Video Coding by Applying the Extension of EZW to 3D and the Liftting Scheme

José Salvado
Electronics Depart., EST/IPCB
Castelo Branco, Portugal

Leonel Sousa
DEEC-IST / INESC-ID
Lisboa, Portugal

***Abstract* – This paper proposes a video codec based on the discrete wavelet transform (DWT). The video sequences are first splitted in groups of frames (GOF). Each GOF is then decomposed in the time domain, and in the spatial domain (for each frame) by applying the DWT. We use the lifting scheme to compute the DWT coefficientes. The coefficients of the 3D structure resulting from both DWT (in time and space domain) are than coded through the extension of EZW coding to 3D and the bitstream is finnaly passed through a Huffman coder, to achieve better compression ratios.**

The structure of the video codec and some experimental results are presented. Experimental results evaluate the quality of the coded images and the compression ratio achieved with the proposed codec.

***Keywords* – Video coding, wavelet transform, lifting scheme, EZW extension to 3D.**

I. Introduction

It is well known that the information in image and video signals is highly correlated. Video coding standards, such as MPEG (Motion Pictures Experts Group), H.261 and H.263, use the discrete cosine transform (DCT) [1] to exploit the spatial redundancy, applying the DCT to blocks of pixels, with fixed size, usually 8×8. However, correlation among pixels located at the bound regions in each block is not exploited and artificial frontiers arise in the image, causing a visual effect designated by “blocking effect”, particularly noted for high compression ratios.

The discrete wavelet transform (DWT) is another type of transform used to decorrelate signals, which registered a growing interest in the past few years. Unlike the DCT case, the DWT is applied to the all image, thus preventing from the “blocking effect”.

In the sense of multiresolution analysis (MRA) theory and wavelet filter banks [2,3], one can calculate the DWT coefficients through successive application of filters. The lifting scheme is an efficient method to compute the DWT coefficients [4] with the following advantages:

- the method is generic and much simpler;
- it can be up to two times faster, but still $O(n)$;
- it allows in-place computation;
- one obtains DWT^{-1} by reversing DWT steps.

The main idea associated to the lifting scheme is the construction of simple MRA blocks, going forward to construct successive and more complex MRA blocks. As in the MRA case, the signal decomposition by DWT application results in a hierarchical tree of coefficients.

Among various wavelet and subband coding schemes, we refer in particular the well known Embedded Zero-Tree Coding of Wavelet Coefficients (EZW) proposed by Shapiro [5], and the extension of the zero-tree method to 3D, for video coding purposes, proposed by Chen et. al. [6] (they used quadrature mirror filters - QMF).

In this paper, we propose a video codec that exploits the temporal and spatial redundancies by using the DWT in both domains. using the lifting scheme to calculate both DWTs. We code the DWT coefficients extending EZW to 3D, with adaptative thresholding, and the resulting bitstream is than passed through a Huffman coder. The paper is organized as follows. Section II presents the main aspects of the DWT and section III describes the video codec. The codec is evaluated in section IV and finally conclusions are presented in the last section.

II. Wavelet Transform through Lifting Steps

From MRA theory [2,3,7], wavelets and scaling functions are orthogonally related in each level of decomposition. In the case of orthogonal wavelets with compact support, there are two filters, $\{h_k\},\{g_k\}\in l^2$ (l^2 refers to $L^2(\mathbb{R})$ for discrete sequences), such that at any scale, any new function is obtained as linear combination of the basis functions. For the family of Daubechies orthogonal and biorthogonal wavelets with compact support [7], these filters have finite impulsive response (FIR) and can have linear phase characteristics. Also from MRA theory, one knows that the discrete wavelet transform (1D) can be represented by a bank of filters. An input signal $x[n]$ is passed through the low-pass and high-pass analysis filters, $\tilde{h}_k$ and $\tilde{g}_k$, respectively, followed by decimation by a factor of two (the odd samples are discharged). For a multilevel decomposition scheme, the output of each low-pass filter serves as the input for the new filters in the next level of decomposition. This process is applied iteratively, and after L levels it provides an approximation signal $a_L[n]$ with resolution reduced by a factor of 2^L, and a detail signal $d_L[n],\ldots,d_1[n]$. In the reconstruction side the process in inverted. Perfect reconstruction is possible if the analysis and synthesis filters verify the perfect reconstruction and non-aliasing conditions [3]. The same applies to the 2D DWT, as it is an extension of the unidimensional transform which can be applied independently to the and columns of the image.

With the lifting scheme, for any pair of coefficients a new pair arise, an approximation and a detail coefficient. Considering the filters represented by its Z transform, and its polyphase representation, one can assemble the sinthesys polyphase matrix and dual polyphase matrix, $\mathbf{P}(z)$ and $\tilde{\mathbf{P}}(z)$, respectively. Perfect reconstruction property is given by (where $\mathbf{I}$ is the identity matrix):

$$\mathbf{P}(z)\tilde{\mathbf{P}}(z^{-1})^T = \mathbf{I} \quad (1)$$

With this MRA scheme, starting with one filter pair (h,g), and using primal and dual lifting, it is possible to obtain the new filters in each stage of MRA [7], providing that the filter pair (h,g) is complementary, i.e., the determinant of its corresponding polyphase matrix is unitary. If the pair (h,g) is complementary, so is $(\tilde{h},\tilde{g})$. The algorithm for factoring the wavelet transform into lifting steps is based on the Euclidean algorithm for factoring Laurent polynomials, and primal and dual lifting theorems. Factoring the polyphase matrixes one obtains:

$$\mathbf{P}(z) = \prod_{j=1}^{m} \begin{bmatrix} 1 & s_j(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_j(z) & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & k^{-1} \end{bmatrix} \quad (2)$$

$$\tilde{\mathbf{P}}(z) = \prod_{j=1}^{m} \begin{bmatrix} 1 & 0 \\ -s_j(z^{-1}) & 1 \end{bmatrix} \begin{bmatrix} 1 & -t_j(z^{-1}) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k^{-1} & 0 \\ 0 & k \end{bmatrix} \quad (3)$$

Thus, any orthogonal wavelet transform associated to finite impulse response filters can be computed through the lifting scheme, starting with the lazy wavelet, followed by primal and dual lifting steps, eventually affected by scaling constants, as depicted in fig. 1 (similar to filter banks). We used the lifting scheme to calculate both DWTs, in time and in space domain, to support the video codec described in next section. For the DWT in time we used Daubechies 2 orthogonal wavelet filters. For the DWT in space, experimental evaluation was made in [8], and better results were obtained with biorthogonal (4.4) wavelet, correspomding to 9/7 filter pair.

III. Wavelet Transform based Video Codec

The video coding scheme based on 3D wavelet decomposition was implemented using four functional blocks: DWT in time, DWT in space, EZW 3D coding and Huffman coding, as depicted in figure 2. The frames of the input sequence are first segmented in groups of frames (GOF). For practical reasons, GOF must be a power of 2. Here, we consider $\mathrm{GOF} = 4$.

In the first block of the coder the temporal correlation is explored by performing the DWT (in time domain), using

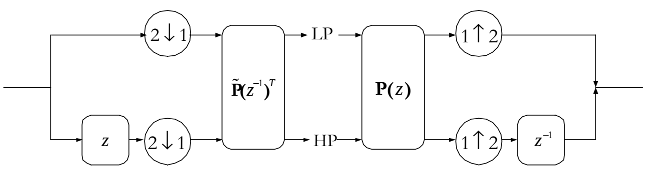

Fig. 1 – Polyphase representation of the wavelet transform

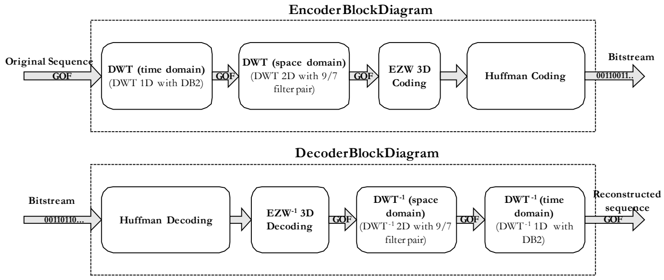

Fig. 2 – Structure of the wavelet transform based video codec

Daubechies 2 wavelet, with 2 levels of decomposition. The transformed GOF is then transformed in the space domain by using the 9/7 filter pair with 3 decomposition levels. From the application of both DWTs (in time and space domain), it results in a 3D tree structure of DWT coefficients shown in figure 3.

Due to the DWT characteristics, the coefficients located at the higher levels, in the first frame, are more significant than those located in lower levels and/or high order frames. The significance of the coefficients decreases as the level decreases inside each transformed frame, and as the frame order increases within the GOF. To code these coefficients, we extend the EZW to 3D, here designated by EZW 3D. The EZW coding scheme is strongly based in the high probability that the coefficients of the 3D structure, at the higher levels, are more significant than those located at lower levels. The scanning order of the coefficients is fixed, assuring that the coefficients located in the lower bands (higher levels) are scanned prior to those located at the higher bands (lower levels). This is performed from the first to the last frame, and inside each frame, form higher to lower levels, taking into account the dependencies of the coefficients in the tree.

With reference to a threshold, four symbols are used: positive POS (if the coefficient is found significant and positive); negative or NEG (if the coefficient is found significant and negative); zero-tree-root, or ZTR (if the coefficient is found insignificant, also as all its dependents) and isolated zero or IZ (if the coefficient is insignificant but there are dependents that are significant).

During EZW 3D coding two lists are maintained: a dominant list, which is of first-in first-out type (FIFO), and a subordinate list. In the dominant pass, the coefficients are scanned, and significant coefficients are coded, which corresponds to add their value and coordinates to the dominant list, and to fill with zero the

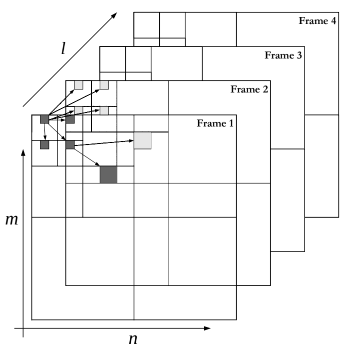

Fig. 3 – 3D hierarchical structure of DWT coefficients

corresponding position in the matrix (preventing from coding in a future dominant pass). After completing the dominant pass, a subordinate pass is performed, refining the coding, by using a threshold that is half of the one used in the dominant pass. After a complete scanning, the threshold is halved and a new dominant pass and subordinate pass are performed. Using a pseudo-code notation, the coding and decoding schemes are presented in figure 4.

The EZW 3D coded bitstream in than coded by a Huffman coder (we use JPEG Baseline Huffman coding).

IV. Codec Evaluation and Experimental Results

Test video sequences, in QCIF format, were coded, in order to evaluate the performance of the coder and the quality of the coded images. These sequences are classified within MPEG [9], and are usually used to evaluate H.263/MPEG video coders. To avoid degradation at the boundary regions, the sequences are symmetrically extended [10].

The evaluation of the codec is made through objective parameters, such as the compression ratio, the peak signal-to-noise ratio (PSNR), but we also subjectively evaluate the quality of the coded images by visual inspection. Compression ratios approximately varying from 10 to 70 are achieved, as shown in figure 5. Figures 6 and 7 show PSNR results in dBs for the luminance component. It is clearly shown that PSNR decreases as the threshold increases and also that there is a period of four (the GOF dimension) in PSNR curves, for the threshold values considered. This effect is caused by the organization of the coefficients of DWT (3D) and due to the errors introduced in EZW 3D coding. The corresponding visual effect is a higher degradation of the first image in the GOF, which is more visible when observing the sequence image by image. However, from a subjective point of view, when observing the sequence at 25 or 30 f.p.s. this effect is almost imperceptible, also because we compensate this effect by applying heuristic methods [8]. The subjective quality of the images can be evaluated by observing and comparing the original and the coded image, as shown in figure 8. One can observe that the Akiyo frame coded denotes acceptable quality. The same subjective quality was observed to other video sequences.

V. Conclusions

The application of wavelet transform to video coding is addressed in this paper. Its is shown that the DWT applied in time and space domains, concentrates energy of the video sequences in lower bands. The organization of the DWT coefficients facilitates the implementation of efficient coding schemes, with bit-rate scalability and control. Computer efficiency can be obtained by using the lifting scheme to calculate the DWT coefficients, with the advantage of in-place computation. The EZW expansion

```
EZW3D_Code(thrsh_min){ I_k=0; max_cof=0; do{
    max_cofI_k = max(I_k(i,j));
    if (max_cofI_k>max_cof) max_cof=max_cofI_k; I_k++;
     }while ( I_k < GOF );
   thrsh_init = 2^⌊log_2 max_coef⌋;
   thrsh_cur = thrsh_init; do {
     for(I_k=0; I_k<GOF, I_k++){
       if(max_cofI_k>=thrsh_cur)
       Cod_dominant_pass(I_k,I_k+1):}
       Cod_subordinate_pass(thrsh_curr/2);
       thrsh_cur=thrsh_cur/2;
    } while (thrsh_curr > thrsh_min);}

Cod_dominant_pass(I_k,I_k+1){ fifo_init();
   while(fifo_empty==false){ get_coded_coeff ();
     if (coded_coeff != ZTR){ code_new_coeff ();
       put_coded_coeff_in_fifo();
       if (coded_coeff==POS)||(coded_coeff==NEG){
         add_abs_coeff_to_subordinate_list();
         zerofill_coded_coeff_position ();} } } }

Cod_subordinate_pass (subord_thresh){ do{ if
coef_sub_list > subord_thresh{
       add_"1"_to_bitstream;
       coef_sub_list -= subord_thresh;}
     else add_"0"_to_bitstream;
   }while(!end_of_subordinate_list)}

EZW3D_Decode(thrsh_cur,thrsh_min,maxcof[GOF]){
   I_k=0; max_cof=0; do{ for(I_k=0; I_k<GOF, I_k++){
       max_coefI_k =maxcoef[I_k];
if((max_coefI_k>=thrsh_cur)&&(thrsh_cur>=thrsh_min))
      Dec_dominant_pass (I_k,I_k+1);}
      Dec_subordinate_pass(thrsh_cur/2);
      thrsh_cur = thrsh_cur/2;
   }while (thrsh_cur > thrsh_min);}

Dec_dominant_pass (I_k,I_k+1){  initialize_fifo();
   while(fifo_empty==false){
     get_coeff_from_bitstream ();
     if (coeff_from_bitstream!=ZTR){
       add_to_fifo();
if((coef==POS)||(coef==NEG))put_coef_in_matrix();}}
}

Dec_subordinate_pass (subord_thresh){
 get_element_from_subord_list (); do{
     bit=get_from_bitstream();
     if ((bit==1)&&(matrix_coeff==POS))
        matrix_coeff += subord_thrsh;
     else if ((bit==1)&&(matrix_coeff==NEG))
        matrix_coeff -= subord_thrsh;
   }while(!end_of_subordinate_list)}
      Cod_subordinate_pass (thrsh_cur/2);
       thrsh_cur = thrsh_cur/2;
   }while (thrsh_curr > thrsh_min);}
```

Fig. 4 – Pseudo-code for EZW 3D coding and decoding schemes

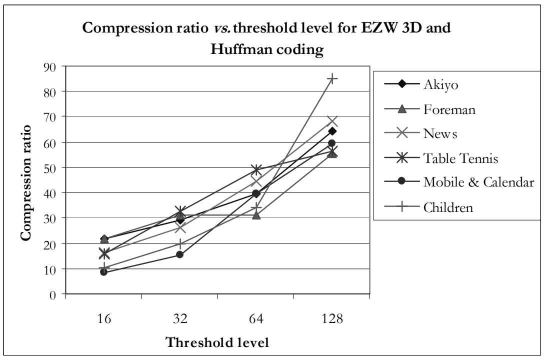

Fig. 5 – Variation of the compression ratio with the threshold level.

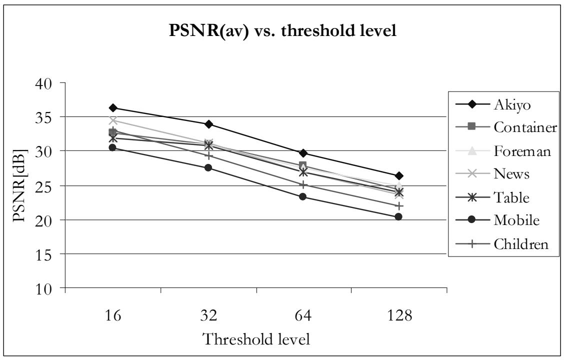

Fig. 6– Variation of the compression ratio with the threshold level.

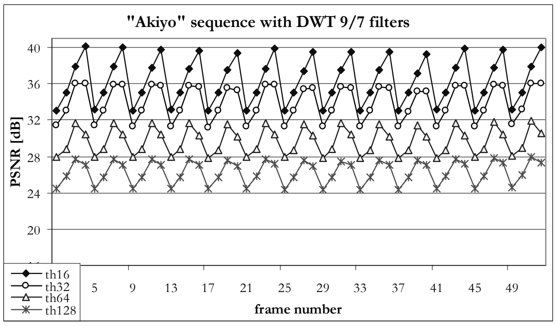

Fig. 7 – PSNR results for luminance component, for "Akiyo" video test sequence (QCIF).

to 3D is an important tool in effective coding. When combined with Huffman coding it allows compression ratios from 10 up to 70, depending on the chosen threshold. Higher compression ratios are possible by motion estimation techniques. However, in this case, motion estimation in the GOF must be done taking the first frame as reference.

Research work is being done in order to integrate motion estimation in the codec. Special attention is also given to the computing efficiency, namely in EZW 3D block, which appears to be the most time consuming.

REFERENCES

[1] V. Bhaskaran and K. Konstantinides, "Image and Video Compression Standards – Algorithms and Architectures", Kluwer Academic Press, 2nd Edition, 1996.

Fig. 8 – Akiyo frame 26: original image (top) and coded with threshold 32 (bottom).

[2] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, pp. 674 – 693, 1989.

[3] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design", IEEE Transactions on Signal Processing, vol 40, pp. 2207 – 2232, 1992.

[4] I. Daubechies and W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps", Joseph Fourier Analysis and Applications, vol. 4, Nº 3, pp. 245 – 267, 1998.

[5] J. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", IEEE Transactions on Signal Processing, vol. 41, Nº 12, pp. 3445 – 3462, December 1993.

[6] Y. Chen, and W. Pearlman, "Three-Dimensional Subband Coding of Video Using the Zero-Tree Method", Proc. of SPIE´s 1996 Int. Symp. on Visual Communications and Image Processing, vol. 2727, March 1996.

[7] I. Daubechies, "Ten Lectures on Wavelets", SIAM – Society for Industrial and Applied Mathematics, 1992.

[8] J. Salvado, "Video Coding by 3D Decomposition Based on the Wavelet Transform", MSc Thesis (only available in Portuguese language), Instituto Superior Técnico, November 2001.

[9] MPEG, "MPEG-4 Video Verification Model version 18.0", Doc ISO/MPEG N3908, pp. 15-16, Pisa, January 2001.

[10] C. Brislawn, "Classification of Nonexpansive Symmetric Extension Transforms for Multirate Filter Banks", Applied and Computational Harmonic Analisys, vol. 3, Nº 2, pp. 148 - 157, April 1993.