From MPEG-1 to MPEG-2
- Improvement in quality
  - from VCR to TV to HDTV
- No CD-ROM based constraints
  - higher data rates
    - MPEG-1: about 1.5 Mbit/s
    - MPEG-2: 2-100 Mbit/s

Evolution
- 1994: International Standard
- Also later known as H.262
- Prominent role for digital TV in DVB (digital video broadcasting)
  - commercial MPEG-2 realizations available
MPEG-2 System

Steps
1. Audio and video combined to "Packetized Elementary Stream (PES)"
2. PES(es) combined to "Program Stream" or "Transport Stream"

Program stream:
- Error-free environment
- Packets of variable length
- One single stream with one timing reference

Transport stream:
- Designed for "noisy" (lossy) media channels
- Multiplex of various programs with one or more time bases
- Packets of 188 byte length

Conversion between Program and Transport Streams possible
11. MPEG-4

Goals

MPEG-4 (ISO 14496) originally:
- Targeted at systems with very scarce resources
- To support applications like
  - Mobile communication
  - Videophone and E-mail
- Max. data rates and dimensions (roughly):
  - Between 4800 and 64000 bits/s
  - 176 columns x 144 lines x 10 frames/s

Largely covered by H.263, therefore re-orientation:
- Goal to provide enhanced functionality
to allow for analysis and manipulation of image contents

MPEG-4: Schedule for Standardization
- 1993: Work started
- 1997: Committee Draft
- 1998: Final Committee Draft
- 1998: Draft International Standard
MPEG-4: Goals (cont.)

1: support composite multimedia i.e. find standardized ways to
   - Represent units of aural, visual or audiovisual content
     - "audio/visual objects" or AVOs

- object coding independent of other objects, surroundings and background
- Compose these objects together
  - i.e. creation of compound objects that form audiovisual scenes
- Multiplex and synchronize the data associated with AVOs
  - for transportation over network channels providing QoS (Quality-of-Service)

2: support synthetic objects
   - computer-gen. (VR), synthesized (txt2speech), model-based ("face")

3: support truly interactive applications (more than play/pause/rewind..)
   - Interact with the audiovisual scene generated at the decoder's site
MPEG-4: Scope

Definition of

- "System Decoder Model"
  - specification for decoder implementations
- Description language
  - binary syntax of an AVO's bitstream representation
  - scene description information
- Corresponding concepts, tools and algorithms, especially for
  - content-based compression of simple and compound audiovisual objects
  - manipulation of objects
  - transmission of objects
  - random access to objects
  - animation
  - scaling
  - error robustness
Targeted bit rates for video and audio:

- **VLBV core**
  - "Very Low Bit-rate Video"
  - 5 - 64 Kbit/s
  - image sequences with up to CIF resolution and up to 15 frames/s

- **Higher-quality video**
  - 64 Kbit/s - 4 Mbit/s
  - quality like digital TV

- **Natural audio coding**
  - 2 - 64 Kbit/s
MPEG-4: Video and Image Encoding

Encoding / decoding of
- Rectangular images and video
  - coding similar to MPEG-1/2
  - motion prediction
  - texture coding
- Images and video of arbitrary shape
  - as done in conventional approach
    - 8x8 DCT or shape-adaptive DCT
    - plus coding of shape and transparency information

Encoder
- Must generate timing information
  - speed of the encoder clock = time base
  - desired decoding times and/or expiration times
    - by using time stamps attached to the stream
- Can specify the minimum buffer resources needed for decoding
MPEG-4: Composition of Scenes

Scene description includes:
- Tree to define hierarchical relationships between objects
- Objects' positions in space and time
  - by converting the objects' local coordinate system into a global coordinate system
- Attribute value selection
  - e.g. pitch of sound, color, texture, animation parameters

Description based on some VRML concepts
- VRML = „Virtual Reality Modelling Language“

Interaction with scenes
- e.g. change viewing point, drag object, start/stop streams, select language
MPEG-4: Composition of Scenes

Scene description includes:
- Tree to define hierarchical relationships between objects

- Objects' positions in space and time
  - by converting the objects' local coordinate system into a global coordinate system
- Attribute value selection
  - e.g. pitch of sound, color, texture, animation parameters

Description based on some VRML concepts
- VRML = "Virtual Reality Modelling Language"

Interaction with scenes
- e.g. change viewing point, drag object, start/stop streams, select language
MPEG-4: Layered Networking Architecture

**Display / Recording**

**Codec**

**Coding / Decoding**
- e.g. video or audio frames
- or scene description commands

**Access Units**

**Adaptation Layer**

**Elementary Streams**
- + stream type info, sync. info, QoS req.,...

**Flexible Multiplexing**
- e.g. multiple elementary streams with similar QoS requirements

**Multiplexed Streams**

**TransMux Layer**

**Network or Local Storage**

**Codec**
For example, one can replace the object *Man* by object *Woman* from the library [scene (c)]. Concurrent playout of VO₀ and VO₁, followed by an object VO₃ at the time 𝑡₄. Here, VO₂ and AO₀ can act as the background audio-visual object throughout