An Overview of MPEG4

Thanks for slides preparation of Dr. Shawmin Lei, Sharp Labs of America
And, Mei-Yun Hsu
February 1999

Material Sources

- The MPEG-4 Tutuorial, San Jose, March 1998
  - MPEG-4: Context and Objectives - Rob Koenen
  - Natural Video in MPEG-4 - Thomas Sikora
  - MPEG-4 Natural Video Tools - Touradj Ebrahimi
- Thanks to all of them

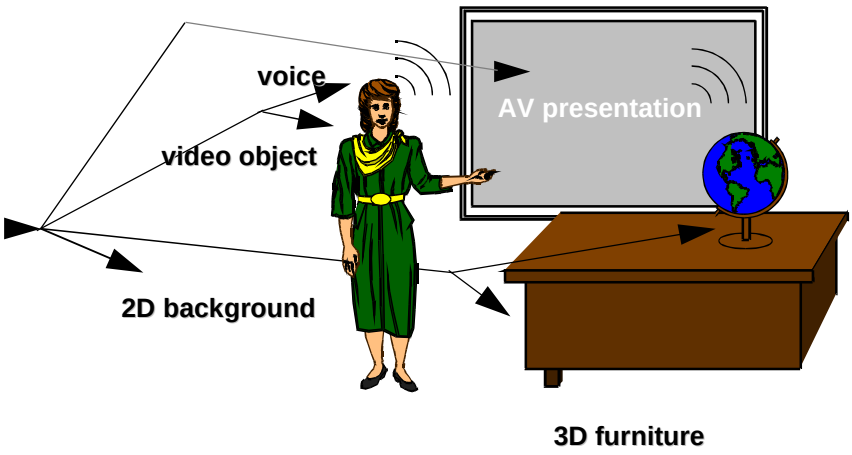
Goal of MPEG-4

- MPEG-4: '**Coding of Audio-Visual Objects**'
- *One* generic toolbox for *many different* kinds of applications (e.g. both conversational, interactive, and broadcast)
- Support for *a* **new kind of interactivity**: based on content and meaning
- Compression no longer the only reason for doing coding!
  - 'Low Bitrate', although still important, is not MPEG4's only focus !

MPEG-4: Coding of Audiovisual Objects

- Audiovisual Scene is composed of '**Objects**' (A&V)
- '**Compositor**' puts objects in scene (A&V, 2&3D)
- Objects can be of different nature
  - **natural** or **synthetic** A&V, text & graphics, animated faces, arbitrary shape or rectangular
- Coding scheme can differ for individual objects
- Principle is **independent** of bitrate!
  - from low bitrates to (virtually) lossless quality

Example: MPEG-4 Audiovisual Scene



More about the goals of MPEG-4 (1)

- **Interactivity**
  - Content based
  - Random access (in time & to objects)
- **Integration** of natural and synthetic material
  - Separate activity SNHC within MPEG (Synthetic-Natural Hybrid Coding)
  - Mixing synthetic and natural objects together in the same scene
  - Virtual Environments

More about the goals of MPEG-4 (2)

- Accessing information **anywhere**:
  - Access on **mobile** networks (efficient coding and low bitrates still important)
  - Ability to cope with **error-prone** environments
  - Access across **different** networks
  - **Scalability** based on (audio/visual) objects
  - Different quality, priority, error protection for **different objects** possible
- Intellectual Property Rights (**IPR**)
  - identification (V.1) and protection (V.2)

The different parts of MPEG-4 Standard

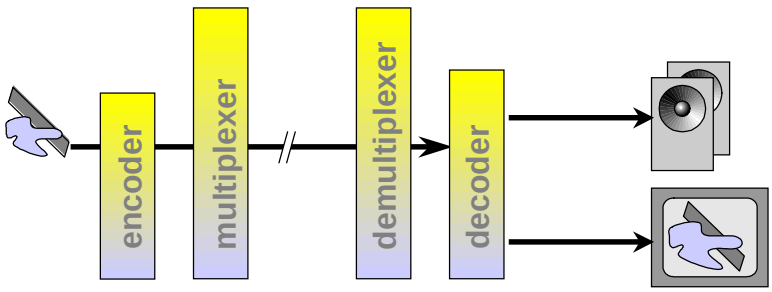
- **Delivery** (DMIF: Delivery MM Integration Framework)
  - set-up of connection channels (broadcast & interactive)
  - network becomes transparent to application
- **Systems**
  - Scene Description: composition of different objects in the scene
    - **BIFS**: **B**inary **F**ormat for **S**cene description
    - Build on several concepts from VRML
  - Buffering, Multiplexing, Timing
  - Interaction
  - 'Intermedia Format'

Different parts of MPEG-4 Standard (2)

- Visual
  - coding of natural, and synthetic (mostly moving) ***images***
- Audio
  - coding of natural and synthetic ***sounds***

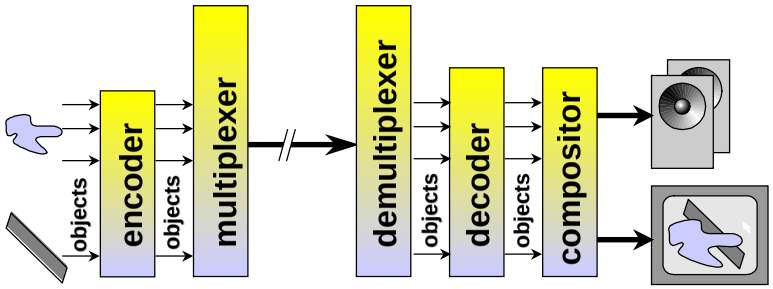
- SNHC (Synthetic-Natural Hybrid Coding)
  - develops coding for ***synthetic*** data types
  - not separate part of standard, but technology is integrated in Audio and Visual parts

Overview MPEG-4 System

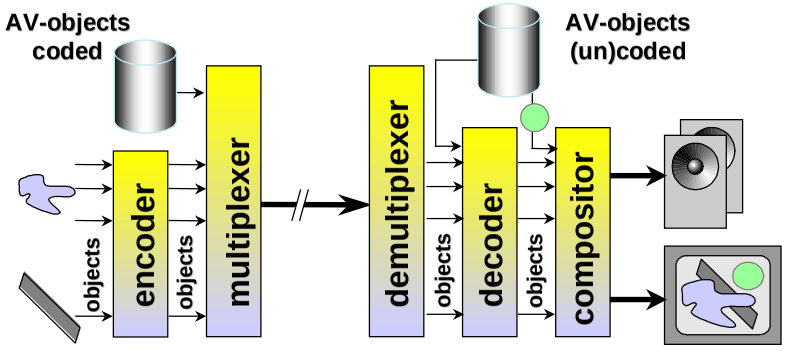


(conventional system)

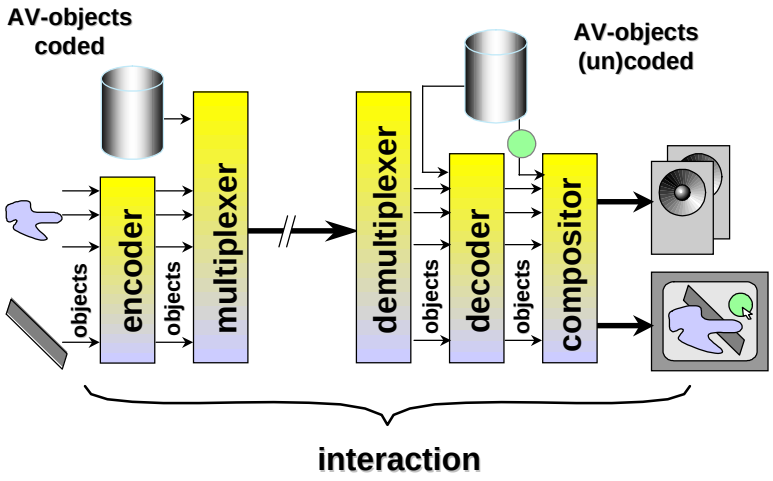
Overview MPEG-4 System



Overview MPEG-4 System

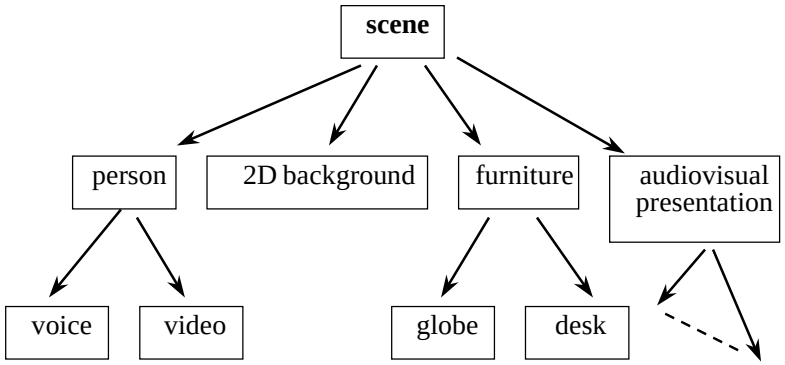


Overview MPEG-4 System



@NTUEE DSP/IC Lab

MPEG-4 Systems: scene composition



@NTUEE DSP/IC Lab

Intellectual Property Rights Management

- MPEG-4 Version 1: **Identification**
  - using existng registration Systems (e.g. ISBN)
- MPEG-4 Version 2: **Protection**
  - The persistence of content identification in modified MPEG-4 objects
  - Content Protection
- MPEG4 does not standardize IPMP (Intellectual Property Management and Protection).
- MPEG4 does standardize **IPMP interface**, which consists of:
  - IPMP-Descriptors (IPMP-Ds)
  - IPMP-Elementary Stream (IPMP-ES)

Profiles

- Useful subsets of the toolbox
- Allow interworking & conformance tests
- In all areas: Systems, Visual, Audio, Delivery (DMIF)
  - which objects can be combined in a scene?
- MPEG does not prescribe combinations of A, V, S, D

Five Profiles for Natural Video

- **Simple Visual Profile:**
  - Provide efficient and error resilient coding of rectangular video objects
  - Suitable for applications on mobile networks
- **Simple Scalable Visual Profile:**
  - Add support for coding of temporal and spatial scalable objects
- **Core Visual Profile:**
  - Add support for coding of arbitrary-shaped and temporally scalable objects to the Simple Visual Profile
- **Main Visual Profile:**
  - Add support for coding of interlaced, semi-transparent, and sprite objects to the Core Visual Profile
- **N-Bit Visual Profile:**
  - Add support for coding video objects having pixel-depths ranging from 4 to 12 bits to the Core Visual Profile

@NTUEE
DSP/IC Lab

17

MPEG-4 Schedule

- November '97: Committee Draft (CD)
- March '98: Final CD (FCD)
- October '98: Draft International Standard
- February '99: International Standard

- Version 2 will follow Version 1 with all phases one year later
- Version 2 will add new 'Profiles', with new functionality (V2 = V1+ new Profiles)

@NTUEE
DSP/IC Lab

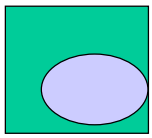
18

MPEG-4 Video Standard

- **MPEG-4 Video Provides Tools for a Number of Functionalities**
- **Integrated Approach (Core and Extensions)**
- **Based on DCT Technology (except for Still Texture Coding)**

Functionality (Core)

- **Coding Efficiency**
  - 5 kbits/s - 5 Mb/s
  - Resolution: Small - TV
  - Progress/Interlace
- **Error Resilience/Robustness**
  - Mobile Environments
- **Scalability** (Spatial/Temporal)



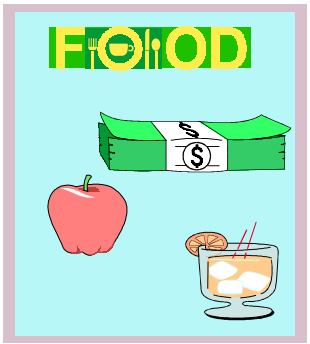
Frame-based

Content-Based Coding of Video

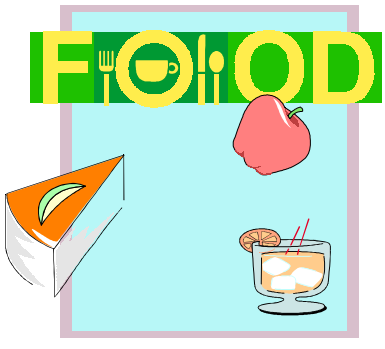
- **Content-Based Coding Allows the User to Access Arbitrarily-Shaped Objects in a Coded Scene**
- **Content-Based Coding Enables High Interaction With Scene Content**
- **Manipulation of Scene Content on Bitstream Level**

Object Manipulation

- **Original Decoded**

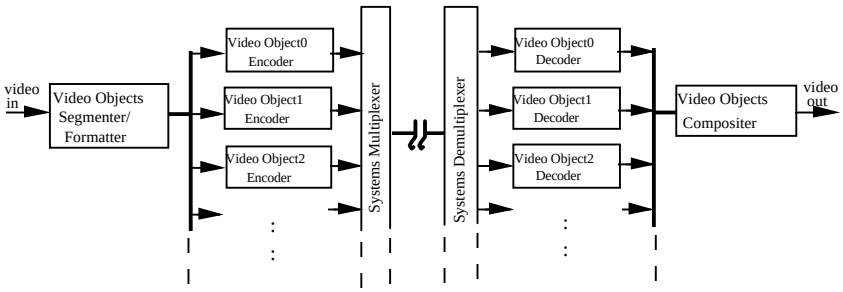


- **Decoded and Manipulated**



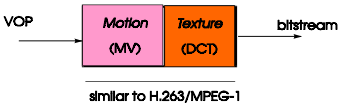
Content-Based Layering of Video

- Each Video Object in a Scene is Coded and Transmitted Separately

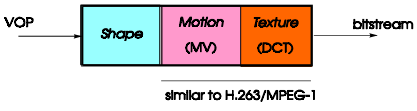


MPEG-4 Core and Extension

MPEG-4 Core Coder

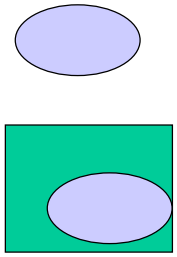


Extended MPEG-4 Core Coder

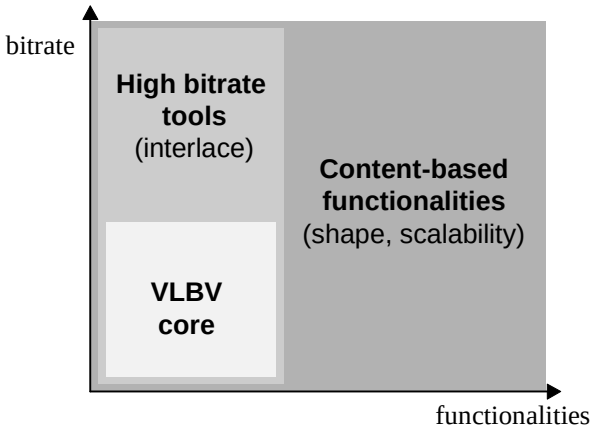


MPEG-4
Content-Based Functionalities

- **Shape Coding**
- **Sprites**
- **Scalability (Content Based)**
- **Error Resilience/Robustness**
- **Scalable Texture Coding (Wavelets)**



Classification of MPEG4 Video Tools



MPEG-4 Tools - Summary -

- **I-VOP**
- **P and B Prediction**
- **Interlace Prediction**

- **Scalability**
  - *Temporal (rectang.)*
  - *Spatial (rectang.)*
  - *Temporal (object)*

- **Binary Shape Coding**

- **Error Resilience**
  - *rectang./object*

MPEG-4 Tools (Cont.)

- **12 bit Video**

- **Texture Coding**
  (Wavelets)
  - *rectang./object*
  - *Coding Efficiency*
  - *SNR Scalability*
  - *Spatial Scalability*

- **Static Sprites**
  - *Basic Sprites*
  - *Low Latency Sprites*

- **Computational**
  Graceful Degradation

Video Object Coding Outline

- Data structure used in visual part of MPEG-4
- Block diagram of natural video decoding
- List of major natural video tools
- Shape coding tools
- Motion compensation tools
- Texture coding tools
- Scalable coding tools
- Error resilience tools
- Sprite coding tools
- 12-bit and Interlaced coding tools

Data structure in visual part of MPEG-4

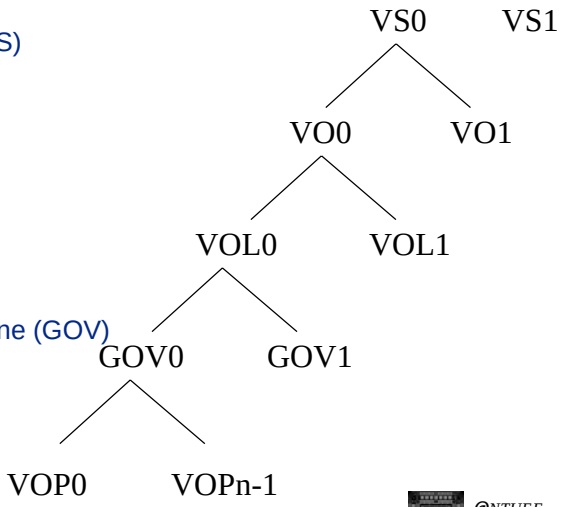
Visual object Sequence (VS)

Visual Object (VO)

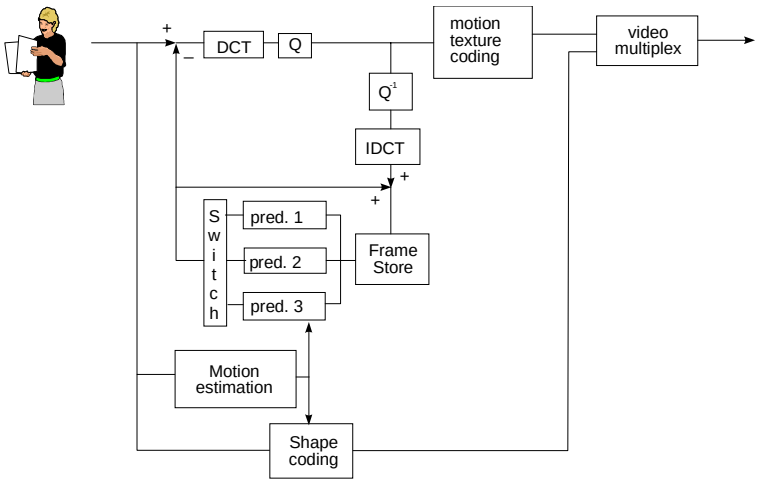
Visual Object Layer (VOL)

Group of Visual Object Plane (GOV)

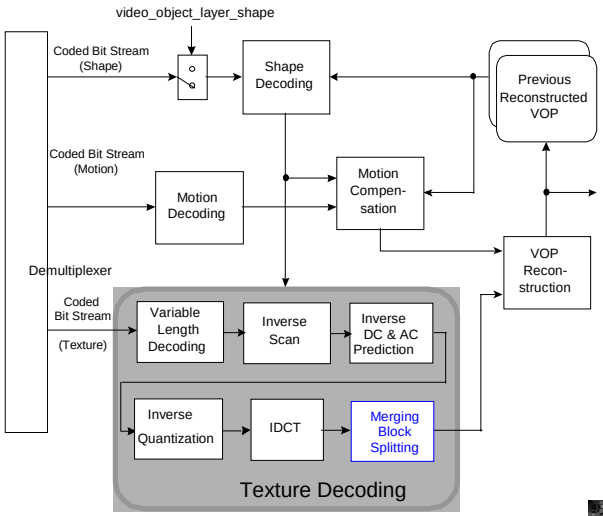
Visual Object Plane (VOP)



Basic Block Diagram of MPEG4 Video Coder



Simplified block diagram of natural video decoding

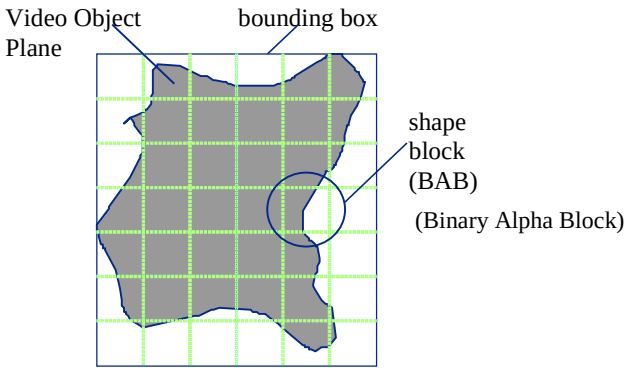


List of major natural video tools

- **Binary shape**
- **Padding**
- **Motion compensation**
  - Overlapped motion compensation
  - Advanced motion compensation
- **Quantization**
  - Method 1
  - Method 2
- **AC/DC prediction**
  - Non-linear
- **Scanning**
  - Type 1
  - Type 2
- **I, P, B modes**
- **Temporal scalability**
  - Slice synchronization
  - Extended header code
- **Spatial Scalability**
  - Data partitioning
  - Reversible VLC
- **Error resilience**
- **Static sprites**
- **Interlaced coding**
- **12-bit video**
- **Static texture**

Shape coding tool (1/3)

Every VOP is coded by dividing it into smaller macroblocks

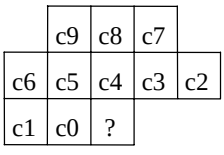


Shape coding tool (2/3)

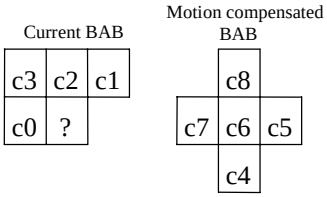
- Coding modes
  - Opaque
  - Transparent
  - No-update
  - Intra Context based Arithmetic Encoding
  - Inter Context based Arithmetic Encoding
- Lossless
- Lossy
  - Motion compensation without update
  - sub-sampling by factor 2 or 4

Shape coding tools - CAE

- Context based Arithmetic Encoding
  - Intra

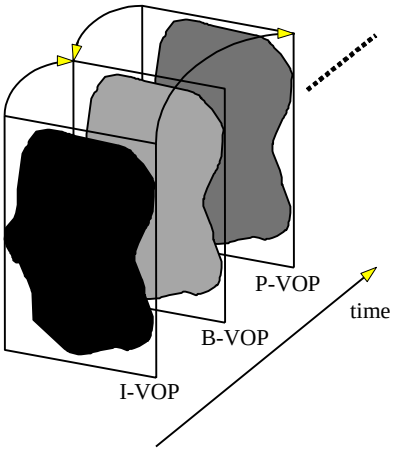


  - Inter

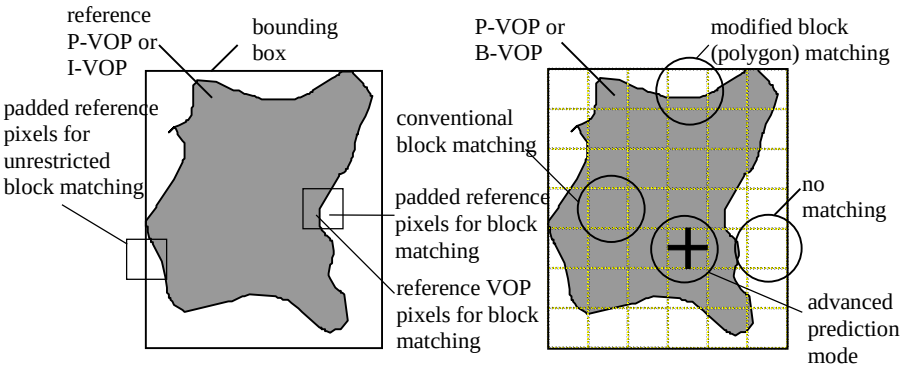


Motion compensation tools (1/3)

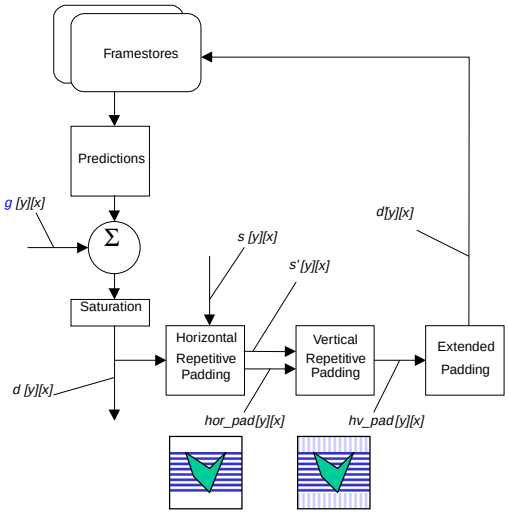
Motion compensated coding modes (I, B, P)

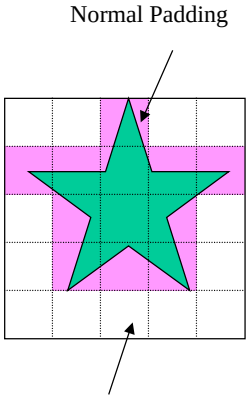


Motion compensation tools - Motion vector computation (2/3)

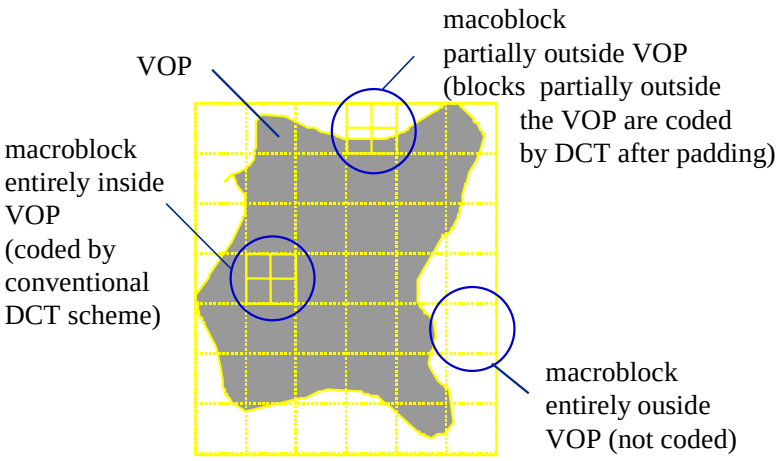


Motion compensation tools - padding (3/3)

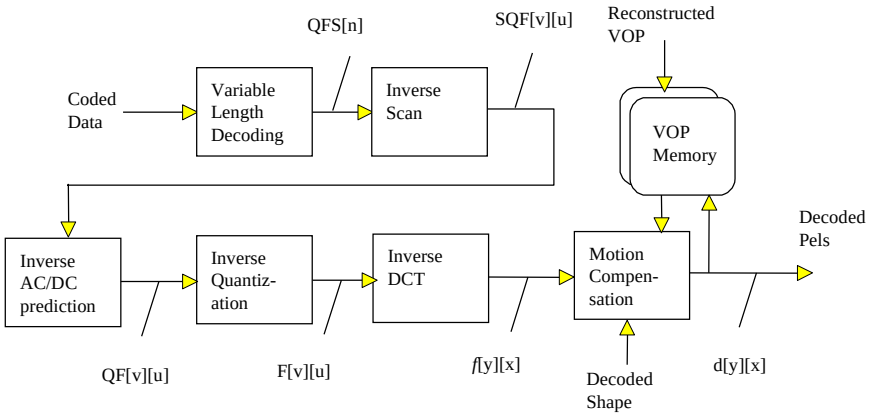




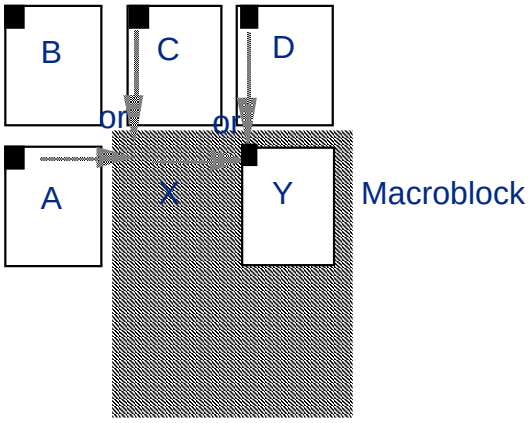
Texture coding tools (1/2)



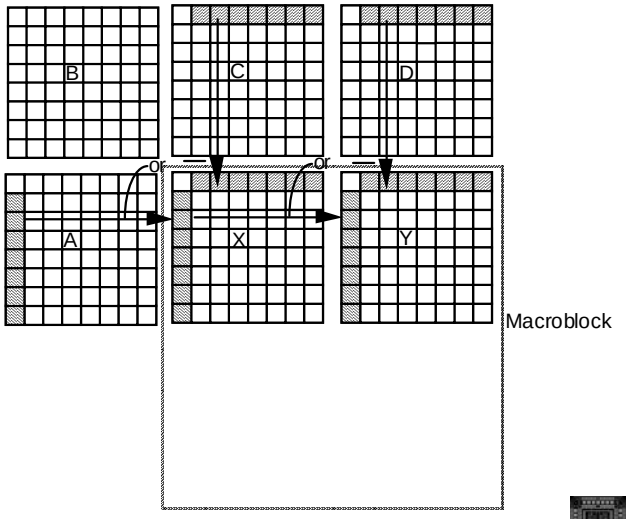
Texture coding tools (2/2)



Adaptive DC prediction



Adaptive AC prediction



Coefficients scanning

0	1	2	3	10	11	12	13
4	5	8	9	17	16	15	14
6	7	19	18	26	27	28	29
20	21	24	25	30	31	32	33
22	23	34	35	42	43	44	45
36	37	40	41	46	47	48	49
38	39	50	51	56	57	58	59
52	53	54	55	60	61	62	63

Alternate-Horizontal scan

0	4	6	20	22	36	38	52
1	5	7	21	23	37	39	53
2	8	19	24	34	40	50	54
3	9	18	25	35	41	51	55
10	17	26	30	42	46	56	60
11	16	27	31	43	47	57	61
12	15	28	32	44	48	58	62
13	14	29	33	45	49	59	63

Alternate-Vertical scan

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

zig-zag scan

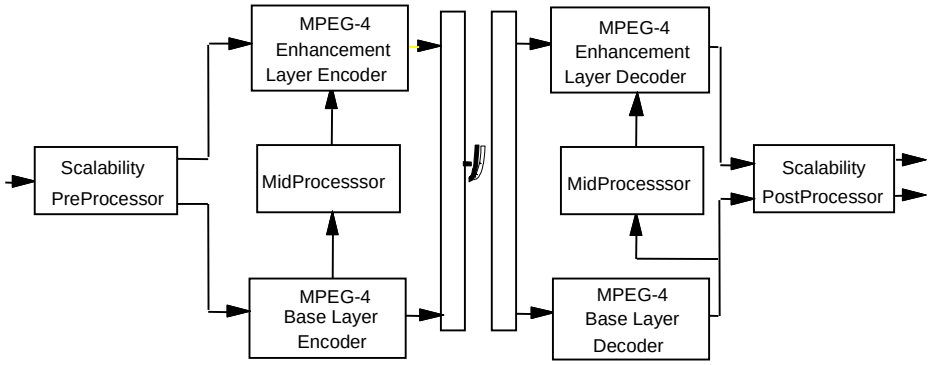
Quantization

- Method 1: Similar to that of H.263
- Method 2: Similar to that of MPEG-2
- Optimized non-linear quantization of DC coefficients
- Quantization matrices and loading mechanism

Scalability

- Object scalability
  - Achieved by the data structure used and the shape coding
- Temporal scalability
  - Achieved by generalized scalability mechanism
- Spatial scalability
  - Achieved by generalized scalable mechanism

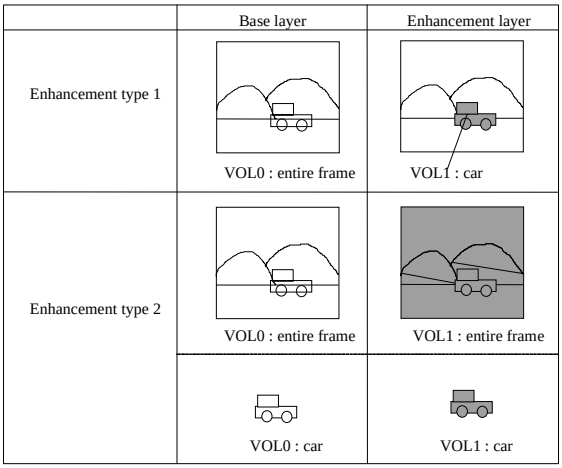
Scalable coding general scheme



Temporal scalability

- The temporal scalability is achievable for both rectangular frames and arbitrarily shaped VOPs
- The base layer is encoded conventional MPEG-4 video
- The enhancement layer is encoded using one of the following two mechanisms:
  - Type 1
  - Type 2

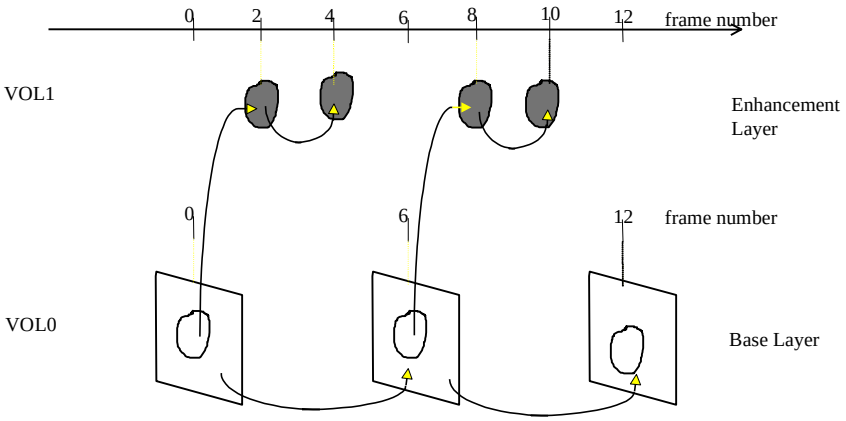
Temporal enhancement types



: region to be enhanced by an enhancement layer

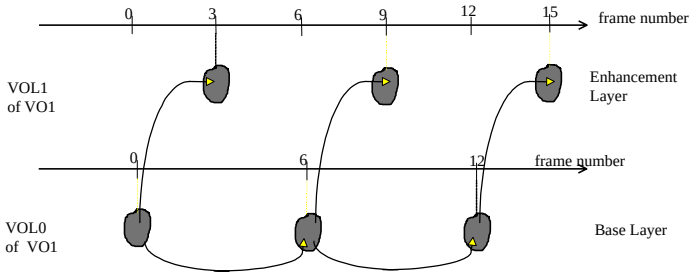
Temporal scalability Type 1

Only a portion of the VOP in the base layer is enhanced



Temporal Scalability Type 2

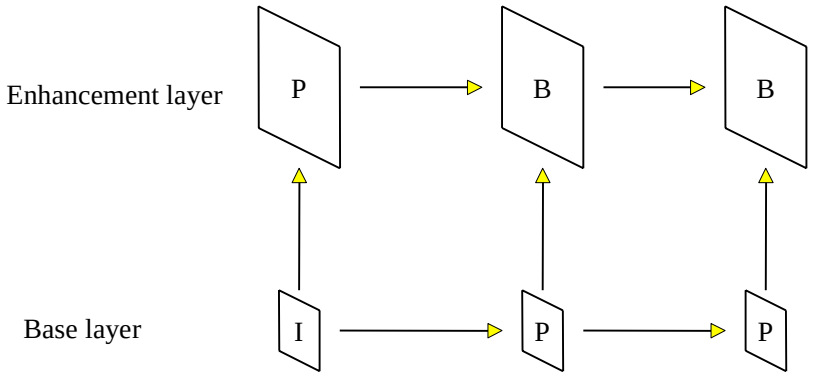
The entire VOP in the base layer is enhanced



Spatial scalability

- The base layer is coded as conventional MPEG-4 video
- The enhancement layer is encoded using prediction mechanims from the base layer

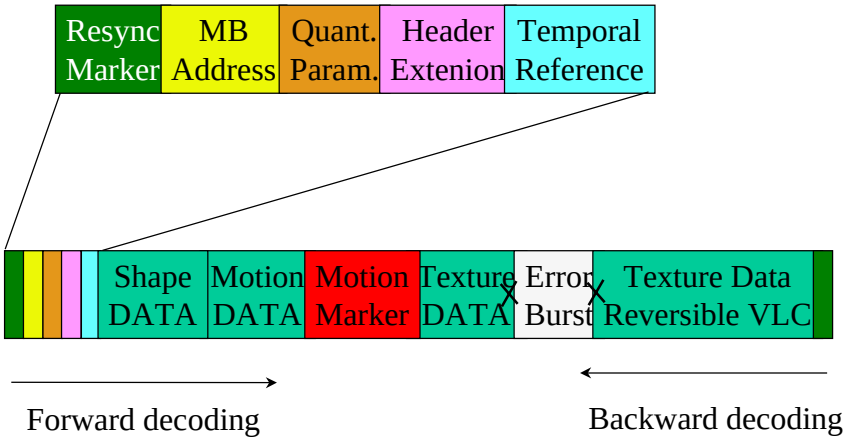
Spatial scalability



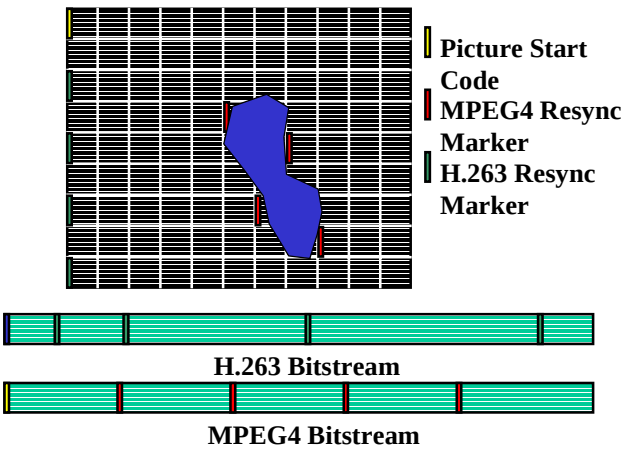
Error resilience tools (1/3)

- Resynchronization markers
  - Spatial Resynchronization: GOB Start Code
  - Periodic Resynchronization Markers
- Extended header code
  - A single bit, when enable, indicates additional resynchronization information for VOP header
- Data partitioning
  - Separate motion and texture information
  - Enable better error concealment
- Reversible VLCs

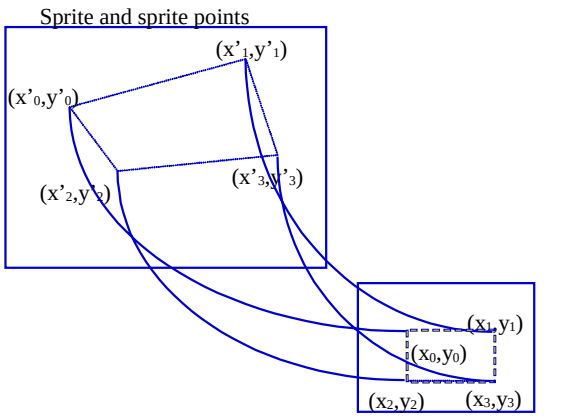
Error resilience tools (2/3)



Error resilience tools (3/3)

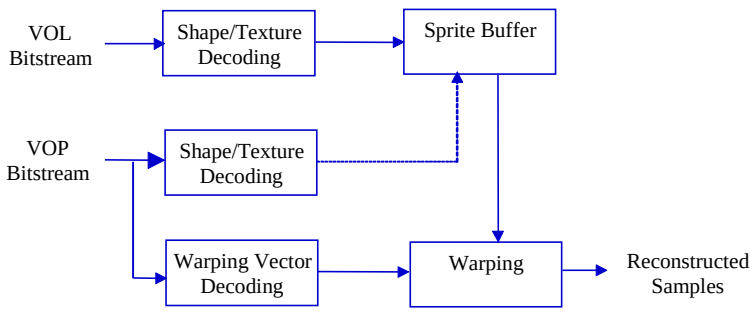


Static sprite coding tools (1/3)

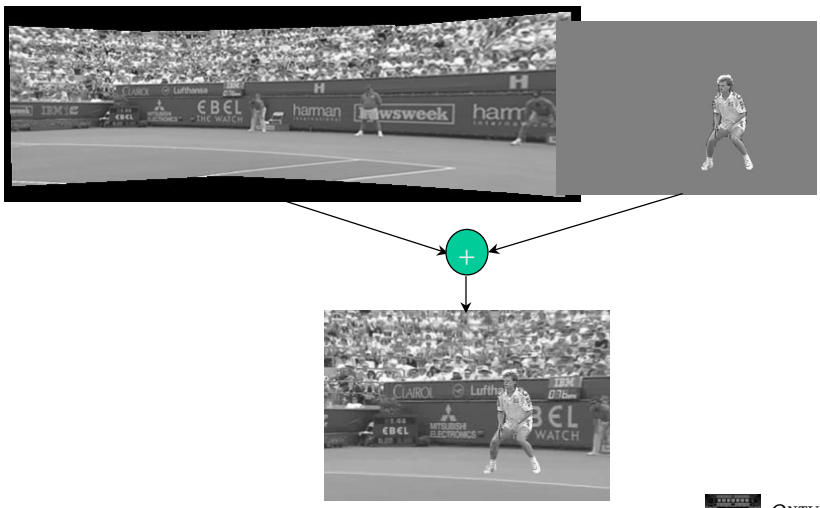


Static sprite coding tools (2/3)

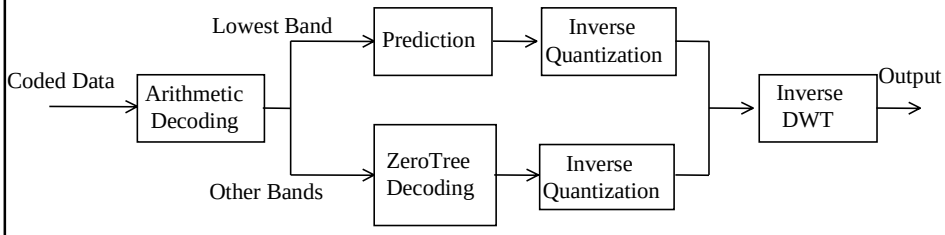
- Basic sprite coding
- Low latency sprite coding
- Scalable sprite coding



Static sprite coding tools (3/3)



Shape adaptive wavelet coding (1/4)



Shape adaptive wavelet coding (2/4)

- Generalization of the wavelet transform to arbitrarily shaped VOP
  - number of transformed coefficients in the VOP = number of pixels in the VOP
- Generalization of zero-tree coding
  - no extra bit necessary for pixels outside the VOP

Shape adaptive wavelet coding - SNR scalability(3/4)

Shape adaptive wavelet coding - spatial scalability(4/4)

12-bit video coding tool

- Allows compression of video data with precision of up to 12-bits / pixel
- The syntax, semantics, and coding tools are extended:
  - bit-precision
  - extended DC VLC tables
  - extended quantization mechanism
  - Insertion of marker bits to avoid start code emulations

Interlaced coding mode

- Allows all option in progressive also for interlaced.
- Motion compensation for field or frames similar to that of MPEG-2
- Modified AC/DC prediction
- Field DCT
- Interlaced I, P, and B VOP coding
- Modified prediction for motion coding
- Modified scan rules
- 10% more efficient in compression efficiency compared to MPEG-2

SNHC Visual: Areas of Work

- Facial Animation
  - Facial Definition Parameters (FDP)
  - Facial Animation Parameters (FAP)
  - Face Interpolation Technique (FIT)
- 2D Animated Meshes
  - Triangular meshes
- Scalable Textures
  - Wavelet-based, ZTE (variation of EZW) + DC prediction
- View Dependent Scalability
  - Use a back-channel
- Body Animation (in MPEG4 Version 2)

Visual Tools in Version 1

- Intra Coding Mode (I-VOP)
- Inter Prediction Mode (P-VOP)
- AC/DC Prediction
- Slice Resynchronization
- Data Partitioning
- Reversible VLC
- 4MV, Unrestricted MV
- Binary Shape Coding
- H.263/MPEG-2 Quantization Tables
- P-VOP based temporal scalability Rectangular Shape
- P-VOP based temporal scalability Arbitrary Shape
- Bi-directional Prediction Mode (B-VOP)
- OBMC
- Interlaced tools
- Grayscale Alpha Shape Coding
- Static Sprites (includes low latency mode)
- Spatial Scalability Rectangular Shape
- Temporal Scalability Rectangular Shape
- Temporal Scalability Arbitrary Shape
- 4- to 12-bit pixel depth
- Scalable Wavelet Texture
- 2D Dynamic Mesh with uniform topology
- 2D Dynamic Mesh with Delaunay topology
- Facial Animation Parameters

Visual Tools in Version 2

- Quarter Pel Prediction
- Global Motion Compensation
- Boundary Block Merging
- Shape Adaptive DCT
- Newpred
- Object based Spatial Scalability
- Multiple Auxiliary Components
- Wavelet Tiling
- Scalable Shape Coding for Still Texture
- Dynamic Resolution Conversion

Conclusions

- MPEG4 is a very rich standard
- It provides many tools
- Most saliently, content-based (or object-based) functionalities
- Challenges: How to use it?
  - Killer applications?
  - How to implement it? Too complex?