MPEG-4 Video Compression

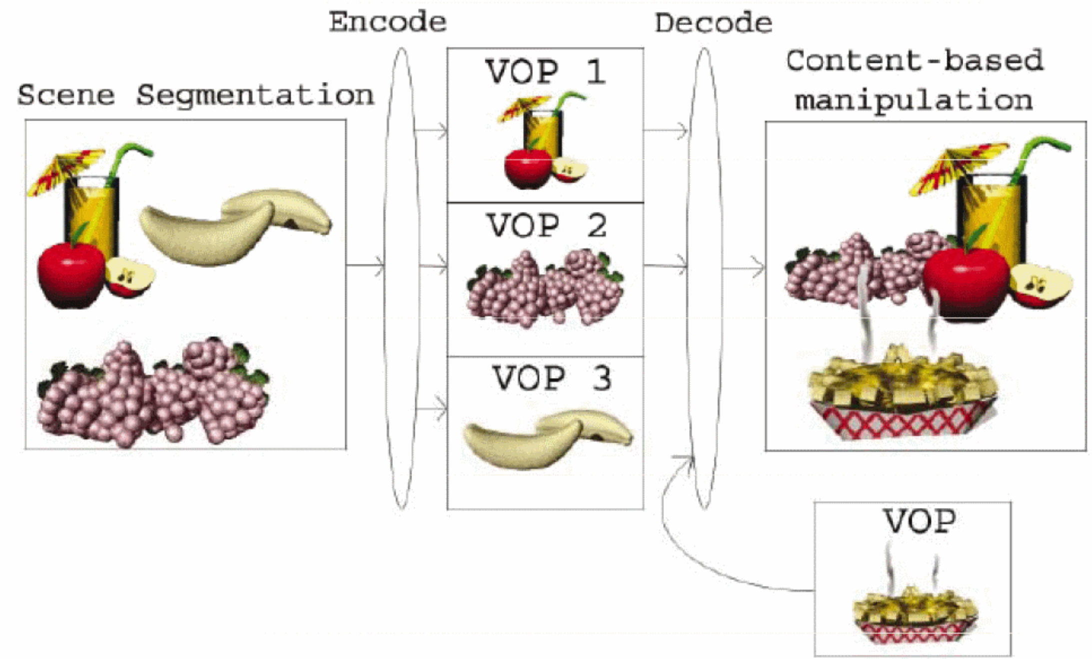
A newer standard than MPEG-2. Besides compression, great attention is paid to issues about **user interactivity**.

We look at **key ideas** here.

- **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.

Object based coding (1)

Composition and manipulation of MPEG-4 videos.



Object based coding (2)

Compared with MPEG-2, MPEG-4 is an entirely new standard for

- **Composing** media objects to create desirable audiovisual scenes.
- **Multiplexing** and **synchronising** the bitstreams for these media data entities so that they can be transmitted with guaranteed Quality of Service (QoS).
- **Interacting** with the audiovisual scene at the receiving end.

MPEG-4 provides a set of advanced coding modules and algorithms for audio and video compressions.

We have discussed MPEG-4 Structured Audio and we will focus on video here.

Object based coding (3)

The hierarchical structure of MPEG-4 visual bitstreams is very different from that of MPEG-2: it is very much **video object-oriented**:

Video-object Sequence (VS)
Video Object (VO)
Video Object Layer (VOL)
Group of VOPs (GOV)
Video Object Plane (VOP)

Object based coding (4)

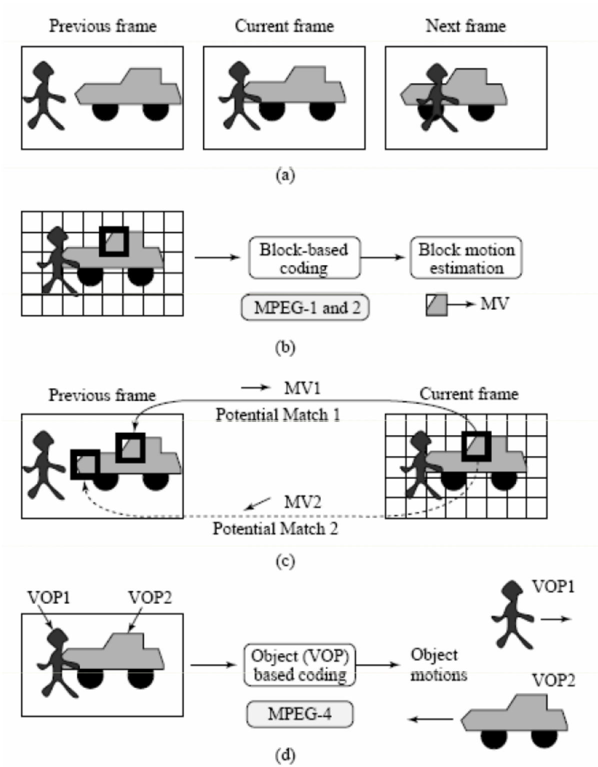
- **Video-object Sequence (VS)**: delivers the complete MPEG4 visual scene; may contain 2D/3D natural or synthetic objects.
- **Video Object (VO)**: a particular object in the scene, which can be of arbitrary (non-rectangular) shape corresponding to an object or background of the scene.
- **Video Object 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.

VOP-based vs. Frame-based Coding

- MPEG-1 and MPEG-2 do **not** support the VOP concert; 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.

See the example in the next page:

VOP-based vs. Frame-based Coding



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.

VOP-based Motion Compensation (MC)

- MC-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 16×16 luminance and 8×8 chrominance images.

VOP-based Motion Compensation (MC)

- Let $C(x+k, y+l)$ be pixels of the MB in target in target VOP, and $R(x+i+k, y+j+l)$ be pixels of the MB in Reference VOP.
- A **Sum of Absolute Difference (SAD)** for measuring the difference between the two MBs can be defined as:

$$SAD(i,j) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \cdot Map(x+k, y+l).$$

  N — the size of the MB

  $Map(p,q) = 1$ when $C(p,q)$ is a pixel within the target VOP otherwise $Map(p,q) = 0$.

- The vector (i,j) that yields the **minimum SAD** is adopted as the motion vector (u,v).

Coding of Texture and Shape

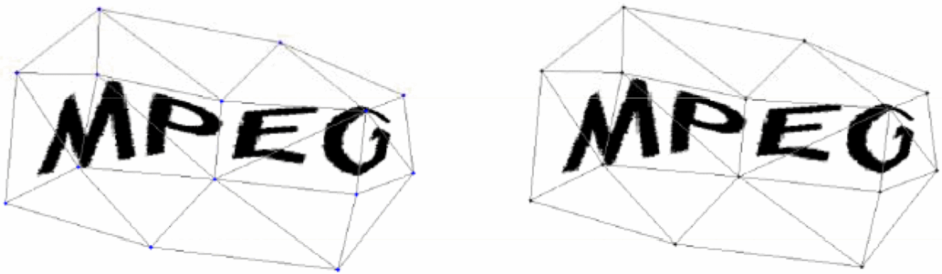
- Texture Coding (luminance and chrominance):
  - I-VOP: the gray values of the pixels in each MB of the VOP are directly coded using **DCT followed by VLC (Variable Length Coding)**, such as Huffman or Arithmetic Coding.
  - P-VOP/B-VOP: MC-based coding is employed — the **prediction error** is coded similar to I-VOP.
  - Boundary MBs need appropriate treatment. May also use improved Shape Adaptive DCT.

Coding of Texture and Shape (cont.)

- 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.

Synthetic Object Coding: 2D Mesh

- 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)

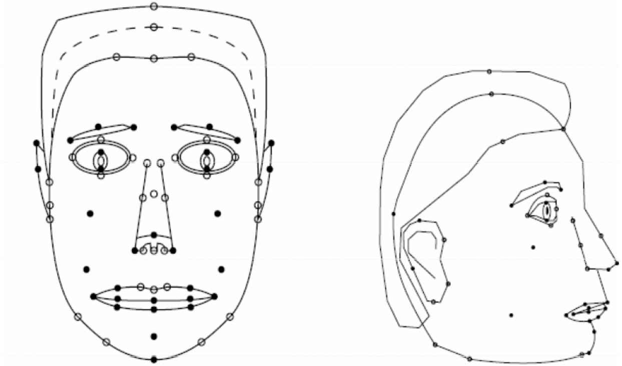
Synthetic Object Coding: 3D Model

- 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**: teleconferecing, 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.

Synthetic Object Coding: Face Object

Face Object Coding and Animation

- MPEG-4 adopted a generic default face model, developed by VRML Consortium.
- **Face Animation Parameters (FAPs)** can be specified to achieve desirable animation.
- **Face Definition Parameters (FDPs)**: feature points better describe individual faces.

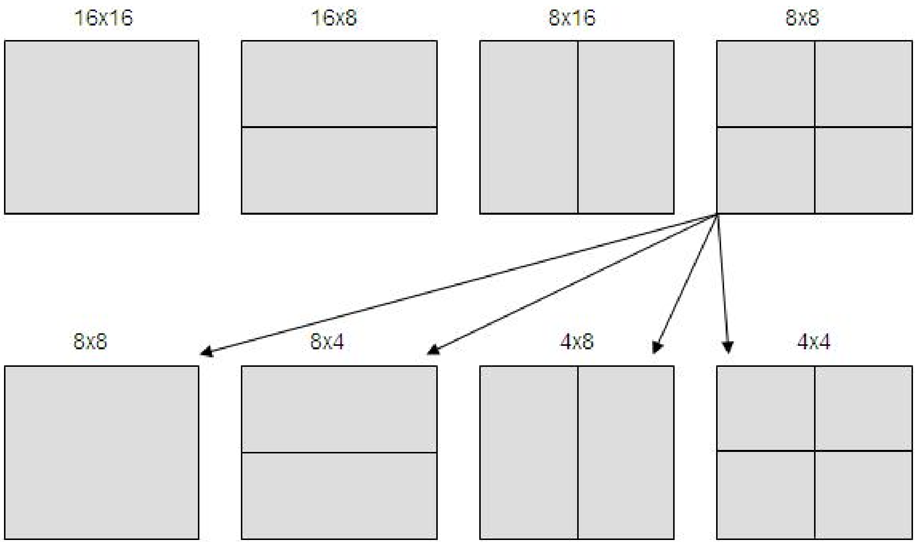


MPEG-4 Part 10/H.264

- 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 one of the leading candidates to carry **High Definition TV (HDTV)** video content on many applications, e.g. Blu-ray.
- Involves various technical improvements. We mainly look at improved inter-frame encoding.

MPEG-4 AVC: Flexible Block Partition

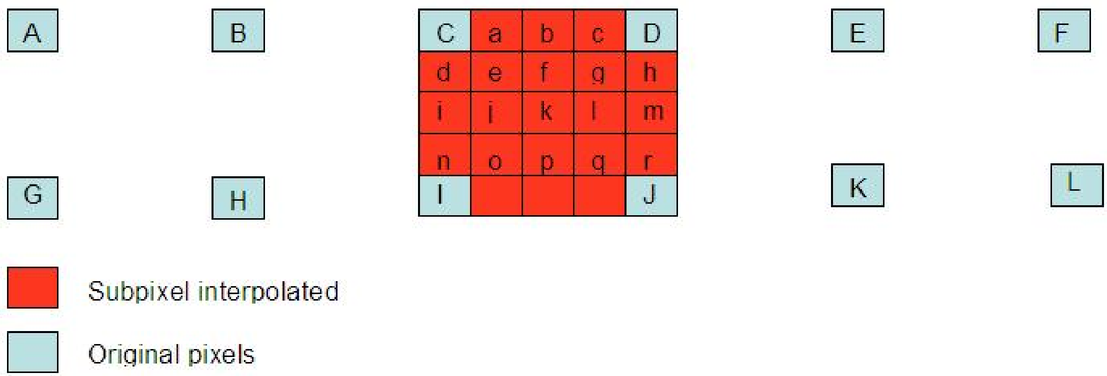
Macroblock in MPEG-2 uses 16×16 luminance values. MPEG-4 AVC uses a tree-structured motion segmentation down to 4×4 block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4). This allows much more accurate motion compensation of moving objects.



MPEG-4 AVC: Up to Quarter-Pixel Resolution Motion Compensation

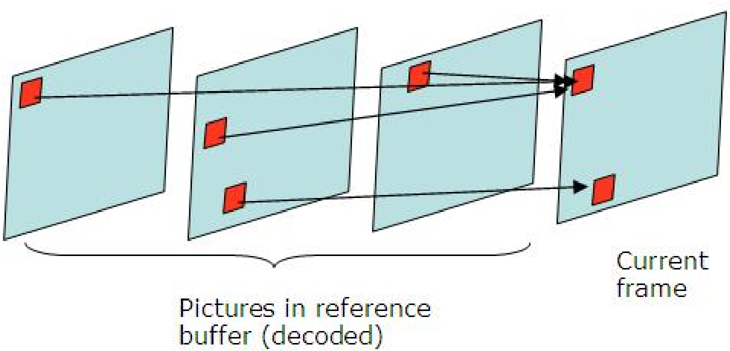
Motion vectors can be up to half-pixel or quarter-pixel accuracy. Pixels at quarter-pixel position are obtained by bilinear interpolation.

- Improves the possibility of finding a block in the reference frame that better matches the target block.



MPEG-4 AVC: Multiple References

- 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.



Further Reading

- Overview of the MPEG-4 Standard
- The H.264/MPEG4 AVC Standard and its Applications