T2 – MPEG

Universitat Pompeu Fabra Barcelona

Some previous concepts that YOU SHOULD ALREADY KNOW

Pixel

Resolution

Aspect Ratio

Video

Universitat Pompeu Fabra Barcelona

Pixel

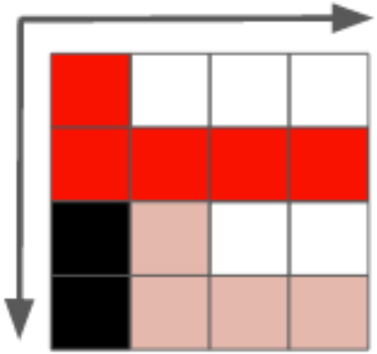
Pixel (picture element)

Represents the intensity (usually a numeric value) of a given color.

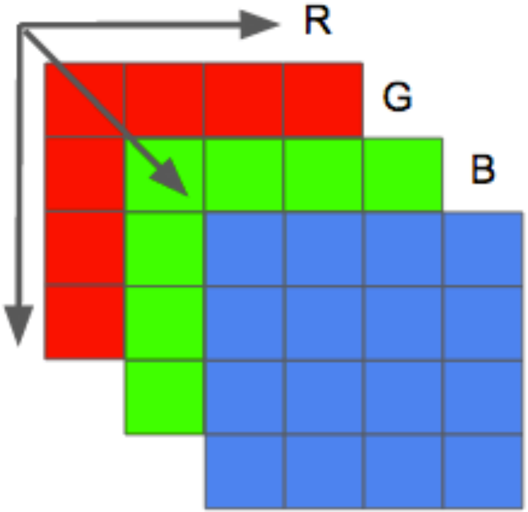
For example:

Red Pixel: 0 of green + 0 of blue + maximum of red

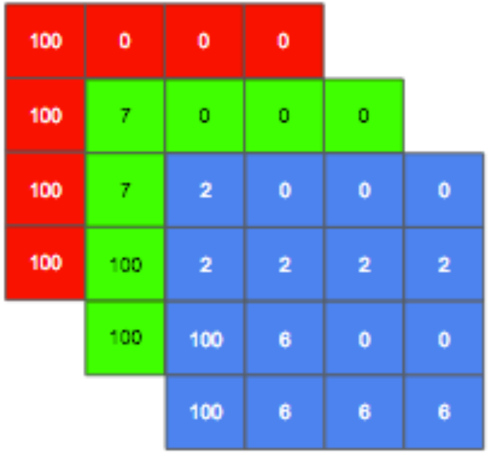
Pink Pixel: 192 of green + 203 of blue + 255 of red



2D



3D

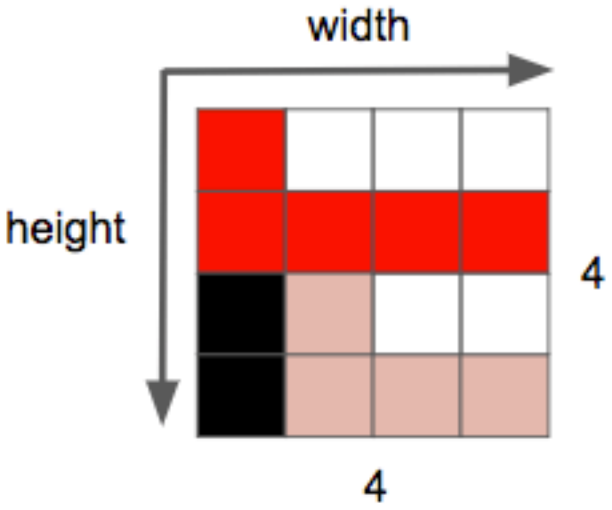


color intensity

Resolution

Universitat Pompeu Fabra Barcelona

Resolution: number of pixels in 1 dimension
Example: 360x240



Aspect Ratio:

Describes the proportional relationship between width and height of an image or pixel

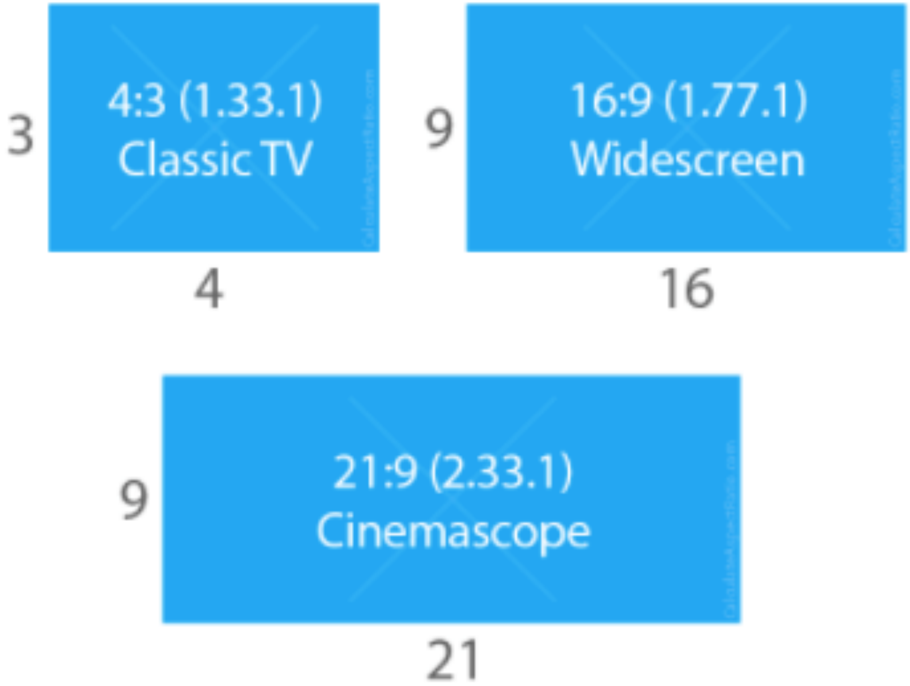
Universitat Pompeu Fabra Barcelona

Display Aspect Ratio (DAR):

Describes the proportional relationship between width and height (or shape) of an image or video

Universitat Pompeu Fabra Barcelona

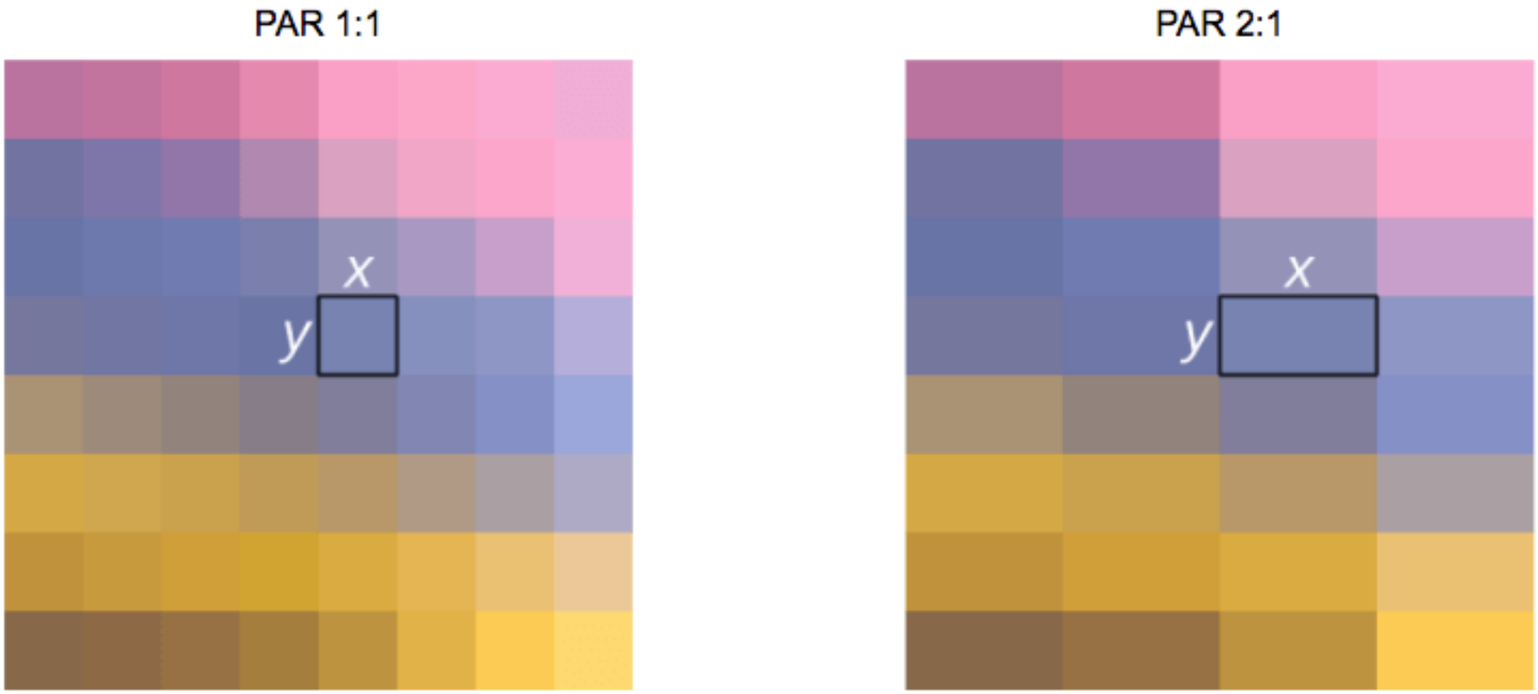
Display Aspect Ratio



Pixel Aspect Ratio (PAR):

Describes the proportional relationship between width and height (or shape) of individual pixels

Pixel Aspect Ratio (PAR):

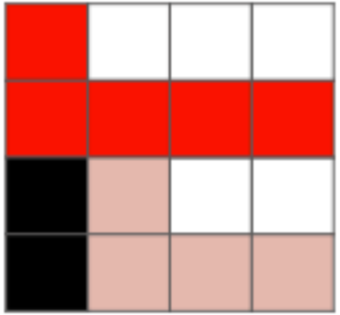


VIDEO

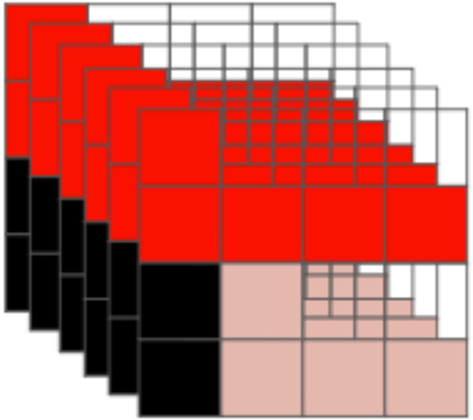
We can define a video as a succession of *n* frames in time which can be seen as another dimension, *n* is the frame rate or frames per second (FPS).

Universitat Pompeu Fabra Barcelona

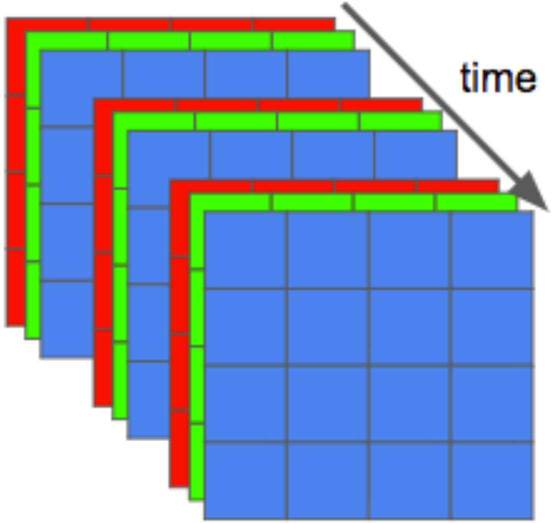
VIDEO



a single frame



30 frames per sec
(FPS)
framerate



4D

Bit rate

The number of bits per second needed to show a video is its bit rate.

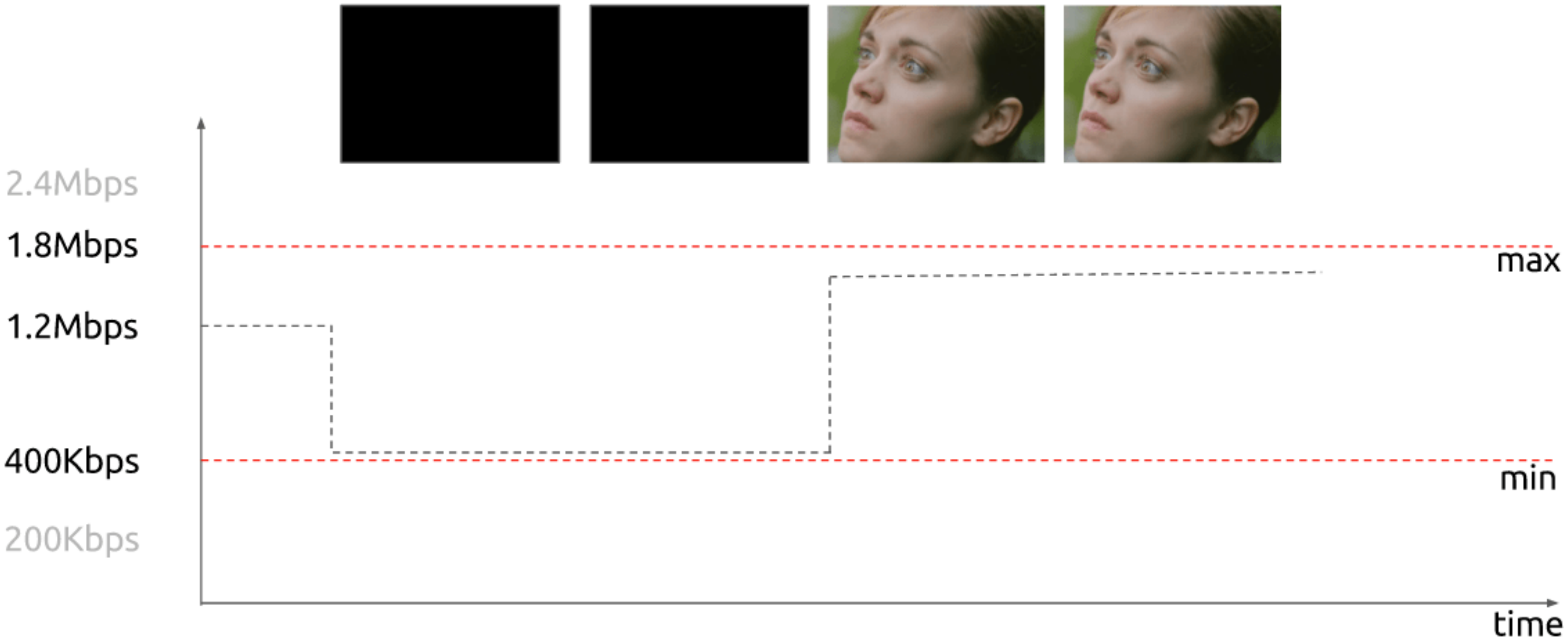
bit rate = width * height * bit depth * frames per second

For example:
Video with 30 fps, 24 bits/pixel, 480x240 resolution

Will need 82,944,000 bits per second or 82.944 Mbps (30x480x240x24) if we don't employ any kind of compression.

- When the bit rate is nearly constant it's called constant bit rate (CBR)

- When the bit rate is variable it’s called variable bit rate (VBR)


2.4Mbps
1.8Mbps
1.2Mbps
400Kbps
200Kbps
max
min
time

Universitat Pompeu Fabra Barcelona

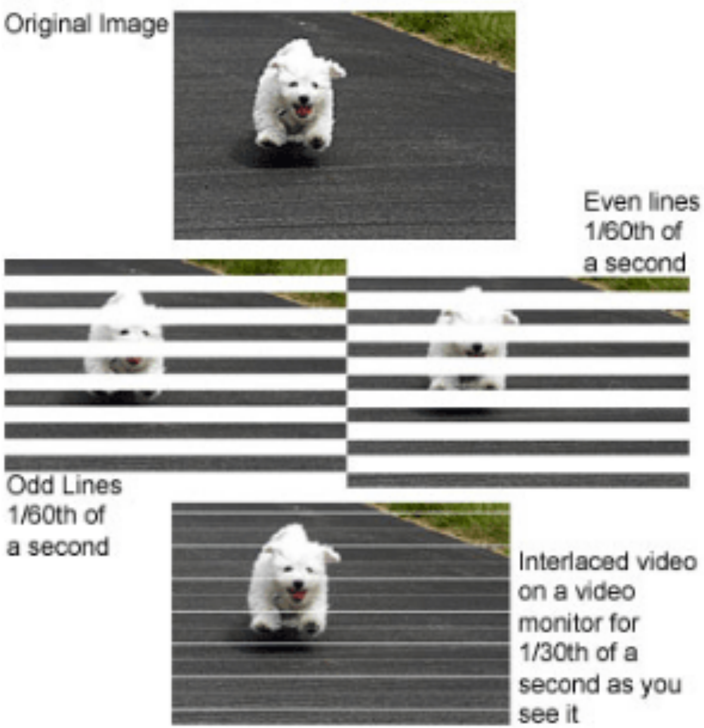
Back in the days

Interlaced video: first technique for doubling the perceived frame rate of a video display without consuming extra bandwidth.

It sends half of the screen in 1 frame and the other half in the next frame

Nowadays

Progresive scan video: way of displaying, storing, or transmitting moving images in which all the lines of each frame are drawn in sequence.



Interlaced scan image

progressive scan image

Universitat Pompeu Fabra Barcelona

MPEG

Moving
Picture
Experts
Group

Universitat Pompeu Fabra Barcelona

Established by ISO in 1988

Standard MPEG created in 1993

It's a lossy compression method

It is designed to compress VHS-quality raw digital video and CD audio down to 1.5 Mbit/s (26:1 and 6:1 compression ratios respectively)

Universitat Pompeu Fabra Barcelona

(This was a VHS)

(360x240 pixels only)

Universitat Pompeu Fabra Barcelona

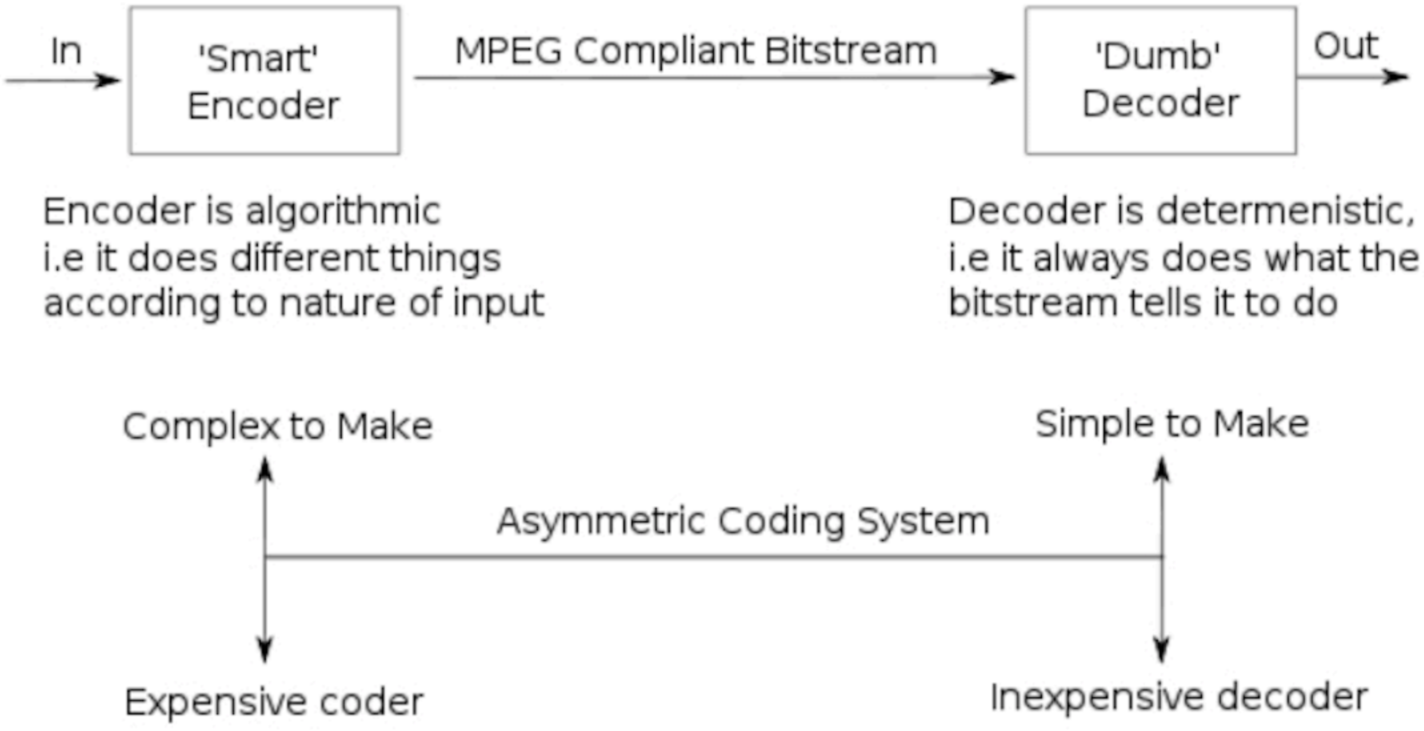
It's also a standard format for compressing and storing digital video & audio

MPEG1: 352x240 @ 30 fps (slightly below VCR video)

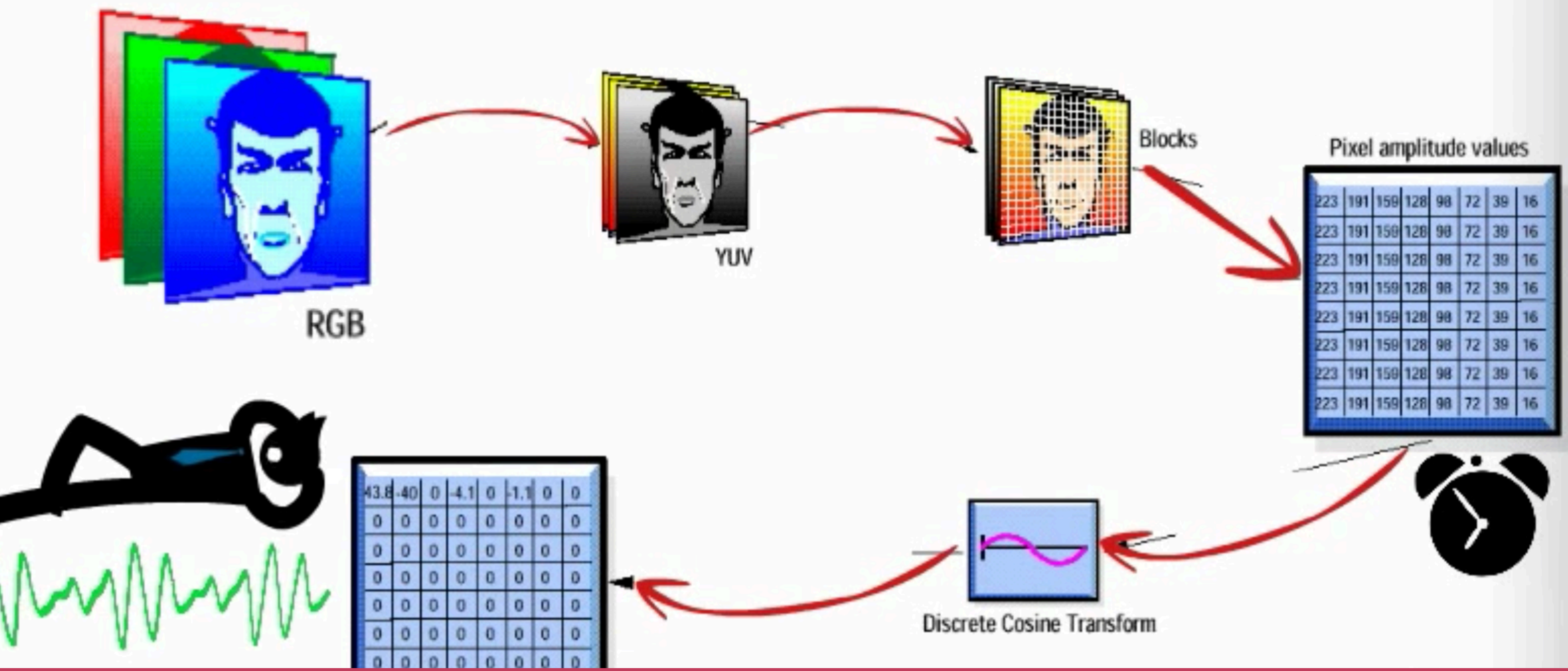
It allowed the Video CD exist

VIDEO CD

Universitat Pompeu Fabra Barcelona

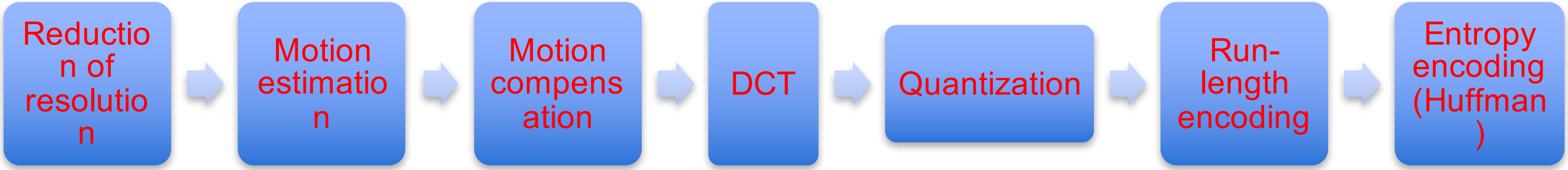


Universitat Pompeu Fabra Barcelona



Universitat Pompeu Fabra Barcelona

Workflow:


Reductio n of resolutio n
Motion estimatio n
Motion compens ation
DCT
Quantization
Run-length encoding
Entropy encoding (Huffman)

Reduction of resolution:

- Video downsize (if required)
- RGB -> YUV (as we saw in JPG)

Universitat Pompeu Fabra Barcelona

Y: Luminance

Cb: Chroma blue difference

Cr: Chroma red difference

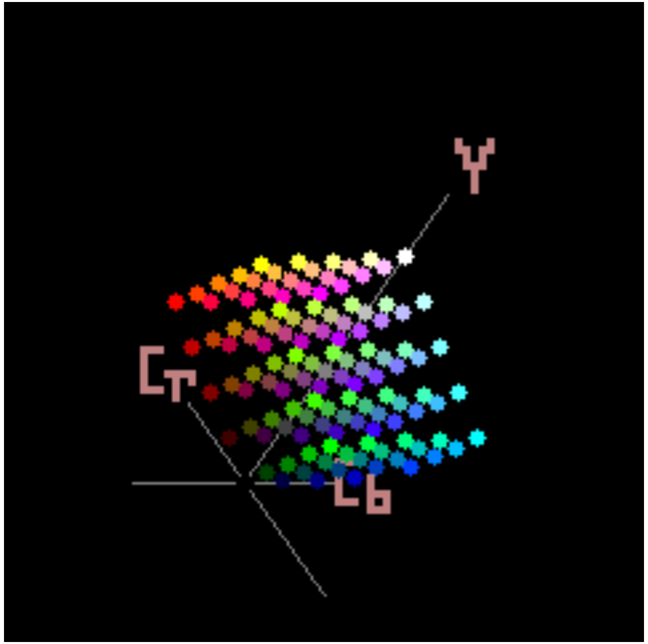
Universitat Pompeu Fabra Barcelona

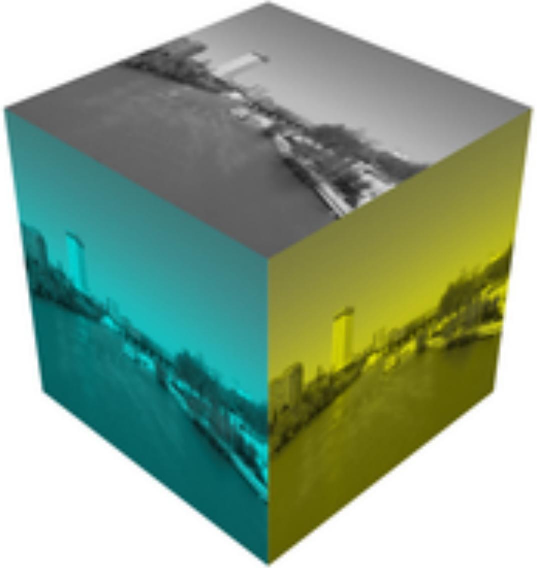
$$Y' = K_R \cdot R' + K_G \cdot G' + K_B \cdot B'$$

$$P_B = \frac{1}{2} \cdot \frac{B' - Y'}{1 - K_B}$$

$$P_R = \frac{1}{2} \cdot \frac{R' - Y'}{1 - K_R}$$

Universitat Pompeu Fabra Barcelona





Universitat Pompeu Fabra Barcelona

YCbCr formula from RGB

```
Y  =         0,257 * R + 0,504 * G + 0,098 * B + 16
Cb = U = -0,148 * R - 0,291 * G + 0,439 * B + 128
Cr = V =  0,439 * R - 0,368 * G - 0,071 * B + 128
```

```
B = 1,164 * (Y - 16)                          + 2,018 * (U - 128)
G = 1,164 * (Y - 16) - 0,813 * (V - 128) - 0,391 * (U - 128)
R = 1,164 * (Y - 16) + 1,596 * (V - 128)
```

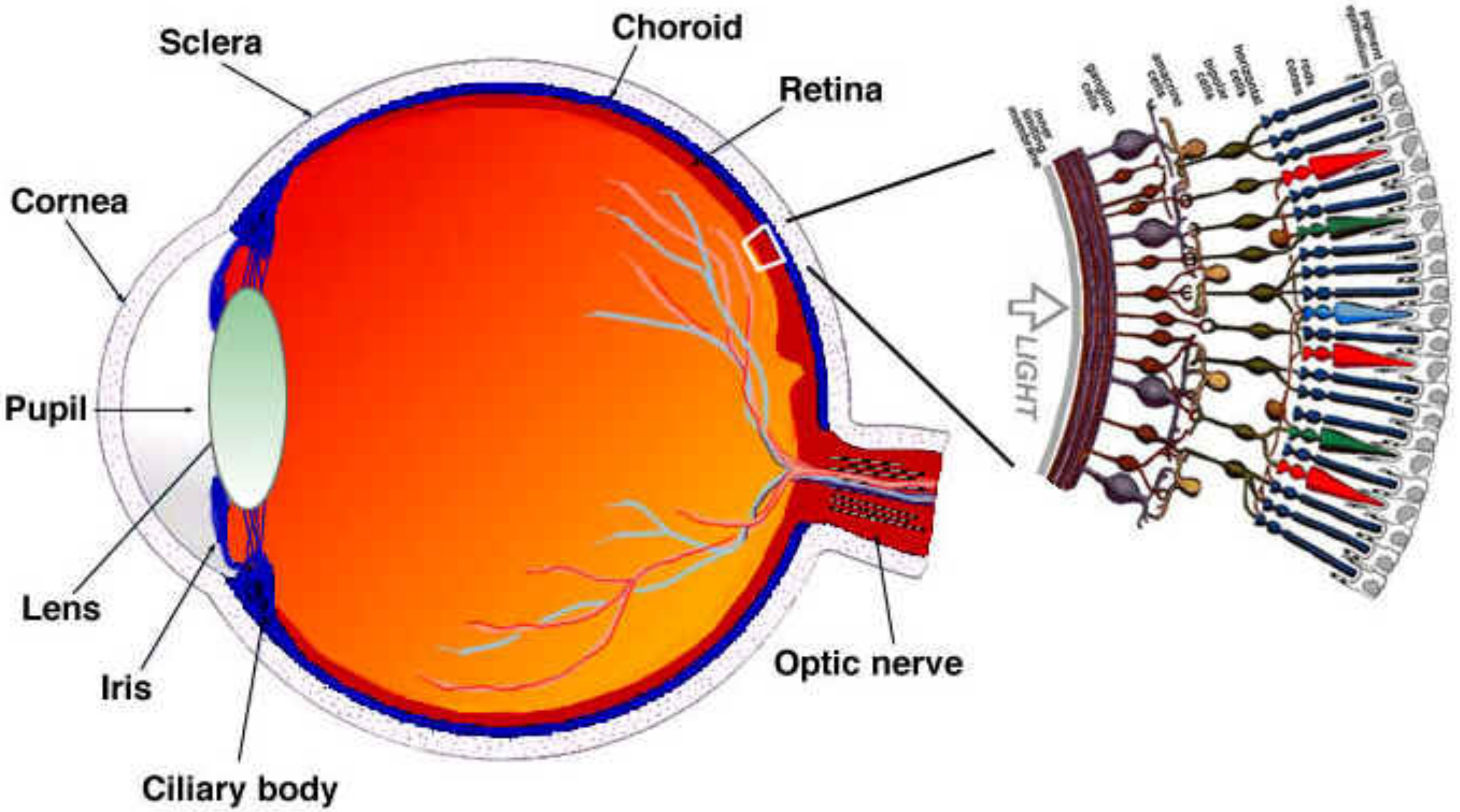
Universitat Pompeu Fabra Barcelona

Chroma subsampling

It's the practice of encoding images by implementing less resolution for chroma information than for luma information, taking advantage of the human visual system's lower acuity for color differences than for luminance.

Universitat Pompeu Fabra Barcelona

Simplistic explanation of how our eyes work The eye is a complex organ, it is composed of many parts but we are mostly interested in the cones and rods cells. The eye contains about 120 million rod cells and 6 million cone cells


Sclera
Choroid
Retina
Cornea
Pupil
Lens
Iris
Ciliary body
Optic nerve
LIGHT

To oversimplify, let's try to put colors and brightness in the eye's parts function. The rod cells are mostly responsible for brightness while the cone cells are responsible for color, there are three types of cones, each with different pigment, namely: S-cones (Blue), M-cones (Green) and L-cones (Red).

Since we have many more rod cells (brightness) than cone cells (color), one can infer that we are more capable of distinguishing dark and light than colors.

Y (luma)

U (chroma blue)

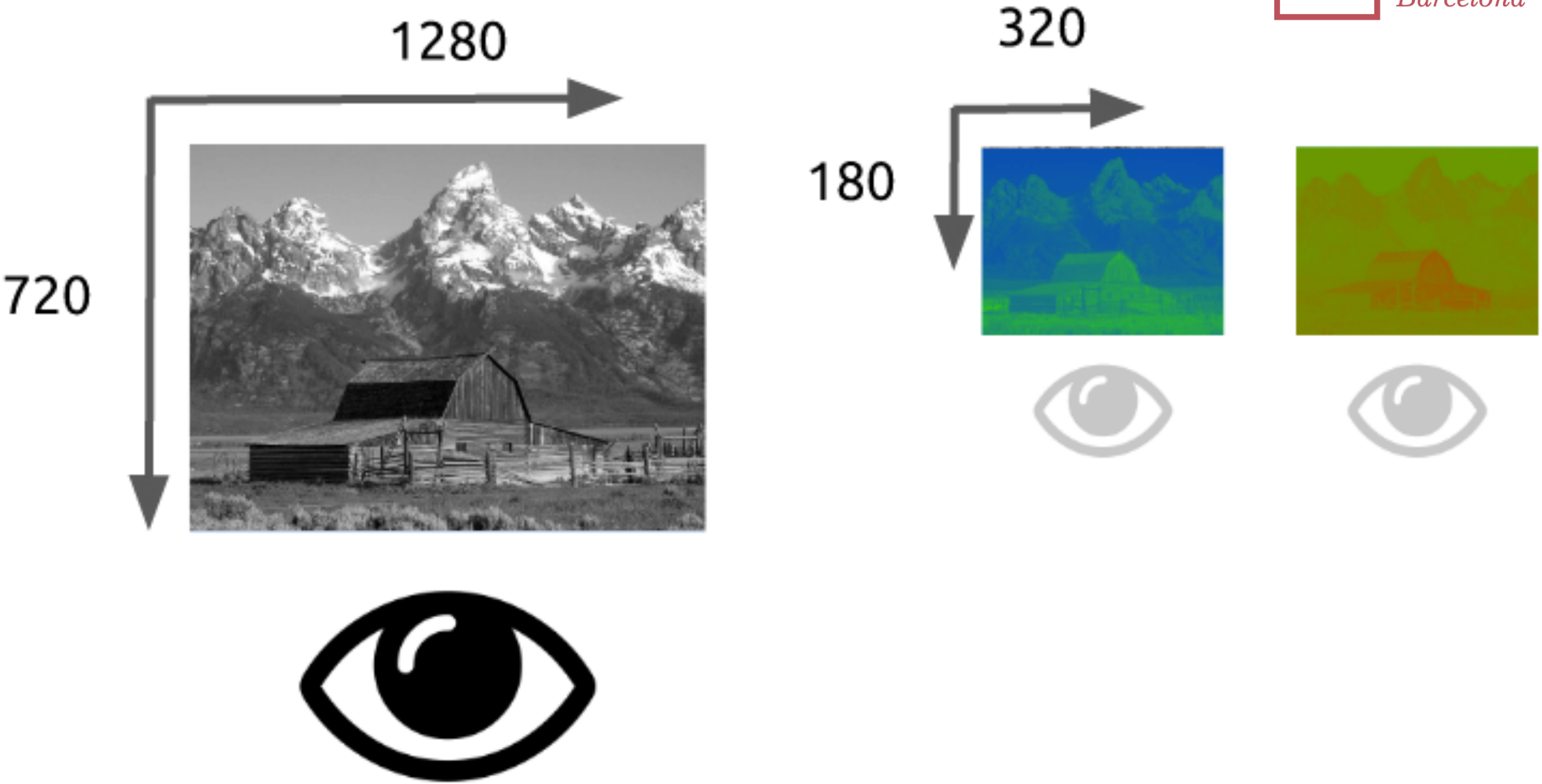
V (chroma red)

Universitat Pompeu Fabra Barcelona

Chroma subsampling is the technique of encoding images using less resolution for chroma than for luma.

Universitat Pompeu Fabra Barcelona

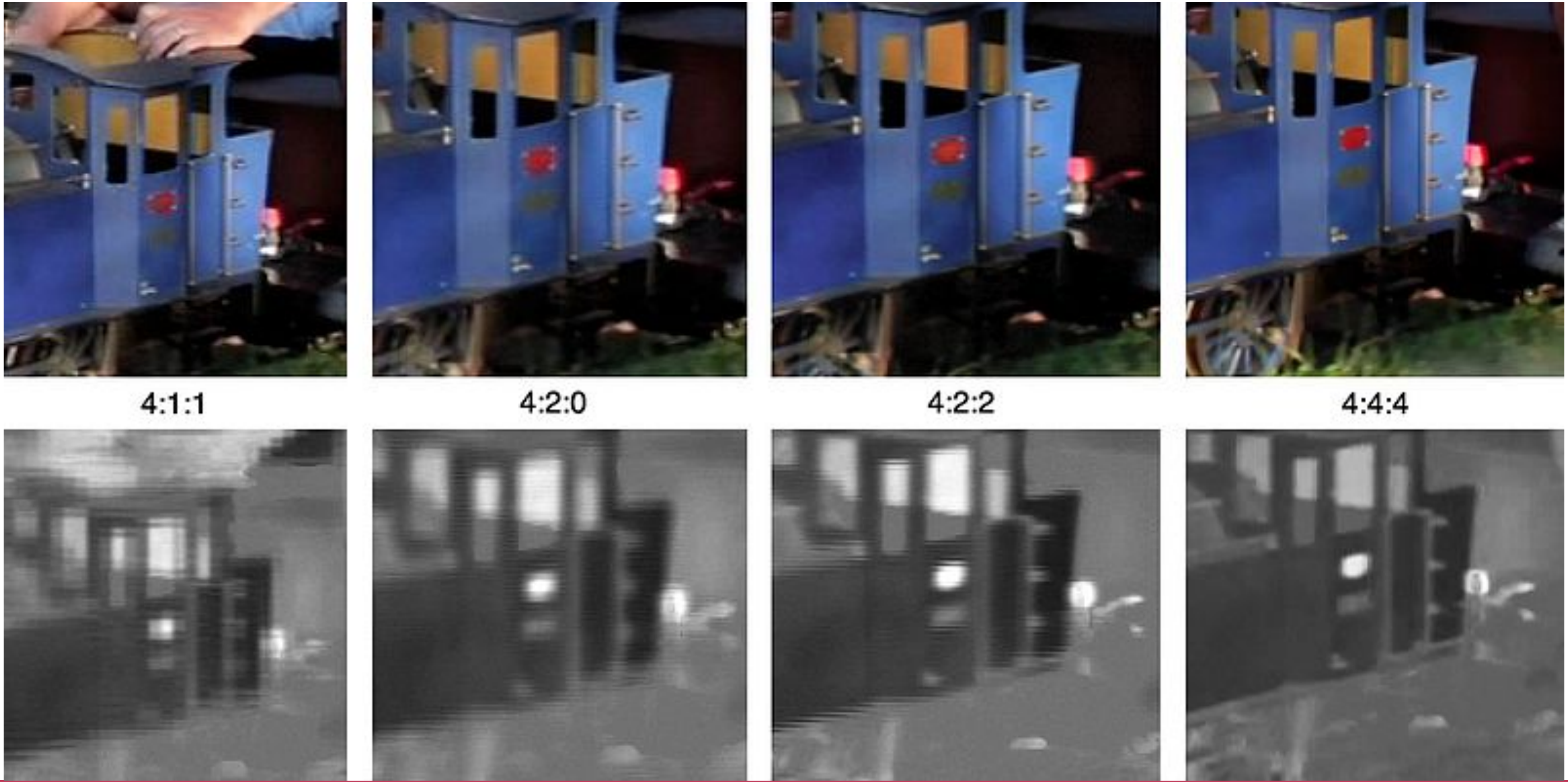
These schemas are known as subsampling systems and are expressed as a 3 part ratio - a:x:y which defines the chroma resolution in relation to a *a x 2* block of luma pixels.


1280
720
320
180

Universitat Pompeu Fabra Barcelona

- *a* is the horizontal sampling reference (usually 4)
- *x* is the number of chroma samples in the first row of a pixels (horizontal resolution in relation to a)
- *y* is the number of changes of chroma samples between the first and seconds rows of a pixels.

Common schemes used in modern codecs are: 4:4:4 (no subsampling), 4:2:2, 4:1:1, 4:2:0, 4:1:0 and 3:1:1.


4:1:1
4:2:0
4:2:2
4:4:4

Motion estimation

For explaining this, first we go with the 3 type of frames we find and how they work in MPEG

Suppose we have a movie with 30fps, here are its first 4 frames.

We can see lots of repetitions within frames like the blue background, it doesn't change from frame 0 to frame 3. To tackle this problem, we can abstractly categorize them as three types of frames.

Universitat Pompeu Fabra Barcelona

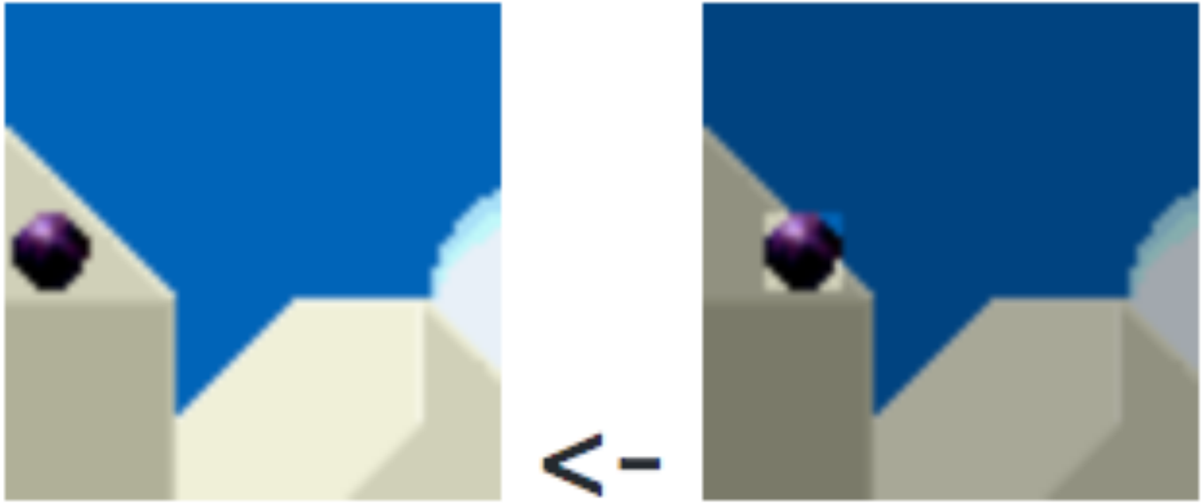
I Frame (intra, keyframe)

An I-frame (reference, keyframe, intra) is a self-contained frame. It doesn't rely on anything to be rendered, an I-frame looks similar to a static photo. The first frame is usually an I-frame but we'll see I-frames inserted regularly among other types of frames.

P Frame (predicted)

A P-frame takes advantage of the fact that almost always the current picture can be rendered using the previous frame. For instance, in the second frame, the only change was the ball that moved forward. We can rebuild frame 1, only using the difference and referencing to the previous frame.

P Frame (predicted)

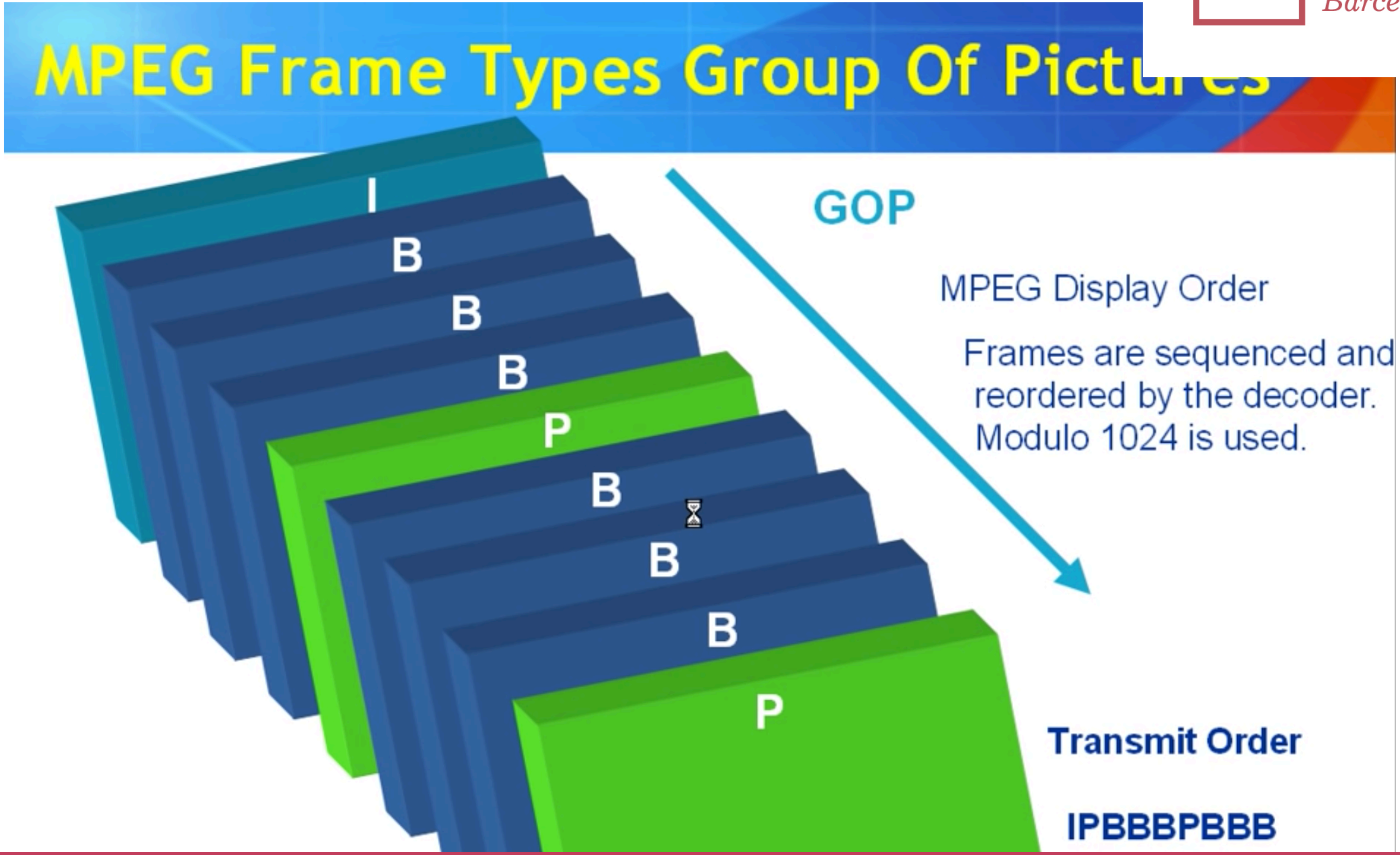


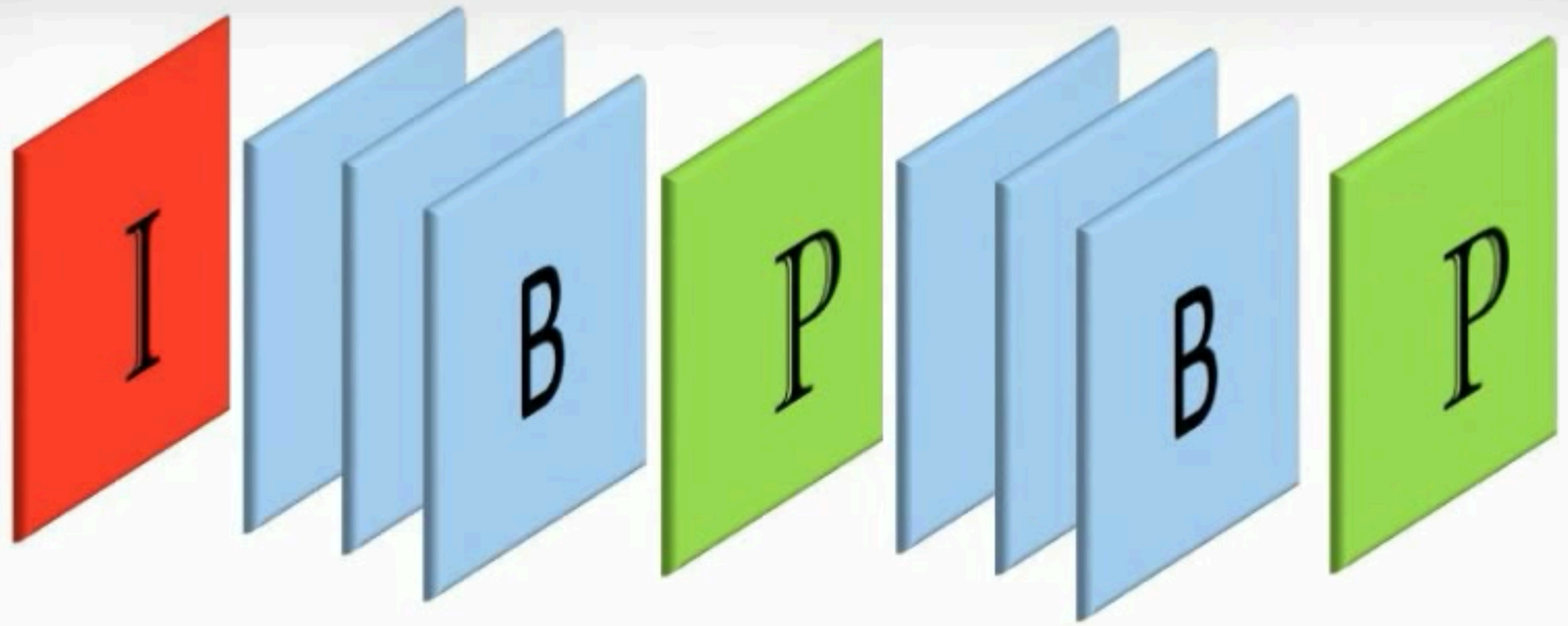
Universitat Pompeu Fabra Barcelona

B Frame (Bi-directional predicted frame)

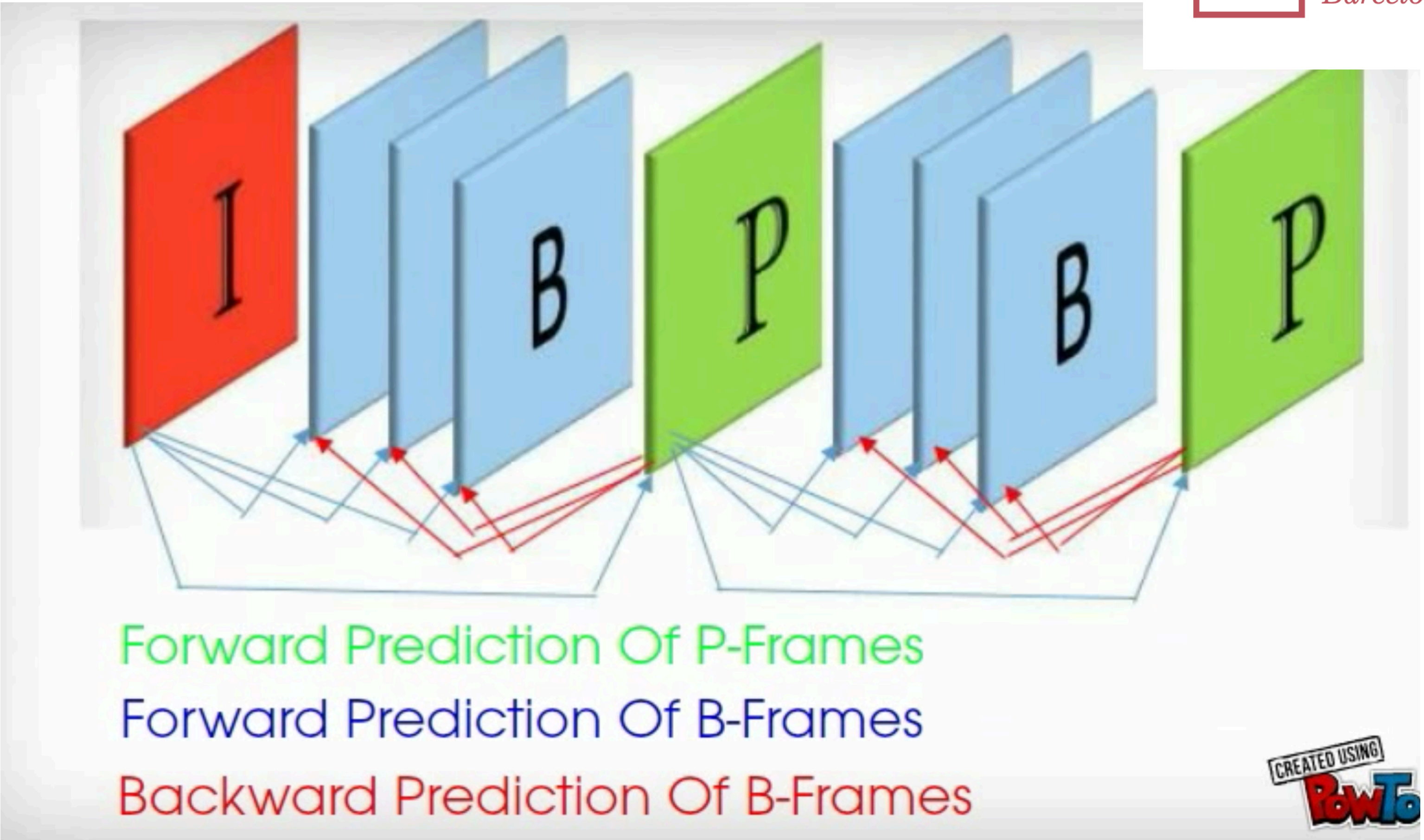
Using the same technique, we can not only move to the previous frame but also move to the following one. We'll be predicting the past and the future frame.

B Frame (Bi-directional predicted frame)

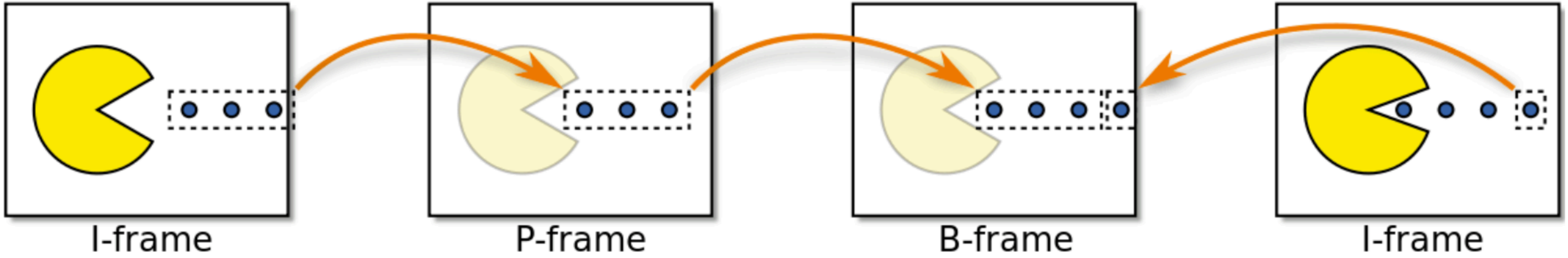

MPEG Frame Types Group Of Pictu
I
B
B
B
P
B
B
B
P
GOP
MPEG Display Order
Frames are sequenced and
reordered by the decoder.
Modulo 1024 is used.
Transmit Order
IPBBBPBBB


I
B
P
B
P

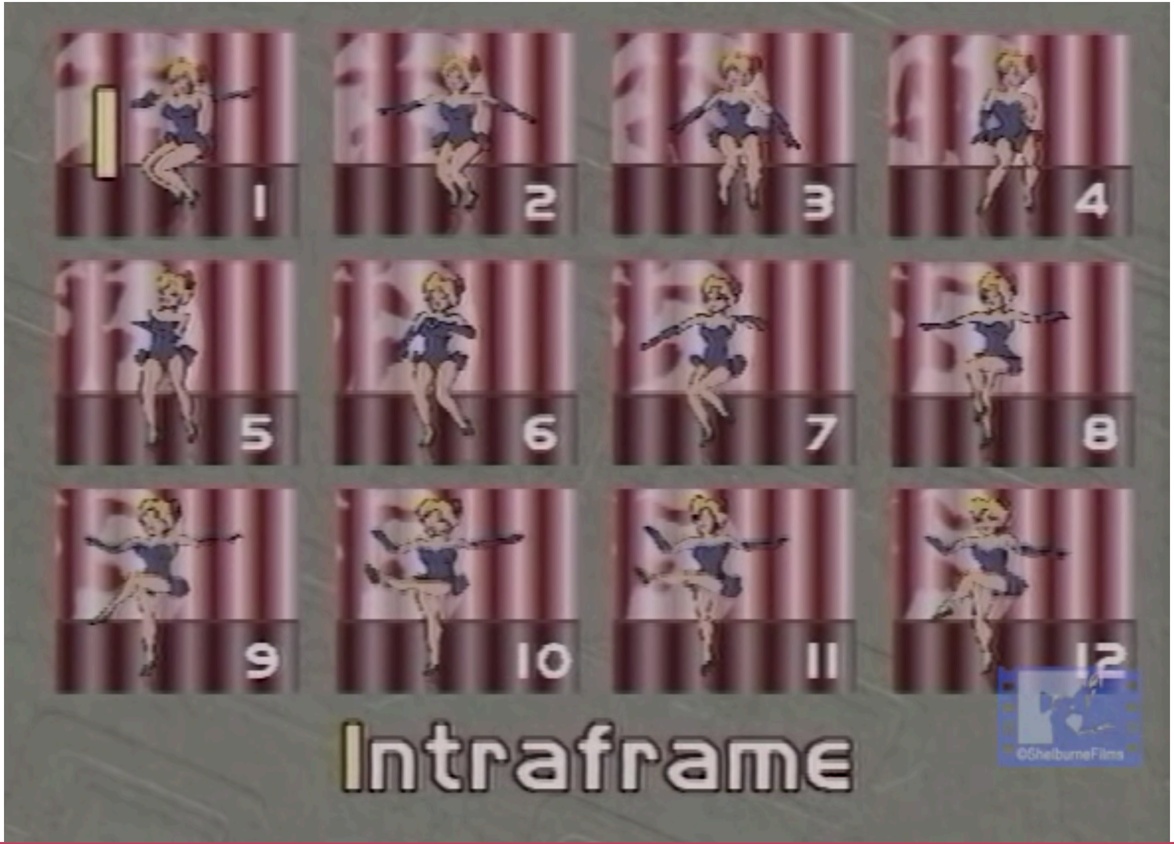
Universitat Pompeu Fabra Barcelona



These frames types are used to **provide better compression**. We'll look how this happens in the next section, but for now we can think of **I-frame as expensive while P-frame is cheaper but the cheapest is the B-frame.**



Universitat Pompeu Fabra Barcelona




I
1
2
3
4
P
5
6
7
8
P
9
10
11
12
Predicted frame

Universitat Pompeu Fabra Barcelona

Universitat Pompeu Fabra Barcelona

What happens when there’s info lost? We can’t recover the frames correctly (specially I Frames)

Universitat Pompeu Fabra

Motion estimation:

Motion estimation examines the movement of objects in an image sequence to try to obtain vectors representing the estimated motion

Motion estimation:

It’s to generate the motion vector(s).

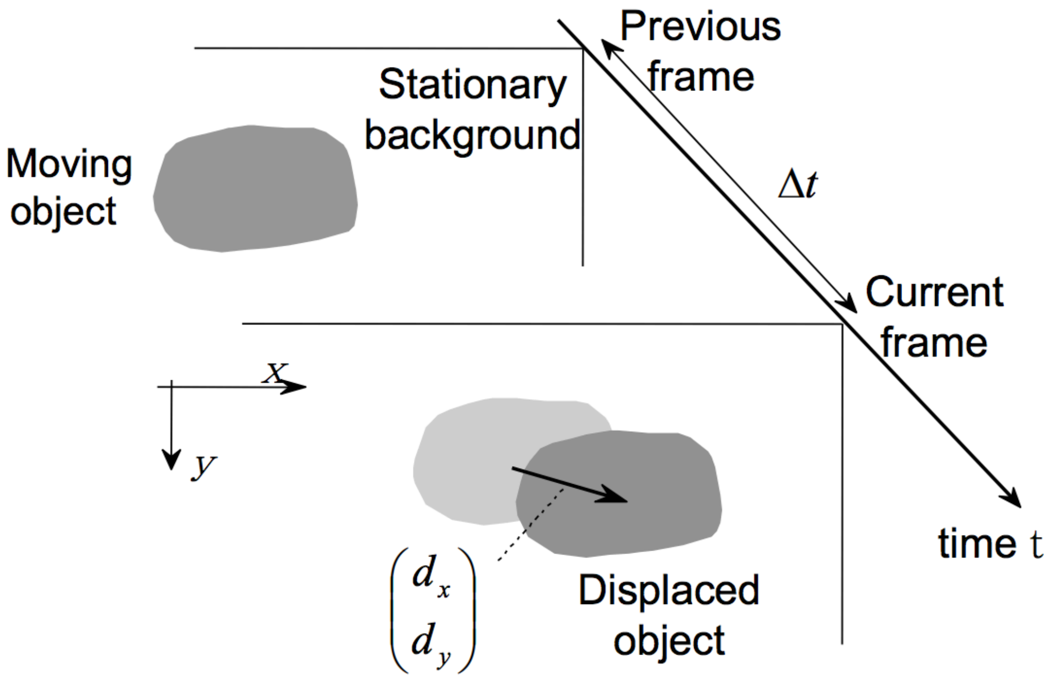
It's in the encoder only, since decoder only consumes the motion vector data.

Universitat Pompeu Fabra Barcelona

IMPORTANT!!! Difference between motion estimation & motion compensation:

Motion compensation is the use of the motion estimation information to achieve compression. If you can describe the motion, then you only have to describe the changes that occur after compensating for that motion.

WATCH OUT because some authors/sources mix both concepts, or consider 'motion estimation' as a part of the motion compensation process



Prediction for the luminance signal $s[x,y,t]$ within the moving object:

$$\hat{s}[x,y,t] = s'(x-d_x, y-d_y, t-\Delta t)$$

Universitat Pompeu Fabra Barcelona

Representation of Video Signal

Decoded video signal is given as

$$s'[x,y,t] = \hat{s}[x,y,t] + u'[x,y,t]$$

Motion-compensated prediction signal

$$\hat{s}[x,y,t] = s'(x-d_x, y-d_y, t-\Delta t)$$

$$d_x = \sum_{i=0}^{N-1} a_i \cdot \varphi_i(x,y), \quad d_y = \sum_{i=0}^{N-1} b_i \cdot \varphi_i(x,y)$$

Transmitted motion parameters

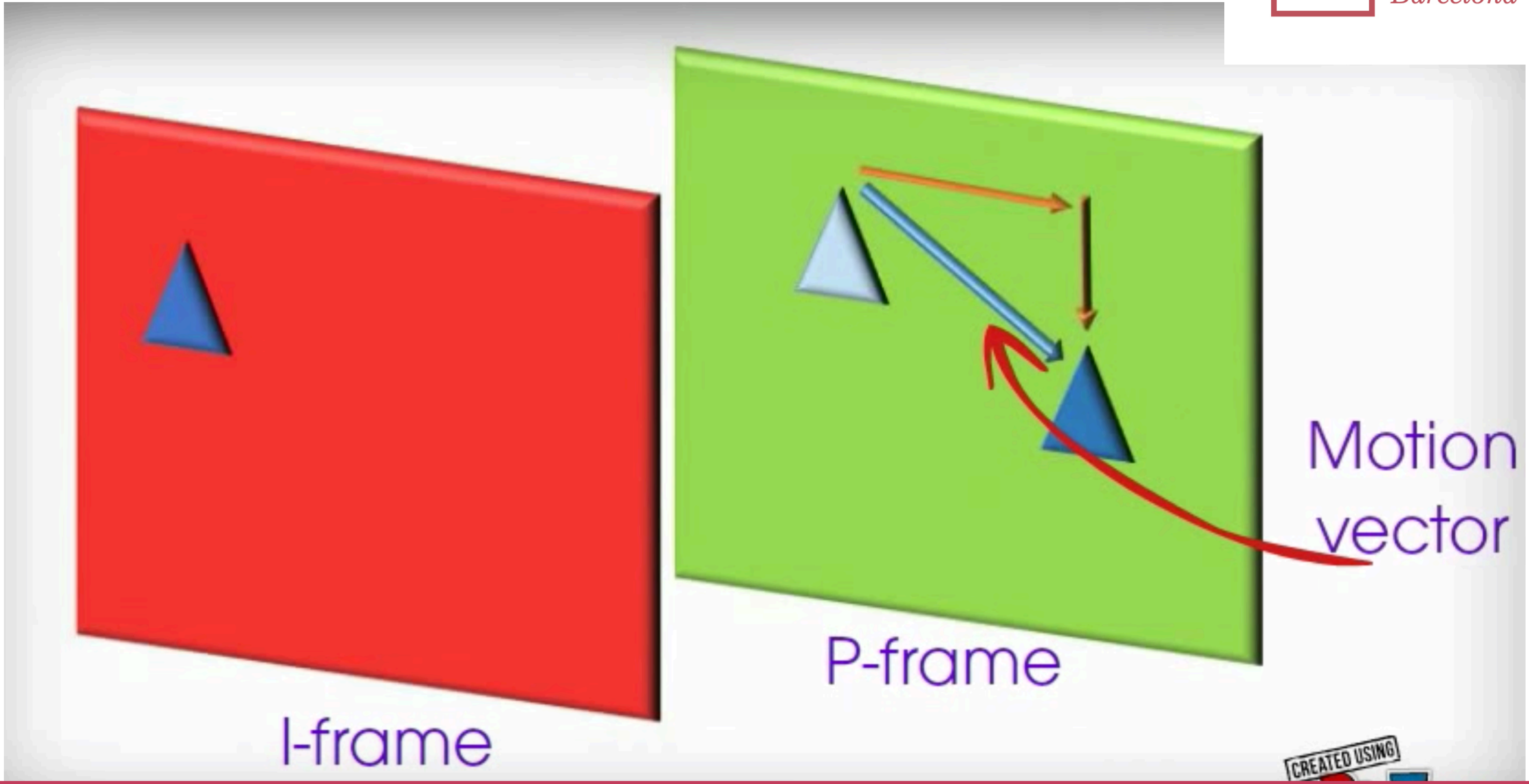
$$R_m = f(\mathbf{a},\mathbf{b}), \quad \mathbf{a} = (a_0, \ldots), \quad \mathbf{b} = (b_0, \ldots)$$

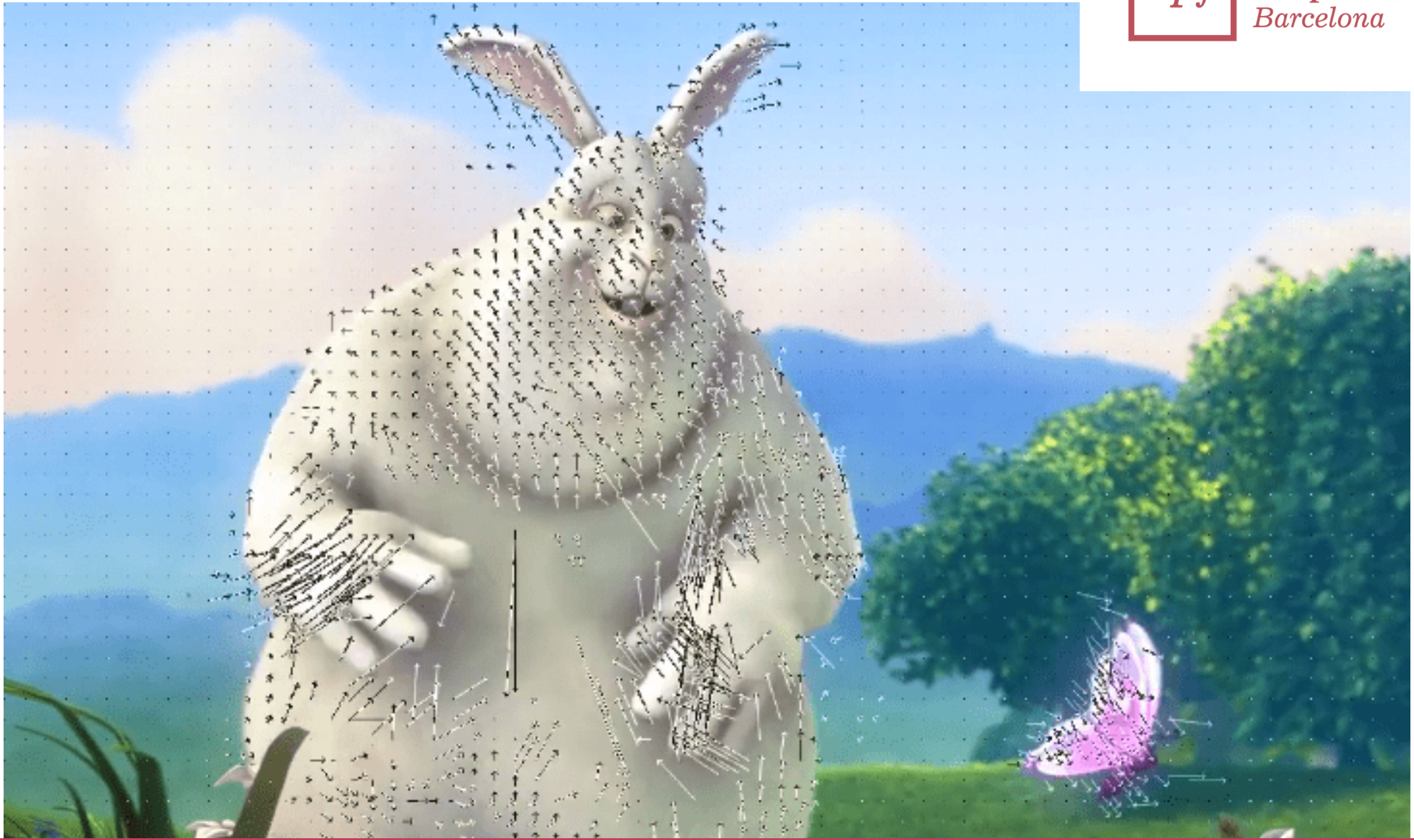
Prediction residual signal

$$u'(x,y) = \sum_{j=0}^{M-1} c_j \cdot \phi_j(x,y)$$

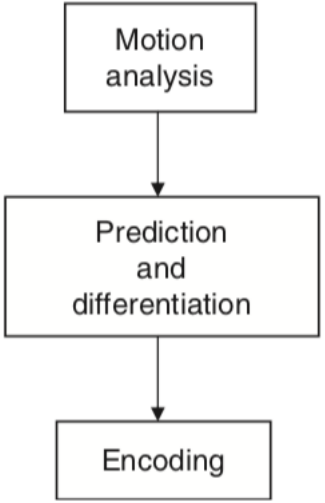
Transmitted residual parameters

$$R_u = f(\mathbf{c}), \quad \mathbf{c} = (c_0, \ldots)$$


Motion
vector
P-frame
I-frame
CREATED USING



Motion compensation:



Motion compensation:

1 - motion analysis stage

Displacement vectors for every pixel or a set of pixels in image planes from sequential images are estimated (REMEMBER MOTION ESTIMATION)

Motion compensation:

2 – Prediction and differentiation

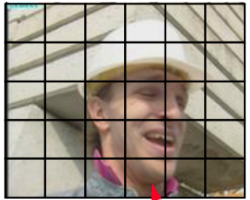
Present frame is predicted by using estimated motion vectors and the previous frame. The prediction error is then calculated

Motion-Compensated Prediction: Example

Frame 1 $s[x,y,t-1]$ (previous)

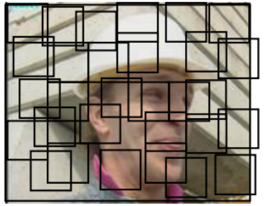
Frame 2 $s[x,y,t]$ (current)

Partition of frame 2 into blocks (schematic)



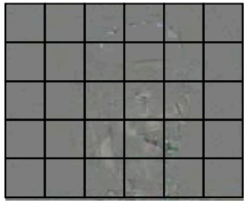
Size of Blocks

Accuracy of Motion Vectors



Referenced blocks in frame 1

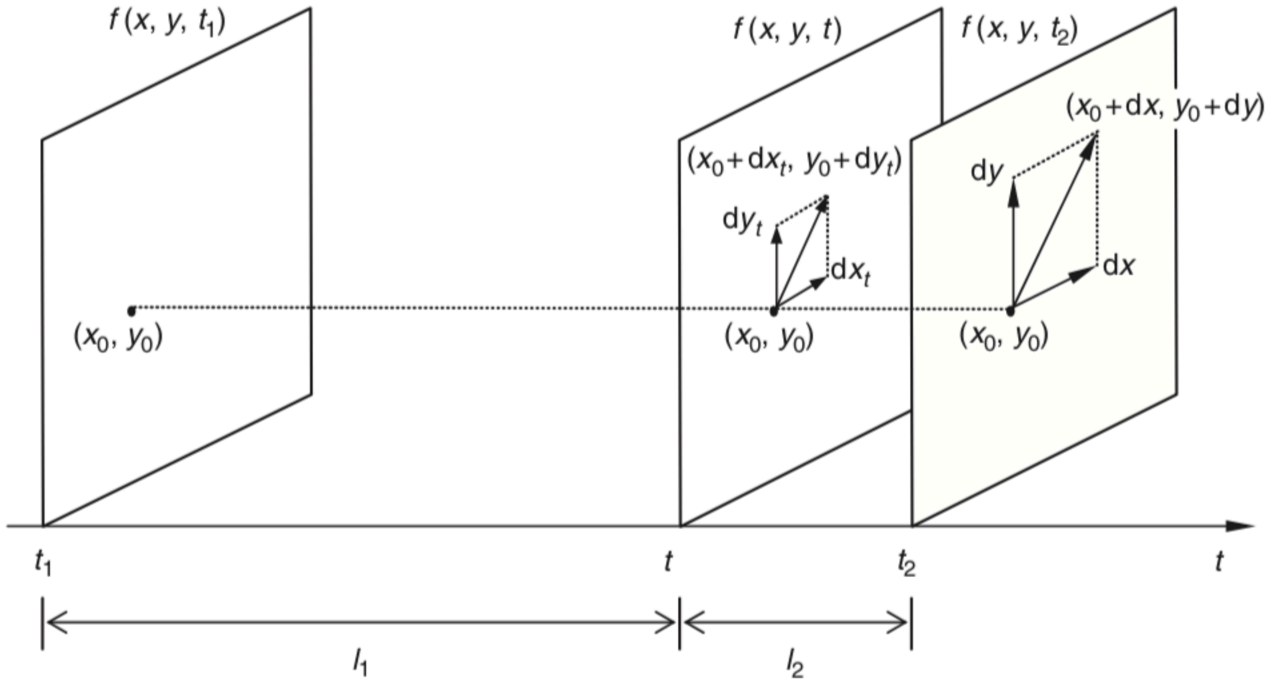
Frame 2 with displacement vectors



Difference between motion-compensated prediction and current frame $u[x,y,t]$

Universitat Pompeu Fabra Barcelona

Motion compensation interpolation



Universitat Pompeu Fabra Barcelona

Block matching

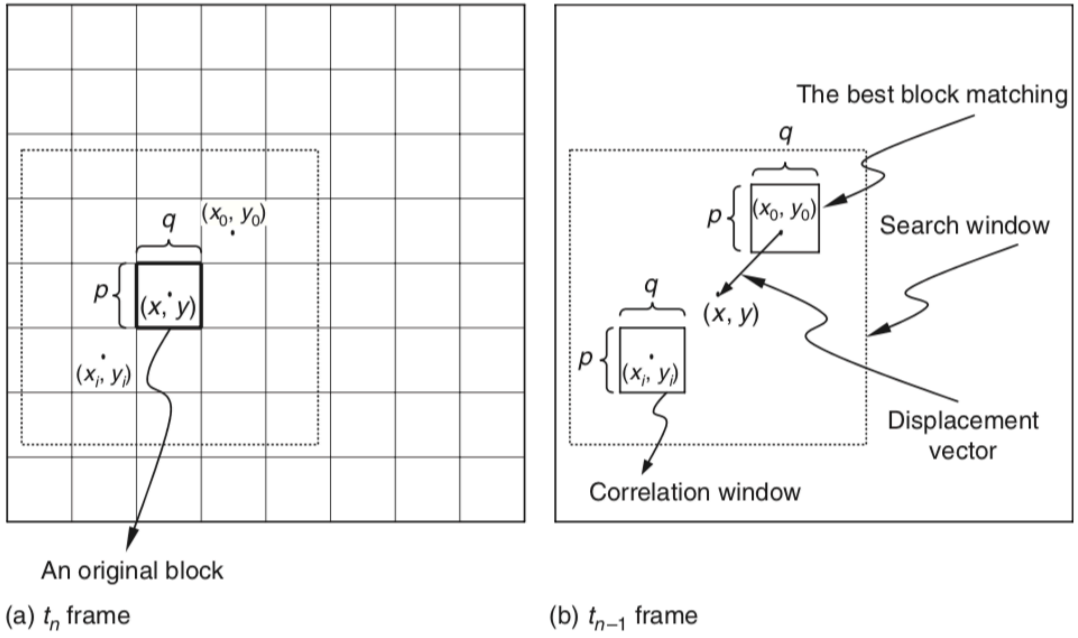
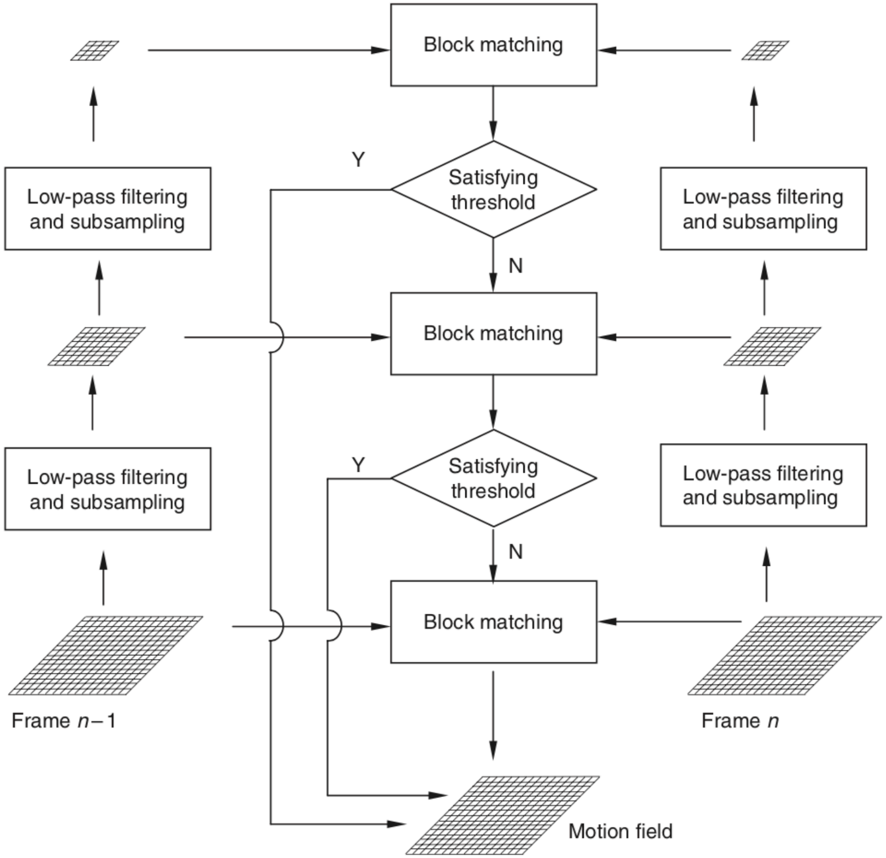
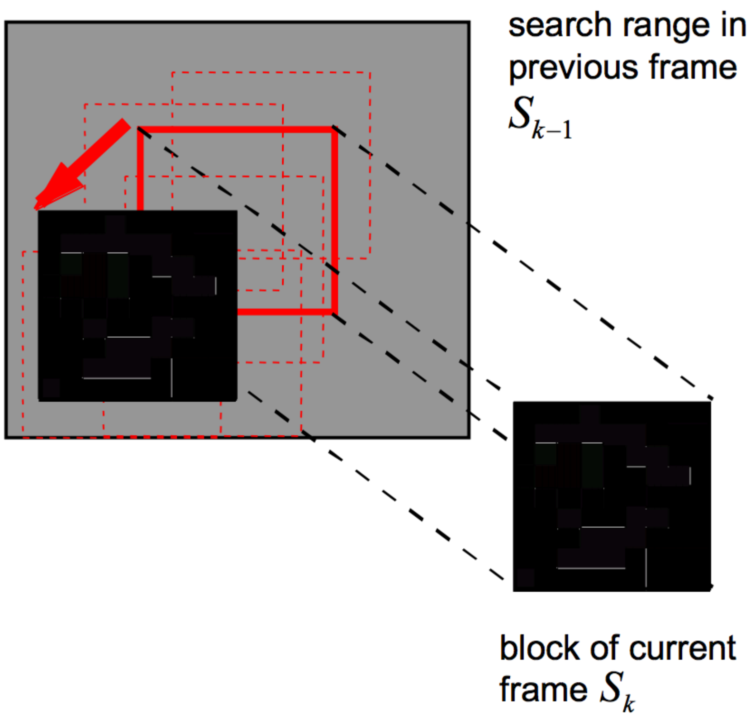


FIGURE 11.1
Block matching.


Block matching
Y
Satisfying threshold
N
Low-pass filtering and subsampling
Low-pass filtering and subsampling
Block matching
Y
Satisfying threshold
N
Low-pass filtering and subsampling
Low-pass filtering and subsampling
Block matching
Frame n–1
Frame n
Motion field

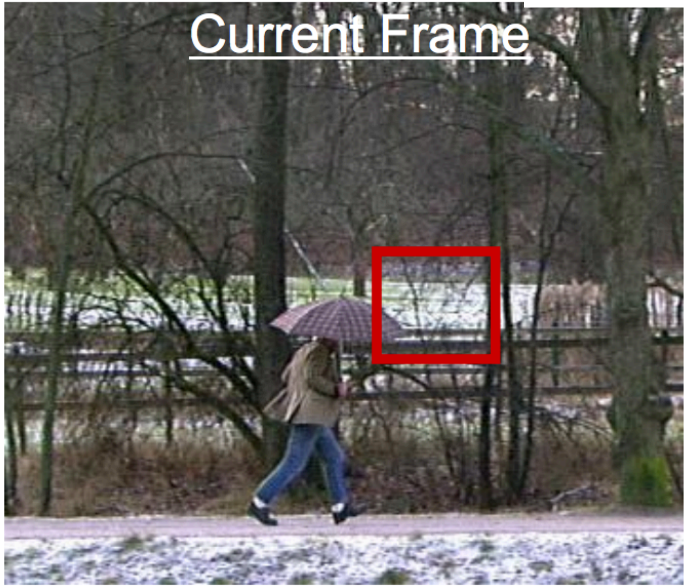
Universitat Pompeu Fabra Barcelona

Block-matching Algorithm



- Subdivide current frame into blocks.
- Find one displacement vector for each block.
- Within a search range, find a "best match" that minimizes an error measure.
- Intelligent search strategies can reduce computation.

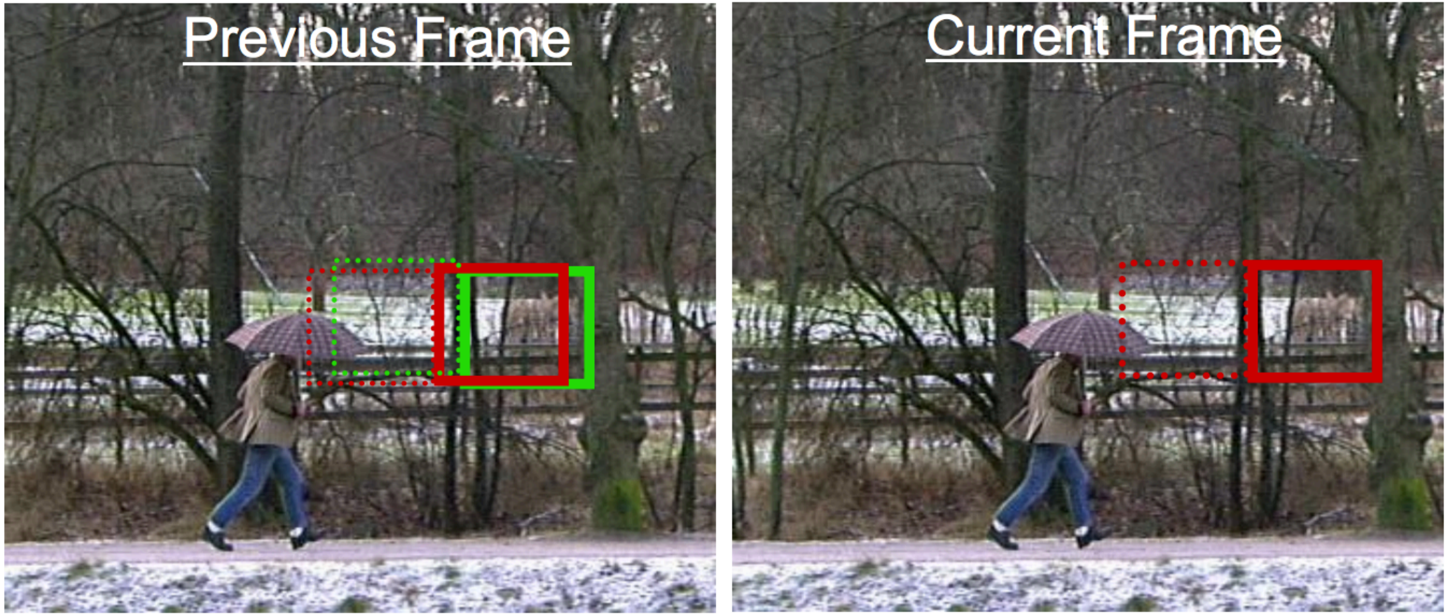
Block-matching Algorithm



Measurement window is compared with a shifted block of pixels in the other image, to determine the best match

Block of pixels is selected as a measurement window

Block-matching Algorithm



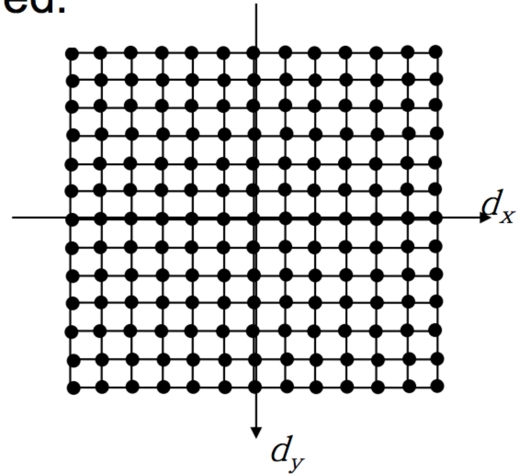
. . . process repeated for another block.

Universitat Pompeu Fabra Barcelona

Block-matching: Search Strategie

Full search

- All possible displacements within the search range are compared.

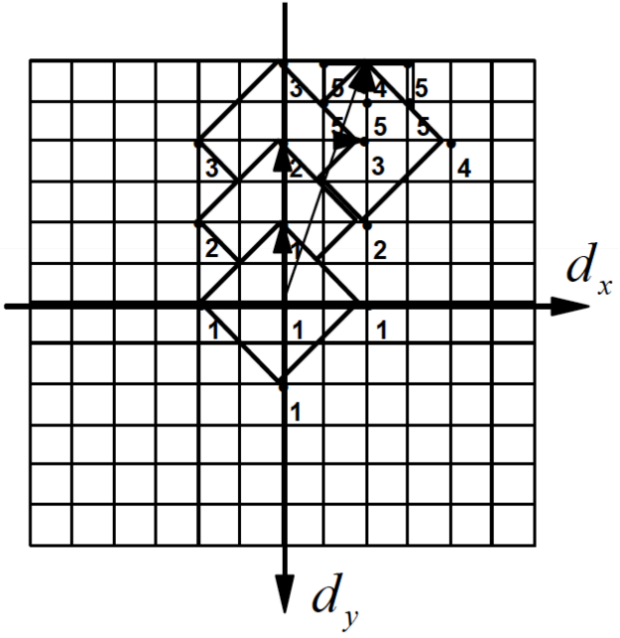


- Computationally expensive
- Highly regular, parallelizable

Blockmatching: Search Order I

2D logarithmic search [Jain + Jain, 1981]

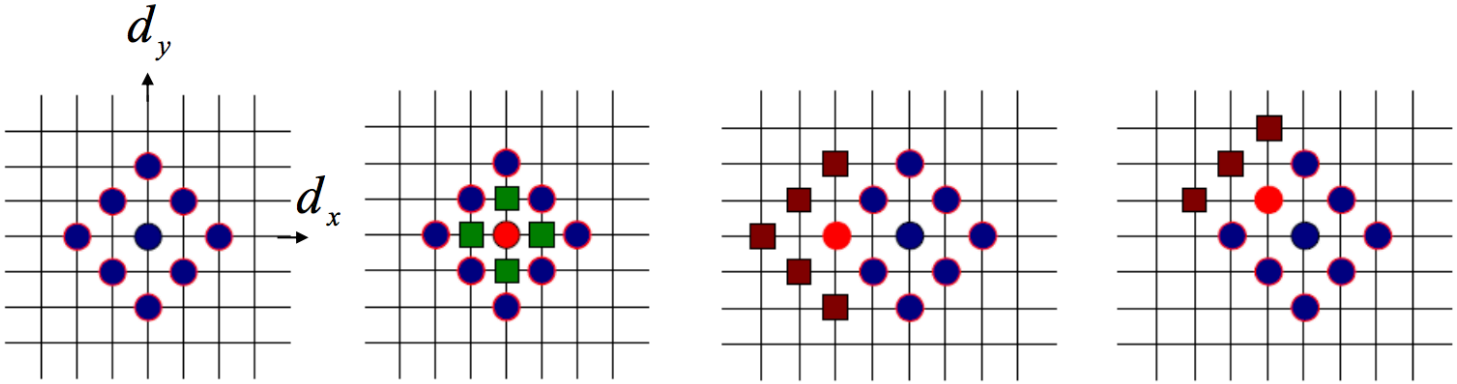
- Iterative comparison of error measure values at 5 neighboring points
- Logarithmic refinement of the search pattern if
  - best match is in the center of the 5-point pattern
  - center of search pattern touches the border of the search range



Universitat Pompeu Fabra Barcelona

Blockmatching: Search Order II

Diamond search [Li, Zeng, Liou, 1994] [Zhu, Ma, 1997]



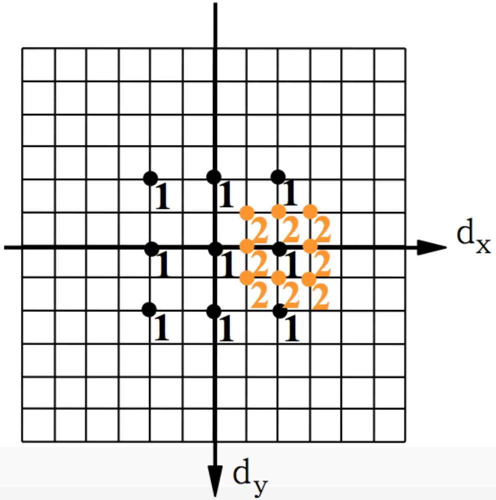
Start with large diamond pattern at (0,0)

If best match lies in the center of large diamond, proceed with small diamond

If best match does not lie in the center of large diamond, center large diamond pattern at new best match

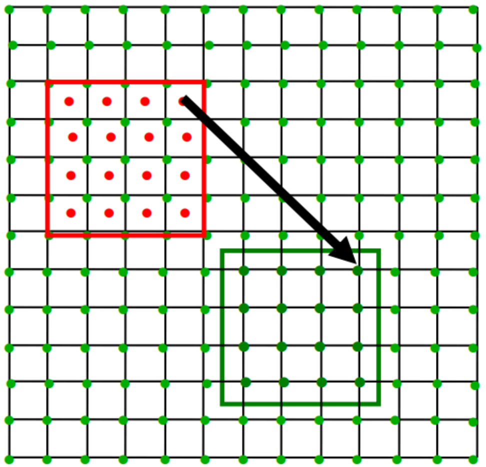
Sub-pel Translational Motion

- Motion vectors are often not restricted to only point into the integer-pel grid of the reference frame
- Typical sub-pel accuracies: half-pel and quarter-pel
- Sub-pel positions are often estimated by „refinement“

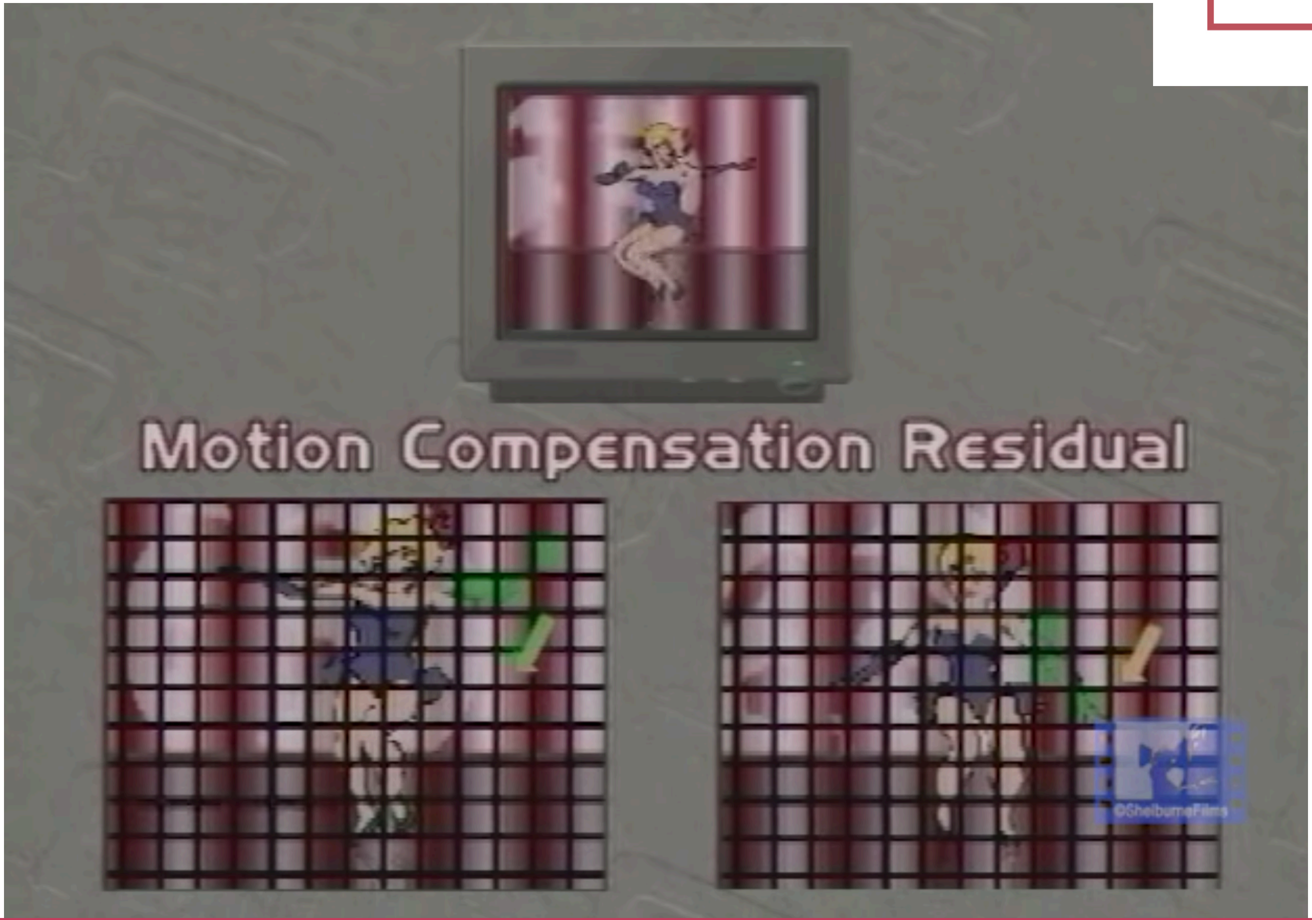


Sub-pel Motion Compensation

- Sub-pel positions are obtained via interpolation
- Example: half-pixel accurate displacements



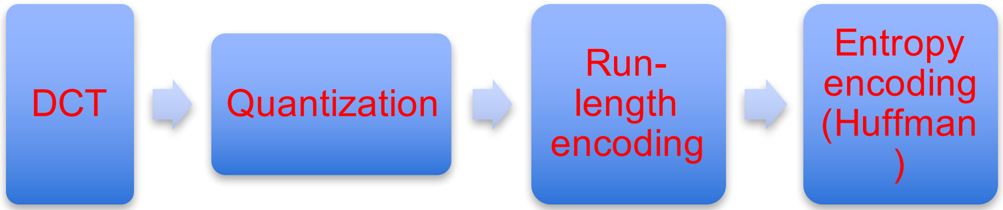
$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} 4.5 \\ 4.5 \end{pmatrix}$$


Motion Compensation Residual

Motion compensation:

3 – Encoding

We already know this part :-)



Example in real life of MPEG encoder

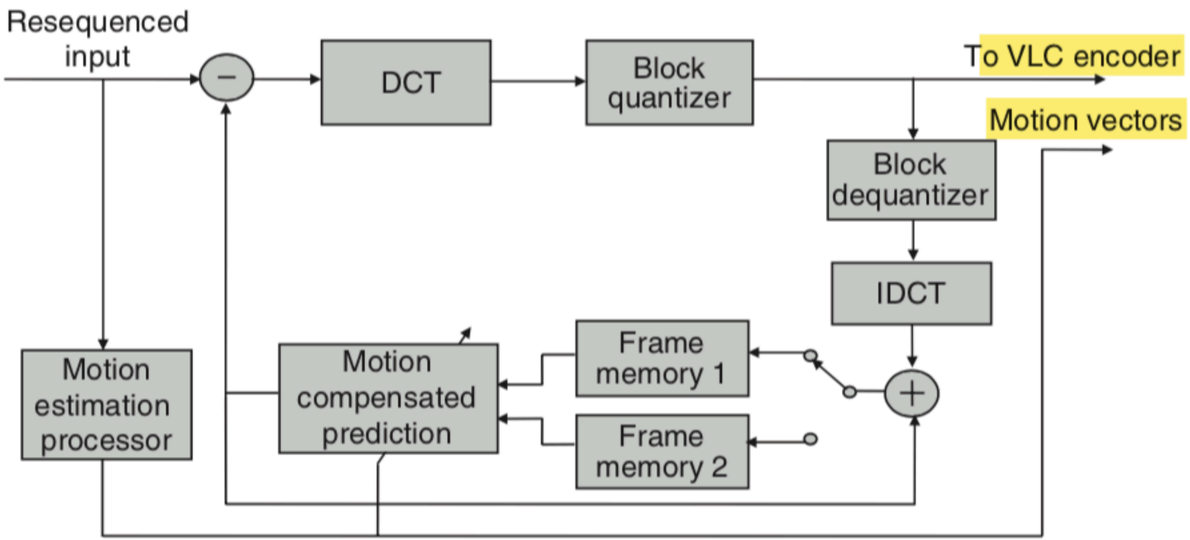


FIGURE 16.2
Typical MPEG-1 encoder structure.

Universitat Pompeu Fabra Barcelona

Remember: the idea was to encode harder and decode easier

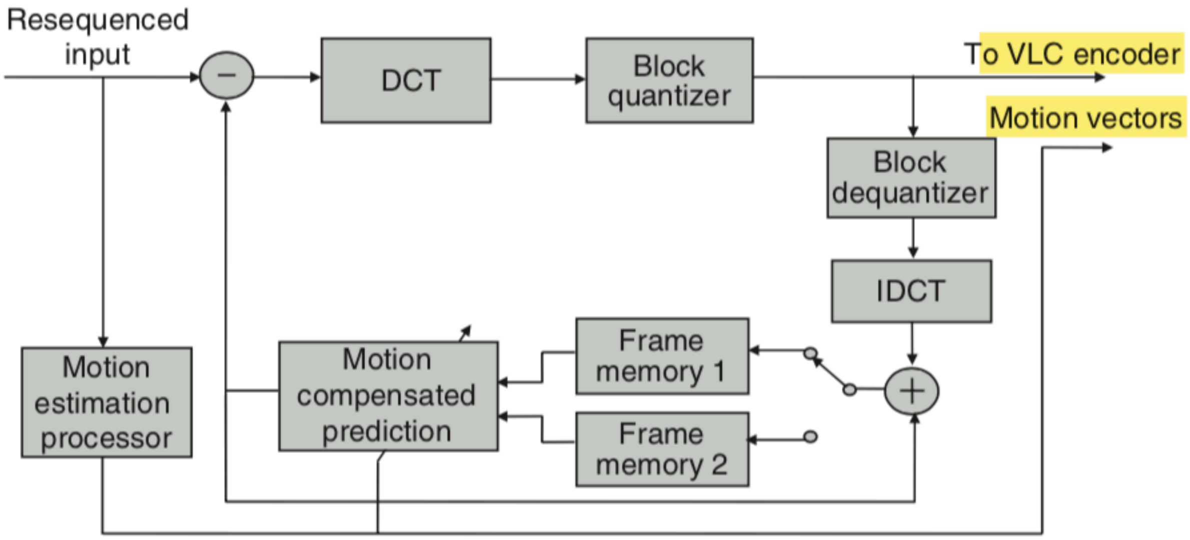


FIGURE 16.2

How does audio work inside MPEG?

Universitat Pompeu Fabra Barcelona

Own codec based on psychoacustics

MPEG-1 Layer I (.mp1) is defined in ISO/IEC 11172-3, which first version was published in 1993.

Sampling rates: 32, 44.1 and 48 kHz
Bitrates: 32, 64, 96, 128, 160, 192, 224, 256, 288, 320, 352, 384, 416 and 448 kbit/s[6]

MPEG-2

Moving
Picture
Experts
Group

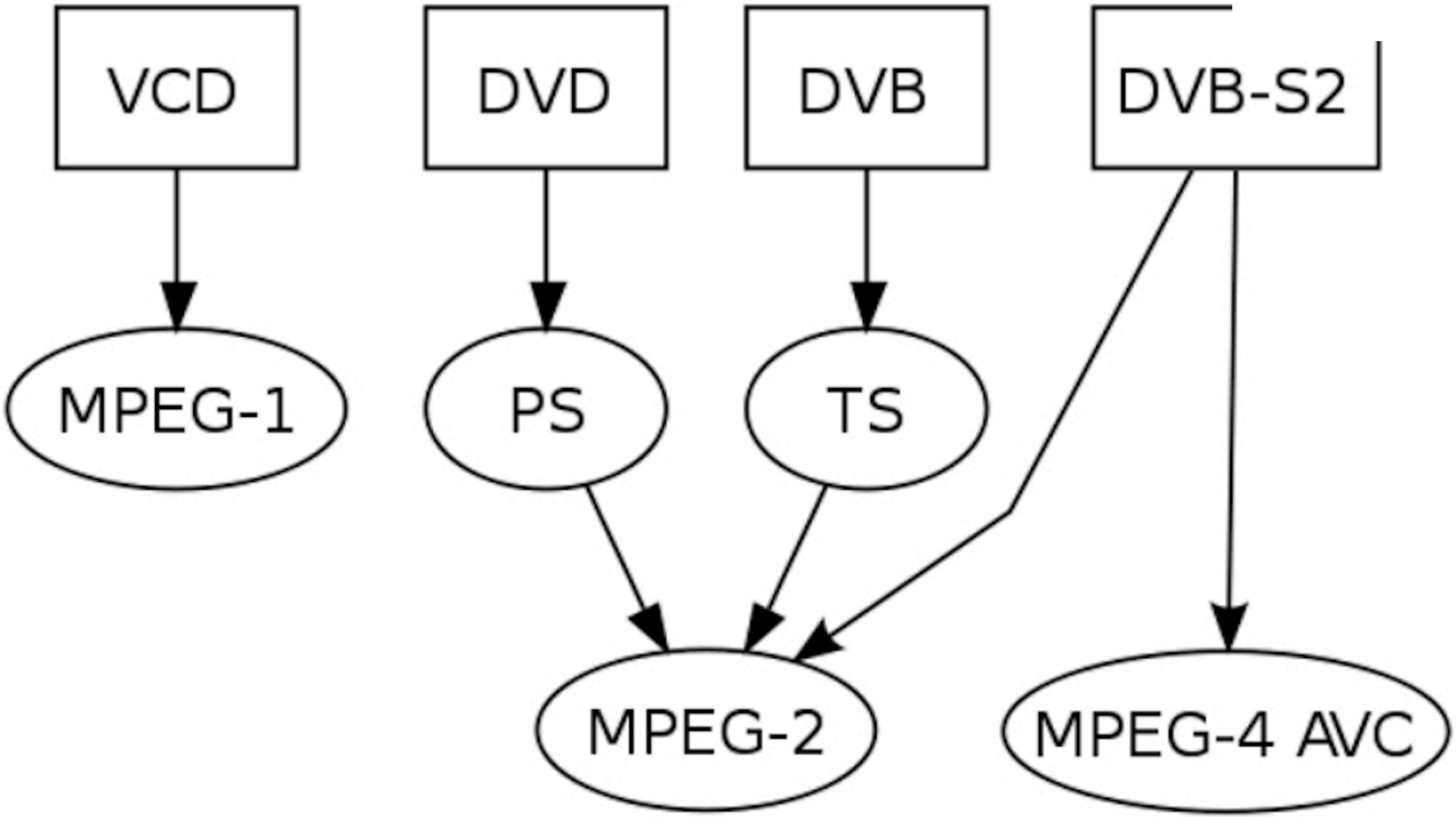
Universitat Pompeu Fabra Barcelona

It’s the same technology as MPEG-1 but with some enhancements

The major MPEG standards include the following: Simplyinfo.net

- MPEG-2: Offers resolutions of 720x480 and 1280x720 at 60 fps, with full CD-quality audio. This is sufficient for all the major TV standards, including NTSC, and even HDTV. MPEG-2 is used by DVD-ROMs. MPEG-2 can compress a 2 hour video into a few gigabytes. While decompressing an MPEG-2 data stream requires only modest computing power, encoding video in MPEG-2 format requires significantly more processing power.

Universitat Pompeu Fabra Barcelona



Universitat Pompeu Fabra Barcelona

·Field/frame prediction modes for supporting the interlaced video input

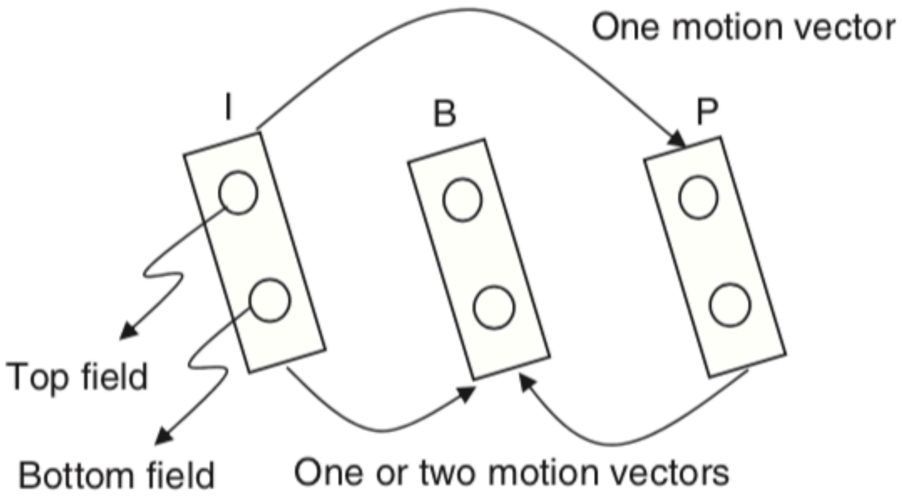


FIGURE 16.11
Frame-based prediction of MPEG-1 video coding.

·Field/frame prediction modes for supporting the interlaced video input

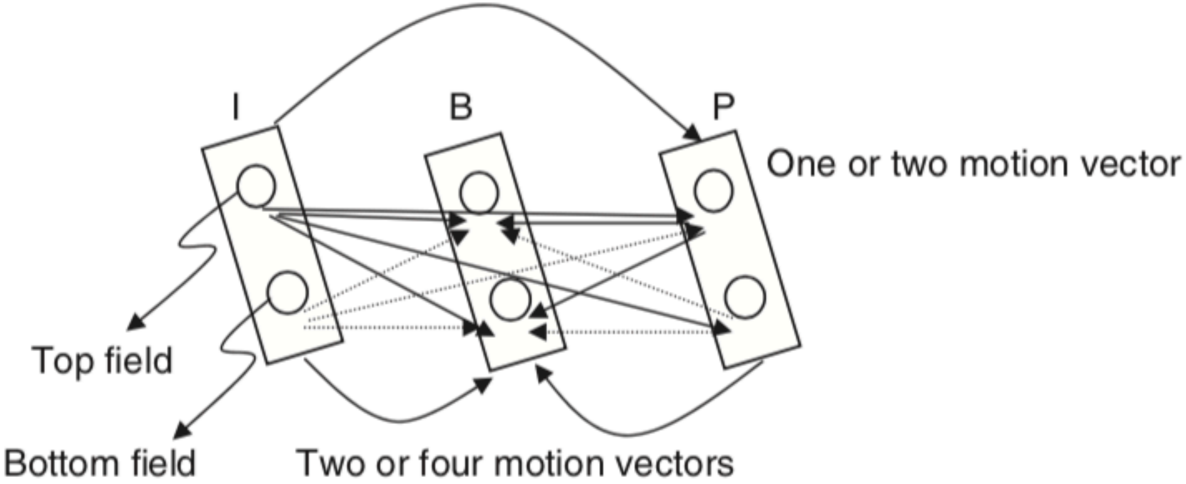


FIGURE 16.12
Field-based prediction of enhanced option of MPEG-2 video coding.

Universitat Pompeu Fabra Barcelona

·Field/frame prediction modes for supporting the interlaced video input

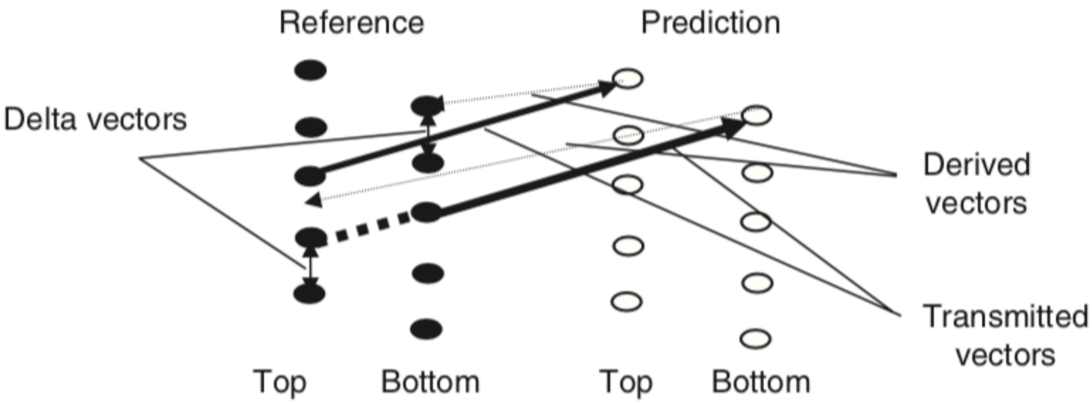


FIGURE 16.13
Dual prime prediction in MPEG-2 video coding.

Universitat Pompeu Fabra Barcelona

▪Field/frame DCT coding syntax

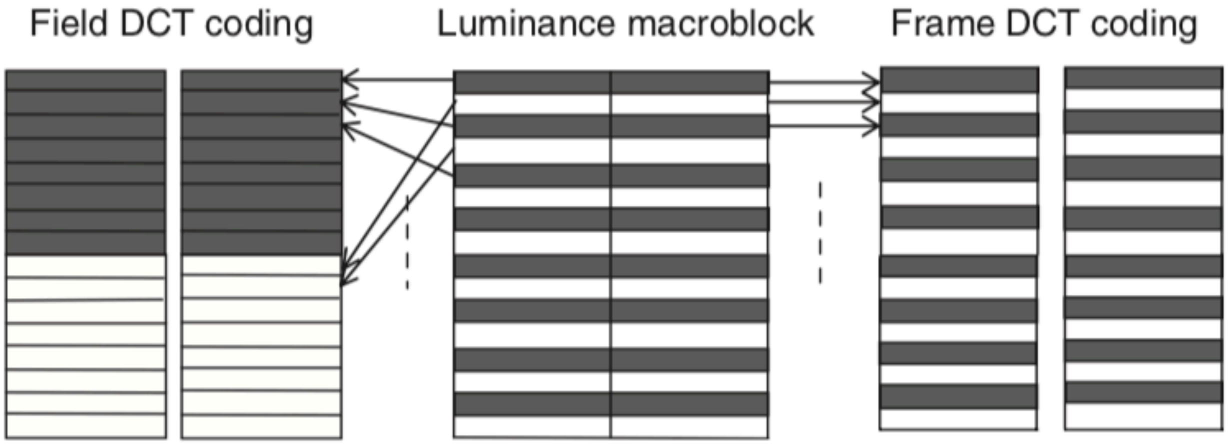


FIGURE 16.14
Frame and field discrete cosine transform (DCT) for interlaced video.

·Downloadable quantization matrix and alternative scan order

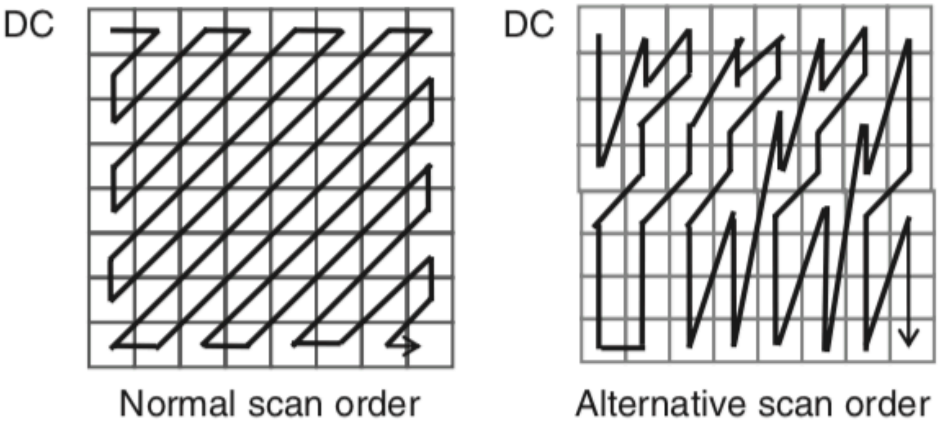


FIGURE 16.15
Two zigzag scan methods for MPEG-2 video coding.

▪Scalability extension

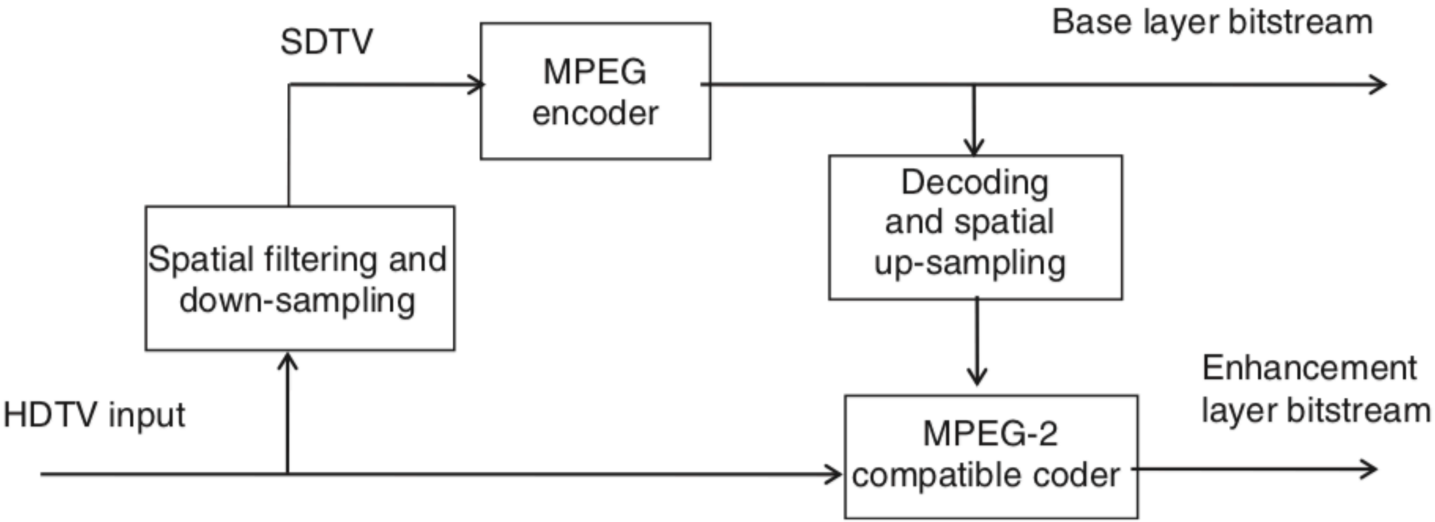


FIGURE 16.17
Block diagram of spatial scalability encoder.

Universitat Pompeu Fabra Barcelona

▪Scalability extension

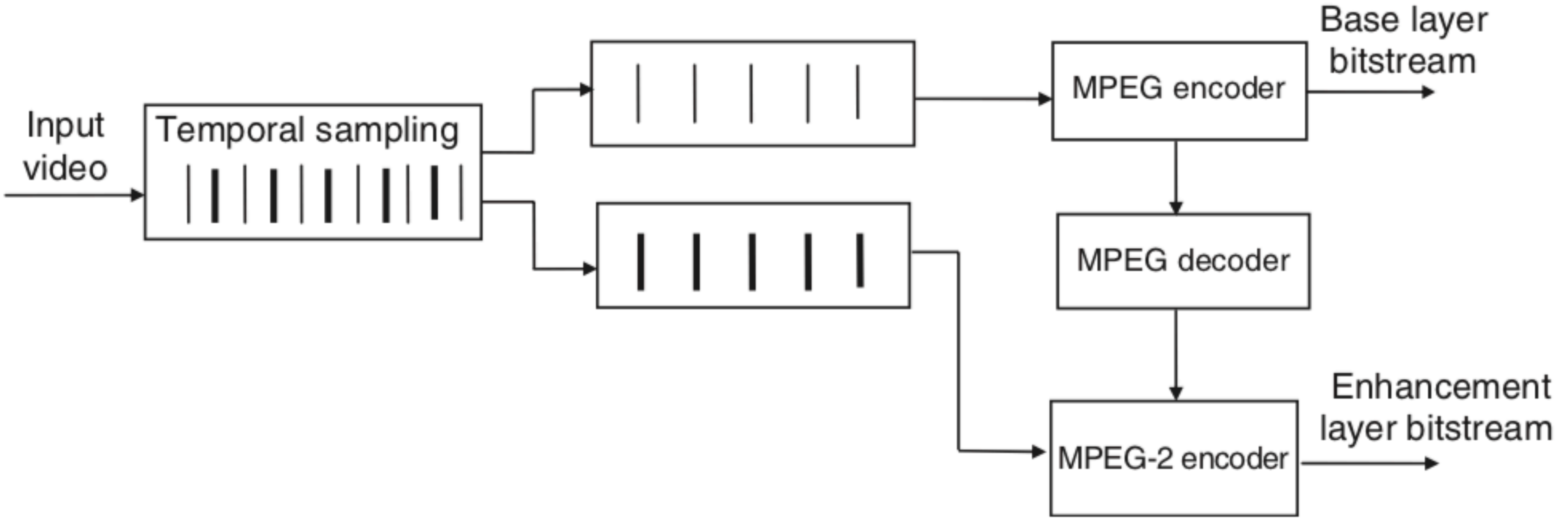


FIGURE 16.18
Block diagram of temporal scalability encoder.

Previous enhancement items are all coding performance improvements that are related to the support of interlaced material

Universitat Pompeu Fabra Barcelona

Non-compression enhancements:

- Syntax to facilitate 3:2 pull-down in the decoder
- Pan and scan codes with 1=16 pixel resolution
- Display flags indicating chromaticity, subcarrier amplitude, and phase (for NTSC=PAL=SECAM source material)

How does audio work?

Same as MPEG1, but also new improved versions of the codec:

The MPEG-2 standard includes several extensions to MPEG-1 Audio, and these are known as MPEG-2 BC – backwards compatible with MPEG-1 Audio. MPEG-2 Audio is defined in ISO/IEC 13818-3

MPEG Multichannel – Backward compatible 5.1-channel surround sound

Sampling rates: 16000, 22050, and 24000 Hz

Bitrates: 8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 144 and 160 kbit/s

Universitat Pompeu Fabra Barcelona

Sampling rates are exactly half that of those originally defined for MPEG-1 Audio.

They were introduced to maintain higher quality sound when encoding audio at lower-bitrates.

The even-lower bitrates were introduced because tests showed that MPEG-1 Audio could provide higher quality than any existing (circa 1994) very low bitrate (i.e. speech) audio codecs

And finally… MPEG 2 Part 7 a.k.a. …

MPEG2 Real life applications

Digital Versatile Disc (almost obsolete)

Allowed Dimensions (resolution):

- 720 × 480, 704 × 480, 352 × 480, 352 × 240 pixel (NTSC)
- 720 × 576, 704 × 576, 352 × 576, 352 × 288 pixel (PAL)

Allowed Aspect ratios (DAR):

- 4:3 (for letterboxed widescreen and non-widescreen frames)
- 16:9 (for anamorphic widescreen[dvdaspect 1])

Allowed frame rates

- 29.97 interlaced frame/s (NTSC)
- 23.978 progressive frame/s (for NTSC 2:3 pull-down to
- 29.97[dvdrates 1])
- 25 interlaced frame/s (PAL)

Digital Versatile Disc

For audio it has a lot of restrictions plus it's able to use more modern codecs

Universitat Pompeu Fabra Barcelona

HDV

Universitat Pompeu Fabra Barcelona

HDV

Universitat Pompeu Fabra Barcelona

HDV is a format for recording of high-definition video on DV cassette tape. The format was originally developed by JVC and supported by Sony, Canon, and Sharp. The four companies formed the HDV consortium in September 2003.

Universitat Pompeu Fabra Barcelona

Conceived as an affordable high definition format for digital camcorders, HDV quickly caught on with many amateur and professional videographers due to its low cost, portability, and image quality acceptable for many professional productions.

HDV (in use)

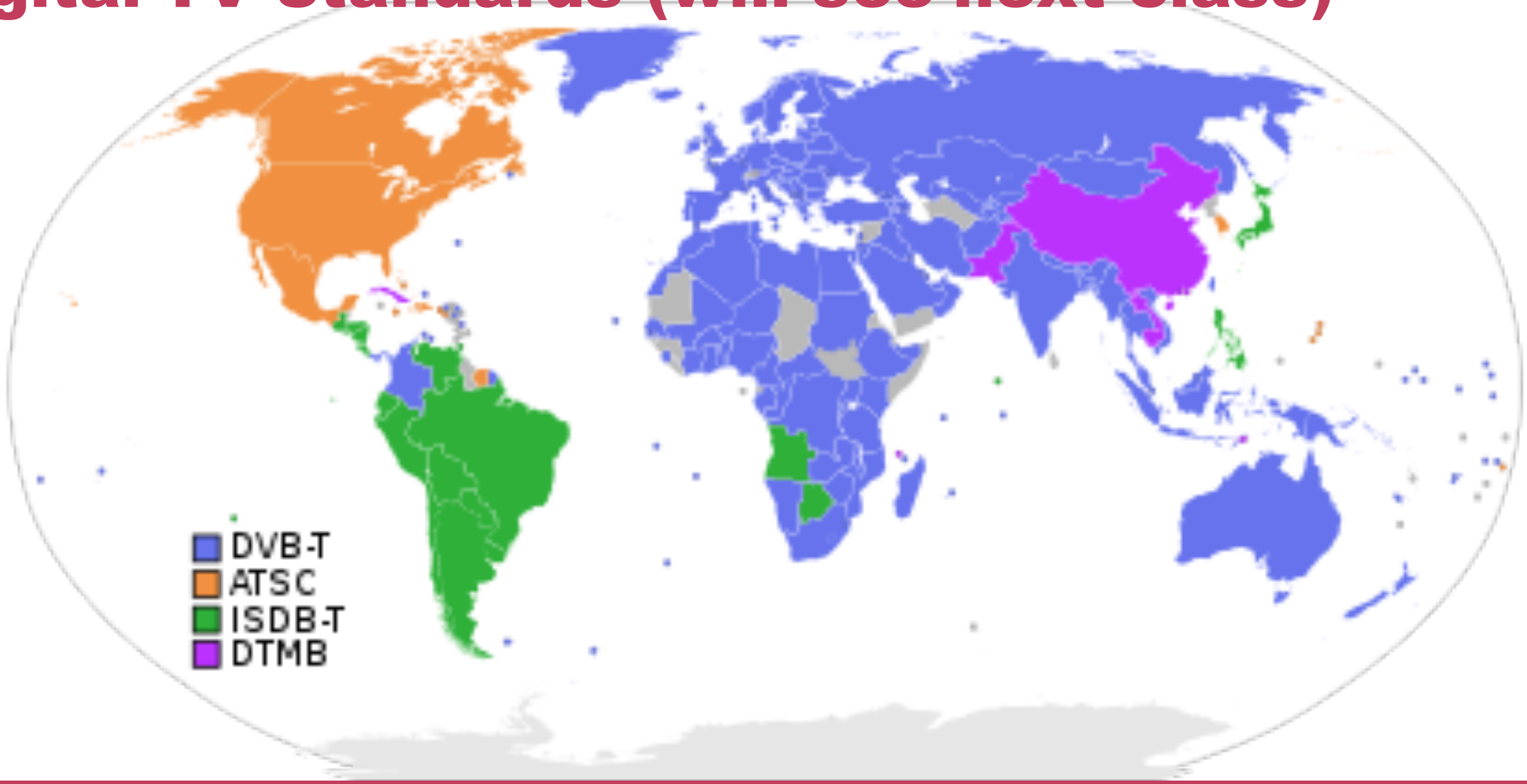

60
SP MODE
90
LP MODE
MiniDV
DVM60
SONY
MiniDV 63
DIGITAL HD VIDEO
DVM63
for HDV / DV

MOD and TOD (obsolete)

XDCAM (in use)

Universitat Pompeu Fabra Barcelona

Digital TV Standards (will see next class)



Early Blue-ray Disc (in use)

Thanks

franciscojavier.brines@upf.edu