A RATE CONTROL STRATEGY FOR EMBEDDED WAVELET VIDEO CODERS IN AN MPEG-4 FRAMEWORK

Eduardo A. B. da Silva *Rogério Caetano*

PEE/COPPE/DEL/EE
Universidade Federal do Rio de Janeiro
Cx. P. 68504, Rio de Janeiro, RJ
21945-970, BRAZIL
eduardo@lps.ufrj.br caetano@lps.ufrj.br

ABSTRACT

Embedded wavelet encoders possess a number of interesting features for video encoding applications. Among those, we can mention the ability to encode a picture with precise control over the bit-rate. This differs significantly from the DCT-based methods adopted in the MPEG-4 standard, in which the rate cannot be directly set, being controlled by the quantizer step size instead. In this paper we investigate rate-control strategies for embedded wavelet encoders in an MPEG-4 framework, that take advantage of the precise control that can be obtained for the bit-rate in such coders. We also investigate the use of successive approximation vector quantization in them for replacing the traditionally used successive approximation scalar quantization. The results are encouraging, showing that both the use of vector quantization and the adoption of an adequate rate-control strategy can improve both objective and subjective quality of video sequences encoded using embedded wavelet encoders.

1. INTRODUCTION

Embedded wavelet coders have been very successful in coding still images. They are part of a number of standards, including JPEG-2000 [1] and the texture coding part of MPEG-4 [2]. Nevertheless, for interframe video encoding, the classical DCT-based methods are in general preferred. However, considering the excellent performance that embedded wavelet coders have for still images, it is worthy to investigate their behavior in video coding. This assumes particular importance if one considers that MPEG-4 syntax allows the specification of alternative encoders for particular applications via its Systems Description Language (MSDL). In this paper, we investigate aspects related to the use of such encoders as replacements for the DCT encoding of the motion compensated frame difference in an VM-8 encoder [2] .

Among the most well known embedded wavelet encoders, one can cite the EZW [3] and the SPIHT [4] encoders. One of their main advantages is the absence of blocking effects due to the use of wavelet transforms instead of the DCT. Despite that, EZW-like interframe encoders failed to give signal-to-noise ratio improvements over DCT-based ones during MPEG-4 tests.

Another advantage of such coders is that they have precise control over the bit-rate of every frame. In DCT-base methods the rate can only be indirectly controlled by the quantizer step-size, that is, a larger step-size implies a smaller signal-to-noise ratio and vice versa.

For example, in MPEG-4 VM-8, the bit-rate control strategy assumes a quadratic relation between the quantizer step size and the rate spent in transmitting a frame. It is updated by a least-squares fitting as each new frame is encoded. Given the number of bits remaining to encode the sequence, one updates the quantizer step size according to this relation. Although efficient, this is based on a number of ad-hoc assumptions and is clearly suboptimal in a rate-distortion sense.

This paper investigates rate-allocation strategies for embedded wavelet encoders. Unlike previous works [5], which assume that the R-D curve for a single frame difference is exponential, we have shown that the best model for the R-D of an embedded wavelet encoder is piecewise linear. This enables us to perform rate-allocation using Lagrangian optimization at low computational complexity. In addition, we propose an iterative procedure to deal with the problem of frame dependency. We also investigate the use of the successive approximation vector quantization (SA-W-VQ) algorithm [6] instead of EZW. We show that the proposed improvements generate an embedded wavelet video encoder that outperforms the MPEG-4 VM-8 by a comfortable margin. The rest of the paper is divided as

0-7803-5796-5/99/$10.00 © 1999 IEEE

follows: section 2 describes the modifications to the MPEG-4 VM-8 necessary to use EZW and SA-W-VQ instead of the DCT. Section 3 analyses the rate-control problem, and proposes the new rate-control strategy. Section 4 presents the experimental results and section 5 states the conclusions.

2. MODIFICATIONS TO MPEG-4 VM-8

Wavelet encoders

In this implementation we have employed two types of embedded wavelet encoders. The EZW [3], which is based on successive approximation *scalar* quantization, and the SA-W-VQ, which is based on successive approximation *vector* quantization [6]. In both coders, we have used a 2-stage biorthogonal wavelet transform as recommended in [7]. In the SA-W-VQ case, the vectors were formed by dividing the wavelet coefficients into 2×2, 4×2 or 4×4 blocks. The orientation codebooks used were the first shells of the root lattices D_4, E_8 and Λ_{16}, related to the solution of the sphere packing problem [8] at dimensions 4, 8 and 16, respectively.

It is important to point out that both schemes (EZW and SA-W-VQ) are built on the EZW framework. It is true that we could have used a more sophisticated framework, like the one of the SPIHT coder [4], which would certainly improve both the scalar and vector quantization results [9]. However, this is of secondary importance, since the main emphasis of the paper is on the rate-control strategy. This will be confirmed in sections 3 and 4, where it is shown that the effectiveness of the proposed rate-control strategy is independent of the specific type of embedded wavelet encoder used.

Other modifications

The video encoder used in this work is based on the MPEG-4 VM-8. The main difference is that the DCT, quantization and run-length encoding followed by Huffman coding have been replaced by an embedded wavelet encoder followed by an adaptive arithmetic coder. The respective inverse operations have also been correspondingly replaced. These requires some further modifications to the MPEG-4 VM-8, which will be described below. The modifications to the rate-control scheme will be dealt with in section 3.

One modification has to do with the fact that a wavelet transform is no longer applied to independent image blocks, but to the image as a whole. The image is divided in macroblocks only for the purposes of motion estimation and compensation. In a coded frame, either all of its macroblocks are intra-frame or all of its macroblocks are inter-frame (restricted to forward prediction - P frames). We have also chosen the option to consider each frame composed of only one rectangular VOP, identical to the frame itself.

Another important point is that it is necessary to always turn on the overlapped motion compensation option. The reason for this lies in the fact that it is very undesirable to have a "blocky" interframe error when a wavelet transform is applied to it. The discontinuities in the block boundaries of such an error signal generate high frequency coefficients having high energy, which greatly reduce the performance of wavelet encoders [7]. Since, with overlapped motion compensation, the prediction of a macroblock uses motion information from neighboring blocks, the motion field is smoother, and the prediction error is almost deprived of blockiness. It is important to point out that with overlapped motion compensation turned off, the performance of any interframe wavelet encoder in an MPEG-4 framework would be poor.

Since one of the goals of this work has been to investigate the performance of embedded wavelet encoders for interframe encoding, the first frame of a sequence was never encoded in the simulations; in this way, any difference in performance of the encoders is related only to the interframe encoder.

3. RATE CONTROL

Embedded wavelet encoders like EZW and SA-W-VQ have the capability to set up precisely the bit-rate of each frame. In constant bit-rate (CBR) systems, if one looks at this problem purely from the rate-control point of view, the trivial solution would be to divide equally the bit-rate among all frames. This solution would in principle have the advantage of obviating the need of a buffer in order to smooth out bit-rate variations. However, very often one needs to find out how many bits should be allocated to each frame in order to obtain the highest average signal-to-noise ratio (or, alternatively, subjective image quality) of the entire sequence for a given rate budget. This optimal bit allocation problem can be stated more formally as [10]:

Given a set of N frames $\{f_i,\ i = 1,\ldots,N\}$ and an average target rate of R_{target} bits/frame, we have to encode frame f_i with rate R_i, yielding distortion D_i such that:

$$D = \frac{1}{N}\sum_{i=1}^{N} D_i \text{ is minimum} \tag{1}$$

$$R = \frac{1}{N}\sum_{i=1}^{N} R_i \leq R_{\text{target}} \tag{2}$$

Considering that the rate-distortion (R-D) functions of each frame f_i are convex, this optimization problem can be solved via Lagrangian optimization. In this case the problem can be restated as:

$$\text{minimize } J = D + \lambda R \tag{3}$$

such that the restriction in eq. 2 holds.

The advantage of this formulation is that, if λ is fixed, the rates that minimize eq. 3, $\{R_i(\lambda),\ i = 1,\ldots,N\}$ can be found by solving only a non-restricted optimization problem, which is in general much easier to solve than the restricted optimization problem in eqs. 1 and 2. Having the rates $R_i(\lambda)$, one has just to search in the λ space for the solution corresponding to a λ_0 such that

$$\frac{1}{N}\sum_{i=1}^{N} R_i(\lambda_0) = R_{\text{target}} \tag{4}$$

The Lagrangian optimization problem in eq. 3 can be further simplified if we make the extra assumption that the R-D characteristics of frame f_i are independent of the particular point (R_j, D_j) in which frame j is being coded, $\forall j \neq i$. Then eq. 4 is equivalent to solving:

$$\text{minimize } J_i = D_i + \lambda R_i,\ i = 1,\ldots,N \tag{5}$$

If we know an analytical expression for $D_i(R_i)$, $i = 1,\ldots,N$, then the minimization problem in eq. 5 can be first solved as a function of λ, and then one finds λ such that eq. 5 is satisfied. In [5], Cheng et. al. investigated the use of an EZW-like encoder in interframe coding, and developed an optimum rate allocation strategy supposing that the R-D characteristics of each frame are exponential. However, we argue here that the exponential model is neither the best nor the simplest model for such R-D characteristics. Our argument is based on the fact that, in embedded wavelet encoders, the wavelet coefficients are transmitted by bitplanes.[1] Within a certain bitplane, for each bit sent, there is a fixed reduction in the total distortion, that is, $\Delta D/\Delta R = k$, and therefore the R-D characteristics are linear. Within the next bitplane, $\Delta D/\Delta R = k/2$, and the R-D is linear with half the slope of the previous one. From the above, one can conclude that the function $D_i(R_i)$ can be modeled as being piecewise linear instead of exponential. This can be verified in figure 1, where the R-D characteristics of the motion compensated frame difference between frames 003 and 000 of the Mother-and-daughter sequence when encoded by an SA-W-VQ encoder using E_8 as the orientation codebook and an α of 0.55. It is important to note that an exponential model would only fit to a curve as in figure 1 if the number of bitplanes is large, which seldom is the case for interframe coding, specially in low bit-rate applications.

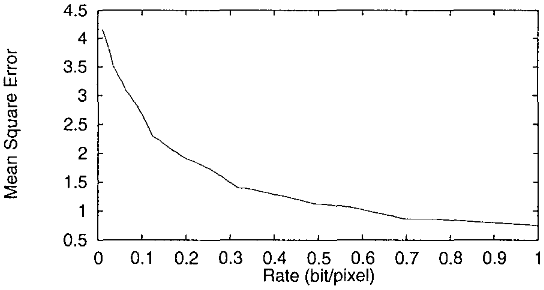

Figure 1: R-D characteristics of an interframe prediction error encoded by an embedded wavelet encoder.

An important characteristic of the behavior depicted in figure 1 is that the "breakpoints" of the piecewise linear curve lie in the boundary of the dominant and subordinate phases of a given pass [3, 6]. Therefore in order to estimate this R-D characteristic, the encoder has to just encode the frame for a certain number of passes and decode it just for the rates corresponding to the breakpoints.

Noting that to minimize J_i in eq. 5 corresponds to finding the tangent to the R-D curve of frame f_i with inclination $-\lambda$ [10], the following simple algorithm can be used to find the optimum rate allocation among N frames:

1. Given λ, find, for each frame f_i, the tangent to the R-D curve with inclination $-\lambda$; the tangency point is $(R_i(\lambda), D_i(\lambda))$;
2. Compute $R(\lambda)$ according to eq. 2;
3. If $R(\lambda) = R_{\text{target}}$ then the optimal rate allocation is given by R_i, $i = 1,\ldots,N$; else, vary λ and go to step 1.

The main difficulty with the above algorithm is that to find the value of λ which gives a rate allocation satisfying eq. 4 can be a quite time-consuming task if the number of different values of λ is large. However, since in our case the R-D characteristics of the frames are piecewise linear, we just need to execute passes 1 and 2 of the algorithm for the values of λ which correspond to the negatives of the set of slopes of all the linear segments of the R-D curves of every frame.

The frame dependency problem

A very important assumption made when deriving the rate allocation strategy above was that the R-D curves of each frame were independent of the operating point

---

[1] In the SA-W-VQ algorithm one considers the set of codevectors in a given pass as a "vector bitplane" [6].

of the others. Unfortunately, in interframe video encoders this is not the case, because the prediction error corresponding to frame f_i depends on how frame f_{i-1} has been encoded, which completely violates the above assumption. In fact, for each (R_i, D_i) point in frame f_i, there will be a different R-D *function* for frame f_{i+1}. This is the so-called *dependency problem* [10].

The strategies to solve this problem are in general extremely complex. In [11], Ortega and Ramchandran propose feasible solutions for the case when the number of points on each R-D curve is small enough. However, their solutions are unsuitable for embedded wavelet encoding schemes, that provide for each frame difference a very large number of points on the R-D curve. A different approach is the one employed in [5], where the influence of frame f_i in frame f_{i+1} is restricted to a linear dependency of the variance of frame f_{i+1} on the coding error of frame f_i.

Here, we propose a novel strategy for dealing with the frame dependency problem. In our case, we apply the rate allocation algorithm described in this section to N consecutive frames of a video sequence in several iterations. Having the reconstructed frames for iteration $n-1$, the rate allocation for iteration n is computed, and so the reconstructed frames for iteration n are obtained. This process proceeds until the change in the signal to noise ratio performance is below a threshold. One should note that one advantage of this approach is that there is no need to explicitly model the frame dependency. Simulation results presented in section 4 show that this process usually converges in less than 4 iterations.

4. EXPERIMENTAL RESULTS

The MPEG-4 VM-8 adaptation described in section 2 has been simulated using the rate control strategy proposed in section 3. We have used the test sequences Mother-and-daughter, Silent and Hall-monitor, with 300, 450 and 330 QCIF[2] frames at 30 frames/s, subsampled in time by a factor of 3 to generate 10 frames/s sequences. The bit-rate in all experiments was 64kbit/s.

Table 1 compares the average peak signal to noise ratio (PSNR) of the different MPEG-4 adaptations. DCT refers to MPEG-4 VM-8, EZW refers to the EZW encoder and Λ_{16}, E_8 and D_4 to the SA-W-VQ encoder using the referred codebooks. The columns labelled "constant rate" refer to the use of the same bitrate for every frame of the wavelet encoders, while the columns labelled "rate control" refer to the wavelet encoders employing the rate control strategy proposed in section 3, using 3 iterations of the rate-control algorithm in a group of 40 frames (GOF=40). In the DCT case, the MPEG-4 VM-8 one is used. One can see from this table that, without the proposed rate-control, the performance of the EZW-based encoder is in general inferior to the one of MPEG-4 VM-8. The use of rate-control improves somewhat its performance, but its average performance is still inferior to the one of the MPEG-4 VM-8. On the other hand, one can see that, even without rate control, the performance of the SA-W-VQ-based video encoder is superior to the one of the MPEG-4 VM-8.

	constant rate			rate control		
	Mother	Silent	Hall	Mother	Silent	Hall
DCT	-	-	-	39.26	35.77	39.30
EZW	38.25	36.04	38.49	38.70	36.49	39.31
Λ_{16}	39.61	37.52	39.80	39.90	37.56	40.07
E_8	39.46	37.32	39.82	39.71	37.46	40.15
D_4	39.22	37.09	39.79	39.48	37.32	40.11

Table 1: Comparison between the average PSNR (dB) of the different MPEG-4 versions.

Table 2 shows results for the proposed rate-control strategy for the embedded wavelet video encoders, for the images Mother-and-daughter and Silent. The rate control has been carried out in groups of 1 (equivalent to constant rate), 5, 10 and 40 frames (GOFs), and with 1, 2, 3, and 4 iterations (see section 3). These results show that in most cases there is little improvement in going from 3 to 4 iterations of the algorithm. Also, one can observe that there is a small tendency for the average PSNR to grow with the GOF size; however, a GOF of 10 seems to be a good compromise. It is important to point out that, although a larger GOF size implies a large encoding/decoding delay, it does not imply an increase in complexity. Also, the complexity grows linearly with the number of iterations of the rate-control algorithm. One can also see from table 2 that the proposed rate-control strategy provides gains in PSNR in the vast majority of cases.

Figure 2 shows the PSNR plotted against frame number for the sequence Silent, coded with the EZW encoder, both without rate control and with a rate control using a GOF of 40 and 3 passes.

5. CONCLUSIONS

In this paper we have proposed a novel rate-control strategy for use in embedded wavelet video encoders based on an MPEG-4 framework. We have shown that the best model for the R-D characteristics of the frame difference encoded using embedded wavelet encoders is

---

[2] It is important to point out that the QCIF frames have had its dimensions changed to 160 by 128 in order to avoid boundary problems with the wavelet encoders.

Mother	GOF=1	GOF=5				GOF=10				GOF=40			
		1 iter	2 iter	3 iter	4 iter	1 iter	2 iter	3 iter	4 iter	1 iter	2 iter	3 iter	4 iter
EZW	38.25	38.44	38.73	38.82	38.74	38.20	38.62	38.74	38.73	38.28	38.68	38.70	38.80
Λ_{16}	39.61	39.65	39.69	39.71	39.76	39.66	39.75	39.76	39.74	39.72	39.82	39.90	39.87
E_8	39.46	39.47	39.59	39.61	39.61	39.48	39.58	39.64	39.63	39.53	39.66	39.71	39.71
D_4	39.22	39.26	39.39	39.40	39.40	39.27	39.34	39.42	39.44	39.40	39.44	39.48	39.50
Silent													
EZW	36.04	36.79	36.58	36.60	36.61	36.31	36.65	36.68	36.71	36.15	36.44	36.49	36.63
Λ_{16}	37.52	37.57	37.60	37.60	37.62	37.60	37.68	37.68	37.67	37.61	37.61	37.56	37.54
E_8	37.32	37.32	37.46	37.52	37.51	37.38	37.49	37.50	37.50	37.34	37.43	37.46	37.47
D_4	37.09	37.18	37.31	37.32	37.33	37.23	37.33	37.32	37.33	37.10	37.26	37.32	37.32

Table 2: Average PSNR for different wavelet encoders and different rate control parameters for the sequences Mother-and-daughter and Silent.

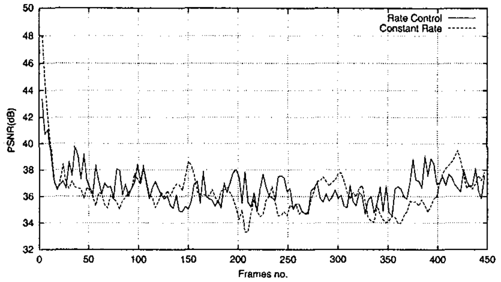

Figure 2: Plot of PSNR versus frame number for the Silent sequence, coded using EZW encoder.

piecewise linear, and propose an algorithm for rate allocation. The frame dependency problem has been tackled by applying iteratively the proposed algorithm to a group of frames.

Simulation results have shown that embedded wavelet encoders can be successfully used in interframe video encoding, specially if employing successive approximation vector quantization. Also, the proposed rate-control strategy was show to be effective, with the advantage of not requiring any arbitrary assumption about neither the R-D characteristics the encoders nor the frame dependency.

Considering that the MPEG-4 VM-8 used as the basis of these coders has been "optimized" for use with a DCT-based encoding method, the good results shown here suggest us that embedded wavelet video encoders can also be very effective for interframe video coding.

6. REFERENCES

[1] P. J. Sementelli, A. Bilgin, J. H. Kasner, and M. W. Marcellin, "Wavelet TCQ: Submission to JPEG 2000 (invited paper)," in *Proc. of Applications of Digital Image Processing, SPIE*, (San Diego, CA), July 1998.

[2] ISO/IEC JTC1/SC29/WG11, "MPEG-4 video verification model version 8.0," July 1997.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 41, pp. 3445–3462, December 1993.

[4] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.

[5] J. Li, P.-Y. Cheng, and C.-C. J. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 696–702, August 1997.

[6] E. A. B. da Silva, D. G. Sampson, and M. Ghanbari, "A successive approximation vector quantizer for wavelet transform image coding," *IEEE Transactions on Image Processing, Special Issue on Vector Quantization*, vol. 5, pp. 299–310, February 1996.

[7] D. G. Sampson, E. A. B. da Silva, and M. Ghanbari, "Low bit rate video coding using wavelet vector quantization," *IEE Proceedings Part-I: Vision, Image and Signal Processing*, vol. 142, pp. 141–148, June 1995.

[8] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.

[9] D. Mukherjee and S. K. Mitra, "Vector set-partitioning with seccessive refinement voronoi lattice VQ for embedded wavelet image coding," in *1998 IEEE International Conference on Image Processing*, (Chicago, Illinois), October 1998.

[10] K. Ramchandran and A. Ortega, "Rate-distortion methods for image and video compression," *IEEE Signal Processing*, vol. 15, pp. 23–50, November 1998.

[11] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Transactions on Image Processing*, vol. 3, pp. 533–545, September 1994.