые для взаимодействия с сетями коммутации каналов, вошли в состав этих рекомендаций. Рис. П. 2.1 дает общее представление о составе стандарта и взаимосвязи его компонентов.

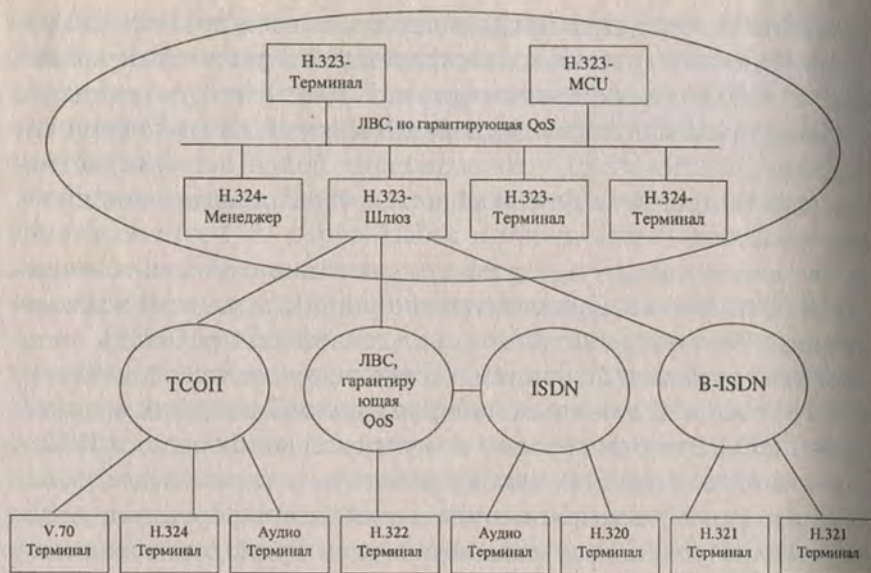


Рис. П. 2.1. Схема взаимодействия H.323-системы с другими стандартными H.3xx-системами

H.323 определяет 4 основных компонента для сетевой конференц-системы: терминал, шлюз, конференц-менеджер и модуль групповых конференций (MCU).

Терминал. Терминалом далее называется всякое оконечное сетевое устройство, которое обеспечивает возможность двунаправленной коммуникации в реальном времени. На рис. П. 2.1 представлены возможные компоненты H.323-терминала. Все H.323-терминалы должны обеспечивать аудиокоммуникации. Прием/передача видеoinформации и режим документ-конференции являются необязательными (опциональными) функциями. Рекомендации H.323 определяют режимы работы, необходимые для взаимодействия различных аудио, видео и (или) документ-терминалов.

Все H.323-терминалы для оценки возможностей канала связи должны поддерживать функции управления логическим каналом, определенные рекомендациями H.245. Посколь-

ку стандарт H.245 является чрезвычайно громоздким, так как описывает многочисленные возможные варианты реализации функций управления, то несколько производителей оборудования для конференц-связи объединились и разработали более компактную версию этого стандарта — рекомендации H.245.1. (H.245 profile 1).

H.323-терминалы должны также обязательно поддерживать упрощенную версию протокола Q.931 для сигнализации и вызова, содержать модуль, называемый RAS (Registration/Admission/Status), обеспечивающий функции контроля доступа, регистрации участников и определение их текущего состояния, а также иметь возможность реализации протокола RTP/RTCP для передачи аудио- и видеоинформации по сетям с коммутацией пакетов.

Опционально терминал может поддерживать видеообмен, документ-конференции по протоколам серии T.120 и выполнять функции многовходового моста для организации групповых конференций (MCU).

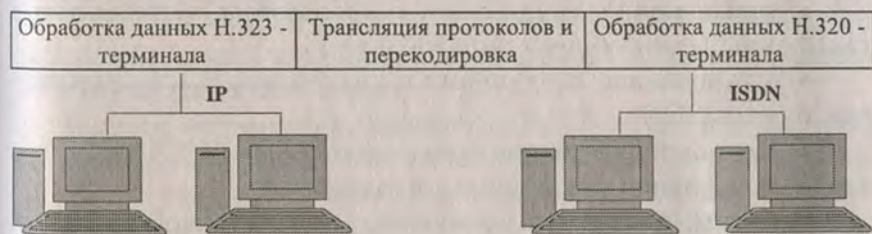


Рис. П. 2.2. Структура H.323-системы

Шлюз. Шлюз является необязательным элементом в H.323-системе. Это устройство обеспечивает целый ряд сервисов, включая обмен информацией между H.323-терминалом и терминалами, определяемыми другими ITU-стандартами серии H для электронных конференций. Реализация такой функции требует трансляцию формата кадров и преобразование коммуникационных процедур (например, H.225.0 в H.221 и H.245 в H.242 при обмене информацией H.323-терминала с H.320-

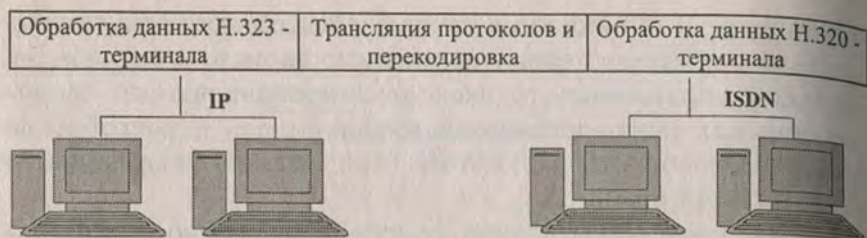


Рис. П. 2.3. Функции шлюза между ISDN и ЛВС-терминалами

терминалом). Кроме этого, шлюз выполняет перекодировку аудио- и видеопотоков, а также обеспечивает функцию установления и разрыва соединения между ЛВС и сетями с коммутацией каналов. Рисунок П. 2.3 иллюстрирует функции шлюза между H.323- и H.320-терминалами.

В общем случае задачей шлюза является взаимное отражение свойств и характеристик оконечного конференц-оборудования ЛВС и терминалов сетей с коммутацией каналов. Основные функции шлюза следующие:

- ◆ установление соединения с аналоговым терминалом в телефонной сети общего пользования;
- ◆ установление соединения с удаленным H.320-терминалом в сети ISDN;
- ◆ установление соединения с удаленным H.324-терминалом в телефонной сети общего пользования.

Заметим, что шлюз не нужен, если нет необходимости установления соединений из ЛВС с удаленными терминалами в сетях с коммутацией каналов, ибо H.323-терминалы имеют механизм установления непосредственных соединений. Для этого используются процедуры протоколов H.245 и Q.931.

С соответствующими транскодерами H.323-шлюз обеспечивает взаимодействие H.323-терминала с терминалами, определенными рекомендациями H.320, H.321 (АТМ ЛВС), H.322 (ЛВС с гарантированным QoS) и V.70.

Большое число возможных функций шлюза остается нестандартизованными. Например, количество H.323-терминалов, которые могут одновременно устанавливать соединения

через шлюз, количество соединений с сетями коммутации каналов, функции перекодировки аудио/видеопотоков, возможности поддержки многоточечных конференций и т. д. Все эти характеристики могут отличаться в продуктах разных производителей. Тем не менее стандартизованных функций шлюза достаточно, чтобы рекомендации H.323 сыграли роль средства объединения стандартных систем конференц-связи, ориентированных на работу в различных сетях.

Конференц-менеджер. Конференц-менеджер в H.323-системе реализует две важные функции, позволяющие сохранить связность корпоративной сети передачи данных для конференц-связи. Первое из них — это функция трансляции ЛВС-имен терминалов и шлюзов в IP- или IPX-адреса в соответствии со Спецификацией RAS. Вторая функция — это управление полосой пропускания канала, выделенной для конференц-связи. Эта процедура также определена в спецификации RAS. Управление полосой пропускания осуществляется посредством оценки уровня графика, порожденного всеми реализуемыми одновременно конференциями, и сопоставления его с установленным порогом. Удовлетворение запроса на установление новой конференции возможно лишь при условии не нарушения этого порога.

Группа терминалов, шлюзов и MCU, управляемая одним конференц-менеджером, носит название H.323-зоны (рис. П. 2.4).

Хотя конференц-менеджер является логически отдельным устройством, в большинстве реальных H.323-систем он интегрируется с терминалом, шлюзом или MCU. Поскольку H.323-зона определяется соответствующим конференц-менеджером, то все компоненты H.323-системы должны обладать способностью отключения своих внутренних управляющих функций. Это, очевидно, необходимо для организации всех устройств H.323-системы в единую зону.

Напомним, что конференц-менеджер не является обязательным компонентом H.323-системы. Однако если он реализован, то терминалы должны использовать сервисы, предоставляемые им. Эти сервисы, как определено в RAS, являются обязательными и включают преобразование адресов,

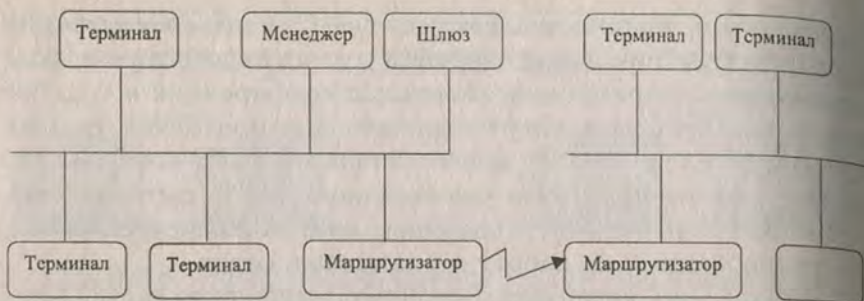


Рис. П. 2.4. Пример Н.323-зоны

контроль доступа, контроль полосы канала и управление зоной. Конференц-менеджер может участвовать и в организации многоточечных соединений, перенаправляя информацию контрольных каналов (протокол Н.245) на контроллер групповых соединений. Обязательные и опциональные функции конференц-менеджера и их определения приведены в табл. П. 2.1 и табл. П. 2.2.

Трансляция адресов. Трансляция ЛВС-имен в IP-адреса, используя таблицу, которая обновляется посредством регистрационных сообщений. Другие методы обновления содержания таблицы также допустимы.

Контроль доступа. Проверка прав доступа к ЛВС, использующая сообщения "Запрос доступа", "Подтверждение" и

Таблица П. 2.1

Трансляция адресов	Трансляция ЛВС-имен в IP-адреса, используя таблицу, которая обновляется посредством регистрационных сообщений. Другие методы обновления содержания таблицы также допустимы.
Контроль доступа	Проверка прав доступа к ЛВС, использующая сообщения "Запрос доступа", "Подтверждение" и "Отказ". Контроль доступа может быть нуль-функцией, при которой доступ не ограничивается.
Контроль полосы канала	Обрабатываются сообщения "Запрос полосы", "Подтверждение", "Отказ". Эта функция может базироваться на процедуре управления полосой. Контроль полосы может быть и нуль-функцией, при которой принимаются все запросы на изменение полосы.
Управление зоной	Конференц-менеджер реализует описанные выше функции для терминалов, шлюзов и MSU, которые зарегистрированы внутри зоны.

Таблица II. 2.2

Обработка сигнальной информации вызова	В двухточечной конференции конференц-менеджер может обрабатывать сигнальную информацию протокола Q.931. В противном случае он может направлять эту сигнальную информацию непосредственно взаимодействующим терминалам.
Опознание вызова	Конференц-менеджер может запретить вызовы от терминалов, базируясь на информации протокола Q.931. Критериями запрета могут быть: ограничение доступа к (от) определенному (определенного) терминалу, ограничение доступа к терминалу в определенный интервал времени и т.д. Эти критерии не входят в сферу определения H.323.
Управление пропускной способностью канала	Конференц-менеджер может не разрешить вызов от терминала, если требуемая пропускная способность канала не может быть реализована. Это функция остается активной и в течение сеанса связи, если терминалом запрашивается дополнительный ресурс.
Управление вызовами	Конференц-менеджер может формировать список текущих вызовов для того, чтобы определить, что вызываемый терминал занят, или для того, чтобы обеспечить информацией функцию управления пропускной способностью канала.

“Отказ”. Контроль доступа может быть нуль-функцией, при которой доступ не ограничивается.

Контроль полосы канала. Обрабатываются сообщения “Запрос полосы”, “Подтверждение”, “Отказ”. Эта функция может базироваться на процедуре управления полосой. Контроль полосы может быть и нуль-функцией, при которой принимаются все запросы на изменение полосы.

Управление зоной. Конференц-менеджер реализует описанные выше функции для терминалов, шлюзов и MCU, которые зарегистрированы внутри зоны.

Модуль управления многоточечными (групповыми) конференциями. Модуль управления многоточечными конференциями MCU обеспечивает установление взаимосвязи между тремя и более терминалами. В соответствии с рекомендациями H.323 в MCU как обязательный элемент входит групповой контроллер (Multi point Controller, MC) и, если необходимо, один

или более групповых процессоров (Multipoint Processor, MP). Групповой контроллер реализует определенную в рекомендациях H.245 функцию оценки общего объема аудио/видеотрафика, порождаемого всеми активными терминалами. Этот контроллер также индицирует многоадресные аудио/видеопотоки. Групповой контроллер не выполняет обработку медиа потоков. Эту задачу решает групповой процессор, который мультиплексирует, коммутрует и при необходимости обрабатывает битовые аудио/видеопотоки. Контроллер и процессор могут быть реализованы как самостоятельные компоненты или входить в состав других компонентов H.323-системы.

Обработка сигнальной информации вызова. В двухточечной конференции конференц-менеджер может обрабатывать сигнальную информацию протокола Q.931. В противном случае он может направлять эту сигнальную информацию непосредственно взаимодействующим терминалам.

Опознание вызова. Конференц-менеджер может запретить вызовы от терминалов, базируясь на информации протокола Q.931. Критериями запрета могут быть: ограничение доступа к (от) определенному (определенного) терминалу, ограничение доступа к терминалу в определенный интервал времени и т. д. Эти критерии не входят в сферу определения H.323.

Управление пропускной способностью канала. Конференц-менеджер может не разрешить вызов от терминала, если требуемая пропускная способность канала не может быть реализована. Это функция остается активной и в течение сеанса связи, если терминалом запрашивается дополнительный ресурс.

Управление вызовами. Конференц-менеджер может формировать список текущих вызовов для того, чтобы определить, что вызываемый терминал занят, или для того, чтобы обеспечить информацией функцию управления пропускной способностью канала.

Многоточечные конференции. В рекомендациях H.323 предусмотрены несколько вариантов реализации групповой

конференции (рис. П. 2.5 и рис. П. 2.6). Рекомендации определяют концепцию централизованной, децентрализованной и гибридной конфигурации многоточечной конференции.

Централизованная схема. Ее реализация требует наличия специализированного мостового устройства — модуля управления MCU. В такой схеме все терминалы адресуют свои потоки данных и аудио/видеоинформацию к MCU в режиме соединения “точка—точка”. Групповой контроллер MC централизованно управляет конфигурацией, используя для этого

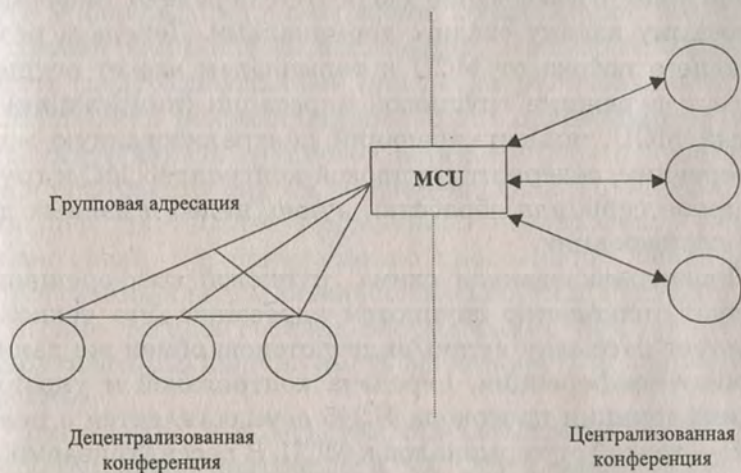


Рис. П. 2.5. Модель групповых конференций

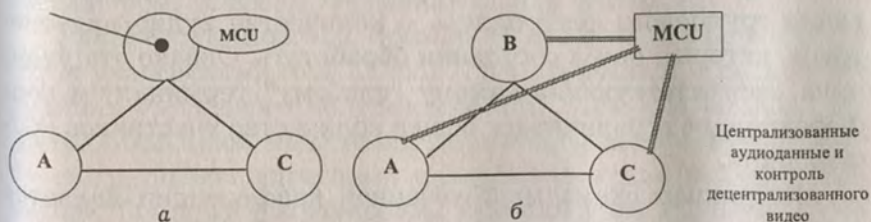


Рис. П. 2.6. Децентрализованная (а) и гибридная (б) схема групповых конференций

функции контроля и управления рекомендацией H.245, посредством которых, в частности, передается к МС и информация о характеристиках каждого терминала.

Групповой процессор МР осуществляет мультиплексирование аудиопотоков, тиражирование потоков данных, коммутацию или мультиплексирование видеопотоков и результирующий поток направляет на терминалы, участвующие в конференции. При необходимости групповой процессор осуществляет перекодировку аудио/видео последовательностей и выполняет согласование скоростей передачи информации по каждому каналу связи с терминалами. Передача результирующего потока от МСУ к терминалам может осуществляться и в режиме групповой адресации (multicasting). Типичный МСУ, поддерживающий централизованную модель конференции, содержит групповой контроллер МС и групповые процессоры для обработки аудио, видео и данных документ-конференции.

Децентрализованная схема групповой конференции использует технологию групповой адресации. Эта технология реализует рассылку аудио/видеопотоков; обмен же данными документ-конференции, передача контрольной и управляющей информации протокола H.245 осуществляется в режиме "точка—точка" от терминалов к МСУ. В рассматриваемой модели групповой конференции принимающие терминалы осуществляют обработку поступающих аудио/видеопотоков от всех терминалов, участвующих в конференции. Терминалы используют контрольную информацию H.245 для информирования группового контроллера о количестве аудио/видеопотоков, которые они в состоянии обработать. Однако эта величина, соответствующая самому "слабому" терминалу в конференции, не ограничивает общее количество участников конференции.

Гибридная схема многоточечной конференции является комбинацией рассмотренных выше схем. В ней информация контрольных каналов H.245, данные документ-конференции и аудио- (или видео-) поток направляются на модуль управ-

ления групповой конференции (MCU). Оставшийся видео- (или аудио-) поток рассылается непосредственно на терминалы в режиме групповой адресации.

Смешанная схема. Рекомендации H.323 определяют и такую схему организации групповой конференции, в которой одна группа терминалов функционирует по централизованной Схеме, а другая их часть реализует децентрализованную схему, при этом модуль управления MCU выполняет функции соединительного моста. Терминалы в этом случае не нуждаются в информации о смешанной конференции; каждый из них должен лишь определить режим, в котором он должен работать (централизованный или децентрализованный).

Поддержка рекомендациями H.323 одновременно режимов двухточечной и групповой адресации обеспечивает гибкость в реализации систем конференцсвязи. Групповая адресация позволяет более эффективно использовать емкость каналов связи, но одновременно предъявляет более высокие требования к вычислительной мощности клиентских станций. Кроме того, и сетевое оборудование (коммутаторы и маршрутизаторы) должно иметь возможность поддержки этого режима.

Архитектура H.323-систем предполагает наличие лишь одного группового контроллера MC. Хотя теоретически число участников конференции может быть очень большим, однако их реальное число не должно превышать 10—20. Групповой контроллер может быть интегрирован с конференц-менеджером, шлюзом, терминалом или MCU.

Рассмотрим простой пример групповой конференции с тремя участниками (рис. П. 2.5). Пусть терминал В выполняет функции группового контроллера MC. При реализации децентрализованной модели все терминалы используют технологию групповой адресации для обмена аудио/видеоинформацией. Групповые процессоры каждого узла выполняют обработку входящей аудио/видеоинформации и отображение ее на терминале участника конференции. Такая схема минимизирует потребность в специализированном сетевом оборудовании.

довании. Однако сеть должна быть способной поддерживать режим групповой адресации. Отметим, что расширение числа участвующих в такой конференции также весьма затруднительно.

Наличие отдельного МСУ позволяет реализовать централизованную схему групповой конференции. При этом аудиопоток, данные и контрольная информация передаются через модуль управления, а обмен видеоданными осуществляется в режиме групповой адресации, что уменьшает требуемую пропускную способность каналов связи. Одним из преимуществ централизованной модели является то, что все H.323-терминалы могут работать в режиме "точка—точка" с МСУ. Последний и свой выходной поток может поправлять к терминалам в таком же режиме, что не требует никаких специальных усовершенствований сетевой инфраструктуры. Вместе с тем если сеть поддерживает режим групповой адресации, то выход МСУ может направляться к терминалам по этой экономящей полосе пропускания технологии.

Групповая конференция, в которой участвуют ЛВС-терминалы и терминалы из внешних сетей, реализуется более эффективно, если модуль управления МСУ интегрирован со шлюзом.

Организация обмена информацией в рекомендациях H.323. Общие положения

Рекомендации H.323 определяют коммуникационный поток как смесь аудио-, видео- и контрольных сигналов. Обязательными компонентами коммуникации являются аудиосигналы, сигналы установления соединения протокола Q.931, контрольные сигналы протокола H.245 и процедур RAS. Видеоданные, данные документ-конференции и традиционные сетевые приложения являются опциональными компонентами коммуникационного обмена. Для кодирования аудио- и видеoinформации может использоваться целый ряд алгоритмов. Информация о приемлемом в данной сессии алгоритме, необходимая для согласования кодера передатчика и декоде-

ра приемника, передается в потоке сигнализации H.245. Стандарт H.323 допускает и асимметричную схему, когда терминал использует разные алгоритмы для кодирования и декодирования медиаинформации. H.323-терминал может посылать (принимать) более одного аудио/видеопотока.

Контрольные каналы. Функции контроля являются критически важными для работы H.323-терминала. Они включают в себя:

- ◆ сигнализацию установления соединения;
- ◆ сигнализацию возможности обмена информацией;
- ◆ сигнализацию вызова команд и индикацию их выполнения;
- ◆ сообщения открытия и описание логического канала.

Все процедуры контроля образуют в своей совокупности уровень контроля в стеке процедур терминала. Все аудио/видеосигналы и контрольные сообщения передаются через этот уровень; последний форматирует битовый поток и передает его на сетевой интерфейс. Входной поток проходит обратную цепь преобразований.

Уровень контроля, выполняя логическую фрагментацию выходного потока, производит нумерацию кадров последовательности, контроль и коррекцию ошибок. Уровень контроля функционирует в соответствии с протоколами Q.931, H.245 и RTP/RTCP. Этот уровень содержит три канала:

- ◆ канал контроля протокола H.245;
- ◆ канал сигнализации протокола Q.931;
- ◆ канал процедур RAS.

Канал контроля протокола H.245 является надежным (с коррекцией ошибок) каналом. Он передает все контрольные сообщения, необходимые для работы любого компонента системы, включая сообщения о возможности обмена информацией, сообщения установления (открытия) и закрытия логического канала, сообщения контроля графика, общие команды и индикаторы. Обмен сообщениями о возможностях терминала является одним из фундаментальных положений ITU-рекомендаций. Рекомендации H.245 определяют механизмы отдельного описания возможностей терминала по приему и передаче сообщений.

Для установления соединения между двумя терминалами или терминала и MSU используется канал сигнализации протокола Q.931.

Канал RAS-сигнализации обеспечивает регистрацию и проверку прав доступа участников конференции, передает команды изменения полосы пропускания сетевого канала связи, выделенной для конференц-связи, обеспечивает реализацию процедур обмена данными между терминалом и конференц-менеджером. Заметим, что канал RAS-сигнализации формируется только при наличии в системе конференц-менеджера, что не является обязательным.

Аудиокомпонента коммуникации. Аудиокомпонента коммуникационного потока представляет собой оцифрованное речевое сообщение, подвергнутое процедуре сжатия в соответствии с определенным алгоритмом. Рекомендации H.323 определяют возможность использования любого, одобренного ИТУ, алгоритма компрессии. H.323-терминалы обязаны поддерживать стандарт цифровой обработки речевых сообщений G.711. Поддержка остальных стандартов серии G.7xx является опциональной. Заметим, что стандарты этой серии отражают разную степень компромисса "качество — полоса канала — вычислительная мощность — задержка". Алгоритм сжатия, определенный в стандарте G.711 (PCM), ориентирован на коммуникационный канал с пропускной способностью 64 кбит/с, что является приемлемой величиной для терминалов в ЛВС. Вместе с тем потребность связи с терминалами в других, более узкополосных, сетях ведет к необходимости использовать алгоритмы стандарта G.723, реализующие более высокую степень сжатия аудиоинформации. В настоящее время кодек стандарта G.723 становится доминирующим в H.323-терминалах

Видеоконпонента коммуникации. Возможность приема/передачи изображения является факультативной функцией терминала. Если эта функция реализована, то поддержка требований рекомендации H.261 является обязательной, а рекомендации H.263 — опциональной. Как отмечалось выше,

стандарт H.261 является общим для всех систем, соответствующих рекомендациям ИТУ серии H.3xx. Процедуры кодирования, предусмотренные рекомендациями H.263, являются расширением рекомендации H.261 и обеспечивают более высокое качество подвижного видео для низкоскоростных каналов. Видеоинформация, преобразованная в битовую последовательность, передается со скоростью, не превышающей порога, определенного в момент соединения.

Рекомендации H.263 определяют 5 размеров изображения (табл. П. 2.3).

Возможность взаимодействия систем с кодеками H.261 и H.263 реализуется благодаря обязательности форматов QCIF.

Компонента документ-конференции. Документ-конференция является факультативной функцией H.323-систем. Будучи реализованной, она предоставляет возможности использования виртуальной аудиторной доски (White Board), совместной работы с приложениями и обеспечивает пересыпку файлов в фоновом режиме.

Рекомендации H.323 определяют возможность документ-конференции на основе реализации функций стандартов семейства T.120, который описывает документ-конференцию в режимах "точка—точка" и "многоточка", а также механизмы взаимодействия терминалов на сетевом, транспортном и верхних уровнях стека сетевых протоколов.

Функции рекомендаций T.120 в H.323-системе могут реализовываться как в программном обеспечении клиента, так и

Таблица П. 2.3

Формат и размер изображения в пикселах H.261 H.263

Формат	Размер изображения в пикселах	H.261	H.263
Sub-QCIF	128×96	опционально	обязательно
QCIF	176×144	обязательно	обязательно
CIF	352×288	опционально	опционально
4 CIF	704×576	не определено	опционально
16 CIF	1408×1152	не определено	опционально

в модуле управления групповыми конференциями MCU. В рекомендациях T.120 предусмотрена и возможность использования технологии групповой адресации, что может быть полезно для уменьшения величины трафика, порождаемого обменом документами в ходе конференций со многими участниками.

IP-сети и мультимедиа конференции

Стек протоколов H.323-систем представлен на рис. П. 2.7. Из него видно, что рекомендации H.323 предполагают использование как “надежных” (с коррекцией ошибок), так и “ненадежных” соединений. Контрольная информация передается по “надежному” каналу, поскольку сигналы управления и контроля должны быть приняты без ошибок и в том порядке, в каком они передавались. Аудио- и видеоинформация по своему характеру допускает прием с некоторым уровнем ошибок и поэтому, для ее транспортировки могут быть использованы более эффективные по пропускной способности “ненадежные” соединения.

“Надежная” передача сообщений использует каналы с установлением соединения. В стеке сетевых протоколов этот тип соединения обеспечивается протоколом TCP. “Надежный” канал гарантирует сборку пакетов на приемном конце в их

Данные			Контроль и сигнализация	Аудио	Видео
Не стандарт. приложения	T.126	T.127	H.245 Q.931	G.711 G.722 G.723 G.728 G.729	H.261 H.263
T.124			RAS	RTP RTCP	
T.122					
T.125					
T.123					
TCP			UDP		
IP					
LAN					

Рис. П. 2.7. Стек протоколов H.323-систем

исходной последовательности, исправление возможных ошибок и восстановление потерянных пакетов. Однако такой канал не обеспечивает изохронность потока и обладает меньшей пропускной способностью в сравнении с каналом без установления соединения. Рекомендации H.323 требуют использования “надежного” канала для передачи контрольных сигналов протокола H.245, для передачи данных документ-конференции (T.120) и сигнализации установления конференц-связи по протоколу Q.931.

Обстоятельное рассмотрение технологии работы приложений реального времени в сетях с коммутацией пакетов будет проведено в следующей главе. Здесь же дадим лишь краткий обзор средств, обеспечивающих передачу трафика реального времени в таких сетях. “Ненадежный” канал, обеспечиваемый протоколом UDP, используется для передачи аудио/видеопотоков и для транспортировки информации RAS-процедур. В многоточечных конференциях с множественными аудио /видеопотоками “ненадежные” каналы протокола UDP используют технологию групповой адресации и протокол RTF (Real-Time Protocol), разработанный инженерной группой Internet для управления потоками мультимедиа информации реального времени. В свою очередь, RTP работает поверх IP-групповой адресации (если она используется) и поверх UDP. Заголовок пакета UDP дополняется полем временной метки и порядковым номером пакета. При буферизации пакетов приемником эти поля позволяют исключить дублированные пакеты, восстановить требуемый порядок их следования, произвести синхронизацию потоков аудио и видеоинформации, что в конечном итоге дает возможность достижения эффекта непрерывности при воспроизведении, несмотря на вариации задержек в доставке каждого пакета.

Поскольку H.323-приложения базируются на RTP, они могут работать в виртуальной сети Mbone — надстройке Internet для мультимедиа приложений реального времени.

В группу протоколов, обеспечивающих работу H.323-систем в IP-сетях, входят также протоколы RTCP и RSVP. Про-

токол RTCP контролирует и управляет параметрами качества обслуживания, распространяет информацию об участниках конференции, периодически рассылает всем участникам информацию о качестве функционирования системы (частота кадров, уровень ошибок и т. д.).

Канал с характеристиками, необходимыми для проведения сессии конференц-связи, весьма трудно обеспечить в условиях Internet или большой корпоративной сети. Задача формирования и поддержания такого канала возлагается на протокол RSVP (Resource Reservation Protocol). Этот протокол определяет механизмы, позволяющие приемнику мультимедиаинформации запросить необходимую ему полосу пропускания у всех сетевых устройств на пути к передающему узлу и сообщить последнему свои требования к характеристикам потока, который он сможет принять. Хотя RSVP не является официально частью рекомендаций H.323, большинство современных конференц-систем для IP-сетей его поддерживают, поскольку резервирование полосы канала является критически важным условием реализации конференц-связи. Естественно, что этот протокол должны поддерживать все компоненты H.323-системы (терминалы, шлюзы, модули управления групповыми конференциями). Должны его поддерживать и все сетевые устройства — маршрутизаторы и коммутаторы.

Литература

1. Синепол В. С., Цикин И. А. Системы компьютерной видеоконференц-связи. Серия изданий "Связь и бизнес". — М.: ООО "Мобильные коммуникации", 1999.

РЕКОМЕНДАЦИИ Н.324.КВКС В НИЗКОСКОРОСТНЫХ СЕТЯХ (1)

Настоящие рекомендации определяют технические требования для коммуникационных мультимедиа-систем, функционирующих в обычных телефонных сетях общего пользования. Н.324-терминал обеспечивает связь с другим терминалом в реальном времени и дает возможность обмена речевыми сообщениями, подвижным изображением, данными в их любой комбинации. Между терминалами может поддерживаться как симплексный, так и дуплексный каналы. Рекомендации определяют возможность и многоточечных конференций с помощью модуля управления групповыми конференциями MCU. Требования к MCU и другим нетерминальным устройствам рекомендациями Н.324 не определяются.

Мультимедиа-терминал, соответствующий рекомендациям Н.324, может быть интегрирован с РС (рабочей станцией) или реализован в виде самостоятельного устройства (например, видеофона).

Рекомендации Н.324 определяют также возможности взаимодействия с Н.320-терминалами в ISDN и с Н.323-терминалами в IP-сетях.

Архитектура Н.324-системы.

Обобщенная схема мультимедиа Н.324-системы представлена на рис. П. 3.1. В нее входят:

- ◆ терминальное оборудование;
- ◆ телефонный модем;
- ◆ телефонная сеть;
- ◆ модуль управления групповыми конференциями;
- ◆ видео- и аудиоустройства;
- ◆ приложения по обработке данных, графики и т. п.

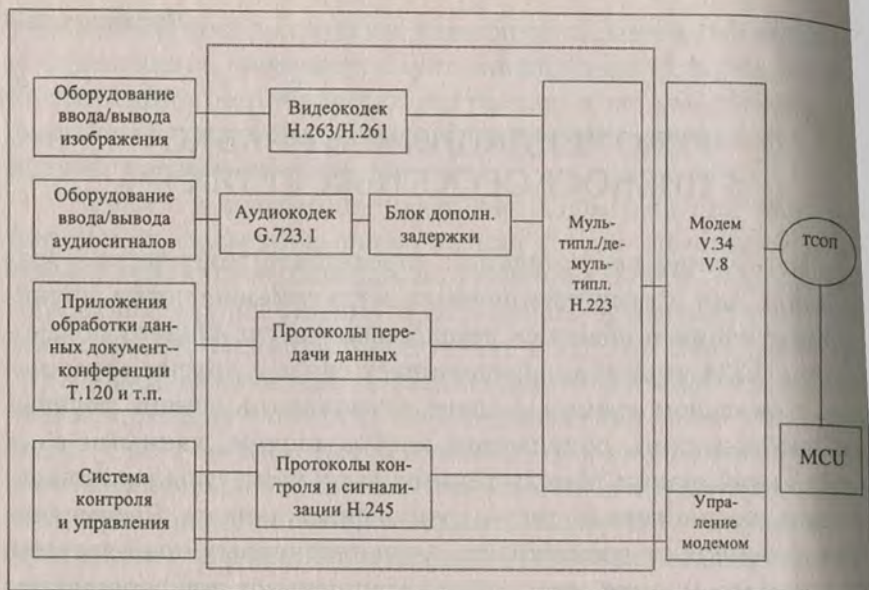


Рис. П. 3.1. Функциональная схема H.324 системы

Рекомендации не требуют обязательного наличия в системе всех перечисленных составляющих. Системы, находящиеся внутри пунктирного прямоугольника на рис. П. 3.1, являются предметом рекомендаций H.324. Остальные элементы приведенной схемы или определяются другими рекомендациями, или вообще не являются стандартизируемыми.

В число устройств, регламентируемых рекомендациями H.324, входят:

- ◆ видекодек (рекомендации H.261 или H.263) для цифровой обработки подвижных видеоданных;
- ◆ аудиокодек (рекомендации G.723.1), кодирующий сигнал от микрофона для передачи и декодирующий его для воспроизведения. В системе может быть предусмотрено устройство дополнительной задержки принятого аудиопотока, служащее для синхронизации аудио и видеорядов;
- ◆ блок протоколов обработки данных обеспечивает работу приложений, соответствующих рекомендациям T.120 —

виртуальной рабочей доски, передачи неподвижного изображения (графики), обмена файлами, доступа к базам данных, удаленного управления устройствами (камерой, микрофонами и т. п.);

♦ блок контрольных протоколов реализует требования рекомендаций H.245 и обеспечивает всю необходимую для работы терминала сигнализацию. Он используется для обмена сообщениями о возможностях терминала, вызова и индикации выполнения команд, передачи сообщений об открытии логического канала и описания его состава;

♦ блок мультиплексирования/демультиплексирования (рекомендации H.223) осуществляет преобразование отдельных медиапотоков (аудио, видео, данные) в единый битовый поток для передачи его к другому терминалу и осуществляет обратное преобразование входного (принятого) битового потока. Кроме этого, он выполняет логическое фрагментирование исходящих данных, нумерует кадры последовательности, детектирует и исправляет ошибки;

♦ телефонный модем (стандарт V.34) конвертирует цифровую последовательность, полученную от мультиплексора, в аналоговый сигнал с параметрами, необходимыми для передачи его по телефонной сети, и выполняет обратное преобразование входного аналогового сигнала.

Функциональные требования

Обязательные и опциональные элементы H.324-системы. В состав обязательных элементов системы входят V.34-модем, H.223-мультиплексор и блок H.245-контроля и сигнализации.

Терминалы, располагающие аудиовозможностями, должны поддерживать аудиокодеки стандарта G.723.1. Терминалы с возможностью приема/передачи подвижного изображения должны иметь видеокодеки, соответствующие рекомендациям H.261 и H.263. Реализация документ-конференции обязательно должна быть основанной на рекомендациях семейства T.120. Заметим, что не запрещено реализовывать до-

полнительно и кодеки других стандартов. Информация о наличии опциональных элементов системы передается между терминалами посредством специальных сообщений в соответствии с рекомендацией Н.245. Механизмы этих рекомендаций обеспечивают и согласование терминалов, участвующих в обмене информацией, в части используемых ими кодеков.

Информационные потоки

Мультимедиа информационные потоки включают:

- ◆ видеопоток, несущий подвижное цветное изображение. Доля пропускной способности канала, отводимая для видеопотока, может изменяться в соответствии с потребностями аудиопотока и графика данных документ-конференции;

- ◆ аудиопоток, являющийся потоком реального времени. Однако, как отмечено выше, он, как правило, подвергается дополнительной задержке для синхронизации с видеопотоком;

- ◆ поток данных документ-конференции, который может нести неподвижное изображение, документы, файлы, данные приложений пользователя и т. п.;

- ◆ поток управляющей информации, которой обмениваются терминалы на этапах установления соединения, конференции и разрыва соединения в соответствии с рекомендациями Н.245. В таблице П. 3.1. приведен пример распределения общей пропускной способности телефонной линии между перечисленными информационными потоками.

Мультиплексирование

Для транспортировки видео/аудиоданных и контрольной информации необходимо установление соответствующих логических каналов. Эти каналы формируются процедурами, определенными в рекомендациях Н.245. Логические каналы являются однонаправленными и независимыми в каждом из направлений передачи. Количество логических медиа каналов любого типа может быть произвольным, исключая канал контрольной информации, который должен быть один. Метод мультимедиа

Таблица П. 3.1

**Полная пропускная способность линии (кбит/с).
Логические каналы**

	Полная пропускная способность линии (кбит/с)	Логические каналы			
		аудио (кбит/с)	видео (кбит/с)	накладные расходы (кбит/с)	данные (кбит/с)
Распределение пропускной способности линии	9.6	4.8	4.3	0.5	сколько остается
	14.4	4.8	8.9	0.7	сколько остается
	28.8	6.8	20.6		
					сколько остается
Характеристика канала		Фиксированная скорость	Варьируемая скорость	Варьируемая скорость	Варьируемая скорость
Приоритет		высший	низший	высший	средний

типлексирования, используемый для объединения медиапотоков в единый битовый поток, должен соответствовать положениям рекомендаций H.223. H.223-мультиплексор (рис. П. 3.2) предполагает наличие уровня мультиплексирования, осуществляющего объединение логических каналов в единый битовый поток, и уровня адаптации, реализующего контроль ошибок и нумерацию кадров. На уровне мультиплексирования информация логического канала разбивается на пакеты, которые называются MUX-PDU (MUX-Packet Data Unit). Границы этих пакетов определяются специальными флагами, и их длина кратна целому числу байт (при необходимости пакет дополняется нулевыми битами для доведения его длины до целого числа байт).

Каждый пакет MUX-PDU содержит однобайтовый заголовок и переменное число информационных байт. В заголовок включается код мультиплексирования, определяющий принадлежность информационных байт пакета различным логическим каналам. Каждый такой пакет может иметь раз-

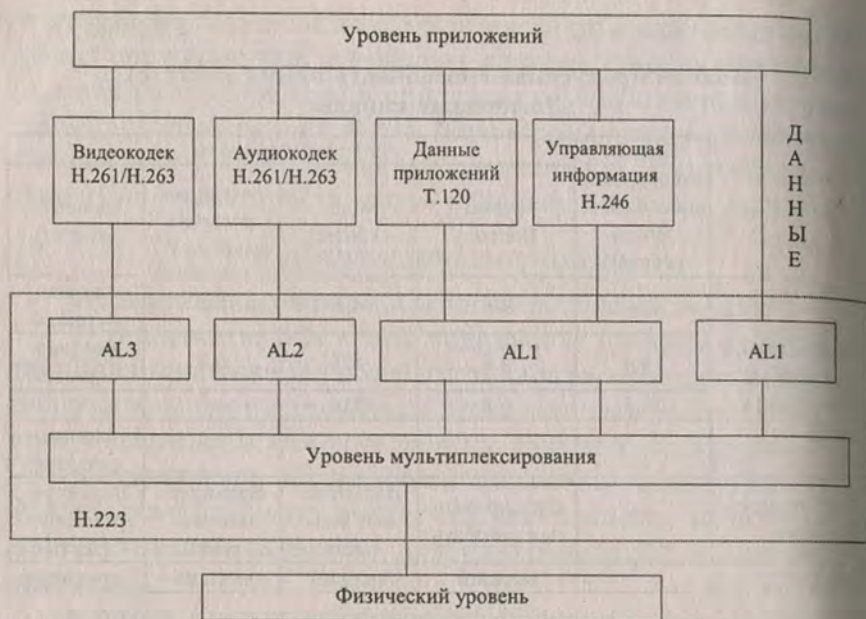


Рис. П.3.2. Протокольная структура H.323-мультиплексора

ный код мультиплексирования и, следовательно, содержать различное сочетание информационных каналов.

Каждому логическому каналу присваивается номер — целое число из диапазона 0—65535. Номер канала назначается произвольно, исключая контрольный канал, которому присваивается нулевой номер. Нумерация каналов необходима исключительно для установления соответствия логического канала с записью в таблице мультиплексирования.

Таблицы мультиплексирования для каждого направления передачи являются независимыми, и их содержание пересылается от передающего терминала к приемному посредством H.245-сообщения “Multiplex Entry Send”. Нулевая запись таблицы, соответствующая контрольному каналу, не пересылается. Она используется для начального обмена информацией о возможностях терминалов и передачи исходного состояния таблицы.

Н.324-терминал обязан отвечать на Н.245-запрос "Flow Control Command", который устанавливает ограничение на допустимую часть пропускной способности канала, отводимую одному или нескольким логическим каналам, или мультиплексу в целом. При ограничении пропускной способности одного или нескольких каналов другие, не имеющие ограничений, логические каналы могут автоматически увеличить свою производительность.

На уровне мультиплексирования реализуется только контроль поля CRC (Cyclic Redundancy Check) в заголовке пакета. Контроль и коррекция ошибок каждого логического канала проводится на уровне адаптации, и эта задача может решаться различными методами, в том числе и посредством повторной передачи ошибочных пакетов.

Уровень адаптации мультиплексора состоит из трех подуровней — AL1, AL2 и AL3. Подуровень AL1 предназначен для обработки информации, передаваемой с переменной скоростью, включая и неформатированные потоки байтов, которые интерпретируются как один пакет неопределенной длины. Подуровень AL2 обеспечивает аудиоканал и включает в пакет значение 8-битного поля CRC и порядковый номер пакета. Подуровень AL3 обслуживает видеопоток и обеспечивает возможность повторной передачи пакетов.

Логический блок информации, которой мультиплексор обменивается с кодеками, с протоколом обработки данных или контрольными протоколами обозначается AL-SDU (Adaptation Layer Service Data Unit — блок данных сервисного канала уровня адаптации).

Логические каналы, открываемые мультиплексором, могут быть сегментируемыми или несегментируемыми, что определяется в момент открытия канала. В общем случае сегментируемые каналы применяются для передачи информации с неравномерным характером трофика (контрольные сигналы, видео, данные), в то время как несегментируемые каналы используют для передачи аудиоинформации.

Приемник должен, используя механизмы протокола H.245, сообщать о своих возможностях обработки потоков, соответствующих различным уровням адаптации, а передатчик обязан информировать о своих уровнях адаптации и типах используемых каналов для каждого уровня адаптации.

Блок контрольных протоколов

Канал контроля устанавливается по схеме “из конца — в конец” и служит для передачи всей контрольной и управляющей информации, включая информацию о возможностях терминалов, об открытии и закрытии логических каналов, о содержимом таблицы мультиплексирования, об уровне трафика, а также общие команды и информационные сообщения (индикаторы). Как отмечалось выше, существует лишь один контрольный канал для каждого направления передачи, и он функционирует в соответствии с рекомендациями H.245. Контрольный канал должен быть открыт от момента установления цифрового соединения до его закрытия.

Рекомендации H.245 задают четыре класса сообщений: запросы, ответы, команды и индикаторы. Запрос требует определенного действия от приемника, включая и немедленную отсылку ответа. Команды также требуют определенных действий, но отсылка ответа не предполагается. Индикаторы — это информационные сообщения, и никаких действий они не предполагают. H.324-терминал обязан выполнять все определенные рекомендациями H.245 запросы и команды, а также должен безошибочно передавать индикаторы, отражающие его состояние. Кроме этого, терминал должен проводить синтаксический разбор всех H.245-сообщений группы “Multimedia System Control PDU”. При обращении к неподдерживаемым терминалом запросам, командам или ответам им должно отсылаться сообщение “Function Not Supported” (функция не поддерживается).

Обмен информацией о возможностях терминалов производится процедурами, описанными в рекомендациях H.245. Эта информация формируется отдельно для характеристики при-

емных и передающих способностей терминала и отражает его способность обработки входных потоков. Передающий терминал должен ограничить свои возможности до уровня, заданного характеристиками приемного терминала. Отсутствие информации о приемных возможностях терминала говорит о способности терминала только передавать информацию.

Информация о передающих возможностях терминала А необходима принимающему терминалу В для того, чтобы последний смог запросить наилучший для него (из возможных) режим передатчика терминала А. Отсутствие информации о передающих способностях терминала свидетельствует об отсутствии вариантов выбора, но не об отсутствии способности к передаче.

Описания возможностей терминала группируются в специальные структуры данных. Каждая запись в структуре "Alternative Capably Set" отражает один из возможных режимов работы терминала. Например, Alternative Capably Set = (G.711, G.723.1, G.728) означает, что терминал может работать в любом из перечисленных режимов, но не более чем в одном.

Структуры Alternative Capably Set группируются в структуры "Simultaneous Capabilities", отражающие одновременно реализуемые возможности терминала. Например, структуры Alternative Capably Set = {H.261, H.263} и {G.711, G.723.1, G.728} означают, что терминал может работать с одним из видеокодеков и одним из аудиокодеков из числа перечисленных в первой и второй структурах соответственно. Структура Simultaneous Capability Set = [{H.261}, {H.261, H.263UG711, G.723 1, G.728}] означает, что терминал располагает двумя видеоканалами, один из которых поддерживает кодек только стандарта H.261, а второй располагает двумя кодеками, любой из которых может быть использован; аудиоканал у терминала лишь один, и он может функционировать с одним из трех перечисленных кодеков.

Полное описание возможностей терминала содержится в множестве "Capability Descriptor", каждый элемент которо-

го состоит из одной структуры Simultaneous Capability Set и переменной "Capability Descriptor Number". Терминал может направить несколько описателей типа Capability Descriptor, что будет означать возможность его работы в более чем одном режиме. Например, если кроме приведенной выше структуры, образующей Capability Descriptor N1, терминал направляет описатель Capability Descriptor N2 [{H.262}, {G.711}], то это будет означать, что он может работать и в режиме с H.262-видеокодеком, но лишь совместно с аудиокодеком G.711.

Терминалы могут динамически изменять свои возможности в ходе конференции, направляя дополнительный описатель Capability Descriptor. Это изменение возможностей терминала может быть проведено в любой момент сессии.

Каждый логический канал несет информацию от передатчика к приемнику и идентифицируется назначением уникального номера каждому каналу. Логические каналы открываются и закрываются посредством H.245-процедур и при этом генерируют сообщения "Open Logical Channel" и "Close Logical Channel". При открытии логического канала сообщение Open Logical Channel полностью описывает его содержание, т. е. содержит информацию о типе данных, используемых алгоритмах их обработки, о подуровне адаптации мультиплекса и т. п., другими словами, все необходимое для того, чтобы приемник мог правильно интерпретировать содержание логического канала.

Логический канала в H.324-системе является однопольным, т. е. асимметричный режим работы канала является допустимым. Однако если приемник оказывается способным работать только в симметричном режиме, то он должен отразить это в сообщении о своих возможностях. Ряд типов данных, включая данные документ-конференции и видеопоток, проходящий через подуровень AL3, органически требуют двунаправленного канала для своей работы. Для удовлетворения этой потребности могут быть открыты два логических канала, по одному в каждом направлении. Номера этих каналов будут различны, поскольку системы нумерации каналов независимы для каждого направления.

Контрольный канал должен быть сегментируемым и использовать нулевой логический канал. Все H.324-терминалы должны поддерживать передачу контрольных сообщений H.245 через уровень адаптации AL1 H.223-мультиплексора по протоколу SRP (Simple Retransmission Protocol, простой протокол повторной передачи), гарантирующему надежность передачи посредством повторной передачи ошибочных кадров. Протокол SRP принадлежит к уровню звена передачи данных для H.245.

Видеоканалы

Все H.324-терминалы, обладающие возможностями приема/передачи подвижного видео, должны поддерживать два стандарта кодирования. H.261 и H.263. Форматы изображения, допускаемые для этих кодеков, приведены в табл. П.3.1.

Видеокодеки должны иметь возможность обрабатывать видеопотоки на максимальной скорости, которую может обеспечить H.223-мультиплексор (максимальная скорость V.34-модема для одиночной линии, удвоенная максимальная скорость для двух линий и т. д.).

Максимальный формат изображения, минимальное количество отбрасываемых кадров изображения, выбранный алгоритм обработки и ряд других параметров устанавливаются в ходе обмена сообщениями о возможностях терминала. После этого кодер может передавать все, что согласуется с возможностями декодера. При открытии логического канала приемник получает информацию о предельных значениях параметров видеопотока, которые могут быть реализованы на этом канале. Очевидно, что реальные значения параметров видеопотока не должны превышать этих предельных значений. Например, логический канал открывался для передачи изображения в формате CIF. Реальные значения формата изображения могут быть CIF, QCIF и SQCIF, но не 4CIF или 16CIF. Информация о реальных значениях параметров каждого кадра изображения содержится в заголовке кадра.

Н.324-терминал должен поддерживать требуемые кодеки в сегментируемом логическом канале и передавать видеопоток через подуровень адаптации мультимплексора AL3. Реализация возможности повторной передачи должна быть предусмотрена в кодере.

В условиях, когда видеокommunikация поддерживается только в одном направлении (прием или передача), протокол подуровня AL3 все равно должен получить обратный логический канал. Следовательно, логический канал, используемый подуровнем AL3, должен открываться Н.245-процедурой, создающей ассоциированные логические каналы для обоих направлений передачи.

Хотя уровень AL3 обеспечивает механизм повторной передачи для коррекции обнаруженных ошибок, принимающий терминал может отказаться от повторной передачи. Такое решение может быть принято на основании:

- ◆ величины сетевой задержки в передаче пакетов;
- ◆ уровня ошибок;
- ◆ участия терминала в групповой конференции;
- ◆ участия терминала в обмене с другими Н.324-терминалами. В этот перечень могут входить и ряд других факторов.

Аудиоканалы

Терминалы, располагающие аудиовозможностями, должны осуществлять низкоскоростное и высокоскоростное кодирование, определенное в рекомендациях G.723.1. Кодеки должны также обеспечивать и передачу фреймов молчания. Синтаксис кадров аудиопотока предусматривает передачу сигналов от передатчика к приемнику о выбранной первым скорости кодирования и о режиме передачи кадров молчания. В соответствии со стандартом S.723.1 переключение скорости кодирования может осуществляться на основе содержания каждого фрейма, требуемого уровня качества аудиовоспроизведения, пропускной способности канала.

Каждый аудиоканал является независимым. Механизмы объединения аудиоканалов в пары стереоканалов или другие синхронизированные группы являются еще предметом изучения и настоящими рекомендациями не предусматривается.

Видеокodeки H.261 и H.263 порождают более значительную задержку видеопотока в сравнении с задержкой аудиопотока, возникающей в ходе цифровой обработки звуковых сигналов в кодеке G.723.1. В силу этой разности задержек возникает эффект рассинхронизации видео- и аудиопотоков. Синхронизация этих потоков не является обязательной функцией H.324-терминала, но для ее осуществления может вводиться дополнительная задержка аудиопотока в приемном терминале. С этой целью передающий терминал посредством H.245-индикатора Skew Indication сообщает усредненную величину разности времен обработки аудио- и видеосигналов. Промежуточные устройства сети, такие как MCU или межсетевые адаптеры, могут внести свой вклад в величину рассинхронизации и обязаны соответственно изменить значение переменной сообщения Skew Indication. Принимающий терминал может использовать эту информацию для обеспечения синхронного воспроизведения аудио- и видеорядов. В этом и заключается, собственно, функция блока компенсации задержки аудиоканала.

H.324-терминал должен передавать аудио-A1-SDU-пакеты с частотой f_a , определяемой используемым кодеком. Передача каждого нового пакета начинается в момент из интервала $[n \cdot T_a, n \cdot T_a + 10 \text{ мсек}]$, где n — порядковый номер пакета, а $T_a = 1/f_a$. Таким образом, продолжительность аудио-пакетов не превышает 10 мсек.

Мультиплексирование потоков G.723.1-кодека реализуется на подуровне адаптации AL2 мультиплексора. Процедура нумерации пакетов является опциональной для подуровня AL2, но ее использование в условиях, когда джиттер меньше длительности аудиокадра, не представляется оправданным.

Все фрейм-ориентированные кодеки для передачи своих информационных потоков должны использовать несегменти-

руемые логические каналы. При этом приемник сообщает количество аудиокадров, которое он способен принять в одном канальном кадре, а передатчик может посылать любое число аудиокадров в каждом пакете, но не больше, чем определяет приемник.

Каналы передачи данных

В число стандартизированных приложений обработки данных входят следующие:

- ◆ двухточечная и групповая аудиографические конференции, реализуемые в соответствии с семейством стандартов T.120, включающие возможности доступа к базам данных, передачу неподвижных изображений, передачу файлов в фоновом режиме и т. п.;

- ◆ двухточечная передача неподвижного изображения (рекомендации T.84);

- ◆ двухточечная передача файлов (рекомендации T.434);

- ◆ поддержка протокола H.221 для управления приложениями в реальном времени, в том числе протокола H.281 для управления удаленной видеокамерой;

- ◆ обработка произвольных данных пользователя от внешнего порта.

Для рассматриваемых систем КВКС обработка данных перечисленных приложений является факультативной возможностью. Они могут располагаться на внешнем компьютере или на каких-то иных самостоятельных устройствах, присоединяемых к H.324-терминалу по интерфейсу V.24, или могут быть интегрированы с терминалом (рис. П. 3.3).

Приложения могут использовать перечисленные ниже канальные протоколы. В их числе:

- ◆ V.14 с буферизацией для асинхронной передачи данных без коррекции ошибок;

- ◆ LAPM/V.42 (Link Access Procedures for Modems) — для асинхронной передачи данных с коррекцией ошибок;

- ◆ HDLC-туннелирование;

- ◆ прозрачный режим передачи данных посредством нефрагментирующих или самофрагментирующих протоколов.

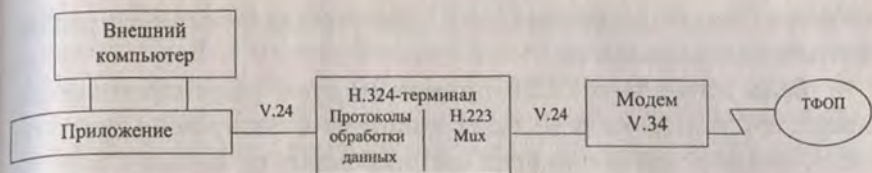


Рис. П. 3.3. Интерфейс приложений обработки данных

Протоколы обработки данных работают с H.223 логическими каналами и не оказывают непосредственного влияния на состояние физических линий терминала. Терминал может поддерживать одновременную работу нескольких приложений и, соответственно, могут быть одновременно открыты несколько логических каналов для передачи данных.

H.324-терминал, предлагающий любой из перечисленных выше протоколов, должен обеспечить реализацию его процедур, используя сегментируемый логический канал и подуровень адаптации AL1 H.223-мультиплексора. Могут использоваться и другие подуровни адаптации (AL2, AL3), если приемник обладает такой возможностью.

Протокол V.14 с буферизацией используется для преобразования асинхронного потока символов и команды BREAK, направляемых на интерфейс V.24, в синхронный битовый поток.

Протокол V.42 в режиме LAPM, по существу, решает те же задачи, что и протокол V.14, но обеспечивает еще и коррекцию ошибок передачи. Следует отметить, что этот протокол требует наличия обратного канала. Следовательно, логический канал для него должен открываться H.245-процедурами, предназначенными для создания ассоциированных двунаправленных каналов

HDLC-туннелирование является механизмом реализации протокола HDLC на асинхронных последовательных линиях. HDLC-кадры прибывают на V.24-интерфейс от приложений обработки данных. Если интерфейс V.24 функционирует в синхронном режиме, то из HDLC-кадров удаляются дополняющие нулевые биты, и все содержание кадра от начального до конечного флагов, исключая их, помещается в один ка-

нальный кадр подуровня AL1 для передачи его на уровень мультиплексирования.

Если интерфейс V.24 функционирует в асинхронном режиме, то HDLC-кадры, приходящие на интерфейс, кодируются как последовательность асинхронных символов с использованием техники "байт-стаффинг" вместо обычной для HDLC процедуры вставки нулевых битов. Поскольку обычный последовательный порт персонального компьютера не поддерживает синхронный режим работы, то последний режим оказывается практически весьма важным. Перекодированные HDLC-кадры принимаются терминалом посредством системы процедур, описанных в приложении В Рекомендаций H.324. После выполнения этих процедур все содержимое HDLC-кадра (без флагов начала и конца кадра) помещается в один фрагментируемый канальный пакет подуровня AL1 для передачи его на уровень мультиплексирования.

В прозрачном режиме передачи байты, приходящие на V.24 интерфейс, должны помещаться непосредственно в байты нефрагментируемых пакетов канального уровня с сохранением исходного порядка следования битов. Никакие процедуры фрагментации или другие преобразования исходного потока при этом не выполняются.

Теперь кратко опишем приложения обработки данных, которые выполняются на внешнем компьютере, связанном с H.324-терминалом по интерфейсу V.24. Приложения, интегрированные с терминалом, могут не выполнять процедуры, относящиеся к поддержке интерфейса V.24

Телеконференции T.120. Семейство рекомендаций T.120, подробно рассматриваемое в следующем разделе, определяет приложения двухточечных и групповых телеконференций, включающих доступ к базам данных, аннотирование и передачу графической информации, совместное использование приложений. Фоновую передачу файлов и ряд других приложений, обеспечивающих возможности совместной работы распределенных групп, H.324-терминалы, которые реализуют описываемые приложения, должны использовать базисный

набор протоколов доступа к ТФОП, определенный рекомендациями Т.123. Исключение составляет процедура обработки HDLC-кадров, поступающих на интерфейс V.24 от приложений Т.120, которые обрабатываются по методу туннелирования HDLC-кадров, описанному выше. Так как Т.120 требует для функционирования своих приложений наличия и обратного канала передачи, то логические каналы для них должны открываться Н.245-процедурой открытия ассоциированных двунаправленных каналов.

Процедуры рекомендаций Т.84 (SPIFF — Still Picture Interchange File Format) обеспечивает передачу неподвижного изображения непосредственно от источника (цифровая камера), присоединенного к передающему терминалу по интерфейсу V.24, к приемнику (например, цифровой принтер), присоединенному к принимающему терминалу по интерфейсу V.24 “Непосредственная передача” в данном случае означает пересылку данных без использования других приложений. Для согласования форматов изображений, поддерживаемых обоими конечными устройствами, используются механизмы протокола Н.245. Передача данных реализуется по протоколу LAPM/V.42.

Рекомендации Н.224 определяют протокол управления устройствами в реальном времени. В настоящее время существует лишь одно стандартизированное приложение такого рода, осуществляющее управление удаленной камерой (рекомендации Н.281). Терминалы, поддерживающие Н.224-приложения, должны использовать HDLC-туннелирование как протокол передачи данных. Допускается существование лишь одного логического канала, работающего с этим приложением. При этом предполагается, что этот логический канал работает на скорости передачи данных 4800 бит/с.

Терминальные процедуры

Коммуникационное взаимодействие Н.324-терминалов реализуется следующей последовательностью этапов:

◆ фаза А — установление телефонного соединения между участниками;

- ◆ фаза В — начальное согласование ролей по телефону;
- ◆ фаза С — установление цифрового соединения;
- ◆ фаза D — инициализация терминалов;
- ◆ фаза E — проведение конференц-обмена;
- ◆ фаза F — завершение сессии конференц-связи;
- ◆ фаза G — дополнительный сервис.

Фаза А — установление обычного телефонного соединения. Вызывающий терминал должен затребовать соединение в соответствии с процедурами аналоговой телефонии, принятой в конкретной стране.

Фаза В является процедурой обычного телефонного общения между двумя участниками будущей конференции, в ходе которого согласуется распределение ролей терминалов на этапе установления цифрового соединения.

Фаза С — установление соединения на уровне цифрового канала. Основным содержанием данного этапа является обнаружение модема принимающего терминала и выяснение его характеристик. Если вызывающему терминалу не удается обнаружить V.34-модем, то установление цифрового канала оказывается невозможным. При этом терминал, в зависимости от своей конфигурации, может:

- ◆ перейти в обычный телефонный режим;
- ◆ разорвать соединение;
- ◆ перейти в другой, более подходящий для детектируемого модемарежим.

Если же V.34-модем обнаруживается и цифровое соединение удастся установить, то терминал переходит к фазе D. Фазы В и С реализуются в соответствии с рекомендациями V.8 или V.8.bis.

Фаза D — инициализация терминалов. После установления цифрового соединения для обеспечения синхронизации канала посылаются как минимум 16 HDLC-флагов. После этого должна быть проведена инициализация коммуникационного взаимодействия между терминалами. Для этого используется H.245-контрольный канал. Поскольку таблица мультиплексирования еще не заполнена, то приемнику отсылается конт-

рольное сообщение, использующее нулевую запись таблицы мультиплексирования. Затем осуществляется обмен сообщениями о возможностях терминалов (сообщение Terminal Capability Set) и соглашениями о роли терминалов (Master Slave Determination), а также персональными номерами терминалов, отражающими их тип (Terminal Type). Если эти процедуры обмена не выполняются, они должны быть повторены, как минимум, еще дважды, прежде чем терминал прекратит попытки соединения и перейдет к фазе G.

После успешного обмена сообщениями о возможностях терминалов выполняются H.245-процедуры установления логических каналов для разных медиапоток. Таблица мультиплексирования может быть отослана до или после открытия логических каналов, но передача информации по этим каналам не может быть начата раньше, чем эта таблица будет полностью определена.

Фаза E — обмен конференц-сообщениями. В ходе сессии конференц-связи должно поддерживаться выполнение H.245-процедур изменения атрибутов логических каналов, режима приема и т. п.

В течение сессии возможны изменения модемом скорости передачи данных и даже кратковременные разрывы соединения. В таких ситуациях терминал не должен переходить в стадию реинициализации (фаза D); он должен сохранить свое состояние, определяемое фазой E, и реализовать стандартные процедуры коррекции ошибок, определяемые в рекомендациях H.223.

Если терминал регистрирует несанкционированный, невозможный разрыв цифрового канала или телефонного соединения, он должен немедленно перейти к фазе G, аналоговому телефонному режиму или к режиму разрыва соединения, пропуская фазу F.

Фаза F — завершение сессии конференц-связи. Любой из терминалов может инициировать окончание сессии. Иницилирующий окончание сессии терминал должен выполнить следующие процедуры:

◆ существующий в каждом логическом видеоканале на момент завершения сессии поток должен быть прерван лишь после завершения передачи полного кадра; после этого логический канал может быть закрыт;

◆ должны быть закрыты все существующие каналы, несущие аудиопотоки, и данные приложений;

◆ должно быть передано 1—1.245-сообщение “End Session Command” и прекращена отсылка всех H.245-сообщений;

◆ после получения подтверждения приема сообщения “End Session Command” от удаленного терминала инициирующий терминал переходит к фазе G.

Фаза G — дополнительный сервис и освобождение вызова. Эта фаза определяет возможные действия терминала при несанкционированном разрыве соединения, при изменении режима цифрового канала и заключительные операции при штатном завершении конференции.

Групповые конференции

H.324-терминалы могут участвовать в групповой конференции посредством установления соединения с модулем управления групповыми конференциями MCU. Схема такой конференции в ее общем виде представлена на рис. П. 3.4. MCU посредством рассылки H.245-сообщения “Multi point Mode Command”, содержащего только желаемый режим передачи, может обеспечить установление общего режима вещания у всех терминалов, участвующих в сессии. Все терминалы обязаны выполнить установки, содержащиеся в указанном сообщении.

Аналогичным образом посредством H.245-сообщения “Flow Control Command” MCU может установить те скорости работы модемов на каждой линии, которые могут быть обработаны соответствующими приемными устройствами.

В многоточечной конференции каждый терминал может отсылать разные значения в H.223-сообщениях “Skew Indication”, которые используются для синхронизации взаимосвязанных аудио- и видеорядов. Так как MCU может вно-

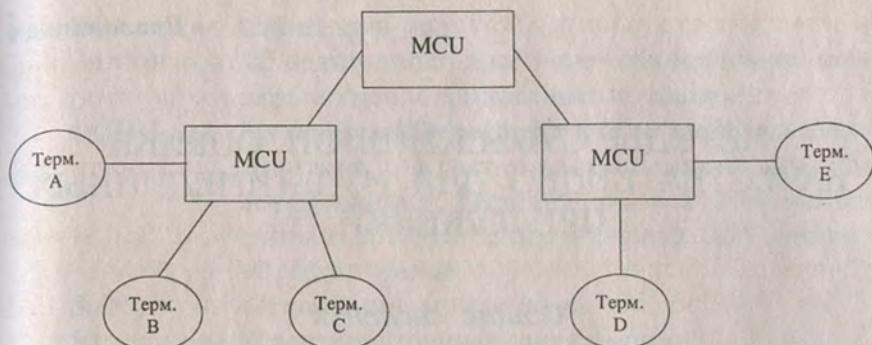


Рис. П. 3.4. Схема организации многоточечной конференции

свить свой вклад в величину рассинхронизации этих потоков, MCU должен рассылать уточненные значения этого сообщения. Следует отметить, что механизмы каскадного соединения MCU представляют собой достаточно сложную проблему и находятся еще в стадии разработки.

Литература

1. Синепол В. С., Цикин И. А. Системы компьютерной видеоконференцсвязи. Серия изданий "Связь и бизнес". — М.: ООО "Мобильные коммуникации", 1999.

МЕТОДЫ СУБЪЕКТИВНОЙ ОЦЕНКИ КАЧЕСТВА ВИДЕО ДЛЯ МУЛЬТИМЕДИЙНЫХ ПРИЛОЖЕНИЙ (1)

Общие сведения

Эти Рекомендации вводятся для определения неинтерактивных методов субъективной оценки качества изображения цифрового видео, закодированного на низких и средних битовых скоростях (до 2 Мбит/с), для таких приложений как видеотелефония, видеоконференция, сохранение и восстановление. Описанные методы подходят также для оценки влияния ошибок при передаче по таким видеосистемам. Они могут использоваться для нескольких различных целей, включая, но не ограничивая выбор алгоритмов, сравнение характеристик видеосистем и оценку уровня качества при видеосоединении.

Сокращения

- | | | |
|-------------|---|---|
| ACR | — | Определение абсолютной категории. |
| CCD | — | Прибор с зарядовой связью. |
| CI | — | Доверительный интервал. |
| CIF | — | Общий промежуточный формат (формат изображения, определенный в Рекомендациях H.261 для видеотелефонов: 352 строк × 288 пикселов). |
| CRT | — | Электронно-лучевая трубка. |
| DCR | — | Определение ухудшения категории. |
| %GOB | — | Процент хорошо или лучше (соотношение оценок хорошо и отлично). |
| LCD | — | Жидкокристаллический дисплей. |
| MOS | — | Средняя экспертная оценка. |

PC	— Сравнение пар.
%POW	— Процент плохо или хуже (соотношение оценок неудовлетворительно и плохо).
QCIF	— Четверть-CIF (вариант формата CIF с уменьшенным вчетверо разрешением) формат изображения, определенный в Рекомендациях H.261 для видеотелефона: 176 строк × 144 пикселей.
S/N	— Отношение сигнал/шум.
SI	— Пространственная информация
SIF	— Стандартный промежуточный формат [формат изображения, определенный ISO 11172 (MPEG-1): 352 строк × 288 пикселей × 25 фрейм/сек and 352 строк × 240 пикселей × 30 фрейм/сек].
SP	— Одновременное представление.
Std	— Стандартная девиация.
TI	— Временная информация.
VTR	— Видеомагнитофон.

Термины и определения

Gamma гамма: параметр, который описывает различие между уровнями серого на дисплее. Связь между яркостью и напряжением входного сигнала нелинейна, при возрастании напряжения как экспонента гаммы. Для компенсации этой нелинейности в камере обычно применяется коэффициент коррекции, который является обратной функцией гамма. Гамма также имеет влияние на передачу цвета.

Тесты по оптимизации: субъективные тесты, которые обычно выполняются при разработке или стандартизации новых алгоритмов или систем. Целью таких тестов является оценка характеристик новых средств, для оптимизации исследуемых алгоритмов или систем.

Тесты по квалификации: субъективные тесты, которые обычно выполняются для сравнения характеристик коммерческих систем или оборудования. Эти тесты должны выпол-

няться при условиях тестирования, которые в наибольшей степени отражают реальные условия использования.

Spatial perceptual information пространственная воспринимаемая информация (SI): мера, которая в общем случае показывает количество пространственных деталей изображения. Она обычно выше для большинства пространственно заполненных сцен. Она не призвана быть мерой энтропии и не связана с понятием информации, определенной в теории связи.

Temporal perceptual information временная воспринимаемая информация (TI): мера, которая в общем случае показывает количество временных изменений видеопоследовательности. Она обычно выше для последовательностей с интенсивным движением. Она не призвана быть мерой энтропии и не связана с понятием информации, определенной в теории связи.

Transparency (fidelity) прозрачность (точность воспроизведения): концепция, описывающая характеристику кодека или системы по отношению к идеальной системе передачи без ухудшения.

Можно определить два типа прозрачности. Первый тип описывает, насколько хорошо обработанный сигнал совпадает с входным сигналом или идеальным сигналом, используя математический критерий. Если разницы нет, система полностью прозрачна. Второй тип описывает, насколько хорошо обработанный сигнал совпадает с входным сигналом или идеальным сигналом, для человека. Если разница не может быть замечена при любых экспериментальных условиях, система воспринимаемо прозрачна. Термин прозрачности без подробной ссылки на критерий будет использоваться для систем, которые воспринимаемо прозрачны.

Replication репликация: Повторение тех же условий (с тем же материалом источника) для того же субъекта.

Достоверность субъективного теста:

- внутренняя ("Внутри субъекта") достоверность относится к согласованности между повторными оценками субъекта при одних и те же условиях теста;

• общая (“между субъектами”) достоверность относится к согласованности между оценками различных субъектов при одних и тех же условиях теста.

Применимость субъективного теста: согласованность между средним значением оценок, полученных в тесте, и истинным значением, которое предполагалось измерить в тесте.

Эталонные условия: макетные условия, добавленные к тестовым условиям, чтобы привязать к ним оценки, полученные в разных экспериментах.

Явный эталон (эталон источника): условия, используемые экспертами в качестве эталона для выражения своего мнения, когда используется метод ухудшения категории. Этот эталон показывается первым в каждой паре последовательностей. Обычно формат явного эталона является форматом, используемым на входе тестируемого кодека (например, ITU-R BT.601-4, CIF, QCIF, SIF и т.д.). В данных Рекомендациях слова “явный” и “источник” будут опускаться, когда по контексту ясно значение слова “эталон”.

Скрытый эталон: условия, используемые экспертами в качестве эталона для выражения своего мнения о материале, когда используется метод определения абсолютной категории. Если экспериментатором предлагается скрытый эталон, он должен быть хорошо известен эксперту (например, реальная система телевидения).

Источник сигнала

Чтобы управлять характеристиками источника сигнала, тестовые последовательности должны быть определены в соответствии с целью теста и записаны с помощью цифровой системы записи. Когда экспериментатор заинтересован в сравнении результатов от разных лабораторий, необходимо использовать общий набор последовательностей источника, чтобы не было разногласий.

Условия записи. Источник(и) света (электрические или флуоресцентные лампы) могут размещаться выше или сбоку

от камеры. При размещении света нужно помнить, что верхний свет более типичен для освещения офисов, и он должен использоваться в сценах, изображающих деловую обстановку. Студийные лампы и другие необычные источники использовать не следует. Условия освещенности комнаты в поле зрения могут меняться от 100 lux до 10 000 lux внутри помещений. Должно быть учтено изменение (частота переменного тока) света (флюоресцентного освещения), так как это может вызвать мерцание в записываемой видеопоследовательности. Условия освещенности, цвет стен, отражающая способность поверхностей и т. п. должны внимательно контролироваться и фиксироваться.

Записывающая система. Камера. Последовательности изображений должны записываться на камеру типа CCD высокого качества. Отношение сигнал/шум входного видеосигнала может сильно влиять на характеристики кодека.

Для определения видеовхода должны быть определены следующие пункты:

- динамический диапазон YUV сигналов;
- коэффициент коррекции гамма (должен быть 0.45);
- полосы/наклоны характеристик фильтров;
- чувствительность камеры в условиях очень низкой освещенности и характеристики АРУ (автоматической регулировки усиления) при ее использовании.

В соответствии с Рекомендациями МККР 567-3 часть С, раздел 3.2.1 [2], должно быть измерено взвешенное отношение сигнал/шум. Взвешенное отношение сигнал/шум должно быть больше 45 дБ (среднеквадратичное значение).

Нестабильность или джиттер синхросигналов может вызвать шумовой эффект. Для устройства синхронизации камеры требуется минимальная стабильность 0.5×10^{-6} .

Могут использоваться системы с фиксированным или переменным фокусным расстоянием. Для настольных терминалов разумной является величина глубины фокуса от 30 см до 120 см, тогда как для многопользовательских систем более подходящей может быть глубина фокуса от 50 см до беско-

нечности. Для обеспечения изменения освещенности в комнате записи следует использовать регулируемую диафрагму или фильтры с нейтральной оптической плотностью. Камера должна иметь автоматический баланс белого, чтобы можно было выполнить адаптацию к цветовой температуре источника света. Коррекция температуры белого может изменяться в диапазоне от 2700°К (использование внутри помещений с электрическими лампочками) до 6500°К (дневная температура при облачном небе).

Видеосигнал и формат записи. Видеосигналы источника, выдаваемые камерой, должны быть дискретизированы в соответствии с Частью А публикации [3]. Чтобы предотвратить искажения источника сигнала, он должен быть сохранен в цифровом формате, например, на компьютере или на ленте формата D-1 4:2:2.

Характеристики сцены. Выбор тестовых сцен является важным вопросом. В частности, критичными параметрами являются объемная и временная воспринимаемая информация сцены. Эти параметры играют критическую роль при определении величины возможной видеокompрессии и, следовательно, уровня ухудшения, который допускается при передаче сцены по каналам цифровой передачи с фиксированной скоростью. Хорошие и важные тестовые сцены должны быть выбраны так, чтобы их объемная и временная информация была согласована с видеоуслугой, для передачи которой был предназначен цифровой канал передачи. Набор тестовых сцен должен охватывать весь диапазон объемной и временной информации, которая интересует пользователей тестируемых устройств.

Подробности о характеристиках тестовых последовательностей и примеры подходящих тестовых сцен даны в Приложении А, D и E.

Число последовательностей должно определяться в соответствии с программой эксперимента. Чтобы не утомить экспертов и достичь минимальной достоверности результатов, для последовательностей следует выбрать, по меньшей мере, четыре различных типа сцен (т. е. различных тем).

Следующие подразделы представляют методы численного определения объемной и временной информации тестовых сцен. Эти методы для оценки объемной и временной информации тестовых сцен применимы для тестирования качества видео в настоящее время и в будущем. Расположение видеосцен в объемно-временной матрице важно, так как качество передаваемой видеосцены (особенно, после пропуска через низкоскоростной кодек) часто сильно зависит от этого расположения. Меры объемной и временной информации, представленные здесь, могут быть использованы для обеспечения соответствующего покрытия объемно-временного плана.

Меры объемной и временной информации, данные ниже, имеют по одному значению для каждого фрейма по полной тестовой последовательности. Это дает в результате временной ряд значений, который в общем будет меняться. Меры воспринимаемой информации, данные ниже, удаляют эту изменчивость с максимальной функцией (максимальным значением для последовательности). Сама по себе изменчивость может быть исследована, например, по значениям объемно-временной информации фрейм-за-фреймом. Использование распределения информации по тестовой последовательности также позволяет дать лучшую оценку сцен с отрезками сцены.

Измерение объемной воспринимаемой информации. Объемная воспринимаемая информация SI имеет в основе фильтр Собела. Каждый видеофрейм (план яркости) в момент времени n (F_n) сначала фильтруется в фильтре Собела [$Sobel(F_n)$]. Затем вычисляется стандартная девиация по пикселям (std_{space}) в каждом пропущенном через фильтр Собела фрейме. Эта операция повторяется для каждого фрейма в видеопоследовательности и дает в результате временной ряд объемной информации сцены. Выбирается максимальное значение во временном ряду (max_{time}) для представления содержимого объемной информации сцены. Этот процесс может быть представлен в виде формулы как:

$$SI = \max_{time} \{std_{space} [Sobel(F_n)]\}$$

Измерение временной воспринимаемой информации.

Временная воспринимаемая информация TI основана на свойстве различного движения $M_n(i, j)$, которое является разницей между значениями пикселей (плана яркости) на том же месте в пространстве, но в последовательные моменты времени или фреймы. $M_n(i, j)$ как функция времени (n) определяется как:

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j),$$

здесь $F_n(i, j)$ — пиксел в i -ом ряду и j -ом столбце n -го фрейма во времени.

Мера временной информации TI вычисляется как максимум по времени (\max_{time}) стандартной девиации по пространству ($\text{std}_{\text{space}}$) от $M_n(i, j)$ по всем i и j .

$$TI = \max_{\text{time}} \{ \text{std}_{\text{space}} [M_n(i, j)] \}$$

Больше движения в соседних фреймах приведет к большему значению TI .

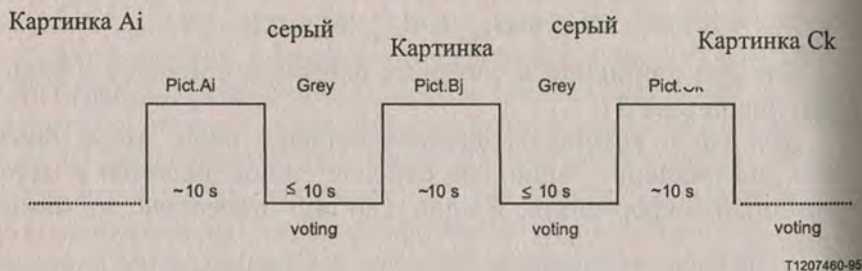
Для сцен, которые содержат отрезки сцен, могут быть даны два значения: одно, где отрезок сцены включен в меру временной информации, и одно, где оно исключено из измерения.

Методы тестирования и программа

Измерение воспринимаемого качества образов требует использования методов субъективных шкал. Условием для таких измерений является наличие смысла, т. е. то, что существует связь между физическими характеристиками “воздействия”, в данном случае видеопоследовательности, представляемой субъекту при тесте, и величиной и природой ощущения, вызванного воздействием. Несколько экспериментальных методов было оценено для различных целей. Здесь рекомендуются три метода для приложений, использующих скорость соединения до 2 Мбит/с. Итоговый выбор одного из этих методов для определенного приложения зависит от нескольких факторов, таких как содержания, цели и где в процессе выполнения теста он выполняется.

Определение абсолютной категории (ACR). Метод определения абсолютной категории — это проверка категории, где тестовые последовательности представлены по одной в каждый момент времени и оцениваются независимо по шкале категорий. Этот метод также называется методом одиночного воздействия. Метод требует, чтобы после каждого представления субъекта просили оценить качество показанной последовательности.

Пример представления воздействия во времени показан на рис. П.4.1, где: voting — голосование; Ai — последовательность A при тестовом условии I; Bj — последовательность B при тестовом условии j; Ck — последовательность C при тестовом условии k.



Ai sequence A under test condition i
 Bj sequence B under test condition j
 Ck sequence C under test condition k

Рис. П.4.1. Представление воздействия при методе ACR

Время голосования должно быть меньше или равно 10 с, в зависимости от используемого механизма голосования. Время представления может быть увеличено или уменьшено в соответствии с содержанием тестового материала.

Следует использовать следующую пятибалльную шкалу для определения общего качества: 5 — отлично; 4 — хорошо; 3 — посредственно; 2 — неудовлетворительно; 1 — плохо.

Если требуется большая различимость, можно использовать 9-балльную шкалу. Примеры подходящих численных

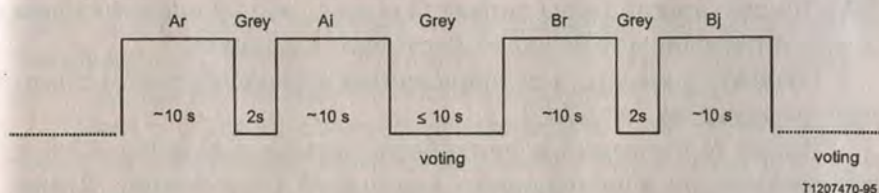
или непрерывных шкал даны в Приложении В. Это приложение также дает примеры определения других размерностей, нежели общее качество. Такие размерности могут быть полезны для получения большей информации о различных воспринимаемых факторах качества, когда общее качество почти одинаково для тестируемых систем, хотя системы явно воспринимаются как разные.

Для метода АСR необходимое число репликаций получается при повторении тех же тестовых условий в разные моменты времени теста.

Определение категории ухудшения (DCR). Определение категории ухудшения требует, чтобы тестовые последовательности представлялись парами: первые воздействия, представляемые в каждой паре, всегда являются эталоном источника, а вторые воздействия — тот же источник, представленный после прохождения тестируемой системы. Этот метод также называется методом шкалы ухудшения при двойном воздействии.

При преобразовании используются форматы изображения (например, CIF, QCIF, SIF); может быть полезно показывать на дисплее одновременно тестовую последовательность и эталон. Руководство по процедуре представления обсуждается в Приложении С.

Пример представления воздействия во времени проиллюстрирован на рис. П.4.2, где: A_i — последовательность А при



A_i sequence A under test condition i
 Ar, Br sequences A and B respectively in the reference source format
 Bj sequence B under test condition j

Рис. П.4.2. Представление воздействия при методе DCR

тестовом условии I ; A_i, B_i — последовательность A и B соответственно в формате эталона источника; V_j — последовательность V при тестовом условии j .

Время голосования должно быть меньше или равно 10 с, в зависимости от используемого механизма голосования. Время представления может быть уменьшено или увеличено в соответствии с содержанием материала.

В этом случае экспертов просят определить ухудшение второго воздействия по отношению к эталону.

Для определения ухудшения следует использовать следующую пятибалльную шкалу: 5 — незаметно; 4 — заметно, но не раздражает; 3 — немного раздражает; 2 — раздражает; 1 — очень раздражает.

Для метода DCR необходимое число репликаций получается при повторении тех же тестовых условий в разные моменты времени теста.

Метод попарного сравнения (РС). Метод попарного сравнения требует, чтобы тестовые последовательности представлялись парами, которые состоят из одной и той же последовательности, пропущенной сначала через одну тестируемую систему, а затем через другую.

Тестируемые системы (A, B, C , и т.д.) в общем случае объединены во всех возможных $n(n-1)$ комбинациях AB, BA, CA и т. д. Поэтому все пары последовательностей должны быть показаны в обоих возможных порядках (например, AB, BA). После каждой пары делается оценка, какой элемент пары предпочтителен в контексте тестового сценария.

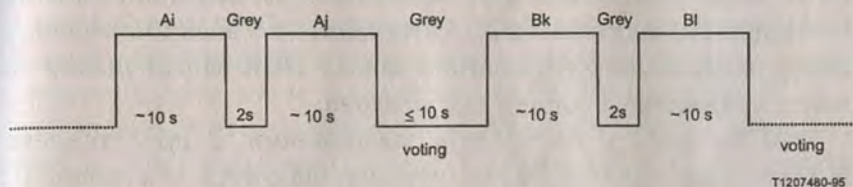
Отрезок времени для представления воздействия проиллюстрирован на рис. П.4.3.

Время голосования должно быть меньше или равно 10 с, в зависимости от используемого механизма голосования. Время представления должно быть около 10 с и может быть уменьшено или увеличено в соответствии с содержанием материала.

При преобразовании используются форматы изображения (например, CIF, QCIF, SIF); может быть полезно показывать на дисплее одновременно тестовую последовательность

и эталон. Руководство по процедуре представления обсуждается в Приложении С.

Для метода РС число репликаций в общем случае несущественно, так как сам метод требует повторного представления одних и тех же условий, хотя и в различных парах.



A_i, A_j — последовательность А при i -ом и j -ом тестовом условии соответственно;

B_k, B_l — последовательность В при k -ом и l -ом тестовом условии соответственно.

Рис. П. 4.3. Представление воздействия в методе попарного сравнения РС

Варианты метода РС используют шкалу категорий для дальнейшего измерения разницы между парами последовательностей (см. [4] и [5]).

Сравнение методов. Важным моментом при выборе метода тестирования является фундаментальная разница между методами, которые используют явные эталоны (например, DCR) и методами, которые не используют явных эталонов (например, ACR и РС). Этот второй класс методов не тестирует прозрачность или точность воспроизведения.

Метод DCR не следует использовать при тестировании точности воспроизведения передачи по отношению к источнику сигнала. Часто это важный фактор при оценке систем высокого качества. DCR долго был основным методом, определенным в [4], для оценки телевизионных изображений, типичное качество которых представляет очень высокие уровни видеотелефонии и видеоконференций. Другие методы также могут использоваться для оценки систем высокого каче-

ства. Специальные комментарии о шкале DCR (не воспринимается/воспринимается) имеют значение, когда определение экспертом ухудшения является важным фактором. Поэтому, когда важно проверить точность воспроизведения по отношению к сигналу источника, следует использовать метод DCR. DCR также применим при оценке систем высокого качества в отношении мультимедиа. Разделение не воспринимаемого/воспринимаемого ухудшения в шкале DCR обеспечивает это, как и сравнение с качеством эталона.

ACR прост и быстр при выполнении, а представление воздействия похоже на то, что используется системой. Поэтому ACR хорошо подходит для квалификационных тестов. Принципиальное достоинство метода РС — это высокая степень различимости, которое имеет особое значение, когда несколько тестируемых вещей почти равны по качеству.

Когда в одном тесте нужно оценить большое число пунктов, процедура, основанная на методе РС, будет очень длительной. В этом случае сначала с ограниченным числом экспертов может быть проведен тест ACR или DCR, а затем — тест РС только по тем пунктам, которые получили приблизительно одинаковую оценку.

Эталонные условия. Результаты оценки качества часто зависят не только от действительного качества видео, но и от других факторов, таких как общий диапазон качества при тестовых условиях, опыт и ожидания экспертов и т. д. Чтобы управлять некоторыми из этих эффектов, может быть добавлено несколько искусственных тестовых условий, которые используются в качестве эталонных.

Описание эталонных условий и процедуры их подготовки даны в Рекомендациях Р.930 [6]. Ввод сигнала источника в качестве эталонного условия при тесте РС особенно рекомендуется, когда ухудшение, вводимое тестируемым пунктом, очень мало.

Уровень качества эталонного условия должен, по меньшей мере, покрывать диапазон качества тестируемого пункта.

Разработка эксперимента. Можно использовать различные методы разработки эксперимента, такие как сложные случайные, Латинский, Греко-Латинский и метод Еврейских площадей, метод повторяющихся блоков и т.д. (см. [7]), выбор которых должен определяться целью эксперимента.

На разработчика эксперимента возлагается выбор метода разработки, чтобы выполнить требования по стоимости и точности. Программа может также зависеть от того, какие условия представляют особый интерес в данном тесте.

Рекомендуется включать в эксперимент, по меньшей мере, 2, если возможно, 3 или 4 повторения (т. е. повтора идентичных условий). Есть несколько причин для использования повторений, важнейшей из которых является то, что "в рамках изменчивости субъекта" можно вести измерения с использованием повторяющихся данных. Для тестирования достоверности субъекта можно использовать один и тот же порядок представлений при идентичных условиях. Если используется другой порядок представления, результирующее изменение в экспериментальных данных складывается из влияния порядка и колебаний самого субъекта.

Повторения делают возможным вычисление индивидуальной достоверности субъекта и, если необходимо, отбрасывание недостоверных результатов нескольких субъектов. Оценка внутренней стандартной девиации каждого субъекта и стандартной девиации между субъектами является следующей предпосылкой для выполнения корректного анализа вариаций и обобщения результатов. Кроме того, эффект обучения в рамках теста также приводит к сбалансированности.

Дальнейшее улучшение для учета эффекта обучения выполняется благодаря включению периода тренировки, где в начале каждого теста представляются минимум 5 условий. Эти условия должны быть выбраны так, чтобы быть репрезентативными по отношению к представлениям, показываемым дальше в ходе теста. Предварительные представления не принимаются во внимание при статистическом анализе результатов теста.

Процедуры оценки

Условия просмотра. Тест должен проводиться при следующих условиях просмотра в соответствии с [8]:

Расстояние для просмотра	от 4Н до 8Н (см. Примечание)
Максимальная яркость	от 70 cd/m ² до 200 cd/m ²
Отношение контрастности экрана без подсветки фона	от 30 до 50
Отношение освещенности фона к максимальной яркости экрана	~0.25
Освещенность	около 500 lux
Основная цветность	белый

Примечание. Н показывает высоту изображения. Расстояние для просмотра должно учитывать не только размер экрана, но и тип экрана, тип приложения и цель эксперимента.

Для данной высоты экрана расстояние для просмотра, удобное для наблюдателя, увеличивается, когда наблюдаемое качество падает. По этому пункту предпочтительное расстояние для просмотра должно быть определено заранее для квалификационных тестов. При тестах по оптимизации должно использоваться более критичное расстояние (т. е. 4Н—6Н).

Лучше использовать полный экран для показа последовательностей. Тем не менее когда по некоторым причинам последовательность должна быть показана на части экрана, цвет фона экрана должен быть 50% серого, соответствующего $Y = U = V = 128$ (U и V без знака).

Для сравнения результатов теста все условия просмотра должны быть зафиксированы и должны быть одинаковыми в лабораториях для одного типа тестов.

Обрабатывающая и воспроизводящая системы. Существует два метода получения тестовых изображений из записей источника:

- при передаче или повторном воспроизведении видеозаписей в реальном времени через тестируемые системы, когда эксперты наблюдают и оценивают;

- при офлайн-обработке записей источника в тестируемом устройстве и записывании выхода для получения нового набора записей.

Во втором случае нужно использовать цифровой видеоманитофон для минимизации ухудшений, которые можно получить в процессе записывания. В любом случае нужно принимать во внимание, что ухудшения, вносимые низкоскоростным кодированием, обычно более очевидны, чем ухудшения, вносимые при модуляции видеоманитофонами с профессиональным качеством, такими как D2, MII и BetacamSP.

Можно использовать электронно-лучевую трубку или жидкокристаллический монитор. Размер и тип монитора должны соответствовать исследуемым приложениям.

Мониторы должны быть отрегулированы в соответствии с процедурами, определенными в [9].

Эксперты. Возможное число экспертов в тестовом просмотре — от 4 до 40. 4 — абсолютный минимум по статистическим соображениям, и одновременно это число редко превышает 40.

Вообще, в эксперименте должно участвовать минимум 15 зрителей. Они не должны заниматься оценкой качества изображений в своей основной деятельности и не должны быть опытными экспертами. Тем не менее в прошлом и группы из 4-8 субъектов давали показательные результаты.

Перед тестом эксперты обычно должны быть проверены на нормальную остроту зрения и нормальное цветовосприятие. По остроте зрения не должно быть ошибок при проверке по стандартной таблице 20/30 [10]. Таблица должна быть размещена на расстоянии, с которого будет вестись тестовый просмотр. По цветовосприятию из 12 иллюстраций не должно быть пропущено больше 2.

Инструкции экспертам и обучение. Перед началом эксперимента экспертам должен быть выдан сценарий предполагаемого использования тестируемой системы. Кроме того,

в письменной форме должны быть выданы описание типа оценки, шкалы мнений и представление воздействия. В предварительных опытах должен быть представлен диапазон ухудшений. Это могут быть видеопоследовательности, но иные, нежели используемые в действительном тесте. Не обязательно, чтобы худшее качество, проявленное в тренировочном наборе, соответствовало бы самой низкой субъективной оценке по шкале. Вопросы по процедуре или о смысле инструкций должны быть заданы только в начале теста.

Возможные предлагаемые тексты с инструкциями даны в Приложении Е.

Статистический анализ и публикация результатов

О результатах нужно сообщать со всеми подробностями. Для каждой комбинации тестовых переменных должны быть даны средние значения и стандартная девиация статистического распределения оценок.

По данным должна быть вычислена субъективная достоверность и сообщен метод оценки субъективной достоверности. Некоторые критерии субъективной достоверности даны в [4] и [11].

Информативным для анализа является кумулятивное распределение оценок. Так как кумулятивные распределения не чувствительны к линейности, они могут быть чрезвычайно полезны для данных, линейность которых сомнительна, как у полученных при использовании методов ACR и DCR, совместно со шкалой категорий без классификации (т. е. оценка категории).

Данные могут быть оформлены, например, как для ACR представленные в табл. П.4.1.

Для оценки существенности тестовых параметров должна использоваться классическая техника анализа вариаций. Если для оценки качества видео нужна оценка как функция от параметра, для интерпретации данных может быть полезна техника совпадения кривых.

Информативная таблица с кумулятивным распределением оценок для метода ACR

Условие	Всего оценок	Отлично	Хорошо	Удовлетворительно	Не удовлетворительно	Плохо	MOS	CI	Std	%GOB	%POW

Условие: отметки показывают комбинации тестовых последовательностей.

Всего оценок: число оценок, собранных для этого условия.

Отлично ... Плохо: число каждой из оценок.

В случае попарного сравнения метод вычисления положения каждого воздействия на шкале, где разница между воздействиями соответствует разнице в предпочтении, описана в Руководстве по телефонетрии, раздел 2.6.2С [12].

Приложение А

Детали, относящиеся к характеристикам тестовых последовательностей.

Фильтр Собела [13]. Фильтр Собела выполняется сверткой двух узлов 3×3 по видеофрейму и нахождением квадратного корня суммы квадратов результатов этих свертки.

Для $y = \text{Sobel}(x)$, пусть $x(i,j)$ обозначает пиксел входного изображения в i -ом ряду и j -ом столбце. $Gv(i,j)$ будет результатом первой свертки и задается как:

$$Gv(i,j) = -1 * x(i-1,j-1) - 2 * x(i-1,j) - 1 * x(i-1,j+1) + \\ + 0 * x(i,j-1) + 0 * x(i,j) + 0 * x(i,j+1) + \\ + 1 * x(i+1,j-1) + 2 * x(i+1,j) + 1 * x(i+1,j+1)$$

Таким же образом, $Gh(i,j)$ будет результатом второй свертки и задается как:

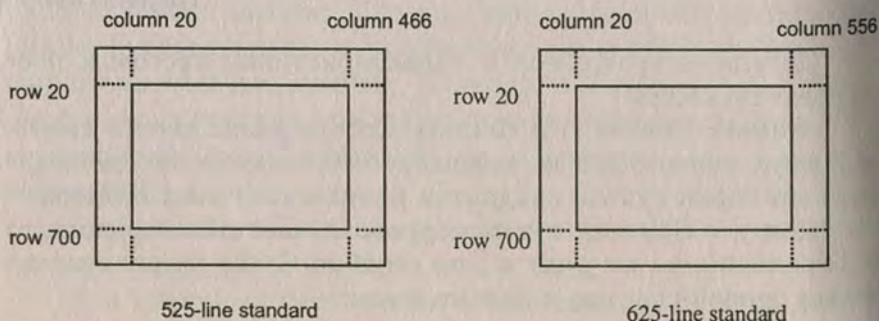
$$Gh(i,j) = -1 * x(i-1,j-1) + 0 * x(i-1,j) + 1 * x(i-1,j+1) + \\ -2 * x(i,j-1) + 0 * x(i,j) + 2 * x(i,j+1) + \\ -1 * x(i+1,j-1) + 0 * x(i+1,j) + 1 * x(i+1,j+1).$$

Поэтому выход пропущенного через фильтр Собела изображения в i -ом ряду и j -ом столбце задается как :

$$y(i,j) = \{[Gv(i,j)]^2 + [Gh(i,j)]^2\}^{0.5}.$$

Вычисление выполняется для всех $2 \leq i \leq N-1$ и $2 \leq j \leq M-1$, где N — число рядов, а M — число столбцов.

Рекомендуется, чтобы вычисление выполнялось по фрагменту изображения видеофрейма, чтобы предотвратить нежелательные угловые эффекты, потому что верхние углы видеофрейма обычно не видны на электронно-лучевой трубке. Это можно выполнить при использовании подходящего фрагмента изображения, как показано для примера на рис. П. А.1 для форматов 625- и 525-строки ITU-R BT.601-4 [3], где *row* — ряд, *column* — столбец, 525-line standard — стандарт на 525 строк.



T1207490-95

Рис. П.А.1. Фрагменты изображения для вычисления пространственной SI и временной TI-информации для форматов 525- и 625-строк ITU-R BT.601-4 [3]

Как использовать SI и TI для выбора тестовой последовательности. При выборе тестовой последовательности полезно сравнить относительную объемную и временную информацию в различных имеющихся последовательностях. В общем случае сложности по сжатию относятся непосредственно к объемной и временной информационной последовательности.

Если в данном тесте предполагается использовать небольшое число тестовых последовательностей, может быть важно выбрать последовательности, которые перекрывают большую часть плана объемно-временной информации (см. рис. П.А.2, где: по вертикали — элементы объемной информации; по горизонтали — элементы временной информации). В случае, где предполагается использовать 4 тестовых последовательности в тесте, можно было бы выбрать последовательность из каждого из четырех квадрантов плана объемно-временной информации.

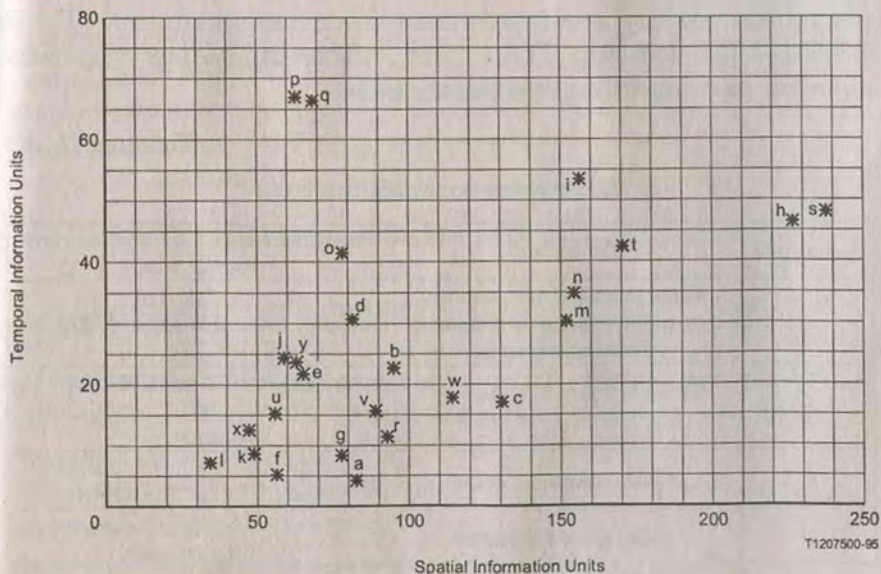


Рис. П.А.2. Объемно-временной рисунок для набора примеров тестовых сцен

В противоположность этому если пытаться выбрать тестовые последовательности, которые были бы одинаково сложны при кодировании, то желателен выбор последовательностей, которые имеют похожие значения SI и TI.

Примеры. Рисунок П.А.2 показывает относительное количество объемной и временной информации для некоторых репрезентативных тестовых сцен и то, как они могут быть размещены на плане объемно-временной информации.

По оси $TI = 0$ (вдоль границы рисунка) находятся спокойные сцены и сцены с ограниченным движением (такие как l, f и a). Рядом с вершиной рисунка находятся сцены с небольшим количеством движения (такие как p, q и i).

Вдоль оси $SI = 0$ (в левом углу рисунка) находятся сцены с минимальными объемными деталями (такие как l, k, x, u и f). Рядом с правым углом рисунка находятся сцены с максимальными объемными деталями (такие как h и s).

Значения SI и TI были получены с использованием выше приведенных уравнений и видео, которое было дискретизировано по объему в соответствии со спецификациями ITU-R BT.601-4 [3]. Таблица П.А.1 дает список примеров тестовых сцен по категории содержания сцены.

Таблица П. А.1

Категории содержания сцен

Категория	Описание	Наименование сцены и буквенное обозначение
A	Один человек, в основном голова и плечи, ограниченные детали и движение	vtc1nw(f), susie(j), disguy(k), disgal(l)
B	Один человек с графикой и/или больше деталей	vtc2mp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w), inspec(x)
C	Более одного человека	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
D	Графика с выделением	washdc(c), cirkit(s), rodmap(t), filter(u), ysmite(v),
E	Сильное движение объекта и/или камеры (примеры вещания TV)	flogar(h), fball(i), fedas(y)

Дополнительные шкалы оценок

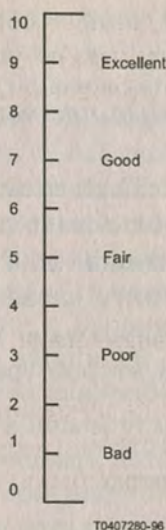
Шкалы. Особенно для оценки низкоскоростных видеокодеков часто необходимо использовать шкалы с большим числом градаций, чем 5. Подходящая для этих целей шкала — это 9-балльная шкала, где пять словесно определенных категорий качества, как рекомендовано в 6.1, используются в качестве меток для каждой второй градации на шкале, как показано на рис. П.В.1.

9	Отлично
8	
7	Хорошо
6	
5	Удовлетворительно
4	
3	Неудовлетворительно
2	
1	Плохо

Рис. П.В.1. Численная шкала качества с 9 градациями

Дальнейшее расширение этой шкалы показано на рис. П.В.2, где конечные точки словесно определены как точки привязки, которые сами не используются для определения оценки. Число 10 обозначает качество воспроизведения, которое достаточно близко к оригиналу. Дальнейшего улучшения не требуется. Число 0 обозначает качество воспроизведения, которое не имеет сходства с оригиналом. Худшее качество невозможно вообразить.

В этом словесном определении используется некоторый эталон (например, на рис. П.В.2 оригинал используется как



The number 10 denotes a quality of reproduction that is perfectly faithful to the original. No further improvement is possible.
 The number 0 denotes a quality of reproduction that has no similarity to the original. A worse quality cannot be imagined.

Рис. П.В.2. Численная шкала качества с 11 градациями

эталон). Этот эталон может быть явным или скрытым, и это будет отчетливо показано на фазе обучения. См. также [5] и [12] Раздел 2.6 Шкала а).

Для обоих типов шкалы ответ субъектов может записываться в виде чисел, которые записываются на листе ответов, или как пометка на самой шкале (в этом случае для каждого оцениваемого условия придется предусмотреть отдельную шкалу на листе ответов). Когда требуются численные ответы, экспертов следует просить использовать числа с десятичными долями (например, 2.2 вместо 2), но они могут еще использовать и только целые числа. Следует заметить, что может быть трудно перевести названия категорий шкалы на разные языки. При этом отношения между категориями могут стать отличными от тех, что были в исходном языке [14].

Дополнительной возможностью является использование непрерывной шкалы.

Так как непрерывные данные обычно округляются до некоторой разумной точности, чтобы упростить сбор данных, можно использовать шкалу голосования, подобную показанной на рис. П.В.3.

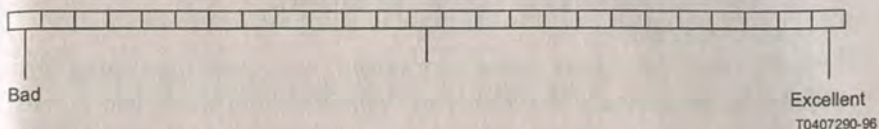


Рис. П.В.3. Квази-непрерывная шкала для оценки качества

Это уменьшит отклонения из-за интерпретации отметок. Каждая зона может соответствовать определенному численному значению, и данные могут быть собраны без двусмысленностей.

Дополнительные размерности оценок. Если системы, которые оцениваются в тесте, оцениваются как довольно равные по общему качеству, и получают при этом похожие оценки, может быть предпочтительнее выделить дополнительные компоненты качества на отдельных шкалах для каждого условия. При этом возможно получить информацию о специфических характеристиках, где объекты тестирования воспринимаются существенно различными, даже если общее качество фактически то же самое. Результаты этих дополнительных тестов могут дать ценную диагностическую информацию о тестируемых системах.

Примеры размерностей оценок, которые могут оцениваться для определения факторов, вносящих вклад в воспринимаемое глобальное качество, перечислены ниже, вместе с пояснением, какой вклад в качество вносит фактор — позитивный или негативный:

- яркость (positive);
- контраст (positive);
- цветовоспроизведение (positive);
- четкость контуров (positive);
- стабильность фона (positive);

скорость повторного собирания изображения? (positive);
дрожание (negative);
эффект "Размывания" (negative);
эффект "Комаров" "Снег" (negative);
двойное изображение/тени (negative);
ореол (negative).

Недавние исследования показали, что эти факторы могут быть объединены в расчетное глобальное качество путем присвоения соответствующих весов каждому фактору и сложения их вместе [15]. Для отдельной оценки размерностей общего качества видео можно использовать специальный опросник. Примеры вопросов, которые можно задать после представления каждого тестового условия, представлены на рис. П.В.4.

Опросный лист

Выразите ваше мнение пометкой на нижеследующих шкалах.

1. Как Вы оцениваете цвета изображения?

Bad
Плохо

Excellent
Отлично

2. Как Вы оцениваете контрастность изображения?

Bad

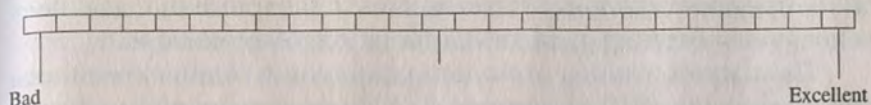
Excellent

3. Как Вы оцениваете границы изображения?

Bad

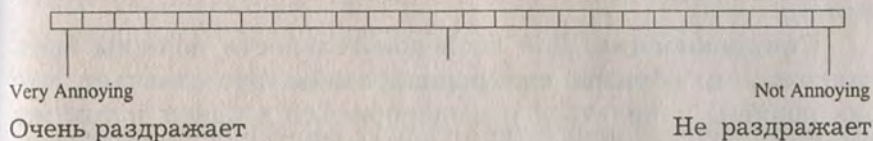
Excellent

4. Как Вы оцениваете непрерывность движения?



5. Заметили Вы какое-либо мерцание в последовательности? Yes No

Если заметили мерцание, опишите интенсивность на шкале



6. Заметили Вы размывание в последовательности? Yes No

Если заметили размывание, опишите его на шкале

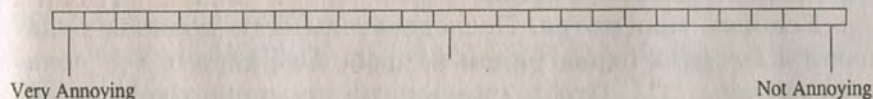


Рис. П.В.4. Опросный лист

Примечание. Когда используются эти шкалы, все категории качества/ухудшения, принимаемые во внимание (например, непрерывность движения, мерцание, размывание и т. д.), должны быть проиллюстрированы на фазе обучения.

Приложение С

*Одновременное представление
пар последовательностей*

Введение. Когда система, которая оценивается в тесте, использует уменьшенный формат изображения, такой как CIF,

QCIF, SIF и т.д., и используются методы DCR или PC, может быть предпочтительнее показывать одновременно две последовательности из каждой пары на одном мониторе.

Преимуществами при использовании одновременного представления (SP) являются: 1) SP уменьшает время теста; 2) если используются подходящие размерности, субъектам проще оценить разницу между воздействиями; 3) так как при одинаковых тестовых условиях число презентаций равно половине, внимание экспертов обычно выше при использовании SP.

Синхронизация. Две последовательности должны быть надлежащим образом синхронизированы, это означает, что они должны начинаться и заканчиваться в одном и том же фрейме и что вывод на экран должен быть синхронизирован. Это не мешает тому, чтобы могли сравниваться последовательности, кодированные с разными битовыми скоростями, что достигается применением подходящего временного восстановления дискретизации.

Условия просмотра. Последовательности должны показываться в двух окнах рядом на фоне 50% серого, как показано на рис. П.С.1. Чтобы уменьшить движение глаз для перевода внимания между двумя окнами, дистанция просмотра должна быть $8H$, где H показывает высоту картинка. Размер монитора по диагонали должен быть не менее 14 дюймов.



T1207510-95

Рис. П.С.1. Относительное положение двух последовательностей при SP

Представления. При DCR эталон должен быть размещен всегда с одной и той же стороны (например, слева), и эксперт должен быть осведомлен об относительном расположении эталонного и тестового условий.

При РС все пары последовательностей должны выводиться на экран в обеих возможных очередностях (например, АВ, ВА). Это означает, что последовательности, которые были выведены слева, сейчас показываются справа, и наоборот.

Приложение D

Тестовые последовательности

Выбор подходящих тестовых последовательностей — это ключевой момент при планировании субъективной оценки. Когда результаты тестов, выполненных с различными группами экспертов или в разных лабораториях, должны быть коррелированы, важно, чтобы имелся общий набор тестовых последовательностей.

Первый набор таких последовательностей описан в табл. П.Д.1.

Таблице П.Д.1

Тестовые последовательности для оценки качества видео в приложениях мультимедиа

Последовательность	Категория	Описание	Формат источника	SI	TI
washdc	D	Карта Вашингтона с движением руки и карандаша	Betacam SP (525-строк)	130.5	17.0
3inrow	C	Человек у стола, панорама	Betacam SP (525-строк)	81.7	30.8
vtc1nw	A	Сидящая женщина, читающая новости	Betacam SP (525-строк)	56.2	5.3
susie	A	Молодая женщина у телефона	ITU-R BT.601-4 525-/625-строк	58.7	24.6
flower garden	E	Пейзаж, панорама	ITU-R BT.601-4 525-/625-строк	227.0	46.4
smity2	B	Продавец у прилавка с журналами	Betacam SP (525-строк)	154.5	35.1

В этой таблице для каждой последовательности дана следующая информация:

- категория (определенная в табл. П.А.1);
- краткое описание сцены;
- формат источника (625- или 525-строк, формат ITU-R BT.601-4 или Betacam SP);
- значения объемной и временной информации (определенные в 5.3.1 и 5.3.2 соответственно).

Все последовательности, перечисленные в табл. П.Д.1, относятся к общим предметам и могут быть свободно использованы для оценок и показов. Некоторые последовательности, предложенные относящимися к ММКР лабораториями, описаны в CCIR Report 1213 [16].

Другие последовательности лабораторий ММКР могут подходить для использования с частными приложениями, такими как основанные записи видео и восстановления.

Набор тестовых последовательностей все еще исследуется. Набор тестовых последовательностей, приведенный в табл. П.Д.1, может быть улучшен или расширен, как минимум, двумя способами:

1) должны быть включены последовательности, представляющие более широкий диапазон приложений (например, мобильная видеотелефония, дистанционное обучения и т.д.);

2) формат источника для каждой последовательности должен быть ITU-R BT.601-4 [3] в обеих — 525- и 625-строчных версиях.

Приложение E

Инструкции по тестам просмотра

Следующие положения могут использоваться в качестве основы для инструкций экспертам, участвующим в экспериментах по методам ACR, DCR или PC.

Помимо этого, инструкции должны давать информацию о приблизительной продолжительности эксперимента, предварительных опытах и других деталях, полезных экспертам. Этой информации здесь нет, так как она зависит от конкретного исполнения.

ACR.

Доброе утро и благодарю вас за то, что вы пришли.

В этом эксперименте вы увидите на экране перед вами короткие видеопоследовательности. Всякий раз, когда последовательность показана, вам нужно оценить ее качество с использованием одного из пяти уровней по следующей шкале: 5 — отлично; 4 — хорошо; 3 — удовлетворительно; 2 — не удовлетворительно; 1 — плохо.

Просмотрите внимательно содержание видеопоследовательности перед принятием решения об оценке.

DCR.

Доброе утро и благодарю вас за то, что вы пришли.

В этом эксперименте вы увидите на экране перед вами короткие видеопоследовательности. Каждая последовательность будет представлена дважды с быстрым повтором: в каждой паре обработана только вторая последовательность. В конце представления каждой пары вы должны оценить ухудшение второй последовательности по отношению к первой. Вы выразите свою оценку с использованием следующей шкалы: 5 — не воспринимаемо; 4 — воспринимаемо, но раздражает; 3 — немного раздражает; 2 — раздражает; 1 — очень раздражает.

Просмотрите внимательно содержание видеопоследовательностей перед принятием решения об оценке.

PC.

Доброе утро и благодарю вас за то, что вы пришли.

В этом эксперименте вы увидите на экране перед вами короткие видеопоследовательности. Каждая последовательность будет представлена дважды с быстрым повтором: каждый раз — через другой кодек. Порядок последовательностей

и комбинация кодеков в парах меняются случайным образом. В конце представления каждой пары вы должны выразить свое предпочтение, поставив “галочку” в одном из квадратов (см. рис. П.Е.1).

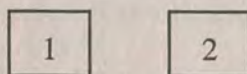


Рис. П.Е.1

Вы должны пометить квадрат 1, если вы предпочли первую последовательность, и квадрат 2, если вы предпочли вторую последовательность в паре.

Просмотрите внимательно содержание видеопоследовательностей перед принятием решения об оценке.

Литература

1. Методы субъективной оценки качества видео для мультимедийных приложений. Серия Р: Качество телефонной передачи. Аудиовизуальное качество в услугах мультимедиа. Рекомендации Р. 910. Женева, 1996.

2. CCIR Recommendation 567-3: Transmission performance of television circuits designed for use in international connections, *Recommendations of the CCIR*, Volume XII, 1990. Характеристики передачи телевизионных сетей, разработанные для использования в международном вещании.

3. ITU-R Recommendation BT.601-4 (1994), *Encoding parameters of digital television for studios*. Параметры кодирования цифрового телевидения для студий.

4. ITU-R Recommendation BT.500-6 (1994), *Method for the subjective assessment of the quality of television pictures*. Методы субъективной оценки качества телевизионного изображения.

5. ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*. Методы субъективного определения качества передачи.

6. ITU-T Recommendation P.930 (1996), Principles of a reference impairment system for video. Принципы эталонного ухудшения системы для видео.

7. KIRK (R.E.): Experimental Design — Procedures for the Behavioural Sciences, 2nd Editions, Brooks/Cole Publishing Co., California, 1982.

8. ITU-R Recommendation BT.812 (1994), *Subjective assessment of the quality of alphanumeric and graphic pictures in Teletext and similar services*. Субъективная оценка качества цифро-буквенных и графических изображений в телетексте и подобных услугах.

9. ITU-R Recommendation BT.814-1 (1994), *Specifications and alignment procedures for setting of brightness and contrast of displays*. Спецификации и процедуры регулировки для установки яркости и контрастности дисплеев.

10. Gramh-Field Catalogue Number 13-1240.

11. IEC Publication 268-13, *Sound System equipment: listening tests on loudspeakers*. Звуковое оборудование системы: тесты по прослушиванию громкоговорителей.

12. CCITT: *Handbook on Telephonometry*, Geneva, 1993. Руководство по телефонометрии.

13. GONZALEZ (R.C.), WINTZ (P.): Digital Image Processing, 2nd Edition, Addison-Wesley Publishing Co., Reading, Massachusetts, 1987.

14. VIRTANEN (M.T.), GLEISS (N.), GOLDSTEIN (M.): On the use of Evaluative Category Scales in Telecommunications, HFT 1995, *Human Factors in Telecommunication Conference*, Melbourne, 1995.

15. RACE Industrial Consortium Project 1018 HIVITS, WP B5, Picture Quality Measurement, 1988.

16. CCIR Report 1213: Test pictures and sequences for subjective assessments of digital codecs, *Reports of the CCIR*, Annex to Volume XI, Part 1, 1990. Тестовые изображения и последовательности для субъективной оценки цифровых кодеков.

Главный редактор — Ю. А. Сулимов
Художник — М. А. Хавторин
Корректор — Л. М. Волкова
Верстка — Н. П. Якушина

Ответственный за выпуск — Л. М. Волкова

Владимир Михайлович Артюшенко
Олег Иванович Шелухин
Михаил Юрьевич Афонин

Цифровое сжатие видеoinформации и звука

Учебное пособие

Санитарно-эпидемиологическое заключение

№ 77.99.02.953.Д.002261.04.01 от 26.04.2001.

Лицензия ИД № 06473 от 19 декабря 2001 г.

Подписано в печать 16.07.2003. Формат 60×84 1/16.

Печать офсетная. Бумага газетная.

Печ. л. 26,75. Тираж 1500 экз. Заказ № 2690.

Издательско-торговая корпорация «Дашков и К».

129347, Москва, Ярославское ш., д. 142, к. 732.

Для писем: 129347, Москва, п/о И-347

Тел./факс: (095) 183-93-01, 182-11-79, 182-01-58

E-mail: sales@dashkov.ru – отдел продаж

ivc.market@relcom.ru; office@dashkov.ru – офис; <http://www.dashkov.ru>

Отпечатано в соответствии с качеством предоставленных диапозитивов
в ФГУП «Производственно-издательский комбинат ВИНТИ»

140010, г. Люберцы Московской обл., Октябрьский пр-т, 403. Тел. 554-21-86

Издательско-торговая корпорация «Дашков и К^о»

Предлагает учебную литературу, изданную в 2003 году
(знаком * обозначены учебники и учебные пособия с грифом
Министерства образования РФ)

- Анализ финансово-хозяйственной деятельности (учебник)*. Чечевицына Л. Н.
- Безопасность жизнедеятельности (учебник)*. Под ред. Арустамова Э. А.
- Бухгалтерский учет на предприятиях сервиса (учебное пособие)*. Тимофеева М. С. и др.
- Бухгалтерский учет на производственных предприятиях (учебник)*. Керимов В. Э.
- Бюджетная система Российской Федерации (учебник)*. Вахрин П. И.
- Гражданское право. Общая часть (учебник)*. Бунич Г. А. и др.
- Гражданское право. Особенная часть (учебник)*. Бунич Г. А. и др.
- Деловые игры в учебном процессе. Трайнев В. А.
- Идентификация и фальсификация товаров (учебное пособие). Чепурной И. П.
- Инвестиции (учебник)*. Вахрин П. И.
- Коммерческая деятельность (учебник)*. Панкратов Ф. Г.
- Коммерческая деятельность (практикум). Панкратов Ф. Г. и др.
- Коммерческое товароведение продовольственных товаров (учебник). Круглякова Т. В., Кругляков Т. Н.
- Коммерция и технология торговли (учебник)*. Дашков Л. П., Памбухчиянц В. К.
- Коммерческое товароведение (учебник)*. Теплов В. И. и др.
- Лизинг (учебник)*. Горемыкин В. А.
- Логистика (учебник)*. Гаджинский А. М.
- Логистика (практикум). Гаджинский А. М.
- Маркетинг (учебник)*. Годин А. М.
- Маркетинг в отраслях и сферах деятельности (учебник)*. Под ред. Алексунина В. А.
- Наследственное право (учебник)*. Гуцин В.В.
- Основы экологии (учебное пособие). Валова В. Д.
- Организация и технология коммерческой деятельности (учебник)*. Памбухчиянц О. В.
- Организация, технология и проектирование торговых предприятий (учебник)*. Дашков Л. П., Памбухчиянц В. К.
- Паблик рилейшнз (учебно-практическое пособие). Антипов К. В. и др.
- Правоохранительные органы (учебник)*. Под ред. Н. А. Петухова, Г. И. Загорского
- Природопользование (учебник)*. Под ред. Арустамова Э. А.
- Прогнозирование и планирование в условиях рынка (учебное пособие)*. Владимирова Л. П.
- Производственный менеджмент (учебник)*. Фатхутдинов Р. А.
- Рекламная деятельность (учебник)*. Под ред. Панкратова Ф. Г.
- Религиоведение (учебник)*. Лобазова О. Ф.
- Россия в системе мировых цивилизаций (учебник)*. Кулешов С. В.
- Сервисное право (учебник). Гуцин В. В. и др.
- Социология (учебник)*. Под ред. Павленка П. Д.
- Статистика (учебник)*. Годин А. М.

- Технология розничной торговли (учебник). Памбухчианц О. В.
- Технология товаров (продовольственных) (учебник). Хлебников В. И.
- Товароведение и экспертиза кондитерских товаров (учебник)*. Чепурной И. П.
- Товароведение и экспертиза вкусовых товаров (учебник)*. Чепурной И. В.
- Управленческий учет (учебник)*. Керимов В. Э.
- Финансовое право (учебник). Мелихова Л. И.
- Финансы предприятия (учебник)*. Шуляк П. Н.
- Финансы. Денежное обращение. Кредит (учебник)*. Вахрин П. И., Нешитой А. С.
- Финансы (учебник)*. Вахрин П. И.
- Финансовый анализ в коммерческих и некоммерческих организациях (учебное пособие)*. Вахрин П. И.
- Ценообразование (учебно-практическое пособие). Шуляк П. Н.
- Экологические основы природопользования (учебник)*. Арустамов Э. А. и др.
- Экономика недвижимости (учебник)*. Горемыкин В. А.
- Экономика. Общий курс (учебник). Войтов А. Г.
- Экономика организации предприятия (учебник)*. Раицкий К. А.
- Экономическая история (учебник)*. Толмачева Р. П.
- Экономическая теория (учебник)*. Носова С. С.
- Экономическая теория (учебник)*. Океанова З. К.

Всегда в наличии свыше 3000 наименований учебной и деловой литературы московских и региональных издательств по издательским ценам

**Оптовая и мелкооптовая продажа книг осуществляется в книжном салоне
Издательско-торговой корпорации «Дашков и К°»**

по адресу: г. Москва, ул. Проходчиков, д. 2 (помещение
"Нового драматического театра", служебный вход),
а также у наших торговых представителей:

Фирма «Юрайт-книга» – г. Москва, городок им. Баумана, 3, к. 4, стр. 10,
тел./факс (095) 742-72-12, 165-91-27, e-mail: sales@urait.ru.

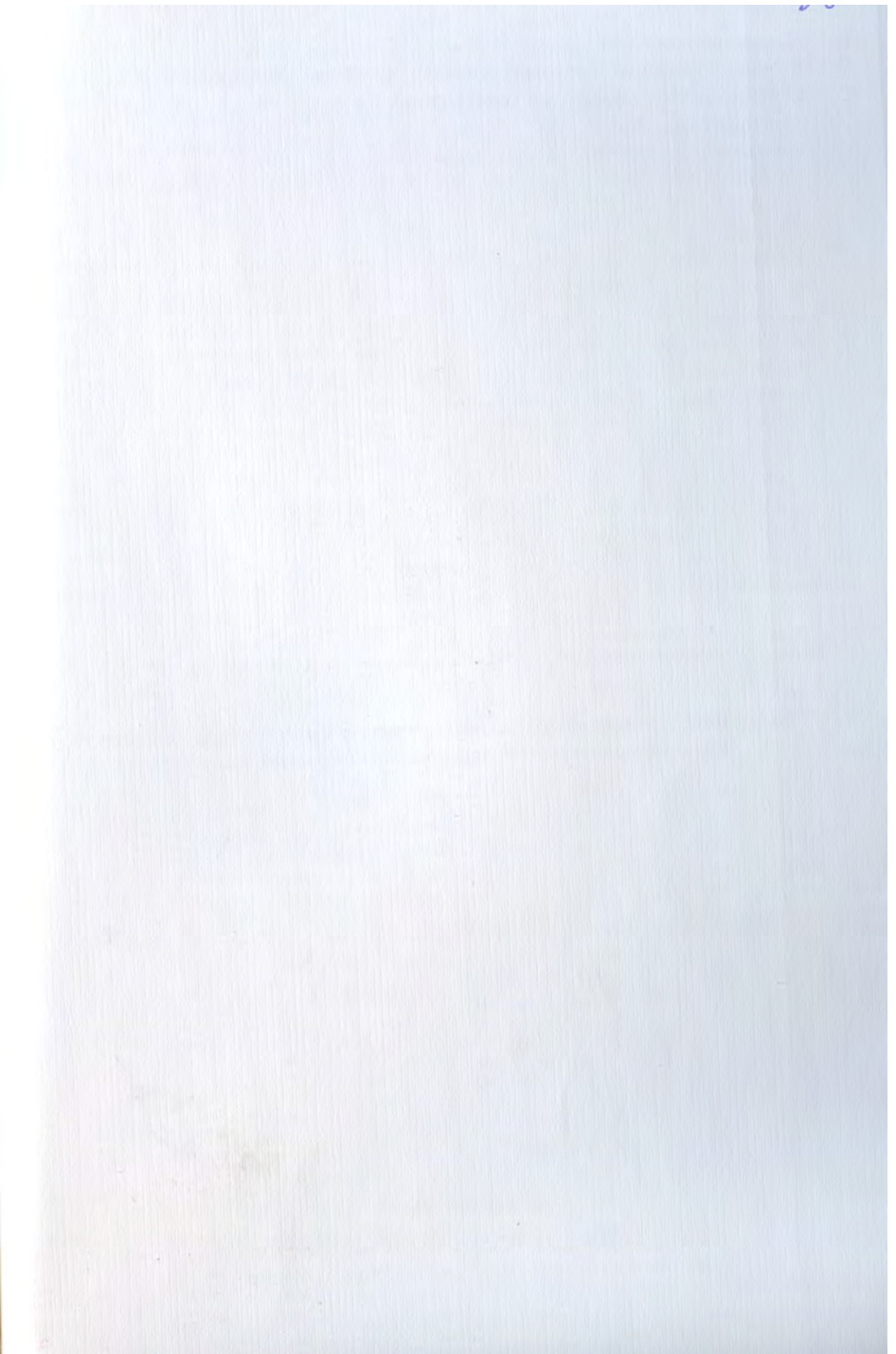
Фирма «Бизнес-пресса» – г. Санкт-Петербург, ул. Разъезжая, д. 37/39,
тел. (812) 164-74-82, 164-56-01, e-mail:sales@bizpressa.spb.ru.

ПБОЮЛ «Дашков С.Б.» – г. Тула, пр. Ленина, д. 108, оф. 203,
тел. (0872) 33-22-43,
тел./факс (0872) 35-13-93. E-mail: dashkov@tula.net

*Приглашаем к взаимовыгодному сотрудничеству книготорги,
библиотечные коллекторы, магазины, библиотеки, учебные
заведения и региональных представителей.*

*Справки по тел.: (095) 183-03-98, 183-93-01, 182-11-79 (продажа),
182-01-58 (редакция)*

**E-mail: sales@dashkov.ru – отдел продаж;
http://www.dashkov.ru; offis@dashkov.ru – офис**



ЦИФРОВОЕ СЖАТИЕ ВИДЕОИНФОРМАЦИИ И ЗВУКА



Владимир
Михайлович
АРТЮШЕНКО

Академик МАИ, АПК и Нью-Йоркской академии наук. Окончил МИИТ (1984); д.т.н., профессор; заведующий кафедрой «Информатика и компьютерный сервис» МГУС. Автор более 250 научных работ. Область научных интересов – информационные системы.



Олег
Иванович
ШЕЛУХИН

Заслуженный деятель науки Российской Федерации. Окончил МИИТ (1973), МГУ им. М.В.Ломоносова (1981); д.т.н., профессор; заведующий кафедрой «Радиотехника и радиотехнические системы» МГУС. Автор более 250 научных работ. Область научных интересов – радио и телекоммуникации.



Михаил
Юрьевич
АФОНИН

Окончил с отличием радиотехнический факультет МТУСИ (2000). Аспирант кафедры «Радиотехника и радиотехнические системы» МГУС. Автор 10 научных работ. Область научных интересов – кодирование аудио и видео информации.

ISBN 5-94798-258-7



9 785947 982589