

Pré-Processamento em Codificação de Imagens por Sub-Bandas para Redução de Irrelevância Visual

MARCELO MOLLICA JOURDAN^{1,2}
ABRAHAM ALCAIM²
PAULO ROBERTO ROSA LOPES NUNES¹

¹IBM Brasil-Centro Científico Rio
Av. Presidente Vargas, 844 - VB22
20071-001 Rio de Janeiro - RJ
nunes@vnet.ibm.com

² CETUC - Centro de Estudos em Telecomunicações
Pontifícia Universidade Católica - PUC-Rio
Rua Marquês de São Vicente 225
22453-900 Rio de Janeiro, RJ, Brasil
jourdant,alcaim@cetuc.puc-rio.br

Abstract. In low bit-rate coding of images, it has been shown that compression ratio can be greatly improved by eliminating information irrelevant to the human eye. In this paper we describe a technique for accomplishing this task. The image spectrum is decomposed in 10 sub-bands and, for each band, blocks of pixels are discarded if they do not satisfy thresholds of average energy. The benefits of this technique are reduction of pixels to be coded (typically to 15-50 % of the original), low computational complexity and very little side information generated.

1 Introdução

Exigências cada vez maiores de taxas de compressão mais altas na codificação de imagens aliadas à necessidade de boa qualidade de reconstrução têm forçado a uma mudança no paradigma (taxa fixa)-(qualidade variável) para (taxa variável)-(qualidade fixa).

Para atingir esses objetivos é importante observar que o sistema visual humano não é suficientemente acurado na percepção do ruído introduzido pela codificação quando este se encontra na mesma região espectral, espacial ou temporal do sinal. A esse fenômeno dá-se o nome de mascaramento do ruído [Jayant et al. (1993)].

A incorporação dos conceitos acima aos resultados previstos pela teoria da taxa-distorção para a codificação de uma fonte de informação, permitem definir a Entropia Perceptual como um limite mínimo para a taxa de codificação levando-se em conta não a minimização do erro médio quadrático, mas sim a reconstrução perfeita para o olho humano. Um codificador ideal no sentido da Entropia Perceptual codificaria todos os pixels de uma imagem exatamente no nível do JND ("Just Noticeable Distortion").

Para se compreender o projeto de um codificador perceptual eficiente, consideremos a Fig. 1. Nessa figura [Jayant et al. (1993)] podemos ver, de maneira esquematizada, as duas grandezas que um codi-

ficador eficiente deve minimizar: redundância e irrelevância. A redundância (eixo vertical), caracteriza-se pela correlação entre amostras do sinal de entrada e é reduzida através de técnicas como predição, codificação por transformada ou quantização vetorial. Já a irrelevância (eixo horizontal) está relacionada ao mascaramento do ruído pelo sinal e é minimizada codificando-se o sinal perto do nível de JND.

Essas duas operações podem ser efetuadas de maneira separada ou conjunta, como no caso de um quantizador vetorial, que explora a correlação entre amostras adjacentes e leva em conta o mascaramento do ruído pelo sinal através do processo de quantização.

A proposta deste trabalho consiste de uma técnica simples e eficiente para redução de irrelevância visual em codificadores de imagens por sub-bandas. Essa redução é alcançada através da eliminação local de informações que não satisfaçam a limiares mínimos de percepção. Esses limiares são determinados empiricamente e variam de acordo com a sensibilidade do olho humano para as várias faixas de frequência bi-dimensionais.

A Seção 2 deste trabalho descreve a decomposição do sinal em sub-bandas de frequência. Esse processamento está intimamente ligado à técnica de redução de irrelevância apresentada neste trabalho

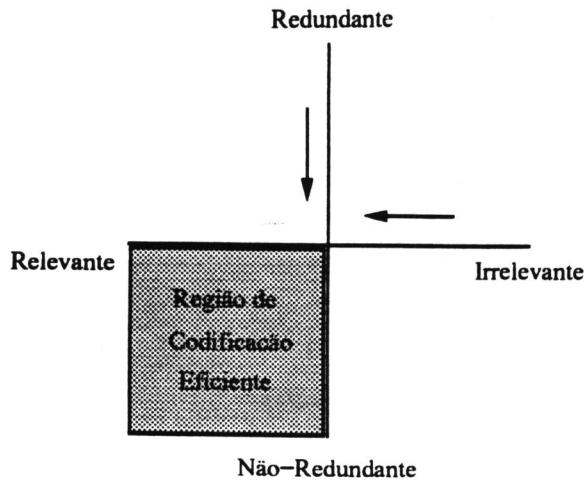


Figura 1: Redundância e Irrelevância.

e deixa os dados de forma adequada para a etapa de redução de redundância. A Seção 3 apresenta a técnica de redução de irrelevância visual proposta neste trabalho e a Seção 4 discute os resultados de simulações. A redução da taxa de informação lateral é tratada na Seção 5, comparando-se três técnicas de compactação: Huffman com predição, Lempel-Ziv e Quad-Trees. Finalmente, na seção 6 são apresentadas as principais conclusões deste trabalho.

2 Decomposição em Sub-bandas

A decomposição em sub-bandas espectrais não reduz a entropia da fonte [Senoo-Girod (1992)], mas apresenta várias características interessantes, como: as sub-bandas geradas são ortogonais entre si, apresentam menor correlação intra-banda, têm características estatísticas diferentes e importâncias subjetivas decrescentes com a frequência. Tais fatos permitem a geração de quantizadores eficientes (Seção 1) de acordo com cada sub-banda.

Para esse trabalho foi escolhida a decomposição em forma de pirâmide com três níveis, num total de 10 sub-bandas (Fig. 2). A decomposição em pirâmide, dentre outras formas de decomposição, é tida como das mais eficientes por aproveitar melhor as altas correlações das bandas mais baixas em frequência.

A decomposição do sinal foi feita de maneira separável com filtros QMF de 9 coeficientes [Woods (1991)] [Senoo-Girod (1992)]. Os filtros QMF ("Quadrature Mirror Filters") são adequados para a filtragem de um sinal em duas partes, uma passa-baixa e a outra passa-alta. As respostas em frequência dos dois

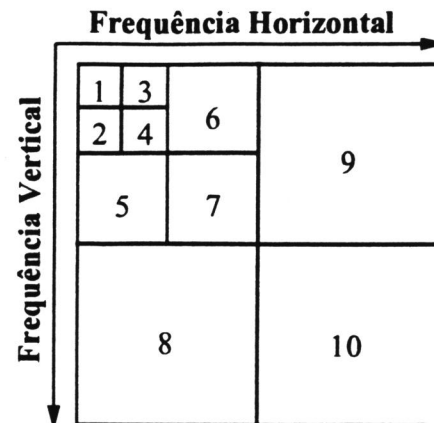


Figura 2: Decomposição em 10 Sub-Bandas.

filtros são simétricas. A reconstrução em frequência do sinal é quase perfeita, com cancelamento total da componente de "alias" introduzida na síntese. A filtragem se dá primeiro nas linhas e depois nas colunas, gerando, para cada imagem de entrada, 4 sub-imagens. De maneira recursiva, através da divisão das sub-bandas de mais baixa frequência em 4 sub-imagens, chega-se a um total de 10 componentes.

A convolução linear de uma seqüência de N coeficientes com um filtro de ordem M gera uma seqüência com $N+M-1$ coeficientes. Tal fato vai de encontro ao objetivo de diminuição da taxa de bits de informação. Em [Woods-O'Neil (1986)], foi sugerido o uso de convolução circular ao invés de linear para a decomposição em sub-bandas, o que pode ser implementado [Smith-Eddins (1990)] repetindo-se a imagem periodicamente. Concatena-se a borda esquerda da imagem à sua borda direita e a borda superior à inferior. Ao final da filtragem, a soma do número de amostras distintas nas sub-bandas é igual ao número de pixels da imagem original. A desvantagem desta técnica é que, caso a imagem apresente valores de luminância muito diferentes entre suas bordas, a região de transição entre bordas terá um conteúdo acentuado de altas frequências, o que gerará distorção após filtragem e quantização dessa região [Karlsson-Vetterli (1989)].

Em [Smith-Eddins (1990)] foi proposta uma solução que evita a distorção nas bordas: a "Extensão Simétrica". Ela consiste na colocação de "espelhos" nas bordas da imagem, gerando uma seqüência simétrica ao redor delas, o que evita as transições de alta frequência. Usando filtros QMF de fase linear, como os usados neste trabalho, as operações de filtragem e decimação geram saídas também simétricas. Nova-

mente, não há expansão no número de amostras.

3 Técnicas de Pré-Processamento

Dentre as 10 sub-imagens geradas pela decomposição em sub-bandas, podemos dizer que a sub-banda de número 1 é considerada informação fundamental, pois contém a componente DC do sinal, enquanto as bandas superiores podem ser consideradas de "realce", com importância decrescente com o aumento de frequência.

As regiões de detalhe na imagem original, como bordas ou texturas, se refletem, dentro das sub-bandas, como áreas de alta energia. Os pixels dessas regiões precisam ser bem codificados para uma boa reconstrução do ponto de vista de um observador humano. Por outro lado, as áreas de baixa energia dentro das sub-bandas podem ser codificadas com baixo número de bits (ou mesmo nenhum bit).

Uma técnica usual para se atingir os objetivos acima envolve o uso de algoritmos de alocação de bits [Woods-O'Neil (1986)]. Entretanto, esses algoritmos baseiam-se na minimização de um critério matemático de erro, não utilizando características próprias da percepção humana, como a existência de limiares mínimos de sensibilidade.

Com base nas observações acima, propõe-se a divisão das sub-bandas de números 2 a 10 em pequenos blocos e a eliminação dos blocos que não satisfaçam a um critério de limiares mínimos de sensibilidade pré-determinado. Um mapa de bits para cada sub-banda indica os blocos a serem codificados dentro daquela sub-banda. Os quesitos importantes para um aproveitamento eficiente desta técnica são: baixa complexidade computacional (escolha do critério de seleção), baixo número de bits por pixel para codificar os mapas de bits (escolha do tamanho dos blocos) e qualidade de reconstrução quase transparente (escolha do critério e dos limiares).

Foram considerados dois critérios para seleção dos blocos a serem eliminados: O primeiro é o critério de energia média em que eliminam-se os blocos cuja soma dos quadrados de seus pixels seja menor que um certo limiar, ou seja, se

$$\sum_{i \in B} x_i^2 < L_j \tag{1}$$

então o bloco não é codificado, onde B é o bloco em questão, x_i é um pixel dentro deste bloco e $j \in [2, 10]$ é o índice da sub-banda. O segundo critério é o de quadrado máximo: o bloco é eliminado se todos os quadrados dos pixels forem menores que um limiar, ou seja,

$$\max_i |x_i|^2 < L_j \tag{2}$$

Para escolha do melhor critério, fixou-se inicialmente limiares de sensibilidade para um dos critérios. Em seguida, procedeu-se da seguinte forma: para cada sub-banda, mantendo-se as outras intactas, foram variados os limiares até que a porcentagem de pixels mantidos fosse exatamente igual nos dois critérios. Foram comparados os valores da Razão Pico-Ruído (PSNR) para cada sub-banda, assim como os valores globais, quando o pré-processamento é feito em todas as sub-bandas. Na Tabela 1 vemos os resultados para a imagem iboatsy (576x736). Foram usados blocos de tamanho 4x4. Nota-se que, embora os dois critérios sejam aproximadamente equivalentes, o critério de energia média é ligeiramente superior em termos de PSNR. Além disso, ele apresenta menor variação para diferentes imagens, e por esses motivos foi empregado ao longo deste trabalho.

Sub-Banda	Porcent. de pixels mantidos	PSNR Energia Média	PSNR Quadrado Máximo
1	100.00	66.37	66.37
2	73.67	58.95	58.79
3	76.33	60.46	60.41
4	66.43	59.74	59.47
5	58.57	57.57	57.46
6	64.07	57.92	57.73
7	44.50	55.36	55.08
8	16.03	47.21	46.96
9	38.86	47.46	47.24
10	0.26	47.18	47.17
1-10	29.17	42.00	41.79

Tabela 1: Energia Média x Quadrado Máximo.

Uma vez escolhido o critério de energia média para seleção dos blocos a serem codificados, surge a questão de como determinar dimensões apropriadas para os blocos em cada sub-banda. À medida em que são diminuídas as dimensões dos blocos, são gerados blocos mais homogêneos. Isso diminui o número de pixels a serem codificados e, por conseguinte, a taxa de bits usada para a informação principal. Essa redução de taxa é proporcional ao número de pixels que deixam de ser codificados quando da diminuição do tamanho dos blocos. Por outro lado, a taxa de bits para a informação lateral é aumentada com a diminuição das dimensões dos blocos. Há, então, um compromisso entre a redução de taxa para a informação principal obtida com a redução das dimensões dos blocos e o aumento de taxa para informação lateral. Técnicas para compactação dessa informação serão examinadas na Seção 5.

Se desejarmos altas taxas de compressão para as imagens, como por exemplo taxas de bits da ordem de 0.3 bits/pixel, vê-se que a taxa de bits para codificação da informação lateral deverá ser muito baixa, de forma a permitir uma quantização apropriada da informação principal. Nesse sentido, pensou-se, inicialmente, no uso de blocos 4x4 em todas as sub-bandas, o que resultaria em uma taxa para informação lateral de 0.0625 bit/pixel.

Analisando a Fig. 2, constatamos que as bandas 2, 5 e 8, contêm, principalmente, informações sobre bordas horizontais, e as bandas 3, 6 e 9 informações sobre bordas verticais. Para as primeiras, torna-se interessante o uso de blocos 2x4, e para as outras, blocos 4x2, uma vez que esses formatos de blocos aproveitam de maneira mais eficiente a orientação das bordas. Para as bandas 4, 7 e 10, que contêm informações sobre bordas diagonais, são mantidos os blocos 4x4 a fim de se evitar a complexidade de blocos de forma irregular. A mudança da dimensão dos blocos de 16 para 8 pixels é compensadora se a taxa de bits para cada sub-banda for diminuída. Isso pode ser verificado através da equação 3. O lado esquerdo dessa equação soma a taxa de bits da informação principal com a da informação lateral no caso de blocos 4x4. O lado direito faz o mesmo para blocos 2x4 ou 4x2.

$$\alpha\beta + \frac{f_{16}}{16} \geq \alpha\beta\gamma + \frac{f_8}{8} \quad (3)$$

Nessa equação, α é a porcentagem de pixels mantidos na sub-banda usando-se blocos de dimensão 16, γ é um fator que leva em conta a redução dessa porcentagem quando se usa blocos de dimensão 8, β é o número de bits/pixel usados para codificar os pixels mantidos e f_8 e f_{16} são fatores de compactação para a informação lateral. Supondo que $f_8 = f_{16} = f$ para todas as sub-bandas, teremos que a mudança de dimensão dos blocos é válida se

$$\gamma < 1 - \frac{f}{16\alpha\beta} \quad (4)$$

Valores típicos são: $f=0.5$

Bandas 2 e 3: $\alpha = 0.75$, $\beta = 1.50$, $\gamma < 0.97$

Bandas 5 e 6: $\alpha = 0.65$, $\beta = 1.00$, $\gamma < 0.95$

Bandas 8 e 9: $\alpha = 0.20$, $\beta = 0.50$, $\gamma < 0.69$

Na Tabela 2, podemos ver os resultados de simulação para a imagem barby (576x736). Vê-se que os blocos de dimensão 8 são vantajosos apenas para as bandas 2,3,5,6. Por esse motivo, as dimensões dos blocos de acordo com os indicados na Tabela 3. A taxa total para armazenamento da informação lateral é de 0.071 bit/pixel. Note que embora essa taxa seja um pouco maior que quando se usa apenas blocos 4x4, a taxa total de bits será reduzida.

Sub-Banda	Porcent. de pixels mantidos Dim. 16	Porcent. de pixels mantidos Dim. 8	γ
2	80.91	74.52	0.92
3	92.51	87.44	0.94
5	61.47	57.73	0.94
6	75.36	69.35	0.92
8	14.19	13.31	0.94
9	47.44	43.04	0.91

Tabela 2: Dimensão 16 x Dimensão 8.

Sub-Banda	Dimensões dos blocos	Limiares normalizados
1	-	-
2	2x4	1.2
3	4x2	1.2
4	4x4	1.4
5	2x4	1.8
6	4x2	1.8
7	4x4	2.2
8	4x4	9
9	4x4	9
10	4x4	12

Tabela 3: Dimensões dos blocos por sub-banda.

4 Resultados de simulação

O passo seguinte consistiu na determinação dos limiares para cada sub-banda, a partir de 7 imagens de treinamento. Para cada dessas imagens, procedeu-se da seguinte forma: deixando todas as sub-bandas intactas menos uma, processou-se esta última variando-se o limiar para eliminação de blocos e reconstruindo a imagem. No ponto em que a qualidade de reconstrução deixava de ser transparente ou semi-transparente, interrompia-se o processo. O procedimento foi repetido para todas as sub-bandas (2-10) e imagens. Os limiares para cada sub-banda foram escolhidos como sendo a mediana dos limiares das imagens de treinamento para cada sub-banda. Dessa forma, espera-se que a metade das imagens a serem codificadas o sejam com uma taxa ligeiramente acima do necessário e o contrário aconteça para a outra metade. Na Tabela 3 estão listados os valores dos limiares obtidos.

Na Tabela 4 podemos ver a porcentagem de pixels mantidos por sub-banda e total para uma imagem fora da seqüência de treinamento. A Fig. 3

apresenta a imagem original e a reconstruída. Nota-se que a reconstrução é de excelente qualidade no aspecto subjetivo, tendo sido mantidos apenas 30.9 % dos pixels.

Sub-Banda	Porcentagem dos pixels mantidos
1	100.00
2	82.00
3	81.40
4	72.94
5	64.91
6	61.68
7	46.62
8	32.90
9	25.69
10	0.54
1-10	30.86

Tabela 4: Dimensões dos blocos por sub-banda.

5 Compactação da Informação Lateral

Nessa seção serão examinados três algoritmos para compactação da informação lateral.

O algoritmo do tipo QUAD-TREE [Dash (1993)] é muito utilizado para imagens com grandes áreas planas, como é o caso dos mapas de bits a serem comprimidos. Ele consiste na divisão da imagem em quatro quadrantes. Os sub-quadrantes que forem totalmente planos são marcados como tal. Os outros são novamente divididos, de forma recursiva, em quatro novos quadrantes. O processo é repetido até que não sejam possíveis novas divisões. Os algoritmos do tipo Lempel-Ziv [Ziv-Lempel (1978)] substituem uma sequência encontrada no sinal de entrada por um código que depende de sequências já encontradas e do novo símbolo adicionado. O terceiro tipo de algoritmo considerado foi o de Huffman [Huffman (1952)], com predição linear, em que os símbolos são substituídos por códigos com comprimentos em bits inversamente proporcionais à sua probabilidade de ocorrência. A tabela com os códigos foi gerada a partir das 7 imagens de treinamento, não sendo necessária a sua transmissão.

Na Tabela 5 vemos o resultado da aplicação dos três algoritmos na informação lateral gerada pela imagem Lena (512x512). Vê-se que o fator de compressão aumenta com a sub-banda. Dessa tabela verifica-se também que o algoritmo de Huffman é bastante superior aos demais, reduzindo a taxa ne-

cessária para informação lateral para 0.035 bit/pixel (metade da original). A porcentagem de pixels de informação principal mantidos para essa imagem é de 22%. Na Fig. 4 vemos as imagens original e reconstruída.

Sub-Banda	Algoritmo utilizado e fator de compressão		
	QUAD-TREE	LEMPEL-ZIV	HUFFMAN
2	1.00	1.00	0.87
3	1.00	1.00	0.75
4	1.00	1.00	1.00
5	1.00	0.99	0.76
6	1.00	0.89	0.67
7	1.00	0.90	0.59
8	0.98	0.55	0.43
9	0.99	0.62	0.44
10	0.37	0.34	0.24
2-10	0.86	0.65	0.49

Tabela 5: Compactação da Informação Lateral.

6 Conclusões

Este trabalho apresentou uma técnica de pré-processamento em codificação de imagens por sub-bandas para redução de irrelevância visual. Essa técnica utiliza um critério de limiar de energia média sobre blocos de dimensões distintas para diferentes sub-bandas. Os blocos que não satisfazem tais limiares não são codificados. Os blocos foram dimensionados de modo a obter um balanço adequado entre a taxa necessária para codificar a informação principal e a usada para a informação lateral. Para essa parcela, foram testados três algoritmos de compactação, sendo constatado que o de Huffman apresenta os melhores resultados.

Os benefícios alcançados pela técnica proposta são: baixa complexidade computacional, boa reconstrução perceptual das imagens com baixo número de pixels mantidos (tipicamente de 15 a 50%) e baixa taxa para codificação da informação lateral (tipicamente 0.035 bit/pixel).

7 Referências

N. Jayant, J. Johnston, R. Safranek, Signal Compression Based on Models of Human Perception, *Proceedings of the IEEE* 81 (1993) 1385-1422.

Z. Xie, T. G. Stockham, Jr., Toward the unification of three visual laws and two visual models

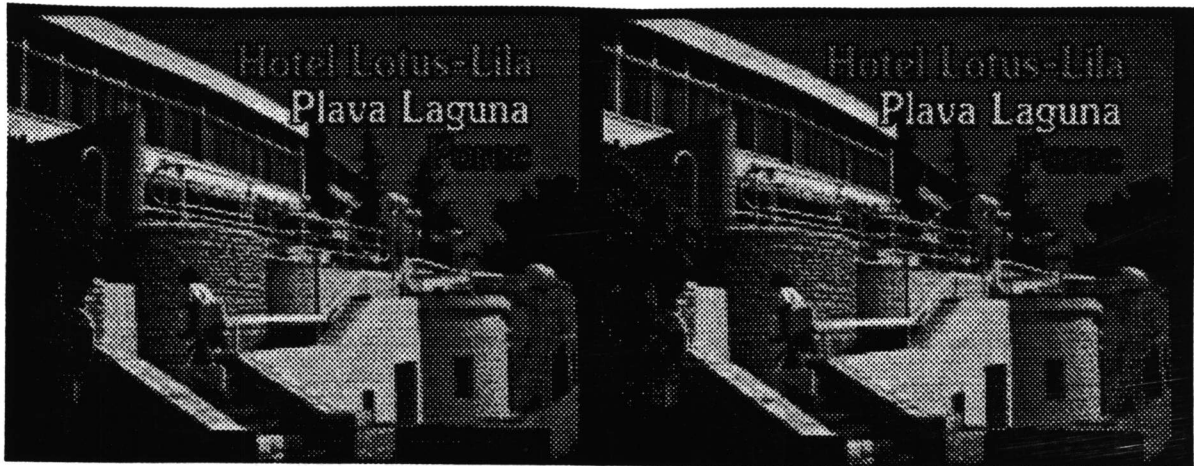


Figura 3: Imagem original e reconstruída (576x736).



Figura 4: Imagem original e reconstruída (512x512).

in brightness perception, *IEEE Transactions on Systems, Man, and Cybernetics* 19 (1989) 379–387.

R. Gray, Vector quantization, *IEEE ASSP Magazine* April (1984) 4–29.

J. Woods, S. D. O'Neil, Subband Coding of Images, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 5 (1986) 1278–1287.

J. Woods, *Subband Image Coding*, Kluwer Academic Publishers, 1991.

T. Senoo, Bernd Girod, Vector quantization for entropy coding of image subbands, *IEEE Transactions on Image Processing* 4 (1992) 526–532.

G. Karlsson, M. Vetterli, Extension of finite length signals for sub-band coding, *Signal Processing* 17 (1989) 161–168.

M. Smith, S. Eddins, Analysis/Synthesis techniques for sub-band image coding, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 8 (1990) 1446–1486.

R. Bamberger, S. Eddins, V. Nuri, Generalized symmetric extension for size-limited multirate filter banks, *IEEE Transactions on Image Processing* 1 (1994) 82–87.

R. Safranek, J. Johnston, A perceptually tuned sub-band image coder with image dependent quantization and pos-quantization data compression, *Proc. ICASSP* (1989).

R. K. Dash, Image processing using quadrees, *Dr. Dobb's Journal* July (1993).

J. Ziv, A. Lempel, Compression of individual se-

quences via variable-rate coding, *IEEE Transactions on Information Theory* 5 (1978) 530-536

D. A. Huffman, A method for the construction of minimum redundancy codes, *Proc. IRE* 40 (1952) 1098-1101