3-Band Motion-Compensated Temporal Structures for Scalable Video Coding

Christophe Tillier, Béatrice Pesquet-Popescu, *Member, IEEE*, and Mihaela van der Schaar, *Member, IEEE*

***Abstract*—Recent breakthroughs in motion-compensated temporal wavelet filtering have finally enabled implementation of highly efficient scalable and error-resilient video codecs. These new wavelet codecs provide numerous advantages over nonscalable conventional solutions techniques based on motion-compensated prediction, such as no recursive predictive loop, separation of noise and sampling artifacts from the content through use of longer temporal filters, removal of long range as well as short range temporal redundancies, etc. Moreover, these wavelet video coding schemes can provide flexible spatial, temporal, signal-to-noise ratio and complexity scalability with fine granularity over a large range of bit rates, while maintaining a very high coding efficiency. However, most motion-compensated wavelet video schemes are based on classical two-band decompositions that offer only dyadic factors of temporal scalability. In this paper, we propose a three-band temporal structure that extends the concept of motion-compensated temporal filtering (MCTF) that was introduced in the classical lifting framework. These newly introduced structures provide higher temporal scalability flexibility, as well as improved compression performance compared with dyadic Haar MCTF.**

***Index Terms*—Three-band (3-band) bidirectional temporal filtering, 3-band temporal decomposition, motion-compensated temporal filtering (MCTF), nonlinear lifting, scalable video compression, video coding.**

I. Introduction

THE spatio-temporal ("2D+t" or "3D") subband coding (3D-SBC) schemes constitute an alternative approach to hybrid coding concepts usually used in today's video standards. 3D-SBC schemes are based on the fact that a subband decomposition applied along the temporal axis of a video sequence leads to an efficient energy concentration on low-pass temporal subbands. Subsequently, the temporal subbands are coded using conventional subband coding techniques that provide spatial as well as SNR (embedded) scalabilities [1]–[3]. Consequently, the 3D-SBC schemes provide temporal/spatial/SNR scalability, achieved naturally from the subband structure, thereby enabling efficient multimedia transmission over heterogeneous networks to diverse devices [4].

The 3D-SBC schemes are based on open-loop motion-compensated (MC) temporal multiresolution decompositions [5]–[7]. Lifting implementation of these decompositions [8], [9] represents a powerful tool for spatio-temporal optimizations and numerous improvements based on this idea have been recently proposed, concerning for example weighted spatial filtering in occluded areas [10] or the increase of motion estimation accuracy [11]. Also, while initially 3D-SBC were implemented using Haar temporal filters, the proposed lifting implementations enabled the use of longer temporal filters (e.g., 5/3 filters) in the temporal direction, in order to exploit long-term temporal redundancies present within the sequence [9], [12]–[15]. Longer temporal filters (i.e., the 9/7 filters) using motion threading have also been proposed in the 3D-ESCOT algorithm [16]. Unconstrained prediction using multiple reference frames is another interesting technique that has been proposed in [17], but in this case the scheme does not involve an update step.

To allow for easy adaptation to different transmission bit-rates and device characteristics, flexible spatio/temporal/SNR scalabilities should be possible. In particular, temporal downsampling by a factor different than 2, e.g., 3, should be possible since this enables adaptation to a wider range of network conditions and user demands. For example, if a client asks for sequences at 30 and 10 fps, this can be satisfied with the proposed three-band (3-band) temporal structures. In closed-loop predictive schemes, this feature is achieved by simply taking one out of, e.g., three frames of the original sequence and encoding them in the base layer. Then, the temporal enhancement layer is obtained by predicting the remaining frames from the base layer. The dyadic structure of the wavelet transform used in 3D-SBC coders [5]–[7], [18] enables only temporal downsampling factors of 2 and does not allow for temporal scalability by a factor 3. However, general *nonlinear M*-band lifting schemes with perfect reconstruction have been introduced in [19] and lifting implementations of *M*-channel filterbanks designed [20], [21]. These works do not consider the application of such structures for performing the temporal transform in the framework of motion compensated subband coding. In this case, the introduction of the motion estimation/compensation in the structure needs a particular attention. In [22], [23], [13], and [24], 3-band motion-compensated temporal decomposition structures that allow nondyadic scalability factors for video sequences have been proposed.

In this paper, we extend the simple operators introduced in [22], [25] to more flexible long-term motion-compensated temporal filters and analyse the invertibility of this structure. If our first scheme can be considered as a 3-band equivalent of the Haar MC wavelet decomposition, the new scheme introduced in this paper is the 3-band equivalent of the 5/3 MC temporal filterbank [26]. It involves bidirectional MC in both the predict and

Manuscript received March 4, 2004; revised November 11, 2005. This work was presented in part at the IEEE ICIP 2003 and the IEEE ICASSP 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Trac D. Tran.

C. Tillier and B. Pesquet-Popescu are with the Signal and Image Processing Departement, Ecole Nationale Supérieure des Télécommunications and CNRS UMR 5141, Paris, France (e-mail: tillier@tsi.enst.fr; pesquet@tsi.enst.fr).

M. van der Schaar is with the Department of Electrical Engineering, University of California at Los Angeles (UCLA), Los Angeles, CA 90095-1594 USA (e-mail: mihaela@ee.ucla.edu).

Digital Object Identifier 10.1109/TIP.2006.877411

update operators. However, while the simple Haar-like 3-band scheme still belongs to the classical lifting framework, where the inversion is straightforwardly guaranteed by the structure, the proposed scheme does no longer appertain to this framework. We perform, therefore, a theoretical analysis of the invertibility of this 3-band structure and determine conditions for perfect reconstruction.

By combining this structure with dyadic decompositions at various temporal levels, it becomes possible to achieve more flexible temporal downsampling of video, with e.g., factors of 3 *and* 2. Finally, we show that classical hybrid coding solutions become a particular case of this framework. Simulation results show improved coding efficiency compared with dyadic multiresolution analysis (MRA), as well as with the 3-band Haar-like scheme and MPEG-4 hybrid codec.

The paper is organized as follows. In the next section, we briefly review the generic 3-band MCT lifting scheme, and in Sections II-A and II-B, we discuss two particular cases of this framework. In Section III, we introduce the new 3-band structure, analyse its invertibility and discuss optimal parameters. Section IV discusses the temporal scalability properties of the proposed schemes. In Section V, we present simulation results illustrating the coding performance and the scalability features of the proposed structure. Concluding remarks and hints for future research are given in Section VI.

II. Haar-Like Three-Band Scheme

First, let us introduce some notations: the frames in the sequence will be denoted by $(x_t(\boldsymbol{n}))$, where t is the temporal index and $\boldsymbol{n}$ is a spatial variable that takes values in $\{1,\ldots,M\}\times\{1,\ldots,N\}$. In the temporal wavelet decomposition, we denote by h_t the detail ("high-frequency") subband frames and by l_t the approximation ("low-frequency") subband frames. Below we only describe one transform level, but extensions to multiresolution decompositions by subsequent decompositions of the approximation subband can be derived in a straightforward manner.

The 3-band decomposition scheme using uni-directional prediction operators is depicted in Fig. 1. The corresponding equations describing this analysis structure are

$$h_t^+ = x_{3t+1} - P^+[(x_{3t})_{t\in\mathbb{N}}], \quad t\in\mathbb{N} \tag{1}$$
$$h_t^- = x_{3t-1} - P^-[(x_{3t})_{t\in\mathbb{N}^*}], \quad t\in\mathbb{N}^* \tag{2}$$
$$l_t = x_{3t} + U^+[(h_t^+)_{t\in\mathbb{N}}] + U^-[(h_t^-)_{t\in\mathbb{N}}], \quad t\in\mathbb{N}. \tag{3}$$

Each detail subband is obtained by prediction from a single polyphase component of the input sequence. Therefore, the perfect reconstruction is achieved by the inversion property of the lifting scheme.

Note that we have two (forward and backward) prediction operators: P^+ and P^-, leading to two detail subbands, and two update operators for computing the approximation subband: U^+ and U^-. As in the case of two-band temporal decompositions, these operators will be nonlinear due to the temporal prediction that involves motion estimation and compensation (ME/MC).

Let us now analyze the benefits of this structure for two simple cases.

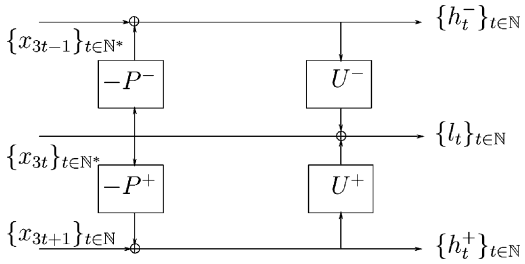

Fig. 1. Three-band temporal lifting scheme with uni-directional prediction operators.

A. No Motion Estimation/Compensation

First, consider the scheme with no ME/MC and the simplest predictors, i.e., identity operators. In this case, the generic equations reduce to

$$h_t^+ = x_{3t+1} - x_{3t}, \quad h_t^- = x_{3t-1} - x_{3t} \tag{4}$$
$$l_t = x_{3t} + U^+[h_t^+] + U^-[h_t^-]. \tag{5}$$

Very simple linear functions can be considered also for the update operators $U^+[h_t^+] = \alpha h_t^+$, $U^-[h_t^-] = \alpha h_t^-$, where the positive constant α can be determined such that l_t is obtained by low-pass filtering the input sequence (i.e., its frequency response should cancel at the normalized frequency 1/2). This leads to $\alpha = 1/4$ and to an approximation subband obtained by filtering three consecutive frames, as follows:

$$l_t = \frac{1}{2}x_{3t} + \frac{1}{4}x_{3t+1} + \frac{1}{4}x_{3t-1}. \tag{6}$$

The corresponding transfer function of the filter bank described by (4) and (6) are, respectively

$$H^+(z) = z-1,\ H^-(z) = z^{-1}-1,\ L(z) = \frac{1}{2} + \frac{1}{4}(z+z^{-1}).$$

These are the shortest operators that can be employed for building a 3-band structure and the scheme obtained in this manner can be seen as the 3-band equivalent of the Haar decomposition. It will be called in the sequel the "Haar-like 3-band scheme".

Moreover, nonlinear filters can be designed [27], even for such a simple scheme, leading to an improved quality of the updated frame. Such functions will behave as linear functions for small absolute value of the detail signals and will not use these prediction errors for updating when they are too large, with a smooth transition between the two states. This kind of adaptive behavior can also be implemented for motion-compensated operators (as discussed in the next section), by performing an adaptive update depending on the content characteristics (e.g., based on the MC temporal prediction error).

B. Nonlinear Spatio-Temporal Operators

If we insert ME/MC in the previous scheme, the two detail signals can be computed at the same spatial positions as the frames x_{3t-1} and x_{3t+1}, while the approximation frame can be determined at the same spatial position as frame x_{3t} (see Fig. 2). In this case, the frame x_{3t} is used as reference for both forward and backward ME, and we obtain the detail frames h_t^+ and h_t^-

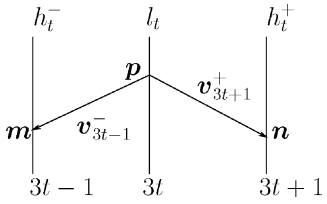

Fig. 2. Temporal filtering in the Haar-like 3-band scheme.

synchronous with x_{3t+1} and x_{3t-1}, respectively, while the approximation frame l_t will be synchronous with x_{3t}. Note that different choices can be made for the temporal synchronisation between the original frames and the resulting subband frames (for example, l_t synchronous with x_{3t-1}, h_t^- synchronous with x_{3t} and h_t^+ synchronous with x_{3t+1}). We have preferred the current scheme for symmetry reasons (the temporal "distance" involved in the two predict operators is the same).

Using the notations in Fig. 2, we denote by $\boldsymbol{v}_t^+$ the forward motion vector and by $\boldsymbol{v}_t^-$ the backward motion vector used to predict frame t. Following [5], a pixel $\boldsymbol{p}$ in frame $3t$ will be called simple, respectively, multiple connected forward/backward, if it is used for the prediction of one, respectively, several pixels in the frames $\{3t+1, 3t-1\}$ (for example, in Fig. 2, the pixels $\boldsymbol{m}$ in frame $3t-1$ and $\boldsymbol{n}$ in frame $3t+1$ are predicted using $\boldsymbol{p}$). A pixel in the frames $3t$ will be called unconnected forward/backward if it is not used for the prediction of one pixel in the frames $\{3t+1, 3t-1\}$.

The equations describing the temporal filtering become (the index corresponds to the temporal coordinate, while the function argument is used to denote the spatial position)

$$\begin{aligned} h_t^+(\boldsymbol{n}) &= x_{3t+1}(\boldsymbol{n}) - x_{3t}(\boldsymbol{n}-\boldsymbol{v}_{3t+1}^+) = x_{3t+1}(\boldsymbol{n}) - x_{3t}(\boldsymbol{p}) \\ h_t^-(\boldsymbol{m}) &= x_{3t-1}(\boldsymbol{m}) - x_{3t}(\boldsymbol{m}-\boldsymbol{v}_{3t-1}^-) = x_{3t-1}(\boldsymbol{m}) - x_{3t}(\boldsymbol{p}) \\ l_t(\boldsymbol{p}) &= x_{3t}(\boldsymbol{p}) + \frac{1}{4}\left[h_t^+(\boldsymbol{p}+\boldsymbol{v}_{3t+1}^+) + h_t^-(\boldsymbol{p}+\boldsymbol{v}_{3t-1}^-)\right] \\ &= x_{3t}(\boldsymbol{p}) + \frac{1}{4}\left[h_t^+(\boldsymbol{n}) + h_t^-(\boldsymbol{m})\right]. \end{aligned} \quad (7)$$

The advantage of choosing such a correspondence between the detail and approximation frames is that the unconnected or multiple connected pixels resulting from MCTF are located in the same approximation frame. Hence, the low-pass filtering optimization for multiple connections using techniques as in [8] is much simplified. Indeed, the unconnected pixels are processed in a similar manner as for the dyadic decomposition: if a pixel $\boldsymbol{p}$ in frame $3t$ is backward connected with a pixel $\boldsymbol{m}$ of frame $3t-1$ and it has no connections to frame $3t+1$, the update equation (7) reads

$$l_t(\boldsymbol{p}) = x_{3t}(\boldsymbol{p}) + \frac{1}{2}h_t^-(\boldsymbol{m}).$$

Similarly, for a pixel $\boldsymbol{p}$ in the frame $3t$ that is forward connected with $\boldsymbol{n}$ in frame $3t+1$ and with no backward connections, the low-pass filtering is performed as follows:

$$l_t(\boldsymbol{p}) = x_{3t}(\boldsymbol{p}) + \frac{1}{2}h_t^+(\boldsymbol{n}).$$

If the pixel $\boldsymbol{p}$ is unconnected, its value is retained after low-pass filtering, i.e., $l_t = x_{3t}(\boldsymbol{p})$.

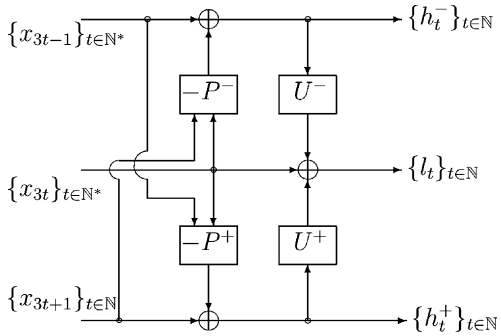

Fig. 3. Three-band decomposition structure with bidirectional predict operators.

C. Motion Vector Redundancy

Due to the simplicity of the predictors, this scheme can be seen as the 3-band equivalent of the Haar MRA. However, the update operator is bidirectional, and compared to a dyadic Haar temporal decomposition or to a hybrid coding scheme, the new structure requires a larger number of motion vector fields to be estimated and encoded at a given level. The "redundancy factor" is in this case 4/3 (meaning that the number of displacement fields is increased by a third with regard to (w.r.t.) the Haar case) at a given decomposition level. For J decomposition levels and a group of frames (GOF) size of N frames, a Haar MRA requires $\sum_{j=1}^{J} N/2^j = N(1-1/2^J)$ motion vector fields, while the 3-band scheme involves $\sum_{j=1}^{J} 2N/3^j = N(1-1/3^J)$ motion field estimations. In particular, if $N = 2^J$, the Haar MRA will lead to $N-1$ motion fields, and if $N = 3^J$ the 3-band scheme will also need $N-1$ motion fields. Therefore, after the entire temporal decomposition, the ME complexity does not exceed that of hybrid coding or Haar MRA.

Moreover, as we have shown in [24] and [28], one can take advantage of the redundancy between motion vectors at different temporal levels to reduce the computational complexity. For example, for two-band (2-band) structures, by turning off the motion estimation for two over four temporal levels, savings up to a factor of 16 in complexity have been achieved, with a loss in PSNR of less than 1 dB. Moreover, different tradeoffs between complexity and quality are also possible.

III. Bidirectional Predict-Update 3-Band Schemes

A. Proposed Structure

The quality of the detail subbands depends on the performance of the predictor involved in the lifting step. Until now, we have predicted the current frame by motion compensating the previous (or future) frame. However, it is well known from hybrid coding that bidirectional prediction is very useful, especially in areas where occlusions occur. Hence, we use bidirectional predictors for the two detail subbands. The proposed 3-band decomposition scheme with bidirectional predict operators is presented in Fig. 3.

The corresponding equations describing this analysis structure are

$$h_t^+ = x_{3t+1} - P^+[(x_{3t})_{t\in\mathbb{N}}, (x_{3t-1})_{t\in\mathbb{N}^*}], \quad t\in\mathbb{N} \quad (8)$$

$$h_t^- = x_{3t-1} - P^-[(x_{3t})_{t\in\mathbb{N}^*}, (x_{3t+1})_{t\in\mathbb{N}}], \quad t\in\mathbb{N}^* \quad (9)$$

$$l_t = x_{3t} + U^+[(h_t^+)_{t\in\mathbb{N}}] + U^-[(h_t^-)_{t\in\mathbb{N}}], \quad t\in\mathbb{N}. \quad (10)$$

Note that in this scheme, the two predict operators involve frames from two polyphase components. At the decoder side, inverting the update step is straightforward, as for the classical lifting case [29] by simply reverting the scheme and changing the sign of the operator, but the inversion of the predict step is not guaranteed by the structure, especially since the two predict operators can involve various time delays. Indeed, x_{3t} can be computed by inverting the update step, while x_{3t+1} requires x_{3t}, h_t^+ and x_{3t-1}. However, the last frame is not available and it requires for its computation x_{3t}, h_t^- and x_{3t+1}. The perfect reconstruction of the scheme is therefore not obvious and it will be discused in Section III-D also taking into account motion compensation.

B. Bidirectional 3-Band Structure Without ME/MC

For simplicity, we constrain in the sequel the prediction to be performed only from one future and one previous frame (which corresponds to the B-frames case in hybrid coding). Without considering ME/MC, the detail subbands are obtained by the following equations:

$$h_t^+ = x_{3t+1} - \beta\, x_{3t+2} - (1-\beta)\, x_{3t} \tag{11}$$
$$h_t^- = x_{3t-1} - \beta\, x_{3t-2} - (1-\beta)\, x_{3t} \tag{12}$$

where $\beta \in [0,1]$ is a weighting factor. Note that (11) and (12) obviously appear as specific cases of (8)–(10). The case $\beta = 0$ corresponds to the situation analysed in Section II-A (Haar-like 3-band scheme). Relations (11) and (12) correspond to filtering operations followed by decimation by a factor 3. Let $H^+(z)$ respectively, $H^-(z)$ denote the z-transform of the previous two filters, we can remark that for any $\beta \in [0,1]$, we have $H^+(1) = H^-(1) = 0$, meaning that we have indeed two high-pass filters. The parameter β can be tuned to take into account irregularities along motion trajectories, but the two detail subbands remain symmetric w.r.t. the central frame.

In order to obtain the approximation subband, the update filter can be the same as the one used in the case of simpler predictors

$$l_t = x_{3t} + \alpha h_t^+ + \alpha h_t^-, \quad \alpha \in \mathbb{R}. \tag{13}$$

Again, the positive constant α can be determined so that l_t results from a low-pass filtering from the input sequence. This leads to $\alpha = 1/4$, independently of the value of β.

C. Bidirectional 3-Band Structure With ME/MC

If we want to include the ME/MC in the two predictors, we need to consider forward/backward motion vectors, as illustrated in Fig. 4.

In this case, the analysis relations become

$$\begin{aligned} h_t^+(\boldsymbol{n}) &= x_{3t+1}(\boldsymbol{n}) - \beta\, x_{3t+2}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-) \\ &\quad - (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+) \end{aligned} \tag{14}$$
$$\begin{aligned} h_t^-(\boldsymbol{m}) &= x_{3t-1}(\boldsymbol{m}) - \beta\, x_{3t-2}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^+) \\ &\quad - (1-\beta)\, x_{3t}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^-). \end{aligned} \tag{15}$$

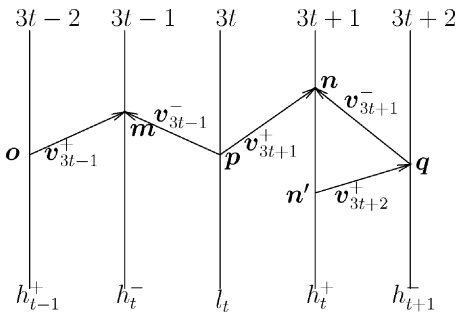

Fig. 4. Temporal prediction in a group of frames using the bidirectional predict-update operators.

By taking into account the MC in (13), the low-pass temporal subband can be expressed, for all pixels $\boldsymbol{p}$ bidirectionally connected, as

$$l_t(\boldsymbol{p}) = x_{3t}(\boldsymbol{p}) + \alpha h_t^+(\boldsymbol{p} + \boldsymbol{v}_{3t+1}^+) + \alpha h_t^-(\boldsymbol{p} + \boldsymbol{v}_{3t-1}^-). \tag{16}$$

Using the expressions (14) and (15) of the details in (16), we get a five-tap low-pass filter

$$\begin{aligned} l_t(\boldsymbol{p}) &= \big[1 - 2\alpha(1-\beta)\big] x_{3t}(\boldsymbol{p}) + \alpha x_{3t+1}(\boldsymbol{p} + \boldsymbol{v}_{3t+1}^+) \\ &\quad + \alpha x_{3t-1}(\boldsymbol{p} + \boldsymbol{v}_{3t-1}^-) - \alpha\beta x_{3t+2}(\boldsymbol{p} + \boldsymbol{v}_{3t+1}^+ - \boldsymbol{v}_{3t+1}^-) \\ &\quad - \alpha\beta x_{3t-2}(\boldsymbol{p} + \boldsymbol{v}_{3t-1}^- - \boldsymbol{v}_{3t-1}^+). \end{aligned} \tag{17}$$

Remark that, due to the MC motion, all the positions to be filtered in the five successive frames are aligned along the same motion trajectory, resulting in a meaningful temporal filtering. Note, however, that this connectivity is limited in time, and no motion threads are explicitely extracted, as in [16].

Note that, as for 2-band schemes or 3-band Haar-like schemes, some pixels in the frame synchronous with the approximation subband (here, $3t$) can be unconnected (not used for prediction) or multiple connected (used to predict two or more pixels). In our case, we have to consider the connections with both detail frames. Thus, a pixel $\boldsymbol{p}$ in the frame $3t$ can be for example multiple connected with pixels in frame $3t+1$ and unconnected with the frame $3t-1$. The update equation will be different according to the different cases. For multiple connections on one or both sides, more complicated strategies can be applied for a spatio-temporal filtering, as for MC 5/3 2-band filterbanks [26].

D. Invertibility of the Proposed Scheme

As the proposed analysis structure does not stem from a classical lifting scheme, perfect reconstruction needs to be proved. An additional difficulty to establish the invertibility of the decomposition comes from the nonlinearity of the operators, introduced by motion compensation.

In order to derive the invertibility of the scheme, we introduce the following notations:

$$\begin{aligned} \widetilde{h}_t^+(\boldsymbol{n}) &= h_t^+(\boldsymbol{n}) + (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+) \\ &= x_{3t+1}(\boldsymbol{n}) - \beta\, x_{3t+2}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-) \end{aligned} \tag{18}$$
$$\begin{aligned} \widetilde{h}_t^-(\boldsymbol{n}) &= h_t^-(\boldsymbol{n}) + (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t-1}^-) \\ &= x_{3t-1}(\boldsymbol{n}) - \beta\, x_{3t-2}(\boldsymbol{n} - \boldsymbol{v}_{3t-1}^+) \end{aligned} \tag{19}$$

and also the so-called "polyphase components"

$$\begin{cases} x_t^1 &= x_{3t+1} \\ x_t^2 &= x_{3t+2} \end{cases} \Longrightarrow \begin{cases} x_{t-1}^1 &= x_{3t-2} \\ x_{t-1}^2 &= x_{3t-1} \end{cases}.$$

In order to simplify the analysis, we first consider the situation where $\boldsymbol{v}_{3t+1}^- = \boldsymbol{v}_{3t-1}^+ = 0$. Using this assumption, (18) and (19) become

$$\widetilde{h}_t^+(\boldsymbol{n}) = x_t^1(\boldsymbol{n}) - \beta x_t^2(\boldsymbol{n}),\ \widetilde{h}_t^-(\boldsymbol{n}) = x_{t-1}^2(\boldsymbol{n}) - \beta x_{t-1}^1(\boldsymbol{n}).$$

In each of the previous relations, the spatial positions are aligned. It is possible therefore to apply a temporal z-transform along this trajectory, as follows:

$$\begin{bmatrix} \widetilde{H}_z^+(\boldsymbol{n}) \\ \widetilde{H}_z^-(\boldsymbol{n}) \end{bmatrix} = \begin{bmatrix} 1 & -\beta \\ -\beta z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} X_z^1(\boldsymbol{n}) \\ X_z^2(\boldsymbol{n}) \end{bmatrix}. \quad (20)$$

This system can be inverted when $\beta \neq \pm 1$ and yields

$$\begin{bmatrix} X_z^1(\boldsymbol{n}) \\ X_z^2(\boldsymbol{n}) \end{bmatrix} = \frac{1}{(1-\beta^2)z^{-1}} \begin{bmatrix} z^{-1} & \beta \\ \beta z^{-1} & 1 \end{bmatrix} \begin{bmatrix} \widetilde{H}_z^+(\boldsymbol{n}) \\ \widetilde{H}_z^-(\boldsymbol{n}) \end{bmatrix}. \quad (21)$$

Applying an inverse z-transform, we get the temporal domain relations

$$x_{3t+1}(\boldsymbol{n}) = \frac{1}{1-\beta^2}\left[\widetilde{h}_t^+(\boldsymbol{n}) + \beta \widetilde{h}_{t+1}^-(\boldsymbol{n})\right] \quad (22)$$

$$x_{3t+2}(\boldsymbol{n}) = \frac{1}{1-\beta^2}\left[\beta \widetilde{h}_t^+(\boldsymbol{n}) + \widetilde{h}_{t+1}^-(\boldsymbol{n})\right]. \quad (23)$$

The above formulas were obtained under the hypothesis that $\boldsymbol{v}_{3t+1}^- = \boldsymbol{v}_{3t-1}^+ = 0$, which allowed us to perform the inversion of the scheme. However, the inversion is also possible *without* making the assumption of null motion vector fields. Starting from (18) and (19), and considering the expression

$$\frac{1}{1-\beta^2}\left[\widetilde{h}_t^+(\boldsymbol{n}) + \beta \widetilde{h}_{t+1}^-(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-)\right] = \frac{1}{1-\beta^2}\left[x_{3t+1}(\boldsymbol{n}) - \beta^2 x_{3t+1}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^- - \boldsymbol{v}_{3t+2}^+)\right]$$

we remark that it is possible to obtain a reconstruction formula for x_{3t+1} from the above equation, if the forward and backward motion vector fields between two successive frames are identical, with opposite sense, i.e.,

$$\boldsymbol{v}_{3t+1}^- = -\boldsymbol{v}_{3t+2}^+. \quad (24)$$

Indeed, we have

$$x_{3t+1}(\boldsymbol{n}) = \frac{1}{1-\beta^2}\left[\widetilde{h}_t^+(\boldsymbol{n}) + \beta \widetilde{h}_{t+1}^-(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-)\right]. \quad (25)$$

A similar expression can be found in order to compute x_{3t+2}. Indeed, consider

$$\frac{1}{1-\beta^2}\left[\beta \widetilde{h}_t^+(\boldsymbol{n} + \boldsymbol{v}_{3t+1}^-) + \widetilde{h}_{t+1}^-(\boldsymbol{n})\right] = \Big\{\beta\big[x_{3t+1}(\boldsymbol{n} + \boldsymbol{v}_{3t+1}^-) - \beta x_{3t+2}(\boldsymbol{n})\big] + x_{3t+2}(\boldsymbol{n}) - \beta x_{3t+1}(\boldsymbol{n} - \boldsymbol{v}_{3t+2}^+)\Big\}.$$

Again, under the assumption $\boldsymbol{v}_{3t+1}^- = -\boldsymbol{v}_{3t+2}^+$, the above expression allows to invert the lifting scheme and to get $x_{3t+2}(\boldsymbol{n})$ from the detail subbands

$$x_{3t+2}(\boldsymbol{n}) = \frac{1}{1-\beta^2}\left[\beta \widetilde{h}_t^+(\boldsymbol{n} + \boldsymbol{v}_{3t+1}^-) + \widetilde{h}_{t+1}^-(\boldsymbol{n})\right]. \quad (26)$$

Note, however, that different update filters can be used to obtain the approximation subband. For instance, in [23], we derived a subset of this framework that uses only prediction from past frames (thus, the inversion can be performed within the lifting formalism) and does not employ an update step. However, in this case, temporal aliasing can occur in the approximation subband (which will correspond in this case to a simple subsampling of the original sequence).

The synthesis algorithm is straightforward:

- From transmitted l_t, h_t^-, h_t^+ and (16), compute x_{3t}.
- From h_t^-, h_t^+, x_{3t}, and (18)–(19), compute $\widetilde{h}_t^-$, $\widetilde{h}_t^+$.
- Using $\widetilde{h}_t^-$, $\widetilde{h}_t^+$, from (25)–(26) we get x_{3t+1}, x_{3t+2}.

This yields the following synthesis equations (for connected pixels denoted generically here by n):

$$\begin{aligned} x_{3t}(\boldsymbol{n}) &= l_t(\boldsymbol{n}) - \alpha\left[h_t^+(\boldsymbol{n} + \boldsymbol{v}_{3t+1}^+) + h_t^-(\boldsymbol{n} + \boldsymbol{v}_{3t-1}^-)\right] \\ x_{3t+1}(\boldsymbol{n}) &= \frac{1}{1-\beta^2}\left[h_t^+(\boldsymbol{n}) + \beta h_{t+1}^-(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-)\right. \\ &\quad + (1-\beta)x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+) \\ &\quad \left. + \beta(1-\beta)x_{3t+3}(\boldsymbol{n} - \boldsymbol{v}_{3t+2}^- + \boldsymbol{v}_{3t+2}^+)\right] \\ x_{3t+2}(\boldsymbol{n}) &= \frac{1}{1-\beta^2}\left[\beta h_t^+(\boldsymbol{n} + \boldsymbol{v}_{3t+1}^+) + h_{t+1}^-(\boldsymbol{n})\right. \\ &\quad + (1-\beta)x_{3t+3}(\boldsymbol{n} - \boldsymbol{v}_{3t+2}^-) \\ &\quad \left. + \beta(1-\beta)x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+ + \boldsymbol{v}_{3t+1}^-)\right]. \end{aligned}$$

Thus, at the synthesis, we see that the proposed scheme can be completely inverted,[1] under the assumption (24). This hypothesis is not true for every pair of motion connected pixels between the frames $3t+1$ and $3t+2$, but a large percentage of them satisfy it, as can be seen from Fig. 5. The solution that we have chosen (to limit the complexity and transmission overhead) is to compute only one of these two motion vector fields, for example $\boldsymbol{v}_{3t+2}^+$. Considering this solution, bidirectionally connected pixels in the frame $3t+1$ are obtained by backward prediction with $\boldsymbol{v}_{3t+1}^- = -\boldsymbol{v}_{3t+2}^+$ [see (14) and (25)]. For the pixels not connected with the frame $3t+2$, we apply a simple uni-directional prediction from the frame $3t$, using the vector $\boldsymbol{v}_{3t+1}^+$.

Note that the constraint (24) is very similar in spirit with the direct mode in MPEG-4, allowing to perform a simplified bidirectional prediction. Moreover, analogous to MPEG-4, depending on the available bitrate and required quality, the encoder could switch between a bidirectional ME/MC mode performing the estimate of both MVs $\boldsymbol{v}_{3t+1}^-$ and $\boldsymbol{v}_{3t+2}^+$ (and computing a combination of these estimates satisfying the desired linear constraint) and this "direct" mode.

[1]Note that the synthesis equations do not correspond to simply inverting the order of operations and changing the signs, which confirms the fact that the structure is not part of the classical lifting framework.

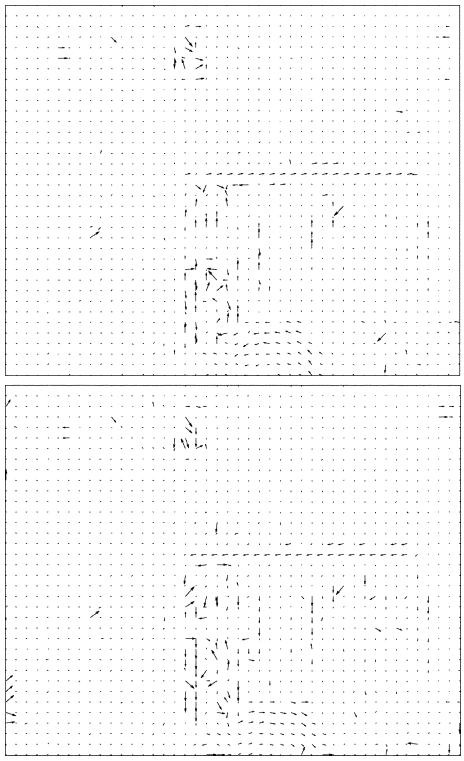

Fig. 5. Forward and backward motion vector fields between the first two frames in the CIF sequence "mobile."

TABLE I
IMPACT ON THE GLOBAL COMPRESSION PERFORMANCE OF THE MV SEARCH OPTIMIZED METHOD, AT DIFFERENT BITRATES (IN kbs)

Sequence	400	800	1200	1600	2000
mobile	27.49	32.11	34.22	35.67	36.89
mobile opti	28.24	32.26	34.25	35.72	36.95
foreman	34.08	37.50	39.92	40.75	41.81
foreman opti	34.42	37.57	39.38	40.75	41.81
stefan	X	28.81	31.64	33.64	35.38
stefan opti	22.95	29.37	31.95	33.98	35.52

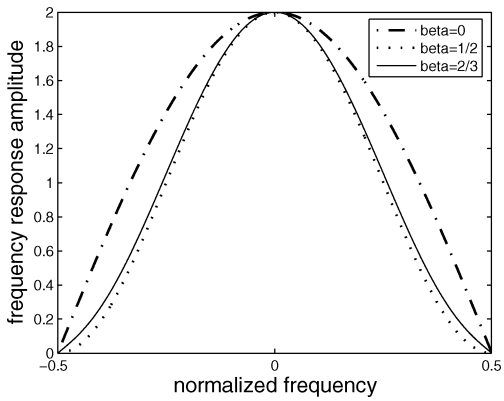

Fig. 6. Frequency response of high-pass filters for different values of β.

In the motion-compensated subband coding framework, this kind of motion invertibility was also discussed in [30] for motion composition. The same hypothesis of invertibility was used for a theoretical analysis of the Haar MCTF in [31].

Moreover, in order to optimize the nonlinear predict operators we have adopted an optimizing technique on the MV fields aimed at minimizing a criterion directly related to the coding efficiency of the detail frames. Detailed in [26] and [32], this iterative method basically consists in a joint search algorithm of the two MV (for example, $\boldsymbol{v}_{3t+1}^{-}$ and $\boldsymbol{v}_{3t+1}^{+}$), instead of separately estimating them. Note that this kind of approach was also used for interpolative prediction of video [33]. Additionally, this also leads to smoother MVs, which are easier to encode. Here, the optimization criterion needs to be adapted to the mathematical expression allowing to compute the detail frames. For example, the two parameter minimization problem related to the h_t^- frame reads as shown in the equation at the bottom of the page, where d can be any distortion measure (quadratic, absolute error, etc.), W^+ (respectively, W^-) is the forward (respectively, backward) search window in x_{3t-2} (respectively, x_{3t}) and $\mathcal{B}$ is the block of pixels in the current frame x_{3t-1}, which has to be predicted.

The impact on the global compression performance of this optimization method is illustrated in Table I. Note that the high-motion "stefan" sequence cannot be decoded at 400 kbs when a classical separate ME is used (marked by an X in Table I), and due to the reduction of the MV cost by the optimized method, the decoding becomes possible at low bitrates.

E. Structure Optimization

1) Parameter Choices: The parameters α and β can be selected to optimize the frequency characteristics of the low-pass, respectively, high-pass filters, but also to take into account the temporal distance between the frames employed for this nonlinear filtering operation.

For example, by varying $\beta \in [0, 1]$, we get for the high-pass filters the frequency responses in Fig. 6. Note that the most selective filters are obtained for $\beta = 1/2$, which corresponds to a symmetric temporal prediction based on adjacent frames. This is in accordance with the parameters usually employed for bidirectional prediction of B-frames in hybrid coding. However, the selectivity of these filters is deduced here in a *linear* framework, by neglecting the ME/MC. We show in Section V that for coding purposes the optimal value for β can be different for nonlinear spatio-temporal operators.

The least selective filters are those obtained for $\beta = 0$, i.e., corresponding to the Haar-like scheme. Moreover, note that with the structure proposed in (11)–(12), the magnitude of the two high-pass filters is identical.

The parameter α can be derived from (17) by setting the condition to have a low-pass filter, i.e., $L(-1) = 0$, where $L(z)$ is

$$(\widehat{\boldsymbol{v}}_{3t-1}^{+}, \widehat{\boldsymbol{v}}_{3t-1}^{-}) = \arg\min_{\boldsymbol{v}_{3t-1}^{+} \in W^{+}, \boldsymbol{v}_{3t-1}^{-} \in W^{-}} \sum_{\boldsymbol{m} \in \mathcal{B}} d\left[x_{3t-1}(\boldsymbol{m}) - \beta x_{3t-2}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^{+}) - (1-\beta) x_{3t}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^{-})\right]$$

the transfer function of the low-pass filter. In this case, independently of the value of β, we get $\alpha = 1/4$.

If, in addition, we choose $\beta = 1/2$, the parameters we get in this case are the same as those of the dyadic 5/3 biorthogonal filters. The structure developed in this section can be seen therefore as the 3-band nonlinear equivalent of the 5/3 biorthogonal multiresolution analysis.

2) Normalizing Constants: Since the quantization is performed identically in all the subbands, a re-normalization of the temporal subbands is necessary in order to be as close as possible to the orthonormal situation. The normalized filters will be obtained as

$$\check{l}_t = k_l\, l_t,\ \check{h}_t^+ = k_h\, h_t^+,\ \check{h}_t^- = k_h\, h_t^- \tag{27}$$

where l_t, h_t^+, and h_t^- are the filters defined in Sections II-A and III. Due to the symmetry of the scheme, we consider equal normalizations for h_t^+ and h_t^-.

Two conditions could be however pertinent for the goal we consider. On one hand, we would like to preserve the unitary norm for the impulse responses of the filters involved in the 3-band structure. On the other hand, an orthonormal structure preserves the energy of an input sequence. In particular, by considering the quantization error in each detail and approximation frame as i.i.d. variables, the sum of reconstruction errors of three consecutive frames should be equal to the sum of quantization errors of the approximation and detail frames

$$\sigma^2_{x_{3t-1}} + \sigma^2_{x_{3t}} + \sigma^2_{x_{3t+1}} = \sigma^2_{\check{l}_t} + \sigma^2_{\check{h}_t^+} + \sigma^2_{\check{h}_t^-} \tag{28}$$

where σ_a^2 denotes the variance of the frame a.

We first analyze the 3-band Haar-like structure. The first approach leads to $k_l = \sqrt{8/3}(\simeq 1.63)$ and $k_h = 1/\sqrt{2}(\simeq 0.707)$.

If the second approach is applied, from (27) and (4), (6) we obtain

$$x_{3t} = \frac{\check{l}_t}{k_l} - \frac{1}{4k_h}\left(\check{h}_t^+ + \check{h}_t^-\right) \tag{29}$$

$$x_{3t+1} = \frac{\check{l}_t}{k_l} + \frac{3}{4k_h}\,\check{h}_t^+ - \frac{1}{4k_h}\,\check{h}_t^- \tag{30}$$

$$x_{3t-1} = \frac{\check{l}_t}{k_l} + \frac{3}{4k_h}\,\check{h}_t^- - \frac{1}{4k_h}\,\check{h}_t^+ \tag{31}$$

and

$$\sigma^2_{x_{3t}} = \frac{1}{k_l^2}\,\sigma^2_{\check{l}_t} + \frac{1}{16k_h^2}\left(\sigma^2_{\check{h}_t^+} + \sigma^2_{\check{h}_t^-}\right) \tag{32}$$

$$\sigma^2_{x_{3t+1}} = \frac{1}{k_l^2}\,\sigma^2_{\check{l}_t} + \frac{9}{16k_h^2}\,\sigma^2_{\check{h}_t^+} + \frac{1}{16k_h^2}\,\sigma^2_{\check{h}_t^-} \tag{33}$$

$$\sigma^2_{x_{3t-1}} = \frac{1}{k_l^2}\,\sigma^2_{\check{l}_t} + \frac{9}{16k_h^2}\,\sigma^2_{\check{h}_t^-} + \frac{1}{16k_h^2}\,\sigma^2_{\check{h}_t^+}. \tag{34}$$

Now, using (28), we get $k_l = \sqrt{3}(\simeq 1.73)$ and $k_h = \sqrt{11}/4(\simeq 0.82)$. In Table II, we compare the coding efficiency obtained by these two normalization techniques on several sequences and at different bitrates. One can remark the almost equivalent performances of the two methods, which is not surprising, since the numerical values of the constants obtained by the two approaches are quite close. Due to its simplicity, our further investigations are based on the first technique.

TABLE II
PSNR DIFFERENCE IN dB BETWEEN THE TWO NORMALIZATION TECHNIQUES ($\mathrm{PSNR_I} - \mathrm{PSNR_{II}}$) FOR "MOBILE," "STEFAN," AND "FOREMAN" CIF SEQUENCES AT 30 FPS

Bitrate (kbs)	400	800	1200	1600	2000
mobile	-0.06	-0.04	0.01	-0.02	0.05
stefan	0.005	0.002	-0.05	-0.02	0.01
foreman	0.09	0.04	-0.07	-0.07	0.02

Let us now consider the 5/3-like 3-band structure. According to the first criterion and using (14), (15), and (17) with $\alpha = 1/4$, the conditions we impose are as follows:

$$k_l^2\left\{\left[1 - \frac{1}{2}(1-\beta)\right]^2 + 2\cdot\frac{1}{4^2} + 2\cdot\left(\frac{\beta}{4}\right)^2\right\} = k_h^2\left(1 + \beta^2 + (1-\beta)^2\right) = 1. \tag{35}$$

This leads to

$$k_l = \sqrt{\frac{8}{3}}\cdot\frac{1}{\sqrt{1+\frac{4}{3}\beta+\beta^2}},\ k_h = \frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{1-\beta+\beta^2}} \tag{36}$$

and for $\beta = 0$, the constants found for the Haar-like 3-band scheme are obtained. Using the values in (36), the PSNR variation (at 2 Mbs) of the reconstructed sequence "mobile" as a function of $\beta \in [0, 0.9)$, for one decomposition level is illustrated in Fig. 8. Note, first, that the optimal value is not $\beta = 1/2$ (corresponding to symmetric predictors and update operator), but $\beta = 0.21$. Second, we remark an important coding performance decrease for β values closer to 1. Even though our tests have shown that the energy of the detail frames is minimal for $\beta = 0.5$, the reconstruction error rapidly increases for larger values of β, which explains both phenomena. Indeed, let us compute the reconstruction error depending on β. We start by developing the synthesis equations as follows:

$$\begin{aligned}
x_{3t} &= l_t - \alpha h_t^+ - \alpha h_t^- \\
x_{3t+1} &= \frac{1}{1-\beta^2}[(1-\beta)\,l_t + \beta\,(1-\beta)\,l_{t+1} \\
&\quad + (1-\alpha\,(1-\beta))\,h_t^+ - \alpha\,(1-\beta)\,h_t^- \\
&\quad - \alpha\beta\,(1-\beta)\,h_{t+1}^+ + \beta\,(1-\alpha\,(1-\beta))\,h_{t+1}^-] \\
x_{3t-1} &= \frac{1}{1-\beta^2}[\beta\,(1-\beta)\,l_t + (1-\beta)\,l_{t+1} \\
&\quad + \beta\,(1-\alpha\,(1-\beta)) - \alpha\beta\,(1-\beta)\,h_t^- \\
&\quad - \alpha\,(1-\beta)\,h_{t+1}^+ + (1-\alpha\,(1-\beta))\,h_{t+1}^-].
\end{aligned}$$

Supposing now that the quantization errors of the different subbands are not correlated and also that the quantization step is the same in all the subbands, we get

$$\sigma^2_{x_{3t}} = \left(1+2\alpha^2\right)\sigma^2,\ \sigma^2_{x_{3t+1}} = \sigma^2_{x_{3t-1}} = \frac{1+\beta^2}{(1-\beta^2)^2}\left(\left(1+\alpha^2\right)(1-\beta)^2 + (1-\alpha\,(1-\beta))^2\right)\sigma^2$$

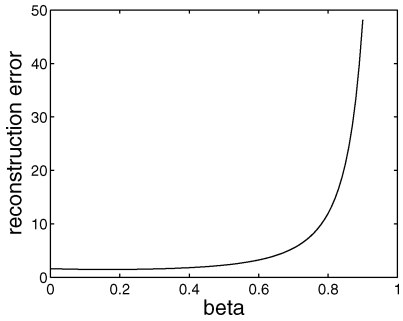

Fig. 7. Variation of $\sigma^2_{x_{3t-1}}$ and $\sigma^2_{x_{3t+1}}$ for $\beta \in [0, 0.9)$.

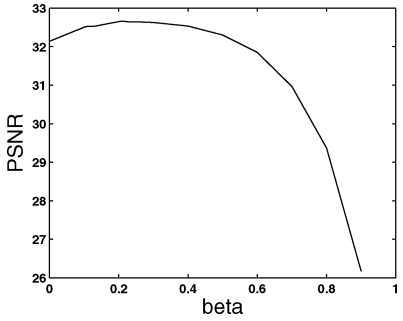

Fig. 8. PSNR of the reconstructed sequence "mobile" as a function of $\beta \in [0, 0.9)$.

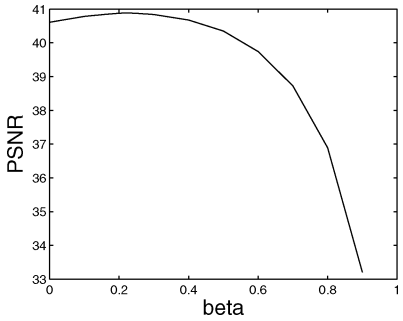

Fig. 9. PSNR of the reconstructed sequence "foreman" as a function of $\beta \in [0, 0.9)$.

where σ^2 denotes the variance of the subband frames due to the quantization. By taking $\alpha = 1/4$, as found before, the reconstruction error of the frames x_{3t+1} and x_{3t-1} is illustrated in Fig. 7. The strong increase when $\beta > 0.5$ explains the PSNR decrease for these values. Here, the value minimizing the reconstruction error is about 0.17. The other factor influencing the optimal value[2] of β is, as already said, the frequency selectivity of the filters. This influence is also highly dependent on the precision of the motion compensation. Compared with the results we presented in [25], where only integer pel ME/MC was used, one can remark the bidirectional filters can bring up to 1 dB compared with mono-directional ones (less selective), while for a 1/8th pel precision, this difference due to the selectivity reduces to maximum 0.3 dB.

As a matter of fact, the optimal value of β, and of the normalizing constants depending on it, are actually content-dependent. This is illustrated in Fig. 9 (optimum β is 0.23) and Fig. 10 (optimum β is 0.14), under the same simulation conditions as for

[2]In the sequel, we shall call "optimal value" for β the value leading to the best coding performance in terms of PSNR.

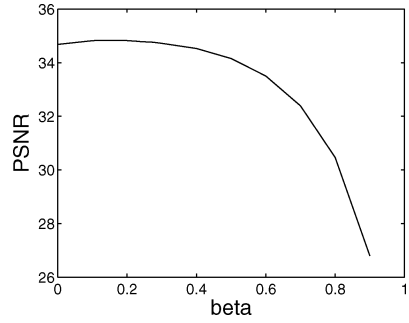

Fig. 10. PSNR of the reconstructed sequence "stefan" as a function of $\beta \in [0, 0.9)$.

TABLE III
OPTIMAL VALUES OF β AT THREE DECOMPOSITION LEVELS (β_1, β_2, AND β_3) AND THE VALUE OPTIMIZED GLOBALLY OVER THE THREE LEVELS (β_g) FOR DIFFERENT CIF SEQUENCES

Sequence	β_1	β_2	β_3	β_g
mobile	0.21	0.13	0.13	0.19
foreman	0.23	0.11	0.10	0.15
stefan	0.14	0.10	0.11	0.13

TABLE IV
PSNR VALUES (dB) WHEN USING OPTIMAL β (β_o) AT EACH OF THE THREE DECOMPOSITION LEVELS AND WHEN THE GLOBALLY OPTIMIZED VALUE (β_g) IS USED FOR DIFFERENT CIF SEQUENCES AND BITRATES (IN kbs)

Sequence	400	800	1200	1600	2000
$mobile_g$	28.24	32.26	34.25	35.72	36.95
$mobile_o$	28.28	32.28	34.28	35.72	36.97
$foreman_g$	34.42	37.57	39.38	40.75	41.81
$foreman_o$	34.45	37.58	39.45	40.75	41.81
$stefan_g$	22.95	29.37	31.95	33.98	35.52
$stefan_o$	22.96	29.38	31.95	33.98	35.52

"mobile." Also remark that the optimal β can vary for different temporal levels since at coarser temporal decomposition levels the filtered frames are farther apart, resulting in a worse temporal prediction. The variation of optimal β at different temporal decomposition levels is illustrated in Table III, where it is compared with the value of β optimized globally over three temporal decomposition levels. The compression performance of the codec can vary by at most 0.07 dB when using optimal values at each level, compared with the codec using β_g at all levels, as illustrated in Table IV. This table also illustrates a slight variation of the β parameter with the percentage of multiple connected and unconnected pixels, which increases with the temporal level, leading to a decrease of the optimum β with the temporal level.

Note also that the above analysis which allowed us to deduce the values of the normalizing constants was conducted under the simplifying hypothesis of linear filtering without taking into account the motion compensation in the temporal prediction.

As a conclusion of this subsection, in practice, two solutions are possible for the choice of optimal β parameters: either a two-pass encoding algorithm is allowed (for off-line encoding in applications where processing time is not critical and the quality is the main objective), or the value (or values, if different parameters are used at different level) of β has to be fixed in advance, independently of the input sequence. In the former case, the first pass of the encoding algorithm can explore the optimal values of

TABLE V
COMPARISON OF THE PSNR VALUES WHEN USING $\beta = 0$ AND $\beta = 0.15$ FOR DIFFERENT CIF SEQUENCES AND BITRATES (IN kbs)

Sequence	600	800	1200	1600	2000
$mobile\ \beta = 0$	30.64	32.12	33.86	35.44	36.59
$mobile\ \beta = 0.15$	30.59	32.25	34.28	35.71	36.96
$foreman\ \beta = 0$	35.98	37.33	39.07	40.55	41.58
$foreman\ \beta = 0.15$	36.22	37.57	39.38	40.75	41.81
$stefan\ \beta = 0$	27.77	29.44	31.90	33.85	35.44
$stefan\ \beta = 0.15$	27.46	29.37	31.95	34.00	35.51

β at each temporal level and transmit it together with the coded sequence. In the latter case, the fixed value(s) will be obviously suboptimal, but the encoding time is not increased. However, from the set of sequences we tested, one can deduce that optimizing β for each temporal level will not lead to a high improvement in coding performance and therefore a global value can be applied without too much loss. As practical guidelines for the choice of β, one can remark from the previous examples that the global optimal value is between 0.1 and 0.2. If different parameters need to be fixed at different levels, the value of β decreases at coarser temporal levels. We provide here a possible solution for a global β, leading to better coding performance than $\beta = 0$ for the three test sequences we considered, and being at less than 0.1 dB from the optimal PSNR at all bitrates. This value is $\beta = 0.15$, and we provide in Table V the rate-distortion comparison with $\beta = 0$.

F. Comparison With MPEG-Like Video Coding

The classical IBBPBBP… structure in hybrid coding can be seen as a subset of the above framework. In this case, the two consecutive B-frames represent the detail subbands, while the I and P-frames contribute to the approximation subband.

The equivalent 3-band structure can be described by the following equations (not taking into account MC):

$$h_t^+ = x_{3t+1} - \beta x_{3t} - \gamma x_{3t+3} \tag{37}$$

$$h_t^- = x_{3t-1} - \beta x_{3t} - \gamma x_{3t-3} \tag{38}$$

$$l_t = x_{3t} - x_{3t-3} \tag{39}$$

where β and γ are weighting factors between forward and backward prediction, generally taken equal to 1/2. The last equation defines the P frames, the I frames being the "initial condition," for $t = 0$. The h frames correspond to the B-frames of the hybrid scheme (actually, to the prediction error resulting from the bidirectional prediction).

An important remark is that the approximation subband is obtained by unstable recursive temporal filtering, compared with the stable low-pass filtering of our proposed open-loop schemes. This instability leads to the well known drift phenomenon, where the errors accumulate at the decoder, producing in time a desynchronisation with the encoder. In classical hybrid coding however, this problem is alleviated by the periodic introduction of I-frames.

One can also note the slightly different number of motion vectors between the open-loop subband scheme and the hybrid coding scheme. Indeed, the recursive filtering of the later one introduces one more motion vector field in the GOF structure, predicting a P frame from an I or P previous one.

Fig. 11. Part of the approximation frames at the third temporal decomposition level obtained with a Haar-like 3-band MRA (left) and with the bidirectional 3-band scheme (right) for the sequence "stefan" in CIF format.

IV. TEMPORAL SCALABILITY

The main advantage of the proposed schemes lies in the possibility to achieve an increased flexibility in temporal scalability. Indeed, by combining these 3-band or 5-band decompositions with two-band temporal filter banks, subsampling factors of containing powers of 3, 5, and 2 are achieved.

Concerning the quality aspects of temporal scalability, unlike temporal Haar two-band decomposition, the approximation is less affected by shadow effects around the contours, thanks to the bidirectional update. Moreover, compared with a direct time-subsampling of the sequence, the approximation subband resulting from the 3-band scheme takes better into account motion and contrast or intensity variations. Visually, temporal aliasing in the subsampled sequence can result in a slightly jerkyness effect, which is removed in the temporal filtered subband. This effect is amplified when the number of temporal scalability levels increases.

In order to compare the temporal scalability properties of the two 3-band structures, we have as before decomposed over three temporal levels. In Fig. 11, the approximation frames at the third level obtained with the Haar-like scheme and with the bidirectional 3-band scheme are compared. The blurring introduced by the former temporal MRA is much more important than the artefacts resulting from the processing with the latter one. To compare the temporal scalability performance in terms of compression efficiency, several options are available for computing the PSNR of the decoded temporal approximation subband (the same as in the case of a two-band MCTF scheme): either the reference sequence is considered with respect to the original (not encoded) temporal subband, or the reference is the temporally downsampled original sequence. The latter comparison can be misleading, for several reasons: do we need to subsample the even or the odd frames? The subsampling leads to temporal aliasing, which is precisely avoided by the low-pass filtering involved in obtaining the approximation subband which, in this sense, will make a better reference. In the former case, the dynamic range of the coefficients in the approximation subband

TABLE VI
RATE-DISTORTION COMPARISON OF DYADIC AND TRIADIC SCHEMES FOR "MOBILE" CIF 30 FPS SEQUENCE (PSNR IN dB)

Bitrate (kbs)	3B Haar	3B bidir	2B Haar 4 lev	2B Haar 5 lev	2B 53 4 lev	2B 53 5 lev	MPEG4
400	28.48	28.24	25.76	26.82	26.22	27.38	25.36
600	30.64	30.66	28.82	29.57	29.36	30.24	27.61
800	32.12	32.26	30.65	31.12	31.16	31.72	28.90
1200	33.86	34.25	33.06	33.37	33.46	33.82	30.71
1600	35.44	35.72	34.70	35.06	35.07	35.47	32.05
2000	36.59	36.95	36.18	36.31	36.41	36.76	33.10

TABLE VII
RATE-DISTORTION COMPARISON OF DYADIC AND TRIADIC SCHEMES FOR "FOREMAN" CIF 30 FPS SEQUENCE (PSNR IN dB)

Bitrate (kbs)	3B Haar	3B bidir	2B Haar 4 lev	2B Haar 5 lev	2B 53 4 lev	2B 53 5 lev	MPEG4
400	34.30	34.42	33.40	33.42	33.82	33.85	31.46
600	35.98	36.22	35.33	35.28	35.87	35.82	34.09
800	37.33	37.57	36.62	36.57	37.14	37.06	35.45
1200	39.07	39.38	38.58	38.51	38.99	38.92	37.23
1600	40.55	40.75	40.12	40.06	40.37	40.31	38.41
2000	41.58	41.81	41.25	41.19	41.46	41.40	39.28

frames is larger than that of the original frames (and varies depending on the number of temporal decomposition levels and on the temporal filters used for the analysis), which would require the use of a different definition for the PSNR measure. Moreover, the reference in this case is different from one encoder to the other, and in such a comparison one can have a very good PSNR but a reference frame of very bad subjective quality.

Excepting the visual comparison of the approximation subband (see for example [34]), an objective way to assess the performance in temporal scalability of two subband schemes is the PSNR of the reconstructed sequence at the original framerate, taking as reference the original sequence. In this context, an additional advantage of the proposed 5/3 3-band scheme is the possibility to perform a high quality motion-compensated temporal upconversion of a factor 3 by applying the synthesis scheme with null details. Indeed, the update step involves long-term motion-compensated filtering, leading to a more "fluid" upscaled sequence than a simple repetition of the original frames. In order to enhance the details in the sequence, one can use a stochastic modeling of the spatio-temporal wavelet coefficients, allowing to efficiently predict missing high-frequency subbands [35].

V. SIMULATION RESULTS

A structure with a variable number of temporal decomposition levels has been considered for simulations. As we established in our previous work in [36], the depth of the temporal decomposition can be adjusted depending on the motion activity in the sequence (a structure with two decomposition levels is more appropriate for example for a high activity sequence, while filtering over three levels will better capture the temporal correlation in a slowly moving sequence). Motion estimation is done within the MC-EZBC framework [5], [2] through the hierarchical variable size block matching (HVBSM) algorithm with block sizes varying from 64×64 to 4×4. Search range is first initialized at $[-2; 2]$, is increased if no good match can be found and is doubled at each temporal level. A motion-compensated prediction with fractional-pel accuracy [37] of 1/8th pel was performed. Motion vector fields encoding and bitrate allocation are done within the MC-EZBC framework; MVF are encoded as quad-tree maps and motion vector values are encoded with a zero-order arithmetic coder, in raster-scan order.

The temporal subbands are further spatially decomposed using biorthogonal 9/7 filters and then encoded using the 3D-EZBC algorithm [2]. Note that other encoding techniques can also be successfully applied on the spatio-temporal coefficients [38], [39]. The color CIF sequences have been encoded once and then the full bitstream was decoded at different bitrates.

In Tables VI and VII, we compare the coding efficiency of the proposed scheme with the equivalent codec based on the Haar-like 3-band structure and also with a 2-band Haar MCTF and a 2-band 5/3 MCTF [26]. For the purpose of comparison, we also added the results obtained with an MPEG-4 codec [40]. The two 3-band decompositions are performed over three temporal levels, which leads to a group of pictures of 27 frames. The performance of the dyadic temporal decomposition is tested for GOFs of size 16 (four decomposition levels) and 32 (five levels). The MPEG-4 GOF contains 27 frames between every two Intra (I) frames, and 2 B frames between any I or P frames, being from this point of view similar to a three-level 3-band scheme (see Section III-F). The optimal value for β, which is $\beta = 0.19$ for "mobile" and $\beta = 0.15$ for "foreman" was used in the 3-band scheme with bidirectional operators.

Note that the proposed bidirectional 3-band MCTF outperforms by about 0.3 dB the Haar-like 3-band scheme and by 0.4–1 dB the dyadic Haar and 5/3 structures on a sequence with irregular motion, like "foreman" at bitrates higher than 400 kbs. The improvement compared with the 3-band Haar-like scheme is related to the bidirectional nature of its update and predict operators. At very low bitrates (400 kbs), the 3-band Haar-like scheme takes advantage of a reduced number of MVF and overpasses the bidirectional codec. Compared with the dyadic schemes, the improved performance of the proposed temporal structure can be explained by a reduced number of motion vector fields to encode, and also a higher number of temporal detail frames per temporal level. One can also remark that our proposed 3-band schemes overpass by 2–3 dB the MPEG-4 codec. On a sequence with uniform motion like "mobile," the

TABLE VIII
RATE-DISTORTION COMPARISON OF A HAAR-LIKE 3-BAND SCHEME USING THE UPDATE LIFTING STEP AND THE SAME SCHEME WITHOUT THE UPDATE (CIF SEQUENCES, 30 FPS)

Bitrate (kbs)	400	800	1200	1600	2000
Foreman with update step	34.30	37.33	39.07	40.55	41.58
Foreman without update	33.44	36.32	37.90	39.37	40.37
Stefan with update step	24.82	29.44	31.91	33.85	35.44
Stefan without update	24.02	28.53	31.01	32.82	34.50
Mobile with update step	28.48	32.12	33.86	35.44	36.59
Mobile without update	27.25	30.74	32.52	34.14	35.38

TABLE IX
RATE-DISTORTION COMPARISON OF THE TEMPORAL SCALABILITY FEATURES FOR THE HAAR-LIKE AND THE BIDIRECTIONAL 3-BAND SCHEMES (CIF SEQUENCES, 30 FPS)

Bitrate (kbs)	400	800	1200	1600	2000
Foreman Haar-like 3-band	29.07	30.16	30.73	31.09	31.30
Foreman 3-band bidir	29.74	30.91	31.52	31.92	32.18
Stefan Haar-like 3-band	21.58	22.90	23.58	23.99	24.21
Stefan 3-band bidir	21.89	23.30	24.06	24.52	24.79
Mobile Haar-like 3-band	23.72	24.97	25.59	25.96	26.20
Mobile 3-band bidir	24.38	25.81	26.48	26.93	27.22

PSNR difference ranges from 0.1–0.5 dB with the Haar-like 3-band scheme (except at low bitrates, where the MV cost gives advantage to the monodirectional codec) and up to 1.6 dB with the dyadic MCTF.

Compared with our previous results, presented in [25], where the motion precision was only integer pel (and no iterative algorithm was used to compute the forward and backward motion vector fields), one can remark an improvement of up to 5 dB. This is related both to the improved prediction provided by fractional pel accuracy, but also to the difficulty to perform "motion composition" in integer pel, as discussed in Section III-D. In Table VIII, we compare the coding performance of a scheme without the update lifting step (the approximation subband is just a subsampled version of the original sequence) with a scheme involving the update step. One can remark a difference of 0.8–1.1 dB in favor of the latter structure. We compare in Table IX the coding performance of the two 3-band schemes in a context of temporal scalability. As explained in Section IV, the sequence is encoded at full bitrate and full framerate and then the part of the bitstream corresponding to the last temporal level is cut. Reconstruction is performed to the original framerate, by introducing null coefficients for the missing temporal detail frames and the PSNR is computed w.r.t. the original sequence. The table presents the results for a decomposition over three temporal levels and a reconstruction of only two of them. The difference of 0.3–1 dB is related to better performance of the bidirectional prediction operators as well as to the longer update interpolator used in the 5/3 3-band structure, compared to the Haar-like structure. Note that this upsampling was performed without using neither the coefficients in the first temporal detail frames, nor the MVFs at the first temporal decomposition level. As shown in [26], the performance can be increased by exploiting the MV information for a MC interpolation.

VI. CONCLUSION

In this paper, we have presented several extensions to the MCTF framework, allowing to provide temporal scalability factors that are not powers of two. We have mainly proposed and analysed a three-band motion-compensated temporal subband decomposition for scalable video compression which is the 3-band nonlinear counterpart of the dyadic 5/3 multiresolution analysis. Under some mild conditions on motion invertibility, we have proven the perfect reconstruction of this scheme which differs from the conventional lifting formalism employed until now for MCTF. By taking advantage of parametrized bidirectional prediction and update operators, the proposed structure outperforms the Haar-like 3-band scheme and the dyadic MCTF schemes, while also providing more flexible temporal scalability. Other schemes with nondyadic factors have been also proposed, and it was highlighted that combinations of different subband structures at different temporal levels based on the content characteristics can enable even more flexibility in frame-rate conversion as well as an improved coding efficiency. Moreover, we have shown how classical hybrid video coding and MCTF structures can be unified in a common framework of motion-compensated subband filtering.

The proposed structures can benefit from a context-based adaptativity at the block level: not only the prediction mode could be chosen (intra, uni-directional, bidirectional), but also the filter coefficient β could be optimized in this framework. This will lead to a reduced flexibility temporal scalability, but possible also to a higher coding efficiency.

Future work will be oriented towards improving the compression performance and generalizing the 3-band motion-compensated spatio-temporal structures to M-band decompositions. For example, using a method very similar to that introduced in this paper, one can build a 5-band motion compensated lifting scheme, resulting in one approximation and four detail subbands. Another example would consist in introducing motion estimation/compensation in the lifting structures proposed in [20], [21]. The interest for more channels can be two-fold. First, to allow complete freedom in the choice of the scalability factor (e.g., allowing temporal subsampling with factors of 5, 7, and combinations of such factors: for example, a 3-band scheme followed by a 5-band scheme leads to a reduction of a factor 15 in the framerate). Second, this enables the creation of approximation subbands using a reduced number of temporal decompositions. Depending on the sequence characteristics, motion model etc, this structure can provide higher coding performance. Also, depending on the desired frame-rates/temporal scalabilities a particular structure could be more efficient. This feature benefits certain applications, like for example video surveillance, where the motion activity is very low in most cases.

ACKNOWLEDGMENT

The authors would like to thank the reviewers and the Associate Editor for their valuable comments, which helped to improve the quality of the paper.

REFERENCES

[1] *Information Technology – JPEG 2000 Image Coding System*, ISO/IEC 15444-1, 2000.

[2] 3D MC-EZBC Software Package MPEG CVS repository.
[3] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Very low bit-rate embedded video coding with 3-D set partitioning in hierarchical trees (3D-SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 1365–1374, 2000.
[4] U. Horn and B. Girod, "Scalable video transmission for the internet," *Comput. Networks ISDN Syst.*, vol. 29, pp. 1833–1842, 1997.
[5] S. J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Process.*, vol. 8, pp. 155–167, 1999.
[6] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Process.*, vol. 3, pp. 559–589, 1994.
[7] S. T. Hsiang and J. W. Woods, "Invertible three-dimensional analysis/synthesis system for video coding with half-pixel accurate motion compensation," in *Proc. VCIP 99, SPIE*, 1999, vol. 3653, pp. 537–546.
[8] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Salt Lake City, UT, May 2001, vol. 3, pp. 1793–1796.
[9] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001, vol. 2, pp. 1029–1032.
[10] T. Russert and K. Hanke, "Optimized quantization in interframe wavelet coding," presented at the Shanghai MPEG Meeting Shanghai, China, Oct. 2002, doc. m9003.
[11] J. W. Woods, P. Chen, and S.-T. Hsiang, "Exploration experimental results and software," presented at the Klagenfurt MPEG Meeting Klagenfurt, Germany, Jul. 2002, doc. m8524.
[12] Y. Zhan, M. Picard, B. Pesquet-Popescu, and H. Heijmans, "Long temporal filters in lifting schemes for scalable video coding," presented at the Klagenfurt MPEG Meeting Klagenfurt, Germany, Jul. 2002, doc. m8680.
[13] J.-R. Ohm, "Complexity and delay analysis of MCTF interframe wavelet structures," presented at the Klagenfurt MPEG Meeting Klagenfurt, Germany, Jul. 2002, doc. m8520.
[14] M. Flierl and B. Girod, "Investigation of motion-compensated lifted wavelet transforms," in *Proc. Int. Picture Coding Symp.*, St. Malo, France, Apr. 2003, pp. 59–62.
[15] M. Flierl, "Video coding with lifted wavelet transforms and frame-adaptive motion compensation," *Lecture Notes Comput. Sci.*, vol. 2849, pp. 243–251, 2003.
[16] J. Xu, Z. Xiong, S. Li, and Y. Zhang, "Three-dimensional embedded subband coding with optimized truncation (3D-ESCOT)," *Appl. Comput. Harmon. Anal.*, vol. 10, pp. 290–315, 2001.
[17] D. Turaga and M. van der Schaar, "Unconstrained temporal scalability with multiple reference and bi-directional motion compensated temporal filtering," presented at the Fairfax MPEG Meeting Fairfax, VA, 2002.
[18] A. Secker and D. Taubman, "Highly scalable video compression using a lifting-based 3D wavelet transform with deformable mesh motion compensation," in *Proc. IEEE Int. Conf. Image Processing*, New York, Oct. 2002, vol. 3, pp. 749–752.
[19] F. J. Hampson and J.-C. Pesquet, "*M*-band nonlinear subband decompositions with perfect reconstruction," *IEEE Trans. Image Process.*, vol. 7, no. , pp. 1547–1560, 1998.
[20] T. Tran, "*M*-channel linear phase perfect reconstruction filter bank with rational coefficients," *IEEE Trans. Circuits Syst. I*, vol. 49, no. , pp. 914–927, 2002.
[21] Y. J. Chen, S. Oraintara, and K. Amaratunga, "*M*-channel lifting-based design of paraunitary and biorthogonal filter banks with structural regularity," in *Proc. IEEE Int. Conf. Circuits Systems*, May 2003, pp. IV 221–IV 224.
[22] C. Tillier and B. Pesquet-Popescu, "3D, 3-band, 3-tap temporal lifting for scalable video coding," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, vol. 3, pp. 779–782.
[23] M. van der Schaar and D. S. Turaga, "Unconstrained motion compensated temporal filtering (UMCTF) framework for wavelet video coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Hong Kong, Apr. 2003, vol. 3, p. 84.
[24] D. S. Turaga, M. van der Schaar, and B. Pesquet-Popescu, "Complexity scalable motion compensated wavelet video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, 15, no. 8, pp. 982–993, Aug. 2005.
[25] C. Tillier, B. Pesquet-Popescu, and M. van der Schaar, "Highly scalable video coding by bidirectional predict-update 3-band schemes," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, May 2004, vol. 3, pp. 125–128.
[26] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *Signal Process.: Image Commun., Special Issue on Wavelet Video Coding*, pp. 577–600, 2004.
[27] A. Benazza-Benyahia, J.-C. Pesquet, and H. Krim, "A nonlinear diffusion-based three-band filter bank," *IEEE Signal Process. Lett.*, vol. 10, pp. 360–363, 2003.
[28] D. S. Turaga, M. van der Schaar, and B. Pesquet-Popescu, "Differential motion vector coding with application to spatial scalable coding," in *Proc. Image and Video Communications and Processing*, Santa Clara, CA, Jan. 2003, vol. 5022, SPIE.
[29] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 245–267, 1998.
[30] J. Konrad, "Transversal versus lifting approach to motion-compensated temporal discrete wavelet transform of image sequences: Equivalence and tradeoffs," presented at the SPIE Conf. Visual Comm. Image Processing Jan. 2004.
[31] M. Flierl, P. Vandergheynst, and B. Girod, "Video coding with lifted wavelet transforms and complementary motion-compensated signals," presented at the SPIE Conf. Visual Comm. Image Processing San Jose, Jan. 2004.
[32] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Iterative predict optimization in MCTF video," presented at the MPEG Jul. 2003, doc. m9929.
[33] S.-W. Wu and A. Gersho, "Joint estimation of forward and backward motion vectors for interpolative prediction of video," *IEEE Trans. Image Process.*, vol. 3, pp. 684–687, 1994.
[34] S. S. Tsai, H.-M. Hang, and T. Chiang, Exploration Experiments on the Temporal Scalability of Interframe Wavelet Coding. Shanghai, China, Oct. 2002, SO/IEC JTC1/SC 29/WG 11 doc.M8959.
[35] G. Feideropoulou and B. Pesquet-Popescu, "Stochastic modelling of the spatio-temporal wavelet coefficients. Application to quality enhancement and error concealment," *J. Appl. Signal Process.*, vol. 12, pp. 1931–1942, 2004.
[36] M. van der Schaar and D. S. Turaga, "Content-adaptive filtering in the UMCTF framework," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Hong Kong, Apr. 2003, vol. 3, pp. 621–624.
[37] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–612, 1993.
[38] J. Viéron, C. Guillemot, and S. Pateux, "Motion compensated $2\mathrm{D} + \mathrm{t}$ wavelet analysis for low rate FGS video compression," presented at the Int. Thyrrhenian Workshop on Digital Communications 2002 Capri, Italy, Sep. 2002.
[39] C. Parisot, M. Antonini, and M. Barlaud, "3D scan based wavelet transform for video coding," in *Proc. 4th IEEE Workshop on Multimedia Signal Proc.*, 2001, pp. 403–408.
[40] *Video Reference Software, Version: Microsoft-FDAMI-2.3-001213*, ISO/IEC 14496 (MPEG-4), 2000.

Christophe Tillier was born on August 15, 1977, in Paris, France. He received the engineering degree in electrical engineering from Ecole Nationale Supériéure de l'Electronique et de ses Applications (ENSEA), Cergy, France, in 1999, the Agrégation degree in applied physics from Ecole Normale Supérieure (ENS), Cachan, France, and the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France.

From 2001 to 2005, he was a Teacher at Université Paris XII, Créteil, France. He is currently an Associate Professor in Multimedia at ENST. His research interests include scalable and robust video coding and multimedia applications.

Béatrice Pesquet-Popescu (M'06) received the engineering degree in telecommunications from the "Politehnica" Institute, Bucharest, Romania, in 1995, and the Ph.D. degree from Ecole Normale Supérieure (ENS), Cachan, France, in 1998.

She was a Research and Teaching Assistant at the Université Paris XI, Paris, France, in 1998 and then joined Philips Research France in 1999, where she worked for two years as a Research Scientist in scalable video coding. Since October 2000, she has been an Associate Professor in Multimedia at Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France. Her current research interests include scalable and joint source-channel video coding, adaptive wavelets, and multimedia applications. She holds 20 patents in wavelet-based video coding and has authored more than 100 book chapters, journal, and conference papers in the field.

Dr. Pesquet-Popescu received the EURASIP Best Student Paper Award in the IEEE Signal Processing Workshop on Higher-Order Statistics in 1997 a Young Investigator Award granted by the French Physical Society in 1998. She was a Guest Editor of the EURASIP *Journal on Applied Signal Proccessing* Special Issue on Video Analysis and Coding for Robust Transmission. She is a Member of the Technical Committee on Multimedia Signal Processing of the IEEE Signal Processing Society.

Mihaela van der Schaar (M'04) received the Ph.D. degree from Eindhoven University of Technology, Eindhoven, The Netherlands, in 2001.

She is an Assistant Professor in the Electrical Engineering Department, University of California at Los Angeles (UCLA). Prior to this, she was a Senior Researcher at Philips Research in both The Netherlands and the U.S., where she led a team of researchers working on multimedia compression, networking, communications, and architectures. In 2003, she was also an Adjunct Assistant Professor at Columbia University, New York. From 2003 to 2005, she was an Assistant Professor in the Electrical and Computer Engineering Department, University of California at Davis. She has published extensively on multimedia compression, processing, communications, networking, and architectures and holds 22 granted U.S. patents, with several more pending. Since 1999, she has been an active participant to the ISO Motion Picture Expert Group (MPEG) standard, to which she made more than 50 contributions and for which she received three ISO recognition awards. She was also chairing the *ad-hoc* group on MPEG-21 Scalable Video Coding and co-chairing the MPEG *ad-hoc* group on Multimedia Test-bed for three years.

Dr. van der Schaar was an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA and the *SPIE Electronic Imaging Journal* from 2002 to 2005. Currently, she is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY and an Associate Editor of *IEEE Signal Processing Letters.* She received the National Science Foundation Career Award in 2004 and the IBM Faculty Award in 2005.