

Codificador de Vídeo Baseado na Transformada de Ôndulas 3D

José Salvado¹, Leonel Sousa²

¹ Escola Superior de Tecnologia de Castelo Branco, Av. do Empresário, 6000-767 Castelo Branco, PORTUGAL, jsalvado@est.ipcb.pt

² Instituto Superior Técnico (DEEC) e Instituto de Engenharia de Sistemas e Computadores (SIPS), R. Alves Redol, 9, 1000-029 Lisboa, PORTUGAL, las@inesc-id.pt

Neste artigo apresenta-se um codificador de vídeo baseado na transformada de ôndulas, com interface do tipo Windows®, desenvolvido em Borland C++ Builder v5.0. O codificador tira partido das características da estrutura hierárquica 3D de coeficientes da DWT, resultante da sua aplicação no domínio do tempo e do espaço. O utilizador escolhe a sequência a codificar (armazenada em ficheiro), o número de tramas, o nível de decisão, e a ôndula a considerar na DWT no espaço, podendo obter resultados do desempenho do codificador, e guardar esses valores em ficheiros compatíveis com Microsoft® Excel. Pode, também, visualizar alguns resultados intermédios do processo de codificação, nomeadamente a decomposição por aplicação da DWT no tempo e no espaço, conferindo a este trabalho, também, um cariz didáctico.

1. Introdução

A transmissão e o armazenamento de vídeo (sequência de imagens com referência temporal) apresenta requisitos exigentes. Por exemplo, para transmitir uma sequência de vídeo com resolução 288×352 pixels, 8 bits por pixel, 25 imagens por segundo, é necessário um ritmo de 20,275 Mbits/s. Por outro lado, nestas condições, um CD-ROM com capacidade 650 MB apenas pode armazenar cerca de 32 segundos de vídeo. Para reduzir os requisitos de armazenamento e adequar os ritmos de transmissão à largura de banda dos canais de comunicação actualmente disponíveis, é comum usarem-se técnicas de codificação (ou compressão), que visam sobretudo obter representações mais compactas da informação.

A informação presente nos sinais de imagem e de vídeo é fortemente correlacionada. Nas imagens explora-se a correlação espacial, enquanto que no vídeo explora-se, também, a correlação temporal. Os codificadores de imagem e de vídeo têm-se baseado na Transformada de Co-seno Discreta (DCT) (e.g. JPEG, MPEG, H.263) [1], aplicando-se a DCT a blocos de imagem de dimensão fixa, normalmente de 8×8 . Este esquema, apesar de eficiente, cria descontinuidades artificiais que originam um efeito visual designado por “efeito de bloco”, e que é particularmente visível para factores de compressão elevados.

A transformada de ôndulas¹ discreta (DWT) também permite a descorrelação de sinais, e tem merecido alguma atenção nos últimos anos. Por exemplo, a mais recente norma para codificação de imagem (JPEG2000) é baseada na DWT [2,3]. Porém, ao contrário da DCT, a DWT aplica-se à totalidade da imagem, evitando o aparecimento do “efeito de bloco”, e melhorando a qualidade subjectiva das imagens.

As relações entre a DWT, os esquemas de decomposição em sub-bandas e a análise multiresolução (MRA) são bem conhecidas [4,5]. A DWT, devido às suas características, concentra energia nas sub-bandas de menor frequência (níveis mais altos), e os seus coeficientes apresentam uma organização hierárquica em árvore, que facilita a realização de esquemas de codificação que exploram o grau de importância dos coeficientes, tal como em [6] (para imagem) mas que pode ser expandido a 3D, para aplicações em vídeo [8].

O presente artigo encontra-se organizado da seguinte forma. Na secção seguinte apresentam-se os conceitos fundamentais da DWT, a sua interligação com a MRA e bancos de filtros. Na secção 3 refere-se a formação da estrutura 3D de coeficientes da DWT por aplicação da DWT no tempo e no espaço, e refere-se brevemente o método a codificação hierárquica dos seus coeficientes. A aplicação desenvolvida é referida na

¹ *Wavelets*, em Inglês.

secção 4 e apresentam-se alguns resultados experimentais na secção 5. Por fim, na última secção apresentam-se as conclusões gerais.

2. Transformada de Ôndulas e Análise Multiresolução

Uma ôndula é uma função $\psi \in L^2(\mathbb{R})$, escalável, isto é, pode ser deslocada ou dilatada, possui boa localização no tempo e na frequência, sendo definida pela equação:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right), \quad a \neq 0, b \in \mathbb{R}$$

Esta equação estabelece as translações e as dilatações da ôndula de base, ou ôndula mãe (*mother wavelet*), através dos parâmetros de escala e de translação, a e b , respectivamente. O parâmetro escala está inversamente associado à frequência, isto é, a escalas baixas correspondem frequências altas e vice-versa. As ôndulas ortogonais e biortogonais de Daubechies [9] verificam estas propriedades e apresentam suporte compacto, sendo este um factor importante em processamento de sinais, na medida em que corresponde à realização de filtros com resposta impulsiva finita (filtros FIR). Com base nos valores discretos dos parâmetros de escala e de translação, $a = 2^{-j}$ e $b = k2^{-j}$, respectivamente, define-se a transformada de ôndulas discreta (DWT):

$$w_{k,j} = W\psi(2^{-j}, k2^{-j}) = \int_{-\infty}^{\infty} g(x) \cdot \overline{\psi\left(\frac{x-k2^{-j}}{2^{-j}}\right)} \cdot dx$$

Nas expressões anteriores, $\bar{\psi}$ é o complexo conjugado de ψ , e $\frac{1}{\sqrt{a}}$ é um parâmetro de normalização de energia nas várias escalas, que se organizam em oitavas.

A DWT permite a análise (e a síntese) de sinais num plano tempo-frequência (escala), com vantagens face à transformada de Fourier local, na medida em que permite obter resolução ajustável.

Análise Multiresolução

O método de análise multiresolução [4] é uma das aplicações mais importantes da transformada de ôndulas em processamento digital de sinal. A ideia fundamental associada à MRA é a decomposição de uma função $f \in L^2(\mathbb{R})$, em várias funções mais simples, por aproximação em escalas sucessivas, permitindo a sua representação com vários níveis de detalhe. Deste modo, a aproximação obtida para a escala mais alta (baixa frequência), em conjunto com os detalhes nas restantes escalas (frequências mais altas), permitem obter uma representação eficiente do sinal original. Na figura 1 ilustra-se a decomposição de espaços na MRA, V_j e W_j , que se designam, respectivamente, por espaço de aproximação e espaço de detalhe, sendo ortogonais entre si e complementares em V_{j+1} .

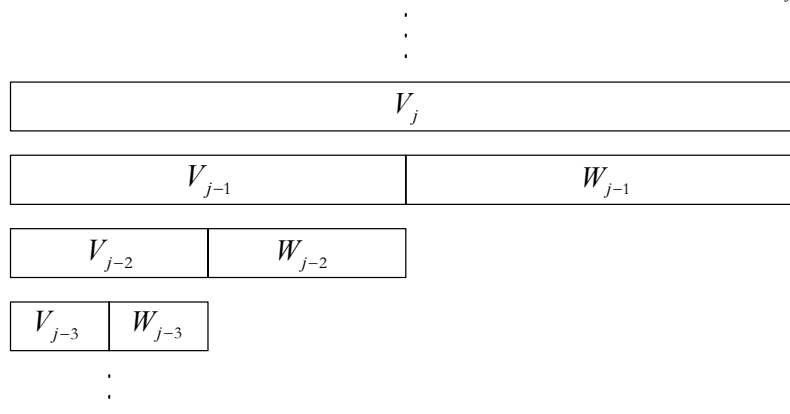


Figura 1. Esquema de decomposição e organização dos espaços de aproximações e de detalhe na MRA.

Os espaços V_j e W_j são gerados, respectivamente, por bases $\{\phi_{k,j} : 2^{j/2} \phi(2^j x - k), k \in \mathbb{Z}\}$ e $\{\psi_{k,j} : 2^{j/2} \psi(2^j x - k), k \in \mathbb{Z}\}$, em que ϕ é a função de escala, e ψ a ôndula, sendo ortogonais entre si nas diferentes escalas, aqui referenciadas por j .

DWT por Bancos de Filtros

Dado o encadeamento e a ortogonalidade dos espaços na MRA, existem duas seqüências $\{h_k\}, \{g_k\} \in l^2$, em que $g_k = (-1)^k h_{1-k}$, e que permitem representar as funções ϕ e ψ através dos coeficientes dos filtros h_k e g_k (até uma dada escala):

$$\phi(2^j x) = \sqrt{2} \sum_k h_k \cdot \phi(2^{j+1} x - k) \text{ e } \psi(2^j x) = \sqrt{2} \sum_k g_k \cdot \phi(2^{j+1} x - k)$$

Em que $\sqrt{2}$ é o valor do factor de normalização de energia no caso particular das ôndulas de Daubechies.

Por aplicação de pares de filtros passa-baixo (h_k) e passa-alto (g_k), para análise e para síntese, $\{\tilde{h}(z), \tilde{g}(z)\}$ e $\{h(z), g(z)\}$, respectivamente, é possível implementar a DWT e a DWT⁻¹ (1D). O cálculo da DWT 2D relativa a sinais bidimensionais (e.g imagens) é feita por extensão da DWT 1D aplicada a seqüências separáveis, primeiro segundo as linhas, e depois segundo as colunas, ou vice-versa, como se ilustra na figura 2.

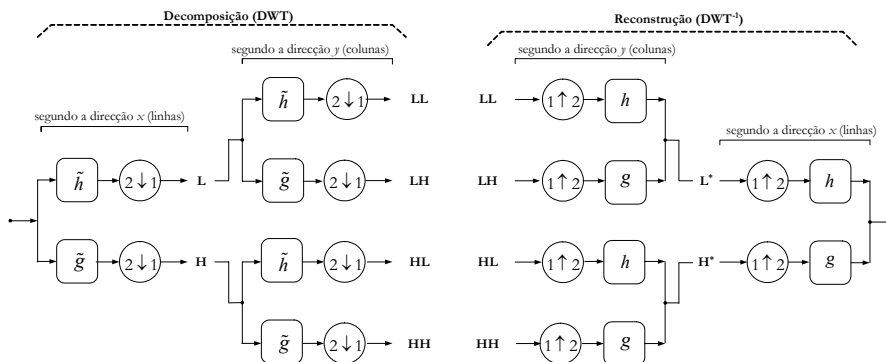


Figura 2. Decomposição em sub-bandas e reconstrução de sinais bidimensionais, através da DWT 2D.

Na figura 3 apresenta-se um exemplo da decomposição (análise) de uma imagem com três níveis (escalas) de decomposição. Para uma imagem de dimensões $M \times N$, a região correspondente ao nível mais alto, referente à localização dos coeficientes de aproximação, tem dimensões $\frac{M}{8} \times \frac{N}{8}, \left(\frac{M}{2^3} \times \frac{N}{2^3}\right)$.



Figura 3. Exemplo da aplicação da DWT 2D a uma imagem, por decomposição em 3 níveis.

Os coeficientes da DWT 2D apresentam relações de dependência hierárquica (ver fig. 4), sendo o grau de importância dos coeficientes maior no nível (escala) mais alto, decrescendo de importância à medida que a escala diminui.

O esquema de codificação EZW proposto por Shapiro [6] faz o aproveitamento da organização em árvore dos coeficientes da DWT 2D, e da elevada probabilidade de que os coeficientes mais altos se situem nos níveis mais altos. A ordem de varrimento dos coeficientes é fixa (ver fig. 4), sendo realizada de modo que os coeficientes dos níveis mais altos (coeficientes mais significativos) são codificados antes dos coeficientes dos níveis mais baixos (menos importantes). Aplicando a DWT a sinais de tridimensionais obtém-se uma estrutura de coeficientes da DWT com dependências hierárquicas a 3D, que pode ser aplicada à codificação de vídeo [7], podendo os seus coeficientes ser codificados através da extensão do algoritmo EZW a 3D [8], aqui designado por EZW3D.

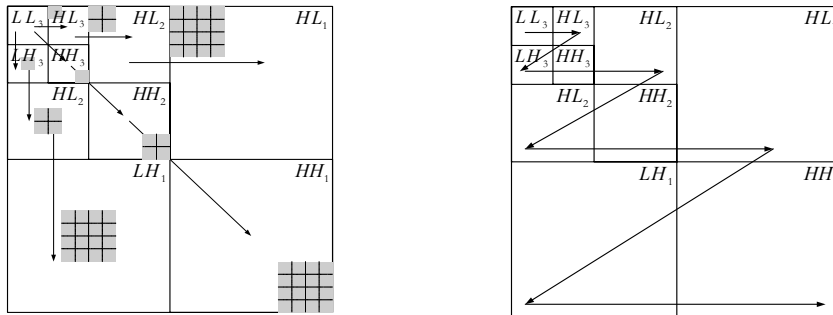


Figura 4. DWT 2D com 3 níveis: relações de dependência hierárquica (esquerda) e ordem de varrimento dos coeficientes (direita).

3. Codificador de Vídeo e Resultados Experimentais

O codificador desenvolvido tem por base a estrutura hierárquica 3D de coeficientes da DWT. A sequência de vídeo de entrada é separada, formando grupos de tramas, aos quais se aplica a DWT, no domínio do tempo e do espaço (3D). Por razões práticas de implementação, o número de tramas deve ser uma potência de 2. Neste caso considera-se o grupo formado por 4 tramas.

Primeiro calcula-se a DWT (1D) com dois níveis de decomposição, no domínio do tempo, de modo a explorar a correlação temporal. Em seguida, aplica-se a DWT 2D no espaço a cada uma das tramas transformadas no tempo, com três níveis de decomposição, de modo a explorar a correlação espacial. A DWT no tempo e no espaço concentra energia nos níveis mais altos, donde resulta uma estrutura de coeficientes em árvore, como se ilustra na figura 5.

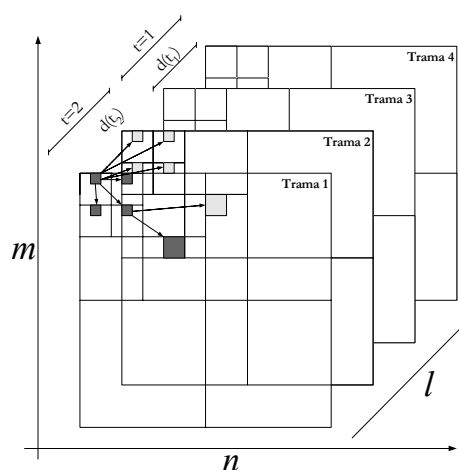


Figura 5. DWT 3D com decomposição em p níveis no domínio do tempo, e q níveis no domínio do espaço.

Para grupos de tramas de dimensão $L \times M \times N$, os coeficientes mais significativos situam-se predominantemente numa região de dimensões $L_h = \frac{L}{4}$, $M_h = \frac{M}{8}$, $N_h = \frac{N}{8}$. As coordenadas dos coeficientes são :

$$\begin{aligned}
(l, m, n) &\rightarrow \{(l + L_h, m, n), (l, m + M_h, n), (l, m, n + N_h), \\
&\quad (l + L_h, m + M_h, n), (l + L_h, m, n + N_h), (l, m + M_h, n + N_h), \\
&\quad (l + L_h, m + M_h, n + N_h)\} \quad \text{se } (l, m, n) \in R \text{ (raiz da arvore)} \\
(l, m, n) &\rightarrow \{(2l, 2m, 2n), (2l + 1, 2m, 2n), (2l, 2m + 1, 2n), (2l, 2m, 2n + 1), \\
&\quad (2l + 1, 2m + 1, 2n), (2l + 1, 2m, 2n + 1), (2l, 2m + 1, 2n + 1), \\
&\quad (2l + 1, 2m + 1, 2n + 1)\} \quad \text{se } (l, m, n) \in I \text{ (nivel entemedio)} \\
(l, m, n) &\rightarrow \{ \} \quad \text{se } (m, n) \in B \text{ (terminação da arvore)}
\end{aligned}$$

O varrimento dos coeficientes na estrutura 3D faz-se da primeira para a última trama, e dentro de cada uma delas, dos níveis mais altos para os mais baixos, segundo as relações de dependência hierárquica, e respeitando o seu grau de importância (significado), como se indica na figura 4.

Na codificação dos coeficientes, atendendo ao seu grau de importância (significado) face a um dado valor de decisão, podem resultar quatro símbolos: **positivo (POS)**, se o coeficiente é significativo e positivo; **negativo (NEG)**, se o coeficiente é significativo mas de sinal negativo; **zero isolado (IZ)**, se o coeficiente não é considerado significativo, mas pelo menos um dos dependentes é significativo; **raiz da árvore (ZTR)**, se o coeficiente e todos os seus dependentes são considerados não significativos.

Aplicação Desenvolvida

Para implementar este esquema de codificação de vídeo com base na transformada de ôndulas, desenvolveu-se uma aplicação em ambiente Windows® utilizando o ambiente de desenvolvimento Borland C++ Builder v5.0 [10], cuja interface com o utilizador se apresenta na figura 6.

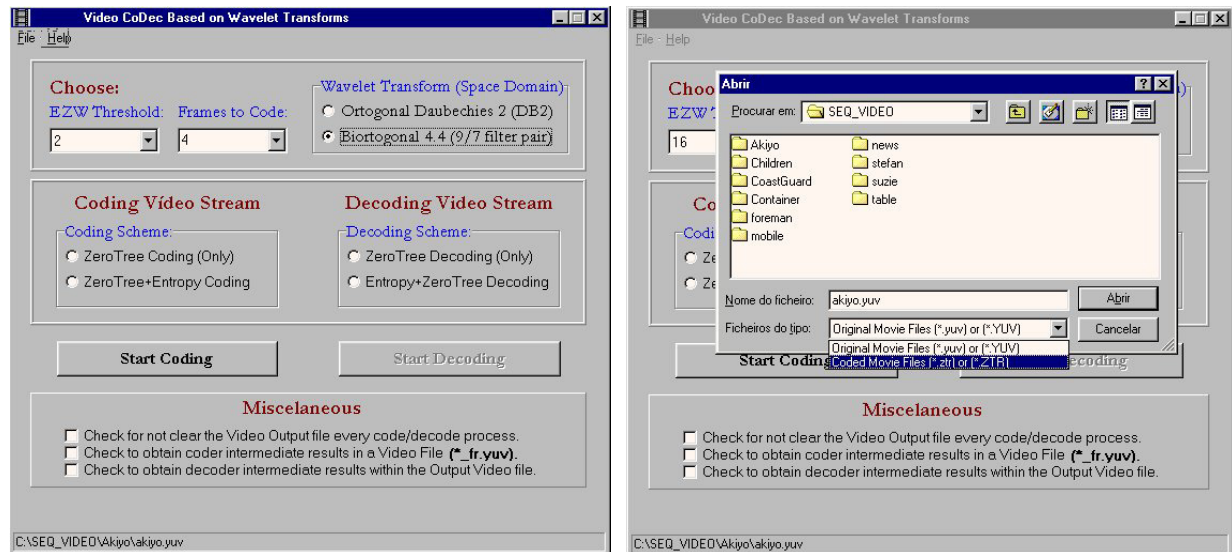


Figura 6. Aspectos da interface com o utilizador do codificador de vídeo desenvolvido.

O codificador desenvolvido apenas permite codificar seqüências de vídeo no formato QCIF (*quarter* CIF), cujas imagens têm uma resolução espacial 144×176. Usam-se as componentes YUV que se encontram relacionadas com o sistema RGB (*Red-Green-Blue*) de cores primárias da seguinte forma:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ -0,148 & -0,289 & 0,437 \\ 0,615 & -0,515 & -0,100 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Para a DWT no tempo usa-se a ôndula ortogonal Daubechies 2, com 2 níveis de decomposição. Para a DWT no domínio do espaço, de modo a comparar os resultados, usam-se duas situações distintas. A ôndula ortogonal Daubechies 2 e as ôndulas biortogonais (4,4), com 3 níveis de decomposição em qualquer dos casos.

No codificador, seleccionando o menu **File**, através da janela de diálogo (ver fig. 6), o utilizador selecciona o ficheiro contendo a sequência de vídeo a codificar (ficheiros em binário com extensão **.yuv**). O ficheiro seleccionado, assim como toda a informação relativa à sua localização, são mostradas na parte inferior da janela principal da aplicação (ver fig. 6). Posteriormente, através das caixas de diálogo, o utilizador decide, também, o número de tramas a codificar (múltiplo de 4), o nível de decisão a considerar na codificação, o tipo de filtros de ôndula a considerar na DWT no espaço, e o tipo de codificação pretendido: apenas EZW 3D ou EZW 3D + Huffman. A atribuição de valores aos parâmetros faz-se de uma forma muito simples, através da selecção de entre um conjunto de valores pré-definidos, fazendo deslizar o cursor e activando o valor pretendido. Para iniciar o processo de codificação basta seleccionar o respectivo botão, que apenas fica activo se o tipo de ficheiro de vídeo for o correcto (*.yuv) e se a sua abertura para leitura tiver sido realizada com sucesso.

Para realizar a descodificação, também através do menu **File** e da janela de diálogo (fig. 6), o utilizador selecciona o ficheiro contendo a *stream* a descodificar (com extensão ***.ztr**), e inicia a descodificação seleccionando o botão **Start Decoding**. A informação relativa aos parâmetros escolhidos no processo de codificação, tais como o número de tramas codificadas, o nível de decisão e o tipo de ôndula usado na DWT no espaço, está contida na *stream* resultante do codificador, não sendo necessária qualquer intervenção para sincronizar o descodificador.

Porém, por uma questão de verificação da informação, as caixas de diálogo relativas a esses parâmetros são seleccionadas (automaticamente) após a detecção das condições utilizadas. No início o utilizador é ainda inquirido acerca da obtenção de resultados do desempenho. Em caso afirmativo, surge uma janela de diálogo para a escolha dos referidos parâmetros, dos quais apenas estão disponíveis, o factor de compressão e a relação sinal ruído de pico (PSNR). Os resultados obtidos para o factor de compressão (relação entre a dimensão da sequência original e a sequência codificada) são apresentados numa janela simples. Os resultados relativos a PSNR são representados gráficamente numa nova janela, podendo ser guardados em memória, num ficheiro compatível com a aplicação Microsoft® Excel (**.xls**).

Através do menu **Miscellaneous**, está também disponível um conjunto de funcionalidades que permitem obter informação sobre os passos intermédios no processo intermédio na codificação. As opções disponíveis permitem visualizar o resultado da formação da estrutura de coeficientes da DWT, tais como a DWT no tempo e a DWT no espaço, e compará-las com as imagens originais do grupo de tramas considerado (ver figs. 7 a 9).



Figura 7. Imagens originais das tramas 5 a 8 (2º grupo de tramas) da sequência “Coastguard”.



Figura 8. Imagens resultantes da DWT no tempo (2 níveis) das tramas 5 a 8 da sequência “Coastguard”.

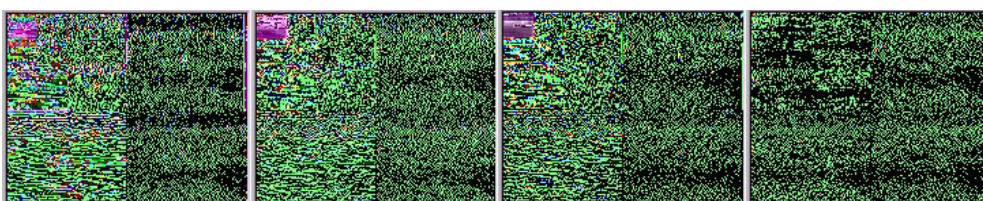


Figura 9. Imagens resultantes da DWT 2D (3 níveis) de cada trama transformada no domínio do tempo.

Na figura 8 regista-se uma maior intensidade nos coeficientes da primeira trama, face ao valor dos coeficientes das restantes tramas, que decrescem de valor à medida que o número da trama aumenta, estando os coeficientes menos significativos localizados na última trama. Este efeito resulta da característica da DWT na concentração de energia nos níveis mais altos. A primeira trama contém os coeficientes de aproximação no segundo nível, a segunda trama os coeficientes de detalhe referentes ao nível 2; as restantes tramas contêm os coeficientes de detalhe relativos ao nível 1. Na figura 9 regista-se a concentração de energia no nível mais alto, no canto superior esquerdo. Esta região concentra os coeficientes de aproximação, sendo as restantes regiões preenchidas pelos coeficientes de detalhe referentes aos restantes níveis.

Avaliação dos Resultados Experimentais

De modo a avaliar a qualidade das imagens, codificou-se um conjunto de sequências que estão classificadas no seio do grupo MPEG [11], e que são normalmente utilizadas para avaliar codificadores de vídeo. Os resultados obtidos na avaliação da aplicação indicam, em geral, melhor qualidade nas imagens codificadas usando a DWT no espaço com as ôndulas biortogonais (4,4) do que com a ôndula DB2. Este facto evidencia as melhores características das ôndulas biortogonais para este tipo de codificação. Nas figuras 10 e 11 apresentam-se alguns resultados experimentais considerando vários níveis de decisão. Verifica-se que a qualidade geral das imagens codificadas é aceitável, mesmo para factores de compressão mais elevados.

5. Conclusões

Neste artigo apresentou-se um codificador de vídeo desenvolvido, em ambiente Windows®, baseado na aplicação da transformada de ôndulas a 3D. Referiram-se as suas principais funcionalidades, e apresentaram-se alguns resultados experimentais, obtidos por codificação de algumas sequências de teste normalmente usadas na avaliação do desempenho de codificadores. Verifica-se que as imagens apresentam, em geral, uma qualidade subjectiva aceitável, mesmo para factores de compressão mais elevados. Os resultados obtidos na avaliação da aplicação indicam, em geral, melhores características das ôndulas biortogonais neste tipo de codificação. No codificador desenvolvido, a possibilidade de visualizar alguns resultados intermédios do processo de codificação é um aspecto importante, na medida em que permite demonstrar a decomposição em sub-bandas, através da DWT, conferindo-lhe, portanto, um cariz didático.



Figura 10. Imagens 153 a 156 da sequência “Foreman”: imagens originais (no topo), codificadas com nível 16 (ao meio) e codificadas com nível 64 (ao fundo).



Figura 11. Imagens 12, 14, 16 e 18 da sequência “Children”: imagens originais (no topo), codificadas com nível 32 (ao meio), e codificadas com nível 128 (ao fundo).

Referências

- [1] V. Bhaskaran, K. Konstantinides, “Image and Video Compression Standards”, *Kluwer Academic Press*, 2ª edição, 1996.
- [2] M. Marcellin, M. Gormish, A. Bilgin, M. Poliek, “An Overview of JPEG-2000”, *Proc. of IEEE Data Compression Conference*, pp. 523 – 541, 2000.
- [3] B. Usevitch, “A Tutorial on Modern Lossy wavelet Image Compression: Foundations of JPEG 2000”, *IEEE Signal Processing Magazine*, vol. 18, Nº 5, pp. 22 – 35, 2001.
- [4] S. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 7, pp. 674 – 693, 1989.
- [5] M. Vetterli, C. Herley, “Wavelets and Filter Banks: Theory and Design”, *IEEE Trans. on Signal Processing*, Vol. 4, Nº 8, 1995.
- [6] J. Shapiro, “Embedded Image Coding using Zerotrees of Wavelet Coefficients”, *IEEE Trans. on Signal Processing*, Vol. 41, Nº 12, pp. 3445 – 3462, 1993.
- [7] J. Salvado, “Codificação de Vídeo por Decomposição 3D Baseada na Transformada de Ondulas”, Tese de Mestrado, *IST/UTL*, Novembro de 2001.
- [8] Y. Chen, W. Pearlman, “Three-Dimensional Subband Coding of Video Using the Zero-Tree Method”, *Proc. SPIEs, 1996 Int. Symposium on Visual Communications and Image Processing*, Vol. 2727, 1996.
- [9] I. Daubechies, “Ten Lectures on Wavelets”, *SIAM – Society for Industrial and Applied Mathematics*, 1992.
- [10] J. Hollingworth, D. Butterfield, et. al. “C++ Builder 5.0 Developer’s Guide”, *SAMS*, 2000.
- [11] MPEG, “MPEG-4 Video Verification Model version 18.0”, doc. ISO/MPEG N3908, pp. 15–16, Pisa, 2001 (<http://www.mpeg.org>).