Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2007, Article ID 31319, 12 pages
doi:10.1155/2007/31319

Research Article

A Motion-Compensated Overcomplete Temporal Decomposition for Multiple Description Scalable Video Coding

Christophe Tillier, Teodora Petrişor, and Béatrice Pesquet-Popescu

Signal and Image Processing Department, École Nationale Supérieure des Télécommunications (ENST), 46 Rue Barrault, 75634 Paris Cédex 13, France

Received 26 August 2006; Revised 21 December 2006; Accepted 23 December 2006

Recommended by James E. Fowler

We present a new multiple-description coding (MDC) method for scalable video, designed for transmission over error-prone networks. We employ a redundant motion-compensated scheme derived from the Haar multiresolution analysis, in order to build temporally correlated descriptions in a $t+2D$ video coder. Our scheme presents a redundancy which decreases with the resolution level. This is achieved by additionally subsampling some of the wavelet temporal subbanbds. We present an equivalent four-band lifting implementation leading to simple central and side decoders as well as a packet-based reconstruction strategy in order to cope with random packet losses.

1. INTRODUCTION

With the increasing usage of the Internet and other best-effort networks for multimedia communication, there is a stringent need for reliable transmission. For a long time, the research efforts have been concentrated on enhancing the existing error correction techniques, but during the last decades an alternative solution has emerged and is gaining more and more popularity. This solution mainly answers the situation in which immediate data retransmission is either impossible (network congestion or broadcast applications) or undesirable (e.g., in conversational applications with very low delay requirements). We are referring to a specific joint source-channel coding technique known as *multiple-description coding* (MDC). A comprehensive presentation of MDC is given in [1].

The MDC technique leads to several correlated but independently decodable (preferably with equivalent quality) bitstreams, called *descriptions*, that are to be sent over as many independent channels. In an initial scenario, these channels have an on-off functioning: either the bitstream is flawlessly conveyed or it is considered unusable at the so-called *side decoder* end if an error had occurred during the transmission. According to this strategy, some amount of redundancy has to be introduced at the source level such that an acceptable reconstruction can be achieved from any of the bitstreams. Then, the reconstruction quality will be enhanced with every bitstream received.

The application scenario for MDC is different from the one of scalable coding, for example. Indeed, the robustness of a scalable system relies on the assumption that the information has been hierarchized and the base layer is received without errors (which can be achieved, e.g., by adding sufficient channel protection). However, if the base layer is lost, the enhancement layers cannot be exploited and nothing can be decoded. The MDC framework has a complementary approach, trying to cope with the channel failures, and thus allowing the decoding of at least one of the descriptions, when the other is completely lost.

An ingredient enabling the success of an MDC technique is the path diversity, since its usage balances the network load and reduces the congestion probability.

In wireless networks, for instance, a mobile receptor can benefit from multiple descriptions if these arrive independently, for example on two neighbor access points; when moving between these access points, it might capture one or the other, and in some cases both. Another way to take advantage of MDC in a wireless environment is by splitting in

frequency the transmission of the two descriptions: for example, a laptop may be equipped with two wireless cards (e.g., 802.11a and 802.11g), each wireless card receiving a different description. Depending on the dynamic changes in the number of clients in each network, one of them may become overloaded and the corresponding description may not be transmitted.

In wired networks, the different descriptions can be routed to a receiver through different paths by incorporating this information into the packet header [2]. In this situation, a description might contain several packets and the scenario of on-off channels might no longer be suitable. The system should, in this case, be designed to take into consideration individual or bursty packet losses rather than a whole description.

An important issue that concerned the researchers over the years is the amount of introduced redundancy. One has to consider the tradeoff between this redundancy and the resulting distortion. Therefore, a great deal of effort has been spent on defining the achievable performances with MDC ever since the beginning of this technique [3, 4] and until recently, for example, [5]. Practical approaches to MDC include scalar quantization [6], correlating transforms [7], and frame expansions [8]. Our work belongs to the last category and we concentrate on achieving a tunable low redundancy while preserving the perfect reconstruction property of our scheme [9].

In this paper, we present an application of multiple-description coding to robust video transmission over lossy networks, using redundant wavelet decompositions in the temporal domain of a $t + 2D$ video coder.

Several directions have already been investigated in the literature for MD video coding. In [10–13], the proposed schemes mainly involve the spatial domain in hybrid video coders such as MPEG/H.26x. A very good survey on MD video coding for hybrid coders is given in [14].

Only few works investigated the design of MDC schemes allowing to introduce source redundancy in the temporal domain, although the field is very promising. In [15], a balanced interframe multiple-description coder has been proposed starting from the popular DPCM technique. In [16], the reported MDC scheme consists in temporal subsampling of the coded error samples by a factor of 2 so as to obtain 2 threads at the encoder, which are further independently encoded using prediction loops that mimic the decoders (two side prediction loops and a central one).

Existing work for $t + 2D$ video codecs with temporal redundancy addresses three-band filter banks [17, 18] and temporal or spatiotemporal splitting of coefficients in 3D-SPIHT sytems [19–21]. Here, we focus on a two-description coding scheme for scalable video, where temporal and spatial scalabilities follow from a classical dyadic subband transform. The correlation between the two descriptions is introduced in the temporal domain by exploiting an oversampled motion-compensated filter bank. An important feature of our proposed scheme is its reduced redundancy which is achieved by an additional subsampling of a factor of two of the resulting temporal details. The remaining details are then distributed in a balanced manner between the two descriptions, along with the nondecimated approximation coefficients. The global redundancy is thus tuned by the number of temporal decomposition levels. We adopt a lifting approach for the temporal filter-bank implementation and further adapt this scheme in order to design simple central (receiving both descriptions) and side decoders.

This paper relies on some of our previous work which is presented in [22]. Here, we consider an improved version of the proposed scheme and detail its application to robust video coding. The approximation subbands which participate in each description are decorrelated by an additional motion-compensated transform, as it will be explained in Section 5. Moreover we consider two transmission scenarios. In the first one, we tackle the reconstruction when an entire description is lost or when both descriptions are received error-free, and in the second one we discuss signal recovery in the event of random packet losses in each description. For the random-loss case, we compare our results with a temporal splitting strategy, as in [2], which consists in partitioning the video sequence into two streams by even/odd temporal subsampling and reconstructing it at half rate if one of the descriptions is lost.

An advantage of our scheme is to maintain the scalability properties for each of the two created descriptions, allowing to go further than the classical on-off channel model for MDC and also cope with random packet losses on the channels.

The rest of the paper is organized as follows. In Section 2 we present the proposed strategy of building two temporal descriptions. Section 3 gives a lifting implementation of our scheme together with an optimized version well suited for Haar filter banks. We explain the generic decoding approach in Section 4. We then discuss the application of the proposed scheme to robust video coding in Section 5 and the resulting decoding strategy in Section 6. Section 7 gives the simulation results for the two scenarios: entire description loss and random packet losses in each description. Finally, Section 8 concludes the paper and highlights some directions for further work.

2. TEMPORAL MDC SCHEME

The strategy employed to build two temporal descriptions from a video sequence is detailed in this section. We rely on a temporal multiresolution analysis of finite energy signals, associated with a decomposition onto a Riesz wavelet basis.

Throughout the paper, we are using the following notations. The approximation subband coefficients are denoted by a and the detail subband coefficients by d. The resolution level associated with the wavelet decomposition is denoted by j, whereas J stands for the coarsest resolution. The temporal index of each image in the temporal subbands of the video sequence is designated by n and the spatial indices are omitted in this section in order to simplify the notations.

The main idea of the proposed scheme consists in using an oversampled decomposition in order to get two wavelet representations. The superscript symbols I and II distinguish

the coefficients in the first basis from those corresponding to the second one. For example, $d^{\mathrm{I}}_{j,n}$ stands for the detail coefficient in representation I at resolution j and temporal index n. Then a secondary subsampling strategy is applied along with distributing the remaining coefficients into two descriptions. The redundancy is reduced by this additional subsampling to the size of an approximation subband (in terms of number of coefficients).

Let $(h_n)_{n\in\mathbb{Z}}$ (resp., $(g_n)_{n\in\mathbb{Z}}$) be the impulse responses of the analysis lowpass (resp., highpass) filter corresponding to the considered multiresolution decomposition.

For the first $J-1$ resolution levels, we perform a standard wavelet decomposition, which is given by

$$a^{\mathrm{I}}_{j,n} = \sum_k h_{2n-k} a^{\mathrm{I}}_{j-1,k} \tag{1}$$

for the temporal approximation subband, and by

$$d^{\mathrm{I}}_{j,n} = \sum_k g_{2n-k} a^{\mathrm{I}}_{j-1,k} \tag{2}$$

for the detail one, where $j \in \{1,\ldots,J-1\}$.

We introduce the redundancy at the coarsest resolution level J by eliminating the decimation of the approximation coefficients (as in a shift-invariant analysis). This leads to the following coefficient sequences:

$$\begin{aligned} a^{\mathrm{I}}_{J,n} &= \sum_k h_{2n-k} a^{\mathrm{I}}_{J-1,k}, \\ a^{\mathrm{II}}_{J,n} &= \sum_k h_{2n-1-k} a^{\mathrm{I}}_{J-1,k}. \end{aligned} \tag{3}$$

Each of these approximation subbands is assigned to a description.

In the following, we need to indicate the detail subbands involved in the two descriptions. At the last decomposition stage, we obtain in the same manner as above two detail coefficient sequences (as in a nondecimated decomposition):

$$\begin{aligned} d^{\mathrm{I}}_{J,n} &= \sum_k g_{2n-k} a^{\mathrm{I}}_{J-1,k}, \\ d^{\mathrm{II}}_{J,n} &= \sum_k g_{2n-1-k} a^{\mathrm{I}}_{J-1,k}. \end{aligned} \tag{4}$$

Note that the coefficients in representation II are obtained with the same even-subsampling, but using the shifted versions of the filters h and g: h_{n-1} and g_{n-1}, respectively.

In order to limit the redundancy, we further subsample these coefficients by a factor of 2, and we introduce the following new notations:

$$\hat{d}^{\mathrm{I}}_{J,n} = d^{\mathrm{I}}_{J,2n}, \tag{5}$$

$$\check{d}^{\mathrm{II}}_{J,n} = d^{\mathrm{II}}_{J,2n-1}. \tag{6}$$

At each resolution, each description will contain one of these detail subsets.

Summing up the above considerations, the two descriptions are built as follows.

Description 1. This description contains the even-sampled detail coefficients $(\hat{d}^{\mathrm{I}}_{j,n})_n$ for $j \in \{1,\ldots,J\}$, and $(a^{\mathrm{I}}_{j,n})_n$, where, using the same notation as in (5),

$$\hat{d}^{\mathrm{I}}_{j,n} = d_{j,2n}. \tag{7}$$

Description 2. This description contains the odd-sampled detail coefficients $(\check{d}^{\mathrm{I}}_{j,n})_n$ for $j \in \{1,\ldots,J-1\}$, $(\check{d}^{\mathrm{II}}_{J,n})_n$, and $(a^{\mathrm{II}}_{J,n})_n$, where, similarly to (6),

$$\check{d}^{\mathrm{I}}_{j,n} = d_{j,2n-1}. \tag{8}$$

Once again, we have not introduced any redundancy in the detail coefficients, therefore the overall redundancy factor (evaluated in terms of number of coefficients) stems from the last level approximation coefficients, that is, it is limited to $1+2^{-J}$.

The choice of the subsampled detail coefficients at the coarsest level in the second description is motivated by the concern of having balanced descriptions [9].

3. LIFTING-BASED DESIGN OF THE ENCODER

3.1. Two-band lifting approach

Since the first $J-1$ levels are obtained from a usual wavelet analysis, in the following we will be interested mainly in the last resolution level. The corresponding coefficients in the two descriptions are computed as follows:

$$a^{\mathrm{I}}_n = \sum_k h_{2n-k} x_k, \tag{9a}$$

$$\hat{d}^{\mathrm{I}}_n = \sum_k g_{4n-k} x_k, \tag{9b}$$

$$a^{\mathrm{II}}_n = \sum_k h_{2n-1-k} x_k, \tag{9c}$$

$$\check{d}^{\mathrm{II}}_n = \sum_k g_{4n-3-k} x_k, \tag{9d}$$

where, for simplicity, we have denoted by x_k the approximation coefficients at the $(J-1)$th level and we have omitted the subscript J.

We illustrate our scheme in Figure 1, using a one-stage lifting implementation of the filter bank. The p and u operators in the scheme stand for the predict and update, respectively, and γ is a real nonzero multiplicative constant. Note that the lifting scheme allows a quick and memory-efficient implementation for biorthogonal filter banks, but especially it guarantees perfect reconstruction. For readability, we display a scheme with only two levels of resolution, using a basic lifting core.

3.2. Equivalent four-band lifting implementation

The two-band lifting approach presented above does not yield an immediate inversion scheme, in particular when using nonlinear operators, such as those involving motion estimation/compensation in the temporal decomposition of the

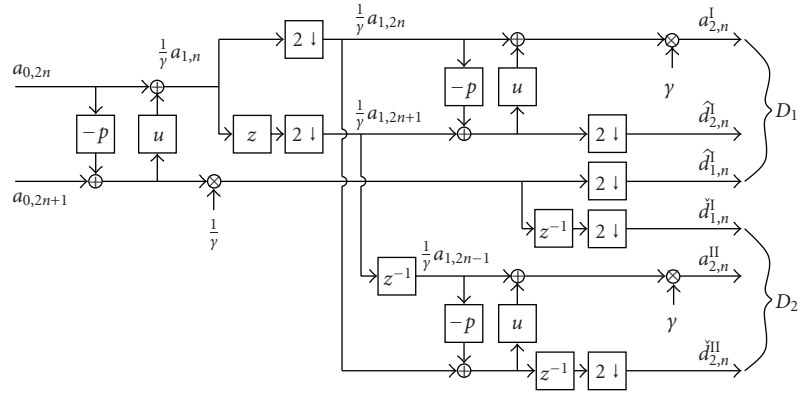

Figure 1: Two-band lifting implementation of the proposed multiple-description coder for the last two resolution levels.

video. This is the motivation behind searching an equivalent scheme for which global inversion would be easier to prove. In the following, we build a simpler equivalent lifting scheme for the Haar filter bank, by using directly the four-band polyphase components of the input signal, instead of the two-band ones. Let these polyphase components of $(x_n)_{n\in\mathbb{Z}}$ be defined as

$$\forall i \in \{0,1,2,3\}, \quad x_n^{(i)} = x_{4n+i}. \tag{10}$$

For the first description, the approximation coefficients can be rewritten from (9a), while the detail coefficients are still obtained with (9b), leading to

$$\begin{aligned} \hat{a}_n^{\mathrm{I}} &= a_{2n}^{\mathrm{I}} = \sum_k h_{4n-k}x_k, \\ \check{a}_n^{\mathrm{I}} &= a_{2n-1}^{\mathrm{I}} = \sum_k h_{4n-2-k}x_k, \\ \hat{d}_n^{\mathrm{I}} &= \sum_k g_{4n-k}x_k. \end{aligned} \tag{11}$$

Similarly, for the second description, we express the approximation subband from (9c) and keep the details from (9d):

$$\begin{aligned} \hat{a}_n^{\mathrm{II}} &= \sum_k h_{4n-1-k}x_k, \\ \check{a}_n^{\mathrm{II}} &= \sum_k h_{4n-3-k}x_k, \\ \check{d}_n^{\mathrm{II}} &= \sum_k g_{4n-3-k}x_k. \end{aligned} \tag{12}$$

Note that the coefficients in the two descriptions can thus be computed with an oversampled six-band filter bank with a decimation factor of 4 of the input signal, which consequently amounts to a redundant structure.

In the sequel of this paper, we will focus on the Haar filter banks, which are widely used for the temporal decomposition in $t+2D$ wavelet-based video coding schemes.

To go further and find an equivalent scheme for the Haar filter bank, note that the two-band polyphase components of the input signal, $x_{2n} = a_{J-1,2n}$ and $x_{2n+1} = a_{J-1,2n+1}$, are first filtered and then subsampled (see Figure 1). However, for the Haar filter bank, recall that the predict and update operators are, respectively, $p = \mathrm{Id}$ and $u = (1/2)\,\mathrm{Id}$ (and the constant $\gamma = \sqrt{2}$). Since these are both instantaneous operators, one can reverse the order of the filtering and downsampling operations. This yields the following very simple expressions for the coefficients in the first description:

$$\hat{a}_n^{\mathrm{I}} = \frac{x_{4n}+x_{4n+1}}{\sqrt{2}} = \frac{x_n^{(0)}+x_n^{(1)}}{\sqrt{2}}, \tag{13a}$$

$$\check{a}_n^{\mathrm{I}} = \frac{x_{4n-2}+x_{4n-1}}{\sqrt{2}} = \frac{x_{n-1}^{(2)}+x_{n-1}^{(3)}}{\sqrt{2}}, \tag{13b}$$

$$\hat{d}_n^{\mathrm{I}} = \frac{x_{4n+1}-x_{4n}}{\sqrt{2}} = \frac{x_n^{(1)}-x_n^{(0)}}{\sqrt{2}}, \tag{13c}$$

and in the second:

$$\hat{a}_n^{\mathrm{II}} = \frac{x_{4n}+x_{4n-1}}{\sqrt{2}} = \frac{x_n^{(0)}+x_{n-1}^{(3)}}{\sqrt{2}}, \tag{14a}$$

$$\check{a}_n^{\mathrm{II}} = \frac{x_{4n-2}+x_{4n-3}}{\sqrt{2}} = \frac{x_{n-1}^{(2)}+x_{n-1}^{(1)}}{\sqrt{2}}, \tag{14b}$$

$$\check{d}_n^{\mathrm{II}} = \frac{x_{4n-2}-x_{4n-3}}{\sqrt{2}} = \frac{x_{n-1}^{(2)}-x_{n-1}^{(1)}}{\sqrt{2}}. \tag{14c}$$

In Figure 2, we schematize the above considerations.

4. RECONSTRUCTION

In this section, we give the general principles for decoders design considering the generic scheme in Figure 2. The next sections will discuss the application of the proposed scheme to robust video coding and more details will be given about the central and side decoders in the video coding schemes. Some structure improvements that lead to better reconstruction will also be presented.

In the generic case, our aim is to recover x_n, the input signal, from the subsampled wavelet coefficients. The components involved in the basic lifting decomposition can be perfectly reconstructed by applying the inverse lifting schemes. However, since we have introduced redundancy, we benefit from additional information that can be exploited at the

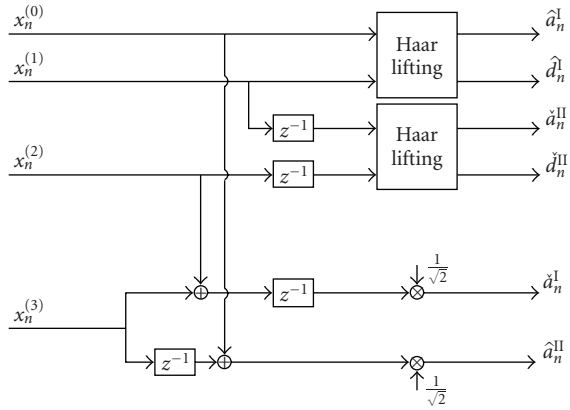

Figure 2: Redundant four-band lifting scheme.

reconstruction. Let us denote the recovered polyphase components of the signal by $\widetilde{x}_n^{(i)}$.

4.1. Central decoder

We first discuss the reconstruction performed at the central decoder. The first polyphase component of x_n is obtained by directly inverting the basic lifting scheme represented by the upper block in Figure 2. The polyphase components reconstructed from $\hat{a}_n^{\mathrm{I}}$ and $\hat{d}_n^{\mathrm{I}}$ are denoted by $y_n^{(0)}$ and $y_n^{(1)}$. Thus, we obtain

$$\widetilde{x}_n^{(0)} = y_n^{(0)} = \frac{[\hat{a}_n^{\mathrm{I}}] - [\hat{d}_n^{\mathrm{I}}]}{\sqrt{2}}, \tag{15}$$

where $[\hat{a}_n^{\mathrm{I}}]$ and $[\hat{d}_n^{\mathrm{I}}]$ are the quantized versions of $\hat{a}_n^{\mathrm{I}}$ and $\hat{d}_n^{\mathrm{I}}$, analogous notations being used for the other coefficients. Obviously, in the absence of quantization, we have $x_n^{(0)} = y_n^{(0)}$ and $x_n^{(1)} = y_n^{(1)}$.

Similarly, the third polyphase component is reconstructed by directly inverting the second two-band lifting block in Figure 2:

$$\widetilde{x}_n^{(2)} = z_{n+1}^{(2)} = \frac{[\check{a}_{n+1}^{\mathrm{II}}] + [\check{d}_{n+1}^{\mathrm{II}}]}{\sqrt{2}}, \tag{16}$$

where the polyphase components reconstructed from $\check{a}_n^{\mathrm{II}}$ and $\check{d}_n^{\mathrm{II}}$ are denoted by $z_n^{(1)}$ and $z_n^{(2)}$.

The second polyphase component of x_n can be recovered as the average between the reconstructed subbands from the two previous lifting blocks:

$$\begin{aligned}\widetilde{x}_n^{(1)} &= \frac{1}{2}\left(y_n^{(1)} + z_{n+1}^{(1)}\right) \\ &= \frac{1}{2\sqrt{2}}([\hat{a}_n^{\mathrm{I}}] + [\hat{d}_n^{\mathrm{I}}] + [\check{a}_{n+1}^{\mathrm{II}}] - [\check{d}_{n+1}^{\mathrm{II}}]).\end{aligned} \tag{17}$$

The last polyphase component of the input signal can be computed as the average between the reconstructions from $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$. Using (13b) and (14a), we get

$$\begin{aligned}\widetilde{x}_n^{(3)} &= -\frac{1}{2}\left(y_{n+1}^{(0)} + z_{n+1}^{(2)}\right) + \frac{1}{\sqrt{2}}([\check{a}_{n+1}^{\mathrm{I}}] + [\hat{a}_{n+1}^{\mathrm{II}}]) \\ &= -\frac{1}{2\sqrt{2}}([\hat{a}_{n+1}^{\mathrm{I}}] - [\hat{d}_{n+1}^{\mathrm{I}}] + [\check{a}_{n+1}^{\mathrm{II}}] + [\check{d}_{n+1}^{\mathrm{II}}]) \\ &\quad + \frac{1}{\sqrt{2}}([\check{a}_{n+1}^{\mathrm{I}}] + [\hat{a}_{n+1}^{\mathrm{II}}]).\end{aligned} \tag{18}$$

4.2. Side decoders

Concerning the side decoders, again from Figure 2, we note that from each description we can partially recover the original sequence by immediate inversion of the scheme. For instance, if we only receive the first description, we can easily reconstruct the polyphase components $x_n^{(0)}, x_n^{(1)}$ from the first Haar lifting block. The last two polyphase components $x_n^{(2)}$ and $x_n^{(3)}$ are reconstructed by assuming that they are similar:

$$\widetilde{x}_n^{(2)} = \widetilde{x}_n^{(3)} = \frac{[\check{a}_{n+1}^{\mathrm{I}}]}{\sqrt{2}}. \tag{19}$$

Similarly, when receiving only the second description, we are able to directly reconstruct $x_n^{(1)}, x_n^{(2)}$ from the second Haar lifting block, while $x_n^{(0)}$ and $x_n^{(3)}$ are obtained by duplicating $\hat{a}_{n+1}^{\mathrm{II}}$:

$$\widetilde{x}_{n+1}^{(0)} = \widetilde{x}_n^{(3)} = \frac{[\hat{a}_{n+1}^{\mathrm{II}}]}{\sqrt{2}}. \tag{20}$$

5. APPLICATION TO ROBUST VIDEO CODING

Let us now apply the described method to robust coding of video sequences. The temporal samples are in this case the input frames, and the proposed wavelet frame decompositions have to be adapted to take into account the motion estimation and compensation between video frames, which is an essential ingredient for the success of such temporal decompositions. However, as shown in the case of critically sampled two-band and three-band motion-compensated filter banks [23–25], incorporating the ME/MC in the lifting scheme leads to nonlinear spatiotemporal operators.

Let us consider the motion-compensated prediction of a pixel $\mathbf{s}$ in the frame $x_n^{(1)}$ from the frame $x_n^{(0)}$ and denote by $\mathbf{v}$ the forward motion vector corresponding to $\mathbf{s}$. Writing now (13a)–(13c) in a lifting form and incorporating the motion into the predict/update operators yield

$$\begin{aligned}\hat{d}_n^{\mathrm{I}}(\mathbf{s}) &= \frac{x_n^{(1)}(\mathbf{s}) - x_n^{(0)}(\mathbf{s}-\mathbf{v})}{\sqrt{2}}, \\ \hat{a}_n^{\mathrm{I}}(\mathbf{s}-\mathbf{v}) &= \sqrt{2}x_n^{(0)}(\mathbf{s}-\mathbf{v}) + \hat{d}_n^{\mathrm{I}}(\mathbf{s}), \\ \check{a}_n^{\mathrm{I}}(\mathbf{s}) &= \frac{x_{n-1}^{(2)}(\mathbf{s}) + x_{n-1}^{(3)}(\mathbf{s})}{\sqrt{2}}.\end{aligned} \tag{21}$$

One can also note that several pixels $\mathbf{s}_i$, $i \in \{1\ldots,N\}$, in the current frame $x_n^{(1)}$ may be predicted by a single pixel in the reference frame $x_n^{(0)}$, which is called in this case multiple

connected [26]. Then, for the pixels $\mathbf{s}_i$ and their corresponding motion vectors $\mathbf{v}_i$, we have $\mathbf{s}_1 - \mathbf{v}_1 = \cdots = \mathbf{s}_i - \mathbf{v}_i = \cdots = \mathbf{s}_N - \mathbf{v}_N$. After noting that the update step may involve all the details $\hat{d}_n^{\mathrm{I}}(\mathbf{s}_i)$, $i \in \{1, \ldots, N\}$, while preserving the perfect reconstruction property, we have shown that the update step minimizing the reconstruction error is the one averaging all the detail contributions from the connected pixels $\mathbf{s}_i$ [27]. With this remark, one can write (21) as follows:

$$\hat{d}_n^{\mathrm{I}}(\mathbf{s}_i) = \frac{x_n^{(1)}(\mathbf{s}_i) - x_n^{(0)}(\mathbf{s}_i - \mathbf{v}_i)}{\sqrt{2}}, \quad i \in \{1, \ldots, N\}, \tag{22a}$$

$$\hat{a}_n^{\mathrm{I}}(\mathbf{s}_i - \mathbf{v}_i) = \sqrt{2}x_n^{(0)}(\mathbf{s}_i - \mathbf{v}_i) + \frac{\sum_{\ell=1}^{N} \hat{d}_n^{\mathrm{I}}(\mathbf{s}_\ell)}{N}, \tag{22b}$$

$$\check{a}_n^{\mathrm{I}}(\mathbf{s}) = \frac{x_{n-1}^{(2)}(\mathbf{s}) + x_{n-1}^{(3)}(\mathbf{s})}{\sqrt{2}}, \tag{22c}$$

and with similar notations for multiple connections in the second description:

$$\check{d}_n^{\mathrm{II}}(\mathbf{s}_i) = \frac{x_{n-1}^{(2)}(\mathbf{s}_i) - x_{n-1}^{(1)}(\mathbf{s}_i - \mathbf{v}_i)}{\sqrt{2}}, \quad i \in \{1, \ldots, M\}, \tag{23a}$$

$$\check{a}_n^{\mathrm{II}}(\mathbf{s}_i - \mathbf{v}_i) = \sqrt{2}x_{n-1}^{(1)}(\mathbf{s}_i - \mathbf{v}_i) + \frac{\sum_{\ell=1}^{M} \check{d}_n^{\mathrm{II}}(\mathbf{s}_\ell)}{M}, \tag{23b}$$

$$\hat{a}_n^{\mathrm{II}}(\mathbf{s}) = \frac{x_n^{(0)}(\mathbf{s}) + x_{n-1}^{(3)}(\mathbf{s})}{\sqrt{2}}. \tag{23c}$$

Since for video coding efficiency, motion prediction is an important step, we propose an alternative scheme for building the two descriptions, in which we incorporate the motion estimation/compensation in the computation of the second approximation sequence ($\hat{a}_n^{\mathrm{I}}$, resp., $\check{a}_n^{\mathrm{II}}$). This scheme is illustrated in Figure 3. Per description, an additional motion vector field needs to be encoded. In the following, this scheme will be referred to as 4B_1MV. In the case of the 4B_1MV scheme, if we denote by $\mathbf{u}$ the motion vector predicting the pixel $\mathbf{s}$ in frame $x_{n-1}^{(3)}$ from $x_{n-1}^{(2)}$ and by $\mathbf{w}$ the motion vector predicting the pixel $\mathbf{s}$ in frame $x_n^{(0)}$ from $x_{n-1}^{(3)}$, the analysis equations for $\hat{a}_n^{\mathrm{I}}$ and $\check{a}_n^{\mathrm{II}}$ can be written as

$$\check{a}_n^{\mathrm{I}}(\mathbf{s} - \mathbf{u}) = \frac{x_{n-1}^{(3)}(\mathbf{s}) + x_{n-1}^{(2)}(\mathbf{s} - \mathbf{u})}{\sqrt{2}}, \tag{24}$$

$$\hat{a}_n^{\mathrm{II}}(\mathbf{s} - \mathbf{w}) = \frac{x_{n-1}^{(3)}(\mathbf{s} - \mathbf{w}) + x_n^{(0)}(\mathbf{s})}{\sqrt{2}} \tag{25}$$

for the connected pixels (here, only the first pixel in the scan order is considered in the computation), and

$$\begin{aligned} \check{a}_n^{\mathrm{I}}(\mathbf{s}) &= \sqrt{2}x_{n-1}^{(2)}(\mathbf{s}), \\ \hat{a}_n^{\mathrm{II}}(\mathbf{s}) &= \sqrt{2}x_{n-1}^{(3)}(\mathbf{s}) \end{aligned} \tag{26}$$

for the nonconnected pixels.

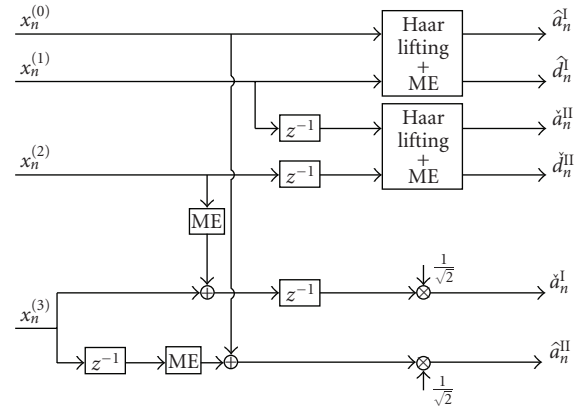

Figure 3: Four-band lifting scheme with motion estimation on the approximation subbands.

Furthermore, a careful analysis of the video sequences encoded in each description revealed that the two polyphase components of the approximation signals that enter each description are temporally correlated. This suggested us to come up with a new coding scheme, illustrated in Figure 4, where a motion-compensated temporal Haar transform is applied on $\hat{a}_n^{\mathrm{I}}$ and $\check{a}_n^{\mathrm{I}}$ (resp., on $\check{a}_n^{\mathrm{II}}$ and $\hat{a}_n^{\mathrm{II}}$). Compared to the original structure, two additional motion vector fields have to be transmitted. The scheme will thus be referred to as 4B_2MV. In Figure 5 the temporal transforms involved in two levels of this scheme are represented. One can note the temporal subsampling of the details on the first level and the redundancy at the second level of the decomposition.

6. CENTRAL AND SIDE VIDEO DECODERS

The inversion of (22a) and (22b) is straightforward by the lifting scheme, allowing us to reconstruct the first two polyphase components. Using the same notations as in Section 4, the reconstructed polyphase components from the first description are as follows:

$$\begin{aligned} \tilde{x}_n^{(0)}(\mathbf{s}_i - \mathbf{v}_i) &= \frac{1}{\sqrt{2}}\left(\left[\hat{a}_n^{\mathrm{I}}(\mathbf{s}_i - \mathbf{v}_i)\right] - \frac{1}{N}\sum_{\ell=1}^{N}\left[\hat{d}_n^{\mathrm{I}}(\mathbf{s}_\ell)\right] \right), \\ \tilde{x}_n^{(1)}(\mathbf{s}_i) &= \frac{1}{\sqrt{2}}\left(\left[\hat{a}_n^{\mathrm{I}}(\mathbf{s}_i - \mathbf{v}_i)\right] + 2\left[\hat{d}_n^{\mathrm{I}}(\mathbf{s}_i)\right] - \frac{1}{N}\sum_{\ell=1}^{N}\left[\hat{d}_n^{\mathrm{I}}(\mathbf{s}_\ell)\right] \right). \end{aligned} \tag{27}$$

When analyzing the reconstruction of the connected pixels in the first two polyphase components, one can note that it corresponds to the inverse lifting using the average update step.

A similar reasoning for the second description allows us to find the reconstruction of the sequence from the received

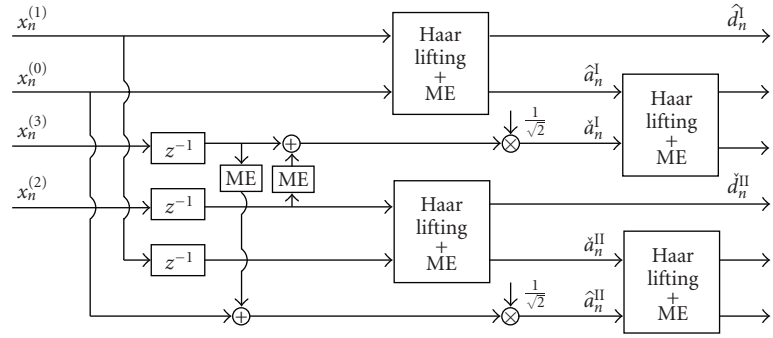

Figure 4: Four-band lifting scheme with motion estimation and Haar transform on the approximation subbands.

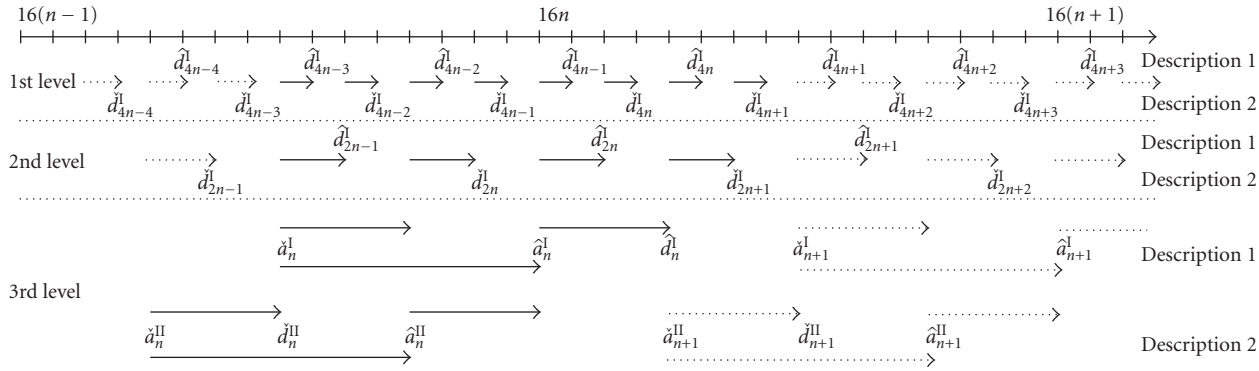

Figure 5: 4B_2MV scheme over 3 levels (GOP size = 16). Motion-compensated temporal operations are represented by arrows (solid lines for the current GOP, dashed lines for the adjacent GOPs).

frames $\check{a}_n^{\mathrm{II}}$, $\check{d}_n^{\mathrm{II}}$, and $\hat{a}_n^{\mathrm{II}}$. By inverting (23a) and (23b), we obtain

$$\begin{aligned}\widetilde{x}_n^{(1)}(\mathbf{s}_i-\mathbf{v}_i) &= \frac{1}{\sqrt{2}}\left(\left[\check{a}_{n+1}^{\mathrm{II}}(\mathbf{s}_i-\mathbf{v}_i)\right]-\frac{1}{M}\sum_{\ell=1}^{M}\left[\check{d}_{n+1}^{\mathrm{II}}(\mathbf{s}_\ell)\right]\right),\\ \widetilde{x}_n^{(2)}(\mathbf{s}_i) &= \frac{1}{\sqrt{2}}\left(\left[\check{a}_{n+1}^{\mathrm{II}}(\mathbf{s}_i-\mathbf{v}_i)\right]+2\left[\check{d}_{n+1}^{\mathrm{II}}(\mathbf{s}_i)\right]\right.\\ &\qquad\left.-\frac{1}{M}\sum_{\ell=1}^{M}\left[\check{d}_{n+1}^{\mathrm{II}}(\mathbf{s}_\ell)\right]\right).\end{aligned} \tag{28}$$

For the nonconnected pixels, we have

$$\begin{aligned}\widetilde{x}_n^{(0)}(s_i) &= \frac{1}{\sqrt{2}}\left[\hat{a}_n^{\mathrm{I}}(s_i)\right],\\ \widetilde{x}_n^{(1)}(s_i) &= \frac{1}{\sqrt{2}}\left[\check{a}_{n+1}^{\mathrm{II}}(s_i)\right].\end{aligned} \tag{29}$$

As it can be seen, $x_n^{(1)}$ can be recovered from both descriptions, and the final central reconstruction is obtained as the mean of these values. Also, one can note that by knowing $x_{n-1}^{(2)}$ (resp., $x_n^{(0)}$) from the first (resp., second) description, it is possible to reconstruct $x_{n-1}^{(3)}$, by reverting to (24) and (25).

As for the side decoders of the initial scheme, the solution for the first description is given by (27) and

$$\widetilde{x}_n^{(2)}(\mathbf{s}) = \widetilde{x}_n^{(3)}(\mathbf{s}) = \frac{1}{\sqrt{2}}\left[\check{a}_{n+1}^{\mathrm{I}}(\mathbf{s})\right], \tag{30}$$

while for the second description it reads

$$\widetilde{x}_{n+1}^{(0)}(\mathbf{s}) = \widetilde{x}_n^{(3)}(\mathbf{s}) = \frac{1}{\sqrt{2}}\left[\check{a}_{n+1}^{\mathrm{II}}(\mathbf{s})\right], \tag{31}$$

in addition to $\widetilde{x}_n^{(1)}$ and $\widetilde{x}_n^{(2)}$ obtained with (28).

For the 4B_1MV scheme, the additional motion compensation involved in the computation of the approximation sequences requires reverting the motion vector field in one of the components. Thus, we have

$$\begin{aligned}\widetilde{x}_{n-1}^{(2)}(\mathbf{s}) &= \frac{\left[\check{a}_n^{\mathrm{I}}(\mathbf{s})\right]}{\sqrt{2}},\\ \widetilde{x}_{n-1}^{(3)}(\mathbf{s}) &= \frac{\left[\check{a}_n^{\mathrm{I}}(\mathbf{s}-\mathbf{u})\right]}{\sqrt{2}}\end{aligned} \tag{32}$$

for the first side decoder and

$$\begin{aligned}\widetilde{x}_{n-1}^{(3)}(\mathbf{s}) &= \frac{\left[\hat{a}_n^{\mathrm{II}}(\mathbf{s})\right]}{\sqrt{2}},\\ \widetilde{x}_n^{(0)}(\mathbf{s}) &= \frac{\left[\hat{a}_n^{\mathrm{II}}(\mathbf{s}-\mathbf{u})\right]}{\sqrt{2}}\end{aligned} \tag{33}$$

for the second one.

For the scheme 4B_2MV, the temporal Haar transform being revertible, no additional difficulties appear for the central or side decoders.

Note that the reconstruction by one central and two side decoders corresponds to a specific application scenario, in which the user receives the two descriptions from two different locations (e.g., two WiFi access points), but depending on its position, it can receive both or only one of the descriptions. In a more general scenario, the user may be in the reception zone of both access points, but packets may be lost from both descriptions (due to network congestion, transmission quality, etc.). In this case, the central decoder will try to reconstruct the sequence by exploiting the information in all the received packets. It is therefore clear that an important issue for the reconstruction quality will be the packetization strategy. Even though the complete description of the different situations which can appear in the decoding (depending on the type of the lost packets) cannot be done here, it is worth noting that in a number of cases, an efficient usage of the received information can be employed: for instance, even if we do not receive the spatiotemporal subbands of one of the descriptions, but only a packet containing its motion vectors, these vectors can be exploited in conjunction with the other description for improving the fluidity of the reconstructed video. We also take advantage of the redundancy existing at the last level to choose, for the frames which can be decoded from both descriptions, the version which has the best quality, and thus to limit the degradations appearing in one of the descriptions.

7. SIMULATIONS RESULTS

The Haar lifting blocks in Figure 4 are implemented by a motion-compensated lifting decomposition [23]. The motion estimation is performed using hierarchical variable size block-matching (HVBSM) algorithm with block sizes ranging from 64×64 to 4×4. An integer-pel accuracy is used for motion compensation. The resulting temporal subbands are spatially decomposed with biorthogonal 9/7 Daubechies wavelets over 5 resolution levels. Spatiotemporal coefficients and motion vectors (MVs) are encoded within the MC-EZBC framework [26, 28], where MV fields are first represented as quad-tree maps and MV values are encoded with a zero-order arithmetic coder, in raster-scan order.

First, we have tested the proposed algorithm on several QCIF sequences at 30 fps. In Figure 6, we compare the rate-distortion performance of the nonrobust Haar scheme with that of the MDC central decoder on the "Foreman" video test sequence. The bitrate corresponds to the global rate for the robust codec (both descriptions). Three temporal decomposition levels have been used in this experiment ($J = 3$). We can observe that even the loss of one description still allows for acceptable quality reconstruction especially at low bitrates and also that the global redundancy does not exceed 30% of the bitrate.

Figure 7 illustrates the central rate-distortion curves for different levels of redundancy and, together with Figure 6, shows the narrowing of the gap with respect to the nonredundant version when the number of decomposition levels increases.

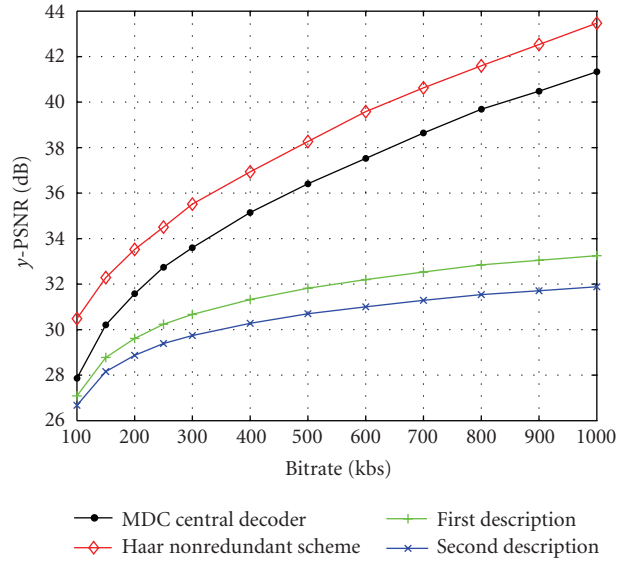

Figure 6: Central and side rate-distortion curves of the MDC scheme compared with the nonrobust Haar codec ("Foreman" QCIF sequence, 30 fps).

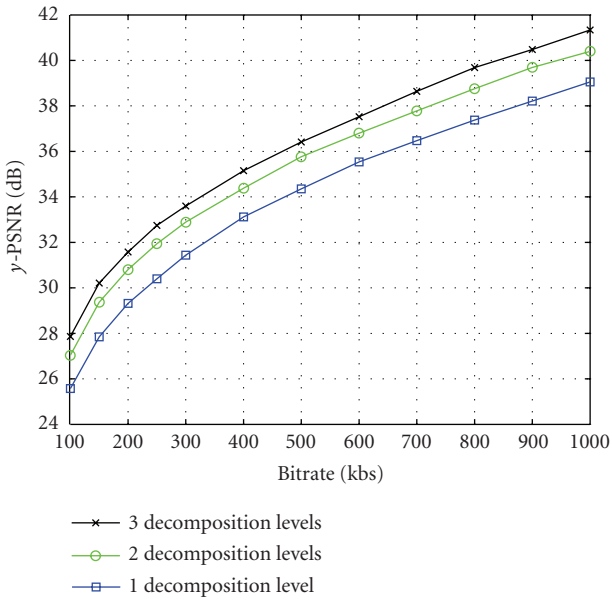

Figure 7: Rate-distortion curves at the central decoder for several levels of redundancy.

The difference in performance between the two descriptions is a phenomenon appearing only if the scheme involves three or more decomposition levels, since it is related to an asymmetry in the GOF structure of the two descriptions when performing the decimation. Indeed, as illustrated in Figure 5, when the first description is lost, some of the motion information in the second description cannot be used

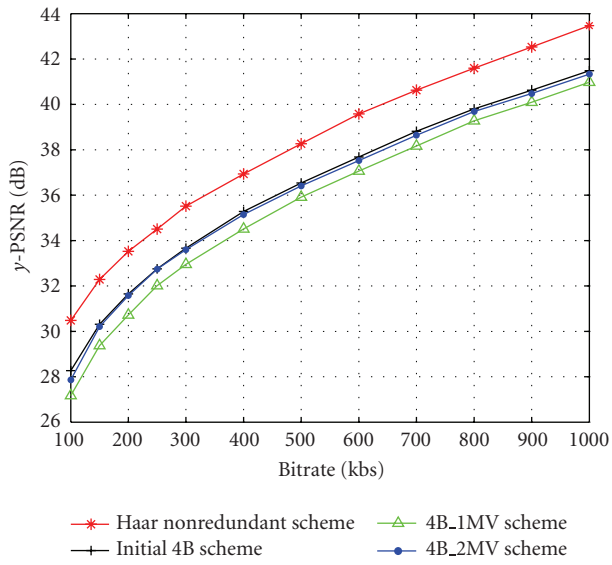

FIGURE 8: Rate-distortion curves for different reconstruction strategies, central decoder ("Foreman" QCIF sequence, 30 fps).

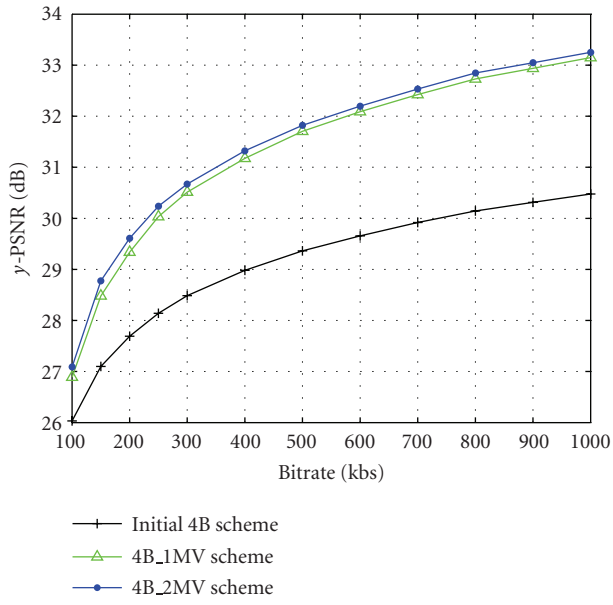

FIGURE 9: Rate-distortion curves for different reconstruction strategies, first side decoder ("Foreman" QCIF sequence, 30 fps).

to improve the reconstruction, while this does not happen when loosing the second description.

In Figures 8-9, we present the rate-distortion curves for the central and side decoders, in the absence of packet losses. The performance of the scheme without ME/MC in the computation of the approximation sequences $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ is compared with the 4B_1MV and 4B_2MV schemes.

One can note that the addition of the ME/MC step in the computation of $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ does not lead to an increase in the coding performance of the central decoder, since the expected gain is balanced by the need to encode an additional MV field. On the other hand, the final MC-Haar transform leads to much better results, since instead of two correlated approximation sequences, we now only have transformed subbands. For the side decoders however, the introduction of the motion-compensated average in the computation of $\check{a}_n^{\mathrm{I}}$ and $\hat{a}_n^{\mathrm{II}}$ leads to a significant improvement in coding performances (increasing with the bitrate from 1 to 2.5 dB), and the MC-Haar transform adds another 0.3 dB of improvement.

In a second scenario, we have tested our scheme for transmission over a packet loss network, like Ethernet. In this case, the bitstreams of the two descriptions are separated in packets of maximal size of 1500 bytes. For each GOP, separate packets are created for the motion vectors and for each spatiotemporal subband. If the packet with motion vectors is lost, or if the packet with the spatial approximation subband of the temporal approximation subband is lost, then we consider that the entire GOP is lost (it cannot be reconstructed).

We compare our scheme with a nonredundant MCTF one and also with another temporal MDC scheme, consisting in a temporal splitting of the initial video sequence. Odd and even frames are separated into two descriptions which are encoded with a Haar MCTF coder (Figure 10 illustrates the motion vectors and temporal transforms for this structure).

The coding performance as a function of the packet loss rate is illustrated in Figures 11 and 12 for the "Foreman" and "Mobile" video test sequences at 250 kbs. As expected, when there is no loss, the nonredundant coding is better than both MDC schemes (which have comparable performances). However, as soon as the packet loss rate gets higher than 2%, our scheme overpasses by 0.5–1 dB the temporal splitting and the nonrobust coding by up to 4 dB.

Moreover, we have noticed that the MDC splitting scheme exhibits a flickering effect, due to the fact that a lost packet will degrade the quality of one over two frames. In our scheme, this effect is not present, since the errors in one description have limited influence thanks to the existing redundancies, and also to a different propagation during the reconstruction process.

Figure 13 presents the influence of the average update operator, with gains of about 0.2 dB over the entire range of packet loss rates. Finally, we have compared in Figure 14 the rate-distortion curves of the temporal splitting and the proposed MDC schemes for a fixed packet loss rate (10%). One can note a difference of 0.5–1.3 dB at medium and high bitrates (150–1000 kbs) and slightly smaller at low bitrates (100 kbs). It is noticeable that the PSNR of the reconstructed sequence is not monotonically increasing with the bitrate: a stiff increase in PSNR until 250 kbs is followed by a "plateau" effect which appears at higher bitrates. This is due to the loss of the information in the spatial approximation of the temporal approximation subband. Indeed, for low bitrates, this spatiotemporal subband can be encoded into a single packet, so for uniform error distribution, the rate-distortion curve increases monotonically. At a given threshold (here, it happens at about 250 kbs for packets of 1500 bytes), the

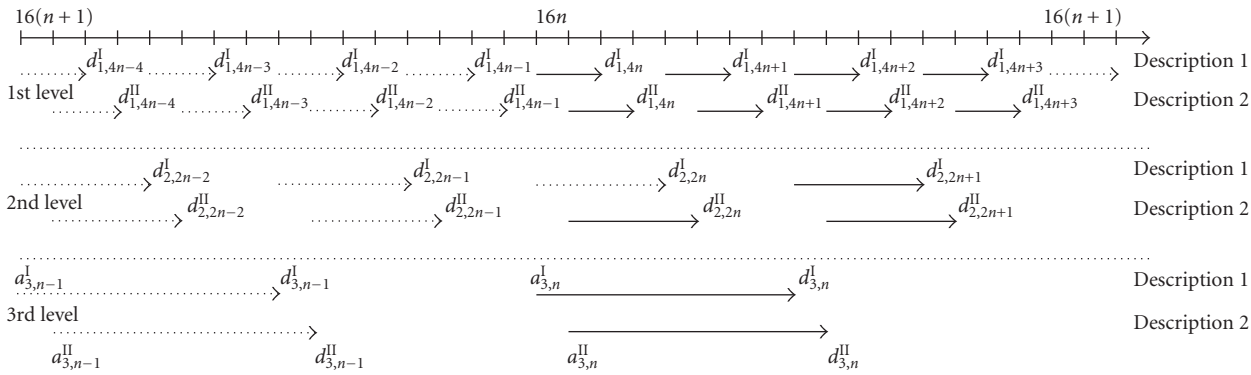

Figure 10: Three levels of decomposition in the temporal splitting scheme.

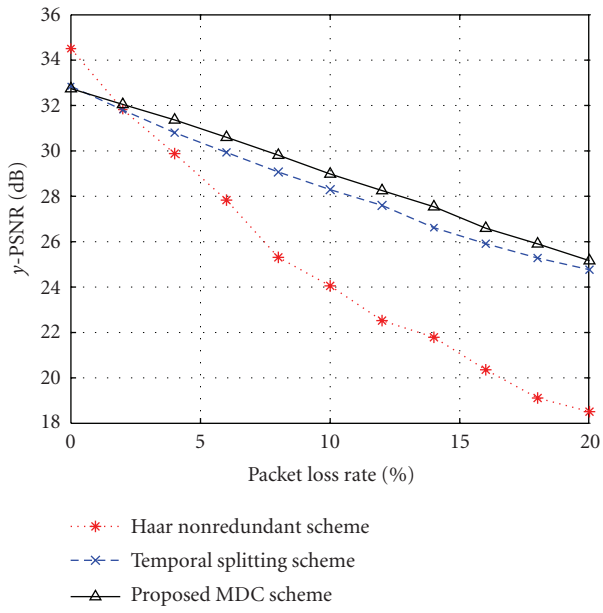

Figure 11: Distortion versus packet loss rate ("Foreman" QCIF sequence, 30 fps, 250 kbs).

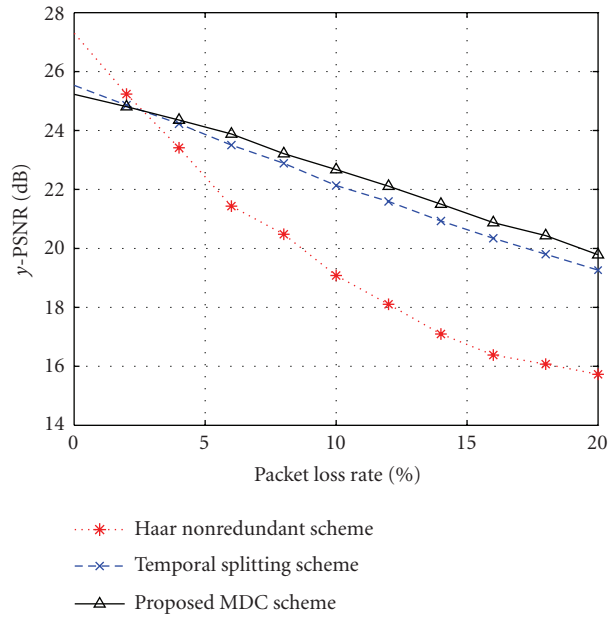

Figure 12: Distortion versus packet loss rate ("Mobile" QCIF sequence, 30 fps, 250 kbs).

approximation subband has to be coded into two packets. Moreover, we considered that if any of these two packets is lost, the GOF cannot be reconstructed. Therefore, we see a drop in performance. From this point, with the increasing bitrate, the performance improves till a new threshold where the subband needs to be encoded into three packets and so on. A better concealment scheme in the spatial domain, allowing to exploit even a partial information from this subband, would lead to a monotonic increase in performance.

8. CONCLUSION AND FUTURE WORK

In this paper, we have presented a new multiple-description scalable video coding scheme based on a motion-compensated redundant temporal analysis related to Haar wavelets.

The redundancy of the scheme can be reduced by increasing the number of temporal decomposition levels. Reversely, it can be increased either by reducing the number of temporal decomposition levels, or by using nondecimated versions of some of the detail coefficients. By taking advantage of the Haar filter bank structure, we have provided an equivalent four-band lifting implementation, providing more insight into the invertibility properties of the scheme. This allowed us to develop simple central and side-decoder structures which have been implemented in the robust video codec.

The performances of the proposed MDC scheme have been tested in two scenarios: on-off channels and packet losses, and have been compared with an existing temporal splitting solution.

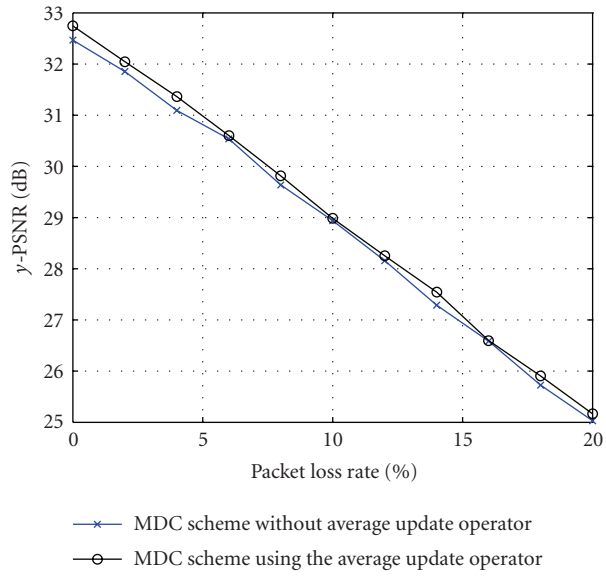

Figure 13: Influence of average update operator on the performance ("Foreman" QCIF sequence, 30 fps).

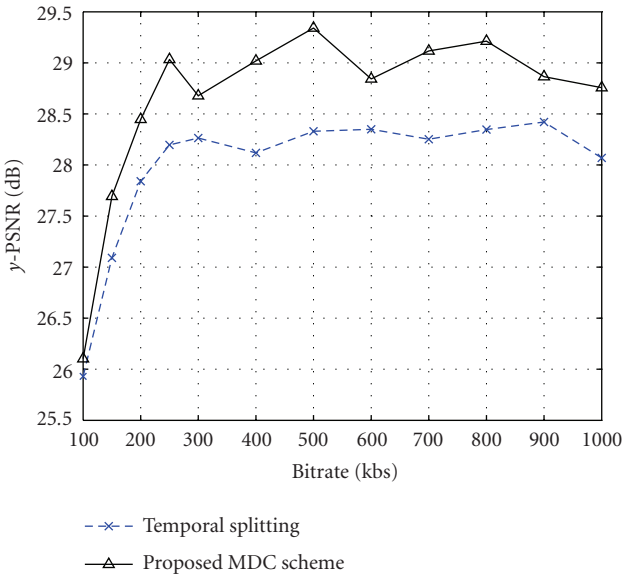

Figure 14: Rate-distortion curves at 10% packet loss rate ("Foreman" QCIF sequence, 30 fps, 10% packet losses).

Note that the presented scheme builds the descriptions in the temporal domain of the video, but it can be combined with structures introducing the redundancy in the spatial domain, for which many more solutions have been proposed in the literature. The increased flexibility thus achieved may be exploited to better adapt the packetization to different situations of network losses and also to improve the reconstruction at different levels.

ACKNOWLEDGMENT

Part of this work was funded by the ANR under the Grant ANR-05-RNRT-019 (DIVINE Project).

REFERENCES

[1] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, 2001.

[2] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Visual Communications and Image Processing*, vol. 4310 of *Proceedings of SPIE*, pp. 392–409, San Jose, Calif, USA, January 2001.

[3] L. Ozarow, "On a source-coding problem with two channels and three receivers," *The Bell System Technical Journal*, vol. 59, no. 10, pp. 1909–1921, 1980.

[4] A. E. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, 1982.

[5] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, 2003.

[6] V. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[7] Y. Wang, M. T. Orchard, V. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Transactions on Image Processing*, vol. 10, no. 3, pp. 351–366, 2001.

[8] J. Kovačević, P. L. Dragotti, and V. K. Goyal, "Filter bank frame expansions with erasures," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1439–1450, 2002.

[9] T. Petrişor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet, "Comparison of redundant wavelet schemes for multiple description coding of video sequences," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 5, pp. 913–916, Philadelphia, Pa, USA, March 2005.

[10] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, "A robust codec for transmission of very low bit-rate video over channels with bursty errors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1403–1412, 2000.

[11] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple-description video coding using motion-compensated temporal prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 193–204, 2002.

[12] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1211–1225, 2003.

[13] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 321–334, 2005.

[14] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.

[15] V. A. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proceedings of IEEE International Conference on Image Processing (ICIP '99)*, vol. 3, pp. 812–816, Kobe, Japan, October 1999.

[16] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 438–452, 2002.

[17] M. van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 3, pp. 489–492, Barcelona, Spain, September 2003.

[18] C. Tillier, B. Pesquet-Popescu, and M. van der Schaar, "Multiple descriptions scalable video coding," in *Proceedings of 12th European Signal Processing Conference (EUSIPCO '04)*, Vienna, Austria, September 2004.

[19] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Network-adaptive video streaming using multiple description coding and path diversity," in *Proceedings of International Conference on Multimedia and Expo (ICME '03)*, vol. 2, pp. 653–656, Baltimore, Md, USA, July 2003.

[20] S. Cho and W. A. Pearlman, "Error resilient compression and transmission of scalable video," in *Applications of Digital Image Procedding XXIII*, vol. 4115 of *Proceedings of SPIE*, pp. 396–405, San Diego, Calif, USA, July-August 2000.

[21] N. Franchi, M. Fumagalli, G. Gatti, and R. Lancini, "A novel error-resilience scheme for a 3-D multiple description video coder," in *Proceedings of Picture Coding Symposium (PSC '04)*, pp. 373–376, San Francisco, Calif, USA, December 2004.

[22] T. Petrişor, C. Tillier, B. Pesquet-Popescu, and J.-C. Pesquet, "Redundant multiresolution analysis for multiple description video coding," in *Proceedings of IEEE 6th Workshop on Multimedia Signal Processing*, pp. 95–98, Siena, Italy, September-October 2004.

[23] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01)*, vol. 3, pp. 1793–1796, Salt Lake, Utah, USA, May 2001.

[24] C. Tillier and B. Pesquet-Popescu, "3D, 3-band, 3-tap temporal lifting for scalable video coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 779–782, Barcelona, Spain, September 2003.

[25] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 577–600, 2004.

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

[27] C. Tillier, B. Pesquet-Popescu, and M. van der Schaar, "Improved update operators for lifting-based motion-compensated temporal filtering," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 146–149, 2005.

[28] "3D MC-EZBC Software package," http://www.cipr.rpi.edu/ftp_pub/personal/chen/MC_EZBC.zip.