




# SCC5909 – Fundamentos de Multimídia

## Aula 2

**Prof.: Dr. Rudinei Goularte**  
 (rudinei@icmc.usp.br)  
 Instituto de Ciências Matemáticas e de Computação - ICMC  
 Sala 4-234




# Conteúdo

- Áudio digital
- Imagens digitais
- Vídeo digital

2




# 1. Áudio Digital



# 1.1 - O quê é som?

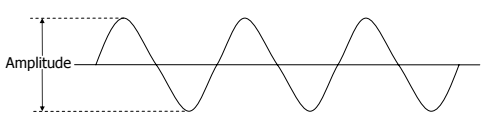
- Som é um fenômeno físico.
- Onda mecânica.

4




# 1.2 - Características físicas do som

- Amplitude



- Amplitude = Intensidade
  - Está relacionada ao volume do som. Quanto maior a amplitude, mais alto ouvimos o som.

5



# 1.2 - Características físicas do som

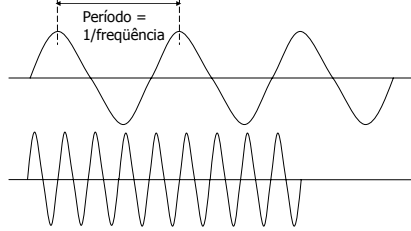
- Amplitude = Intensidade
  - Medida em decibéis (dB).

Intensidade	Exemplos típicos
0dB	Limite da audição
25dB	Estúdio de gravação
50dB	Escritório
70dB	Conversação típica
90dB	Home audio
120dB	Limiar da dor
140dB	Show de rock

6

## 1.2 - Características físicas do som

### Frequência



- Frequência determina altura do som (altura  $\neq$  volume).
- Frequências altas = altura maior = sons agudos.
- Frequências baixas = altura menor = sons graves.

7

## 1.2 - Características físicas do som

Categoria	Intervalo de Frequência
Infra-som	0 - 20 Hz
Som Audível	20 Hz - 20 KHz
Ultra-som	20 KHz - 1GHz
Hipersom	1 GHz - 10 GHz

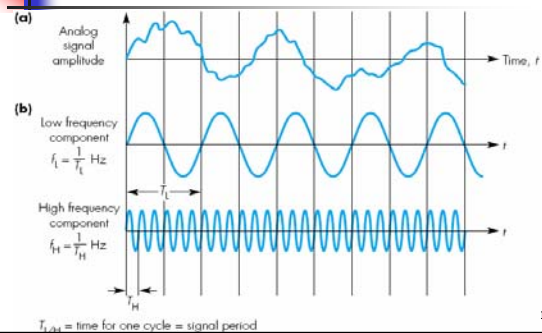
8

## 1.3 - Princípios de Digitalização

- Duas transformações:
  - Eletrônica: conversão de ondas mecânicas em sinais elétricos.
  - Digital: conversão de sinais elétricos em bits.
- Transformação eletrônica
  - Sinal de áudio.
  - Analógico.

9

## 1.3 - Princípios de Digitalização



10

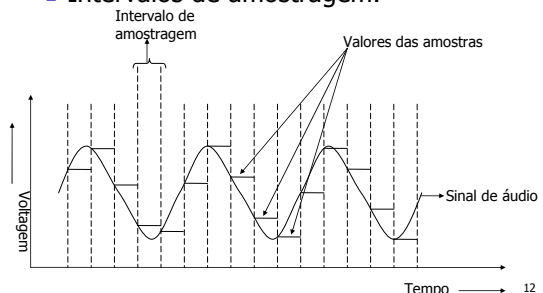
## 1.3 - Princípios de Digitalização

- Conversão analógico-digital.
  - Voltagem e tempo.
- Amostragem: realiza uma leituras periódicas e instantâneas da voltagem em espaços de tempo uniformes.
- Quantização: converte os valores analógicos amostrados em valores digitais.

11

## 1.4 - Amostragem

### Intervalos de amostragem.



12

## 1.4 - Amostragem

- O quanto deve ser amostrado?
  - Reconstruir exatamente o sinal = infinitas amostras.
  - Poucas amostras = sinal distorcido.

13

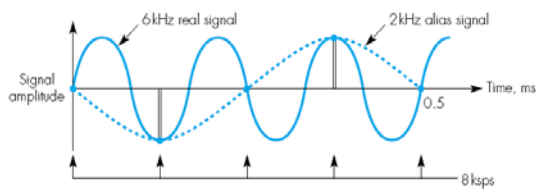
## 1.4 - Amostragem

- O quanto deve ser amostrado?
  - Teorema de Nyquist: "Para obter uma representação precisa de um sinal analógico, sua amplitude deve ser amostrada a uma **taxa** mínima igual ou superior ao dobro da componente de mais alta frequência presente no sinal". (taxa de Nyquist).

14

## 1.4 - Amostragem

- Aliasing.



15

## 1.4 - Amostragem

- Filtros anti-aliasing.
  - Removem componentes de alta frequência.
- Em sistemas multimídia:
  - A largura de banda do canal é normalmente menor que a largura de banda do sinal.
    - A taxa de amostragem é determinada pela largura de banda do canal.
    - A taxa de Nyquist será baseada na frequência mais alta suportada pelo canal.

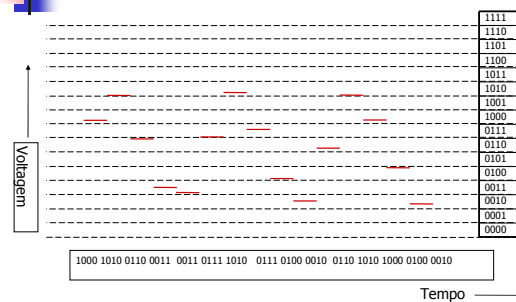
16

## 1.5 - Quantização

- Processo pelo qual os valores analógicos das amostras tomadas da amplitude do sinal são convertidos em valores digitais.
- Para reconstruir exatamente o sinal:
  - Necessidade de um número infinito de bits.
  - Usando um número finito de bits:
    - Representa-se cada amostra através de um número correspondente de níveis discretos.

17

## 1.5 - Quantização



18

## 1.5 - Quantização

- Amostragem e Quantização
  - Número de amostras x número de níveis.
    - Compromisso.
  - Quantização resulta em distorções.
  - Como descobrir o número ótimo de bits por amostra?

19

## 1.6 - Digitalização

- Taxas comuns de amostragem:
  - 8.000Hz, 11.025Hz, 22.050Hz e 44.100Hz (CD).
- Quantidades comuns de bits por amostra:
  - 4, 8, 16 e 24.
- Canais de som:
  - 1 (mono), 2 (estéreo), 3, 5, 7, ...
- Exemplo: Qualidade de CD:
  - Amostras a 44.100Hz (4,1 KHz), 16 bits por amostra e 2 canais de som (estéreo).

20

## 1.6 - Digitalização

- Circuito que realiza amostragem e quantização:
  - Conversor analógico-digital (*analog to digital converter* – ADC).
  - Caminho inverso: DAC. Usado na reprodução de áudio digital.
- Normalmente implementado em hardware.



21

## 1.6 - Digitalização

- Após a captura
  - os dados amostrados e quantizados devem ser "guardados" em algum formato – mídia de representação.
  - WAV e MP3, por exemplo.

22

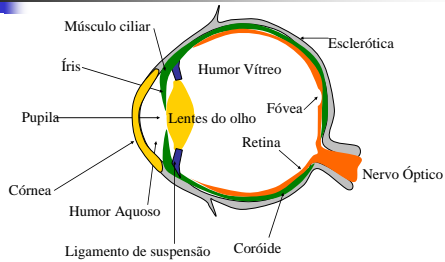
## 1.7 – Compressão de Áudio

- (Tema de Seminário)
- PCM, ADPCM – Voz
- Codificação Perceptiva – Voz e Música
  - Modelo Psicoacústico.

23

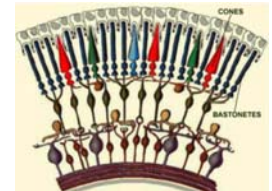
## 2. Imagens Digitais

## 2.1 - Sistema Visual Humano



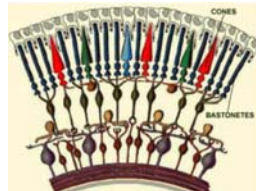
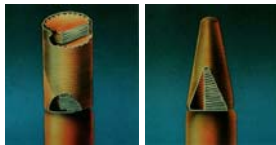
## 2.1 - Sistema Visual Humano

- Cones: fotoreceptores para cor.
  - 6 a 7 milhões.
  - Concentrados na fóvea.
- Bastonetes: fotoreceptores para intensidade luminosa.
  - 75 a 150 milhões.
  - Espalhados por toda a retina.



## 2.1 - Sistema Visual Humano

- Cones:
  - Teoria tristimulus da visão (*tristimulus theory of vision*)
  - Possuem três tipos de fotopigmentos:
    - Azul, Verde e Vermelho.
    - Sensibilidade: 430nm, 530nm e 560nm.
    - Porcentagem de cones: 4%, 32% e 64%.



## 2.1 - Sistema Visual Humano

Característica do SVH	Implicação p/ Multímídia
SVH é mais sensível à luminância do que à crominância.	Resolução da crominância pode ser reduzida sem grande prejuízo perceptível.
SVH é mais sensível a alto contraste.	Arestas (bordas) são importantes para a aparência da imagem (mudança na luminância).
SVH é mais sensível a baixas frequências espaciais.	Regiões com alta frequência espacial podem ser descartadas (menos bordas).
A resposta do SVH varia de pessoa para pessoa.	Múltiplos observadores devem ser usados em testes subjetivos de qualidade.

## 2.2 - Representação de Imagens

- Representação de imagens:
  - Na memória do computador:
    - Matriz de pixels armazenadas no frame-buffer.
  - Em arquivos:
    - Geradas por computador:
      - Gráficos.
    - Digitalizadas:
      - Documentos.
      - Imagens.

## 2.2 - Representação de Imagens

- Gráficos:
  - Dois modos de representação:
    - Mapas de bits
      - Arquivos maiores.
      - Não necessita de interpretadores.
      - BMP, TIFF (tagged image file format), ...
    - Comandos de alto nível.
      - Arquivos menores (somente instruções).
      - Necessita de interpretadores.
      - SVG (Scalable Vector Graphics), ...

## 2.2 - Representação de Imagens

- Imagens
  - Adquiridas por scanners ou câmeras.
  - Imagens de tom contínuo.
    - Monocromáticas: 8 bits por pixel.
    - Coloridas: de 8, 16, 24 ou 32 bits por pixel.
  - Conteúdo do frame-buffer em um arquivo.
    - Normalmente aplica-se compressão.
    - Diversos formatos
      - GIF, PNG, JPEG, ...

## 2.3 - O quê é compressão de imagens?

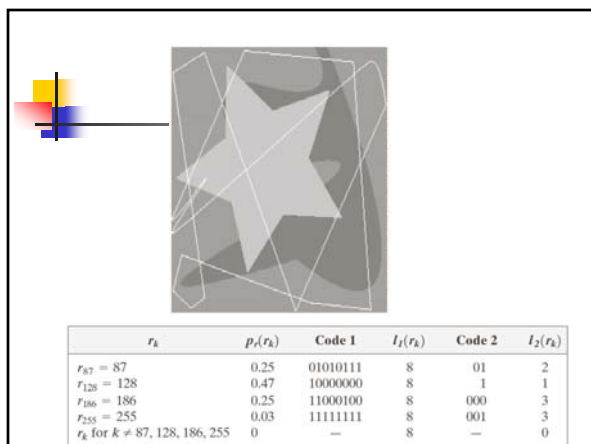
- “O termo *compressão de imagens* refere-se ao processo de reduzir a quantidade de dados necessários para representar uma imagem com uma qualidade subjetiva aceitável.”
- Dados x informação

## 2.4 - Redundâncias em Imagens

- Dados de imagem são altamente redundantes.
  - Remover redundâncias ajuda a alcançar compressão.
  - Redundâncias são matematicamente quantificáveis.
- Redundâncias em imagens:
  - Redundância Estatística.
    - Também conhecida como redundância de codificação.
  - Redundância Espacial.
    - Também conhecida como redundância interpixel.
  - Redundância Psicovisual.
    - Utiliza conceitos do HVS.

## 2.4 - Redundâncias em Imagens

- Redundância Estatística.
  - Função de Densidade de Probabilidade (pdf).
    - Texto.
  - Valores dos pixels em uma imagem tem pdf não uniforme.
  - Métodos de codificação estatística podem ser usados para compressão.



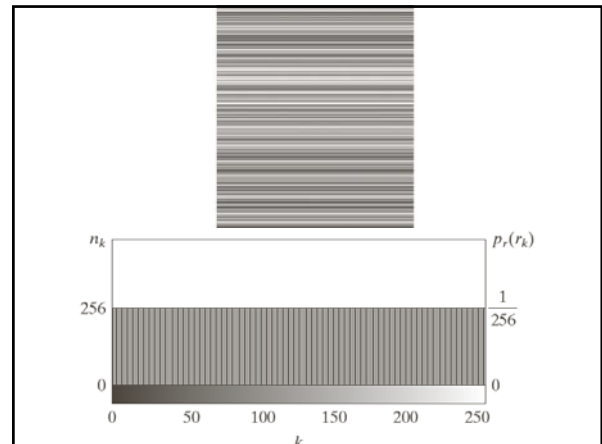
## 2.4 - Redundâncias em Imagens

- Redundância Estatística.
  - Codificação estatística.
    - *Variable Length Coding.*
      - Códigos menores para símbolos (valores) mais frequentes.
    - Lossless.
  - Huffman, codificação aritmética.

## 2.4 - Redundâncias em Imagens

### ■ Redundância Espacial

- Se refere à correlação entre pixels vizinhos em uma imagem.
- Relação geométrica ou estrutural entre os objetos em uma imagem.



## 2.4 - Redundâncias em Imagens

### ■ Redundância Espacial

- O valor de um pixel pode ser razoavelmente "adivinhado" por meio dos valores de seus vizinhos.
- Para remover redundância espacial:
  - Codificação por diferença, codificação run-length.

## 2.4 - Redundâncias em Imagens

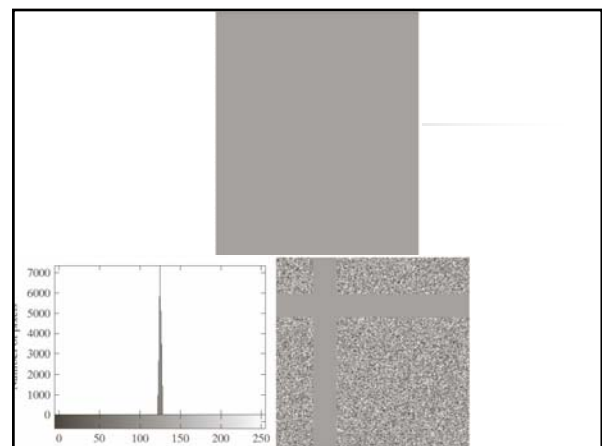
### ■ Redundância Psicovisual

- Percepção de brilho.
  - Olho não responde com igual sensibilidade a toda informação visual.
  - Algumas informações tem mais importância relativa que outras.
    - Informação psicovisual redundante.
- Diferente das outras redundâncias.
  - Está associada com informação visual de fato.
  - Então como é possível eliminá-la?

## 2.4 - Redundâncias em Imagens

### ■ Redundância Psicovisual

- Sua eliminação implica em perda de informação visual quantitativa (real).
  - Daí o nome *quantização*.
- É uma operação irreversível.



## 2.4 - Redundâncias em Imagens

### ■ Redundância Psicovisual

- Propriedades do sistema visual humano (HSV).
  - Maior sensibilidade a distorções em áreas suaves (com baixa frequência espacial).
  - Maior sensibilidade a distorções em áreas escuras de imagens.
  - Em imagens coloridas, maior sensibilidade a mudanças na luminância do que na crominância.

## 2.4 - Redundâncias em Imagens

### ■ Técnicas podem ser combinadas!

- Estatísticas + Espaciais + propriedades do HSV
- Vantagem?

## 2.5 - O Padrão JPEG

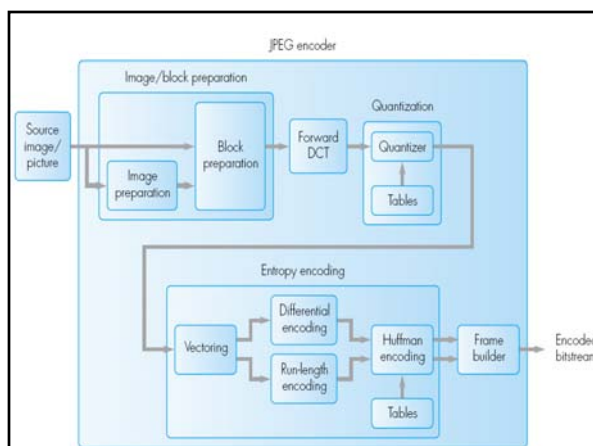
### ■ Joint Photographic Experts Group.

- ISO, CCITT e IEC.
- Padrão para codificação de imagens estáticas de tons contínuos.
- Possui 4 modos de operação:
  - Sequencial (*baseline mode*).
  - Progressivo.
  - Sem perdas.
  - Hierárquico.

## 2.5 - O Padrão JPEG

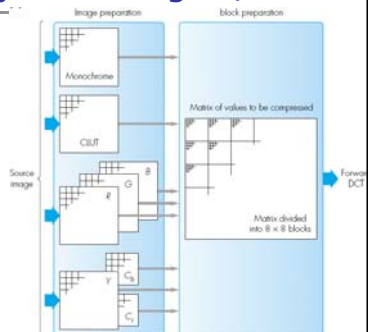
### ■ Modo sequencial

- É um método de compressão com perdas.
- Possui 5 etapas principais:
  - Preparação da imagem/bloco.
  - DCT.
  - Quantização.
  - Codificação.
  - Construção do quadro.



## Preparação da imagem/bloco

- Imagem é dividida em blocos de 8 x 8 pixels.
- Isso permite aplicação mais eficiente da DCT.

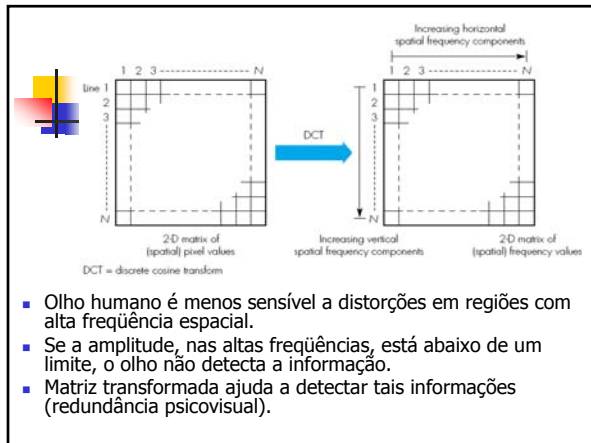






## Transformada DCT

- Transformada Discreta de Cossenos (DCT).
- Transformadas:
  - Transformam a informação de um formato (domínio) para outro.
- Transformada DCT aplicada a imagens:
  - Transforma matriz (imagem) em matriz de frequências espaciais.
  - Não produz perdas.



## Transformada DCT

- Transformada Discreta de Cossenos (DCT)
  - Todos os blocos, um a um, são submetidos à DCT.

$$F[i, j] = \frac{1}{4} C(i) C(j) \sum_{x=0}^7 \sum_{y=0}^7 P[x, y] \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16}$$

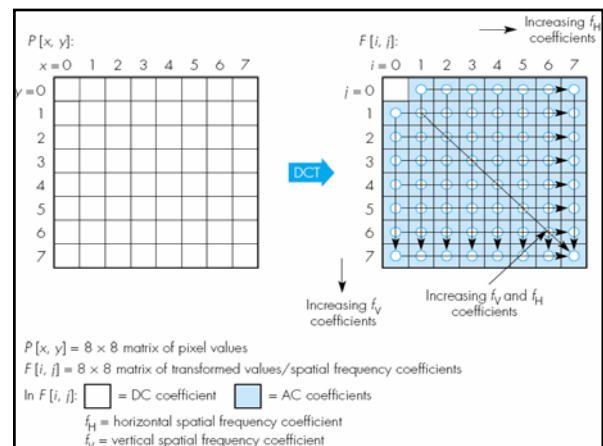
onde  $C(i)$  e  $C(j) = 1/\sqrt{2}$  para  $i, j=0$   
 $= 1$  para todos os outros valores de  $i$  e  $j$ .  
 $x, y, i$  e  $j$  todos variam de 0 a 7.

## Transformada DCT

- Transformada Discreta dos Cossenos Inversa (IDCT)

$$P[x, y] = \frac{1}{4} \sum_{i=0}^7 \sum_{j=0}^7 C(i) C(j) F[i, j] \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16}$$

onde  $C(i)$  e  $C(j) = 1/\sqrt{2}$  para  $i, j=0$   
 $= 1$  para todos os outros valores de  $i$  e  $j$ .  
 $x, y, i$  e  $j$  todos variam de 0 a 7.



## Transformada DCT

- Após DCT:
  - As regiões da imagem que possuem uma única cor geram matrizes com coeficientes DC idênticos (ou próximos) e poucos coeficientes AC diferentes.
  - As regiões da imagem que possuem transições de cores geram matrizes com coeficientes DC distintos e muitos coeficientes AC diferentes.
- Tamanho do bloco na imagem.
- Implicações para compressão.

## Transformada DCT

- DCT (2-D) é computacionalmente intensiva.
  - 1 bloco: 64 mult \* 64 adic = 4096 operações.
- DCT é:
  - Separável: DCT 1-D.
  - Simétrica: repetição de cálculos.

## Transformada DCT

- Algoritmos rápidos (1-D)

$$F_x = \frac{C(x)}{2} \sum_{i=0}^7 f_i \cos\left(\frac{(2i+1)x\pi}{16}\right) \quad f_i = \sum_{x=0}^7 \frac{C(x)}{2} F_x \cos\left(\frac{(2i+1)x\pi}{16}\right)$$

- DCT 1-D aplicada às linhas e colunas.
  - Separável.

## Transformada DCT

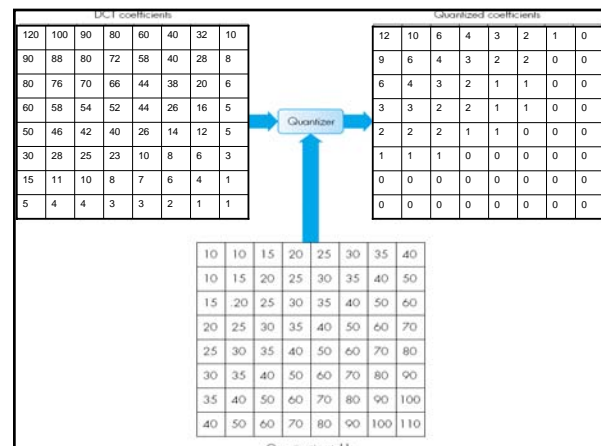
Table 7.2 Comparison of 1-D DCT algorithms (8 point DCT)

Source	Multiplications	Additions
'Direct'	64	64
Chen <sup>3</sup>	16	26
Lec <sup>4</sup>	12	29
Loeffler <sup>5</sup>	11	29
Arai <sup>6</sup>	5	28

## Quantização

### Quantização

- Olho humano:
  - Boa resposta para coeficientes DC.
  - Baixa resposta para coeficientes AC.
- Busca reduzir a quantidade de dados.
  - Limite da amplitude para frequências: divide os valores da matriz transformada pelos valores correspondentes em uma tabela pré-definida.
  - Isso diminui os valores dos coeficientes proporcionalmente à posição dos mesmos na matriz.
  - Ocorre perda. No caso ideal, não perceptível.



## Quantização

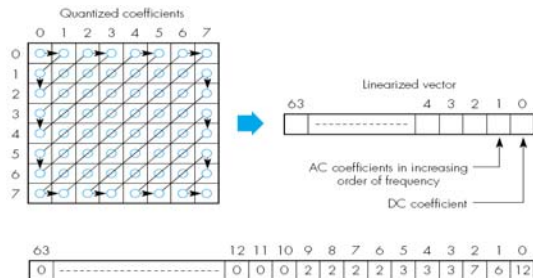
- Tabelas de quantização:
  - JPEG define duas tabelas *default*
    - Uma para luminância.
    - Uma para crominância.
  - JPEG permite a utilização de tabelas personalizadas.

## Codificação por Entropia

- Envolve quatro passos:
  - Vetorização.
  - Codificação por diferença.
  - Codificação por carreira (*run-length*).
  - Codificação Estatística (método de Huffman).

## Codificação por Entropia

- Vetorização (zig-zag scan)



## Codificação por Entropia

- Codificação por diferença
  - Codificação dos coeficientes DC.
  - DCs possuem alto grau de correlação (redundância espacial).
    - São blocos adjacentes na imagem.

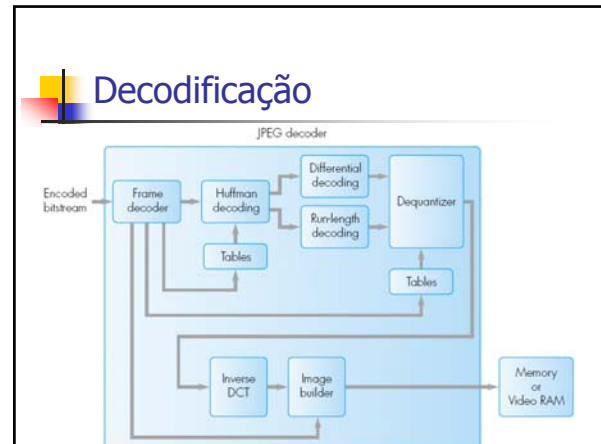
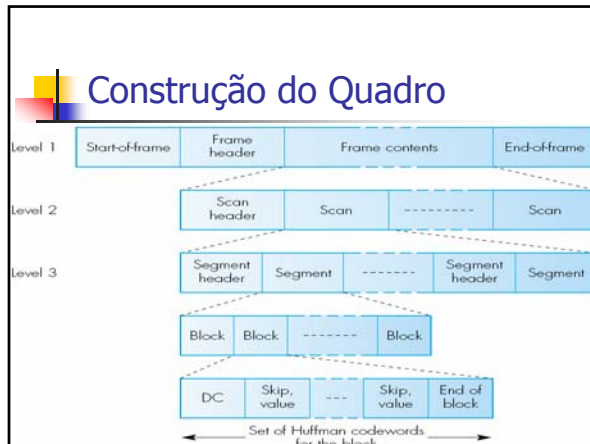
## Codificação por Entropia

- Codificação *run-length*
  - Codificação dos coeficientes AC.
  - Vetor de coeficientes possui longas cadeias de zeros.

0	1	2	3	4	5	6	7	8	9	10	11	12	63
12	6	7	3	3	3	2	2	2	2	0	0	0	...

## Codificação por Entropia

- Codificação estatística
  - JPEG usa Huffman.
  - A codificação estatística é aplicada no resultado das codificações dos DCs e ACs.
  - Vetor possui cadeias de bits – apropriado para codificação estatística.
- Quantas tabelas usar?



- ### Considerações Sobre JPEG
- JPEG:
    - Padrão abrangente.
    - Alcança boas taxas de compressão para imagens de tons contínuos. (até 20:1).
    - Desempenho diminui em imagens com muita transição de cores.
    - Baseado em particularidades do sistema visual humano.

- ### JPEG2000
- Uma das maiores limitações do JPEG:
    - Desempenho degrada em baixas taxas de dados (*bitrate*).
  - Emprega transformada Wavelet.
  - Melhorias:
    - Codificação estática/dinâmica de alta qualidade de uma região específica.
    - Recuperação de erros.
    - Desempenho: até 50:1 (níveis de cinza), 100:1 (cor).

## 3. Vídeo Digital

- ### 3.1 - Princípios
- Codificação de vídeo.
    - Processo de compressão e descompressão de sinais digitais de vídeo.

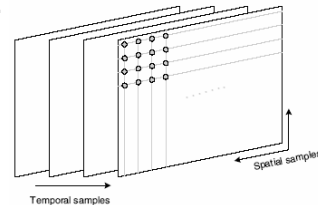
### 3.1 - Princípios



### 3.1 - Princípios

#### ■ Captura

- Envolve amostragem espacial (uma área retangular da cena) e temporal (uma série de quadros).



### 3.1 - Princípios

#### ■ Captura

- Cada amostra espaço-temporal é representada como um conjunto de números que descreve o brilho e a cor da amostra.
- Responsável por obter a amostra: CCD.
  - Sensor fotossensível de câmeras.

### 3.1 - Princípios

#### ■ Amostragem espacial

- Saída do CCD = sinal analógico de vídeo.
- Amostragem = obter valores do sinal em um ponto no tempo.
- Formato mais comum de amostragem = grid.

### 3.1 - Princípios

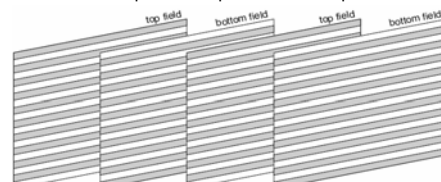
#### ■ Amostragem temporal

- Vídeo é capturado tomando amostras retangulares do sinal em intervalos regulares.
- O 'play back' da série de amostras produz a sensação de movimento.
- Quanto maior a taxa de amostragem, mais suave o movimento parece. Contudo, mais amostras são capturadas e armazenadas.
- Taxas:
  - < 10 fps – very low bit rate. Movimentos não naturais.
  - Entre 10 e 20 – não 'capta' corretamente movimentos rápidos.
  - Entre 25 e 30 – padrão de TV.
  - Entre 50 e 60 – qualidade muito boa. Muitas amostras.

### 3.1 - Princípios

#### ■ Quadros e campos

- Amostragem progressiva – produz quadros completos.
- Amostragem entrelaçada – produz uma série de campos entrelaçados.
  - Dois campos: linhas pares e linhas ímpares.



### 3.1 - Princípios

- Quadros e campos
  - Dois campos = 1 quadro. Cada campo contém metade da informação do quadro.
  - Vantagem: é possível enviar o dobro de campos por segundo que quadros por segundo, com a mesma taxa de dados, produzindo movimentos suaves.
  - Desvantagens: ruído.

### 3.1 - Princípios

- Quadros e campos

■ Top field

Bottom field



### 3.1 - Princípios

- Espaços de cores
  - Espaço de cor refere-se ao método escolhido para representar luminância e cor em cada amostra espacial de vídeo.
  - Os mais comuns para vídeo colorido:
    - RGB
    - YCbCr

### 3.1 - Princípios

- RGB
  - Necessita de três valores para indicar a proporção relativa das cores primárias.
  - Bom para captura e exibição de imagens.



### 3.1 - Princípios

- Subamostragem de crominância
  - Motivação.
  - Como fazer?

### 3.1 - Princípios

- YCbCr
  - YCbCr: MPEG e JPEG; YIQ: NTSC; YUV: PAL.
  - É um modo mais eficiente de se representar cor.
  - Baseado no HVS (Human Visual System).
  - Luminância (Y) e Crominância (Cb, Cg e Cr).
  - $Y = k_r R + k_g G + k_b B$



Cr, Cg and Cb components

## 3.1 - Princípios

### ■ YCbCr

- $Y = k_r R + k_g G + k_b B$
- $Cr = R - Y$
- $Cg = G - Y$
- $Cb = B - Y$

- YCbCr tem 4 componentes e RGB tem 3!
  - Eficiência?

## 3.1 - Princípios

### ■ YCbCr

- $Cr + Cg + Cb$  é uma constante!
  - $Cg$  não é armazenado.
- $Kr + Kg + Kb = 1$ .
- $G$  pode ser extraído de YCbCr.
  - $G$  não é representado. Menos informação.
- $Cr$  e  $Cb$  são representados com resolução menor que  $Y$  (HVS).

## 3.1 - Princípios

### ■ YCbCr

- Normalmente converte-se de RGB p/ YCbCr antes de armazenar dados de vídeo.
- Fórmulas padronizadas:
  - $Y = 0.299R + 0.587G + 0.114B$
  - $Cb = 0.564(B - Y)$
  - $Cr = 0.713(R - Y)$
  - $R = Y + 1.402Cr$
  - $G = Y - 0.344Cb - 0.714Cr$
  - $B = Y + 1.772Cb$

## 3.1 - Princípios

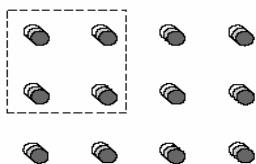
### ■ Formatos de amostragem YCbCr

- 4:4:4
- 4:2:2
- 4:2:0

## 3.1 - Princípios

### ■ Formatos de amostragem YCbCr

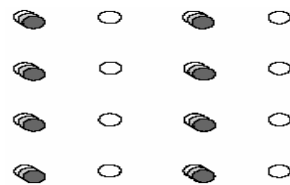
- 4:4:4
  - Um componente  $Y$ , um  $Cr$  e um  $Cb$  para cada pixel.



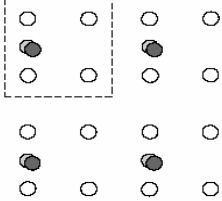
## 3.1 - Princípios

### ■ Formatos de amostragem YCbCr


- 4:2:2 (YUY2)
  - Para cada 4  $Y$  (na horizontal) existem 2  $Cr$  e 2  $Cb$ .



### 3.1 - Princípios

- Formatos de amostragem YCbCr
  - 4:2:0 (YV12)
    - Para 4 Y, 1 Cr e 1 Cb.
 

### 3.2 - Codificação

- Codificação
 


### 3.2 - Codificação

- Modelo Espacial
  - Codificação preditiva
  - Codificação por transformada
  - Quantização
  - Codificação por entropia
    - Estatística
    - Diferencial

### 3.2 - Codificação

- Codificação
  - Modelo Temporal
    - Predição temporal (do quadro anterior e/ou posterior)
      - Estimativa de movimento
      - Compensação de movimento

### 3.2 - Codificação

- Codificação
  - Modelo Temporal

### 3.3 – Tipos de Quadros

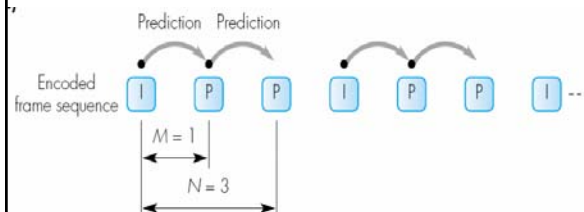
- Dois tipos básicos de quadros:
  - Quadros codificados independentemente e quadros *predicted*.
    - *Intracoded frames* ou *I-frames* ou quadros I.
    - *Predicted frames*:
      - *Predictive* ou *P-frames* ou quadros P.
      - *Bidirectional* ou *B-frames* ou quadros B.



### 3.3 – Tipos de Quadros

- Quadros I.
  - São codificados sem nenhuma referência a outros quadros.
  - Cada quadro é tratado como uma imagem independente sendo Y, Cb e Cr codificados usando o algoritmo JPEG.
  - Aparecem no *stream* de saída em intervalos regulares.
    - $N = \text{GOP (group of pictures) span}$ : número de quadros (3 a 12) entre dois quadros I sucessivos.

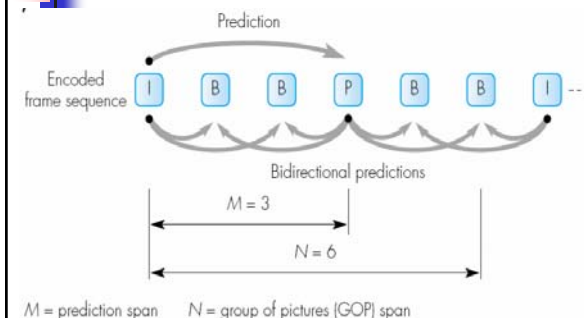
### 3.3 – Tipos de Quadros



### 3.3 – Tipos de Quadros

- Quadros P
  - São codificados em relação ao conteúdo de um quadro I ou de um quadro P anterior.
  - Usam combinação de estimativa e compensação de movimento - alcançam maiores taxas de compressão do que quadros I.
  - Propagam erros - número de quadros P entre quadros I é limitado.
  - $M = \text{prediction span}$  - número de quadros entre um quadro P e o quadro I ou P imediatamente anterior.
  - Desempenho: taxa de compressão entre 20:1 e 30:1.

### 3.3 – Tipos de Quadros



### 3.3 – Tipos de Quadros

- Quadros B
  - São codificados em relação ao conteúdo de um quadro I ou de um quadro P anterior e de um posterior.
  - Envolve o processamento de 3 quadros: o quadro I ou P anterior, o quadro atual e o quadro I ou P posterior. Todos não codificados.
  - Aumento no tempo (*delay*) para codificação e decodificação. É o tempo de esperar o próximo quadro I ou P.
  - Provêem alta taxa de compressão: entre 30:1 e 50:1.
  - Não propagam erros.

### 3.3 – Tipos de Quadros

#### ■ Quadros B

- Decodificação:
  - **I**B**B**PBBPBB**I**...
  - **I**PBBPBB**I**BB...

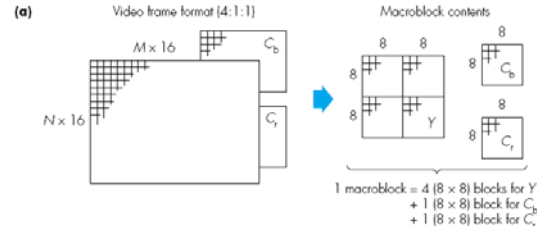
### 3.4 Estimativa e Compensação de Movimento



- Modelo: a partir de um quadro de referência.
  - Crítério de matching.
  - Estimativa de movimento.
- Residual = Modelo – Quadro atual.
  - Compensação de movimento.

### 3.4 Estimativa e Compensação de Movimento

- Block matching.** A imagem é dividida em macroblocos.
  - Y, Cb e Cr = matrizes de 16x16, 8x8 e 8x8 pixels.



### 3.4 Estimativa e Compensação de Movimento

#### Block matching

<table><tr><td>1</td><td>3</td><td>2</td></tr><tr><td>6</td><td>4</td><td>3</td></tr><tr><td>5</td><td>4</td><td>3</td></tr></table>	1	3	2	6	4	3	5	4	3	<table><tr><td>1</td><td>3</td><td>2</td><td>4</td><td>5</td></tr><tr><td>6</td><td>4</td><td>2</td><td>3</td><td>2</td></tr><tr><td>5</td><td>4</td><td>2</td><td>2</td><td>3</td></tr><tr><td>4</td><td>4</td><td>3</td><td>3</td><td>1</td></tr><tr><td>4</td><td>6</td><td>7</td><td>4</td><td>5</td></tr></table>	1	3	2	4	5	6	4	2	3	2	5	4	2	2	3	4	4	3	3	1	4	6	7	4	5	<table><tr><td>(-1,1)</td><td>(0,1)</td><td>(1,1)</td></tr><tr><td>(-1,0)</td><td>(0,0)</td><td>(1,0)</td></tr><tr><td>(-1,-1)</td><td>(0,-1)</td><td>(1,-1)</td></tr></table>	(-1,1)	(0,1)	(1,1)	(-1,0)	(0,0)	(1,0)	(-1,-1)	(0,-1)	(1,-1)
1	3	2																																											
6	4	3																																											
5	4	3																																											
1	3	2	4	5																																									
6	4	2	3	2																																									
5	4	2	2	3																																									
4	4	3	3	1																																									
4	6	7	4	5																																									
(-1,1)	(0,1)	(1,1)																																											
(-1,0)	(0,0)	(1,0)																																											
(-1,-1)	(0,-1)	(1,-1)																																											
Current block	Reference area	Positions (x,v)																																											

- Erro quadrático médio – *Mean Squared Error* (MSE)
  - Posição (0,0)

$$\{(1-4)^2 + (3-2)^2 + (2-3)^2 + (6-4)^2 + (4-2)^2 + (3-2)^2 + (5-4)^2 + (4-3)^2 + (3-3)^2\} / 9 = 2.44$$

### 3.4 Estimativa e Compensação de Movimento

Table 6.1 MSE values for block matching example

Position (x, y)	(-1, -1)	(0, -1)	(1, -1)	(-1, 0)	(0, 0)	(1, 0)	(-1, 1)	(0, 1)	(1, 1)
MSE	4.67	2.89	2.78	3.22	2.44	3.33	0.22	2.56	5.33

- The "best" match : posição (-1, 1) – Melhor predição ("modelo") é a região apontada por (-1, 1).
  - Vetor de movimento: (-1, 1)

### 3.4 Estimativa e Compensação de Movimento

- Calculate the energy of the difference between the current block and a set of neighbouring regions in the reference frame.
  - Select the region that gives the lowest error (the 'matching region').
  - Subtract the matching region from the current block to produce a difference block.
  - Encode and transmit the difference block.
  - Encode and transmit a 'motion vector' that indicates the position of the matching region, relative to the current block position (in the above example, the motion vector is (-1, 1)).
- Steps 1 and 2 above correspond to *motion estimation* and step 3 to *motion compensation*. The video decoder reconstructs the block as follows:
- Decode the difference block and motion vector.
  - Add the difference block to the matching region in the reference frame (i.e. the region 'pointed' to by the motion vector).

### 3.4 Estimativa e Compensação de Movimento

- Crítérios de comparação:

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (C_{ij} - R_{ij})^2$$

- Mean squared error

$$MAE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}|$$

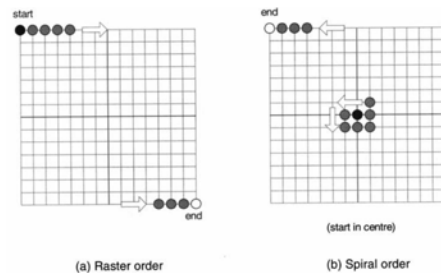
- Mean absolute error

$$SAE = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}|$$

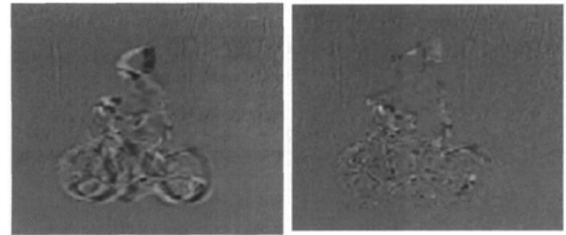
- Sum of absolute errors/differences

### 3.4 Estimativa e Compensação de Movimento

#### Full Search:



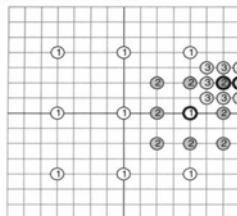
### 3.4 Estimativa e Compensação de Movimento



### Busca Rápida

#### Three-Step Search (TSS)

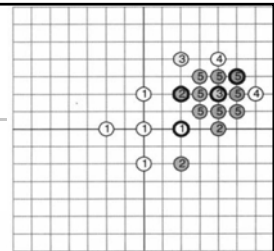
1. Search location (0, 0).
2. Set  $S = 2^{N-1}$  (the step size).
3. Search eight locations  $\pm S$  pixels around location (0, 0).
4. From the nine locations searched so far, pick the location with the smallest SAE and make this the new search origin.
5. Set  $S = S/2$ .
6. Repeat stages 3–5 until  $S = 1$ .



### Busca Rápida

#### Busca Logarítmica

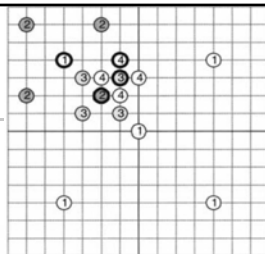
1. Search location (0, 0).
2. Search four locations in the horizontal and vertical directions,  $S$  pixels away from the origin (where  $S$  is the initial step size). The five locations make a '+' shape.
3. Set the new origin to the best match (of the five locations tested). If the best match is at the centre of the '+',  $S = S/2$ , otherwise  $S$  is unchanged.
4. If  $S = 1$  then go to stage 5, otherwise go to stage 2.
5. Search eight locations immediately surrounding the best match. The search result is the best match of the search origin and these eight neighbouring locations.



### Busca Rápida

#### Cross-Search

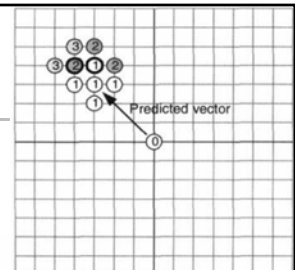
1. Search location (0, 0).
2. Search four locations at  $\pm S$ , forming an 'X' shape (where  $S = 2^{N-1}$  as for the TSS).
3. Set the new origin to be the best match of the five locations tested.
4. If  $S > 1$  then  $S = S/2$  and go to stage 2; otherwise go to stage 5.
5. If the best match is at the top left or bottom right of the 'X', evaluate four more points in an 'X' at a distance of  $\pm 1$ ; otherwise (best match is at the top right or bottom left) evaluate four more points in a '+' at a distance of  $\pm 1$ .



### Busca Rápida

#### Nearest Neighbours Search

1. Search the (0, 0) location.
2. Set the search origin to the predicted vector location and search this position.
3. Search the four neighbouring positions to the origin in a '+' shape.
4. If the search origin (or location 0, 0 for the first iteration) gives the best match, this is the chosen search result; otherwise, set the new origin to the position of the best match and go to stage 3.



## Busca Rápida

- Comparação entre os métodos

**Table 6.3** Motion estimation algorithm comparison, five frames: search window =  $\pm 15$

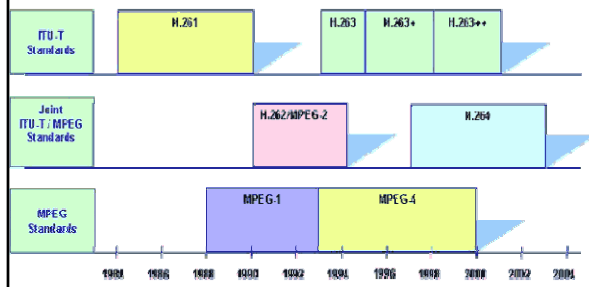
Algorithm	Total SAE (uncompensated)	Total SAE (compensated)	Number of comparison operations
Full search	1 326 783	897 163	$99.1 \times 10^6$
Three-step search	...	914 753	$3.6 \times 10^6$

## 3.4 Estimativa e Compensação de Movimento

- Estimativa sub-píxel.
- Escolha dos quadros de referência.

## 3.5 - Padrões para Compressão de Vídeo

- Evolução dos padrões:



## 3.6 - Outras Mídias de Representação

- Mais comuns:
  - AVI (microsoft)
  - MOV (apple)
- Servem como “pacotes” de dados.
- Usam codificadores diversos para comprimir vídeo.
  - Indeo, Cinepack, DivX, ADPCM

## 3.7 – Qualidade de Vídeo

- Testes Subjetivos
  - ITU-T BT.500-10
    - Protocolos
    - Testadores diversos x sequências diversas
    - Escala de notas

## 3.7 – Qualidade de Vídeo

- Testes Objetivos
  - Ainda sem padrão.
  - Aproximação amplamente utilizada: PSNR
    - Peak Signal-to-Noise Ratio


$$\text{PSNR}_{\text{dB}} = 10 \log_{10} \frac{(2^n - 1)^2}{\text{MSE}}$$

## 3.7 – Qualidade de Vídeo




■ Imagem Original.

## 3.7 – Qualidade de Vídeo




■ PSNR: 33.2 dB

## 3.7 – Qualidade de Vídeo



■ PSNR: 26.5 dB

## 3.7 – Qualidade de Vídeo



■ PSNR: 37.5 dB

## Para Saber Mais

- **Áudio digital:**
  - Mandal, M. K.; Multimedia Signals and Systems. Kluwer Academic Publishers, 2003. Capítulo 2, seção 2.1.
  - Halsall, F.; Multimedia Communications: Applications, Networks, Protocols, and Standards, Addison-Wesley Publishing, 2001. ISBN: 0201398184. Capítulo 2, seções 2.1, 2.2 e 2.5; Capítulo 4, seção 4.2.
- **Imagens:**
  - Gonzales & Woods. Digital Image Processing. 3rd ed. Prentice-Hall, 2008. Capítulo 8, seção 8.1.
  - Halsall, F. Multimedia Communications: Applications, Networks, Protocols, and Standards, Addison-Wesley Publishing, 2001. ISBN: 0201398184. Capítulo 2, seção 2.4 e capítulo 3, seções 3.2 e 3.4.
  - Pennebaker & Mitchell. JPEG Still Image Data Compression Standard. Van Nostrand Reinhold, 1993.
- **Vídeo:**
  - Richardson, L. E. G. Video Codec Design, Wiley, 2002.
  - Halsall, F. Multimedia Communications: Applications, Networks, Protocols, and Standards, Addison-Wesley Publishing, 2001. ISBN: 0201398184. Capítulo 3, seção 4.3.
  - H.261 e H.263: [http://www.compression-links.info/H.261\\_H.263](http://www.compression-links.info/H.261_H.263)
  - Padrões MPEG: <http://www.chiariglione.org/mpeg/>