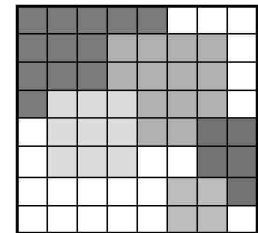
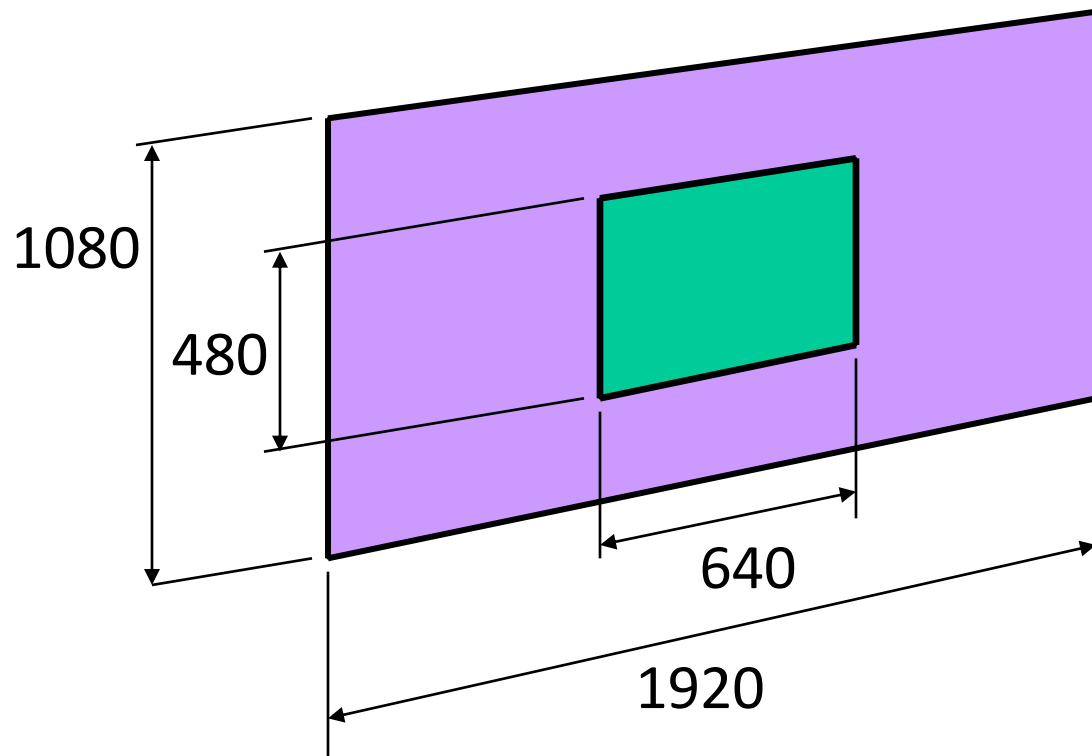


Compressão de Imagens: Padrão JPEG

PTC3547 – Codificação e Transmissão Multimídia
Guido Stolfi – 04/2018



- Justificativas para Compressão de Imagens
- Codificador JPEG
- Transformada Discreta de Cossenos
- Quantização
- Compactação de Dados
- Artefatos da compressão JPEG
- Outros processos de compressão de imagens
 - KLT e Wavelets



- Luminância: 640×480 pontos $\times$ 30 quadros por segundo $\times$ 8 bits
= **73,7 Mb/s**
- Crominância: 320×240 pontos $\times$ 2 componentes $\times$ 30 quadros por segundo $\times$ 8 bits = **36,8 Mb/s**
- Áudio “qualidade CD”: $44,1$ kHz $\times$ 16 bits $\times$ 2 canais = **1,41 Mb/s**
- Total: **112 Mb/s**

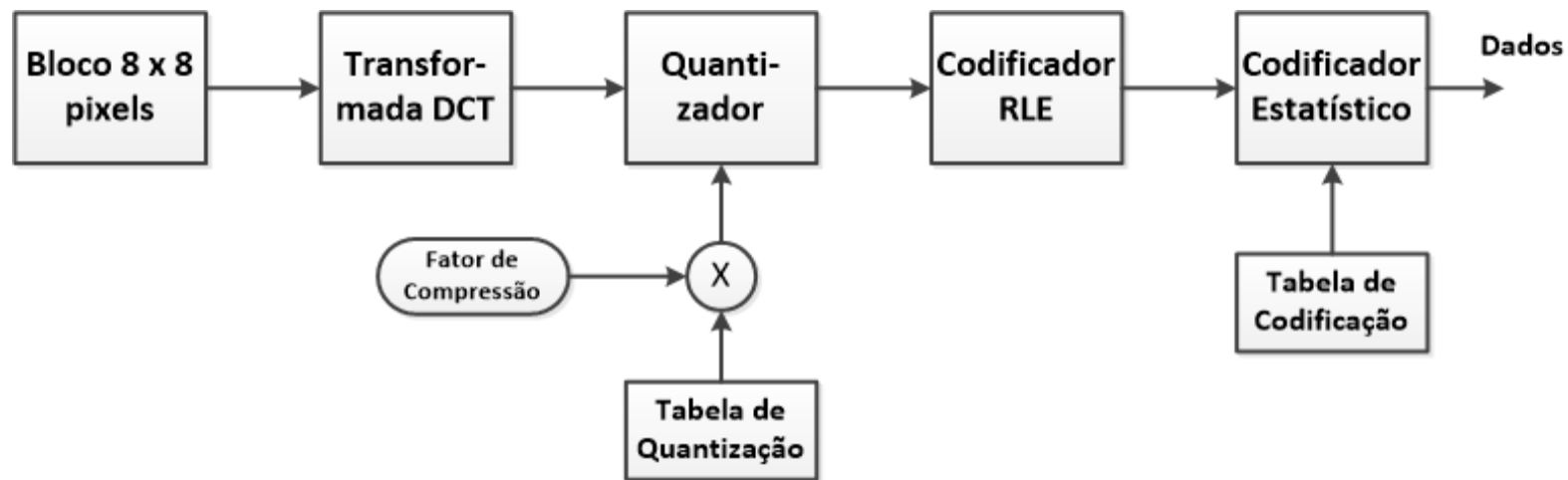
- Luminância: 1920 x 1080 pontos x 30 quadros por segundo x 10 bits = **622 Mb/s**
- Crominância: 960 x 540 pontos x 2 componentes x 30 quadros por segundo x 10 bits = **311 Mb/s**
- 5.1 canais de áudio, 16 bits, 48 kHz = **3,9 Mbit/s**
- Total: **934 Mb/s**

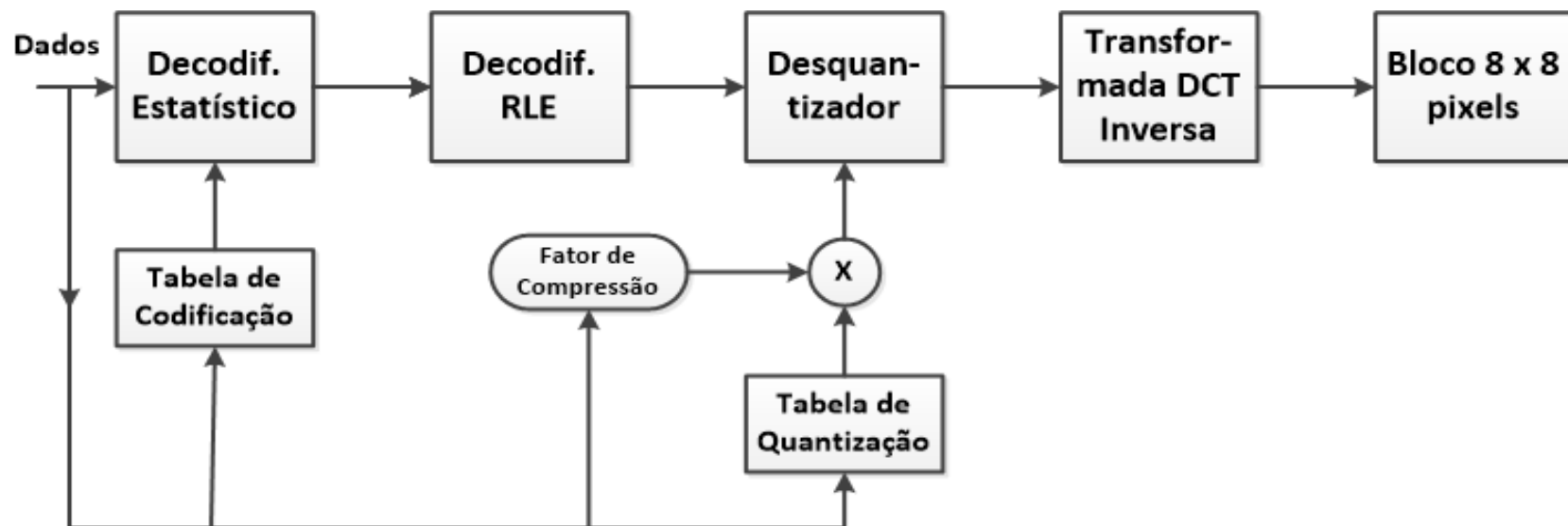
- **Cabo Coaxial** (Modulação 64-QAM)
 - (Relação Sinal/Ruído Maior que 30 dB)
 - Taxa Bruta: 43 MBits/s
 - Taxa Líquida: 38 MBits/s
- **Rádiodifusão / UHF** (Modulação 8-VSB)
 - (Relação Sinal/Ruído Maior que 15 dB)
 - Taxa Bruta: 32 MBits/s
 - Taxa Líquida: 20 MBits/s
- **Satélite DTH** (*Direct to Home*-Modulação QPSK)
 - Taxa Bruta: 50 MBits/s por transponder
 - Taxa Líquida: 40 MBits/s

(Taxa Bruta = Taxa Líquida + Códigos de Correção de Erros)

- Sinal reconstruído não é exatamente igual ao sinal transmitido;
- Objetivo: eliminar informação a ser transmitida, desde que o erro de reconstrução seja imperceptível *em condições normais de visualização*.

- JPEG = Joint Photographic Experts Group
 - Comitê CCITT / ISO – Recomendação JTC1/10918-2
- Padrão de Compressão de **Imagens Naturais Estáticas**
- Compressão **Com Perdas** (*baseline*)
- Taxa de Compressão Indeterminada **A Priori**
- Baseada em:
 - Transformada DCT
 - Quantização
 - Codificação RLE
 - Codificação Entrópica (Huffman)





Bits / pixel	Taxa de Compressão	Qualidade da imagem reconstruída
6	2 : 1	Indistinguível do original
3,6	3,3 : 1	Qualidade de estúdio / produção
1,8	6,6 : 1	Excelente para uso doméstico
0,75	16 : 1	Boa qualidade
0,25	50 : 1	Imagem usável
0,08	150 : 1	Imagem reconhecível

(Imagens a cores, 4 : 2 : 0)

Desempenho do Padrão JPEG

Original: 200 x 150 pixels (25 x 19 blocos), 90000 bytes (24 b/p)

Descontado overhead do cabeçalho JPEG: 720 bytes



6,44 b/p (3,7:1)



1,18 b/p (20:1)



0,55 b/p (44:1)



0,38 b/p (62:1)

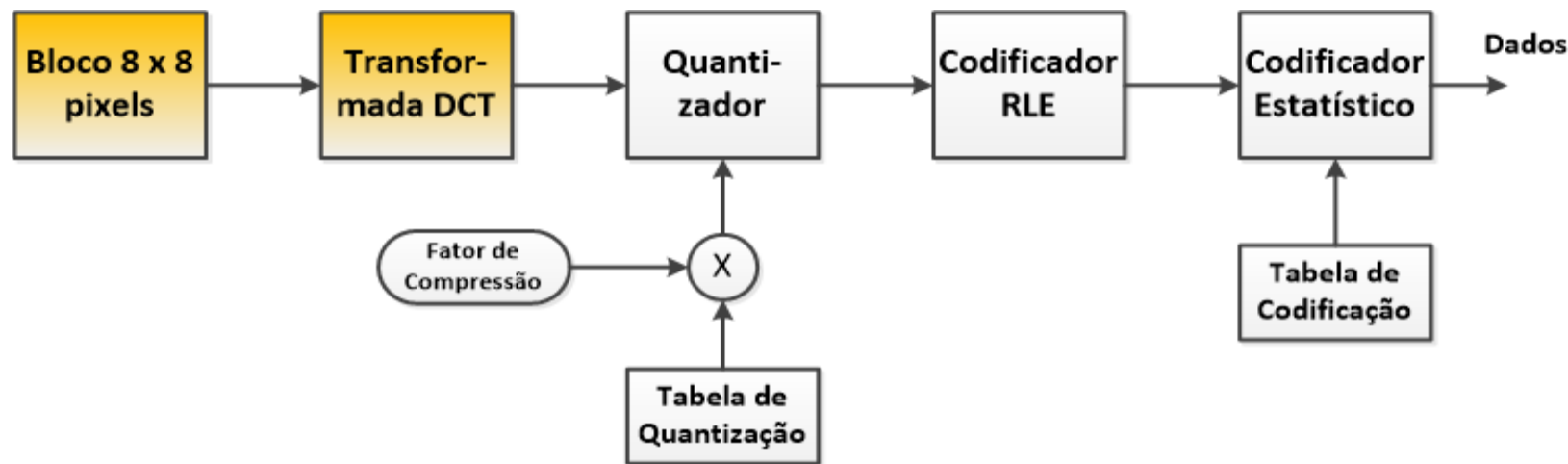


0,28 b/p (85:1)



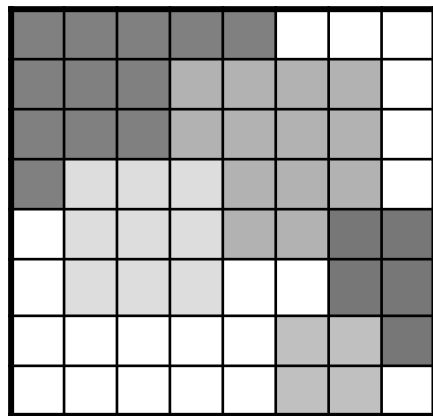
0,17 b/p (140:1)

Transformada Discreta de Cossenos (DCT)

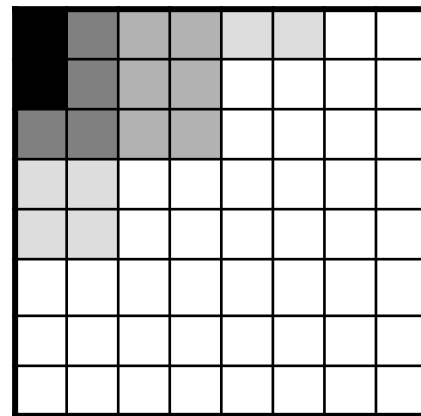


$$F(u, v) = \frac{C(u)}{2} \frac{C(v)}{2} \sum_{y=0}^7 \sum_{x=0}^7 f(x, y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right]$$

$$\begin{cases} C(k) = \frac{1}{\sqrt{2}} & \text{se } k = 0 \\ C(k) = 1 & \text{se } k > 0 \end{cases}$$



DCT

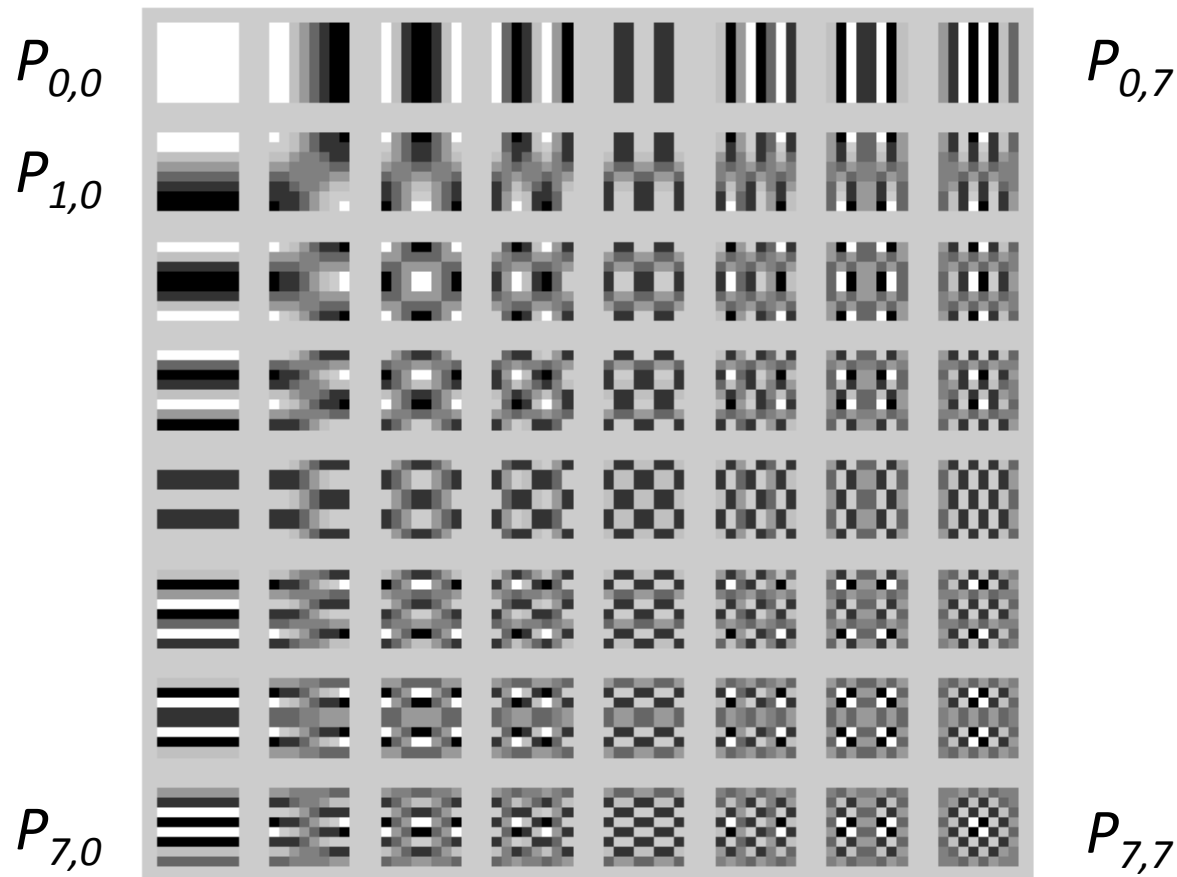


$f(x,y)$ = 64 elementos
de imagem (8x8 pixels)

$F(u,v)$ = 64 componentes
de frequências espaciais

- Cada coeficiente $F(u,v)$ expressa a semelhança (correlação) entre o bloco de 8x8 pixels e uma determinada função primitiva correspondente:

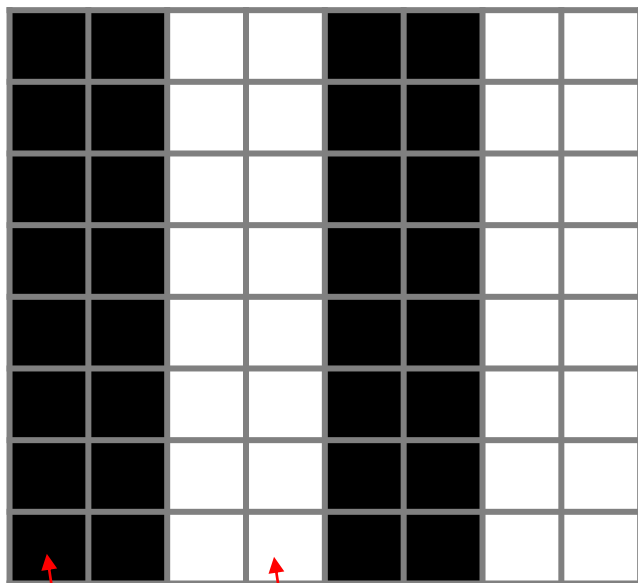
$$P_{u,v}(x, y) = \cos\left(\frac{(x+1)u\pi}{16}\right) \times \cos\left(\frac{(y+1)v\pi}{16}\right)$$



$$f(x, y) = \frac{1}{2} \left(\sum_{u=0}^7 \sum_{v=0}^7 C(u)C(v)F(u, v) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right)$$

$$DCT^{-1} = \sum_u \sum_v F(u, v) \cdot P(u, v)$$

Exemplo de Transformada DCT

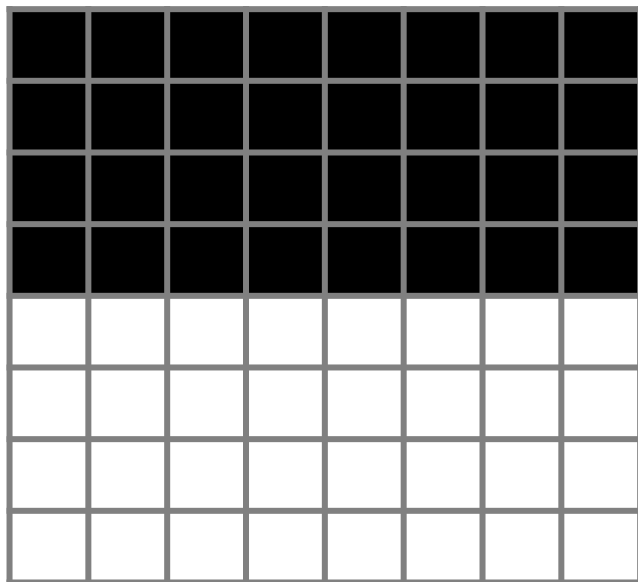


-128

+127

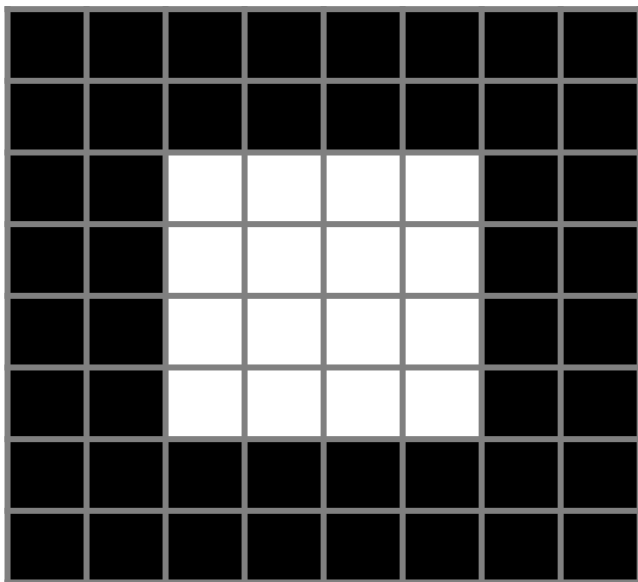
-4	-383	0	-784	0	524	0	76
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Exemplo de Transformada DCT



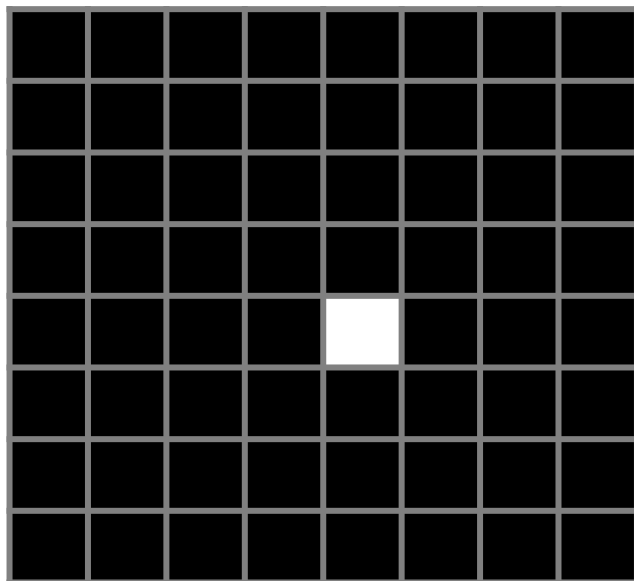
-4	0	0	0	0	0	0	0
-924	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
325	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-217	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
184	0	0	0	0	0	0	0

Exemplo de Transformada DCT



-514	0	-471	0	0	0	195	0
0	0	0	0	0	0	0	0
-471	0	435	0	0	0	-180	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
195	0	-180	0	0	0	75	0
0	0	0	0	0	0	0	0

Exemplo de Transformada DCT



-992	-9	-42	25	32	-37	-17	44
-9	2	11	-7	-9	10	5	-12
-42	11	54	-33	-42	49	23	-58
25	-7	-33	20	25	-29	-14	35
32	-9	-42	25	32	-37	-17	44
-37	10	49	-29	-37	44	20	-52
-17	5	23	-14	-17	20	9	-24
44	-12	-58	35	44	-52	-24	61

- Descorrelacionar os elementos da imagem
- Representar a imagem em termos de componentes de frequência espacial para posterior processamento, de acordo com as características da visão humana
- Simplicidade de processamento (valores reais)

$$\mathbf{F} = \mathbf{C} \times \mathbf{B} \times \mathbf{C}^T$$

onde

$$C_{i,j} = \begin{cases} \frac{1}{2\sqrt{2}} & \text{se } i = 0 \\ \frac{1}{2} \cos\left(\frac{(j+1/2)i\pi}{16}\right) & \text{se } 0 \leq i \leq 7 \end{cases}$$

$C_{8 \times 8} =$

.354	.354	.354	.354	.354	.354	.354	.354
.490	.416	..278	.098	-.098	-.278	-.416	-.490
..462	.191	-.191	-.462	-.462	-.191	.191	.462
.416	-.098	-.490	-.278	.278	.490	.098	-.416
.354	-.354	.354	-.354	.354	-.354	.354	-.354
.278	-.490	.098	.416	-.416	-.098	.490	-.278
.191	-.462	.462	-.191	-.191	.462	-.462	.191
.098	-.278	.416	-.490	.490	-.416	.278	-.098

DCT unidimensional:

$$F(u) = \frac{C(u)}{2} \sum_{x=0}^7 f(x) \cos \left[\frac{(2x+1)u\pi}{16} \right]$$

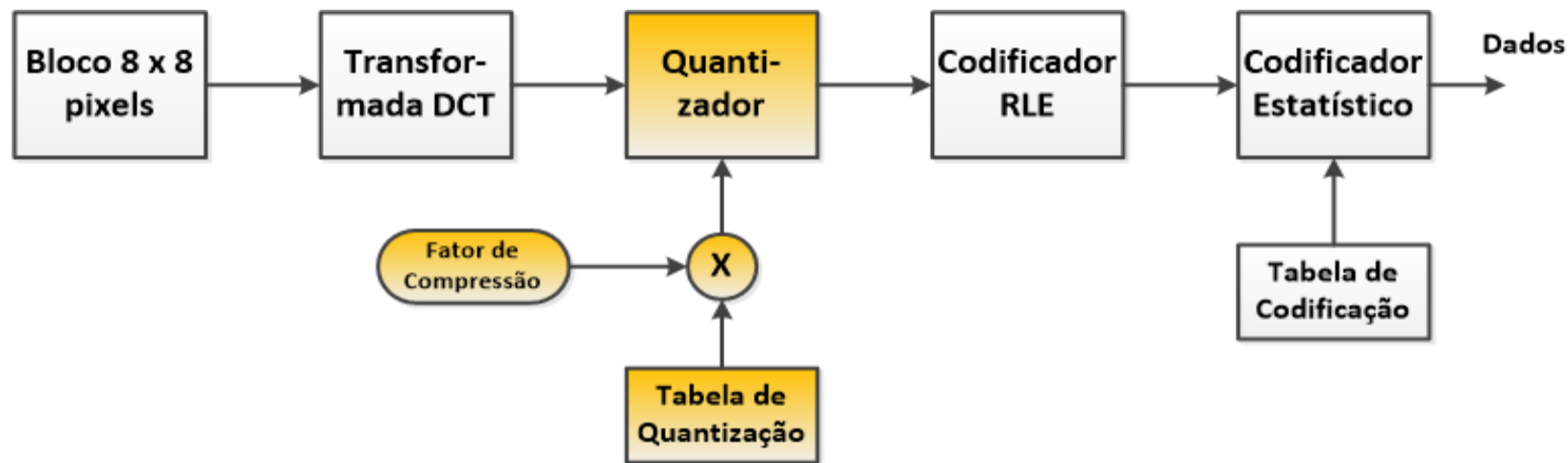
DCT "horizontal" em uma matriz 8 x 8:

$$g(u, y) = \frac{C(u)}{2} \sum_{x=0}^7 f(x, y) \cos \left[\frac{(2x+1)u\pi}{16} \right]$$

DCT "vertical" aplicada sobre os coeficientes DCT obtidos:

$$\begin{aligned} F(u, v) &= \frac{C(v)}{2} \sum_{y=0}^7 g(u, y) \cos \left[\frac{(2y+1)v\pi}{16} \right] \\ &= \frac{C(u)}{2} \frac{C(v)}{2} \sum_{y=0}^7 \sum_{x=0}^7 f(x, y) \cos \left[\frac{(2x+1)u\pi}{16} \right] \cos \left[\frac{(2y+1)v\pi}{16} \right] \end{aligned}$$
$$\begin{cases} C(k) = \frac{1}{\sqrt{2}} & \text{se } k = 0 \\ C(k) = 1 & \text{se } k > 0 \end{cases}$$

- 1024 multiplicações e 896 somas (8x8);
- **Algoritmos rápidos:** 54 multiplicações e 468 somas e deslocamentos (DCT 8x8).



$$F_Q(u, v) = \text{round} \left[\frac{F(u, v)}{k \cdot Q(u, v)} \right] \text{ onde } k = \text{fator de compressão}$$

$Q(u, v) =$

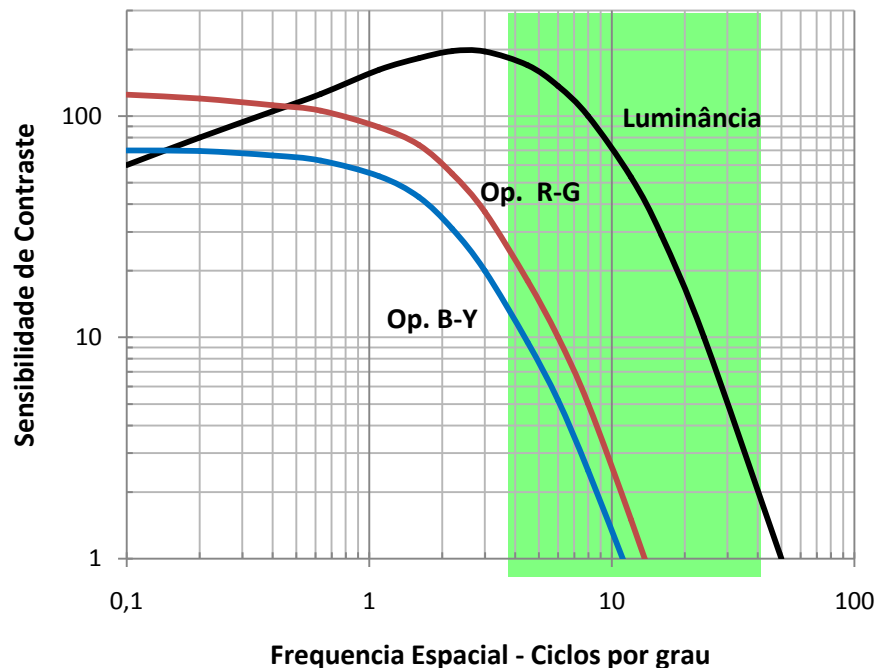
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
79	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(Ex. Luminância)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(Ex. Crominância)

MTF da
visão



- Imagem R', G', B' pode ser mapeada em componentes:

$$Y = 0.3R' + 0.6G' + 0.1B' - 128$$

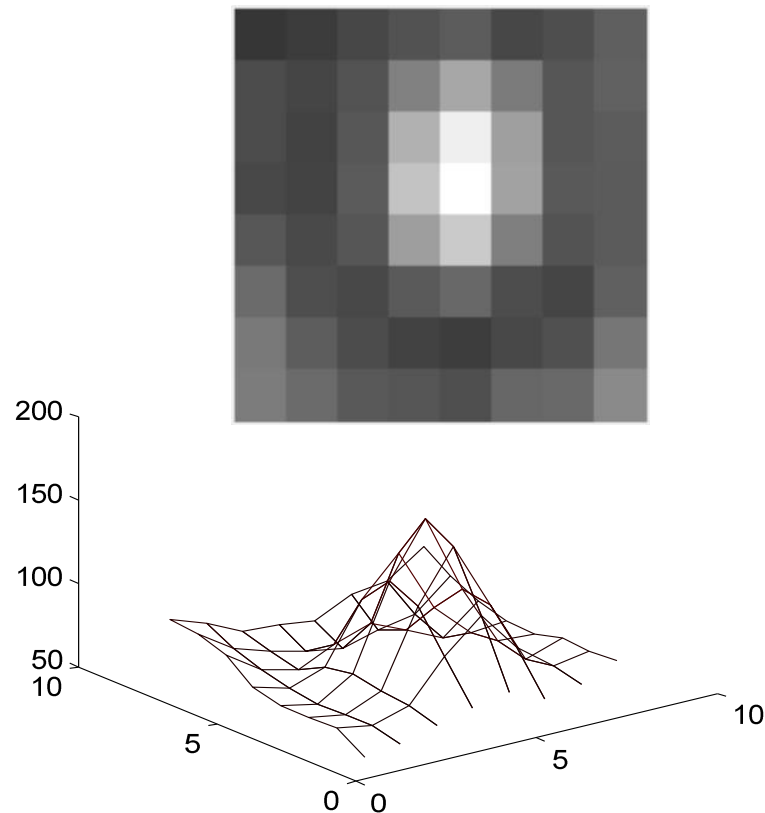
$$U' = \frac{B' - Y'}{2} + 0.5$$

$$V' = \frac{R' - Y'}{1.6} + 0.5$$

- Componentes U' e V' são sub-amostradas pela metade (h e v).

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

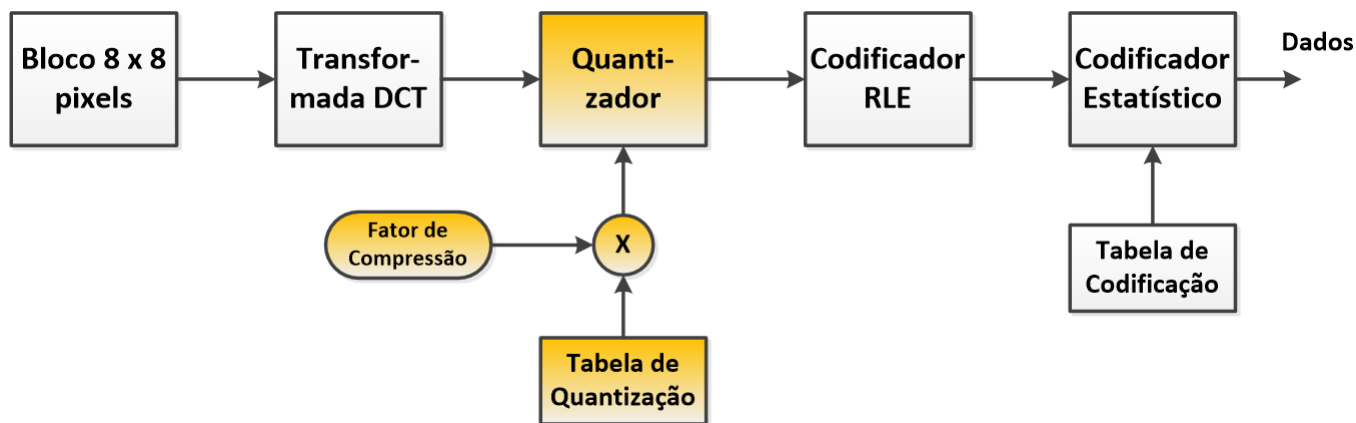
Bloco de 8 x 8 pixels



52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

DCT
→

-415	-29	-62	25	55	-20	-1	3
7	-21	-62	9	11	-7	-6	6
-46	8	77	-25	-30	10	7	-5
-50	13	35	-15	-9	6	0	3
11	-8	-13	-2	-1	1	-4	1
-10	1	3	-3	-1	0	2	-1
-4	-1	2	-1	2	-3	1	-2
-1	-1	-1	-2	-1	-1	0	-1



$$F_Q(u, v) = \text{round} \left[\frac{F(u, v)}{k \cdot Q(u, v)} \right] \text{ onde } k = \text{fator de compressão}$$

$Q(u, v) =$

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
79	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(Ex. Luminância)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(Ex. Crominância)

-415	-29	-62	25	55	-20	-1	3
7	-21	-62	9	11	-7	-6	6
-46	8	77	-25	-30	10	7	-5
-50	13	35	-15	-9	6	0	3
11	-8	-13	-2	-1	1	-4	1
-10	1	3	-3	-1	0	2	-1
-4	-1	2	-1	2	-3	1	-2
-1	-1	-1	-2	-1	-1	0	-1

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
79	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

$$\text{round} \left[\frac{F(u, v)}{2 \cdot Q(u, v)} \right] =$$

-26	-1	-3	1	1	0	0	0
0	-1	-2	0	0	0	0	0
-2	0	2	-1	0	0	0	0
-2	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

Coeficientes da DCT reconstruídos

-26	-1	-3	1	1	0	0	0
0	-1	-2	0	0	0	0	0
-2	0	2	-1	0	0	0	0
-2	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

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
79	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

×

416	-22	-60	32	48	0	0	0
0	-24	-56	0	0	0	0	0
-56	0	64	-48	0	0	0	0
-56	0	44	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

DCT⁻¹

Erro de Reconstrução

50	45	53	66	64	53	60	78
55	54	69	95	99	81	69	71
64	62	83	121	135	111	78	61
74	64	79	118	139	117	79	56
82	62	62	91	111	98	73	59
86	62	53	69	80	74	67	70
87	69	60	67	68	61	67	83
87	75	72	76	70	59	70	92

-2	-10	-8	0	-6	-8	-4	5
-8	-5	3	5	-10	-4	0	-1
2	3	15	8	-9	7	12	-12
11	6	8	-4	-15	11	9	-13
15	1	-6	-13	-15	10	5	-11
7	-3	-7	-1	3	6	9	-5
2	-2	-4	8	13	0	2	0
0	-4	3	8	5	-17	-8	-2

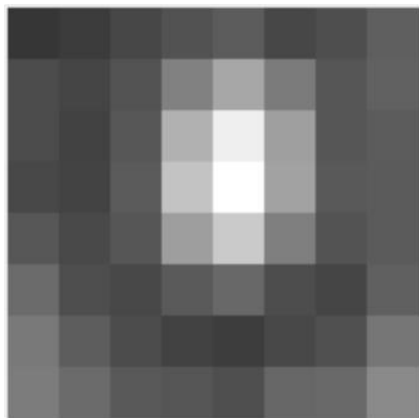
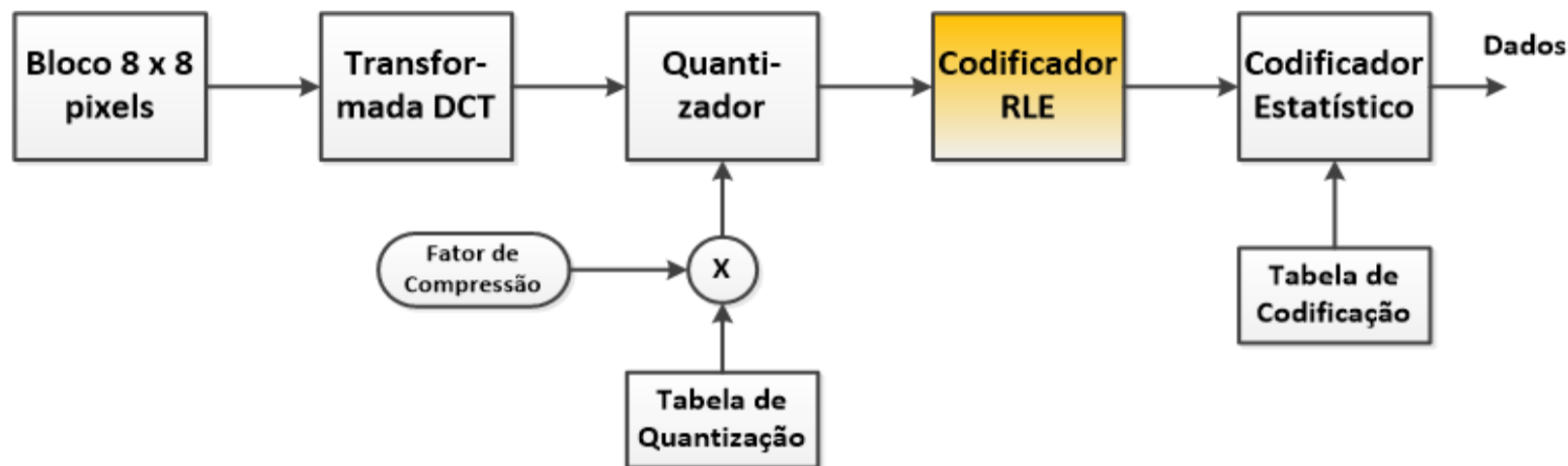


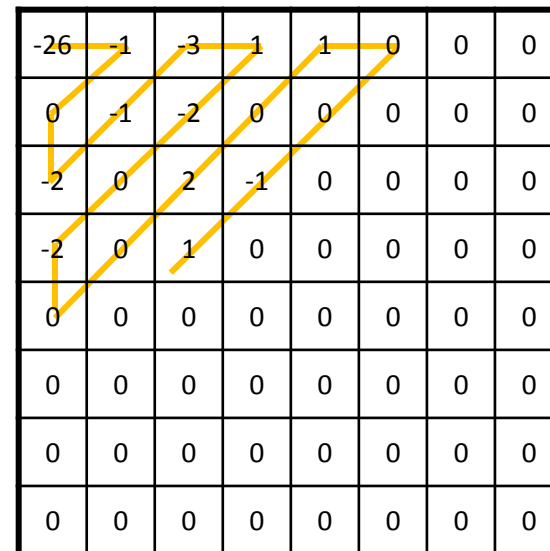
Imagem Original



Imagem Reconstruída

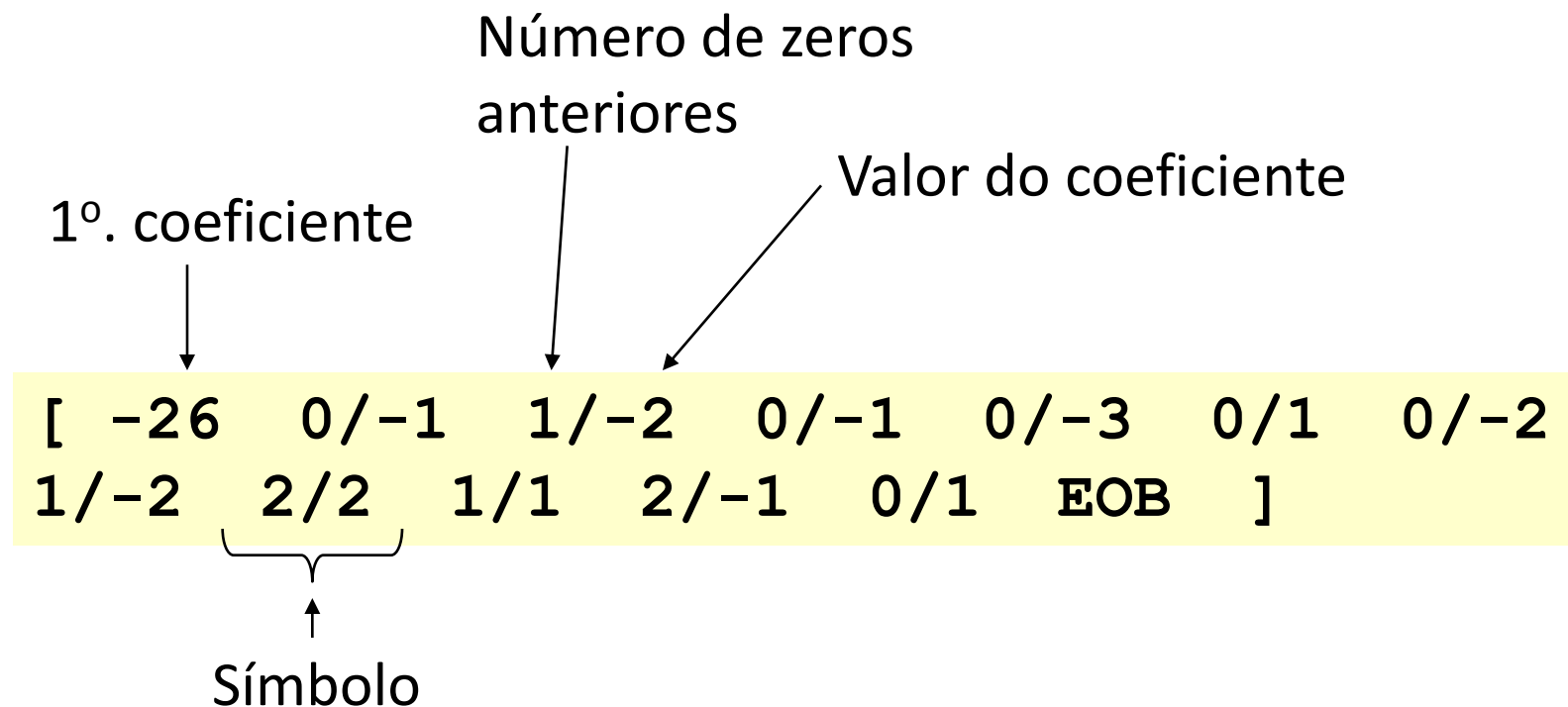
Codificação R.L.E. (Run-Length Encoding)



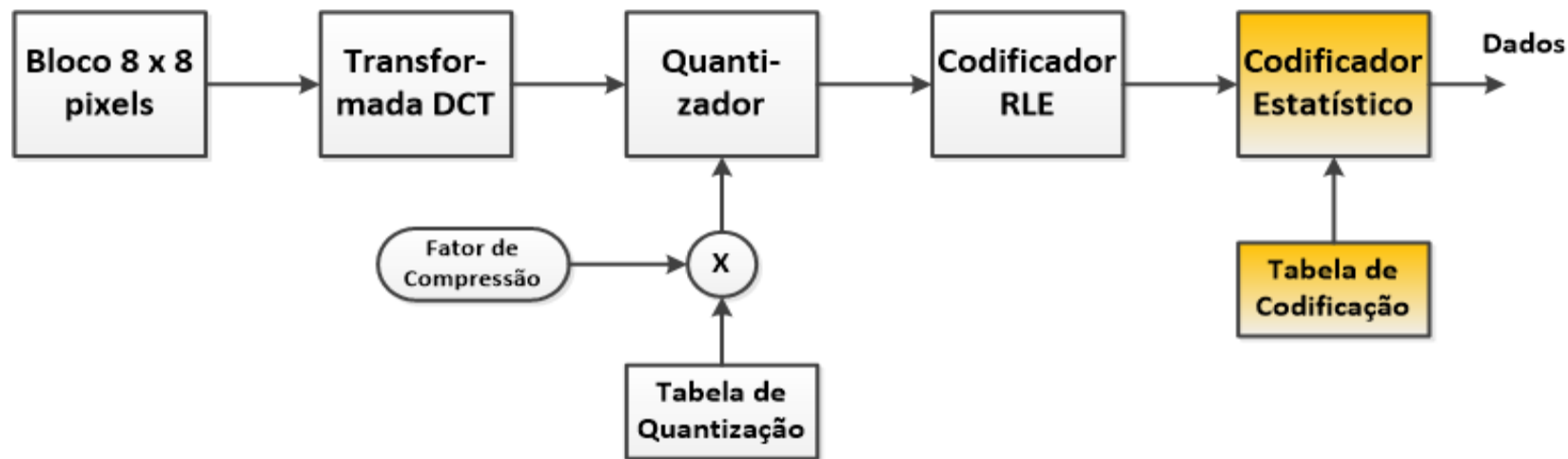


(20 símbolos)

Codificação RLE (Run-Length Encoding)



(13 símbolos)



- Símbolos de Comprimento Variável
- Comprimento do símbolo é tanto menor quanto maior a sua probabilidade de ocorrência
- Sequência de símbolos concatenados (*bit stream*) é univocamente decodificável

- Primeiro Coeficiente ($F(0,0)$) : Codificação Diferencial (com relação ao Bloco anterior)
- Valor de $F(0,0) \Rightarrow \langle \text{Prefixo} : \text{Mantissa} \rangle$
- **Prefixo** definido pela Categoria (ordem de grandeza) do valor codificado
- **Mantissa** define o valor do coeficiente e o seu sinal
- **Prefixo** e **Mantissa** têm comprimento variável.

- Demais Coeficientes (“*AC*”): Tabela de Codificação (Dicionário Huffman Modificado)
- Código = <Prefixo : Mantissa>
- Número de zeros + Categoria do Coeficiente $\Rightarrow$ <Prefixo>
- Valor do Coeficiente $\Rightarrow$ <Mantissa>
- Códigos especiais para Fim de Bloco (EOB) e sequências de mais de 15 zeros

Faixa de Valores	Categoria
0	0
-1; +1	1
-3, -2; +2, +3	2
-7 a -4; +4 a +7	3
-15 a -8; +8 a +15	4
-31 a -16; +16 a +31	5
-63 a -32; +32 a +63	6
-127 a -64; +64 a +127	7
-255 a -128; +128 a +255	8
-511 a -256; +256 a +511	9
-1023 a -512; +512 a +1023	A
-2047 a -1024; +1024 a +2047	B

Prefixos para Codificação de F(0,0)

Categoria	Prefixo	Comprimento total (bits)	Comprimento Mantissa
0	010	3	0
1	011	4	1
2	100	5	2
3	00	5	3
4	101	7	4
5	110	8	5
6	1110	10	6
7	11110	12	7
8	111110	14	8
9	1111110	16	9
A	11111110	18	10
B	111111110	20	11

- Ex.: Valor = +200
- Representável pela Categoria 8
(faixa de valores: -255 a -128;
+128 a +255)
- Prefixo = **111110**
- Mantissa = 8 bits (vide tabela) =
1100 1000
- Concatenando ambos:
+200 = **11111011001000**

Valor	Código	
-255	0000	0000
-254	0000	0001
-253	0000	0010
...		
-128	0111	1111
+128	1000	0000
+129	1000	0001
...		
+200	1100	1000
...		
+255	1111	1111

Tabela de Codificação para Coeficientes AC

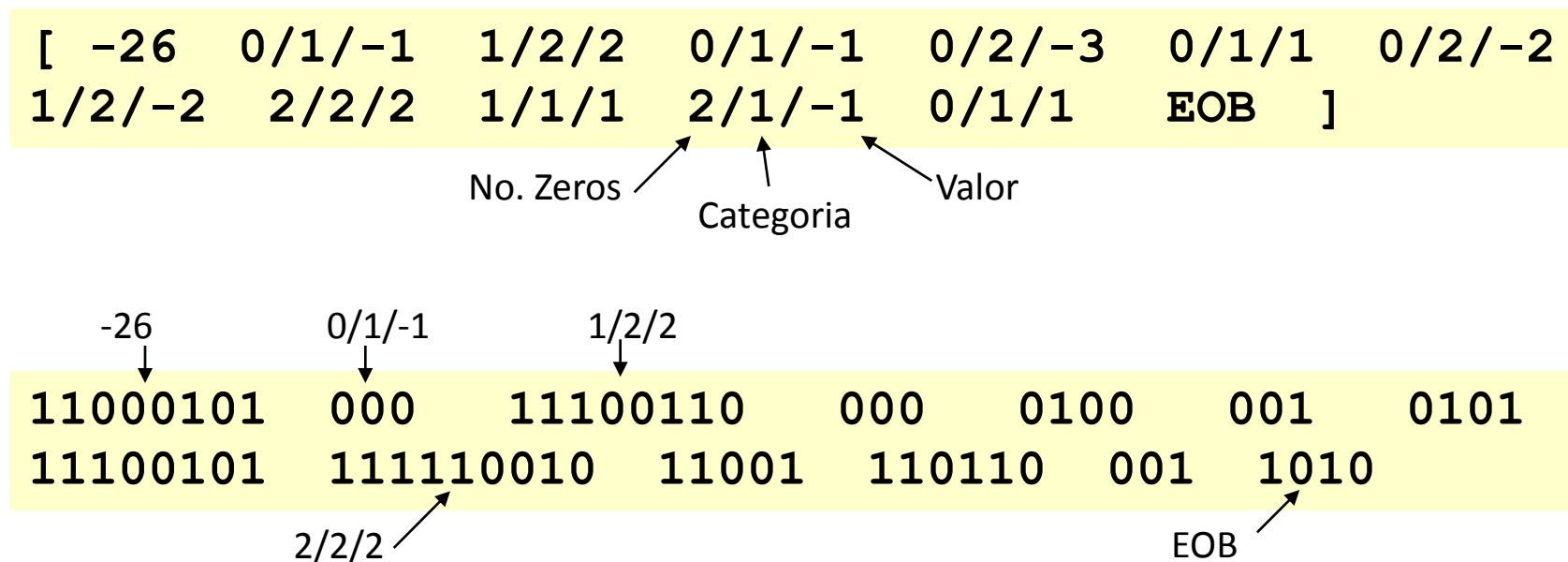
Zeros/ Categoria	Prefixo	Comprim. total
0/1	00	3
0/2	01	4
0/3	100	6
0/4	1011	8
0/5	11010	10
0/6	111000	12
0/7	1111000	14
0/8	111110110	18
0/9	111111110000010	25
0/A	111111110000011	26
1/1	1100	5
1/2	111001	8
1/3	1111001	10

Zeros/ Categoria	Prefixo	Comprim. total
3/1	111010	7
3/2	111110111	11
3/3	11111110111	14
3/4	111111110010000	20
...		
13/1	11111111010	12
13/2	1111111111100011	18
13/3	1111111111100100	19
...		
15/7	1111111111111011	23
15/8	1111111111111100	24
15/9	1111111111111101	25
15/A	1111111111111110	26

(Trechos)

- Fim de Bloco (EOB) : **1010**
- Extensão de Zeros (sequência de 16 zeros, correspondente ao par [15 / 0]) : **111111110111**

Codificação Huffman por Tabela de Códigos



Bloco original: $8 \times 8 \times 8 = 512$ bits (8 bits por pixel)

Bloco comprimido: 68 bits (1,06 bits/pixel)

Compressão de 7,5:1 !



“Lena”, © Playboy 1972

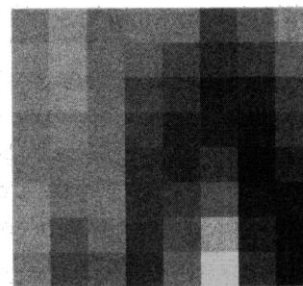
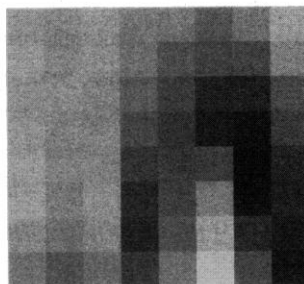
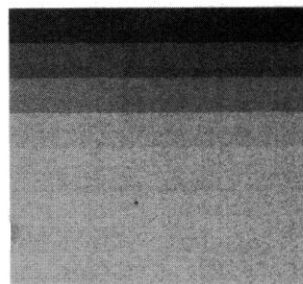
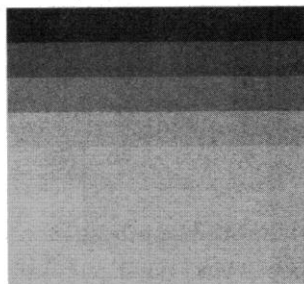




Imagem Original (57x49)



Imagem Comprimida ~20:1



- Transformada de Walsh
- Transformada de Hadamard
- Transformada de Karhunen-Loève

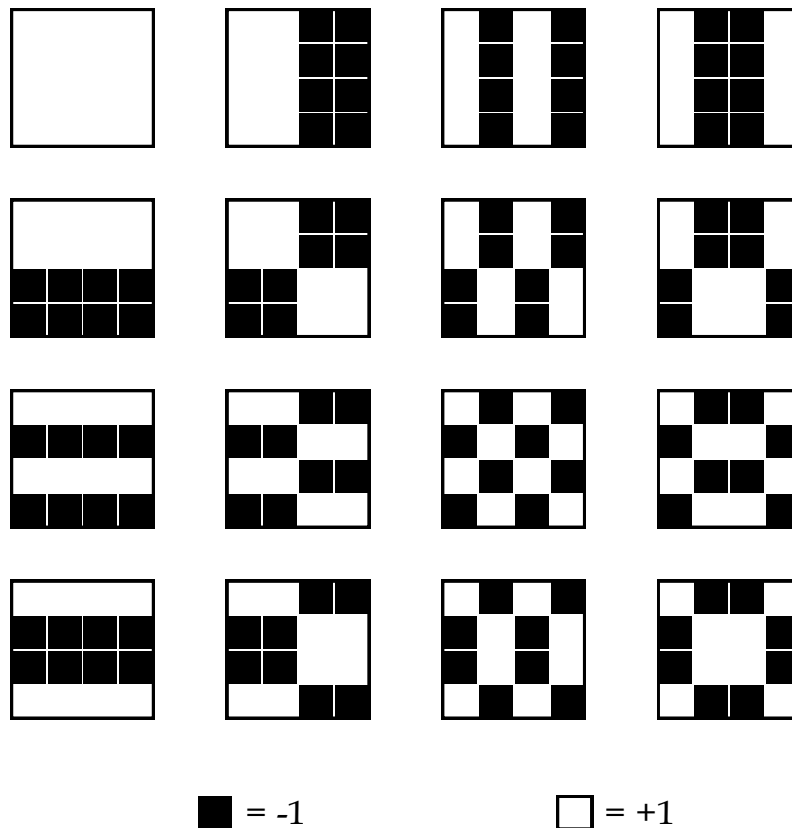
$$W(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \prod_{i=0}^{n-1} (-1)^{b_i(x)b_{n-1-i}(u) + b_i(y)b_{n-1-i}(v)}$$

$$H(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) (-1)^{\sum_{i=0}^{n-1} b_i(x)b_i(u) + b_i(y)b_i(v)}$$

onde $b_k(z)$ é o k-ésimo bit da representação binária de z .

- Funções Primitivas são binárias (+1 / -1)

Primitivas da Transformada de Walsh (4x4)



Seja uma imagem representada por um conjunto de vetores estocásticos (p. ex., blocos de n pixels de uma imagem) da forma

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

e seja $\mathbf{m}_x = E[\mathbf{x}]$ o vetor médio desse conjunto;

A *Matriz de Covariança* desse conjunto de vetores é dada por:

$$\mathbf{C}_x = E[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T] \quad (n \times n)$$

Sejam $\mathbf{e}_i$ e λ_i , $i = 1, 2, \dots, n$

os n autovetores e correspondentes autovalores da matriz de covariância,
ordenados em ordem decrescente de modo que $\lambda_k \geq \lambda_{k+1}$

e seja $\mathbf{A}$ a matriz composta pelos autovetores correspondentemente ordenados:

$$\mathbf{A} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$$

Então a expressão

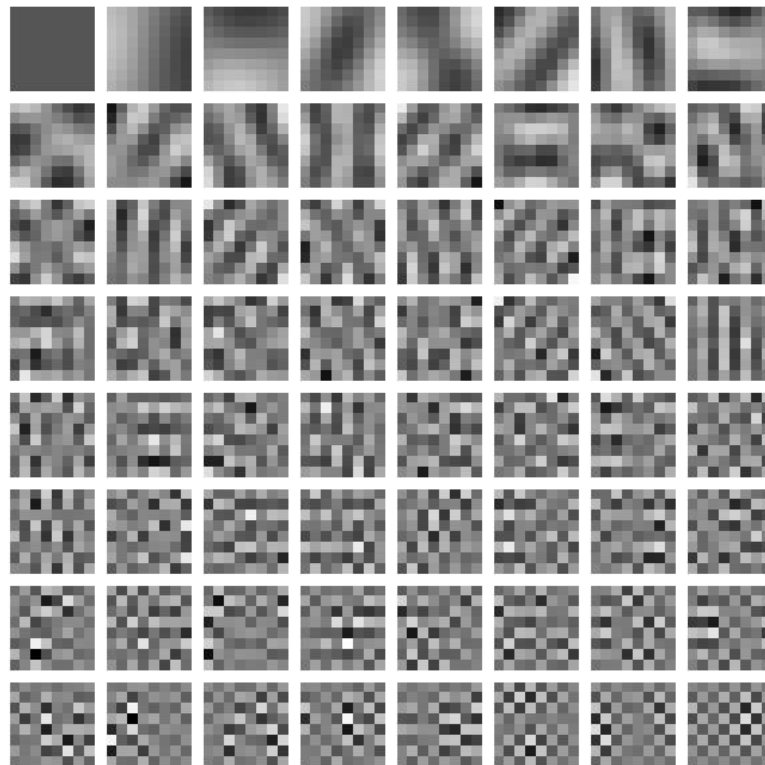
$$\mathbf{y} = \mathbf{A} \mathbf{x} - \mathbf{m}_x$$

corresponde à *Transformada de Karhünen-Loève* baseada no conjunto de vetores $\mathbf{x}$.

- Propriedades da KLT:
 - Elementos de $\mathbf{y}$ são descorrelacionados;
 - Erro quadrático médio da reconstrução é minimizado;
 - No entanto, as funções-base (matriz $\mathbf{A}$) dependem do conteúdo da imagem.



Exemplo: Primitivas para
imagem “Lena”, blocos 8 x 8

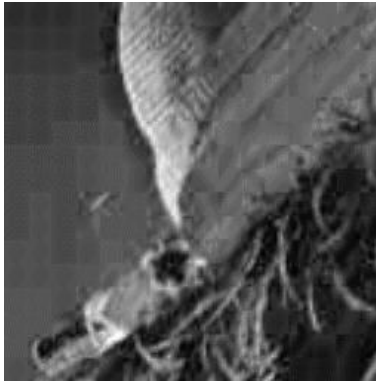




Original



10:1



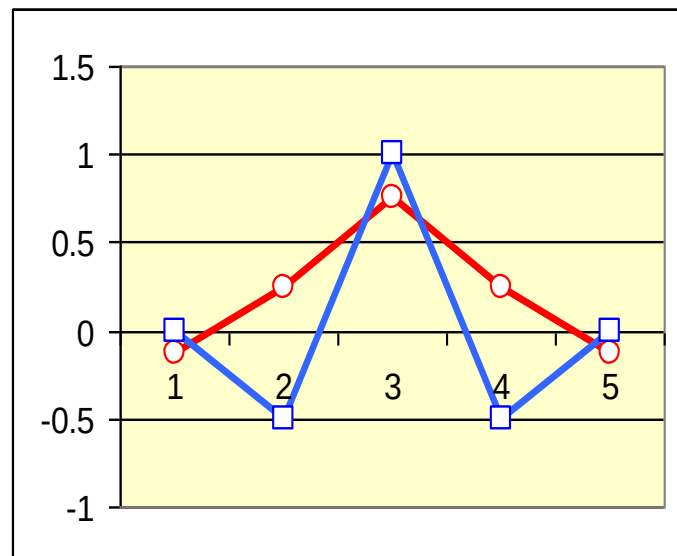
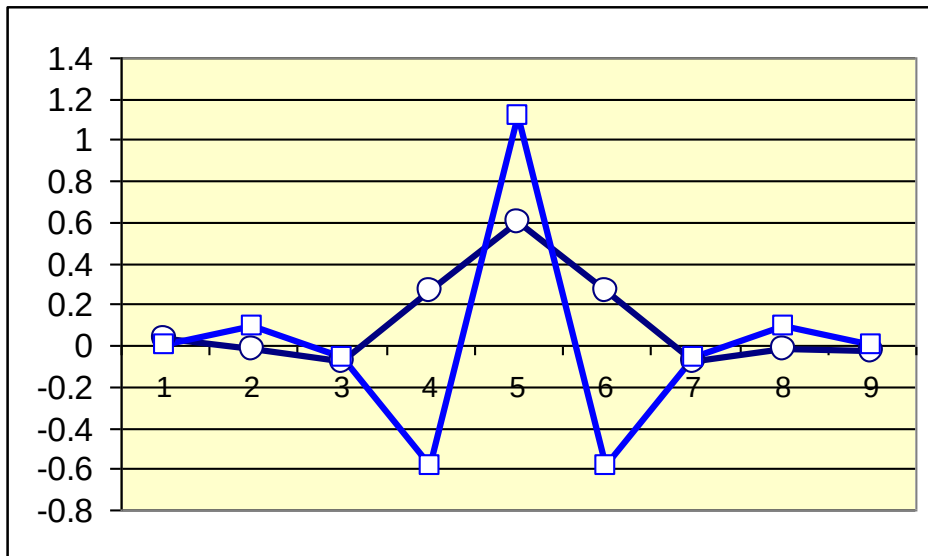
20:1



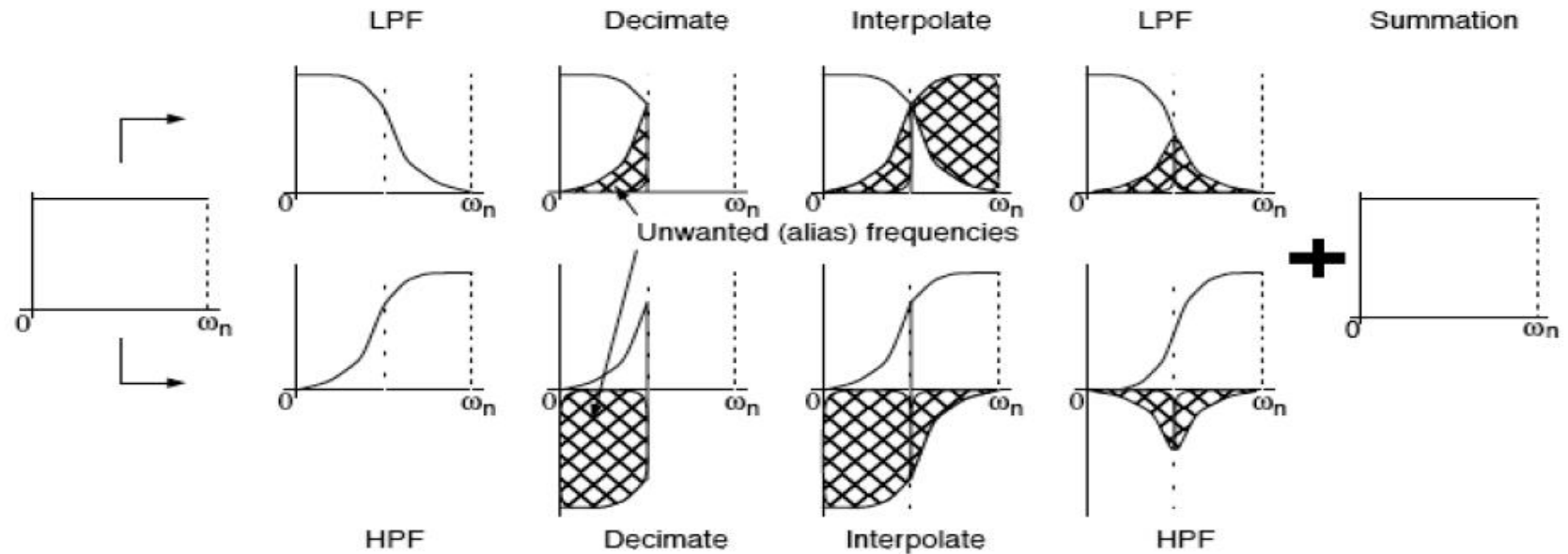
40:1

- ISO/IEC 15444-1:2000
- Compressão por “Wavelets”
- Codificação Aritmética
- Segmentação: retângulos, planos, regiões de interesse (ROI)
- Desempenho: arquivos 20-30% menores para mesma qualidade

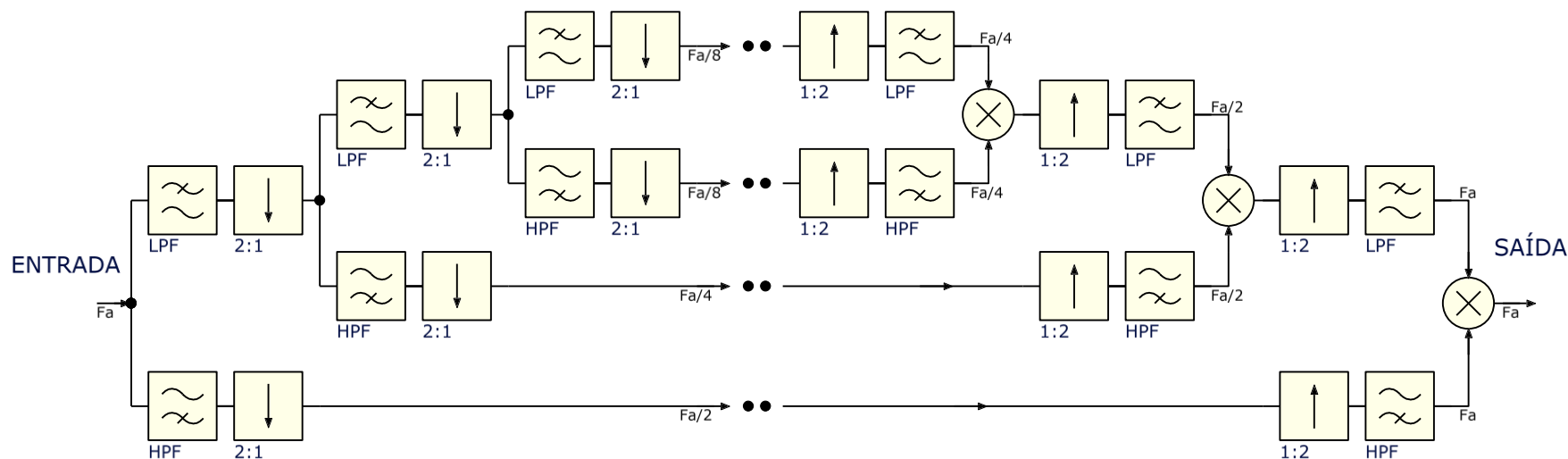
- Filtro passa-baixas + passa-altas + decimação 2:1
- Independente no sentido H e V
- Daubechies (irreversível) ou LeGall (reversível)



Transformada “Wavelet” unidimensional



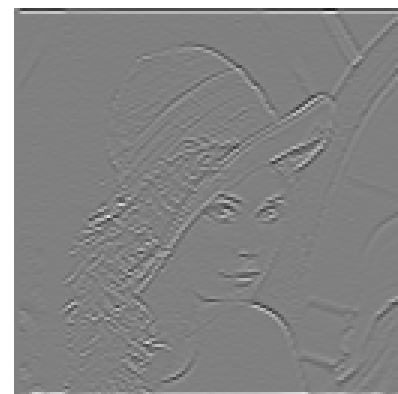
Transformada “Wavelet” em várias camadas



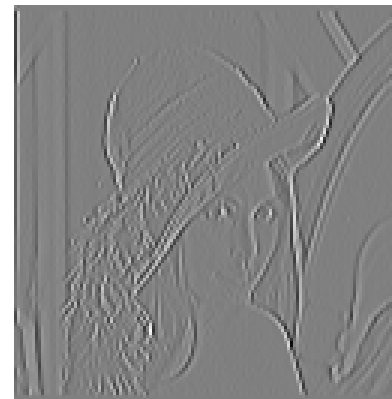
JPEG2000: Filtragem “Wavelet” Bidimensional



Original (128 x 128)



Passa-baixas ←



→ Passa-altas

Passa-baixas



Passa-altas ↑



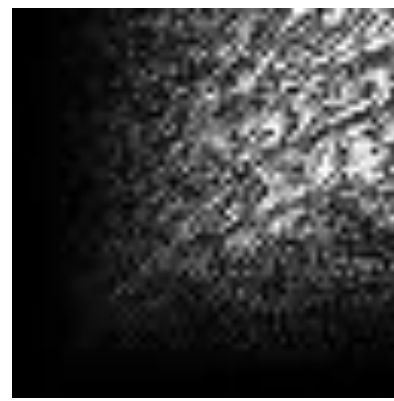
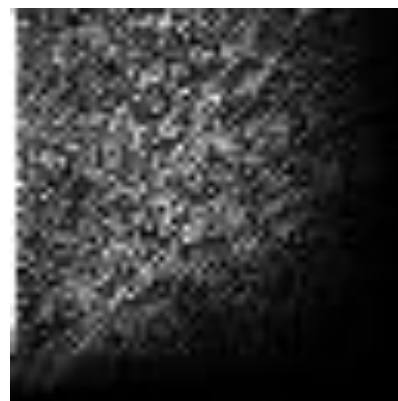
JPEG2000: Espectro pós Filtragem “Wavelet”



Espectro original



Passa-baixas

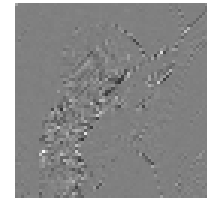
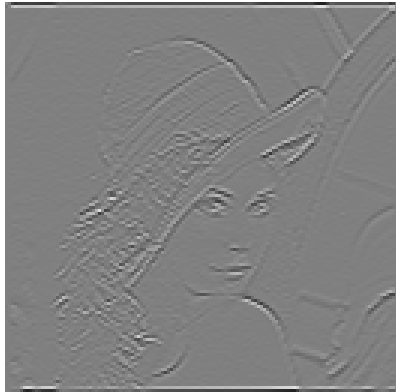
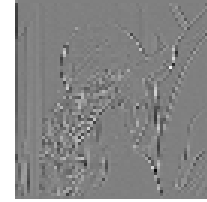
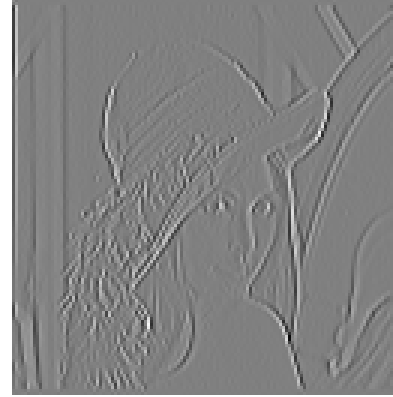


Passa-altas ↑

Passa-baixas ←

→ Passa-altas

JPEG2000: Decimação 2:1



JPEG2000: Filtragem “Wavelet” (1 passo)



Original (128 x 128)



Após filtragem e
decimação 2:1
(128 x 128)

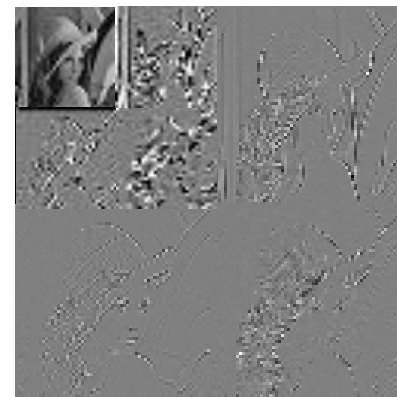
JPEG2000: Filtragem “Wavelet” (2 passos)



Original (128 x 128)



128 x 128 amostras
 $\frac{1}{2}$ banda

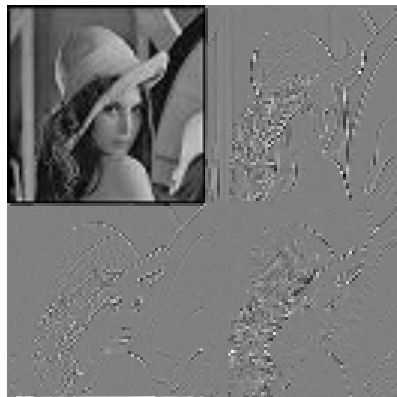


128 x 128 amostras
 $\frac{1}{2}$ e $\frac{1}{4}$ banda

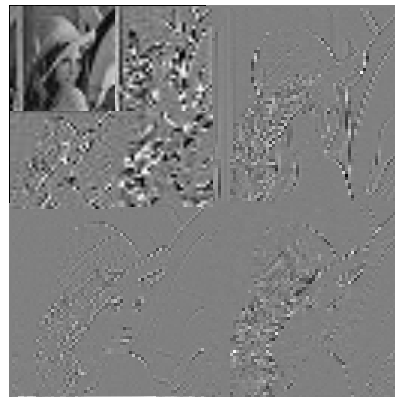
JPEG2000: Filtragem “Wavelet” (3 passos)



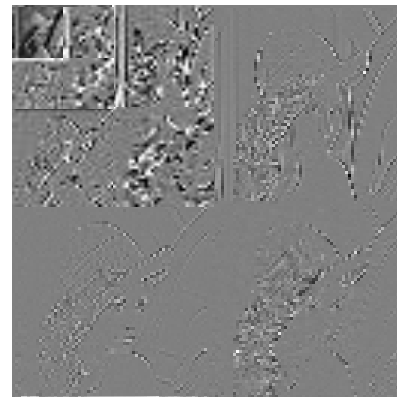
Original (128 x 128)



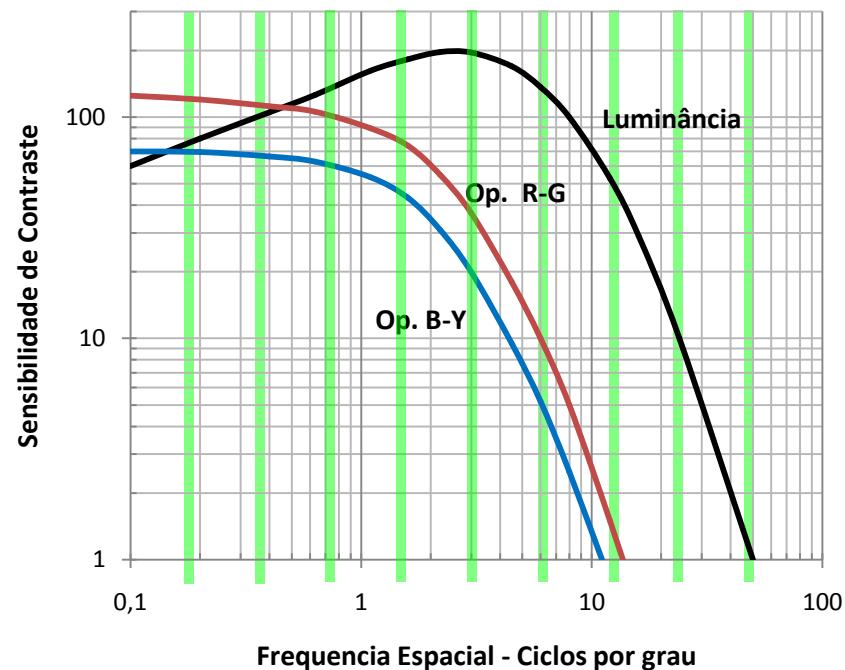
128 x 128 amostras
 $\frac{1}{2}$ banda



128 x 128 amostras
 $\frac{1}{2}$ e $\frac{1}{4}$ banda



128 x 128 amostras
 $\frac{1}{2}$, $\frac{1}{4}$ e $\frac{1}{8}$ banda





Original (512x512)
(8 bits por pixel)



Compressão 32:1
(0,25 bits/pixel)



Compressão 128:1
(0,0625 bits/pixel)



JPEG



JPEG2000

0,25 bits/pixel

