MPEG-4: Overview

Multimedia
Naresuan University

Sources
- Chapters 1 and 2, The MPEG-4 Book, F. Pereira and T. Ebrahimi
- Some slides are adapted from NTNU, Odd Inge Hillestad.
MPEG-1 and MPEG-2

- **MPEG-1 (1992)**
  - Designed for video on CD-ROM at 1.2 Mbit/s and optimized for coding of non-interlaced (progressive) video
  - Supports basic interactivity with stored video stream such as **fast forward, reverse, random access**

- **MPEG-2 (1994)**
  - Optimized for interlaced video at **4-30** Mbit/s (TV quality Video **4-10** Mbit/s, HDTV **15-30** Mbit/s)
  - Compatible with MPEG-1
Interlaced Video:

Video programs using the NTSC, PAL, and SECAM standards are interlaced: Each frame consists of two fields displayed in two passes.

A TV displays the first field of alternating lines over the entire screen, and then displays the second field to fill in the alternating gaps left by the first field.

Non-Interlaced Video—Most personal computers display using progressive scan, in which all lines in a frame are displayed in one pass from top to bottom before the next frame appears.
What is MPEG-4?

  - International standard for efficient representation and communication of interactive multimedia presentations.
  - Represents individual *media objects* in an audiovisual scene.
    - Video, audio, pictures, 2D/3D synthetic graphics, text.
  - Main features:
    - Content-based interactivity.
    - Object-based coding and scalability.
    - Highly compressed, synchronized and error resilient bitstreams.
MPEG-4 Version

- **Version 1**
  - May ‘99: International Standard

- **Version 2**
  - adds new ‘Profiles’, with new tools and functionalities (V2 = V1+ new tools + new profiles)

- **Version 3**
  - Studio profile

- **Version 4**
  - Streamed video profile
The Power of Examples

Envivio

- An end-to-end provider of MPEG-4 solutions

- Functionality
- Interactivity
- Coding Efficiency (AVC @ 800 kbps)
Generation Video Coding

- Coding of rectangular video frames
  - MPEG-1, MPEG-2, H.263
  - Motion-compensated block-based DCT coding

- Advantages
  - Simple matrix representation
  - Easy to display decoded video
  - Easy to capture
  - Low complexity in encoder/decoder

- Drawbacks
  - Not appropriate for interaction with visual elements
  - Differs from human visual system mechanisms
  - Limited potential
Generation Video Coding

- The *interactive audiovisual scene*
  - What do we *see*? Visual media objects
  - What do we *hear*? Audio objects
  - What can we *do*? Interactive features in the scene

Message extraction: Identifying objects
Message coding: Coding object features
Generation Video Coding

- The audiovisual information is now represented as a set of objects called media objects.
- Each media object now represents the message to be coded.
  - An object is described by a set of properties/frames
  - Example-video object: shape+texture+motion

- Advantages
  - Easier to interact with objects in the audiovisual scene.
  - Similar to human visual system mechanism
  - Big potential in various applications.

- Drawbacks
  - The object segmentation/tracking problem
  - Needs an interface to conventional displays (composition)
  - Higher complexity
Object-based Video Coding

- Each Video Object in a Scene is Coded and Transmitted Separately
Relation between MPEG standards

- Data representation pyramid

- Objects features extraction
- Objects Formation and tracking
- Semantic based representation
- Object based representation
- Pixel-based representation

- MPEG-1
- MPEG-2
- MPEG-4
- MPEG-7
MPEG-4 New Functionalities

- Streaming AV over mobile networks as MPEG-4 provides excellent error resilience capabilities and coding at low bitrates—Robust transmission in error-prone environments (Wireless systems)
- More freedom to flexibly interact with what is within scenes.
- Support integration of natural and synthetic AV media.
- Identification, Protection of Intellectual Property and Rights on content.
- MPEG-4 provides High Coding Efficiency and Coding of Multiple Concurrent Data Stream
MPEG-4 New Functionalities Cont..

- Coding of arbitrarily shaped objects
- Efficient compression of video sequences and still images over a wide range of bit rates
  - 5 kbits/s-50 Mb/s
  - Resolution: Small -TV
  - Progressive/Interlace
- Scalability (Spatial, temporal and SNR Fine Granularity Scalability)
- The MPEG-4 Visual Texture Coding provides an efficient tool for the compression of still image
ISO/IEC 14496 - MPEG-4 (1)

- Part 1: MPEG-4 Systems
  - Tools to represent a dynamic audiovisual scene, interactivity, and the synchronization of different media objects.
  - Systems Decoder Model, PM4 file format, MPEG-J, XMT,...
- Part 2: MPEG-4 Visual
  - Natural and synthetic audio compression tools.
- Part 3: MPEG-4 Audio
  - Natural and synthetic audio compression tools
- Part 4: Conformance
  - Specifies how tests can be designed to verify that the encoded bitstreams and decoders are compliant with ISO/IEC 14496.
- Part 5: Reference Software
  - “Im1” for systems, reference codecs for Visual & Audio
Part 6: DMIF
- Delivery Multimedia Integration Framework
- Specifies an application interface (DIA) for access to MPEG-4 content.
- Abstracts the underlying network and transport protocols for an MPEG-4 application.

Part 7: Optimized Reference Software

Part 8: MPEG-4 on IP Framework
- Specifies how MPEG-4 content is delivered over IP networks.

Part 9: Reference Hardware
- Reference hardware implementation in VHDL.
ISO/IEC 14496 - MPEG-4 (3)

- Part 10: Advanced Video Coding (AVC)
  - State of the art video compression (finished 2003)
  - A joint effort between MPEG and ITU-T (H.264)
Profiles @ Levels

- MPEG-4 can be seen as a toolbox with 6 dimensions:
  - Visual, audio, graphics, scene graph, ODs, MPEG-J.
- Profiles define subsets of tools for each dimension
  - Gives the functionality for an MPEG-4 application
- Levels limit the complexity of the tools at a given profile
Example of Profile and Level

- **Visual Profile:**
  - The *simple* profile accepts only objects of type *Simple* and was created with low-complexity applications in mind.

- **Usage:**
  - Mobile (audio) visual services,
  - Very-low-complexity video on the Internet,
  - Small camera devices recording moving video to disk or memory chips.

- **Four levels for the *Simple* profiles with bit rates from 64 kbit/s in Level 0 to 384 kbit/s in Level 3.**
MPEG-4 ecosystem

Creation Contribution
- video
- audio
- interaction
- 2D images
- animation
- 3D images
- BITS
- text

Production

Encode

Playout Storage
- Packaged
- MPEG-4 server
- IPMP DRM

Distribution
- terrestrial
- satellite
- wireless
- Cable
- Recall

Consumption
- TV
- PC
- DVD
- CellPhone PDA
- Game Console
- Radio
An End-to-End Walkthrough
MPEG-4 Systems Overview

- An audiovisual scene is represented by:
  - Object Descriptors (ODs)
    - Contains information about the media objects (metadata)
    - Used as an interface to the media objects
  - Scene Description
    - Defines the spatial position and temporal behavior of the objects in the audiovisual scene. *(Scene Graph)*
    - Defines the *interactive* possibilities in the scene.
    - BIFS (BInary Format for Streams)
  - Audiovisual object data
    - The coded representation of the media objects
MPEG-4 Terminal Architecture

- An interactive audiovisual scene is represented by:
  - Object Descriptors
  - Scene Description
  - Audiovisual Object Data
- OD, BIFS, AV data are all conveyed in Elementary Streams (ES).
- A sync layer embraces ESs to enable synchronization.
  - Time stamps
  - All ESs can be synchronized
- FlexMux
  - Simple multiplexing tool
  - Used when network or transport protocol lacks multiplexing facilities.
MPEG-4 Architecture

- **MPEG-4 Systems**
  - Provides synchronization and multiplexing of Elementary Streams, BIFS and interactivity tools, MP4

- **MPEG-4 Visual and MPEG-4 Audio**
  - Provides compression tools

- **DMIF Application Interface**
  - Provides transparent access and delivery of content.
Access to an MPEG-4 scene
Alternative media representations

- Scene Description
  - Hierarchic build-up
  - Nodes and Fields
  - The media objects are referenced with OD_IDs

- Object descriptor
  - Contains Elementary Stream Descriptors
  - Media object data is referenced with ES_IDs

- Alternative representation
  - One media object can have more than one ES for its representation!
Access to scalable MPEG-4 content
The tools of MPEG-4 Systems

- Object Descriptor Framework
- BIFS, scene representation
- XMT
  - A textural representation format for BIRS and ODs
  - Based on XML
  - XMT-O for high-level authoring (~SMIL)
  - XMT-A for low-level conformity with BIFS (~X3D)
- MP4 file format
  - A generic exchange format for MPEG-4 presentations.
- FlexMux—a simple multiplexing tool
  - Can be used over the transport layer if the underlying protocols lack the necessary multiplexing facilities.
  - Low overhead and low delay
The tools of MPEG-4 Systems (2)

- **Sync Layer (SL)**
  - For synchronization of Elementary Streams
  - A flexible, configurable encapsulation tool
  - ESs is a stream of Access Units (Aus)
  - SL stamps Aus with time codes (DTS & CTS)

- **FlexTime**
  - A tool for synchronizing ESs from sources with different time bases >> We can sync media from different locations!

- **MPEG-J**
  - Java API to control an MPEG-4 player and the audiovisual scene. Change the scene interactively based on:
    - Terminal capabilities, User preferences, Network monitoring
MPEG-4 Visual

- MPEG-4 Visual natural coding tools
- MPEG-4 Visual SNHC—synthetic coding tools
  - FBA—Face and Body Animation
  - 2D mesh coding
  - 3D mesh coding
  - AFX—Animation Framework eXtension
MPEG-4 Audio

- Natural speech coding
  - CELF (~6-12 kbps, 10-24 kbps WB)
  - HVXC (~1.2 to 4 kbps)

- MPEG-4 audio coding
  - AAC (Advanced Audio Coding) ~30-50% better than MP3
  - AAC-LD (Low Delay ~ 20 ms)
  - Lossess Audio Coding (ongoing work..)

- SNHC Audio
  - Synthetic speech and audio
Evolution of MPEG-4 Video compression

Bitrate

1998  2000  2002

Simple Profile  Advanced Simple  MPEG-4 AVC/H.26L

MPEG-4 progress
The position of MPEG-4
MPEG-4 Applications

3D Shopping Portal: the user enters a 3D virtual shop; navigates it; examines products by a video animation or modeled in 3D.
MPEG-4 Applications (2): Interactive Broadcasting

(a) A news program enhanced with synchronized graphics. The user can interact with the graphics and navigate in the news menu to have access to background information.

(b) The graphics are used in a sports event to display enhanced data about a soccer match: real-time tracking and highlighting of player or team.
MPEG-4 Applications (3): Interactive Broadcasting

The application exploits MPEG-4 shape-coded video along with streamed graphics to show personalized advertisement during a sport event.
The terminal could be a multimedia PC (equipped with camera, microphone, and speakers). The main idea here is that of a shared communication space.
MPEG-4 Applications (4): Multimedia Conferencing

The terminal could be a multimedia PC (equipped with camera, microphone, and speakers). The main idea here is that of a shared communication space.
MPEG-4 Applications (4): Multimedia Conferencing

The terminal could be a multimedia PC (equipped with camera, microphone, and speakers). The main idea here is that of a shared communication space.