Cyclostationary error analysis and filter properties in a 3D wavelet coding framework

Riccardo Leonardi, Alberto Signoroni*

Department of Electronics for Automation, University of Brescia, via Branze, 38 I-25123 Brescia, Italy

Received 25 August 2005; received in revised form 1 June 2006; accepted 2 June 2006

Abstract

The reconstruction error due to quantization of wavelet subbands can be modeled as a cyclostationary process because of the linear periodically shift variant property of the inverse wavelet transform. For \( N \)-dimensional data, \( N \)-dimensional reconstruction error power cyclostationary patterns replicate on the data sample lattice. For audio and image coding applications this fact is of little practical interest since the decoded data is perceived in its wholeness, the error power oscillations on single data elements cannot be seen or heard and a global PSNR error measure is often used to represent the reconstruction quality. A different situation is the one of 3D data (static volumes or video sequences) coding, where decoded data are usually visualized by plane sections and the reconstruction error power is commonly measured by a PSNR\( [n] \) sequence, with \( n \) representing either a spatial slicing plane (for volumetric data) or the temporal reference frame (for video). In this case, the cyclostationary oscillations on single data elements lead to a global PSNR\( [n] \) oscillation and this effect may become a relevant concern. In this paper we study and describe the above phenomena and evaluate their relevance in concrete coding applications. Our analysis is entirely carried out in the original signal domain and can easily be extended to more than three dimensions. We associate the oscillation pattern with the wavelet filter properties in a polyphase framework and we show that a substantial reduction of the oscillation amplitudes can be achieved under a proper selection of the basis functions. Our quantitative model is initially made under high-resolution conditions and then qualitatively extended to all coding rates for the wide family of bit-plane quantization-based coding techniques. Finally, we experimentally validate the proposed models and we perform a subjective evaluation of the visual relevance of the PSNR\( [n] \) fluctuations in the cases of medical volumes and video coding.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Wavelet coding; 3D data; Biorthogonal filters; Cyclostationary models; Bit-plane quantization; Quality assessment

1. Introduction

1.1. Motivations and objectives

Quality assessment of reconstructed data is a crucial task in the data coding environment. For audiovisual applications, it is widely recognized that objective quality measures, based on simple reference and decoded data differences, are not well
matched with respect to the human perceived (subjective) quality. Anyway, such objective measures play an essential role in coding system design, implementation and use. First of all, as they are easy to compute, they are used for monitoring and controlling the data quality; secondly, being usually mathematically linked to the approximations (quantization error) introduced in the transformed data domain they are also used for rate-distortion optimizations techniques. MSE-based quality measures (e.g. the peak signal-to-noise ratio, hereinafter indicated as $P_{SNR}$) are also commonly used as benchmarks for coding technique comparisons.

Being a global measure, the MSE loses the possibility to be linked to data domain localized reconstruction quality issues. Conversely, pure data element-based error measures (e.g. made on each pixel position) are inconsistent and ask for aggregating criteria that can lead to a reconstruction quality interpretation. To this end, signal to error dependencies and coding artifacts strength are usually analyzed and conveyed into quality measures according to a reconstructed data analysis inspired to human sensorial system limitations [19] or structural distortion principles [32].

In the case of wavelet data coding, another reconstruction quality aspect can be evidenced which comes from the linear and periodically time-variant (LPTV) nature of the inverse discrete wavelet transform (IDWT). This modifies the statistical properties of the quantization error introduced in the transformed domain in a way that a stationary error in the wavelet domain becomes cyclostationary (CS) in the data domain, after IDWT [31]. For multidimensional data, the CS statistical behavior can be associated to the data sampling grid and measured on a hypercubic periodic pattern characterized by the time-varying periodicity along each data dimension. Moreover, the related effect on the reconstruction quality assessment depends on the data fruition mode, or rather on which and how portions (aggregates of data elements) of the decoded data are presented to the user. In this perspective, the reconstruction quality aggregating criteria (in the sense described above) can be considered as deriving from the relationships between the data fruition mode and the CS patterns. Thus, it becomes meaningful to model and predict objective quality oscillations among portions of data according to their fruition and then evaluate the related subjective impact, as will be clarified below.

For example, $P_{SNR}$ oscillations can be observed on hyperplanes intersecting the data, e.g. the rows of an image or on the slices of a volume. In these cases, our interest focuses on how and to what extent the CS property generates oscillatory phenomena of the $P_{SNR}$ by moving the hyperplane along the remaining orthogonal directions. This point is not commonly considered in image coding: no particular attention is dedicated to fluctuations of the $P_{SNR}$ between consecutive rows, as the perception of the coded information remains 2D. Instead, in the case of the slices of a volume, measuring the $P_{SNR}[n]$ on each slice $n$ is a common procedure.

In static 3D data (e.g. computed tomography CT or magnetic resonance MR data-sets), adjacent slices are commonly seen side by side or in a movie loop, with low frame-rate or manual frame navigation. In this case limiting $P_{SNR}$ oscillations significantly contributes to ensure a suitable reliability level to the content reproduction. This is especially true in the field of diagnostic imaging, where 3D images are widespread used and compression is, at the same time, a strong necessity and a concern. Moreover, as (for diagnostic purposes) the slicing can be done in different directions (usually the three orthogonal axes, i.e. axial, coronal and sagittal, are considered), the oscillatory behavior should be monitored along all such directions.

For 2D + t video data, only the temporal direction is of interest for PSNR fluctuation evaluations. Because of the band-pass temporal response of the human visual system, objective quality fluctuations in decoded video frames are quite well tolerated under certain perceptual threshold. However, for low bit-rates or high-quality requirements, quality fluctuations become may become relevant issues, causing visible effects usually perceived as localized or cyclic, object-based or global flicker artifacts. Fluctuation effects generated in hybrid schemes at low bit-rates can be observed in wavelet-based schemes where considerations on motion compensated prediction (MCP) should be reformulated in terms of motion compensated temporal filtering (MCTF) [17] and can be put in relation with the models presented here. Flickering have been observed in wavelet-based intra-frame coders too [3] for high-quality applications (such as D-Cinema). Even if not recognized in [3], due to the LPTV nature of the wavelet transform, moving objects have different coefficient maps and, as it will be
shown, different pixelwise reconstruction error energy patterns, causing reconstructed objects to appear slightly different (with possible temporal flicker effects) in decoded frames.

1.2. Relations to other works

The PTV property of the wavelet transform is well known [31]. The analysis of the effects of this property in terms of propagation of the subband quantization error on the reconstructed data has been made by Uzun and Haddad in [30] as an extension of the work of Westerink et al. [34]. A main objective of these works was to find wavelet filter banks which minimize the reconstruction error in a MSE sense. These approaches, and in general other works on subband coding systems which incorporate quantization models (e.g. [11,7,8,27]), consider error spectral density functions and derive statistical expectation parameters, following a z-transformed domain approach. Likewise, Reichel et al. [21] analyze the non-linear effects of integer wavelet transforms on the reconstruction error. In the present work a different approach has been adopted: the analysis is entirely performed in the data domain in order to directly model the non-stationary properties of the reconstruction error and their relations with structural filter properties and visual quality.

As far as the coding schemes are concerned, a certain number of papers describe effective wavelet compression schemes for 3D data [23,26,25,14,4,36] while state-of-the-art wavelet-based video codecs are reviewed in [15]. Most of the existing 3D data coders adopt similar bit-plane-based quantization strategies. We are then motivated to find results of general validity and not tailored on a specific coding system.

1.3. Outline

This paper proposes a quantitative analysis and a multidimensional model that describes the reconstruction error oscillation due to wavelet-based compression and it shows how to obtain its reduction. In fact, a particular property of some linear phase biorthogonal filters emerges causing a substantial oscillation reduction. A time domain formulation of the problem is proposed here so as to describe the oscillation pattern. Section 2 addresses the problem in one dimension (1D), under high-resolution hypothesis, from one to $L$ levels of wavelet decomposition. Section 3 extends the model from 1D to the multidimensional case (especially 3D). The model is then adapted to handle the characteristics of real codecs working at various bit-rates (Section 4). Model validation, experimental results and discussions are finally provided in Section 5.

2. Reconstruction error properties

In this section, some well known fundamentals of multirate filter banks are shortly reviewed. A 1D framework is first used to formulate a data domain statistical analysis of the reconstruction error. After considering a single stage 2-channel decomposition, an extension to the case of multilevel DWT is presented. Finally, the focus is shifted to the case of linear phase biorthogonal wavelet basis. In order to easily characterize the error contributions, the polyphase multirate filter bank representation (see e.g. [31]) will be adopted.

2.1. Reconstruction error in a single synthesis stage

Fig. 1(a) shows the well-known 2-channel wavelet decomposition with the associated two sources of quantization noise in the subband domain. The
additive quantization error signals, introduced with respect to the $i$th subband, are denoted by $q_i[n]$, with $i = 0, 1$. Without such noise components, the overall system acts as a pure delay $b$ with gain $a$. At the output of the IDWT, implemented by filters $G_i$, the reconstruction error $r[n]$ adds up to the perfect reconstructed signal. In Fig. 1(b) only the quantization error processing chain is shown. Each subband error $q_i[n]$ contribution to $r[n]$ is denoted by $r_i[n]$. The statistical properties of $q_i[n]$ and the LPTV nature of each channel of the synthesis filter bank determine a CS behavior of each $r_i[n]$, thus of the whole $r[n]$. The polyphase representation of the synthesis bank is shown in Fig. 1(c), where

$g_i[n] = (2 \uparrow)(e_i^0)[n] + (2 \uparrow)(e_i^1)[n - 1],$

$G_i(z) = E_i^0(z^2) + z^{-1}E_i^1(z^2).$  

(1)

Here various error components can be considered according to their contribution to the reconstruction error $r[n]$, namely $r = r_0 + r_1 = r_0^0 + r_0^1 + r_1^0 + r_1^1$, where $r_0^0 \triangleq r_0^1 + r_1^0$ and $r_1^1 \triangleq r_0^1 + r_1^1$ represent the overall contribution of the polyphase components at even and odd sample positions, respectively, that is $r_0^0[n] = 0$ for $(n \mod 2) = 1$ and $r_1^1[n] = 0$ for $(n \mod 2) = 0$.

To better understand the synthesis filter bank effects on the quantization error, the analysis starts from high-resolution quantization hypotheses: for each channel $i$, $q_i[n]$ can be considered the realization of a uniformly distributed ergodic white noise $q_i[n]$ and, $\forall i, \tilde{n}$ the random variables $q_i[\tilde{n}]$ can be assumed to be i.i.d., with $f_{q_i}(z) = (1/\Delta)\text{rect}(\Delta z)$ (hence having the same expected value $\eta_q = E[q_i[n]] = 0$ and the same variance $\sigma_q^2 = E[q_i^2[n]] = \Delta^2/12$).

In the simple case of Fig. 1(c) it is easy to verify that, due to the LPTV behavior of the IDWT, the 2D ergodic input process is converted in a 1D CS output process $r[n]$, with period 2, having the following characteristics: $\eta_r = 0$ and,

$\sigma_r^2[n] = \begin{cases} 
\sigma_{r_0}^2 & \text{for } (n \mod 2) = 0, \\
\sigma_{r_1}^2 & \text{for } (n \mod 2) = 1.
\end{cases}$

By linearity, the reconstruction error can be decomposed as follows:

$\sigma_{r_0}^2 = \sigma_{q_0}^2 + \sigma_{q_1}^2 = \sigma_{q_0}^2 \mathcal{F}^2(e_0^0[n]) + \sigma_{q_1}^2 \mathcal{F}^2(e_0^1[n]),$

(2)

$\sigma_{r_1}^2 = \sigma_{q_0}^2 + \sigma_{q_1}^2 = \sigma_{q_0}^2 \mathcal{F}^2(e_1^0[n]) + \sigma_{q_1}^2 \mathcal{F}^2(e_1^1[n]),$

(3)

where $\mathcal{F}^2(\cdot)$ is the energy operator and $e_j^i[n]$ represents each polyphase component.

2.2. Variance oscillation strength

The absolute difference $|\sigma_{r_0}^2 - \sigma_{r_1}^2|$ is a first indicator of the entity of the cyclostationarity implications on the reconstruction error. Obviously, it is desirable to make this difference small.

Considering (2) and (3), and supposing $\sigma_{q_0}^2 = \sigma_{q_1}^2$, for the above difference to be equal to zero it is sufficient that the sum of the energies of the low- and high-pass corresponding (same phase) polyphase components give the same value. In general, this is not under direct control and minimization of the $|\sigma_{r_0}^2 - \sigma_{r_1}^2|$ should be achieved by proper design of the filter bank. However, a very simple solution can be found, thanks to the properties of linear phase biorthogonal filters. These filters are commonly used for visual data compression. They can have either odd (e.g. the popular 9/7 filters [2]) and even-lengths (e.g. the 10/18 filters [29] or the 22/14 ones [33]). It is well known that linear phase filters exhibit symmetric or antisymmetric impulse response. Let us now understand how this property is conveyed to the polyphase representation of a linear phase filter bank. Odd-length symmetric filters (Type I and III linear phase FIR) have sample-centered symmetry axes, while even-length ones (Type II and IV) have bin-centered symmetry axes. Thus, it is easy to notice (see Fig. 2) that in the odd-length case the polyphase splitting generates two symmetric filters (an odd-length and an even-length one) while, for even-length filters, two non-symmetric filters are generated, one being the mirror image of the other (with respect to the original center of symmetry). At this point, it can be observed that for even-length filters the CS behavior of the output noise $r[n]$ for a 2-channel filter bank is “inherently eliminated” because $\sigma_{r_0}^2 = \sigma_{r_1}^2$ and $f_{r_0}(z) = f_{r_1}(z)$, whereas this does not hold for the odd-length filter case. This clear-cut distinction is valid for the considered case of 1D signals and 1-level wavelet decomposition, but it will be a determinant even for N-dimensional and L-levels cases.

1From now on the term “shift-variance” is preferred instead of “time-variance” because of the type of the considered data. However, for clarity, the acronym LPTV is maintained.

2In the adopted notation the superscripts correspond to the reference positions within the period of time variance (or the reference number of the polyphase components), while subscripts are used to identify each subband channel. Moreover, $(2 \uparrow)(\cdot)$ denotes the upsampling operator and its argument.
2.3. Reconstruction error in a L-levels IDWT

The analysis of the output noise characteristics is now considered for a L-level wavelet decomposition. In Fig. 3(a) a dyadic L-level IDWT is shown (with \( L = 3 \)). For the purpose of establishing the statistical behavior of the output noise component, it is convenient to use the equivalent single-adder \((L+1)\)-channel multirate scheme (Fig. 3(b)), where \( F_0(z) = G_0(z^4)G_0(z^2)G_0(z) \).

\[
F_1(z) = G_1(z^4)G_0(z^2)G_0(z), \quad F_2(z) = G_1(z^2)G_0(z) \quad \text{and} \quad F_3(z) = G_1(z).
\]

The polyphase representation (Fig. 3(c)) allows for an immediate interpretation of the reconstruction error components. The value of \( \sigma_r^2 \) at position \( n \) can be obtained by a generalization of (2) and (3) which considers a higher number of subband channels with the appropriate number of polyphase components:

\[
\sigma_r^2[n] = \sum_{l=0}^{L} \sigma_{r_l}^2[n] = \sum_{l=0}^{L} \sigma_{r_{n \mod \lambda(l)}}^2,
\]

being \( \sigma_{r_l}^2 = \sigma_{r_l}^2 \cdot \mathcal{F}^2(e'[n]l) \), with \( l = 0, \ldots, L \) and \( j = 0, \ldots, \lambda(l) - 1 \), the \( j \)th polyphase component of \( \sigma_{r_l}^2[n] \). In (4), at each position \( n \) the right polyphase component \( j = n \mod \lambda(l) \) is selected for each subband channel \( l \). The usage of \( \delta(l) \) allows to handle the low-pass channel 0 with the same resolution level of the high-pass channel 1.

The random process \( r[n] \) which represents the reconstruction error results CS with period \( 2^L \), in fact \( r[n] \) represents a sum of CS random processes \( r_l[n] \) with periodicity \( 2^l \), with \( l = 0, \ldots, L \). Fig. 4 shows the sequence of \( \sigma_{r_l}^2[n] \) values for a 16 sample signal, with \( L \) set to 3 and with a common subband error variance \( \sigma_q^2 = D^2/12 \) with \( D = 8 \). During the analysis stage the lower resolution subbands exhibit a larger gain with respect to the higher resolution ones. This gain is compensated during the synthesis phase.
Fig. 4. CS $\sigma^2[n]$ oscillation magnitude at various resolution levels $l$, in the case of (a) 9/7 and (b) 10/18 filters, for a 16-sample signal and a 3-level IDWT.

Fig. 5. Overall MSE and $P_{SNR}$ fluctuations for a 3-level 1D-IDWT.
stage so that the wavelet domain quantization error is subject to a de-emphasis which becomes more substantial as the resolution level decreases. This can be seen in Fig. 4 where the average error contribution decreases towards low-resolution subbands and so does its oscillation amplitude. Therefore, the first decomposition level error contribution is dominant in defining the oscillation pattern of the CS behavior. In Fig. 5, for the same biorthogonal wavelet filters, the MSE\([n] = \sigma_1^2[n]\) and the corresponding \(P_{\text{SNR}}[n]\) are calculated and plotted for each position \(n\) in the original signal domain. As expected, for the \(L = 3\) case of Fig. 5 and in general for a \(L\)-level IDWT, there is a smaller oscillation amplitude in the case of even-length filters. It can also be observed that the CS oscillation pattern exhibits a \(2^L\) (\(L = 3\)) periodicity.

The \(P_{\text{SNR}}\) difference between the \(j\)th and the \((j + 1)\)th position for a \(L\)-level IDWT can be written as

\[
\Delta P_{\text{SNR}}[j] = 10 \log \frac{255^2}{\sigma_j^2} - 10 \log \frac{255^2}{\sigma_{j+1}^2} = 10 \log \frac{\sigma_{j+1}^2}{\sigma_j^2},
\]

where the \(j\)th MSE value can be expressed as \(\sigma_j^2 = \sum_{l=0}^{L} \sigma_{l,j}^2\), with \(l = 0, \ldots, L\), \(\sigma_{l,j}^2\) being the contribution of the \(l\)th subband. It can also be noted that for a reference quantization noise \(\sigma_0^2\) which remains the same in each subband, there is no dependency of the \(P_{\text{SNR}}\) fluctuation with respect to it. \(\Delta P_{\text{SNR}}[j]\) is only a function of the polyphase filter coefficients; e.g. for a 1-level IDWT case (see (2), (3)):

\[
|\Delta P_{\text{SNR}}[j]| = 10 \log [\mathcal{G}^2(e_j^0[n]) + \mathcal{G}^2(e_j^1[n])] - 10 \log [\mathcal{G}^2(e_0^0[n]) + \mathcal{G}^2(e_0^1[n])], \quad \forall j.
\]

### 3. 3D extension of the model

The above formulation provides the basic elements to model the reconstruction error in the context of \(N\)-dimensional data coding. This section concentrates on a set of structural properties of \(r[n]\) in 2D and 3D spaces. As it is common in almost all coding applications separable implementations of the wavelet transform are considered. Moreover, in order to avoid burdening the notation and without limiting the expressive power of the model, the wavelet basis are assumed to be the same for each filtering direction. Then, in the following expressions, there is no need to specify direction-related subscripts for the \(z\) variable.

#### 3.1. 2D and 3D CS patterns

For \(N\)-dimensional data, periodic behavior of the reconstruction error statistics can be observed on the signal Cartesian reference system. In particular, the CS periodicities produce some elementary patterns (e.g. tiles or bricks) which partition the whole data space.

The 2D case is dealt first, and the error reconstruction variance at each pixel position calculated. To this end a separable 2D IDWT dyadic tree and its separable single-adder implementation are considered. Both schemes are depicted in Fig. 6 considering a 3-level decomposition. For each subband channel in Fig. 6(b) two 1D filters operate in cascade along the vertical and horizontal dimensions, respectively. A general expression of such filters, for \(l = 1 \ldots L\), is given by

\[
F_{l,a}(z) = \prod_{i=1}^{L} G_{a,i[l-1]}(z^{2^{(i)})}
\]

with \(a \in \mathbb{B} = \{0, 1\}\), \(\gamma(i) = 2^{L-i}\). For \(l = 0\), \(F_{0,0}(z) = F_{1,0}(z)\) (because the \(c_0\) and \(d_1\) coefficients belong to the same scale). In the first synthesis stage of Fig. 6(b), the vertical filtering introduces the CS behavior described by (2), (3) on each subband column. It can be easily observed that, at the input of the horizontal stage (before the upsamplers), i.i.d. hypotheses on the error still hold along the row direction, but with alternating non-uniform pdfs on each row due to the previous column filtering. The horizontal stage introduces a CS behavior in each row, which combined to the one introduced in the column, results in a square tiled 2D CS pattern. This simple mechanism is shown in Fig. 7 in the case of 1-level IDWT (which determines a \(2 \times 2\) CS pattern). By analogy to the 1D case, increasing the IDWT number of levels to \(L\), the CS pattern becomes a \(2^L \times 2^L\) matrix which constitutes the basic tile of the 2D cyclostationarity. In other terms, considering the reconstruction error variance, \(\sigma_i^2[n_0, n_1] = \sigma_i^2[n_0, n_1]\) where \(n_0 = n_0 \mod 2^L\), \(n_1 = n_1 \mod 2^L\).

The reconstruction error variance values for each spatial (tile) position can be calculated by the following extension of (4):

\[
\sigma_j^2[n_0, n_1] = \sigma_j^2_{i_0(0,0)} + \sum_{l=1}^{L} \sum_{s \in \mathbb{B}^2} \sigma_j^2_{i_l s}
\]
where \( \bar{n}_0 = n_0 \mod \lambda(l) \), \( \bar{n}_1 = n_1 \mod \lambda(l) \), \( \lambda(l) = 2^{L+1-(l+1)L} \) and, given \( a, b \in \mathbb{B} = \{0, 1\} \),
\[
\mathbb{B}^2 = \{(a, b)\} \backslash \{(0, 0)\}.
\]

In (7) the first term refers to the lowest resolution \( l = 0 \) low-pass subband coefficients \( s = (0, 0) \), while the second term represents the contribution of the three detail subbands \( s \) at each resolution level \( l \in (1, L) \). For each CS tile position, every subband contribution \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} \) to the total reconstruction error variance is given by
\[
\sigma^2_{r,\bar{n}_0,\bar{n}_1} = \sigma^2_{q,\bar{n}_0,\bar{n}_1} \cdot P^0_{0,0} P^1_{1,1},
\]
(8)
where \( P^0_{0,0} = G^2_0(\varepsilon^0_{1,0}[n]) \), and \( E^q_{l,a}(z) \) is the \( n \)th polyphase component of \( F_{l,a}(z) \) defined in (6). Eq. (8) considers the general case of a different quantization error variance \( \sigma^2_{q,\bar{n}_0,\bar{n}_1} \) associated to each subband.

A symmetry property can be demonstrated for the \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} \) CS tile pattern which is relevant to the present analysis:

**Proposition 1.** Given \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} \) a \( 2^L \times 2^L \) matrix (CS tile) with elements \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} \) derived from (7), then \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} \) is symmetric, i.e. \( \sigma^2_{r,\bar{n}_0,\bar{n}_1} = (\sigma^2_{r,\bar{n}_0,\bar{n}_1})^T \).

**Proof.** Eq. (7) can be rewritten as
\[
\sigma^2_{r,\bar{n}_0,\bar{n}_1} = \sum_{l=1}^{L} \sigma^2_{q,\bar{n}_0,\bar{n}_1},
\]
where, using (8),
\[
\sigma^2_{q,\bar{n}_0,\bar{n}_1} = \sum_{s \in \mathbb{B}^2} \sigma^2_{q,\bar{n}_0,\bar{n}_1} \cdot P^0_{0,0} P^1_{1,1} = \sigma^2_{q,\bar{n}_0,\bar{n}_1} (P^0_{1,0} P^1_{1,1} + P^0_{1,1} P^1_{1,0} + P^0_{0,0} P^1_{1,1}).
\]
It can be observed that, thanks to the separability in (8) and to the presence of complementary elements
in $\mathbb{R}^{3+}$, there are symmetric matrices at each IDWT decomposition level, namely $\hat{\sigma}_{l_0,h_0}^2 = \hat{\sigma}_{h_0,l_0}^2$ and $\hat{\sigma}_{l_0,h_0}^2 = \hat{\sigma}_{l_1,h_1}^2$ for $l = 1, \ldots, L$. This is true even if, in general, the single subband components of the reconstruction error variance are not symmetric, that is $\hat{\sigma}_{l_0,h_0}^2 \neq \hat{\sigma}_{l_1,h_1}^2$.

Then, at each decomposition level $l = 0, \ldots, L$ the fundamental $2^{l+1}[l] \times 2^{l+1}[l]$ CS matrices $\hat{\sigma}_{l}^2$, with elements $\hat{\sigma}_{l_0,h_0}^2$, are symmetric by construction.

Using these matrices as tiles, (7) can also be rewritten in the following form:

$$\hat{\sigma}_{l}^2 = \hat{\sigma}_{l}^2 \otimes \hat{\sigma}_{l}^2 + \sum_{l=1}^{L} \hat{\sigma}_{l}^2 \otimes \hat{\sigma}_{l}^2,$$

where $\otimes$ is the Kronecker product, while $\hat{\sigma}_{l}^2$ is a $(2^{l+1}[l]) \times (2^{l+1}[l])$ matrix containing all “ones”.

As operator $\otimes$ preserves the symmetry and all the matrices in (9) are symmetric, $\hat{\sigma}_{l}^2$ is also symmetric being a sum of symmetric matrices. □

The 3D extension follows easily. Using the same notation,

$$\hat{\sigma}_{l}^2[n_0, n_1, n_2] = \hat{\sigma}_{l}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2] = \hat{\sigma}_{l_0,h_0}^2[0,0,0,0] + \sum_{l=1}^{L} \sum_{s \in \mathbb{R}^{3+}} \hat{\sigma}_{l}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2],$$

with

$$\hat{\sigma}_{l}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2] = \hat{\sigma}_{l}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2] = \hat{\sigma}_{l}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2],$$

and the following proposition holds:

**Proposition 2.** Given $\hat{\sigma}_{r,s}^2$, a $2^L \times 2^L \times 2^L$ 3D array (CS brick), with elements $\hat{\sigma}_{l_0,h_0}^2[\tilde{h}_0, \tilde{n}_1, \tilde{n}_2]$ derived from (10), then $\hat{\sigma}_{r,s}^2$ is symmetric, i.e. $\hat{\sigma}_{r,s}^2 = (\hat{\sigma}_{r,s}^2)^T = (\hat{\sigma}_{r,s}^2)^T\tau_2$, where $(\cdot)^T_1$ and $(\cdot)^T_2$ are the two possible transposition operators with respect to the main diagonals of the tensor; i.e. $\hat{\sigma}_{r,s}^2 = (\hat{\sigma}_{r,s}^2)^T_1 = (\hat{\sigma}_{r,s}^2)^T_2$ and $\hat{\sigma}_{r,s}^2 = (\hat{\sigma}_{r,s}^2)^T_1 = (\hat{\sigma}_{r,s}^2)^T_2$.

The proof of this proposition can be derived in analogy to the proof of Proposition 1. The analysis can be easily extended in a similar way to higher dimension problems.

Propositions 1 and 2 virtually allow to establish the spatial orientation invariance of the CS patterns which in turn determines, in the 3D case, the independence of the $P_{SNR}$ oscillation from the slicing direction, as shown in the next section.

### 3.2. $P_{SNR}$ oscillation along various volume slicing directions

The description of the 3D cyclostationarity with a symmetric and periodic cubic pattern allows to quantify the fluctuations of the $P_{SNR}$ evaluated on the images obtained by slicing a volume (or 3D dataset) along one of its perpendicular axes. In fact, the expected MSE values along a certain slicing direction can be estimated by averaging the punctual (voxel-related) MSE values of the various slices of the CS brick. This measure is called slice-based MSE (sMSE). For example, by slicing along the $n_0$ direction, the $2^L$ periodical sequence is obtained:

$$s\text{MSE}_i^n[n] = \frac{1}{2\pi} \sum_{n_1=0}^{2^L-1} \sum_{n_2=0}^{2^L-1} \sigma_{l}^2[n_1,n_2].$$

Thanks to Proposition 2, it can be easily observed that

$$s\text{MSE}_i^n[n] = s\text{MSE}_i^{l_1}[n] = s\text{MSE}_i^{l_2}[n], \quad \forall n,$$

where $s\text{MSE}_i^{l_1}[n] = (1/2^L) \sum_{n_1=0}^{2^L-1} \sum_{n_2=0}^{2^L-1} \sigma_{l_1}^2[n_1,n_2]$, and $s\text{MSE}_i^{l_2}[n] = (1/2^L) \sum_{n_1=0}^{2^L-1} \sum_{n_2=0}^{2^L-1} \sigma_{l_2}^2[n_1,n_2]$.

Similarly a slice-based peak signal-to-noise ratio ($sP_{SNR}$) can be defined as

$$sP_{SNR}[n] = 10 \log \frac{255^2}{s\text{MSE}_i^n[n]}.$$  

Fig. 8 shows the estimated $sP_{SNR}[n]$ values using (14) for six different biorthogonal filters, where it is assumed to have the same quantization error $\sigma_{l_1}^2 = \sigma_{l_2}^2$ on each subband. The set of biorthogonal filters is representative of the most relevant wavelet kernels used in coding applications. It contains three odd-length and three even-length filter pairs. There are considered:

(a) the best known B-spline 9/7 filter bank [2,6] and, from the same family, the 9/3 filters [6, p. 277];
(b) two longer even-length kernels, the 10/18 of [29] and the 22/14 [33]; both have demonstrated, despite their length, competitive performance when used for 2D images and 3D medical data coding [14,12];
(c) two well-known short filters: the Haar and the 5/3 B-spline kernels [6, p. 277], because of their
role in the current research on wavelet video coding. In fact short filters are useful for an effective MCTF, which must be used instead of the rectilinear third dimension (temporal) filtering for high-performance scalable video coding applications [18, Chapter 13].

Fig. 8 suggests the following considerations:

- As expected, odd-length filters produce a substantially higher $sP_{\text{SNR}}[n]$ fluctuation with respect to the even-length ones.
- The odd-length filters 5/3 and 9/3, with their short (three tap) low-pass synthesis stage, are particularly critical; they present a $sP_{\text{SNR}}[n]$ oscillation of about 2.8 and 1.8 dB, respectively.
- Comparing the most popular filters for still image coding, the maximum $sP_{\text{SNR}}[n]$ oscillation for the 9/7 filters is about 0.95 dB versus a 0.07 dB of the 10/18 and 22/14 ones. In terms of reconstruction error power, 9/7 filters show a relative peak-to-peak $s\text{MSE}[n]$ oscillation (with respect to the mean value) of about 25%, while the same factor is close to 1%, 5% for 10/18 and 22/14 filters.
- Compared to the 1D case, there is an increased predominance of the finest detail subband on the overall oscillation trend (see Fig. 4). This is due to the normalization (de-emphasis) factors applied to the lower detail subbands. This originates from an augmented error de-emphasis (due to subband power normalization) effect, which grows in power with the dimensionality of the problem.
- The plane-based measure of $sP_{\text{SNR}}[n]$ actually averages the local expected $P_{\text{SNR}}[n_0,n_1,n_2]$ value on each voxel (or frame pixel). Thus, the $s\text{SNR}$ oscillation between adjacent planes is less than the local $P_{\text{SNR}}$ pattern fluctuation in the vicinity of a voxel. This observation would suggest to use even-length linear phase filters also to maintain a greater error homogeneity in a spatial neighborhood.
- Finally, filters do not only differ in terms of $sP_{\text{SNR}}[n]$ oscillation amplitude but also in terms of average $sP_{\text{SNR}}[n]$ value, and this fact can find a broader confirmation for N-dimensional signals. In other words, a filter ranking can be observed in terms of averaged quantization–reconstruction

![Fig. 8. Quantification of the $sP_{\text{SNR}}[n]$ on adjacent cutting planes using (14). The direction \(n\) represents any of the three orthogonal directions of the volumetric data-set. Six biorthogonal filters are compared, for a 3-level 3D-IDWT.](image_url)
error power transfer. This could be seen as a contributing factor to the multivariate problem of determining the goodness of a wavelet basis for coding purposes. Usually, the coding performance of wavelet filters are complex to quantify on real data and justifications are usually made in terms of approximation power or energy compaction attributes. This actually refer back to what said in the Introduction (Section 1.1) about the quality assessment problem, and again it can be observed that data-centered criterions can be used in combination to reconstruction error-centered ones (as the above ranking) in order to better understand the filter peculiarities for visual data coding. For example, the Haar basis shows the lowest error power transfer, but it has very poor approximation capabilities. On the other hand, it is interesting to see that even if 22/14 and 10/18 visual performance could be hard to compare, there clearly is an error power transfer disparity in favor of the 22/14 filter pair. Considerations about the apparently better average behavior of the 9/7 filters will be also made (Section 5.1.1).

The above considerations will be further discussed in Section 5 by comparing the values derived from (12) and shown in Fig. 8 with respect to simulations on real data and appropriate rate-distortion ranges. Before doing this, it is necessary to analyze what happens to the oscillation model when, by increasing the compression ratio, high-resolution hypotheses degrade.

4. 3D embedded wavelet coding and bit-plane quantization

Until now, some simplifying hypothesis on \( \sigma^2_{q_{1,s}} \) have been made, with \( s \in \mathbb{B}^3, t \in \{1,2,3\} \). Then, there is the need to reconsider the CS quality oscillation model in a real context of interest. As later clarified, the i.i.d. hypothesis must be discussed on an appropriate range of bit-rates from an intra- and inter-subband point of view. In this section, a degradation model is presented which is able to describe what happens to the CS reconstruction error when diminishing the coding rate from high-resolution to low bit-rate conditions. Our specific interest consists in finding a model for the quantization error pdf modifications in order to interpret the trends of the \( \sigma^2_{q_{1,s}} \) values when the bit-rate decreases. This makes it possible to understand the \( \sigma^2_q[n] \) and \( sPSNR[n] \) fluctuation modifications. The properties of the quantization error are determined, at various bit-rates, by the coding technique and by the data to be coded that affect the statistics of the wavelet subbands. Because of the intrinsic variability of the data statistics, and even though observations will be restricted to natural image and volumes, the degradation model will only qualitatively describe the desired phenomena. Nevertheless, its validity extends on a wide class of bit-plane-based wavelet coders.

In Section 4.1 some bit-plane quantization fundamentals and prerequisites are derived for the degradation model description of Section 4.2.

4.1. Preliminary considerations on \( \sigma^2_{q_{1,s}} \)

State-of-the-art wavelet coders (based on zero-trees [24,22], zero-blocks [9], significant-clusters [5,14] or independent blocks [28]), all make use of a progressive bit-plane quantization (BPQ) strategy. The wavelet structure and the high-order statistical dependency among intra- or inter-subband coefficients are exploited by the significance-map description and/or by a context-based arithmetic coding to obtain a progressive or scalable coded bit-stream. The bit-plane quantizer is an almost uniform quantizer with the only exception of a double sized zero-centered dead-zone. The corresponding quantization law can be expressed as

\[
Q(x, \Delta) = \begin{cases} 
0 & \text{if } |x| < \Delta, \\
\text{sgn}(x)(|x/\Delta| + 1/2)\Delta & \text{if } |x| \geq \Delta,
\end{cases}
\]

(15)

where, for the \( b \)th bit-plane, \( \Delta = 2^b \). When applied to the whole subband structure the progressive BPQ starts from \( \Delta = 2^B \), with

\[
B = \left[ \max_{n,l,s} \{ \log_2 c_0[n], \log_2 d_{l,s}[n] \} \right]
\]

(16)

and refines the wavelet coefficients representation by halving the value of \( \Delta \) for each added bit-plane. Mallat [16] verified that at low bit-rates and for natural images the BPQ improves the R-D performance compared to the uniform one. Because of the zero-concentrated wavelet coefficients pdf shape, a 2.4 step size around the zero value considerably reduces the number of significant coefficients and consequently the associated bit-rate, and the expected distortion is smaller than the one obtained by
uniform $\Delta$ quantization (with the same bit-rate reduction).

Pdf's of subband coefficients obtained from natural images have been observed to follow, with a good approximation, a generalized Gaussian distribution [2]:

$$G(x) = a \exp\left(-|bx|^\gamma\right)$$ (17)

with

$$a = \frac{b^\gamma}{2\Gamma(1/\gamma)}$$ and $$b = \frac{1}{\delta_x} \sqrt{\Gamma(3/\gamma)/\Gamma(1/\gamma)}$$ (18)

where $\Gamma(\cdot)$ is the “Gamma function” [20]. From the source pdf (17) and the quantization law, it is well known how to precisely determine the quantization error pdf. Given $f_X(x)$ the pdf of a subband source, and $\{I_k, k \in \{1, \ldots, K\}\}$ the set of events of the type $I_k : X \in (x_{k-1}, x_k]$ with probability $P_k = \text{Prob}[I_k]$, the quantization error pdf can be calculated using the total probability theorem:

$$f_q(q) = \sum_{k=1}^K P_k f_q(q|I_k),$$ (19)

where the $f_q(q|I_k)$ is the zero-centered version of $f_X(x|I_k)$ with respect to the $I_k$ center value $x_k$,

$$f_q(q|I_k) = f_X(q + x_k|I_k)$$ (20)

with

$$f_X(x|I_k) = \begin{cases} f_X(x)/P_k, & I_k, \\ 0 & \text{otherwise}. \end{cases}$$ (21)

For high-resolution conditions one can confidently assume $f_q(q) = 1/\Delta \cdot \text{rect}(\Delta)$. Otherwise data-error dependencies must be taken into account. The distribution (17), for $\gamma$ less than unity, exhibits a sharp peak around the origin. Then, if the quantizer belongs to the midtread or dead-zone class (as the BPQ), this distribution peak falls, averaged with the other intervals (see (19)), into the $f_q(q)$; conversely, with midrise quantizers the $f_q(q)$ remains uniform by construction because each $f_q(q|I_k)$ has its symmetrical counterpart. In addition, when using dead-zone quantizers, the non-uniformity due to the zero-centered bin leads to an increased $f_q(q)$ support (range) with respect to the uniform $\Delta$ quantization case. In the BPQ case the error range is twice the uniform quantizer one (with same $\Delta$). Then, for a BPQ, $|q|_{\text{max}} = \Delta$. However, $|q| > \Delta/2$ only for near-zero coefficients and this is reassuring from a coding and artifact perception point of view.

Fig. 9 illustrates the above observations, the $f_X(x)$ is processed using (19), (20) in three different situations. The $f_q(q)$ is sharp when the central interval is predominant in the summation (19), i.e. when the $\Delta$ is relatively large with respect to the $f_X(x)$ peak. By
reducing the $\Delta$, the $f_q(q)$ flattens (tends to uniform) because a smaller fraction of the central peak is picked off and its contribution is no more over-riding in (19).

The three conditions (a)–(c) of Fig. 9 may represent either a random variable with pdf $f_X(x)$ and quantized with $A_a$, $A_b$ and $A_c$, or three different RVs pdf (17) with ($\sigma_X$, $\gamma_X$), ($\sigma_Y$, $\gamma_Y$), ($\sigma_Z$, $\gamma_Z$) quantized with the same $\Delta$. These three RVs could also be associated to three subbands of the same wavelet transform, or from another point of view, to the same subband for three different input data.

For an easier understanding of what follows, a shape factor $\varphi = \sigma_a/\Delta$ can be defined. Its value is $1/\sqrt{12}$ for uniform distributions while Fig. 9 reveals that $\varphi_a < \varphi_b < \varphi_c$.

In (8) as well as in (11) an alternative expression of $\sigma^2_{q_jb}$ can be used:

$$\sigma^2_{q_jb} = \varphi^2_{A_{j,b}} \cdot \Delta^2,$$

(22)

where

$$\varphi_{A_{j,b}} = \varphi(\Delta, \sigma_{q_jb}, \gamma_{q_jb}).$$

(23)

According to the observations made in Fig. 9 and for a certain data to be coded, Eqs. (22)–(23) allow to better see the dependencies of $\sigma^2_{q_jb}$ from the quantization $\Delta$.

4.2. Reconstruction error modeling for a bit-plane-based wavelet coding

Limitations introduced by the i.i.d. hypothesis (see Section 2.1) on the subband-based quantization error variances $\sigma^2_{q_jb}$ which act in the CS model (see Eqs. (8), (11)) are now reconsidered. In particular, the independence assumption is weakened and observation will be made on what happens to the subband-wise identical distribution (i.d.) hypothesis from high to low bit-rates.

Statistical independence is only an ideal assumption. In order to model the output reconstruction error as a weighted sum of the subband quantization error (e.g. see (2) and (3)), it is sufficient to have uncorrelated subband quantization error samples. Thus the original hypothesis can be reduced to a wide sense CS (WSCS). Under the well-known conditions of sufficient signal dynamics, low over-sampling rate and lack of periodicity, a uniform quantizer produces a highly uncorrelated error even when using few quantization levels [10]. Due to the properties of the wavelet transform, the subbands contain critically sampled coefficients with a low degree of correlation, $\rho_j(n) \to 0$ for $n \neq 0$. However, at low bit rates a great amount of near-zero coefficients falls into the dead-zone of the BPQ. This leads to the evidence that most quantization error samples coincide with the near-zero wavelet coefficients. This fact does not pose great problems in terms of dependencies as the wavelet coefficients are highly uncorrelated, but even if a “zero-valued coefficient” correlation is present, its contribution to the overall correlation $\rho_j(1)$ remains moderate. Thus, it is reasonable to consider uncorrelated quantization error in the different subbands.

The second issue is the i.d. part of the initial assumptions. As anticipated by the considerations made in Section 4.1 the i.d. assumption is no longer valid when the bit-rate drops. In these cases, the interest is on a qualitative evaluation of the WSCS model modifications deriving from the bit-rate reduction. For the sake of clarity, the model degradations will be observed on a discrete set of points each one reflecting the end of a certain quantization bit-plane $b \in [1, B]$ associated to the quantization step $A_b$ with $B$ corresponding to the least significant bit-plane. If $B$ is sufficiently large, one can assume that, with $A_b$, the high-resolution i.i.d. conditions are valid for the whole subband set. In such a high-resolution case, $\varphi_{l,b}$ is worth about $1/\sqrt{12}$ and it becomes easy to calculate the reconstruction error properties, e.g. by using (10)–(14). Specifically, by considering two adjacent bit-planes, i.e. $A_{b+1} = A_b/2$, one can easily verify the classical 6 dB difference e.g. between the $sP_{SNR}[n]$ levels along any slicing direction:

$$sP_{SNR}[n, b + 1] = sP_{SNR}[n, b] + 6.02 \text{ dB}. \quad (24)$$

The 6 dB value relies on the fact that, with a fixed shape-factor (22), all the $\sigma^2_{q_jb}$ values are scaled by the same value as $A_b$ is halved from bit-plane $b$ to bit-plane $b + 1$. This situation changes by decreasing $b$, when a certain set of subbands modify their quantization error shape factor $\varphi$.

As it can be expected (and experimentally verified in the next section), with progressive or fine grain scalable coders, quantization parameter modifications between adjacent bit-planes can be considered gradual.

In the next section it is shown that this is not always guaranteed and the quantization model could yet be degraded at the finest quantization level.
value (see Eq. (23)). In particular, it is possible to observe that:
\[ \varphi^2(A_b, l, s) \leq \varphi^2(A_{b+1}, l, s), \quad (25) \]
\[ \varphi^2(A_b, l, s) \geq \varphi^2(A_b, l + 1, s). \quad (26) \]
Equality holds when both terms of (25) or (26) are in high-resolution conditions; when these conditions degrade the following two observations can be made, actually constituting the proposed degradation model description:

(a) Relation (25) actually determines a shrinking of the average \( sP_{SNR} \) distance among adjacent bit-planes, with respect to the 6 dB rule of (24). This trend persists until the very last bit-planes (with \( b \) near to 1) when other degenerations occur for extremely low bit-rates, i.e. when the reconstructed data is useless.

(b) The main effect of (26) is to introduce a certain decaying gradient of the shape-factor linked to an increase of the decomposition level \( l \). Looking at (11) and (22) it is possible to understand how (26) determines a smaller contribution from the high-level detail subbands to the CS oscillation pattern of the reconstruction error variance matrix \( \tilde{\sigma}_r \), with respect to the case of high-resolution conditions (in this latter case, see Section 2.3 and Fig. 4, the contribution of the high-level subbands is dominant on the CS pattern). In short, looking at the \( sP_{SNR} \) fluctuation of (14), the presence of (26) influences its amplitude and causes a reduction whose entity will be experimentally evaluated in the next section (see Section 5.1.2).

5. Experimental observations and results

A series of experimental results are proposed in order to evaluate, based on real data, the reconstruction error models presented above. A first part is devoted to the quantitative validation of the \( sP_{SNR} \) fluctuation model and of its degradation. Then, some visual effects of the \( sP_{SNR} \) oscillations are shown and discussed. Finally, a brief discussion on possible further application of the model is proposed. The 3D data that has been used are mainly medical volumetric data-sets, coming from MR and CT 3D scanners. Visual results on a test video sequence will be presented too. The quality measures are extracted from single observations (intra-data-set), exploiting the presence of a large number of voxels. In fact, thanks to a statistically significant number of samples contained in a volume slice and assuming, as it can be made, the ergodicity of the single components of the CS cubic pattern (cycloergodicity), it is possible to verify the proposed statistical analysis with single realization analyses. In particular, estimated values of \( \tilde{m}SE_r[n] \) can be compared with the measured \( \tilde{m}SE_r[n, x, \tilde{x}] \), where for example:

\[
\tilde{m}SE_r[0, n, x, \tilde{x}] = \frac{1}{D_1D_2} \sum_{n_1=0}^{D_1-1} \sum_{n_2=0}^{D_2-1} ((x[n, n_1, n_2] - \tilde{x}[n, n_1, n_2])^2 \quad (27)
\]

with \( D_i=0,1,2 \) the size of the volume in voxels.

If not differently indicated, the progressive bit-plane-based coding algorithm used for the following experiments is the 3D version of the embedded morphological dilation coding (EMDC) algorithm [14,13].

5.1. Model verification

5.1.1. Error fluctuations and filter properties

The main objective of the following experiment is to verify the fluctuation model of Section 3 and highlight the differences between even- and odd-length biorthogonal wavelet filters. Due to their widely recognized good coding performances, the 9/7 and the 10/18 filter banks has been selected for this test. Fig. 10(a)–(d) shows the \( sP_{SNR} \) measures associated to the considered filters for the MR volume MR-BRAIN SAG 256 \times 256 \times 128. The diagrams refer to the four resulting combinations when considering two coding-rates and two slicing directions. The selected slicing directions are the original one, being perpendicular to the sagittal plane (128 slices along \( n_2 \)), and the coronal one (256 slices along \( n_0 \)). The sagittal anatomical symmetry leads to a U-shaped distortion main trend along the \( n_2 \) axis which is essentially due to the actual space occupied by the useful MR brain signal with respect to the noisy background, at each slicing position. The \( sP_{SNR} \) oscillation due to the 9/7 filters is clearly visible. The same oscillatory behavior appears along the \( n_0 \) slicing direction as shown in Fig. 10(c) and (d) where a subset of slices is considered to allow a more detailed view. Even along \( n_1 \) the same behavior can be observed. The \( sP_{SNR} \) oscillation is regular, very similar to the one in Fig. 8 and close to 0.6–0.8 dB for the 9/7 filters; whereas using the 10/18 filters the oscillation
contracts to less than 0.1 dB, blurring itself into the data dependent entropic fluctuations. Moreover, the oscillating behavior does not depend on a particular choice of the coding rate or phase. To show this, a first test point corresponding to the completion of 10 out of 14 bit-planes\(^5\) has been selected and compared to a nearby second one (0.4 coded bpv) falling in a non-specific bitstream location between the ninth and the 10th bit-plane. Fig. 10(a) and (c) are related to the first test point, while Fig. 10(b) and (d) correspond to the second one. Fig. 10 also discovers that 10/18 filters perform better than 9/7 filters not only for the oscillation amplitude but also in terms of mean quality value. This contrasts with the high-resolution predictions shown in Fig. 8. As already anticipated this is probably due to better approximation properties of 10/18 filters with respect to the data considered here. Such properties shape the reconstruction error statistics and allows the observed higher performance.

5.1.2. Bit-planes variation and model degradation

Experimental evidence of the oscillation model degradation described in Section 4 is now presented. In particular, it is shown how some features of the data to be coded have an influence on the \(sP_{SNR}\) behavior. Fig. 11(a) depicts the MR-BRAIN-Cor volume case (\(xyz\) dimension: 256 × 256 × 64 voxels). 9/7 filters and five levels of 3D wavelet decomposition have been used. Bit-planes \(b\) from 1 to 12 are considered and the \(sP_{SNR}\) values after each bit-plane coding completion observed. The \(sP_{SNR}\) signal was calculated along the MR scanning axis \(n_2\) (or \(z\)). For the higher bit-planes 11 and 12 the high-resolution conditions can be recognized in the expected \(sP_{SNR}\) fluctuations (see Section 3.2) and the 6 dB distance of the curves. Owing to (25), in the bit-plane range from 5 to 10, the \(sP_{SNR}\) curves become closer. The CS oscillation diminishes but

\(^5\)The actual DWT coefficient range is strongly data and filter-dependent. In the present case the uncoded 8 bit per voxel (bpv) volume generated DWT coefficients which integer part needs 14 bits to be represented.
Fig. 11. $sP_{\text{SNR}}$ curves as a function of the bit-plane number $b$, obtained along the $n_2$ (or z) axis of the volumes (a) MR-BRAIN_COR and (b) CT-ABDOMEN, using 9/7 filters.
another pseudo-periodic fluctuation gains ground, with a sort of symmetry center on slice 34. This effect does not find an explanation in our model nor in some anatomical symmetry (the considered MR-BRAINCOR data-set is coronal, i.e. acquired along the frontal direction), but it is rather due to an interpolation-related effect. In fact, the 64 slices volume has been intentionally obtained from a 141 slices original one. This subsampling implies an entropic alteration of the slices which are governed by the resampling ratio. Interesting here is not studying the details of this effect, but observing that in this case, as a consequence of (26), the CS fluctuations gradually decrease and another effect may become dominant.

Another example is given in Fig. 11(b), where a 256 × 256 × 64 CT-ABDOMEN volume has been decomposed (on five levels) still using 9/7 filters. In this case, only CS-related fluctuations take place because original (non-interpolated) data has been used. Obviously the short-range CS variations are superimposed to the trend determined by the sensibility of the coding algorithm to the data information content (entropy) on every slicing plane. This long-range fluctuation trend is a passive outcome and does not affect the visual quality of salient information because the coding effort is not made intentionally selective. In Fig. 11(b), it can be observed that as far as the bit-plane $s_P^{SNR}$ distance is concerned, the high-resolution conditions are not satisfied from the beginning, while the amplitude of the CS fluctuations have a value around 0.8 dB (well matched with the high-resolution prediction) for several $b$ values (from 9 to 12). Then, the fluctuation model degrades in a different way for the two considered, equally sized, volumes. This evidences and confirms the hypothesis of a data-dependent model degradation, where high-frequency content or noise level are determinant features. In general, and also in our case, CT data-set are less noisy compared to the MR ones. As a consequence CT-related high-level detail subband histograms are more concentrated on the zero peak, and this determines a populated set of subbands for which (25) and (26) holds with inequality, even for near-lossless coding conditions. As already mentioned, the $s_P^{SNR}$ oscillation reduces its amplitude but does not remain negligible until the eighth bit-plane is reached. In our experience the rates of interest for the lossy coding of MR and CT data-sets falls between the 9th and the 11th bit-plane, thus at the peak of oscillation.

5.2. Visual perception of the $s_P^{SNR}$ oscillations

Volumetric static (medical) data-sets are usually visualized on wide and high-definition displays that reproduce a lightbox where slices are placed side by side and form a single sight. Another usual way to see such data is in a movie loop where slices are played with a low reproduction rate or, more commonly, manually run. Hence, it is important to assess in which measure $s_P^{SNR}$ oscillations are related to unpleasant visual effects when 3D static data are displayed as stated above. Moreover, due to the relevance of the subject, it is our interest to approach the thickening case of wavelet-based video coding. Due to the intrinsic difficulty in giving and synthesizing numerical results about visual perception, observations and examples that, according to our experience, can be taken as representative of different issues and situations are here presented. For clarity, the main conclusions of the visual analysis are anticipated prior to the presentation and discussion of some tests.

- First of all, no $s_P^{SNR}$-related visual effect has never been seen when using even-length wavelet filters (such as 10/18 or 22/14).
- Using odd-length 9/7 filters, potentially visible effects arise, their visibility depending on light conditions, display devices and viewers expertise.
- When short odd-length filters (see Fig. 8) are used, the visual effects of the induced $s_P^{SNR}$ oscillation are noticeable, even to the non-expert eye. Objectionable effects have been observed both in the case of static volumes and video coding.

5.2.1. Static medical volumes

In Fig. 12 four contiguous slices of the 256 × 256 × 128 MR data-set MR-BRAINCOR are shown. Slices are taken along the $n_1$ (i.e. $y$) direction. The 3D wavelet transform is calculated using the 9/3 filter bank along each direction. The coding rate is 0.27 bpv (compression ratio CR = 30). Fig. 13 refers to the same decoded volume but sliced along the $n_0$ (i.e. $x$) direction. Differences between upper and lower images can be quite easily perceived on display and on printed paper (with adequate printing quality). Upper images are more detailed while the lower ones are more blurred. Even if, at first, this quality difference apparently could not be considered significative, it was observed that when experts (physicians or image processing professionals) or
non-expert viewers were allowed to look through the various slice of the whole volume with a manual back and forth movie facility, they were all able, in few instants, to notice a quality alternation between consecutive slices. In Fig. 14 (two upper signals) the $sP_{SNR}$ signals measured on the decoded data-set according to the two considered slicing directions are shown. In order to give a more clear figure the viewing area has been restricted to a slice interval (from slice nr. 70 to nr. 140).

In the case of medical data our general suggestion is to avoid using 9/7 filters. In our view, producing coded data with a significant $sP_{SNR}$ oscillation can decrease the overall quality and reliability of the data itself. For example, with respect to an agreed compression level, apparently hidden visual effects could reveal themselves in some detailed data observations, or results generated by some 3D post-processing tasks (on coded data) could be altered by the $sP_{SNR}$ fluctuations.

5.2.2. Wavelet coding of moving pictures

The visual analysis is completed by considering a representative case of wavelet video coding. Wavelet-based video coding intrinsically enables a series of interesting features such as spatio-temporal and quality scalability that are of particular interest for today’s multimedia and networking applications. Moreover, efficient MCTF solutions [35,15] allowed scalable wavelet video coding to reach performances comparable to the ones of the best video coding standards (AVC-H.264). Typical decoded videos generated by wavelet video codecs present periodic oscillations in the time direction (here the only one direction of interest). The bottom trace in Fig. 14 reports the $sP_{SNR}$ (i.e. the frame by frame $P_{SNR}$) in the case of a decoded test video sequence; the wavelet-based scalable video codec described in [1] has been used on the Harbour sequence (in QCIF and YUV 4:2:0 format) at the coding rate of 92 kbps. The 3D decomposition is not purely dyadic.
but a complete temporal decomposition (MCTF with 5/3 integer lifting wavelet decomposition) is followed by a 9/7 dyadic wavelet spatial decomposition performed on each temporal subband frame. Such decomposition structure is typical in wavelet video coding systems, where the 5/3 filters are commonly used for their recognized good matching between a reliable motion estimation (very similar to the bidirectional motion estimation of hybrid schemes) and the availability of related efficient MCTF implementations [18]. The $sP_{SNR}$ oscillation of Fig. 14 evidences the presence of CS phenomena of a significant entity (compatible with the use of 5/3 filters). However, because of the differences (hereinafter itemized) between the assumptions made in this work and the video coding architecture, the oscillation pattern differs from the one issued by the 3D model. Fig. 15 shows how two consecutive frames (original frames and the relating decoded ones are presented) can visually differ due to the $sP_{SNR}$ variations. Because of the low bit-rate (92 kbps), decoding artifacts are clearly visible both in texture and in color (some blue to red shifts). The quality disparity between the two consecutive frames is evident (the PSNR jump in this case is close to 5 dB indeed) and the two selected frames are representative of the worse situation. However, quality alternation is actually visible along the whole video when manual frame selection or very low frame-rate reproductions are used. At normal reproduction frame-rates, flickering effects can be perceived.

A quantitative analysis of the video coding case study is beyond the scope of this work. A series of factors can apparently influence the CS pattern oscillation (e.g. the decomposition structure, the MCTF, different basis for spatial and temporal filtering, different subband weighing), however, the

![Fig. 13. Visual results on the MR-BRAINCor volume along the x slicing direction.](image-url)
more significant one is the common use of an integer-to-integer (non-linear) lifting wavelet decomposition.\footnote{The use of integer-to-integer wavelet transforms allows to preserve memory and decrease the computational cost. These both are critical factors when considering the processing of a huge quantity of data, as in the case of video coding.} This has been demonstrated to introduce CS error fluctuations [21]. In this case, fluctuations are due to the non-linear rounding noise propagation effects produced into the lifting computation structure. Even if in [21] CS fluctuations are not described in the data domain, it can be inferred that when integer-to-integer transforms are used two cyclostationarity contributions (due to linear and non-linear effects) adds up to produce the overall CS oscillation pattern.

5.3. Further model exploitation

The given model and the presented experiments aim to explain and show in detail the CS reconstruction quality oscillatory phenomena that arise in wavelet coding schemes. The “risks” in using odd-length filters have been highlighted and, as a natural consequence, the even-length filter usage is suggested. Wavelet coders can alternatively exploit the model results in order to know how to shape and reduce the quality oscillations, especially when odd-length filters are used. To obtain oscillation reductions, wavelet coefficients should be slightly weighted before quantization, in a way that alteration of the coding principles and performance were not introduced. Depending on coding system or quality needs such modality can be generally adopted in wavelet coding systems. Inverse weighting must be done at the decoder side, before IDWT. Weights can be applied either on entire subbands or on single coefficients (according to a periodic pattern fitting the CS one). In such a way Eqs. (2) and (3) can be rewritten in the form:

\begin{alignat}{1}
\sigma_{r0}^2 &= \sigma_{q0}^2 + \sigma_{r1}^2 = \sigma_{q0}^2 G^2(e_0[n]) + \sigma_{q1}^2 G^2(e_0[n]), \\
\sigma_{r1}^2 &= \sigma_{q0}^2 + \sigma_{r1}^2 = \sigma_{q0}^2 G^2(e_1[n]) + \sigma_{q1}^2 G^2(e_1[n]),
\end{alignat}

\[ (28) \]

\[ (29) \]

with \( \sigma_{qj}^2 \), \( w_i \), and \( w_i^j \) the power weight for subband \( i = 0, 1 \) and polyphase component \( j = 0, 1 \).

---

Fig. 14. \( s_{PSNR} \) oscillations on the MR-brainCor volume and on a test video.
In the general multidimensional case a weighting pattern $w_{l,s}^{(\bar{n}_0, \bar{n}_1, \ldots)}$ could be designed in order to obtain

$$\sigma^2_{\bar{q},l,s} = w_{l,s}^{(\bar{n}_0, \bar{n}_1, \ldots)} \cdot \sigma^2_{\bar{q}}$$

(30)

that, e.g. in the 3D case, can be used in (10) and (11).

There are various way to design and produce the $w_{l,s}^{(\bar{n}_0, \bar{n}_1, \ldots)}$ pattern. Such weighting must be validated on specific data according to application-dependent quality requirements. These aspects are not further explored here. As stated the weights should not alter the coding mechanism nor the coding performance in a sensible way. As visual performance could be improved by a perceptual subband weighting, one could entail a joint weights optimization to improve the visual performance while decreasing the objective fluctuations.

6. Conclusions

In this paper, we focused on the problem of the presence of cyclostationary oscillations of the reconstruction error power, especially when this error is measured on consecutive slices of a 3D wavelet (de)coded data-set. The work is motivated by the application relevance of such problem in a 3D wavelet coding environment both for volumetric (medical applications) and video (with or without MCTF) data-sets. Being the volume slices (video frames) at adjacent spatial (temporal) positions quite similar, an appreciable $sP_{SNR}$ oscillation may be critical at various coding rates, causing objectionable artifacts and/or lowering the reliability of the coding process.

In the first part of the paper, we considered the LPTV properties of a multidimensional $L$-level wavelet transform and we formulated, in the data
domain and under high-resolution conditions, a multidimensional CS error model. We observed and quantified the periodic oscillation patterns of the error variance on the lattice of data elements and associated them to the properties of biorthogonal wavelet filters. We found a peculiar property of even-length symmetric (biorthogonal) FIR filters that nullifies the error variance oscillation for 1D data and 1-level of wavelet decomposition. Then we observed how this favorable condition could be beneficial also on multidimensional and multilevel cases.

We carried out the error model by deriving and quantifying the error power oscillations when, instead of being referred to single data positions, they are associated (by statistical averaging) on portions of data elements. In particular, because of their application relevance, we considered the 2D slices obtained from a 3D data-set according to any of the three orthogonal slicing directions. The favorable case of even-length filters employment is confirmed on such averaged error measures as well. Conversely, we found out that odd-length filters, especially the shortest ones, produce significant oscillations (in a range of 1 to 3 dB).

Simplifying high-resolution hypothesis was used to highlight the fundamental CS nature of the reconstruction error and to predict and quantify the $P_{SNR}$ oscillation pattern in their coding rates range of validity.

In the second part of the paper we proposed a reconstruction error model degradation analysis for medium to low bit-rates. We based our analysis on an observation of the quantization error pdfs on different wavelet subbands when a bit-plane quantization is applied (as it is the case of almost all current wavelet coding techniques). Even if in this case only qualitative deductions are possible, we got a sufficient degree of detail to predict and interpret the phenomenon of a gradual reduction of the CS oscillations towards low bit-rates and some model reliance on data content aspects.

In the final part of the paper we presented some experimental results which confirm the high-resolution model outcomes and verify the model degradation hypotheses on real 3D medical data-sets. We also studied in which measure the $sP_{SNR}$ fluctuations generate perceptible visual quality disparities among the reconstructed slices of a static volume or the frames of a decoded video. The results show that a $sP_{SNR}$ oscillation of about 1 dB is visible on static data-sets then representing a detrimental factor especially in diagnostic imaging application fields. For these reasons the usage of even-length filters (showing good coding performance) is to prefer in order to make the coding process more reliable and the coding rate determination free from the oscillation issues.

We also evidenced CS oscillations along the temporal direction in the more complex case of MCTF wavelet video coding. The typical usage of short odd-length filter for bidirectional motion compensation generates (with possible superposition of others effects) important $sP_{SNR}$ disparities that are clearly visible at reduced frame-rates.

References