Joint Source-Channel Coding with Partially Coded Index Assignment for Robust Scalable Video

Georgia Feideropoulou, Maria Trocan, James E. Fowler, *Senior Member, IEEE*, Béatrice Pesquet-Popescu, *Member, IEEE*, and Jean-Claude Belfiore, *Member, IEEE*

***Abstract*— A scalable video coder consisting of motion-compensated temporal filtering coupled with structured vector quantization plus a linear mapping of quantizer indices that minimizes simultaneously source and channel distortions is presented. The linear index assignment takes the form of either a direct, uncoded mapping or a coded mapping via Reed-Muller codes. Experimental results compare the proposed system to a similar scheme using unstructured vector quantization as well as to a prominent scalable video coder protected by more traditional convolutional codes. The proposed system consistently outperforms the other two schemes by a significant margin for very noisy channel conditions.**

***Index Terms*— joint source-channel coding, scalable video coding**

I. Introduction

Shannon's separability theorem is often used to justify the independent design of source- and channel-coding subsystems. However, in real-time video systems, the separability principle may not be applicable due to the high complexity for both the source and channel coders potentially entailed by the theorem. Consequently, there has been increasing interest in joint source-channel coding (JSCC) to provide efficient performance with complexity lower than tandem schemes.

Many prior JSCC techniques can be partitioned into two main categories: 1) *source-optimized channel coding*, wherein channel coding is optimized with respect to the source; and 2) *channel-optimized source coding*, wherein source coding is optimized with respect to the channel. In source-optimized channel coding, a quantizer—most generally, a vector quantizer—is designed for a noiseless channel. In the absence of explicit channel coding, vector quantization (VQ) can be made robust by applying a good index assignment (IA) to map quantization indices to channel codewords so as to minimize the impact of channel noise (e.g., [1]). On the other hand, when an explicit channel coding is used, careful attention is paid to optimally partition given resources between the source and channel coder (e.g., [2–5]). In channel-optimized source coding, the VQ and IA are simultaneously optimized for a specific channel such that very efficient clean-channel performance is obtained while providing robustness in the presence of noise (e.g., [6]).

G. Feideropoulou, M. Trocan, B. Pesquet-Popescu, and J.-C. Belfiore are with École Nationale Supérieure des Télécommunications (ENST), Paris, France.

J. E. Fowler is with the Department of Electrical & Computer Engineering and the GeoResources Institute (GRI), Mississippi State University, Starkville, MS.

This work was funded in part by the EC under grant IST-1-507113 (DANAE project), by the French Centre National de la Recherche Scientifique, and by the US National Science Foundation under Grant No. CCR-0310864.

In both the source-optimized channel coding and channel-optimized source coding categories of JSCC, the traditional approach is to cascade the channel code after the source code, such that the channel code adds redundancy to the transmission to combat channel errors and effectively increases the end-to-end transmission rate. An alternative category of JSCC, which can be considered to be *channel-constrained source coding*, was introduced in [7]. In such an approach, VQ is trained for minimum quantization distortion under constraints arising from the channel. The main result is that the channel distortion of a binary symmetric channel (BSC) is minimized if the source codebook can be expressed as a linear transform [8], that is, if the IA labeling is linear. Such linear IA includes direct mapping of VQ indices to channel codewords as well as coded IA wherein the VQ indices are mapped through a channel code. The use of the channel code in this latter approach effectively constrains the VQ source codewords to reside in the space of channel codewords. This marks a substantial departure from the traditional use of channel coding to add redundancy—and, consequently, increased transmission rate—as is the case in schemes that concatenate source and channel coding (e.g., [2–5]).

In [9], linear transforms constructed from lattice constellations with "maximum component diversity" were used to build structured VQ codebooks which minimized simultaneously the source and channel distortions for Gaussian sources. In this paper, we develop a JSCC scheme in the channel-constrained source-coding category for the coding of video wherein the source distribution is not Gaussian. Specifically, we describe a scalable video-coding system constructed from $t+2D$ motion-compensated temporal filtering (MCTF) coupled with JSCC using the structured VQ of [9]. The VQ indices are mapped to channel codewords either directly in an uncoded form, or through coded IA based on Reed-Muller codes, with the encoder adaptively deciding between the coded and uncoded IA on a subband-by-subband basis. Consequently, with coded IA, the source codewords themselves are constrained to belong to the channel code, and there is no rate increase due to the incorporation of the channel code. We compare our proposed coding scheme to a source-optimized channel-coding technique featuring unstructured VQ of MCTF coefficients coupled with the IA mapping of [1], as well as to the more traditional approach to error resilience consisting of concatenating a source coder (the prominent MCTF-based coder MC-EZBC [10]) with a channel coder (convolutional codes). We find that the proposed JSCC system consistently outperforms the other two schemes as the channel noise level increases.

II. Index Assignment for Gaussian Sources

Let vector $\boldsymbol{x}$ be the input to a vector quantizer which produces an n-bit binary codeword, the quantization index of the vector. The source codebook can then be viewed as a function of $\boldsymbol{b} = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T \in \{+1, -1\}^n$, where $\boldsymbol{b}$ represents the IA of $\boldsymbol{x}$. Under the assumption of a maxentropic quantizer, the total distortion is $D = D_s + D_c$, where D_s is the source distortion due to quantization, and D_c is the channel distortion dependent on the IA.

In [8], the channel distortion of a BSC is proved to be minimized by IA in the form of a linear labeling, while, in [9], a linear labeling that minimizes simultaneously the source and channel distortions is constructed. In the case of a zero-mean Gaussian source, this linear labeling is constructed using a subset of lattice constellations with "maximum component diversity." Specifically, let $\boldsymbol{U}_n$ be an $n \times n$ generator matrix of a maximum-component-diversity lattice constellation as described in [11]. Its construction is based on number-field theory, and it is expressed by the standard embeddings in $\boldsymbol{R}^n$ of the ideal ring of the totally real subfield of cyclotomic fields. The rows and the columns of $\boldsymbol{U}_n$ are denoted by L_{in} and C_{nj}, respectively, where $1 \le i, j \le n$. If J is some subset of $\{1, \ldots, n\}$, then $C_{nj}(J)$ is the j^{th} column of $\boldsymbol{U}_n(J)$, which is a matrix of only the rows of $\boldsymbol{U}_n$ corresponding to the indices in J. Using $\boldsymbol{U}_n$, one can linearly map $\mathrm{BPSK}_n = \{-1, +1\}^n$ onto a new set $\boldsymbol{U}_n \cdot \mathrm{BPSK}_n$. Allowing n to increase while J remains fixed, we get a codebook $S_n(J)$ with codewords $\boldsymbol{y}^{(l)}$, $\boldsymbol{y}^{(l)} = \sum_{j=1}^{n} b_j^{(l)} C_{nj}(J)$, where $\boldsymbol{b}^{(l)} = \begin{bmatrix} b_1^{(l)} & \cdots & b_n^{(l)} \end{bmatrix}^T \in \mathrm{BPSK}_n$, and $1 \le l \le 2^n$. In order to obtain a family of matrices $\boldsymbol{U}_n$ such that $S_n(J)$ is an asymptotically Gaussian source dictionary that minimizes D_s as $n \to \infty$, $\boldsymbol{U}_n$ must be orthogonal with coefficients going uniformly to 0 as $n \to \infty$ [9]. In this case, the linear mapping $\boldsymbol{b} \in \mathrm{BPSK}_n \to \big(\boldsymbol{G}_{d,n}\boldsymbol{b} \in S_n(J)\big)$, where $d \times n$ matrix $\boldsymbol{G}_{d,n} = \boldsymbol{U}_n(J)$, $d = |J|$, allows the construction of a source dictionary that is asymptotically Gaussian. Similar properties are achieved by selecting columns of the matrix $\boldsymbol{U}_n$, and we shall denote the $n \times r$ matrices constructed this way as $\boldsymbol{G}'_{n,r}$, where $r \le n$.

The above discussion assumes that the uncoded IA $\boldsymbol{b}$ is transmitted directly on the channel. In the alternative case that an error-correcting code is used, $\boldsymbol{b}$ ranges in $\boldsymbol{m}(\boldsymbol{c})$ where $\boldsymbol{c}$ is one of the 2^k possible binary codewords belonging to the (n, k) linear code $\mathcal{C}$. The function $\boldsymbol{m}(\cdot)$ maps $\boldsymbol{c} = \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix}^T$ onto $\boldsymbol{m}(\boldsymbol{c}) = \begin{bmatrix} m(c_1) & \cdots & m(c_n) \end{bmatrix}^T$, where $m(0) = 1$ and $m(1) = -1$. The codebook for this coded case has codevectors $\boldsymbol{y}^{(l)}$ given by $\boldsymbol{y}^{(l)} = \boldsymbol{G}_{d,n}\boldsymbol{b}^{(l)}$, where $\boldsymbol{b}^{(l)} = \boldsymbol{m}(\boldsymbol{c}^{(l)})$, $\boldsymbol{c}^{(l)} \in \mathcal{C}$, and $1 \le l \le 2^k$.

III. Coding of Spatio-Temporal Subbands

We now apply the JSCC scheme described above to a scalable video coder. The resulting system first applies $t+2D$ MCTF in the form of a motion-compensated temporal wavelet transform applied to a group of frames (GOF) followed by a spatial wavelet transform of the temporal subbands. Next, an optimal bit-allocation procedure allocates rate among the spatio-temporal subbands, after which the spatio-temporal coefficients are vector quantized. Finally, a linear IA mapping between the source codebook and the coded symbols sent on the channel is applied to provide resilience to channel noise.

A. Index Assignment for non-Gaussian Sources

Because the coefficients of the $t+2D$ MCTF subbands are not Gaussian, the coding scheme of Sec. II cannot be applied directly. However, the marginal distribution of the subband coefficients has been shown to be well-modeled by a mixture of two Gaussians [12]; thus, we classify vectors drawn from the spatio-temporal subbands into two vector classes and apply the IA approach of Sec. II to each class independently.

For vectors from the temporally lowpass (approximation) frames, it was observed in [12] that classification according to vector magnitude, such that the vectors are partitioned into a low-variance and a high-variance class, results in an approximately Gaussian distribution within each class. Similarly, vectors from the temporal highpass (detail) frames are classified into two classes using the stochastic model of spatio-temporal dependencies introduced in [12]. This permits accurate classification based on only the coefficients already decoded, without requiring transmission of side information. Following this model, we assume that the conditional probability of a coefficient is Gaussian with variance depending on a set of its spatio-temporal neighbors; i.e., the conditional probability of coefficient x is $f(x|\sigma_x^2) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right)$, where the variance is $\sigma_x^2 = \sum_i w_i \left|p_i(x)\right|^2 + \alpha$, such that $p_i(x)$ are coefficients neighboring x in the same spatio-temporal subband, w_i are weight parameters, and α is an offset parameter. The spatio-temporal neighborhood is a set of causal coefficients that will have already been received by the decoder when the current coefficient is decoded. Estimation of the parameters (w_i and α) of this model is done as in [12].

B. Quantization and Bit Allocation

For each vector class described above, we design a VQ codebook by minimizing cost $\gamma = \mathrm{E}\big[\min_l \big\|\boldsymbol{x} - \beta \boldsymbol{G}_{d,n}\boldsymbol{b}^{(l)}\big\|^2\big]$, where, in the case that the IA is uncoded, $\boldsymbol{b}^{(l)}$ ranges over the set of 2^n possible codewords of a BPSK_n, and, in the case that the IA includes an error-correcting code, $\boldsymbol{b}^{(l)}$ ranges over the set of 2^k possible codewords of an (n, k) code $\mathcal{C}$. β is a parameter which scales the lattice constellation $\boldsymbol{G}_{d,n}$ to the source dynamics. In order to find β, as well as the codebook with vectors $\boldsymbol{y}^{(l)} = \beta \boldsymbol{G}_{d,n}\boldsymbol{b}^{(l)}$, an iterative optimization algorithm (similar to that of shape-gain VQ) is used. A similar optimization is applied when using the matrix $\boldsymbol{G}'_{n,r}$ for VQ; in this case, an (r, k) code $\mathcal{C}'$ is used for the coded IA.

The channel distortion D_c is minimized due to the linearity of the IA labeling, and its value is fixed for a given channel-noise variance. Consequently, an iterative bit-allocation algorithm is applied to allocate VQ rate among the spatio-temporal subbands in an optimal fashion. This bit-allocation algorithm, which originates in [13], takes into account a nonnegativity constraint on the rate allocated to each subband. The algorithm indicates the size of the $\boldsymbol{G}_{d,n}$ or $\boldsymbol{G}'_{r,n}$ matrix which minimizes the end-to-end distortion. The choices of $\boldsymbol{G}_{d,n}$ or $\boldsymbol{G}'_{r,n}$

are, however, limited in practice by computational-complexity issues and dependences between the spatio-temporal coefficients. That is, it is known that the spatio-temporal coefficients exhibit strong correlation with their spatial or spatio-temporal neighbors; thus, in order to exploit these relationships, the dimensions d in $\boldsymbol{G}_{d,n}$ or n in $\boldsymbol{G}'_{n,r}$ should be a power of 4. However, in order to attain high or low coding rates, we permit these values to be 2 if need be. In addition to keep the complexity low, we limit the dimensions n in $\boldsymbol{G}_{d,n}$ and r in $\boldsymbol{G}'_{n,r}$ to be no greater than 16.

C. Partially Coded Index Assignment

We initially applied the VQ and IA described above in an uncoded fashion, i.e., without the use of any error-correcting codes in the IA mapping. However, when transmitting over a Gaussian channel with low SNR, we remarked that for some subbands, especially those with high energy, the total distortion D was very high compared to the source distortion D_s obtained when the channel was noiseless. We conclude that, in this situation, the channel distortion D_c must be dominant. In order to improve performance, we replace the uncoded IA with coded IA incorporating an error-correcting code for these subbands. We choose Reed-Muller codes due to their symmetry, their widespread use in lattice construction, and their error-correcting capability.

For coded IA, we restrict the mapping space to be the space of the binary vectors belonging to the Reed-Muller code. We additionally constrain the source-coding rate to be the same rate as dictated by the bit-allocation algorithm in the uncoded case. We then choose the (η, k) Reed-Muller code in light of the trade-off between the following considerations: 1) the error-correction capability of the code; 2) the blocklength η of the code must be $\eta = n$ of $\boldsymbol{G}_{d,n}$, or $\eta = r$ of $\boldsymbol{G}'_{n,r}$, as appropriate, in order that the bitrate does not increase; and 3) the dimension of the code k should be close to n or r so that the number of 2^k possible codewords is close to the 2^n or 2^r possible codewords of the uncoded case, in order to minimize the increase to the source distortion. In this way, the end-to-end distortion decreases without changing either the source-coding rate of the uncoded case or the total bitrate.

The encoding algorithm consists of the following steps. In each subband, we calculate D_s in a noiseless environment as well as the end-to-end distortion D for the given noisy channel as it would be obtained with an uncoded IA. If the difference between D and D_s is high (which means that D_c is significant), we restrict the IA to be codewords of a Reed-Muller code selected with the considerations discussed above. Otherwise, the IA maps directly to uncoded codewords. At the decoder side, soft-decision decoding via the Viterbi algorithm with the BCJR trellis [14] is applied to the coded codewords, while hard-decision decoding is applied to the uncoded codewords. We note that the encoder sends a small amount of side information to the decoder (β and the sizes of $\boldsymbol{G}_{d,n}$ and $\boldsymbol{G}'_{n,r}$ for each vector class for each subband, as well as the coded/uncoded state for each subband); it is assumed that this side information is highly protected so as to arrive at the decoder uncorrupted.

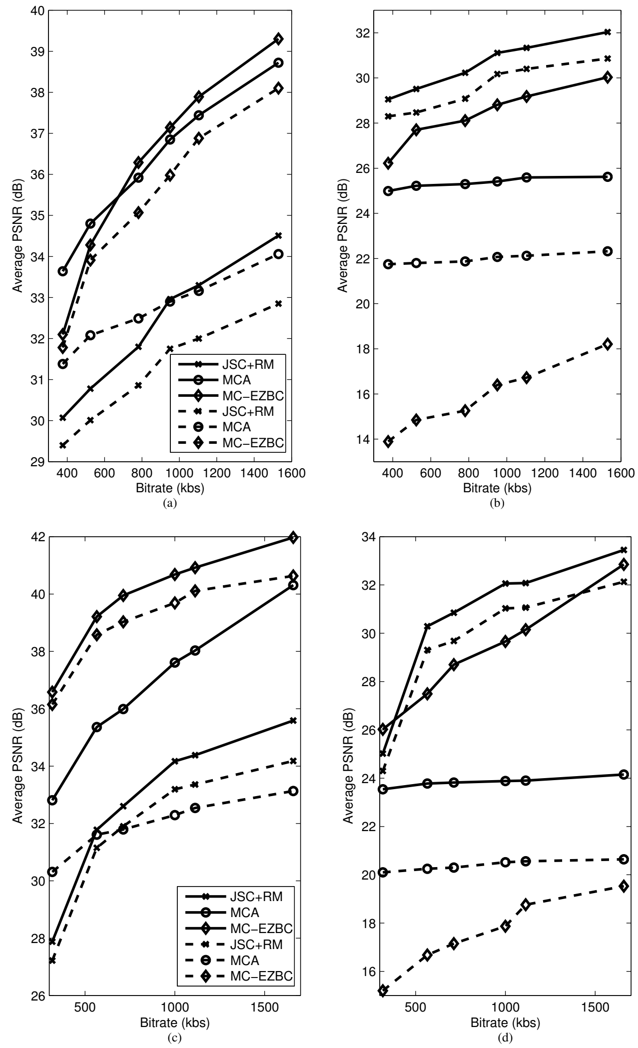

Fig. 1. (a) Foreman, channel SNRs of ∞ (solid lines) and 6.75 dB (dashed lines); (b) Foreman, channel SNRs of 4.33 dB (solid lines) and 3.00 dB (dashed lines); (c) Hall Monitor, channel SNRs of ∞ (solid lines) and 6.75 dB (dashed lines); (d) Hall Monitor, channel SNRs of 4.33 dB (solid lines) and 3.00 dB (dashed lines).

IV. Simulation Results

To experimentally evaluate the effectiveness of the system described in the previous section, we perform simulations using CIF test sequences at 30 fps. The video-coding system uses a Haar MCTF decomposition applied on GOFs of 16 frames, with 4 temporal and 2 spatial resolution levels. The spatial transform uses the popular biorthogonal 9/7 filters.

As we remarked in Sec. III-C, a Reed-Muller code is incorporated into the IA for subbands of high energy. Our bit-allocation algorithm dictates that these high-energy subbands are coded using $\boldsymbol{G}_{d,16}$ or $\boldsymbol{G}'_{n,16}$. Thus, in consideration of the three trade-offs discussed in Sec. III-C, we choose the $\mathrm{RM}\{2,4\}$ with $\eta = 16$ and $k = 11$. However, compared to the uncoded case, we expect the source distortion to increase, as now the dimension of the mapping space has decreased to 2^{11} possible codewords, instead of 2^{16} as in the uncoded case. On the other hand, the end-to-end distortion of the entire scheme in a noisy environment will decrease significantly compared to the uncoded case, and the total bitrate remains the same.

We compare our proposed scheme (which we denote as "JSC+RM") to two other video coders using the same scalable MCTF transform structure. The first technique belongs to the class of source-optimized channel coding. In this scheme, a VQ source coder is designed to minimize source distortion for the noiseless channel, while a good, albeit suboptimal, IA is applied to increase the error resiliency of the quantizer. For the VQ source coder, we apply the locally optimal generalized Lloyd algorithm (GLA) to produce unstructured VQ codebooks with locally minimal source distortion D_s. The GLA VQ codebooks are of the same dimensions as dictated by the bit-allocation algorithm of our proposed scheme.We then follow with the Minimax Cover Algorithm (MCA) [1], which is an IA using a minimax error criterion that is designed against worst-case performance without sacrificing average performance. We refer to this coder as "MCA."

The second video coder to which we compare corresponds to an implementation of MCTF concatenated with traditional error-control coding. We employ the prominent MC-EZBC [10] coder, and, to provide error resilience, we packetize the MC-EZBC bitstream while applying rate punctured convolutional (RPC) codes to the resulting packets. Specifically, each packet contains the information corresponding to a single spatial resolution level from a single temporal subband frame. Hence, the packets have unequal length and are coded unequally by RPC codes. If the decoder fails to decode a received packet, the packet is dropped. The RPC codes have $R_p = 2/3, 3/4$, and $7/8$ with memory $m = 6$ and mother code $R = 1/2$ [15]. The most important information is protected by $2/3$ codes, the medium spatio-temporal frequencies by $3/4$ codes, and the finest details by $7/8$ punctured codes. We refer to this second coder as "MC-EZBC."

Fig. 1 presents the results obtained using the three different coding schemes for the "hall-monitor" and "foreman" sequences at different bitrates over a Gaussian channel with four different channel-noise levels. Note that, for the noiseless channel, the IA is entirely uncoded for our JSC+RM scheme. In Fig. 1 we observe that both MCA and MC-EZBC yield performance superior to JSC+RM when the channel is noiseless. This is as expected, as these two algorithms are designed for noiseless channels. In particular, we expect MCA to outperform JSC+RM due to the unstructured nature of the codebooks generated by GLA, whereas the JSC+RM codebooks are highly structured. On the other hand, when the channel becomes very noisy (e.g., channel SNRs of 4.33 dB and 3 dB), the performances of MCA and MC-EZBC drop dramatically while JSC+RM remains quite close to its noiseless performance. Indeed, JSC+RM consistently outperforms both MCA and MC-EZBC for the very noisy channel.

V. Conclusions

In this paper, we presented an approach to the JSC coding of scalable MCTF video. The proposed system is based on structured VQ coupled with linear IA in the form of uncoded IA, or coded IA via Reed-Muller codes. We compared the performance of the proposed system to that of a source-optimized channel coding using unstructured VQ codebooks without an explicit channel coder, as well as to that of the prominent MCTF-based MC-EZBC coder protected unequally with RPC codes, in the more traditional paradigm of concatenated source and channel codes. As the channel noise increases, the proposed coding system retains end-to-end distortion performance close to that of the noiseless channel as well as consistently outperforms the other two schemes for very low channel SNR.

As a final observation, we note that, at the encoder, the complexity of our scheme is similar to that of the unstructured-VQ coder, except we avoid the IA post processing of [1] in the creation of VQ codebooks. No additional encoder complexity occurs due to the use of coded IA, since the VQ source codewords belong to the space of channel codewords, unlike concatenated source-channel schemes that require subsequent channel-coding processing. On the other hand, at the decoder, the complexity of our proposed scheme is comparable to that of concatenated schemes as Viterbi decoding is required in both cases.

References

[1] L. C. Potter and D.-M. Chiang, "Minimax nonredundant channel coding," *IEEE Transactions on Communications*, vol. 43, no. 4, pp. 804–811, April 1995.

[2] M. Bystrom and T. Stockhammer, "Dependent source and channel rate allocation for video transmission," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 258–268, January 2004.

[3] P. G. Sherwood and K. Zeger, "Progressive image coding for noisy channels," *IEEE Signal Processing Letters*, vol. 4, no. 7, pp. 189–191, July 1997.

[4] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 340–356, March 2000.

[5] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1412–1424, September 1997.

[6] S. Gadkari and K. Rose, "Vector quantization with transmission energy allocation for time-varying channels," *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 149–157, January 1999.

[7] M. Skoglund, "On channel-constrained vector quantization and index assignment for discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 45, no. 17, pp. 2615–2622, November 1999.

[8] P. Knagenhjelm and E. Agrell, "The Hadamard transform—A tool for index assignment," *IEEE Transactions on Information Theory*, vol. 42, pp. 1139–1151, July 1996.

[9] J.-C. Belfiore, X. Giraud, and J. Rodriguez-Guisantes, "Optimal linear labeling for the minimization of both source and channel distortion," in *Proceedings of the IEEE International Symposium on Information Theory*, Sorrento, Italy, June 2000, p. 404.

[10] P. Chen and J. W. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 10, pp. 1183–1194, October 2004.

[11] X. Giraud, E. Boutillon, and J.-C. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Transactions on Information Theory*, vol. 43, pp. 938–952, May 1997.

[12] G. Feideropoulou, B. Pesquet-Popescu, J.-C. Belfiore, and J. Rodriguez, "Non-linear modeling of wavelet coefficients for a video sequence," in *Proceedings of the International Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, June 2003.

[13] G. Feideropoulou, B. Pesquet-Popescu, and J.-C. Belfiore, "Bit allocation algorithm for joint source-channel coding of $t + 2D$ video sequences," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, Philadelphia, PA, March 2005, pp. 177–180.

[14] R. J. McEliece, "On the BCJR trellis for linear block codes," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1072–1091, July 1996.

[15] D. Haccoun and G. Bégin, "High-rate punctured convolutional codes for Viterbi and sequential decoding," *IEEE Transactions on Communications*, vol. 37, no. 11, pp. 1113–1125, November 1989.