MODELING THE MOTION CODING ERROR FOR MCWT VIDEO CODERS

Marie Andrée Agostini, Thomas André, Marc Antonini and Michel Barlaud

CReATIVe Research Group - I3S Laboratory, CNRS - University of Nice-Sophia Antipolis
2000 route des Lucioles - BP121 - 06903 Sophia Antipolis Cedex - France
{agostini, andret, am, barlaud}@i3s.unice.fr

ABSTRACT

In motion-compensated wavelet based video coders, a very precise motion estimation is necessary. However, a motion vectors field of high precision is expensive in binary resources and requires a great place in the bitstream compared to the wavelet coefficients. Thus, we need to reduce the cost of the motion information. To this end, we propose an approach based on a scalable lossy coding of high-precision motion vectors. It allows to optimize the trade-off between motion bit-rate and wavelet coefficients bit-rate, strongly reduces the motion cost, and thus, increases the coder performances at low bit-rate.

Obviously, this lossy motion coding has an impact on the decoded sequence. In this paper, we evaluate this impact by establishing a theoretical distortion model for the motion coding error. This model will allow to realize an optimal model-based bit-rate allocation between wavelet subbands and motion vectors. The experimental validation of the model gives satisfactory results.

1. INTRODUCTION

In front of the explosion of the quantity and quality of visual data, video compression has been for a few years in full expansion. Recently, the algorithms performances have considerably increased, in particular with the latest standards, like MPEG-4 [1] or H.264 [2].

Recent works showed that $t + 2D$ wavelet based video coders [3], with motion-compensated temporal lifting scheme [4], allow good scalability support [5] and almost reach the performances of the hybrid coders [3, 6]. In order to further increase the video coding efficiency at low bit-rate, it is necessary to improve the motion vectors processing. This crucial problem was explored in [7], [8]. Today, most of video coders optimize the rate-distortion trade-off for a given rate, by varying for example the size of the blocks or the precision of the estimates. But these methods are not well-adapted to low rates, because the motion information becomes proportionally more significant, and, besides, the scalability is difficult to obtain for the motion.

In a recent work [9], we have shown that encoding in a scalable way precise motion information with controlled losses allows to reduce the motion cost with good coding performances at low bit-rate. Moreover, it is possible to realize an optimal bit-rate allocation between the motion information and the wavelet coefficients. The approach of [9] consists in estimating motion vectors with a very high precision, and then in quantizing them by optimizing a rate / distortion criterion.

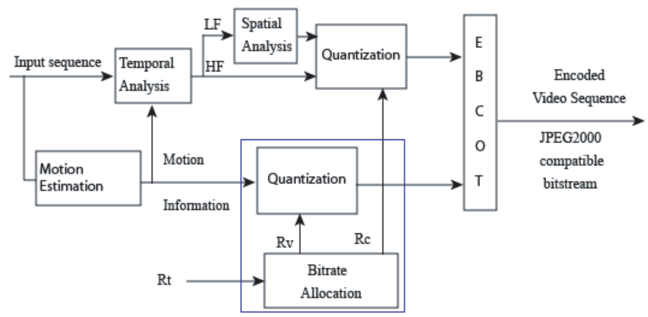

Fig. 1. General structure of the encoder (R_c is the bit-rate of the subbands, R_v the one of the motion vectors and R_t the total bit-rate).

Obviously, the introduction of the loss on the motion have an impact on the decoded sequence. In this paper, we propose to evaluate analytically this impact. To this purpose, we establish a theoretical distortion model for the motion coding error. Indeed, if we are able to model analytically the distortion of the reconstructed sequence as a fucntion of the quantization distortion introduced on the motion vectors and the wavelet coefficients, it will be possible to realize a model-based bit-rate allocation, with the main objective of optimizing the trade-off between motion vectors bit-rate and wavelet subbands bit-rate.

The general principle of our coder and the motion coding approach are presented in section 2. The proposed model of motion coding error is described in section 3. Finally, experimental validation results are presented in section 4. Conclusion and further works are presented in section 5.

2. LIFTED MCWT VIDEO CODING

We present in this section the motion-compensated wavelet transform (MCWT) coder on which our work is based, and we focus on the motion vectors coding method.

2.1. General principle

Fully scalable, our video encoder is based on a lifted motion-compensated wavelet transform. The general structure is described in the figure 1.

Motion compensation is essential for an efficient decorre-

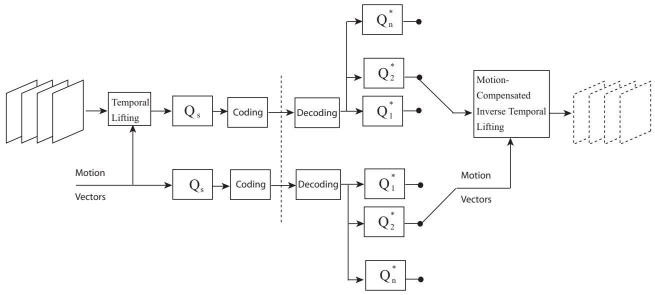

Fig. 2. Open-loop coding of motion vectors: the wavelet subbands and the vectors are scalable. Motion bit-rate can thus be perfectly adapted to the subbands bit-rate.

lation of the video sequences [3] and motion estimation is a crucial problem in video compression. Nevertheless, a vectors field of good quality can be very expensive in binary resources compared to the wavelet coefficients.

2.2. Motion information coding

The cost of the motion vectors can be very significant, which is not desirable, especially at low rate. The challenge is thus to reduce this cost. For this purpose, the proposed method uses a precise motion estimator and quantizes with loss the vectors in order to reduce their cost, while controlling the rate-distortion trade-off on the reconstructed sequence.

In order to remain fully scalable and to preserve the quality of the motion-compensated temporal filtering, we have to encode the motion vectors in open-loop [9]: full-precision vectors are thus used for motion compensation at the coder (figure 2). Motion vectors are quantized using an uniform scalar quantizer whose quantization step q controls the motion rate-distortion trade-off. Then, the quantized vectors are encoded using the MQ-Coder of an EBCOT encoder [6] and embedded in a JPEG2000-compliant bitstream.

At the decoder side, the bitstream is decoded using EBCOT and the quantized decoded vectors are rescaled by the quantization step q. Then, the motion compensation and the inverse temporal wavelet transform are performed using the quantized decoded vectors.

2.3. Bit-rate allocation algorithm

The quantization step q controls the quantized motion vectors rate R_v and the distortion introduced by the motion coding. Moreover, the subbands quantization at rate R_c also introduces an error on the reconstructed sequence.

In a general way, coding is optimal if the rate-distortion points are located on the convex hull of the rate-distortion curve. For this purpose, in our approach, we search empirically the best rate R_v, i.e. the best quantization step q, which minimizes the reconstruction Mean Square Error (MSE) for a desired total rate R_t.

The rates distribution between the motion vectors and the wavelet coefficients is thus done in an optimal way by the following algorithm (the input-output distortion D is measured during optimization):

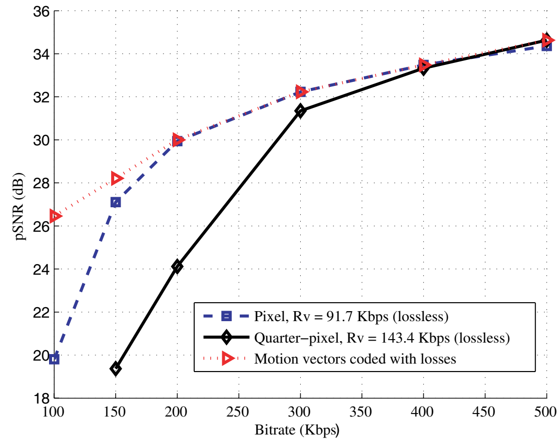

Fig. 3. Performance comparison on the sequence “Foreman”.

Step 0: R_t is given (total bit-rate)
Step 1: Find R_v^* (by varying q) such that:

$$R_v^* = \underset{R_v \; and \; R_c \,=\, R_t \,-\, R_v}{\arg\min} D(R_v)$$

Step 2: Compute $R_c^* = R_t - R_v^*$
End

2.4. Efficiency of the method

This approach presents interesting performances on CIF and on SD sequences. Figure 3 shows the results in terms of Peak Signal Ratio (pSNR) as a function of the total rate R_t for the first 144 images of the sequence “Foreman”, with three temporal decomposition levels using the (2,0) lifting scheme (equivalent to a truncated 5/3 lifting scheme). The curve with the triangular markers is obtained by applying the algorithm proposed in section 2.3 for a quarter-pixel motion estimator. We also present the curves obtained with a lossless coding of the vectors estimated at pixel and quarter-pixel precisions. At low rate, these curves show that quantizing motion vectors of high precision allows to obtain better performances than lossless motion coding.

3. MODELING OF THE MOTION VECTORS QUANTIZATION NOISE

We have established a theoretical distortion model, for one decomposition level, which describes the impact of the motion vectors coding on the decoded sequence.

The motion quantization noise model is based on the input-output distortion computation (MSE between signals x and $\widetilde{x}$). Its expression is given by:

$$\begin{aligned} D &= \frac{1}{K}\frac{1}{NM}\sum_{k=0}^{\frac{K}{2}-1}\sum_{\mathbf{p}}[(x_{2k}(\mathbf{p}) - \widetilde{x}_{2k}(\mathbf{p}))^2 \\ &+ (x_{2k+1}(\mathbf{p}) - \widetilde{x}_{2k+1}(\mathbf{p}))^2] \end{aligned} \quad (1)$$

where N is the number of rows and M the number of columns for one image of the sequence, and K the size of the sequence. Introducing the signal power,

$$\mathbf{Pn}\left(x_k\right) = \frac{1}{NM}\sum_{\mathbf{p}} x_k^2\left(\mathbf{p}\right)$$

and assuming that for one temporal wavelet decomposition level, the first part of equation (1) is equal to zero for a (2,0) lifting scheme, the distortion D can be written as:

$$D = \frac{1}{K}\sum_{k=0}^{\frac{K}{2}-1} \mathbf{Pn}\left(x_{2k+1} - \widetilde{x}_{2k+1}\right). \tag{2}$$

Let us establish some notations: in the following, we denote by $\widehat{B}_k$ the "Backward" quantized motion and $\widehat{F}_k$ the "Forward" quantized motion. We have: $\widehat{B}_k = Q\left(\mathbf{v}_{k\to k-1}\right) = \widehat{\mathbf{v}}_{k\to k-1}$ and $\widehat{F}_k = Q\left(\mathbf{v}_{k\to k+1}\right) = \widehat{\mathbf{v}}_{k\to k+1}$, with $Q\left(.\right)$ the quantization operator.

If we denote respectively $x_k^{B_{k+1}} = x_k\left(\mathbf{p} + B_{k+1}\left(\mathbf{p}\right)\right)$ and $x_k^{F_{k-1}} = x_k\left(\mathbf{p} + F_{k-1}\left(\mathbf{p}\right)\right)$ the "Backward" and "Forward" motion-compensated pixels, the motion-compensated pixels with the quantized motion vectors can be written as:

$$\widetilde{x}_k^{\widehat{B}_{k+1}} = \widetilde{x}_k(\mathbf{p}+\widehat{B}_{k+1}(\mathbf{p})) \text{ and } \widetilde{x}_k^{\widehat{F}_{k-1}} = \widetilde{x}_k(\mathbf{p}+\widehat{F}_{k-1}\left(\mathbf{p}\right)).$$

Let us remind the (2,0) lifting scheme analysis equations on one decomposition level:

$$\left\{\begin{array}{l} h_k\left(\mathbf{p}\right) = x_{2k+1}\left(\mathbf{p}\right) - \frac{1}{2}(x_{2k}^{B_{2k+1}} + x_{2k+2}^{F_{2k+1}}) \\ l_k\left(\mathbf{p}\right) = x_{2k}\left(\mathbf{p}\right) \end{array}\right. \tag{3}$$

With quantized motion vectors, the synthesis equations are given by:

$$\left\{\begin{array}{l} x_{2k}\left(\mathbf{p}\right) = l_k\left(\mathbf{p}\right) \\ \widetilde{x}_{2k+1}\left(\mathbf{p}\right) = h_k\left(\mathbf{p}\right) + \frac{1}{2}(\widetilde{x}_{2k}^{\widehat{B}_{2k+1}} + \widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}) \end{array}\right. \tag{4}$$

where $\widetilde{x}_{2k+1}\left(\mathbf{p}\right)$ represents the synthesized pixel with the quantized motion vectors.
By using the first equation of system (3) and the second equation of system (4), and because the quantizer works in open-loop, we can write the following relation:

$$\begin{aligned} x_{2k+1}\left(\mathbf{p}\right) - \widetilde{x}_{2k+1}\left(\mathbf{p}\right) = {} & \frac{1}{2}(x_{2k}^{B_{2k+1}} - \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}) \\ & + \frac{1}{2}(x_{2k+2}^{F_{2k+1}} - \widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}) \end{aligned} \tag{5}$$

Let us assume that the reconstruction errors "Backward" $\epsilon_B = x_{2k}^{B_{2k+1}} - \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}$ and "Forward" $\epsilon_F = x_{2k+2}^{F_{2k+1}} - \widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}$ due to the motion quantization are decorrelated.
By combining the equations (2) and (5), we can write:

$$\begin{aligned} D = {} & \frac{1}{4K}\sum_{k=0}^{\frac{K}{2}-1} (\mathbf{Pn}(x_{2k}^{B_{2k+1}} - \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}) \\ & + \mathbf{Pn}(x_{2k+2}^{F_{2k+1}} - \widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}})) \end{aligned}$$

Or equivalently,

$$\begin{aligned} D = {} & \frac{1}{K}\sum_{k=0}^{\frac{K}{2}-1} [\frac{1}{4}\mathbf{Pn}(x_{2k}^{B_{2k+1}}) + \frac{1}{4}\mathbf{Pn}(\widetilde{x}_{2k}^{\widehat{B}_{2k+1}}) \\ & - \frac{1}{2}\left\langle x_{2k}^{B_{2k+1}}, \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}\right\rangle + \frac{1}{4}\mathbf{Pn}(x_{2k+2}^{F_{2k+1}}) \\ & + \frac{1}{4}\mathbf{Pn}(\widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}) - \frac{1}{2}\left\langle x_{2k+2}^{F_{2k+1}}, \widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}\right\rangle] \end{aligned}$$

with the scalar products for the "Backward" vectors defined by (with similar notation for the "Forward" vectors):

$$\left\langle x_{2k}^{B_{2k+1}}, \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}\right\rangle = \frac{1}{NM}\sum_{\mathbf{p}} x_{2k}^{B_{2k+1}} \times \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}$$

Assuming that the image is stationary at time k, the scalar product becomes: $\left\langle x_{2k}^{B_{2k+1}}, \widetilde{x}_{2k}^{\widehat{B}_{2k+1}}\right\rangle = \Gamma_{x_{2k}}(\eta_{B_{2k+1}})$, with $\Gamma_{x_{2k}}$ the autocorrelation function of the signal x_{2k} and $\eta_{B_{2k+1}} = B_{2k+1} - \widehat{B}_{2k+1}$ the quantization error on the "Backward" motion vectors (similar notations for the "Forward" motion).
If the motion quantization errors $\eta_{B_{2k+1}}$ and $\eta_{F_{2k+1}}$ are small (asymptotical hypothesis or equivalently high bit-rate), we have:

$$\mathbf{Pn}(x_{2k}^{B_{2k+1}}) \approx \mathbf{Pn}(\widetilde{x}_{2k}^{\widehat{B}_{2k+1}}) \approx \mathbf{Pn}\left(x_{2k}\right)$$
$$\mathbf{Pn}(x_{2k+2}^{F_{2k+1}}) \approx \mathbf{Pn}(\widetilde{x}_{2k+2}^{\widehat{F}_{2k+1}}) \approx \mathbf{Pn}\left(x_{2k+2}\right).$$

Finally, the distortion D, which expresses the reconstruction error due to the motion quantization, can be simplified as:

$$\begin{aligned} D = {} & \frac{1}{2K}\sum_{k=0}^{\frac{K}{2}-1} [\mathbf{Pn}\left(x_{2k}\right) - \Gamma_{x_{2k}}(\eta_{B_{2k+1}}) \\ & + \mathbf{Pn}\left(x_{2k+2}\right) - \Gamma_{x_{2k+2}}(\eta_{F_{2k+1}})] \end{aligned}$$

Moreover, if we suppose that the sequence is stationary inside one GOP and that the "Backward" and "Forward" motion vectors are estimated symmetrically [10] ($B_{2k+1} = -F_{2k+1}$), we can write:

$$\mathbf{Pn}\left(x_{2k}\right) \approx \mathbf{Pn}\left(x_{2k+2}\right)$$
$$\Gamma_{x_{2k}}(\eta_{B_{2k+1}}) \approx \Gamma_{x_{2k+2}}(\eta_{F_{2k+1}}).$$

The distortion can therefore be simplified in:

$$D \approx \frac{1}{K}\sum_{k=0}^{\frac{K}{2}-1} [\mathbf{Pn}\left(x_{2k}\right) - \Gamma_{x_{2k}}(\eta_{B_{2k+1}})] \tag{6}$$

$$\approx \frac{1}{K}\sum_{k=0}^{\frac{K}{2}-1} [\mathbf{Pn}\left(x_{2k+2}\right) - \Gamma_{x_{2k+2}}(\eta_{F_{2k+1}})] \tag{7}$$

These equations mean that the knowledge of images x_{2k} and x_{2k+2} over a GOP allows to estimate the distortion introduced by the motion quantization. Indeed, this distortion is simply function of the x_{2k} (or x_{2k+2}) images power and of the x_{2k} (or x_{2k+2}) images autocorrelation function (which depends of the motion quantization errors η).

4. EXPERIMENTAL VALIDATION

To validate the proposed model, we compare the results obtained experimentally for the input-output distortion with

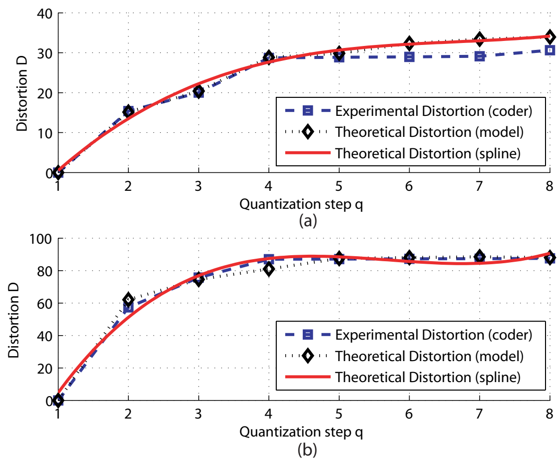

Fig. 4. Validation of the distortion theoretical model on the sequences "Foreman" (a) and "City" (b).

q	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Foreman	1.7	1.8	0.7	3.3	10.4	11	10.8
Rv (Kbps)	46.7	33.3	32	25.9	22.7	22.5	22.3
City	7.6	1.6	6.3	1.4	1.4	1.6	0.7
Rv (Kbps)	67.1	48.9	47.9	38	31.5	26.2	23.4

Table 1. Errors (in %) between the theoretical distortion model and the experimentation for the sequences "Foreman" and "City" for each motion quantization step q, with the corresponding motion bit-rates Rv in Kbps.

those obtained by applying the theoretical distortion formulas (6) or (7). In figure 4, we present the results for the sequences "Foreman" (a) (144 first images) and "City" (b) (48 first images) on one decomposition level with a (2,0) lifting scheme, where the pixelic motion vectors are coded with different quantization steps q and where the wavelet subbands are coded losslessly. These curves represent the distortion D as a function of the quantization step q. Moreover, we use Smoothing-B splines to model analytically the theoretical curves. The errors in percentage between theory and experimentation for each q are summarized in table 1.

These results show that, for these two sequences, the theoretical and experimental curves are very close and follow the same progression. We observe no more than 5 % of error on average. Therefore, the theoretical distortion model for the motion vectors coding error provides a good approximation.

5. CONCLUSION

We have presented in this paper a distortion model of motion coding error for video coders based on temporal motion-compensated lifting scheme.

Indeed, we have already shown that it is necessary to optimize the rate-distortion trade-off between motion information and wavelet coefficients. To this end, we quantize with losses motion vectors estimated with a high subpixelic precision. This approach allows to improve the quality of the reconstructed video sequence at low bit-rate.

We proposed a theoretical model to evaluate the impact of the losses introduced by the motion information coding on the decoded signal. Experimental validation is satisfactory. The generalization to several temporal decomposition levels and the introduction of the wavelet subbands quantization noise will be included in a future work.

Our final objective is to realize an optimal model-based bit-rate allocation between the wavelet subbands and the motion information: for a desired total bit-rate R_t, it will be possible to find analytically which optimal rates to choose for the motion vectors and the wavelet coefficients in order to have a minimal distortion at decoding. This will decrease the computational complexity of the bit-rate allocation at coding and obviously improve the whole coder performances.

6. REFERENCES

[1] MPEG4 Video Group, *Coding of audio-visual objects: Video, ISO/IEC JTC1/SC29/WG11 N2202*, Mar. 1998.

[2] Gary J. Sullivan, Pankaj Topiwala, and Ajay Luthra, "The h264 / avc advanced video coding standard : Overview and introduction to the fidelity range extensions," in *SPIE Conference on Applications of Digital Image Processing XXVII, special Session on Advances in the New Emerging Standard : H264 / AVC*, 2004.

[3] J-R. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. On Image Processing*, vol. 3, no. 5, November 1994.

[4] J. Viéron, C. Guillemot, and S. Pateux, "Motion compensated 2D+t wavelet analysis for low rate fgs video compression," in *Proc. of Tyrrhenian Intern. Workshop on Digital Comm.*, Capri, Italy, 2002.

[5] G. Pau, C. Tillier, B. Pesquet-Popescu, and H. Heijmans, "Motion compensation and scalability in lifting-based video coding," *EURASIP Signal Processing: Image Communication, special issue on Wavelet Video Coding*, pp. 577–600, 2004.

[6] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Transaction on Image Processing*, vol. 12, no. 12, pp. 1530–1542, Dec. 2003.

[7] L. A. Da Silva Cruz and J. W. Woods, "Adaptive motion vector quantization for video coding," in *IEEE Intern. Conf. on Image Processing*, Vancouver, Canada, 2000, pp. 867–870, vol 2.

[8] R. L. Joshi, T. R. Fischer, and R. H. Bamberger, "Lossy encoding of motion vectors using entropy-constrained vector quantization," in *IEEE Intern. Conf. on Image Processing*, Washington, USA, 1995, pp. 3109, vol 3.

[9] M. A. Agostini, T. André, M. Antonini, and M. Barlaud, "Scalable motion coding for video coders with lifted MCWT," in *Proc. International Workshop on Very Low Bit-rate Video-coding (VLBV)*, Costa Rei, Sardinia - Italy, 2005.

[10] T. André, M. Cagnazzo, M. Antonini, M. Barlaud, N. Božinović, and J. Konrad, "(N,0) motion-compensated lifting-based wavelet transform," in *Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, Montreal, Canada, May 2004.