Mid-term Exam

• Take-home exam
  – Will assign it on October 29 (Wednesday) through email (please watch your UMD email at 5:00 pm on Wednesday, no excuse for the email)
  – Start to work on your exam at 5:00 pm on Wednesday (Programing questions; you do not need to come to the classroom, no class on Wednesday and the class is used for the take-home exam)
  – Due at 5:00 pm Nov 6 (Wednesday). Submit your answer sheet before the class.
  – Midterm Project presentation on Nov 10 (Monday)
  – Individual work is required. Do not copy.
MPEG-4: Overview

- Flexible Multimedia Communications
- 5kbps - 50Mbps
- Audio object compression
- Video object compression
- Synthetic Audio/Speech and Video
- Systems: multiplexing and flexible composition

Slide courtesy of Hung Nguyen
Functionalities beyond MPEG-1/2

- Interaction with individual objects: The displayed scene can be composed by the receiver from coded objects
- Scalability of contents
- Error resilience
- Coding of both natural and synthetic audio and video

Slide courtesy of Hung Nguyen
MPEG-4 : System

- No need to know
  - What the terminal device
  - How to receive the stream

- Sync Layer
  - Object Descriptors: Relate to elementary streams to a media object (MO)
  - Synchronization Layer has two types of timing information
    - Convey the speed of the encoder to the decoder
    - The time stamp to portions of the encoded Audio Visual data

Slide courtesy of Hung Nguyen
Overview of MPEG-4 System

Scene segmentation and depth layering

Layered encoding

contour motion texture  
contour motion texture  
contour motion texture

bitstream layer 1
bitstream layer 2
bitstream layer 3

Separate decoding

AV-objects

Slide courtesy of Hung Nguyen
MPEG-4: System

- Delivery of streaming data

Slide courtesy of Hung Nguyen
MPEG-4

- **Main application domain:**
  - Audio and video associated with interactive multimedia applications over the internet
  - Other various types of entertainment networks
- **Standard contains features to enable the user**
  - Passively access a video sequence e.g. start/stop/pause commands
  - Also access and manipulate the individual elements that make up each scene within the video
Differences Between MPEG-4 and Other Standards

- MPEG-4 has a number of content-based functionalities
- Prior to compression the scene is defined in the form of a background and one or more foreground audio-visual objects (AVOs)
- Each AVO is defined in terms of one or more video objects and/or audio objects
  - Stationary car in a scene may be defined using a single video object
  - Person who is talking may be defined by both a video and an audio object
  - Each video and audio object may be comprised of many video and audio sub-objects
    - Head with moving lips
Object Descriptors

- Each audio and video object has a separate object descriptor
  - Allows the viewer to manipulate the object prior to its being decoded and played out
- The language used to describe and modify objects is called the **binary format for scenes (BIFS)**
  - Contains commands for:
    - Object deletion
    - Modification of the video object shape, color etc.
    - Animation of the video object in real time
    - Similar commands for audio objects
- Multiple versions of the same AVO
  - A base version and others at various levels of enhancement detail
  - This type of compression is called scaling and allows an AVO to be played out at a rate and resolution matching those of the viewing terminal
Content-based Video Coding Principles

Original frame/scene

VOP = video object plane

VOP 0

VOP 1

VOP 2
Example of Editing, Reusability and Compositing of Video Objects

Slide courtesy of Hung Nguyen
Example of Editing, Reusability and Compositing of Video Objects

Slide courtesy of Hung Nguyen
Example of Scalability of Video Objects and Optimizing Regions of Interest

Slide courtesy of Hung Nguyen
Example of Scalability of Video Objects and Optimizing Regions of Interest

Slide courtesy of Hung Nguyen
Example of Creativity and Interactivity with Video Objects

Slide courtesy of Hung Nguyen
Example of 2-D Interactivity

Slide courtesy of Hung Nguyen
Example of Creating a 3-D illusion from a 2-D Presentation

Slide courtesy of Hung Nguyen
Example of Interactive Virtual Reality

Slide courtesy of Hung Nguyen
MPEG-4: Visual Objects

Classification of Video Coding Tools

- High bit rate tools (interlace)
  1. High spatial resolution
  2. High frame rate
  3. Broadcast TV
  4. High coding efficiency

- VLBV core
  1. Low spatial resolution
  2. Low frame rate
  3. Conventional rectangular size
  4. High coding efficiency

- Content-based functionalities (shape, scalability)
  1. Separate coding of content
  2. Content-based interactive
  3. Object manipulation
  4. Hybrid coding of natural as well as synthetic objects

Slide courtesy of Hung Nguyen
MPEG-4: Visual Objects

VLBC Core and the Generic MPEG-4 Coder

Motion (MV)  Texture (DCT)

MPEG-4 VLBV Core Coder (similar to H.263/MPEG-1)

Shape  Motion (MV)  Texture (DCT)

Generic MPEG-4 Coder

Slide courtesy of Hung Nguyen
MPEG-4: Visual Objects

- Natural Video
  - Conventional and Content-Based Functionality

Slide courtesy of Hung Nguyen
MPEG-4: Synthetic Visual Objects

- Parametric descriptions of
  - A synthetic description of human face and body (body in Version 2)
  - Animation streams of the face and body (body in Version 2)
- Static and Dynamic Mesh Coding with texture mapping
- Texture Coding for View Dependent application
MPEG-4: Synthetic Visual Objects – Facial Animation

- Facial Definition Parameter
- Facial Animation Parameter
- Facial animation needs
  - The Facial Description Parameter (FDP) in Binary Format Scene
  - The Face Animation Table within FDPs
    - Example: FAP could say ‘open_jaw’
  - The Face Interpolation Technique in BIFS
    - Example: Missing FAP
MPEG-4: Synthetic Visual Objects

- Parametric descriptions of
  - A synthetic description of human face and body (body in Version 2)
  - Animation streams of the face and body (body in Version 2)

- Static and Dynamic Mesh Coding with texture mapping

- Texture Coding for View Dependent application
MPEG-4: Synthetic Visual Objects – Facial Animation

- Facial Definition Parameter
- Facial Animation Parameter
- Facial animation needs
  - The Facial Description Parameter (FDP) in Binary Format Scene
  - The Face Animation Table within FDPs
    - Example: FAP could say ‘open_jaw’
  - The Face Interpolation Technique in BIFS
    - Example: Missing FAP
MPEG-4: Synthetic Visual Objects – Animated Meshes

- 2D animated meshes
  - Only triangular meshes
- View dependent scalability
  - Taking into account the viewing position in the 3D virtual world
  - This object is sent and computed both at the encoder and the decoder side

Slide courtesy of Hung Nguyen
Example of MPEG-4 Visual Objects

Slide courtesy of Hung Nguyen
MPEG-4: Scene Description

- How objects are grouped together and positioned in space & time
- Attribute Value Selection
- Other transforms on media objects

Sprite: region of a VO present in scene throughout a video segment
MPEG-4: Intellectual Property Rights (IPR)

- IPR generally bestows on its owners the right to exclude others (with certain limited exceptions) from the use or re-use of their intellectual property without a license from the IPR owner.
- Any intellectual property which is delivered, either freely or for commercial gain, through an MPEG-4 application is afforded such protection.
- More information on MPEG-4 and IPR on http://www.chiariglione.org/mpeg/faq/mp4-sys/sys-faq-ipmp.htm#IPMP-What-is-IP

Slide courtesy of Hung Nguyen
Audio and Video Compression

- Audio, depending on the available bit rate
  - G723.1 CELP for interactive multimedia applications
  - Dolby AC-3 or MPEG Layer 2 for interactive TV over entertainment networks

- Video
  - Each VOP is identified and defined and then encoded separately (a difficult image processing task)
  - Each resulting object is bounded by a rectangle
  - Each is then encoded based on its shape, motion and texture
  - Since VOPs that move often occupy only a small portion of the scene, much higher compression ratios over other standards can be obtained
MPEG-4 VOP Encoder Schematic
Transmission Format

- All information relating to a frame/scene is transmitted over a network in the form of a transport stream (TS)
  - Consists of a multiplexed stream of packetized elementary streams (PES)
- The compressed audio and video information relating to each AVO in the scene is called and elementary stream (ES)
  - Carried in the payload field of a PES packet
  - Each PES packet also contains a type field used by the FlexMux layer to id and route the PES packet to the related synch block in the synchronization layer
  - The elementary stream descriptor (ESD) is used by the synchronization layer to route each ES to its related decoder
  - Each PES packet also contains time stamps to ensure each packet is passed to the decoder at the correct time instants
- Composition and rendering block takes the decompressed data relating to each AVO together with the scene descriptor information and composes each frame for output to the display screen
MPEG-4 Decoder Schematic
MPEG-4: Applications

- Scalable & Interactive Application on the WWW
- “Configurable TV”: choose the elements
  - subtitles (+language)
  - stock values
  - inset of soccer match
- Animated talking head with speech synthesis
- Interactive DVD applications
Interactive content can include anything that enhances the viewer's appreciation of TV— for example, local news headlines, sports scores, program lineups, links to Web sites, advertiser logos, actor/actress profiles, and plot summaries. Interactive TV content is limited only by the imagination.