

COMPRESSÃO DE IMAGENS: PADRÃO JPEG

PTC3547 - CODIFICAÇÃO E TRANSMISSÃO MULTIMÍDIA

Guido Stolfi – EPUSP – 05/2022

1. Compressão de Imagens Estáticas

Iniciaremos os estudos de métodos de compressão de imagens analisando o processo JPEG, largamente usado para armazenamento de fotografias e desenhos. Os conceitos básicos deste processo serão expandidos posteriormente para englobar técnicas mais recentes de compressão de vídeo, como o MPEG e H.264.

Entre 1985 e 1990, foi organizado um comitê conjunto da ISO (International Organization for Standardization) e do CCITT (International Telegraph and Telephone Consultative Committee), hoje ITU-T, parte da ITU (International Telecommunications Union), com a finalidade de desenvolver um padrão de compressão de imagens. Este comitê recebeu a denominação de **JPEG** (Joint Photographic Experts Group).

O padrão JPEG, sintetizado na Recomendação JTC1 / 10918-2, especifica vários processos de compressão para imagens estáticas, incluindo um método de compactação (sem perdas) baseado em DPCM com preditores múltiplos. Já o método básico "com perdas", que é o mais utilizado, baseia-se na Transformada Discreta de Cossenos (DCT) para reduzir a correlação espacial entre as amostras, seguida de quantização variável e codificação estatística (Huffman). A figura 1.1 mostra um diagrama de blocos do processo JPEG básico.

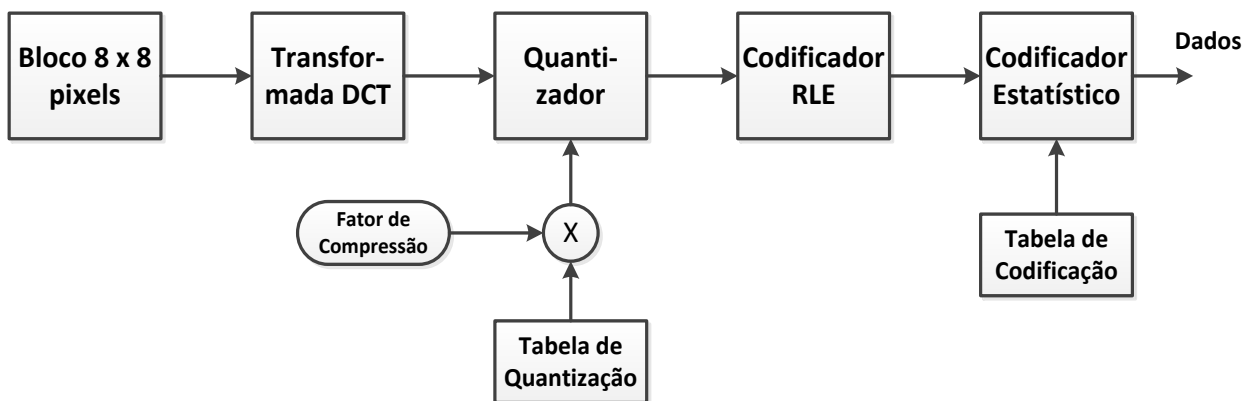


Figura 1.1 - Codificador JPEG

A **DCT**, assim como outras transformadas espaciais semelhantes usadas em outros processos de compressão, decompõe um bloco da imagem original na forma de uma soma ponderada de funções-base, ou primitivas, que são ortogonais entre si. Nesta etapa, a informação espacial contida na imagem é analisada e reagrupada em coeficientes que são relacionados ao conteúdo espectral da imagem.

As perdas (bem como a maior parte da compressão efetiva da informação) do processo são determinadas pela etapa de quantização, e a taxa de compressão pode ser aumentada, em detrimento da qualidade, ajustando-se um fator de escala no quantizador.

As etapas de codificação R. L. E. (*Run Length Encoding*) e Huffman otimizam a quantidade de bits usados para transmitir a informação quantizada.

O desempenho do padrão JPEG pode ser avaliado pela qualidade obtida para várias taxas médias de compressão. Na tabela abaixo, várias taxas são comparadas, tendo como referência imagens coloridas com estrutura de amostragem 4 : 2 : 0:

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

Devemos ressaltar que o padrão JPEG é destinado à compressão de imagens “reais”, ou seja, com tonalidades contínuas, como as obtidas por processos fotográficos a partir de objetos reais. Não é adequado para compressão de imagens geométricas, de alto contraste e resolução (textos impressos e desenhos técnicos, por exemplo).

1.1 Blocos 8×8 e Transformada DCT

Se a imagem for colorida, descrita por componentes R' , G' , B' (com correção gama), a etapa inicial consiste em um mapeamento para uma representação em componentes Y' , U' , V' , através das seguintes expressões:

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

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

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

As componentes U' e V' podem ser sub-amostradas de modo a reduzir a resolução espacial pela metade, nos sentidos vertical e horizontal. A imagem é a seguir decomposta em blocos de 8×8 pixels, seja no plano da luminância como nos da crominância, sendo que estes blocos são processados independentemente uns dos outros. O dimensionamento dos blocos (8 × 8) leva em conta a resposta em frequência espacial da visão humana: supondo que a taxa de amostragem espacial foi adequadamente escolhida, a distância entre 2 pixels deve estar perto do limite de resolução visual (1 a 2 minutos de grau). Nestas condições, a dimensão do bloco corresponde a detalhes de aproximadamente 8 a 16 minutos de grau, para os quais a visão humana atinge sensibilidade máxima.

A transformada DCT de um bloco de imagem $f(x,y)$ é definida por:

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

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

A função inversa é dada por

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

o que equivale a:

$$f(x, y) = \sum_u \sum_v F(u, v) \cdot P(u, v)$$

onde $P(u, v)$ são as funções-base ou primitivas da DCT. Os coeficientes $F(u, v)$ correspondem às quantidades de cada função primitiva que devem ser combinadas para obtermos o bloco original da imagem. A fig. 1.2 mostra as 64 primitivas da DCT, na forma de imagens elementares 8×8 .

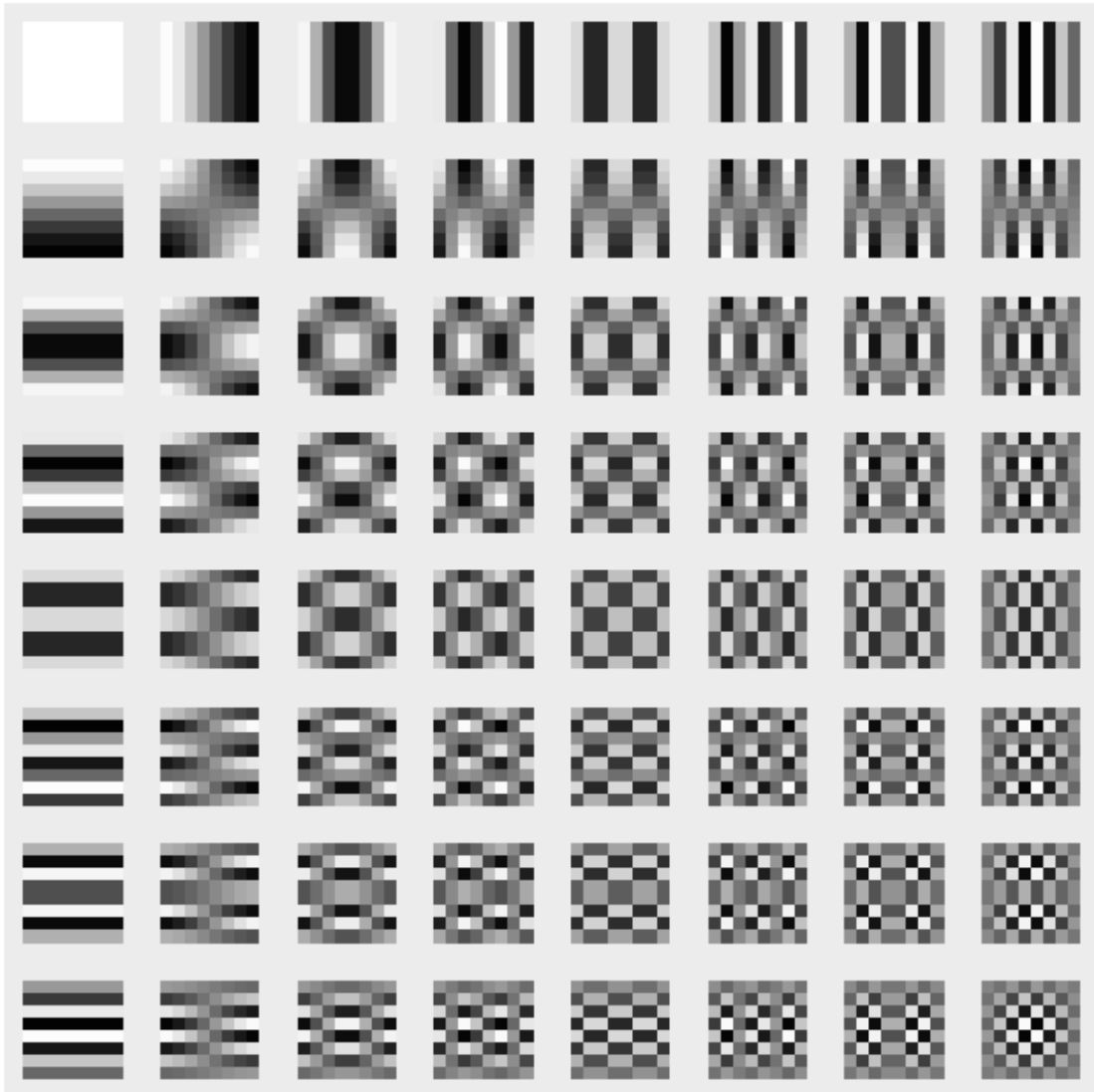


Fig. 1.2 - Primitivas, ou Funções-Base da DCT

A DCT é *separável*, ou seja, pode ser efetuada sequencialmente, primeiro no sentido horizontal e depois no vertical. Uma forma de implementar a DCT utilizando esta propriedade é através de um produto matricial:

$$DCT = C \times B \times C^T$$

onde B é um bloco de 8×8 pixels da imagem original e C é uma matriz de transformação, definida por:

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

$$C_{8 \times 8} =$$

.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

1.2 Quantização

Os coeficientes resultantes da DCT são quantizados usando tabelas arbitrárias $Q(u,v)$ e arredondados para o inteiro mais próximo:

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

As tabelas da Fig. 1.3, fornecidas como exemplo no padrão JPEG, são comumente usadas. Estes coeficientes foram obtidos experimentalmente, e levam em conta a resposta visual em frequência espacial para detalhes de luminância e croma. As tabelas de quantização são definidas pelo usuário durante a compressão da imagem, e devem ser anexadas ao arquivo comprimido.

Após a quantização, a tendência para imagens típicas é de que a maior parte dos coeficientes seja zero, restando apenas alguns coeficientes de baixa frequência espacial (concentrados no canto superior esquerdo).

Os coeficientes quantizados são reordenados por um padrão em zigue-zague, que tende a agrupar os coeficientes de maior importância no início da sequência (fig. 1.4).

1.3 Codificação Entrópica

O coeficiente quantizado $F(0,0)$ é processado separadamente; para os demais coeficientes, após a reordenação é feita uma codificação RLE (Run-Length Encoding), onde são criados símbolos consistindo do par ordenado:

{ quantidade de zeros / valor do coeficiente }

Quando na sequência reordenada de símbolos não há mais coeficientes não nulos dentro do bloco, a sequência é terminada por um símbolo especial (**EOB** - End of Block).

É usada finalmente uma codificação Huffman modificada, para traduzir os símbolos gerados pela RLE em padrões binários de comprimento variável. Nesta codificação, os valores dos coeficientes são representados por uma categoria e uma mantissa:

- a categoria especifica a ordem de grandeza do coeficiente e define a quantidade de bits necessária para representar a mantissa;
- a mantissa fornece o valor do coeficiente, dentro da faixa correspondente à categoria.

Os símbolos assumem então a forma:

{ quantidade de zeros / categoria / mantissa }

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

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

**Sugestão
(JPEG)**

8	6	5	8	12	20	26	31
6	6	7	10	13	29	30	28
7	7	8	12	20	29	35	28
7	9	11	15	26	44	40	31
9	11	19	28	34	55	52	39
12	18	28	32	41	52	57	46
25	32	39	44	52	61	60	51
36	46	48	49	56	50	52	50

9	9	12	24	50	50	50	50
9	11	13	33	50	50	50	50
12	13	28	50	50	50	50	50
24	33	50	50	50	50	50	50
50	50	50	50	50	50	50	50
50	50	50	50	50	50	50	50
50	50	50	50	50	50	50	50
50	50	50	50	50	50	50	50

**“Paint”
(Microsoft)**

Luminância

Crominância

Fig. 1.3 – Exemplos de Tabelas $Q(u,v)$ para Quantizador JPEG

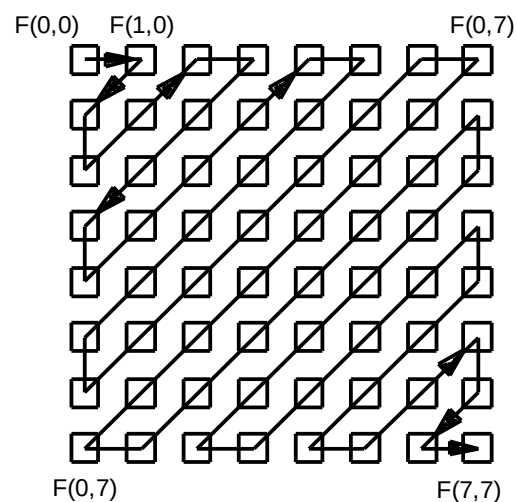


Fig. 1.4 - Sequência de Codificação dos Coeficientes da DCT

O código de Huffman utiliza uma tabela para codificar o par ordenado {quantidade de zeros + categoria} como um símbolo de comprimento variável, composto de um prefixo e de uma mantissa, cujo comprimento está implícito no prefixo.

A taxa de compressão efetiva do JPEG não é conhecida *a priori*, pois depende dos parâmetros estatísticos específicos da imagem. Atuando no parâmetro k (fator de compressão) conseguem-se taxas de compressão de 20:1 ou mais, mantendo degradação aceitável.



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



1,18 b/ pixel (20:1)



0,55 b/ pixel (44:1)



0,38 b/ pixel (62:1)



0,28 b/ pixel (85:1)



0,17 b/ pixel (140:1)

Fig. 1.5 – Exemplos de Codificação com Taxas Variadas

1.4 – Dimensionamento da DCT para JPEG

A escolha de uma matriz 8 x 8 para efetuar a extração dos coeficientes de frequência espacial está relacionada com a resposta em frequência espacial da visão humana.

O processo de quantização, em conjunto com a DCT, permite aplicar fatores de compressão diferentes, em função da frequência, dentro de uma escala de 8:1. Supondo que a maior frequência espacial esteja posicionada no limite da percepção visual, o que é de se esperar em condições normais de visualização, os coeficientes da última linha e coluna da matriz de quantização terão controle sobre a sensibilidade de contraste nessa frequência, que estaria em torno de 40 ciclos por grau (vide figura 1.6).

O coeficiente Q(0,0) teria controle sobre uma frequência igual a 1/8 desse valor, em torno de 5 ciclos por grau. Nesta faixa de frequências (região assinalada em verde na figura) temos a maior variação da resposta da visão, onde então será maior o benefício de podermos aplicar pesos diferentes a cada frequência na matriz de quantização.

Tamanhos maiores (p. ex. 16 x 16) entrariam no patamar da resposta em frequência da visão, onde não haveria vantagem em aplicar coeficientes de quantização diferentes.

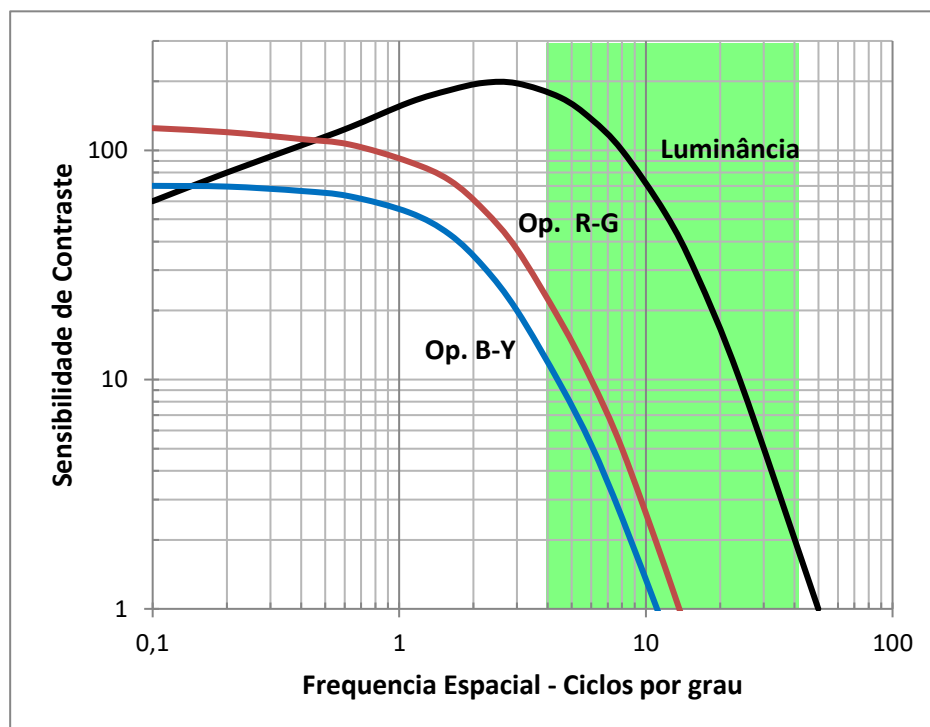


Fig. 1.6 – Resposta em Frequência Espacial (MTF) f_v da Visão Humana

Os valores dos coeficientes são classificados em categorias de acordo com sua ordem de grandeza, conforme a Tabela 4.1. Assim sendo, temos:

[0/2/-3 0/1/1 0/2/-3 0/2/-2 0/3/-6 0/2/2 0/3/-4 0/1/1 0/3/-4
0/1/1 0/1/1 0/3/5 1/2/2 2/1/-1 0/2/2 5/1/-1 0/1/-1 EOB]

Consultando o "Codebook" JPEG (Tabela 4.3), indexado pelo par {zeros/categoria}, os símbolos acima, incluindo o coeficiente DC já codificado, são convertidos para:

11000101 0100 001 0100 0101 100001 0110 100011 001 100011 001 001
100101 11100110 110110 0110 11110100 000 1010

Como resultado final, temos que o bloco original (512 bits) foi representado por uma sequência de 93 bits, o que significa uma taxa de compressão de 5.5 : 1, ou 1.45 bits/pixel. O bloco analisado é um exemplo de imagem "difícil" para o JPEG, pois contém um pico de luminância restrito ao centro do bloco. Em situações normais, a maior parte dos blocos apresentará apenas 2 ou 3 coeficientes quantizados, proporcionando taxas de compressão muito maiores.

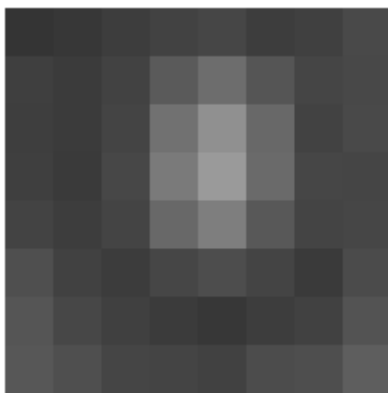
As tabelas abaixo mostram o bloco reconstruído a partir dos coeficientes quantizados, bem como o erro de reconstrução (diferença para o bloco original). Para este exemplo, o erro R.M.S. é da ordem de 5.9 níveis de intensidade.

58	64	67	64	59	62	70	78
56	55	67	89	98	88	74	69
60	50	70	119	141	116	80	64
69	51	71	128	149	115	77	68
74	53	64	105	115	84	65	72
76	57	56	74	75	57	57	74
83	69	59	60	61	61	67	78
93	81	67	62	69	80	84	84

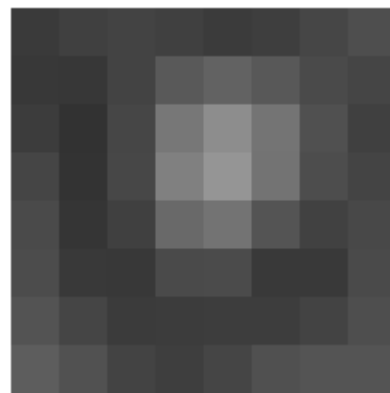
Imagem reconstruída

-6	-9	-6	2	11	-1	-6	-5
7	4	-1	1	11	-3	-5	3
2	9	-2	-6	-3	-12	-14	9
-6	7	0	-4	-5	-9	-7	1
-7	8	4	-1	11	4	3	-2
3	8	4	-4	2	11	1	1
2	2	5	-1	-6	0	-2	5
-6	-2	2	6	-4	-4	-6	10

Erro de reconstrução



Bloco Original



Bloco Reconstruído

Apesar dos erros de reconstrução, o processo preservou os elementos de maior importância dentro do bloco. A diferença pode ser imperceptível, considerando que a imagem seja visualizada em condições normais, ou seja, onde o tamanho aparente de cada pixel esteja no limite da acuidade visual.

Quando o fator de compressão é muito elevado, o processo JPEG apresenta um efeito de descontinuidade nos contornos dos blocos, denominado de "blocagem", causado pelo fato de que o tratamento dado a cada bloco é independente dos demais. A "blocagem" pode ser reduzida pela aplicação de filtros na imagem reconstruída, ou pela superposição parcial dos blocos durante a codificação, embora estes recursos não estejam incorporados no padrão JPEG.

A figura 2.1 mostra uma imagem fotográfica típica, com resolução de 512 x 512 pixels. A mesma imagem, após codificação/decodificação JPEG, é mostrada na figura 2.2. Para esta imagem em particular, a taxa de compressão obtida foi de 22:1.

Na figura 2.3 podemos observar o efeito de "blocagem" no detalhe ampliado à direita. Em condições normais de visualização, no entanto, estas irregularidades não são facilmente observáveis.

A figura 2.4 mostra exemplos de blocos "fáceis" e "difíceis" (ou seja, capazes de taxas de compressão efetiva respectivamente alta e baixa) que podem ser encontrados em imagens reais.



Fig. 2.1 – Imagem original ("Lena")



Fig. 2.2 – Imagem Comprimida por JPEG (aprox. 22:1)

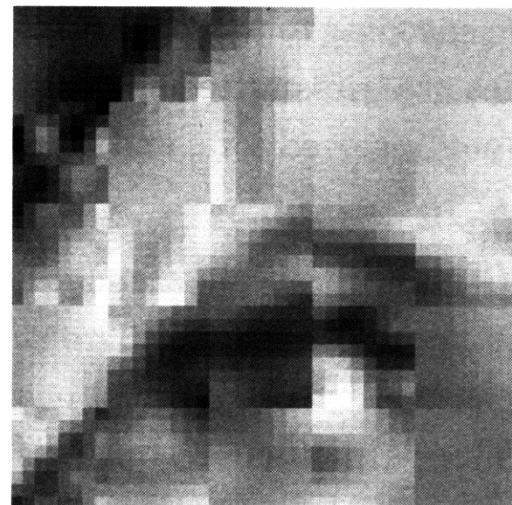


Fig. 2.3 – Detalhes Ampliados da Imagem “Lena”: Original e Comprimido

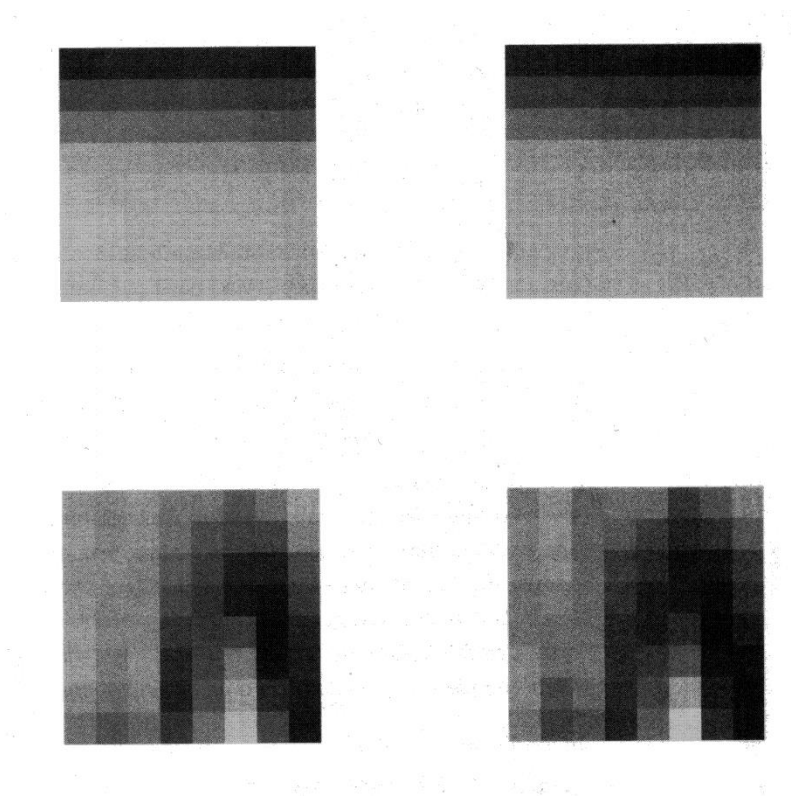


Fig. 2.4 – Blocos Típicos de Imagens Reais: Originais (esq.) e Comprimidos (Dir.)

3. Outras Transformações para Decorrelação Espacial

A transformada DCT possui propriedades excelentes no que diz respeito à eliminação de redundâncias espaciais, sendo essa a razão pela qual é utilizada no JPEG. Entre essas propriedades, podemos destacar:

- Separabilidade
- Coeficientes reais (não exige aritmética complexa)
- Existência de algoritmos rápidos (cerca de 50 multiplicações e 400 somas / deslocamentos para DCT 8×8).
- Minimização de efeitos de descontinuidades nas bordas dos blocos

Entre outras transformações de interesse para compressão de imagens, temos as transformadas de Hadamard, de Walsh e de Karhunen-Loève (ou Hotelling).

A transformada de Walsh utiliza primitivas (funções base) binárias (figura 3.1), e é obtida por:

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

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

A transformada de Hadamard é semelhante à de Walsh (para N igual a potências de 2, as funções base são as mesmas, porém reordenadas), sendo dada por:

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

As funções-base destas transformadas, sendo binárias, favorecem o processamento da imagem uma vez que dispensam multiplicações.

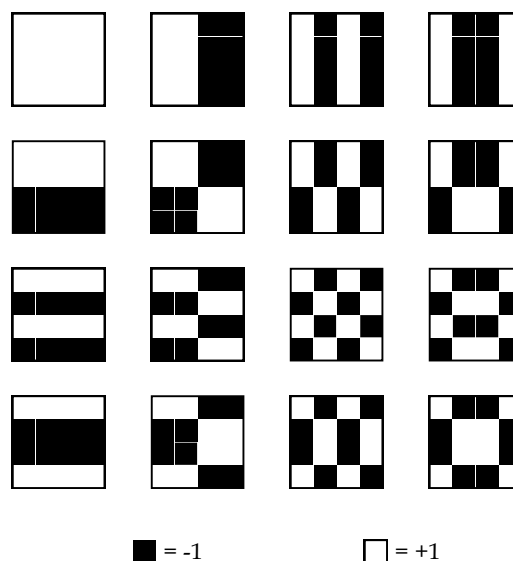


Fig. 3.1 - Funções Base para Transformada de Walsh 4×4

Já a transformada de Karhunen-Loève (KLT) é ótima, no sentido de que minimiza a correlação entre os coeficientes. No entanto, as funções-base dependem das características da imagem, dificultando a aplicação prática desta transformação. Suponhamos que uma imagem seja representada por um conjunto de vetores aleatórios (por exemplo, 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 Covariância 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 então $\mathbf{e}_i$ e $\lambda_i, i = 1, 2, \dots, n$ 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 Karhunen-Loève baseada no conjunto de vetores $\mathbf{x}$.

Algumas propriedades da KLT são:

- ⇒ Elementos de $\mathbf{y}$ são descorrelacionados;
- ⇒ Erro quadrático médio da reconstrução é minimizado;
- ⇒ Funções-base (Matriz $\mathbf{A}$) dependem da imagem analisada.

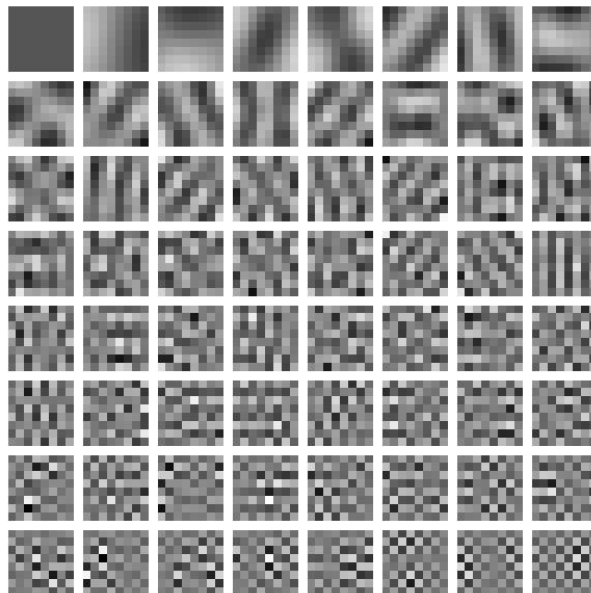


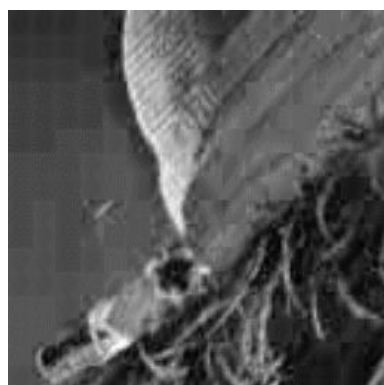
Fig. 3.2 – Exemplo de primitivas KLT para a imagem “Lena”



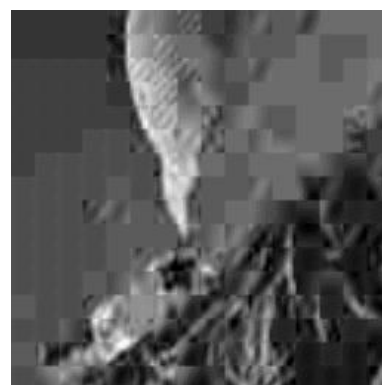
a) Original



b) Compressão 10:1



c) Compressão 20:1



d) Compressão 40:1

Fig. 3.3 - Erros de Reconstrução pela Transformada KLT

4. Tabelas de Codificação JPEG

A representação dos coeficientes quantizados no JPEG, por codificação entrópica, é composta de um prefixo de comprimento variável, seguido por um campo de codificação do valor do coeficiente (também de comprimento variável).

Os prefixos definem categorias de comprimento para as palavras codificadas, e correspondem à quantidade de bits necessária para representação do valor do coeficiente quantizado. Por exemplo, a categoria 5 reserva 5 bits para codificar os 32 valores compreendidos no intervalo $[-31$ a -16 e $+16$ a $+31]$.

A tabela 4.1 mostra as categorias correspondentes a cada faixa de valores representáveis dos coeficientes (DC ou AC). Podem ser representados coeficientes DC de -2047 a $+2047$, e coeficientes AC de -1023 a -1 e de $+1$ a $+1023$.

Faixa de Valores	Categoria p/ coef. DC	Categoria p/ coef. AC
0	0	N/A
-1, 1	1	1
-3, -2, 2, 3	2	2
-7, ..., -4, 4, ..., 7	3	3
-15, ..., -8, 8, ..., 15	4	4
-31, ..., -16, 16, ..., 31	5	5
-63, ..., -32, 32, ..., 63	6	6
-127, ..., -64, 64, ..., 127	7	7
-255, ..., -128, 128, ..., 255	8	8
-511, ..., -256, 256, ..., 511	9	9
-1023, ..., -512, 512, ..., 1023	A	A
-2047, ..., -1024, 1024, ..., 2047	B	N/A

Tabela 4.1 - Categorias de Codificação de Coeficientes DC e AC p/ JPEG

Categoria	Prefixo	Comprimento Total	Bits na 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

Tabela 4.2 - Prefixos para Coeficientes DC

A tabela 4.2 fornece os prefixos e o comprimento total dos símbolos codificados para as categorias correspondentes aos coeficientes DC (correspondentes a $F(0,0)$ após quantização). Por exemplo: um coeficiente DC de valor +200 é representável pela categoria 8 (que cobre a faixa de valores de -255 a -128 e de +128 a +255).

Na tabela 4.2, vemos que o prefixo correspondente é 111110 e o comprimento total é 14 bits; restam portanto 8 bits para codificar o valor 200 dentro da faixa coberta por aquela categoria. Dentro desta faixa os valores são codificados da seguinte maneira:

Valor	Código
-255	00000000
-254	00000001
-253	00000010
...	
-128	01111111
+128	10000000
+129	10000001
...	
+255	11111111

Desta forma, o valor 200 será codificado como **11001000**; concatenando-o com o prefixo, temos então que $+200 = \mathbf{111110\ 11001000}$.

Já para os coeficientes AC, a tabela 4.3 traz, para cada combinação de {quantidade de zeros / categoria de representação do valor do coeficiente}, o correspondente prefixo e o comprimento total do símbolo. Por exemplo, para uma sequência de 2 coeficientes nulos seguidos de um coeficiente igual a -17 (ou seja, $\{0 / 0 / -17\}$) temos, pela tabela 4.1, que a categoria 5 é adequada para representar o valor -17, através de uma mantissa de 5 bits.

Esta mantissa será representada por uma sequência 01110. Na tabela 4.3, o prefixo correspondente à combinação (2 / 5), ou seja, dois zeros e um valor da categoria 5, é **1111111110001010**. Acrescentando a mantissa do coeficiente, chegamos a **1111111110001010 01110**.

zeros/ categ.	Prefixo	comp. total	zeros/ categ.	Prefixo	comp. total	zeros/ categ.	Prefixo	comp. total
0/0	1010 (= EOB)	4	5/4	111111110100000	20	A/8	111111111001110	24
0/1	00	3	5/5	1111111110100001	21	A/9	1111111111001111	25
0/2	01	4	5/6	1111111110100010	22	A/A	1111111111010000	26
0/3	100	6	5/7	1111111110100011	23	B/1	111111010	10
0/4	1011	8	5/8	1111111110100100	24	B/2	1111111111010001	18
0/5	11010	10	5/9	1111111110100101	25	B/3	1111111111010010	19
0/6	111000	12	5/A	1111111110100110	26	B/4	1111111111010011	20
0/7	1111000	14	6/1	1111011	8	B/5	1111111111010100	21
0/8	1111110110	18	6/2	11111111000	13	B/6	1111111111010101	22
0/9	1111111110000010	25	6/3	1111111110100111	19	B/7	1111111111010110	23
0/A	1111111110000011	26	6/4	1111111110101000	20	B/8	1111111111010111	24
1/1	1100	5	6/5	1111111110101001	21	B/9	1111111111011000	25
1/2	111001	8	6/6	1111111110101010	22	B/A	1111111111011001	26
1/3	1111001	10	6/7	1111111110101011	23	C/1	1111111010	11
1/4	111110110	13	6/8	1111111110101100	24	C/2	1111111111011010	18
1/5	11111110110	16	6/9	1111111110101101	25	C/3	1111111111011011	19
1/6	1111111110000100	22	6/A	1111111110101110	26	C/4	1111111111011100	20
1/7	1111111110000101	23	7/1	11111001	9	C/5	1111111111011101	21
1/8	1111111110000110	24	7/2	11111111001	13	C/6	1111111111011110	22
1/9	1111111110000111	25	7/3	1111111110101111	19	C/7	1111111111011111	23
1/A	1111111110001000	26	7/4	1111111110110000	20	C/8	1111111111100000	24
2/1	11011	6	7/5	1111111110110001	21	C/9	1111111111100001	25
2/2	11111000	10	7/6	1111111110110010	22	C/A	1111111111100010	26
2/3	1111110111	13	7/7	1111111110110011	23	D/1	11111111010	12
2/4	1111111110001001	20	7/8	1111111110110100	24	D/2	1111111111100011	18
2/5	1111111110001010	2	7/9	1111111110110101	25	D/3	1111111111100100	19
2/6	1111111110001011	22	7/A	1111111110110110	26	D/4	1111111111100101	20
2/7	1111111110001100	23	A/7	1111111111001101	23	D/5	1111111111100110	21
2/8	1111111110001101	24	8/1	11111010	9	D/6	1111111111100111	22
2/9	1111111110001110	25	8/2	111111111000000	17	D/7	1111111111101000	23
2/A	1111111110001111	26	8/3	1111111110110111	19	D/8	1111111111101001	24
3/1	111010	7	8/4	1111111110111000	20	D/9	1111111111101010	25
3/2	111110111	11	8/5	1111111110111001	21	D/A	1111111111101011	26
3/3	11111110111	14	8/6	1111111110111010	22	E/1	111111110110	13
3/4	1111111110010000	20	8/7	1111111110111011	23	E/2	1111111111101100	18
3/5	1111111110010001	21	8/8	1111111110111100	24	E/3	1111111111101101	19
3/6	1111111110010010	22	8/9	1111111110111101	25	E/4	1111111111101110	20
3/7	1111111110010011	23	8/A	1111111110111110	26	E/5	1111111111101111	21
3/8	1111111110010100	24	9/1	111111000	10	E/6	1111111111110000	22
3/9	1111111110010101	25	9/2	1111111110111111	18	E/7	1111111111110001	23
3/A	1111111110010110	26	9/3	1111111111000000	19	E/8	1111111111110010	24
4/1	111011	7	9/4	1111111111000001	20	E/9	1111111111110011	25
4/2	1111111000	12	9/5	1111111111000010	21	E/A	1111111111110100	26
4/3	1111111110010111	19	9/6	1111111111000011	22	F/0	111111110111 (*)	12
4/4	1111111110011000	20	9/7	1111111111000100	23	F/1	1111111111110101	17
4/5	1111111110011001	21	9/8	1111111111000101	24	F/2	1111111111110110	18
4/6	1111111110011010	22	9/9	1111111111000110	25	F/3	1111111111110111	19
4/7	1111111110011011	23	9/A	1111111111000111	26	F/4	1111111111111000	20
4/8	1111111110011100	24	A/1	111111001	10	F/5	1111111111111001	21
4/9	1111111110011101	25	A/2	1111111111001000	18	F/6	1111111111111010	22
4/A	1111111110011110	26	A/3	1111111111001001	19	F/7	1111111111111011	23
5/1	1111010	8	A/4	1111111111001010	20	F/8	1111111111111100	24
5/2	1111111001	12	A/5	1111111111001011	21	F/9	1111111111111101	25
5/3	1111111110011111	19	A/6	1111111111001100	22	F/A	1111111111111110	26

Tabela 4.3 - Codificação Entrópica (Huffman) para Símbolos AC (luminância) JPEG

(*)-extensão de zeros

5. Codificação JPEG-2000

O padrão ISO-IEC 15444-1:2000, conhecido como JPEG 2000, utiliza transformada por “Wavelets” para efetuar a decomposição espectral da imagem. Esta transformada se baseia na aplicação de filtros complementares, passa-baixas e passa-altas (fig. 5.1), que subdividem o espectro de frequências espaciais da imagem em uma partição binária (logarítmica), ao contrário da DCT, que efetua uma subdivisão linear.

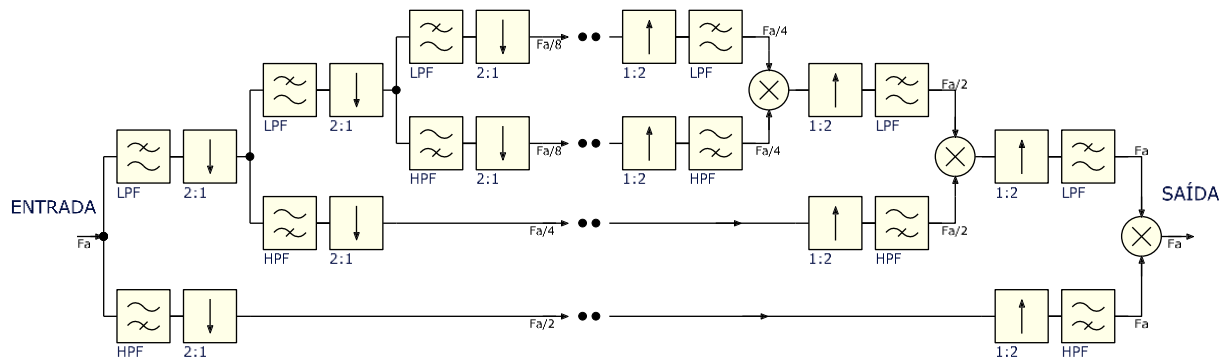


Fig. 5.1 - Esquema de filtros para decomposição e reconstrução da imagem

Um “Wavelet”, no caso, é uma função que representa a resposta ao impulso de um filtro (passa-baixas ou passa-altas), e que permite a análise combinada nos domínios da frequência e do espaço. O padrão JPEG 2000 pode utilizar “Wavelets” de dois tipos: *Daubechies* (irreversíveis) e *LeGall* (reversíveis), conforme fig. 5.2. Observe que a somatória dos coeficientes das respostas ao impulso é 1 (passa-baixas) ou 0 (passa-altas).

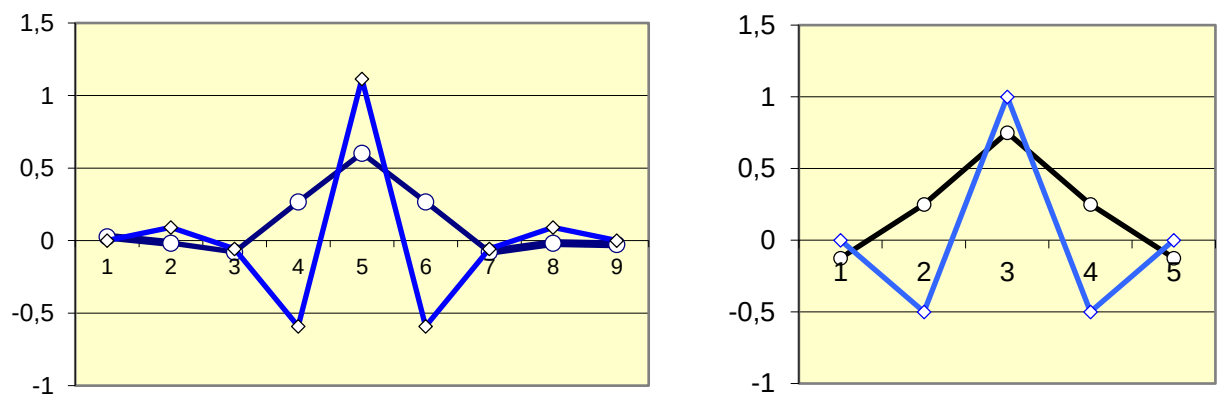


Fig. 5.2 – Wavelets tipo Daubechies (esq.) e LeGall (dir). $\diamond$ = Passa-altas; $\circ$ = Passa-baixas

A aplicação de um par de filtros (passa-altas e passa-baixas) no sentido horizontal, sobre a imagem original, produz duas novas imagens, cada uma com metade do conteúdo espectral na direção H. Aplicando novamente o mesmo par de filtros nessas duas imagens, mas no sentido vertical, obtemos 4 novas imagens, cada uma contendo uma combinação de conteúdo espectral nos sentidos H e V: baixo/baixo, baixo/alto, alto/baixo e alto/alto (figuras 5.3 e 5.4).

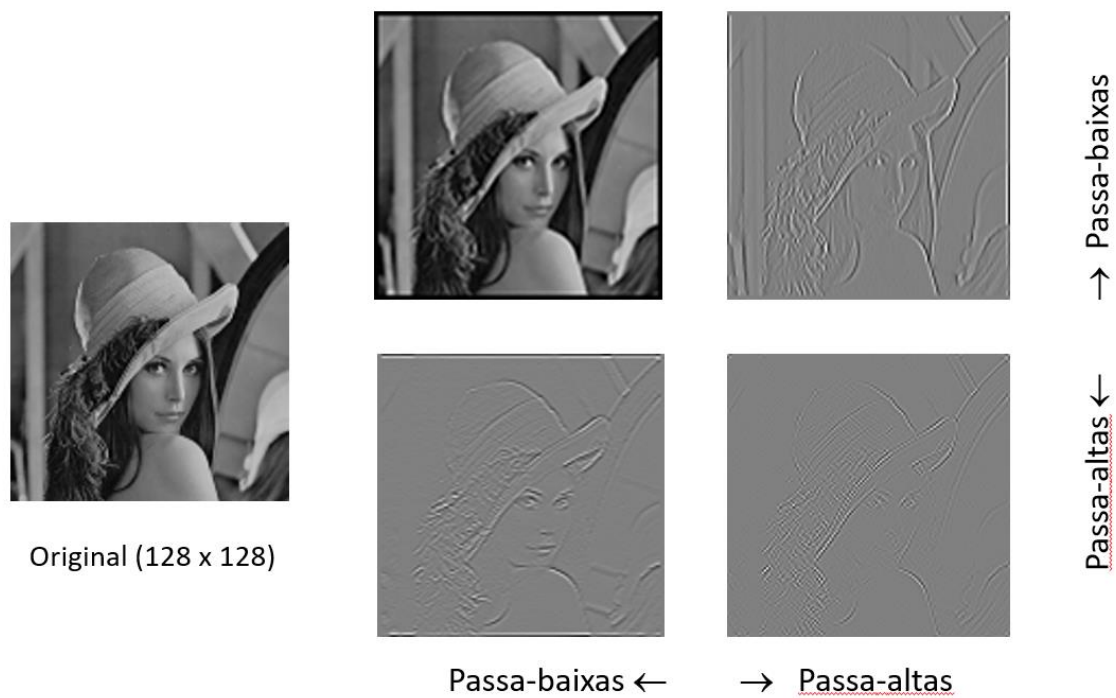


Fig. 5.3 – Filtragem na metade da banda, nas duas dimensões

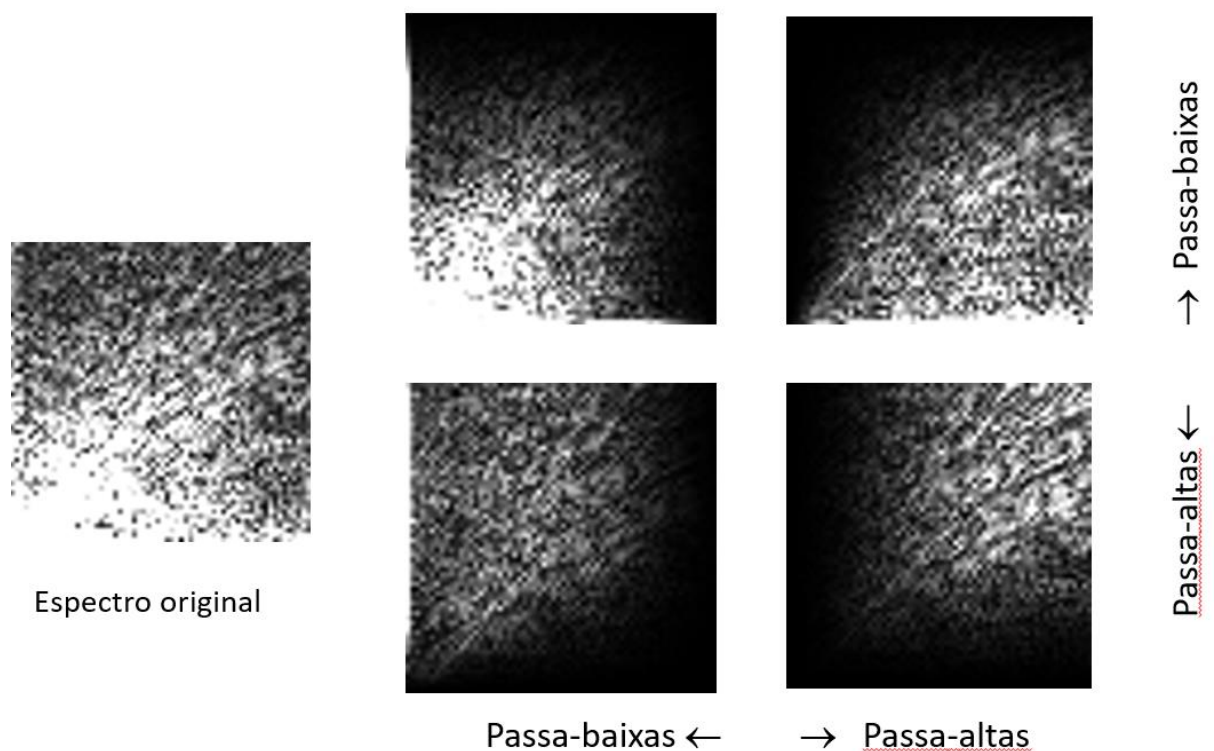


Fig. 5.4 – Espectros bidimensionais – original e imagens filtradas

Uma vez que cada nova imagem possui espectro limitado à metade da frequência de Nyquist correspondente, as 4 imagens obtidas podem ser decimadas na proporção 2:1, resultando assim em 4 sub-imagens (fig. 5.5 e 5.6), contendo cada qual a informação referente à metade do espectro de frequências espaciais (nos sentidos H e V).

Como os filtros não são ideais, vai ocorrer rebatimento espectral. No entanto, pela simetria dos filtros, esse rebatimento será praticamente cancelado na reconstrução da imagem. Note que a quantidade total de amostras das 4 sub-imagens permanece constante após a decimação, ou seja, este processo por si só não causa redução na informação.

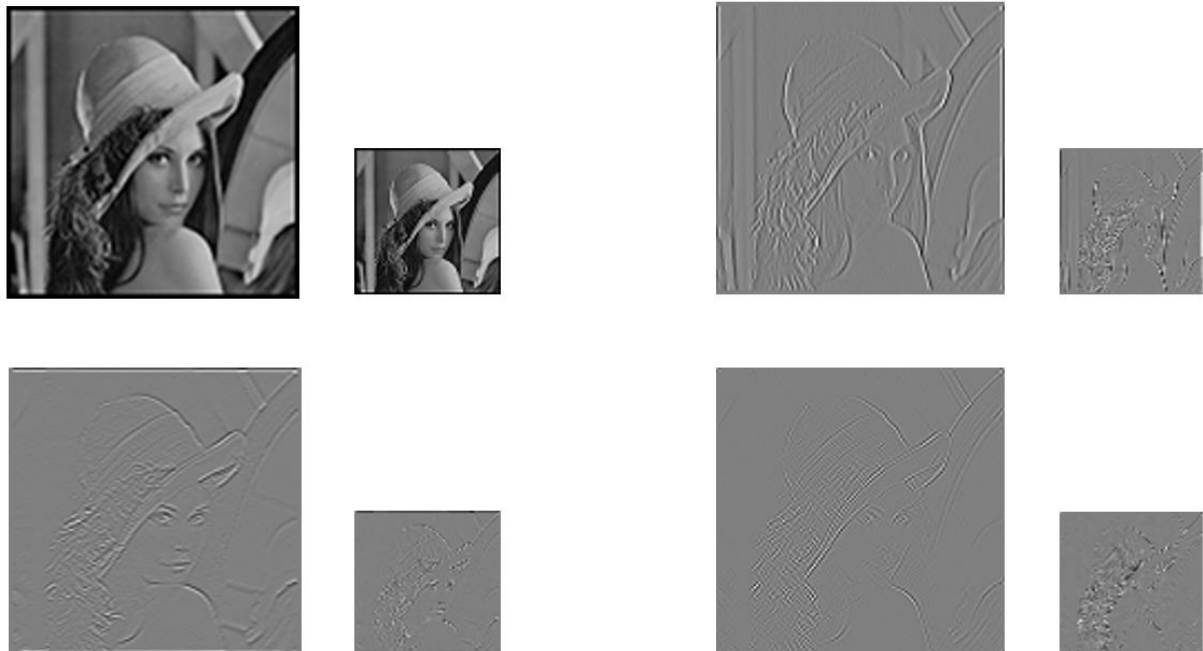


Fig. 5.5 – Decimação 2:1

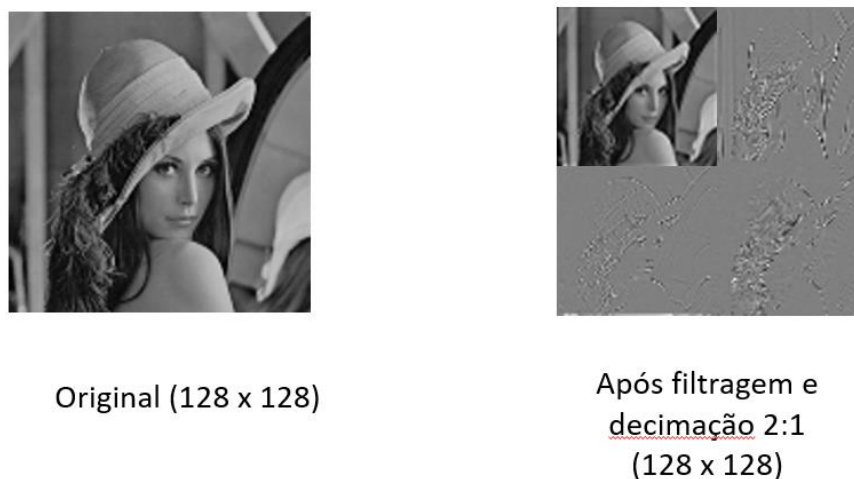


Fig. 5.6 – Resultado da filtragem por "Wavelets" seguida de decimação 2:1

O processo é repetido na sub-imagem correspondente às baixas frequências, nos sentidos horizontal e vertical, ou seja, na parte superior esquerda da imagem resultante da associação das 4 sub-imagens, e assim sucessivamente (fig. 5.7), até eventualmente resultar em um único *pixel*, cujo valor representa a intensidade média de toda a imagem (frequência zero).

Cada sub-imagem obtida em cada etapa será então quantizada, e as amostras não nulas resultantes são compactadas e codificadas em símbolos de comprimento variável.

Além disso, o padrão JPEG 2000 utiliza a chamada “Codificação Aritmética” para compactar os símbolos a serem transmitidos. Pelo fato de operar sobre as probabilidades condicionais de conjuntos de símbolos, esta codificação permite compressão mais eficiente do que a codificação de Huffmann, que opera sobre símbolos independentes entre si.

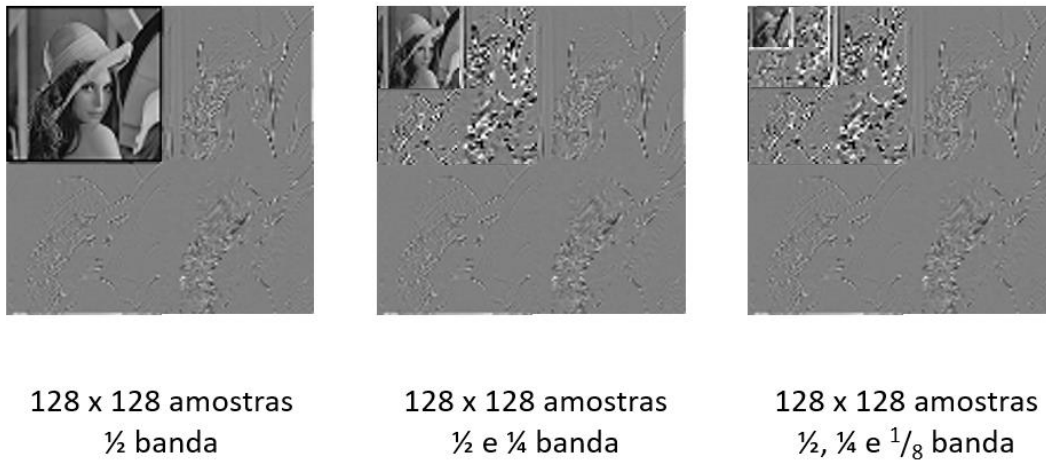


Fig. 5.7 – Etapas sucessivas de filtragem e decimação no JPEG 2000

A codificação JPEG 2000 proporciona economias significativas na taxa de bits necessária, em relação ao JPEG “baseline”, para mesma qualidade aparente. Pelo fato de poder operar com regiões de qualquer formato e dimensão, apresenta-se relativamente livre de efeitos de “blocagem”, como pode ser visto na figuras 5.8. No entanto, a complexidade computacional é substancialmente maior que o JPEG convencional.



Fig. 5.8 – Imagem comprimida 32:1 (0,25 bits por pixel, à esq.) e 128:1 (0,0625 bits/pixel, à dir.) com processo de “wavelets”

6. Aplicação do processo JPEG em imagens em movimento: M-JPEG, DV e DCP

O termo M-JPEG refere-se a um método de codificação de imagens em movimento no qual cada quadro de uma sequência é codificado utilizando o processo JPEG (ou algum algoritmo semelhante). Este método não é padronizado, levando a sistemas incompatíveis entre si, mas que são usados basicamente na manipulação e armazenamento de imagens em sistemas de edição de vídeo.

O padrão DV, utilizado nos gravadores DVC (*Digital Video Cassette*), também é similar ao M-JPEG, mas permite codificação de imagens entrelaçadas. O codificador DV opera com taxa de bits fixa, alterando dinamicamente a matriz de quantização de forma a manter taxas de 12.5 Mb/s (modo SDL – *Standard Definition, Long Play*), 25 Mb/s (Modo SD – *Standard Definition*) ou 50 Mb/s (Modo HD – *High Definition* para HDTV).

O padrão DCP (*Digital Cinema Package*) é usado para distribuir conteúdo cinematográfico, sendo que a informação de vídeo consiste de imagens codificadas em JPEG-2000 em taxas de 24 a 60 fotogramas por segundo. A resolução alcança 2160 x 4096 pixels por quadro, usando quantização de 12 bits e espaço colorimétrico XYZ. Adota uma restrição na taxa de bits, permitindo no máximo 250 Mbps.

7. Referências

- Digital Image Processing* - Rafael C. Gonzales, Richard E. Woods - Addison Wesley, 1992
- Digital Vídeo and HDTV: Algorithms and Interfaces – Charles Poynton – Morgan Kaufmann, 2003
- Video Engineering* - Andrew F. Inglis, Arch C. Luther - Mc Graw-Hill, 1996
- Video Compression* – Peter D. Symes – Mc Graw-Hill, 1998
- The JPEG Still Picture Compression Standard* – Gregory K. Wallace – IEEE Transactions on Consumer Electronics, Vol. 38, 1992
- The JPEG 2000 Still Image Compression Standard* – A. Skodras, C. Christopoulos, T. Ebrahimi – IEEE Signal Processing Magazine, Set. 2001, p. 36 (Figs. 5.2, 5.3 e 5.5)
- Compression in Video and Audio* – John Watkinson – Focal Press, 1995 (Figs. 2.1 a 2.4)
- Digital Video Processing* - A. Murat Tekalp – Prentice Hall, 1995
- Karhunen-Loève Transform* - R. D. Dony - The Transform and Data Compression Handbook, Ed. K. R. Rao and P.C. Yip - CRC Press, 2001 (figs. 3.2 e 3.3)
- Wavelet-Based Image Compression* - James S. Walker - The Transform and Data Compression Handbook, Ed. K. R. Rao and P.C. Yip - CRC Press, 2001 (fig. 5.4)