MPEG-4

MPEG-4

- MPEG-4 will soon be international standard for true multimedia coding.
- MPEG-4 provides very low bitrate & error resilience for Internet and wireless.
- MPEG-4 can be carried in MPEG-2 systems layer.
- MPEG-4 text and graphics can be overlaid on MPEG-2 video for enhanced content: sports statistics and player trajectories.

MPEG-4

- Real audio and video objects
- Synthetic audio and video
- 2D and 3D graphics (based on VRML)

MPEG-4

- Traditional video coding is block-based.
- MPEG-4 provides object-based representation for better compression and functionalities.
- Objects are rendered after decoding object descriptions.
- Display of content layers can be selected at MPEG-4 terminal.

MPEG-4

- User can search or store objects for later use.
- Content does not depend on the display resolution.
- Network providers can re-purpose content for different networks and users.

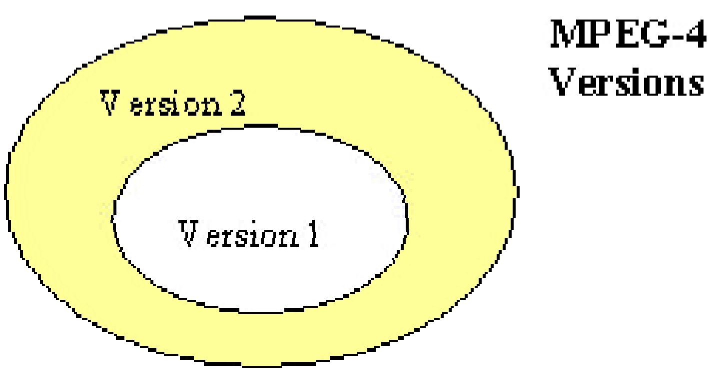
Scope & Features of MPEG-4

- Authors
  - reusability
  - flexibility
  - content owner rights
- Network providers
- End users

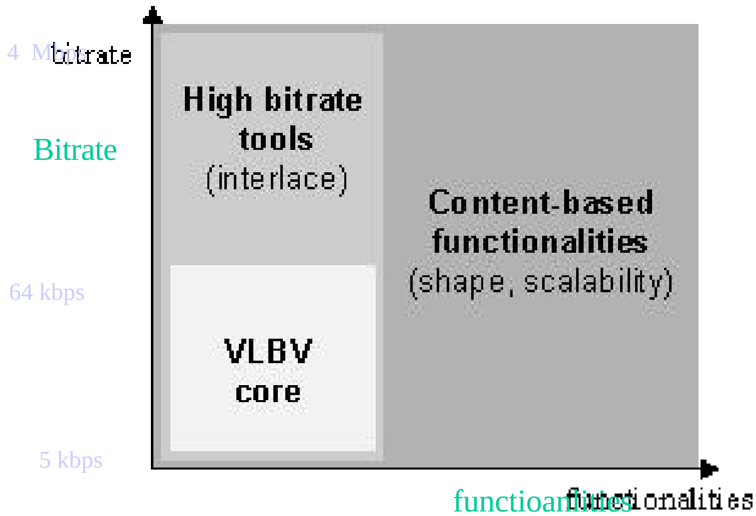
Media Objects

- Primitive Media Objects
- Compound Media Objects
- Examples
  - Still Images (e.g. fixed background)
  - Video objects (e.g., a talking person-without background)
  - Audio objects (e.g., the voice associated with that person)
  - etc

MPEG-4 Versions



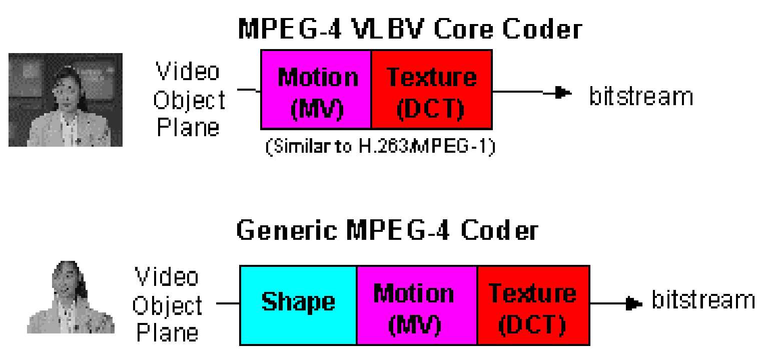
MPEG-4



User Interactions

- Client Side
  - content manipulation done at client terminal
    - changing position of an object
    - making it visible or invisible
    - changing the font size of text
- Server Side
  - requires back channel

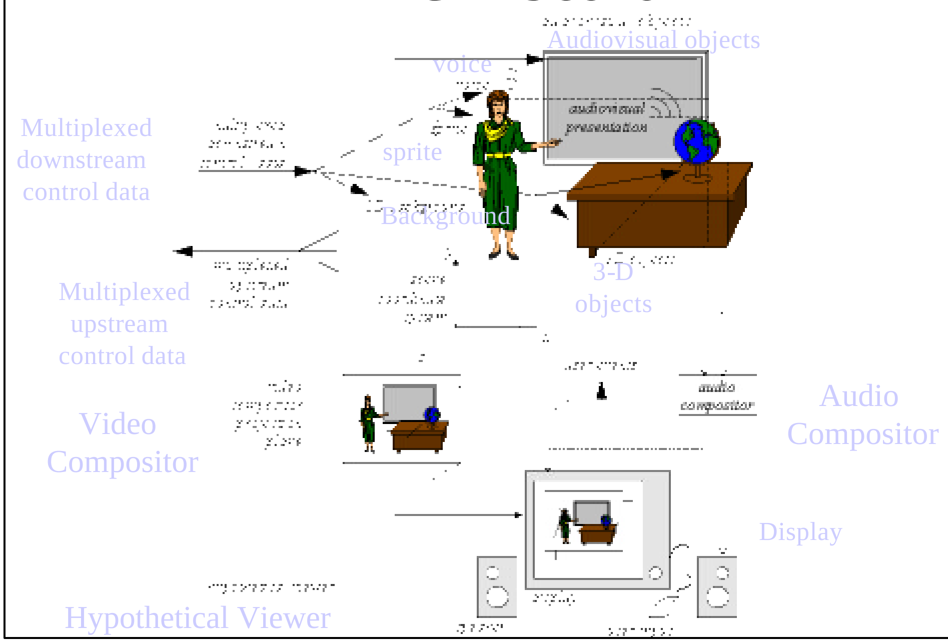
- Efficient representation of visual objects of arbitrary shape to support content-based functionalities
- Supports most functionalities of MPEG-1 and MPEG-2
  - rectangular sized images
  - several input formats
  - frame rates
  - bit rates
  - spatial, temporal and quality scalability



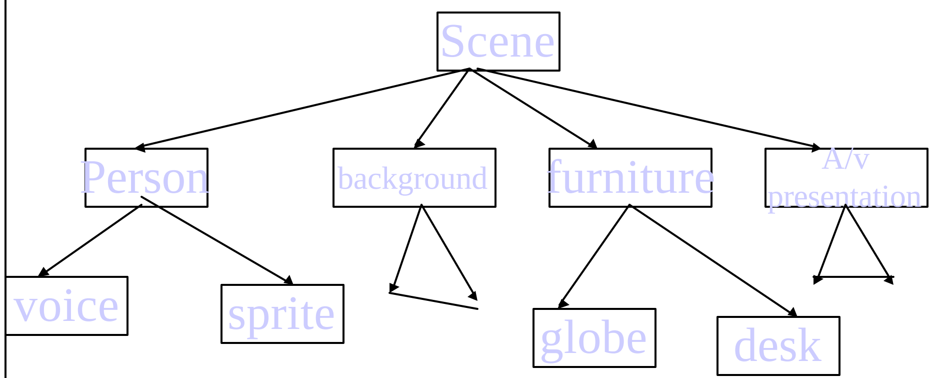
Object Composition

- Objects are organized in a scene graph.
- VRLM based binary format BIF is used to specify scene graph.
- 2-D and 3-D objects, transforms and properties are specified.
- MPEG-4 allows objects to be transmitted once, and displayed repeatedly in the scene after transformations.

MPEG-4 Scene



Scene Graph



Standardized Ways

- To represent “media object”
  – visual or audiovisual
  – synthetic or natural
- To multiplex and synchronize the data associated with media objects for transportation over the network
- Interact with audiovisual scene generated at the receiver’s end.

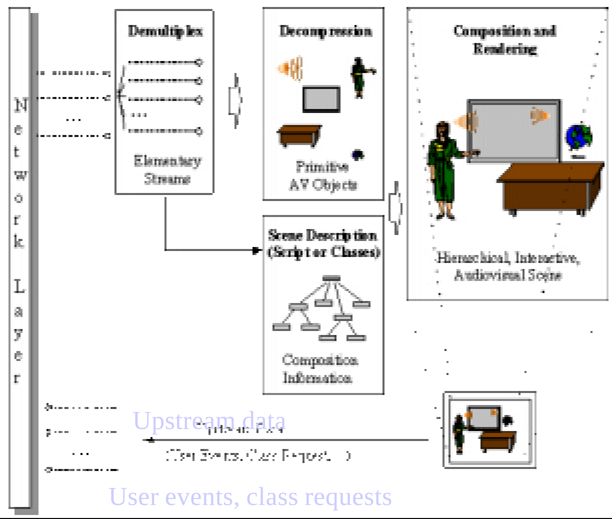
Standardized Ways To

- place a media objects anywhere in a given coordinate system;
- apply transforms to change the geometrical or acoustical appearances of media objects;
- group primitive media objects to form compound media objects;
- apply stream data to media objects to modify their attributes;
- change interactively user’s viewing and listening points anywhere in the scene

Interaction with media objects

- change the viewing/listening point of the scene, e.g., by navigating through a scene;
- drag objects in the scene to a different position;
- trigger a cascade of events by clicking on specific objects, e.g., starting or sopping a video stream;
- select the desired language when multiple language tracks are available;
- more complex behavior

MPEG-4 Terminal



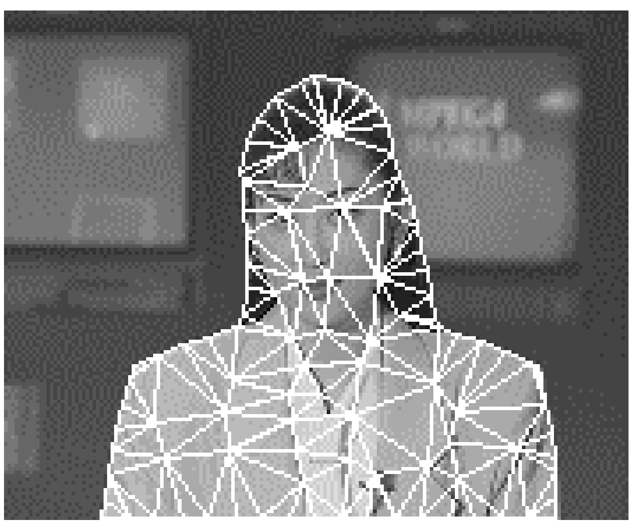
Textures, Images and Video

- Efficient compression of
  - images and video
  - textures for texture mapping on 2D and 3D meshes
  - implicit 2D meshes
  - time-varying geometry streams that animate meshes

Textures, Images and Video

- Efficient random access to all types of visual objects
- Extended manipulation functionalities for images and video sequences
- Content-based coding of images and video
- Content-based scalability of textures, images and video
- Spatial, temporal and quality scalability
- Error robustness and resilience

2-D Mesh Modeling



2-D Mesh Representation of Video Object

- Video Object Manipulation
  - Augmented Reality
  - Synthetic-object-transfiguration/animation
  - Spatio-temporal interpolation (e.g., frame rate up-conversion)
- Video Object Compression
  - transmit texture maps only at keyframes
  - animate texture maps for the intermediate frames

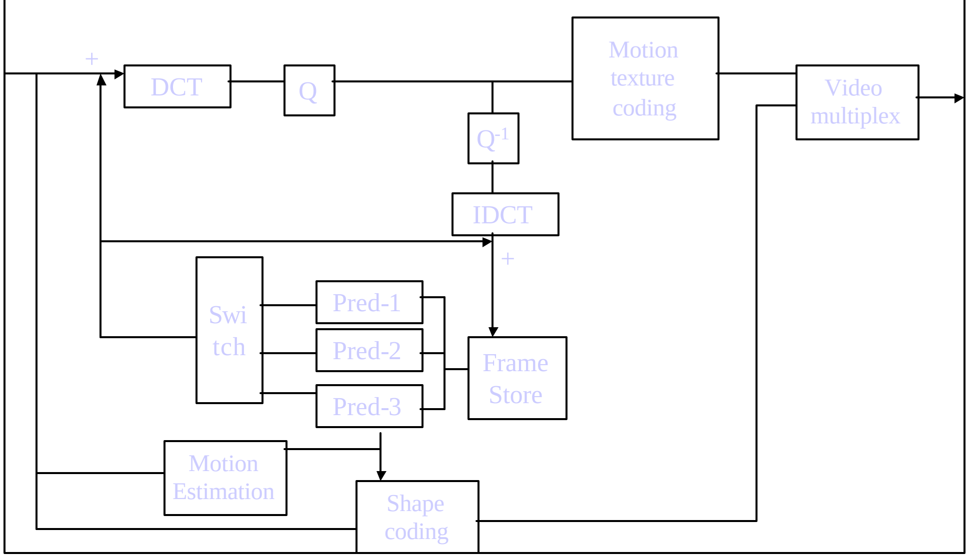
2-D Mesh Representation of Video Object

- Content-Based Indexing
  - Provides vertex-based object shape representation which is more efficient than the bitmap representation of shape-based object retrieval
  - Provides accurate object trajectory information that can be used to retrieve visual objects with specific motion
  - Animated key snapshots as visual synopsis of objects

MPEG-4 Video and Image Coding Scheme

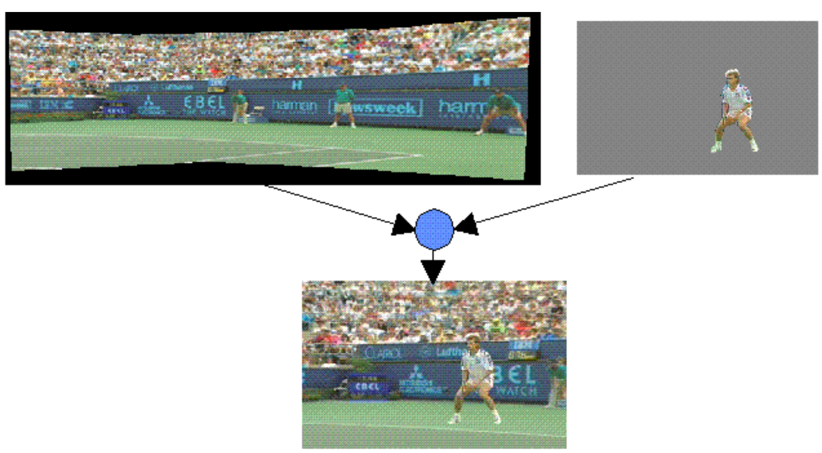
- Shape coding and motion compensation
- DCT-based texture coding
  - standard 8x8 and shape adpated DCT
- Motion compensation
  - local block based (8x8 or 16x16)
  - global (affine) for sprites

MPEG-4 Video Coder



Sprite Panorama

- First compute static “sprite” or “mosaic”
- Then transmit 8 or 6 global motion (camera) parameters for each frame to reconstruct the fame from the “sprite”
- Moving foreground is transmitted separately as an arbitrary-shape video object.



Other Objects

- Text and graphics
- Talking synthetic head and associated text
- Synthetic sound

Face and Body Animtion

- Face animation is in MPEG-4 version 1.
- Body animation is in MPEG-4 version 2.
- Face animation parameters displace feature points from neutral position.
- Body animation parameters are joint angles.
- Face and body animation parameter sequences are compressed to low bit rate.
- Facial expressions: joy, sadness, anger, fear, disgust and surprise.

Face Node

- FAP (Facial Animation Parameters) node
- Face Scene graph
- Face Definition Parameters (FDP)
- Face Interpolation Table (FIT)
- Face Animation Table (FAT)

Face Model

- Face model (3D) specified in VRLM, can be downloaded to the terminal with MPEG-4
- FAT maps FAPS to face model vertices.
- FAPS are quantized and differentially coded
- Typical compressed FAP bitrate is less than 2 kbps

Neutral Face

- Face is gazing in the Z direction
- Face axes parallel to the world axes
- Pupil is 1/3 of iris in diameter
- Eyelids are tangent to the iris
- Upper and lower teeth are touching and mouth is closed
- Tongue is flat, and the tip of tongue is touching the boundary between upper and lower teeth

Facial Animation Parameters (FAPS)

- 2 eyeball and 3 head rotations are represented using Euler angles
- Each FAP is expressed as a fraction of neutral face mouth width, mouth-nose distance, eye separation, or iris diameter.

FAP Groups

Group	FAPS
Visemes & expressions	2
jaw, chin, inner lower-lip, corner lip, mid-lip	16
eyeballs, pupils, eyelids	12
eyebrow	8
cheeks	4
tongue	5
head rotation	3
outer lip position	10
nose	4
ears	4

Visemes and Expressions

- For each frame a weighted combination of two visemes and two facial expressions
- After FAPs are applied the decoder can interpret effect of visemes and expressions
- Definitions of visemes and expressions using FAPs can be downloaded

Phonemes and Visemes

- 56 phonemes
  - 37 consonants
  - 19 vowels/diphthongs
- 56 phonemes can be mapped to 35 visemes

Visems

Viseme_select	phonemes	example
0	none	na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T, D	think, that
4	t, d	tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	s, z	sir, zeal
8	n, l	lot, not
9	r	red
10	A:	car
11	e	bed
12	I	tip
13	O	top
14	U	book

Facial Expressions

- Joy
  - The eyebrows are relaxed. The mouth is open, and mouth corners pulled back toward ears.
- Sadness
  - The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed.
- Anger
  - The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose teeth.

Facial Expressions

- Fear
  - The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert.
- Disgust
  - The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically.
- Surprise
  - The eyebrows are raised. The upper eyelids are wide open, the lower relaxed. The jaw is open.

FAPs

- Speech recognition can use FAPs to increase recognition rate.
- FAPs can be used to animate face models by text to speech systems
- In HCI FAPs can be used to communicate speech, emotions, etc, in particular noisy environment.

MPEG-4 Decoder