

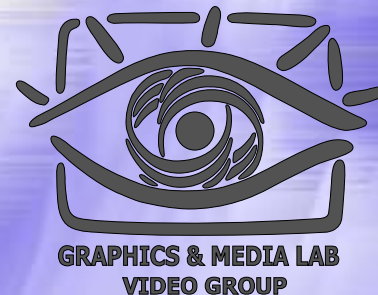
# Стандарт MPEG-4

*Из курса лекций «Сжатие  
медиаданных»*

Дмитрий Ватолин

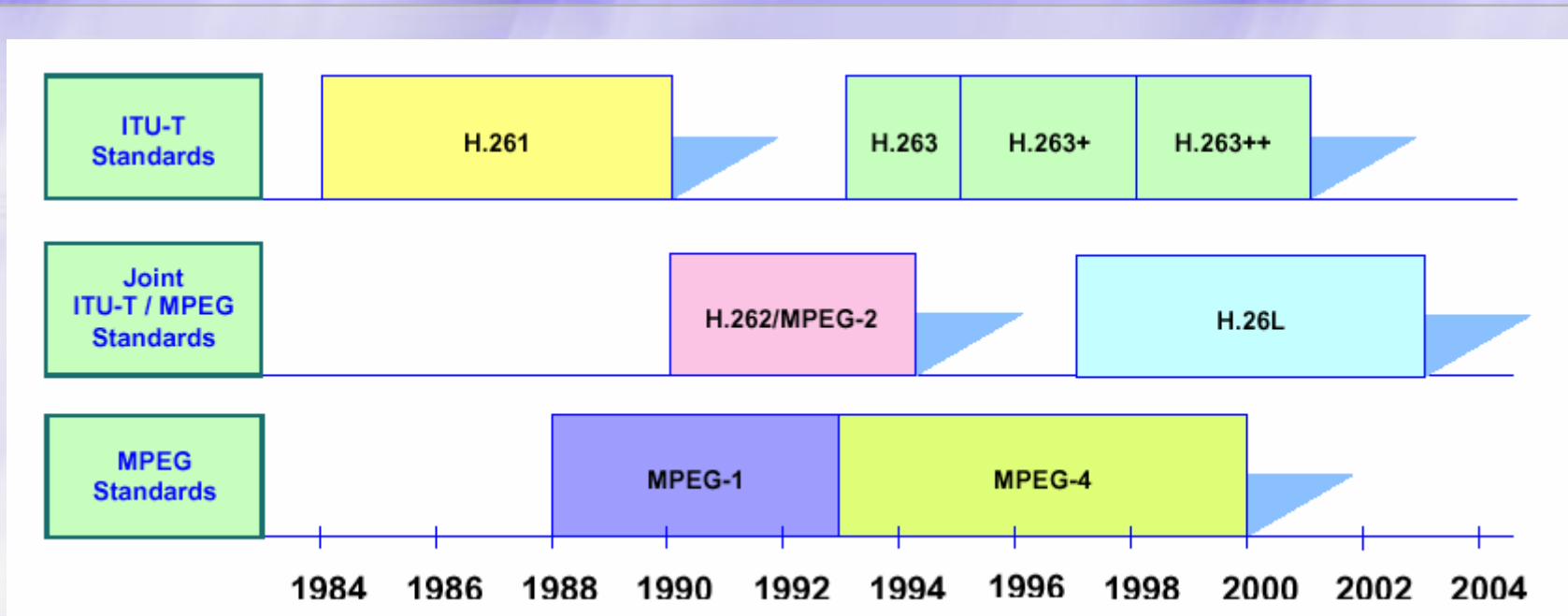
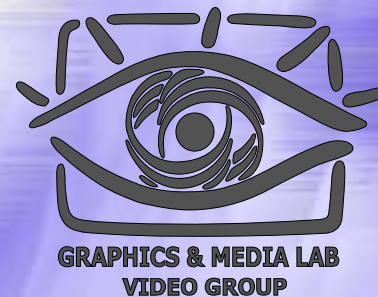
*Video Group  
CS MSU Graphics & Media Lab*

# Содержание



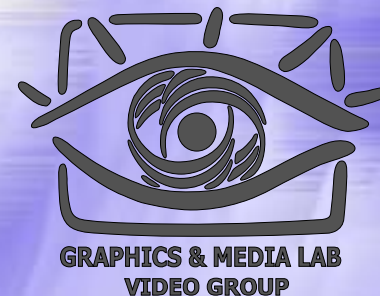
- ◆ Введение в стандарты сжатия видео
- ◆ Общие сведения об MPEG-4
- ◆ Цели MPEG-4
- ◆ Требования к MPEG-4
- ◆ Аудио-визуальные объекты
- ◆ Описание сцены
- ◆ Архитектура системы
- ◆ Профили
- ◆ MPEG-4: кодирование
- ◆ Синтаксическое описание в MPEG-4
- ◆ Инструменты в MPEG-4
- ◆ Формат файла
- ◆ MPEG-J
- ◆ Устойчивость к ошибкам
- ◆ Литература

# Стандарты сжатия видео



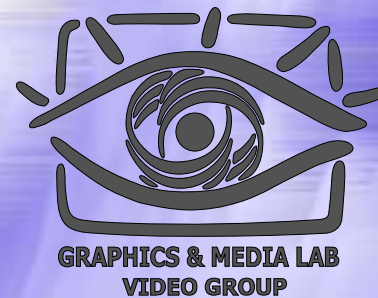
Основные стандарты по годам

# Сравнение стандартов



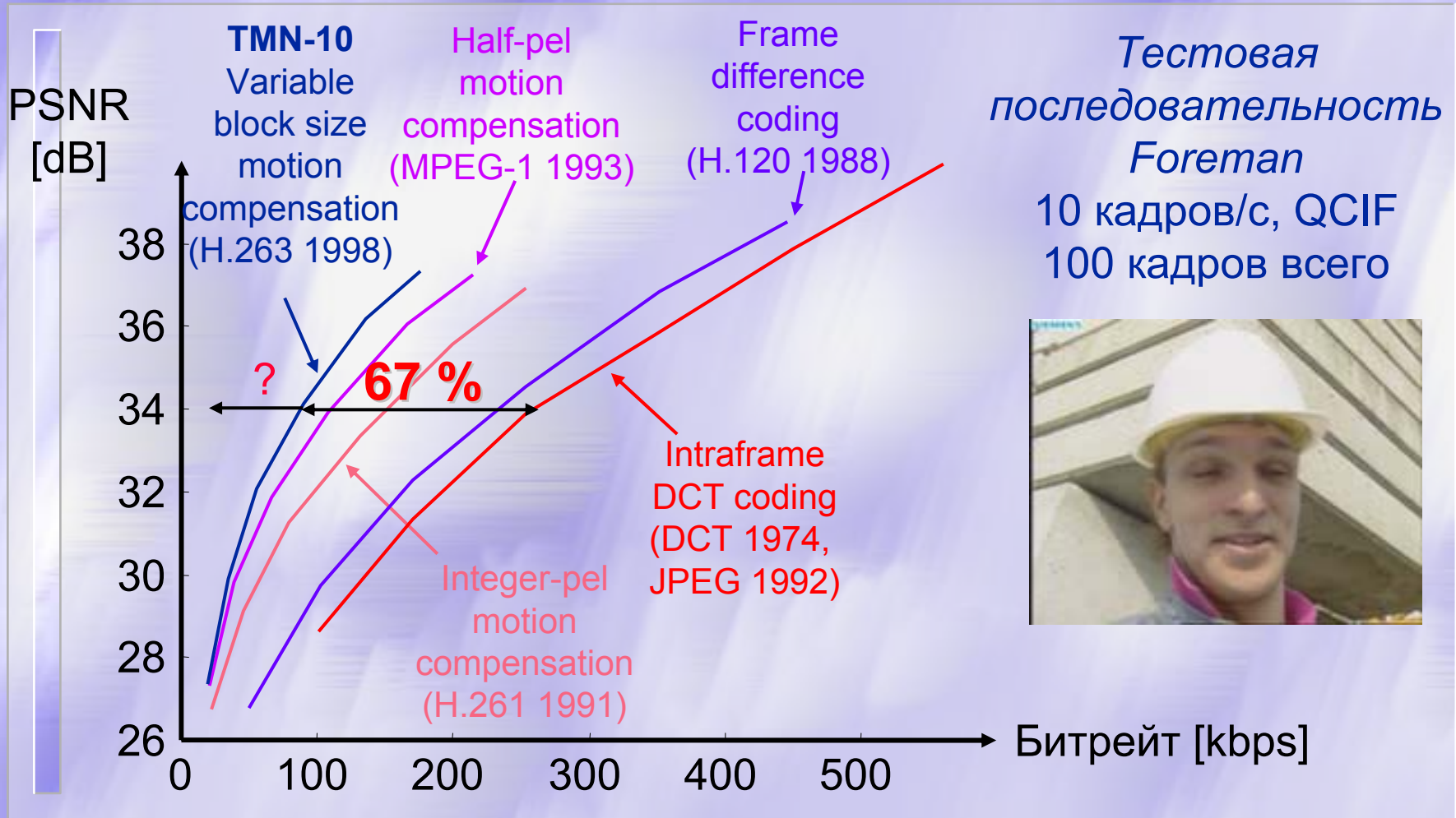
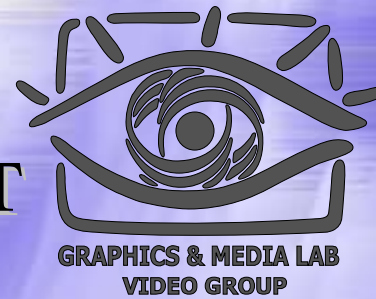
Название	Год	Разрешение и поток	Аудио	Применение
MPEG-1	1992	352x240x30, 352x288x25, 1.5 Мбит/с	MPEG-1 Layer II	Ранние VideoCD
H.261	1993	352x288x30, 176x144x30, 0,04-2 Мбит/с (рх64 Кбит/с, р от 1 до 30)	—	
MPEG-2	1995	Универсальный, 3-15 Мбит/с	MPEG-1 Layer II, Dolby Digital 5.1, DTS	DVD
MPEG-3 не принят	1993-95	Телевидение высокой четкости, 20-40 Мбит/с		HDTV
MPEG-4	1999	Универсальный, 0,0048-20 Мбит/с	MPEG-1 Layer II, MPEG-1 Layer III, Dolby Digital 5.1, DTS	Новые VideoCD

# Дополнительные ВОЗМОЖНОСТИ

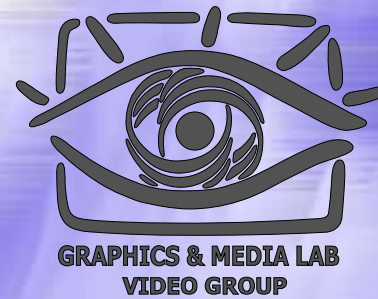


Название	За счет чего сжимает	Дополнительные возможности
MPEG-1	ICT, DCT	
H.261	ICT, DCT, MC	Передача потока данных по r каналам с пропускной способностью 64Кбит (телеконференции по нескольким телефонным линиям).
MPEG-2	ICT, DCT, MC	
MPEG-4	ICT, Wavelet, MC, спрайты, объекты с прозрачным фоном, 3d-рендеринг	Встроенный язык описания BIFS, синтезатор речи, функции анимации лиц, 3D-рендеринг и т.д.

# Развитие стандартов за 10 лет



# MPEG

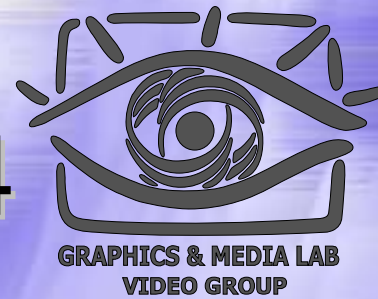


- ◆ *MPEG* : Moving Picture Experts Group: рабочая группа ISO/TEC

*“Компактное представление цифровых видео и аудио сигналов для широкого использования”*

- ◆ MPEG-1: Стандарт сжатия видео для носителей медиаданных
- ◆ MPEG-2: Стандарт сжатия видео для цифрового телевидения

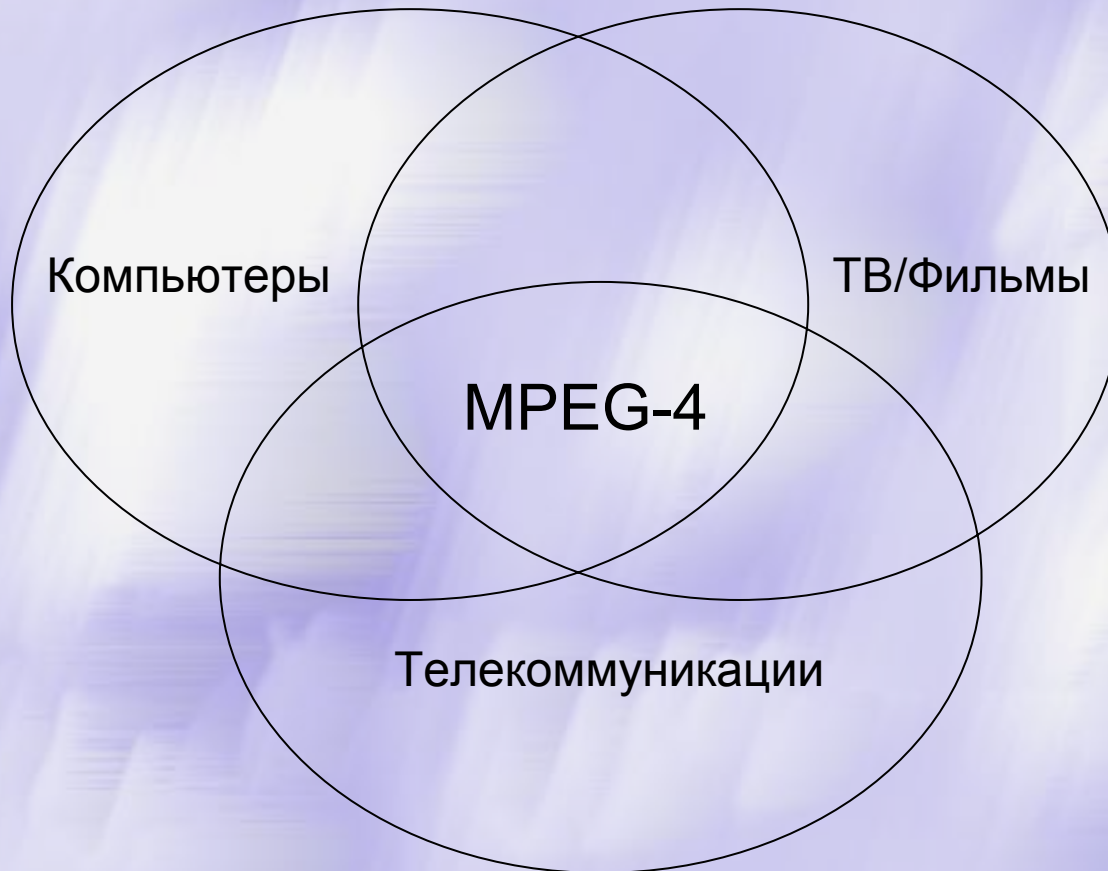
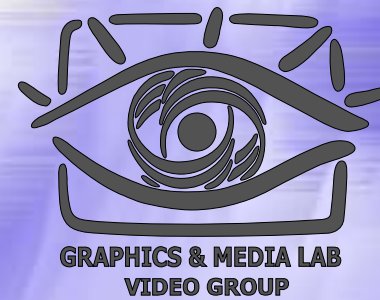
# Что дает стандарт MPEG-4



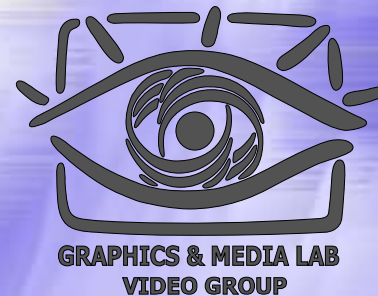
- ◆ Производители: высокая гибкость создания продуктов и их модифицируемость
- ◆ Поставщики сетевого сервиса: предоставление транспортной информации, которая может быть интерпретирована на различных сетевых платформах
- ◆ Конечный пользователь: высокий уровень возможностей продукта без ограничения по производителю



# MPEG-4: применение



# MPEG-4: цели



- ◆ **Интерактивность: взаимодействие с различными аудио-визуальными объектами**
- ◆ **Масштабируемость: адаптация содержания к ширине канала передачи данных**
- ◆ **Использование компонентов: как инструментов, так и данных**

# Цели: интерактивность

- ◆ Интерактивность на стороне клиента:  
манипуляции с описанием сцены и  
аудио-визуальными объектами на ней
- ◆ Поведение аудио-визуальных объектов:  
вызов пользователем или другими  
способами
- ◆ Интерактивность на стороне сервера:  
смена канала по желанию

# Цели: масштабируемость

Масштабируемость подразумевает возможность декодировать лишь часть потока и восстановить последовательность картинок с:

- Уменьшением сложности декодирования (качества)
- Уменьшением пространственного разрешения
- Уменьшением временного разрешения

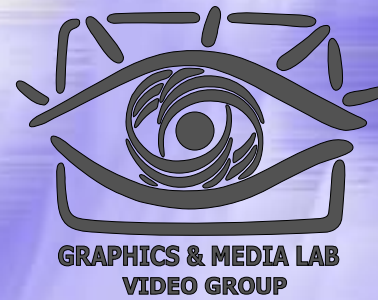
Масштабируемый объект содержит информацию о качестве представления. Когда пропускная способность потока улучшается, параметр качества увеличивается, поднимая качество отображения

# Цели: масштабируемость (продолжение)



- ◆ Масштабируемость имеет ключевое значение во многих приложениях с низким битрейтом (в мобильных устройствах)
- ◆ Применение MPEG-4 возможно для беспроводного видео со скоростью 10Kbps в GSM
- ◆ Низкий битрейт достижим, благодаря масштабируемым объектам

# Цели: использование КОМПОНЕНТОВ



- ◆ Производители могут легко управлять и организовывать компоненты, заново используя существующие, уже декодированные элементы
- ◆ Каждый тип данных может быть раскодирован с применением наиболее эффективных алгоритмов

# MPEG-4: требования

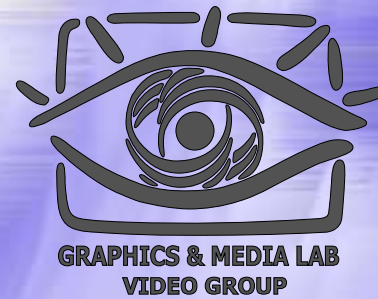
## ◆ Традиционные (MPEG-1,2):

- Поточность
- Синхронизация
- Управление потоком

## ◆ Специфические (MPEG-4)

- Аудио-визуальные объекты
- Описание сцены

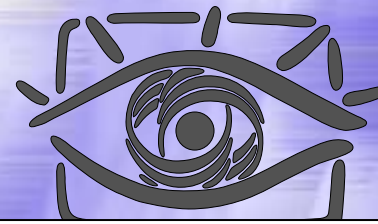
# Аудио-визуальные объекты



- ◆ Представление естественных или синтетических объектов
- ◆ Примеры:
  - Видеопоследовательность (с информацией о формах объектов)
  - Аудиотрек
  - Анимированные 3D фигуры
  - Синтезированная из текста речь
- ◆ Преимущества: интерактивность, масштабируемость, многократное использование объектов (сжатие, CPU)



# Video Object Plane (VOP)



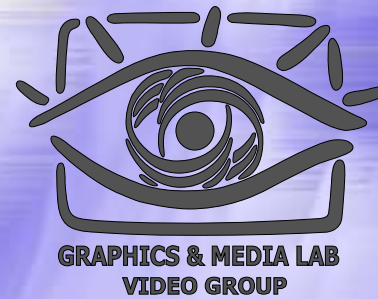
VOP1

VOP2

VOP3

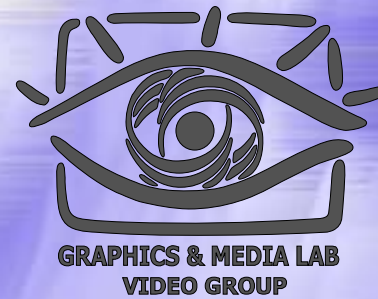


# Описание сцены



Кодирование информации, которая описывает пространственно-временные отношения между аудио-визуальными объектами

# Описание сцены (продолжение)



- ◆ Поместить медиаобъекты в нужном месте в данной координатной системе
- ◆ Применить преобразование, чтобы изменить геометрию или акустические свойства объекта
- ◆ Сгруппировать примитивы в сложные медиаобъекты
- ◆ Применить данные из потоков к объектам, изменяя их атрибуты во времени
- ◆ Изменить в интерактивном режиме точку наблюдения или звук в любом месте сцены

# Пример



Сегментация  
видеообъекта  
и маска  
объекта



VOP1

bitstream  
(VOP1)



VOP2

bitstream  
(VOP2)



VOP3

bitstream  
(VOP3)



Поуровневое  
кодирование

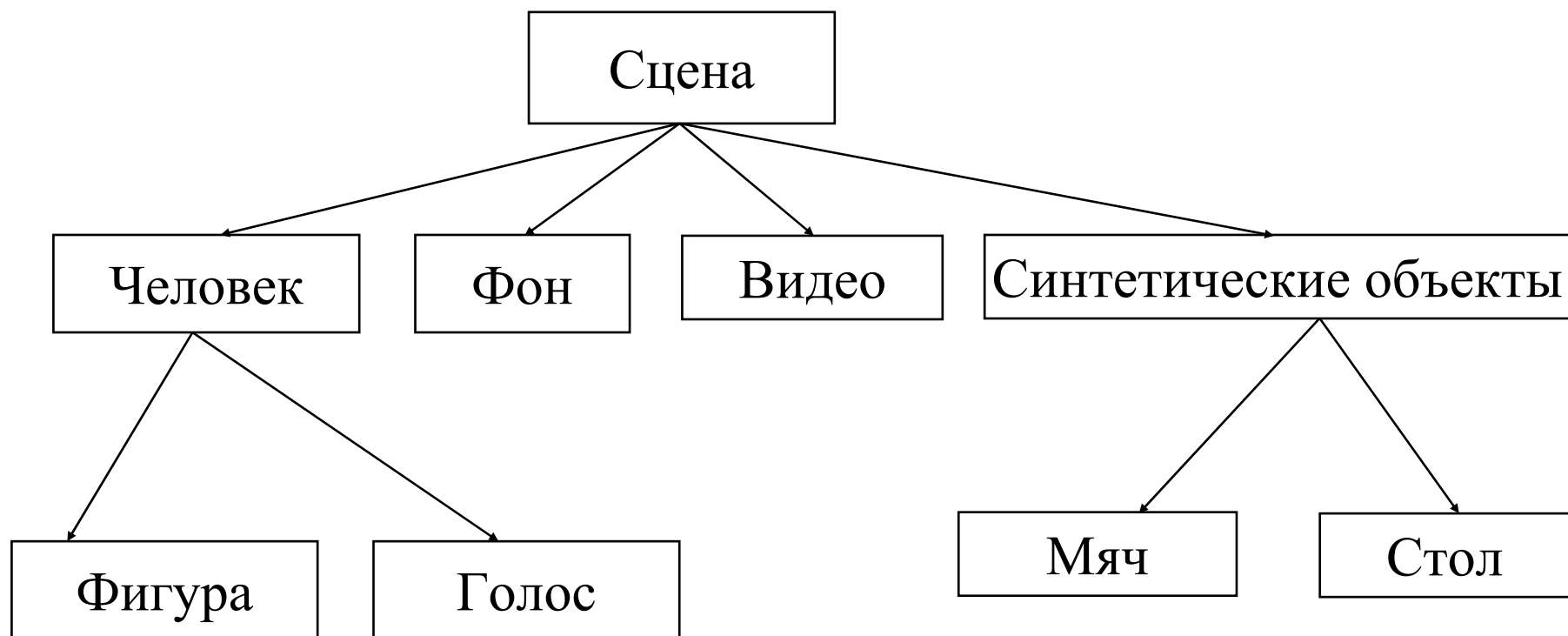
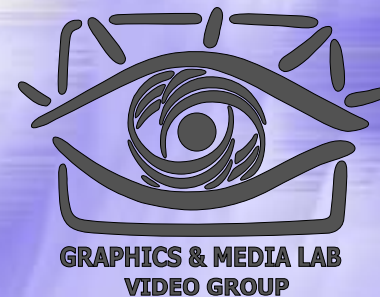
Раздельное  
декодирование

Масштабирование  
на содержательном  
уровне

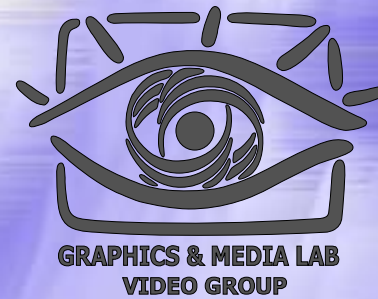


Доступ объектам  
потока и  
управление

# Логическая структура сцены

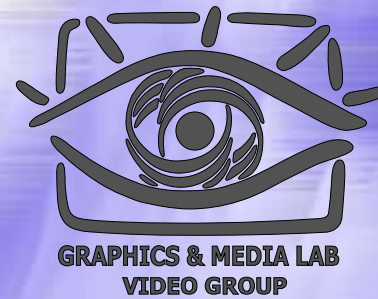


# Описание сцены (продолжение)



- ◆ Основываясь на VRML, MPEG разработал двоичный язык - BInary Format for Scenes (*BIFS*)
- ◆ Стандарт различает параметры, предназначенные для улучшения эффективности кодирования (вектора движения), и те, с помощью которых модифицируют объекты (координаты на сцене)
- ◆ Модификация объектов не предполагает их повторного раскодирования из потока

# Анимированные полигональные объекты



- ◆ Наиболее интересная особенность MPEG-4 – возможность накладывать картинки на полигональные объекты
- ◆ Деформируя эти объекты можно создать впечатление движения (развевающийся флаг; губы, произносящие слова)

# Пример

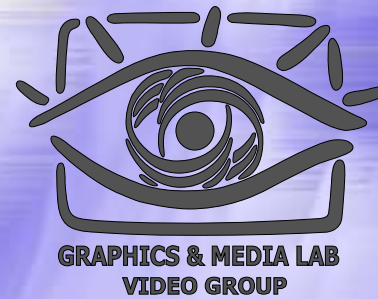




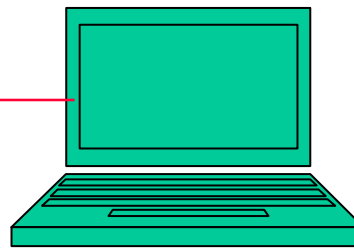
# Архитектура системы

- ◆ Поточковые данные для медиаобъектов
- ◆ Различные уровни архитектуры:
  - Уровень доставки
  - Уровень синхронизации
  - Уровень сжатия
  - Уровень композиции
- ◆ Описание синтаксиса

# Потоковые данные для медиаобъектов



- ◆ Необходимые для медиаобъектов данные конвертируются в один или несколько элементарных потоков *Elementary Streams (ESs)*.
- ◆ Дескриптор объекта (OD) определяет все потоки, ассоциированные с данным объектом
- ◆ OD содержит набор дескрипторов, которые характеризуют ESs (необходимые для декодера ресурсы, время кодирования, и т.п.)



Уровень композиции

Уровень сжатия

ODs

Сцена

Уровень синхронизации

ES

SL

Уровень доставки

DAI

DNI

FlexMux

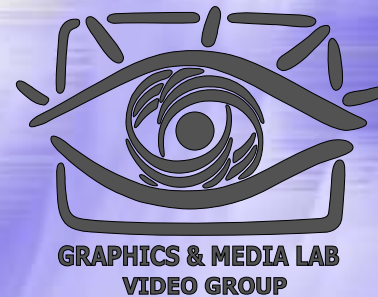
Различные транспортные протоколы

# Уровень доставки

Содержит двухуровневый мультиплексор:

- ◆ *FlexMux*: компонент, определенный в соответствии с *DMIF (Delivery Multimedia Integration Framework)*. Он позволяет группировать ESs с заголовками нижних уровней
- ◆ *TranMux*: предоставляет интерфейс с различными транспортными протоколами (UDP/IP- MPEG-2,.....)

# DMIF



- ◆ Протокол сессии для управления потоками
- ◆ Функциональность:
  - Делает независимым программный интерфейс приложения от способа получения потока (из сети, с медианосителя)
  - Контролирует установление FlexMux каналов
  - Использует сети: IP, ATM, mobile, PSTN, Narrowband ISDN.
  - Поддерживает мобильные сети, разработанные с ITU-T
  - Поддерживает пользовательские команды
  - Управляет информацией уровня синхронизации

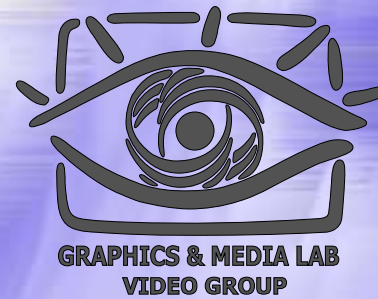
# Уровень доставки: DAI

- ◆ *DAI: DMIF Application Interface* – программный интерфейс, через который расположенные выше слои взаимодействуют с DMIF

# Уровень синхронизации

- ◆ Уровень синхронизации – это гибкое средство пакетизации, позволяющее производить фрагментацию, деление по времени и присоединять информацию к ассоциированным с ним пакетам.
- ◆ Он не предоставляет информации о кадрах (например, длина пакета). Этим занимается уровень доставки

# Уровень синхронизации: функции



- ◆ Определение отмеченных единиц доступа (*Access Units*) – единиц данных, которые полностью содержат единицы представления
- ◆ Каждый пакет – это единица доступа или фрагмент единицы доступа
- ◆ Единицы доступа формируют только семантическую структуру ESs на данном уровне
- ◆ Пометка единиц доступа включением информации о времени для декодирования и уровня композиции
- ◆ Уровень синхронизации извлекает ESs из пакетизированных ESs



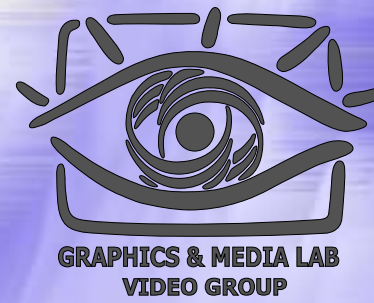
# Уровень сжатия

- ◆ Для соотнесения ESs и медиаобъектов используется дескриптор объекта, передающий информацию о числе и свойствах ESs, соответствующих объекту
- ◆ Описание сцены определяет:
  - Положение объектов в пространстве и времени
  - Динамическое поведение объектов
  - Интерактивные свойства объектов
- ◆ Описание сцены содержит набор уникальных идентификаторов, которые указывают на дескрипторы объектов
- ◆ Описание сцены имеет древовидную структуру и базируется на VRML

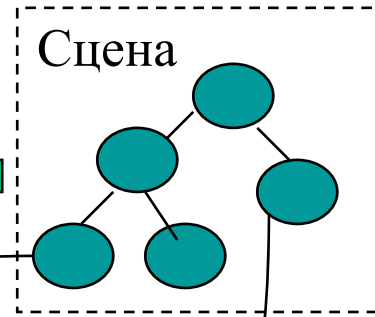
# Уровень композиции

- ◆ Используются описание сцены и декодированные аудио-визуальные объекты для прорисовки финальной сцены
- ◆ MPEG-4 не определяет, как именно производится прорисовка

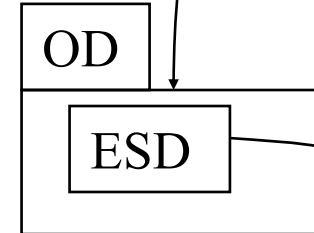
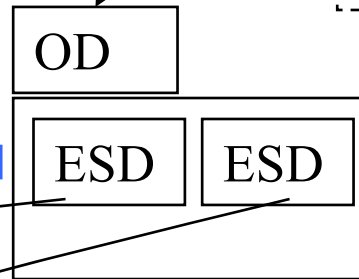
# Пример



ES описания сцены



ES дескриптора объекта



ES  
ES  
ES



ES – элементарный поток  
OD – дескриптор объекта

# Профили (profiles)

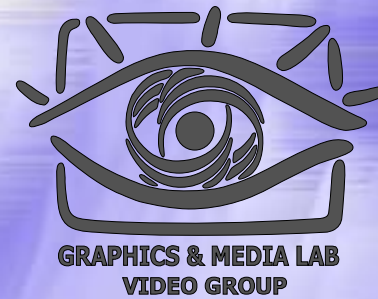
- ◆ Визуальные профили
- ◆ Аудио профили
- ◆ Графические профили
- ◆ Профили графики сцены
- ◆ MPEG-J профили
- ◆ Профили дескриптора объекта

# Визуальные профили



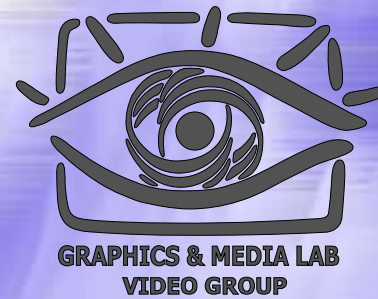
Визуальная часть стандарта определяет профили для реального, синтетического и смешанного видео

# Визуальные профили для реального видео



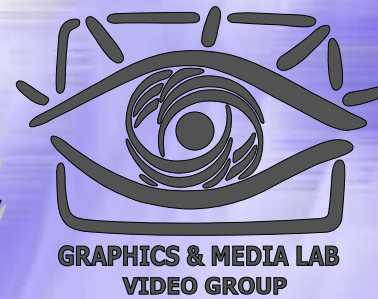
- ◆ **Simple Visual Profile** обеспечивает эффективное, устойчивое к ошибкам кодирование прямоугольных видеообъектов применимый в мобильных сетях PCS и IMT2000.
- ◆ **Simple Scalable Visual Profile** добавляет к предыдущему поддержку временной и пространственной масштабируемости
- ◆ **Core Visual Profile** добавляет поддержку кодирования масштабируемых по времени объектов произвольной формы
- ◆ **Main Visual Profile** добавляет поддержку кодирования интерлейсинговых, полупрозрачных и спрайтовых объектов
- ◆ **N-Bit Visual Profile** добавляет поддержку кодирования видеообъектов, имеющих глубину пикселя в пределах от 4 до 12 бит

# Профили синтетического и смешанного видео



- ◆ **Simple Facial Animation Visual Profile** предоставляет простые средства для анимации модели лица
- ◆ **Scalable Texture Visual Profile** – пространственно масштабируемое кодирование для текстурных объектов
- ◆ **Basic Animated 2-D Texture Visual Profile** – пространственно и SNR масштабируемая полигональная анимация текстурных объектов
- ◆ **Hybrid Visual Profile** сочетает возможность кодирования масштабируемых по времени естественных объектов произвольной формы, с возможностью кодирования синтетических и смешанных объектов

# Новые профили в версии 2

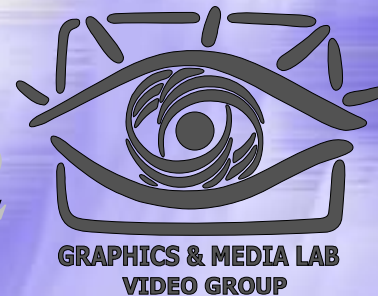


## Реальное видео:

- ◆ **The Advanced Real-Time Simple Profile (ARTS)** – улучшенная устойчивость к ошибкам
- ◆ **The Core Scalable Profile** добавляет поддержку пространственной масштабируемости объектов к Core Profile.
- ◆ **The Advanced Coding Efficiency (ACE)** улучшает эффективность кодирования для объектов произвольной формы



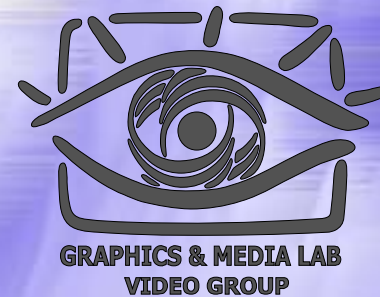
# Новые профили в версии 2



Профили синтетического и смешанного видео:

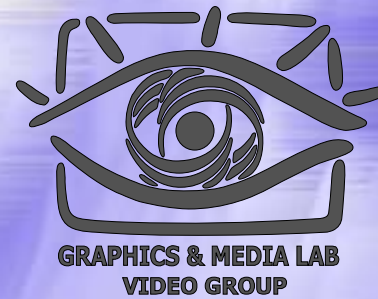
- ◆ **The Advanced Scaleable Texture Profile** поддерживает кодирование текстурных объектов произвольной формы, включая временное масштабирование, клеточное представление с помощью вейвлетов и устойчивость к ошибкам
- ◆ **The Advanced Core Profile** сочетает возможность кодировать видеообъекты произвольной формы с возможностью кодировать неподвижные картинки произвольной формы
- ◆ **The Simple Face and Body Animation Profile** расширяет Simple Face Animation Profile добавлением анимации тела

# В последующих версиях добавлены



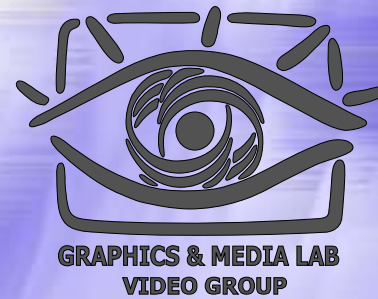
- ◆ **Advanced Simple Profile** по сравнению с **Simple Profile** добавлены важные вещи: B-frames, j пиксельная компенсация движения, добавочная таблица квантования и глобальная компенсация движения
- ◆ **Fine Granularity Scalability Profile** позволяет урезать битовый поток в любой битовой позиции, что дает возможность легко адаптироваться к скорости передачи и декодирования
- ◆ **Simple Studio Profile** – профиль с очень высоким качеством для студийного использования. Видео содержит только I-фреймы. Битрейт достигает 2 Gbps.
- ◆ **Core Studio Profile** добавляет P фреймы к Simple Studio

# Аудиопрофили



- ◆ **Speech Profile**
- ◆ **Synthesis Profile**
- ◆ **Scalable Profile**
- ◆ **Main Profile**
- ◆ **High Quality Audio Profile**
- ◆ **Low Delay Audio Profile**
- ◆ **Natural Audio Profile**
- ◆ **Mobile Audio Internetworking Profile (MAUI)**

# Графические профили



- ◆ **Simple 2-D Graphics Profile**
- ◆ **Complete 2-D Graphics Profile**
- ◆ **Complete Graphics Profile**
- ◆ **The 3D Audio Graphics Profile**

# Профили графики сцены

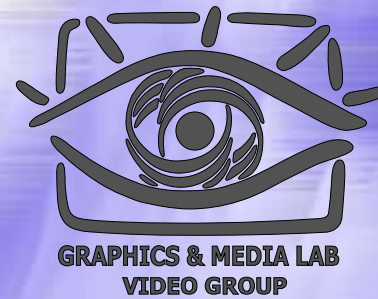


- ◆ **The Audio Scene Graph Profile**
- ◆ **The Simple 2-D Scene Graph Profile**
- ◆ **The Complete 2-D Scene Graph Profile**
- ◆ **The Complete Scene Graph Profile**
- ◆ **The 3D Audio Scene Graph Profile**

# MPEG-J профили

- ◆ **Personal** простой пакет для персональных устройств. Содержит следующие пакеты MPEG-J APIs:
  - Сеть
  - Сцена
  - Ресурсы
- ◆ **Main** включает все MPEG-J APIs

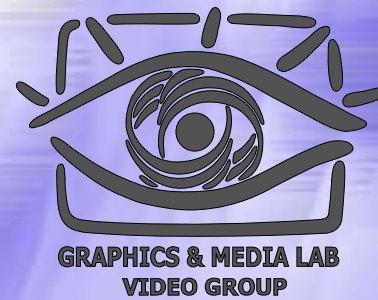
# Профиль дескрипторов объектов



Профиль дескриптора объекта включает следующие инструменты:

- ◆ Object Descriptor (OD) tool
- ◆ Sync Layer (SL) tool
- ◆ Object Content Information (OCI) tool
- ◆ Intellectual Property Management and Protection (IPMP) tool

# Профили (дополнение)



**Базовая линия**

**Compression**

**Error Resilience**

**Scalability**

Обычное кодирование

**Расширение**

**Content-based  
Coding**

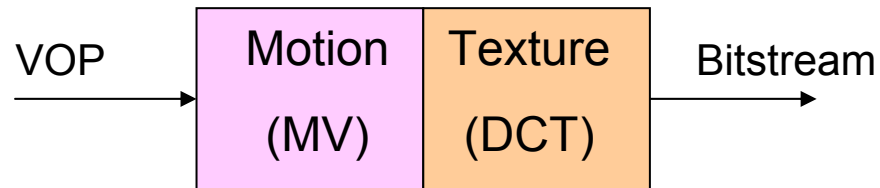
**Still Texture  
Coding**

Объектное кодирование

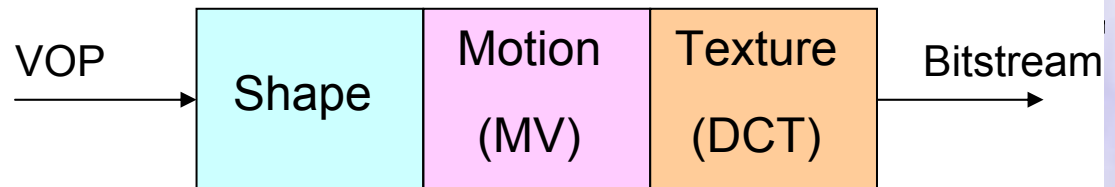


# Профили (дополнение)

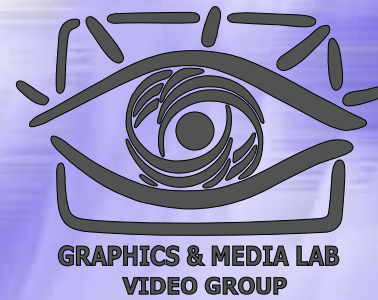
## MPEG-4 Core Coder



## Extended MPEG-4 Core Coder

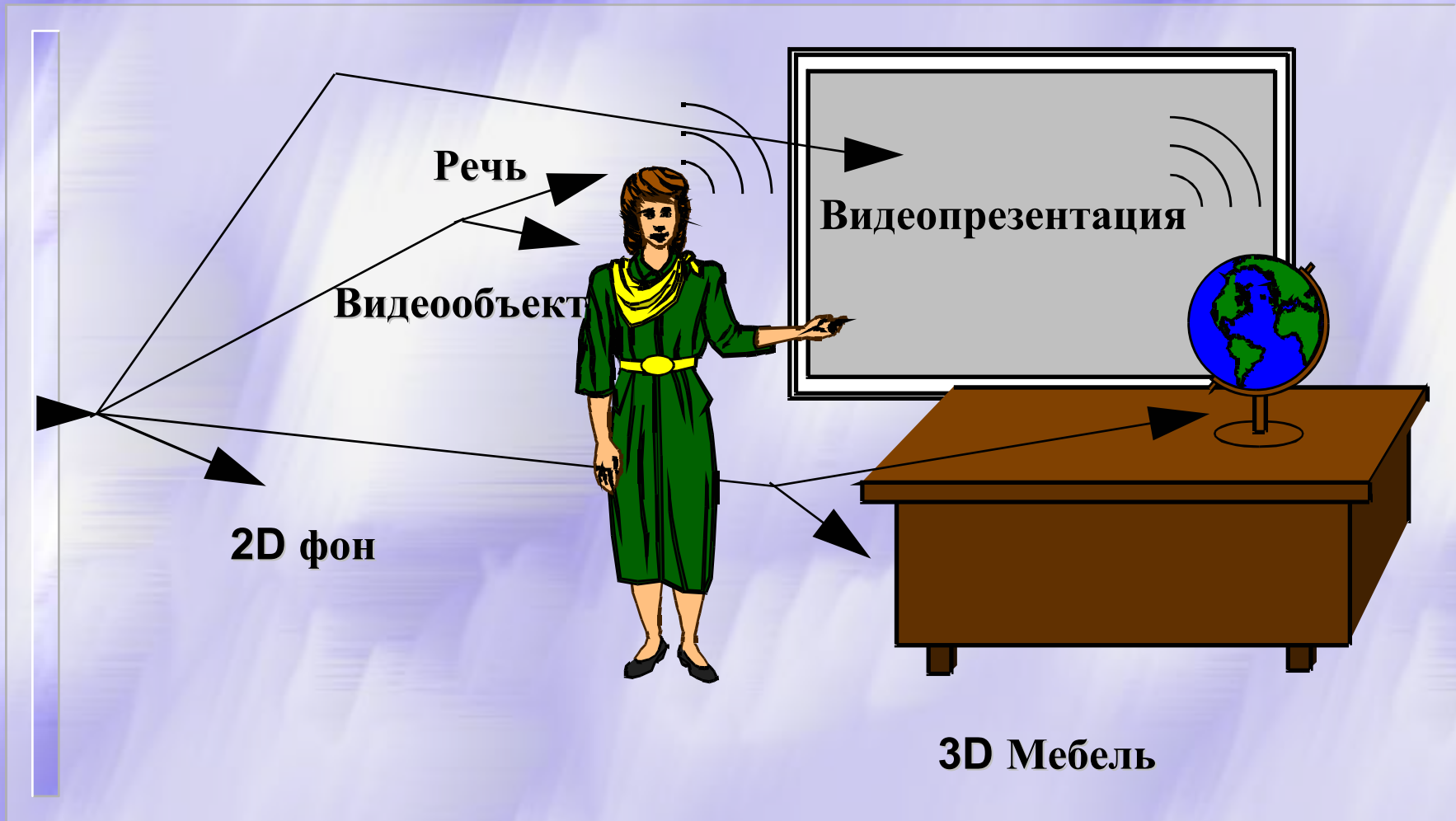


# Содержательное кодирование

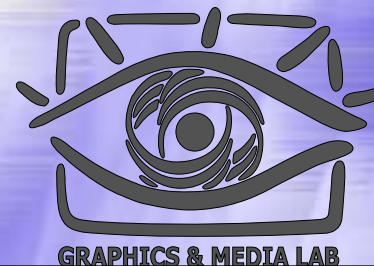


- ◆ Позволяет пользователю получить доступ к объектам на кодированной сцене
- ◆ Предоставляет высокий уровень взаимодействия пользователя с содержимым
- ◆ Манипуляции с содержанием происходят на уровне потоков данных

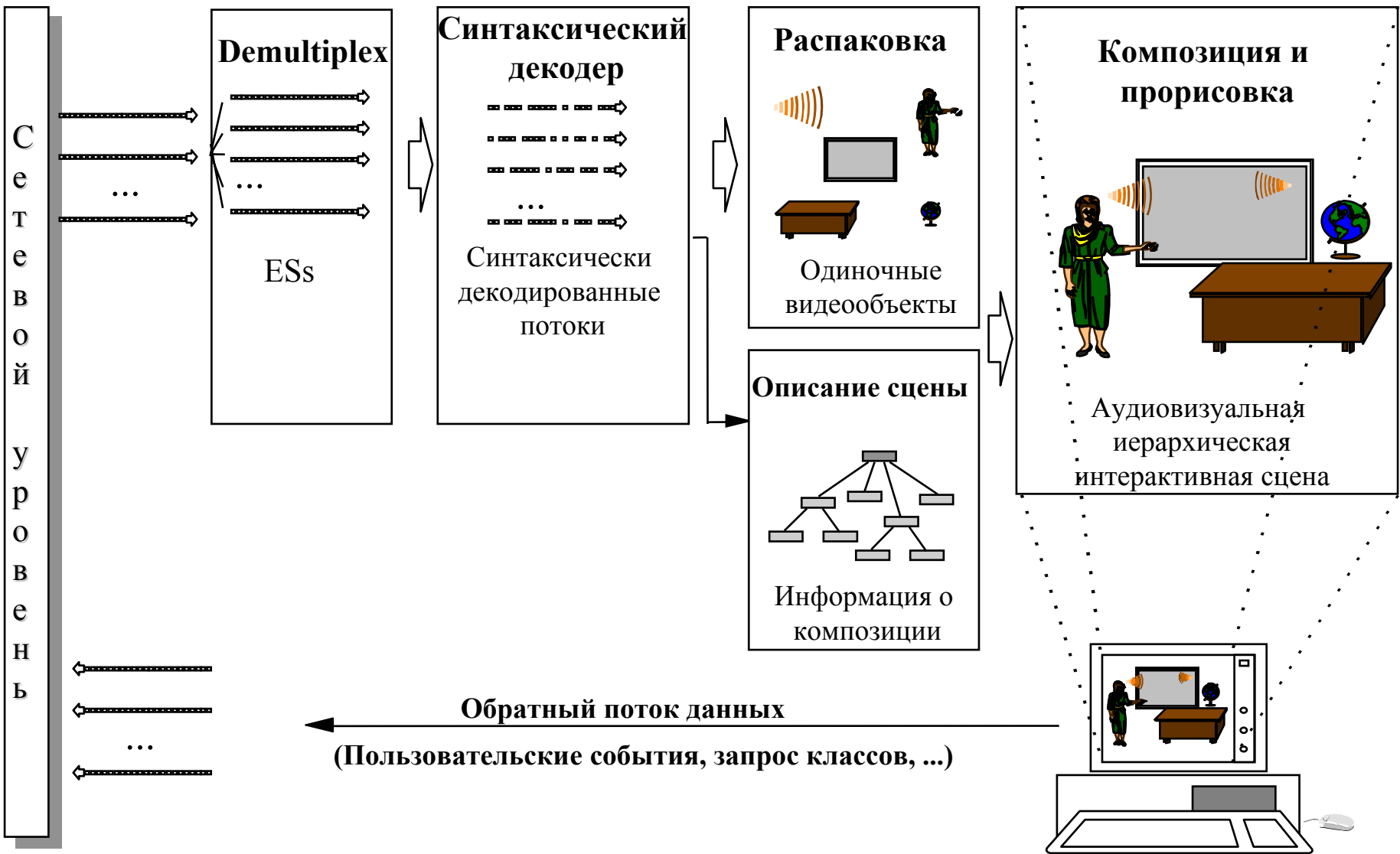
# Сцена в MPEG-4



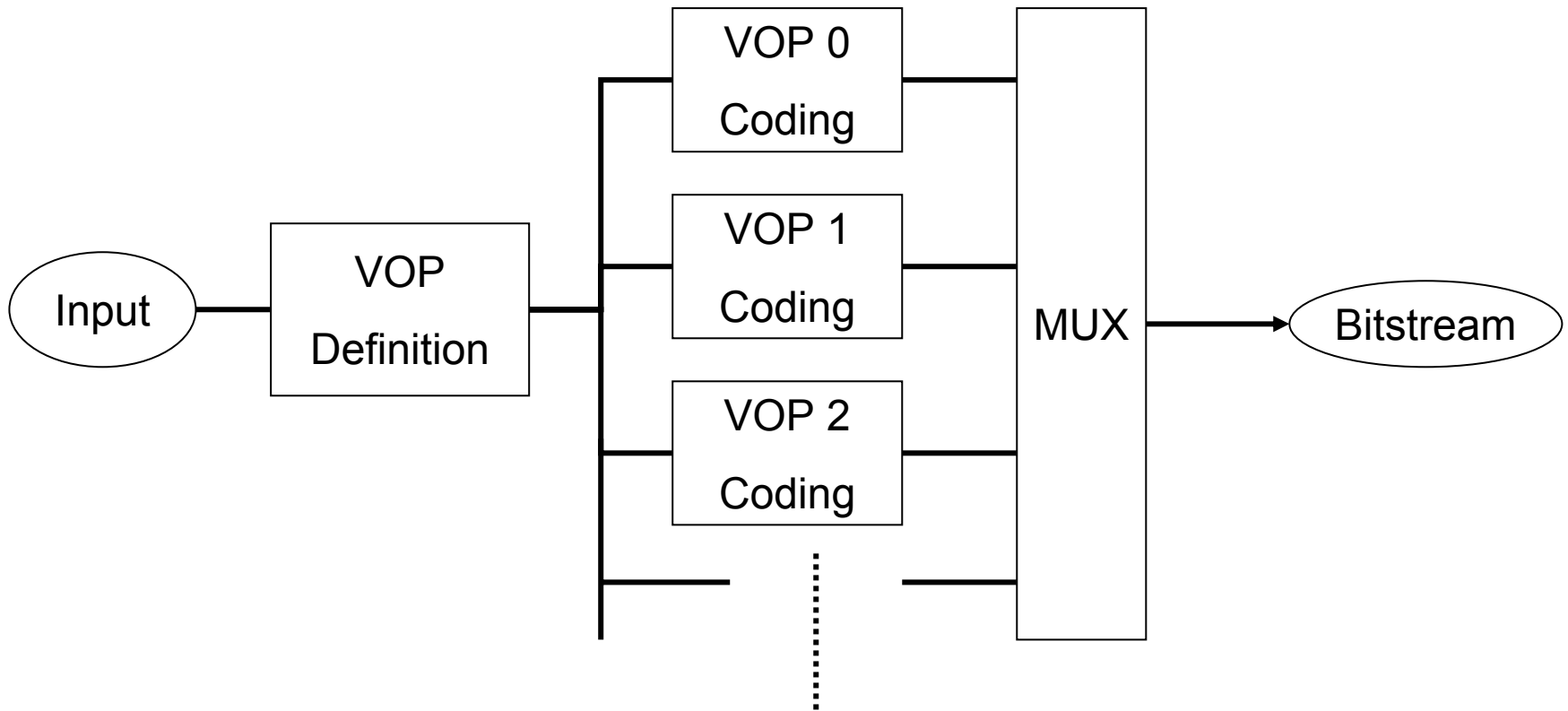
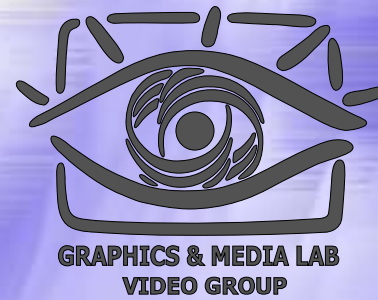
# MPEG-4 терминал



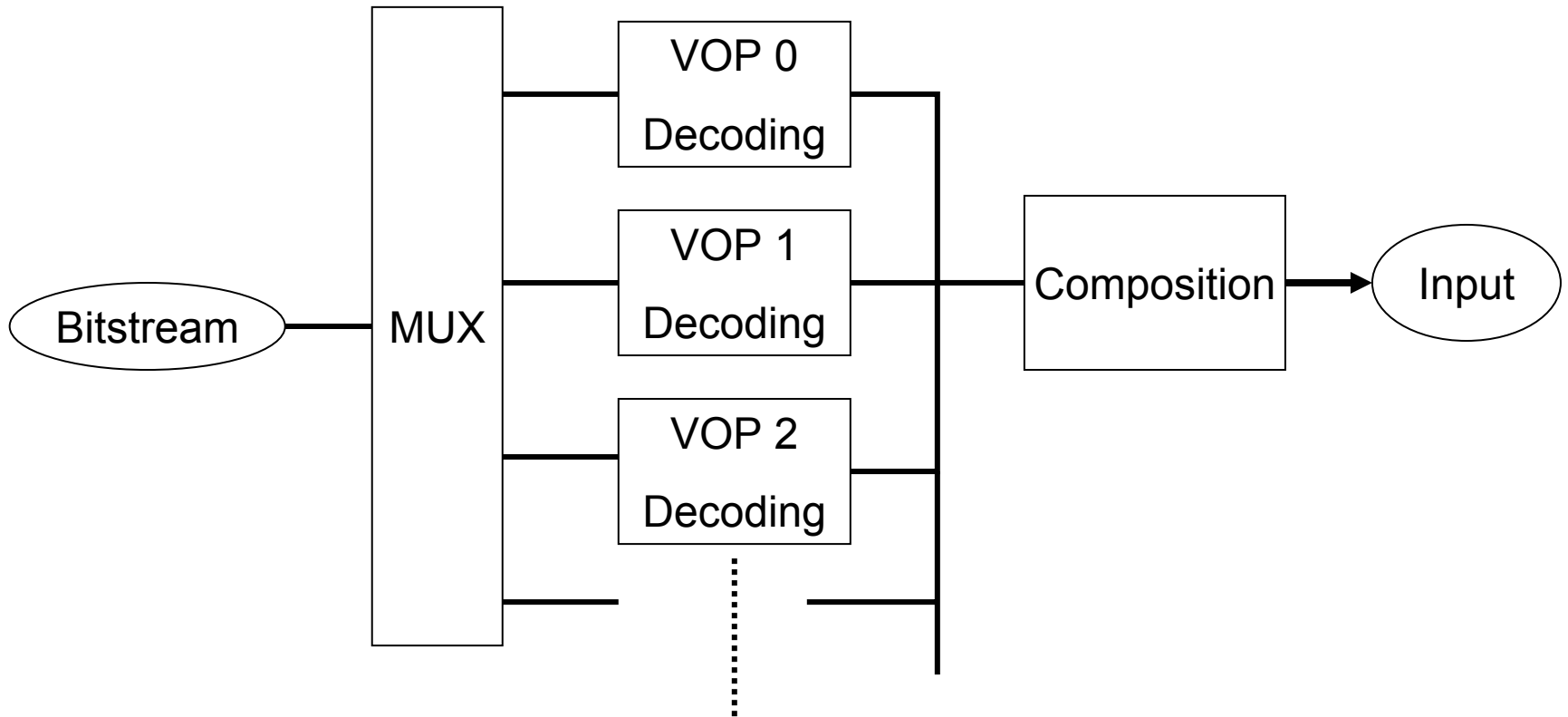
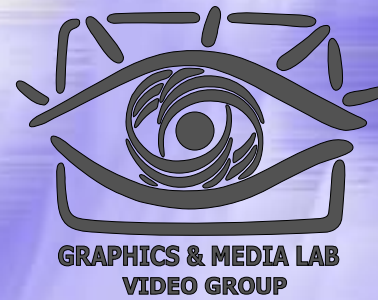
GRAPHICS & MEDIA LAB



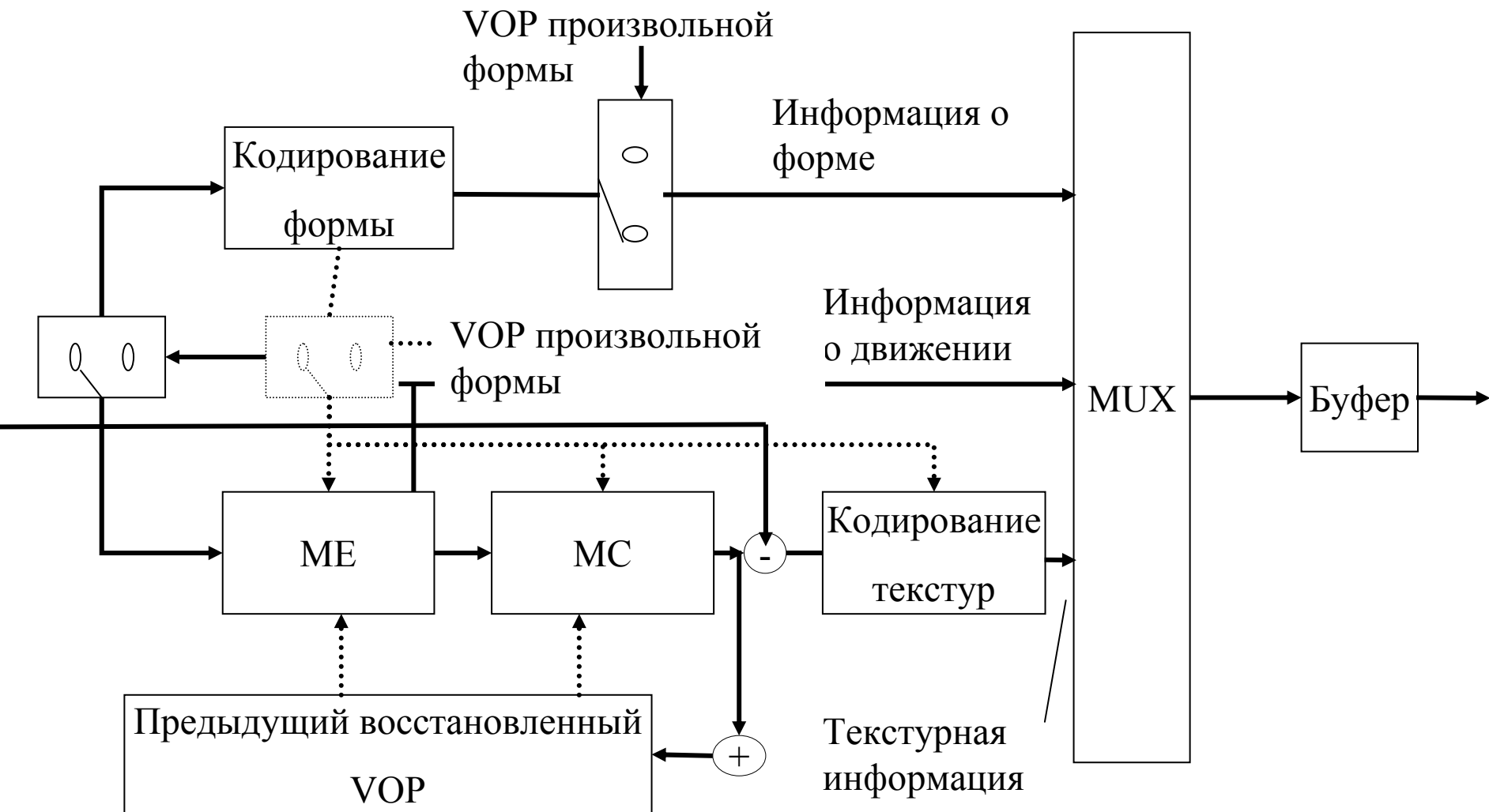
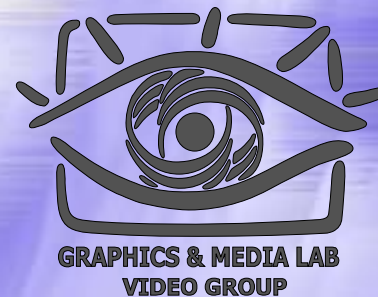
# Структура VOP кодера



# Структура VOP декодера



# VOP кодер



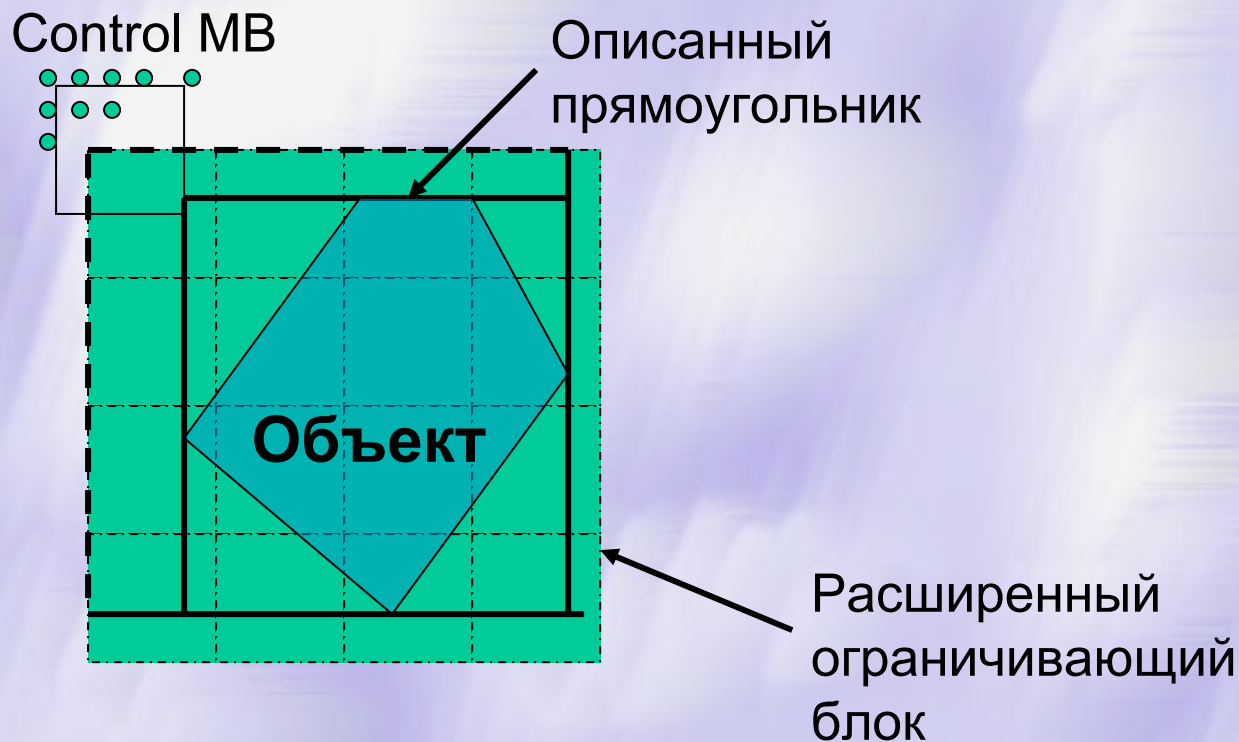
# VOP декодер





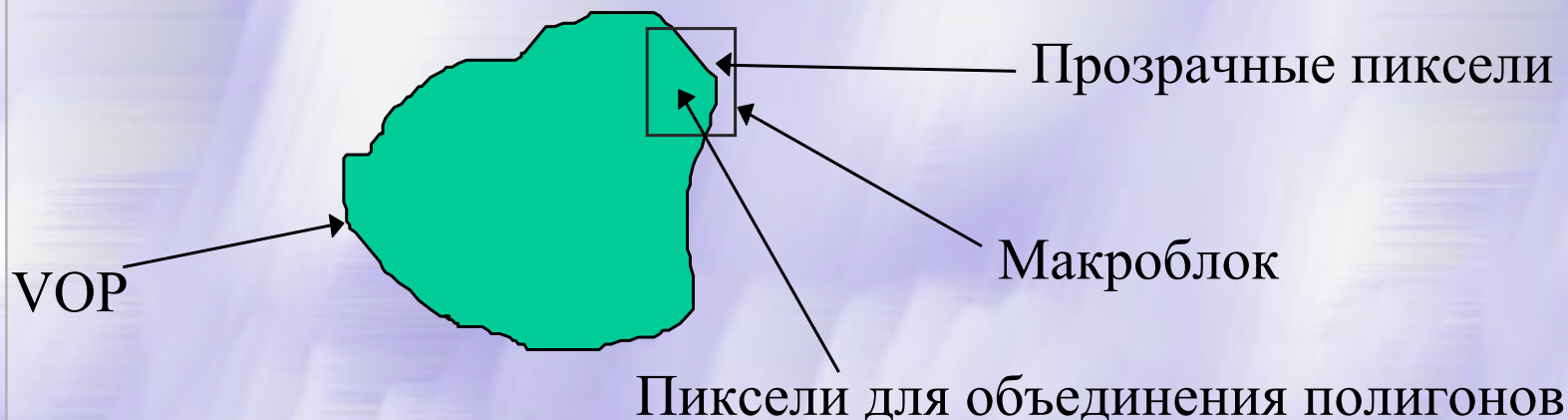
# Формирование VOP

## Минимизация числа обрабатываемых макроблоков (MB)



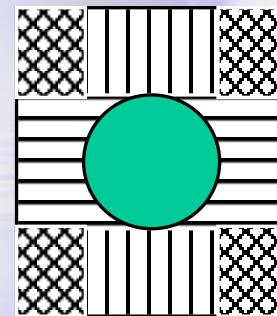
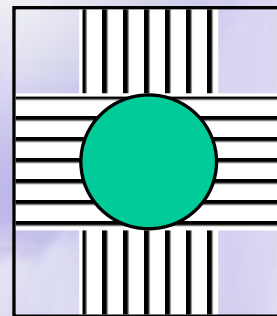
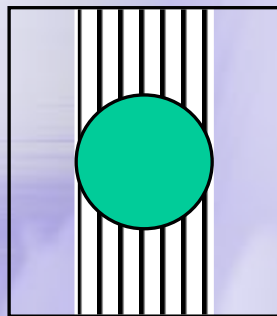
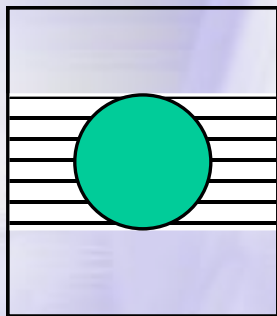
# Кодирование текстур

- ◆ DCT после компенсации движения
- ◆ Объединение полигонов



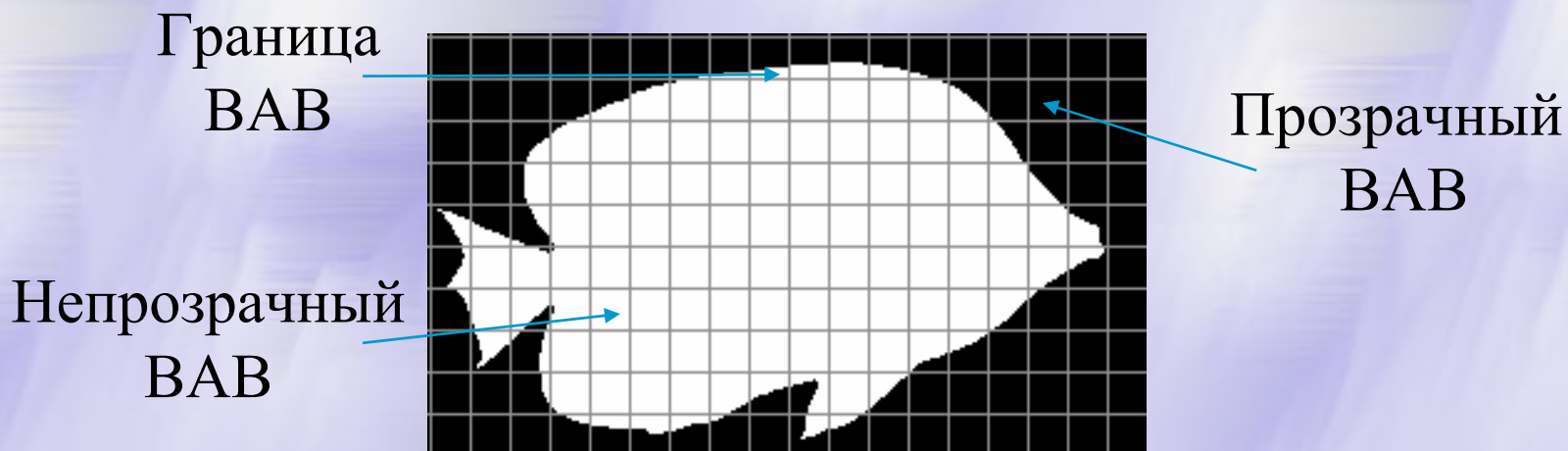
# Заполнение VOP для объединения блоков

- ◆ Повторное заполнение VOP необходимо для ME/МС



# Кодирование формы

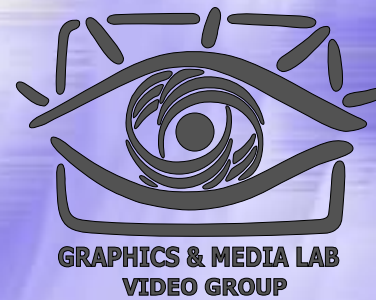
- ◆ Базовая единица кодирования формы  
VOP в MPEG-4 – бинарный альфа блок  
(binary alpha block - BAV)
- ◆ Объединение BAV



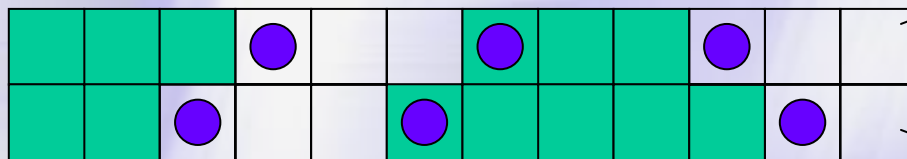
# Кодирование формы

- ◆ Кодирование формы, основанное на битовых картах:
  - Модифицированное MMR кодирование
  - Арифметический контекстный кодер (Context-based Arithmetic Encoder (CAE))
- ◆ Кодирование формы, основанное на контурах:
  - Кодирование по вершинам
  - Кодирование по базовым линиям

# Кодирование по битовым картам



## Модифицированный MMR метод

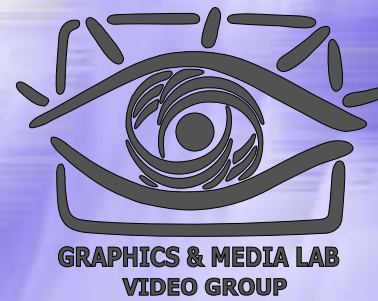


Предыдущая  
линия

Кодируемая  
линия

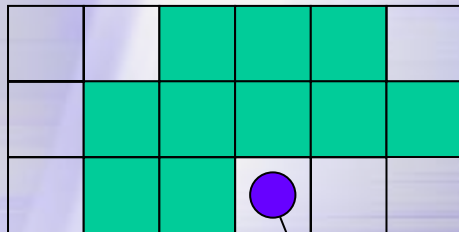
● - пиксель изменения

# Кодирование по битовым картам



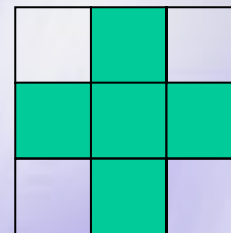
## Арифметический контекстный кодер

### Интра-фрейм

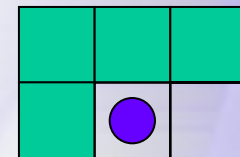


Текущий  
пиксель

### Интер-фрейм

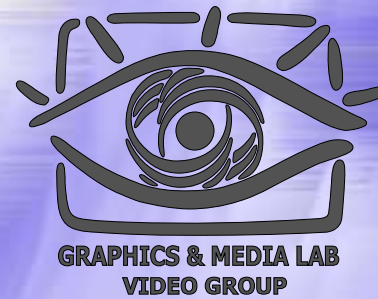


Предыдущий



Текущий

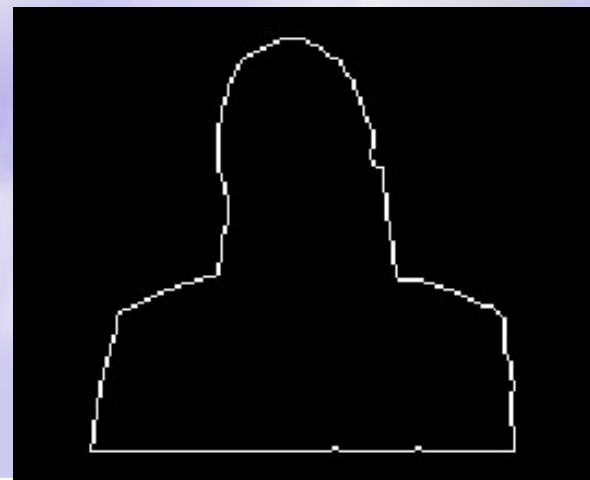
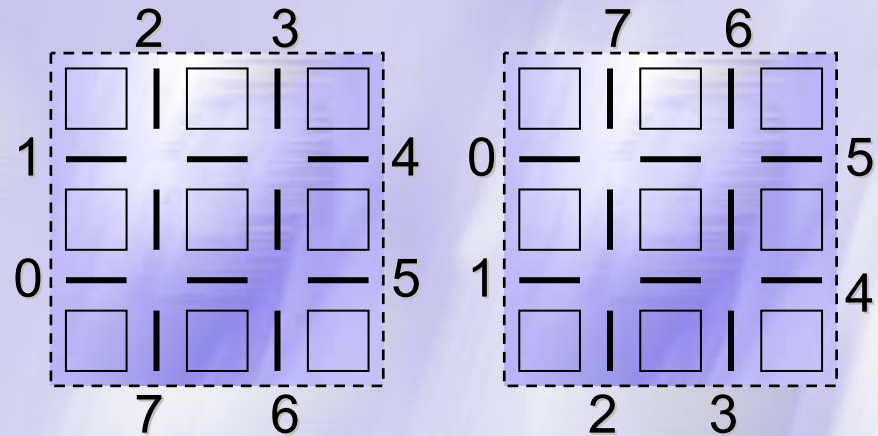
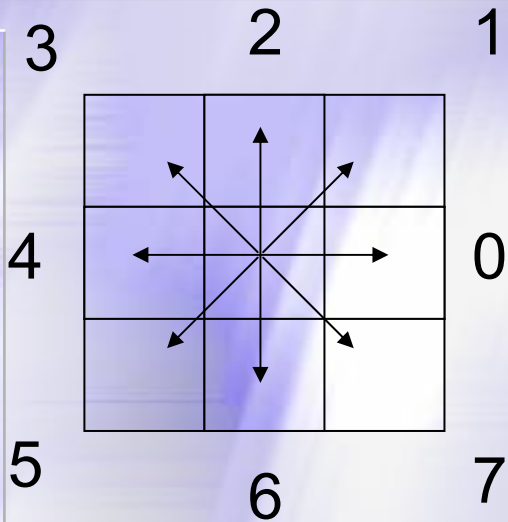
# Кодирование формы по контуру



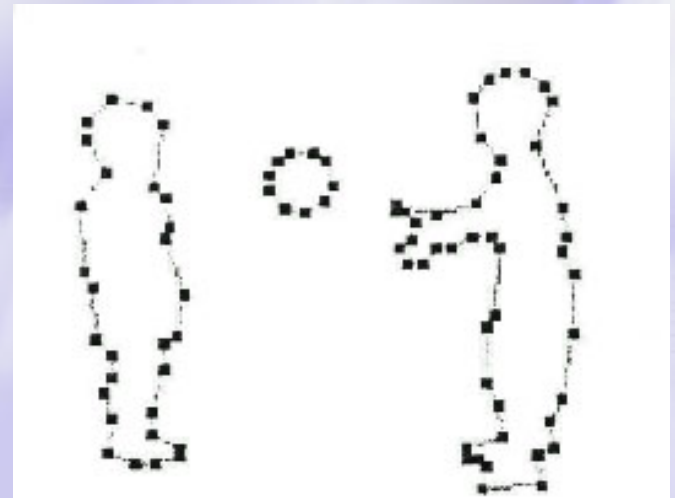
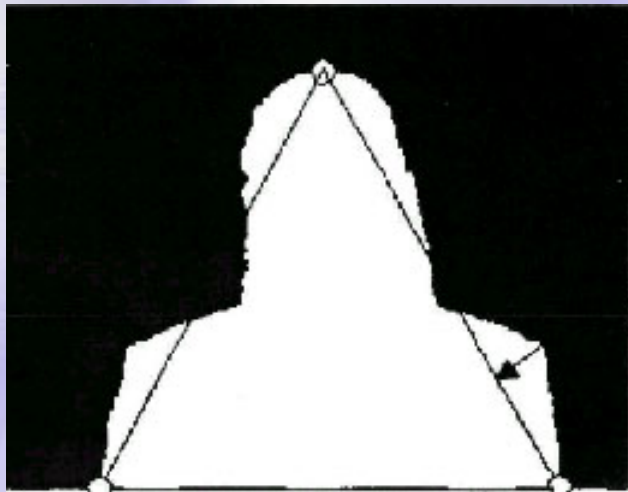
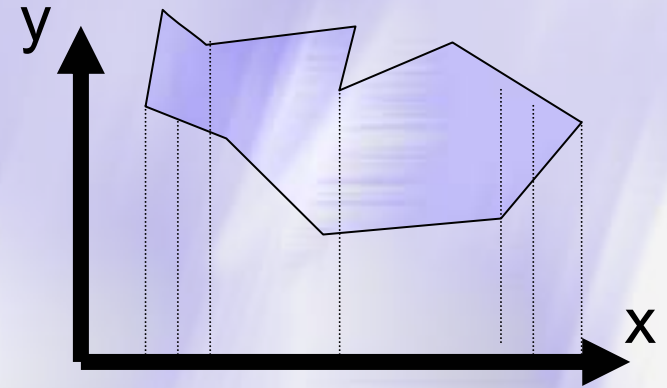
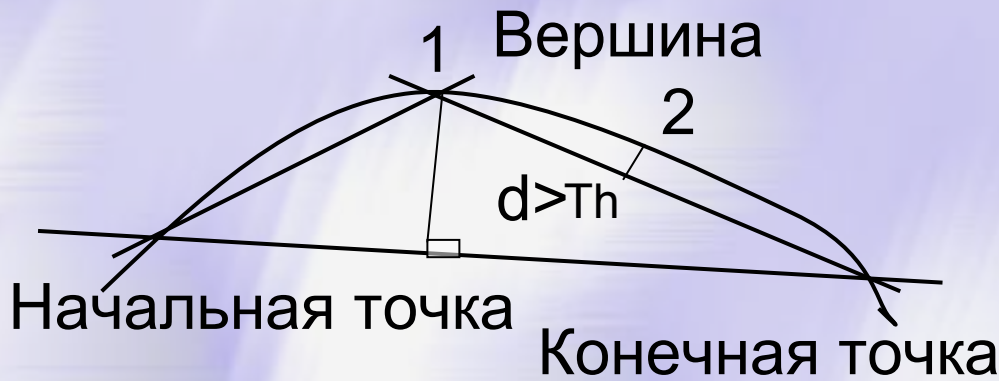
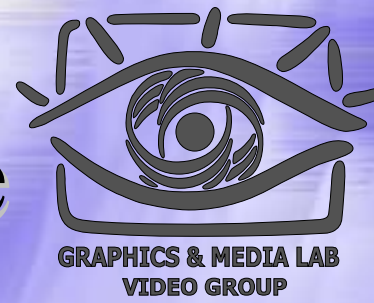
- ◆ Chain code : использует 4 или 8 направлений, чтобы описать контур
- ◆ MGCC code : многосеточный цепной код
- ◆ Vertex code : находит подобный многогранник
- ◆ Baseline code : используется проекция фигуры на ось  $X$ , кодируется расстояние (координата  $y$ )



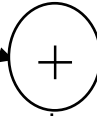
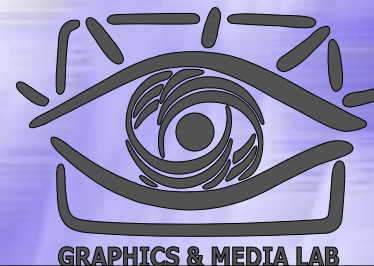
# Chain code и MGCC



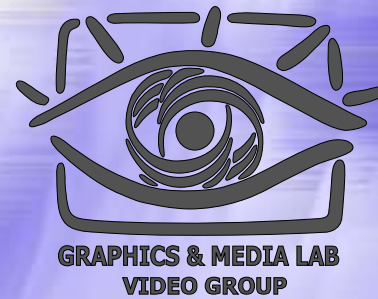
# Vertex code и Baseline code



# Кодирование со спрайтами



# Синтаксическое описание



- ◆ MPEG-4 определяет язык синтаксического описания (MSDL) для битовых потоков кодированных объектов и потоков описания сцены
- ◆ MSDL – расширение C++, он нужен чтобы описать синтаксическое представление кодируемых объектов

# MPEG-4: инструменты



- ◆ Управление потоками: The Object Description Framework (ODF)
- ◆ Синхронизация: The System Decoder Model (SDM)
- ◆ Механизм представления: Binary Format for Scenes (BIFS)

# Инструменты: ODF

- ◆ Предоставляет механизм связывания описания сцены с элементарными потоками
- ◆ В описании сцены используются уникальные идентификаторы для указания на дескрипторы объектов
- ◆ Дескрипторы объектов передаются в соответствующих элементарных потоках

# Инструменты: SDM

- ◆ Адаптация для MPEG-2 System Target Decoder (описывает ограничения на время и буферы для пакетных элементарных потоков).
- ◆ MPEG-4 не определяет в SDM ограничения на мультиплексирование
- ◆ SDM основывается на конкурентной доставке демультимплексированных элементарных потоков в буфер декодера

# Инструменты: BIFS

- ◆ BIFS - BInary Format for Scenes
- ◆ Используется для описания информации о декомпозиции сцены:
  - Пространственное и временное положение объектов
  - Атрибуты и поведение объектов
  - Отношения между объектами в графе сцены
- ◆ Основывается на VRML



# Протоколы BIFS

- ◆ Сжатие сцены с BIFS (текстовое или битовое)
- ◆ Команды BIFS:
  - Заменить (Replace) всю сцену новой
  - Вставить (Insert) вершину
  - Удалить (Delete) вершину
  - Изменить (Change) значение поля вершины
- ◆ BIFS Anim (используется для анимации сцены) позволяет изменять любые объекты в сцене: точку наблюдения, форму объектов, цвет, освещение

# MPEG-4: версия 2

- ◆ Intellectual Property Management & Protection (IPMP)
- ◆ Улучшенный BIFS
- ◆ MPEG-4 формат файла
- ◆ MPEG-J
- ◆ Кодирование 3D meshes
- ◆ Анимация тела

# Улучшенный BIFS

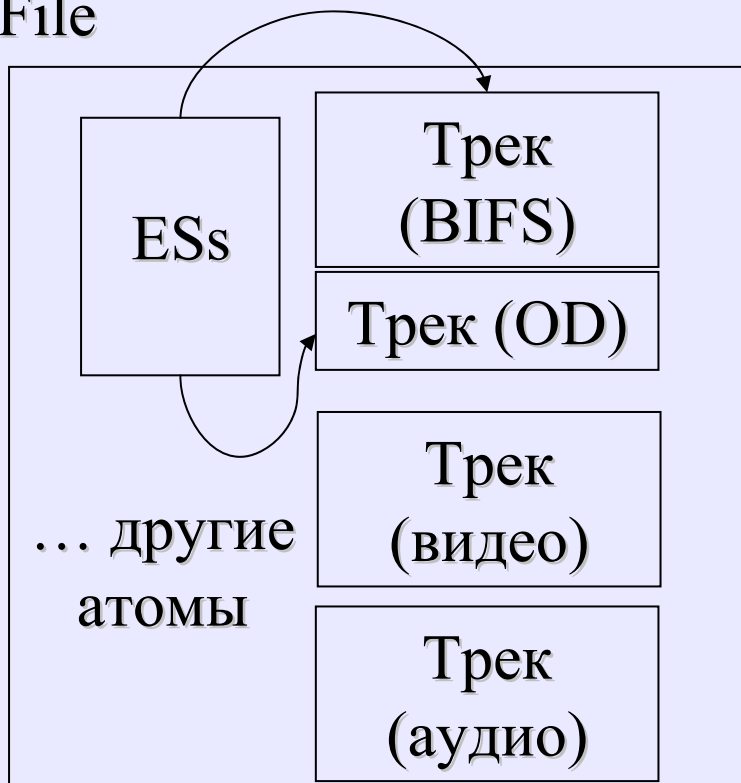
- ◆ Многопользовательский доступ к сцене
- ◆ Улучшенный звук
- ◆ Анимация лица и тела
- ◆ Proto- , Externproto- и Script VRML конструкции
- ◆ Другие VRML функции, не включенные в первую версию

# Формат файла MPEG-4

- ◆ Разработан, чтобы хранить информацию, представленную в MPEG-4, в гибком и расширяемом формате
- ◆ Основывается на формате QuickTime<sup>®</sup>
- ◆ Составлен из объектно-ориентированных структур, названных атомами

# Простой файл обмена

mp4 File

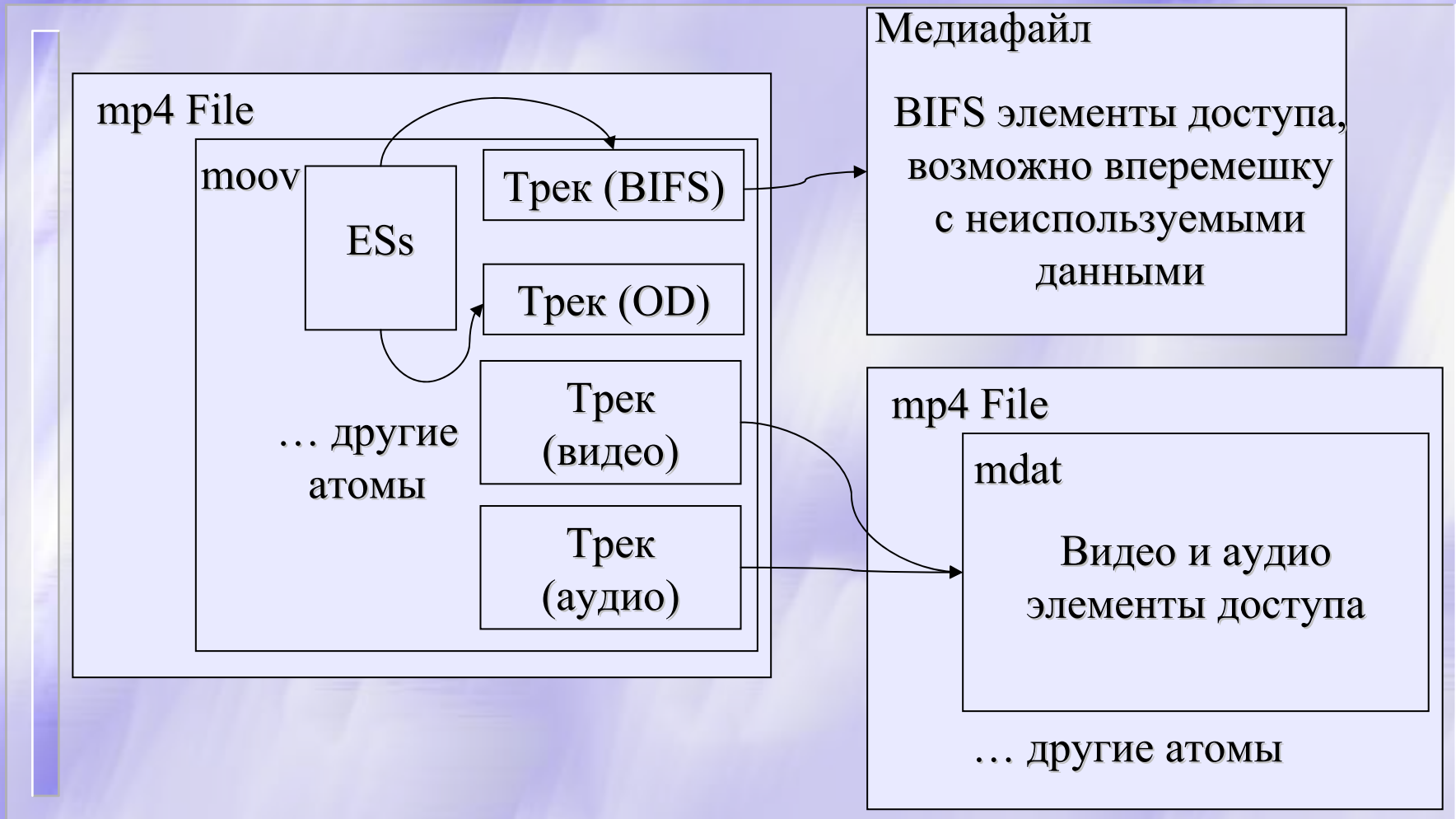
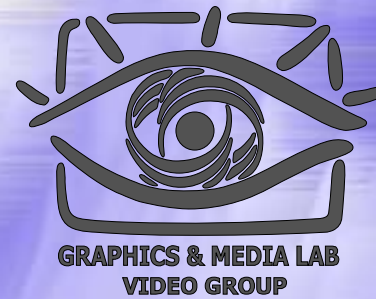


mdat

Слоевые, временные,  
BIFS, OD, видео и  
аудио элементы  
доступа

OD – дескриптор объекта; ESs – элементарные потоки  
BIFS - BInary Format for Scenes

# Более сложный файл со внешними медиаданными

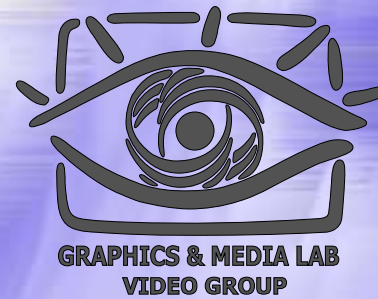


# MPEG-J



- ◆ Спецификация Java API в MPEG-4
- ◆ Позволяет использовать сложный контроль и интеллектуальную обработку для управления аудио-визуальной сессией
- ◆ Java-приложения доставляются в отдельных элементарных потоках и передаются среде исполнения

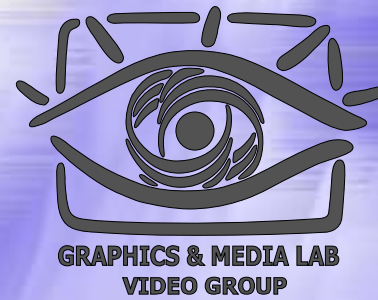
# MPEG-4: устойчивость к ошибкам



MPEG-4 предоставляет набор специальных инструментов для устранения ошибок, возникающих при передаче.



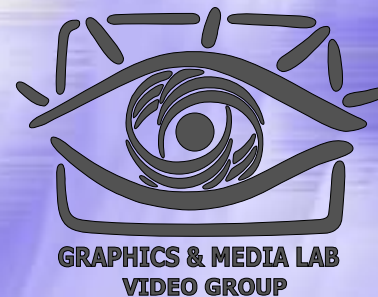
# MPEG-4: устойчивость к ошибкам (2)



Средства для локализации и устранения ошибок:

- Ресинхронизационные маркеры
- Разделение данных
- Защита заголовка
- Обратимая изменяемая длина кода

# Локализация ошибок



- Ресинхронизационные маркеры

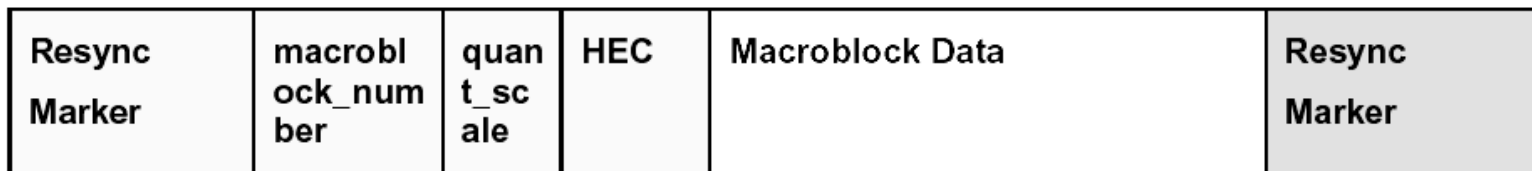


Figure E-1 -- Error Resilient Video Packet

- Разделение данных

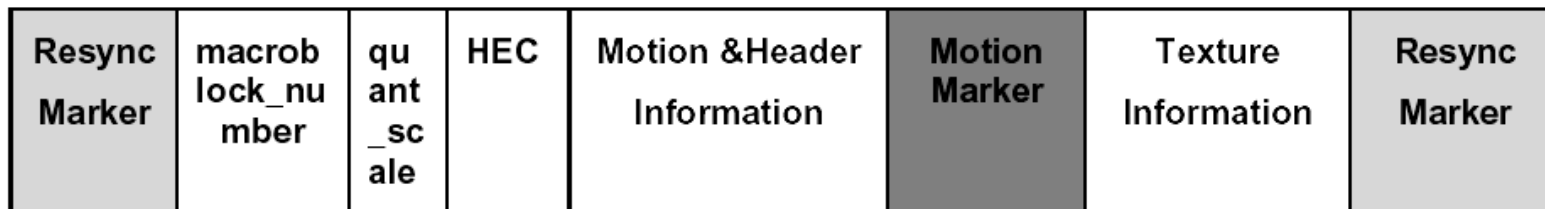
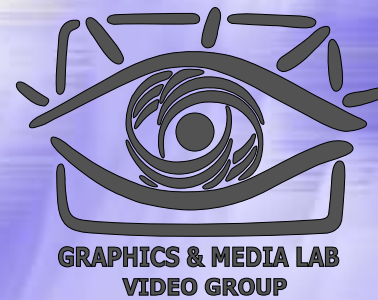
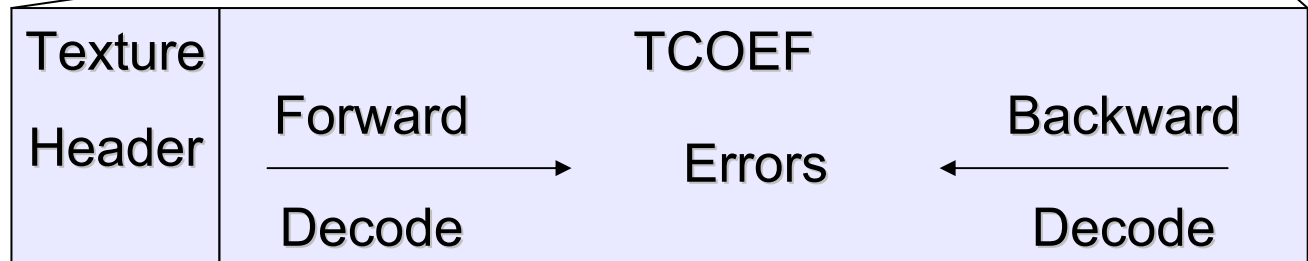


Figure E-2 -- Data Partitioning

# Локализация ошибок (2)



Resync Marker	Macroblock _number	Quant _scale	HEC	Motion & Header Information	Motion Marker	Texture Information	Resync Marker
------------------	-----------------------	-----------------	-----	--------------------------------	------------------	------------------------	------------------





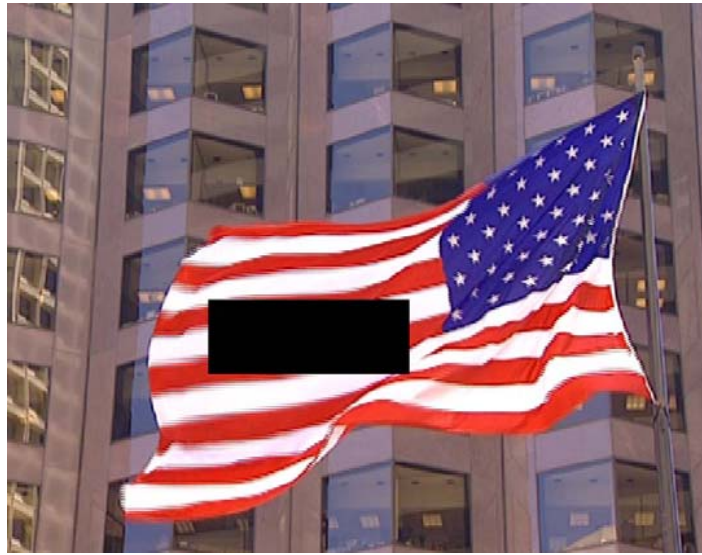
Исходное видео



Границы видеопакетов



Распространение ошибки  
ввода-вывода без видеопакетов



Та же ошибка с видеопакетами

# Литература



- ◆ Mei-Juan Chen “Video Coding”
- ◆ Rob Koenen “MPEG-4 Overview”
- ◆ Steven Gringeri ,Roman Egorov, Khaled Shuaib, Arianne Lewis, Bert Basch “Robust Compression and Transmission of MPEG-4 Video”
- ◆ W.G. Aref “MPEG-4-WWW”
- ◆ Schafer, R. “MPEG-4: a multimedia compression standard for interactive applications and services”