

CONSIDERAÇÕES SOBRE O USO DE FILTRAGEM B-SPLINE APLICADA A UMA ESTRUTURA DE CODIFICAÇÃO DE VÍDEO DE DOIS ESTÁGIOS PARA OPERAÇÃO EM BAIXA TAXA DE BITS

EVÂNIO RAMOS NICOLEIT E RUI SEARA

LINSE: Circuitos e Processamento de Sinais

Departamento de Engenharia Elétrica

Universidade Federal de Santa Catarina

Campus Universitário, 88040-900 – Florianópolis – SC – Brasil

Tel: (0xx48) 331-9504, Fax: (0xx48) 331-9091, evanio@linse.ufsc.br

RESUMO

Este artigo apresenta algumas considerações sobre o uso de filtragem *B-spline* aplicada à codificação de vídeo escalável espacialmente para operação em baixas taxas de bits. Neste contexto, é avaliado o desempenho da técnica *B-spline*, levando-se em conta a ordem dos filtros associados. Resultados de simulações são apresentados comparando-se e discutindo-se o desempenho da codificação. Também é mostrada uma curva taxa-qualidade para a Recomendação H.263, permitindo uma comparação entre o desempenho do referido padrão e o da metodologia proposta.

1. INTRODUÇÃO

A codificação de sinais de vídeo para baixas taxas de bits tem despertado, nos últimos anos, um grande interesse de laboratórios e pesquisadores da área. Isso se deve, principalmente, à grande demanda do mercado de vídeo por aplicações na *internet* e outras com baixas taxas, destacando-se os sistemas de videofonia e videoconferência, levando à necessidade de técnicas de compressão de vídeo cada vez mais eficientes. A grande maioria dos codificadores de vídeo padronizados [1,2] é baseada na transformada cosseno discreta (DCT - *discrete cosine transform*), em um único laço. Contudo, os codificadores baseados nessa técnica introduzem alguns inconvenientes [3], tais como: surgimento de artefatos de blocos e efeitos de recobrimento em altas taxas de compressão. Por outro lado, esquemas baseados em codificação escalável têm recebido uma considerável atenção no contexto de compressão de sinais de vídeo. Isso se deve principalmente às seguintes razões:

- i) por meio da escalabilidade espacial, diferentes resoluções podem ser obtidas, usando os respectivos subconjuntos do *bitstream* completo;
- ii) boas aproximações por interpolação *B-spline* podem ser alcançadas a partir de subimagens;
- iii) redução significativa do efeito de recobrimento, obtida pela escolha adequada de filtros de decomposição e reconstrução [3];

iv) escalabilidade de razão sinal-ruído (SNR) permite reconstruções com diferentes qualidades para as mesmas resoluções espacial e temporal.

A codificação escalável permite que os dados codificados sejam convenientemente direcionados para dois ou mais estágios (ou camadas), em diferentes níveis de importância, de forma que a informação mais relevante esteja no estágio principal. Em uma primeira etapa, o princípio básico atrás dessa filosofia consiste na construção de uma representação grosseira da imagem. Um caminho eficiente para obter essa representação é através de filtragem baseada em funções do tipo *B-spline*. Um grande atrativo para o uso de *splines* deve-se à suavidade que é obtida pelas restrições de continuidade impostas para a função e suas derivadas. Desta forma, obtém-se um alto grau de concordância nos pontos de amostragem no processo de reconstrução da imagem.

Este artigo apresenta algumas considerações sobre o uso de filtragem *B-spline* aplicada à codificação de vídeo escalável espacialmente para operação em baixas taxas de bits. Nesse contexto, é avaliado o desempenho da técnica *B-spline* levando-se em conta a ordem dos filtros associados. A Seção 2 descreve, de forma sucinta, a estrutura de codificação usada como referência na avaliação do desempenho dos filtros *B-spline* em questão. As Seções 3 e 4 apresentam o desenvolvimento matemático da filtragem *B-spline* aplicada ao problema de compressão de imagem, visando à obtenção de adequadas aproximações para o sinal de vídeo. Na Seção 5, são comentados os detalhes de implementação e discutidos os resultados de simulações, comparando o desempenho taxa-qualidade da codificação, considerando as ordens dos filtros *B-spline* testados. Finalmente, na Seção 6, são apresentadas as conclusões acerca deste trabalho.

2. ESTRUTURA DE CODIFICAÇÃO

O esquema de codificação utilizado como referência na avaliação de desempenho dos filtros *B-spline* é mostrado na Fig. 1 [4]. Trata-se de uma estrutura escalável tanto espacialmente quanto em SNR [5,6]. A estrutura consiste

de dois estgios. O primeiro opera em uma baixa resoluo espacial e utiliza o mesmo princpio de codificao do esquema baseado na Recomendao H.263. O segundo realiza um refinamento espacial que detecta a perda de informao do primeiro estgio [7,8].

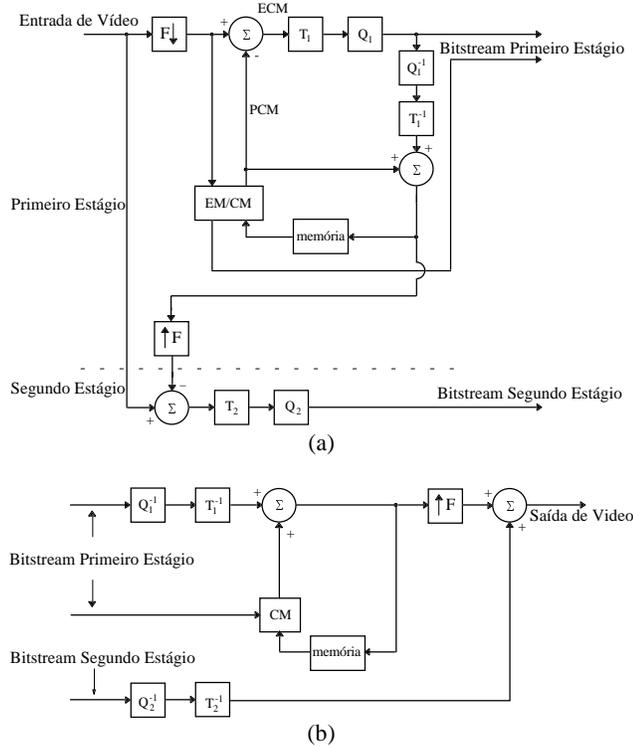


Figura 1. Codificador de vdeo de dois estgios escalvel espacialmente. (a) codificador e (b) decodificador.

O princpio de funcionamento desta estrutura (Fig. 1)  descrito como segue. A primeira etapa consiste de uma filtragem linear passa-baixas do quadro atual, seguida por uma operao de subamostragem (em cada dimenso da imagem por um fator dois) realizada pelo operador $[F\downarrow]$. A predio do quadro atual subamostrado  obtida usando estimo/compenso de movimentos (EM/CM) em uma resoluo espacial reduzida. No processo de codificao, aps a obteno da verso subamostrada codificada da imagem, estima-se uma predio do quadro atual na escala original da imagem. Isso  realizado atravs de um processo de expanso (interpolao e filtragem subsequente), efetuado pelo operador $[\uparrow F]$ [5]. Essa predio  uma verso grosseira da imagem na resoluo original [9]. Em seguida, no segundo estgio, um refinamento espacial  aplicado a essa verso. A imagem erro reconstruda  somada ao quadro atual gerado pelo primeiro estgio, resultando no quadro codificado atual, que  ento reconstrudo da mesma forma no decodificador.

3. FILTRAGEM BASEADA EM B-SPLINES

As *B-splines* (*basis-splines*) podem ser usadas para representar sinais. Estes sinais podem ser caracterizados como uma combinao linear de *B-splines* da seguinte forma:

$$s(x) = \sum_{k \in \mathbf{Z}} c(k) \beta^n(x-k) \quad , x \in \mathbf{R} \quad (1)$$

onde $c(\ell)$ so os coeficientes da representao de $s(x)$ atravs de uma funo *B-spline* centrada de grau n , $\beta^n(x)$ [11]. No contexto deste trabalho, somente esto sendo consideradas *splines* para amostras uniformemente espaadas. Os coeficientes $c(\ell)$ podem ento ser determinados via filtragem digital.

As *B-splines* $\beta^0(x)$ podem ser construdas a partir de $n+1$ operaes de convoluo entre $n+1$ pulsos retangulares [11]:

$$\beta^0(x) = \text{rect}(x) = \begin{cases} 1, & \text{for } |x| \leq 1/2 \\ 0, & \text{for } |x| > 1/2 \end{cases}$$

e

$$\beta^n(x) = \beta^0(x) * \beta^{n-1}(x) = \underbrace{\beta^0(x) * \beta^0(x) * \dots * \beta^0(x)}_{(n+1) \text{ termos}}$$

resultando em funes simtricas e bem comportadas.

Considerando-se o operador *B-spline* discreto, $b_m^n(k)$, obtido por amostragem da funo *B-spline* de grau n expandida por um fator m , tem-se

$$b_m^n(k) = \beta^n\left(\frac{x}{m}\right) \Bigg|_{x=k} \quad , k \in \mathbf{Z} \quad (2)$$

Sejam consideradas escalas didicas, $m = 2^i$. Definindo-se $s_i(k)$ como uma aproximao de mnimo erro do sinal discreto $s(k) = s_0(k) \in \mathbf{L}^2$ [11], onde \mathbf{L}^2 denota o espao de funes com energia finita, $s_i(k)$ pode ser agora representado por:

$$s_i(k) = \sum_{\ell \in \mathbf{Z}} c_m(\ell) b_m^n\left(\frac{k}{m} - \ell\right) \quad , m = 2^i \quad (3)$$

Os coeficientes $c_m(\ell)$ podem ser determinados pelo seguinte produto interno:

$$c_m(k) = \frac{1}{m} \left\langle s(\ell) , (b_m^n)^{-1}\left(\frac{k}{m} - \ell\right) \right\rangle \quad (4)$$

onde $(b_m^n)^{-1}(k)$  operador inverso de $b_m^n(k)$.

Agora, estamos interessados na aproximao do sinal $s(k)$ por um conjunto de coeficientes *B-spline*, $c_m(k)$, que pode

ser determinado, iterativamente, usando-se uma combinação de operações de pré-filtragem e subamostragem, de acordo com um critério de aproximação de erro mínimo. Um critério prático de implementação, para um número reduzido de coeficientes, é que os operadores $[F\downarrow]$ e $[\uparrow F]$ sejam ótimos no sentido do erro quadrático mínimo [12,13].

4. FORMULAÇÃO VIA MÍNIMOS QUADRADOS

O operador inverso, $(b_m^n)^{-1}(k)$, é ótimo no sentido dos mínimos quadrados, com respeito a um operador de síntese, $b_m^n(k)$, quando a energia $\varepsilon_m^2 = \sum_{k \in \mathbf{Z}} [s(k) - \hat{s}(k)]^2$ do erro de aproximação torna-se mínima.

A formulação geral do enfoque baseado em filtragem *B-spline*, com um fator de interpolação inteiro, $m = 2^i$, é:

$$\hat{s}(k) = \sum_{\ell \in \mathbf{Z}} c_m(\ell) \beta^n \left(\frac{k}{m} - \ell \right). \quad (5)$$

O desenvolvimento via mínimos quadrados consiste na obtenção dos coeficientes $c_m(k)$ que minimizam o erro de aproximação. Assim,

$$\varepsilon_m^2 = \sum_{k \in \mathbf{Z}} [s(k) - c_m(k) * b_m^n(k)]^2, \quad (6)$$

onde “*” denota a operação de convolução. Pela utilização da notação de produto interno, $\langle \cdot, \cdot \rangle$, a Eq. (6) é equivalente a:

$$\varepsilon_m^2 = \langle s(k), s(k) \rangle - 2 \langle s(k), c_m(k) * b_m^n(k) \rangle + \langle c_m(k) * b_m^n(k), c_m(k) * b_m^n(k) \rangle. \quad (7)$$

Diferenciando-se ε_m^2 com respeito aos coeficientes $c_m(k)$ e igualando-se a zero, obtém-se:

$$[b_m^n(k) * b_m^n(k)]_{\downarrow m} * c_m(k) = [b_m^n(k) * s(k)]_{\downarrow m}, \quad (8)$$

onde “ $\downarrow m$ ” denota a operação de subamostragem por um fator m . Agora, considerando-se que $\{[b_m^n(k) * b_m^n(k)]_{\downarrow m}\}^{-1}$ existe, a partir de (8) via filtragem inversa, obtém-se:

$$c_m(k) = s_m^n(k) * [b_m^n(k) * s(k)]_{\downarrow m}^{-1}, \quad k \in \mathbf{Z}, \quad (9)$$

onde $s_m^n(k)$ é definido como:

$$s_m^n(k) = \{[b_m^n(k) * b_m^n(k)]_{\downarrow m}\}^{-1}, \quad k \in \mathbf{Z}. \quad (10)$$

As funções de transferência, no domínio z , correspondentes aos operadores $b_m^n(k)$ e $s_m^n(k)$ para $n = 1, \dots, 4$, são mostradas nas Tabelas 1 e 2, respectivamente.

Tabela 1. Funções de transferência de filtros *B-spline* $B_{m=2}^n(z)$, para $n = 1, \dots, 4$.

N	$B_{m=2}^n(z)$
1	$\frac{z^{-1} + 2 + z^1}{2}$
2	$\frac{z^{-2} + 4z^{-1} + 6 + 4z^1 + z^2}{8}$
3	$\frac{z^{-3} + 8z^{-2} + 23z^{-1} + 32 + 23z^1 + 8z^2 + z^3}{48}$
4	$\frac{z^{-4} + 16z^{-3} + 76z^{-2} + 176z^{-1} + 230 + 176z^1 + 76z^2 + 16z^3 + z^4}{384}$

Tabela 2. Funções de transferência para os filtros mínimos quadrados *B-spline* $S_{m=2}^n(z)$ com fator 2 de subamostragem, para $n = 1, \dots, 4$.

N	$S_{m=2}^n(z)$
1	$\frac{4}{z^{-1} + 6 + z^1}$
2	$\frac{64}{z^{-2} + 28z^{-1} + 70 + 28z^1 + z^2}$
3	$\frac{2304}{z^{-3} + 110z^{-2} + 1087z^{-1} + 2212 + 1087z^1 + 110z^2 + z^3}$
4	$\frac{147475}{z^{-4} + 408z^{-3} + 11870z^{-2} + 71729z^{-1} + 126934 + 71729z^1 + 11870z^2 + 408z^3 + z^4}$

Note que o correspondente modelo bidimensional *B-spline* é separável. Dessa forma, são realizadas sucessivas filtragens unidimensionais com operações de subamostragem e interpolação das linhas e colunas da imagem.

5. ASPECTOS DE IMPLEMENTAÇÃO E RESULTADOS DE SIMULAÇÕES

Os resultados descritos nesta seção são baseados na seqüência padrão QCIF *Miss America*, com 176×144 *pixels*, 256 níveis de cinza e amostrada a 30 quadros/s. Algumas definições relativas à implementação da metodologia de codificação se fazem necessárias e são apresentadas a seguir.

Dentro da família das *splines* polinomiais, as *splines* cúbicas são, sem dúvida, as mais populares e utilizadas [10,11]. Contudo, no contexto de compressão de vídeo em baixas taxas de bits, uma investigação acerca da ordem é necessária. Uma razão para tal é que, no processo de codificação, há inserção de ruídos provenientes de certas fontes, tais como: ruído/erro de quantização/codificação e variância ao deslocamento devido aos algoritmos de

estimação/compensação de movimentos. Neste trabalho, são considerados filtros *B-spline* de primeira, segunda, terceira e quarta ordens, com as respectivas respostas ao impulso centradas e simétricas. A escolha dos filtros dos operadores $[F\downarrow]$ e $[\uparrow F]$ é feita segundo o critério dos mínimos quadrados.

Para a implementação da etapa de EM/CM, as imagens são divididas em blocos. O algoritmo BMA (*block-matching algorithm*) [14] é usado para estimar o movimento de cada bloco dentro do quadro anterior, para a sua nova posição dentro do quadro atual. O tamanho de bloco utilizado para as seções de EM/CM é 4×4 pixels. Nesse caso, esta escolha tem representado um bom compromisso entre a taxa de bits requerida e a qualidade da estimação para as imagens subamostradas. O resultado dessa operação, para cada quadro, é um conjunto de 11×9 vetores de deslocamento (VD). A seção CM aplica os VD ao quadro anteriormente codificado para a obtenção de uma imagem de predição com compensação de movimentos (PCM). A janela de busca utilizada no BMA é de ± 7 pixels em ambas as direções (horizontal e vertical). O critério de busca utilizado para encontrar a melhor representação é o MAD (*mean absolute difference*) [14]. O campo de VD é codificado, adotando-se os mesmos códigos utilizados pela Recomendação H.263. Para uma representação eficiente das imagens erro com compensação de movimentos (ECM), essas imagens são transformadas através de decomposição *wavelet* (DWT – *discrete wavelet transform*). A DWT utiliza o banco de filtros Antonini-Daubechies *7/9-tap* [15,16] com três níveis de decomposição para o componente de luminância. Para alcançar um desempenho taxa-qualidade competitivo, a codificação dos coeficientes da DWT é realizada usando o algoritmo EZW (*embedded zerotree wavelet*) [17].

No segundo estágio, as imagens diferença são transformadas através de decomposição DWT em quatro níveis para o sinal luminância, empregando um banco de filtros idêntico ao usado no primeiro estágio. A codificação dos coeficientes da DWT é realizada usando novamente o algoritmo EZW.

A alteração da taxa de bits é obtida pela variação do número de passadas dominante/subordinada através do algoritmo EZW em ambos os estágios, dando-se maior ênfase ao primeiro estágio.

No conjunto de simulações, somente predição do tipo P é utilizada para o modo INTER [2]. Contudo, outros modos de predição de quadro podem ser facilmente incorporados à estrutura proposta para efeito de comparação com a Recomendação H.263.

Para o codificador baseado na Recomendação H.263, foi utilizada a implementação TMN-5 [2,18].

A figura de mérito objetiva, usada para medir a qualidade de reconstrução dos quadros codificados, é a razão sinal ruído de pico (PSNR), definida como:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{255^2}{\text{EQM}} \right), \quad (15)$$

com o erro quadrático médio (EQM), considerando-se R quadros, definido como:

$$\text{EQM} = \frac{1}{M \cdot N \cdot R} \sum_{r=0}^{R-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [I_r(m, n) - \hat{I}_r(m, n)]^2, \quad (16)$$

onde $I_r(m, n)$ representa o nível correspondente ao quadro r original e $\hat{I}_r(m, n)$, o nível correspondente ao quadro r reconstruído, ambos com dimensões $M \times N$ pixels.

Uma vez que se está interessado nas características em regime permanente de codificação, operando sobre uma estrutura preditiva em laço fechado, o primeiro quadro da seqüência não é codificado. Portanto, para o primeiro quadro, a taxa de bits e a PSNR não são incluídas nas medidas taxa-qualidade.

A Fig. 2 apresenta as curvas de desempenho taxa-qualidade para ambos os codificadores, o baseado na Recomendação H.263 e o baseado em filtragem *B-spline* em diferentes ordens. Pode ser constatado que, para taxas de bits próximas a 40 kbits/s, considerando-se ordem unitária, a filtragem em questão apresenta uma PSNR de aproximadamente 2 dB maior do que a do codificador baseado na Recomendação H.263. Pode-se também notar que, à medida que a ordem da filtragem *B-spline* cresce, o desempenho da qualidade de reconstrução decresce. Isso é devido ao alto grau de suscetibilidade da filtragem inversa a erros/ruídos de codificação provenientes do processo de compressão com perdas em baixas taxas de bits. O filtro de mínimos quadrados, $S_m^n(z)$, realiza uma filtragem recursiva (IIR). A implementação da filtragem é baseada em uma cascata de funções de transferência recursivas simétricas (sendo uma causal e a outra anti-causal) [13]. A representação aritmética dos coeficientes de $S_m^n(z)$ demanda uma faixa dinâmica que é crescente à medida que a ordem do filtro cresce. Como uma consequência do aumento da ordem, o sinal reconstruído torna-se mais e mais sensível a erros de codificação/quantização inerentes ao processo de compressão com perdas. Por outro lado, quando se tem um processo de compressão sem perdas, a qualidade das imagens (PSNR), como esperado e mencionado em [19], aumenta à medida que a ordem dos filtros também aumenta. A Tabela 3 mostra os resultados de PSNR para algumas ordens de filtros, usando as operações de decomposição/reconstrução ($[F\downarrow]$ e $[\uparrow F]$) baseadas em filtragem *B-spline* para este caso.

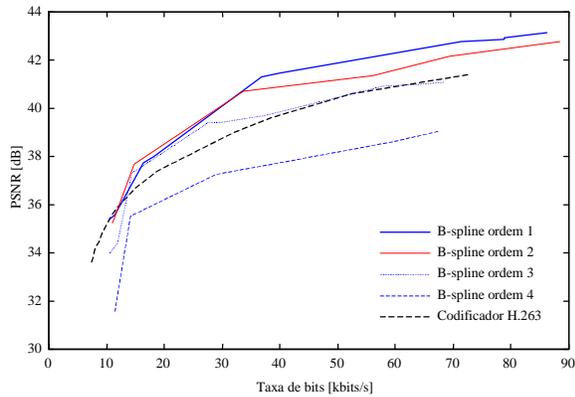


Figura 2. Características taxa-qualidade para os codificadores de dois estágios com filtragem *B-spline* e o baseado na Recomendação H.263.

Tabela 3. Resultados de PSNR para a seqüência *Miss America* usando as operações $[F\downarrow]$ e $[\uparrow F]$ por filtragem *B-spline* para um processo de compressão sem perdas.

Ordem	PSNR
1	38,57
2	39,58
3	39,69
4	39,72

6. CONCLUSÕES

Neste artigo, foram apresentadas algumas considerações sobre o uso de filtragem *B-spline* aplicada a uma estrutura de codificação de vídeo de dois estágios escalável para operação em baixas taxas de bits. A filtragem *B-spline*, utilizada para mudança de níveis de resolução, apresenta um significativo potencial em codificação de vídeo. Foi observado que, à medida que a ordem dos filtros aumenta, o desempenho da qualidade de reconstrução decresce. Isso é devido ao alto grau de suscetibilidade da filtragem inversa a erros/ruídos de codificação provenientes do processo de compressão com perdas em baixas taxas de bits. Resultados de simulações foram apresentados, comparando-se o desempenho taxa-qualidade de codificação, considerando-se a ordem dos filtros *B-spline* usados. Melhores resultados são obtidos para ordens mais baixas.

REFERÊNCIAS

[1] ITU-T. "Recommendation H.261 - Video Codec for Audiovisual Services at $p \times 64$ kbit/s", Mar. 1993.
 [2] ITU-T. "Recommendation H.263 - Video Coding for Low Bitrate Communication", Jul. 1995.

[3] M. K. Mandal, E. Chan, X. Wang e S. Panchanathan, "Multiresolution Motion Estimation Techniques for Video Compression". *Optical Engineering*, vol. 35, no. 1, pp. 128-136, Jan. 1996.
 [4] E. R. Nicoleit e R. Seara, "Codificação de Vídeo Escalável Espacialmente para Operação em Baixas Taxas de Bits". *XIII Congresso Brasileiro de Automática*, Florianópolis, SC, pp. 1393-1397, Set. 2000.
 [5] K. Illgner e F. Müller, "Spatially Scalable Video Compression Employing Resolution Pyramids". *IEEE J. Selected Areas in Communications*, vol. 15, pp. 1688-1703, Dec. 1997.
 [6] R. Mathew e J. F. Arnold, "Layered Coding Using Bitstream Decomposition with Drift Correction". *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 882-891, Dec. 1997.
 [7] H. Kwon, M. Venkatramam e N. M. Nasrabadi, "Very Low Bit-Rate Video Coding Using Variable Block-Size Entropy-Constrained Residual Vector Quantizers". *IEEE J. Selected Areas in Communications*, vol. 15, pp. 1714-1725, Dec. 1997.
 [8] E. R. Nicoleit, R. Seara e J. C. M. Bermudez, "Compressão de Vídeo com Alta Qualidade e Baixa Taxa de Bits Usando Codificação Multi-Estágio". *XV Simpósio Brasileiro de Telecomunicações*, Recife, PE, pp. 590-593, Set. 1997.
 [9] P. J. Burt e E. H. Adelson, "The Laplacian Pyramid as a Compact Image Code". *IEEE Trans. on Communications*, vol. Com-31, no. 4, pp. 532-540, Apr. 1983.
 [10] J. L. Buchanan e P. R. Turner, *Numerical Methods and Analysis*, McGraw Hill, New York, 1992.
 [11] M. Unser, "Splines – A Perfect Fit for Signal and Image Processing". *IEEE Signal Processing Magazine*, pp. 22-38, Nov. 1999.
 [12] M. Unser, A. Aldroubi e M. Eden, "B-Spline Signal Processing – Part I - Theory". *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 821-833, Feb. 1993.
 [13] M. Unser, A. Aldroubi e M. Eden, "B-Spline Signal Processing – Part II – Efficient Design and Applications". *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 834-848, Feb. 1993.
 [14] H. G. Musmann, P. Pirsch e H. J. Grallert, "Advances in Picture Coding". *Proceedings of the IEEE*, vol. 73, no. 4, pp. 523-548, Apr. 1985.
 [15] M. Antonini, M. Barlaud, P. Mathieu e I. Daubechies, "Image Coding Using Wavelet Transform". *IEEE Trans. on Image Processing*, vol. 1, no. 2, pp. 205-220, Apr. 1992.
 [16] J. D. Villasenor, B. Belzer e J. Liao, "Wavelet Filter Evaluation for Image Compression". *IEEE Trans. on Image Processing*, vol. 4, no. 8, pp. 1053-1060, Aug. 1995.

- [17]J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients". *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.
- [18]Telenor R&D, "Tmn (h.263) encoder/decoder, version 1.7". <ftp://bonde.nta.no/pub/tmn>, TMN codec, 1996.
- [19]P. Brigger, F. Müller, K. Illgner e M. Unser, "Centered Pyramids". *IEEE Trans. on Image Processing*, vol. 8, no. 9 pp. 1254-1264, Sept. 1999.