BIT ALLOCATION ALGORITHM FOR JOINT SOURCE-CHANNEL CODING OF $T + 2D$ VIDEO SEQUENCES

Georgia Feideropoulou, Béatrice Pesquet-Popescu* and Jean-Claude Belfiore+*

*ENST Paris, Signal and Image Proc. Dept.
+ENST Paris, Communication and Electronics Dept.
46 rue Barrault, 75634 Paris, France
{feiderop}@tsi.enst.fr

ABSTRACT

A joint source-channel coding scheme of $t + 2D$ decomposed video sequences and an iterative bit allocation are presented. The joint source-channel coding scheme consists of a vector quantization and a linear labelling by lattice constellations minimizing simultaneously the channel and the source distortion. The channel distortion, due to the linear labelling, is minimized and depends on the value of the noise variance and the variance of the source. The iterative algorithm results in an optimal codebook allocation subject to a global bit rate and a nonnegativity constraint. The overall flexible coding scheme is proved to be very efficient in noisy environments.

1. INTRODUCTION

Shannon's theory allows to claim that source coding and channel coding can be treated/optimized separately. However, this is achievable only for large block sizes from the source and the channel sides. This imposes a high complexity which is not convenient for real-time video systems. To avoid this drawback, a joint optimization of the source-channel coding provides a realizable solution. The purpose of a joint source-channel coding approach is to allocate bits between the source and channels coders in an optimal manner, subject to a constraint, which, in most cases, is the overall coding rate. The optimal manner is based on the minimization of the end-to-end distortion subject to the above constraint.

However, especially in video transmission, the information exhibit different layers of importance. In one way, this can be captured by scalable video codecs based on $t + 2D$ wavelet decomposition which provide very high coding efficiency and enable spatio/temporal scalability. In an additional way by applying a bit allocation algorithm which treats efficiently these different layers of importance.

Our joint source-channel coding scheme follows the channel point of view which means that it is based on the minimization of the channel distortion first followed by the minimization of the source distortion. In [1], it has been proved that for binary discrete channels, the channel distortion is minimized if the vector quantization lattice can be expressed as a linear transform of an hypercube. This work was the motivation in [2] to find a set of linear transforms which minimizes the channel distortion, in the same time as the source distortion for Gaussian sources.

Our formulation is based on the extension of the above works to the case of video sources, whose distribution is not Gaussian. The spatio-temporal wavelet coefficients are encoded by a vector quantization based on a linear labelling. "Maximum component diversity" lattice constellations are used to minimize not only the channel distortion but at the same time the distortion of the video source. No additional protection by error-correcting codes is necessary.

As the channel distortion is already minimized and its value is fixed for a given variance of channel noise, an iterative bit allocation algorithm is applied in order to optimally allocate the bits between the subbands resulting in an "optimal" codebook allocation. It takes into account the nonnegativity constraint of the rate allocated to each subband. An early attempt to avoid this problem was presented in [5] where the nonnegativity constraint had been treated. Moreover, in [6] the constraint of a nonnegative integer solution had been proposed.

The paper is organized as follows: in the next section, we present the structure of our joint source-channel coding scheme in Section 3 we develop our bit allocation algorithm and in Section 4 we present some simulation results. Section 5 concludes this paper.

2. STRUCTURE OF AN EFFICIENT JOINT SOURCE-CHANNEL CODING SCHEME

Let a d-dimensional vector $\boldsymbol{x}$ be the input of a vector quantizer, producing a n-bit binary codeword, which is the index of the vector used for signal reconstruction at the receiver. The source codebook can be viewed as as function of $(b_1...b_n) \in \{+1,-1\}^n$ representing the index assignement. Under the assumption of a maxentropic quantizer, the total distortion can be expressed as

$$D = D_s + D_c$$

where D_s is the distortion due to the quantization and D_c is the distortion dependent on the index assignement. In [1], it is proved that for the binary symmetric channel, D_c is minimized by linear labelling. Moreover, in [2] a linear labelling that minimizes the source distortion D_s is constructed. It is fully described by a $d \times n$ $(d \leq n)$ matrix $\boldsymbol{G}_d$. In the case of a stationnary, memoryless zero-mean Gaussian source with variance one and a maxentropic source coding, the linear labelling represented by the matrix $\boldsymbol{G}_d$ must transform an identically distributed random variable into a random variable (the source codebook) which has to mimic the source distribution.

Let $\boldsymbol{M}_n$ be an $n \times n$ generator matrix of a "maximum component diversity" lattice constellation as described in [3]. Its construction is based on the number-field theory and it is expressed by the standard embeddings in $\boldsymbol{R}^n$ of the ideal of ring of totally real subfield of cyclotomic field.

The rows and the columns of $\boldsymbol{M}_n$ are denoted by L_{in} and C_{jn} respectively, where $1 \leq i,j \leq n$. If J is some subset of $\{1,...,n\}$, then $C_{jn}(J)$ is the truncation of the j-th column of $\boldsymbol{M}_n$ according to the subset of indices J. By means of $\boldsymbol{M}_n$ one can map linearly $BPSK_n = \{-1,+1\}^n$ on a new set $\boldsymbol{M}_n.BPSK_n$. Allowing n to increase while J remains fixed, we get a codebook $S_n(J)$ with codewords: $y_n = \sum_{j=1}^{n} b_j C_{jn}(J)$ where $\boldsymbol{b} = (b_1,...,b_n)^t \in BPSK_n$.

In order to obtain a family of matrices $\boldsymbol{M}_n$ such that $S_n(J)$ is an asymptotically Gaussian source dictionnary and it minimizes D_s as $n \to \infty$, it is shown in [2] that $\boldsymbol{M}_n$ must be orthogonal with coefficients going uniformly to 0 as $n \to \infty$. Then the mapping

$$\boldsymbol{b} \in BPSK_n \to (\boldsymbol{G}_d \cdot \boldsymbol{b} \in S_n(J))$$

where $\boldsymbol{G}_d = \boldsymbol{M}_n(J)$ is linear, and allows to build a source dictionnary asymptotically Gaussian.

The matrix $\boldsymbol{G}_d$ can be constructed as any combination of rows of the matrix $\boldsymbol{M}_n$. Similar properties are achieved when selecting columns of the matrix $\boldsymbol{M}_n$ and we shall denote by $\boldsymbol{G}'_r$ the $n \times r$ $(r \leq n)$ matrices constructed this way.

However, the distribution of the video sources is not Gaussian and a direct application of the above vector quantization is not appropriate. Significant modifications have been made. Thus, in order to take into account the non-Gaussian source distribution we classify the coefficients in each subband in two classes and we adapt the quantizer to each class. This classification of vectors of wavelet coefficients in the detail frames is based on a stochastic model of the spatio-temporal dependencies between the wavelet coefficients, which we have introduced in [4]. The vectors of the approximation frames are classified according to their norm. We find the source codebook for each class of vectors by minimizing the following expression:

$$\min_{\boldsymbol{b}} \mathrm{E}||\boldsymbol{x} - \beta \boldsymbol{G}_d \boldsymbol{b}||^2 \tag{1}$$

where $\boldsymbol{b} = (b_1,...,b_n)^t \in BPSK_n$, $\boldsymbol{G}_d$ is the matrix obtained as explained above and β is a parameter which scales the lattice constellation to the source dynamics. In order to find the parameter β and the codebook with vectors $\boldsymbol{y} = \boldsymbol{G}_d\boldsymbol{b}$, an iterative optimization algorithm is applied. A similar optimization is applied when using the matrix $\boldsymbol{G}'_r$ for vector quantization.

3. BIT ALLOCATION ALGORITHM

As we have, already, presented due to the linear labelling the channel distortion D_c is minimized. In the case of hard decision detection, the Gaussian channel with binary inputs is transformed into a binary symmetric channel with probability $Q(1/\sigma_b)$, where σ_b^2 is the variance of the Gaussian noise. Thus, the channel distortion is:

$$D_c = 4 \cdot Q(1/\sigma_b) \cdot \sigma_{VQ}^2$$

where σ_{VQ}^2 is the variance of the quantized source. We checked through simulations that even at low bitrates σ_{VQ}^2 is very well approximated by the variance of the unquantized source.

Consider that a GOF of a video sequence is decomposed into I spatio-temporal subbands. Let N be the total number of coefficients in the GOF and n_j the number of coefficients of the subband j, $1 \leq j \leq I$. σ_j^2 is the variance of the subband j. We suppose that all the coefficients in the same subband are quantized by the same number of bits, thus, let r_j be the bits per coefficient in the subband j.

If we assume that the overload distortion of the quantization is negligible and under the assumption of high resolution approximation, then an approximated model of the quantization error and in consequence of the distortion of a subband j can be expressed as:

$$D_j \approx \sigma_j^2 2^{-2r_i}$$

As the channel distortion is minimized due to the linear labelling, we consider that the distortion that has to be minimized subject to a global bitrate $R = \sum_{j=1}^{I} \frac{r_j n_j}{N}$ is given by:

$$\min_R D = \sum_{j=1}^{I} D_j = \sum_{j=1}^{I} \sigma_j 2^{-2r_i}$$

Corresponding to the above constrained minimization problem, an unconstrained minimization problem using Lagrange multipliers can be stated as:

$$J_\lambda = D + \lambda(\sum_{j=1}^{I} \frac{r_j n_j}{N} - R)$$

However, the direct application of this classical method can lead to very bad results, especially at low bitrates. Indeed, neglecting the practical requirement that $r_j \geq 0$ the "optimal" solution could allow negative values of r_j. Our proposed algorithm avoids the drawback of the negative solutions and changes the criterion in order to take into account the priority of the solution.

Our bit allocation algorithm is based on the following steps:

1. Order the subbands by decreasing variance: $+\infty \geq \sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_I^2 \geq 0$
2. Initialize the index of iterations: $l = 1$
3. Calculate
$$\lambda_l = -2 \ln 2 \; 2^{-\frac{2NR}{M_l}} \prod_{k=1}^{l} (\sigma_k^2)^{\frac{n_k}{M_l}}$$
where $M_l = \sum_{i=1}^{l} n_i$
4. If λ_l satisfies the following inequality:
$$-2 \ln 2 \; \sigma_l^2 < \lambda_l \leq -2 \ln 2 \; \sigma_{l+1}^2$$
then set $l = l + 1$ and go to step 3, else exit.

By this way this algorithm will allocate bits only to the subbands whose variance is smaller than the variance placed at the $l + 1$ position of the order in the first step. The final result is given by:

$$r_k = \begin{cases} \frac{N}{M_l} R + \frac{1}{2} \log_2 \left(\frac{\sigma_k^2}{\prod_{k=1}^{l} (\sigma_k^2)^{\frac{n_k}{M_l}}} \right), & k \in \{1, \cdots, l\} \\ 0, & k \in \{l+1, \cdots, I\} \end{cases}$$

4. SIMULATION RESULTS

We consider a temporal Haar decomposition applied on GOFs of 16 frames, with 4 temporal and 2 spatial resolution levels. Two cases of temporal decomposition are considered. Motion-compensated, with full search block matching algorithm and full pel accurancy, and no motion estimation. The spatial multiresolution analysis is based on the biorthogonal $9/7$ filters.

The bit allocation algorithm presented in the previous section indicates the size of the $\boldsymbol{G}_d$ or $\boldsymbol{G}'_r$ which minimizes the end-to-end distortion. The choices of the $\boldsymbol{G}_d$ or $\boldsymbol{G}'_r$ are, however, limited by the complexity and the dependences of the spatio-temporal coefficients. It is known that the spatio-temporal coefficients exhibit strong relations with their spatial or spatio-temporal neighbors, thus, in order to capture these relations, the dimensions d in $\boldsymbol{G}_d$ or n in $\boldsymbol{G}'_r$ should be even. In addition, in order to keep the complexity low, we have limited the dimensions n in $\boldsymbol{G}_d$ and r in $\boldsymbol{G}'_r$ to 16.

For our tests we considered CIF (352×288) test sequences at 30 fps.

The global bitrates per pixel tested are : $0.1bpp$, $0.16bpp$, $0.33bpp$ and $0.48bpp$. However, due to the additional bits sended in order to indicate the class where the vectors of the approximation frames belong to, and due to the coding of the motion vectors the global bitrate can not be the same for two different GOFs or sequences.

In Table 1 and Table 2 we present the average PSNR of two test sequences ("hall-monitor" and "foreman") on a Gaussian channel under different noise states and under different bitrates. The decoder uses in all cases a hard decision criterion.

Fig. 1 illustrates the reconstructed frames of "hall-monitor" at 566.53 Kbs without motion estimation and of "foreman" at 1104 Kbs with motion estimation, first in a noiseless environment and then after transmission over a Gaussian channel with SNR=6.75 dB and SNR=8.0 dB.

We can notice that our joint source-channel scheme globally presents a good robustness to noise. Moreover, at SNR=8.0 dB the reconstruction quality already approaches the noiseless case. Note also that no additional protection by error-correcting codes is applied. However, for low SNR the scheme could benefit from the use of a linear block code. One can remark the gracefull degradation with the noise level, due to the efficient allocation, and also with the bitrate, due to the scalability of the scheme.

5. CONCLUSION

This paper presented a joint source-channel coding approach of $t + 2D$ decomposed video sequences. An iterative bit allocation algorithm taking into account the nonnegativity constraint was proved efficient to the optimal distribution of the available bits among the spatio-temporal subbands. It was shown that good video quality can be obtained over Gaussian channels at low channel SNR for different bitrates.

6. REFERENCES

[1] P. Knagenhjelm and E. Agrell, "The Hadamard Transform - A Tool for Index Assignement", IEEE Trans. on Information Theory, vol. IT 42, pp 1139-1151, July 1996.

Fig. 1. Reconstructed Frames. First Line: “hall-monitor” at 566.53 Kbs without motion estimation. Second Line: “foreman” at 1104.1 Kbs with motion estimation. Left: reconstructed frame in a noiseless environment. Center: reconstructed frame over a Gaussian channel with SNR=6.75 dB. Right: reconstructed frame over a Gaussian channel with SNR=8.0 dB.

“hall-monitor”				
bitrate(Kbs)	295.5	566.53	1027.7	1550.0
noiseless	26.68	30.53	32.81	34.66
SNR=4.33	23.44	26.01	26.57	26.90
SNR=6.75	26.21	29.75	31.54	32.84
SNR=8.0	26.58	30.38	32.54	34.25
“foreman”				
bitrate(Kbs)	307.4	576.92	983.07	1410.0
noiseless	25.07	27.56	28.81	30.14
SNR=4.33	22.98	24.71	25.04	25.44
SNR=6.75	24.86	27.13	28.22	29.36
SNR=8.0	25.06	27.49	28.66	29.98

Table 1. Average PSNR of the “hall-monitor” and “foreman” CIF sequences, without motion estimation.

“hall-monitor”				
bitrate(Kbs)	317.05	565.04	1113.7	1660.0
noiseless	27.88	31.78	34.38	35.59
SNR=4.33	23.91	26.41	26.96	27.23
SNR=6.75	27.22	30.81	32.68	33.47
SNR=8.0	27.76	31.58	34.05	35.09
“foreman”				
bitrate(Kbs)	376.48	524.20	1104.1	1530.0
noiseless	30.07	30.78	33.30	34.51
SNR=4.33	25.53	25.66	26.27	26.49
SNR=6.75	29.27	29.86	31.72	32.63
SNR=8.0	29.91	30.58	32.98	34.05

Table 2. Average PSNR of the “hall-monitor” and “foreman” CIF sequences, with motion estimation/compensation.

[2] J. C. Belfiore, X. Giraud and J. Rodriguez-Guisantes, “Optimal Linear Labelling for the Minimisation of both Source and Channel distortion”, ISIT 2000, Sorrento, Italy, June 25-30, 2000.

[3] X. Giraud, E. Boutillon and J.-C. Belfiore, “Algebraic Tools to Build Modulation Schemes for Fading Channels” IEEE Trans. on Information Theory, vol. IT 43, pp.938-952, May 1997.

[4] G. Feideropoulou, B. Pesquet-Popescu, J. C. Belfiore and J. Rodriguez,“Non-linear modeling of wavelet coefficients for a video sequence”, Proc of IEEE Int. Workshop NSIP03, Grado, Italy, June 8-11, 2003.

[5] A. Segall, “Bit Allocation and Encoding for Vector Sources”, IEEE Trans. on Information Theory, vol. IT 22, no. 2, pp.162-169, Mar. 1976.

[6] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers”, IEEE Trans. Acoust. Speech Signal. Proc., vol. 36, no. 9, pp: 1445-1453, Sept. 1988.