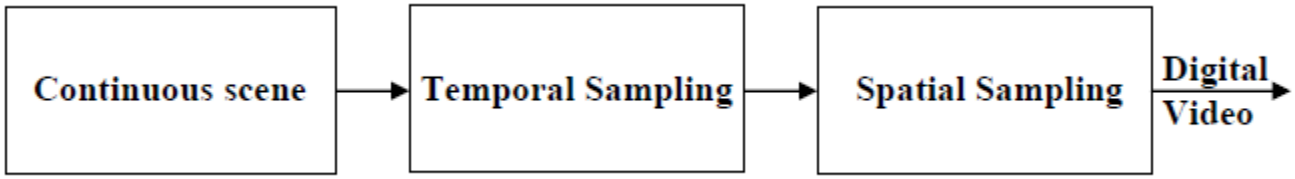
Digital Video Processing

- *Video signal is basically any sequence of time varying images.*
- *In a digital video, the picture information is digitized both spatially and temporally and the resultant pixel intensities are quantized.*



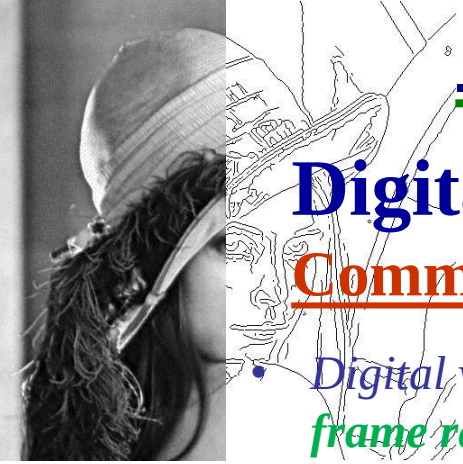
Spatial Sampling

*In the digital representation of the image, the value of each pixel needs to be quantized using some finite precision. In practice, **8 bits** are used per luminance sample*.

Temporal sampling

*A video consists of a sequence of images, displayed in rapid succession, to give an illusion of continuous motion. If the time gap between successive frames is too large, the viewer will observe jerky motion. In practice, most video formats use temporal sampling rates of **24 frames per second** and **above**.*

Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Common Video Formats

- *Digital video frames that are displayed at a prescribed frame rate. For example, **frame rate** of **30 frames/sec** is used in NTSC video.*

- *The **Common Intermediate Format** (CIF) has 352 x 288 pixels, and the **Quarter CIF** (QCIF) format has 176 x 144 pixels.*

Format	Luminance Pixel Resolution	Typical Applications
Sub-QCIF	128 X 96	Mobile Multimedia
QCIF	176 X 144	Video conferencing and Mobile Mulimedia
CIF	352 X 288	Video conferencing
4CIF	704 X 576	SDTV and DVD-Video
16CIF	1408 X 1152	HDTV and DVD-Video

***Pixel Resolution** (dots per lines)*

- *SD Resolution:* **640 × 480** (720p)
- *HD Resolution:* **1280 × 720** (720p) **/** **1920 × 1080** (1080p)

Digital Video Processing

Common Video Formats

• *Each pixel is represented by three components: the **luminance component Y**, and the two **chrominance components Cb and Cr**.*

• ***RGB to YCbCr Conversion***

$$\begin{bmatrix} Y & Cb & Cr \end{bmatrix} = \begin{bmatrix} R & G & B \end{bmatrix} \begin{bmatrix} 0.299 & -0.168935 & 0.499813 \\ 0.587 & -0.331665 & -0.418531 \\ 0.114 & 0.50059 & -0.081282 \end{bmatrix}$$

• ***Video quality** is commonly evaluated by using PSNR in luminance (Y)Channel, which is referred to as the **Y-PSNR** (dB).*

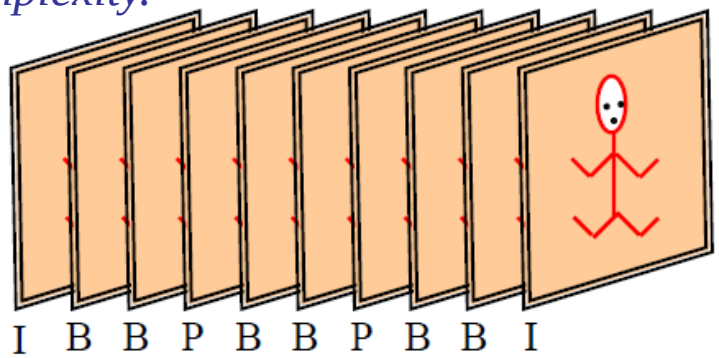
Digital Video Processing

Video Frame Types

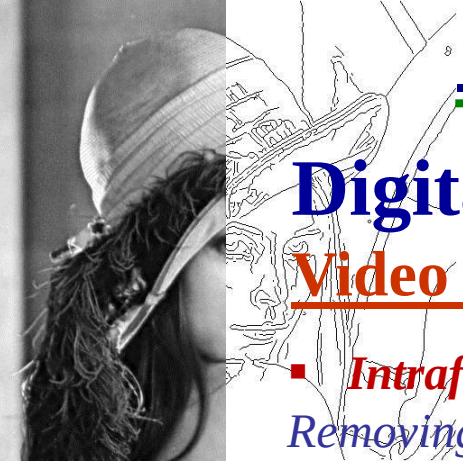
- *Three types of video frames are **I-frame**, **P-frame** and **B-frame**. 'I' stands for **Intra** coded frame, 'P' stands for **Predictive** frame and 'B' stands for **Bidirectional** predictive frame.*

- *'**I' frames** are encoded without any motion compensation and are used as a reference for future predicted '**P'** and '**B'** type frames. '**I'** frames however require a relatively large number of bits for encoding.*

- ***'P' frames** are encoded using motion compensated prediction from a reference frame which can be either 'I' or 'P' frame. 'P' frames are more efficient in terms of number of bits required compared to 'I' frames, but still require more bits than 'B' frames. '**B' frames** require the lowest number of bits compared to both 'I' and 'P' frames but incur computational complexity.*



Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Video Coding

- ***Intraframe coding***

*Removing the **spatial redundancy** with a frame is generally termed as intraframe coding. The spatial redundancy within a frame is minimized by using transform. The commonly used **transform is DCT**.*

- ***Interframe coding***

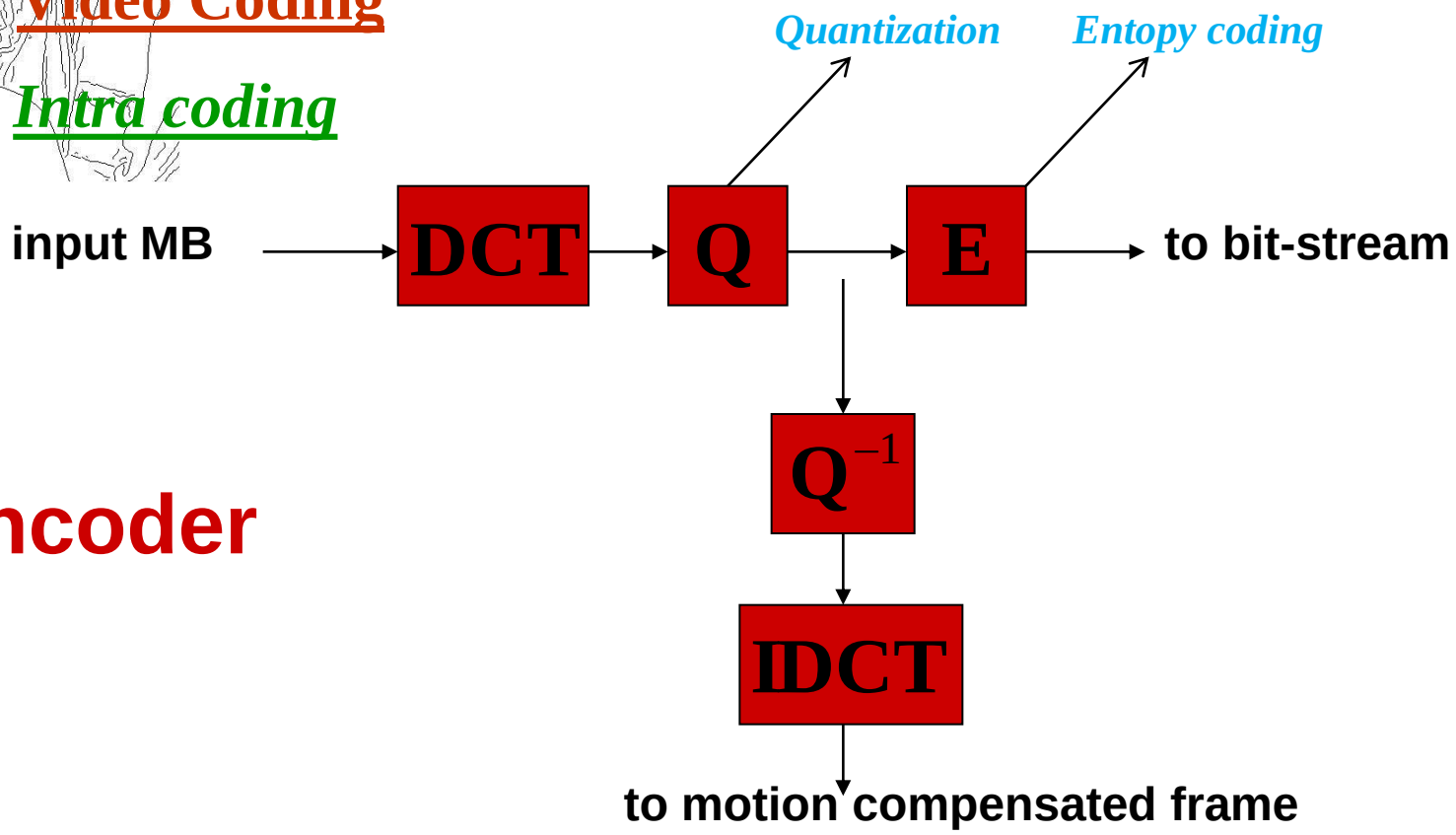
*The **temporal redundancy** between successive frames is removed by interframe coding. Interframe coding exploits the interdependencies of video frames. Interframe coding relies on the fact that adjacent pictures in a video sequence have high temporal correlation.*

- ***Intra (I-coding)***
  - ***MB (Macro Block)** is encoded as is, without motion compensation.*
  - *DCT followed by Q (Quantization), zig-zag, run-length, Huffman Coding.*
- ***Inter (P- and B-coding)***
  - ***Block-matching - motion estimation***
  - *Predictive motion residue from best-match block is DCT encoded (similarly to intra-mode)*
  - *Motion vector is differentially encoded .*

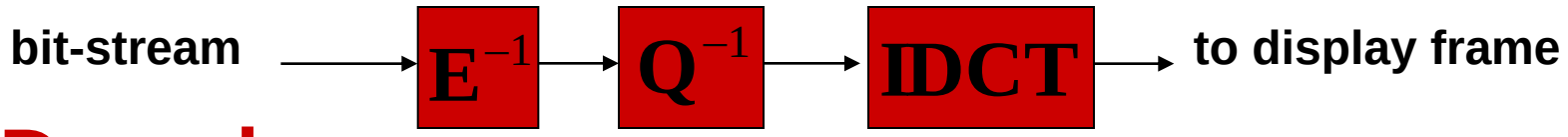
Digital Video Processing

Video Coding

Intra coding



Encoder



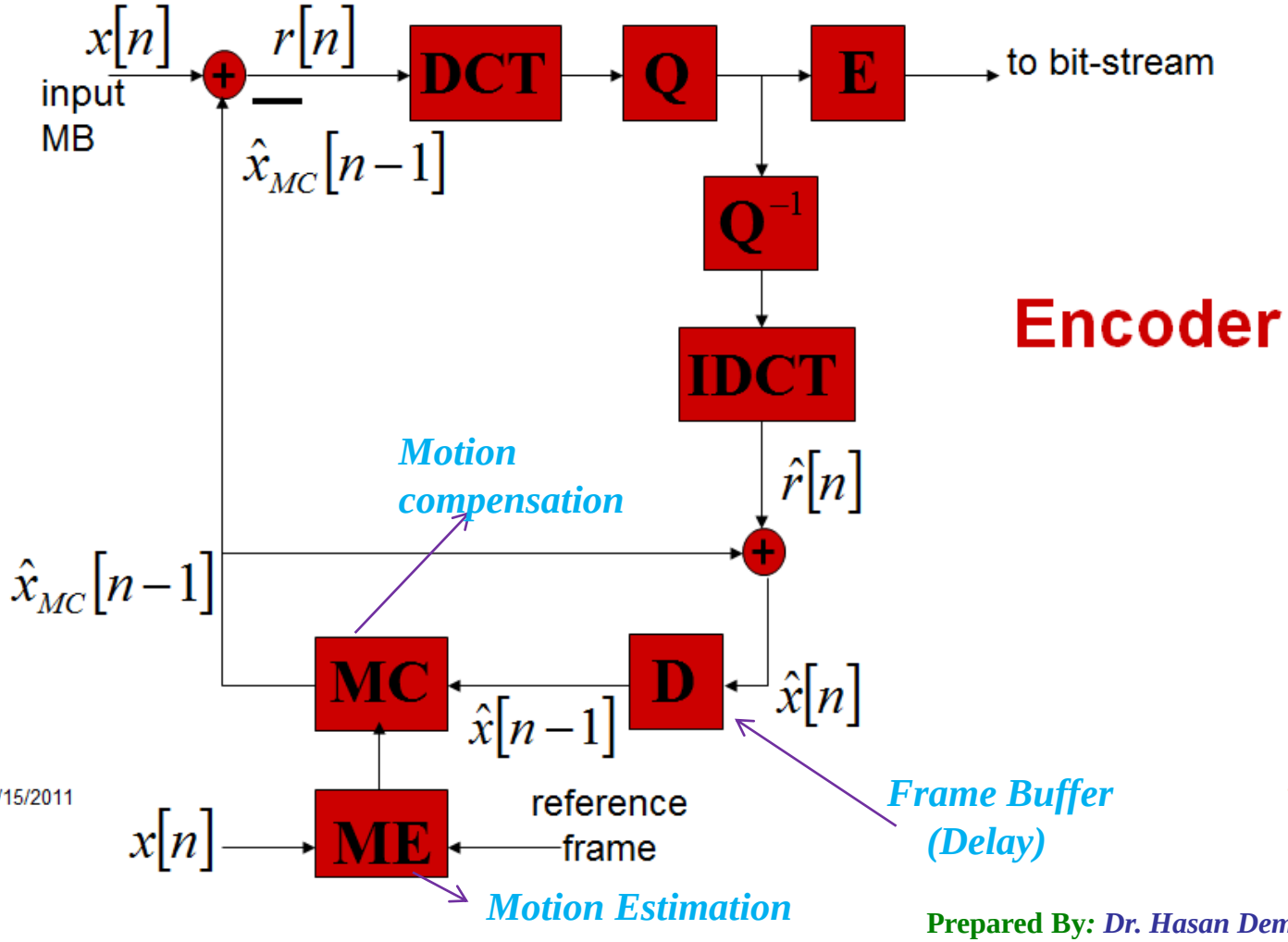
Decoder

Prepared By: *Dr. Hasan Demirel, PhD*

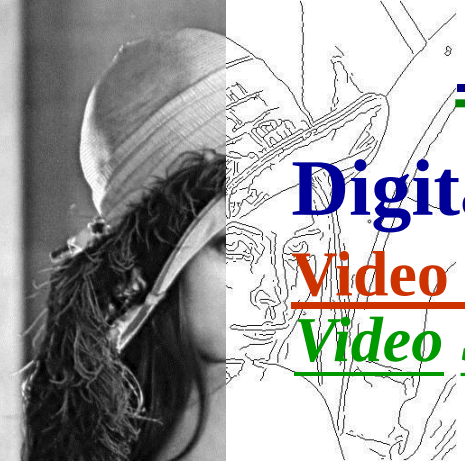
Digital Video Processing

Video Coding

Inter coding



Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Video Coding

Video Sequence and Picture

Intra 0 Inter 1 Inter 2 Inter 3 Inter 4 Inter 5

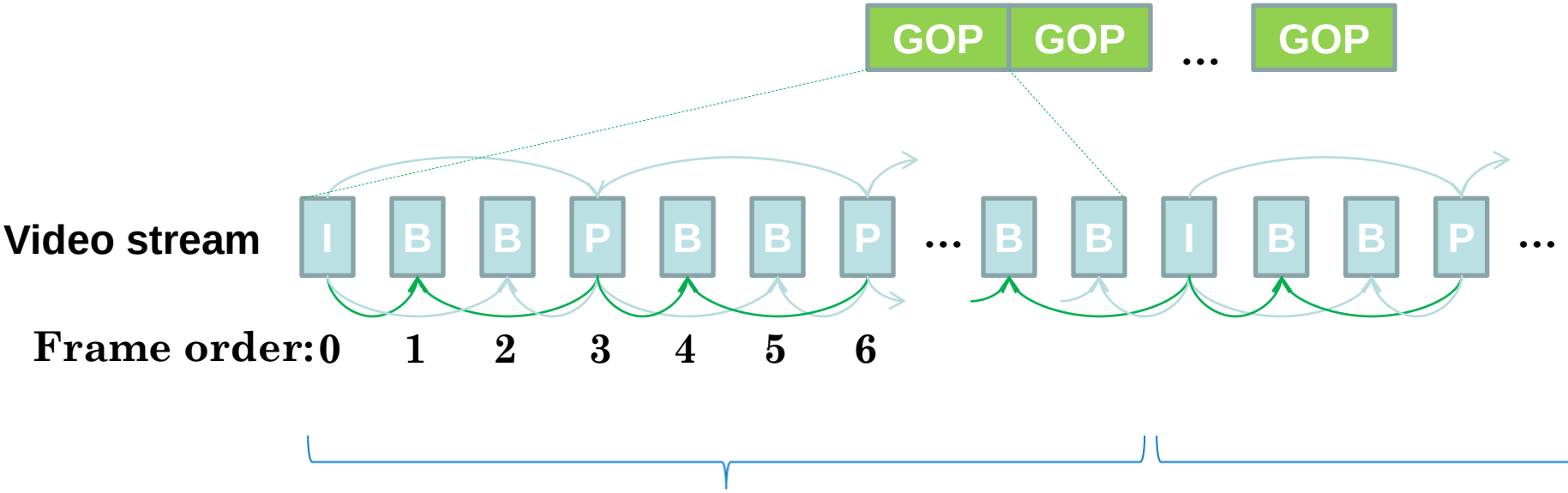
- ***Intra Picture (I-Picture)***
  - *Encoded without referencing others*
  - *All MBs are intra coded*
- ***Inter Picture (P-Picture, B-Picture)***
  - *Encoded by referencing other pictures*
  - *Some MBs are intra coded, and some are inter coded*

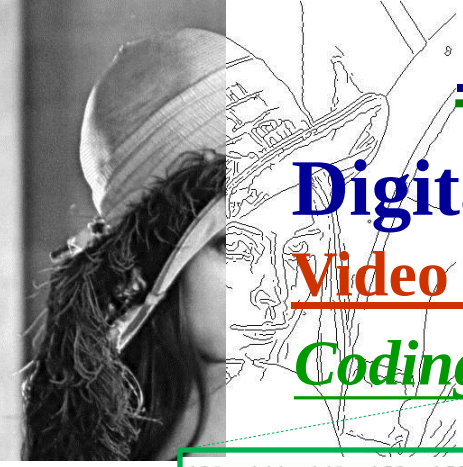
Prepared By: *Dr. Hasan Demirel, PhD*

Digital Video Processing

Video Coding

Group of Pictures (GOP)

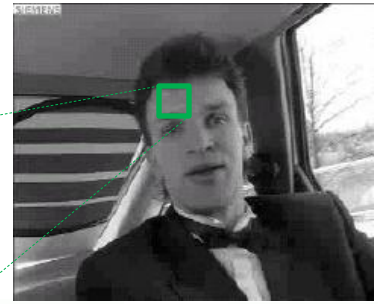




Digital Video Processing

Video Coding

Coding of I-Slice



139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

Original block

DCT

235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3
-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2
-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
1.8	-0.2	1.6	-0.3	-0.8	1.5	1.0	-1.0
-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

Transformed block

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Quantization matrix

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	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

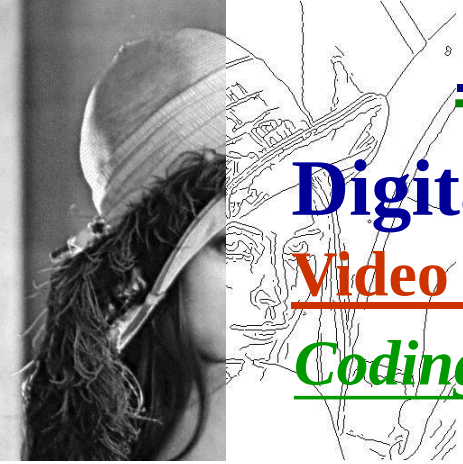
Zig-zag scan

15 0 -2 -1 -1 -1 0 …

Entropy coding

Bit-stream

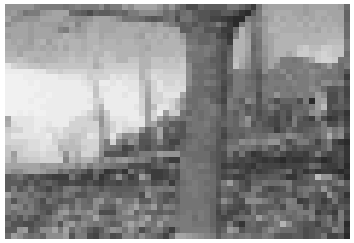
Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

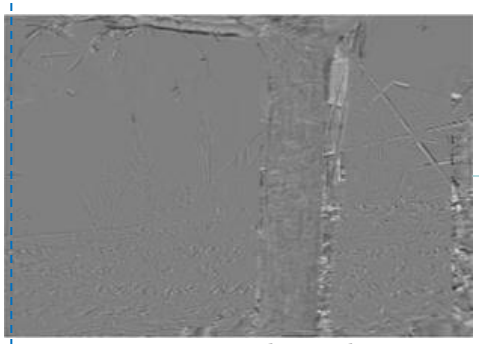
Video Coding

Coding of P-Slice



-

=



Original current frame

=

Residual
(Difference Frame)

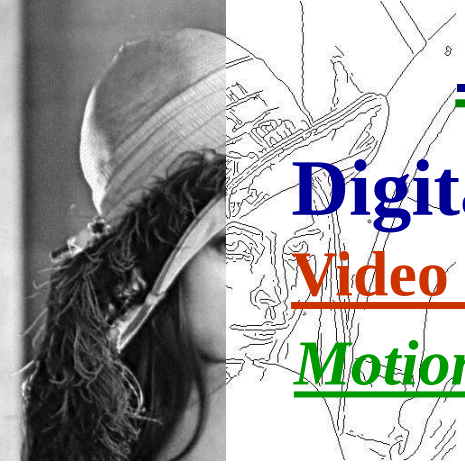
Motion Vectors
+

Motion Estimation

Reconstructed reference frame

Motion Compensation

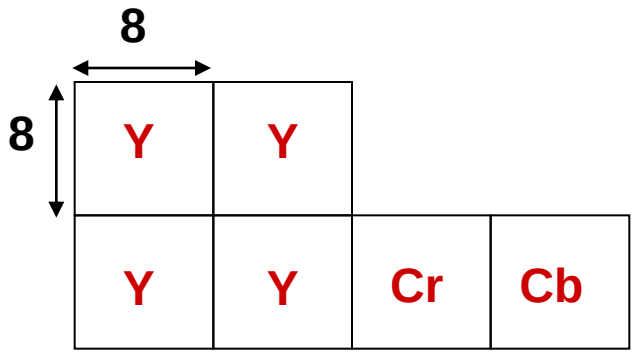
Prepared By: *Dr. Hasan Demirel, PhD*



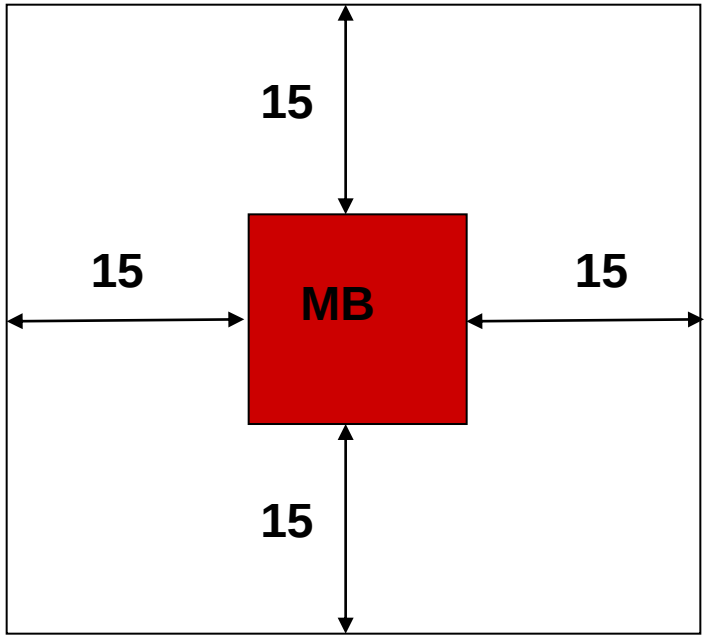
Digital Video Processing

Video Coding

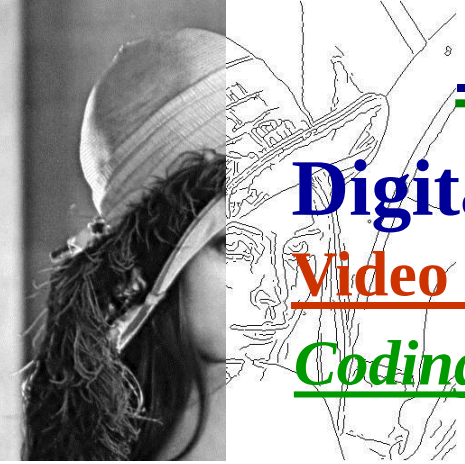
Motion Estimation in H.261



- *Macro-block*
  - *Luminance: 16x16, four 8x8 blocks*
  - *Chrominance: two 8x8 blocks*
  - *Motion estimation only performed for luminance component*
- *Motion vector range*
  - *[-15, 15]*



Search Area in Reference Frame



Digital Video Processing

Video Coding

Coding of Motion Vectors

- *MV has range [-15, 15]*
- *Integer pixel ME search only*
- *Motion vectors are differentially & separably encoded*

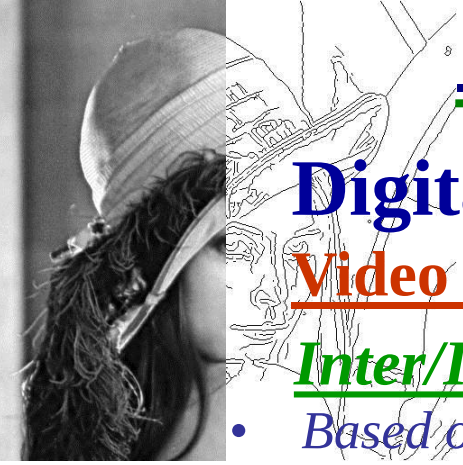
$$MVD_x = MV_x[n] - MV_x[n-1]$$
$$MVD_y = MV_y[n] - MV_y[n-1]$$

- *11-bit VLC (Variable Length Coding) for MVD*
- *Example*

MV = 2 2 3 5 3 1 -1…

MVD = 0 1 2 -2 -2 -2…

Binary: 1 010 0010 0011 0011 0011…

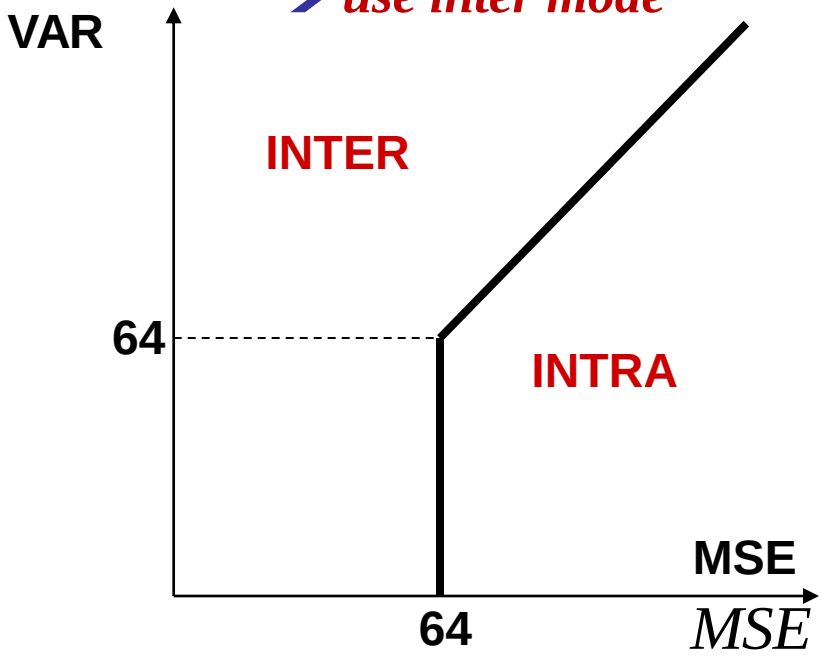


Digital Video Processing

Video Coding

Inter/Intra Switching

- *Based on energy of prediction error*
  - ***High energy**: scene change, occlusions, uncovered areas…*
    ➔ ***use intra mode***
  - ***Low energy**: stationary background, translational motion …*
    ➔ ***use inter mode***



$$VAR = \frac{1}{256}\sum_{MB}\left(c[x,y]-\bar{c}\right)^2$$

$$MSE = \frac{1}{256}\sum_{MB}\left(c[x,y]-r[x+dx,y+dy]\right)^2$$

Digital Video Processing

Video Coding

H.263 Standard

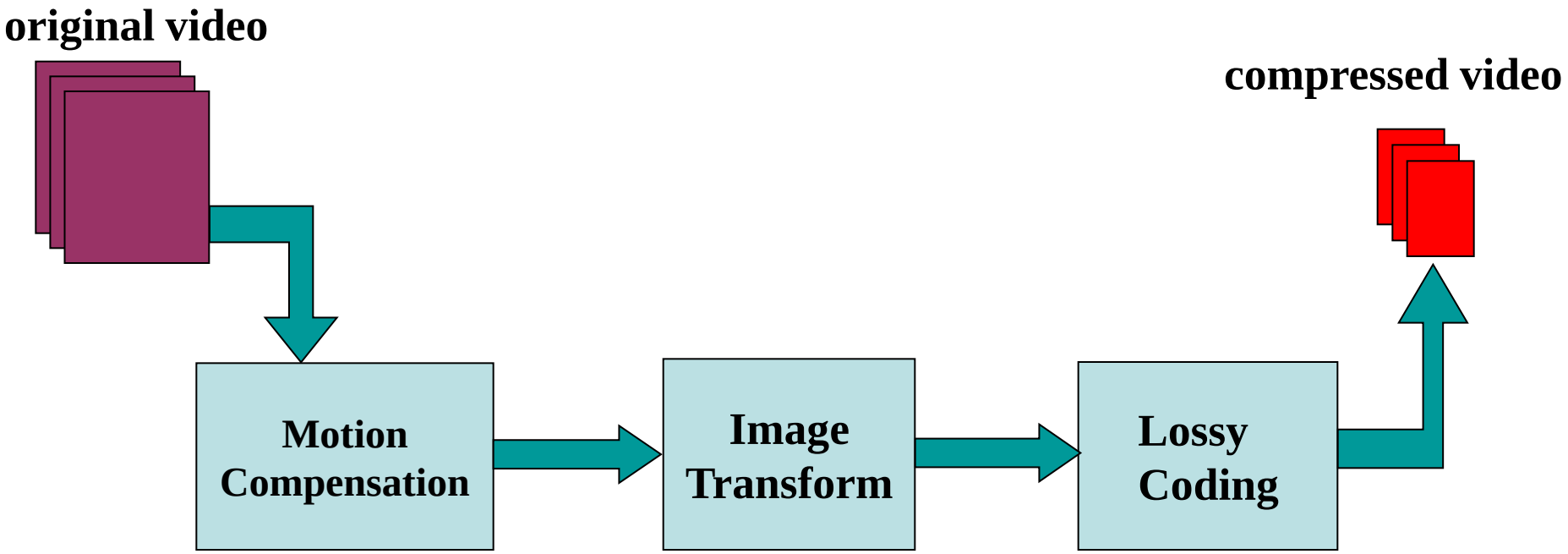
- *Standardization effort started Nov **1993***

- ***Aim***
  - *low bit-rate video communications, **less than 64 kbps***
  - *target PSTN and mobile network: 10-32 kbps*
- ***Near-term***
  - *H.263 and H.263+: established late 1997*
- ***Long-term***
  - *H.26L, H.264: still under investigation*
- ***Main properties***
  - *H.261 with many MPEG features optimized for low bit rates*
  - *Performance: **3-4 dB improvements over H.261** at less than 64 kbps; 30% bit rate saving over MPEG-1.*

Digital Video Processing

Video Coding

H.263 Standard Coder



Prepared By: *Dr. Hasan Demirel, PhD*

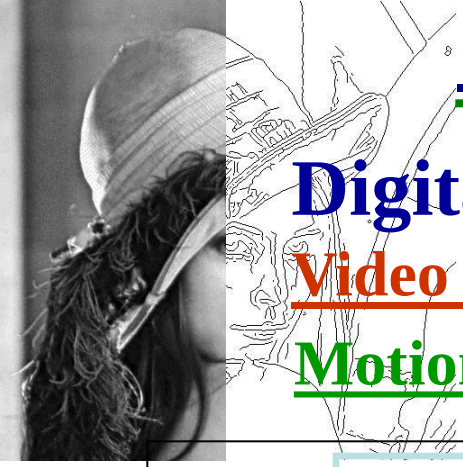
Digital Video Processing

Video Coding

H.263 Standard Coder

H.263 Motion Compensation

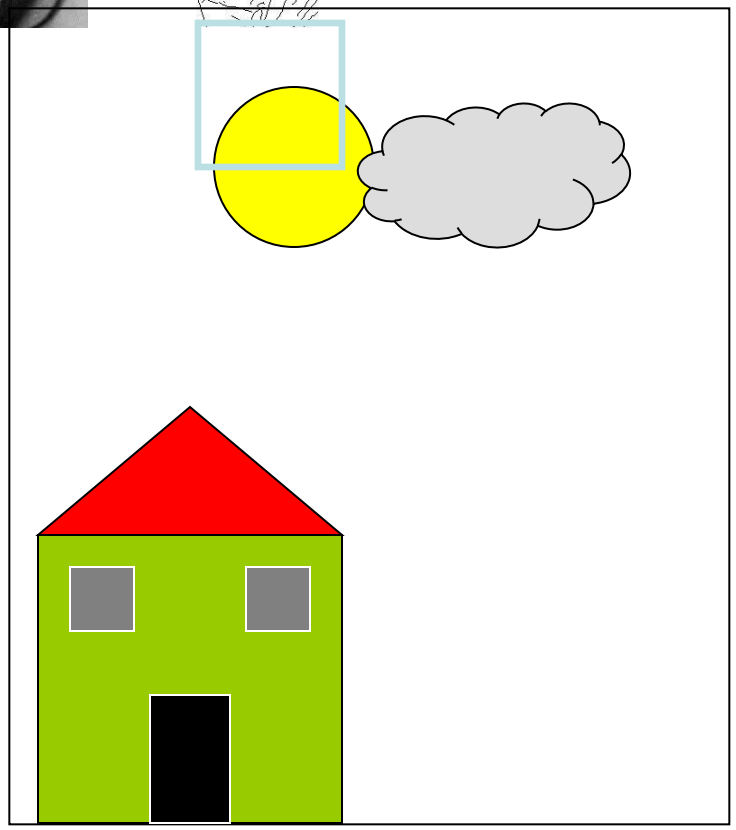
- Image is divided into 16x16 *macroblocks*,
- Each macroblock is matched against nearby blocks in previous frame (called *reference frame*),
  - "Nearby" = within 15-pixel horizontal/vertical range
  - Half-pixel accuracy (with bilinear pixel interpolation)
- Best match is used to *predict* the macroblock,
  - The relative displacement, or *motion vector*, is encoded and transmitted to decoder
- *Prediction error* for all blocks constitute the residual.



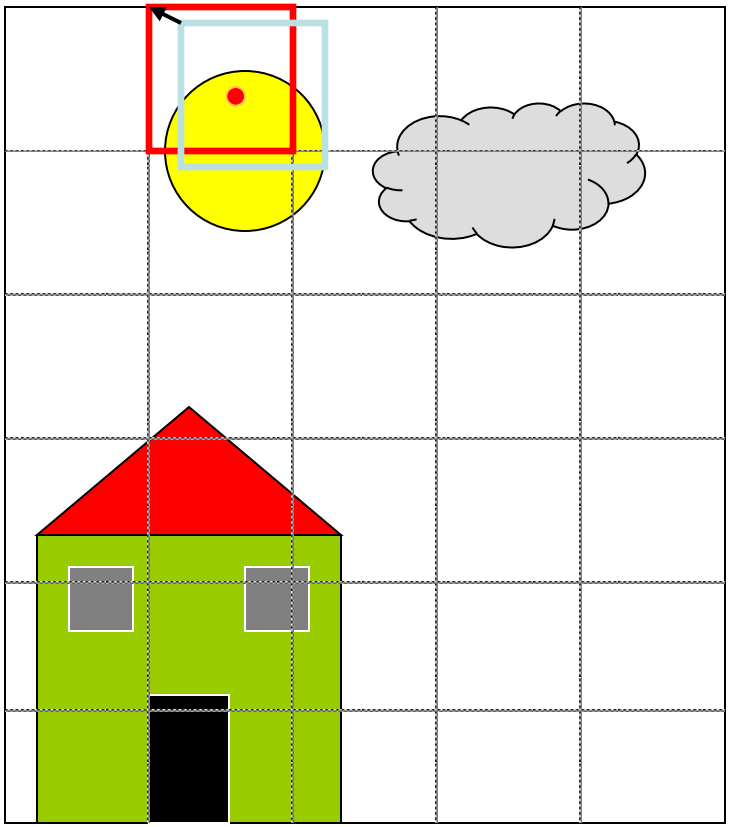
Digital Video Processing

Video Coding

Motion Compensation Example

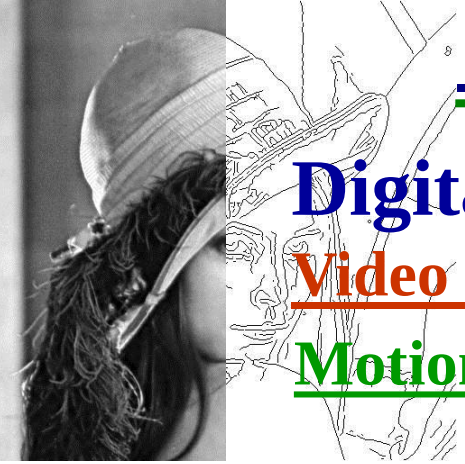


T=1 (reference)



T=2 (current)

Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Video Coding

Motion Compensation Example

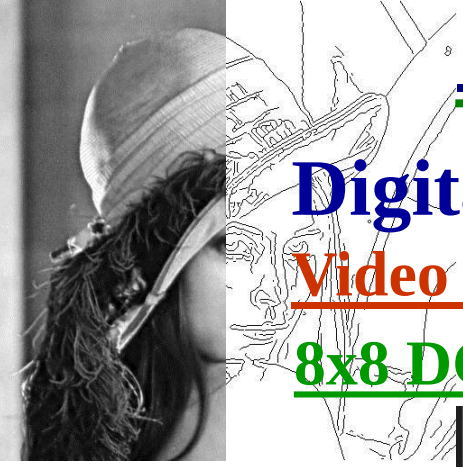
original video

compressed video

H.263 Image Transform

- Residual is divided into 8x8 blocks,
- 8x8 2-d Discrete Cosine Transform (DCT) is applied to each block independently
- DCT coefficients describe *spatial frequencies* in the block:
  - High frequencies correspond to small features and texture
  - Low frequencies correspond to larger features
  - Lowest frequency coefficient, called DC, corresponds to the average intensity of the block

Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Video Coding

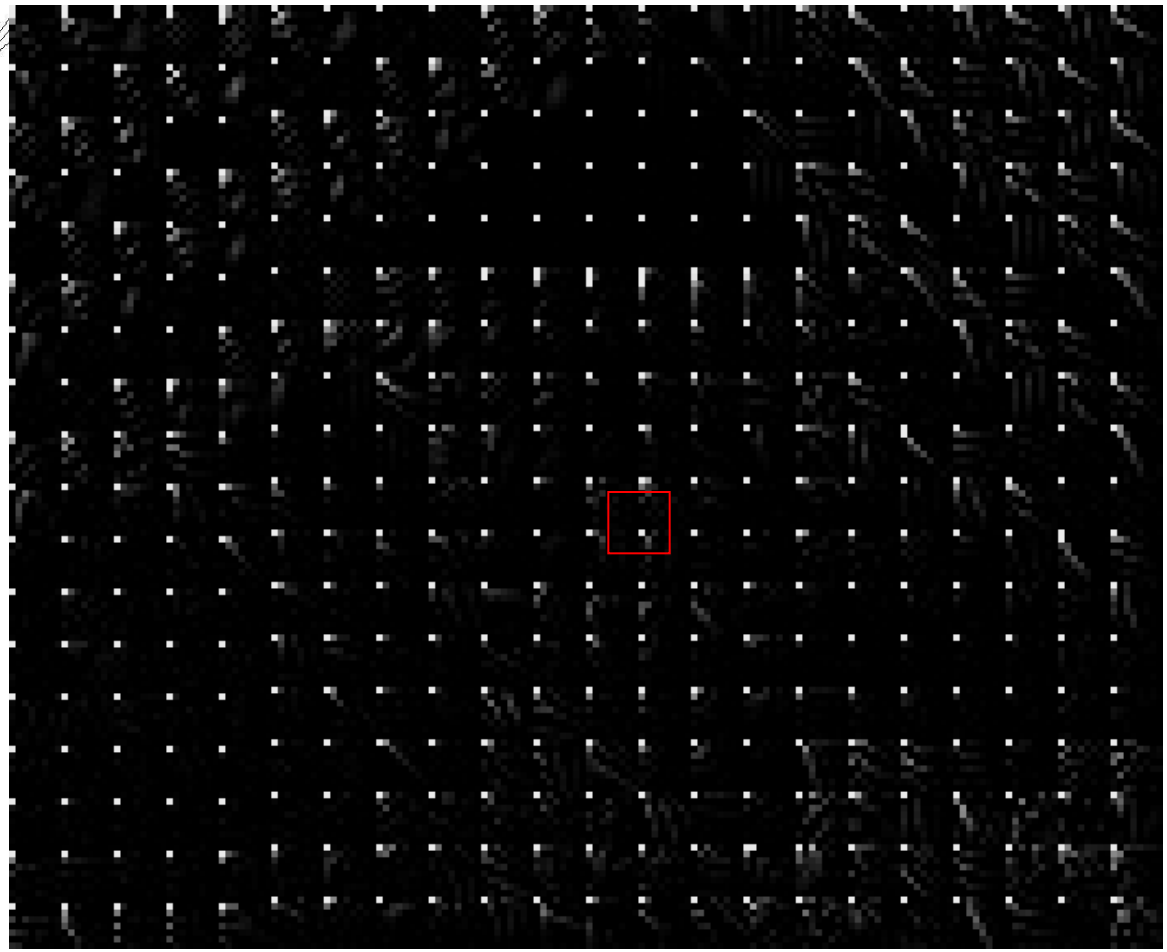
8x8 DCT Example

Prepared By: *Dr. Hasan Demirel, PhD*

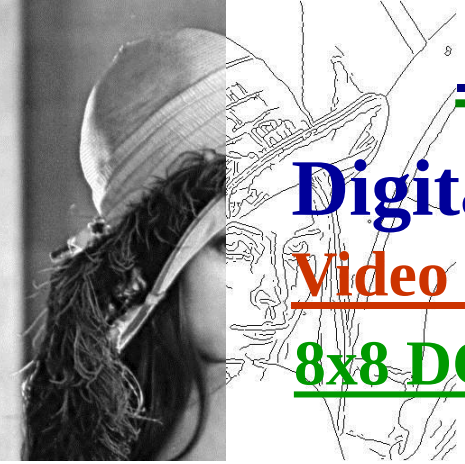
Digital Video Processing

Video Coding

8x8 DCT Example



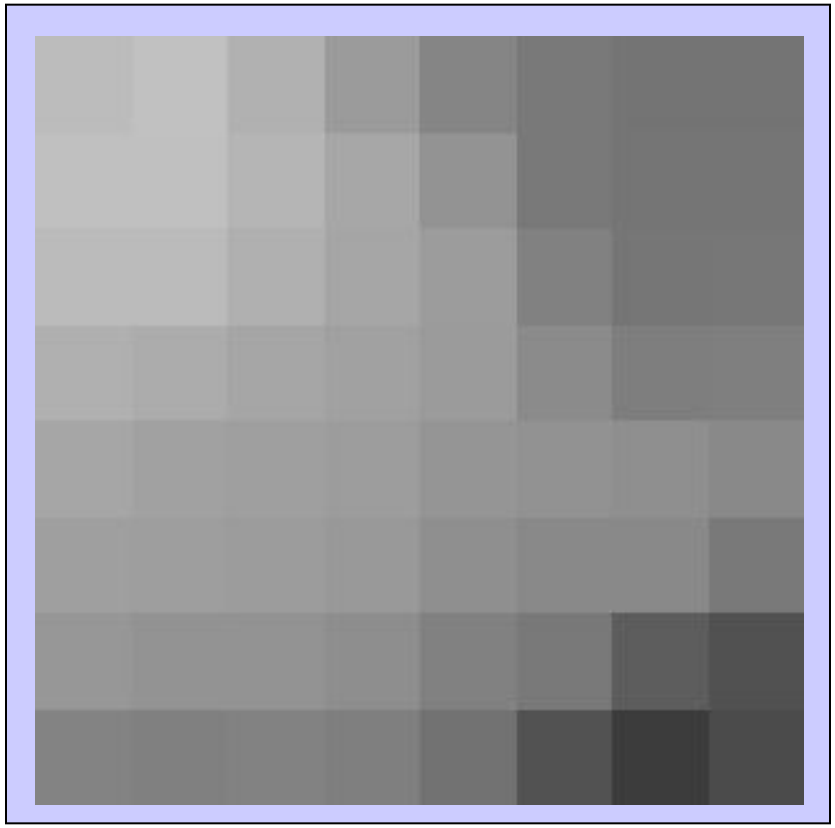
Prepared By: *Dr. Hasan Demirel, PhD*

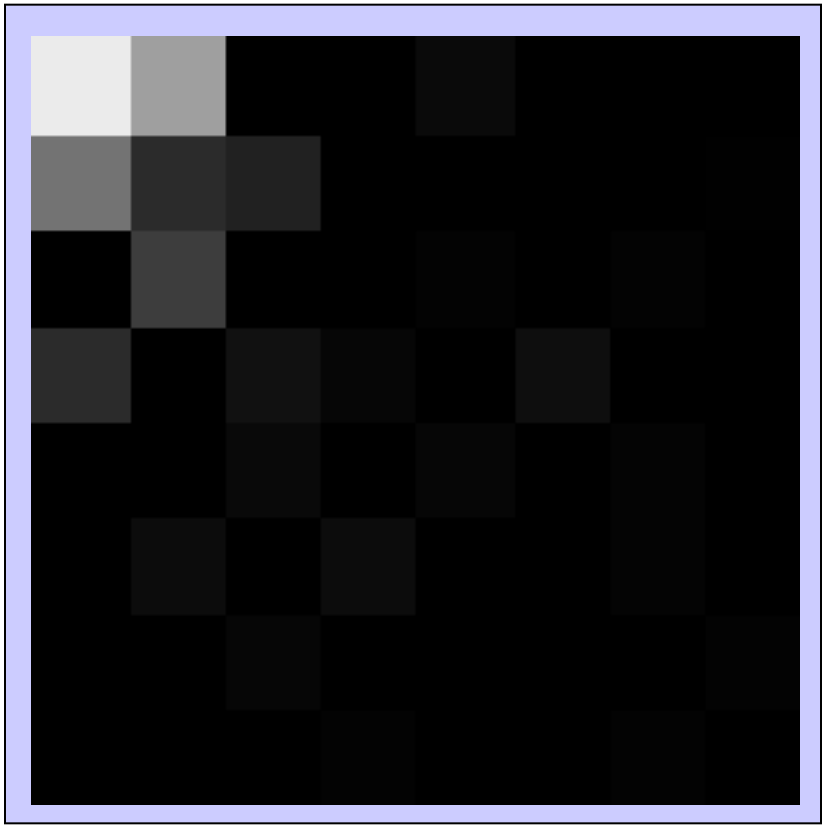


Digital Video Processing

Video Coding

8x8 DCT Example





Prepared By: *Dr. Hasan Demirel, PhD*

Digital Video Processing

Video Coding

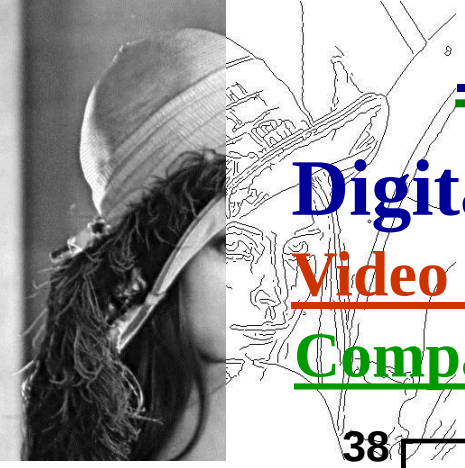
H.263 Standard Coder

original video

compressed video

H.263 Lossy Coding

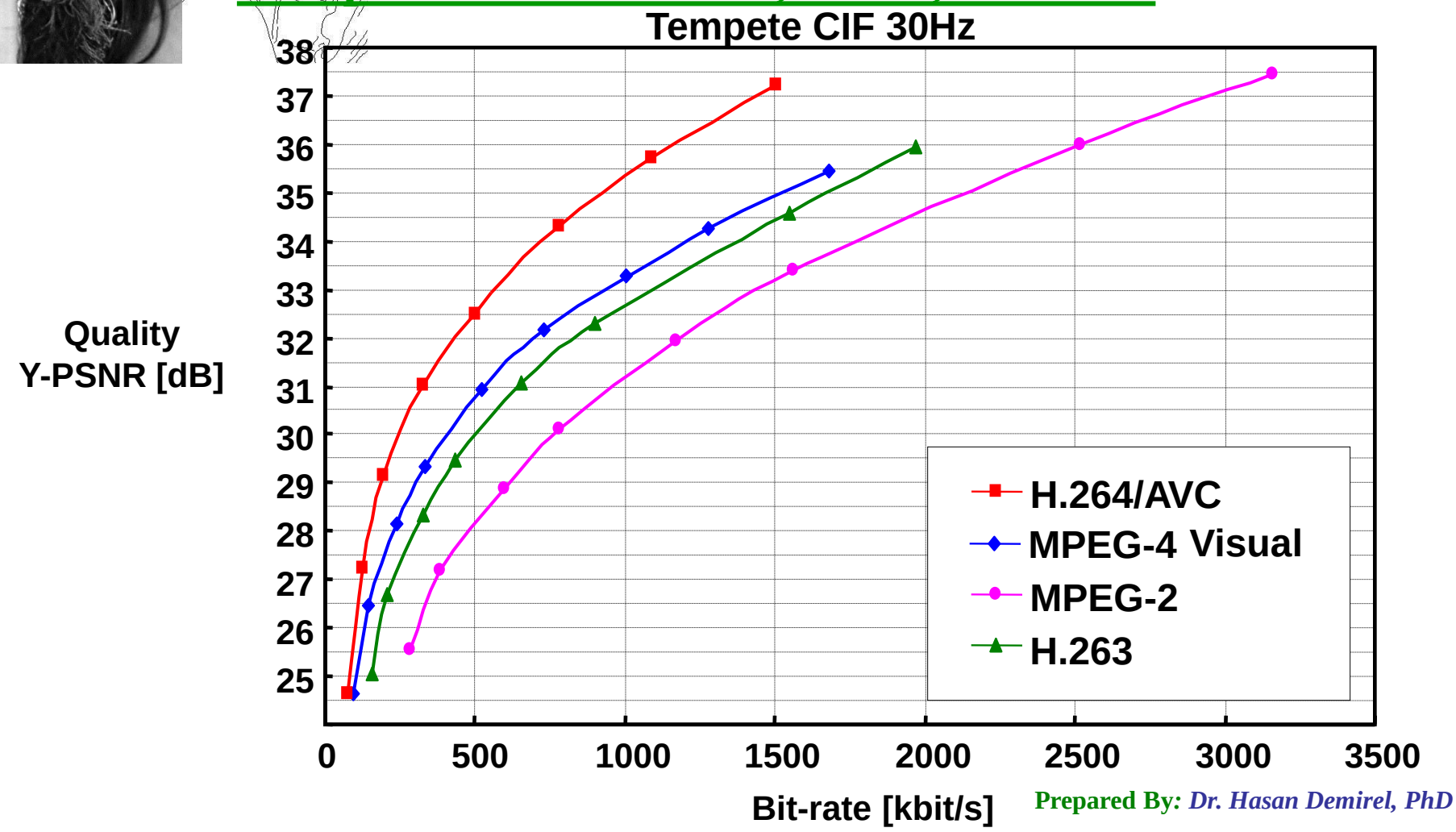
- Transform coefficients are *quantized:*
  - Some less-significant bits are dropped
  - Only the remaining bits are encoded
- For inter-frames, all coefficients get the same number of bits, except for the DC which gets more.
- For intra-frames, lower-frequency coefficients get more bits
  - To preserve larger features better
- The actual number of bits used depends on a *quantization parameter* (QP), whose value depends on the bit-allocation policy
- Finally, bits are encoded using entropy (lossless) code
  - Traditionally Huffman-style code

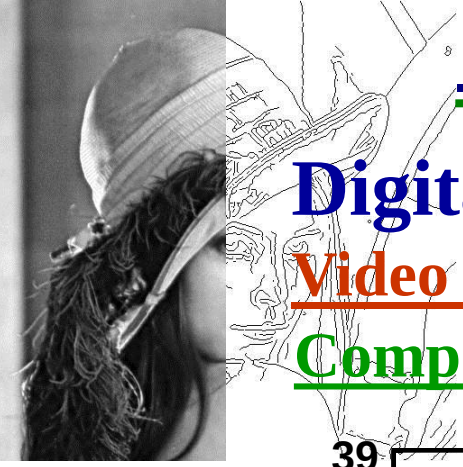


Digital Video Processing

Video Coding

Comparison to MPEG-2, H.263, MPEG-4

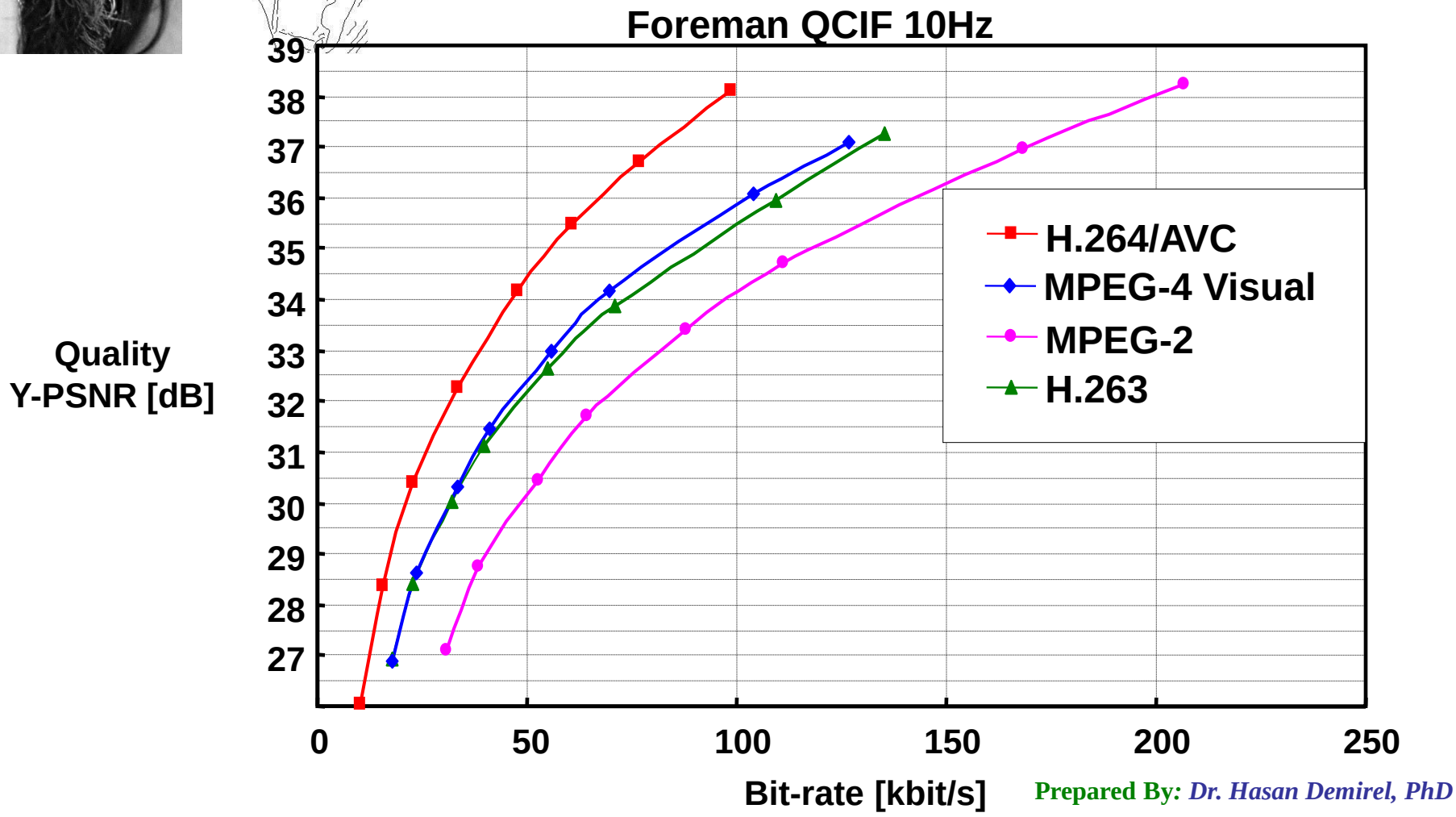




Digital Video Processing

Video Coding

Comparison to MPEG-2, H.263, MPEG-4



Prepared By: *Dr. Hasan Demirel, PhD*

Digital Video Processing

Video Coding Standards

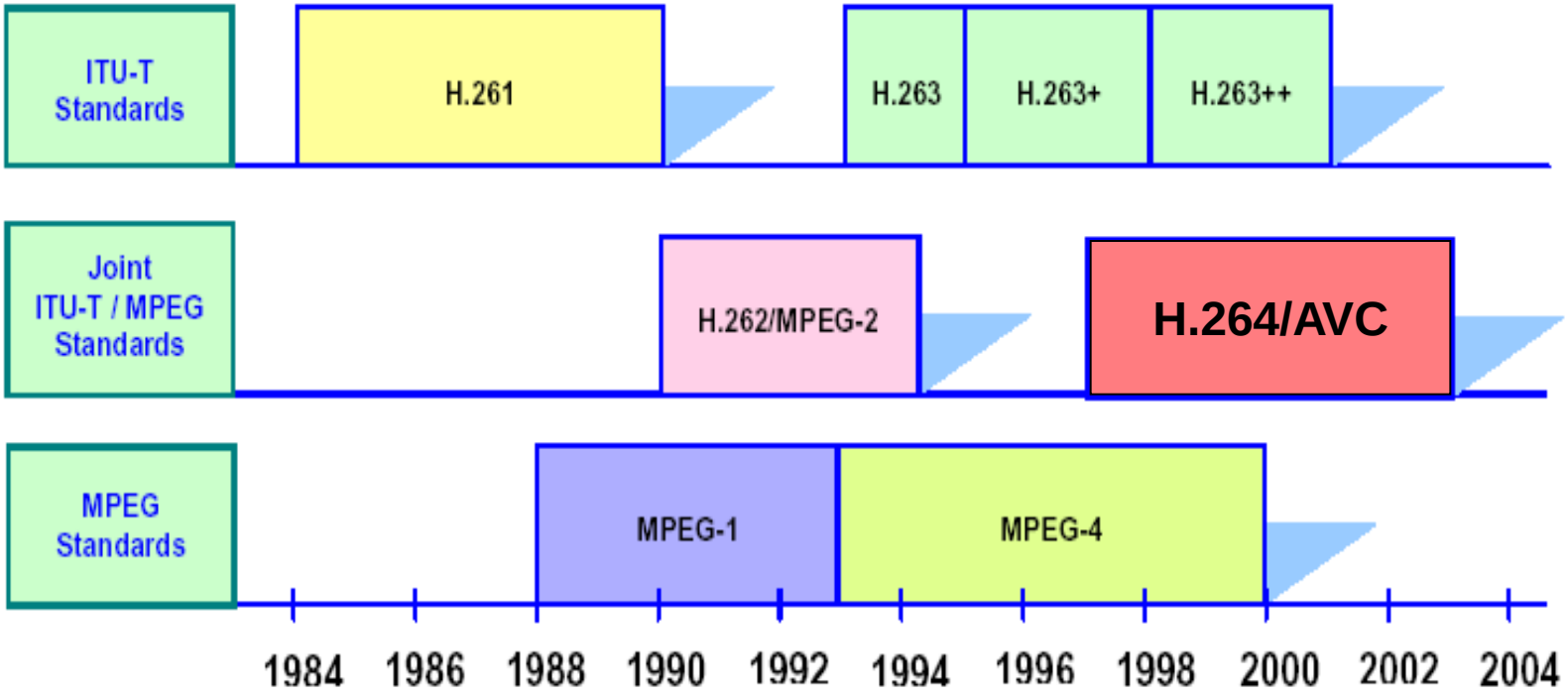


Figure 1. Progression of the ITU-T Recommendations and MPEG standards.

Digital Video Processing

Video Coding Standards

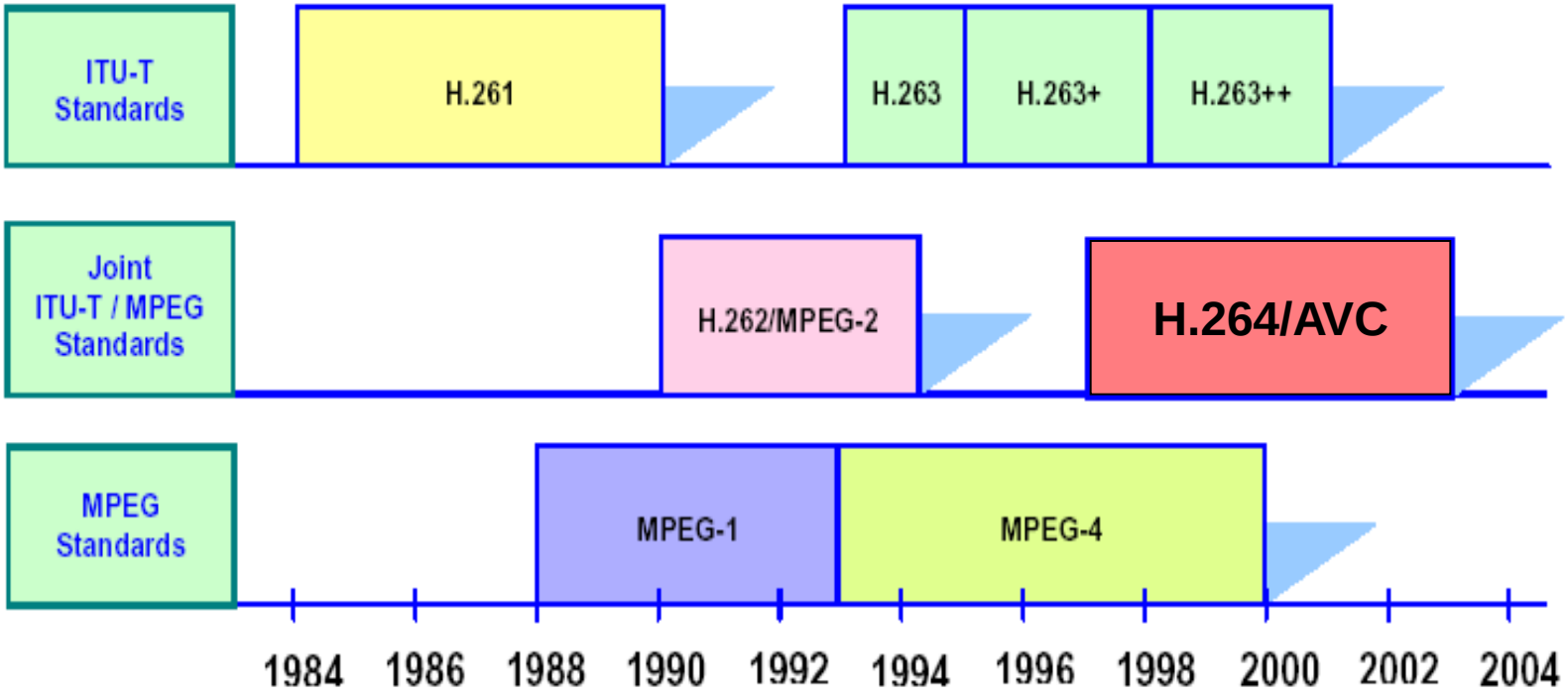


Figure 1. Progression of the ITU-T Recommendations and MPEG standards.

Digital Video Processing

Video Coding Standards

ISO (Int. Organization for Standardization)

MPEG-1 (1992)

1.5Mbps, VCD

MPEG-2 (1996)

2-10Mbps, DVD

MPEG-4 (2000)

8-1024Kbps, videophone

Digital cinema (ongoing)

ITU (Int. Telecommunication Union)

H.261 (1990)

p×64Kbps

H.263

8-64Kbps, videophone

H.263+/++

8-64Kbps, videophone

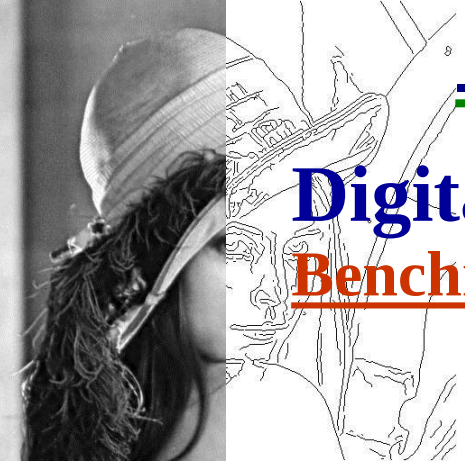
H.264/AVC

windows media player(Microsoft)

real player(Real-Networks)

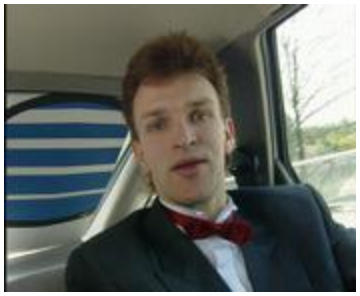
Skype Video

Prepared By: *Dr. Hasan Demirel, PhD*



Digital Video Processing

Benchmark Videos



Carphone

Suzie

Foreman

Prepared By: *Dr. Hasan Demirel, PhD*