MPEG-4

Wolfgang Leister
Knut Holmqvist

Today we'll talk about ...

- MPEG-4 / ISO/IEC 14496 ...
  - ... is more than a new audio-/video-codec
  - ... handles Objects – not pixels
  - ... offers interactivity beyond «fast-forward»

Parts of MPEG-4

- ISO/IEC 14496 has 10 parts:
  1) ***Systems***
  2) ***Visual***
  3) ***Audio***
  4) ***Conformance Testing***
  5) ***Reference Software***
  6) ***Delivery Multimedia Integration Framework (DMIF)***
  7) ***Optimized Software for MPEG-4 tools***
  8) ***4 on IP framework***
  9) ***Reference Hardware Description***
  10) ***Advanced Video Coding (AVC)***

MPEG-4 Overview (1)

- Represents aural, visua, and audio-visual content in «Media Objects» (MO)
- MO can be of natural or synthetic origin
- Describe composition of MO
- Multiplex and synchronize the data associated with MO for transport over network channels providing QoS
- Interact with the audiovisual scene generated at the receiver's end.

MPEG-4 Overview (2)

- Coded representation of MO
- Composition of MO
- Description and synchronization of streaming data for MO
- Delivery of streaming data
- Interaction with MO
- Management and Identification of Intellectual Property

Coded representation of MO

- Still images (e.g., fixed background)
- Video objects (e.g., talking person wo. background)
- Audio objects (e.g., voice, background noise)

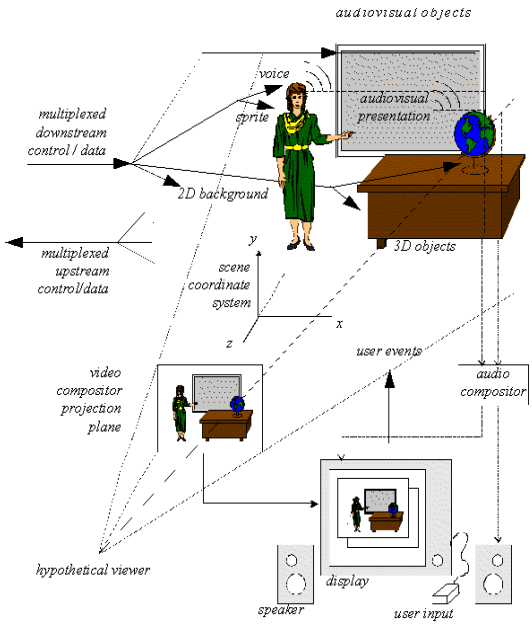
- Natural / synthetic; 2D / 3D
- Text and graphics, talking heads, synthetic sound
- MPEG-4 standardizes:
  - coded representation of a number of MO

Composition of MO

- Describe complex scenes
  - Place MO
  - Transforms to change geometrical / acoustical appearance of MO
  - Group primitive MO to compound MO
  - Apply streamed data to a MO (e.g., sound, moving texture, animation parameters for synthetic face)
  - Change interactively the user's viewing / listening points
  - VRML concepts!

MPEG-4 Scenes

- Organises MO in a hierarchical system
- Leafs are simple media objects like:
  - Still images
  - Video objekts
  - Audio objekts
  - Textures
  - 2D and 3D
  - etc.



Description and synchronization of streaming data for MO

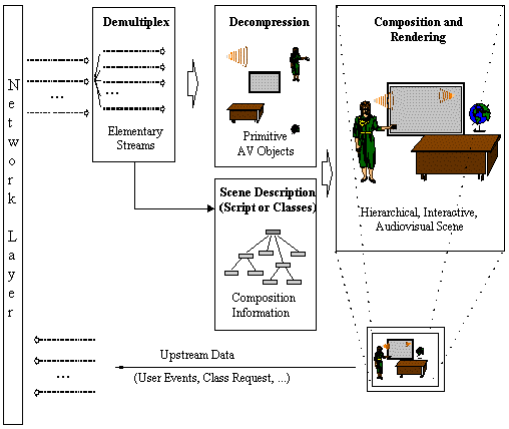
- Object descriptor identifies all streams associated to one MO
  - Hierarchically encoded data
  - Meta-information
  - Intellectual property rights
- Each stream characterized by descriptors
  - Decoder resources
  - Precision of encoded timing information
  - QoS requirements
- Synchronisation by time stamps

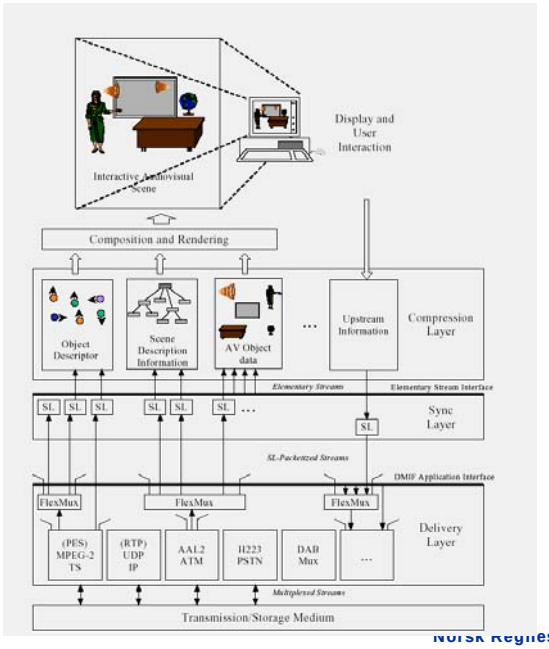
Interaction with MO

- Examples
  - Change viewing / listening point, navigation
  - Drag objects to different position
  - Trigger cascade of events by clicking on specified object
  - Select desired language when multiple tracks are available
  - Trigger more complex kinds of behavior

Major Functionalities in MPEG-4

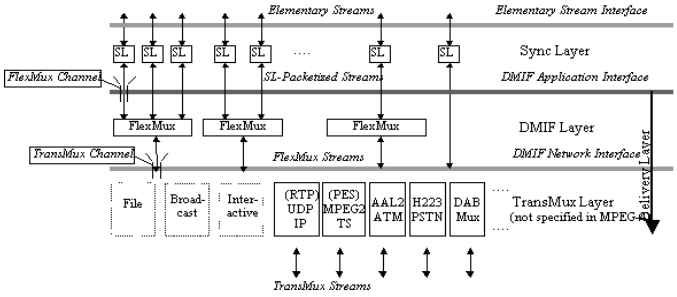
- Transport
  - MPEG-4 does not define transport layers
  - Adaptation defined for
    - Transport over MPEG-2 Transport Stream
    - Transport over IP
- DMIF
  - Transparent interface to interactive peer, broadcast, local storage, ...
  - Control of the establishment of FlexMux channels
  - Support for mobile networks


Network Layer
Demultiplex
Elementary Streams
Decompression
Primitive AV Objects
Scene Description (Script or Classes)
Composition Information
Composition and Rendering
Hierarchical, Interactive, Audiovisual Scene
Upstream Data
(User Events, Class Request, ...)


Display and User Interaction
Interactive Audiovisual Scene
Composition and Rendering
Object Descriptor
Scene Description Information
AV Object data
Upstream Information
Compression Layer
Elementary Streams
Elementary Stream Interface
SL
Sync Layer
SL-Packetized Streams
DMIF Application Interface
FlexMux
Delivery Layer
(PES) MPEG-2 TS
(RTP) UDP IP
AAL2 ATM
H223 PSTN
DAB Mux
Multiplexed Streams
Transmission/Storage Medium

Delivery of streaming data

- DMIF – Delivery Multimedia Integration Framework (part 6 of MPEG-4)
  - Grouping of ES
- TransMux Layer (only interface specified in MPEG-4)



Major Functionalities in MPEG-4

- Systems
  - Description of the relationship between audio-visual components that constitute a scene.
  - BIFS – Binary Format For Scenes
    - Describes spatio-temporal arrangements of MO
    - 2D – 3D – based on VRML
  - Object Descriptors (OD)
    - Describes relationship between ES pertinent to each object.
    - Additional information, e.g., URL

Major Functionalities in MPEG-4

- Systems
  - Standard file format, MP4
  - Interactivity, including client-server based interaction, event model, event handling and routing
  - Java (MPEG-J), «MPEGlets»
  - Interleaving of multiple streams to one single stream (FlexMuxTool)
  - Transport layer independence
  - ...

Major Functionalities in MPEG-4

- Audio
  - General Audio Signals
  - Speech Signals
  - Synthetic Audio
  - Synthesized SpeechScalable

Major Functionalities in MPEG-4

- Visual
  - Pixel based – synthetic scenes
    - 5 kbit/s – 1Gbit/s
    - Progressive, interlaced
    - Sub-QCIF – Studio resolution (4k x 4k)
  - Content based functionalities
  - Scalability of textures, images and video
  - Shape and alpha channel coding
  - Faces and body animation
  - 2D meshes, 3D polygonal meshes

MPEG-4 Extensions

- IPMP Extension
  - Identify intellectual property of MPEG-4 objects
  - Stores unique identifiers issued by e.g., ISAN, ISRC
  - Identify intellectual property by key-value pair
- Animation Framework eXtension, AFX
- Multi User Worlds
- Advanced Video Coding / H.264
- Audio extensions
  - Bandwidth extension
  - Parametric coding

Profiles in MPEG-4

- Visual Profiles

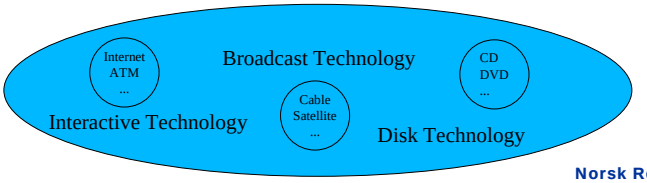
  The Simple Visual Profile, The Simple Scalable Visual Profile, The Core Visual Profile, The Main Visual Profile, The N-Bit Visual Profile, The Simple Facial Animation Visual Profile, The Scalable Texture Visual Profile, ..., The Core Studio Profile
- Audio Profiles
- Graphics Profiles
- Scene Graph Profiles
- MPEG-J Profiles
- Object Descriptor Profile

Transport of MPEG-4

- MPEG-4 on MPEG-2
  - MPEG-4 uses MPEG-2 TS (ISO/IEC 13818-1)
- MPEG-4 over IP
  - Developed jointly with IETF AVT working group
  - RTP payload specifications
  - Part 8 of MPEG-4 (ISO/IEC 14496-8)
  - MPEG-4 sessions over RTP, RTSP, and HTTP
  - Normative mapping functions to MPEG-4 SL packets

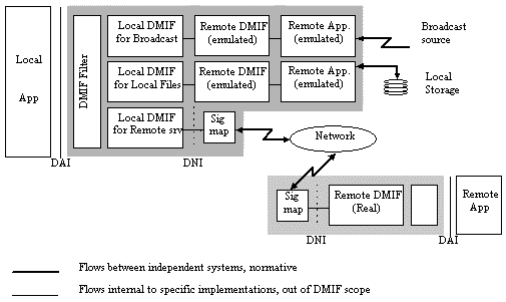
DMIF

- Delivery Multimedia Integration Framework
- Session protocol
- Returns pointers to where to get streamed data
- DMIF is both framework and protocol
  - DMIF-Application Interface (DAI)
  - QoS considered in DMIF design



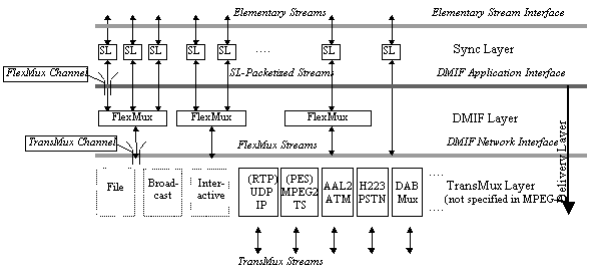
DMIF Communication Architecture

- Remote application vs. broadcast / local storage
- Application un-awareness (not entirely true in broadcast / local storage)
- DNI (DMIF Network Interface)



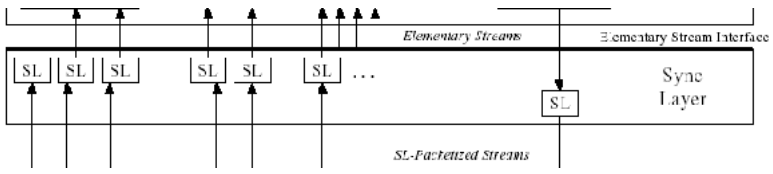
Transport Layer

- MPEG-4 does not define Transport Layer
- Uses existing Transport Protocols
- E.g., RTP, H223, ATM, MPEG-2, ...



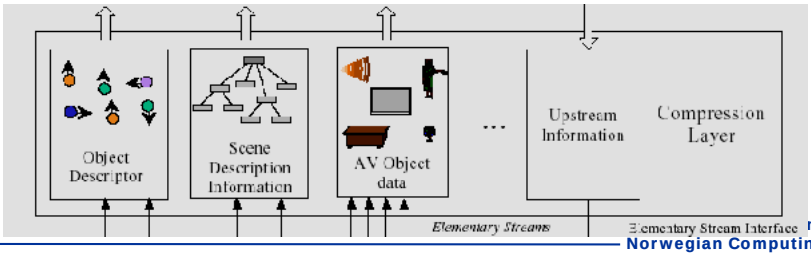
Sync Layer

- Synchronisation and buffering of compressed media
- DMIF Application Interface towards transport layer
- Independent of network



Media Layer

- Decoding (and encoding)
- Elementary Stream IF towards sync layer
- BIFS protokol
- Object descriptors

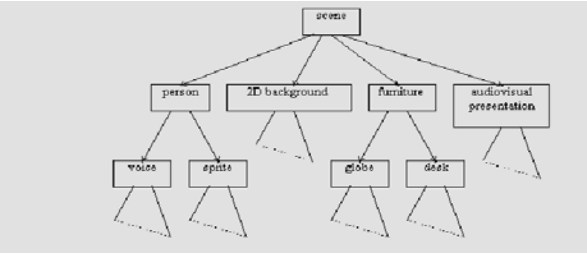


FlexTime Model

- Traditional MPEG-4 Timing model
  - Hard timestamps / reference clocks (as in MPEG-2)
- Advanced Synchronisation Model
  - CoStart, CoEnd, Meet
  - Flexible duration – spring metaphor
    - Minimum length
    - Maximum length
    - Optimal time
  - Compensate for delay, jitter
  - Synchronise nodes of unknown length
  - Synchronise BIFS updates
  - Re-adjust out-of-sync situations

Binary Format for Scene description (BIFS)

- Composition of objects in scenes
- Based on VRML
- Represented as a graph
- Nodes are MOs
- Dynamic tree-structure

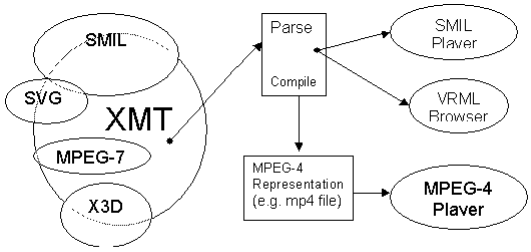


Advanced BIFS

- Advanced sound environment modeling
- Body animation
- Chroma keying
- Inclusion of hierarchical 3D meshes
- Associating interactive commands to media nodes (back channel)
- PROTOs and EXTERNPROTOs

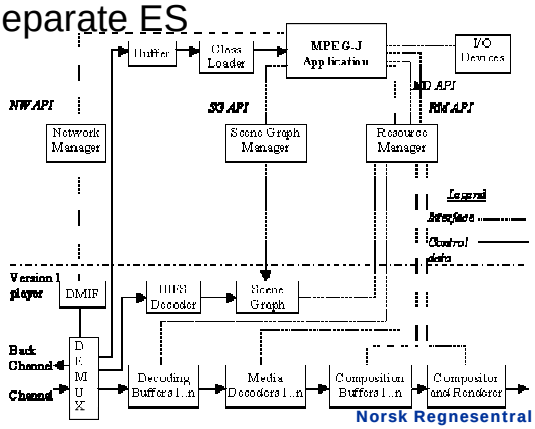
BIFS – Textual Format

- Extensible MPEG-4 Textual format (XMT)
- Interoperable with
  - X3D (by Web3D consortium)
  - SMIL (by W3C consortium)
- XMT-A: XML-based subset of X3D



MPEG-J

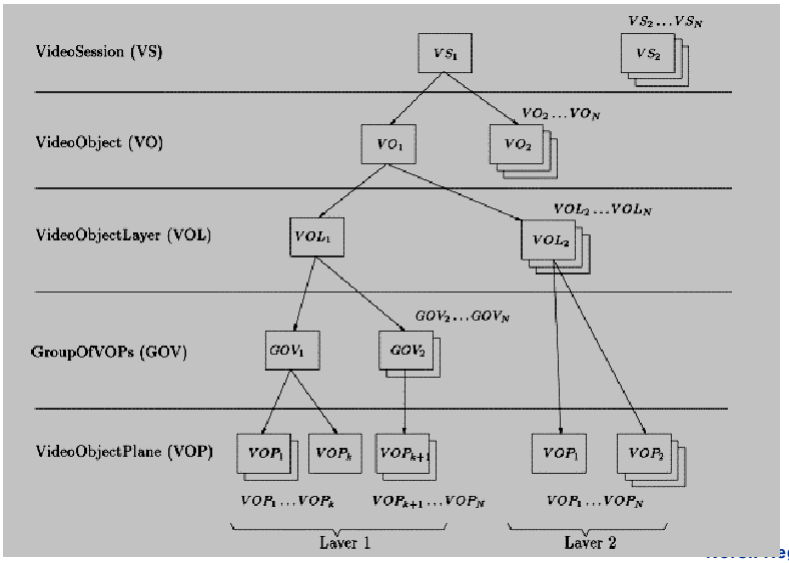
- MPEG-J: programmatic system of MPEG-4
- Parametric: Presentation Engine
- Java Application: separate ES



Natural Video Coding (NVC)

- Coding of silhouettes (shape)
- Motion estimation and -compensation
- Texture coding
- Error handling
- Coding of «sprites»
- Scalability

Structure (NVC)



NVC (details)

- Shape coding
  - Binary mode, described by binary mask
  - Gray value (corresponds alpha channel)
- Motion coding
  - Block size 16x16 or 8x8
  - I-VOP, P-VOP, B-VOP

Texture coding (NVC)

- YCbCr coding
- I-VOP full, second change
- «Standard» DCT and quantization
- Especially for blocks outside the shape

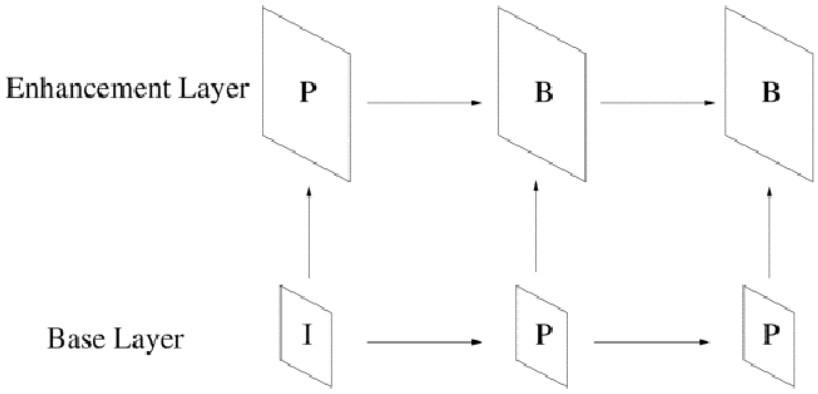
«Sprite» coding (NVC)

- Sprites are regions of a VO that are ther «all the time» (e.g., background, globe)
- Like for I-VOP
- Need to be sendt only once

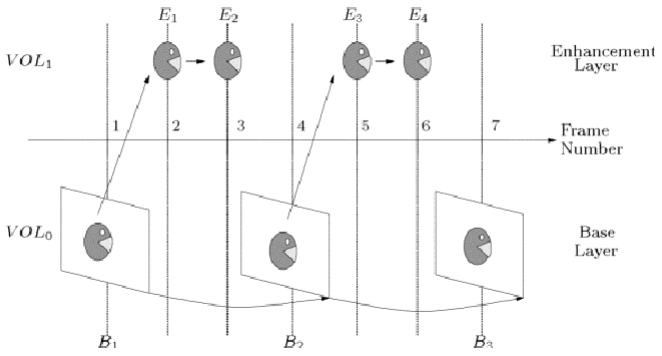
Scalability (NVC)

- Two or more VOL
- Depending on bandwidth more or less is sendt
- Both temporal and «spatial»

Scalability (NVC)



Scalability (NVC)



Audio

- Speech: 2 to 24 kbit/s
- Synthetic speech
- Lip synchronisation
- General audio
- Synthetic audio

Profiles

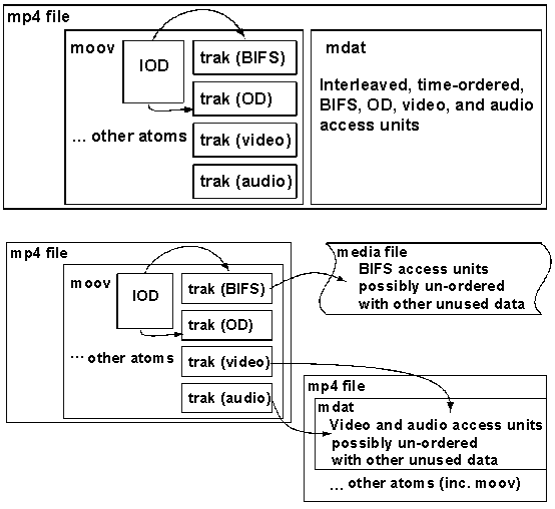
- Profile = part of the standard
- Limits what a decoder must implement
- Like in previous MPEG standards
- For audio (8), video (19), and graphics (4)

File format, MP4

- Based on QuickTime
- Consists of objects: «Atoms»
- Atoms describe a hierarchy of meta data
- Meta data can be included
  - Pointers to content outside file
  - Pointers to content inside file

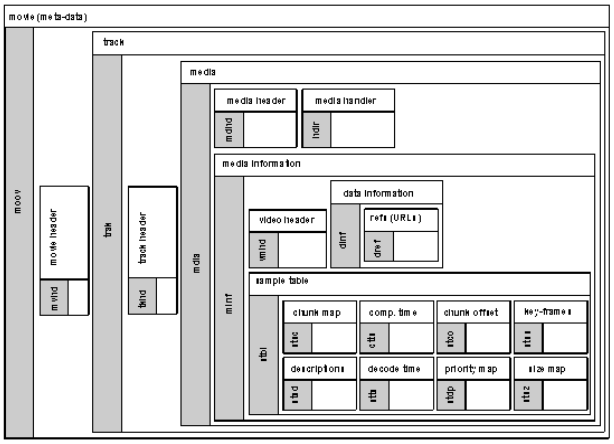
MPEG-4 File Format

- Atoms
  - Unique tag, length
  - Describe hierarchy of metadata with index points, durations, pointers to media data.
  - Movie atom contains collection of atoms



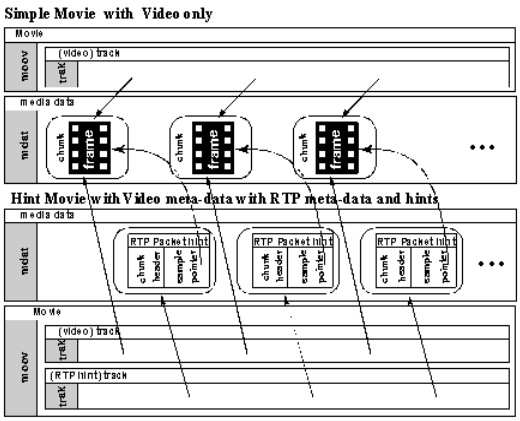
MPEG-4 File Format

- Relation between different objects / atoms



MPEG-4 File Format

- RTP protocol hint tracks to stream simple video movie
- Streaming format
- Streamable format



It is called MPEG-4 – however it isn't ...

- Microsoft MPEG 4 v1-v3 (does not support «QuickTime» file format)
- DivX;-) hack of the hack, Uses MS AVI file format

End of Lecture

Thank you for your attention!