

LABORATOIRE



INFORMATIQUE, SIGNAUX ET SYSTÈMES  
DE SOPHIA ANTIPOLIS  
UMR 6070

# 3D SCAN BASED WAVELET TRANSFORM AND QUALITY CONTROL FOR VIDEO CODING

*Christophe PARISOT, Marc ANTONINI, Michel BARLAUD*

*Projet CREATIVE*

Rapport de recherche  
I3S/RR-2002-06-FR

mars 2002

Submitted in MMSP, Special issue in EURASIP  
Journal on Applied Signal Processing 2002

---

RÉSUMÉ :

MOTS CLÉS : Transformée en ondelettes au fil de l'eau, codage en ondelettes 3D, contrôle de qualité, compression vidéo.

---

ABSTRACT: Wavelet coding has been shown to be better than DCT coding. This method outperforms DCT JPEG codec and moreover allows scalability. 2D DWT can be easily extended to 3D and thus applied to video coding. However 3D subband coding of video suffers from two drawbacks. The first one is the memory complexity required for coding 3D blocks. The second one is the lack of temporal quality resulting in temporal blocking artifacts since DWT is performed on temporal groups of frames. In this paper we propose a new temporal scan based wavelet transform method for video coding combining the advantages of wavelet coding (performance, scalability), with acceptable reduced memory requirements and no additional CPU complexity. We also propose an efficient quality allocation procedure to ensure a constant quality over time.

KEY WORDS : Scan based DWT, 3D subband coding, quality control, video coding.

# 3D SCAN BASED WAVELET TRANSFORM AND QUALITY CONTROL FOR VIDEO CODING

C. Parisot, M. Antonini, M. Barlaud

I3S Laboratory, UMR 6070 (CNRS, University of Nice-Sophia Antipolis)

Bât. Algorithmes/Euclide, 2000, route des Lucioles - BP 121

F-06903 Sophia Antipolis Cedex, France

parisot,am,barlaud@i3s.unice.fr

**Abstract** - Wavelet coding has been shown to be better than DCT coding. This method outperforms DCT JPEG codec and moreover allows scalability. 2D DWT can be easily extended to 3D and thus applied to video coding [1, 2, 3, 4, 5, 6, 7]. However 3D subband coding of video suffers from two drawbacks. The first one is the memory complexity required for coding 3D blocks. The second one is the lack of temporal quality resulting in temporal blocking artifacts since DWT is performed on temporal groups of frames. In this paper we propose a new temporal scan based wavelet transform method for video coding combining the advantages of wavelet coding (performance, scalability), with acceptable reduced memory requirements and no additional CPU complexity. We also propose an efficient quality allocation procedure to ensure a constant quality over time.

**Keywords:** Scan based DWT, 3D subband coding, quality control, video coding.

## 1. Introduction

Quality control is a very important problem in video coding. Although 3D subband coding of video provides encouraging results comparatively to MPEG [4], its generalization suffers from significant memory requirements. One way to reduce memory requirements is to apply the temporal Discret Wavelet Transform (DWT) on 3D blocks coming from a temporal splitting of the sequence. But, this method introduces temporal blocking artifacts which result in undesirable jerks. In this paper, we propose new tools for 3D subband codecs to avoid these jerks and guarantee the output frames a constant quality over time.

Initially, 2D scan based wavelet transforms were first introduced in [8, 9] for on board satellite compression and by Ortega et al. in [10]. In section 2, we extend the 2D scan based method to video coding.

We propose a 3D scan based DWT method which allows the computation of the temporal wavelet decomposition of a sequence with infinite length with finite low memory requirements and no extra CPU complexity. Furthermore, the proposed wavelet transform provides higher quality control than 3D block based video compression schemes (avoiding jerks).

In section 3, we propose an efficient model-based quality control procedure. This bit allocation procedure controls output frames quality over time. This new quality

control procedure takes advantage of the model based rate allocation methods described in [11, 12]. Finally, section 4 presents experimental results obtained with our method.

## 2. 3D video wavelet transform

### 2.1 Principle

The method generally used to reduce memory requirements for large image coding is to split the image and then perform the transform on tiles such as JPEG with 8x8 DCT blocks and JPEG 2000 [13]. Unfortunately, coefficients are computed from periodical or symmetrical extensions of the signal. This results in undesirable temporal blocking visual artifacts. For video coding, the same blocking artifacts in the temporal direction (introduced by temporal tiling) result in jerks.

In this section, we propose a 3D wavelet transform framework for video coding that requires storing a minimum amount of data without any additional CPU complexity [14]. The frames of the sequence are acquired and processed on the fly to generate 3D wavelet coefficients. These wavelet coefficients are stored in memory until they are encoded.

#### Definitions of the temporal coherency and the buffer names

Let us consider a temporal interval (set of input frames). We define the set of its *temporally coherent wavelet coefficients* as the set of all coefficients, in all subbands, obtained by a filter (or convolution of filters) centered on one of the frames of this temporal interval. In this paper, we assume that encoding is allowed only when we have a temporally coherent set of wavelet coefficients. Temporal coherency improves encoder performances since it allows optimal bit allocation for wavelet coefficients of the same temporal interval.

The set of buffers used to perform the temporal wavelet transform will be called *filtering buffers*. These buffers alternately produce low and high frequency temporal wavelet coefficients. In the same way, we call *synchronization buffers* the set of buffers used to store output coefficients before their encoding.

### 2.2 Temporal scan based video DWT and delay

Let us consider the case of a 3D wavelet transform which can be split into a 2D DWT on each frame and an additional 1D DWT in the time direction [15]. In this paper, we focus on an efficient implementation of the temporal wavelet transform and propose a method independent of the choice of the spatial wavelet transform.

Each time a frame is received, we perform its 2D wavelet transform and send it into our scan based temporal wavelet transform system. We consider symmetrical filters with odd length since they are the most widely used in state of the art image compression algorithms [13, 16]. Let  $L = 2S + 1$  be the maximum length of the low-pass and high-pass filters. That is,  $S = \frac{L-1}{2}$ . Border effects will be reduced thanks to a symmetrical extension centered on the first data [17].

## Single stage DWT

Let us consider first a single stage of the wavelet transform. Receiving the  $S + 1^{st}$  2D transformed frame allows the computation of the first low-frequency coefficients. Indeed, the filtering buffer is symmetrically filled up after receiving the  $S + 1^{st}$  transformed frame to avoid side effects. Then, we alternately obtain high and low-frequency coefficients. Therefore, we have to wait  $S + 2$  frames to get one low-frequency and one high-frequency temporal frames. See Fig. 1 for the scheme of a single stage of the temporal wavelet transform in the case of a 5-tap filter.

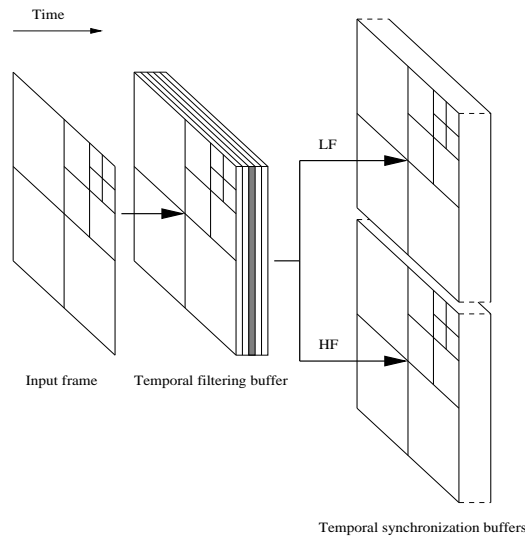


Figure 1: One level temporal scan based wavelet decomposition.

## Multi-stage DWT

Let us consider now the general scheme of an  $N$ -level temporal wavelet decomposition. For simplicity, we focus on the usual dyadic decomposition without additional high-frequency subband decomposition. Let us assume that decomposition levels are indexed from 1 to  $N$ , where level  $j$  corresponds to the coefficients produced by the  $j^{th}$  wavelet decomposition (level 0 is the sequence of all 2D wavelet transformed frames).

Let us compute the delay for a two-level wavelet decomposition. The first stage has to compute  $S + 2$  low-frequency temporal frames to get coefficients in all subbands of the second level. Since low-frequency coefficients are computed prior to high-frequency ones, the first stage has also computed  $S + 1$  high-frequency temporal frames. But, these  $(S + 2) + (S + 1) = 2S + 3$  output frames can only be computed after the delay of  $S$  frames to initialize the filtering buffer of the first stage. Thus, we have to wait  $3S + 3$  frames to get coefficients in all subbands of the last decomposition level.

To compute the delay for an  $N$ -level temporal wavelet decomposition, we define  $d_j$  as the number of frames required at the input of the  $j^{\text{th}}$  filtering buffer to get temporally coherent coefficients in all subbands. The processing of the first set of 3D subbands of wavelet coefficients will be possible after  $D = d_1$  frames have been received. Repeating the previous scheme iteratively, we find that  $d_j = d_{j+1} + (d_{j+1} - 1) + S$ . That is,  $d_j = 2d_{j+1} + S - 1$  for  $j = N-1..1$  and  $d_N = S + 2$ . Solving these equations, we find that the number of input frames required at level  $j$  before the first wavelet coefficients are available for processing is  $d_j = 2^{N-j} (2S + 1) - S + 1$ . Therefore, for an  $N$ -level temporal wavelet decomposition, the number of input frames needed to get a set of temporally coherent coefficients in all subbands is

$$D = 2^{N-1} (2S + 1) - S + 1 \quad (1)$$

Thus, the number of frames needed for the synchronization of the multi-stage decomposition increases exponentially with the number of decomposition levels. Fig. 2 shows the scheme of a three-level wavelet decomposition for  $L = 5$ . Each frame in the figure represents an image of 2D wavelet coefficients. Dark frames in the synchronization buffers are the set of coefficients which will be processed together (quantized and encoded) as soon as we have coefficients in all temporal frequency bands. This set of coefficients is temporally coherent. At the beginning of the sequence, we have to wait  $D$  input frames. Then, sets of temporally coherent coefficients will be available each  $2^N$  input frames.

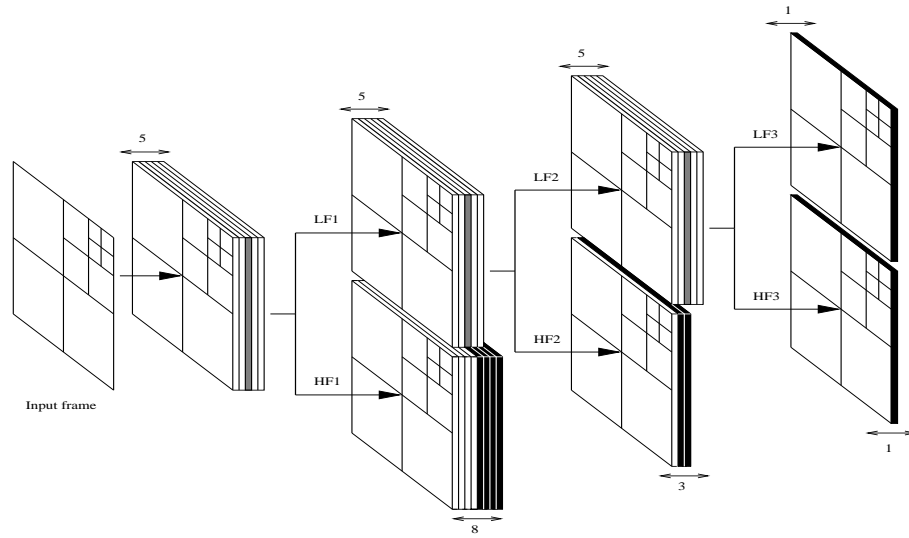


Figure 2: Three-level temporal scan based wavelet decomposition for 5-tap filters.

## 2.3 Memory requirements

The memory requirement is given by the sum of the number of frames in the  $N$  filtering buffers and the number of frames in the synchronization buffers. Namely,  $(2S + 1)N$  frames for the filtering buffers and  $2 + \sum_{j=2}^N (d_j - 1)$  frames for the synchronization buffers. Therefore, the memory requirement of this method is only

$$M = \text{framesize} \times [2^{N-1} (2S + 1) + (N - 1)S + N + 1] \quad (2)$$

for an  $N$ -level temporal wavelet transform. Table 1 and Table 2 show delays and memory requirements for the 9/7 and 5/3 filter banks. Depending on the application, the best compromise has to be found between filter length and the number of decompositions to determine the best frequency analysis with respect to memory or delay constraints. The CPU complexity of our temporal scan based DWT is exactly the same as to perform the usual 1D DWT in the temporal direction knowing the entire sequence.

Number of levels	9/7 DWT	5/3 DWT
1	6 frames	4 frames
2	15 frames	9 frames
3	33 frames	19 frames

Table 1: Number of input frames needed to get the first set of temporally coherent wavelet coefficients.

Number of levels	9/7 DWT	5/3 DWT
1	11 frames	7 frames
2	25 frames	15 frames
3	48 frames	28 frames

Table 2: Memory requirement of the scan based DWT system.

## 3. Model based temporal quality control

The bit allocation for the successive sets of temporally coherent coefficients can be performed with respect to either rate or quality constraints. In this section, we propose a new temporal quality control procedure to ensure constant quality over time. The quality measure is based on the mean squared error (MSE) between the compressed signal and the original one.

### 3.1 Principle of MSE allocation

The purpose of MSE allocation is to determine the optimal quantizers in each sub-band which minimize the total bitrate for a given output MSE. The 9/7 biorthogonal

filter bank is nearly orthogonal. Therefore, the mean squared error between the original image and the decoded one can be computed by a weighted sum of the mean squared quantization errors of each subband. We have

$$MSE_{output} = \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2 \quad (3)$$

with  $\#SB$  the number of 3D subbands,  $\sigma_{Q_i}^2$  the mean squared quantization error for subband  $i$  and  $\{\pi_i\}$  the weights used to take account of the non-orthogonality of the filter bank [18]. The weights  $\Delta_i$  are optional and can be used for frequency selection or distortion measures other than MSE. The output bitrate can be expressed as the following weighted sum:

$$R_{output} = \sum_{i=1}^{\#SB} a_i R_i \quad (4)$$

with  $R_i$  the output bitrate for subband  $i$  and  $a_i$  the weight of subband  $i$  in the total bitrate ( $a_i$  is the ratio of the size of subband  $i$  divided by the size of the sequence).

The subband quantizers are scalar quantizers with optimized deadzone size [12]. They are defined by the size of their zero quantization bin  $z$  and the size of the other quantization bins  $q$ . Therefore, the solution of our constrained problem is obtained thanks to Lagrangian operators by minimizing the following criterion:

$$J = \sum_{i=1}^{\#SB} a_i R_i(z_i, q_i) + \lambda \left( \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2(z_i, q_i) - D_T \right) \quad (5)$$

where  $D_T$  denotes the target output MSE and both  $\{R_i\}$  and  $\{\sigma_{Q_i}^2\}$  depend on the quantization steps and the deadzone sizes. The models used for the bitrate and distortion functions are described in the next subsection.

### 3.2 Rate and distortion models

Three different methods can be used to model the rate and distortion functions  $R_i$  and  $\sigma_{Q_i}^2$ :

- The first one - used in JPEG2000 [13] - consists in pre-quantizing the wavelet coefficients with a small quantization step and encode their bitplanes until the rate or distortion constraint (depending on the application) is verified. In this case, the quantization step of each wavelet coefficient can only be a product of the pre-determined quantization step times an integer power of two. The distortion and bitrate functions are exact but they are computed during the encoding process.
- The second method uses asymptotic models for both the distortion and the bitrate. As the asymptotic rate and distortion functions are simple, the minimum of the rate or distortion allocation criterion can be computed analytically. This

method is therefore the simplest one to get the quantization steps to apply in each subband. However, the asymptotic assumption is only true for high bitrate subbands.

- The third method uses non-asymptotic theoretical models for both rate and distortion [11, 12]. The rate and the distortion depend on the quantization step but also on the probability density function of the wavelet coefficients. Assuming that the probability density model is accurate, this method provides optimal rate distortion performances.

In each 3D subband, the probability density function is unimodal with zero mean. Thus, the wavelet coefficients can be approximated with generalized Gaussians [17, 19]. Therefore, we have

$$p(x) = ae^{-|bx|^\alpha} \quad (6)$$

with  $b = \frac{1}{\sigma} \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$  and  $a = \frac{b\alpha}{2\Gamma(1/\alpha)}$ . We also assume that wavelet coefficients are independent and identically distributed (i.i.d.) in each subband.

Let  $\Pr(m)$  be the probability of the quantization level  $m$ .

$$\Pr(m) = \int_{\frac{z}{2} + (|m|-1)q}^{\frac{z}{2} + |m|q} p(x) dx \quad (7)$$

for  $m \neq 0$  and

$$\Pr(0) = \int_{-\frac{z}{2}}^{+\frac{z}{2}} p(x) dx \quad (8)$$

Then, we can approximate the bitrate by the entropy of the output quantization levels:

$$R = - \sum_{m=-\infty}^{+\infty} \Pr(m) \log_2 \Pr(m) \quad (9)$$

According to [20], the best decoding value for the quantization level  $m$  is the centroid of its quantization bin  $\hat{x}_m = \text{sign}(m) \cdot \frac{\int_{\frac{z}{2} + (|m|-1)q}^{\frac{z}{2} + |m|q} xp(x) dx}{\Pr(m)}$  for  $m \neq 0$  and  $\hat{x}_0 = 0$ .

Then, the mean squared quantization error is given by

$$\sigma_Q^2 = \int_{-\frac{z}{2}}^{+\frac{z}{2}} x^2 p(x) dx + 2 \sum_{m=1}^{+\infty} \int_{\frac{z}{2} + (m-1)q}^{\frac{z}{2} + mq} (x - \hat{x}_m)^2 p(x) dx. \quad (10)$$

For generalized Gaussians, we can show that the bitrate depends only on the shape parameter  $\alpha$  and the ratios  $\frac{z}{\sigma}$  and  $\frac{q}{\sigma}$ . We have  $R = R(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma})$  [12]. We can also show that the quantization distortion can be written as  $\sigma_Q^2 = \sigma^2 D(\alpha, \frac{z}{\sigma}, \frac{q}{\sigma})$  where the function  $D$  depends only on  $\alpha$  and the ratios  $\frac{z}{\sigma}$  and  $\frac{q}{\sigma}$  [12].

Therefore, (5) becomes

$$J = \sum_{i=1}^{\#SB} a_i R\left(\alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i}\right) + \lambda \left( \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D\left(\alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i}\right) - D_T \right) \quad (11)$$



Figure 3: Curves of  $h\left(\frac{q}{\sigma}\right)$  versus  $D\left(\frac{q}{\sigma}\right)$  for different shape parameters  $\alpha$  of the generalized Gaussian distribution.

### 3.3 Optimal quantization for MSE control

Let  $f$  be any multi-variable function and  $x_k$  its  $k^{th}$  variable and define  $\frac{\partial f}{\partial x_k}$  as the derivative of  $f$  with respect to its  $k^{th}$  variable. To find the minimum of  $J$ , we differentiate it with respect to  $z_i$ ,  $q_i$  and  $\lambda$ . This provides the three following equations:

$$a_i \frac{\partial R}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) + \lambda \Delta_i \pi_i \sigma_i^2 \frac{\partial D}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) = 0 \quad (12a)$$

$$a_i \frac{\partial R}{\partial x_3} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) + \lambda \Delta_i \pi_i \sigma_i^2 \frac{\partial D}{\partial x_3} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) = 0 \quad (12b)$$

$$\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) - D_T = 0 \quad (12c)$$

Thus, the quantizers parameters  $\{z_i, q_i\}$  we are looking for must verify the following system of  $2 \times \#SB + 1$  equations and  $2 \times \#SB + 1$  unknowns:

$$\frac{\frac{\partial D}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)}{\frac{\partial R}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)} = \frac{\frac{\partial D}{\partial x_3} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)}{\frac{\partial R}{\partial x_3} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)} \quad (13a)$$

$$\frac{\frac{\partial D}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)}{\frac{\partial R}{\partial x_2} \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right)} = -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \quad (13b)$$

$$\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D \left( \alpha_i, \frac{z_i}{\sigma_i}, \frac{q_i}{\sigma_i} \right) = D_T \quad (13c)$$

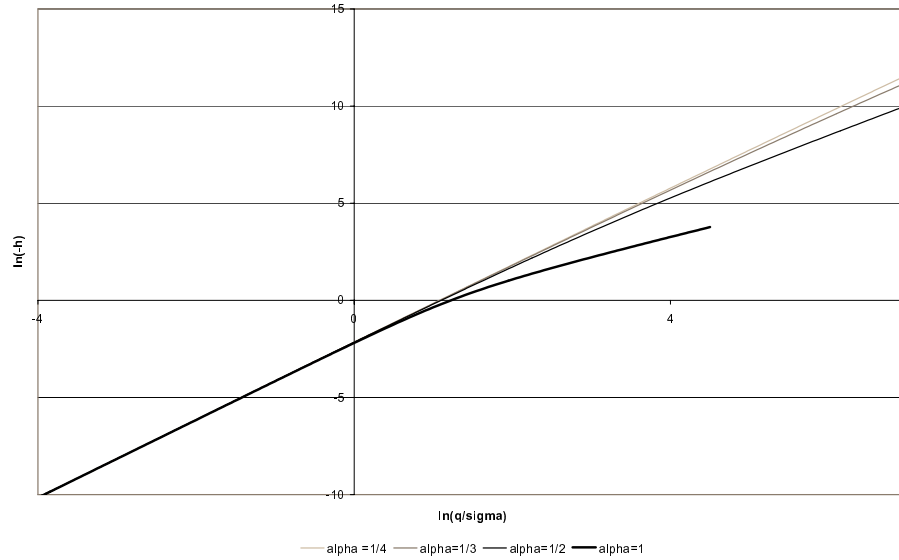


Figure 4: Curves of  $h$  for different shape parameters  $\alpha$  of the generalized Gaussian distribution.

The solution of Eq. (13a) provides the optimal relation between  $z_i$  and  $q_i$  for the MSE allocation procedure. Let us define  $g_{\alpha_i}$  the function such that  $\frac{z_i}{\sigma_i} = g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right)$  verifies Eq. (13a). The function  $g_{\alpha_i}$  is computed numerically and saved for different generalized Gaussian parameters (see Fig. 5).

Inserting this solution into Eq. (13b) and (13c), we get

$$h_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right) = \frac{\frac{\partial D}{\partial x_3} \left( \alpha_i, g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right), \frac{q_i}{\sigma_i} \right)}{\frac{\partial R}{\partial x_3} \left( \alpha_i, g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right), \frac{q_i}{\sigma_i} \right)} = -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \quad (14a)$$

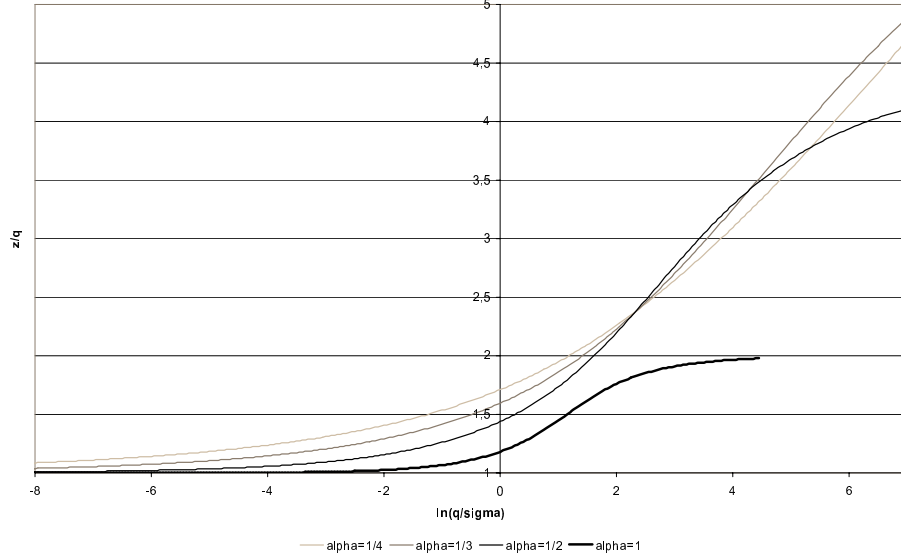


Figure 5: Curves of the optimal  $\frac{z}{q}$  ratio (solution of Eq. (13a)).

$$\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D \left( \alpha_i, g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right), \frac{q_i}{\sigma_i} \right) = D_T \quad (14b)$$

In order to simplify the notations, let us introduce the function  $h$  such that

$$h_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right) = \frac{\frac{\partial D}{\partial x_3} \left( \alpha_i, g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right), \frac{q_i}{\sigma_i} \right)}{\frac{\partial R}{\partial x_3} \left( \alpha_i, g_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right), \frac{q_i}{\sigma_i} \right)} \quad (15)$$

Thus, Eq. (14a) becomes

$$h_{\alpha_i} \left( \frac{q_i}{\sigma_i} \right) = -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \quad (16)$$

Therefore, the quantization steps  $q_i$  can be found by solving the following two equations:

$$\frac{q_i}{\sigma_i} = h_{\alpha_i}^{-1} \left( -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \right) \quad (17a)$$

$$\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D \left( \alpha_i, g_{\alpha_i} \left( h_{\alpha_i}^{-1} \left( -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \right) \right), h_{\alpha_i}^{-1} \left( -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2} \right) \right) = D_T \quad (17b)$$

Note that Eq. (17b) depends only on  $\lambda$ . Thus, solving (17b) provides the optimal  $\lambda$ . Once the optimal  $\lambda$  is known, we can solve the set of equations (17a) to get the optimal quantization steps  $\{q_i\}$ .

Unfortunately, there is no analytical formula for  $h^{-1}$ . To clear up this difficulty, we plot the parametric curve  $[D(\frac{q}{\sigma}); h(\frac{q}{\sigma})]$  for  $\alpha$  fixed (Fig. 3). We deduce from (17a) that this parametric curve is equivalent to the curve  $[D(h^{-1}(-\frac{\alpha_i}{\lambda\Delta_i\pi_i\sigma_i^2})); -\frac{\alpha_i}{\lambda\Delta_i\pi_i\sigma_i^2}]$ . Therefore, the proposed MSE allocation procedure is the following:

1. Lambda is given. For each 3D subband  $i$ , compute  $\ln(\frac{\alpha_i}{\lambda\Delta_i\pi_i\sigma_i^2}) = \ln(-h)$  and read the corresponding normalized mean squared error  $D_i$  from the curves shown in Fig. 3.
2. Compute  $|\sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D_i - D_T|$ . If it is lower than a given threshold, the constraint (17b) is verified and the current  $\lambda$  is optimal. Else, compute<sup>1</sup> a new  $\lambda$  and go back to step 1.
3. For each 3D subband  $i$ , compute  $\ln(\frac{\alpha_i}{\lambda\Delta_i\pi_i\sigma_i^2}) = \ln(-h)$  with the optimal  $\lambda$  and read  $\frac{q_i}{\sigma_i}$  from the curves shown in Fig. 4.  $q_i$  is the optimal quantization step for subband  $i$ .
4. For each 3D subband  $i$ , read the optimal deadzone size from the curves shown in Fig. 5.

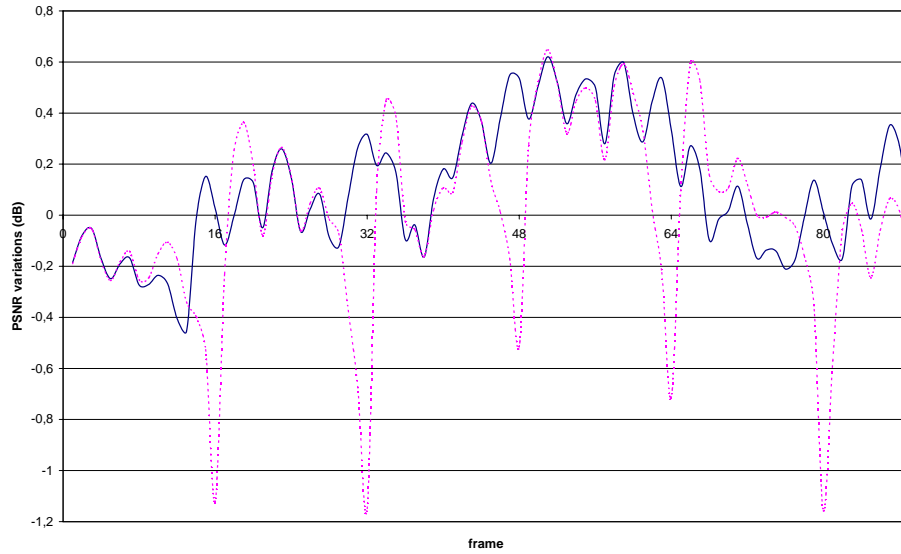


Figure 6: PSNR variations for the 3D scan based temporal DWT (continuous) and the 3D temporal tiling approach (dashed) on the 89 first luminance frames of the sequence Akiyo at 80 kbps (25 fps). 9/7 DWT with two levels of decomposition; bitrate control for groups of 16 frames.

<sup>1</sup>Any method (such as for example dichotomy) can be used.

## 4. Experimental results

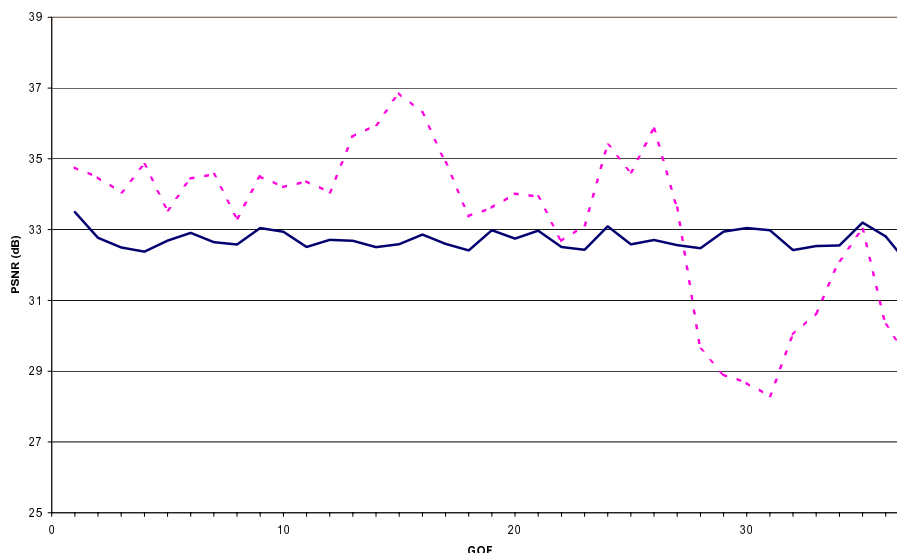


Figure 7: PSNR of each Group of height Frames (GOF) for the proposed quality control procedure (continuous) and a bitrate control procedure (dashed). The sequence is Foreman at 890 kbps (30 fps) in both cases.

To show the efficiency of our 3D scan based wavelet transform method in removing the temporal blocking artifacts (jerks), we have first extended EBWIC [11] to 3D data. The quantization and bit allocation procedures have been changed to take account of the deadzone optimization proposed in this paper. The quantized wavelet coefficients have been encoded using EBCOT's bit plane context-based arithmetic coder [21]. We have first encoded a sequence with the proposed 3D scan based temporal wavelet transform and a bitrate regulation for the temporally coherent coefficients of each group of 16 frames. Then, we have encoded the same sequence with the tiling approach, where the temporal wavelet coefficients and their encoding are performed on independent temporal blocks of 16 frames. Fig. 6 shows a global PSNR improvement of mean 0.11 dB with our approach. Furthermore, we have reduced the PSNR variance from 0.13 to 0.06. The peaks of the tiling approach fit with the artifacts produced at temporal tiles borders (jerks). Regarding the visual quality, the proposed method is also better since the annoying jerks are canceled.

Then, we have replaced the bitrate regulation by our new MSE allocation procedure. Fig. 7 shows that the quality of successive groups of 8 frames is well controlled. The PSNR variations are less than 1 dB with our method while they were up to 9 dB with a bitrate control procedure. The global sequence PSNR is 32.7 dB in both cases. Therefore, our method provides the same global rate distortion performances but ensures constant quality output frames. This results in better visual quality.

## 5. Conclusion

We have proposed a method for efficient quality control in video coding applications. This method combines the advantages of wavelet coding (performance, scalability) with minimum memory requirements and no additional CPU complexity. Compared to temporal tiling approaches often used to reduce memory requirements, our method avoids the jerks coming from temporal tiles. Furthermore, the new model based quality control procedure provides high performance PSNR control of the video output frames.

## References

- [1] G. Karlsson and M. Vetterli, "Three-dimensional subband coding of video", *Proc. of IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pp. 1100-1103, New York, NY, April 1988.
- [2] C.I. Podilchuck, N.S. Jayant and N. Farvardin. "Three-dimensional subband coding of video", *IEEE Trans. on Image Processing*, 1995.
- [3] S.J. Choi and J.W. Woods, "Motion-Compensated 3-D Subband Coding of Video", *IEEE Trans. on Image Processing*, vol. 8, no. 2, February 1999.
- [4] D. Taubman and A. Zakhor, "Multirate 3-D Subband Coding of Video", *IEEE Transactions on Image Processing*, vol. 3, no. 5, September 1994.
- [5] B.-J. Kim and W. A. Pearlman, "An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierarchical Trees (SPIHT)", *Proc. Data Compression Conference*, pp. 251-260, Snowbird, USA, March 1997.
- [6] B. Felts and B. Pesquet-Popescu, "Efficient Context Modeling in Scalable 3D Wavelet-Based Video Compression", *Proc. of IEEE Int. Conf. on Image Processing*, September 2000, Vancouver.
- [7] A. Wang, Z. Xiong, P. A. Chou and S. Mehrotra, "Three-Dimensional Wavelet Coding of Video with Global Motion Compensation", *IEEE Data Compression Conf.*, March 1999, Snowbird.
- [8] P. Charbonnier, M. Antonini and M. Barlaud, "Implantation d'une transformée en ondelettes 2D dyadique au fil de l'eau", *CNES contract report no. 896/95/CNES/1379/00*, Oct. 1995.
- [9] C. Parisot, M. Antonini, M. Barlaud, C. Lambert-Nebout, C. Lamy and G. Moury "On board strip-based wavelet image coding for future space remote sensing missions", *Proc. of IEEE International Geoscience and Remote Sensing Symposium*, July 2000, Honolulu.
- [10] C. Chrysafis and A. Ortega, "Line Based, reduced Memory, Wavelet Image Compression", *IEEE Trans. on Image Processing*, March 2000.
- [11] C. Parisot, M. Antonini and M. Barlaud, "EBWIC: A Low Complexity and Efficient Rate Constrained Wavelet Image Coder", *Proc. of IEEE Int. Conf. on Image Processing*, September 2000, Vancouver.

- [12] C. Parisot, M. Antonini and M. Barlaud, "Optimal Nearly Uniform Scalar Quantizer Design for Wavelet Coding", *Proc. of SPIE Int. Conf. on Visual Communications and Image Processing*, January 2002, San Jose.
- [13] ISO/IEC 15444-1:2000 "JPEG 2000 image coding system – Part 1".
- [14] C. Parisot, M. Antonini and M. Barlaud, "3D Scan Based Wavelet Transform for Video Coding", *Proc. of IEEE Workshop on Multimedia Signal Processing*, October 2001, Cannes.
- [15] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Electrical, Prentice Hall, Englewood Cliffs, NJ, 1995.
- [16] J.D. Villasenor, B. Bellzer and B. Liao, "Wavelet filter evaluation for image compression", *IEEE Trans. on Image Processing*, vol. 4, no. 8, pp. 1053-1060, Aug. 1995.
- [17] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using wavelet transform", *IEEE Trans. on Image Processing*, vol. 1, no. 2, pp. 205-220, 1992.
- [18] B. Usevitch, "Optimal Bit Allocation for Biorthogonal Wavelet Coding", *Proc. of Data Compression Conf.*, April 1996, Snowbird.
- [19] M. Barlaud, "Wavelets in Image Communication", *Elsevier*, 1994
- [20] A. Gersho and R. M. Gray, "Vector quantization and signal compression", *Kluwer Academic Publishers*, 1992.
- [21] D. Taubman, "High performance scalable image compression with EBCOT", *IEEE Trans. on Image Processing*, vol. 9, no. 7, July 2000.