

Introdução à Compressão de Sinais

F. Madeiro e Waslon T. A. Lopes

Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, PE, Brasil.

Universidade Federal de Campina Grande, Campina Grande, PB, Brasil.

E-mail: madeiro,waslon@iecom.org.br

Resumo—Este artigo apresenta fundamentos básicos de compressão de sinais. São apresentados e discutidos os parâmetros de desempenho de um sistema de compressão de sinais, a saber: qualidade dos sinais reconstruídos (distorção), taxa de bits, complexidade e retardo. É definida, com brevidade, uma operação importante em sistemas de compressão de sinais: a quantização. Nesse contexto, o artigo apresenta uma visão geral da quantização vetorial. O uso de transformadas também é abordado. Nesse cenário, são apresentados fundamentos da transformada *wavelet* discreta e discutidas questões relacionadas à aplicação em compressão de imagens. Por fim, são apresentadas algumas direções de pesquisa na área de compressão de sinais.

I. INTRODUÇÃO

DENTRE as vantagens decorrentes da representação digital de sinais, destacam-se o controle eficiente de erros de canal, a aplicação de técnicas criptográficas e a possibilidade de integração de sinais e dados.

A codificação eficiente de sinais (voz, imagem, áudio, vídeo) é essencial para diversas aplicações, tais como, comunicação multimídia, navegação *web*, armazenamento de imagens médicas e documentos de negócios, arquivamento de impressões digitais, transmissão de imagens de sensoriamento remoto obtidas de satélites e comunicações móveis. Nessas aplicações, o propósito fundamental da compressão de sinais é reduzir o número de bits para representá-los adequadamente, de forma a minimizar os requisitos de memória, para armazenamento, e de largura de banda, para transmissão.

Um sinal analógico pode ser convertido para a forma digital por meio da combinação de três operações básicas: amostragem, quantização e codificação [1].

Na operação de amostragem, são tomados apenas valores de amostras do sinal analógico situadas em instantes de tempo uniformemente espaçados. O teorema da amostragem estabelece que para sinais limitados em faixa, por exemplo em B Hz, a taxa de amostragem f_s (em amostras por segundo) deve ser maior ou igual a $2B$. Assim, o intervalo de amostragem T deve ser: $T = \frac{1}{f_s} \leq \frac{1}{2B}$. No processo de quantização, cada valor de amostra é aproximado pelo nível mais próximo dentre um conjunto finito de níveis discretos. Na etapa de codificação, cada nível de quantização é representado por uma palavra-código, que consiste de uma sequência de dígitos binários.

A taxa de transmissão, I , de um sistema de comunicações digitais é o produto da taxa de amostragem, f_s , pelo número de bits por amostra, R , que o sistema utiliza para representar as amplitudes das amostras [2]:

$$I = f_s R \text{ bits por segundo (bit/s, ou bps)} \quad (1)$$

Muito frequentemente, quando B e f_s são conhecidos e fixos, a taxa de transmissão é simplesmente especificada pela taxa

$$R \text{ bits por amostra (bit/a)} \quad (2)$$

A Figura 1 ilustra o princípio da compressão de sinais. Observa-se que a informação dos sinais de interesse é composta das seguintes partes [2]:

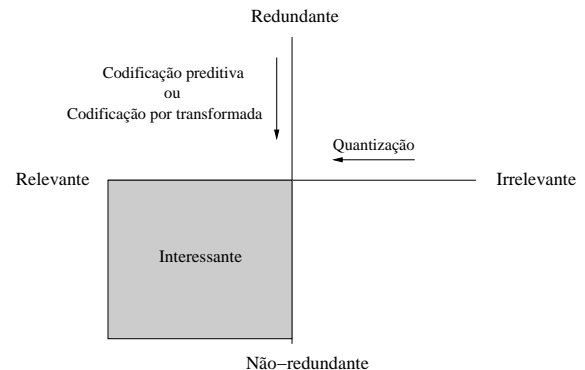


Figura 1. Princípio da compressão de sinais [2].

- componente relevante;
- componente irrelevante, correspondente à parcela supérflua da informação, de que o destino não necessita;
- componente não-redundante;
- componente redundante, que não necessita ser transmitida, uma vez que o receptor tem condições de reconstituí-la automaticamente.

O alvo das técnicas de compressão de sinais é reduzir as componentes redundante e irrelevante da informação, transmitindo apenas o que é imprescindível para o receptor, ou seja, a parte interessante da informação. A componente irrelevante da informação é reduzida com uso de quantização, que introduz perda da informação, de caráter irreversível. Para o propósito de redução da componente redundante da informação, são utilizadas técnicas de predição e de transformação do sinal, que apresentam caráter reversível.

Muito embora alguns sistemas não apresentem grandes limitações de largura de faixa, como é o caso das redes de comunicações por fibra ótica, e apesar de a evolução tecnológica estar continuamente contribuindo para o surgimento de memórias com grandes capacidades de armazenamento, a compressão de sinais desempenha um papel importantíssimo, devido a uma série de fatores [2], [3], tais como: o intenso

tráfego multimídia tem levado ao aumento da demanda no tocante ao armazenamento de voz, música, imagens, vídeo; um maior número de canais de comunicações pode ser multiplexado em sistemas de faixa larga, com uso de técnicas de compressão para reduzir os requisitos de largura de faixa de cada sinal a ser multiplexado; em sistemas de resposta vocal, vocabulários maiores podem ser armazenados com uso de técnicas de compressão.

O artigo encontra-se organizado de acordo com as seções a seguir. A Seção II apresenta uma visão geral de um sistema de comunicação digital. A Seção III contempla, com brevidade, a amostragem de sinais. As Seções IV e V são dedicadas respectivamente à quantização escalar e à quantização vetorial. Parâmetros de desempenho de um sistema de compressão de sinais são apresentados e discutidos na Seção VI. O uso transformadas no âmbito de compressão de sinais é o tema da Seção VII. A transformada *wavelet* discreta, muito usada em diversos sistemas de compressão, é abordada na Seção VIII. Considerações finais e algumas direções de pesquisa são apresentadas na Seção IX.

II. CODIFICAÇÃO DE SINAL – PRELIMINARES

Os termos codificação de fonte e compressão de sinais são utilizados para referir-se a técnicas usadas para obter uma representação digital compacta de um sinal. A codificação de fonte constitui parte importante de um sistema de comunicação digital, apresentado na Figura 2. Enquanto o codificador de fonte tem como alvo minimizar a taxa de bits necessária para a representação adequada de um sinal de entrada, o *modulador-demodulador (modem)* visa maximizar a taxa de bits que pode ser sustentada em um dado canal ou meio de armazenamento sem causar um nível inaceitável de probabilidade de erro de bit. Na codificação de fonte, a taxa de bits é medida em bits por amostra ou bits por segundo (usualmente denotada por b/s ou ainda por bps). Na modulação, é medida em bits por segundo por Hertz ($b/s/Hz$). Na codificação de canal é adicionada redundância à sequência de bits visando proteção contra erros.

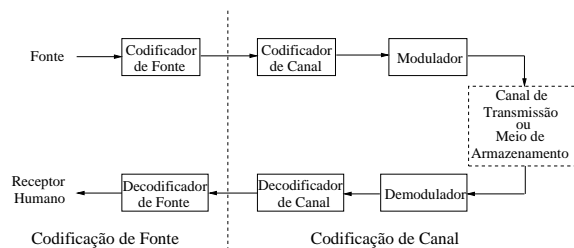


Figura 2. Diagrama de blocos de um sistema de comunicação digital.

O foco deste artigo é a codificação de fonte. Dependendo dos requisitos da aplicação, há duas classes a considerar: codificação sem perda e codificação com perda. Na primeira, também denominada compactação, não há introdução de distorção no sinal decodificado, ou seja, o sinal reconstruído é uma réplica do sinal original. Na segunda, o processo de codificação introduz degradação no sinal, ou seja, o sinal reconstruído é uma versão degradada do sinal original.

III. AMOSTRAGEM

Por meio do processo de amostragem, um sinal contínuo no tempo é transformado em um sinal discreto no tempo.

O Teorema da Amostragem, de Nyquist, que é a base para determinar a taxa de amostragem adequada para um dado sinal, é enunciado a seguir: “Um sinal $f(t)$ limitado em faixa em B Hz (ou seja, $F(w) = 0$ para $|w| > 2\pi B$) pode ser reconstruído exatamente (sem erro) a partir de suas amostras tomadas a uma taxa maior ou igual a $2B$ Hz (amostras por segundo). Em outras palavras, a frequência de amostragem mínima é $f_s = 2B$ Hz”.

Se a frequência de amostragem for inferior à frequência de Nyquist, o sinal não poderá ser recuperado completamente a partir de suas amostras, pois haverá superposição espectral, resultando distorção nas frequências mais altas. Esse fenômeno é conhecido como *aliasing*. Por outro lado, um aumento na frequência de amostragem, além da frequência de Nyquist, implica separação entre os espectros superior à necessária para a recuperação do sinal.

As seções a seguir abordam a quantização, que constitui operação relevante em sistemas de compressão de sinais.

IV. QUANTIZAÇÃO ESCALAR

Quantização é o processo por meio do qual realiza-se a digitalização dos valores de amostras de um sinal. Em outras palavras, o alvo da quantização é a discretização dos valores de amostras. Assim, por exemplo, quando as amostras de um sinal de voz são submetidas a um quantizador escalar com quatro níveis de quantização, obtém-se, na saída do quantizador, um sinal tal que cada amostra apresenta um dentre quatro possíveis valores. Cada um desses valores pode ser codificado, por exemplo, com uma palavra-binária de dois bits. A quantização, no exemplo dado, consiste, basicamente, em substituir cada valor de amostra do sinal original por um dos quatro valores: o nível de quantização é escolhido de tal maneira que a diferença entre o valor original e valor quantizado seja minimizada.

A função do quantizador é realizar um mapeamento $Q: \mathcal{R} \rightarrow \mathcal{C}$, em que \mathcal{C} é um subconjunto do espaço Euclidiano unidimensional \mathcal{R} . O número de elementos de \mathcal{C} é denominado número de níveis do quantizador, denotado por L . Ao final do processo de quantização, cada amostra poderá ser codificada, por exemplo, com uma palavra-binária de $n = \log_2 L$ bits. Daí falar-se em quantizador de n bits. Observa-se que, à medida que o número de níveis de quantização aumenta, o sinal quantizado (saída do quantizador) aproxima-se do sinal original (entrada do quantizador). Em outras palavras, à medida que a taxa de bits (ou número de bits por amostra), n , do quantizador aumenta, a distorção introduzida pela quantização diminui, em virtude de diminuir a diferença entre o sinal original e sua versão quantizada. Uma questão relevante em compressão de sinais é o compromisso entre taxa e distorção. Diz-se que um quantizador Q_1 é melhor que um quantizador Q_2 quando, fixada uma taxa de bits, Q_1 leva a um sinal quantizado mais próximo do sinal original (menor distorção) que o obtido por Q_2 , ou, alternativamente, quando, fixada uma distorção, Q_1 requer uma menor taxa de bits (um menor número de níveis de quantização) que a requerida por Q_2 .

V. QUANTIZAÇÃO VETORIAL

A quantização vetorial [4], que pode ser vista como uma extensão da quantização escalar em um espaço multidimensional, encontra-se fundamentada na Teoria da Distorção Versus Taxa, formulada por Shannon, segundo a qual um melhor desempenho é obtido codificando blocos de amostras (isto é, vetores) em vez de amostras individuais (isto é, escalares). Em diversos sistemas de codificação de voz e imagem, a quantização vetorial tem sido utilizada com sucesso, permitindo elevadas taxas de compressão.

Matematicamente, a quantização vetorial consiste de um mapeamento Q de um vetor de entrada \vec{x} pertencente ao espaço Euclidiano K -dimensional, \mathcal{R}^K , em um vetor pertencente a um subconjunto finito W de \mathcal{R}^K , ou seja,

$$Q : \mathcal{R}^K \rightarrow W, \quad (3)$$

O dicionário $W = \{\vec{w}_i; i = 1, 2, \dots, N\}$ é o conjunto de vetores de reprodução (também denominados vetores-código ou vetores de reconstrução), K é a dimensão do quantizador vetorial e N é o número de vetores-código (ou número de níveis, em analogia com a quantização escalar).

O mapeamento Q introduz um particionamento de \mathcal{R}^K em N células (denominadas regiões de Voronoi) S_i , $i = 1, 2, \dots, N$, tais que:

$$\bigcup_{i=1}^N S_i = \mathcal{R}^K \text{ e } S_i \cap S_j = \emptyset \text{ para } i \neq j, \quad (4)$$

em que cada célula S_i é definida da seguinte maneira:

$$S_i = \{\vec{x} : Q(\vec{x}) = \vec{w}_i\}. \quad (5)$$

Deste modo, o vetor-código \vec{w}_i constitui o vetor representativo de todos os vetores de entrada pertencentes à célula S_i .

Em um sistema de compressão de sinais baseado em quantização vetorial simples, apresentado na Figura 3, um quantizador vetorial pode ser visto como a combinação de duas funções: um codificador de fonte e um decodificador de fonte. Dado um vetor $\vec{x} \in \mathcal{R}^K$, da fonte a ser comprimida, o codificador calcula a distorção $d(\vec{x}, \vec{w}_i)$ entre o vetor de entrada (vetor a ser quantizado) e cada vetor-código \vec{w}_i , $i = 1, 2, \dots, N$ do dicionário W . A regra ótima para codificação é a regra do vizinho mais próximo [5], na qual uma representação binária $b_I = \lceil \log N \rceil$ do índice I é transmitida ao decodificador de fonte se o vetor-código \vec{w}_I corresponder à menor distorção [6]. Em outras palavras, o codificador usa a regra de codificação $\text{enc}(\vec{x}) = I$ se $d(\vec{x}, \vec{w}_I) < d(\vec{x}, \vec{w}_i)$, $\forall i \neq I$. Ao receber a representação binária b_I do índice I , o decodificador de fonte, que dispõe de uma cópia do dicionário W , simplesmente procura pelo I -ésimo vetor-código e produz o vetor \vec{w}_I como a reprodução (versão quantizada) de \vec{x} . Em outras palavras, é utilizada a seguinte regra de decodificação: $\text{dec}(I) = \vec{w}_I$ [7].

A quantização vetorial, portanto, constitui uma técnica de compressão com perdas, visto que o sinal reconstruído é uma versão degradada do sinal original. O erro médio de quantização ao se representar o sinal de entrada por sua versão quantizada é chamado distorção do quantizador. Por outro lado, a taxa de codificação do quantizador vetorial, que

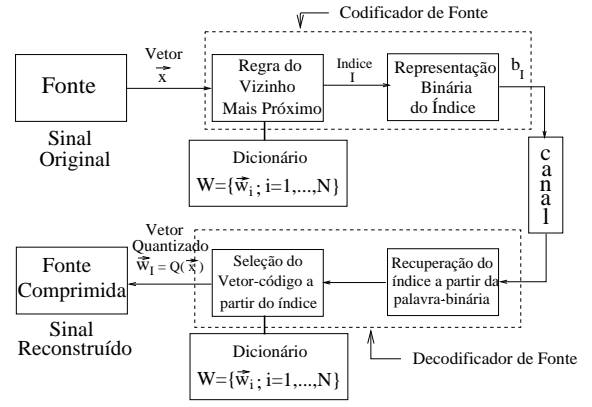


Figura 3. Sistema de compressão de sinais baseado em quantização vetorial simples.

mede o número de bits por componente do vetor, é dada por $R = \frac{1}{K} \log_2 N$. Em codificação de forma de onda de voz, R é expressa em bit/amostra. Em se tratando de codificação de imagens, R é expressa em bits por pixel (bpp).

Uma das questões principais no projeto de quantizadores vetoriais é o compromisso entre taxa e distorção. O alvo a ser perseguido é a obtenção de um dicionário ótimo, que minimize a distorção média introduzida pela aproximação de cada vetor de entrada por um dos vetores-código. O quantizador ótimo deve satisfazer duas condições necessárias: a condição de vizinho mais próximo para partição e a condição de centróide para os vetores-código [3], [8].

Existem diversas técnicas para projeto de dicionários para quantização vetorial, dentre as quais podem ser citadas: algoritmo LBG (Linde-Buzo-Gray) [5]; algoritmo auto-organizativo de Kohonen [9] e outros algoritmos não-supervisionados de redes neurais [10]; *stochastic relaxation* [8]; algoritmos *fuzzy* [11]; algoritmos genéticos [12]; algoritmo PNN (*pair-wise nearest neighbor*) [13].

VI. DESEMPENHO DE UM SISTEMA DE COMPRESSÃO DE SINAIS

O problema geral em compressão de sinal é minimizar a taxa de bits na representação digital do sinal, mantendo os níveis requeridos de qualidade do sinal, complexidade de implementação e retardo de comunicação. Cada um desses aspectos é abordado a seguir.

A. Qualidade de Sinais Reconstruídos

De forma geral, as medidas utilizadas para avaliação da qualidade de sinais enquadram-se em duas classes: medidas de qualidade subjetivas e medidas de qualidade objetivas (e.g. [14]). As primeiras baseiam-se em comparações (realizadas por meio de testes de escuta ou de visualização), entre o sinal original e o sinal processado, realizadas por um grupo de pessoas, que subjetivamente classificam a qualidade do sinal processado segundo uma escala pré-determinada. As medidas objetivas, por sua vez, baseiam-se numa comparação matemática direta entre os sinais original e processado [15].

Medidas de qualidade subjetivas são utilizadas para avaliar de forma plena, definitiva, a qualidade de algoritmos/técnicas

de codificação de sinais. Os testes subjetivos, contudo, são de difícil realização, uma vez que requerem a disponibilidade de um grande número de avaliadores (entre leigos, especialistas e possíveis usuários do sistema de codificação de sinais), envolvem grande volume de sinal processado e necessitam da disponibilidade de laboratórios com condições adequadas ao desenvolvimento das avaliações subjetivas, enfim, são bastante dispendiosos de tempo, implicando custo elevado de realização.

A seguir serão abordadas algumas medidas de qualidade utilizadas para a avaliação de técnicas de compressão de sinais de voz e imagem. Serão descritos o escore médio de opinião (MOS, *mean opinion score*) e os testes de preferência, que constituem alternativas para avaliação de qualidade subjetiva. Também serão apresentadas a relação sinal-ruído (SNR, *signal-to-noise ratio*), que é uma medida objetiva utilizada para avaliação de qualidade de sinais de voz, e a relação sinal-ruído de pico (PSNR, *peak signal-to-noise ratio*), que constitui uma medida objetiva amplamente utilizada para avaliação da qualidade de imagens reconstruídas.

1) *Escore Médio de Opinião*: Uma medida subjetiva bastante utilizada para avaliação de desempenho de sistemas de compressão de voz e imagem denomina-se escore médio de opinião (MOS). No teste MOS, cada avaliador atribui um escore de qualidade ao sinal reconstruído, segundo a escala graduada apresentada na Tabela I [15].

É calculada a média aritmética dos escores obtidos e determinado o valor final da avaliação, ou seja,

$$MOS = \frac{1}{L} \sum_{l=1}^L s_l, \quad (6)$$

em que L é o número de avaliadores utilizados no teste e s_l é o escore atribuído pelo l -ésimo avaliador.

Tabela I
ESCALA PARA O TESTE MOS.

Escore (s)	Qualidade
5	Excelente
4	Boa
3	Razoável
2	Pobre
1	Ruim

Em se tratando de codificação de voz, avaliações por meio de MOS são bem aceitas e algumas vezes complementadas com medidas de inteligibilidade, como, por exemplo, MRT (*modified rhyme test*) e DRT (*diagnostic rhyme test*) [15].

2) *Testes de Preferência*: Os testes de preferência são realizados por comparação entre pares de sinais. Uma forma de realização desse tipo de avaliação subjetiva consiste em atribuir um conceito de acordo com três possíveis resultados de comparação, ou seja [16]:

- Conceito A - A qualidade do primeiro sinal é melhor do que a do segundo;
- Conceito B - A qualidade do segundo sinal é melhor do que a do primeiro;
- Conceito C - A qualidade de ambos sinais não se distingue.

3) *Relação Sinal-Ruído (SNR)*: Sejam $x(n)$ o sinal original, $y(n)$ o sinal processado e $e(n) = x(n) - y(n)$ o sinal erro no instante de tempo n .

A energia contida no sinal original é

$$E_x = \sum_n x^2(n). \quad (7)$$

A energia contida no sinal erro é

$$E_e = \sum_n e^2(n) = \sum_n [x(n) - y(n)]^2. \quad (8)$$

A medida SNR resultante, expressa em dB, é dada por

$$SNR = 10 \log_{10} \frac{E_x}{E_e} = 10 \log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - y(n)]^2}. \quad (9)$$

Como outras medidas de avaliação de qualidade objetiva de voz, podem ser citadas a relação sinal-ruído segmental, a LAR (*log-area ratio*), a medida *log-likelihood* de Itakura [15] e a distorção espectral.

4) *Relação Sinal-Ruído de Pico (PSNR)*: Dentre as diversas medidas objetivas utilizadas para avaliação de qualidade de imagens [14], a relação sinal-ruído de pico (ou razão pico-ruído) apresenta-se como a opção mais utilizada, apesar de serem frequentemente registradas críticas, como relata por exemplo [17], no que diz respeito à correlação de PSNR com resultados de avaliações subjetivas.

A relação sinal-ruído de pico é definida como 10 vezes o logaritmo na base 10 da razão entre o quadrado do valor de pico da amplitude do sinal de entrada, v_p^2 , e o erro médio quadrático (MSE, *mean square error*):

$$PSNR = 10 \log_{10} \left[\frac{v_p^2}{MSE} \right]. \quad (10)$$

Para o caso de uma imagem original codificada a 8,0 bpp (256 níveis de cinza),

$$PSNR = 10 \log_{10} \left[\frac{255^2}{MSE} \right], \quad (11)$$

em que o erro médio quadrático entre as imagens original e reconstruída é definido como

$$MSE = \frac{1}{L \cdot C} \sum_{l=1}^L \sum_{c=1}^C [F(l, c) - \hat{F}(l, c)]^2, \quad (12)$$

em que $F(l, c)$ e $\hat{F}(l, c)$ representam os valores de pixels das imagens original e reconstruída, l designa a l -ésima linha e c denota a c -ésima coluna de uma imagem (matriz) $L \times C$.

B. Taxa de Bits

Um importante parâmetro de algoritmos e sistemas de compressão é a taxa de bits, geralmente expressa em bits por amostra, bits por pixel (bpp), bits por segundo, dependendo do contexto. A taxa de bits por segundo é simplesmente o produto da taxa de amostragem e o número de bits por amostra. A taxa de amostragem é, geralmente, ligeiramente superior a duas vezes a largura de faixa do sinal, conforme estabelecido pelo teorema da amostragem de Nyquist.

A Tabela II ilustra alguns formatos comumente utilizados para áudio [18]. São utilizadas taxas de amostragem típicas de 8 kHz para voz telefônica, 16 kHz para áudio AM, 32 kHz para áudio FM, e 44,1 kHz ou 48 kHz para CD (*compact-disk*) ou áudio DAT (*digital audio tape*), sendo ambos sinais de largura de faixa 20 kHz. Observe que as larguras de faixa respectivas são menores que metade da taxa de amostragem correspondente, seguindo o princípio da amostragem de Nyquist.

Tabela II
FORMATOS DE ÁUDIO DIGITAL.

Formato	Taxa de Amostragem (kHz)	Largura de Faixa (kHz)	Faixa de Frequência
Telefonia	8	3,2	200-3400 Hz
Teleconferência	16	7	50-7000 Hz
Compact Disk (CD)	44,1	20	20-20000 Hz
Digital Audio Tape (DAT)	48	20	20-20000 Hz

C. Complexidade

A complexidade de um algoritmo de codificação está relacionada ao esforço computacional requerido para implementar os processos de codificação e decodificação. Pode ser medida em MIPS (milhões de instruções por segundo) [16]. Outras medidas relacionadas à complexidade são o tamanho físico do codificador, decodificador ou *codec* (codificador mais decodificador), seu custo e o consumo de potência (medida, por exemplo, em milliwatt, mW), sendo este um critério particularmente importante para sistemas portáteis [18]

D. Retardo de Comunicação

O aumento de complexidade em um algoritmo de codificação é geralmente associado a um aumento de atrasos de processamento no codificador e decodificador. A importância do retardo em um sistema de comunicação depende da aplicação. Dependendo do ambiente de comunicação, o atraso total tem que ser mantido em um limite severo, como no caso da utilização em redes telefônicas [18]. No entanto, o retardo de comunicação pode ser visto como irrelevante em aplicações que envolvem comunicação unidirecional [16].

VII. USO DE TRANSFORMADAS NO ÂMBITO DA COMPRESSÃO DE SINAIS

A codificação por transformada é uma estratégia que tem sido amplamente utilizada no cenário de compressão de sinais. A codificação é realizada no domínio da transformada em vez de ser levada a efeito no domínio original do sinal. Em outras palavras, o propósito é quantizar e codificar os coeficientes da transformada em vez dos valores originais das amostras do sinal. Duas transformadas amplamente utilizadas são a DCT (*discrete cosine transform*), usada no codificador JPEG com perda e no MPEG, e a DWT (*discrete wavelet transform*), padronizada no MPEG-4 e no JPEG2000.

Em compressão de vídeo, a codificação por transformada geralmente é combinada com técnicas de previsão (compensação de movimento).

VIII. TRANSFORMADA WAVELET DISCRETA

A transformada *wavelet* discreta (DWT) tem sido utilizada com sucesso em diversas aplicações [19], dentre as quais podem ser citadas: supressão de ruído, solução de equações diferenciais parciais, codificação de canal e compressão de sinais de áudio, vídeo, voz e imagem.

Conforme mostra a Figura 4, a transformada *wavelet* discreta [19] pode ser descrita a partir de um banco de filtros. Os filtros $h_0(n)$ e $h_1(n)$ correspondem aos filtros *wavelet* de análise, enquanto que $g_0(n)$ e $g_1(n)$ correspondem aos filtros *wavelet* de síntese. Um sinal de entrada $x(n)$ é convoluído com o filtro passa-baixa $h_0(n)$ e com o filtro passa-alta $h_1(n)$. O sinal resultante de cada convolução é então submetido a uma decimação (sub-amostragem) de ordem 2. Geram-se, assim, no processo de decomposição (ou análise), o sinal de aproximação $a_1(n)$ e o sinal de detalhe $d_1(n)$. Em outras palavras, $a_1(n)$ contém os coeficientes *wavelet* correspondentes à componente de aproximação do sinal, ao passo que $d_1(n)$ contém os coeficientes *wavelet* referentes à componente de detalhe do sinal. No processo de reconstrução (ou síntese), procede-se uma interpolação (super-amostragem) de ordem 2, seguida de convolução com os filtros passa-baixa $g_0(n)$ e passa-alta $g_1(n)$. Satisfeitas as condições de reconstrução perfeita, o sinal $x(n)$ é reconstruído de forma exata, como a soma das saídas dos filtros *wavelet* de síntese, ou seja, $y(n) = x(n)$.

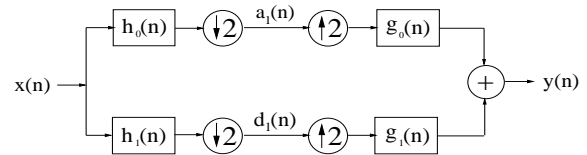


Figura 4. Estrutura básica do banco de filtros e das operações de decimação e interpolação envolvidos na DWT.

Em se tratando de uma decomposição *wavelet* multiresolucional (multinível) [20], o próximo nível de decomposição (resolução mais grosseira) é obtido mediante a convolução do sinal $a_1(n)$ com um par idêntico de filtros $h_0(n)$ e $h_1(n)$, seguida da decimação de ordem 2. Geram-se, assim, os sinais $a_2(n)$ e $d_2(n)$, onde o índice 2 diz respeito ao segundo nível de decomposição. Esse processo pode ser repetido iterativamente, e o número de pares de filtros *wavelet* de análise define o número de níveis de decomposição (ou número de níveis de resolução), denotado por \mathcal{L} , da transformada *wavelet* discreta. Deste modo, são obtidos, ao final de \mathcal{L} níveis de decomposição, um sinal de aproximação $a_{\mathcal{L}}(n)$, com resolução reduzida por um fator de $2^{\mathcal{L}}$ em relação ao sinal de entrada, e os sinais de detalhe $d_{\mathcal{L}}(n)$, $d_{\mathcal{L}-1}(n)$, \dots , $d_1(n)$. Em outras palavras, $d_i(n)$, com $1 \leq i \leq \mathcal{L}$, contém os coeficientes *wavelet* correspondentes às componentes de detalhe no i -ésimo nível de decomposição, enquanto que $a_{\mathcal{L}}(n)$ contém os coeficientes *wavelet* referentes à componente de aproximação no \mathcal{L} -ésimo nível de decomposição. A Figura 5(a) ilustra a estrutura piramidal de uma decomposição *wavelet* de 3 níveis.

Para se realizar a etapa de síntese, ilustrada na Figura 5(b), os sinais $a_i(n)$ e $d_i(n)$, após interpolação de ordem 2, são submetidos, respectivamente, aos filtros *wavelet* de síntese $g_0(n)$ e $g_1(n)$ para permitirem a reconstrução de $a_{i-1}(n)$, que constitui a componente de aproximação no próximo nível de resolução mais alta. Uma vez asseguradas as condições de reconstrução perfeita, $y(n) = x(n)$. O conjunto de operações (convolução e interpolação) que permitem a reconstrução do sinal a partir dos coeficientes *wavelet* $a_L(n)$ e $d_i(n)$, com $1 \leq i \leq L$, constitui a transformada *wavelet* discreta inversa (IDWT, *inverse discrete wavelet transform*).

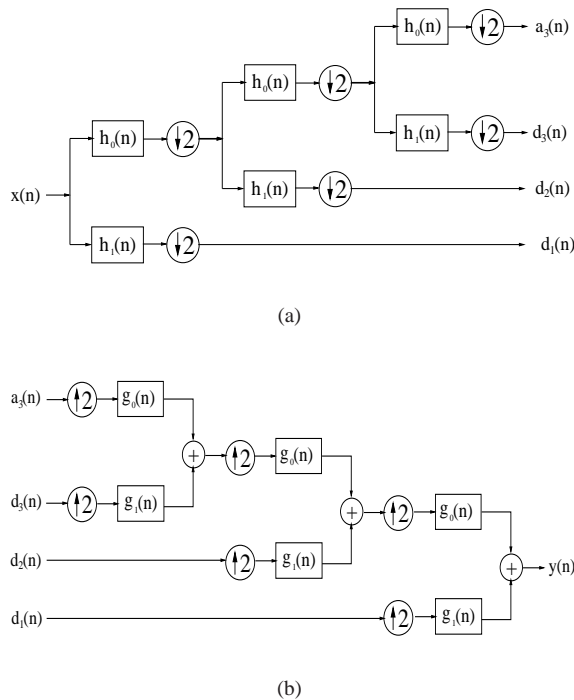


Figura 5. Estrutura piramidal de uma decomposição *wavelet* multiresolucional em 3 níveis: (a) operação de análise ou decomposição; (b) operação de síntese ou reconstrução.

A. Transformada Wavelet Discreta Aplicada a Imagens

Conforme abordado na seção anterior, o princípio básico da transformada *wavelet* discreta (DWT) é a decomposição de um sinal em um conjunto de componentes de aproximação e de detalhe, mediante uso de bancos de filtros *wavelet*. Cada uma destas componentes corresponde a uma faixa de frequências (sub-banda).

Nesta seção, procurando obedecer à notação mais comumente utilizada em transformada *wavelet* aplicada à imagem, o filtro *wavelet* passa-baixa de análise $h_0(n)$ será simplesmente denotado por L (*low-pass filter*), enquanto que o filtro *wavelet* passa-alta de análise $h_1(n)$ será simplesmente denotado por H (*high-pass filter*).

A abordagem mais utilizada para a decomposição *wavelet* multiresolucional hierárquica foi introduzida por Mallat [21]. A transformada *wavelet* discreta é implementada por meio da aplicação de um filtro passa-baixa, L , e de um filtro passa-alta, H , nas direções horizontal e vertical da imagem. Em

seguida, as saídas dos filtros são sub-amostradas por um fator de 2, gerando as sub-bandas LH_1 , HL_1 , HH_1 e LL_1 no nível de decomposição (nível de resolução) 1, que corresponde à escala mais fina. Para se obterem as próximas sub-bandas (em uma escala mais grosseira), o processo é repetido na sub-banda passa-baixa LL_1 para gerar o próximo nível (resolução, escala) da decomposição, ou seja, o nível de resolução 2, com as correspondentes sub-bandas LH_2 , HL_2 , HH_2 e LL_2 . Para se obterem as sub-bandas do próximo nível de resolução, o processo é repetido em LL_2 . O processo continua até que o nível de resolução desejado seja obtido. A Figura 6 mostra uma decomposição *wavelet* de 3 escalas (3 níveis de resolução, decomposição em 3 níveis) de uma imagem 256×256 .

Neste trabalho, é adotada a notação apresentada na Tabela III, em que S_{ij} denota a sub-banda correspondente à j -ésima direção do i -ésimo nível de decomposição. Deste modo, $j = 1$ identifica a direção horizontal (HL), $j = 2$ identifica a direção vertical (LH) e $j = 3$ identifica a direção diagonal (HH), ao passo que $j = 0$ identifica a sub-banda (componente de aproximação) correspondente à resolução mais baixa. Com a notação adotada neste trabalho, portanto, as componentes de detalhe são identificadas por $j = 1, 2, 3$.

A Figura 7 apresenta a disposição das sub-bandas obtidas em uma decomposição *wavelet* de 3 níveis de uma imagem 256×256 .

Tabela III
NOTAÇÃO UTILIZADA.

	Sub-banda	Nível	Direção
HL_1	S_{11}	1	1 (Horizontal)
LH_1	S_{12}	1	2 (Vertical)
HH_1	S_{13}	1	3 (Diagonal)
HL_2	S_{21}	2	1 (Horizontal)
LH_2	S_{22}	2	2 (Vertical)
HH_2	S_{23}	2	3 (Diagonal)
HL_3	S_{31}	3	1 (Horizontal)
LH_3	S_{32}	3	2 (Vertical)
HH_3	S_{33}	3	3 (Diagonal)
LL_3	S_{30}	3	0

A seguir, será apresentada uma breve abordagem de alguns aspectos relacionados à aplicação da DWT em codificação de imagens.

B. Codificação de Imagens Utilizando DWT

A transformada *wavelet* discreta (DWT) apresenta características que justificam sua adequação (e, conseqüentemente, sua ampla utilização) em compressão de imagens, como por exemplo [22]:

- a DWT é uma descrição multiresolucional de uma imagem, de modo que é possível proceder uma decodificação sequencial, de um nível de resolução muito baixo a um nível de resolução mais alto;
- a DWT é bem mais próxima do sistema visual humano que a transformada discreta de cosseno (DCT), a qual é bastante utilizada em vários sistemas de codificação de imagens, como por exemplo o JPEG (*Joint Photographics Experts Group*);

LL_3 32×32	HL_3 32×32	HL_2 64×64	HL_1 128×128 Horizontal
LH_3 32×32	HH_3 32×32		
LH_2 64×64		HH_2 64×64	
LH_1 128×128 Vertical		HH_1 128×128 Diagonal	

Figura 6. Decomposição *wavelet* de 3 níveis de uma imagem 256 × 256. A filtragem passa-baixa é denotada por *L* e a passa-alta por *H*.

S_{30} 32×32	S_{31} 32×32	S_{21} 64×64	S_{11} 128×128 Horizontal
S_{32} 32×32	S_{33} 32×32		
S_{22} 64×64		S_{23} 64×64	
S_{12} 128×128 Vertical			

Figura 7. Sub-bandas obtidas em uma decomposição *wavelet* de 3 níveis de uma imagem 256 × 256. A sub-banda S_{ij} diz respeito à j -ésima direção do i -ésimo nível de decomposição, $i, j \in \{1, 2, 3\}$. Observe que $j = 0$ diz respeito à sub-banda correspondente a *LL*.

- as imagens reconstruídas por meio da DWT não apresentam a incômoda distorção sob a forma de bloqueamento da imagem, típica de técnicas de codificação envolvendo transformada DCT ou quantização vetorial a elevadas taxas de compressão.
- a representação hierárquica da imagem, em níveis de resolução distintos, sob a forma de componentes de aproximação e de detalhe, contriuiu para o desenvolvimento de estratégias de quantização mais “sintonizadas” com as características do sistema visual humano;
- em virtude de permitir uma representação compacta da energia da imagem (grande parte da energia se concentra em um pequeno número de coeficientes *wavelet*), a DWT

favorece a concepção de quantizadores (escalares ou vetoriais) eficientes, que procuram aproveitar a estatística dos coeficientes em cada sub-banda da decomposição *wavelet*.

Diversos fatores afetam o desempenho dos sistemas de codificação de imagens que utilizam DWT, dentre os quais podem ser citados: a escolha das bases (famílias ou filtros) *wavelet*, o tipo de extensão utilizada ao se aplicar a DWT, a estratégia de quantização/alocação de bits adotada, a utilização de codificadores entrópicos e a concepção/aplicação de medidas de distorção sintonizadas com critérios de percepção visual. Em seguida, cada um destes aspectos será brevemente abordado.

A escolha da base *wavelet* para utilização em codificação de imagens constitui objeto de interesse de diversas pesquisas, sendo abordada, sucinta ou profundamente, em vários trabalhos [23], [24]. Importantes características [19], [20] (que influenciam a qualidade da imagem reconstruída) devem ser observadas, como, por exemplo: suavidade, regularidade, tamanho de suporte, seletividade em frequência do filtro, precisão de aproximação e número de momentos de decaimento (ou *vanishing moments*).

O tipo de extensão da imagem [19], [25] (necessária para se efetuar a convolução com os filtros *wavelet*) utilizada ao se aplicar a DWT pode afetar a qualidade da imagem reconstruída, obtida após aplicação da IDWT. Diversos tipos de extensão têm sido utilizados, como, por exemplo: a extensão do tipo *zero-padding*, a extensão circular e a extensão simétrica.

A estratégia de quantização dos coeficientes *wavelet* desempenha papel importantíssimo para o bom desempenho dos sistemas de codificação de imagens que utilizam DWT. Os codificadores empregam tanto a quantização escalar [26], [27] quanto a quantização vetorial [28].

Em se tratando, especificamente, da estratégia de alocação de bits [19], [29], [30], [31], o objetivo é determinar o número de bits b_{ij} alocados para cada sub-banda S_{ij} da decomposição *wavelet*, de modo que a distorção introduzida no processo de quantização seja minimizada, obedecendo à restrição $\sum_i \sum_j b_{ij} \leq b$.

A utilização de codificadores entrópicos (como por exemplo: código de Huffman, código do tipo *run-length* e código aritmético) contribui para redução da taxa de codificação, em virtude de explorar adequadamente a estatística dos símbolos utilizados na representação resultante da quantização.

Apesar de as técnicas de codificação serem levadas a efeito no sentido de minimizar uma medida de distorção objetiva, não necessariamente se observa como resultado a obtenção de uma imagem reconstruída com elevada qualidade subjetiva. É importante mencionar, contudo, que algumas abordagens [32], [33], [34] têm sido utilizadas para introduzir aspectos do sistema visual humano (HVS, *human visual system*) [35], [36] em sistemas de codificação de imagens.

IX. CONSIDERAÇÕES FINAIS

Neste artigo foram apresentados os fundamentos básicos de compressão de sinais. Amostragem e quantização escalar

foram brevemente descritas. Uma visão geral de quantização vetorial foi apresentada. Parâmetros de desempenho de um sistema de compressão de sinais foram sucintamente discutidos. O artigo apresentou, ainda, uma noção de codificação por transformada, com destaque para o uso da transformada *wavelet* discreta.

Como possíveis direções de pesquisa, podem ser citadas: codificação de imagem e vídeo baseada em segmentação – o alvo é fazer com que coleções cada vez maiores de segmentos que pertençam a um mesmo objeto, ou que se movam coerentemente, possam ser descritas por um segmento, de tal maneira que a predição de movimento seja mais precisa; codificação “Panorama”, cuja ideia básica é assumir que imagens em sequências de vídeo possam ser decompostas em fundo estático (*background*) e objetos móveis; codificação de vídeo baseada em objetos, cujo alvo é identificar objetos semânticos nas imagens, tais como pessoas em cenas, e representá-los usando modelos 2-D e 3-D para melhorar a eficiência de codificação; codificação de sinais que obedeça a restrições de redes de comunicações e requisitos do usuário final; codificação robusta, de forma a assegurar robustez a erros de canal e a perdas de pacotes; codificação de sinais no cenário de redes de sensores sem fio.

REFERÊNCIAS

- [1] Haykin, S. *Digital Communications*. John Wiley & Sons, 1988.
- [2] Jayant, N. S. and Noll, P. *Digital Coding of Waveforms*. Prentice-Hall, Englewood Cliffs, NJ, 1984.
- [3] Gersho, A. and Gray, R. M. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston, MA, 1992.
- [4] Gray, R. M. “Vector Quantization”. *IEEE ASSP Magazine*, pp. 4–29, April 1984.
- [5] Linde, Y., Buzo, A., and Gray, R. M. “An Algorithm for Vector Quantizer Design”. *IEEE Transactions on Communications*, Vol. COM-28, No. 1, pp. 84–95, January 1980.
- [6] Ramamurthi, B. and Gersho, A. “Classified Vector Quantization of Images”. *IEEE Transactions on Communications*, Vol. COM-34, No. 11, pp. 1105–1115, November 1986.
- [7] Moayeri, N., Neuhoﬀ, D. L., and Stark, W. E. “Fine-Coarse Vector Quantization”. *IEEE Transactions on Signal Processing*, Vol. 39, No. 7, pp. 1503–1515, July 1991.
- [8] Zeger, K., Vaisey, J., and Gersho, A. “Globally Optimal Vector Quantizer Design by Stochastic Relaxation”. *IEEE Transactions on Signal Processing*, Vol. 40, No. 2, pp. 310–322, February 1992.
- [9] Kohonen, T. “The Self-Organizing Map”. *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464–1480, September 1990.
- [10] Madeiro, F., Vilar, R. M., Fecine, J. M., and Aguiar Neto, B. G. “A Self-Organizing Algorithm for Vector Quantizer Design Applied to Signal Processing”. *International Journal of Neural Systems*, Vol. 9, No. 3, Special Issue on Neural Networks in Brazil: V Brazilian Symposium on Neural Networks, pp. 219–226, June 1999.
- [11] Karayannis, N. B. and Pai, P.-I. “Fuzzy Vector Quantization Algorithms and Their Applications in Image Compression”. *IEEE Transactions on Image Processing*, Vol. 4, No. 9, pp. 1193–1201, September 1995.
- [12] Pan, J. S., McInnes, F. R., and Jack, M. A. “VQ Codebook Design Using Genetic Algorithms”. *Electronics Letters*, Vol. 31, No. 17, pp. 1418–1419, 17th August 1995.
- [13] Equitz, W. H. “A New Vector Quantization Clustering Algorithm”. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 10, pp. 1568–1575, October 1989.
- [14] Eskicioglu, A. M. and Fischer, P. S. “Image Quality Measures and Their Performance”. *IEEE Transactions on Communications*, Vol. 43, No. 12, pp. 2959–2965, December 1995.
- [15] Deller Jr., J. R., Proakis, J. G., and Hansen, J. H. L. *Discrete-time Processing of Speech Signals*. Macmillan Publishing Co., 1993.
- [16] Aguiar Neto, B. G. *Processamento e Transmissão Digital de Voz (Apostila)*. Curso de Pós Graduação em Engenharia Elétrica, Universidade Federal da Paraíba, Campina Grande - PB, 1995.
- [17] Kubrick, A. and Ellis, T. “Classified Vector Quantisation of Images: Codebook Design Algorithm”. *IEEE Proceedings*, Vol. 137, Pt. I, No. 6, pp. 379–386, December 1990.
- [18] Jayant, N. “Signal Compression: Technology Targets and Research Directions”. *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 5, pp. 796–818, June 1992.
- [19] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
- [20] Burrus, C. S., Gopinath, R. A., and Guo, H. *Introduction to Wavelets and Wavelet Transforms*. Prentice-Hall, New Jersey, 1998.
- [21] Mallat, S. G. “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674–696, July 1989.
- [22] Desarte, P., Macq, B., and Slock, D. T. M. “Signal-Adapted Multiresolution Transform for Image Coding”. *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp. 897–904, March 1992.
- [23] Villasenor, J. D., Belzer, B., and Liao, J. “Wavelet Filter Evaluation for Image Compression”. *IEEE Transactions on Image Processing*, Vol. 4, No. 8, pp. 1053–1060, August 1995.
- [24] Tewfik, A. H., Sinha, D., and Jorgensen, P. “On the Optimal Choice of a Wavelet for Signal Representation”. *IEEE Transactions on Information Theory*, Vol. 38, No. 2, pp. 747–765, March 1992.
- [25] Liang, J. and Parks, T. W. “Image Coding Using Translation Invariant Wavelet Transforms with Symmetric Extensions”. *IEEE Transactions on Image Processing*, Vol. 7, No. 5, pp. 762–769, May 1998.
- [26] Xiong, Z., Ramchandran, K., and Orchard, M. T. “Space-Frequency Quantization for Wavelet Image Coding”. *IEEE Transactions on Image Processing*, Vol. 6, No. 5, pp. 677–693, May 1997.
- [27] Shapiro, J. M. “Embedded Image Coding Using Zerotrees of Wavelet Coefficients”. *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, pp. 3445–3462, December 1993.
- [28] Karlekar, J., Poonacha, P. G., and Desai, U. B. “Image Compression Using Zerotree and Multistage Vector Quantization”. *Proceedings of the IEEE International Conference on Image Processing (ICIP’97)*, Vol. 2, pp. 610–613, 1997.
- [29] Davis, G. M. and Nosratinia, A. “Wavelet-based Image Coding: An Overview”. *Applied and Computational Control, Signals and Circuits*, 1999.
- [30] Shohan, Y. and Gersho, A. “Efficient Bit Allocation for an Arbitrary Set of Quantizers”. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, pp. 1445–1453, September 1988.
- [31] Moulin, P. “A Multiscale Relaxation Algorithm for SNR Maximization in Nonorthogonal Subband Coding”. *IEEE Transactions on Image Processing*, Vol. 4, pp. 1269–1281, September 1995.
- [32] Ramos, M. G., Hemani, S. S., and Tamburro, M. A. “Psychovisually-Based Multiresolution Image Segmentation”. *Proceedings of the IEEE International Conference on Image Processing (ICIP’97)*, pp. 66–69, 1997.
- [33] Höntsch, I., Karan, L. J., and Safranek, R. J. “A Perceptually Tuned Embedded Zerotree Image Coder”. *Proceedings of the IEEE International Conference on Image Processing (ICIP’97)*, pp. 41–44, 1997.
- [34] Ran, X. and Farvardin, N. “A Perceptually Motivated Three-Component Image Model – Part II: Applications to Image Compression”. *IEEE Transactions on Image Processing*, Vol. 4, No. 4, pp. 430–447, April 1995.
- [35] Jayant, N., Johnston, J., and Safranek, R. “Signal Compression Based on Models of Human Perception”. *Proceedings of the IEEE*, Vol. 81, No. 10, pp. 1385–1422, October 1993.
- [36] Ran, X. and Farvardin, N. “A Perceptually Motivated Three-Component Image Model – Part I: Description of the Model”. *IEEE Transactions on Image Processing*, Vol. 4, No. 4, pp. 401–415, April 1995.