VIDEOTELEPHONY AND VIDEOCONFERENCE

Fernando Pereira

Instituto Superior Técnico

Digital Video

Video versus Images

- **Still Image Services – No strong temporal requirements; no real-time notion.**
- **Video Services (moving images) – It is necessary to strictly follow critical timing and delay requirements to provide a good illusion of motion; this is essential to provide real-time performance.**

For each image and video service, it is possible to associate a quality target (related to QoS/QoE); the first impact of this target is the selection of the appropriate (PCM) spatial and temporal resolutions to use.

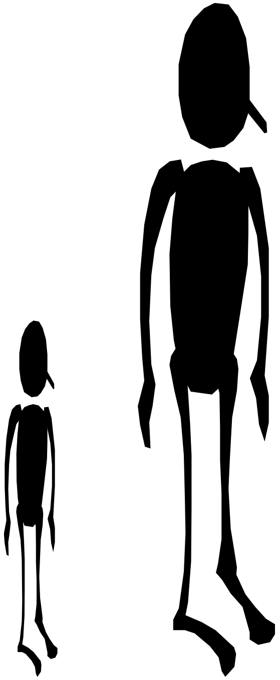
Why Does Video Information Have to be Compressed ?

A video sequence is created and consumed as a flow of images, happening at a certain temporal rate (F), each of them with a spatial resolution of M×N luminance and chrominance samples and a certain number of bits per sample (L)

This means the total rate of (PCM) bits

- and thus the required bandwidth and memory – necessary to digitally represent a video sequence is HUGE !!!

$$(3 \times F \times M \times N \times L)$$



Digital Video: Why is it So Difficult ?

Service	Spatial resolution (lum, chrom)	Temporal resolution	Bit/sample	PCM bitrate
Full HD 1080p	1080 × 1920 1080 × 960	25 imagens/s progressivas	8 bit/amostra	830 Mbit/s
HD Ready 720p	720 × 1280 720 × 640	25 imagens/s progressivas	8 bit/amostra	370 Mbit/s
Standard TV, DVD	576 × 720 576 × 360	25 imagens/s entrelaçadas	8 bit/amostra	166 Mbit/s
Internet *streaming*	288 × 360 144 × 180	25 imagens/s progressivas	8 bit/amostra	31 Mbit/s
Mobile video	144 × 180 72 × 90	25 imagens/s progressivas	8 bit/amostra	7.8 Mbit/s
Music (stereo)	-	44000 amostras/s	16 bit/amostra	1.4 Mbit/s
Speech (GSM)	-	8000 amostras/s	8 bit/amostra	64 kbit/s

Videotelephony: Just an (Easy) Example

- **Resolution: 10 images/s with 288×360 luminance samples and 144 × 188 samples for each chrominance (4:2:0 subsampling format) , with 8 bit/sample**

  $$[(360 \times 288) + 2 \times (180 \times 144)] \times 8 \times 10 = 12.44 \text{ Mbit/s}$$

- **Reasonable bitrate: e.g. 64 kbit/s for an ISDN B-channel**

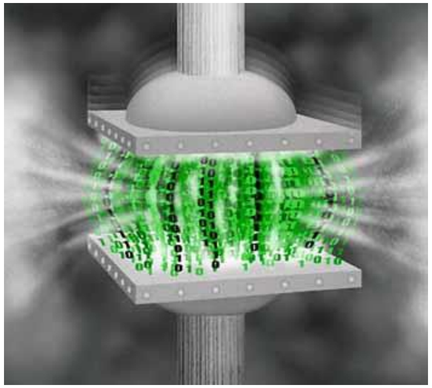
=> Compression Factor: 12.44 Mbit/s/64 kbit/s ≈ 194

The usage or not of compression/source coding implies the possibility or not to deploy services and, thus, the emergence or not of certain services, e.g. Internet video.

Video Coding/Compression: a Definition

Efficient representation (*this means with a smaller than the PCM number of bits*) of a periodic sequence of (correlated) images, satisfying the relevant requirements, e.g. minimum acceptable quality, low delay, error robustness, random access.

And the compression requirements change with the services/applications and the corresponding funcionalities ...



How Big Has to be the Compression 'Hammer' ?

Service	Spatial resolution (lum, chrom)	Temporal resolution	Bit/sample	PCM bitrate	Compressed bitrate	Compression factor
Full HD 1080p	1080 × 1920 1080 × 960	25 imagens/s progressivas	8 bit/amostra	830 Mbit/s	8-10 Mbit/s	80-100
HD Ready 720p	720 × 1280 720 × 640	25 imagens/s progressivas	8 bit/amostra	370 Mbit/s	4-6 Mbit/s	90
Standard TV, DVD	576 × 720 576 × 360	25 imagens/s entrelaçadas	8 bit/amostra	166 Mbit/s	2 Mbit/s	83
Internet *streaming*	288 × 360 144 × 180	25 imagens/s progressivas	8 bit/amostra	31 Mbit/s	150 kbit/s	200
Mobile video	144 × 180 72 × 90	25 imagens/s progressivas	8 bit/amostra	7.8 Mbit/s	100 kbit/s	80
Music (stereo)	-	44000 amostras/s	16 bit/amostra	1.4 Mbit/s	100 kbit/s	14
Speech (GSM)	-	8000 amostras/s	8 bit/amostra	64 kbit/s	13 kbit/s	5

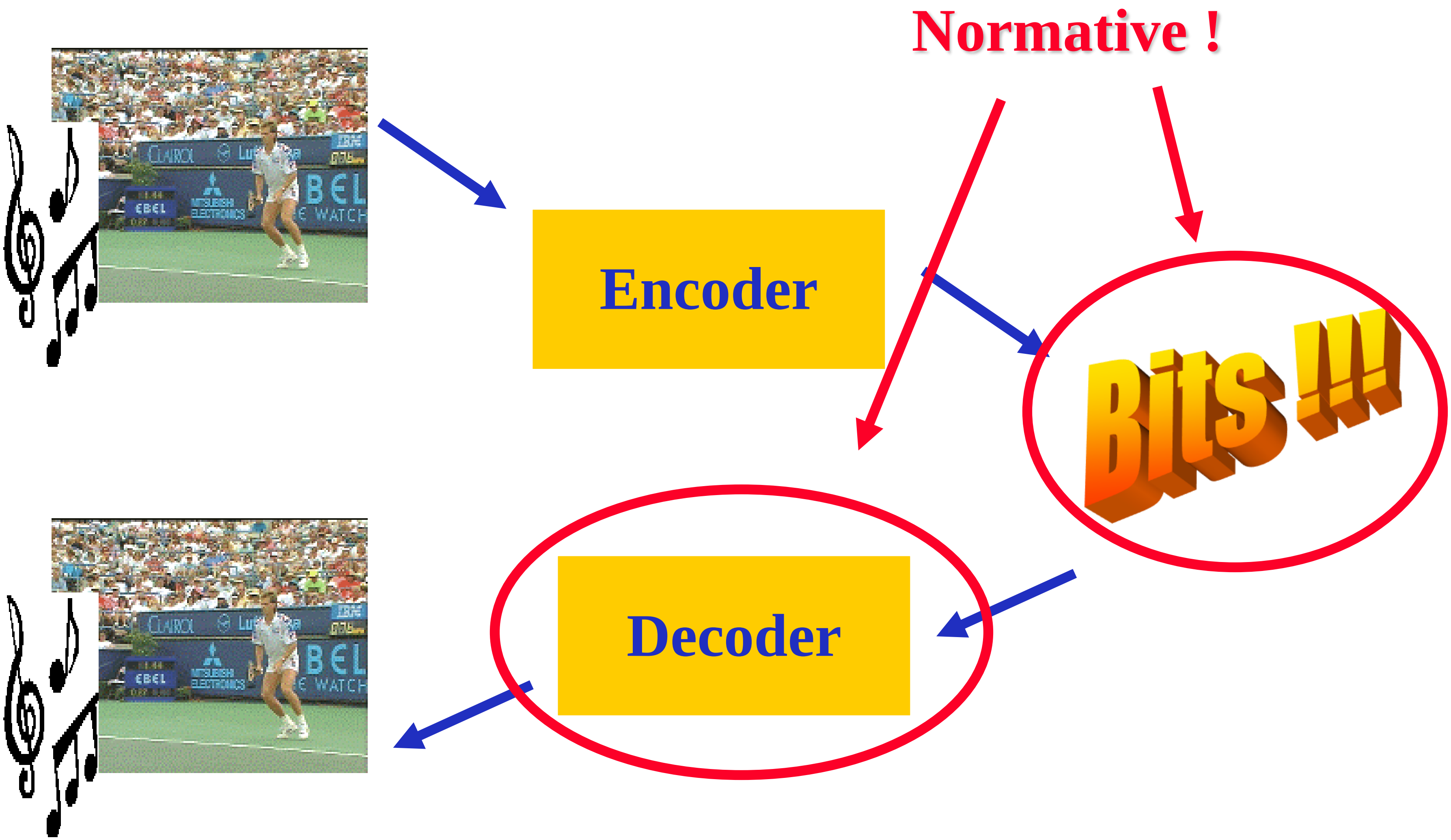
Interoperability as a Major Requirement: Standards to Assure that More is not Less ...

- **Compression is essential for digital audiovisual services where interoperability is a major requirement.**

- **Interoperability requires the specification and adoption of standards, notably audiovisual coding standards.**

- **To allow some evolution of the standards and some competition in the market between compatible products from different companies, standards must specify the minimum set of technology possible, typically the bitstream syntax and the decoding process (not the encoding process).**

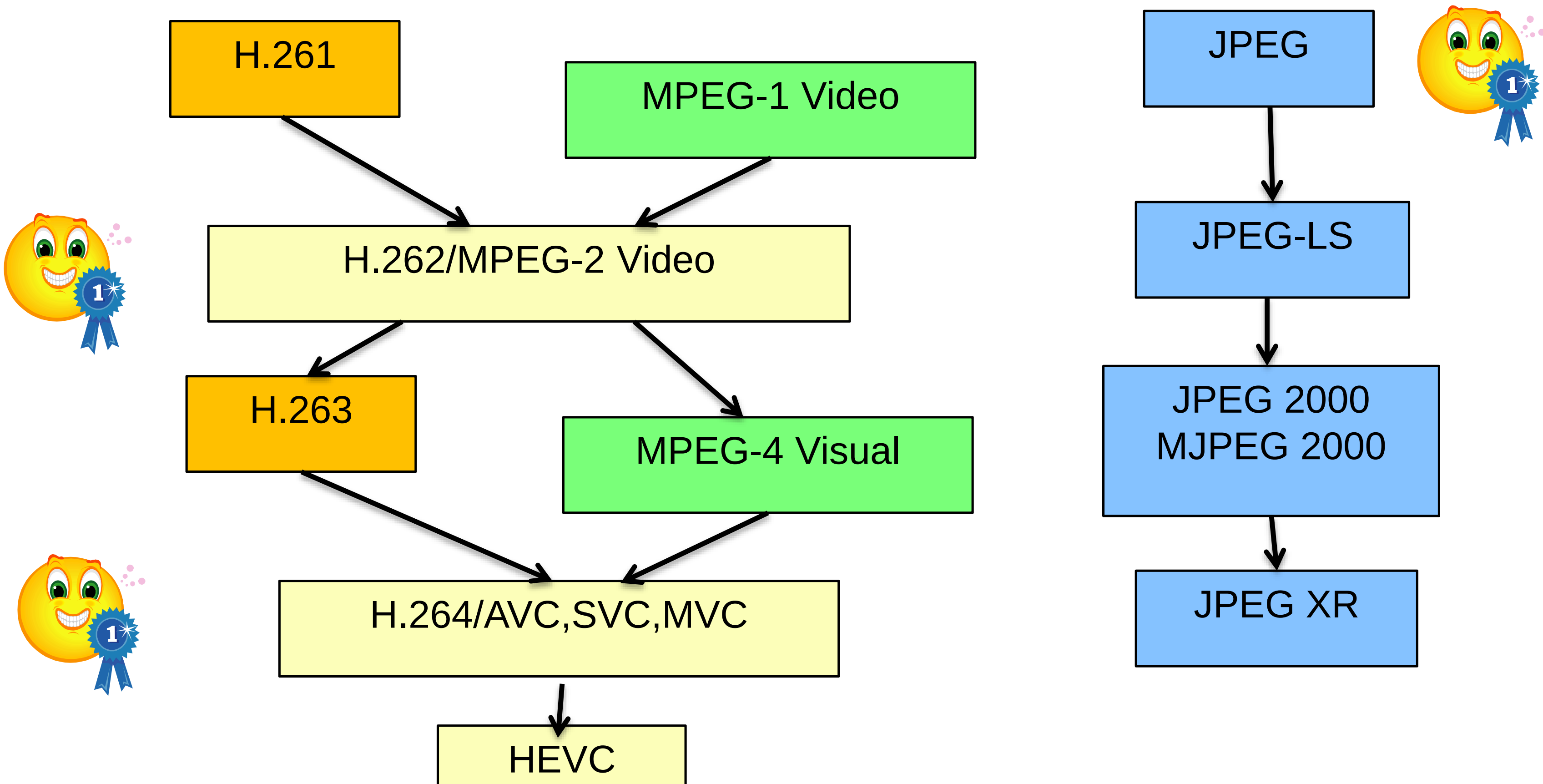
Standards: a Trade-off between Fixing and Inovating



Video Coding Standards …

- **ITU-T H.120 (1984) - Videoconference (1.5 - 2 Mbit/s)**
- **ITU-T H.261 (1988) – Audiovisual services (videotelephony and videoconference) at p×64kbit/s, p=1,…,30**
- **ISO/IEC MPEG-1 (1990)- CD-ROM Video**
- **ISO/IEC MPEG-2 also ITU-T H.262 (1993) – Digital TV**
- **ITU-T H.263 (1996) – PSTN and mobile video**
- **ISO/IEC MPEG-4 (1998) – Audiovisual objects, improved efficiency**
- **ISO/IEC MPEG-4 AVC also ITU-T H.264 (2003) – Improved efficiency**
- **ISO/IEC HEVC also ITU-T H.265 (2013) – Further improved compression efficiency**

The Video Coding Standardization Path …



ITU-T H.320 Terminals

Videotelephony and Videoconference

Videotelephony and Videoconference

Personal (bidirectional) communications in real-time ...

ITU-T H.320 Recommendation: Motivation

The starting of the work towards Rec. H.320 and H.261 goes back to 1984 when it was acknowledged that:

- **There was an increase in the demand for image-based services, notably videotelephony and videoconference.**
- **There was a growing availability of 64, 384 e 1536/1920 kbit/s digital lines as well as ISDN lines.**
- **There was a need to make available image-based services and terminals for the digital lines mentioned above.**
- **Rec. H.120, just issued at that time for videoconference services, was already obsolete in terms of compression efficiency due to the fast developments in the area of video compression.**

Basic ISDN Channels

- **B-Channel - 64 kbit/s – B-channel connections may be performed with circuit-switching, packet-switching or rented lines.**

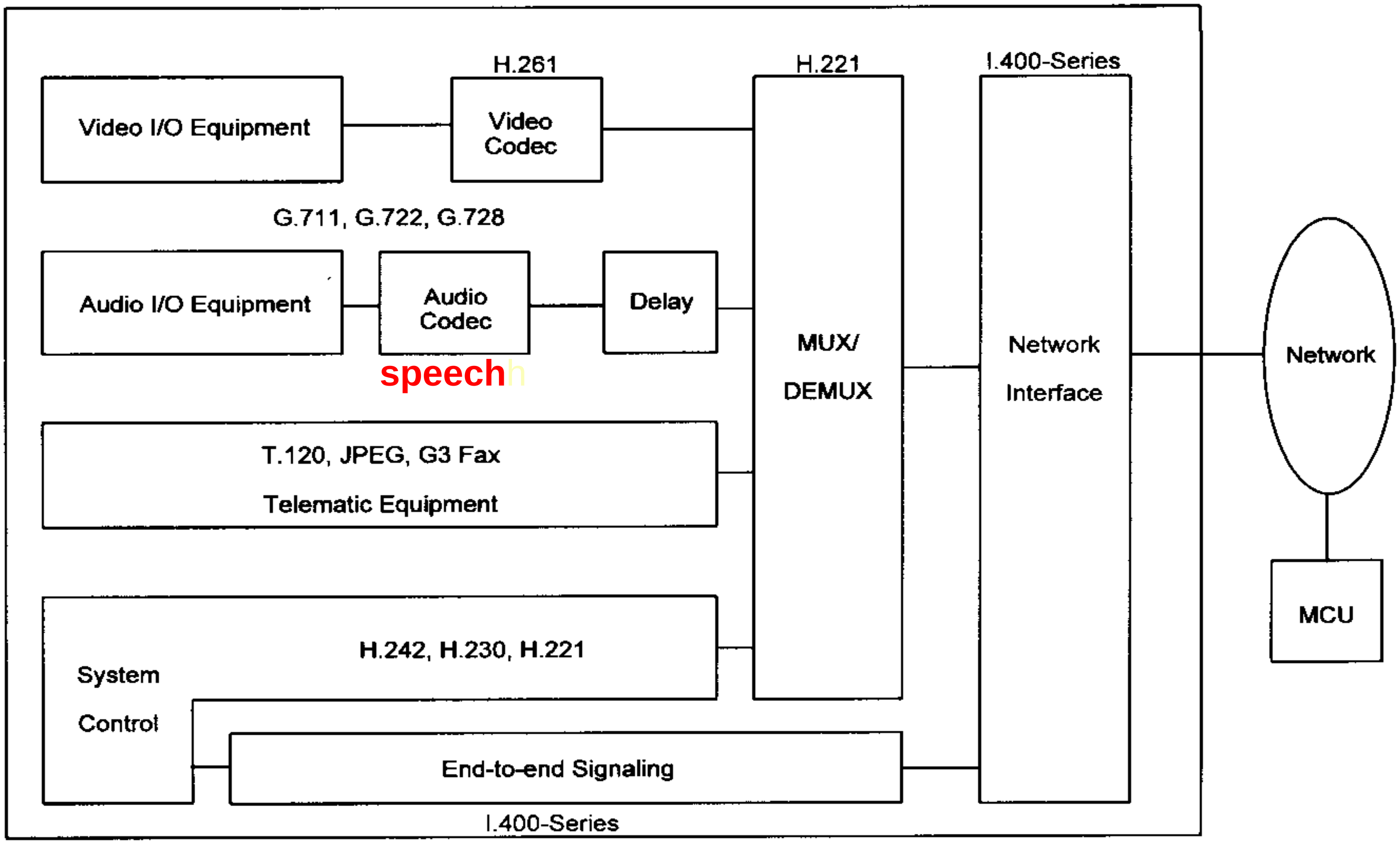
- **D-Channel - 16 ou 64 kbit/s – D-channels have the main function to transport the signalling information associated to B-channels; in the idle periods, they may be used to transmit user data using packet-switching**

- **H-Channel - 384, 1536 ou 1920 kbit/s – H-channels offer connections with higher bitrates.**

Videotelephony and Videoconference: Main Requirements/Features

- Personal communications (point to point or multipoint to multipoint)
- Symmetric bidirectional communications (all nodes involved have the same similar capabilities)
- Critical (low) delay requirements, e.g. Lower than ~ 200 ms
- Low or intermediate quality requirements
- Strong psychological and sociological impacts

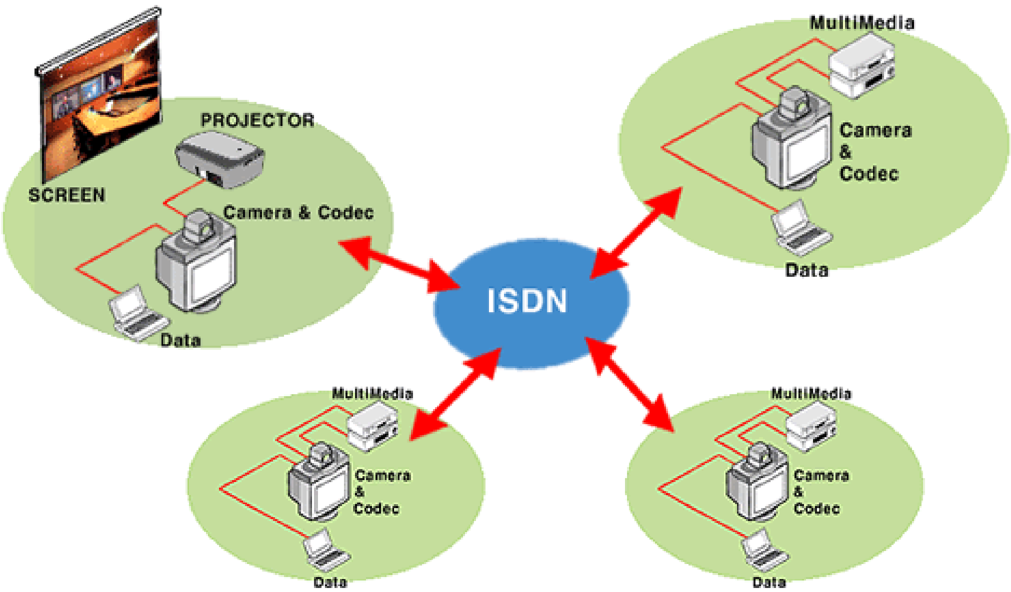
Rec. H.320 Terminal



Video Coding:
Rec. ITU-T H.261

Recommendation H.261: Objectives

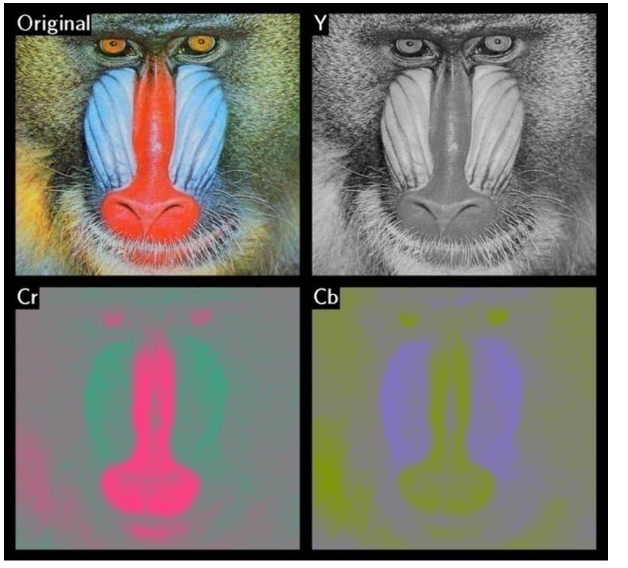
~1985



Efficient coding of videotelephony and videoconference visual data with a minimum acceptable quality using a bitrate from 40 kbit/s to 2 Mbit/s, targeting synchronous channels (ISDN) at p×64 kbit/s, with p=1,...,30.

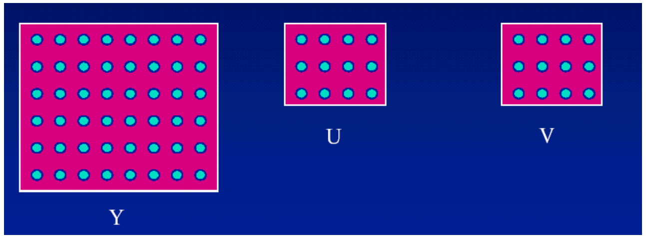
This is the first international video coding standard with relevant market adoption, thus introducing the notion of backward compatibility in video coding standards.

H.261: Signals to Code



- The signals to code for each image are luminance (Y) and 2 chrominances, and C_R or U and V.
- The samples are quantized with 8 bits/sample, according to Rec. ITU-R BT-601:
  - Black = 16; White = 235; Null colour difference = 128
  - Peak colour difference (U,V) = 16 and 240
- The coding algorithm operates over progressive (non-interlaced) content at 29.97 image/s.
- The frame rate (temporal resolution) may be reduced by skipping 1, 2 or 3 images between each coded/transmitted image.

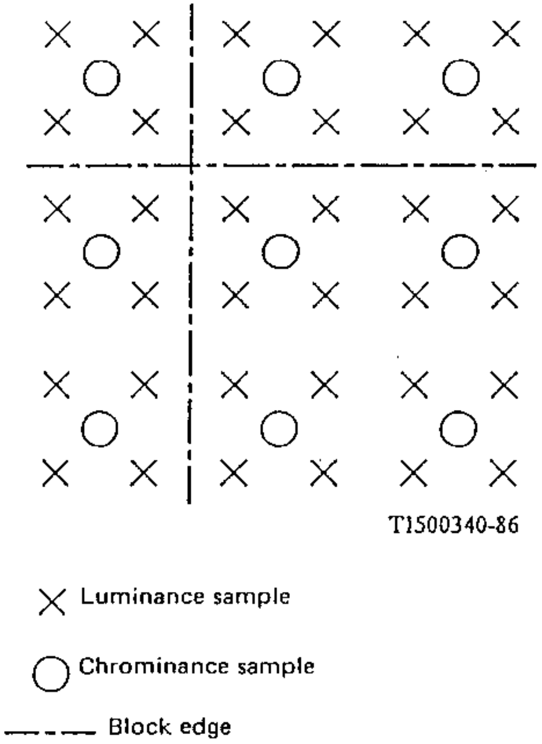
H.261: Image Format



Two spatial resolutions are possible:

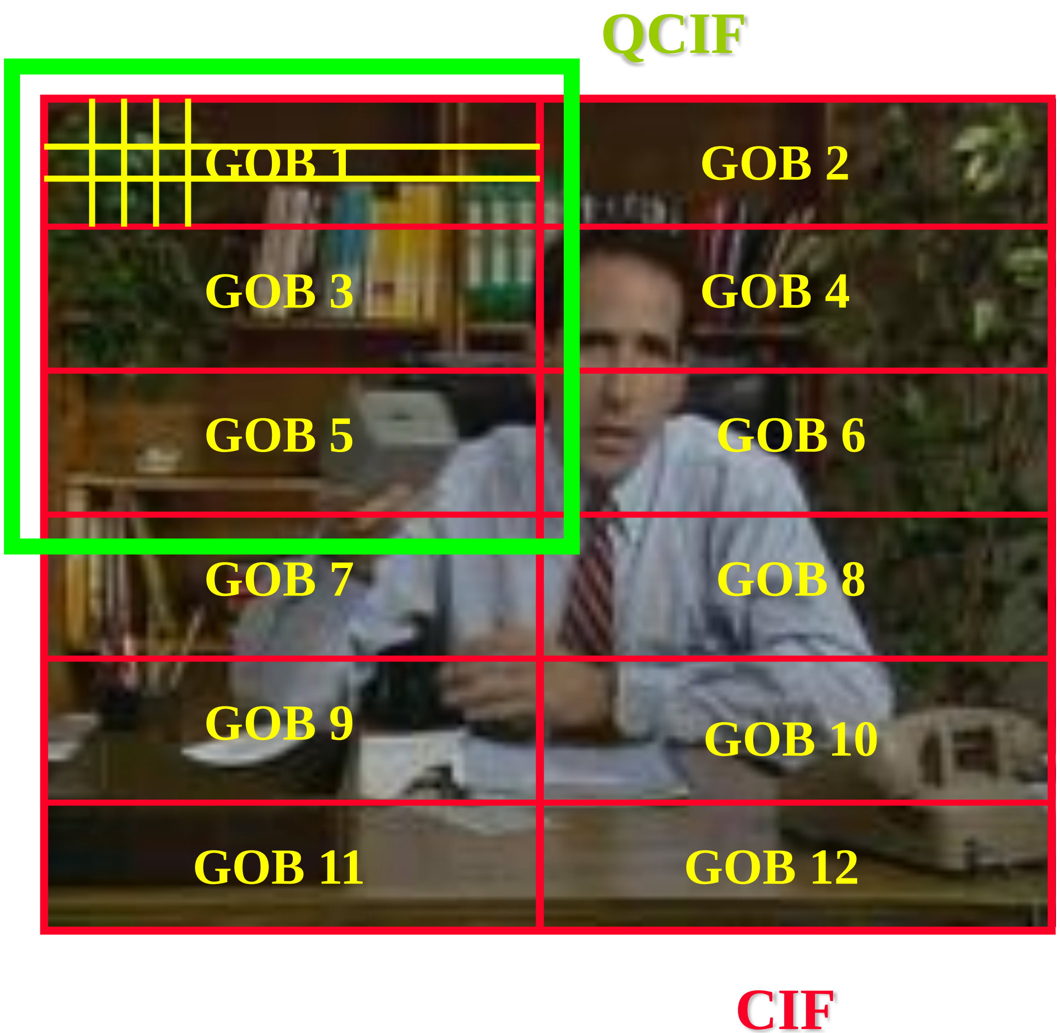
- **CIF (*Common Intermediate Format*)** - 288 × 352 samples for luminance (Y) and 144 × 176 samples for each chrominance (U,V) this means a 4:2:0 subsampling format, with 'quincux' positioning, progressive, 30 frame/s with a 4/3 aspect ratio.

- **QCIF (*Quarter CIF*)** – Similar to CIF with half spatial resolution in both directions this means 144 × 176 samples for luminance and 72 × 88 samples for each chrominance.



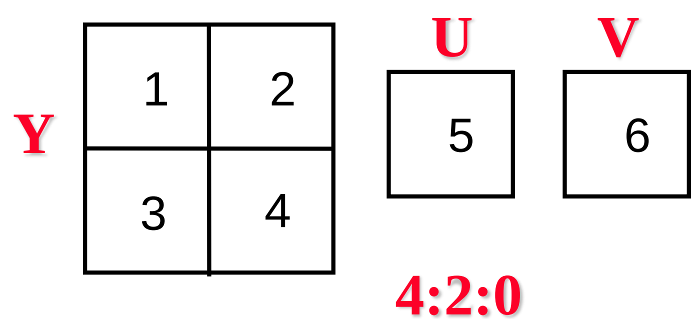
All H.261 codecs must work with QCIF and some may be able to work also with CIF (spatial resolution is set after initial negotiation).

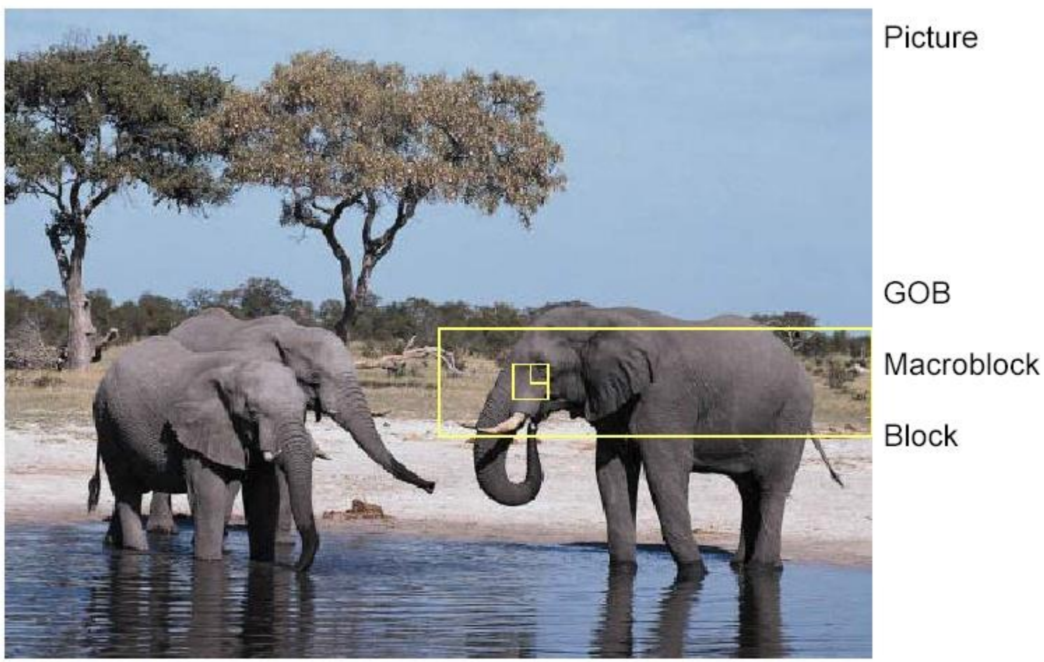
Images, Groups Of Blocks (GOBs), Macroblocks and Blocks

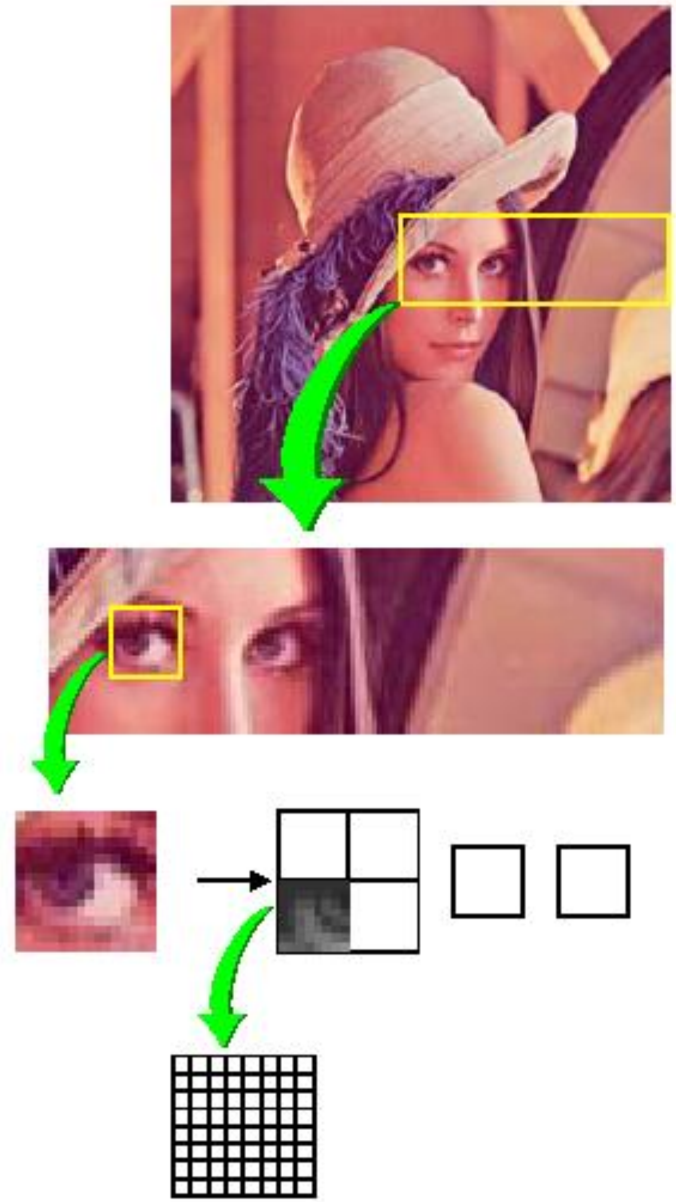


The video sequence is spatially organized according to a hierarchical structure with 4 levels:

- **Images**
- **Group of Blocks (GOB)**
- **Macroblocks (MB) – 16×16 pixels**
- **Blocks - 8×8 samples**




Picture
GOB
Macroblock
Block



H.261: Coding Tools

Lossless ;-)

- **Temporal Redundancy**

  Predictive coding: temporal differences and differences after motion compensation

- **Spatial Redundancy**

  Transform coding (Discrete Cosine Transform, DCT)

- **Statistical Redundancy**

  Huffman entropy coding

Lossy ;-(

- **Irrelevancy**

  Quantization of DCT coefficients

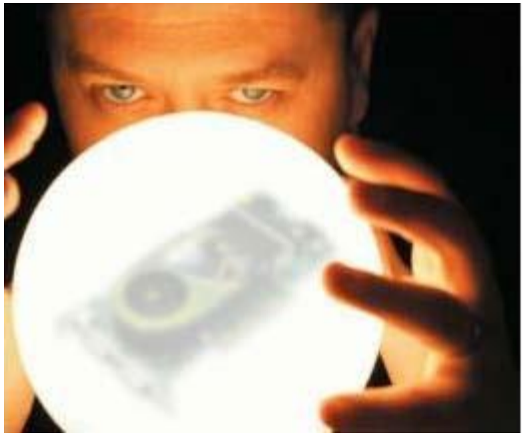
Exploiting

Temporal Redundancy

Temporal Prediction and Prediction Error

- **Temporal prediction is based on the principle that, locally, each image may be represented using as reference a part of some preceding image, typically the previous one.**
- **The prediction quality strongly determines the compression performance since it defines the amount of information to code and transmit, this means the energy of the error/difference signal called *prediction error*.**
- **The lower is the prediction error, the lower is the information/energy to transmit and thus**
  - Better quality may be achieved for a certain available bitrate
  - Lower bitrate is needed to achieve a certain video quality

H.261 Temporal Prediction

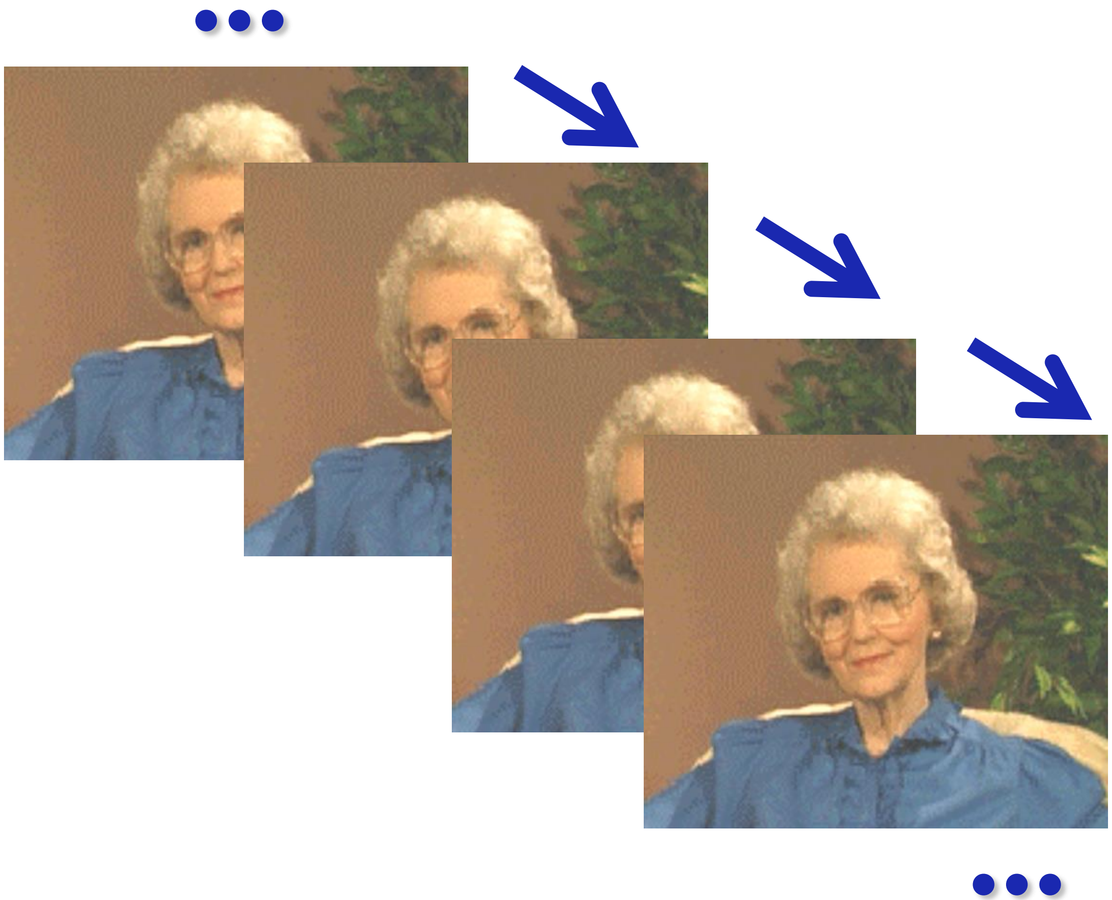


H.261 includes two temporal prediction tools which have both the target to eliminate/reduce the temporal redundancy in the PCM video signal (motion compensation works on top of the temporal differences):

Temporal Differences

Motion Estimation and Compensation

Temporal Redundancy: Sending the Differences



Only the new information in the next image (*this means what changes from the previous image*) is sent !

The previous (decoded) image works as a simple prediction of the current image.

There are no losses in this coding process!

TÉ
LIS

Frame no. 52 original

Frame no. 53 original

Erro sem compensacao de movimento

Frame no. 2 original

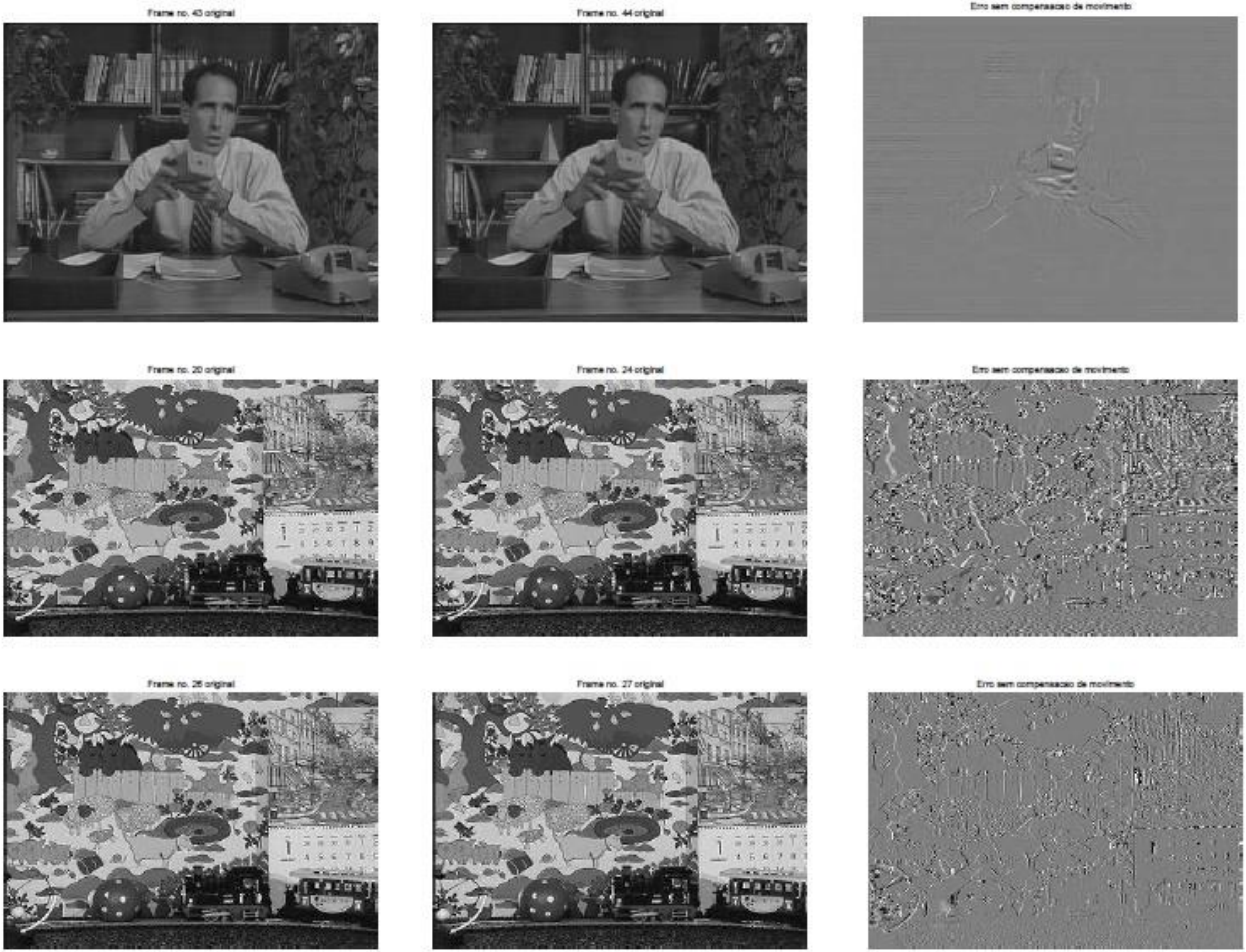
Frame no. 6 original

Erro sem compensacao de movimento

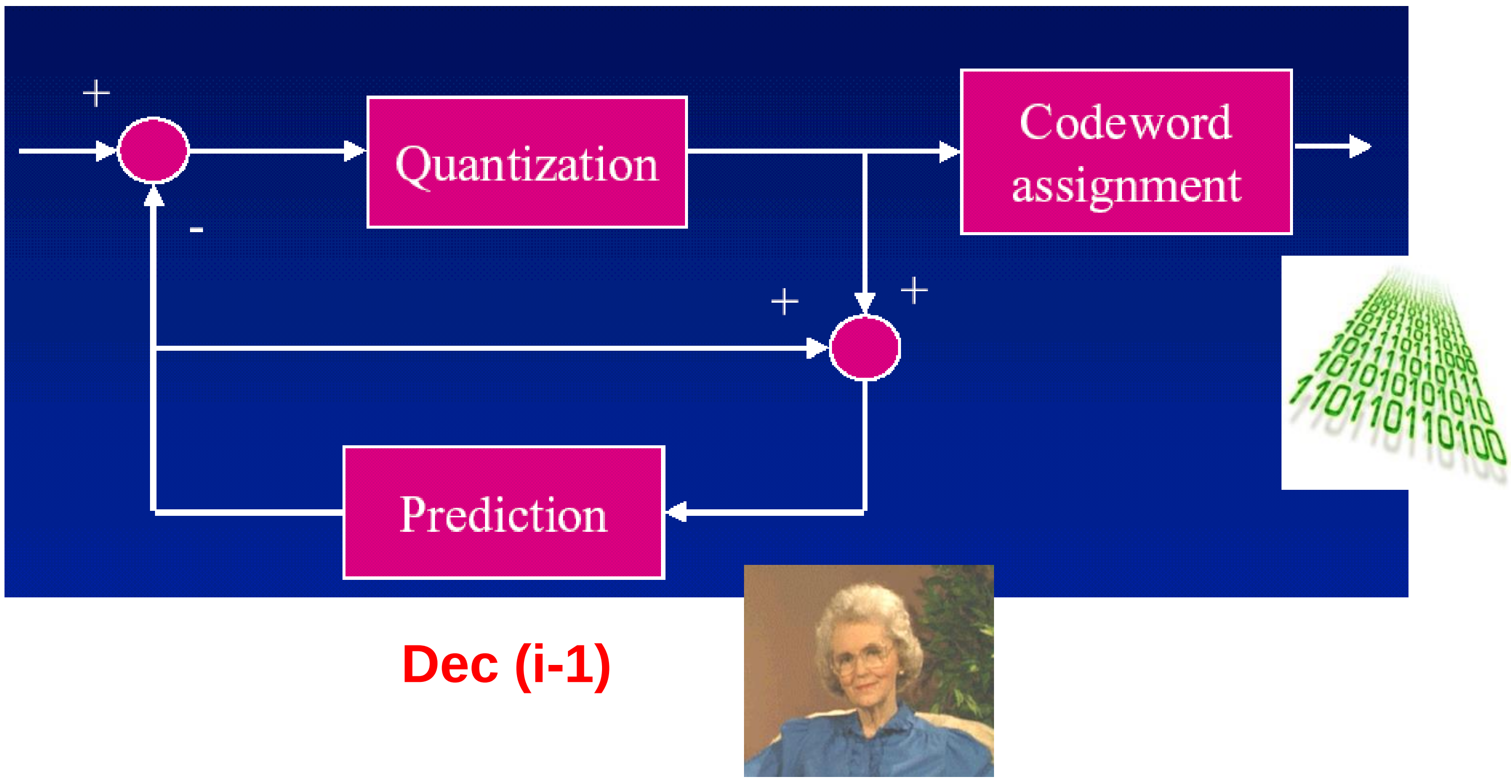
Frame no. 4 original

Frame no. 5 original

Erro sem compensacao de movimento

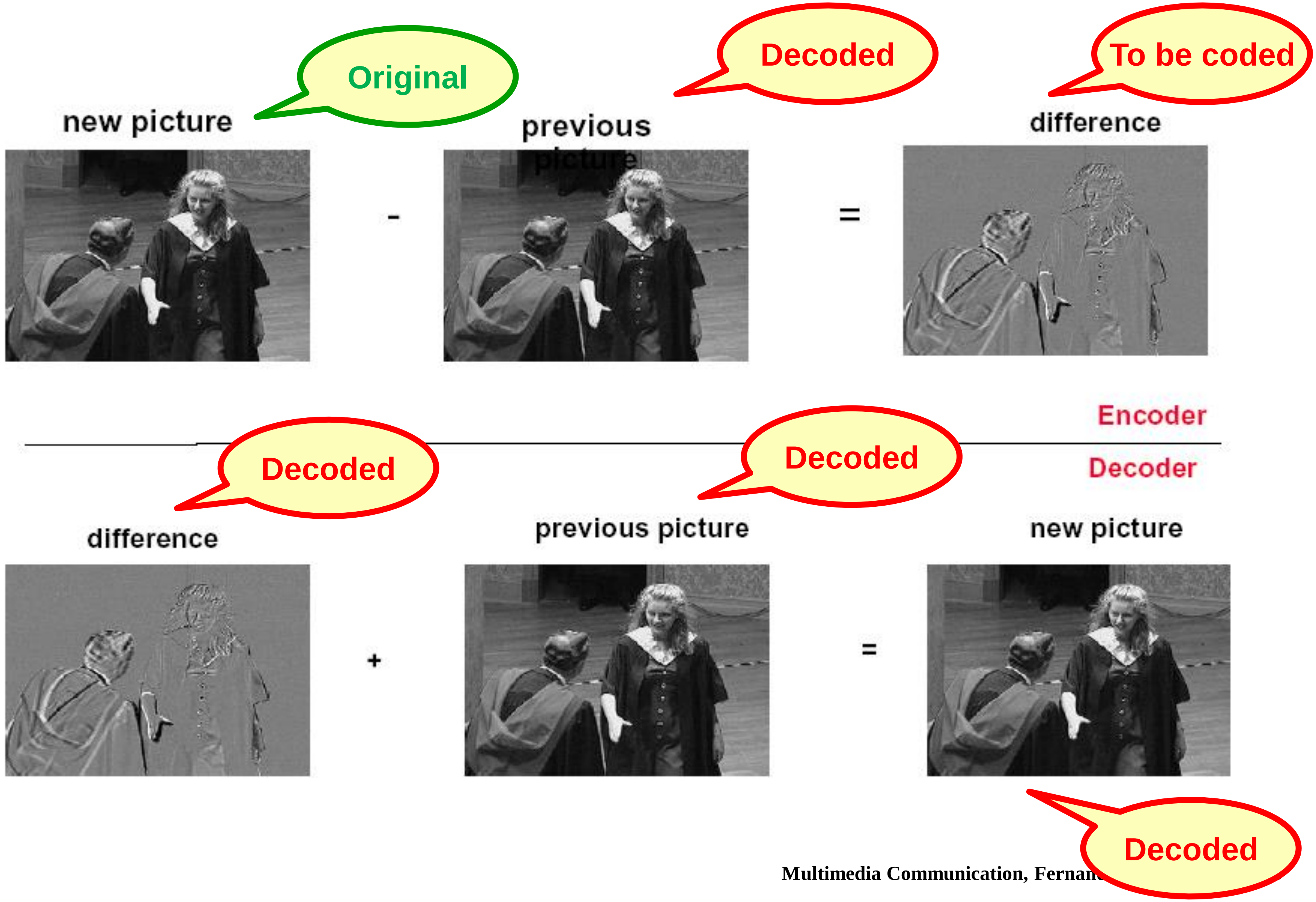

Frame no. 43 original
Frame no. 44 original
Erro sem compensacao de movimento
Frame no. 20 original
Frame no. 24 original
Erro sem compensacao de movimento
Frame no. 26 original
Frame no. 27 original
Erro sem compensacao de movimento

Predictive or Differential Coding: Basic Scheme

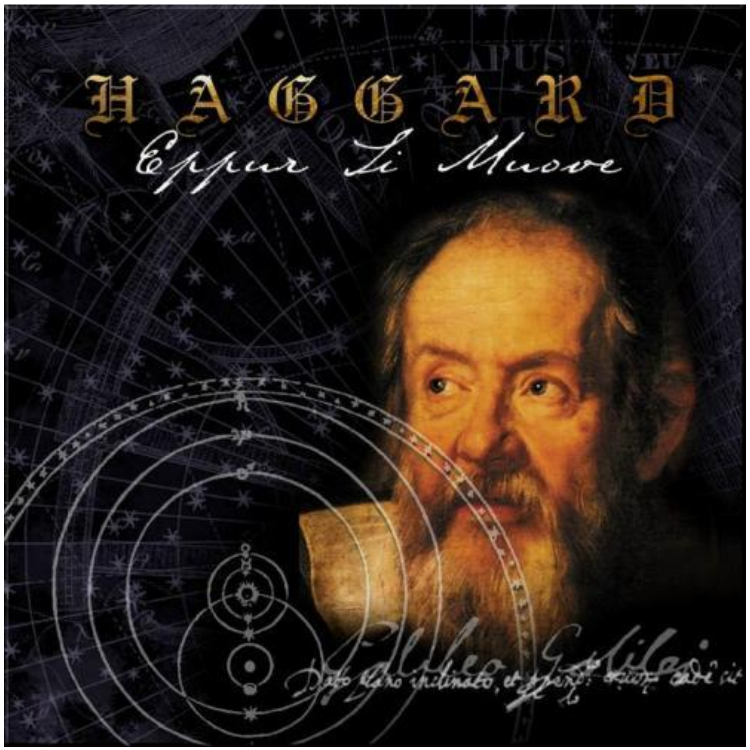


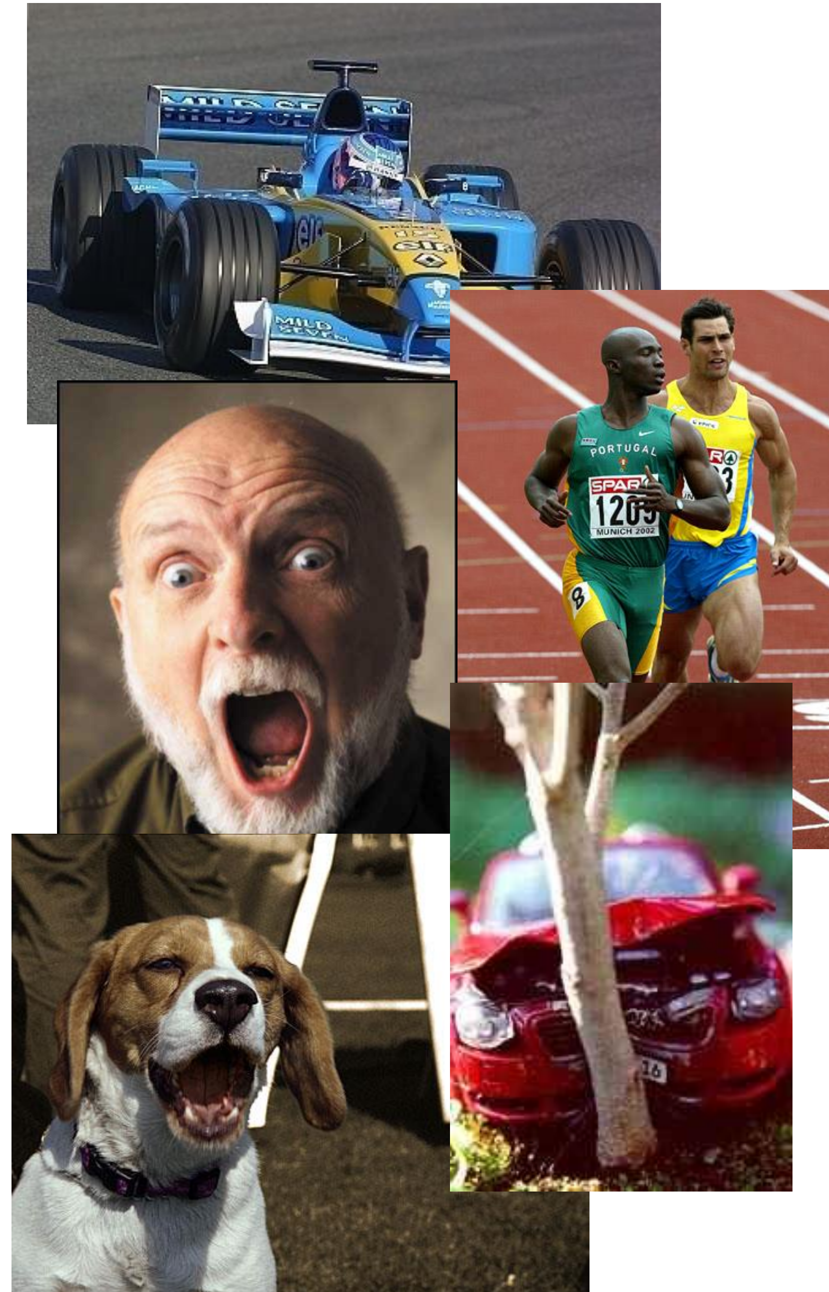
In H.261, there is no quantization in the temporal domain (but there is in the frequency/DCT domain).

Coding and Decoding ...



Eppur Si Muove …





TÉCNICO LISBOA

Motion Estimation and Compensation

Motion estimation and compensation have the target to improve the temporal predictions for each image zone by detecting, estimating and compensating the motion in the image.

- **The motion estimation process is not normative (*as all the encoder tools*) but the so-called *block matching* is the most used technique.**

- **In H.261, motion compensation is made at macroblock (MB) level. The usage of motion compensation for each MB is optional and decided by the encoder.**

Motion estimation implies a very high computational effort. This justifies the usage of fast motion estimation methods trying to reduce the complexity compared to full search motion estimation without significant quality losses (notably for real-time apps).

H.261: Coding Tools

Lossless ;-)

- **Temporal Redundancy**

  Predictive coding: temporal differences and differences after motion compensation

- **Spatial Redundancy**

  Transform coding (Discrete Cosine Transform, DCT)

- **Statistical Redundancy**

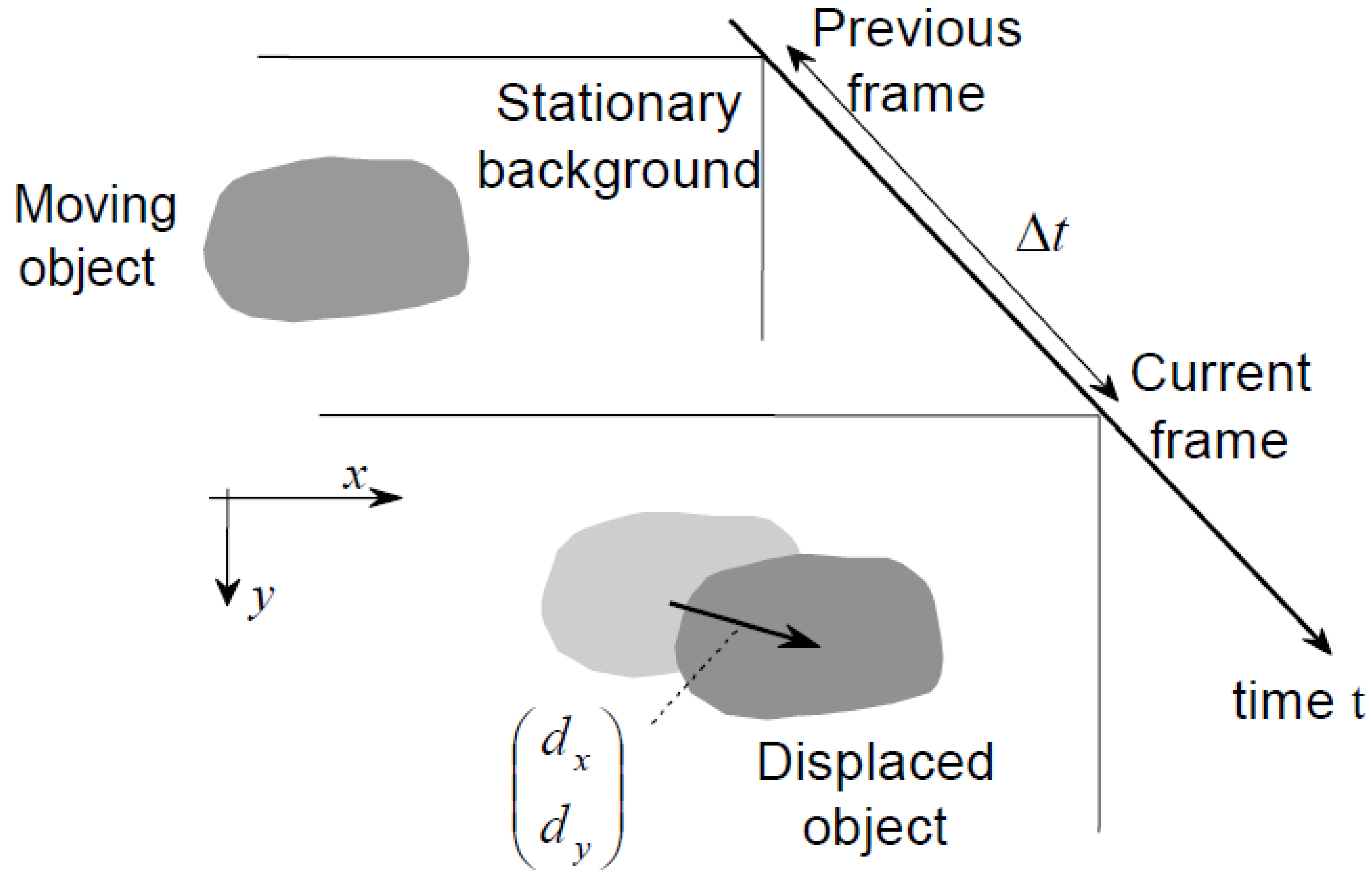
  Huffman entropy coding

Lossy ;-(

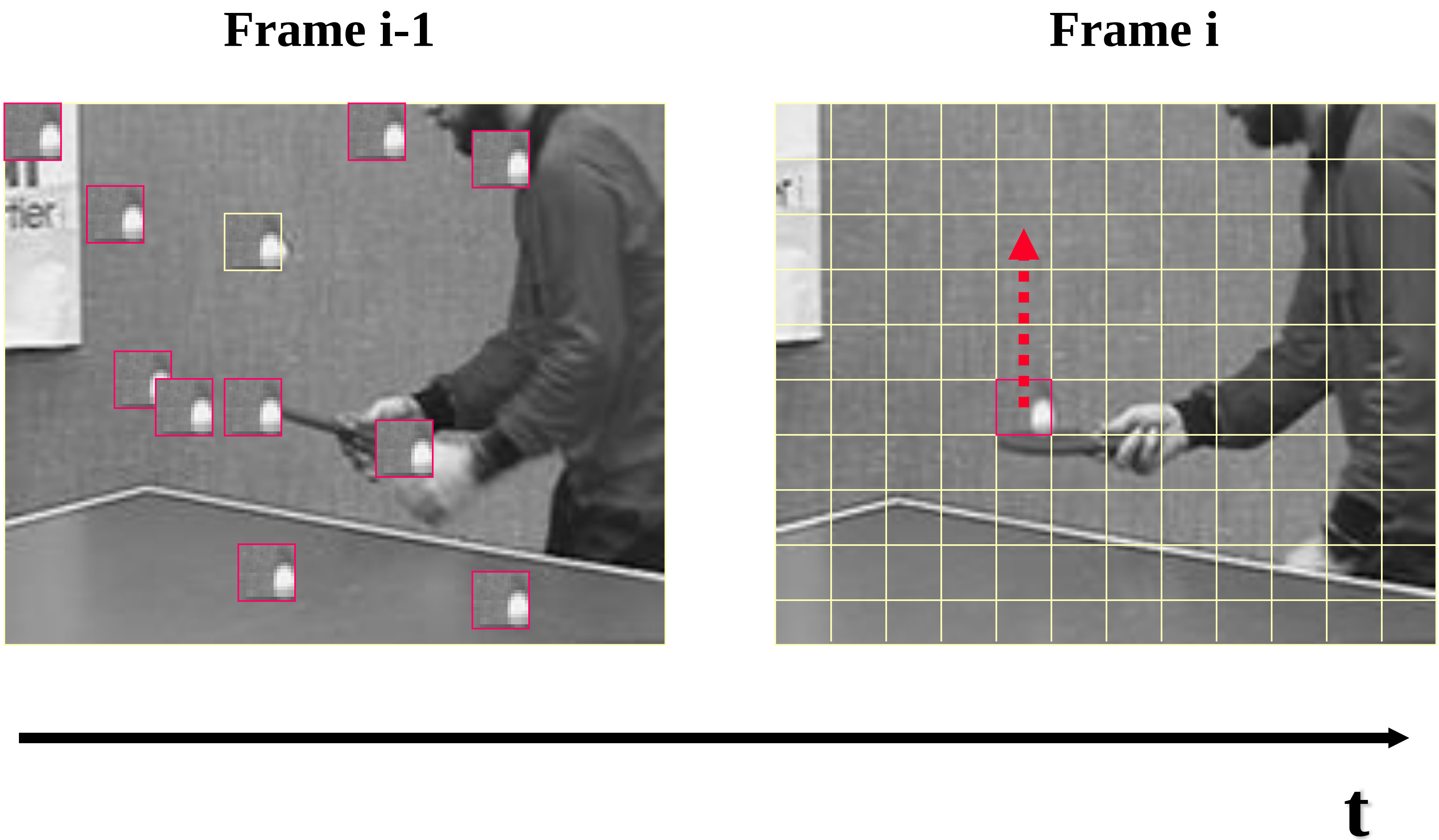
- **Irrelevancy**

  Quantization of DCT coefficients

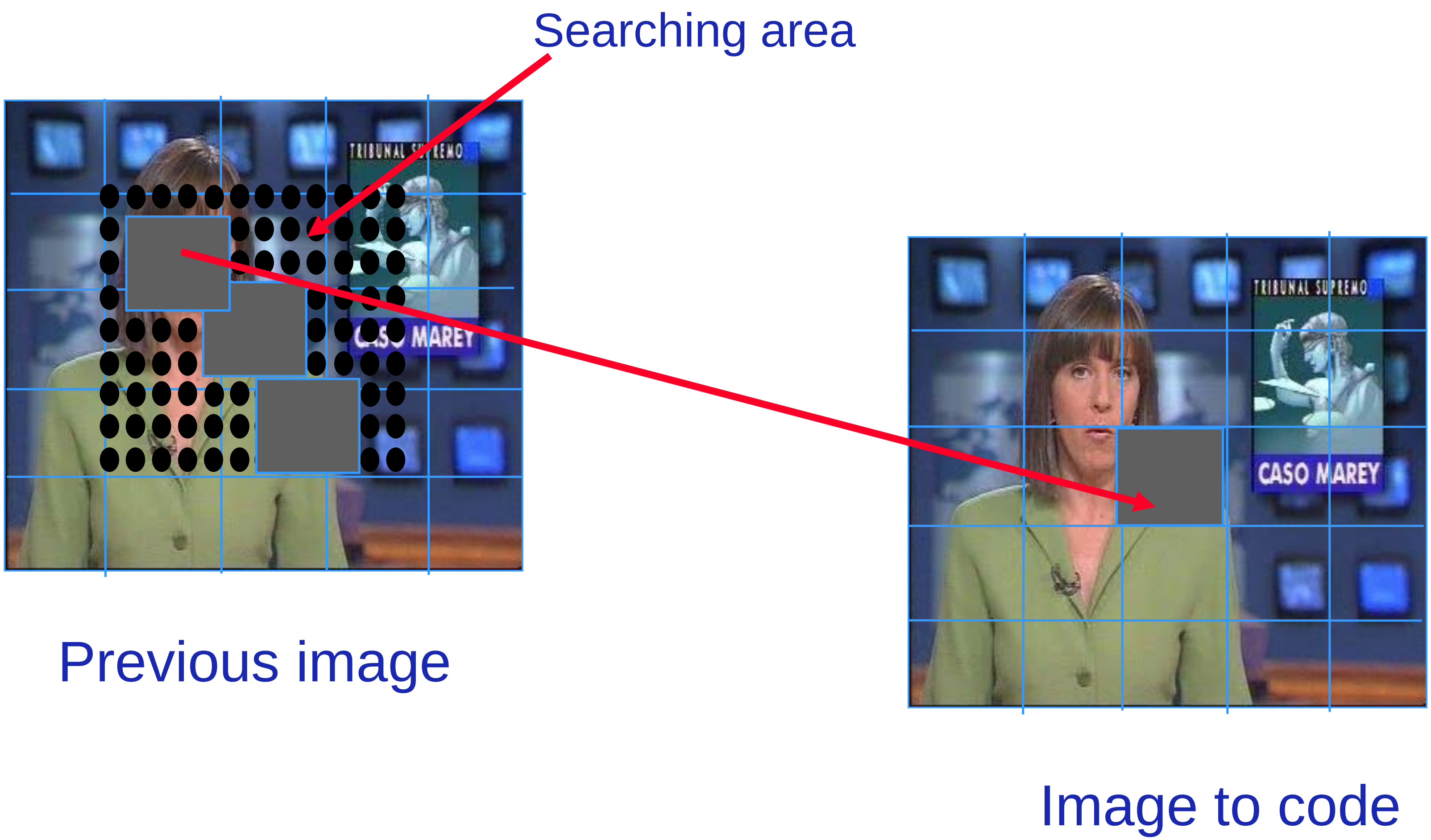
Motion in Action …



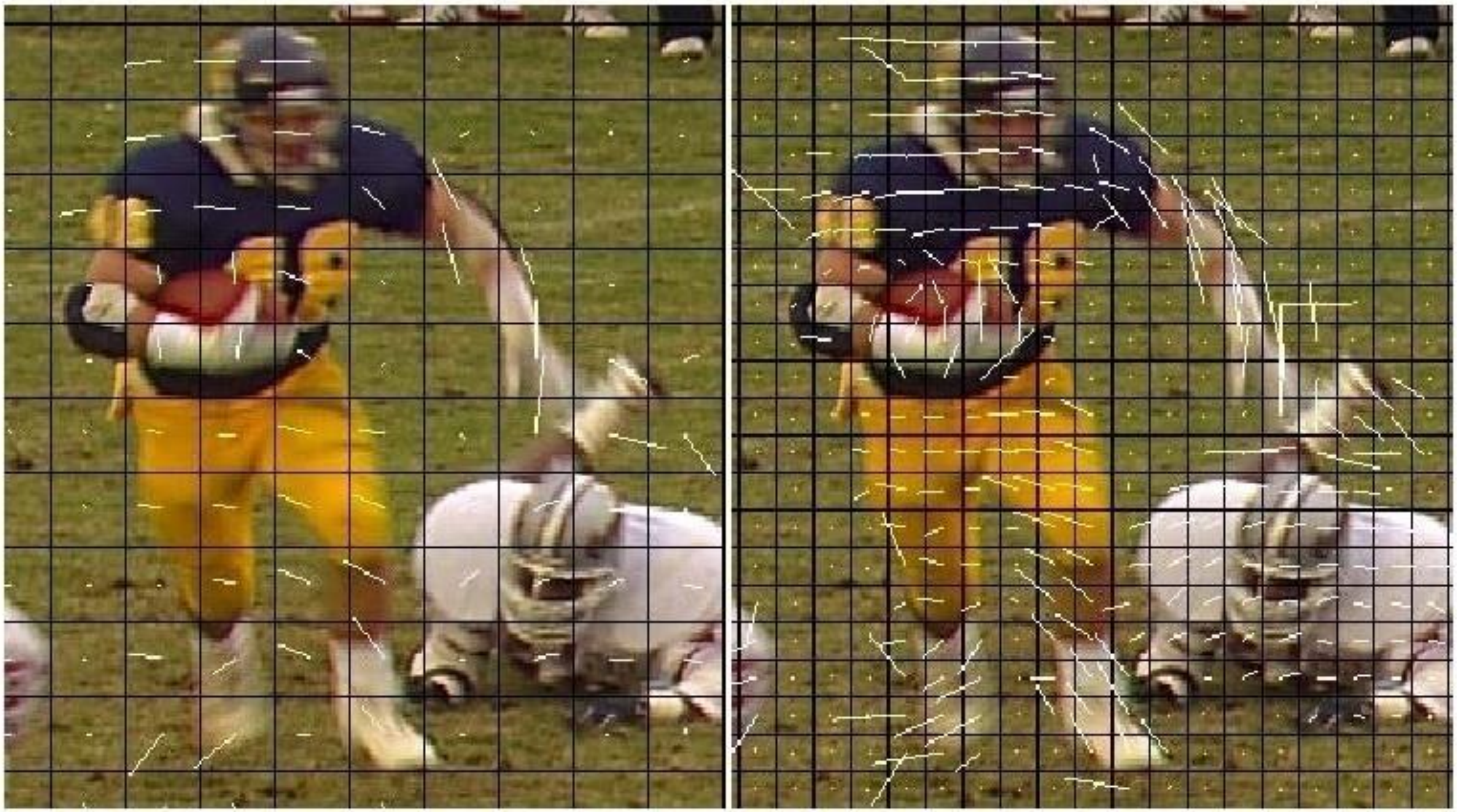
Temporal Redundancy: Motion Estimation



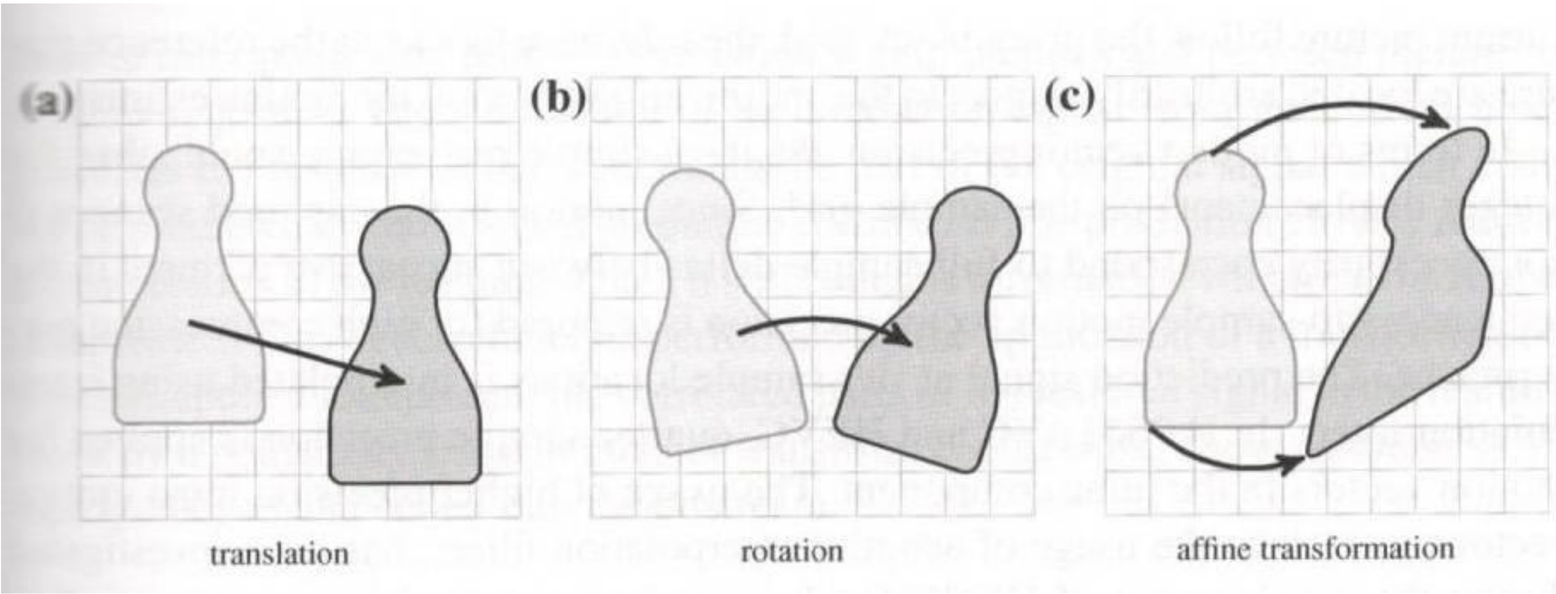
Motion Search: Worthwhile Where ?



Motion Vectors at Different Spatial Resolutions



Motion is More than Translations !

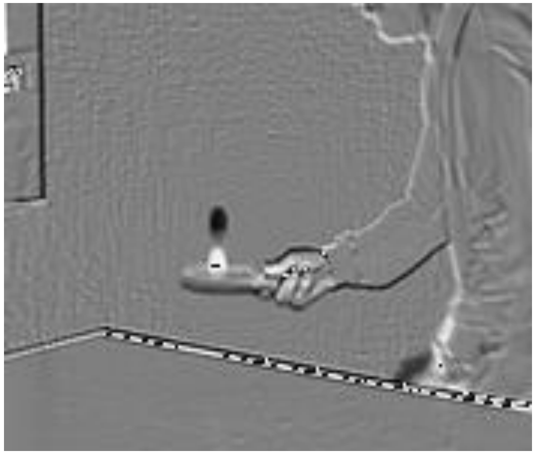


Clearly, a (translational) motion vector cannot represent well many types of motion … But it is still worthwhile !

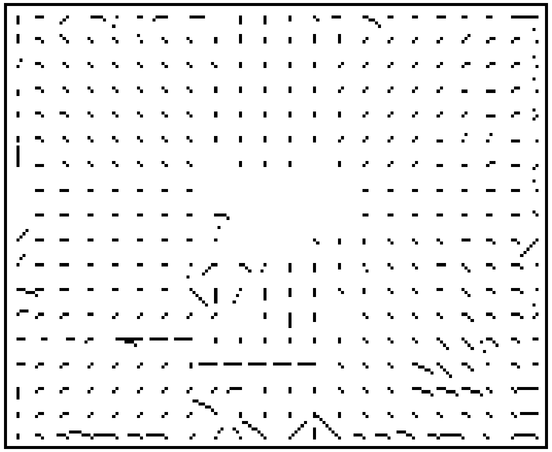
Motion Compensation: an Example

Image i

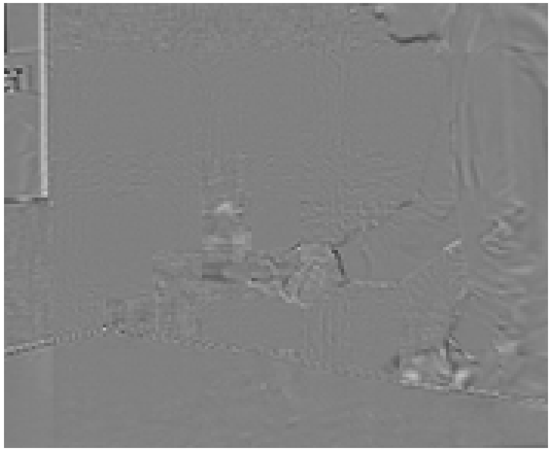
Image i-1



Diff. WITHOUT motion comp.

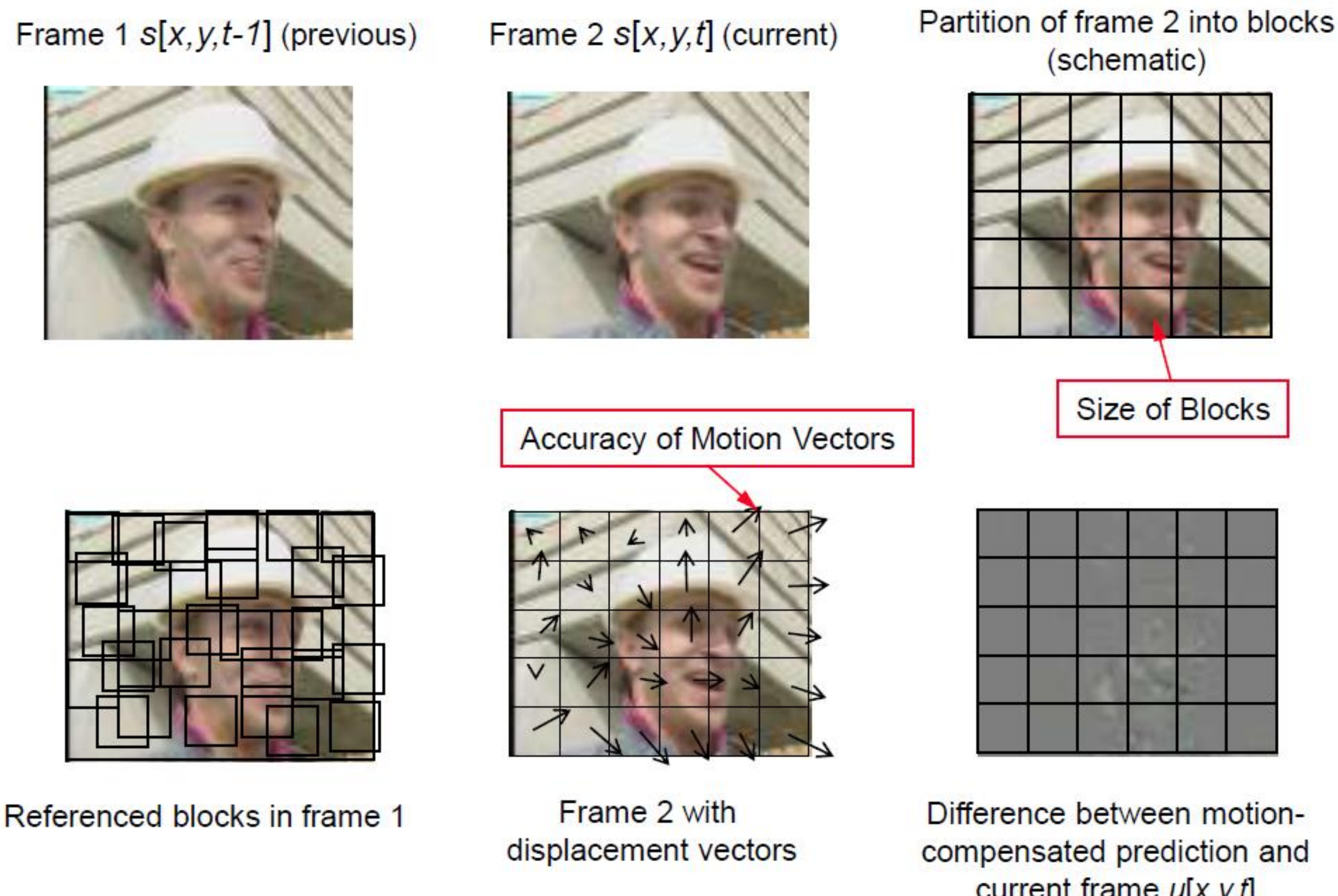


Motion vectors



Differences AFTER motion comp.

Coding with Motion Compensation …



Before and After Motion Compensation …

Frame no. 52 original

Frame no. 53 original

Erro sem compensacao de movimento

Estimacao do frame no. 53

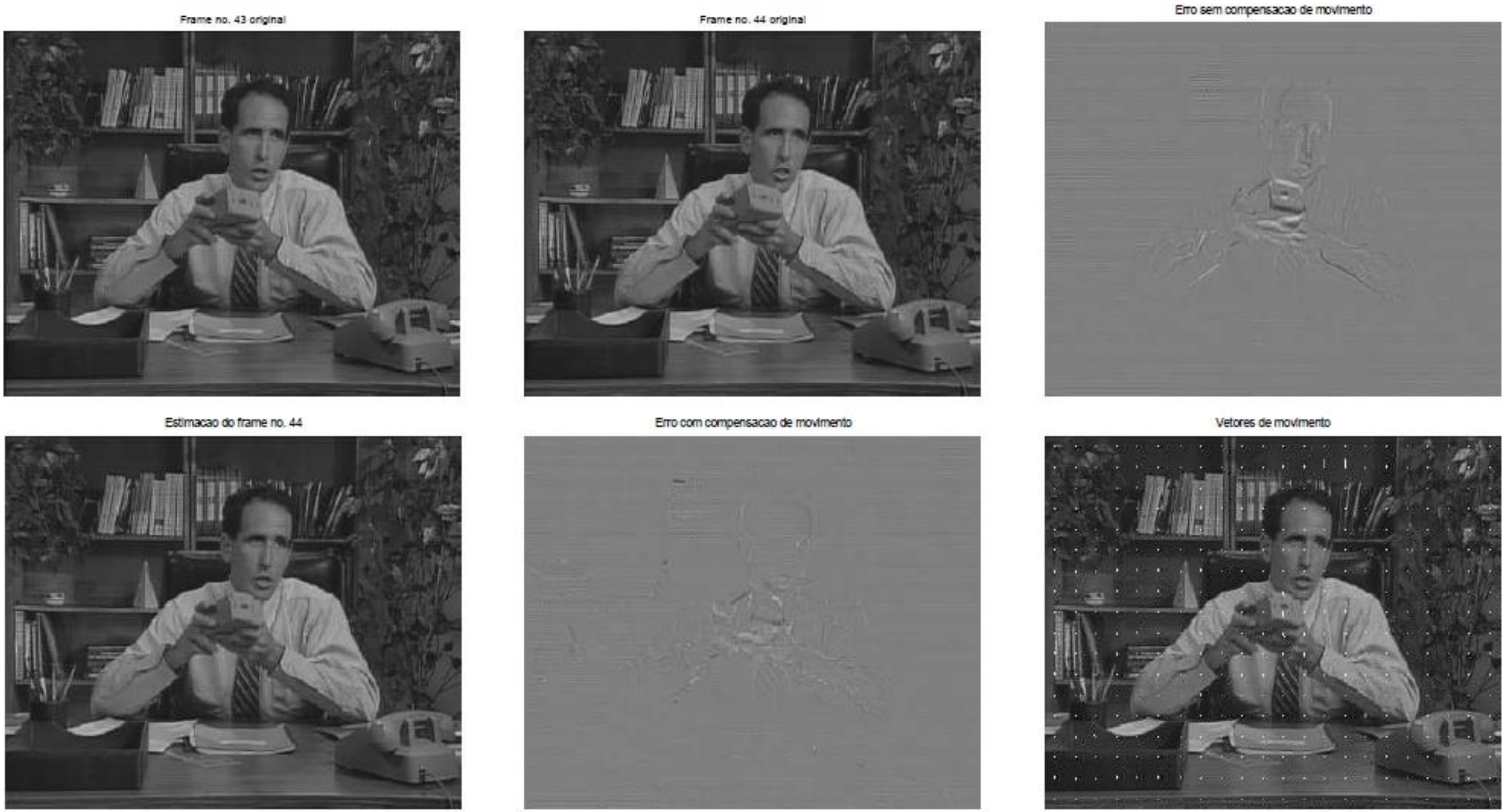
Erro com compensacao de movimento

Vetores de movimento

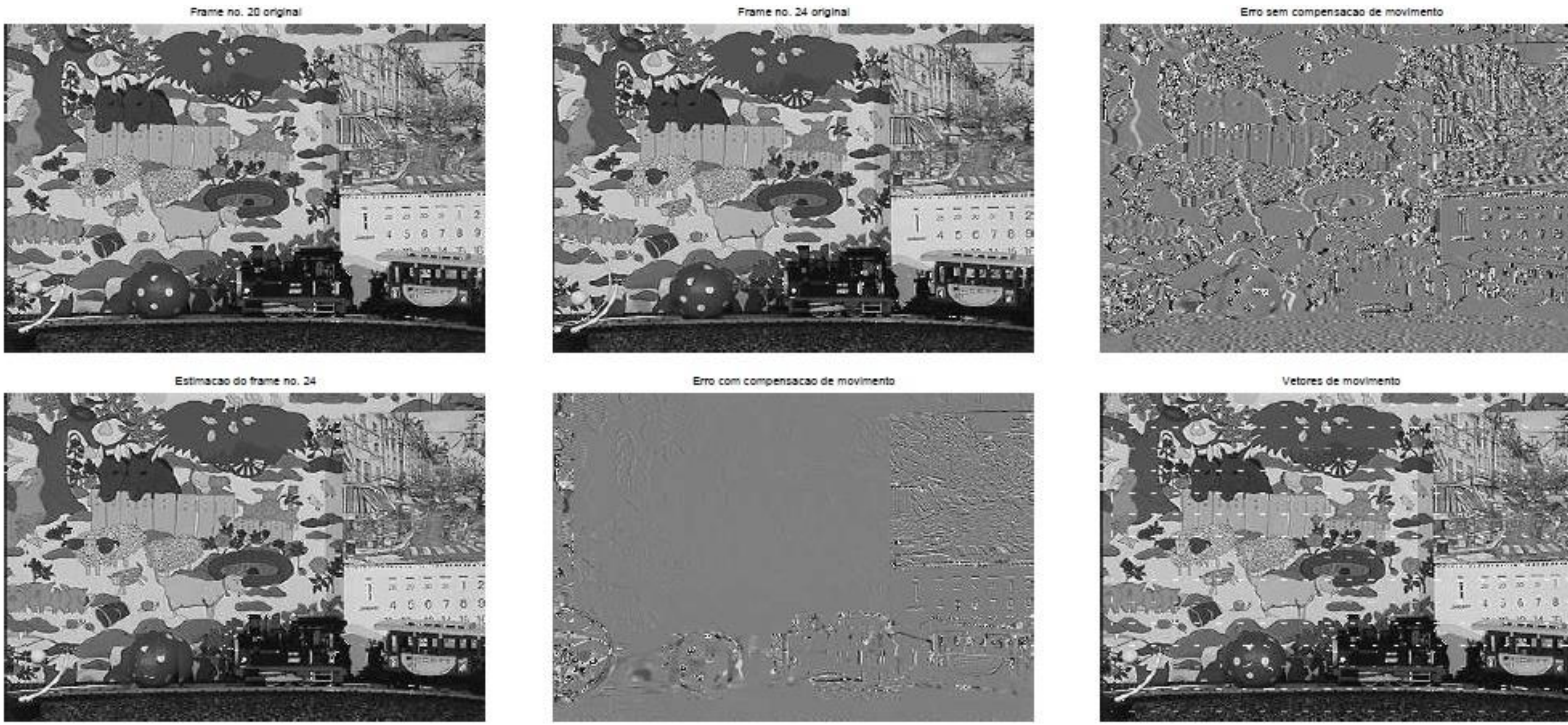
Before and After Motion Compensation …

Before and After Motion Compensation …

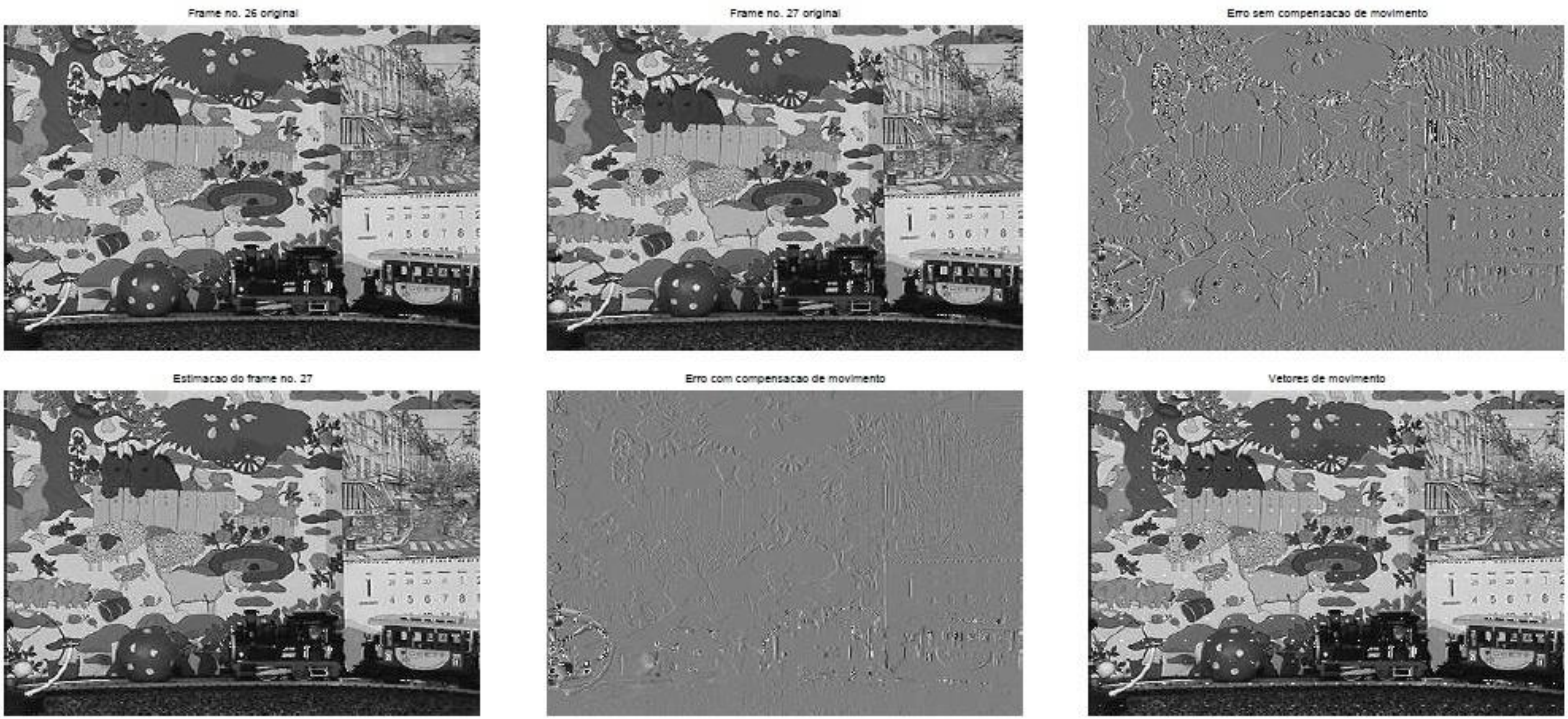
Before and After Motion Compensation …



Before and After Motion Compensation …

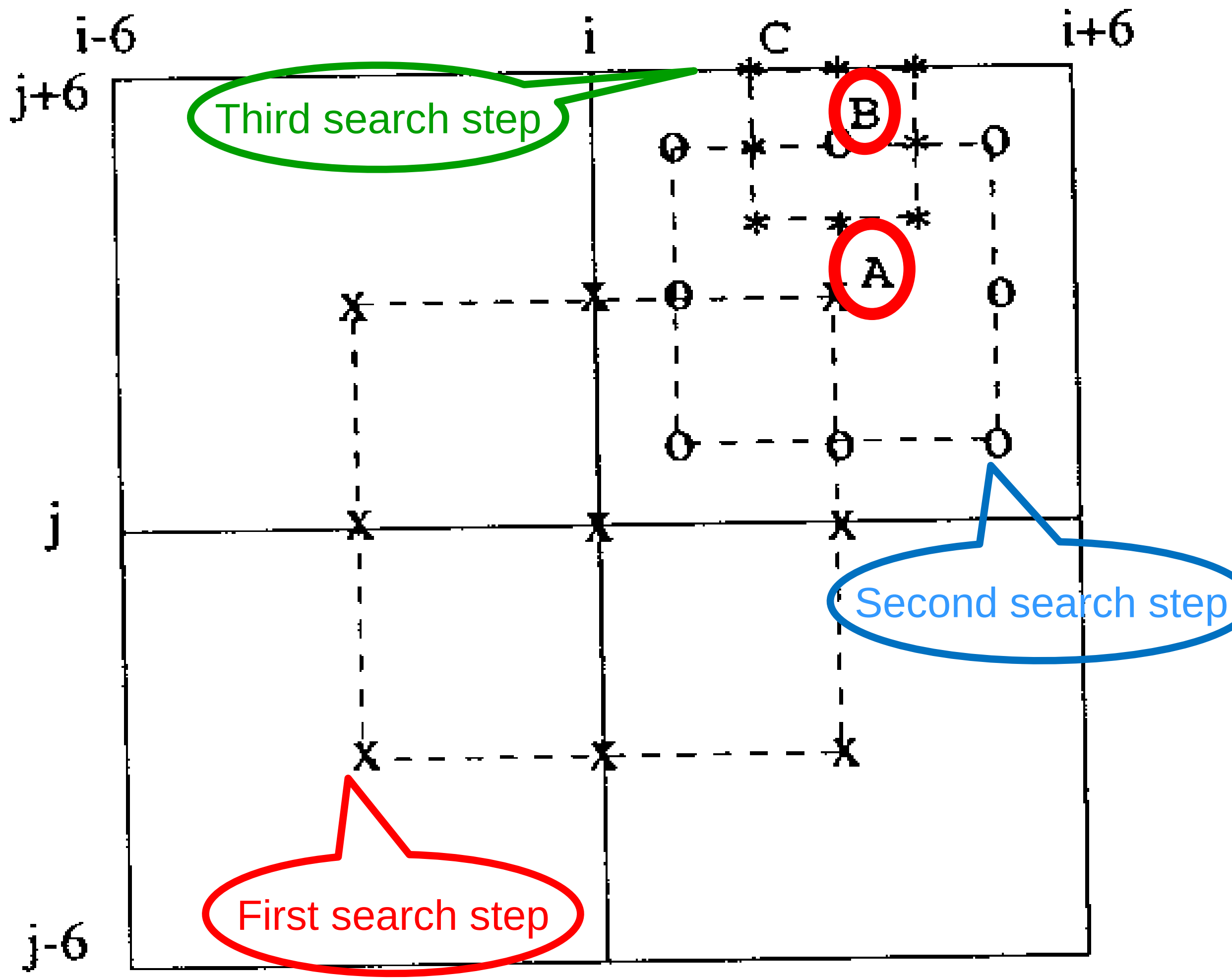


Before and After Motion Compensation …



Fast Motion Estimation: Three Steps Motion Estimation Algorithm

Fast motion estimation algorithms offer much lower complexity than full search at the cost of some small quality reduction since predictions are less optimal and thus the prediction error is (slightly) higher !

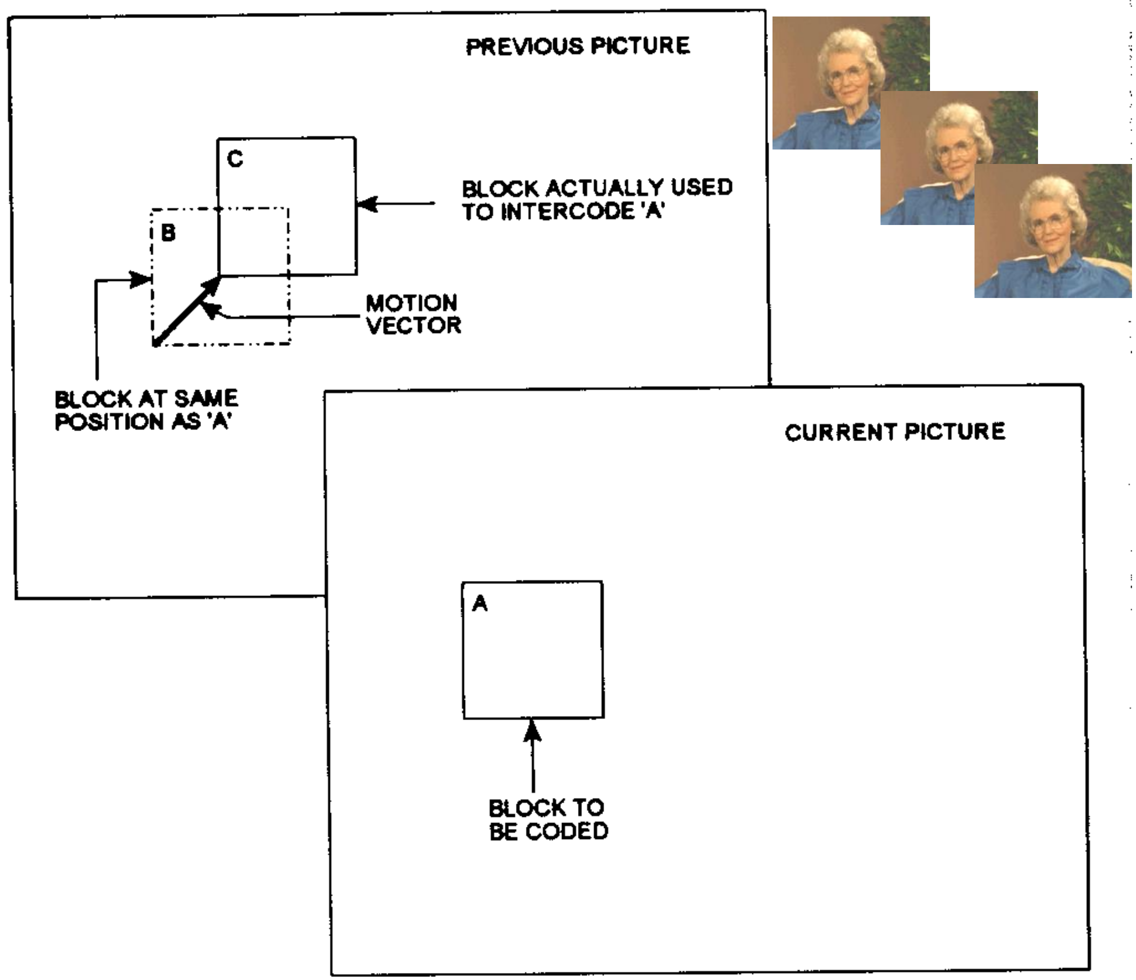


Predicting in Time … With or Without Motion

Two main temporal prediction coding modes are available for each MB:

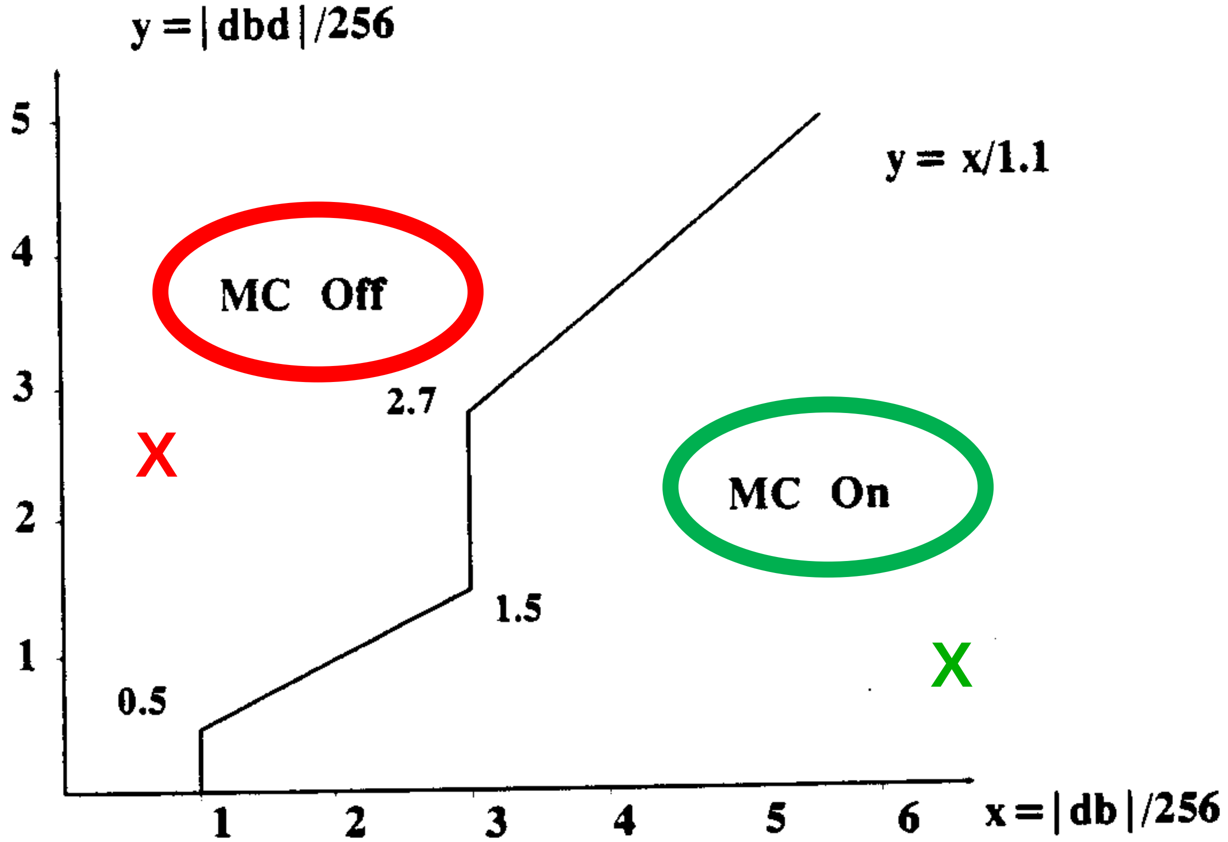
- <u>No motion vector:</u> Prediction from the same position in the previous frame
- <u>Using a motion vector:</u> Prediction from the previous frame

The encoder has to choose the *best compression deal* using some (non-normative) criteria !

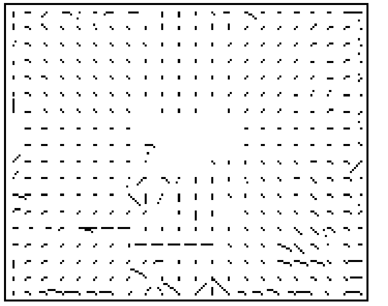


Motion Compensation Decision Characteristic Example (MB level)

db – *difference block*
dbd – *displaced block difference*



H.261 Motion Estimation Rules …



- **Number of MVs -** One motion vector may be transmitted for each macroblock (if the encoder so desires).
- **Range of MVs -** Motion vector components (x and y) may take values from -15 to + 15 pels, in the vertical and horizontal directions, only the integer values.
- **Referenced area -** Only motion vectors referencing areas within the reference (previously coded) image are valid.
- **Chrominance MVs -** The motion vector transmitted for each MB is used for the 4 luminance blocks in the MB. The chrominance motion vector is computed by dividing by 2 and truncating the luminance motion vector.
- **MV Semantics -** A positive value for the horizontal or vertical motion vector components means the prediction must be made using the samples in the previous image, spatially located to the right and below the samples to be predicted.

H.261 Motion Vectors Coding

- To exploit the redundancy between the motion vectors of adjacent MBs (in each image), each motion vector is differentially coded as the difference between the motion vector of the actual MB and its prediction, which in H.261 is the motion vector of the preceding MB.
- The motion vector prediction is null when no redundancy is likely to be present, notably when:
  - The actual MB is number 1, 12 or 23
  - The last transmitted MB is not adjacent to the actual MB
  - The preceding and contiguous MB did not use motion compensation

TÉCNICO LISBOA

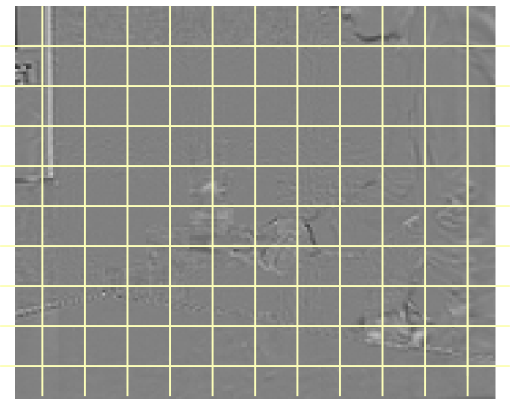
VLC table for MVD

MVD	Code
−16 & 16	0000 0011 001
−15 & 17	0000 0011 011
−14 & 18	0000 0011 101
−13 & 19	0000 0011 111
−12 & 20	0000 0100 001
−11 & 21	0000 0100 011
−10 & 22	0000 0100 11
−9 & 23	0000 0101 01
−8 & 24	0000 0101 11
−7 & 25	0000 0111
−6 & 26	0000 1001
−5 & 27	0000 1011
−4 & 28	0000 111
−3 & 29	0001 1
−2 & 30	0011
−1	011
0	1
1	010
2 & −30	0010
3 & −29	0001 0
4 & −28	0000 110
5 & −27	0000 1010
6 & −26	0000 1000
7 & −25	0000 0110
8 & −24	0000 0101 10
9 & −23	0000 0101 00
10 & −22	0000 0100 10
11 & −21	0000 0100 010
12 & −20	0000 0100 000
13 & −19	0000 0011 110
14 & −18	0000 0011 100
15 & −17	0000 0011 010

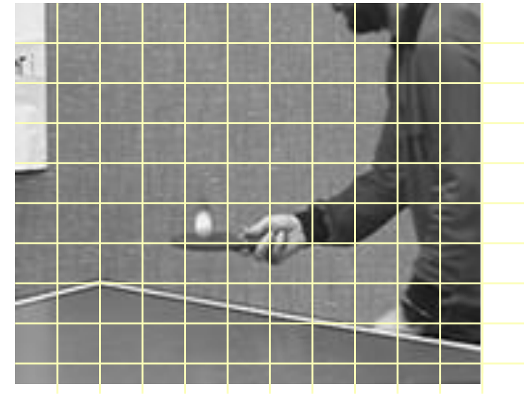
VLC Coding Table for (Differential) Motion Vectors

Inter Versus Intra Coding

In H.261, the MBs are coded either in Inter or Intra coding mode:



- **INTER CODING MODE** – To be used when there is substantial temporal redundancy; may imply the usage or not of motion compensation, i.e. *Inter+MC* and *Inter+noMC*.



- **INTRA CODING MODE** – To be used when there is NO substantial temporal redundancy; no temporal predictive coding is used in this case ('absolute' coding like in JPEG is used to exploit the spatial redundancy).

Exploiting Spatial Redundancy and Irrelevancy

After Temporal Redundancy, Spatial Redundancy

Original

Actual image

\+

\+

\-

Prediction error

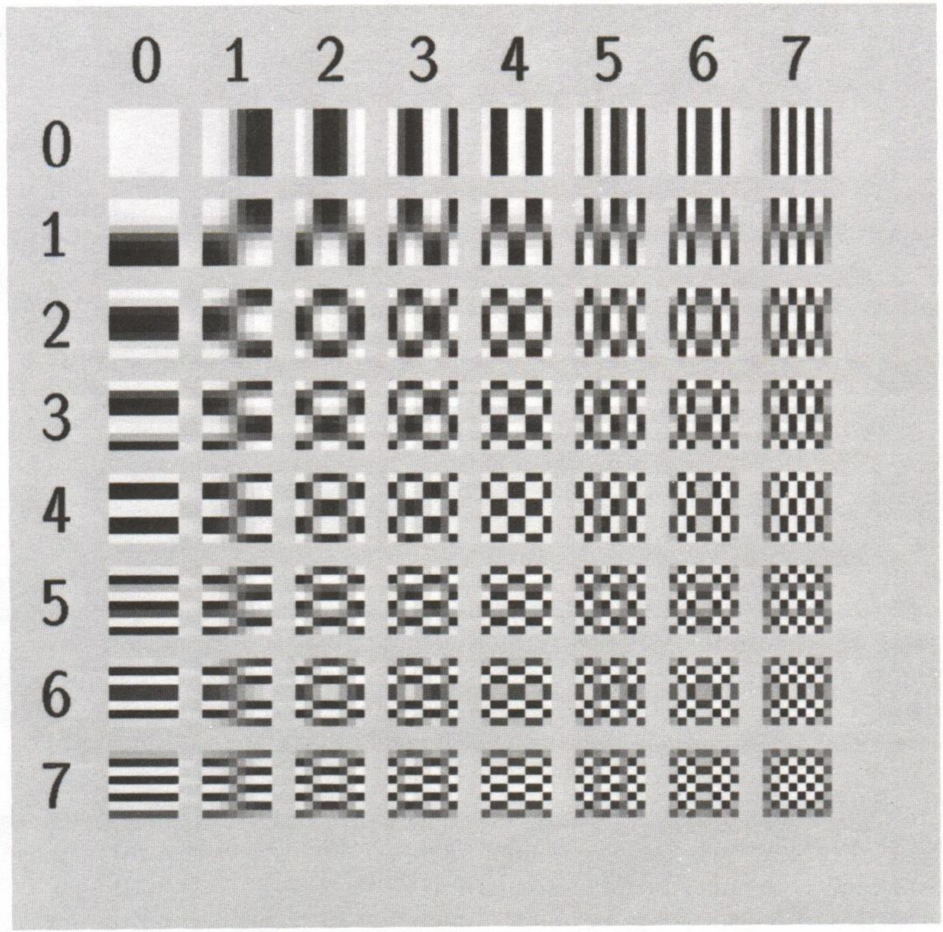
→ DCT Transform

To be coded

Prediction image, motion compensated at MB level

Decoded

Bidimensional DCT Basis Functions (N=8)



The DCT Transform in H.261

- **Block size** - In H.261, the DCT is applied to blocks with 8×8 samples. This value results from a trade-off between the exploitation of the spatial redundancy and the computational complexity involved.

- **Coefficients selection** - The DCT coefficients to transmit are selected using non-normative thresholds allowing the consideration of psychovisual criteria in the coding process, targeting the maximization of the subjective quality.

- **Quantization** - To exploit the irrelevancy in the original signal, the DCT coefficients to transmit for each block are quantized; as a prediction error is coded, an appropriate quantization step is used for all DCT coefficients (with the exception of the Intra MBs DC coefficient which always uses step 8)

- **Zig-Zag scanning** - Since the signal energy is compacted in the upper, left corner of the coefficients' matrix and the human visual system sensibility is different for the various frequencies, the quantized coefficients are zig-zag scanned to assure that more important coefficients are always transmitted before less important ones.

Scalar Quantization

- **The quantization process maps signal amplitudes to a predefined set of representative values. Quantization is an inherently non-linear lossy operation, which cannot be inverted.**

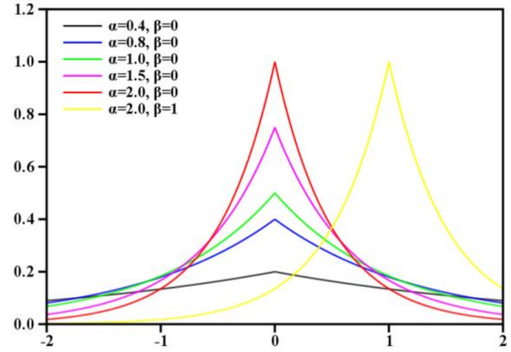
- **If individual values are quantized, the process is called *scalar quantization*.**

- **Quantization inserts signal degradation by removing signal information from the coded representation.**

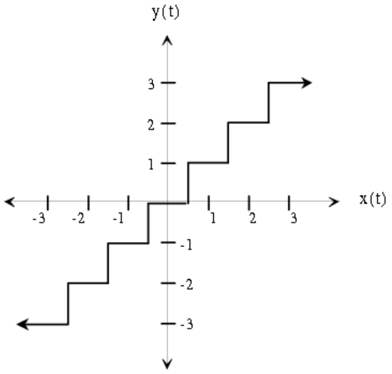
- **The design of the quantizer is driven by the probability distribution of the observed signal amplitudes, balancing the rate needed to encode the quantized values and the distortion introduced by mapping amplitude intervals to a defined reconstruction value.**

- **DCT transform coefficients can be modeled by a Laplacian probability density distribution, leading to quantizer designs used in standards.**

$$\frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$



Quantization at Work



- **Encoder:** The quantization performed at the encoder applies a function C(x) called a *classification rule* that selects an integer-valued class identifier called the *quantization index*.

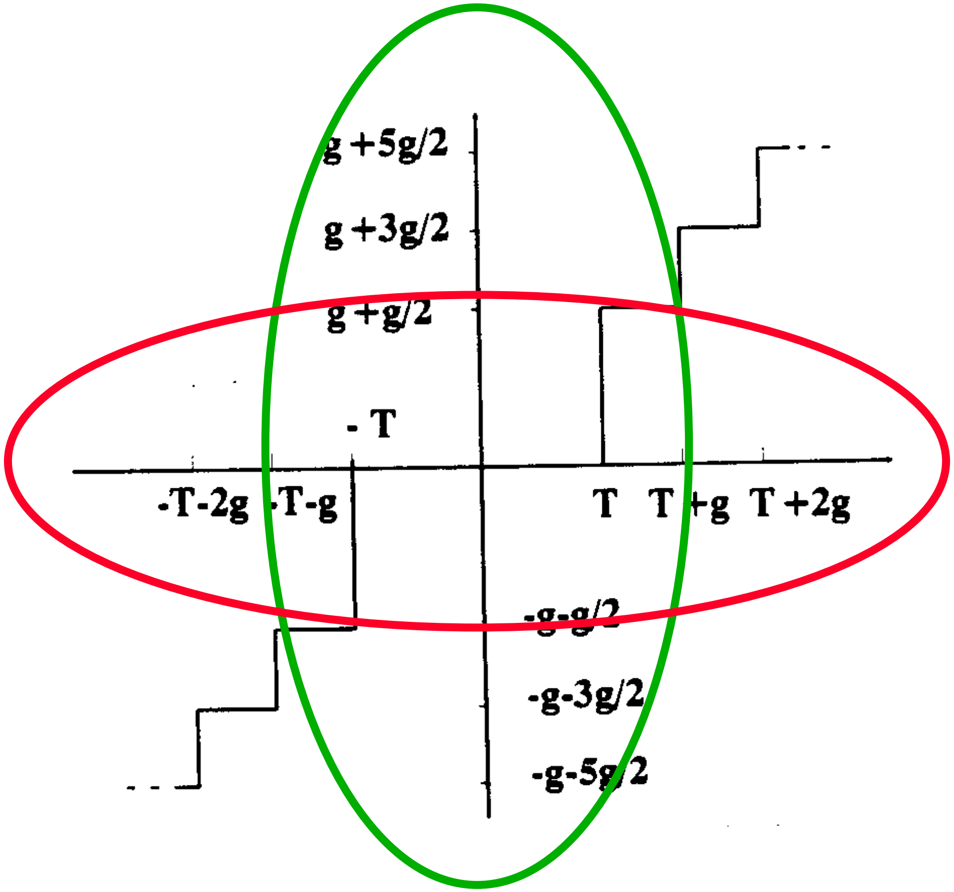
- **Decoder:** A second function, R(k), called a *reconstruction rule* produces a real valued decoded output Q(x) = R(C(x)) called a *reconstruction value*.

- **Solution:** A well know but rather simple quantizer reconstruction rule is the so-called nearly-uniform-reconstruction quantizer (NURQ). The reconstruction rule for a NURQ uses two parameters, a step size, *s*, and a non-zero offset parameter, *p*, and is defined as:

$$R(C(x)) = \text{sign}(C(x)) \times s \times (|C(x)|+p)$$

where *s* is the quantization step and *p* is an offset parameter. A typical value for *p* is ½.

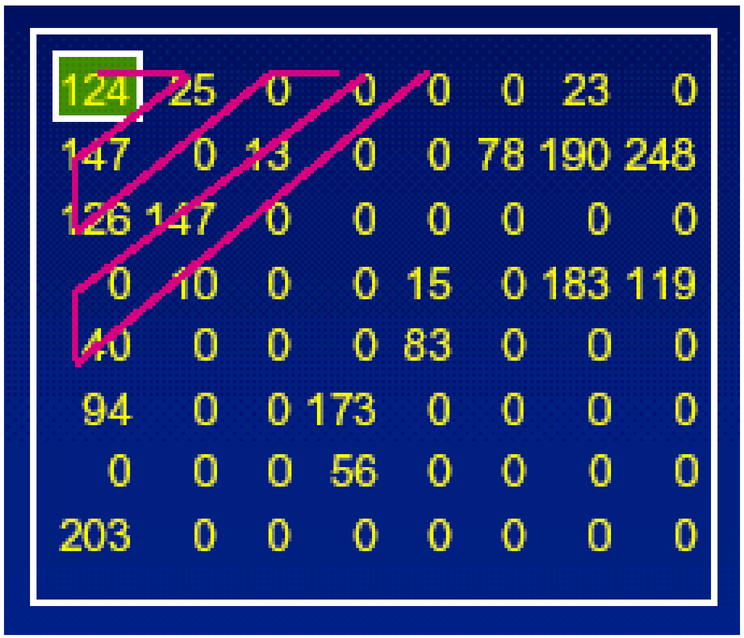
H.261 Quantization



Example quantization function

- H.261 uses as quantization steps all even values between 2 and 62 (31 quantizers available).
- Within each MB, all DCT coefficients are quantized with the same quantization step with the exception of the DC coefficient for Intra MBs which are always quantized with step 8.
- The usage of a same constant quantization step for all the AC DCT coefficients is motivated by the fact that an error (and not absolute sample values) is being coded.
- H.261 normatively defines the regeneration values for the quantized coefficients but not the decision values which may be selected to implement different quantization characteristics (uniform or not).

Serializing the DCT Coefficients



- **The transmission of the quantized DCT coefficients requires to send the decoder two types of information about the coefficients: their <u>position</u> and <u>quantization level</u> (for the selected quantization step).**

- **For each DCT coefficient to transmit, its position and quantization level are represented using a bidimensional symbol**

(*run, level*)

where the *run* indicates the number of null coefficients before the coefficient under coding, and the *level* indicates the quantized level of the coefficient.

Exploiting Statistical Redundancy

Statistical Redundancy: Entropy Coding

Entropy coding

CONVERTS SYMBOLS IN BITS !

Using the statistics of the symbols to transmit to achieve additional (lossless) compression by allocating in a clever way bits to the input symbol stream.

Which code is the best ?

- **A, B, C, D -> 00, 01, 10, 11**
- **A, B, C, D -> 0, 10, 110, 111**

TÉCNICO LISBOA

Huffman Coding

Huffman coding is one of the entropy coding tools allowing to exploit the fact that the symbols produced by the encoder model do not have equal probability.

- **To each generated symbol is attributed a codeword which size (in bits) is 'inversely' proportional to its probability.**
- **The usage of variable length codes implies the usage of an output buffer to 'smooth' the bitrate flow, if a synchronous channel is available.**
- **The increase in compression efficiency is 'paid' with an increase in the sensibility to channel errors.**

75	76	77	78	79	80	81	82
77	78	79	80	81	82	83	84
79	80	81	82	83	84	85	86
81	82	83	84	85	86	87	88
83	84	85	86	87	88	89	90
85	86	87	88	89	90	91	92
87	88	89	90	91	92	93	94
89	90	91	92	93	94	95	96

(a)

Prediction error

76	76	77	79	80	81	82	83
77	77	78	80	81	82	83	84
79	79	80	81	83	84	85	86
81	82	83	84	85	87	88	88
84	84	85	87	88	89	90	91
86	87	88	89	91	92	93	93
88	89	90	91	92	94	95	95
89	90	91	92	93	95	96	96

(f)

Decoded prediction error

684	-19	-1	-2	0	-1	0	-1
-37	0	-1	0	0	0	0	-1
0	0	0	0	0	0	0	0
-4	-1	-1	-1	-1	0	-1	-1
0	0	0	0	0	0	0	0
-2	0	0	-1	0	-1	0	-1
0	0	0	0	-1	-1	-1	-1
-1	-1	-1	0	-1	0	-1	0

(b)

To be coded DCT coefficients

688	-21	0	0	0	0	0	0
-39	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(e)

Decoded DCT coefficients

86	-3	0	0	0	0	0	0
-6	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(c)

Quantized DCT coefficients (levels)

RUN	LEVEL	CODE
0	86	01010110
0	-3	001011
0	-6	001000011
	EOB	10

TOTAL CODE LENGTH = 25

(d)

Coding bits

Combining the Tools …

H.261: Coding Tools

Lossless ;-)

- **Temporal Redundancy**

  Predictive coding: temporal differences and differences after motion compensation

- **Spatial Redundancy**

  Transform coding (Discrete Cosine Transform, DCT)

- **Statistical Redundancy**

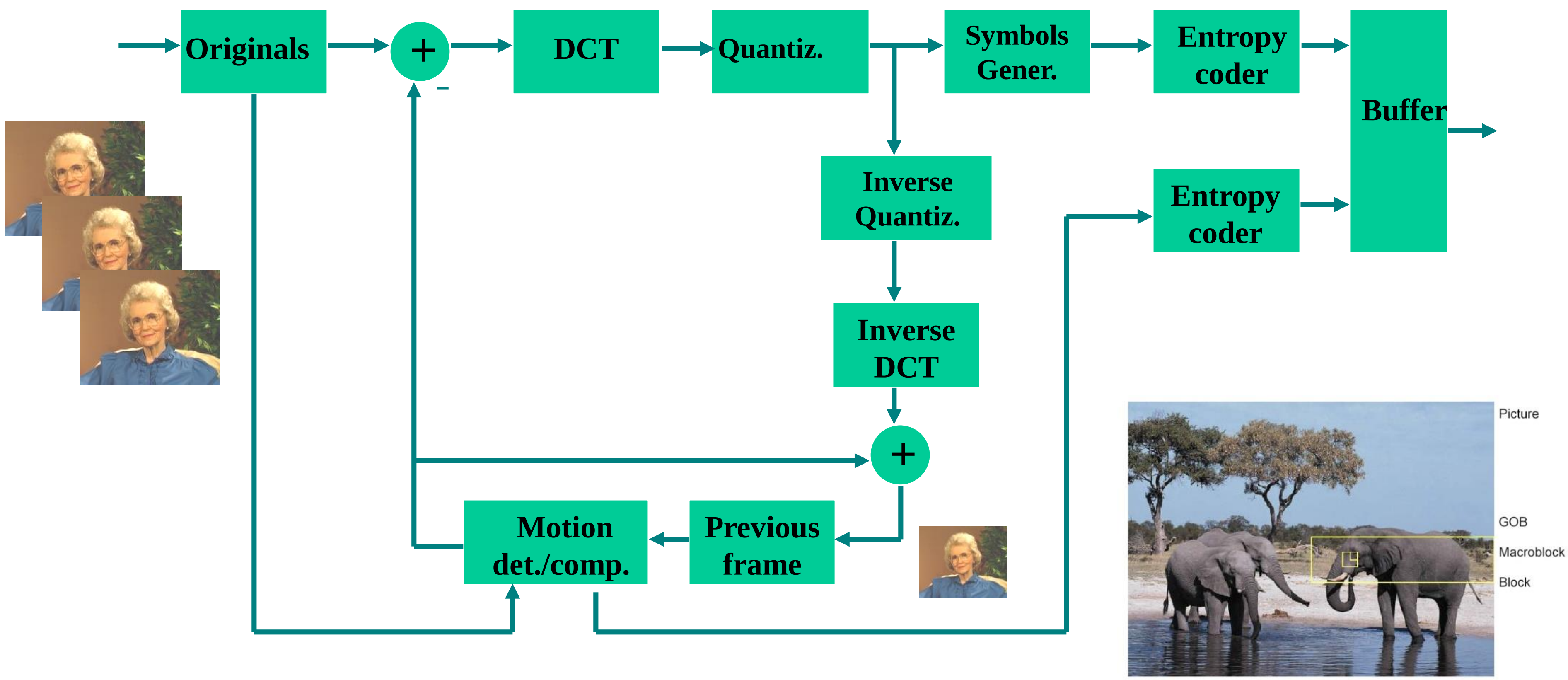
  Huffman entropy coding

Lossy ;-(

- **Irrelevancy**

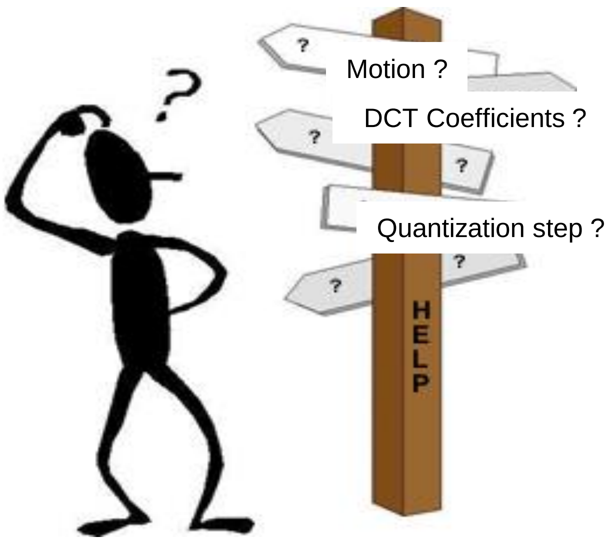
  Quantization of DCT coefficients

Encoder: the Winning Cocktail !



The Importance of Well Choosing !

To well exploit the redundancy and irrelevancy in the video sequence, the encoder has to select:



- **Which coding tools are used for each MB, depending of its characteristics;**
- **Which set of symbols is the best to represent each MB, e.g. motion vector and DCT coefficients.**

While the encoder has the mission to take important decisions and make critical choices, the decoder is a ‘slave’, limited to follow the ‘orders’ sent by the encoder; decoder intelligence is only shown for error concealment.

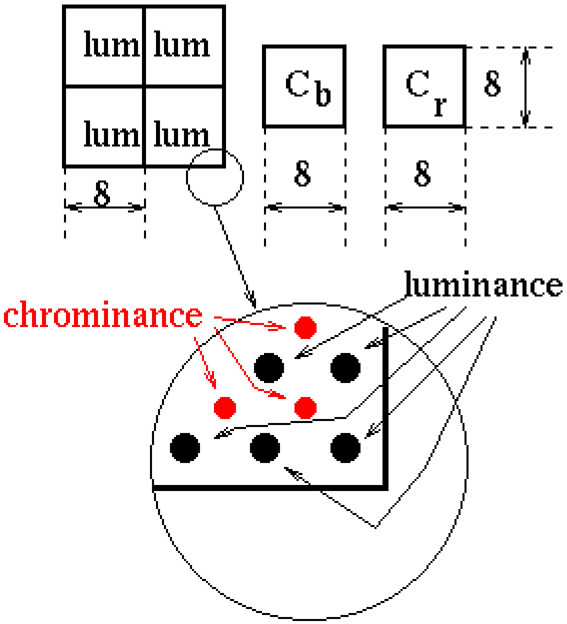
Macroblock Classification: Using the Toolbox

- **Macroblocks are the basic coding unit since it is at the macroblock level that the encoder selects the coding tools to use.**

- **Each coding tool is more or less adequate to a certain type of content and, thus, MB; it is important that, for each MB, the right coding tools are selected.**

- **Since H.261 includes several coding tools, it is the task of the encoder to select the best tools for each MB; MBs are thus classified following the tools used for their coding.**

- **When only spatial redundancy is exploited, MBs are INTRA coded; if also temporal redundancy is exploited, MBs are INTER coded.**



Macroblock Classification Table

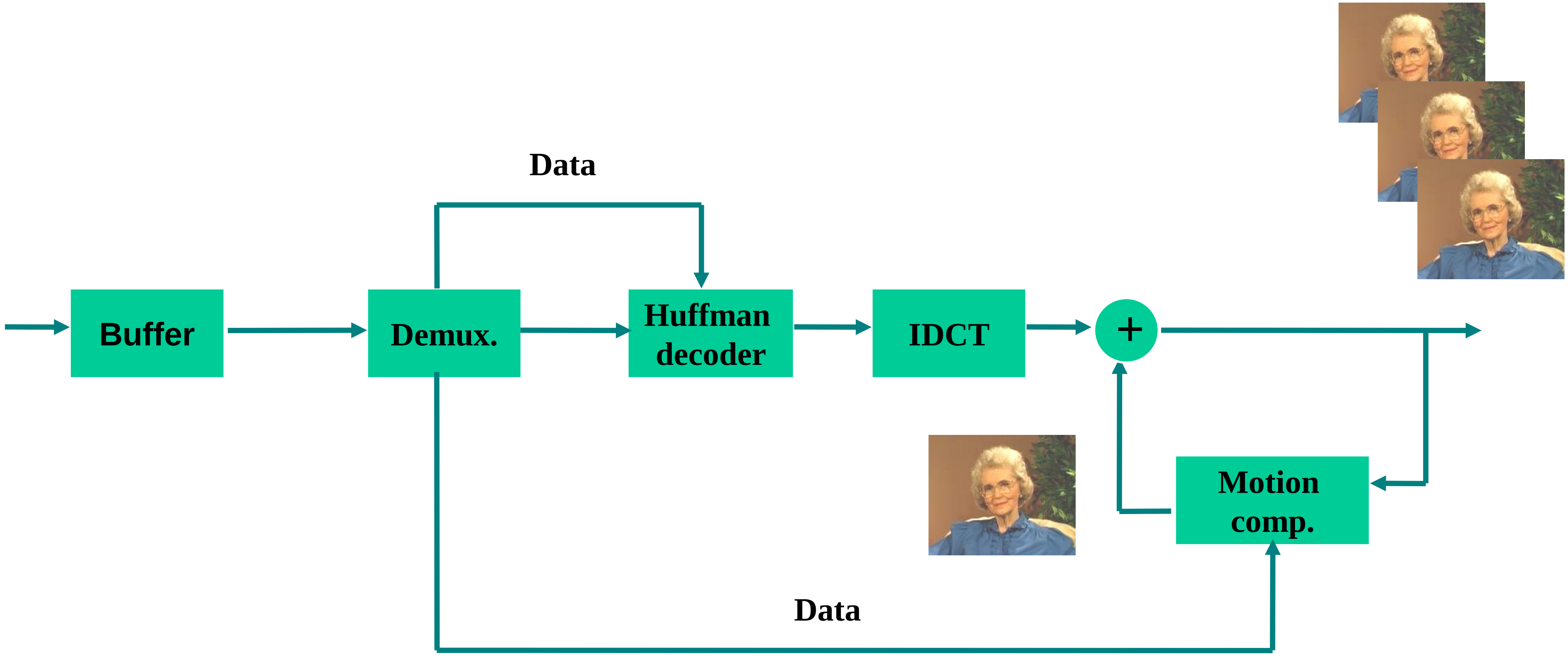
VLC table for MTYPE

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000 001
Inter			x	x	1
Inter	x		x	x	0000 1
Inter + MC		x			0000 0000 1
Inter + MC		x	x	x	0000 0001
Inter + MC	x	x	x	x	0000 0000 01
Inter + MC + FIL		x			001
Inter + MC + FIL		x	x	x	01
Inter + MC + FIL	x	x	x	x	0000 01

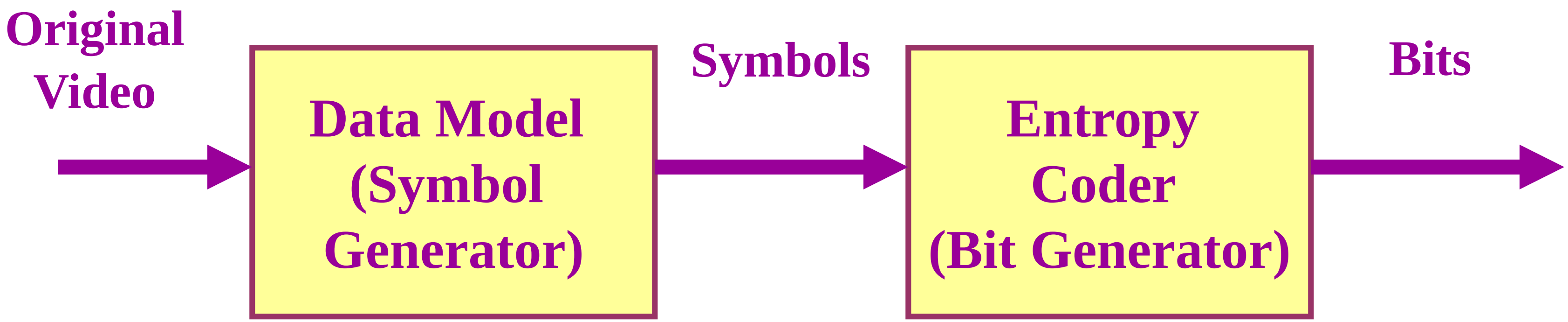
Note 1 – "x" means that the item is present in the macroblock.

Note 2 – It is possible to apply the filter in a non-motion compensated macroblock by declaring it as MC + FIL but with a zero vector.

Decoder: the Slave !



The H.261 Symbolic Model



A video sequence is represented as a sequence of images structured in Groups Of Blocks (GOBs) which are after divided in macroblocks, each of them represented with 1 or 0 motion vectors and/or (Intra or Inter coded) DCT quantized coefficients for 8×8 blocks.

Output Buffer: Absorbing Variations

The production of bits by the encoder is highly non-uniform in time, essentially because:

- **Variations in spatial detail for the various parts of each image**
- **Variations of temporal activity along time**
- **Entropy coding of the coded symbols**

To adapt the variable bitrate flow produced by the encoder to the constant bitrate flow transmitted by the channel, an output buffer is used, which adds some delay.

Bitrate Control

The encoder must efficiently control the way the available bits are spent in order to maximize the decoded quality for the synchronous bitrate/channel available.

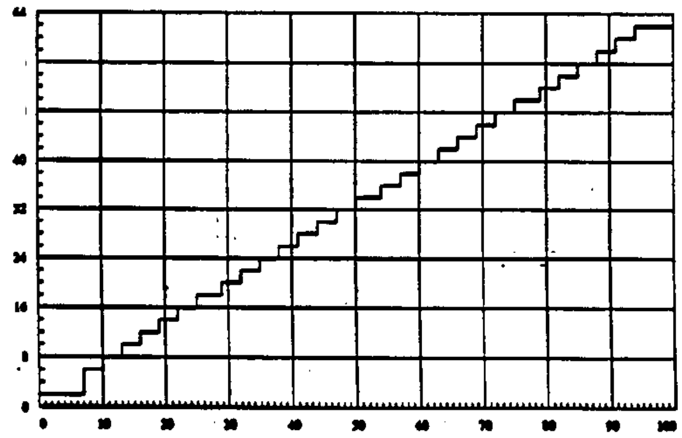
H.261 does not specify what type of bitrate control must be used; various tools are available:

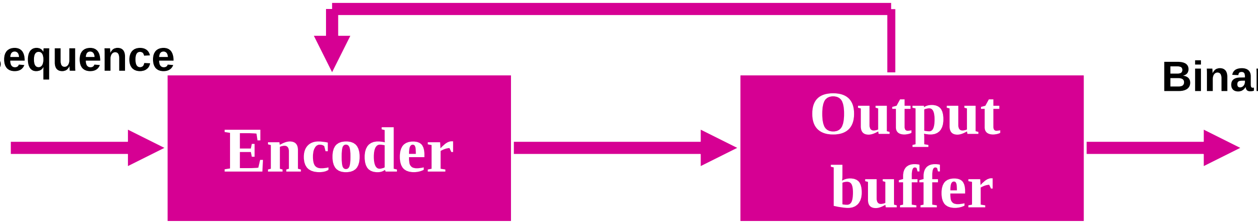
- **Changing the temporal resolution/frame rate**
- **Changing the spatial resolution, e.g. CIF to QCIF and vice-versa**
- **Controlling the macroblock classification**
- **CHANGING THE QUANTIZATION STEP VALUE**

The bitrate control strategy has a huge impact on the video quality that may be achieved with a certain bitrate (and it is not normative) !

Quantization Step versus Buffer Fullness

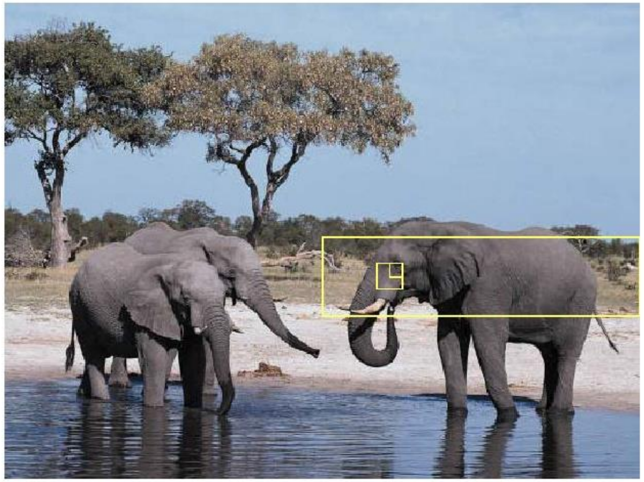
The bitrate control solution recognized as most efficient, notably in terms of the granularity and frequency of the control, controls the quantization step as a function of the output buffer fullness.





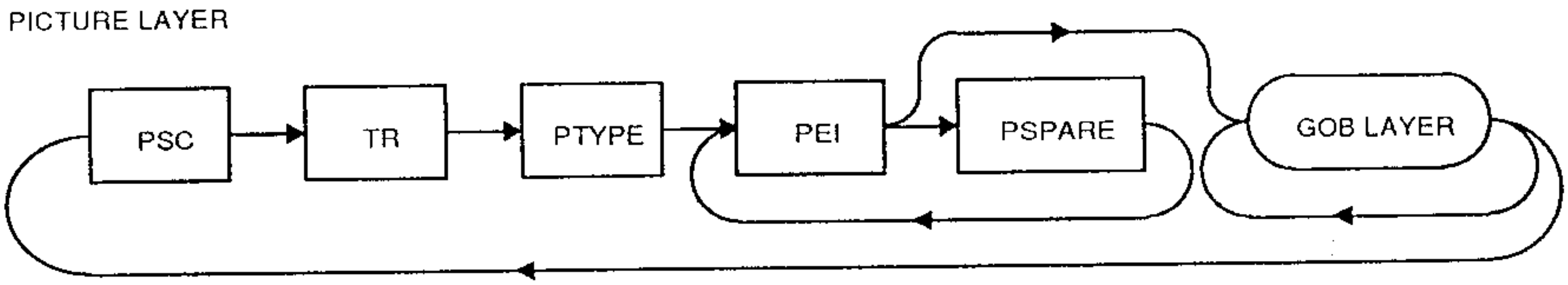
Hierarchical Structure Functions

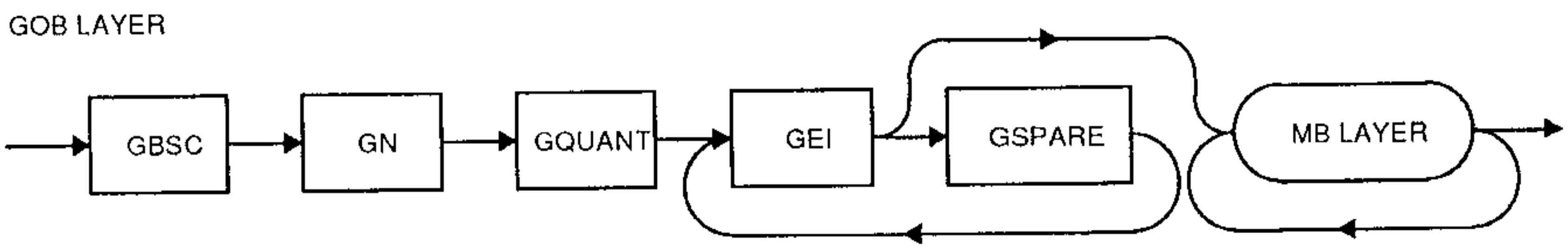
- **Image**
  - Resynchronization (*Picture header*)
  - Temporal resolution control
  - Spatial resolution control



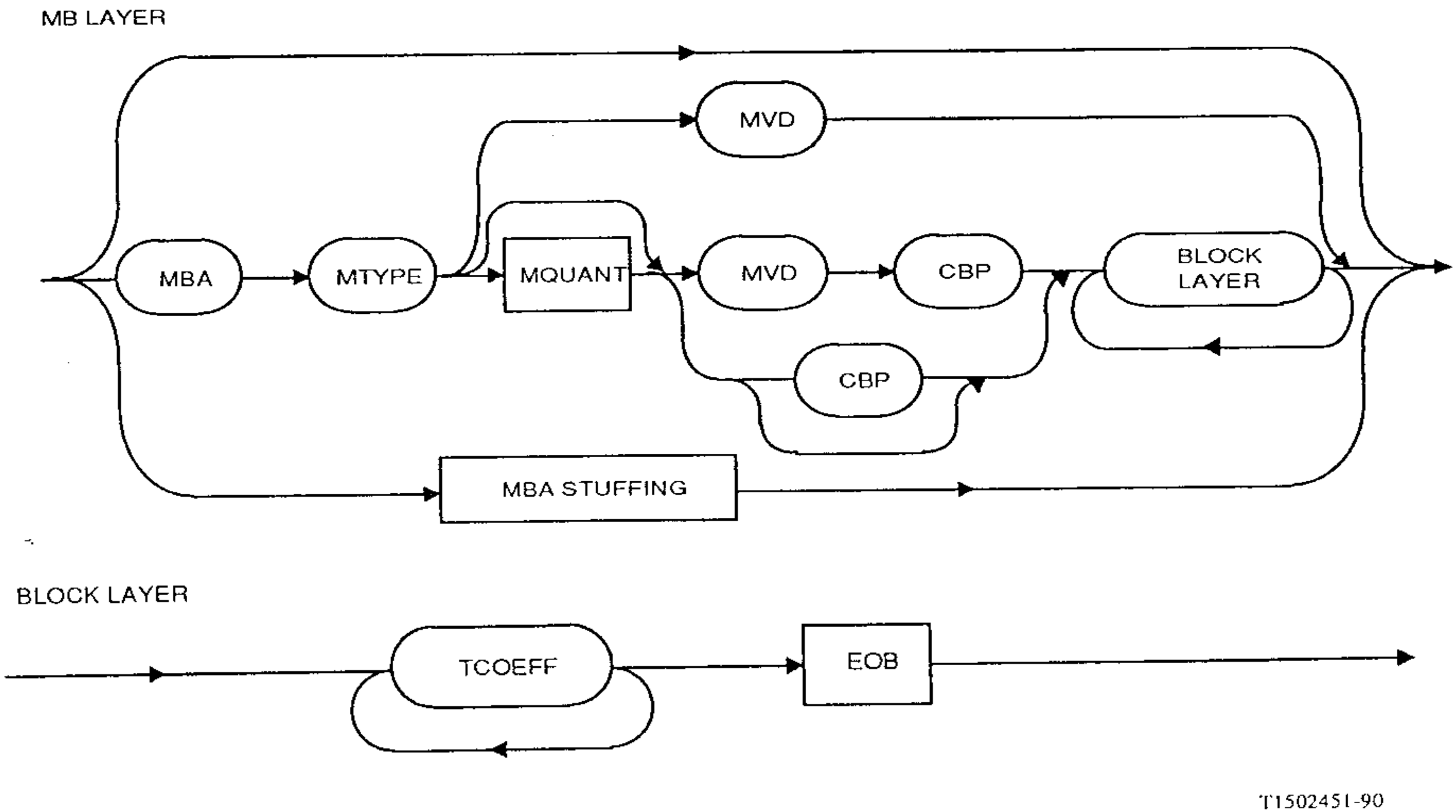
- **Group of Blocks (GOB)**
  - Resynchronization (GOB header)
  - Quantization step control (mandatory)
- **Macroblock**
  - Motion estimation and compensation
  - Quantization step control (optional)
  - Selection of coding tools (MB classification)
- **Block**
  - DCT

Coding Syntax: Image and GOB Levels

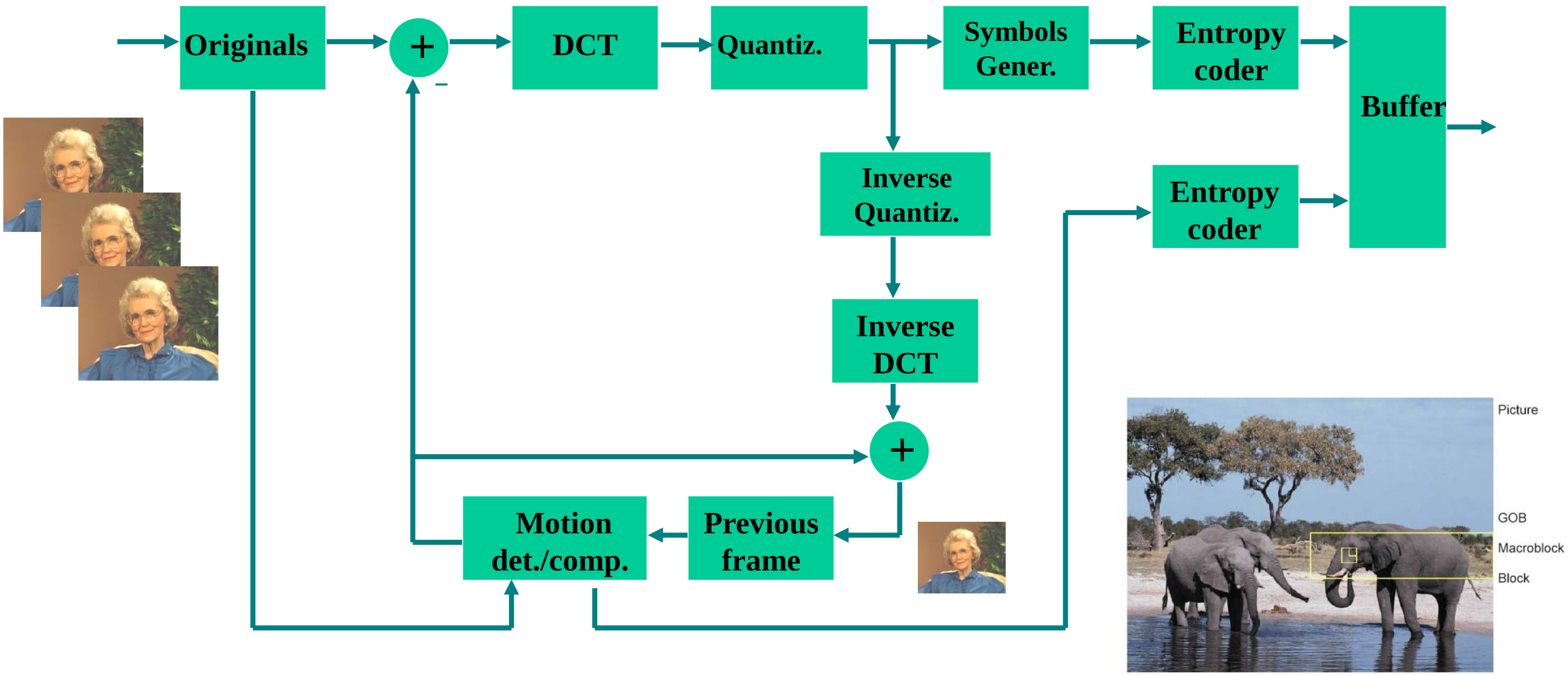




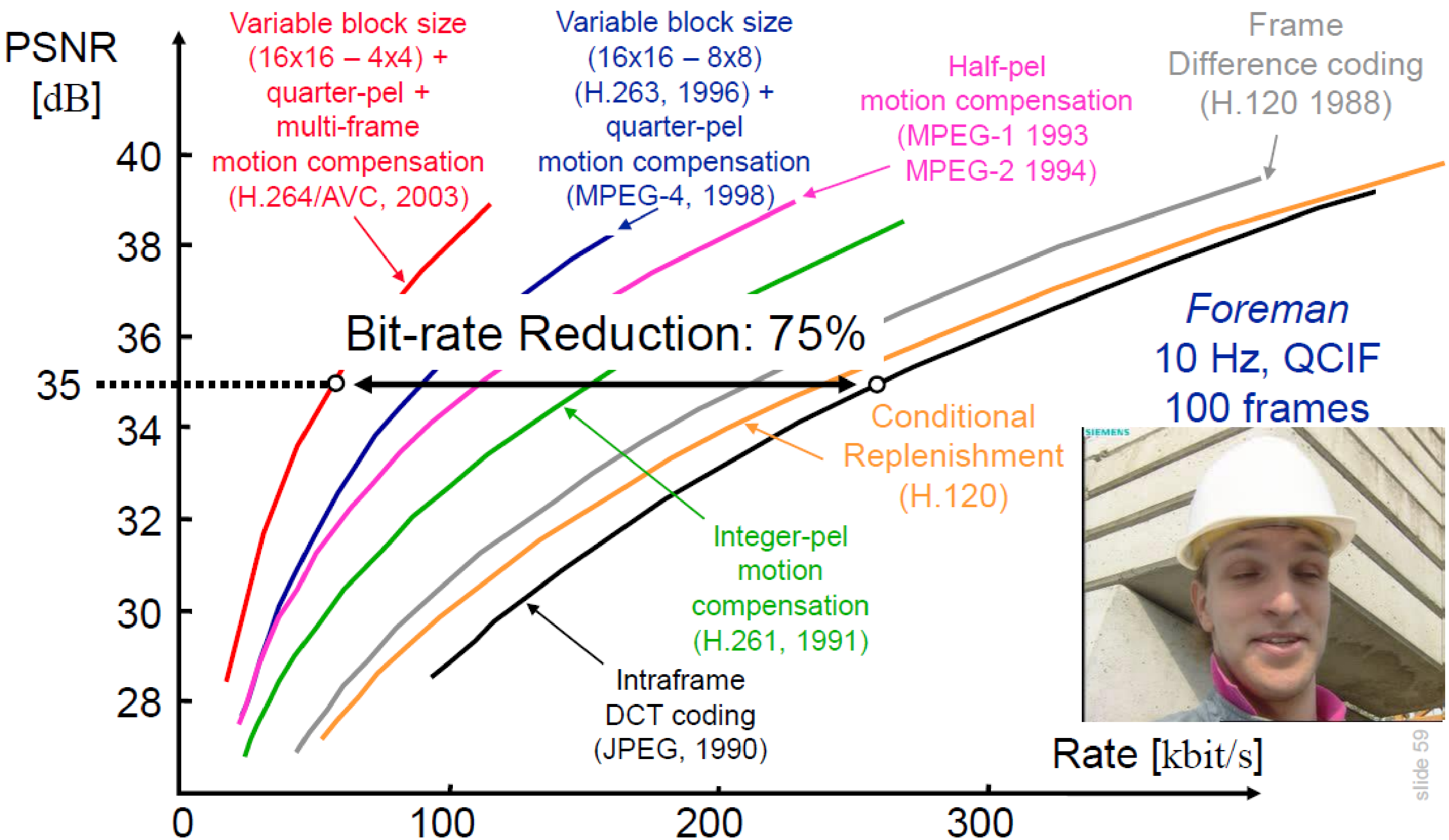
Coding Syntax: MB and Block Levels



Encoder: the Winning Cocktail !



Rate-Distortion (RD) Performance …

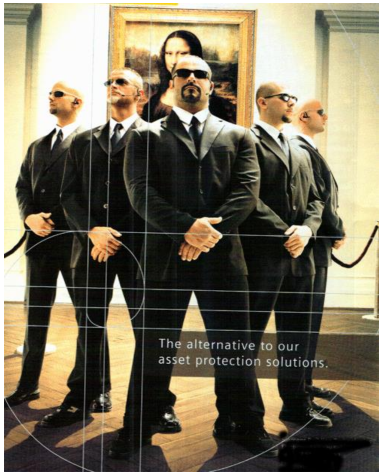


Error Protection for the H.261 Binary Flow

- **Error protection for the H.261 binary flow is implemented by using a BCH (511,493) - *Bose-Chaudhuri-Hocquenghem* – block code (channel coding).**
- **The usage of the channel coding bits (also parity bits) at the decoder is optional.**
- **The syndrome polynomial to generate the parity bits is**

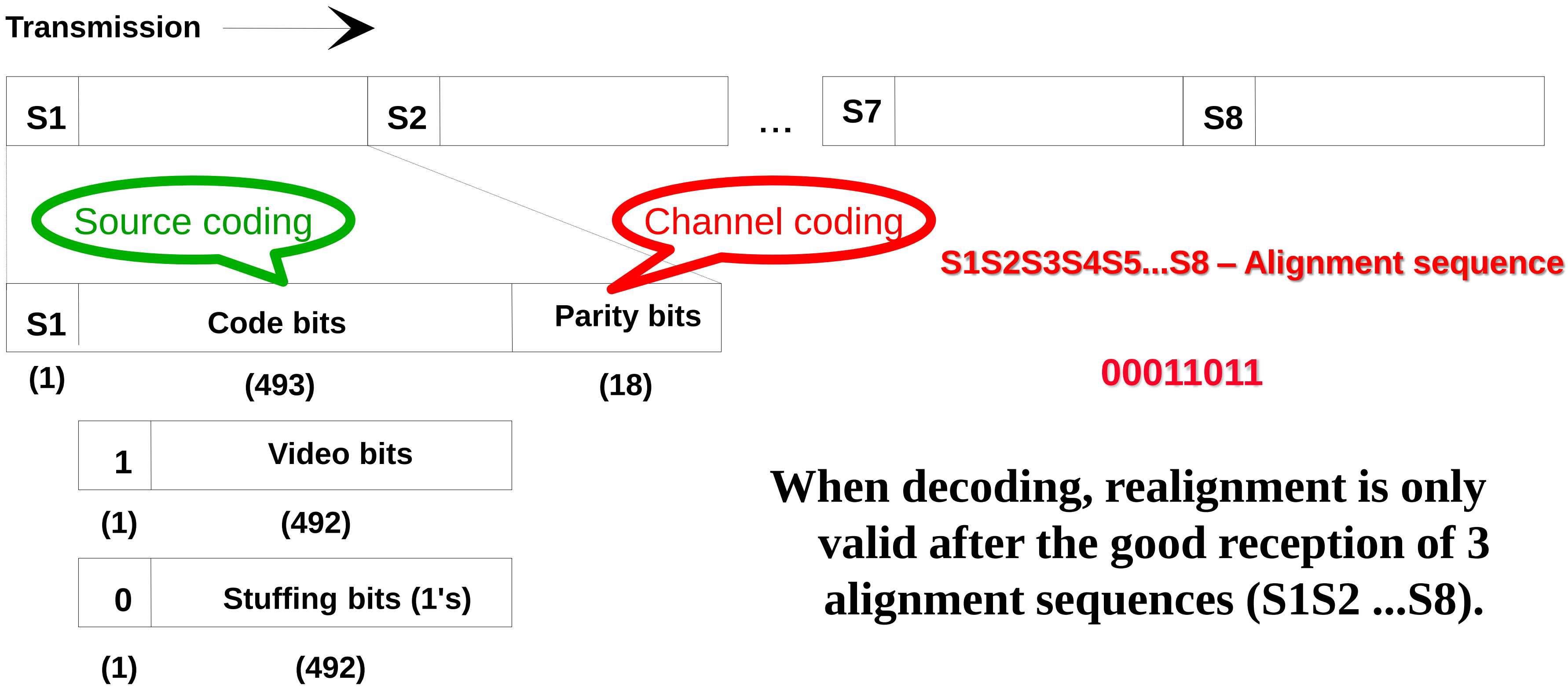
$$g(x) = (x^9 + x^4 + x)(x^9 + x^6 + x^4 + x^3 + 1)$$

Symbols with useful information	Correcting symbols



H.261 Error Protection and Alignment

The final video signal stream structure (multiframe with 512×8 = 4096 bits) is:



When decoding, realignment is only valid after the good reception of 3 alignment sequences (S1S2 ...S8).

Intra Refreshment or Forced Updating

- **Forced updating is achieved by forcing the use at the encoder of the INTRA coding mode.**
- **The update pattern is not defined in H.261 but clearly not too many MBs should be updated in the same frame to avoid strong quality variations (as Intra coded MBs spend more bits for the same quality) .**
- **To control the accumulation of IDCT mismatch error, H.261 recommends that a macroblock should be forcibly updated at least once per every 132 times it is transmitted.**
- **Naturally, forced updating may also be used to stop the propagation of the effect of channel errors.**

Error Concealment

- **Even when channel coding is used, some residual (transmission) errors may end at the source decoder.**
- **Residual errors may be detected at the source decoder due to syntactical and semantic inconsistencies.**
- **For digital video, the most basic error concealment techniques imply:**
  - **Repeating the co-located data from previous frame**
  - **Repeating data from previous frame after motion compensation**
- **Error concealment for non-detected errors may be performed through post-processing.**

Error Concealment and Post-Processing Examples

Final Comments

- **H.261 has been the first video coding international standard with relevant market adoption.**
- **As the first relevant video coding standard, H.261 has established legacy and backward compatibility requirements which have influenced the standards to come after, notably in terms of technology selected.**
- **Many products and services have been available based on H.261.**
- **However, H.261 does not represent anymore the state-of-the-art on video coding (remind this standard is from 1990).**

Bibliography

- **Videoconferencing and Videotelephony,** R. Schaphorst, Artech House, 1996
- **Image and Video Compression Standards: Algorithms and Architectures,** V. Bhaskaran and K. Konstantinides, Kluwer Academic Publishers, 1995
- **Multimedia Communications,** F. Halsall, Addison-Wesley, 2001
- **Multimedia Systems, Standards, and Networks,** A. Puri & T. Chen, Marcel Dekker, Inc., 2000