ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Multimedia Technology

Section # 14: Video coding: MPEG-4

Instructor: George Xylomenos

Department: Informatics

Contents

- The MPEG-4 standard
- Scenes in MPEG-4
- Natural object coding
- Synthetic object coding
- Textures and sprites
- Error recovery and evaluation

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

The MPEG-4 standard

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics

What is MPEG-4? (1 of 2)

- The MPEG-4 standard
  - Goal: reduce (not increase!) bit rates
  - 10x improvement over H.261
- Specific MPEG-4 goals
  - Allow interaction with content
  - Downloading of new algorithms/applications
  - Support scalable coding
  - Independence from underlying network

What is MPEG-4? (2 of 2)

- Differentiation from MPEG-1/2
  - Emphasis on content structure
    - Each frame is a scene consisting of objects
  - Interaction with users
    - Modification of content structure
- Two types of coding
  - Audio-visual object coding
  - Scene coding

MPEG-4 and H.264

- MPEG-4 part 2 (1999, this section)
  - Original video coding standard
  - Emphasis on scene structure
- MPEG-4 part 10 (2003, next section)
  - Totally different from part 2!
  - Same standard as H.264/AVC
  - Emphasis on coding efficiency
  - Not backwards compatible with part 2

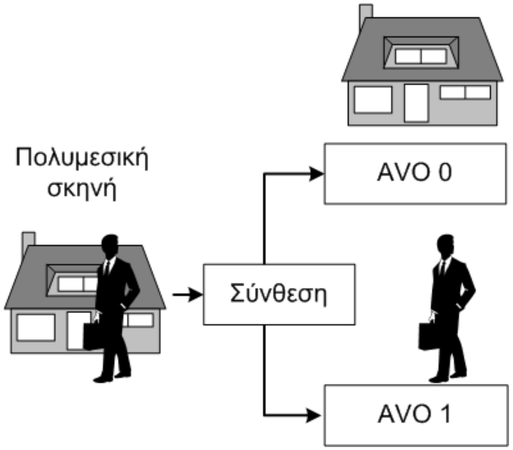
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Scenes in MPEG-4

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

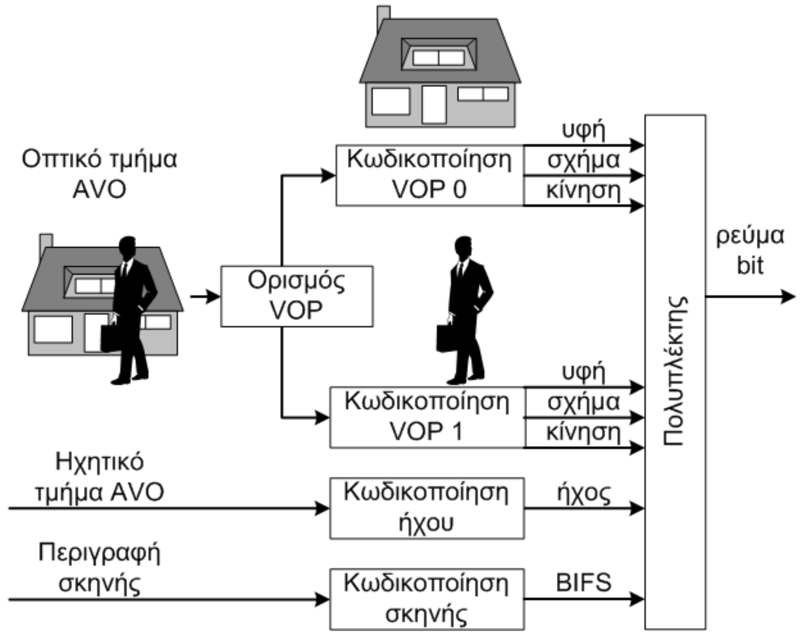
Instructor: George Xylomenos, **Department:** Informatics

Scenes (1 of 3)



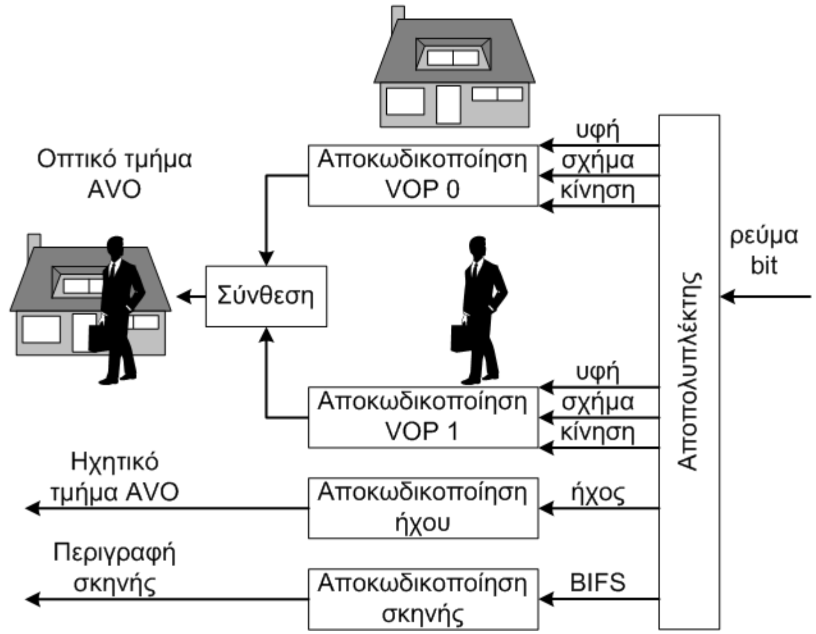
- Scene: a sequence of frames
  - Audio-Visual Objects (AVO)
  - AVO composition using VRML-like language
    - Binary format for scenes (BIFS)
- Video Object Plane (VOP)
  - Representation of visual AVO: shape, texture, motion

Scenes (2 of 3)



- Multiplexer (MUX)
  - Multiplexes all the info into a single bit stream
  - Position, depth and orientation of AVOs
  - Audio and scene description

Scenes (3 of 3)



- Demultiplexer (DEMUX)
  - Recovers AVO shape, texture and motion from stream
  - Decodes AVOs and composes the scene
  - Optional: scene modification (via BIFS)

Scene composition (1 of 2)

- Scene composed from object hierarchy
  - First, the hierarchy is transmitted
  - Then, we can send modifications to nodes
  - May also change the hierarchy
- Interaction options
  - Change of observation point
  - Drag objects to different places
  - Start a sequence of events

Scene composition (2 of 2)

- Representation of new objects
  - Download new object classes
  - Initialize classes and data structures
  - Demultiplexing, synchronization, decoding
- Full hierarchy of a scene
  - Video Sequence (VS): a complete video clip
  - Video Object (VO): consists of layers (for scalability)
  - Video Object Layer (VOL): consists of planes
  - Video Object Planes (VOP): shape/texture/motion

MPEG-4 peculiarities

- Requires structuring a scene
  - Instead of MBs we have irregular areas
  - Each object has a shape (a mask)
  - The texture is what the object looks like
  - Each object can move independently
  - Objects can overlap
- Without objects, it becomes MPEG-2
  - With some additional coding tools

Video planes

- What a scene is really composed of
  - Natural or synthetic
    - Different coding approaches
    - May only have a single object (entire frame!)
- Sprites: background objects
  - Simple movement and distortion
- Textures: projected onto objects

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Natural object coding

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics

Natural objects (1 of 7)

- Coded with MBs and blocks
  - Uses 4:2:0 subsampling (always)
  - Coding of shape, texture and motion
- Shape coding
  - An object is enclosed in a rectangular box
    - Consists of one or more MBs
  - A bitmap shows the exact shape
    - Which part of the MBs is part of the object

Natural objects (2 of 7)

- Binary shape bitmap
  - 1: pixel in object, 0: pixel not in object
  - Dissected into 16 x 16 pixel areas (like an MB)
  - Each area coded separately
  - Special coding for areas with all 1s or 0s
    - Areas completely in or out of the object
  - Prediction from previous and neighboring areas
  - Each prediction has a probability
  - Arithmetic coding of the result

Natural objects (3 of 7)

- Grayscale shape bitmap
  - Similar to binary, but with intermediate values
    - 0 to 255 for each pixel
    - Corresponds to object transparency
  - Transform coded
    - Similar to textures (see below)
    - Lossy coding

Natural objects (4 of 7)

- Motion coding
  - As in MPEG-2 but at the VOP level
    - We can have I-VOPs, P-VOPs and B-VOPs
  - Half pixel accuracy motion vectors
  - Matching based only on pixels inside the shape
    - Differences are only coded for those
    - Pixels outside shape ignored
  - Differential coding of motion vectors
    - Prediction from three neighboring vectors

Natural objects (5 of 7)

- Texture coding
  - Like MPEG-2 but with some additions
  - DCT used in each block
    - The standard defines the accuracy level
  - What about pixels outside the shape?
    - DCT always transforms and entire block
    - In P-VOP/B-VOP we use zeros
    - In I-VOP we use padding values

Natural objects (6 of 7)

- Texture coding
  - Padding values for pixels outside shape
  - For blocks partially inside the object
    - First, horizontal fill with neighboring or average value
    - Then, vertical fill with neighboring or average value
  - For blocks fully outside the object
    - Use values from boundary blocks

Natural objects (7 of 7)

- Texture coding
  - Quantization
    - Fixed quantizer for DC
    - Either table or single value for AC
  - Coefficient scanning
    - Zig-zag, per line or per column
    - Depends on how we predict the DC coefficient
  - Entropy coding
    - Two coding tables, depending on quantization

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Synthetic object coding

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics

Synthetic objects (1 of 6)

- Synthetic objects
  - Grids (2D)
  - Human bodies and faces (3D)
- Representing objects with grids
  - Polygonal 2D areas
  - Broken down into triangles
    - Encoding of vertices and their motion
  - Textures projected over the triangles

Synthetic objects (2 of 6)

- Uniform grids
  - Four types of uniform grids
    - Specific ways of arranging triangles
  - Simple description due to uniformity
    - Number of vertices in each dimension
    - Horizontal and vertical edge length
    - Arrangement of triangles
  - Insufficient for complex objects

Synthetic objects (3 of 6)

- Non-uniform grids
  - Based on Delaunay triangulation
    - Each triangle is embedded into a circle
    - Does not contain vertices of other triangles
  - Edge nodes: periphery of the grid
  - Internal nodes: vertices of the triangles
  - Individual coding of the first node
  - Differential coding of subsequent nodes

Synthetic objects (4 of 6)

- Grid motion
  - In each triangle, the vertices (nodes) move
    - But the triangle structure does not change
  - We start with the first triangle
    - We encode the motion of the two nodes
    - And predict the motion of the third
  - We continue with the either triangles
    - Only need to predict the third node

Synthetic objects (5 of 6)

- Face coding
  - Uses a standardized face model
  - Only specific areas can change
- Face Description Parameters (FDP)
  - Changes to approximate different faces
- Face Animation Parameters (FAP)
  - Changes to approximate different movements
  - Library of changes for speech and expressions
  - Differential encoding of FAPs

Synthetic objects (6 of 6)

- Body coding
  - Variation on face coding
  - Use of standardized body model
  - Only specific areas can change
- Body Description Parameters (BDP)
- Body Animation Parameters (BAP)
- Use of a grid for the surface
  - Textures can be projected to it

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Textures and sprites

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics

Static textures

- Coding of static textures
  - Projected on static objects (2D/3D)
  - Based on wavelet transform
    - Separate coding of each sub-band
    - Recursive division of each sub-band for scalability
    - DC sub-band: prediction from neighbors
    - AC sub-band: prediction from other levels
    - Similar coefficients across levels

Sprites

- Sprites: static images, mostly backgrounds
  - Composed of simpler images
    - Static image coding
  - Transmitted in the beginning
  - A panorama is created in memory
  - Then, only part of it is shown
    - Essentially we have a window over the background

Global motion compensation (1 of 2)

- Some changes modify everything
  - Rotation, zoom, camera movement
    - Some cause identical motion vectors (pan)
    - Some cannot be expressed this way (zoom)
- Encoding global changes for sprites
  - Based on an image distortion model
  - Solves an optimization problem
    - Finds the best rotate/pan/zoom parameters

Global motion compensation (2 of 2)

- Warping and blending
  - Modify the image according to model
  - Blends it with sprite to produce a new one
- Motion trajectory coding
  - Motion of key points in the sprite
- Global vs local motion compensation
  - Both options are examined
  - The most economical one is chosen

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Error recovery and evaluation

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics

Error recovery (1 of 2)

- Low speed/high error rate channels
  - Example: cellular telephony
- Fixed size video packets
  - Easy to resync after errors
  - Size depends on the channel
- Additional control information
  - Separation of motion vectors in packets
  - Frame information in packets

Error recovery (2 of 2)

- Reversible code words
  - Variant on Huffman (VLC->RVLC)
    - Variable length code words
  - Can also be read from end to beginning
  - When we lose sync due to bit errors
    - We start reading from the end of the packet
    - We recover both beginning and end of packet

Did MPEG-4 succeed? (1 of 3)

- MPEG-4 added many new tools
  - New encoding techniques
  - Dynamic extensions
  - Synthetic objects
- Not everything got adopted!
  - Only the encoding techniques were kept
  - And they were not as good as expected

Did MPEG-4 succeed? (2 of 3)

- New coding techniques
  - Improved error resilience: passed to H.264
  - Global motion compensation: too complex
- Dynamic extensions
  - Very ambitious!
  - Introduces security issues
  - Lack of use cases

Did MPEG-4 succeed? (3 of 3)

- Synthetic objects
  - Essentially, encodes animation
  - Can use very low bit rates
  - Many options offered
    - Faces, bodies, grids
  - Did not get a lot of traction
    - Abandoned in H.264
  - But worked fine in Shockwave Flash!

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

End of Section # 14

Class: Multimedia Technology, **Section # 14:** Video coding: MPEG-4

Instructor: George Xylomenos, **Department:** Informatics