

MPEG Standards Evolution

4/20/01

Paul Moroney

Advanced Technology

Motorola Broadband Communications

The MPEG Process

- Motion Picture Experts Group
 - Begun in October 1988, under ISO
 - L. Chiariglione, convenor.
- MPEG 1: 1.5 Mbps target, storage, CD (Emmy)
- MPEG-2: Entertainment television, DVD (Emmy)
- “MPEG-3”: HDTV
- MPEG-4: Object representation
- MPEG-7: Indexing and retrieval
- MPEG-21: Universal Multimedia Access



MPEG-1 and MPEG-2 Review

- Video & audio elementary streams
 - Defined algorithms & syntax
 - Profiles and levels
- PES layer
 - Packetized elementary stream
 - timestamps
- Systems layer
 - Combine audios and videos



Video Compression: Basic Concepts

- Remove redundancy in video content, frame to frame, and within frame
 - Lossless coding
- Reduce total information further
 - Lossy coding
- Hide loss based upon video perception models:
 - Difficult to detect artifacts under high motion
 - Chroma resolution of vision is low

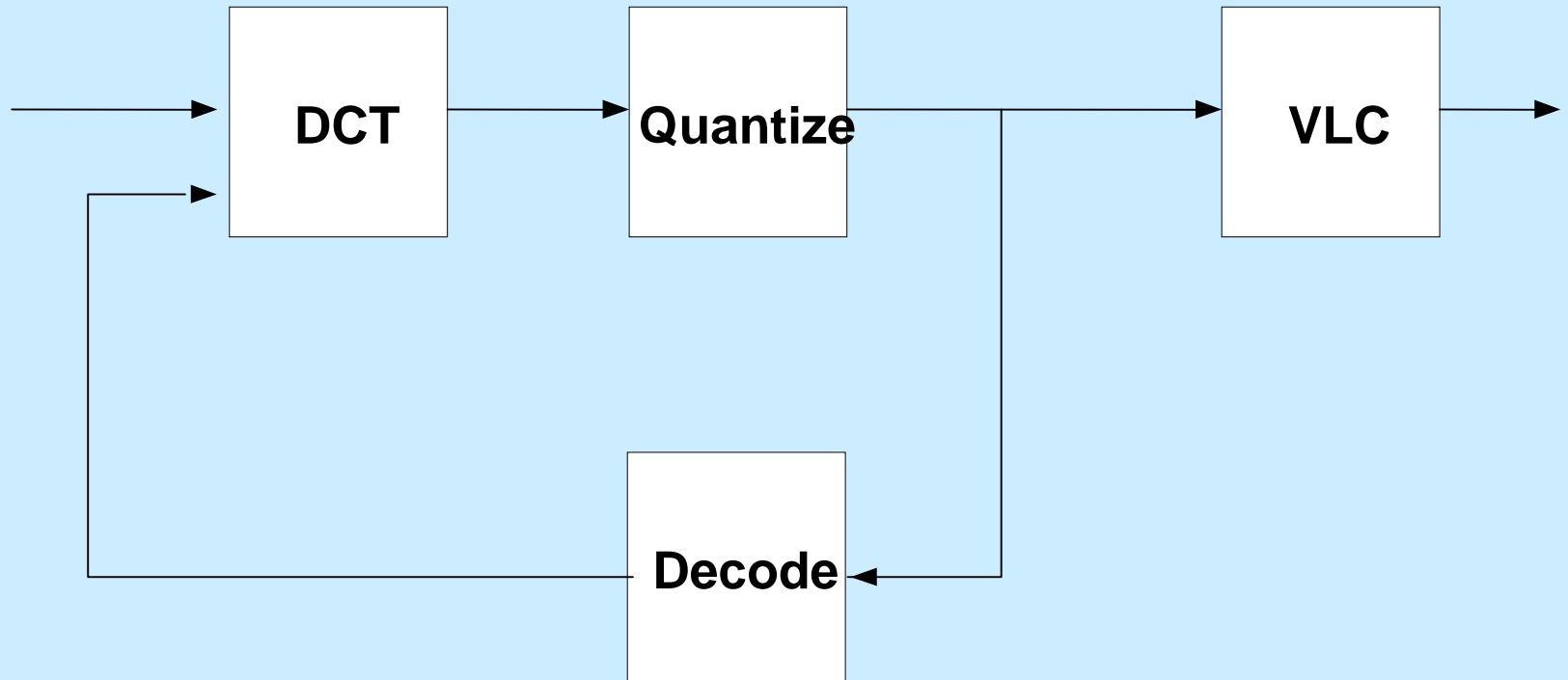


MPEG Video Essentials

- Transform, quantize, code
- 8 by 8 pixel block based approach
 - Compute DCT
 - Scan and Quantize coefficients [lossy]
 - Huffman VLC code [lossless]
- Group bits and headers into layered syntax
 - picture, slice, region, block
- Add Motion estimation [lossless]
 - Find best [macroblock] region in prior frame, only process difference – $\frac{1}{2}$ pel MV

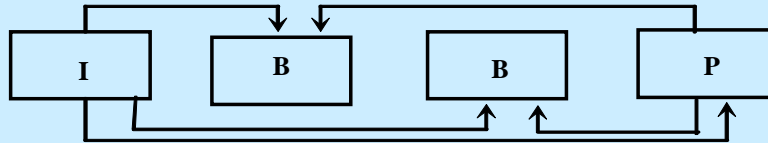


Basic MPEG Encoder



Video Compression

- Bidirectional prediction: Frames may have blocks predicted from a past and a future anchor frame (I- or P- frame)



- Field/frame mode [MPEG-2 only]
 - For interlace content especially, blocks should be formed and processed from every other line, for optimal compression



Field/Frame Mode



Encoding As an Art Form

- Which regions are predicted for best coding?
- Which motion vectors are the best?
- Which regions slices, or frames are best unpredicted, for refresh?
 - Error recovery
 - Acquisition
- How implement rate control [CBR/VBR/Stat mux]
- Which Quant level is best for each block or region?
- Which regions are best as field or frame?
- Which scan is optimal?
- Which Quant table is best?



MPEG Premise (Video and Audio)

- Decoders pure standards based, except for
 - Error recovery
 - Display
- Encoders allow substantial room for creativity and invention, within the context of the standard, and subject to ever changing complexity limits.
 - Attempts to define objective perception quality measures largely fail [PSNR]



MPEG Audio

- Perception follows "Critical Bands," masking related, ear is multi-band analyzer.
- Lossy compression [quantization] driven by perceptual masking phenomena.
- MPEG-1 offers 3 layers of algorithms, of increasing complexity and performance
 - MUSICAM: Layer 2 [sub-band coding]
 - MP3: layer 3
- MPEG-2 adds
 - compatible multi-channel surround
 - NBC mode: AAC



Systems Layer

- MPEG-1 offers storage oriented, file oriented systems layer
 - Program Stream
- MPEG-2 introduces transmission oriented packet layer
 - Transport Stream
 - 188 byte packets, easy to multiplex into multi-service higher rate bit pipe for cable, satellite, and terrestrial broadcast
- Directory Structure defined



Why MPEG-4?

- Broad applicability [10 kbps to 100 Mbps]
- New compression tools are invented over time
- Rich media types
 - Graphics, text, speech, synthetic audio, animation
- Object orientation
 - Standards based interactivity
- Increased flexibility
 - Features, transport, file format



MPEG-4 Rate Targets

- “Equal or better”
- Studio and Cinema
- MPEG-2 application, overhead issue
- MPEG-1 range
- H.263 [conferencing]
- Telephone line applications, wireless
 - Real,
 - Microsoft,
 - Quicktime



MPEG-4 Video Compression

- Performance/complexity tradeoff
 - Over time, more tools are cost effective
- H.263+, H.26L
- Constant improvement
- Key new application:
 - Standards based streaming media

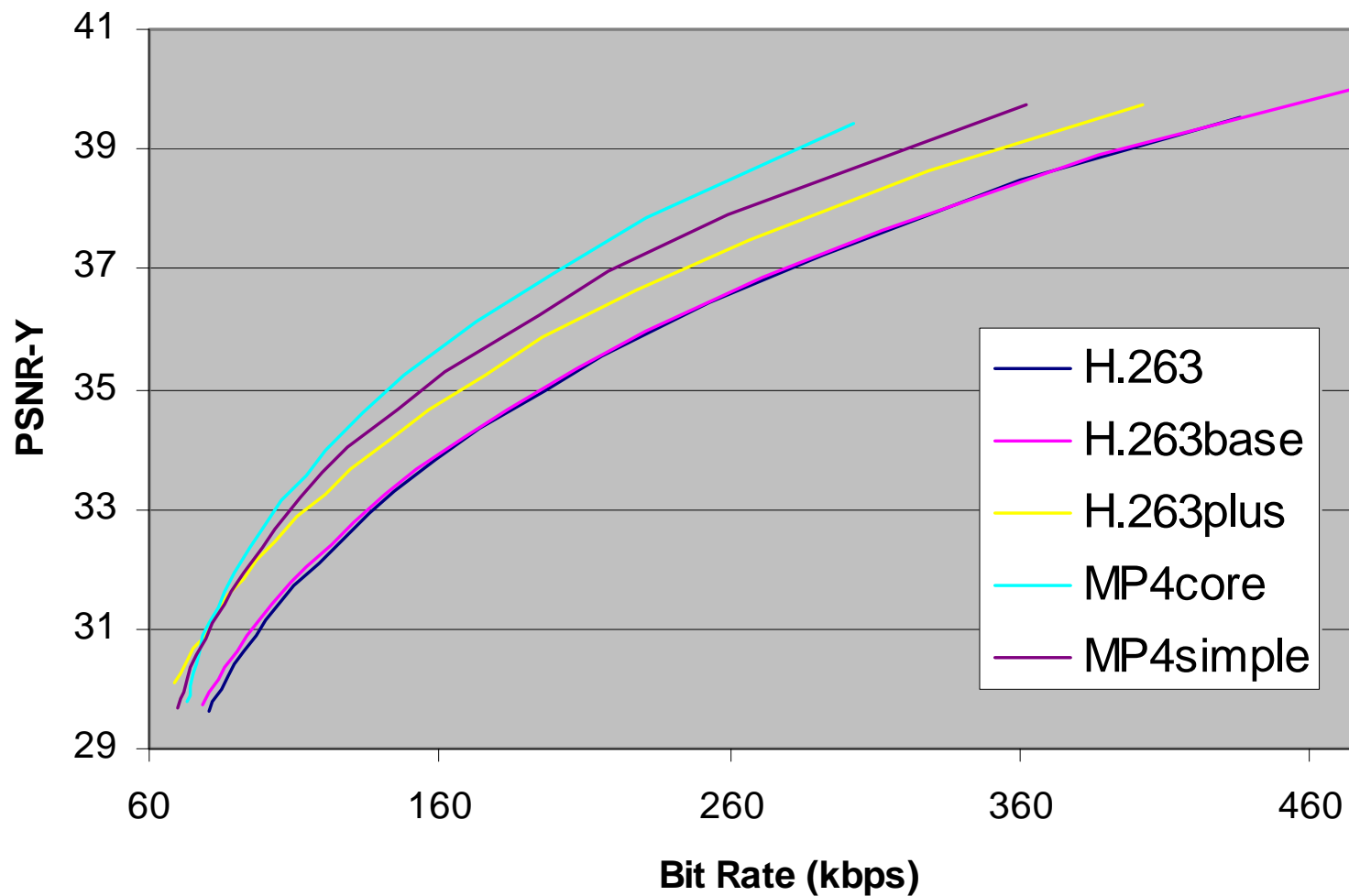


Mpeg-4 Video Tools

- Arbitrary frame rate
- AC coefficient prediction
- Block based motion vectors
- Block based I/P decision
- Motion: Quarter pel, UTMV, GMC
- Global motion compensation
- Formula based quant matrix
- Third scan option
- Short header



Coding Comparison: "News"



MPEG-4 Rich Media Support

- MPEG-4 defines new media types, “natural” and synthetic, and compression algorithms appropriate to them
- Consider rolling credits in DVD movie
 - DCT is poor match
- Far better performance at increase in complexity



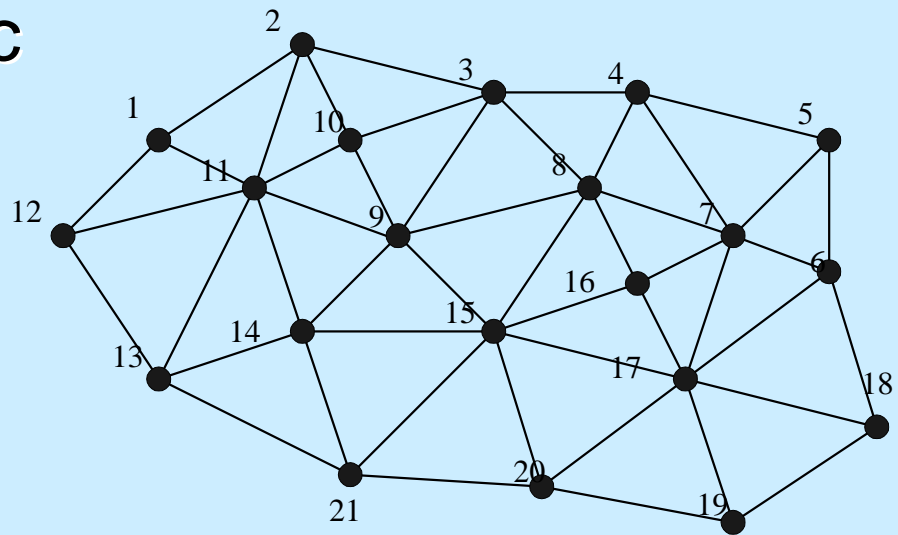
MPEG-4 Media Types

- Still image
 - Zerotree wavelet
- Graphics
 - 2D and 3D polygonal mesh
 - Textures mapped on mesh
- Text
 - Unicode, font, color, streaming



MPEG-4 Geometry Compression

- Minimization of storage and transmission of original mesh
- 2D and 3D topologies.
- Coding of vertex positions, attributes and connectivity
- Static and Dynamic meshes



MPEG-4 Media Types

- Speech
 - Twin VQ, CELP
 - Text-to-speech
- Synthetic audio
 - Download synthesis approach (FM, Wavetable, Models, etc.)
 - Play MIDI protocol, or enhanced “score”
- Animation
 - Face and body models; send only animation commands. [ex: Videophone]

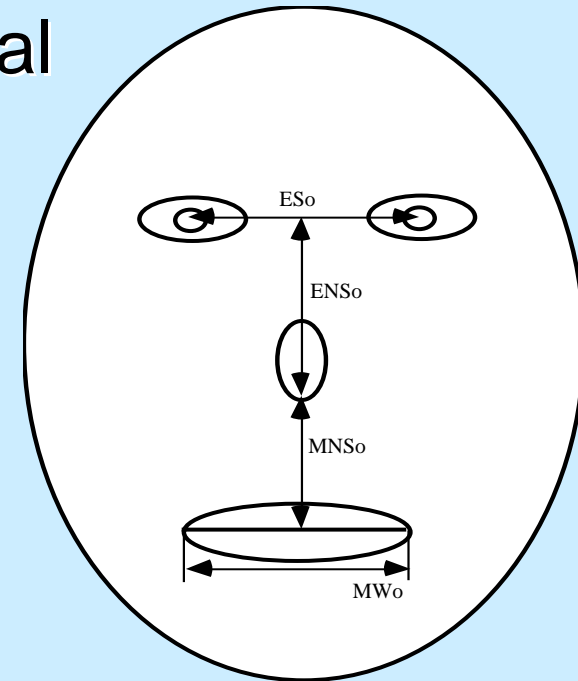


MOTOROLA

Broadband Communications Sector

MPEG-4 Face and Body Animation

- Key Facial Features
- FDP: Definition of a precise facial shape and texture during the setup phase.
- FAP: animation parameters for different facial models without having to initialize or calibrate the facial model.
- Similar idea for human body descriptions (BDP and BAP):
 - larger parameter set.

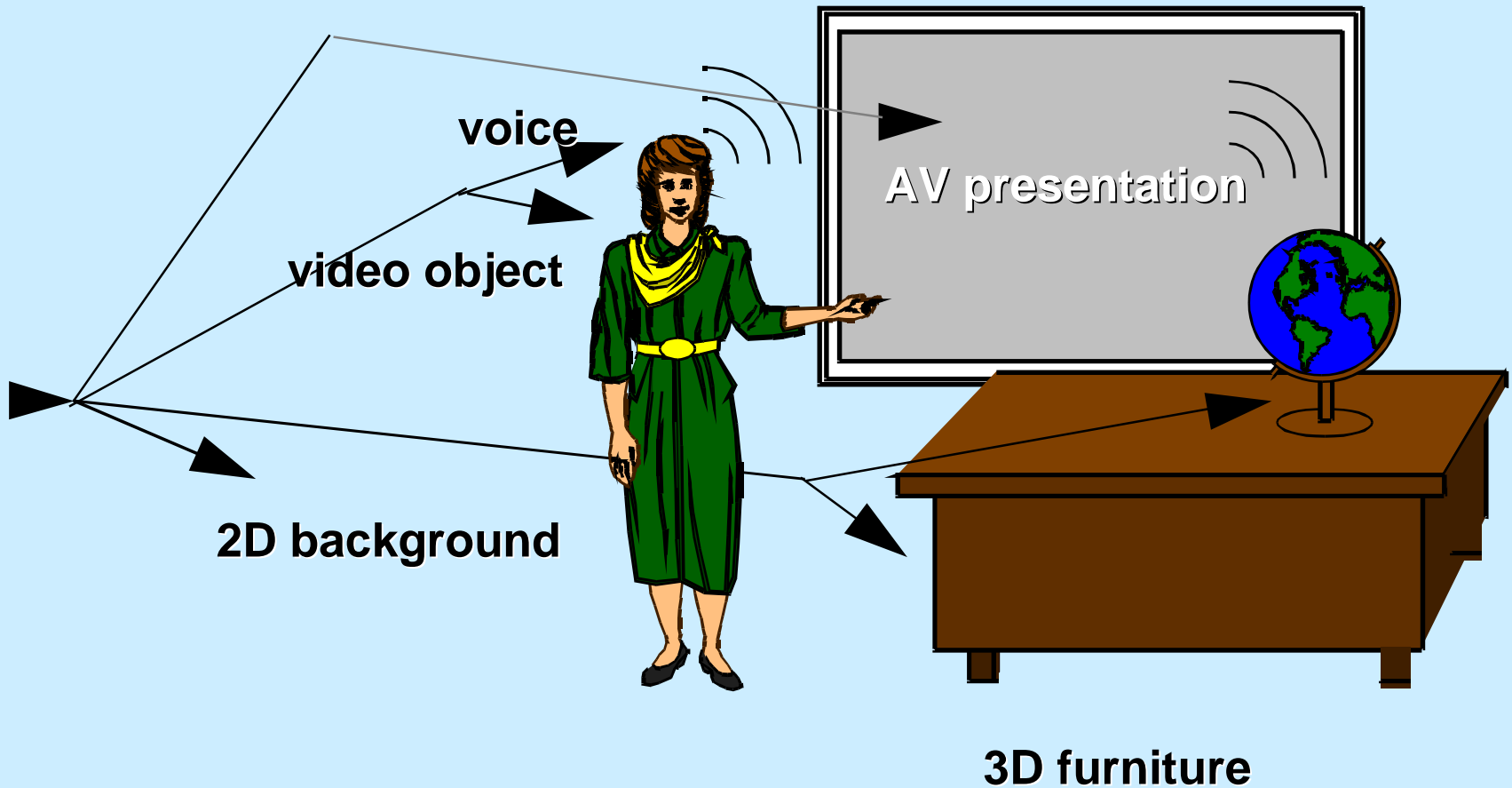


MPEG-4 Objects

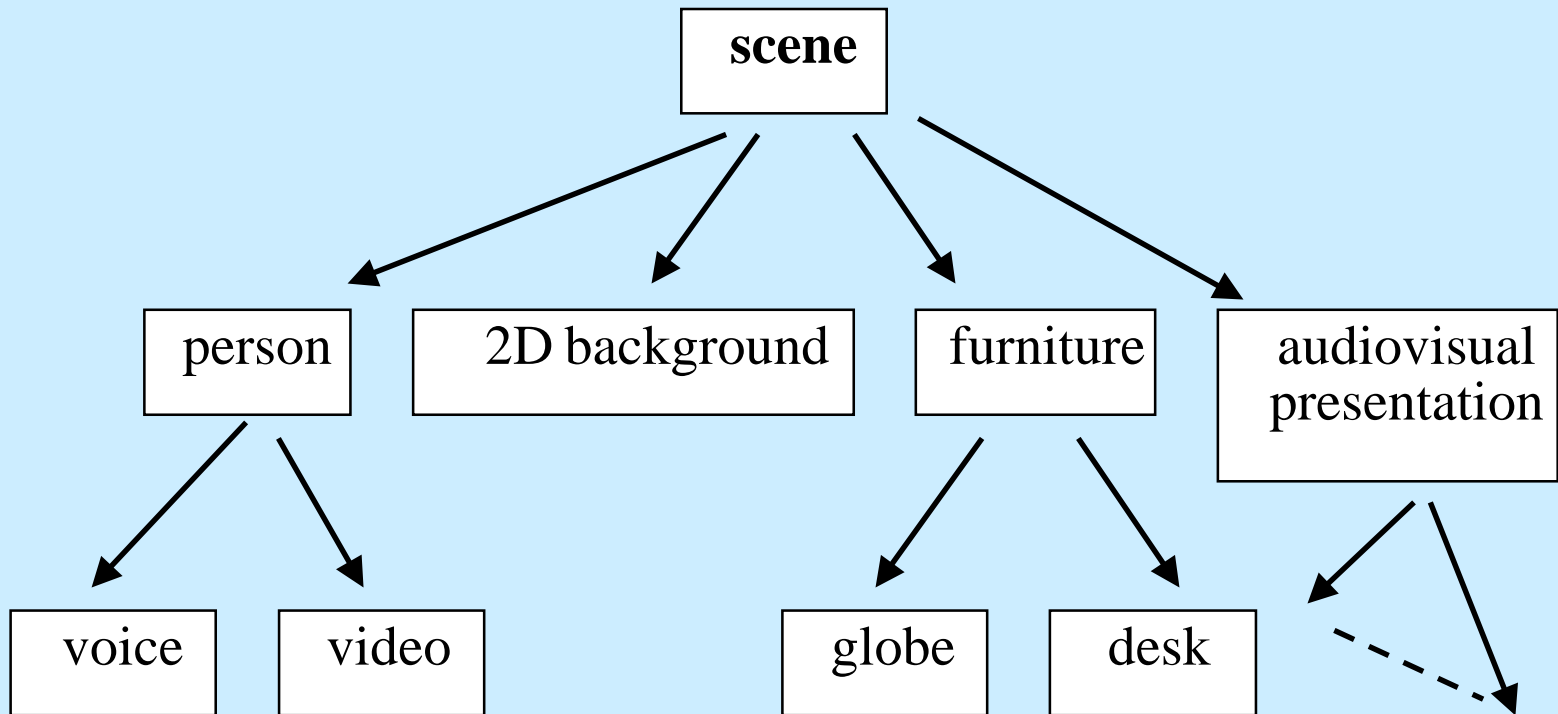
- Major topic for MPEG-4 from 1994
- Enables standards based interactive television to become a reality
- Essence: “Scenes” are composed of various “objects,” which are carried separately, and composited at the decode side. Since separate, interaction is enabled.
- Object technology moves the “studio” into the home!



Example: MPEG-4 Audiovisual Scene



MPEG-4 Scene Composition

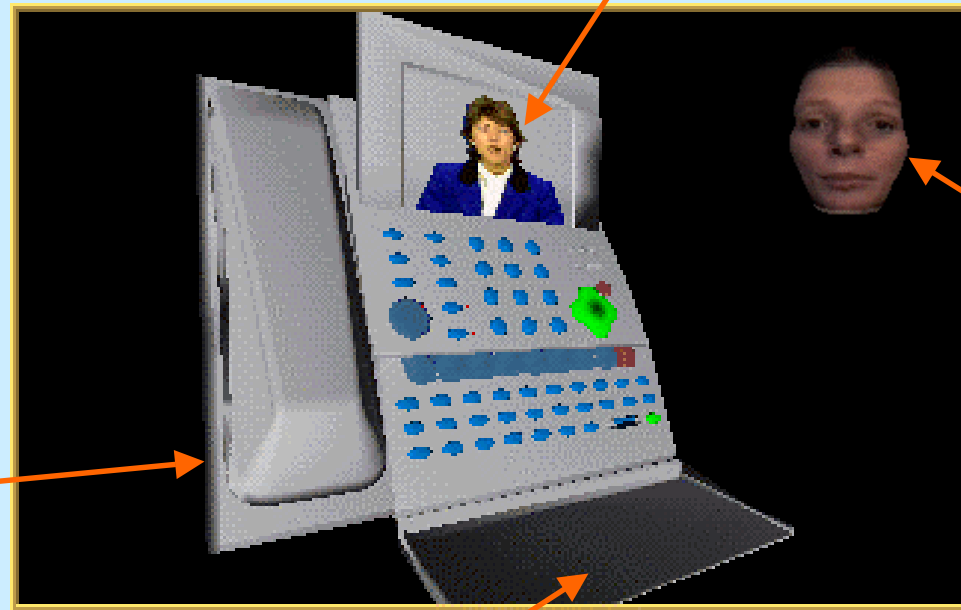


Example 3D Composition

MPEG-4 Video Object

3D face +
Spatialized
audio

3D object



Interactive
Behavior



MOTOROLA
Broadband Communications Sector

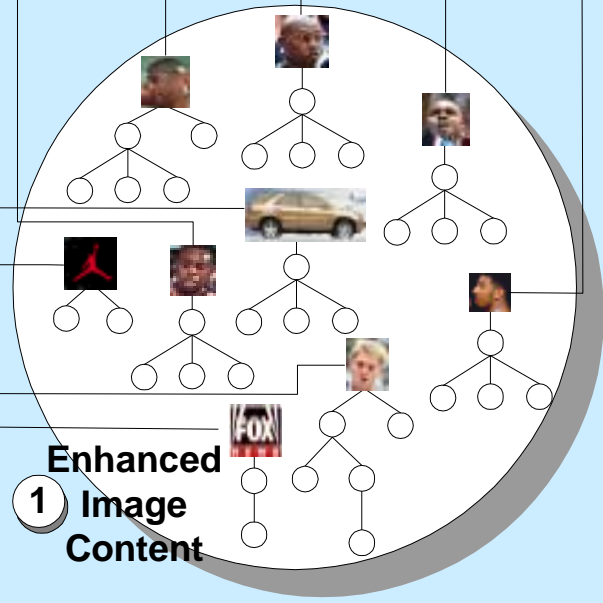
2

BIFS
Specification



Broadcast
MPEG-2
Video
(AC3 Audio)

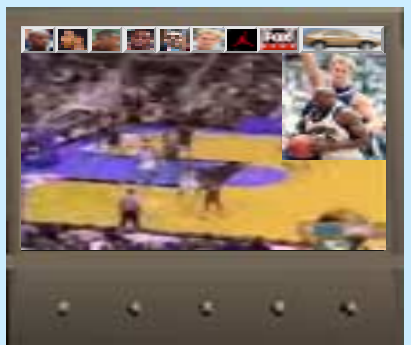
Enhanced
MPEG-4
Audio/Video



1
Enhanced
Image
Content

5
BIFS Scene Graph,
MPEG-2 Media Streams
MPEG-4 Media Streams

6



Interactive Television



MOTOROLA
Broadband Communications Sector

MPEG-4

Mixing Objects with Video

Player Stats



Checking on
you car

Information you
asked to be
informed about

Email Object

Mail: Motorola just announced ...

Incoming Call



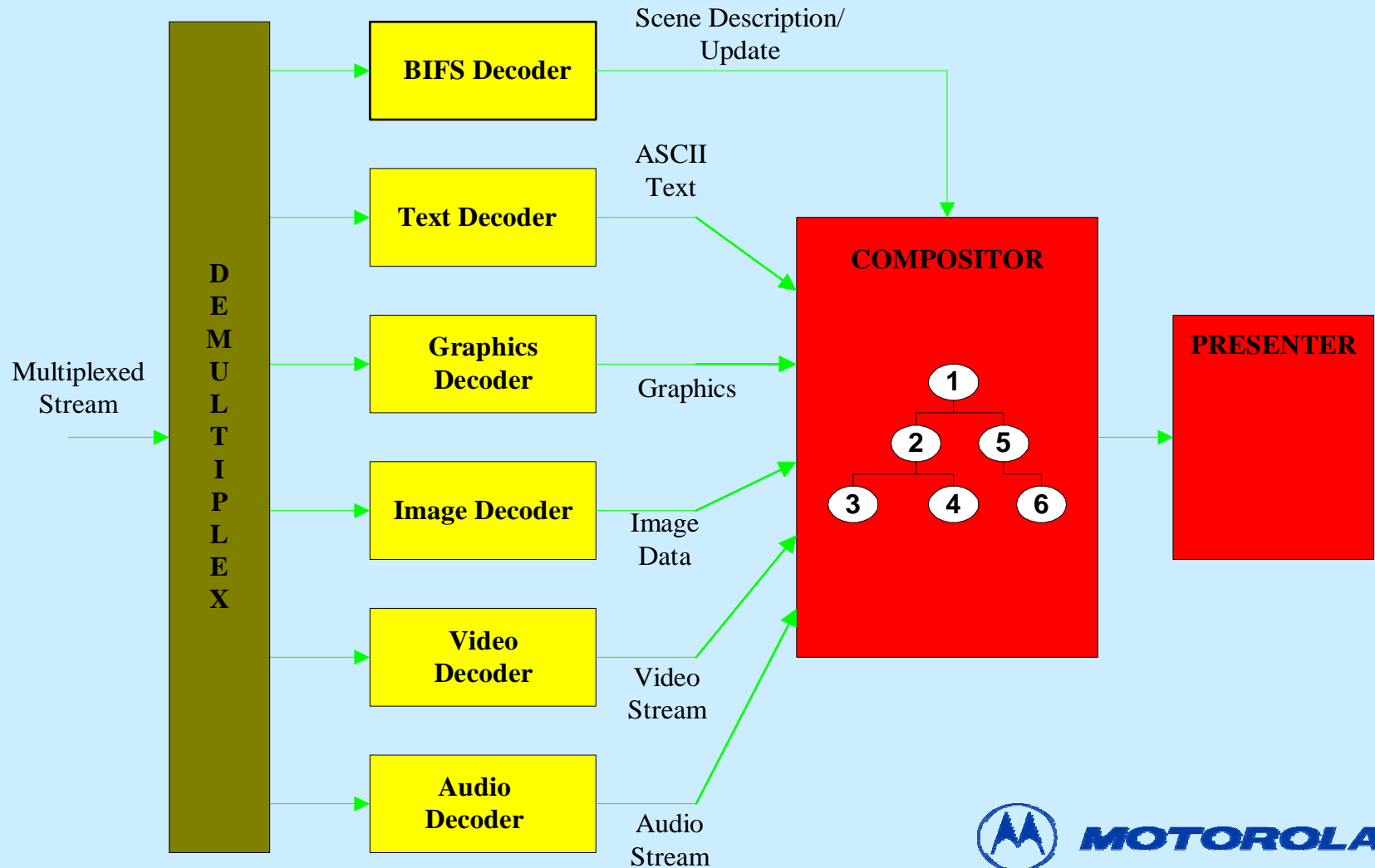
MOTOROLA
Broadband Communications Sector

MPEG-4 Object Flexibility

- BIFS describes scene composition for decoder
 - Based on VRML
- Objects can be of any type, including multiple of each type
- Objects can be coded separately, for example two video objects can have different frame rates, different quality, for optimal compression
- Video objects can have arbitrary shape & relationship
 - Shape coding
 - Alpha blending coding
 - Sprite coding



MPEG-4 Decoder Architecture



MPEG-4 Flexibility

- MP4 file format.
 - Derived from Quicktime format
- Transport flexibility
 - Sync layer [systems]
 - MPEG-4 in MPEG-2 transport
 - MPEG-4 in IP transport
- Scalability
 - Exceeds capability of MPEG-2
- Error resilience



What is MPEG-7?

- MPEG-7 targets multimedia description, so that users can search and retrieve more easily.
- Based upon various features of content:
 - Catalogue [title, creator, etc]
 - Semantic [who, what, when...]
 - Structural [histograms, pitch, ...]



MPEG-7 Does Not:

- Define how to extract features from content
- Describe the search engines, classifiers, and pattern recognition required to find and retrieve information.



MPEG-7 Terminology

- Descriptors: syntax and semantics to represent a feature
 - Coded format
- Description Scheme: defines relationships between descriptors, and/or other description schemes. Expressed in XML schema
- Over 100 complete or in development



Structural Video Features

- Dominant color or texture in a region of image
- Color histogram of image
- Thumbnails
- Shape or edge description [query by sketch]
 - Ex: trademark search
- Motion description
 - Ex: find similar motions in sporting events
 - Ex: video surveillance



Structural Audio Features

- Waveform and spectrum envelopes
 - Ex: search for similar voices
- Spoken content description
 - Words, phonemes, and transitions from automatic speech recognition technology
- Timbre descriptors
- Other examples:
 - Find melody independent of instrument
 - Find music matching portion you whistle or hum.



MOTOROLA

Broadband Communications Sector

Content Management Features

- Title, author, classification, purpose
 - Typically author generated
- Rights holder, access rules “IPMP”
 - May change over time
- Storage format, coding algorithm



What is MPEG-21?

- Began in 2000, still in infancy
- Vision to build “multimedia framework,” so that content can be created, delivered, and consumed, “across wide range of networks and devices in interoperable manner”
- MPEG-21 will fill in gaps where standards do not exist. Collaborative effort
- MPEG-21 thus takes the systems view, examining interfaces, and abstraction layers.
- Ex: Work on management & protection of MPEG-4 objects and MPEG-7 descriptors.
[DRM]



Conclusion

- MPEG continues its long tradition of developing relevant, forward looking standards for multimedia
- MPEG-4 offers flexible, efficient authoring and coding for large variety of media types, at wide range of data rates, supporting true standards based streaming and interactivity
- MPEG-7 brings in the descriptions that will enable whole new classes of multimedia applications
- MPEG-21 will fill in the gaps!

