

## 12. MPEG-4

Prof. Chia-Wen Lin  
Department of CS  
National Chung Cheng University  
886-5-272-0411 ext. 33120  
cwlin@cs.ccu.edu.tw

### MPEG-1, MPEG-2, and MPEG-4

- MPEG-1 (1992):
  - 1.5 Mb/s
  - CD-ROM
- MPEG-2 (1994):
  - 4 Mb/s to 80 Mb/s
  - DVD, Digital TV, HDTV
- MPEG-4 (1999):
  - 5 kb/s - 50 Mb/s
  - Flexible networked multimedia applications

## MPEG-4 Applications

### Multimedia

- Audiovisual communications and messaging
- Multimedia database access
- Remote monitoring, surveillance and control
- Video on LAN, Internet multimedia, Wireless video
- Interactive TV, tele-shopping, home movies
- Collaborative environment, distance learning, virtual reality games
- Audio/video streaming

## MPEG-4

- Initial driving application:
  1. Next generation video coding for very low bit-rate video
  2. Object based manipulation
    - New concepts: Object-based coding
- Current driving application:

**MPEG-4 attempts to become *THE* standard for streaming AV media on the Internet and via wireless networks**
- Error-resilience features, Optimized for low-bit-rate applications, Fine granular scalability coding
- MPEG-4 File Format defines first standard for streaming AV media on the Internet

## Current MPEG-4 Products

- Microsoft supports MPEG-4 video in MS Media Player
- Sharp (JP) introduced MPEG-4 ViewCam in January '99
- Toshiba MPEG-4 chip
- Japan announced the use of MPEG-4 video for wireless services in the IMT2000 project
- PacketVideo Inc. provides technology for streaming MPEG-4 video over Internet and 2nd/3rd generation wireless networks

## MPEG-4 ViewCam

Add compressed MPEG-4 video images to e-mail or Internet home pages for new communication possibilities (US\$ 350)



Internet ViewCam

- Supported by Microsoft Media Player
- MPEG-4 ViewCam enables up to 2 hrs video (1/8 TV size) on 32 MB storage device.

## MPEG-4

- Started in '93
- Originally targeted at < 64 kb/s next generation coding
- July '94, focus shifted from *compression* to *new functionality + compression*
- Address needs for emerging audiovisual applications
  - which are not well supported by current standards
- Video coding for wider range of bit-rates and applications
- Significant emphasis on content oriented functionality
- MPEG-4 decoder must decode H.263

## Working Group Structure

- Requirements
- DSM
- Delivery
- Systems
- Video
- Audio
- Synthetic Natural Hybrid Coding (SNHC)
- Test
- Implementation

## MPEG-4 Parts

- Systems
- Visual
- Audio
- Conformance Testing
- Reference Software
- Delivery Multimedia Integration Framework (DMIF)

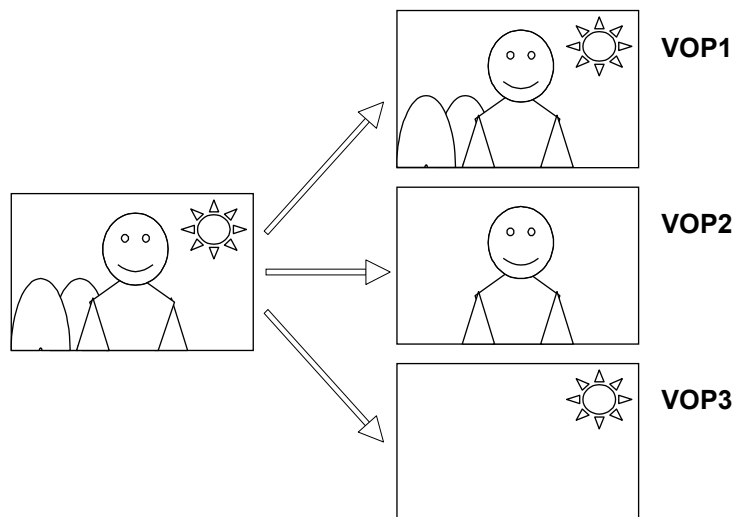
## MPEG-4 Timetable

- 11/94 Call for proposals
  - 11/95 Subjective test
  - 11/96 Working Draft
  - 07/97 Formal Test
  - 11/97 Committee Draft
  - 03/98 Final Committee Draft
  - 10/98 Draft International Standard (DIS)
  - 05/99 International Standard (IS)
- Version 2: adds new profiles and tools  
Version 3: Studio Profile  
Version 4: Streaming Video Profile

## MPEG-4 Functionalities

- Content-Based Interactivity
  - Audio/Visual Object (AVO) manipulation, bitstream editing
  - Graphics (SNHC)
- Compression
- Universal Access
  - Error resilience
  - Content-based scalability

## MPEG-4 Video Object Plane (VOP)



## MPEG-4 VOP Example



## MPEG-4 VOP Example



## Integration of Natural and Synthetic Content



## MPEG-4 Video Standard

### Baseline

**Compression**

**Error Resilience**

**Scalability**

*Conventional coding*

### Extended

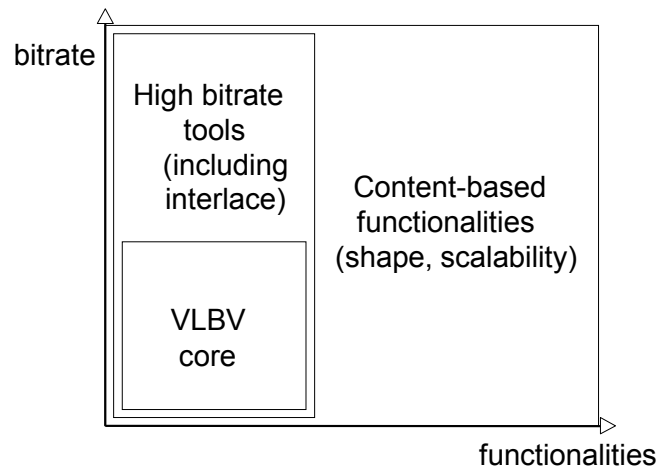
**Content-based  
Coding**

**Still Texture  
Coding**

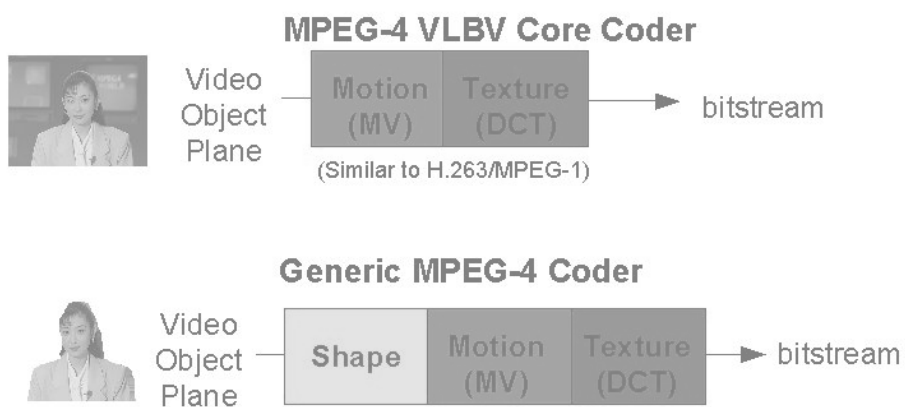
*Object coding*



## Structure of MPEG-4 Video Standard



## MPEG-4 Video: Core & Generic Coder



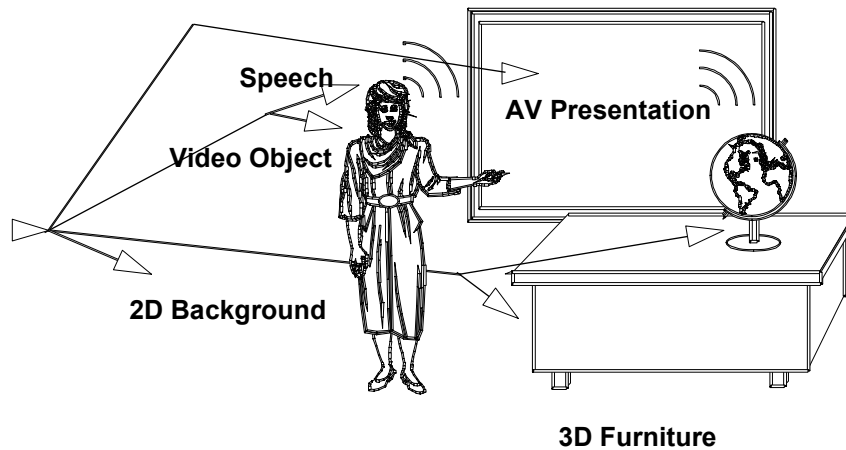
## Content-based Coding

- Allows the user to access Arbitrarily-Shaped Objects in a Coded Scene
- Enables High Interaction with Scene Content
- Manipulation of Scene Content on Bitstream Level

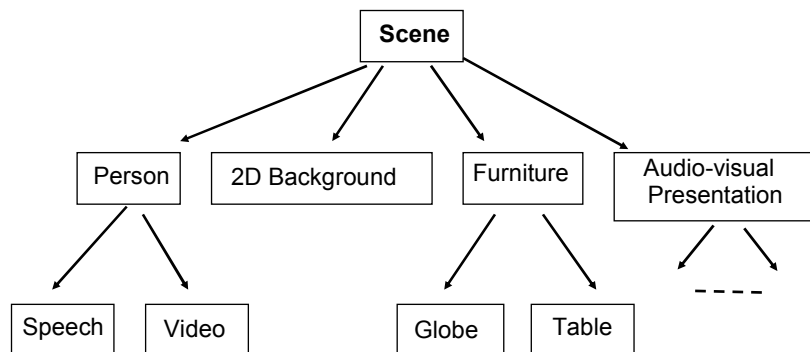
## Coding of AV Objects

- AV scenes consist of “objects”
- Objects can be natural or synthetic (A&V, Text & Graphics, animated faces, arbitrarily shaped or rectangular)
- A “compositor” composes objects in a scene (A&V, 2 & 3D)
- Binary Format for Scenes: BIFS

## An MPEG-4 Scene

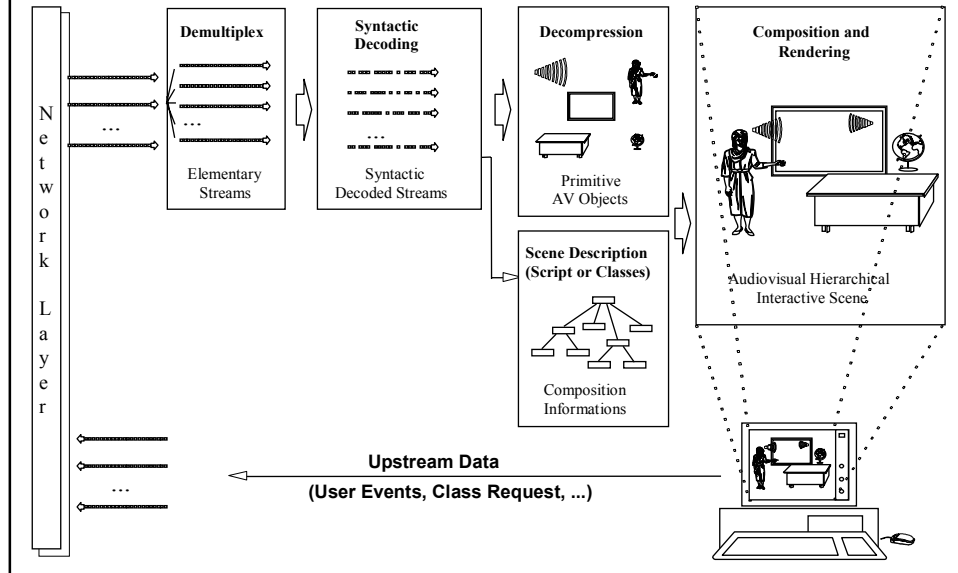


## BIFS Composition of Scenes

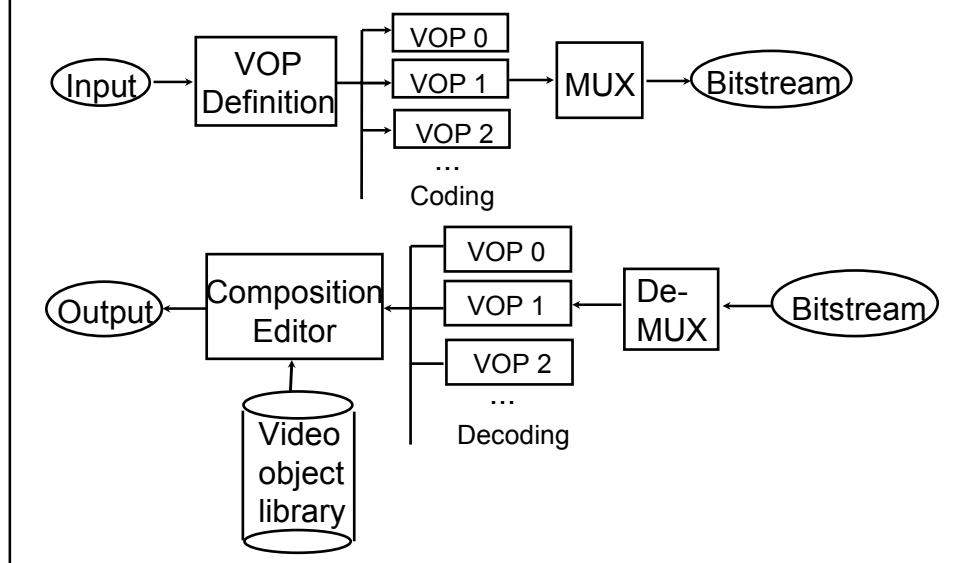


**BIFS (Binary Format for Scenes):** allows placement of objects in a coordinate system, transformation of objects, grouping of objects, use of streamed data to modify objects, change viewing points interactively.

## An MPEG-4 Terminal



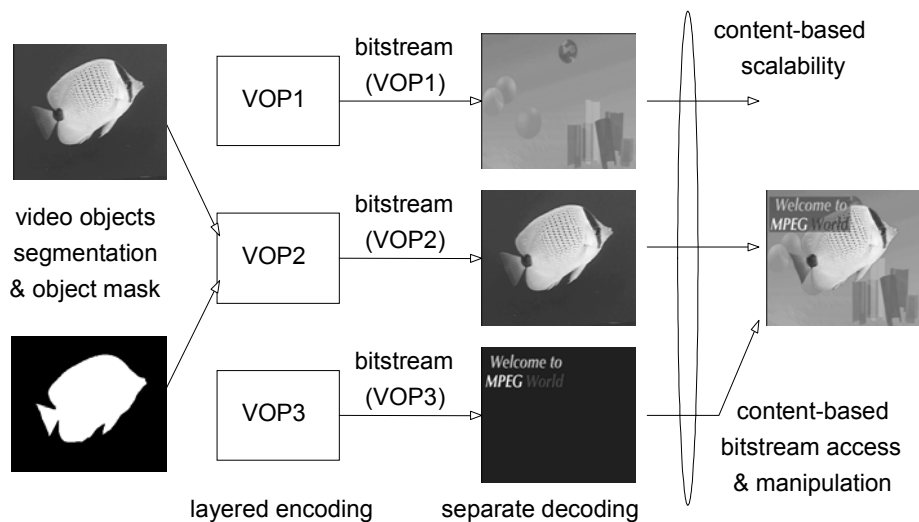
## CODEC Structure



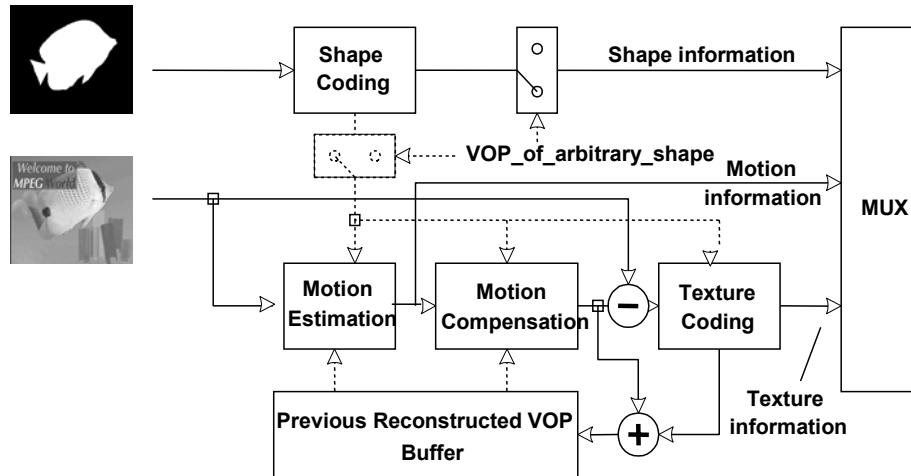
## Visual Object Type

- Video Object
  - Rectangular
  - Arbitrary shaped
  - Sprite
- Face and Body Animation (FBA) Object
  - FDP (Face Definition Parameter)
  - FAP (Face Animation Parameter)
- Mesh Object
  - 2-D Mesh
  - 3-D Mesh
- Still Texture Object

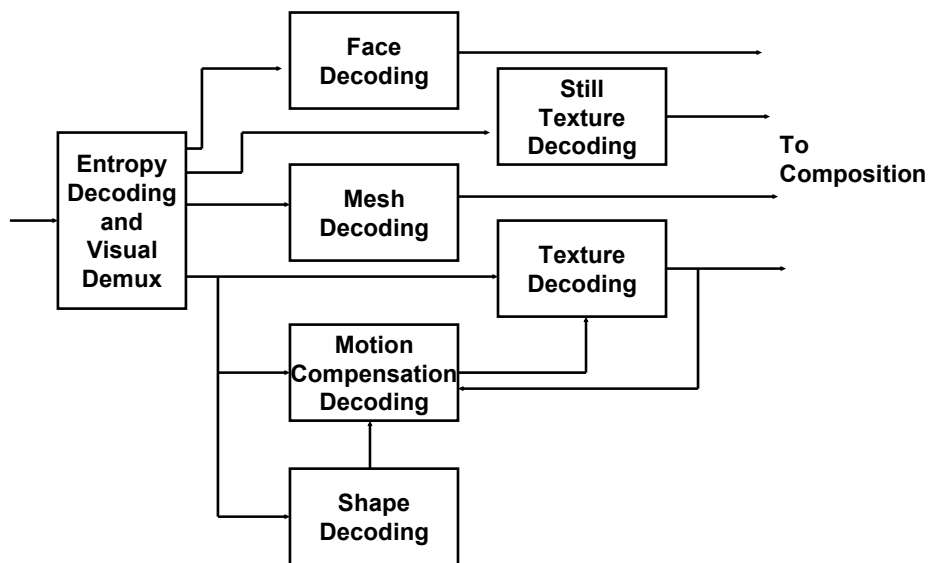
## VOP-Based Encoding



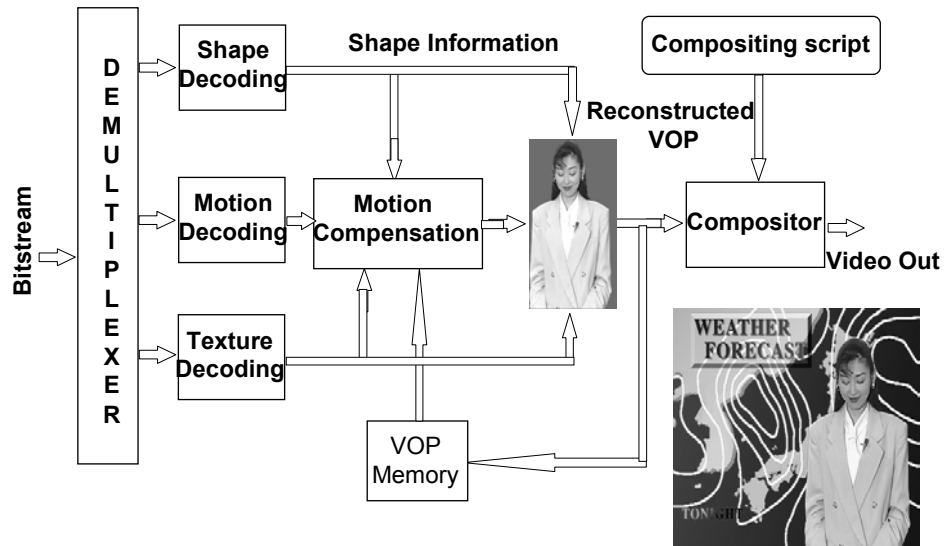
## MPEG-4 Video Encoder



## High Level View of Visual Decoding

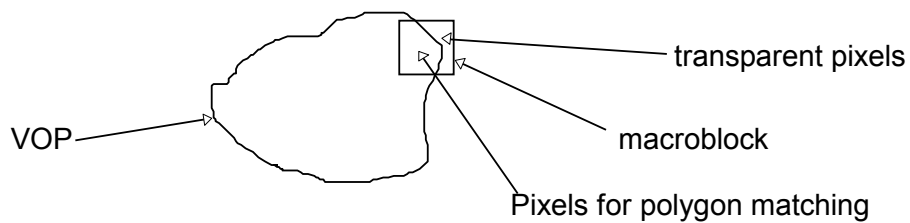


## VOP Decoder

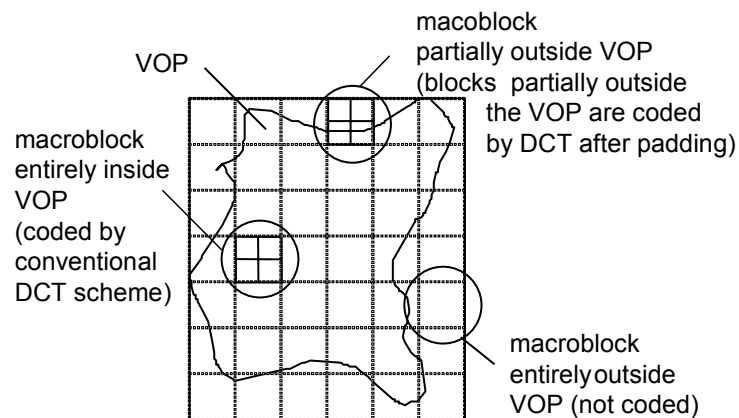


## Texture Coding (1/3)

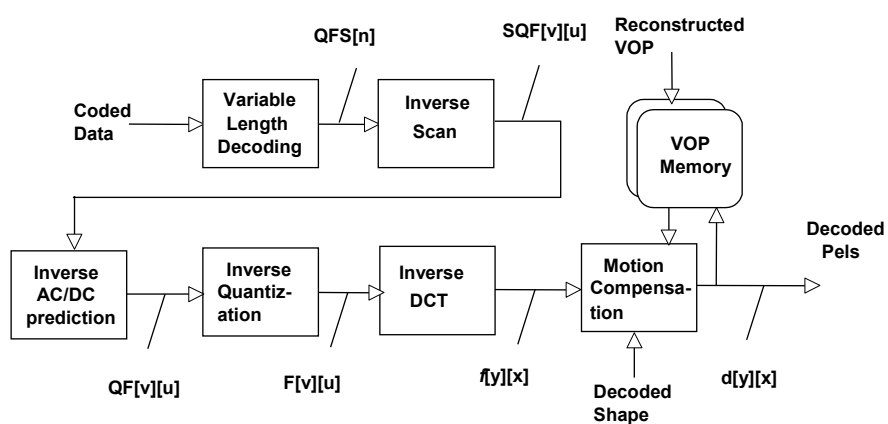
- Motion compensated DCT
  - Very similar to H.263
- Polygon matching



## Texture Coding (2/3)

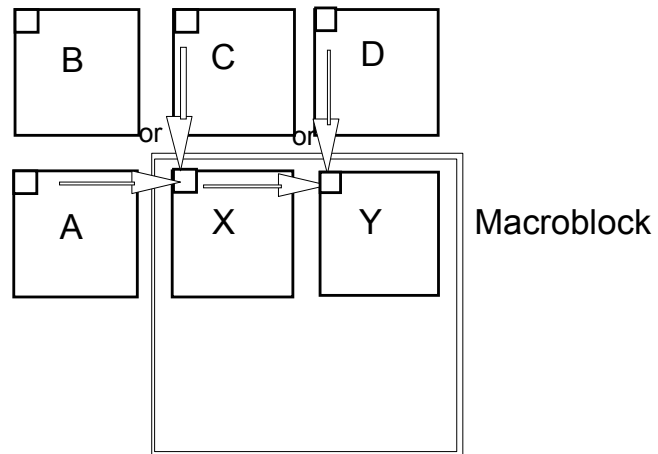


## Texture Coding (3/3)

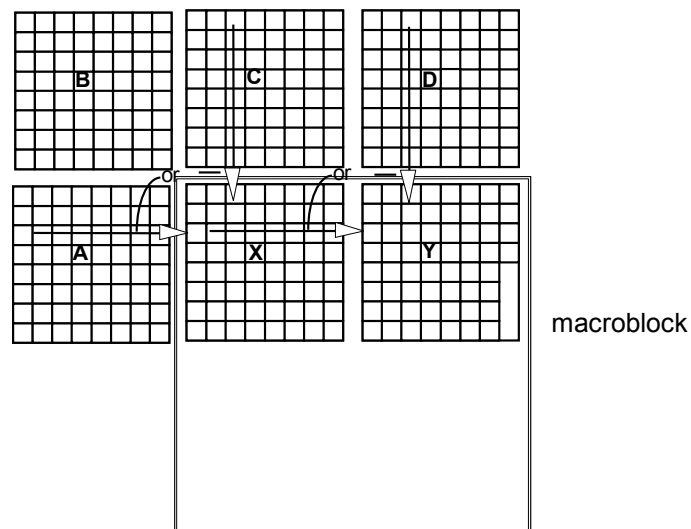




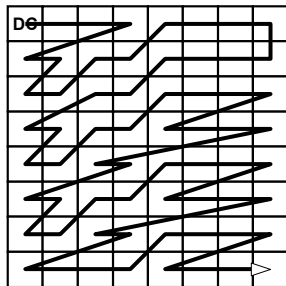
## Adaptive DC Prediction



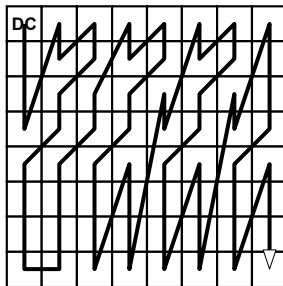
## Adaptive AC Prediction



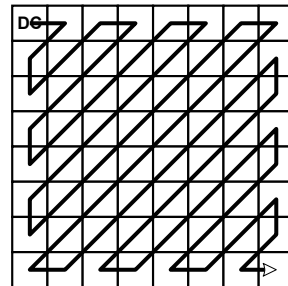
## Scanning Pattern



Alternate-horizontal scan



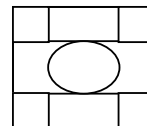
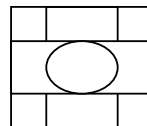
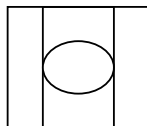
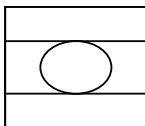
Alternate-Vertical scan



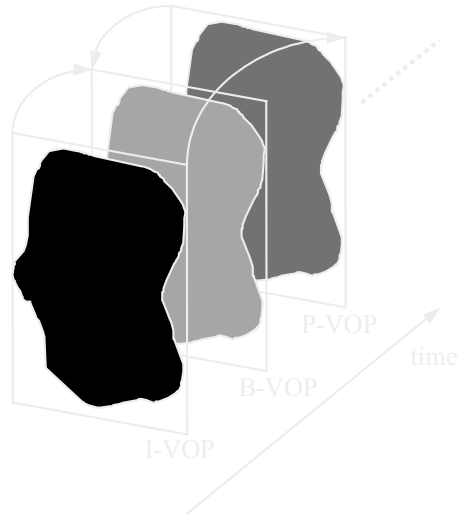
Zig-Zag scan

## VOP Padding For Block Matching

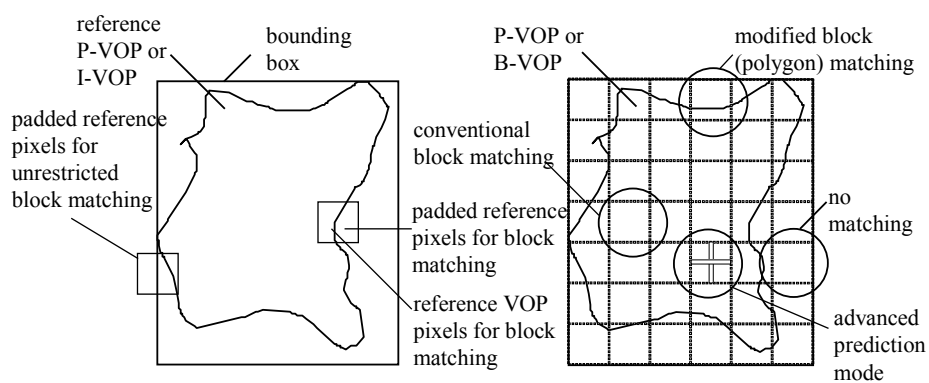
- Repetitive padding is applied to the reference VOP prior to motion estimation/compensation



## Motion Estimation/Compensation (1/3)



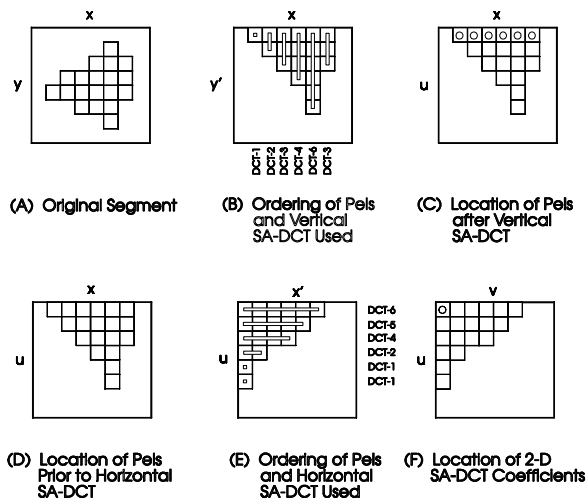
## Motion Estimation/Compensation (2/3)



## Motion Estimation/Compensation (1/3)

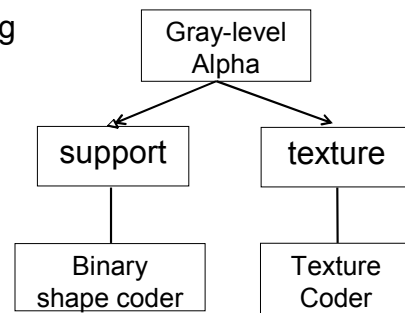


## Shape Adaptive DCT for Texture Coding



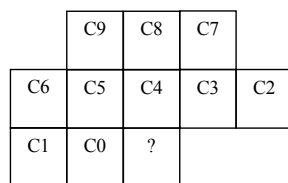
## Shape Coding

- Binary shape
  - Context-based arithmetic encoding (CAE)
- Gray scale alpha plane
  - Motion compensated DCT
    - Similar to texture coding

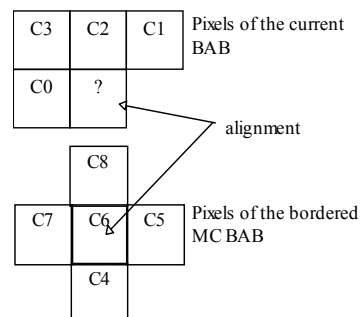


## Block-based CAE

- The context



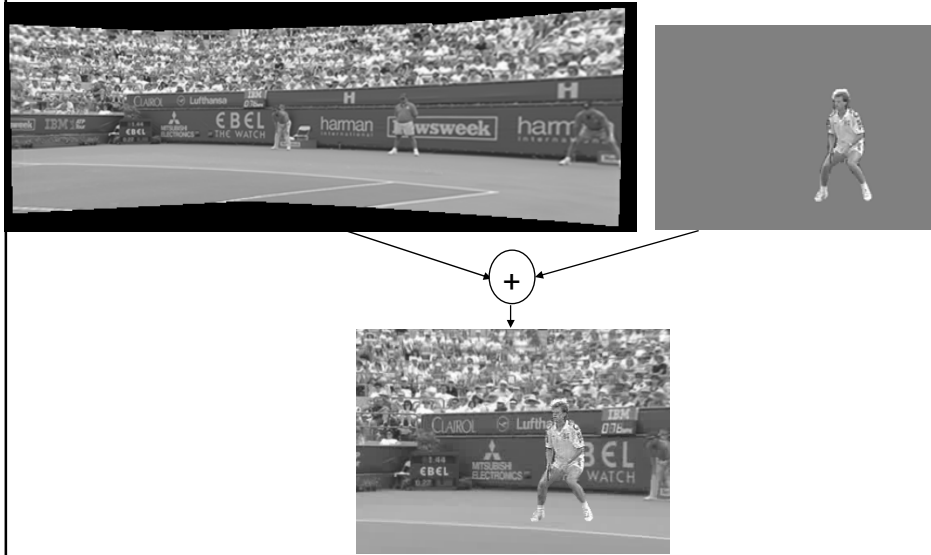
Intra



Inter

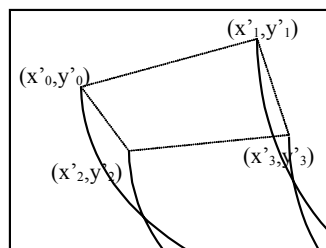
BAB: Binary alpha block

## Sprite Coding



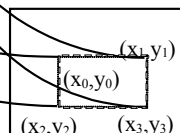
## Warping of Reference Sprite

**Sprite and sprite points**



**Sprite Image**

**Only 2-8 global motion parameters are transmitted per frame**

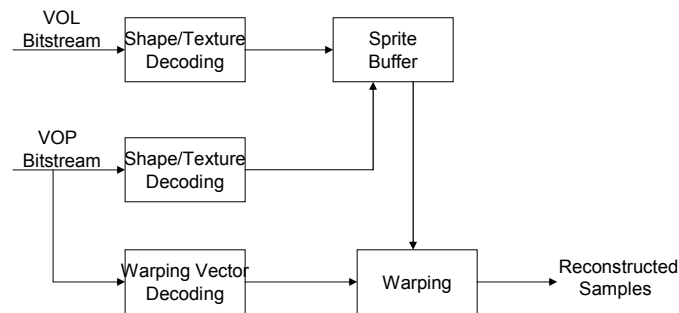


**Actual Frame**

**VOP and reference points**

## Static Sprite Coding Tools

Based sprite coding  
Low latency sprite coding  
Scalable sprite coding



## MPEG-4 Video verification model

- Verification model (VM8)
  - Video object plane (VOP) structure
  - Polygonal matching for motion estimation
  - Padding
  - Motion/Texture coding derived from H.263
  - Binary and gray-scale shape coding
  - B-VOPs (Video Object Plane) derived from H.263 B-pictures and MPEG-1/2 B-pictures

## SNHC

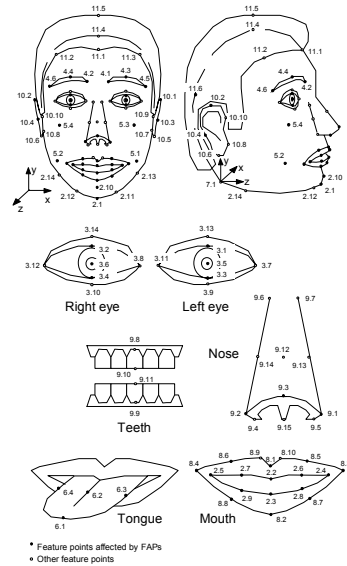
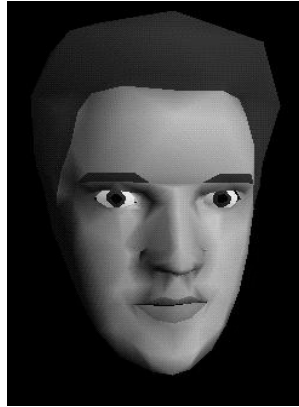
- Synthetic & Natural Hybrid Coding (SNHC)
  - To provide efficient representation and composition of synthetically and naturally generated audiovisual data
  - Target technologies  
Compression of geometry, integration of mixed media, synthetic and spatial audio, facial and body animation
  - Applications  
Virtual environment, conferencing, education/entertainment, media production, and real-time interactive and broadcast media experiences

## MPEG-4 Visual Face Animation

- Application:
  - very low rate video (as low as 100 b/s)
  - user interface to intelligent agents
- Model-based coding of faces
- Models are not normative
- Type of data:
  - Facial Definition Parameter
  - Facial Animation Parameter
  - Rendering



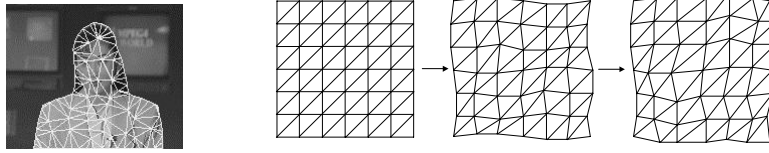
## MPEG-4 Visual Face Animation



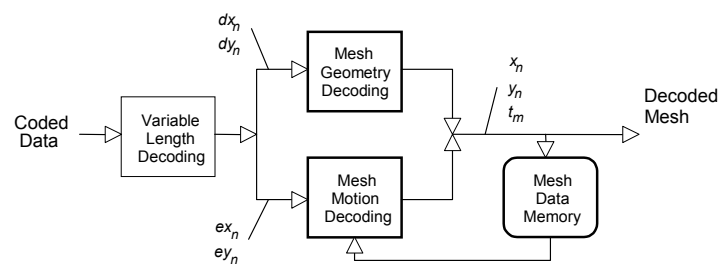
## MPEG-4 Visual Mesh Coding

- Used to code and animate 2D/3D objects
- Mesh is a set of connected polygons
- Image and Video can be rendered onto mesh
- Error robustness
- Mesh + still texture -> synthetic video
- Mesh is modified by motion vectors transmitted for each node

## 2-D Mesh Coding



- Decoding process



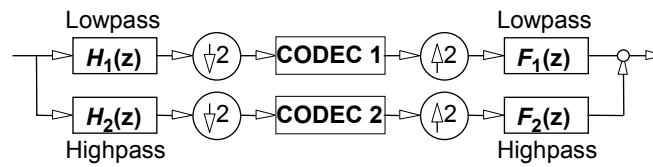
## 3-D Mesh Coding

- Progressive representation
  - Streaming of 3D objects
  - Both spatially and temporally



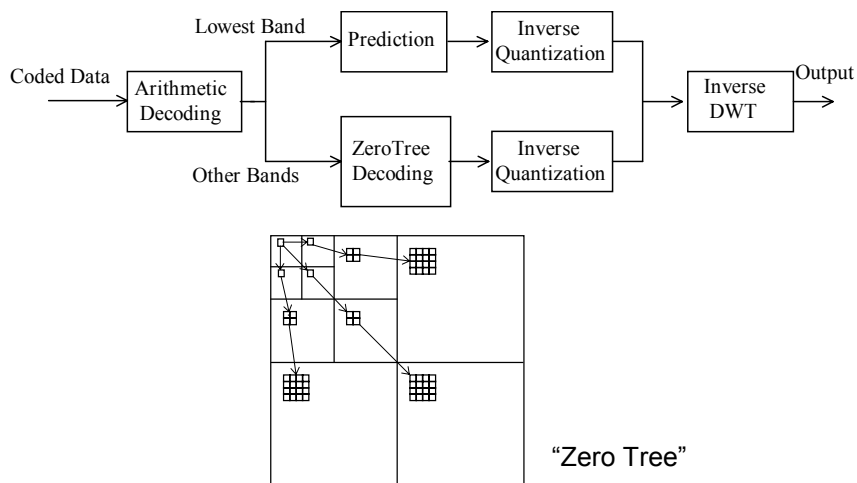
- Indexing and retrieval of 3D meshes
  - Multiresolution databases
  - Related to MPEG-7

## Wavelet for Scalable Texture Coding



- Decompose the signal in the frequency domain
- Critical downsampling maintains the number of samples in the subbands
- For 2D case, separable filters are often used. Decompose into four bands: LL, LH, HL, HH
- Decompose the LL band iteratively

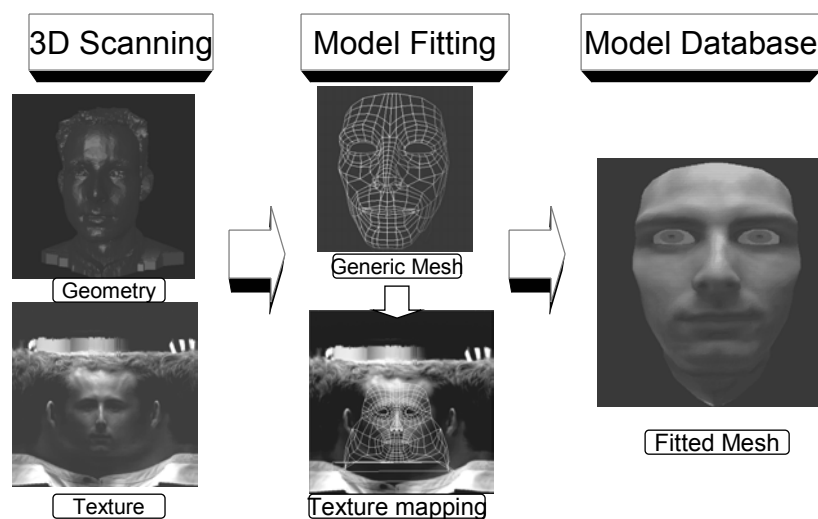
## Wavelet for Scalable Texture Coding (Cont.)

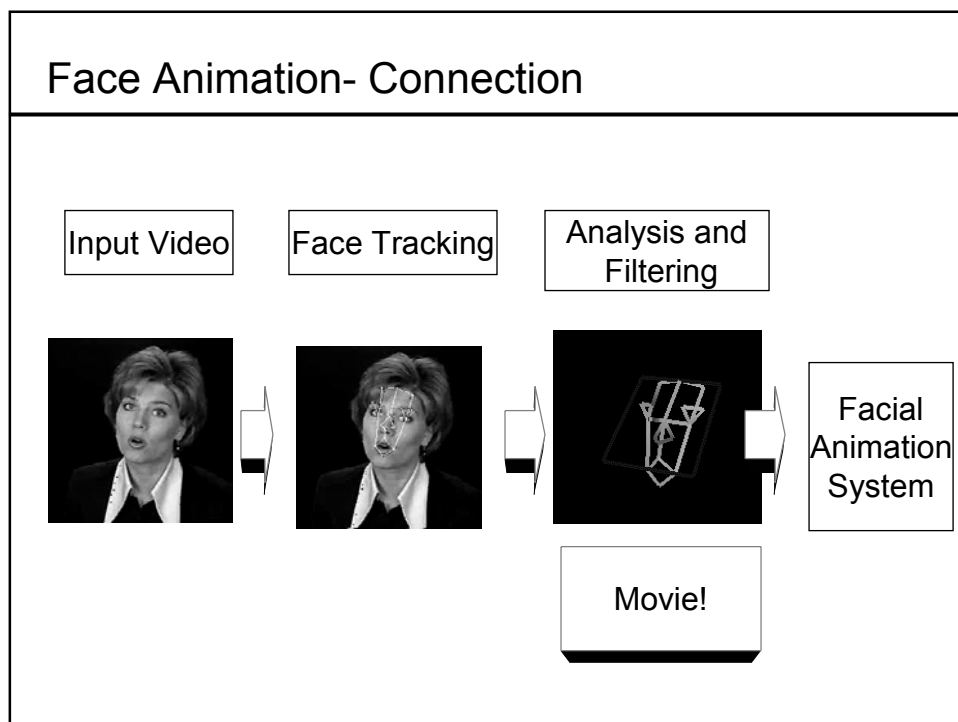
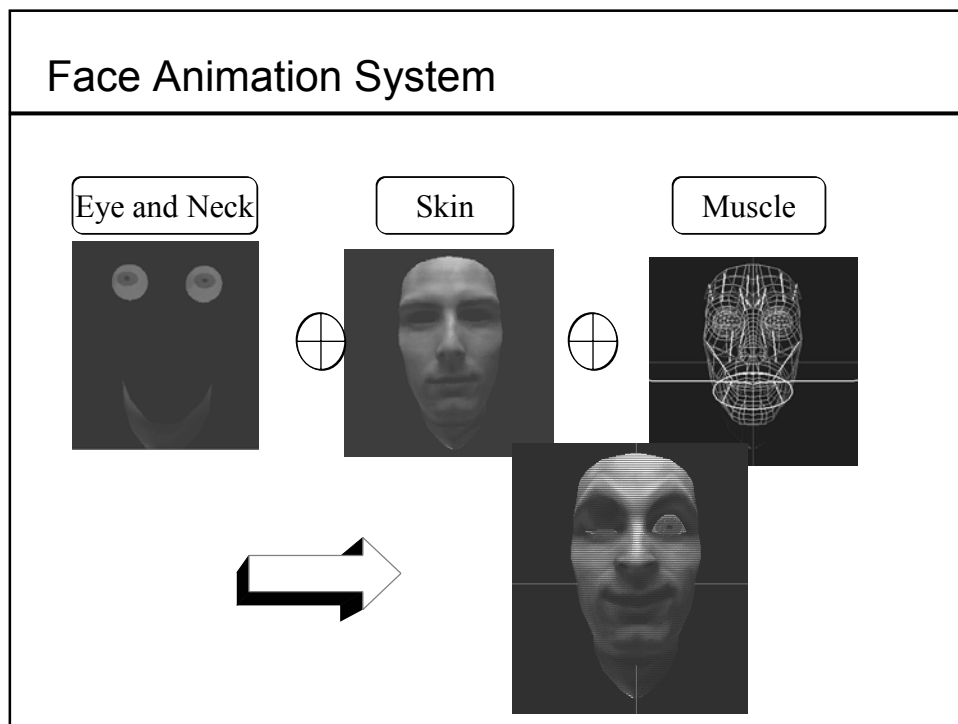


## SNHC VM

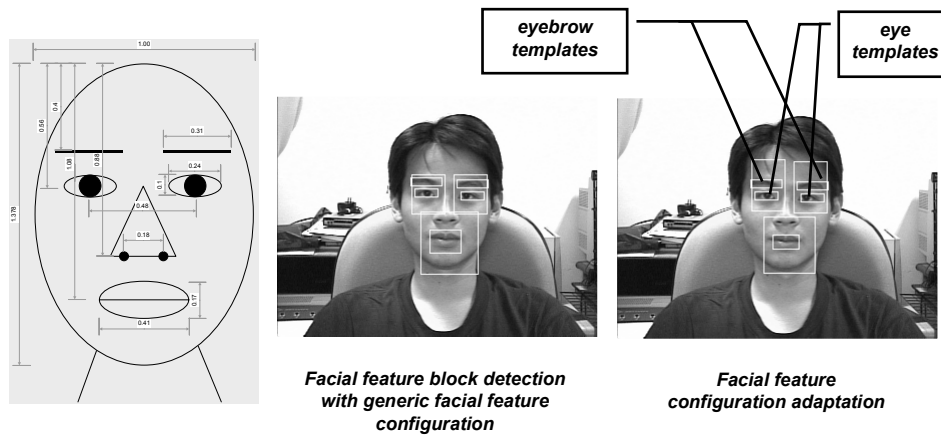
- Verification model
  - Three major parts
    - Face and body parameter description and animation
    - Media integration of text and graphics
    - Text-to-speech synthesis and interfaces with face animation
  - Application Program Interfaces (API) and methods necessary for composition and manipulation of audiovisual objects
  - APIs are intended to interface with the MPEG-4-Systems architecture

## 3-D Face Model Adaptation

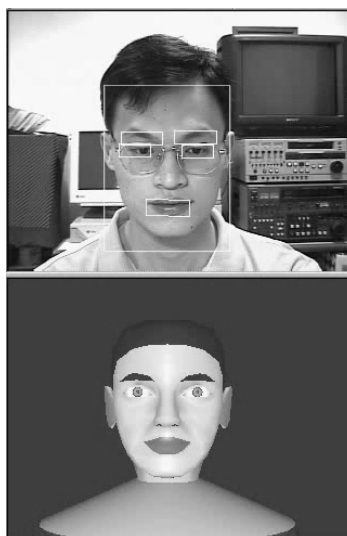




## Feature Block Extraction And Tracking



## Feature Block Tracking and Pose Estimation



## Facial Model Adaptation

- Construct 3-D facial model using two orthogonal views.



## Facial Model Adaptation

- Texture blending and texture mapping on a virtual cylindrical texture map



## Synthetic Avatar w/wo Texture Mapping



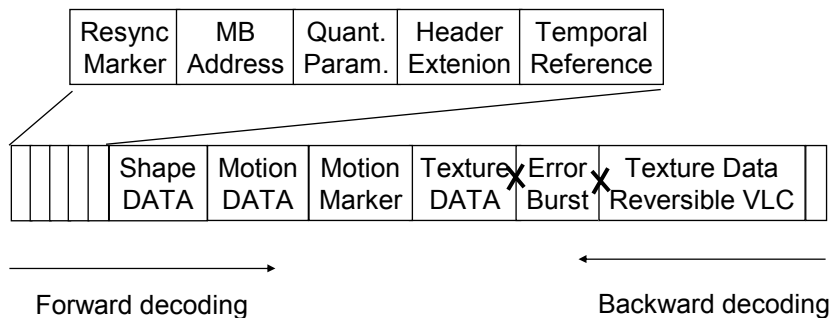
## MPEG-4 Visual Still Texture Coding

- Wavelet image coder
- Intended for rendering of mesh and face objects
- Fine granularity scalability supported
- Error robustness

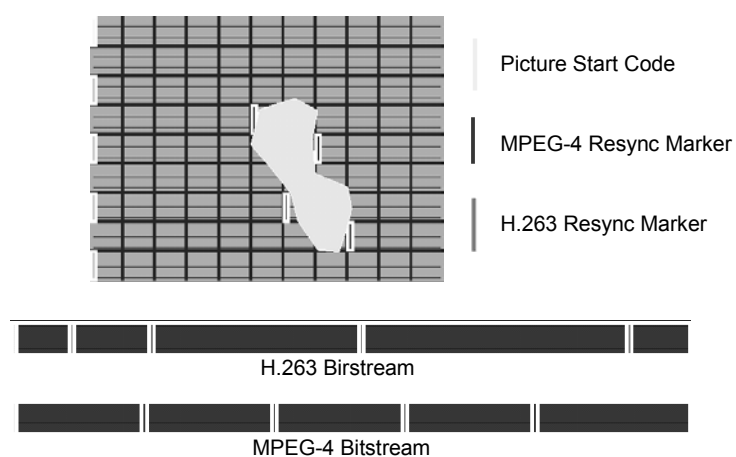


## Error Resilience Tools

- Resync marker can be placed at any MB boundary (resync with MB address) (e.g. resync evenly spaced in bit-stream)
- Separating motion vectors and the texture data with resync
- RVLC (Reversible Variable Length Coding)
- HEC (Header Extension Code)



## Error Resilience Tools- Resync Marker

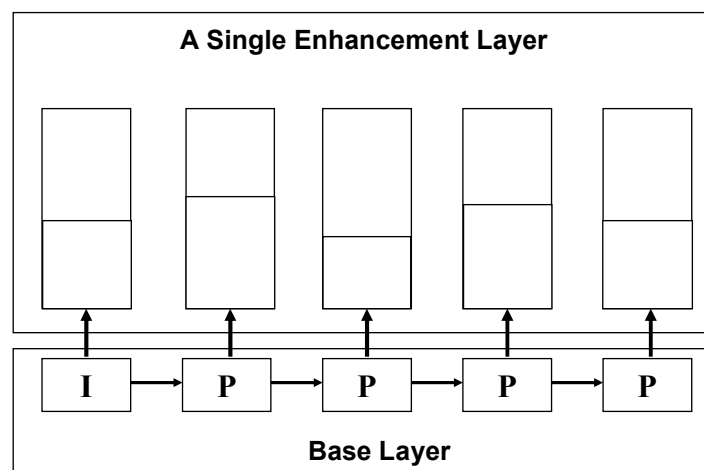


## MPEG-4 Fine Granularity Scalability (FGS)

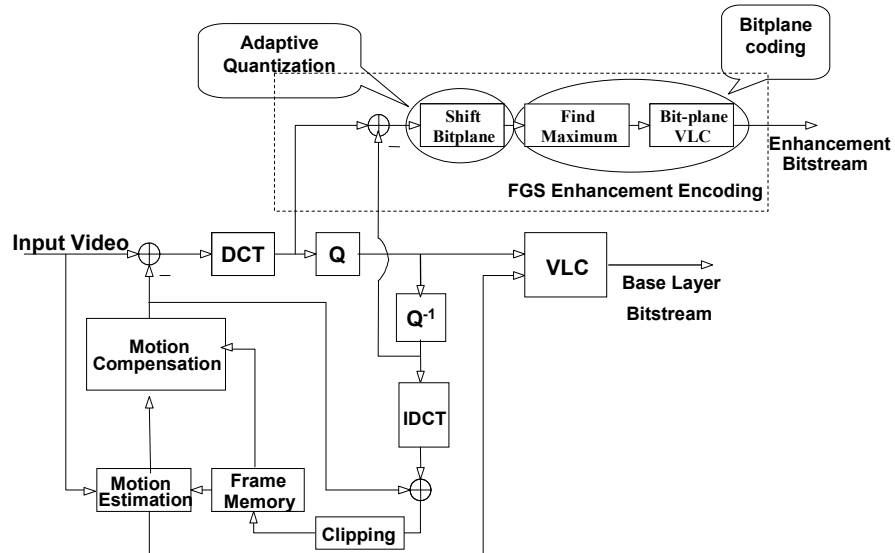
- **Internet applications**
- **broadcast applications over packet networks**

- Low complexity
- Supports both unicast & multicasting capabilities
- Supports various layers of SNR enhancements
- Covers a “range” of bitrates instead of a few discrete bitrates
- Base-layer compatible to MPEG-4
- Error robustness

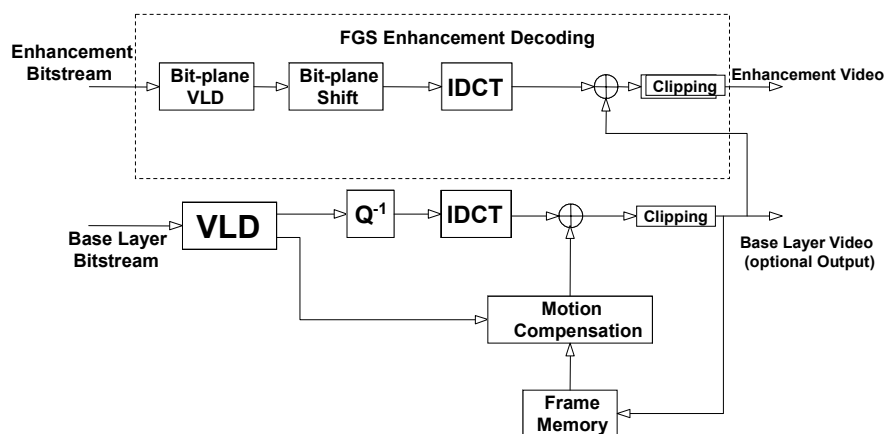
## Rate Adaptation with MPEG-4 FGS



## FGS Encoder



## FGS Decoder



	Simple	Core	Main	Simple Scalable	N-Bit	Hybrid	Basic Animated Texture	Still Scalable Texture	Simple Face	
<b>VISUAL TOOLS:</b>										
Basic (I-VOP, P-VOP, AC/DC Prediction, 4-MV, Unrestricted MV)	X	X	X	X	X	X				
Error Resilience	X	X	X	X	X	X				
Short Header	X	X	X		X	X				
B-VOP		X	X	X	X	X				
P-VOP with OBMC (Texture)										
Method 1/2 Quantization		X	X		X	X				
P-VOP based temporal scalability (rectangular, arbitrary shaped)		X	X		X	X				
Binary Shape		X	X		X	X	X			
Gray Shape			X							
Interlace			X							
Sprite			X							
Temporal Scalability (rectangular)				X						
Spatial Scalability (rectangular)				X						
N-Bit					X					
Scalable Still Texture						X	X	X		
2D Dynamic Mesh with uniform topology						X	X			
2D Dynamic Mesh with Delaunay topology						X				
Facial Animation Parameters						X	X		X	

## Profiles

Profiles:

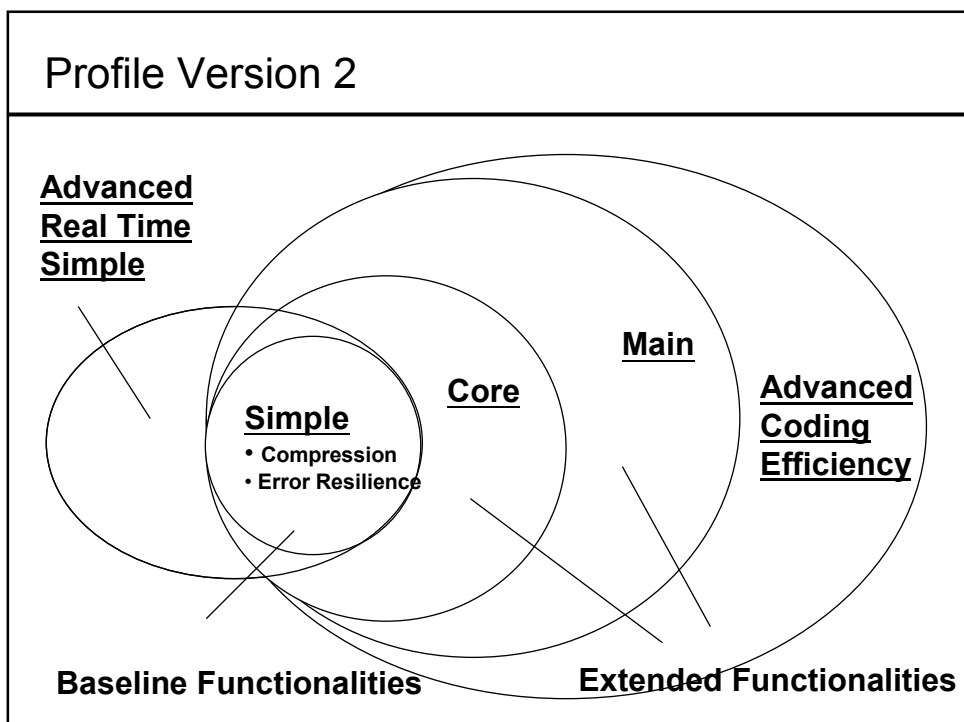
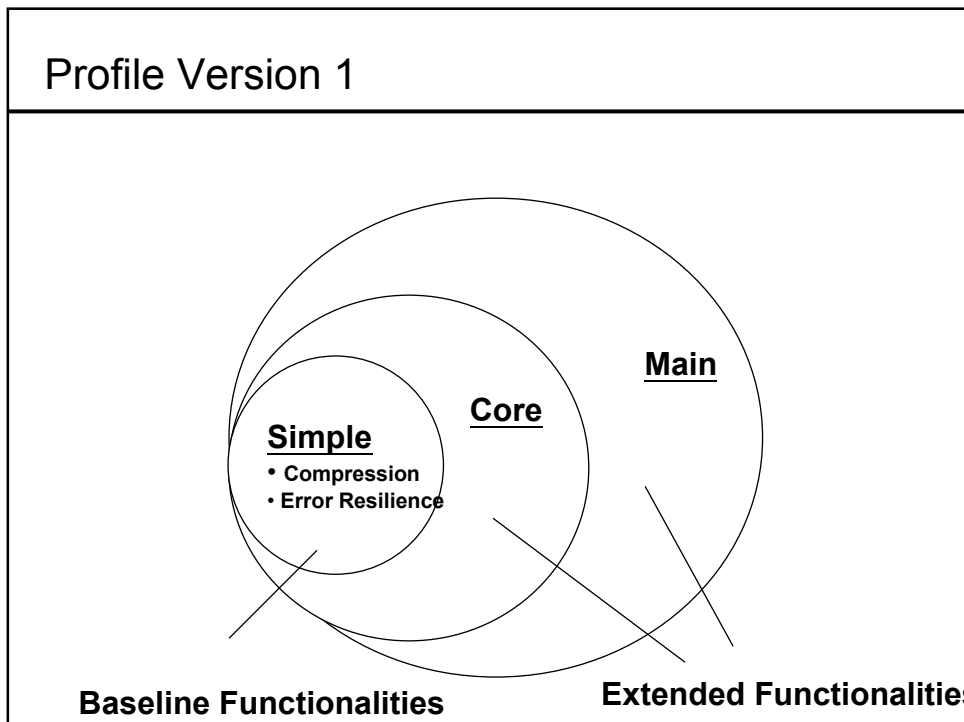
Simple: H.263-like

Simple Scalable: Simple + rectangular scalability

Core: Simple + binary shape + scalability

Main: Core + gray shape + interlace + sprite

...



## Proposed Studio Profile

- Application: Editing systems, Archive systems
- Pixel depth: 8/10 bits
- Rate: up to 600 Mb/s
- Chroma format: 4:2:0, 4:2:2, 4:4:4
- Objects of arbitrary shape, binary shape, gray-scale shape
- Permits lossless transcoding from MPEG-2 to MPEG-4

## Proposed Streaming Video Profile

- Application: Internet, datacasting
- Rate: up to 1 Mb/s
- Chroma format: 4:2:0
- Supports scalability with large number of layers

## Performance Baseline

- **Coding Efficiency (MAIN Profile)**

**(optimized headers, VLC tables)**

- 30 % reduction in bitrate compared to MPEG-1  
(tested between 40-768 kbit/s)

**Coding Efficiency (Advanced Coding Eff. Profile (ACE))**

**(with global motion estimation)**

- 30-50% reduction in bitrate compared to MAIN Profile  
(218 - 1000 kb/s) (critical sequences)

## Standards Comparison

JPEG - still image coding

- DCT + VLC
- simple hardware, low cost

H.261 - video-conferencing (64 kb/s - 1.92 Mb/s)

- DCT + VLC + optional integer-pixel MC
- progressive video

MPEG-1 - storage based applications (1.5 Mb/s)

- DCT + VLC + optional half-pixel bi-directional MC
- progressive video

MPEG-2 - general high quality applications (> 2 Mb/s)

- DCT+VLC+optional half-pixel bi-directional frame/field based MC
- progressive and interlaced video

H.263 - improved H.261 with four optional modes

H.263+ - improved H.263 with 12 new optional modes

- error resilience coding

MPEG-4 - improved H.263 with object-based coding and manipulations

## Delivery of MPEG-4 Content Over ...

### **MPEG-2, IP, MP4, Mobile, ATM, DAB**

<i>Technology</i>	<i>Specified in</i>	<i>Discussion in</i>	<i>Control Plane Data Plane</i>	<i>Glue for Control Plane</i>
<b>4on2</b>	MPEG2 Systems	MPEG Systems	MPEG2 Systems MPEG2 Systems	-
<b>4onIP</b>	IETF	IETF, MPEG4	MPEG4 DMIF IETF	DMIF V1
<b>4onMP4</b>	MPEG4 Systems	MPEG4 Systems	MPEG4 Systems MPEG4 Systems	-
<b>4onMobile</b>	ITU-T	MPEG4 DMIF ITU-T	ITU-T -	DMIF V2
<b>4onATM</b>	ITU-T	MPEG	MPEG4 DMIF -	DMIF V1
<b>4onDAB</b>	ETSI	MPEG4 Systems	MPEG4 DMIF MPEG4 DMIF	DMIF V1

## Universal Multimedia Access (UMA)

- Any Network:
  - wireline or wireless
  - circuit-switched or packet-switched
  - symmetrical or asymmetrical
  - broadband or bandwidth-limited
  - LAN, MAN or WAN
- Any Terminal:
  - various processing capability
  - various storage capability
  - various display capability
- Any Where:
  - seamless ubiquitous access