



Special Module on Media Processing and Communication

Dayalbagh Educational Institute
(DEI)
Dayalbagh Agra

PHM 961

Indian Institute of Technology Delhi
(IITD)
New Delhi

SIV 864



MPEG (Background)

MPEG-1 and 2

MPEG 1: Standard for storage and retrieval of audio and video on storage media

MPEG 2: Standard for digital TV

MPEG-1/2 are frame based. Virtually with no interactivity.



MPEG (Background)

MPEG-4

MPEG-4 is not only aimed to improve compression, but to improve functionality and interactivity.

MPEG-4 targets:

- Digital TV.

- Interactive graphics, computer games.

- Interactive multimedia, WWW.

MPEG-4 addresses the needs of authors, service providers, end users.

An architecture and coding methods for representing rich multimedia content



MPEG-4: Objectives

- ▶ *Interactivity* : Interacting with the different audio-visual objects
- ▶ *Scalability* : Adopting contents to match resources
- ▶ *Reusability* : For both tools and data



MPEG-4: Objectives

- ▶ *Interactivity* :
 - ▶ Client Side Interaction
 - Manipulating scene description and properties of audio-visual objects
 - ▶ Audio-Visual Objects Behavior
 - Triggered by user actions and other events
 - ▶ Client Server Interaction
 - In case a return channel is available



MPEG-4: Objectives

- ▶ **Scalability** :
- ▶ **Scalability** refers to the ability to only decode a part of a bitstream and reconstruct images or image sequences with:
 - Reduced decoder complexity (reduced quality)
 - Reduced spatial resolution
 - Reduced temporal resolution
- ▶ A **scalable object** is the one that has basic-quality information for presentation. When enough bitrate or resources can be assigned, enhancement layers can be added for improving quality.



MPEG-4: Audio Visual Object

- ▶ The representation of a *natural* or *synthetic* object that has an audio and/or visual manifestation
- ▶ **Examples:**
 - Video Sequence (with Shape information).
 - Audio Track
 - Animated 3D face
 - Speech synthesized from text.
- ▶ **Advantages:** Interaction – Scalability – Reusability



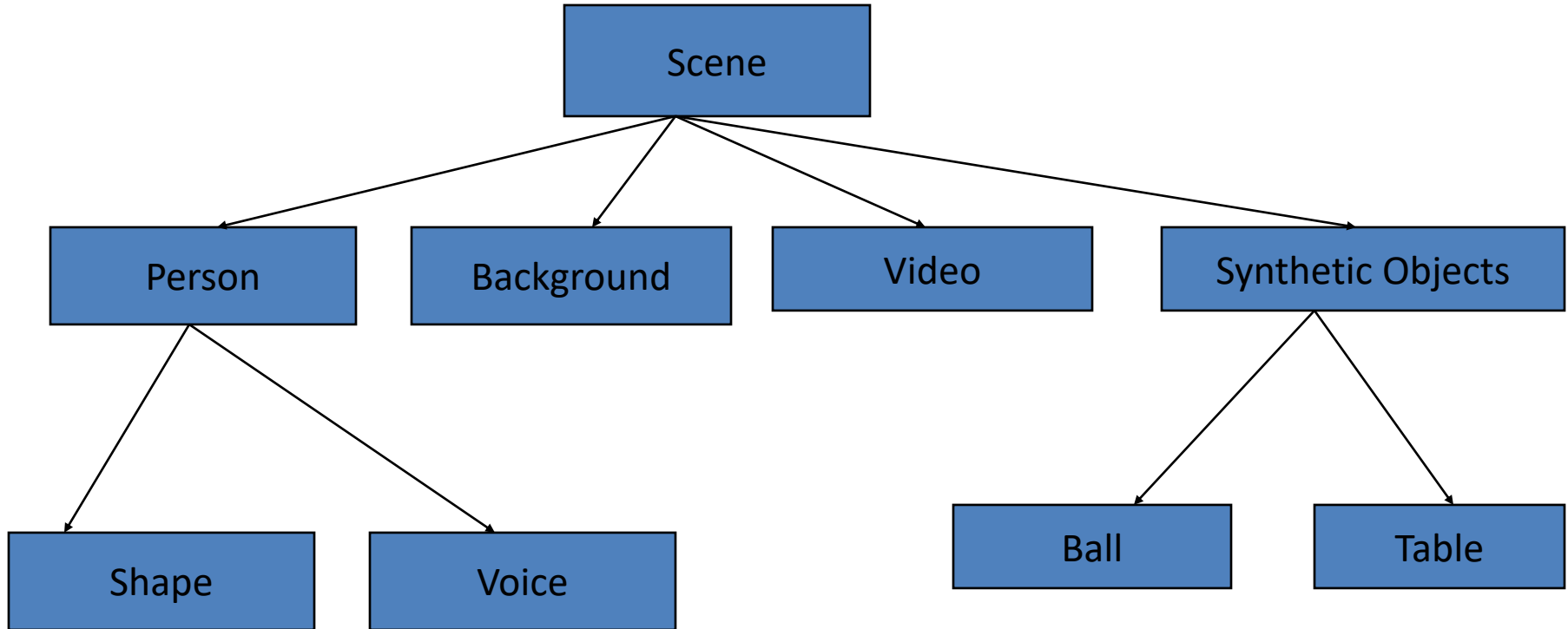
MPEG-4: Scene Description

The coding of information that describes the spatio-temporal relationships between the various audio-visual objects.

- ▶ **the spatial/temporal relationship between the audiovisual objects (2D, 3D, mixed 2D and 3D scene description),**
- ▶ **the behavior and interactivity of the audio-visual objects and scenes,**
- ▶ **protocols to modify and animate the scene in time,**
- ▶ **a binary encoding for the scene.**



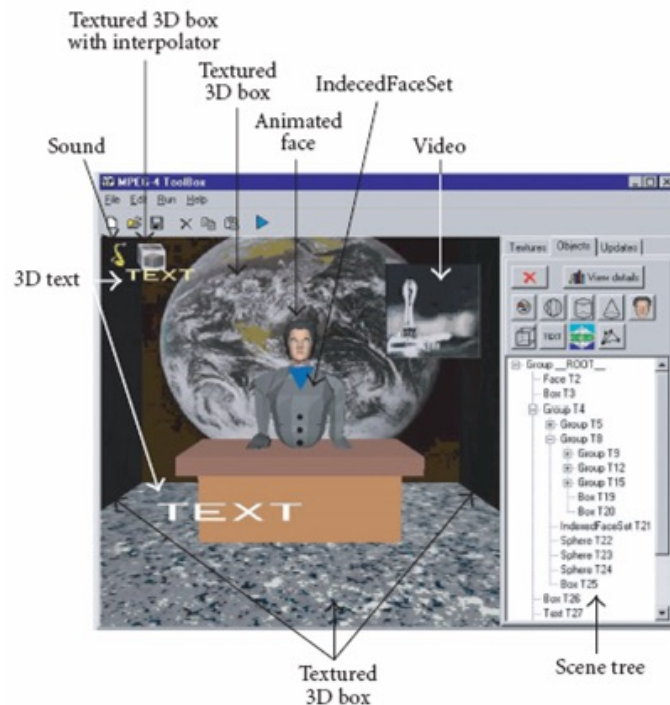
MPEG-4: Scene Description



MPEG-4: Scene Description



2D Audio-visual scene
Audio and Video + Still Images



Source: Ref [2]

<http://www.cse.iitd.ac.in/~pkalra/siv864>

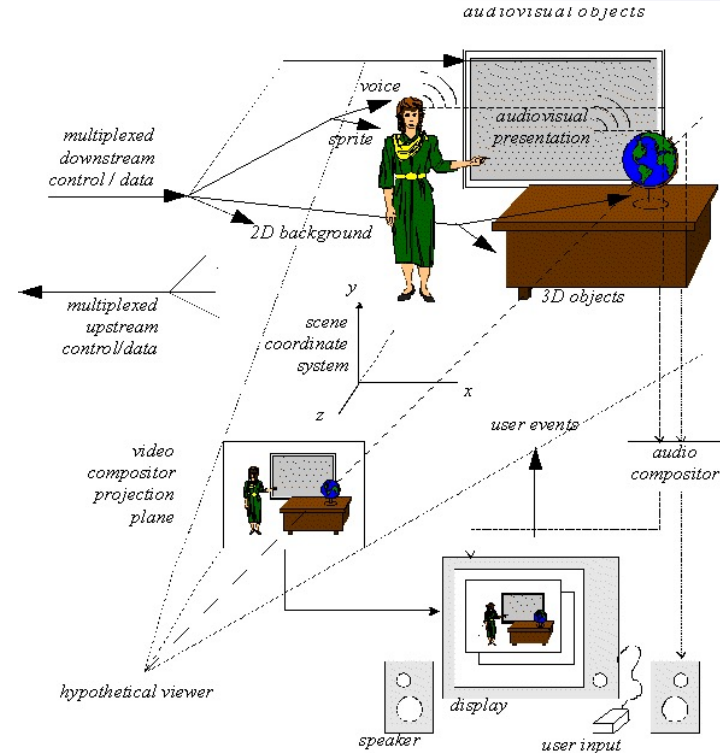
MPEG-4: Scene Description

Composition of Media Objects

In addition MPEG-4 stream includes

- relation between media objects and how they are presented.
- how the media objects are linked to the resources to transmit

Source: Ref [4]





MPEG-4

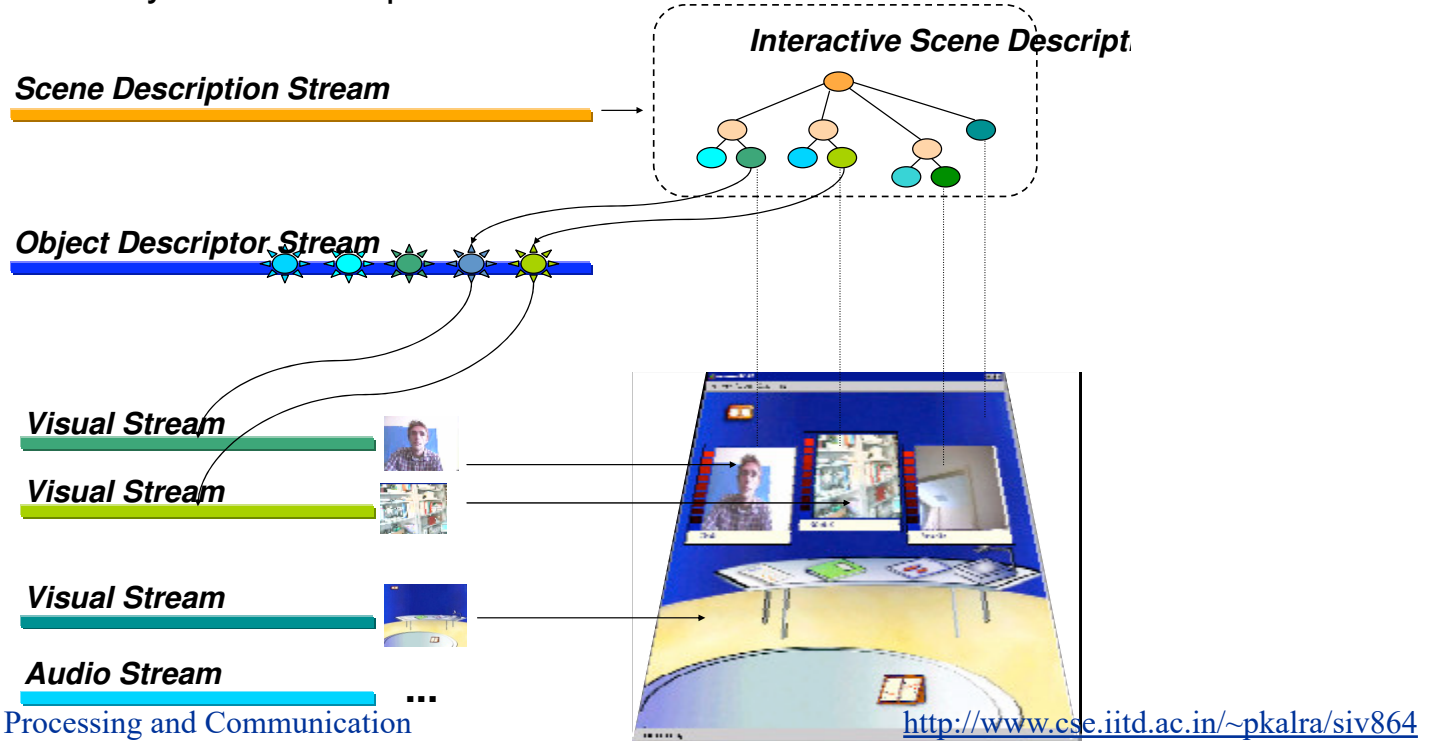
Representation and Scope

- Coded Representation of Interactive Audiovisual Scene,
- Identification and Association of Elementary Streams,
- User interaction.

A coded, streamable representation of audio-visual objects and their associated time-variant data along with a description of how they are combined

MPEG-4

MPEG-4 Systems Principle





MPEG-4: Video Compression

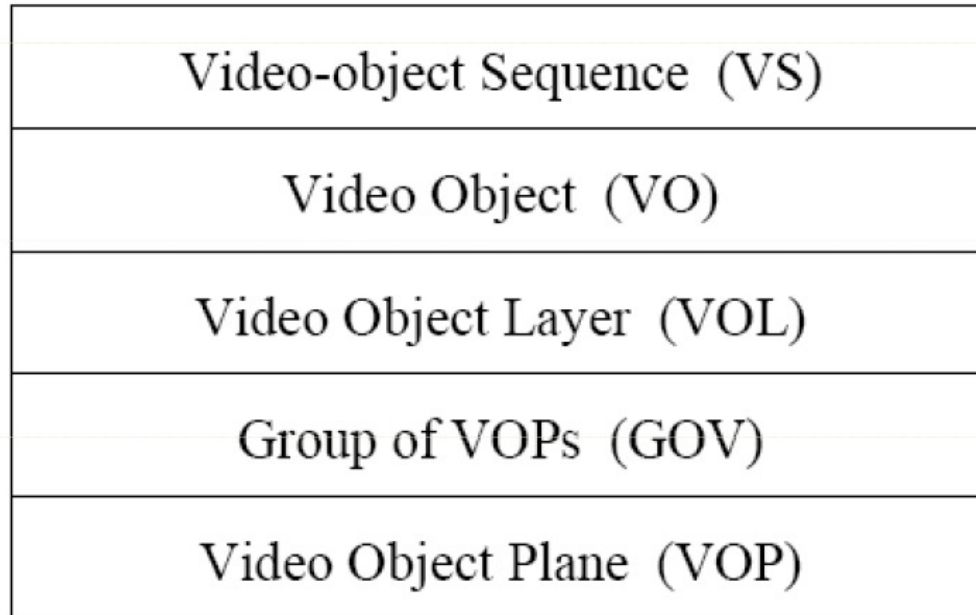
Key ideas:

- Object based coding: offers higher compression ratio, also beneficial for digital video composition, manipulation, indexing and retrieval.
- Synthetic object coding: supports 2D mesh object coding, face object coding and animation, body object coding and animation.
- MPEG-4 Part 10/H.264: new techniques for improved compression efficiency.



MPEG-4: Video Compression

The hierarchical structure of MPEG-4 visual bitstreams is different from that of MPEG-1/2: it is video object-oriented:





MPEG-4: Video Compression

Video-object Sequence (VS): delivers the complete MPEG4 visual scene; may contain 2D/3D natural or synthetic objects.

Video Object (VO): a particular object in the scene, which can be of arbitrary (non-rectangular) shape corresponding to an object or background of the scene.

Video Object Layer (VOL): facilitates a way to support (multi-layered) scalable coding. A VO can have multiple VOLs under scalable (multi-bitrate) coding, or have a single VOL under non-scalable coding.

Group of Video Object Planes (GOV): groups of video object planes together (optional level).

Video Object Plane (VOP): a snapshot of a VO at a particular moment.



MPEG-4: Video Compression

VOP-based vs. Frame-based Coding

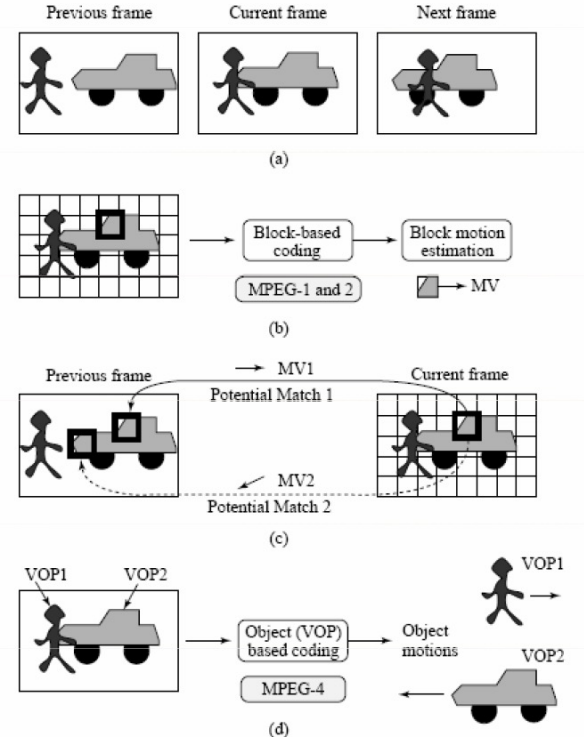
MPEG-1 and MPEG-2 do not support the VOP concept; their coding method is frame-based (also known as block-based).

For block-based coding, it is possible that multiple potential matches yield small prediction errors. Some may not coincide with the real motion.

For VOP-based coding, each VOP is of arbitrary shape and ideally will obtain a unique motion vector consistent with the actual object motion.

MPEG-4: Video Compression

VOP-based vs. Frame-based Coding





MPEG-4: Video Compression

VOP-based coding

MPEG-4 VOP-based coding also employs Motion Compensation technique:

I-VOPs: Intra-frame coded VOPs.

P-VOPs: Inter-frame coded VOPs if only forward prediction is employed.

B-VOPs: Inter-frame coded VOPs if bi-directional predictions are employed.

The new difficulty for VOPs: may have arbitrary shapes. Shape information must be coded in addition to the texture (luminance or chroma) of the VOP.



MPEG-4: Video Compression

Motion Compensation-based VOP coding in MPEG-4 again involves three steps:

1. Motion Estimation
2. MC-based Prediction
3. Coding of the Prediction Error

Only pixels within the VOP of the current (target) VOP are considered for matching in MC. To facilitate MC, each VOP is divided into macroblocks with 16x16 luminance and 8x8 chrominance images.



MPEG-4: Video Compression

Texture Coding (luminance and chrominance):

I-VOP: the gray values of the pixels in each MB of the VOP are directly coded using DCT followed by VLC (Variable Length Coding), such as Huffman or Arithmetic Coding.

P-VOP/B-VOP: MC-based coding is employed the prediction error is coded similar to I-VOP.

Boundary MBs need appropriate treatment. May also use improved Shape Adaptive DCT.C



MPEG-4: Video Compression

Shape Coding (shape of the VOPs)

Binary shape information: in the form of a binary map. A value `1' (opaque) or `0' (transparent) in the bitmap indicates whether the pixel is inside or outside the VOP.

Greyscale shape information: value refers to the transparency of the shape ranging from 0 (completely transparent) and 255 (opaque).

Specific encoding algorithms are designed to code in both cases.



MPEG-4: Video Compression

Sprite Coding

- Sprite consists of those regions of a VO that are present in a scene throughout video segment
 - e.g.. Background Sprite
- Shape and texture component encoded as in I-VOP, may be followed by a set of SVOP
- Can provide high Compression Efficiency



MPEG-4: Video Compression

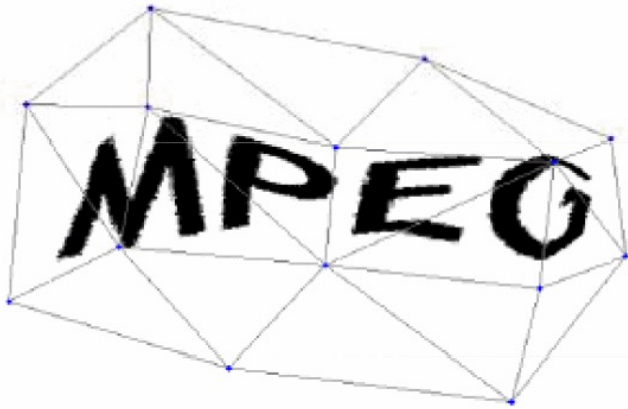
Sprite Coding



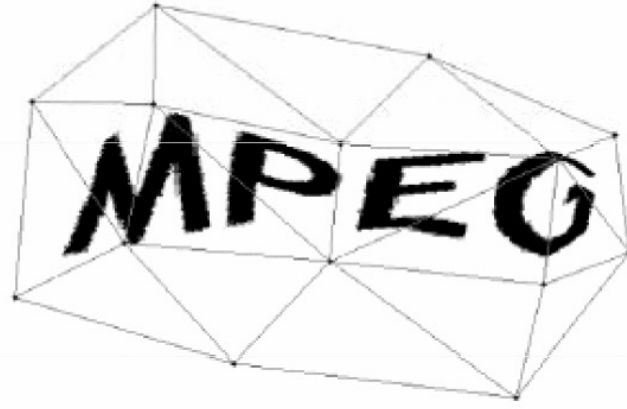


MPEG-4: Synthetic Object Coding

2D Mesh Object: a tessellation (or partition) of a 2D planar region using polygonal patches. Mesh based texture mapping can be used for 2D object animation.



(a)

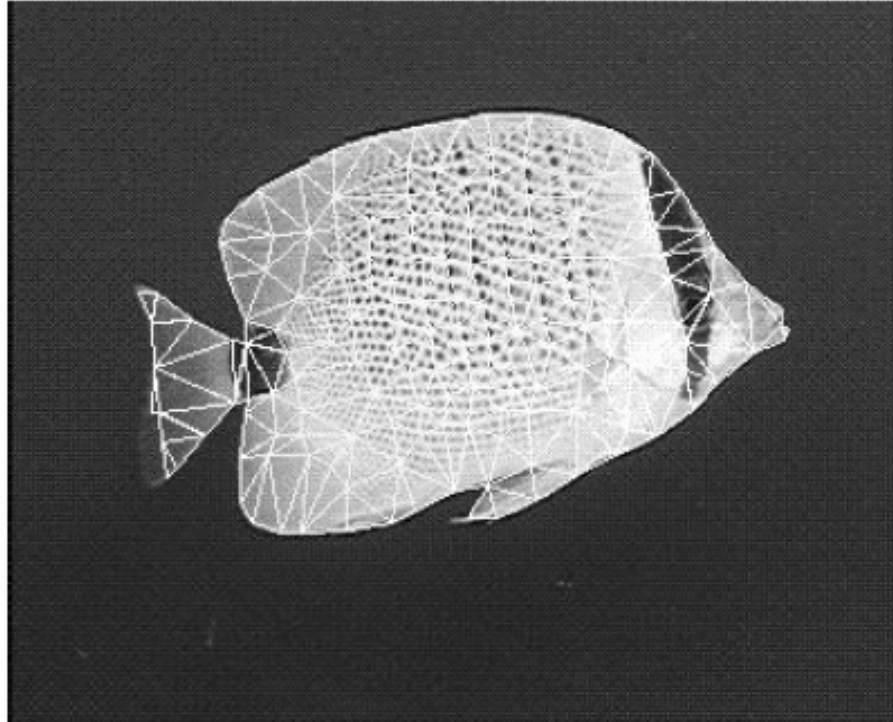
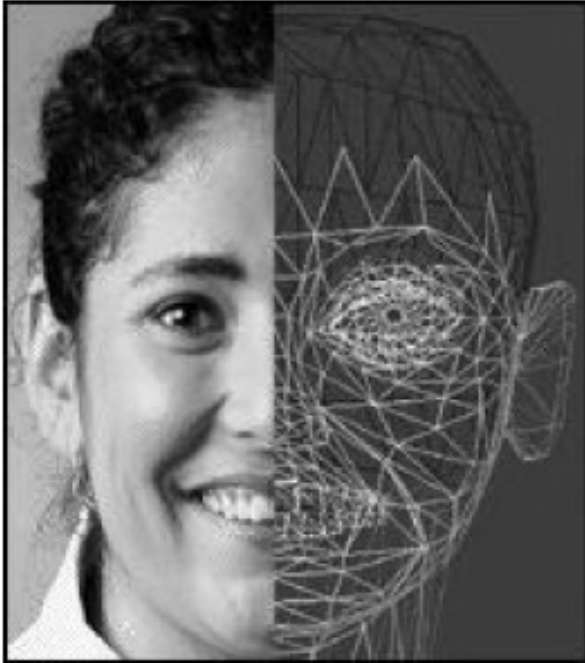


(b)



MPEG-4: Synthetic Object Coding

2D Mesh Object:





MPEG-4: Synthetic Object Coding

MPEG-4 has defined special 3D models for face objects and body objects because of the frequent appearances of human faces and bodies in videos.

Some of the potential applications: teleconferencing, human-computer interfaces, games and e-commerce.

MPEG-4 goes beyond wireframes so that the surfaces of the face or body objects can be shaded or texture-mapped.



MPEG-4: Face Object Coding

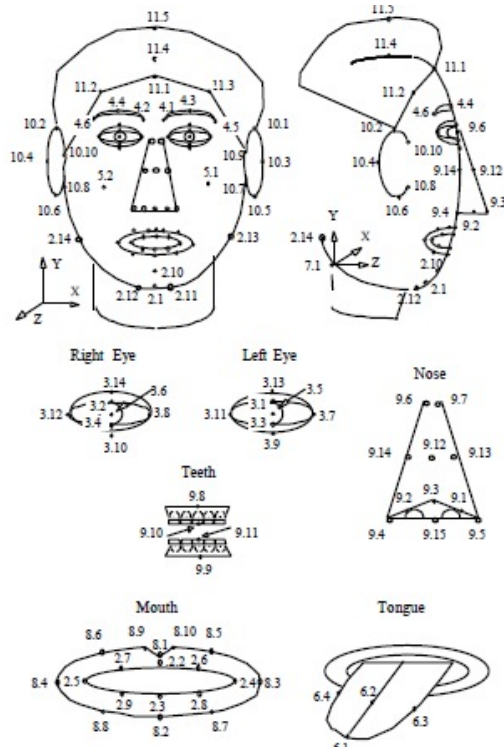
Face Object Coding and Animation

MPEG-4 adopted a generic default face model.

Face Animation Parameters (FAPs) can be specified to achieve desirable animation.

Face Definition Parameters (FDPs): feature points better describe individual faces.

MPEG-4: Face Object Coding



Source: Ref [2]



MPEG-4: Part 10/H.264

Improved video coding techniques, identical standards: ISO MPEG-4 Part 10 (Advanced Video Coding / AVC) and ITU-T H.264.

Preliminary studies using software based on this new standard suggests that H.264 offers up to 30-50% better compression than MPEG-2 and up to 30% over H.263+ and MPEG-4 advanced simple profile.

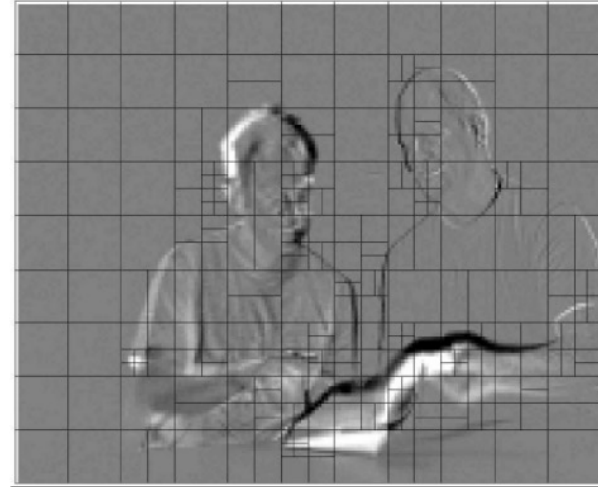
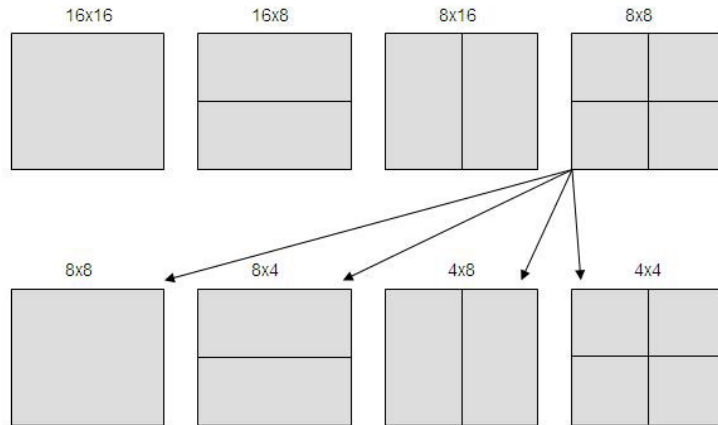
H.264 is currently used to carry High Definition TV (HDTV) video content on many applications.

Involves various technical improvements.

MPEG-4: Part 10/H.264

Macroblock in MPEG-2 uses 16 x16 luminance values.

MPEG-4 AVC uses a tree-structured motion segmentation down to 4x4 block sizes (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4). This allows much more accurate motion compensation of moving objects.



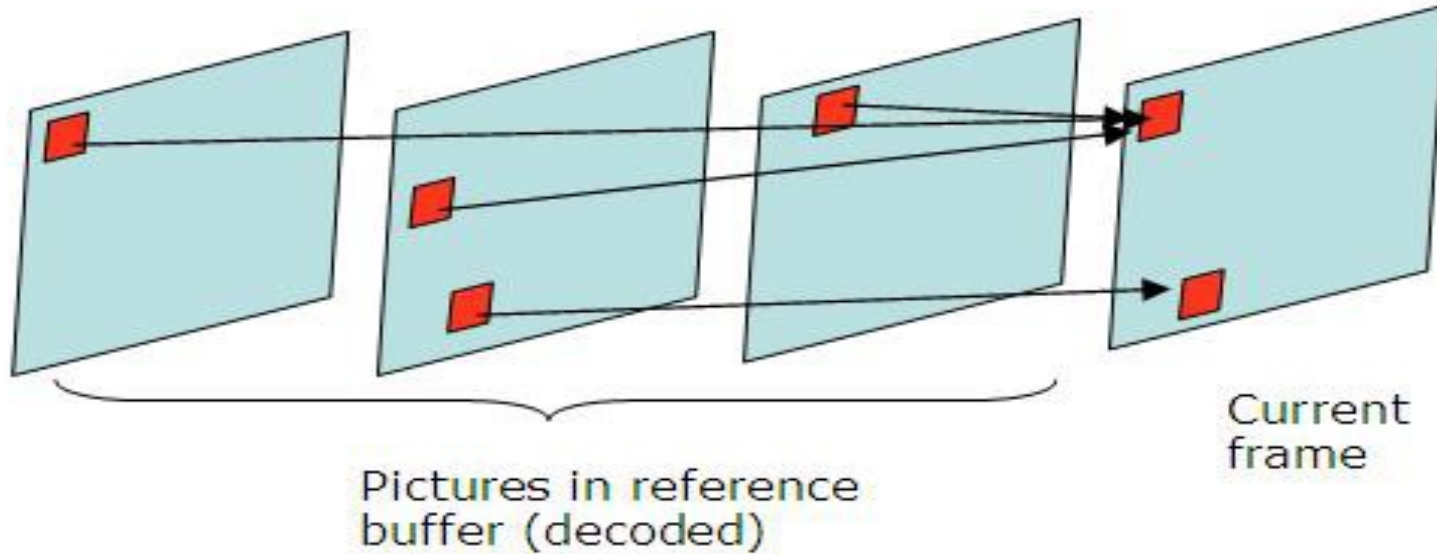


MPEG-4: Part 10/H.264

Multiple references to motion estimation. Allows finding the best reference in 2 possible buffers (past pictures and future pictures) each contains up to 16 frames.

Block prediction is done by a weighted sum of blocks from the reference picture. It allows enhanced picture quality in scenes where there are changes of plane, zoom, or when new objects are revealed.

MPEG-4: Part 10/H.264



Source: Ref [1]



References

1. MPEG 4 Video Course slides of CM3106 by David Marshall and Dr Kirill Sidorov, School of Computer Science & Informatics, Cardiff University, UK.
2. MPEG 4 Overview Course slides on MPEG 4 Overview from Prof Santanu Chaudhury, EE Department IIT Delhi
3. MPEG-4 Natural Video Coding – An Overview, by T Ebrahimi and C Horne
(http://img.ix.it.pt/~fp/cav/Additional_material/MPEG4_video.pdf)
4. Optibase White Paper on A Guide to MPEG-4