EURASIP Journal on Applied Signal Processing 2003:1, 56–65

3D Scan-Based Wavelet Transform and Quality Control for Video Coding

Christophe Parisot
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000,
route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: parisot@i3s.unice.fr

Marc Antonini
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000,
route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: am@i3s.unice.fr

Michel Barlaud
Laboratoire I3S, UMR 6070 (CNRS, Université de Nice-Sophia Antipolis) Bât. Algorithmes/Euclide 2000,
route des Lucioles, BP 121 F-06903 Sophia Antipolis Cedex, France
Email: barlaud@i3s.unice.fr

Received 4 March 2002 and in revised form 17 October 2002

Wavelet coding has been shown to achieve better compression than DCT coding 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 is the amount of memory required for coding large 3D blocks; the second is the lack of temporal quality due to the sequence temporal splitting. In fact, 3D block-based video coders produce jerks. They appear at blocks temporal borders during video playback. 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, no additional CPU complexity, and avoiding jerks. We also propose an efficient quality allocation procedure to ensure a constant quality over time.

Keywords and phrases: scan-based DWT, 3D subband coding, quality control, video coding.

1. INTRODUCTION

Although 3D subband coding of video [1, 2, 3, 4, 5] provides encouraging results compared to MPEG [6, 7, 8, 9], its generalization suffers from significant memory requirements. One way to reduce memory requirements is to apply the temporal discrete wavelet transform (DWT) on 3D blocks coming from a temporal splitting of the sequence. But this block-based DWT method introduces temporal blocking artifacts which result in undesirable jerks during video playback. In this paper, we propose new tools for 3D subband codecs to guarantee the output frames a constant quality over time.

Scan-based 2D wavelet transforms were first suggested for on-board satellite compression in [10, 11] and by Chrysafis and Ortega in [12].

In Section 2, we propose a 3D scan-based DWT method and a 3D scan-based motion-compensated lifting DWT for video coding. The method allows the computation of the temporal wavelet decomposition of a sequence with infinite length using little memory and no extra CPU. 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 the output frames quality over time. This new quality-control procedure takes advantage of the model-based rate allocation methods described in [13].

Finally, Section 4 presents experimental results obtained by 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 8 × 8 DCT blocks or JPEG2000 [14]. Unfortunately, the coefficients are computed from periodic or symmetrical extensions of the signal.

This results in undesirable blocking artifacts. For video coding, the same blocking artifacts in the temporal direction (introduced by temporal splitting) 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 [15]. The frames of the sequence are acquired and processed on the fly.

Definitions of the temporal coherence and the buffer names

We 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 any 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 coherence improves the encoder performance 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 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

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 [16]. In this paper, we focus on an efficient implementation of the temporal wavelet transform and we 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 image compression algorithms [14, 17]. To simplify, we also suppose that the low-pass filter is longer than the high-pass one. Let $L = 2S + 1$ be the length of the low-pass filter with $S \geq 2$. We want to design components that can be easily reused for any wavelet decomposition tree. Therefore, the memory used for the filtering buffers is supposed to be internal and cannot be shared with other filtering buffers nor with the synchronization buffers for wavelet coefficients storage. We propose a method that minimizes the total memory requirements for FIR filtering.

2.2.1 Single-stage DWT

We first consider a single stage of the temporal wavelet transform.

The length of the low-pass filter is L. Therefore, we need L frames of 2D wavelet coefficients in memory to compute one frame of low-frequency temporal wavelet coefficients. The high-pass filter is shorter. Thus, our filtering buffer must contain exactly L frames of 2D wavelet coefficients. Consequently, filtering buffers are FIFO with length L. Figure 1 shows the scheme for a single stage of a 5/3 temporal wavelet decomposition. The filtering buffer contains five frames of 2D wavelet coefficients. The synchronization buffers are used to store output 3D wavelet coefficients until we get a temporally coherent set of 3D wavelet coefficients.

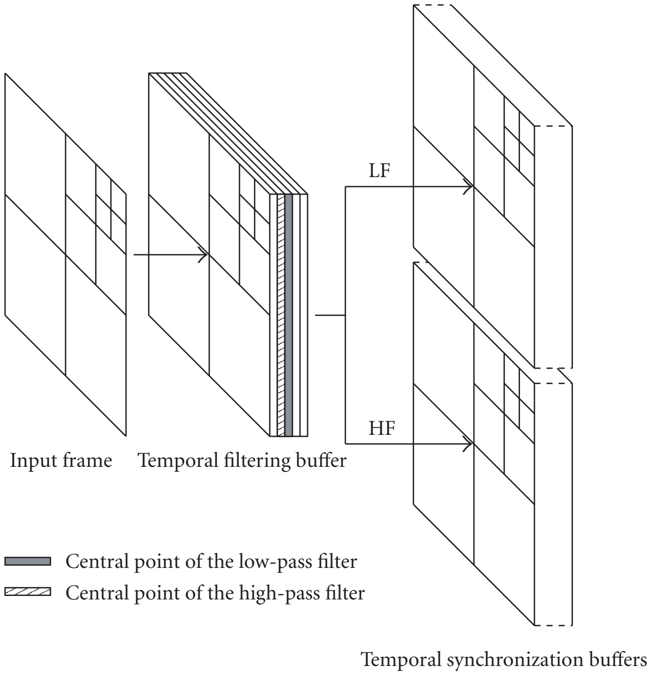
LF
HF
Input frame
Temporal filtering buffer
▆ Central point of the low-pass filter
▨ Central point of the high-pass filter
Temporal synchronization buffers

Figure 1: One-level temporal scan-based wavelet decomposition for the 5/3 filter bank.

When the $(S + 1)$st 2D transformed frame is received, the filtering buffer is symmetrically filled up in order to avoid side effects. The central frame is the 2D wavelet transform of the first image of the sequence. We can compute the first low-frequency temporal coefficients applying the low-pass filter to the central frame of the filtering buffer (gray frame in Figure 1). The first high-frequency temporal coefficients must be computed on the second 2D transformed frame. This frame (hatched frame in Figure 1) and all its necessary neighbours are already present in the filtering buffer since the high-pass filter is shorter than the low-pass one. Therefore, the high-frequency temporal wavelet coefficients can also be computed without additional input frame.

Finally, we have to wait for only $S + 1$ input frames to get one low-frequency and one high-frequency temporal frames of wavelet coefficients. Then, for each pair of input frames, we can compute both low-frequency and high-frequency coefficients. Each pair of low- and high-frequency frames is a set of temporally coherent wavelet coefficients. Therefore, we need $S+1$ input frames to get the first set of temporally coherent wavelet coefficients and $S+1+2(n-1) = S+2n-1$ input frames to get a set of n low-frequency and n high-frequency output frames.

When the input sequence is finished, input frames are replaced by a symmetrical extension using the frames present in the filtering buffer in order to flush it.

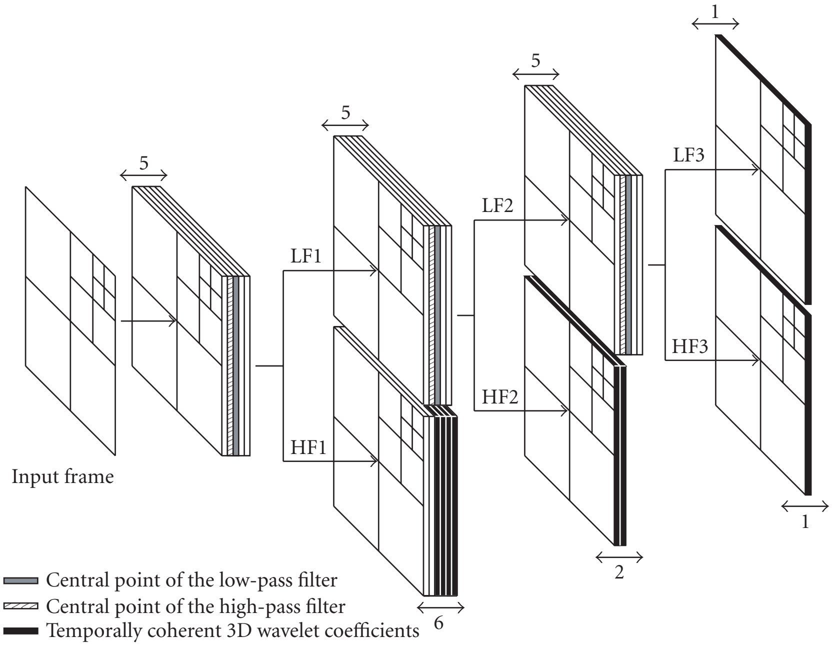
5
5
5
1
LF1
HF1
LF2
HF2
LF3
HF3
Input frame
2
1
6
▬ Central point of the low-pass filter
▨ Central point of the high-pass filter
■ Temporally coherent 3D wavelet coefficients

Figure 2: Three-level temporal scan-based wavelet decomposition for the 5/3 filter bank.

2.2.2 Multistage DWT

We now consider the general scheme of an N-level temporal wavelet decomposition. We focus only on the usual dyadic decomposition without additional high-frequency subband decomposition. We assume that decomposition levels are indexed from 1 to N, where level j corresponds to the coefficients produced by the jth wavelet decomposition (level 0 is the sequence of all 2D wavelet transformed frames).

We compute the encoding delay for a two-level wavelet decomposition. The first stage has to compute $S + 1$ low-frequency temporal frames to get coefficients in both low-frequency and high-frequency subbands of the second level. In the same time, the first stage has also computed $S+1$ high-frequency temporal frames. But, from Section 2.2.1, we know that these $S+1$ low-frequency and $S+1$ high-frequency output frames of 3D wavelet coefficients can only be computed after the delay of $S + 2(S + 1) - 1$ frames. Thus, we have to wait for $3S+1$ frames to get one frame of 3D coefficients in all subbands of the second decomposition level and $S+1$ frames of 3D coefficients in the first level. Notice that for temporal coherence, we need only the first two frames among the $S+1$ of the first 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 jth filtering buffer to get temporally coherent coefficients in all subbands. The processing of the first set of 3D subbands of temporally coherent wavelet coefficients will be possible after $D = d_1$ frames have been received. From Section 2.2.1, we know that $d_j = S + 2d_{j+1} - 1$ for $j \in \{1, \dots, N-1\}$ and $d_N = S+1$. 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 $d_j = (2^{N+1-j} - 1)S + 1$. Therefore, for an N-level temporal wavelet decomposition, the number of input frames needed to get the first set of temporally coherent wavelet coefficients is

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

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

Number of levels (N)	9/7 DWT	5/3 DWT
1	5	3
2	13	7
3	29	15

Thus, the number of frames needed for the synchronization of the multistage decomposition increases exponentially with the number of decomposition levels. Figure 2 shows the scheme of a three-level wavelet decomposition for $S = 2$. Dark frames in the synchronization buffers are the set of coefficients which will be processeded 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 for D input frames. Then, sets of temporally coherent coefficients will be available each 2^N input frames. Table 1 shows the number of input frames needed to get the first set of temporally coherent wavelet coefficients for two widely used filter banks. This table shows that a three-level decomposition introduces an encoding delay of less than one second with the 9/7 filter bank and only half a second with the 5/3 filter bank. In 3D block-based video coders, the delay is equal to the size of the temporal block. As blocks are larger in order to minimize the number of jerks, the delay is more important for 3D block-based wavelet transform video coders.

2.3. Memory requirements

Memory requirements are given by the sum of the number of frames in the N filtering buffers and the number of frames in the synchronization buffers.

The memory requirements for the filtering buffers are equal to $(2S+1)N$ frames.

The synchronization buffers of the last decomposition level must contain one frame of 3D wavelet coefficients for both the low-frequency and high-frequency subbands. For the jth decomposition level ($j < N$), d_{j+1}, low-frequency outputs need to be computed and, in the same time, d_{j+1} high-frequency outputs can be computed. As we know that temporal coherence requires less than d_{j+1} 3D frames of wavelet coefficients at level j, we can decide to delay the computation of the last computable high-frequency coefficients until the new set of temporally coherent 3D wavelet coefficients has been encoded. Once the set of temporally coherent coefficients has been encoded, we compute all the high-frequency coefficients for levels 1 to $N-1$ and send them into the synchronization buffers. Then, the on-the-fly wavelet transform can resume normally. This trick allows to spare one frame in the memory requirements of each synchronization buffer for levels 1 to $N-1$. Thus, the memory requirements for the synchronization buffers are limited to $2+\sum_{j=1}^{N-1}(d_{j+1}-1)$.

We need to store $M_S = (2^N - N - 1)S + 2$ frames of coefficients for all the synchronization buffers. Therefore, the total memory requirements of this method are

$$M = (2^N + N - 1)S + N + 2 \tag{2}$$

frames, for an N-level temporal wavelet transform with filter length $L = 2S+1$. When memory can be shared between filtering buffers and synchronization buffers, the total memory requirements are limited to

$$M = (2^N + N - 1)S + 1 \tag{3}$$

frames. See [18] for complete memory requirements formulae.

Tables 2 and 3 show the total memory requirements for the 9/7 and 5/3 filter banks, respectively, for independent and shared buffers.

Memory requirements increase as an exponential function of the resolution N and as a linear function of the filter length.

Note that, for the same memory requirements (e.g., 48 frames) and three levels of the 9/7 DWT decomposition with a frame rate of 30 fps, the encoding delay for temporal block-based video coders is equal to 1.6 second while it is 0.97 second in our case (from Table 1). Furthermore, block-based video coders have jerks for each group of 48 frames while our method avoids these annoying artifacts.

Table 2: Memory requirements (2), in terms of frames, of the scan-based DWT system including both filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	11	7
2	24	14
3	45	25

Table 3: Memory requirements (3), in terms of frames, of the scan-based DWT system including both filtering and synchronization buffers when memory can be shared between filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	9	5
2	21	11
3	41	21

The CPU complexity of our temporal scan-based DWT is exactly the same as to perform the regular 1D DWT in the temporal direction on the entire sequence.

2.4. Scan-based motion compensated lifting

The main drawback of the 3D scan-based DWT is that it does not take motion compensation into account. 3D motion compensated lifting is an efficient tool to take account of motion in video [4, 6, 9, 19, 20, 21].

Thus, we propose a new 3D scan-based motion compensated lifting scheme [18, 22]. This method combines the benefits of scan-based filtering, block-based coding, and quality control [22].

When filtering and synchronization buffers are independent, the total memory requirements become

$$M = (2^N - N - 1)S + \beta N + 2, \tag{4}$$

where β is a parameter depending on the filter, $\beta = 6$ for the 9/7 Daubechies DWT [23], and $\beta = 4$ for the 5/3 DWT. When memory can be shared between filtering and synchronization buffers, the total memory requirements are limited to

$$M = (2^N - N - 1)S + (\beta - 1)N + 1. \tag{5}$$

Complete memory requirements computation can be found in [18]. The scan-based motion compensated lifting scheme saves memory compared to the regular filter banks implementation. Furthermore, our method does not increase the CPU complexity compared to the usual lifting implementation.

Tables 4 and 5 show the memory requirements for scan-based motion compensated lifting video coders, respectively, for independant and shared buffers.

Thus, the scan-based motion compensated lifting scheme saves 12 to 33% memory (Tables 2 and 4 or Tables 3 and 5) and takes account motion compensation.

Table 4: Memory requirements (4), in terms of frames, of the scan-based motion compensated lifting DWT system including both filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	8	6
2	18	12
3	36	22

Table 5: Memory requirements (5), in terms of frames, of the scan-based motion compensated lifting DWT system including both filtering and synchronization buffers when memory can be shared between filtering and synchronization buffers.

Number of levels (N)	9/7 DWT	5/3 DWT
1	6	4
2	15	9
3	32	18

A 32-frames memory (which is a reasonable GOP memory) allows to implement a 3D scan-based motion compensated lifting with efficient filters (9/7) and three-level decomposition.

The scan-based motion compensated lifting also removes jerks with quality control.

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 both cases, the goal is to find a set of quantizers to apply in each subband, which performances lie on the convex hull of the global rate-distorsion curve [24, 25, 26, 27].

Three different methods can be used to model the rate and distortion.

(i) The first one—used in JPEG2000 [14]—consists in prequantizing the wavelet coefficients with a small predetermined quantization step and encodes their bitplanes until the rate or distortion constraint (depending on the application) is verified. In this method, the quantization step of each wavelet coefficient can only be a product of the chosen quantization step multiplied by an integer power of two. The distortion and bitrate functions are exact but they are computed during the encoding process.

(ii) 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.

(iii) We have proposed to use nonasymptotic theoretical models for both rate and distortion [13]. 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 this section, we propose a new nonasymptotic temporal *quality* control procedure to ensure constant quality over time. The quality measure is based on the mean square error (MSE) between the compressed signal and the original one.

3.1. Principle of the model-based MSE allocation

The purpose of MSE allocation is to determine the optimal quantizers in each subband which minimize the total bitrate for a given output MSE. Since the 9/7 biorthogonal filter bank is nearly orthogonal, the MSE 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

$$\mathrm{MSE}_{\mathrm{output}} = \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2, \tag{6}$$

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 nonorthogonality of the filter bank [28]. The weights Δ_i are optional and can be used for frequency selection or distortion measures. The output bitrate can be expressed as the following weighted sum:

$$R_{\mathrm{output}} = \sum_{i=1}^{\#SB} a_i R_i, \tag{7}$$

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 uniform scalar quantizers. They are defined by their quantization bins q_i. The solution of our constrained problem is obtained thanks to Lagrangian operators by minimizing the following criterion:

$$J(\{q_i\}, \lambda) = \sum_{i=1}^{\#SB} a_i R_i(q_i) + \lambda \left(\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_{Q_i}^2(q_i) - D_T \right), \tag{8}$$

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

3.2. Rate and distortion models

In each 3D subband, the probability density function of the wavelet coefficients is unimodal with zero mean and can be approximated with generalized Gaussian [23, 29]. Therefore, we have

$$p_{\alpha,\sigma}(x) = a e^{-|bx|^{\alpha}}, \tag{9}$$

with $b = (1/\sigma)\sqrt{\Gamma(3/\alpha)/\Gamma(1/\alpha)}$ and $a = b\alpha/2\Gamma(1/\alpha)$. We also assume that wavelet coefficients are independent and identically distributed (i.i.d.) [13] in each subband.

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

$$\Pr(m) = \int_{(|m|-1/2)q}^{(|m|+1/2)q} p_{\alpha,\sigma}(x)dx, \tag{10}$$

for $m \neq 0$ and

$$\Pr(0) = \int_{-q/2}^{+q/2} p_{\alpha,\sigma}(x)dx. \tag{11}$$

From (10) and (11), we can approximate the bitrate R by the entropy of the output quantization levels

$$R = -\sum_{m=-\infty}^{+\infty} \Pr(m)\log_2 \Pr(m). \tag{12}$$

The best coding value for the quantization level m [30] is the centroid of its quantization bin

$$\hat{x}_m = \operatorname{sign}(m) \times \frac{\int_{(|m|-1/2)q}^{(|m|+1/2)q} x p_{\alpha,\sigma}(x)dx}{\Pr(m)}, \tag{13}$$

for $m \neq 0$ and $\hat{x}_0 = 0$.

The mean squared quantization error is given by

$$\sigma_Q^2 = \int_{-q/2}^{+q/2} x^2 p_{\alpha,\sigma}(x)dx + 2\sum_{m=1}^{+\infty}\int_{(m-1/2)q}^{(m+1/2)q} \left(x - \hat{x}_m\right)^2 p_{\alpha,\sigma}(x)dx. \tag{14}$$

Inserting the value of $\hat{x}_m$ into (14), we get

$$\sigma_Q^2 = \sigma^2 - 2\sum_{m=1}^{+\infty} \frac{\left(\int_{(m-1/2)q}^{(m+1/2)q} x p_{\alpha,\sigma}(x)dx\right)^2}{\int_{(m-1/2)q}^{(m+1/2)q} p_{\alpha,\sigma}(x)dx}. \tag{15}$$

Proposition 1. *When $p_{\alpha,\sigma}$ is a generalized Gaussian distribution with standard deviation σ and shape parameter α, there is a family of functions $f_{n,m}$ which verifies*

$$\begin{aligned} \int_{-q/2}^{+q/2} x^n p_{\alpha,\sigma}(x)dx &= \sigma^n f_{n,0}\left(\alpha, \frac{q}{\sigma}\right), \\ \int_{(m-1/2)q}^{(m+1/2)q} x^n p_{\alpha,\sigma}(x)dx &= \sigma^n f_{n,m}\left(\alpha, \frac{q}{\sigma}\right) \quad \forall m > 0 \end{aligned} \tag{16}$$

with

$$\begin{aligned} f_{n,0}\left(\alpha, \frac{q}{\sigma}\right) &= \int_{-(1/2)(q/\sigma)}^{+(1/2)(q/\sigma)} x^n p_{\alpha,1}(x)dx, \\ f_{n,m}\left(\alpha, \frac{q}{\sigma}\right) &= \int_{(m-1/2)(q/\sigma)}^{(m+1/2)(q/\sigma)} x^n p_{\alpha,1}(x)dx. \end{aligned} \tag{17}$$

Proof of Proposition 1 is given in [18].

Therefore, the bitrate R and the quantization distortion σ_Q^2 depend only on the shape parameter α and the ratio q/σ,

$$R = R\left(\alpha, \frac{q}{\sigma}\right), \qquad \sigma_Q^2 = \sigma^2 D\left(\alpha, \frac{q}{\sigma}\right) \tag{18}$$

with

$$\begin{aligned} R\left(\alpha, \frac{q}{\sigma}\right) = &-f_{0,0}\left(\alpha, \frac{q}{\sigma}\right)\log_2 f_{0,0}\left(\alpha, \frac{q}{\sigma}\right) \\ &- 2\sum_{m=1}^{+\infty} f_{0,m}\left(\alpha, \frac{q}{\sigma}\right)\log_2 f_{0,m}\left(\alpha, \frac{q}{\sigma}\right), \end{aligned} \tag{19}$$

$$D\left(\alpha, \frac{q}{\sigma}\right) = 1 - 2\sum_{m=1}^{+\infty} \frac{f_{1,m}(\alpha, q/\sigma)^2}{f_{0,m}(\alpha, q/\sigma)}. \tag{20}$$

3.3. Optimal model-based quantization for MSE control

Therefore, the goal is to find the quantization steps $\{q_i\}$ and λ which minimize

$$J(\{q_i\}, \lambda) = \sum_{i=1}^{\#SB} a_i R\left(\alpha_i, \frac{q_i}{\sigma_i}\right) + \lambda\left(\sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D\left(\alpha_i, \frac{q_i}{\sigma_i}\right) - D_T\right). \tag{21}$$

We differentiate the criterion with respect to q_i and λ. This provides the following equations:

$$\begin{aligned} a_i \frac{\partial R}{\partial \widetilde{q}_i}(\alpha_i, \widetilde{q}_i) + \lambda \Delta_i \pi_i \sigma_i^2 \frac{\partial D}{\partial \widetilde{q}_i}(\alpha_i, \widetilde{q}_i) &= 0, \quad \forall i, \\ \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D(\alpha_i, \widetilde{q}_i) - D_T &= 0, \end{aligned} \tag{22}$$

where $\widetilde{q}_i = q_i/\sigma_i$.

Thus, the quantizers parameters $\{q_i\}$ must verify the following system of $\#SB + 1$ equations and $\#SB + 1$ unknowns:

$$\begin{aligned} \frac{(\partial D/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}{(\partial R/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)} &= -\frac{a_i}{\lambda \Delta_i \pi_i \sigma_i^2}, \quad \forall i, \\ \sum_{i=1}^{\#SB} \Delta_i \pi_i \sigma_i^2 D(\alpha_i, \widetilde{q}_i) &= D_T. \end{aligned} \tag{23}$$

In order to simplify the notation, write

$$h_{\alpha_i}(\widetilde{q}_i) = \frac{(\partial D/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}{(\partial R/\partial \widetilde{q}_i)(\alpha_i, \widetilde{q}_i)}, \tag{24}$$

where

$$h_\alpha(\widetilde{q}) = \frac{A}{B}\ln 2, \tag{25}$$

where $A = \sum_{m=1}^{+\infty}(2(\partial f_{1,m}/\partial\widetilde{q})(\alpha, \widetilde{q}) f_{1,m}(\alpha, \widetilde{q}) f_{0,m}(\alpha, \widetilde{q}) - f_{1,m}(\alpha, \widetilde{q})^2 (\partial f_{0,m}/\partial\widetilde{q})(\alpha, \widetilde{q}))/f_{0,m}(\alpha, \widetilde{q})^2$, $B = (p_{\alpha,1}(\widetilde{q}/2)/2) \times [\ln f_{0,0}(\alpha, \widetilde{q}) + 1] + \sum_{m=1}^{+\infty}(\partial f_{0,m}/\partial\widetilde{q})(\alpha, \widetilde{q})[\ln f_{0,m}(\alpha, \widetilde{q}) + 1]$

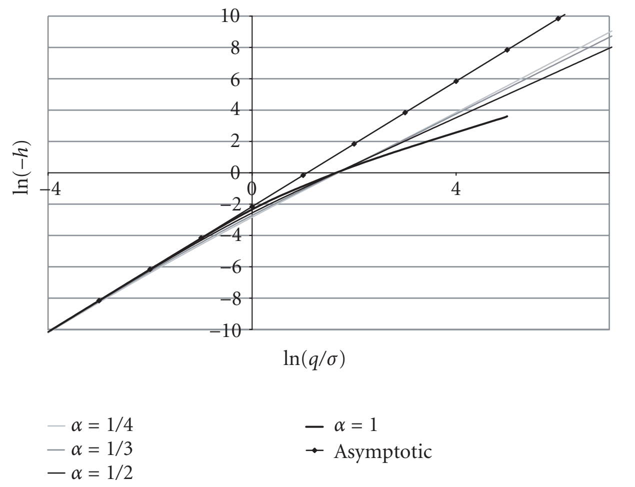

FIGURE 3: Tables of $\ln(-h(q/\sigma))$ for different shape parameters α of the generalized Gaussian distribution.

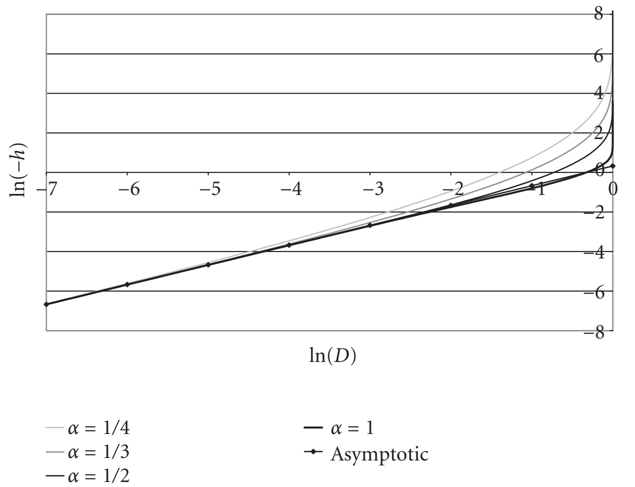

FIGURE 4: Tables of $\ln(-h) = \ln(a_i/\lambda\pi_i\sigma_i^2)$ versus $\ln D$ for different shape parameters α of the generalized Gaussian distribution.

with

$$\frac{\partial f_{n,m}}{\partial \widetilde{q}}(\alpha, \widetilde{q}) = \left[\left(m+\frac{1}{2}\right)^{n+1} p_{\alpha,1}\left(m\widetilde{q}+\frac{\widetilde{q}}{2}\right) - \left(m-\frac{1}{2}\right)^{n+1} p_{\alpha,1}\left(m\widetilde{q}-\frac{\widetilde{q}}{2}\right)\right]\widetilde{q}^{\,n}. \tag{26}$$

Equations (23) become

$$\begin{gathered} h_{\alpha_i}(\widetilde{q}_i) = -\frac{a_i}{\lambda\Delta_i\pi_i\sigma_i^2}, \quad \forall i, \\ \sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D(\alpha_i, \widetilde{q}_i) = D_T. \end{gathered} \tag{27}$$

The solution of the MSE allocation problem can be obtained with the following equations:

$$\sum_{i=1}^{\#SB} \Delta_i\pi_i\sigma_i^2 D\left(\alpha_i, h_{\alpha_i}^{-1}\left(-\frac{a_i}{\lambda\Delta_i\pi_i\sigma_i^2}\right)\right) = D_T, \tag{28}$$

$$\widetilde{q}_i = h_{\alpha_i}^{-1}\left(-\frac{a_i}{\lambda\pi_i\sigma_i^2}\right), \quad \forall i, \tag{29}$$

where h^{-1} is the inverse function of h. The parameter λ can be found from (28), and then (29) provides the optimal quantization steps q_i. Unfortunately, as there is no analytical formula for h^{-1}, the MSE allocation problem will be solved using a parametric approach described below.

3.4. Parametric approach

Equation (29) gives the values of the quantization steps using tables of the function h for different shape parameters α. Figure 3 shows the tables of $\ln(-h_\alpha(\widetilde{q}))$ for $\alpha = 1, 1/2, 1/3$, and 1/4 and the asymptotic curve of equation

$$\ln(-h) = 2\ln\frac{q}{\sigma} + \ln\frac{\ln 2}{6}. \tag{30}$$

To solve (28), we need tables linking D and λ. Using (20) and (25), we plot the parametric curve (with parameter $\widetilde{q}$)

$$\left[\ln D(\alpha, \widetilde{q}); \ln\left(-h_\alpha(\widetilde{q})\right)\right], \tag{31}$$

for a given α. Using (29), this parametric curve is equivalent in each subband to the following parametric curve:

$$\left[\ln D; \ln\left(\frac{a_i}{\lambda\pi_i\sigma_i^2}\right)\right]. \tag{32}$$

Figure 4 shows these tables for $\alpha = 1, 1/2, 1/3$, and 1/4 and the asymptotic curve of equation

$$\ln(-h) = \ln D + \ln(2\ln 2). \tag{33}$$

Thus, we have a relation between D and λ in each subband. The optimal λ is found using the constraint (28). Then, we have a relation between λ and the quantization step q_i in each subband.

3.5. Algorithm of the model-based MSE allocation

The proposed MSE allocation procedure is the following.

(1) Set the initial value of λ to its asymptotic optimum value $\lambda = 1/2D_T \ln 2$.

(2) For each 3D subband i, compute $\ln(a_i/\lambda\Delta_i\pi_i\sigma_i^2) = \ln(-h)$ and read the corresponding normalized MSE D_i using the tables shown in Figure 4.

(3) 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 (28) is verified and the current λ is optimal. Otherwise, compute[1] a new value of λ and go back to step (1).

[1] Several methods (such as dichotomy, bisection, secant method, golden section search) can be used.

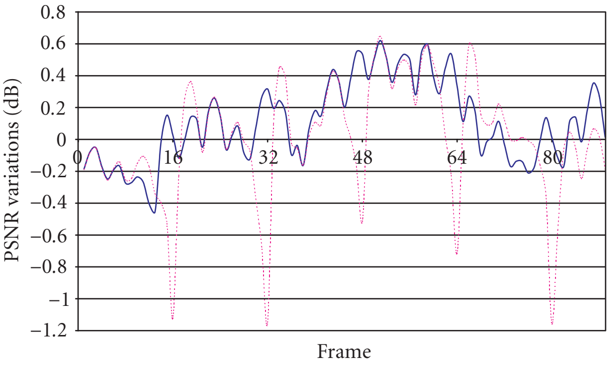

Figure 5: PSNR variations for the 3D scan-based temporal DWT (continuous) and the 3D temporal tiling approach (dashed) on the 89 first luminance frames of the sequence Akiyo at 80 kbps (25 fps). 9/7 DWT with two levels of decomposition; bitrate control for groups of 16 frames.

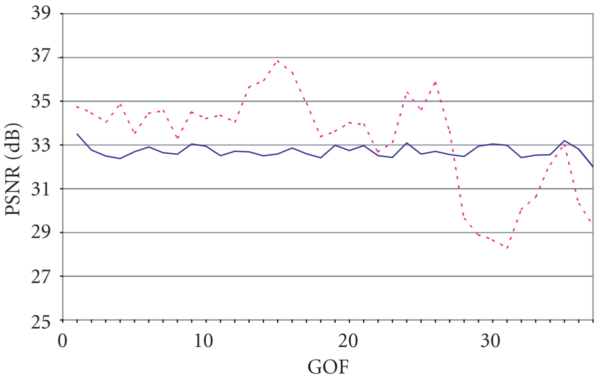

Figure 6: PSNR of each group of height frames (GOF) for the proposed quality control procedure (continuous) and a bitrate control procedure (dashed). The sequence is Foreman at 890 kbps (30 fps) in both cases.

(4) For each 3D subband i, compute $\ln(a_i/\lambda\Delta_i\pi_i\sigma_i^2) = \ln(-h)$ with the optimal λ and read q_i/σ_i using the tables shown in Figure 3. This q_i is the optimal quantization step for subband i.

The tables shown in Figures 3 and 4 are stored for several shape parameters α. They are valid for any video sequence.

4. EXPERIMENTAL RESULTS

To show the efficiency of our 3D scan-based wavelet transform method in removing the temporal blocking artifacts (jerks), we first extended EBWIC [13] to 3D data. The quantized wavelet coefficients have been encoded using JPEG2000's bit-plane context-based arithmetic coder [14]. We 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 encoded the same sequence with the block-based approach, where the temporal wavelet coefficients and their encoding were performed on independent temporal blocks of 16 frames. Figure 5 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 block-based 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 cancelled out.

Then, we replaced the bitrate regulation by our new MSE allocation procedure. Figure 6 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 performance but ensures constant quality output frames. This results in a better visual quality.

5. CONCLUSION

In this paper, we have proposed methods for efficient quality control in video-coding applications.

In Section 2, we have proposed a 3D scan-based DWT method which allows the computation of the temporal wavelet decomposition of a sequence with infinite length using few memory and no extra CPU. Compared to temporal tiling approaches often used to reduce memory requirements, our method avoids temporal tiles artefacts. We have also shown in Section 2.3 that, for the same memory requirements, our method reduces the encoding delay. We have proposed the scan-based motion compensated lifting which results in both saving memory and temporal quality control.

In Section 3, we have proposed a new efficient model-based quality control procedure. This bit allocation procedure controls the output frames quality over time. The extension to scalar quantizers with a deadzone [31, 32, 33] is straightforward.

These methods combine the advantages of wavelet coding (performance, scalability) with minimum memory requirements and low CPU complexity.

REFERENCES

[1] G. Karlsson and M. Vetterli, "Three-dimensional subband coding of video," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 1100–1103, New York, NY, USA, April 1988.

[2] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, no. 2, pp. 125–139, 1995.

[3] B. Felts and B. Pesquet-Popescu, "Efficient context modeling in scalable 3D wavelet-based video compression," in *Proc. IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September 2000.

[4] A. Wang, Z. Xiong, P. A. Chou, and S. Mehrotra, "Three-dimensional wavelet coding of video with global motion compensation," in *Proc. IEEE Data Compression Conference*, pp. 404–414, Snowbird, Utah, USA, March 1999.

[5] J. Xu, S. Li, Y.-Q. Zhang, and Z. Xiong, "A wavelet video coder

using three-dimensional embedded subband coding with optimized truncation (3-D ESCOT)," in *Proc. IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, December 2000.

[6] S. J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 155–167, 1999.

[7] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 572–588, 1994.

[8] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. IEEE Data Compression Conference*, pp. 251–260, Snowbird, Utah, USA, March 1997.

[9] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559–571, 1994.

[10] P. Charbonnier, M. Antonini, and M. Barlaud, "Implantation d'une transformée en ondelettes 2D dyadique au fil de l'eau," CNES contract report 896/95/CNES/1379/00, CNES, October 1995.

[11] C. Parisot, M. Antonini, M. Barlaud, C. Lambert-Nebout, C. Latry, and G. Moury, "On board stripe-based wavelet image coding for future space remote sensing missions," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, pp. 2651–2653, Honolulu, Hawaii, July 2000.

[12] C. Chrysafis and A. Ortega, "Line based, reduced memory, wavelet image compression," *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 378–389, 2000.

[13] C. Parisot, M. Antonini, and M. Barlaud, "EBWIC: A low complexity and efficient rate constrained wavelet image coder," in *Proc. IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September 2000.

[14] ISO/IEC 15444-1:2000, "Information technology—JPEG 2000 image coding system," 2000.

[15] C. Parisot, M. Antonini, and M. Barlaud, "3D scan-based wavelet transform for video coding," in *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 403–408, Cannes, France, October 2001.

[16] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1995.

[17] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. Image Processing*, vol. 4, no. 8, pp. 1053–1060, 1995.

[18] C. Parisot, *Allocations basées modèles et transformée en ondelettes au fil de l'eau pour le codage des images et des vidéos*, Ph.D. thesis, University of Nice-Sophia Antipolis, Nice, France, January 2003.

[19] J.-R. Ohm, "Motion-compensated wavelet lifting filters with flexible adaptation," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[20] J. Viéron, C. Guillemot, and S. Pateux, "Motion compensated 2D+t wavelet analysis for low rate FGS video compression," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[21] T. Wiegand and B. Girod, *Multi-frame Motion-Compensated Prediction for Video Transmission*, Kluwer Academic, Boston, Mass, USA, 2001.

[22] C. Parisot, M. Antonini, and M. Barlaud, "Motion-compensated scan based wavelet transform for video coding," in *Proc. Tyrrhenian International Workshop on Digital Communications*, Palazzo dei Congressi, Capri, Italy, September 2002.

[23] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[24] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.

[25] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distorsion sense," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 160–176, 1993.

[26] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic, Boston, Mass, USA, 1992.

[27] A. Ortega, "Variable bit-rate video coding," in *Compressed Video Over Networks*, M.-T. Sun and A. R. Reibman, Eds., pp. 343–382, Marcel Dekker, New York, NY, USA, 2000.

[28] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. IEEE Data Compression Conference*, pp. 387–395, Snowbird, Utah, USA, April 1996.

[29] M. Barlaud, *Wavelets in Image Communication*, Elsevier, Amsterdam, Netherlands, 1994.

[30] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[31] C. Parisot, M. Antonini, and M. Barlaud, "Optimal nearly uniform scalar quantizer design for wavelet coding," in *Visual Communications and Image Processing*, vol. 4671 of *SPIE Proceedings*, San Jose, Calif, USA, January 2002.

[32] C. Parisot, M. Antonini, and M. Barlaud, "Stripe-based MSE control in image coding," in *Proc. IEEE International Conference on Image Processing*, Rochester, NY, USA, September 2002.

[33] P. Raffy, M. Antonini, and M. Barlaud, "Non-asymptotical distortion-rate models for entropy coded lattice vector quantization," *IEEE Trans. Image Processing*, vol. 9, no. 12, pp. 2006–2017, 2000.

Christophe Parisot graduated and received the M.S. degree in computer vision from the École Supérieure en Sciences Informatiques (ESSI), Sophia Antipolis, France, in 1998. He will receive the Ph.D. degree in image processing from the University of Nice-Sophia Antipolis, France, in 2003. His research interests include image and video compression, quantization, and bit allocation problems.

Marc Antonini received the Ph.D. degree in electrical engineering from the University of Nice-Sophia Antipolis, France, in 1991. He was a Postdoctoral Fellow at the Centre National d'Etudes Spatiales, Toulouse, France, in 1991 and 1992. Since 1993, he has been working with the CNRS at the I3S laboratory both from the CNRS and the University of Nice-Sophia Antipolis. He is a regular reviewer for several journals (IEEE Transactions on Image Processing, Information Theory and Signal Processing, IEE Electronics Letters) and participated in the organization of the IEEE Workshop Multimedia and Signal Processing 2001 in Cannes, France. He also participates in several national research and development projects with French industries, and in several international academic collaborations. His research interests include multidimensional image processing, wavelet analysis, lattice vector quantization, information theory, still image and video coding, joint source/channel coding, inverse problem for decoding, multispectral image coding, and multiresolution 3D mesh coding.

Michel Barlaud received his Thèse d'Etat from the University of Paris XII. He is currently a Professor of image processing at the University of Nice-Sophia Antipolis, and the leader of the Image Processing Group of I3S. His research topics are image and video coding using scan-based wavelet transform, inverse problem using half-quadratic regularization, and image and video segmentation using region-based active contours and PDEs. He is a regular reviewer for several journals, a member of the technical committees of several scientific conferences. He leads several national research and development projects with French industries and participates in several international academic collaborations (Universities of Maryland, Stanford, Boston, Louvain La Neuve). He is the author of a large number of publications in the area of image and video processing and the editor of the book *Wavelets and Image Communication* Elsevier, 1994.