Mobile Programming and Multimedia

Video formats

Prof. Ombretta Gaggi
University of Padua

Video: fundamentals

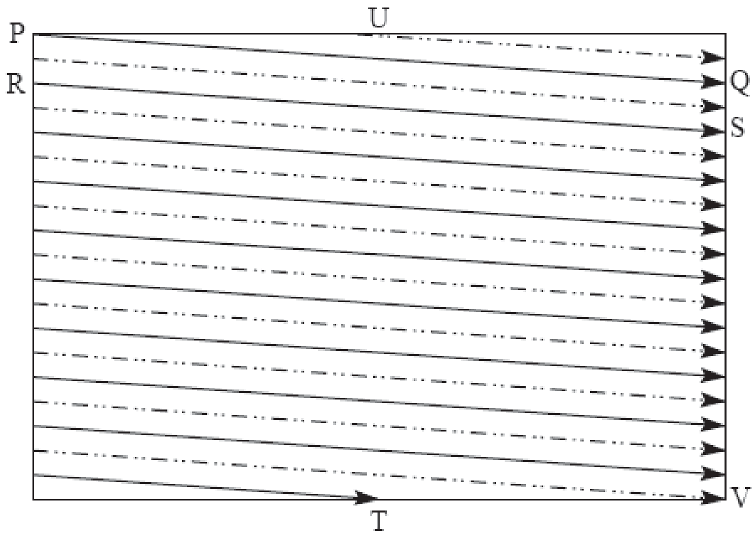
Analog video is encoded as a continuous signal that varies over time

- It can be digitalized, but not further elaborated due to the bi-dimensionality of the images

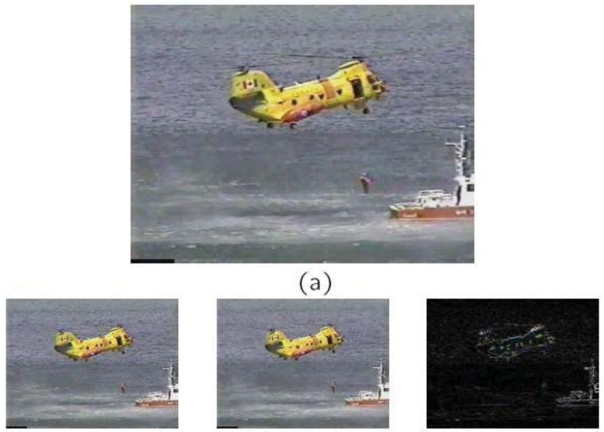
Digital video is a sequence of digital images

- Direct access to every frame
- Nonlinear video editing
- Unnecessary supplementary signals (*blanking*, synchronization, …)

Interlaced video



Interlaced video: example



(a)
(b) (c) (d)

Types of video signals

Video with separated components
- Each primary signal (RGB, YUV) is transmitted as a separated signal
- It allows a better color reproduction due to the absence of interference phenomenon between signals
- Requires high bandwidth and precise synchronization between the three signals

Composite Video
- Luminance and chrominance signals are mixed in a single carrier wave
- Interference between signals

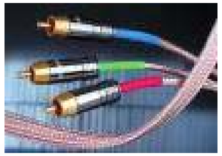
S-Video
- Chrominance signals are mixed in a single carrier wave, while the luminance signal is sent separately

Analog video usually uses a composite signal (always for transmission)

Digital video uses a signal with separated components

Video: wires for different signals

Separated component Video



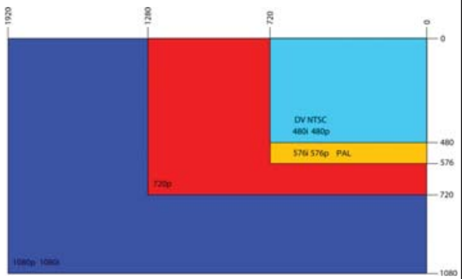
Composite Video

S-video

Video: properties

Color depth
- Recording encodes *true*-images

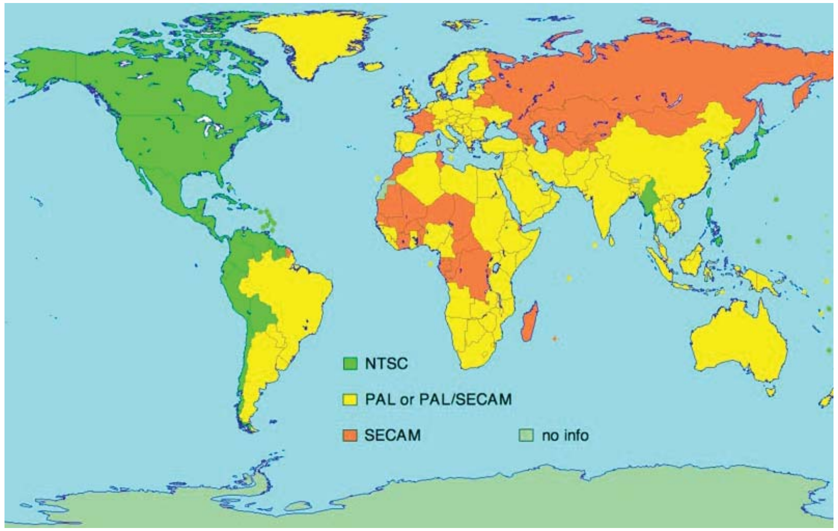
Resolution
- Depends on the standards
- Chrominance information is under-sampled



Frame frequency
- PAL = 25 frames/sec
- NTSC = 29.97 (~ 30) frames/sec
- minimum ~ 15 frames/sec to avoid the perception of snap movements

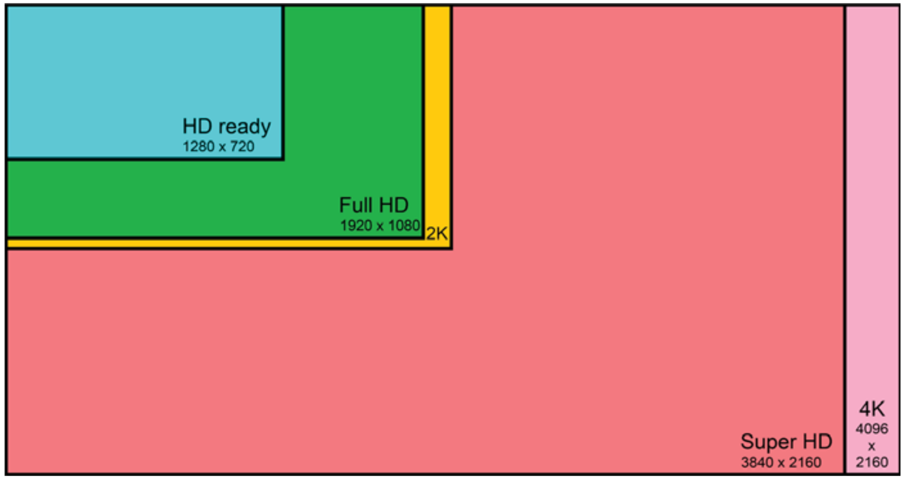
CCIR 601 NTSC	720 x 480 (525)
CCIR 601 PAL	720 x 576 (625)

TV systems

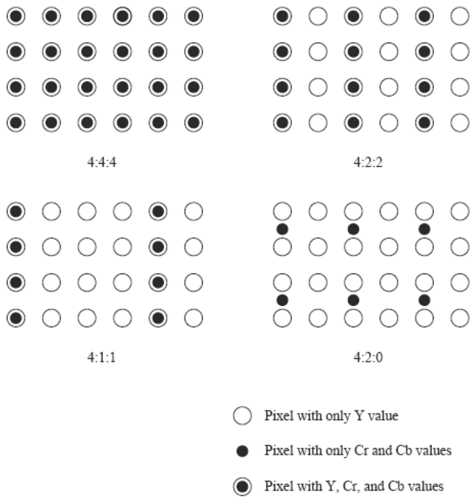


Fonte: www.wikipedia.org

Higher resolutions



Sampling



Video: memory usage

The uncompressed video requires a considerable amount of storage

- High Definition Television (HDTV) requires a bit-rate that can be higher than 1 Gbps

Data must be compressed

- 1 hour of MPEG-1 video with VHS (352 x 288, 25 frames/sec) takes~600 Mbyte (a CD-ROM)

Necessary to use lossy compression techniques

- Elimination of spatial and temporal redundancy
- *intra-frame* and *inter-frame* encoding

Video: transfer time

Video loading from the network has the same problems of images loading, plus …

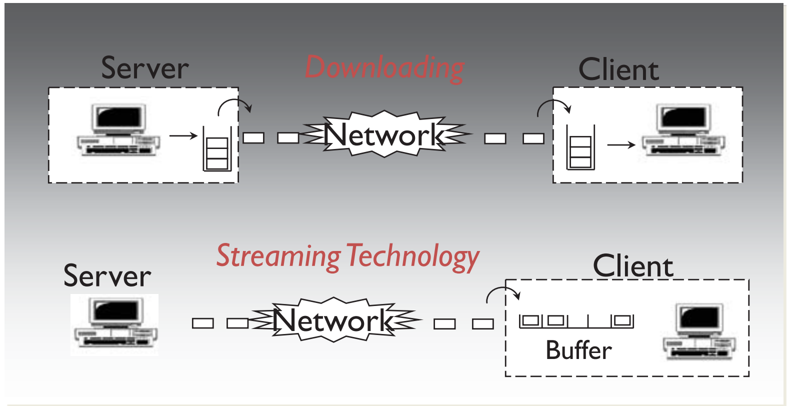
- A video is a temporized and continuous data
- Loading time must be compatible with reproduction time
- Playback must have a constant frame rate

A ***download*** + ***play*** solution is not always acceptable

It is necessary to use ***streaming*** techniques (plaback while transferring data)

Temporization control requires advanced buffering techniques

Streaming technology



Motion JPEG

The first attempt of digital video

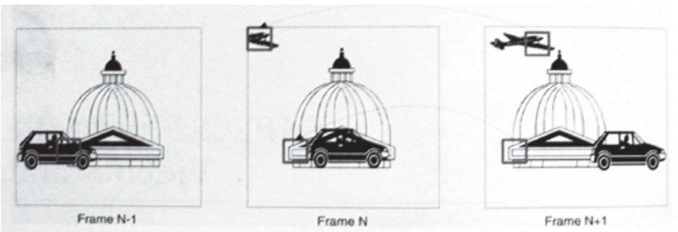
Video signal encoded as a sequence of frames: each frame is encoded as a JPEG image

Does not take advantage of the clear correlation between one frame and the next one

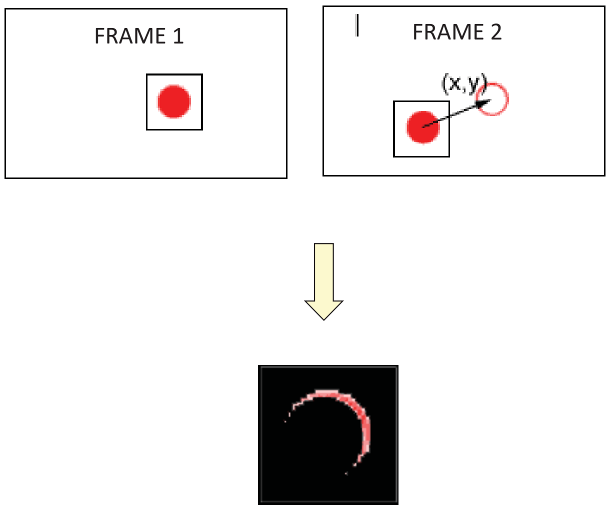
Temporal redundancy

When encoding video frames, it is possible to omit several data because, except for scene changes, there are only a few differences between two images within a small amount of time

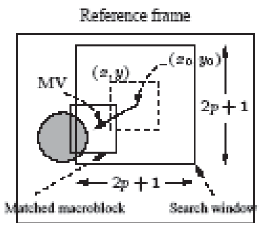
The differences between one frame and the next one usually depend on the movement of some pieces of the frame



Example



Search Algorithms



$$MAD(i,j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)|$$

N – size of the macroblock,
k and l – indices for pixels in the macroblock,
i and j – horizontal and vertical displacements,
$C(x+k, y+l)$ – pixels in macroblock in Target frame,
$R(x+i+k, y+j+l)$ – pixels in macroblock in Reference frame.

$$(u, v) = [\ (i,j) \mid MAD(i,j) \ \text{is minimum}, \ i \in [-p, p], \ j \in [-p, p]\ ]$$

Li & Drew, Fundamentals of Multimedia, 2003

Sequential Search (Full Search)

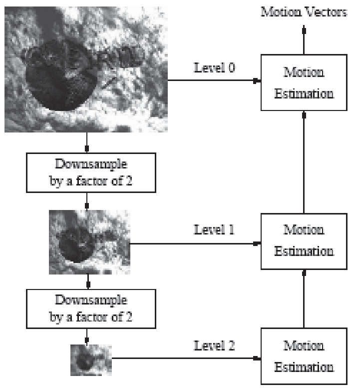
The *Sequential Search algorithm* explores the whole space *(2p+1) x (2p+1)* to find a macroblock similar (minimum MAD) to the considered macroblock

- The target macroblock is compared, bit by bit, with a macroblock centered in every possible position of the research space, and the MAD is calculated
- the difference between the two positions (i.e., the movement) is stored in the motion vectior
- the output is the difference between the target macroblock and the one with minimum MAD

Computationally very expensive: $O(p^2N^2)$

Hierarchical research

The hierarchical research algorithm works using several approximation levels in which initial estimation of the motion vector can be obtained from images with low resolution



Li & Drew, Fundamentals of Multimedia, 2003

H.261 video standard

H. 261, developed by CCITT in 1988-1990

- Developed for videoconferences and video calls using ISDN telephone lines
- Images encoded with CIF (352 x 288) and QCIF (176 x 144) format, 4:2:0
- bit-rate is p x 64 Kb/sec, 1 <= p <= 30

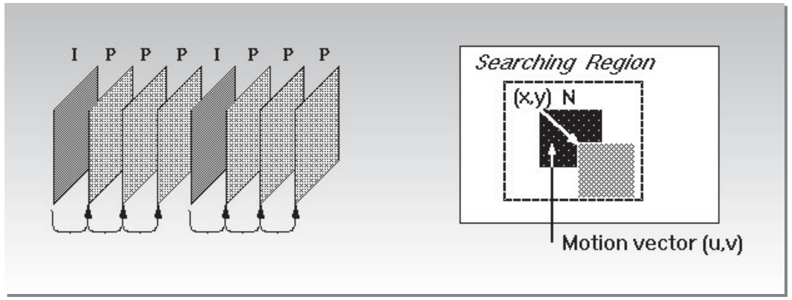
Encoding:

- Encoding and decoding must happen in real-time with a maximum delay of 150ms
- Input frame rate must be 29.97 fps (non-interlaced video), while output frame rate varies between 10 and 15 fps
- Color space *YCbCr* with chrominance components downsampled
- Two different frame types: intra-frames (*I-frames*) and inter-frames (*P-frames*)
- *Intra-frames*: treated as independent images, frames of the video
- *Inter-frames*: encoded using information from other frames

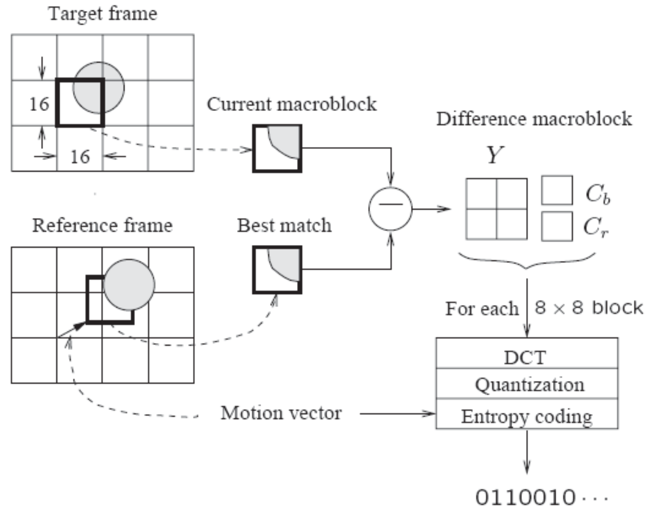
Standard video

H.261 and H.263 encode video frames based on the analysis of the differences with the previous frame

- Only differences are encoded, and the content is rebuilt using comparison
- Motion compensation estimates movements of small portions of the image between subsequent frames, and encodes the difference with the estimation (H.263)

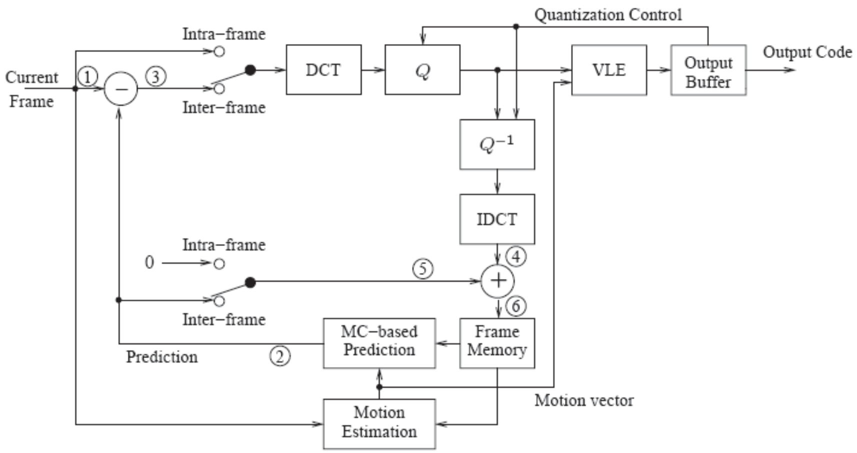


P-frame



Li & Drew, Fundamentals of Multimedia, 2003

H.261 Encoder



Li & Drew, Fundamentals of Multimedia, 2003

H.263 standard

H. 263 (1996): better encoding for low bit-rates

- Images format is variable from 128 x 96 to 1480 x 1152
- Compression algorithm is better and able to encode video flows with a bit-rate lower than 64 Kbps
- Includes several techniques for error corrections

The intraframe encoding works with:

- *PB-frame* to increase frame-rate without increasing bit-rate
- Motion vectors without restrictions
- Advanced prediction: Motion compensation precision reaches ½ pixel

Motion vector

Once calculated the differences between two frames, only direction and movement entity are transmitted (*motion vector*)

H.263 allows the motion vector to refer pixels outside boundaries of the image (*unrestricted motion vector mode*), associating the nearest pixel to the edges of the image, to the one pointed by the MV, external to the image

Integer Pixel Motion Estimation

– Image divided into macroblock (MB) of 16x16 or 8x8
– For each macroblock, a motion vector is calculated, looking for the most similar MB in the previous frame
– Research takes place in the neighborhood of the original position, moving horizontally and vertically for ± 15 pixels, one pixel per time

MPEG, Motion Picture Expert Group (1)

The first MPEG version was released in 1991, and allows compression of a sequence of images and storage on a CD

It allows random video access and fast searches

The compression algorithm is highly complex but strongly asymmetric: it assures a real-time decompression

As H.261 standard, MPEG video works with the *YCbCr* (8 bit) color space, with downsampled chrominance components

Luminance resolution cannot be higher than 768x576 pixels

It does not support interlaced video

MPEG, Motion Picture Expert Group (2)

Different resolutions and refresh frequencies allowed (from 23.98 fps to 60 fps)

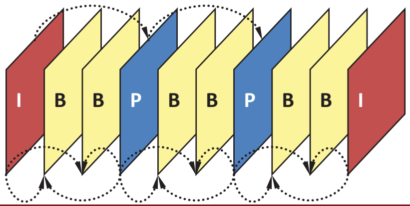
The video information has:

– *Spatial redundancy* → encoding of each single image
  • JPEG encoding
– *Temporal redundancy* → relation between following frames
  • Diversified encoding for each frame

MPEG Compression algorithm

MPEG expands H.261 and H.263 compression algorithms with a more sophisticated scheme of motion estimation

– I frames (*Intra coded frame*) are encoded using a JPEG algorithm, independently but with lower quality
– P frames (*Predictive coded frame*) are encoded based on an estimation referred to the previous I or P frame
– B frames (*Bidirectionally predictive coded frame*) are encoded using two motion estimations related to previous and following frames (bidirectional estimation)



MPEG frames (1)

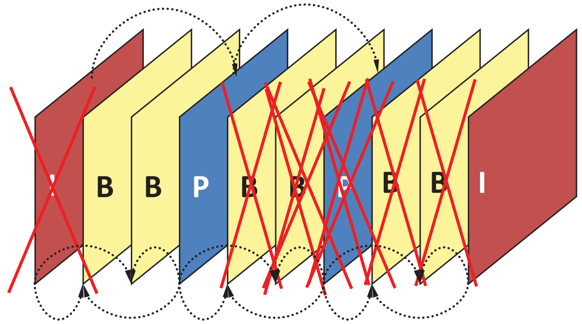
The *Intracoded frames*
- Require higher memory space
- stop errors propagation due to transmission
- Make random access possible

The *Predictive coded frames*
- Differences calculation is based on the absolute value of luminance components
- ''Smaller'' but propagate transmission error

The *Bidirectional predictive coded frames*
- The most complex

More I-frames allow random access in more time points, but increase bit-rate
- IBBPBBPBBIBBPBBPBB…
- There must be one I-frame every 15 frames

MPEG frames (2)



Motion Compensation Prediction (1)

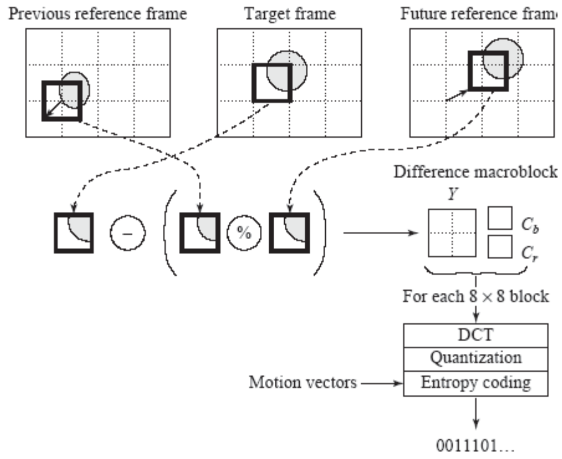
Three phases
- Motion estimation of objects and motion vector creation
- Frames estimation using information collected in the previous phase
- Comparison between the estimated frame and the real one to calculate the error

Only the motion vector and the error estimation are saved

MPEG works with a *half bit* precision:
- Each 16x16 block is expanded, using interpolation, to a virtual 32x32 block
- Search of the new position of the original block inside the macroblock
- Result comes from the interpolation of the virtual 32x32 block with the moved original block
- Research space is ± 512 pixels for half-pixel precision and ± 1024 pixels for whole pixel precision

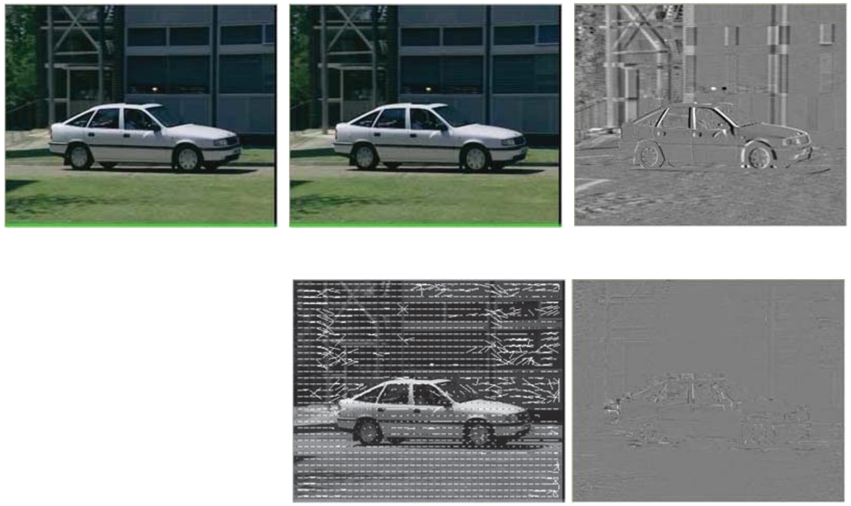
The complexity comes from the research algorithm

Motion Vector for the B-frame



Li & Drew, Fundamentals of Multimedia, 2003

Motion Compensation Prediction (2)



Size of Macroblocks

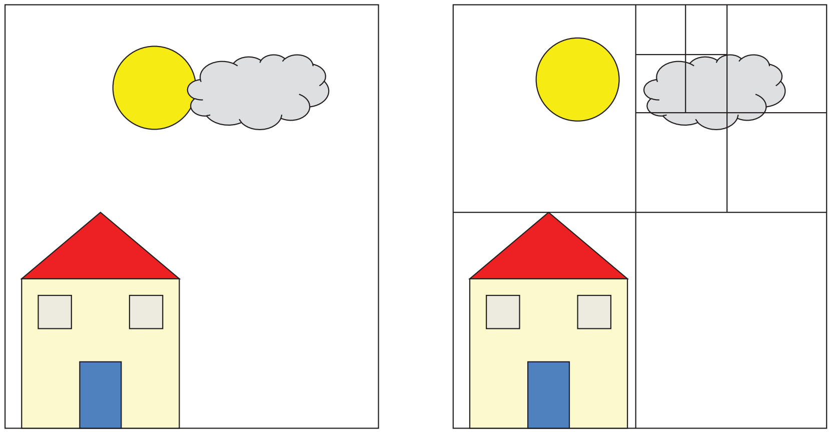
One of the main problems is the size of macroblocks to apply the *motion compensation prediction* algorithm

- Blocks of bigger size → low precision of prediction algorithm
- Blocks of small size → increasing complexity of the algorithm

Blocks with variable dimensions:

- Quad-tree methods
- Binary-tree methods
- H.26L

Quad-Tree Methods



T=1 T=2

Blocks with variable sizes

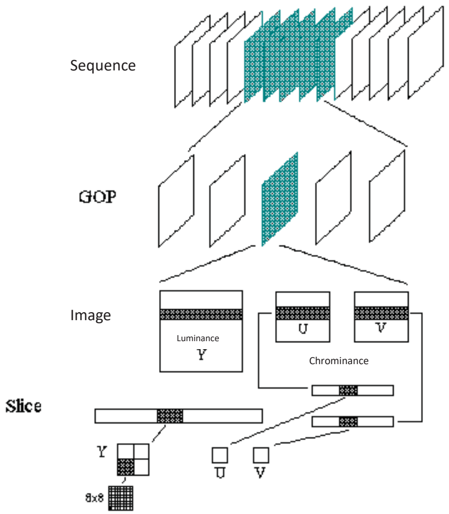
Pros

- Prediction is more accurate
- The more accurate is the prediction, the fewer differences must be encoded

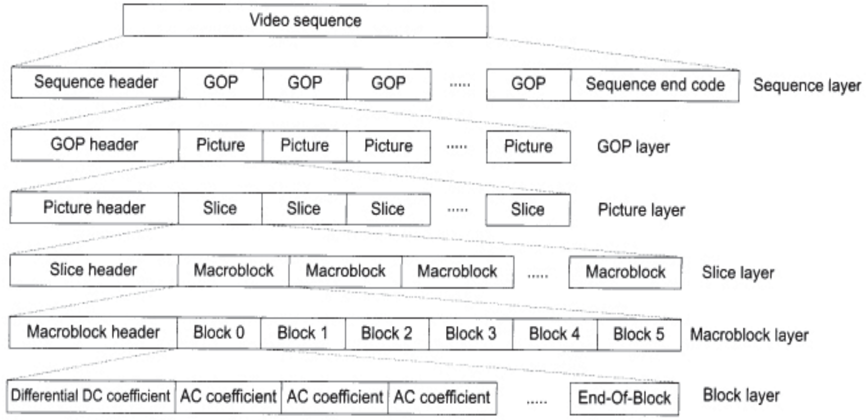
Cons

- Computationally expensive
- The description of the delimitation of the macroblocks (called *regions*) is highly complex

MPEG structure (1)



MPEG structure (2)



Performances & Applications

Considering CIF images (352 x 288), MPEG encoding provides a comparable quality and a compression ratio of about 30:1

It is possible to reach a higher compression ratio but with decreasing quality

Applications:

- video on cd (demo cd, museums,..)
- videogames
- Distance education (but not real-time)
- …

MPEG family (1)

MPEG-1

- CD-ROM video of medium quality
- Quality comparable to the quality of recording on VHS tape
- Decoding do not require specific hardware for standard PCs available on the market

MPEG-2

- High-quality DVD-ROM video (bitrates higher than 4Mbps)
- Quality comparable or higher than old commercial television broadcasting
- Standard format for high-quality *consumer* applications
- It requires specific hardware for decompression or to dedicate the entire PC
- It supports interlaced video

MPEG-2 profiles and levels

Table 11.5: Profiles and Levels in MPEG-2

Level	Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

Table 11.6: Four Levels in the Main Profile of MPEG-2

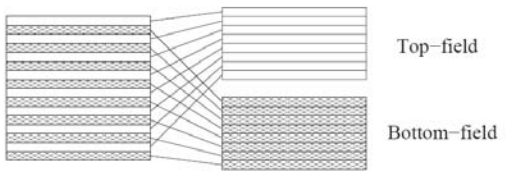
Level	Max Resolution	Max fps	Max Pixels/sec	Max coded Data Rate (Mbps)	Application
High	$1,920 \times 1,152$	60	62.7×10^6	80	film production
High 1440	$1,440 \times 1,152$	60	47.0×10^6	60	consumer HDTV
Main	720×576	30	10.4×10^6	15	studio TV
Low	352×288	30	3.0×10^6	4	consumer tape equiv.

Motion Prediction with MPEG-2 (1)

MPEG-2 supports 5 different motion prediction procedures:

- Frame prediction for frame-picture
- Field prediction for field-picture

Motion Prediction with MPEG-2 (2)



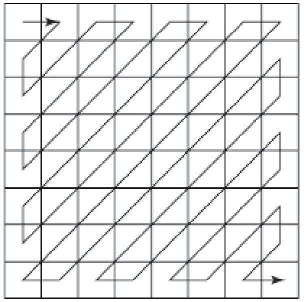
(a) Frame−picture vs. Field−pictures

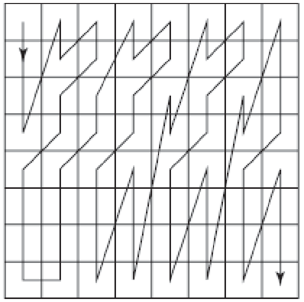
Motion Prediction with MPEG-2 (3)

MPEG-2 supports 5 different motion prediction procedures:

- Frame prediction for frame-picture
- Field prediction for field-picture
- Field prediction for frame-picture
- 16x8 MC for field-pictures
- Dual-prime for P-pictures

Zig-zag scan vs Alternate scan





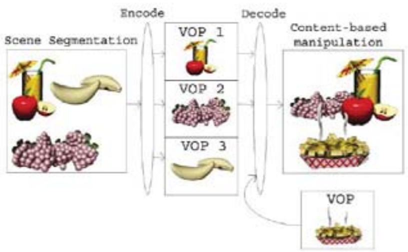
Differences with MPEG-1

- Improved error resistance
- Supports chromatic undersampling 4:2:2 and 4:4:4
- Non-linear quantization
- Higher flexibility of video format

MPEG family (2)

MPEG-4 (1999)

- It allows to integrate video *streams* and objects created independently
- It is optimized for 3 different bitrates: < 64Kbps, 64-384 Kbps, 384-4Mbps
- It allows to index single elements of the scene
- It is intended for applications with complex and interactive multimedia systems
- "…one single technology for playing everywhere…": support for different devices and bandwidths available



MPEG-4

Application examples

UNIVERSITÀ DEGLI STUDI DI PADOVA

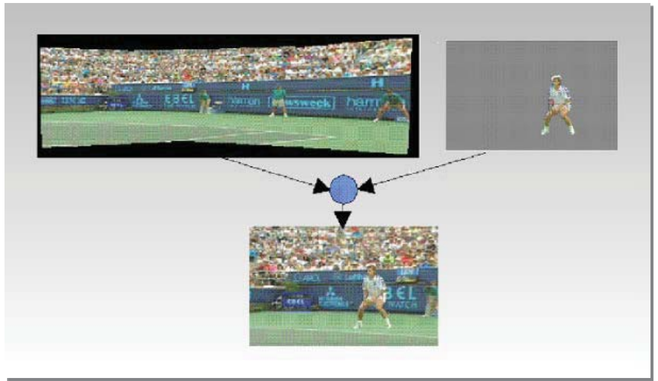
- Video streaming on the Internet
- Videos on smartphones
- Content-based storage and retrieval
- Interactive DVD
- Television production
- Remote monitoring and surveillance
- Infotainment
- Virtual meeting

MPEG-4 Video

UNIVERSITÀ DEGLI STUDI DI PADOVA

An animated scene can be decomposed into two parts

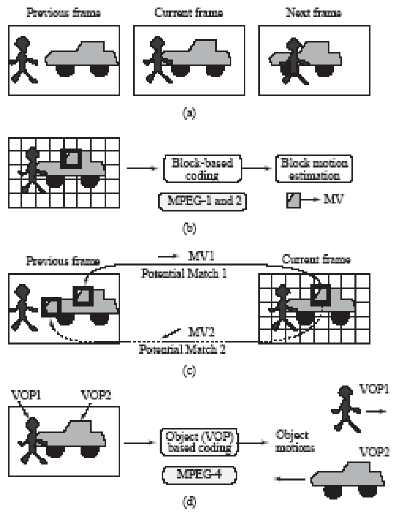
– Background movement is limited to camera movements, therefore it can be encoded as fixed image + coded movements (sprite panorama)



Hierarchical description of a scene with MPEG-4

1. Video-object Sequence (VS): the complete scene; can contain both natural and synthetic objects
2. Video Object (VO): a particular scene object. It can have an arbitrary shape, corresponding to an object or to the background of the scene
3. Video Object Layer (VOL): supports scalable encoding; each VO can have several VOL (scalable encoding) or only one (non-scalable encoding)
4. Group of Video Object Plane (GOV): is an optional level that allows considering sequences of VOP
5. Video Object Plane (VOP): a snapshot of a VO in a particular moment

Frame encoding vs object-oriented encoding



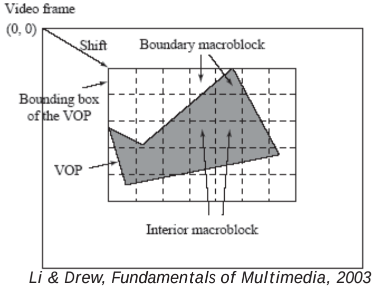
Li & Drew, Fundamentals of Multimedia, 2003

Motion compensation with MPEG-4

The shape of each VOP is arbitrary and must be encoded together with *texture* (using grayscales)

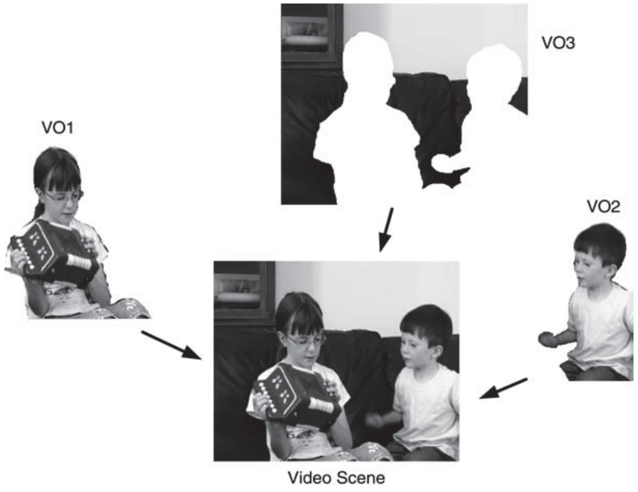
Each VOP is divided into 16x16 blocks, and the motion vector for the global object is calculated

To apply the DCT (that requires squared matrixes), MC uses padding



Li & Drew, Fundamentals of Multimedia, 2003

Video Object composition with MPEG4



Masks

Masks - Detail

Università degli Studi di Padova

Composition with different background

Gray-scale mask

MPEG family (3)

MPEG-7

- Defines how to represent a content descriptor in a standard way
- Associates to objects of a multimedia application a set of descriptors to allow classification and content search
- Defines generic containers for objects of different media of different standards
- Combines descriptions automatically extracted from media with descriptions provided by a human user
- Intended for *information retrieval*
- Defined as standard in September 2001

- *Does not define* how to extract content descriptions and how to use those descriptions

Characteristics and descriptors

Color GoF/GoP Color Scalable Color Color Layout Color Structure Dominant Color	Texture Homogeneous Text. Texture Browsing Edge histogram	Shape Region Shape Contour Shape 3D Shape 2D-3D Multiple View
Motion Camera Motion Motion Trajectory Parametric Motion Motion Activity	Localization Bounding Box Region Locator Spatio-Temporal Locator	Other Face Recognition

MPEG family (4)

- MPEG – 21 (~ 2003)
  - Developed for digital content protection
  - Content description plus rights of whom created the contents
  - Must provide an interface to make media usage easier (search, caching techniques, etc.)