CS 414 – Multimedia Systems Design

Lecture 12 – MPEG-4 and H.264 (Part 7)

Klara Nahrstedt

Spring 2011

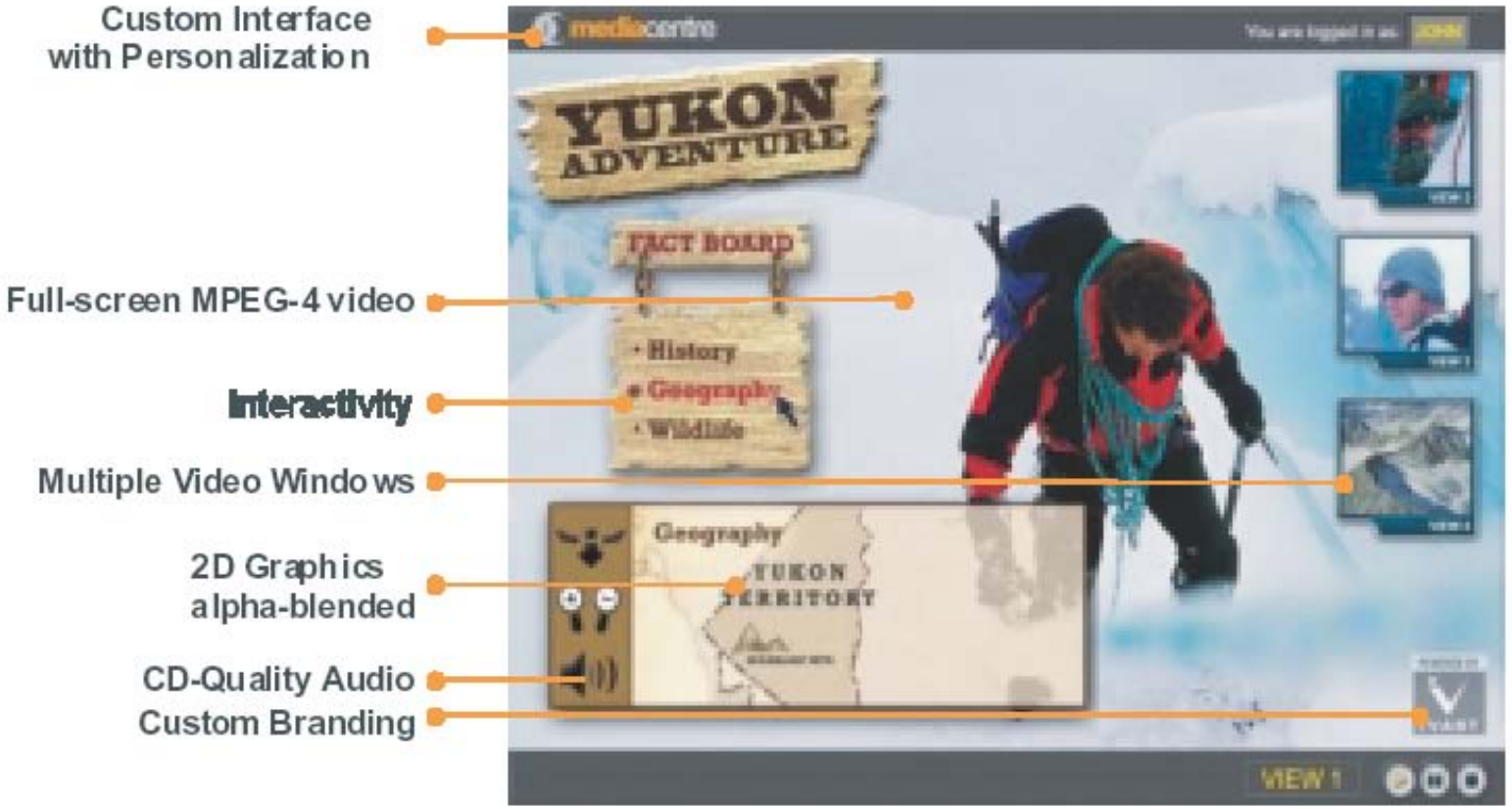
Administrative

- MP1 – deadline February 18
- Homework 1 – posted February 21
- Watch for Android tutorials
  - See posting on newsgroup
  - Organized by ACM SIGSoft

Outline

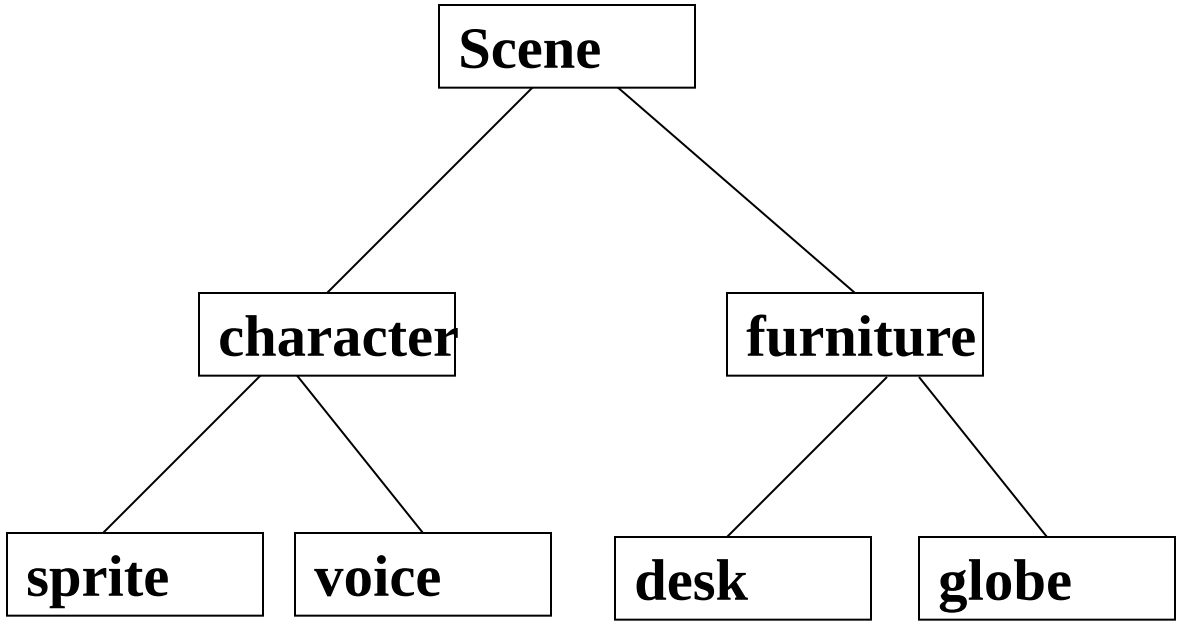
- MPEG-4
- Reading:
  - Media Coding book, Section 7.7.2 – 7.7.5
  - http://www.itu.int/ITU-D/tech/digital-broadcasting/kiev/References/mpeg-4.html
  - http://en.wikipedia.org/wiki/H.264
- Available software
  - Xvid – free software
  - http://www.bing.com/videos/watch/video/xvid-free-download-mpeg-4-video-codec-for-pc

MPEG-4 Example



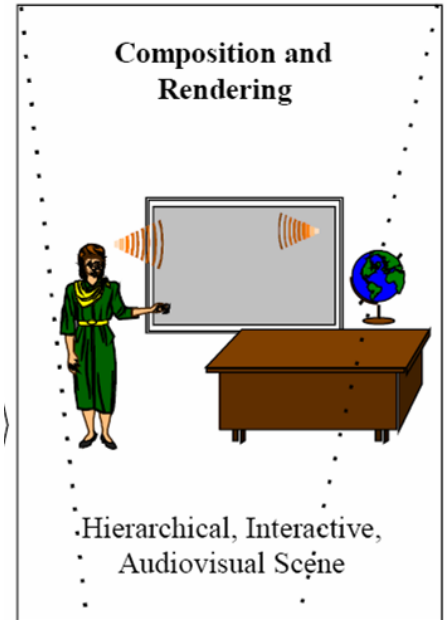
ISO N3536 MPEG4

Composition



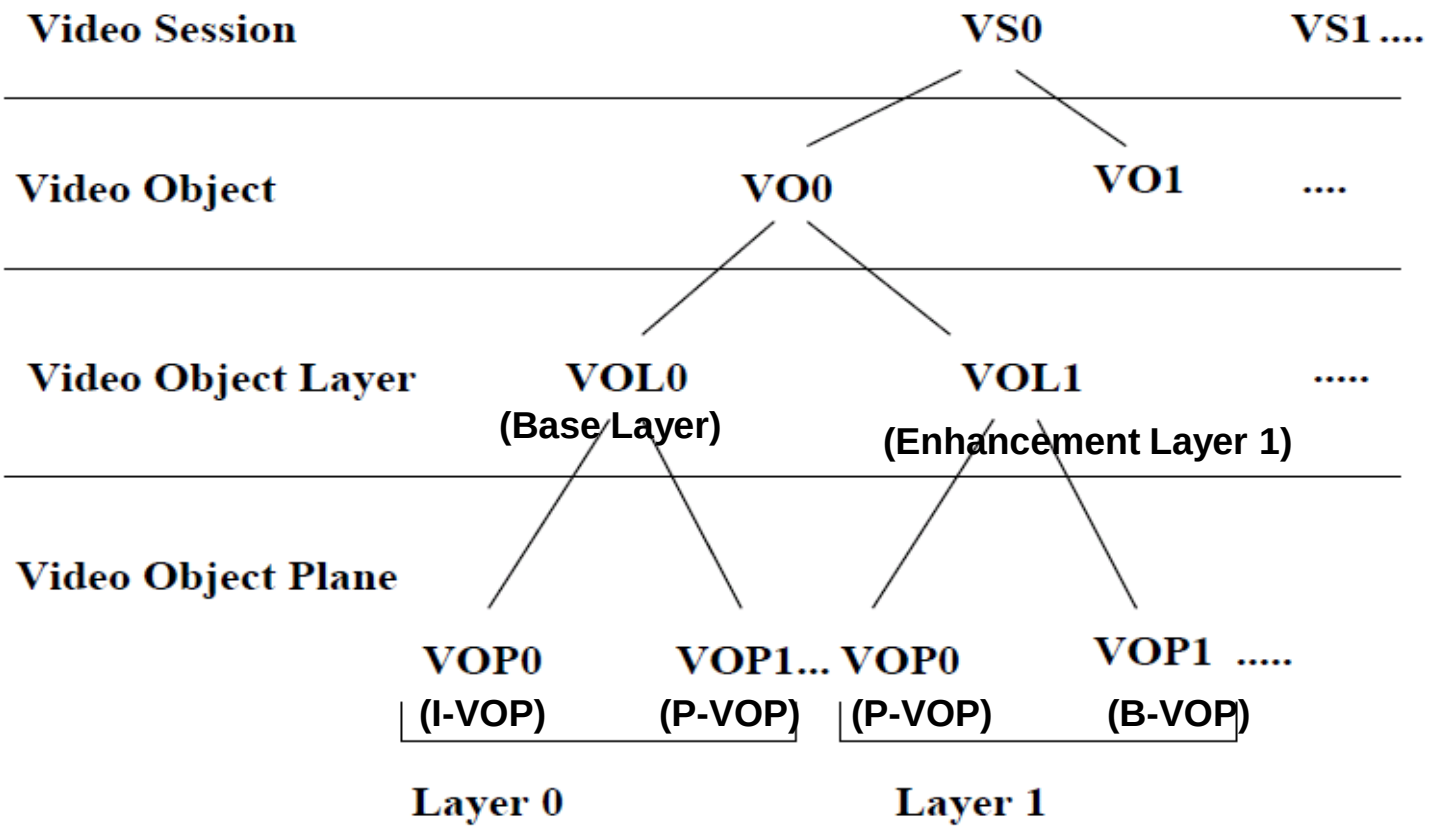
(Video Object VO0)

(Video Object VO1)



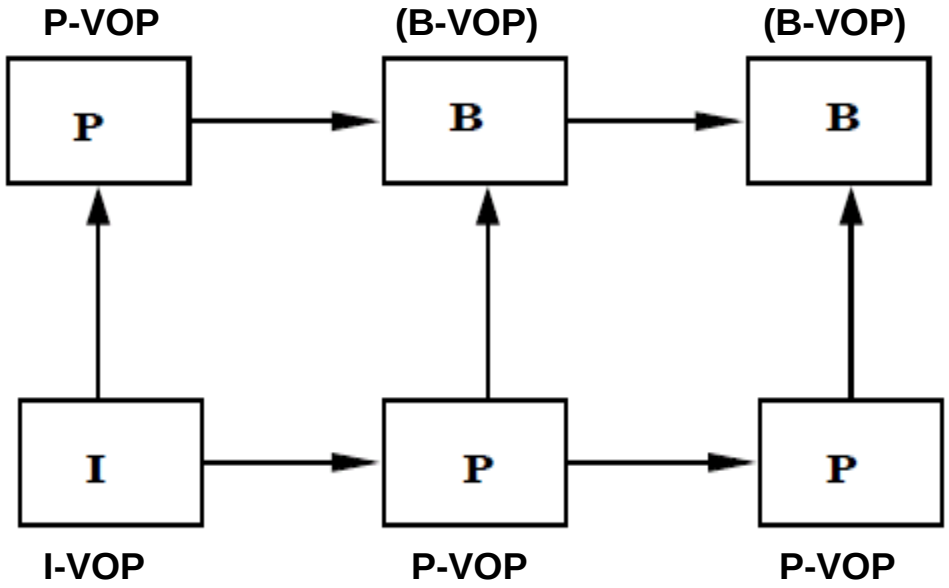
ISO N3536 MPEG4

Video Syntax Structure



New MPEG-4 Aspect: Object-based layered syntactic structure

Spatial Scalability



Enhacement Layer of a VOP

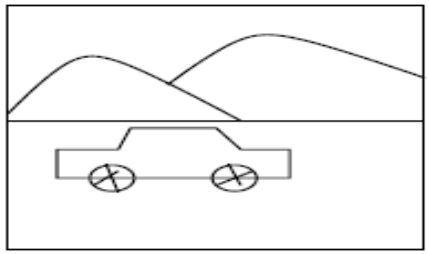
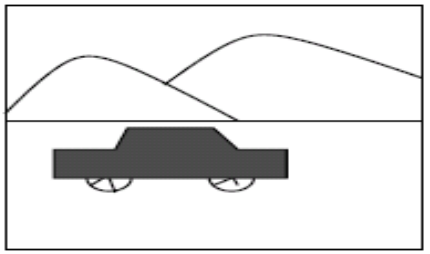
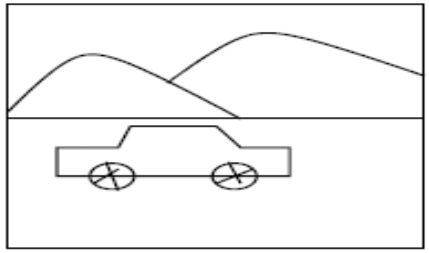
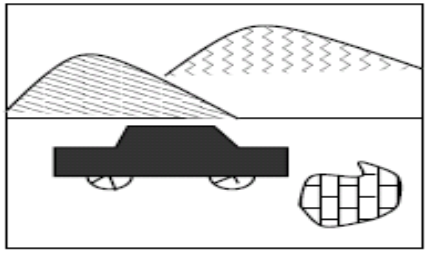
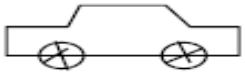
(e.g., VOL1 for VO0)

Base Layer

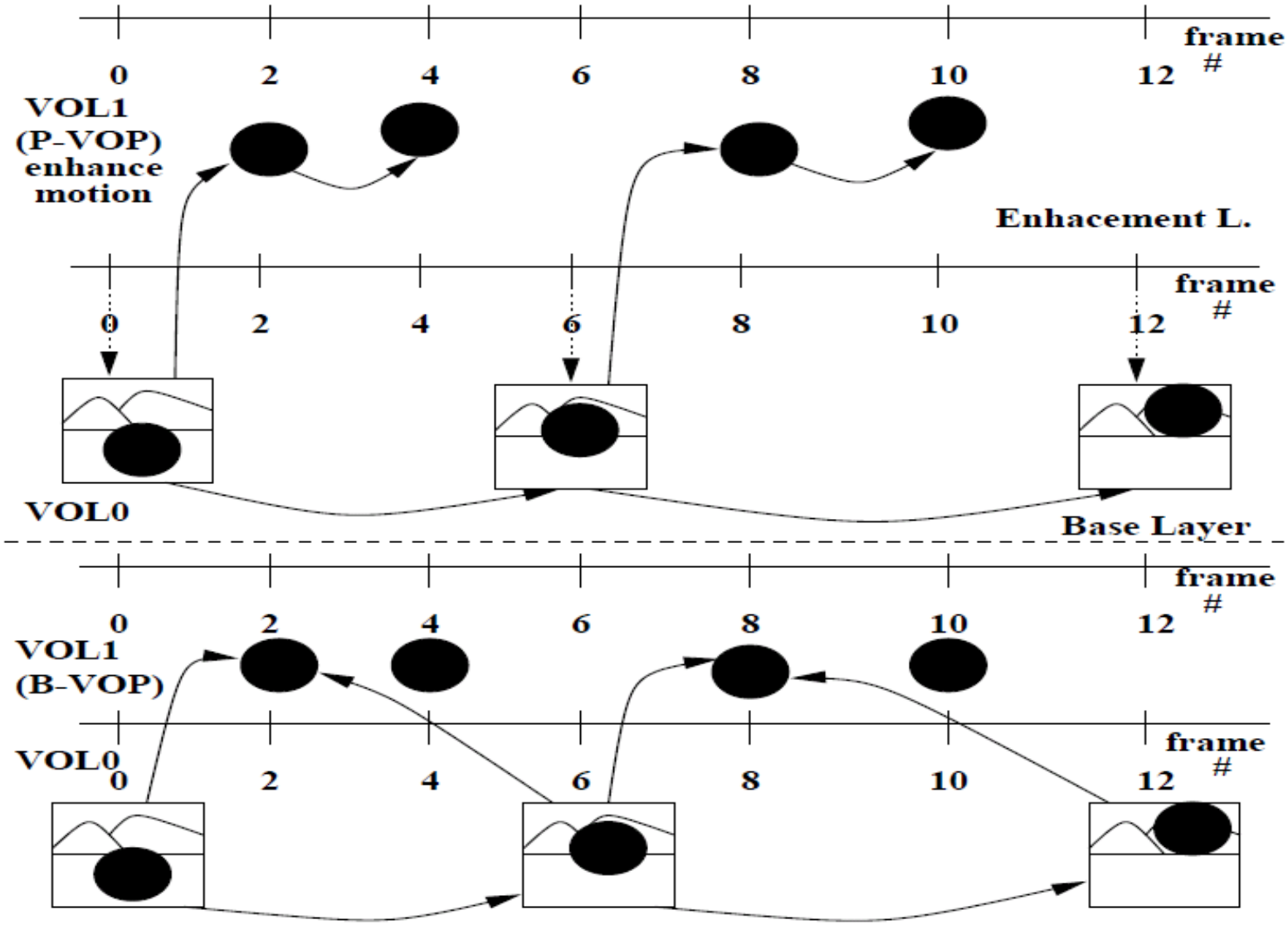
(e.g., VOL0 for VO0)

VOP which is temporally coincident with **I-VOP** in the base layer, is encoded as **P-VOP in the enhancement layer**.
VOP which is temporally coincident with **P-VOP** in the base layer is encoded as **B-VOP in the enhancement layer.**

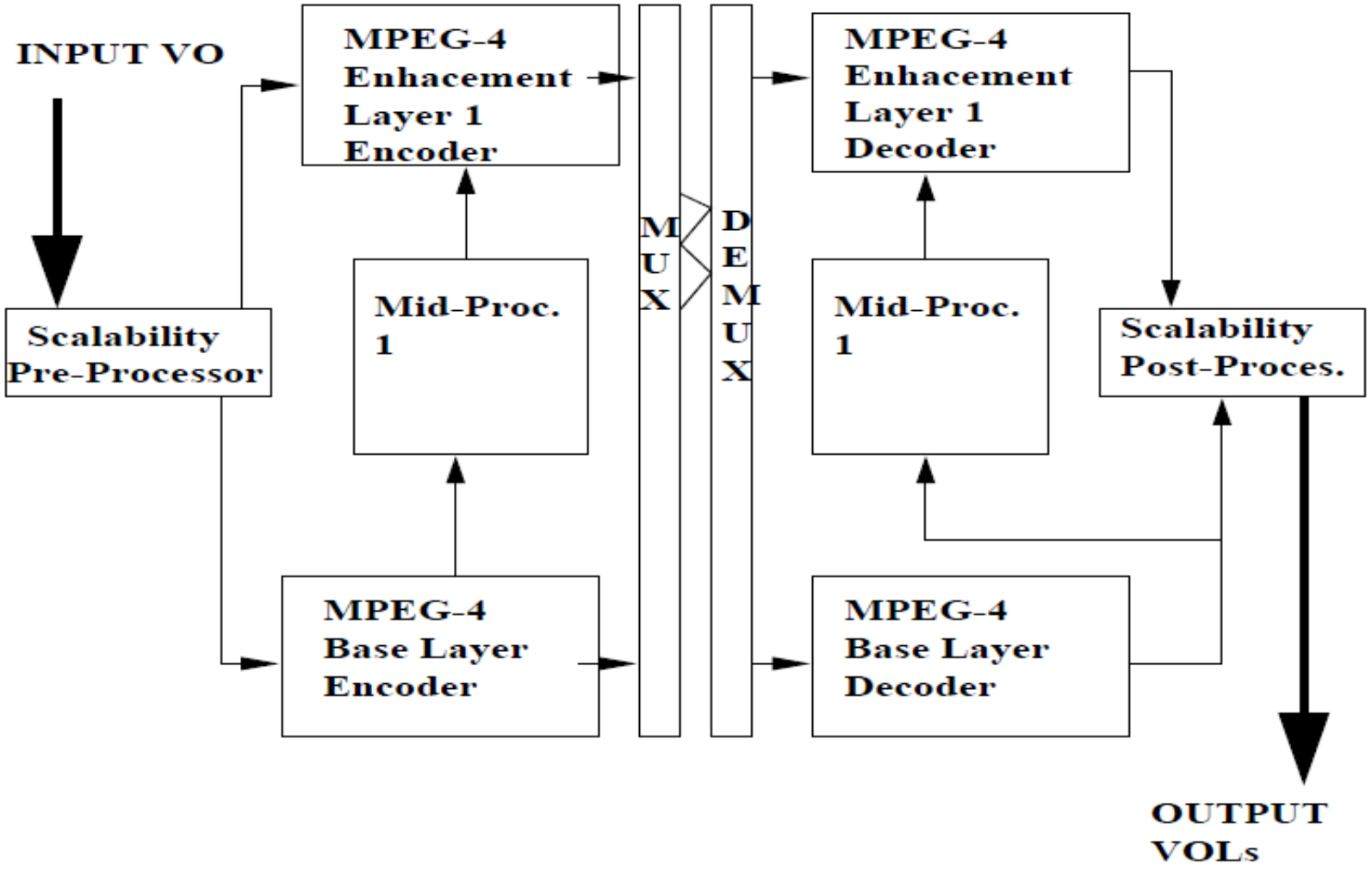
Examples of Base and Enhancement Layers

	Base Layer	Enhacement
Enhacement Type = 1	 VOL0: Entire Frame	 VOL1: Car
Enhancement Type = 2	 VOL0: Entire Frame	 VOL1: Entire Frame
	 VOL0: car	 VOL1: car

Temporal Scalability



High Level Codec for Generalized Scalability



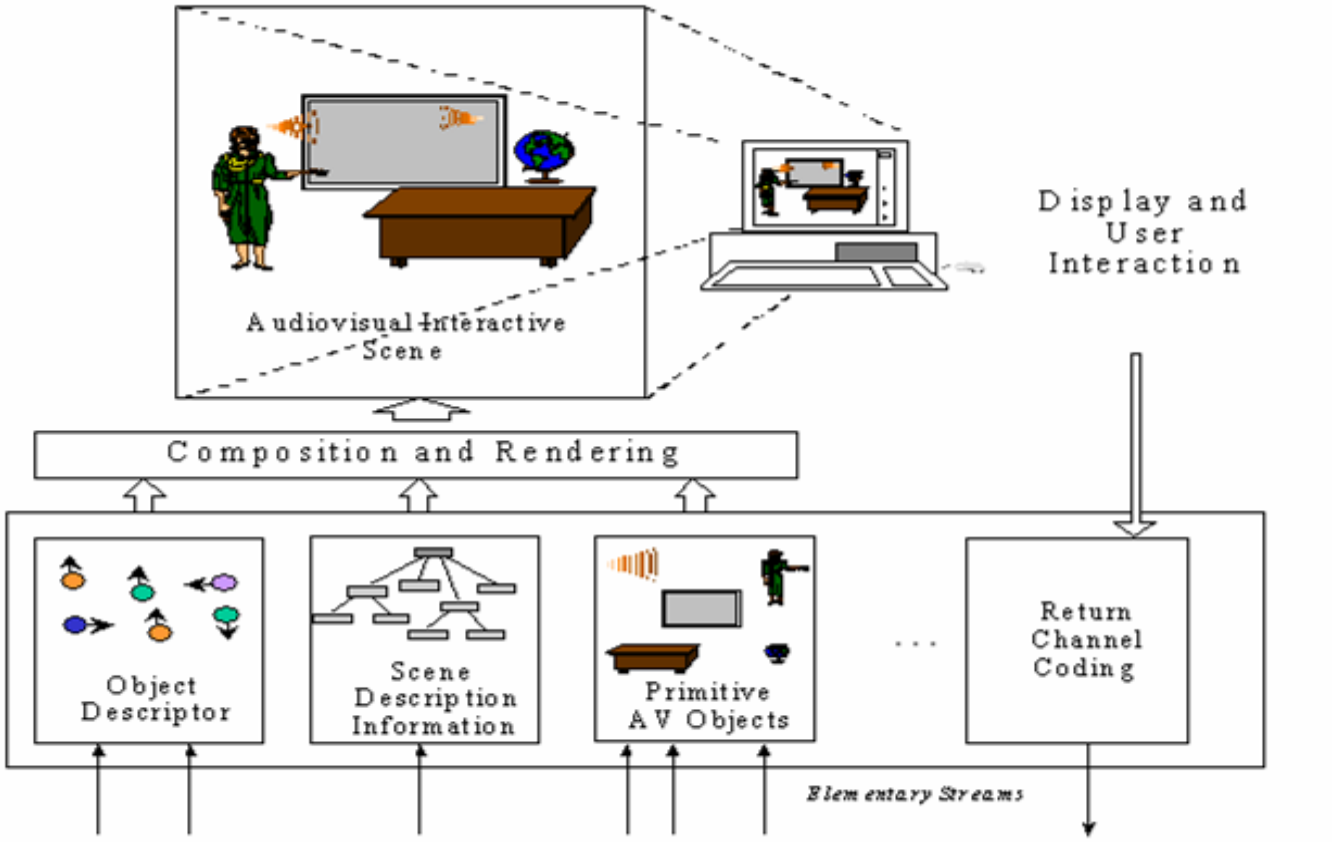
Composition (cont.)

- Encode objects in separate channels
  - encode using most efficient mechanism
  - transmit each object in a separate *stream*
- Composition takes place at the decoder, rather than at the encoder
  - requires a binary scene description (BIFS)
- BIFS is low-level language for describing:
  - hierarchical, spatial, and temporal relations

MPEG-4 Part 11 – Scene Description

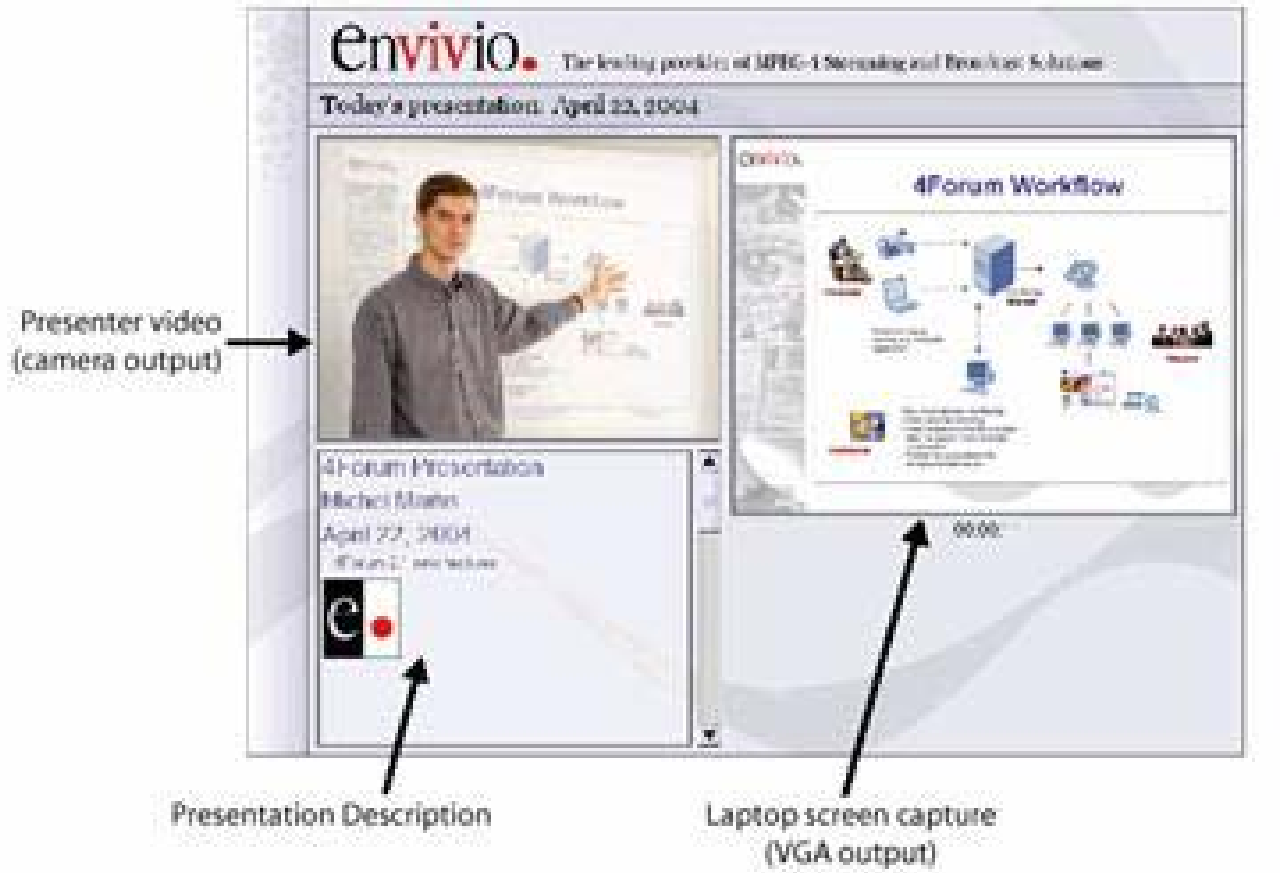
- BIFS – Binary Format for Scenes
  - Coded representation of the spatio-temporal positioning of audio-visual objects as well as their behavior in response to interactions
  - Coded representation of synthetic 2D and 3D objects that can be manifested audibly and/or visibly
- BIFS – MPEG-4 scene description protocol to
  - Compose MPEG-4 objects
  - Describe interaction about MPEG-4 objects
  - Animate MPEG-4 objects
- BIFS Language Framework – XMT (textual representation of multimedia content using XML)
  - Accommodates SMIL, W3C scalable vector graphics and VRML (now X3D)

MPEG-4 Rendering
(Composition at Decoder)



ISO N3536 MPEG4

Integration and Synchronization of Multiple Streams



Source: http://mpeg.chiariglione.org/technologies/mpeg-4/mp04-bifs/index.htm

Interaction as Objects

- Change colors of objects
- Toggle visibility of objects
- Navigate to different content sections
- Select from multiple camera views
  - change current camera angle
- Standardizes content and interaction
  - e.g., broadcast HDTV and stored DVD

Hierarchical Model

- Each MPEG-4 movie composed of tracks
  - each track composed of media elements (one reserved for BIFS information)
  - each media element is an object
  - each object is a audio, video, sprite, etc.
- Each object specifies its:
  - spatial information relative to a parent
  - temporal information relative to global timeline

Synchronization

- Global timeline (high-resolution units)
  - e.g., 600 units/sec
- Each continuous track specifies relation
  - e.g., if a video is 30 fps, then a frame should be displayed every 33 ms.
- Others specify start/end time

MPEG-4 parts

- MPEG-4 part 2
  - Includes Advanced Simple Profile, used by codecs such as Quicktime 6
- MPEG-4 part 10
  - MPEG-4 AVC/H.264 also called Advanced Video Coding
  - Used by coders such as Quicktime 7
  - Used by high-definition video media like Blu-ray Disc

MPEG-4 Audio

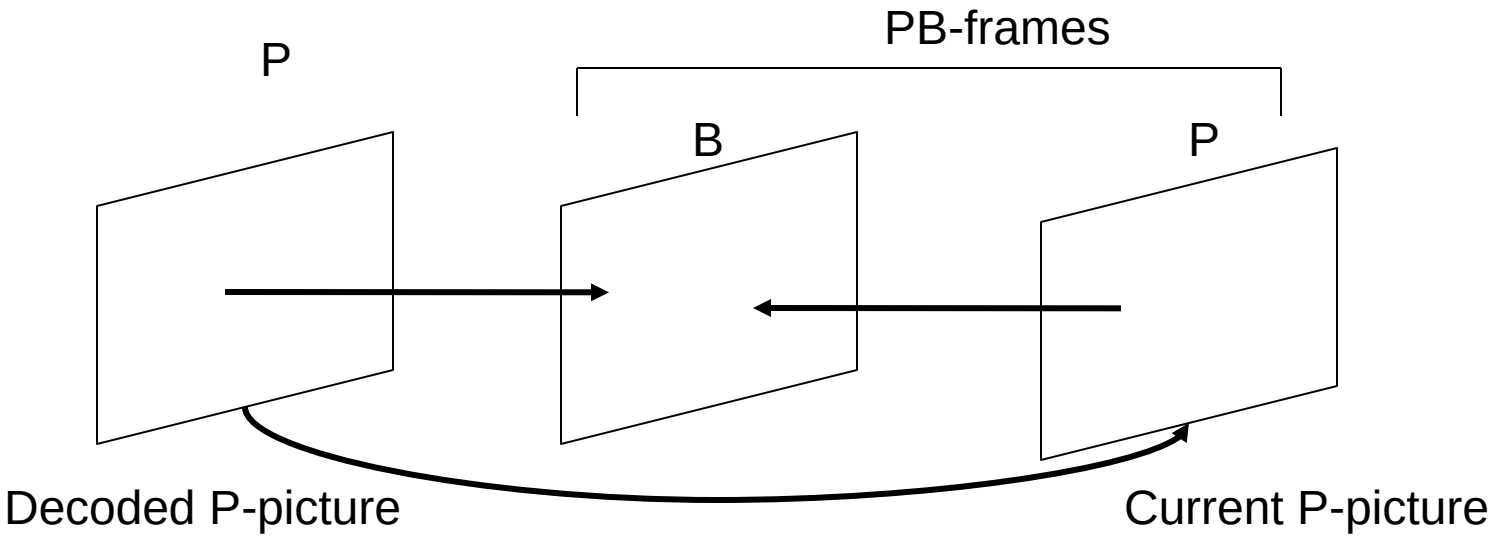
- Bit-rate 2-64kbps
- Scalable for variable rates
- MPEG-4 defines set of coders
  - Parametric Coding Techniques: low bit-rate 2-6kbps, 8kHz sampling frequency
  - Code Excited Linear Prediction: medium bit-rates 6-24 kbps, 8 and 16 kHz sampling rate
  - Time Frequency Techniques: high quality audio 16 kbps and higher bit-rates, sampling rate > 7 kHz

H.26X

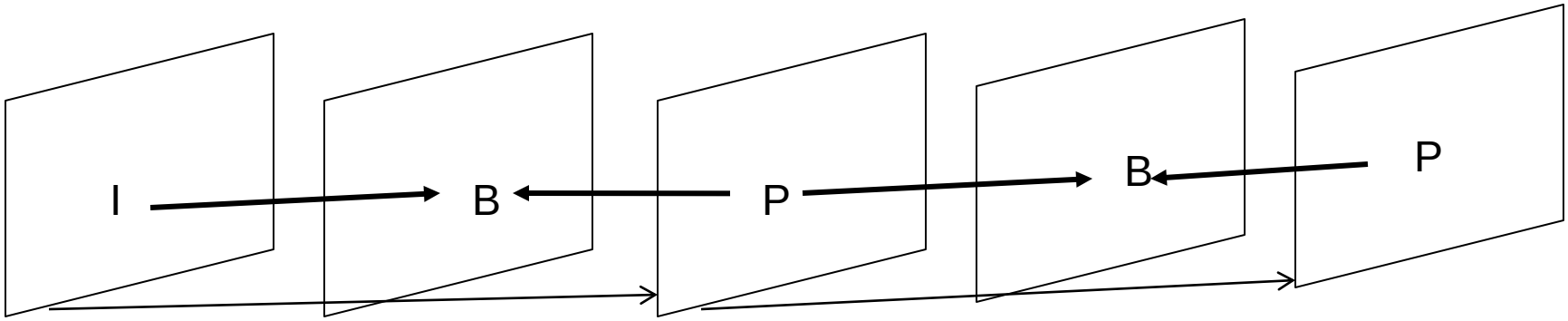
- H.261 – CCITT Recommendation of ITU-T Standard
  - Developed for interactive conferencing applications
  - Symmetric coder - real-time encoding and decoding
  - Rates of p x 64 Kbps for ISDN networks
  - Only I and P frames
- H.263 – established 1996
  - Used for low bit rate transmission
  - Improvements of error correction and performance
  - Takes in PB-frames mode
  - Temporal, Spatial and SNR scalability

H.263 – PB-Frames Mode

- A PB-frames consist of two pictures encoded as one unit.
- PB-frame consists of
  - One P-picture which is predicted from last decoded P-picture
  - One B-picture which is predicted from last decoded P-picture and the P-picture currently being decoded.



Comment on Temporal Scalability



- Temporal scalability is achieved using B-pictures
- These B pictures differ from B-picture in PB-frames
  - they are not syntactically intermixed with subsequent P-picture
- H.263 is used for low frame rate apps (e.g., mobile), hence in base layer there is one B-picture between I and P pictures.

H.264/MPEG-4 AVC Part 10

- Joint effort between
  - ITU- Video Coding Experts Group (VCEG) and
  - ISO/IEC Moving Picture Experts Group (MPEG)
  - 2003 completed
- H.264 – codec
  - Standard for Blu-ray Discs
  - Streaming internet standard for videos on YouTube and iTunes Store
  - web software Adobe Flash Player and Microsoft Silverlight support H.264
  - Broadcast services – direct broadcast satellite television services; cable television services

H.264 Characteristics

- Sampling structure
  - YCbCr 4:2:2 and YCbCr 4:4:4
- Scalable Video Coding (SVC) allows
  - Construction of bit-streams that contain sub-bit-streams that also conform to standard, including only “Base Layer” bit-stream (this can be decoded with H.264 without SVC support)
  - Temporal bit-stream scalability, spatial and quality bit-stream scalability
  - Complete in 2007

H.264 Characteristics

- Multi-view Video Coding (MVC)
  - Construction of bit-streams that represent more than one video of a video scene
    - Example: stereoscopic 3D video coding
  - Two profiles in MVC:
    - Multi-view High Profile (arbitrary number of views);
    - Stereo High Profile (two-view stereoscopic video);
  - Complete in 2009

H.264 Characteristics

- Multi-picture inter-picture prediction
  - Use previously-encoded pictures as references in more flexible way than in pas standards
  - Allow up to 16 reference frames to be used in some cases
    - Contrast to H.263 where typically one or in some cases conventional “B-pictures”, two.
  - Use variable block size from 16x16 to 4x4
  - Use multiple motion vectors per macro-block (one or two per partition where partition can be a block of 4x4)

H.264 Characteristics

- New Transform design features
  - Similar to DCT, but simplified and made to provide exactly-specified decoding
- Quantization
  - Frequency-customized quantization scaling matrices
    - selected by encoder based on perception optimization
- Entropy Encoding
  - Context-adaptive variable-length coding
  - Context-adaptive binary arithmetic coding

Conclusion

- A lot of MPEG-4 examples with interactive capabilities
- Content-based Interactivity
  - Scalability
  - Sprite Coding
- Improved Compression Efficiency (Improved Quantization)
- Universal Accessibility
  - re-synchronization
  - data recovery
  - error concealment
- H.264 – major leap forward towards scalable coding and multi-view capabilities
  - Some controversy on patent licensing