Multimedia Computing

Nuno Correia
http://ctp.di.fct.unl.pt/~nmc
nmc@fct.unl.pt
FCT/UNL, 2018/2019

1

Objectives

- Study the techniques for processing and representing multimedia information considering:
  - Information properties.
  - Development tools.
  - Content analysis and retrieval.
- Build multimedia applications.

2

Multimedia

- Multimedia: several media (*medium*)
- Multimedia Computing: using multimedia data types in computational applications and systems

> *"Multimedia: An application requiring more than two trips to the car to operate.", IMA*

6

Hypermedia

Structuring and connecting heterogeneous information

Text + Hyper = Hypertext
+ Multimedia
= Hypermedia

Main concepts:

- **Node**
- **Link**
- Path
- History

7

Multimedia Computing

1 Multimedia Information

8

Types and Formats

.txt	.bmp	.snd	.cdr
.ps	.au	.mov	.fli
.tex	.wav	.avi	.tbk
.doc	.mid	.ra	.voc
.html	.mp3	.dcr	.qt
.jpg	.mpg	.pcd	.wad
.gif	.m2v	.xls	.asf
.tiff	.dat	.rle	.dcr
.pbm	.ifo	.pdf	.enc
.ppm	.vob	.aiff	.png

9

Static and Dynamic Information

Multimedia information can be:

- **Static**: doesn’t change over time (e.g., text, graphics)
- **Dynamic**: time dependent (e.g., animation)
- **Continuous**: Sequence (stream) of samples with time stamps (e.g., video, audio). Also dynamic with additional storage, transmission and synchronization requirements.

10

Text

- Content/Coding: Character representation.
  - Text semantics
- Presentation: Visual attributes.
  - A A A D A D % A A
- Structure: Text organization (also influences presentation).
  - HTML, text processor

11

Text Coding

- **ASCII** (7 bits)
  – 1963, (ASA, USA)
  – 1967, ISO standard
- **Extended ASCII** (8 bits)

  1980's , standardized as :
  – ISO 8859-1 (ISO Latin1) western europe
  – ISO 8859-2 (ISO Latin2) easter europe
  – ISO 8859-5 (Cirilic), ISO 8859-7 (Modern greek),
  – ISO 8859-8 (Hebrew)
- **EBCDIC** (8 bits, 1964 IBM)

12

Text Coding

- **Unicode** (16 bits, 1986 – 1992 std)

  Consortium: Xerox, Apple ... Adobe, MS, HP, IBM, Oracle, SAP, Sun, Unisys
  – More symbols...
  – ASCII compatible
  – Used in HTML, XML and Java
- **ISO 10646** (32 bits, 1991)
  – Unicode compatible

13

Text

- Codes and description languages: ASCII, SGML, HTML, PostScript,...

- Graphical representations (fonts)
  – Bitmaps: each character is represented by a matrix of pixels. Examples: Laserjet .SFP e .SFL, Tex PK, PXL, and GF.
  – Outlines: each character is represented by a a set of mathematical expressions describing lines. Examples: PostScript Fonts Type 1, Type 2, Type 3 and Type 5, TrueType Fonts.

14

Text

- **Markup Languages**: Instructions describing the way the text should be printed
  – Declarative: Describes actions (Word, Troff, TeX).
  – Procedural: Describes the objects to print (ODA, SGML, modules for TeX and Troff)
- SGML stands for Standard Generalized Markup Language and is an ISO international standard
- SGML is a language for describing document content and structure

15

Text

- PDF and PostScript are languages for output devices
- Provide a device independent common interface
- PostScript is an interpreted language, using RPN notation
  - Example: `12 134 mul`
- Graphics are also supported

16

Audio

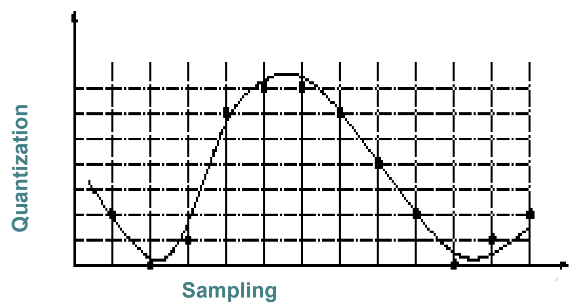
- Applications
  - *Voice Mail*
  - *Sonification*: mapping variables to sound (e.g., the points in a curve are mapped to the amplitude of an audio sample)
  - ...
- Properties
  - Audio perception is time dependent. Pictures or graphics can be experienced “forever”. For audio/music it is difficult to maintain contextual information
  - Important communications channel (probably the most appropriate and efficient to convey emotions)

17

Audio (Psychoacoustics)

- Audio, are changes in pressure that reach the human ear
- These changes are characterized by several variables and, when within certain ranges, are detected
  - **Frequency**: 20 Hz - 20000 Hz, physical measure
  - **Pitch**: perception, related with frequency
  - **Amplitude**: physical measure
  - **Loudness**: perception
  - Interactions between frequency and amplitude

Digital Audio

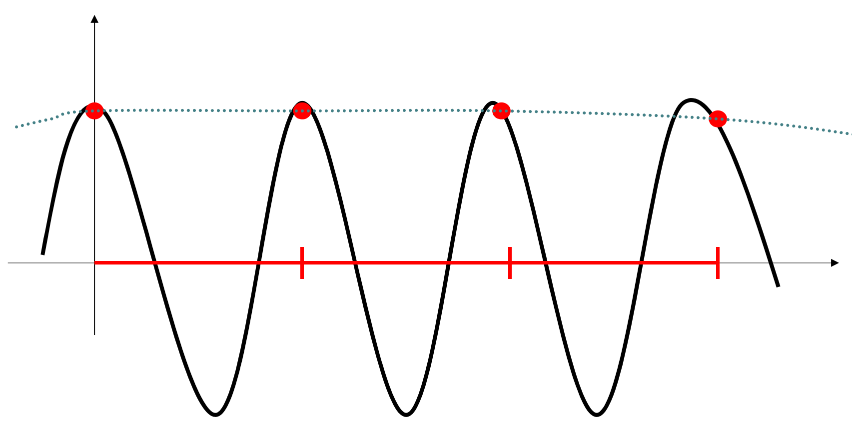


Digital Audio

- A digital signal is defined only in some instants and can only take a limited range of values
- Each of these values is a **sample**, usually taken at fixed intervals at a rate known as **sampling frequency**
- **Sampling theorem**: If a signal has components with frequency f, the sampling frequency must be at least 2f, in order to completely recover the original signal

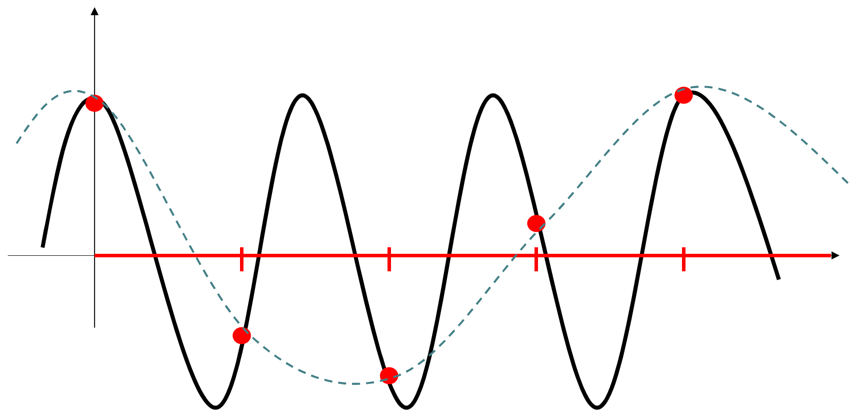
20

Sampling



21

Sampling



22

Digital Audio

- Quantize a signal means to define its values with a certain precision
- The difference between the representation and the original signal is the **quantization noise**
- With more bits a PCM (Pulse Coded Modulation) signal becomes more “clean”
- Audio CDs use 16 bits but people can hear more (aprox. 20 bits)
- DACs and ADCs convert between digital and analogue

23

Compression

- Storage:
  - 8bits and 8KHZ use 8Kb/s
  - stereo, 16 bits PCM, 44.1 KHz (CD) use 176Kb/s or 10Mb/minute
  - 48KHz (professional) use 200Kb/s
- Audio is compressed with **Perceptual Compression** techniques, such as MP3 (MPEG Layer 3) or MPEG-2 AAC (Advanced Audio Coding)
- Explores the characteristics of the human ear
- Reduces the size up to 12 times without reducing quality

24

Digital Audio

Attributes	Quality	Bytes/sec
11025 Hz, 8bits, mono	Reasonable for voice	11025
11025 Hz, 8bits, stereo	Reasonable in stereo	22050
11025 Hz, 16bits, mono	Reasonable with little noise	22050
11025 Hz, 16bits, stereo	Reasonable in stereo, less noise	44100
22050 Hz, 8bits, mono	Good for music and simple sounds	22050
22050 Hz, 8bits, stereo	Good in stereo	44100
22050 Hz, 16bits, mono	Very good, less noise	44100
22050 Hz, 16bits, stereo	Very good in stereo, less noise	88200
44100 Hz, 8bits, mono	High, all types of sound	44100
44100 Hz, 8bits, stereo	High, in stereo	88200
44100 Hz, 16bits, mono	Excellent, less noise	88200
44100 Hz, 16bits, stereo	Excellent in stereo (CD), no noise	176400

25

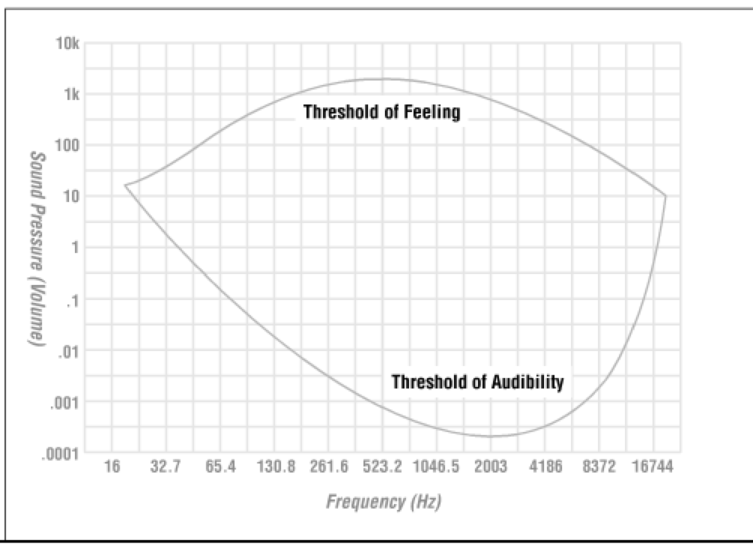
Compression

Perceptual coding techniques:

- **Minimum threshold**: It is not necessary to represent sounds below this threshold.
- **Mask effect**: The stronger sounds “mask” the weaker ones.
- ***Joint Stereo***: Below certain frequencies the spatial origin of the sound is not detected.
- **Huffman coding**: Final coding after applying all the other techniques.

26

Sound Thresholds



27

Storage

- **CD-Audio**: Complete name of current CDs
  - Coding: PCM, Bandwidth: 20KHz, Samples: 16 bits, Duration: 1 Hour
- **CD-ROM**: Uses the CD-Audio format for arbitrary data
  - ISO9660 file system
- **CD-I**: A CD-ROM application. The data (audio, video and text) are interleaved to support synchronized presentation
- **DVD**...
- Cloud

Transforms

A sampled and quantized signal can be transformed to other representation

- More efficient to transmit, store, etc.

- Fourier transform:
  - Fourier coefficients represent the signal in the frequency domain
- Sub-band coding (MPEG):
  - Explores the properties of the human ear. The more relevant frequency bands are coded in more detail.
- Parametric representations:
  - Numeric model (Example: FM)

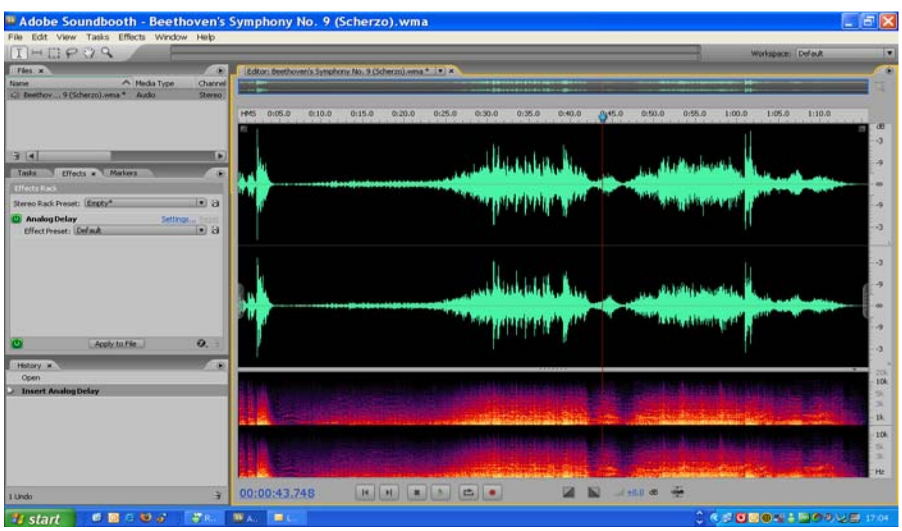
Audio Processing

Consider the signal as a sequence of numbers (samples)

- Example: Replace each number by the average (mode, median) of its neighbors
- Others: low pass filter, high pass filter, reject band filter
- Change the amplitude level is done with a multiplication by a factor
- The signal when delayed and added to itself results in echo
- ...

30

Audio Editing

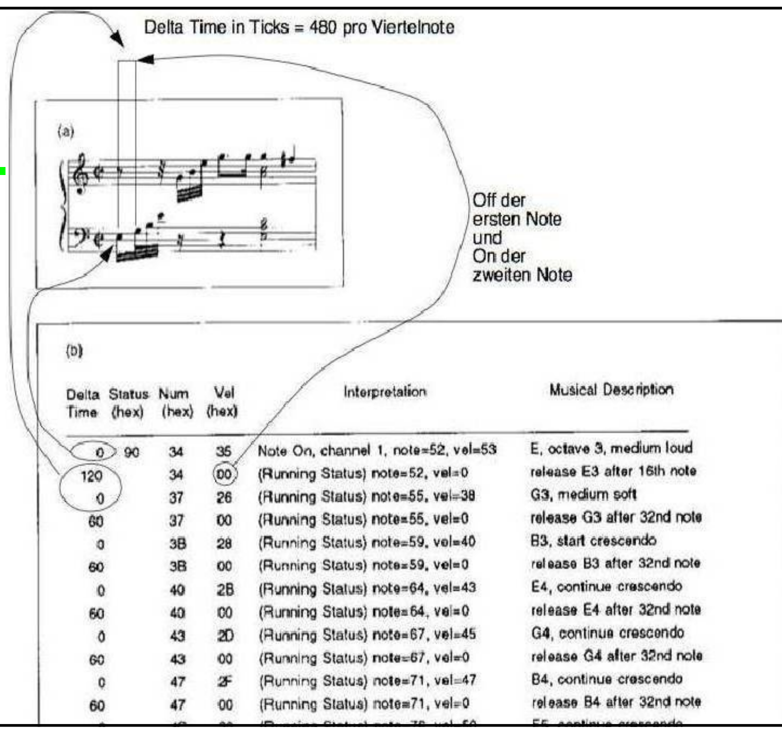


31

Music Representation

- MIDI (Music Instrument Digital Interface): Standard for music representation
- Originally for hardware but currently used in software
- MIDI messages define musical events: note on, off, tone variations, etc.
- Time representation needed for storage: instant for each event depending on the "tempo" information specified at the beginning of the file

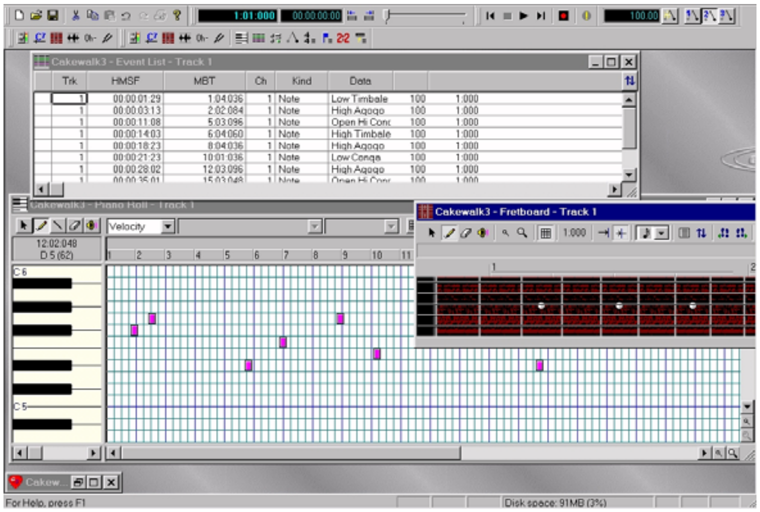
32



Delta Time	Status (hex)	Num (hex)	Vel (hex)	Interpretation	Musical Description
0	90	34	35	Note On, channel 1, note=52, vel=53	E, octave 3, medium loud
120		34	00	(Running Status) note=52, vel=0	release E3 after 16th note
0		37	26	(Running Status) note=55, vel=38	G3, medium soft
60		37	00	(Running Status) note=55, vel=0	release G3 after 32nd note
0		3B	28	(Running Status) note=59, vel=40	B3, start crescendo
60		3B	00	(Running Status) note=59, vel=0	release B3 after 32nd note
0		40	2B	(Running Status) note=64, vel=43	E4, continue crescendo
60		40	00	(Running Status) note=64, vel=0	release E4 after 32nd note
0		43	2D	(Running Status) note=67, vel=45	G4, continue crescendo
60		43	00	(Running Status) note=67, vel=0	release G4 after 32nd note
0		47	2F	(Running Status) note=71, vel=47	B4, continue crescendo
60		47	00	(Running Status) note=71, vel=0	release B4 after 32nd note

33

Music Editing



Speech Synthesis and Recognition

Synthesis: Automatic conversion of text to speech

- Segmentation of the original sequence in morphemes (syntactic language units)
- Match between morphemes and sounds
- Synthesis using human speech models or sampled sounds
- To improve the quality other elements are used: prosody, emphasis

Recognition:

- Segmentation (there are no discrete natural units)
- Coefficients are compared against templates

Raster Image

- Projection of a 3D plane on a 2D plane
- Can be defined as a function of two variables, I(x,y). For each point (x,y), I(x,y) defines the light intensity in that point
- Common image types:
  - Grayscale images
    - I(x,y) defined in the range [a,b]
  - Binary images
    - I(x,y) defined in {0,1}
  - Color images
    - I_R I_G I_B

36

Digital Image and Video

- Represented by an array of values. Each value is a function of the information in the corresponding point in the image.
- An element of the array is a *pixel* (*picture element*).
- Each pixel can include all color components.
- It is common to represent pixels in grayscale with 8 bits and color pixels with three components (RGB) with 8 bits each.
- Video is a sequence of images...

37

Image

- **Vectorial: parametric description of the image content**
  - Described with 2D or 3D objects represented as models
  - Easy to change
  - Maintains structural information
  - Semantic content is preserved
  - Less realistic
- **Bitmap: pixel array**
  - Described with pixels
  - It does not support (easy) corrections
  - It does not maintain structural information
  - The semantic content is not preserved
  - Real world images - photos

38

Raster and Vectorial

- Size
  - Vectorial rectangle: Less than 30 bytes of alphanumeric information (less if compressed)
  - Non-compressed raster rectangle, black and white
    - 5000 bytes (200x200/8)
- Image refresh
  - In a vectorial image, with many objects, redrawing can be slow
- Resize
  - When resizing a vectorial image information is not lost

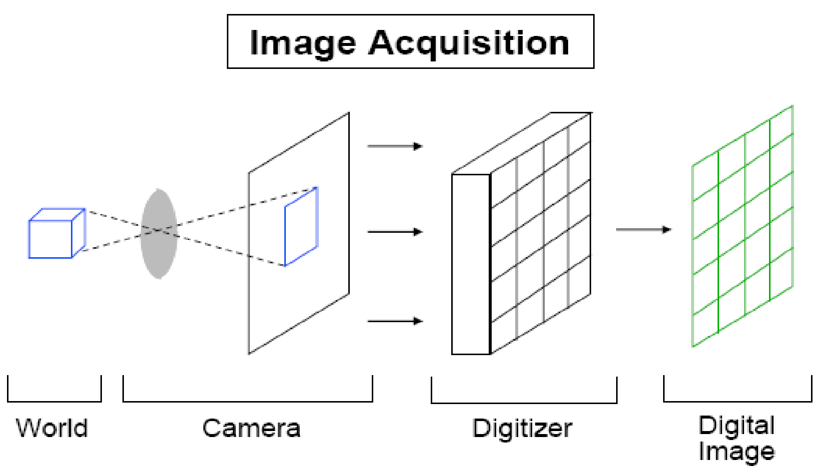
39

Sampling and Quantization

- Steps in the digitizing process (as in audio…)
- **Sampling**: How many samples? How many pixels are needed to represent the real world.
- **Quantization**: How many levels are used for each sample? Directly related with the number of bits for each sample. Example: 8 bits, 256 levels.

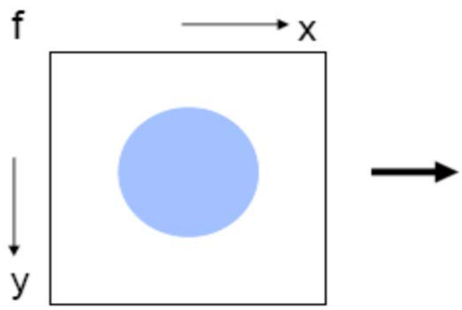
40

Image Acquisition



41

Image Sampling



	1	2	3	4	5
1	100	100	100	100	100
2	100	0	0	0	100
3	100	0	0	0	100
4	100	0	0	0	100
5	100	100	100	100	100

42

Color

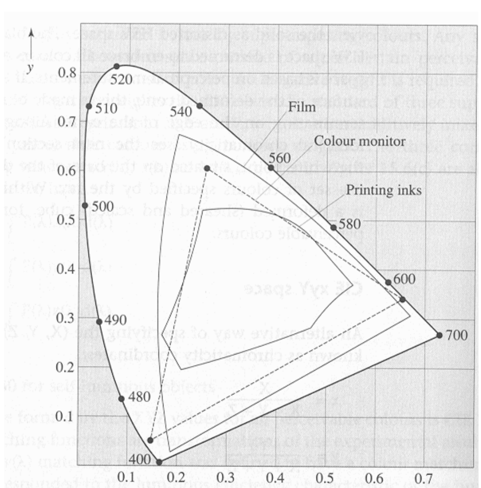
- The eye is sensitive to red, green and blue light (additive primary colors)
  - The human eye and the brain combine these three colors to generate the other colors
  - Orange in a computer monitor is a combination of two frequencies of green and red light and not the spectral frequency seen when looking at a orange object
- The light reflected from a paper sheet can be represented as a set of points using primary colors (in this case subtractive colors)

43

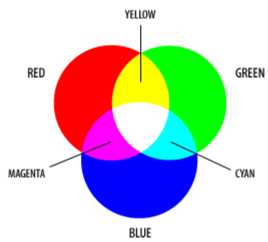
Color Models

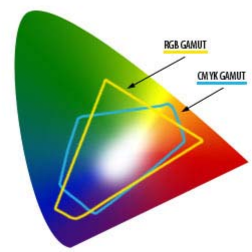
- Models or methodologies used to specify color in a computer
- RGB (red, green, blue) – a color is specified through the amount of red, green and blue that it contains
- HSB (hue, saturation, brightness) and HSL (hue, saturation, lightness)
  - Hue (color) – angle from 0 to 360º of a color circle
  - Saturation – percentage of the color intensity
    - 100% pure color
  - Brightness or lightness – percentage of black or white mixed with the color
    - 100% - white
    - 0% - black
- Other models: CMYK, CIE, YIQ, YUV,

Color Models



Color Models

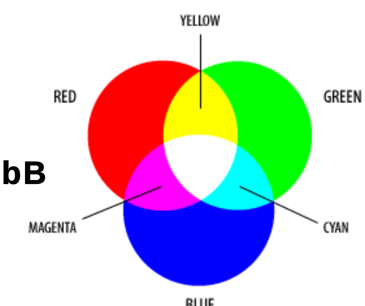




46

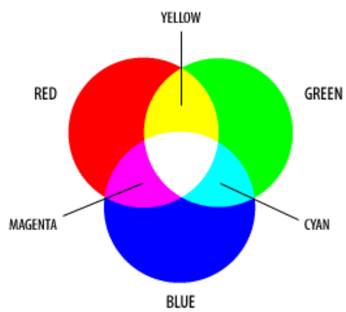
RGB Color Model

- *De facto* standard
- **Additive** primary colors: **R**ed **G**reen **B**lue
- **C** = (r,g,b) = **rR + gG + bB**
- Secondary colors: CMY
- Good for **screens** (light sources) and **scanners** that detect light reflected by documents



47

CMYK Color Model



Secondary Colors RGB (CMY):

Cyan **M**agenta **Y**ellow

color	absorbs	reflects
C	R	B + G
M	G	B + R
Y	B	R + G
blac**K**	all	none

Black was added to the model because CMY does not support it directly

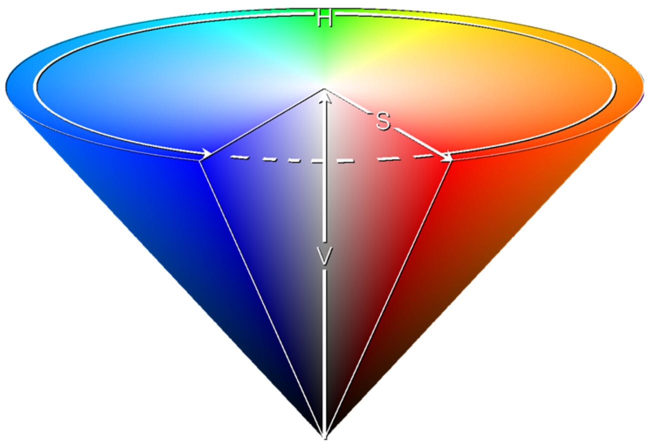
48

HSV Color Model

- HSV (Hue, Saturation, Value) or HSB (H, S, Brightness)
- Correspond to the artistic concepts Tint, Tone and Shade
- Oriented to the user
- Hue (color) - angle of 0 to 360th of a circle of color
  - polar coordinates
- Saturation (Saturation) - percentage of color intensity
  - 100% pure color
- Value, Brightness - percentage of black or white that is mixed with the color
  - 100% - white
  - 0% - Black

49

HSV Color Model



50

YUV Color Model

- Separation between luminance and color
  - The human eye is more sensitive to the light than to the color. Bandwidth can be saved in the color.
- **Y**: "luma", **luminance** Y = 0.2125*R + 0.7154*G + 0.0721*B
  - The human eye is more sensitive to green (G) and less sensitive to blue (B)
- **U**, **V**: **crominance** (B-Y, R-Y)
  - Represent color "differences" : the difference between a color component (B or R) and the luminance.

53

Compression

- Compression level
  - Can be defined based on the number of bits per pixel (e.g., 2 bits per pixel)
- Image quality
  - *Lossy* and Lossless
- Compression and decompression speed
- Applications define the most important criteria
  - In images decompression should be faster
  - In video capture compression is faster

55

Redundancy and Visibility

- **Redundancy**: occurs when the same information is stored (transmitted) more than once:
  - Several adjacent pixels with the same color
  - Adjacent lines that are equal
  - Images with few or no changes (video)
- **Visibility**: it is not necessary to store or transmit what is not visible
  - The human eye cannot distinguish well some colors
  - The color information can have less resolution than the luminance information (used in TV/video)

56

Compression Techniques

- **Truncation**: Reduction of the quantity of information by reducing the number of bits per pixel
  - Examples: RGB 5:5:5, YUV 6:5:5
- **Color Table**: Pixels represent an index in a table that contains more bits per pixel (usually 24)
  - Limited number of colors (8bits => 256 color)
- ***Run-length* coding**: Pixel blocks are replaced by a value and the number of pixels with the same color
  - Good for synthetic images, with large areas of the same color (synthetic images, cartoons, etc.)

57

Interpolative Techniques

- Transmission of part of the pixels and reconstruction of the others using interpolation
- Used for color components of the images
- Good for "real" images
  - Synthetic images can loose some of its elements
- Can be applied to images in a sequence
  - Some images are stored and the others are reconstructed from these

58

Predictive Techniques

The next element (pixel/row/image) is calculated based on the previous element

- DPCM: Adjacent pixels are compared and differences between them are stored (less bits for differences)
  - Not good for abrupt changes
- ADPCM: The values represented by each bit sequence changes, keeping the same number of bits
- Video (MPEG)...

59

Transforms

Convert the signal for other representations that are more efficient

- Statistical coding: Based on the statistical distribution of pixels in the image
  - More bits for less frequent values
  - Example: Huffman coding
- DCT (Discrete Cosine Transform)
  - Applied to a block of adjacent pixels, typically 8x8
  - Results in 64 values that represent the amplitudes of increasing spatial frequencies

60

Huffman Coding

Example: Representation of five symbols

A	**00**	.30	**00**	.30	**1**	.45	**0**	.55
B	**01**	.25	**01**	.25	**00**	.30	**1**	.45
C	**11**	.20	**10**	.25	**01**	.25		
E	**100**	.15	**11**	.20				
D	**101**	.10						

61

BMP

- Raster format for Windows
- Device-independent bitmap (DIB) format
- Allows Windows to display the image in any visualization device
- Named device independent, because pixel color is specified in a way that is independent from the color representation in the device

62

BMP

File Structure

- Bitmap File Header: Information about the type, size, and structure of the image
- Bitmap Information Header: Specifies the dimensions, compression type, and color format
- Colormap
  - Contains as much elements as colors in the image
  - When each pixel has 24 bits the colormap does not exist. The color information will be contained in the data area of the bitmap.
- Pixels
  - Scan lines – Consecutive rows (bottom to top, left to right)
  - The number of bytes in each scanline depends on the format, color, and width in pixels of the image

63

RLE in BMP Format

RLE8

Encoded Bytes	Decoded Bytes
05 10	10 10 10 10 10
00 05 23 65 34 56 45	23 65 34 56 45
0A 0A	0A 0A 0A 0A 0A 0A 0A 0A 0A 0A
00 04 46 57 68 79	46 57 68 79

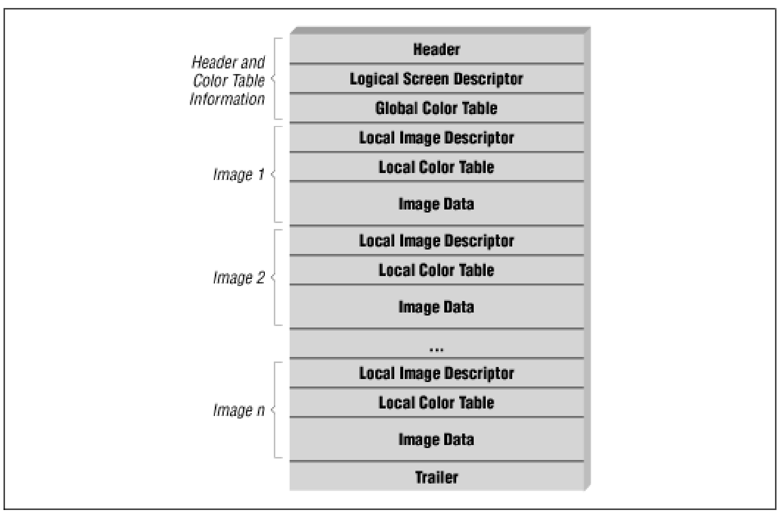
64

GIF

- Graphics Interchange Format (tm)
- Created by Compuserve to store color raster images
- 8 bit images, better for graphics and drawings
- Compression – Variable Length Code LZW
  - Uses a table with codes constructed from patterns in the original data
  - Each pattern is introduced in the table and the code is used to replace the pattern
- Supports interlacing, with four steps
- Two versions: GIF87a and GIF 89a
- A GIF file may contain several images, with different color maps (Local Color Map). It may also include a global color map.

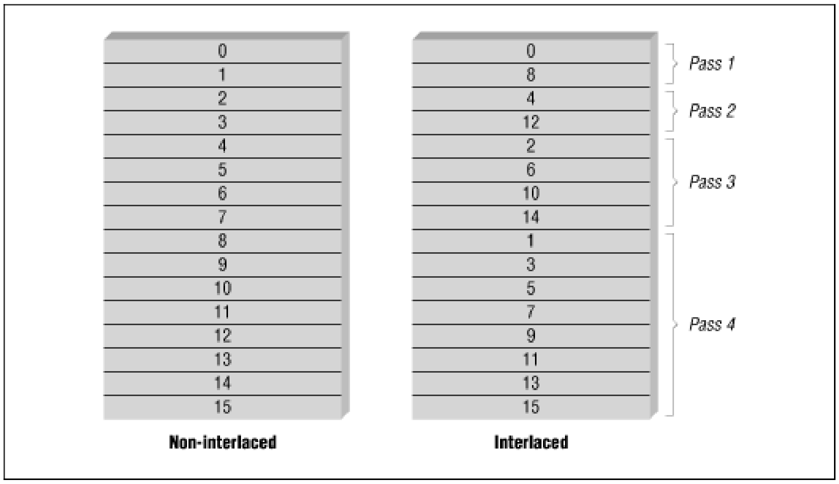
65

GIF



66

GIF



67

Color Table

Byte #	Content
1	Red intensity for color index 0
2	Green intensity for color index 0
3	Blue intensity for color index 0
4	Red intensity for color index 1
5	Green intensity for color index 1
6	Blue intensity for color index 1
...	...

68

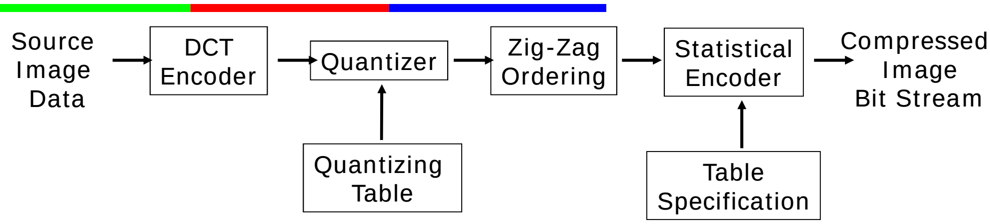
JPEG

Joint Photographic Expert Group

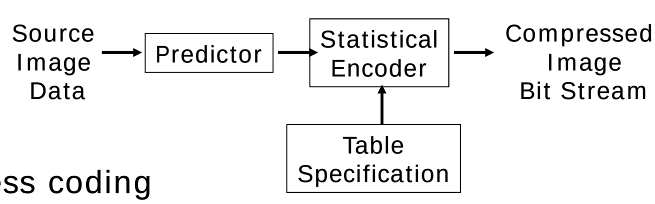
- “Real” images (*continuous-tone imaging*)
- Four modes:
  - Sequential coding
  - Progressive coding
  - Hierarchical coding
  - Lossless coding
- Lossy: DCT (Discrete Cosine Transform) coding
- Lossless: Predictive and statistical coding

69

JPEG



Sequential coding



Lossless coding

70

JPEG Compression

- JPEG takes advantage of the limitations of the human visual system, including the difficulty to distinguish details
  - Even more relevant for color images
- An image that changes a lot in a few pixels has a high spatial frequency
- Images are represented in the spatial frequency domain. The “rate” of change in x and y is calculated and a new “image” is generated with the coefficients that represent this change
- The coefficients that represent small changes (low frequency) are preserved. The smaller coefficients, corresponding to the larger values, are divided by integer values, and will turn into zero (quantization).
- The JPEG quality can be changed by changing the values that are used in the division (quantization table)

71

JPEG Compression

- The DCT application considers that:

1. The image doesn’t change much in small areas, e.g., 8x8 pixel blocks
2. Experiments suggest that humans are less sensitive to the loss of high frequency components
3. The visual acuity – ability to distinguish close lines – is higher in grayscale (black lines) than in color

72

JPEG Compression

- Uncompressed data: 24 bits per pixel
- Lossless compression: average 2:1
- An higher resolution image can be more compressed
- Photographic images in JPEG
  - JPEG 10:1 to 20:1 - without visible loss
    - Storage requirements: 1 or 2 bits per pixel
  - JPEG 30:1 to 50:1 – some problems (could be used in archive indices)
  - JPEG 100:1 – uses the same memory as a thumbnail images (with much more detail)
- Black and white images – the compression ratio is smaller (the eye is more sensitive to brightness variations)
  - Black and white images have 10%-25% reduction when compared with the same color JPEG
  - The level of visible loss corresponds to a compression of 5:1

73

JPEG

Input Array

132	136	138	140	144	145	147	155
136	140	140	147	140	148	155	156
140	143	144	148	150	152	154	155
144	144	146	145	149	150	153	160
150	152	155	156	150	145	144	140
144	145	146	148	143	158	150	140
150	156	157	156	140	146	156	145
148	145	146	148	156	160	140	145

Output Array

172	-18	15	-8	23	-9	-14	19
21	-34	24	-8	-10	11	14	7
-9	-8	-4	6	-5	4	3	-1
-10	6	-5	4	-4	4	2	1
-2	-2	-3	5	-3	3	4	6
-1	-2	-4	6	-4	4	2	-1
-5	-3	-4	5	6	3	1	1
-1	-8	-4	3	2	1	4	0

74

DCT

$$X_{k_1,k_2} = \sum_{n_1=0}^{N_1-1} \left(\sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right]\right) \cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right]$$
$$= \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \cos\left[\frac{\pi}{N_1}\left(n_1+\frac{1}{2}\right)k_1\right] \cos\left[\frac{\pi}{N_2}\left(n_2+\frac{1}{2}\right)k_2\right].$$

The N real numbers $x_0, \ldots, x_{N-1}$ are transformed into the N real numbers $X_0, \ldots, X_{N-1}$

75

JPEG

Quantization array (Quantum(i,j))

4	7	10	13	16	19	22	25
7	10	13	16	19	22	25	28
10	13	16	19	22	25	28	31
13	16	19	22	25	28	31	34
16	19	22	25	28	31	34	37
19	22	25	28	31	34	37	40
22	25	28	31	34	37	40	43
25	28	31	34	37	40	43	46

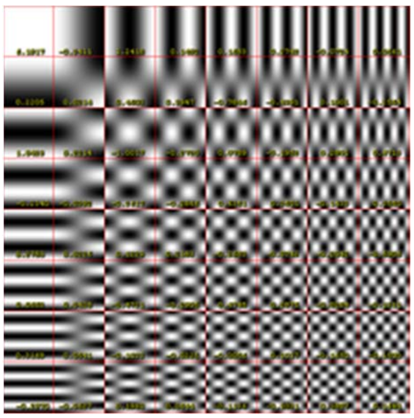
QuantizedCoefficient(i,j)=DCT(i,j)/Quantum(i,j)

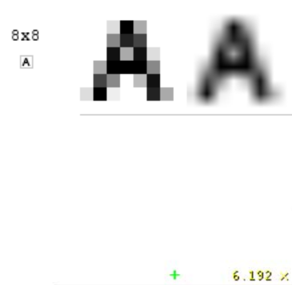
Coefficients array after quantization

43	3	2	0	0	0	0	0
3	3	2	0	0	0	0	0
1	0	0	0	0	0	0	0
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

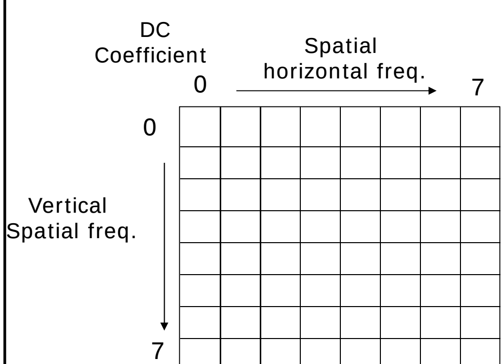
76

$$
\begin{bmatrix}
6.1917 & -0.3411 & 1.2418 & 0.1492 & 0.1583 & 0.2742 & -0.0724 & 0.0561 \\
0.2205 & 0.0214 & 0.4503 & 0.3947 & -0.7846 & -0.4391 & 0.1001 & -0.2554 \\
1.0423 & 0.2214 & -1.0017 & -0.2720 & 0.0789 & -0.1952 & 0.2801 & 0.4713 \\
-0.2340 & -0.0392 & -0.2617 & -0.2866 & 0.6351 & 0.3501 & -0.1433 & 0.3550 \\
0.2750 & 0.0226 & 0.1229 & 0.2183 & -0.2583 & -0.0742 & -0.2042 & -0.5906 \\
0.0653 & 0.0428 & -0.4721 & -0.2905 & 0.4745 & 0.2875 & -0.0284 & -0.1311 \\
0.3169 & 0.0541 & -0.1033 & -0.0225 & -0.0056 & 0.1017 & -0.1650 & -0.1500 \\
-0.2970 & -0.0627 & 0.1960 & 0.0644 & -0.1136 & -0.1031 & 0.1887 & 0.1444
\end{bmatrix}
$$

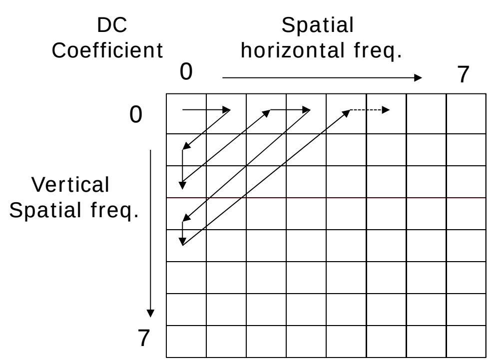




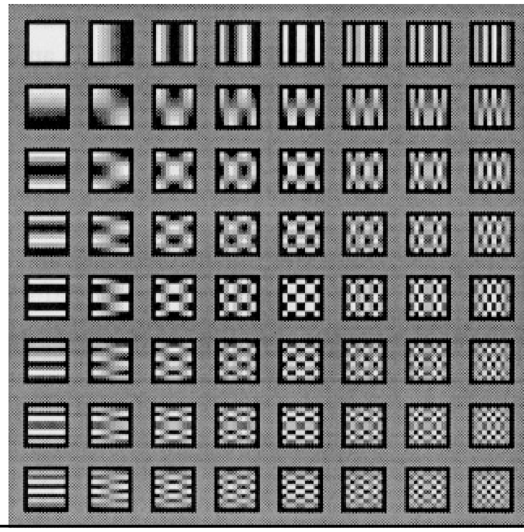
JPEG



Zig-Zag sequence



DCT



80

JPEG

- Images are converted to YUV and sub-sampled
- After the lossy steps the AC coefficients are RLE/RLC encoded
  - This step operates on a 64 values vector resulting from the zig-zag scan
- The encoding replaces values with pairs (RUNLENGTH, VALUE), representing the number of zeros and the value
  - The pair (0,0) marks the last non-zero value
- Example:
  (32, 6, -1, -1, 0, -1, 0, 0, 0, -1, 0, 0, 1, 0, 0, ..., 0)
  (0,6)(0,-1)(0,-1)(1,-1)(3,-1)(2,1)(0,0)

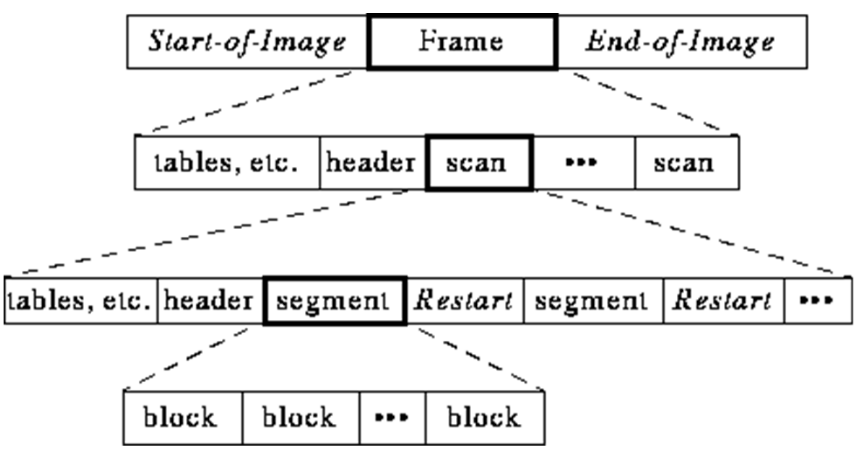
82

JPEG

- The DC components are DPCM encoded
- The values do not change (much) across neighbor blocks
- With DC coefficients 150, 155, 149, 152 e 144
  the result would be 150, +5, -6, +3, -8
- Applied to the entire image
- Huffman coding applied to the bit counting
  (**SIZE**, AMPLITUDE)
- (**8**, 10010110), (**3**, 101), (**3**,001), (**2**, 11), (**4**,0111)
  - E.g., SIZE 2 can be represented with just one bit because it is more frequent

83

JPEG Bitstream



84

JPEG Bitstream

- A “frame" is an image, a "segment" is a group of blocks, a "block" is a 8 x 8 pixels group

- Frame header:
  - (width, height) of image
  - number of components
  - unique ID (for each component)
  - horizontal/vertical sampling factors (for each component)
  - quantization table to use (for each component)
- Scan header
  - Number of components in scan
  - component ID (for each component)
  - Huffman table (for each component)
- Misc. (can occur between headers)
  - Quantization tables
  - Huffman Tables
  - Arithmetic Coding Tables
  - Comments
  - Application Data

85

Color Quantization

- GIF requires cheap visualization hardware: 8 bits per pixel (256 colors)
- Images with more than 256 colors: the computer chooses an appropriate set of colors and the entire image is mapped to those colors (color quantization)
- Lossy process
- The quantization details have more impact in the final image than the errors introduced by JPEG
- Display a JPEG image in hardware with 8 bits or less requires less quantization
  - The speed and quality of the image depend on the algorithms used by the JPEG decoder
  - The variation of the image quality is higher than in 24 bits

86

GIF Quantization

- The color quantization is performed when the GIF image is generated
- The software for GIF visualization does not have to execute quantization while JPEG software always has to (GIF viewers are faster than JPEG viewers)
- In GIF files the quantization is always the one performed when the file was generated
  - Colors may not exist
  - The quantization algorithm used to generate the file may be of poor quality

87

JPEG Quantization

- When an image is visualized in hardware that is different from the one where the image was created JPEG has more quality
- A 24 bit JPEG image can be quantized for the hardware where visualization will occur
- Future improvements in quantization algorithms can be used in existing JPEG images
- Future visualization hardware can use JPEG images, since all color information is kept

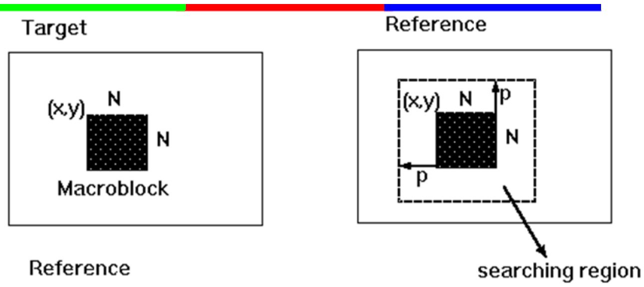
88

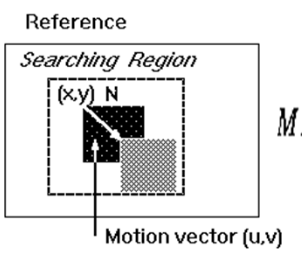
Video Compression

- Temporal redundancy
- Based in Motion Compensation:
  - Motion estimation (search based on motion vectors)
  - Prediction based on motion compensation
  - Prediction error – the difference
- Image is divided in macroblocks of size NxN, usually with N=16 for luminance and N=8 for chrominance

89

Video Compression





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

The objective is to find (u,v) that minimizes the MAE (Mean Absolut Error)

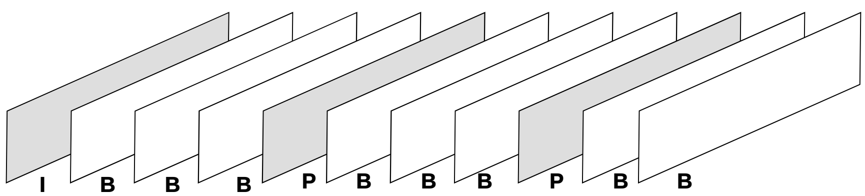
90

MPEG

Motion Picture Expert Group

- Good quality for rates between 1.0 e 1.5 Mbps
- Specification of the format (bitstream) and the decoding process
- Three main images types:
  - **I**: Coded independently from other images
  - **P**: Coded using motion compensation from other I or P images
  - **B**: Interpolated between previous and future I or P images

MPEG



P - Forward Prediction
B - Interpolation (Bi-directional prediction)
I - Possible Random Access Points

MPEG

- I images take more space: similar to a JPEG image
- P images take approx. 1/3 of I images. Include macroblocks of 16x16 with motion compensation information and 8x8 DCT blocks for correction
- B images take between 1/2 and 1/5 of P images. They also include macroblocks and blocks with interpolation parameters and DCT values for correction
- …more B images should be used to achieve more compression

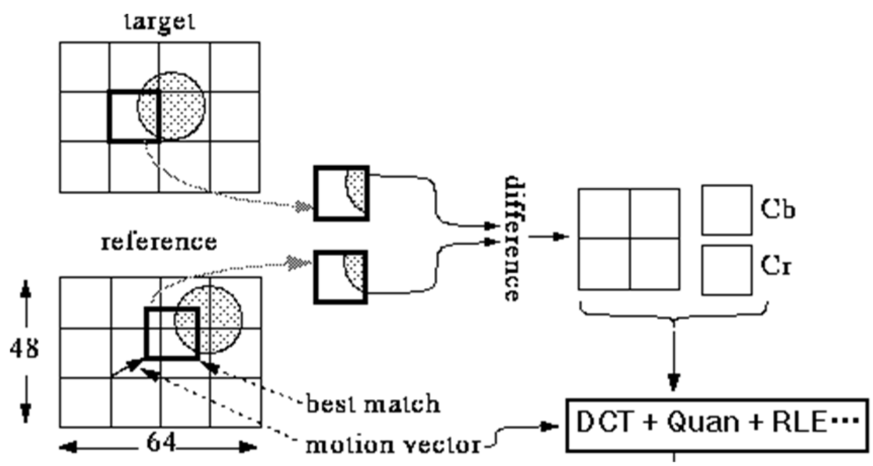
93

MPEG

- I images
  - Macroblocks are independent
- P images
  - Independent macroblocks
  - Motion vector + error terms regarding a previous I or P image
- B images
  - Independent macroblocks
  - Motion vector + error terms regarding a previous I or P image
  - Motion vector + error terms regarding a future I or P image
  - 2 motion vectors + error terms regarding a previous/future I or P image

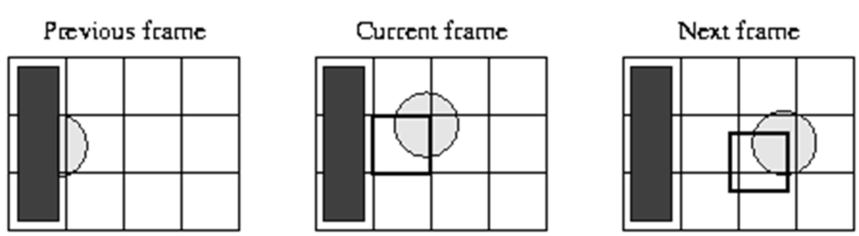
94

Motion Based Coding



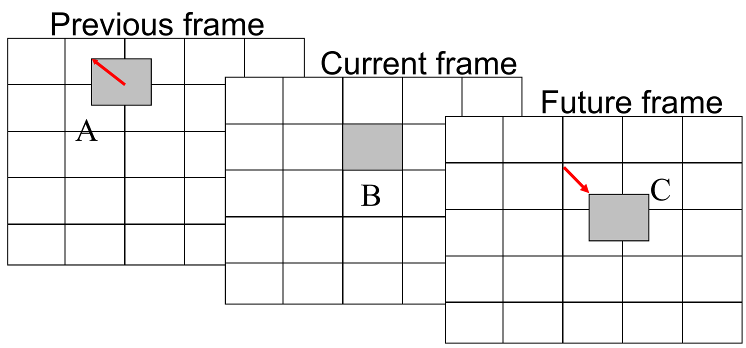
95

MPEG (B Frames)



96

MPEG (B Frames)



B = A⇒ forward prediction
B = C⇒ backward prediction
or B = (A+C)/2 ⇒ interpolation

one motion vector
two motion vectors

97

MPEG-4

- MPEG-4 adopts a object coding paradigm
- Media objects can be natural or synthetic
- Other application areas, including interaction
- Object composition in a scene using BIFS (Binary Format for Scenes)
  – Scene graph similar to VRML (Virtual Reality Modeling Language), with extensions
- Several levels:
  – (1) Video object sequence (VS); (2) Video Object (VO); (3) Video Object Layer (VOL); (4) Group of Video Object Planes (GOV); (5) Video Object Plane

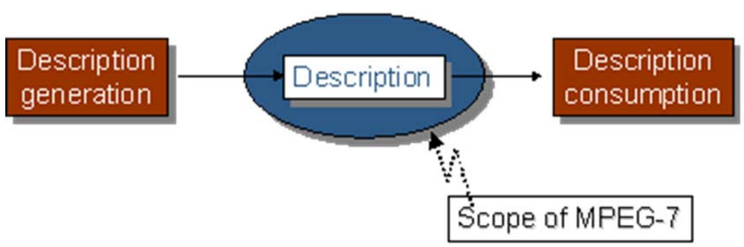
99

MPEG-7

- Multimedia content description
- Used in information retrieval
- Main concepts:
  - Feature: Information feature.
  - Descriptor (D): Syntactic and semantic definition of the feature.
  - Description Scheme (DS): Specification of the structure and relations of Ds and DSs.
  - Description: Set of instances of Ds and DSs describing content.
  - Description Definition Language (DDL): Syntactic rules that combine Ds and DSs.

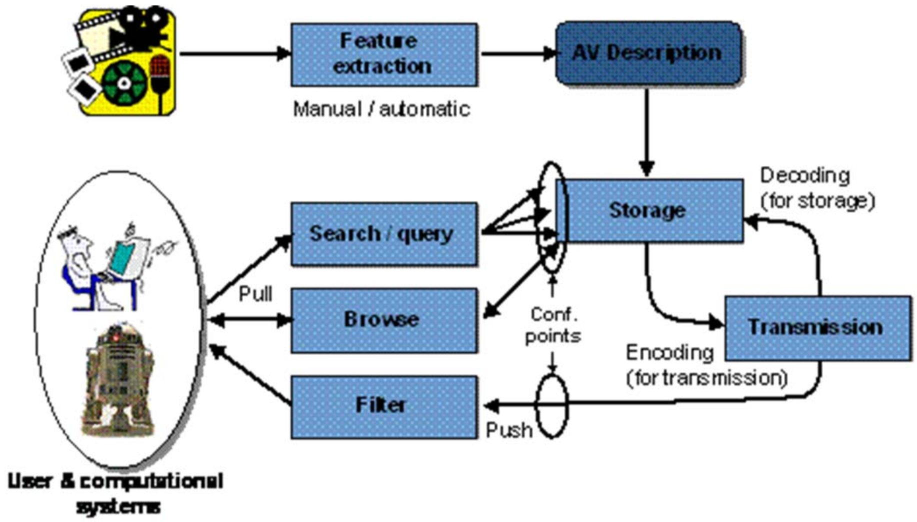
103

MPEG-7



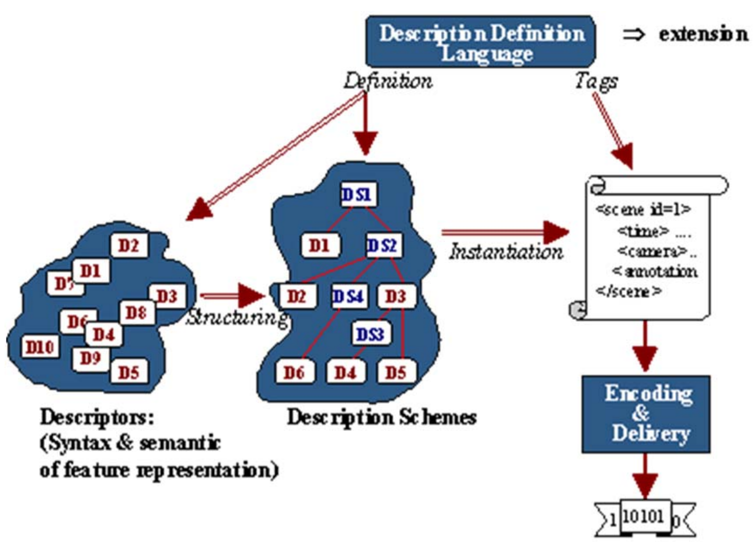
104

MPEG-7



105

MPEG-7



106

MPEG-7 (Descriptor Schemes)

- ...
- Structural Description: Segment DS, a section of an audiovisual object. Subclasses:
  – Audiovisual, Audio, Still region, Moving region, Video
- Conceptual Description: High level content description. For example, a person or an event in a game
- Summaries
- Partitions and Decompositions: View DS, several spatial and frequency descriptions

107

MPEG-7 (Descriptors)

Type	Feature	Descriptors
Visual	Color	DominantColor
		ScalableColor
		ColorLayout
		ColorStructure
		GoFGoPColor (extension of ColorStructure)
	Texture	HomogeneousTexture
		TextureBrowsing
		EdgeHistogram
	Shape	RegionShape
		ContourShape
		Shape3D
	Motion	CameraMotion
		MotionTrajectory
		ParametricMotion
		MotionActivity

Audio	Silence	Silence
	Timbre	InstrumentTimbre
		HarmonicInstrumentTimbre
		PercussiveInstrumentTimbre
	Speech	Phoneme
		Articulation
		Language
	Musical Structure	MelodicContour
		Rhythm
	SoundEffects	Reverberation, Pitch, Contour, Noise

108