Rate Control for an Embedded Wavelet Video Coder

Po-Yuen Cheng, Jin Li, and C.-C. Jay Kuo

***Abstract*— Embedded coders provide a better rate-distortion tradeoff while the coded bit stream can be truncated at any point without a significant perceptible distortion. In this work, we investigate rate control for an embedded wavelet video coder by converting the rate control problem to a bit allocation problem for each frame. Then, a computationally efficient rate control algorithm is derived by exploiting the rate-distortion performance of the embedded wavelet coder and the frame dependency between the reference frame and the predictive frame. Experiments are performed to demonstrate the superior performance of the embedded wavelet video coder with the proposed rate control scheme. It is shown that the proposed rate control strategy outperforms the fixed allocation rate control by 0.1–0.4 dB for a variety of sequences, and the performance gain can be as large as 1.3–2.7 dB around scene changes.**

***Index Terms*— Embedded coding, MPEG, multimedia, rate control, video coding, video compression, wavelet transform.**

I. Introduction

Rate control is essential to video transmission through constant bit rate (CBR) channels such as ISDN, T1, and broadcasting channels. It is also important in other applications such as video editing and video storage with CD-ROM or digital video disk (DVD). The rate control problem can be roughly stated as: the determination of proper coding parameters so that decoded video quality is optimized with respect to a certain fixed channel rate. Since the relationship between coding parameters, coding rate, and decoded video quality is not obvious, it can be complicated to meet the coding rate requirement by adjusting coding parameters. It is even more difficult to optimize the quality besides satisfying the coding rate constraint.

There are several existing rate control schemes for videos coded by the MPEG standard. A well-known example is MPEG Test Model 5 (TM5) [1], which adopts a simple rate control scheme by adjusting the quantization step size based on buffer occupancy. Even though this scheme is easy to implement, the quality of decoded videos is poor since it does not pay any attention to distortion. To improve the video quality, the rate control problem can be formulated as a constrained optimization problem and solved by the Lagrangian or minimax technique [2]. Ortega *et al.* [3], proposed a rate control scheme which used dynamic programming to search for the true global optimal solution. This approach is very complex since it requires the calculation of the rate–distortion (R–D) performance of the coder by repeatedly encoding the source with all quantization steps. Lin *et al.* [4] speeded up the scheme by encoding the source with only a few quantization steps and using interpolation to find the rate–distortion value for other quantization steps. This scheme used the spline interpolation for I frames and the piecewise linear interpolation for P frames. However, the complexity of this rate control algorithm is still high since the source video has to be encoded several times. Frimout [5] and Chen [6] proposed to model the relationship between rate, distortion, and quantization step size in an MPEG macroblock, thus eliminating an actual coding procedure. The resulting schemes measured the variance of the macroblock and used it to predict the entropy and coding rate after quantization. They are computationally efficient, but the performance is poor since the proposed model does not characterize the MPEG coder well. Although motion-compensated predictive coding plays an important role in MPEG, the effect of frame prediction has been seldom taken into consideration in rate control. Lin *et al.* [4] observed from empirical data that the variance of the motion-compensated residue grows linearly with the coding error of the reference frame. However, no analysis of this phenomenon was performed. Other research on rate control can be found in [7]–[9].

Wavelet coders have demonstrated an excellent performance in still image compression as evidenced by a sequence of papers by Shapiro [10], Taubman and Zakhor [11], Said and Pearlman [12], Ramchandran *et al.* [13] and Li *et al.* [14]. In addition to providing a better R–D tradeoff and a more pleasant subjective appearance, the wavelet coder has an embedding property in the sense that the bit stream can be truncated at any point without significant perceptible distortion. Later, Li and Kuo [15] and Xiong and Guleryuz [16] have extended the embedded coder to encode the DCT transformed coefficients. The embedding property greatly simplifies rate control because the coding control parameter is the allocated bit rate for each frame rather than the quantization step size. Compared with rate control for MPEG, the amount of research work on rate control for embedded video coders is relatively small. One simple strategy for rate control with embedded coders is to assign a fixed number of bits to encode each I and P frames, such as specified as the test condition in MPEG-4 core experiment T2-wavelet coding of P pictures. However, this simple strategy does not necessarily achieve the optimal R–D performance. This motivates us to investigate a better rate control strategy for embedded coders. Even though much simpler than the rate control in MPEG-1 and MPEG-4, we feel that it is still valuable to discuss rate control for the embedded coder because the embedded wavelet coder has taken an important role in MPEG-4 standard activity and a nontrivial rate control strategy may significantly improve the R–D performance of the coder. Furthermore, the study may contribute to the development of a better rate control strategy in MPEG as it provides clues of how many bits should be allocated for each frame. As a result, an MPEG-based rate control scheme may be further developed on top of the proposed scheme to adjust the quantization step size to meet the allocated rate.

The paper is organized as follows. We describe the embedded wavelet video coder and formulate the rate-control problem in Section II. The coder uses block-based motion compensation to reduce the temporal redundancy in image sequences and embedded wavelet coding to encode the residue. The R–D characteristic of the coder and the relationship between the reference and predictive frame are studied in Section III. It is shown that the variance of the motion-compensated residue depends linearly on the coding distortion of the reference frame, and computationally efficient rate control algorithms are then derived by using the Lagrangian method. Finally,

Manuscript received September 30, 1996; revised January 31, 1997. This paper was recommended by Guest Editors B. Sheu, C.-Y Wu, H.-D. Lin, and M. Ghanbari. This work was supported in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency, and, in part, by the National Science Foundation Presidential Faculty Fellow (PFF) Award ASC-9350309.

P.-Y. Cheng is with the Applied Technology Division, Physical Optics Corporation, Torrance, CA 90505-5228 USA.

J. Li is with the Digital Video Division, Sharp Laboratories of America, Camas, WA 98607 USA.

C.-C. J. Kuo is with the Integrated Media Systems Center and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2564 USA.

Publisher Item Identifier S 1051-8215(97)05879-5.

experimental results and concluding remarks are given in Sections IV and V, respectively.

II. Problem Formulation

A. Description of the Embedded Video Coder

The embedded video coder in this work uses the block-based motion compensation adopted by the MPEG-2 standard and encodes the residual image with a modified layer zero coder (LZC), which includes pyramid wavelet transform, successive quantization, and adaptive arithmetic coding. We have only implemented the forward predictive P frame, but not the bidirectional predictive B frame. The layered zero coder was first proposed by Taubman and Zakhor [11]; we have modified it slightly to improve its performance. For the details of the coder, we refer to [14]. We select LZC for its outstanding R–D performance and its simplicity in implementation. However, the proposed rate control strategy works for other embedded coding schemes as well.

B. Problem Formulation of Rate Control

The rate control problem for the embedded coder can be formulated as follows. Suppose that the transmitted video can be partitioned into groups of pictures (GOP's), with each GOP starting with an intra-coded frame (I) and followed by $N-1$ predictively coded frame (P), the channel capacity is C b/s, and the duration of one GOP is T s. The allocated rate for one GOP is

$$R_{\mathrm{GOP}} = CT. \tag{1}$$

The rate-control problem for the embedded video coder can then be formulated as the bit allocation among N frames in a GOP

$$\begin{cases} R_1 + R_2 + R_3 + \cdots + R_N = R_{\mathrm{GOP}}, \\ \quad \text{(bit rate constraint)} \\ \min\,(D_1 + D_2 + D_3 + \cdots + D_N), \\ \quad \text{(distortion minimization)} \end{cases} \tag{2}$$

where R_i and D_i, $i = 1, \cdots, N$, are the coding rate and average distortion for each frame, respectively. In contrast with previous work on rate control where the quantization step size is used as the coding control parameter, we now use the coding bit rate R_i directly as the coding control parameter. Because of the embedded property, we can truncate the coding bitstream of frame i at exactly bit rate R_i.

III. Rate Control for Embedded Video Coder

In order to solve the constraint optimization problem (2), we investigate the R–D performance of the residual coder and the relationship between the reference and predictive frames. We then derive the proposed rate control scheme in this section.

A. Rate–Distortion Performance of the Embedded Wavelet Coder

We performed extensive experiments for the evaluation of the R–D characteristic of the embedded wavelet coder with respect to I and P frames. We plot the coding distortion versus the coding rate in Fig. 1 for test videos Flower, Mobile, Tennis, and Cheer of the common image format (CIF). Empirically, the R–D performance of the wavelet video coder can be closely approximated by an exponentially decaying function as

$$D = D_{\max} 2^{-\beta R} = \sigma^2 2^{-\beta R} \tag{3}$$

where $D_{\max}$ is the coding distortion at coding rate $R = 0$, which is also equal to the variance σ^2 of the wavelet coefficients before coding,

TABLE I
Coding Efficiency Parameters of the Proposed Embedded Wavelet Video Coder with Respect to the I and P Frames

	Flower	Mobile	Tennis	Cheer
β_I	2.07	1.65	1.08	1.72
β_P	1.50	1.50	0.86	1.43
β_I/β_P	1.38	1.10	1.27	1.20

and the coding efficiency parameter β characterizes the decaying rate of the distortion as the coding bit rate increases. It is clear that a coder is more efficient if the corresponding β is larger. We calculate the coding efficiency parameter β_I and β_P for I and P frames with a linear mean-squared-error (LMSE) curve fitting of experimental data. The results are listed in Table I. β_I is usually larger than β_P, and the ratio between β_I and β_P is between 1.1 and 1.4.

B. Frame Dependency

We consider the frame dependency caused by motion compensation in this section. Typically, an I frame of a better quality improves the quality for motion compensation, thus reducing the coding bit rate required for the following P frames. Consequently, an optimized rate control scheme should allocate more bits to the I frame.

We denote the pixel of the original reference frame, the encoded reference frame, and its predictively coded frame by $g(i, j)$, $\hat{g}(i, j)$,and $f(i, j)$, respectively. The residue of motion compensation with respect to the original frame can be written as

$$e(i, j) = f(i, j) - g[d(i, j)] \tag{4}$$

where $d(i, j)$ is a displacement function which connects pixel (i, j) in the predictively coded frame to its motion-compensated counterpart in the reference frame. The variance of the motion-compensated residue can be calculated as

$$\sigma_g^2 = E[e^2(i, j)] = E\big(\{f(i, j) - g[d(i, j)]\}^2\big). \tag{5}$$

In video coding, motion compensation is actually based on the encoded reference frame $\hat{g}(i, j)$ since the original reference frame is not available at the decoder. Therefore, the actual residual error is

$$\hat{e}(i, j) = f(i, j) - \hat{g}[d(i, j)] \tag{6}$$

with variance

$$\hat{\sigma}_g^2 = E[\hat{e}^2(i, j)] = E\big(\{f(i, j) - \hat{g}[d(i, j)]\}^2\big). \tag{7}$$

We can rewrite $\hat{e}(i, j)$ as

$$\hat{e}(i, j) = \{f(i, j) - g[d(i, j)]\} + \{g[d(i, j)] - \hat{g}[d(i, j)]\} \tag{8}$$

where the first part is the residue of motion compensation with respect to the original reference frame, while the second part is the coding error of the motion-translated reference frame. We may reasonably assume that these two parts are uncorrelated so that

$$\hat{\sigma}_g^2 = \sigma_g^2 + E\big(\{g[d(i, j)] - \hat{g}[d(i, j)]\}^2\big). \tag{9}$$

Note that the second term of (9) is not the coding distortion of the original reference frame but that of the motion-translated reference frame. These two quantities have nevertheless a strong linear relationship. We use a frame dependency parameter α to characterize this relationship

$$\begin{aligned} E\big(\{g[d(i, j)] - \hat{g}[d(i, j)]\}^2\big) &\approx \alpha D_f \\ D_f &= E\big\{[g(i, j) - \hat{g}(i, j)]^2\big\}. \end{aligned} \tag{10}$$

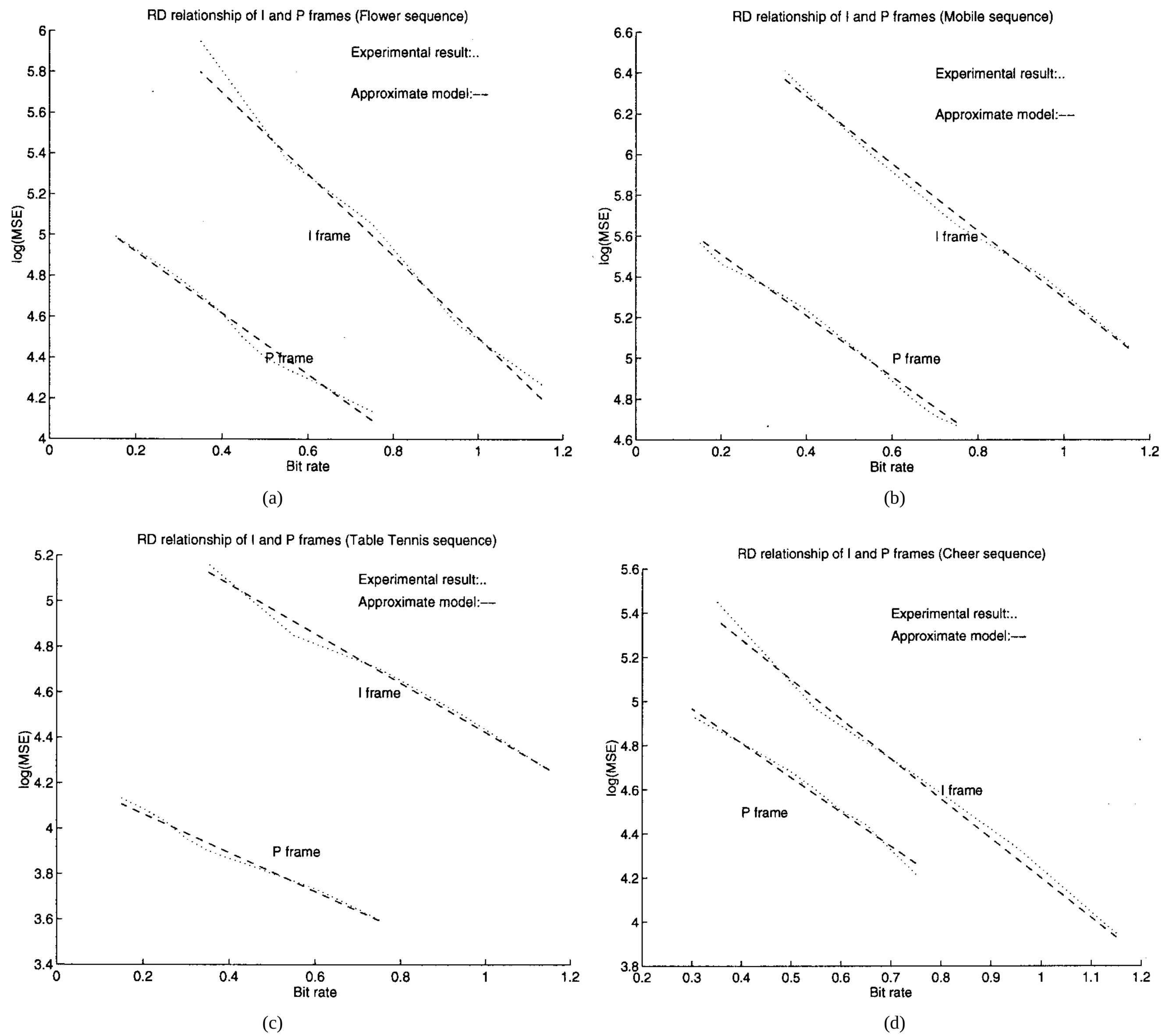


Fig. 1. The R–D characteristics of the embedded wavelet video coder for I and P frames: (a) Flower, (b) Mobile, (c) Tennis, and (d) Cheer.

where D_f is the coding distortion of the reference frame. Combining (10) and (9), we obtain

$$\hat{\sigma}_g^2 = \sigma_g^2 + \alpha D_f. \tag{11}$$

The derived relationship (11) has been confirmed by experiments in which we encode the reference frame at different rates and plot the variance of the motion-compensated residue versus the coding distortion of the reference frame for Flower, Mobile, Tennis, and Cheer sequences in Fig. 2. The affine relationship of (11) can be observed in all experiments. According to the experiments, the frame dependency parameter α ranges from 0.5–0.9. It is close to unity for sequences with high quality motion compensation. However, it decreases significantly if there is a violent motion or scene change, such as that in sequence Cheer. Equation (11) indicates that the variance of the motion-compensated residue grows linearly with the coding distortion of the reference frame offset by the residue of motion compensation with respect to the original frame. Practically speaking, if the reference frame is poorly encoded, the quality of the predictive frame with motion compensation is also poor regardless of the effort spent in motion compensation.

C. Derivation of the Proposed Rate-Control Scheme

With the codec R–D model (3) and the frame dependency relationship (11), we can solve the optimal rate control problem (2). We convert the constrained optimization problem (2) to an unconstrained optimization problem by using the Lagrangian method. That is, our objective is to minimize

$$\begin{aligned} J(R_1, \cdots, R_N) &= \sum_{i=1}^{N} D_i + \lambda \left(\sum_{i=1}^{N} R_i - R_{\text{GOP}} \right), \\ &= D_1 + D_2 + \cdots + D_N \\ &\quad + \lambda(R_1 + R_2 + \cdots + R_N - R_{\text{GOP}}) \end{aligned} \tag{12}$$

where R_{GOP} is the total number of bits assigned to a group of pictures, R_i is the bit rate to be allocated to frame i, and

$$\begin{aligned} D_i &= \hat{D}_{\max,\, i} 2^{-\beta_i R_i} \\ &= \hat{\sigma}_i^2 2^{-\beta_i R_i} \end{aligned} \tag{13}$$

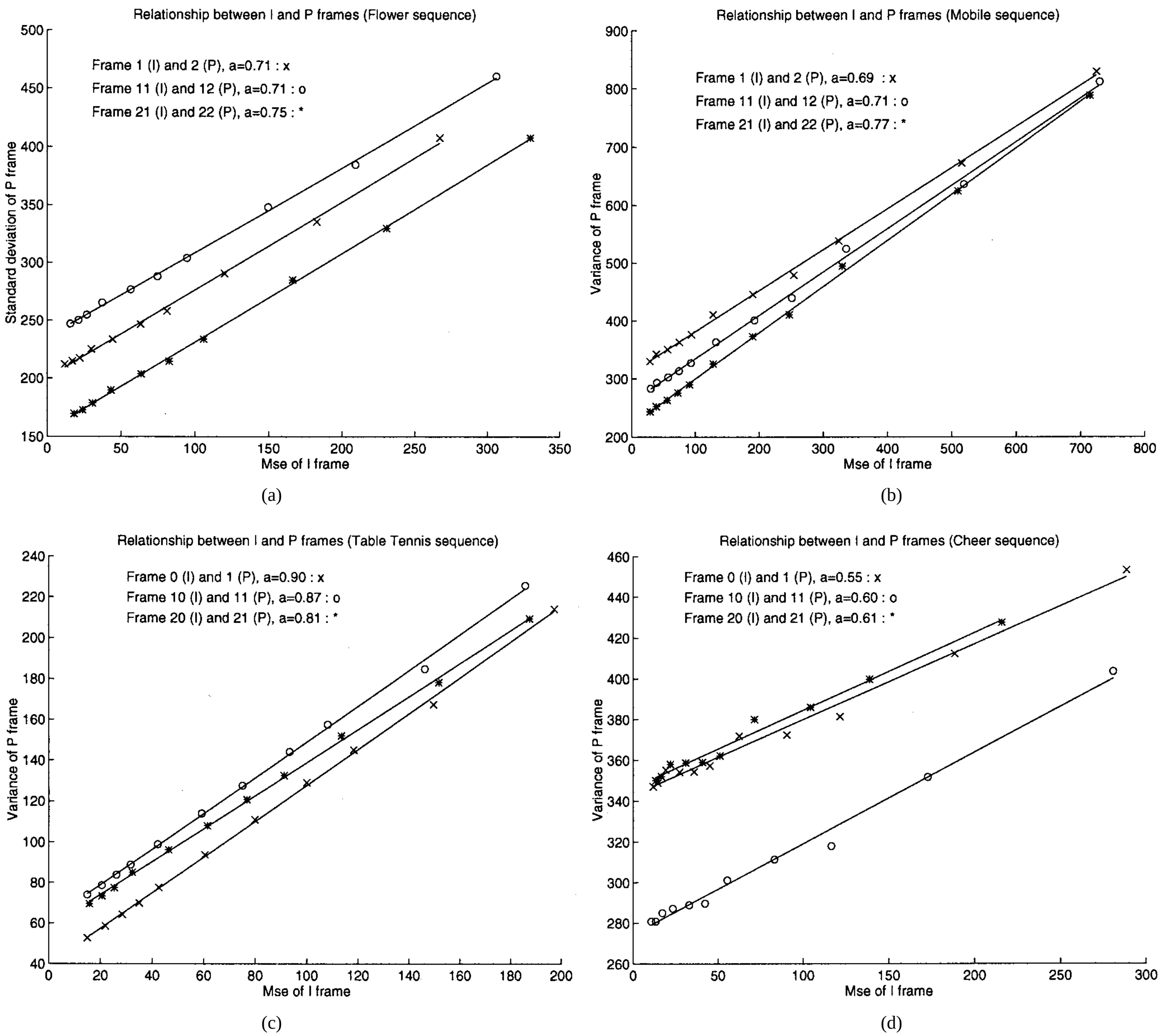


Fig. 2. The relationship between the variance of the motion-compensated residue and the coding distortion of the reference frame. For each sequence, we test three frame pairs: the first and the second frames ("x"), the eleventh and the twelfth frames ("o"), and the twenty-first and the twenty-second frames ("*"): (a) Flower, (b) Mobile, (c) Tennis, and (d) Cheer.

and

$$\hat{\sigma}_i^2 = \sigma_i^2 + \alpha_i D_{i-1} \tag{14}$$

are constraints from (3) and (11), respectively.

At the optimum point, we have

$$\frac{\partial J}{\partial R_i} = 0, \qquad i = 1, 2, \cdots, N. \tag{15}$$

To solve the above problem, let us first examine the coding rate for the last frame, i.e., frame N. We notice

$$\begin{aligned}\frac{\partial D_N}{\partial R_N} &= -\sigma_N^2 2^{-\beta_N R_N} \beta_N \ln 2 \\ &= -D_N(\beta_N \ln 2).\end{aligned} \tag{16}$$

Since frame N is not a reference frame for other frames, we simply substitute (13) into (15) and derive

$$D_N = \frac{K}{\beta_N}$$

and

$$K = \frac{\lambda}{\ln 2}. \tag{17}$$

We introduce a constant K which is equal to $\lambda / \ln 2$. Next, we proceed to frame $N-1$ which is a reference frame for frame N. Since the coding distortion D_N of frame N is related to the coding of frame $N-1$, we have

$$\begin{aligned}\frac{\partial J}{\partial R_{N-1}} &= \frac{\partial D_{N-1}}{\partial R_{N-1}} + \frac{\partial D_N}{\partial R_{N-1}} + \lambda \\ &= 0.\end{aligned} \tag{18}$$

By substituting (13) and (14) into (18), we get

$$\frac{\partial D_{N-1}}{\partial R_{N-1}} F_{N-1} = -\lambda$$

where

$$F_{N-1} = 1 + \alpha_N 2^{-\beta_N R_N}. \tag{19}$$

Similar to (16), we can calculate $\partial D_{N-1}/\partial R_{N-1}$ and obtain

$$D_{N-1} = \frac{K}{\beta_{N-1} F_{N-1}}. \tag{20}$$

Also, from (13), we have

$$2^{-\beta_N R_N} = \frac{D_N}{\hat{\sigma}_N^2} = \frac{D_N}{\sigma_N^2} = \frac{K}{\beta_N \sigma_N^2}. \tag{21}$$

By substituting (21) into (19), we get

$$F_{N-1} = 1 + \alpha_N \frac{K}{\beta_N \sigma_N^2}. \tag{22}$$

Similarly, we can calculate the partial derivatives of J with respective to R_i, $i < N-1$, as

$$\frac{\partial J}{\partial R_i} = \frac{\partial D_i}{\partial R_i} + \frac{\partial D_{i+1}}{\partial R_i} + \frac{\partial D_{i+2}}{\partial R_i} + \cdots + \frac{\partial D_N}{\partial R_i} + \lambda. \tag{23}$$

It is easy to see that frames $i+2, i+3, \cdots, N$ reference frame i through frame $i+1$. Consequently, $D_{i+2}, \cdots, D_N$ are also correlated with R_i through D_{i+1}. It is not difficult to show that

$$\frac{\partial J}{\partial R_i} = \frac{\partial D_i}{\partial R_i} + \frac{\partial D_{i+1}}{\partial R_i} F_{i+1} + \lambda \tag{24}$$

where

$$F_{i+1} = \frac{K}{\beta_{i+1} D_{i+1}}. \tag{25}$$

We notice from (13) and (14) that

$$D_{i+1} = (\sigma_{i+1}^2 + \alpha_{i+1} D_i) 2^{-\beta_{i+1} R_{i+1}} \tag{26}$$

and get

$$\frac{\partial J}{\partial R_i} = \frac{\partial D_i}{\partial R_i} F_i + \lambda$$

with

$$F_i = 1 + \alpha_{i+1} 2^{-\beta_{i+1} R_{i+1}} F_{i+1}. \tag{27}$$

Furthermore, by using (13) and (25), we have

$$F_i = 1 + \frac{\alpha_{i+1} K}{\hat{\sigma}_{i+1}^2 \beta_{i+1}}. \tag{28}$$

By using $\partial J/\partial R_i = 0$, one can calculate $\partial D_i/\partial R_i$ similar to (20) and get

$$D_i = \frac{K}{\beta_i F_i}$$

$$F_i = 1 + \frac{\alpha_{i+1} K}{\hat{\sigma}_{i+1}^2 \beta_{i+1}}.$$

Note that $\hat{\sigma}_{i+1}^2$ depends on D_i with the interframe dependency relation (14), we thus further express D_i as

$$\begin{aligned} D_i &= \frac{-B_i + \sqrt{B_i^2 - 4A_i C_i}}{2A_i}, \qquad i = 1, 2, \cdots, N-2 \\ A_i &= \beta_i \beta_{i+1} \alpha_{i+1} \\ B_i &= \beta_i \beta_{i+1} \sigma_{i+1}^2 + K\alpha_{i+1}(\beta_i - \beta_{i+1}) \\ C_i &= -K\beta_{i+1}\, \sigma_{i+1}^2. \end{aligned} \tag{29}$$

Finally, with the coding distortion for all frames specified by (17), (20), and (29) in a GOP, we can compute the allocated rate for each frame i by (13) as

$$R_i = \begin{cases} \frac{1}{\beta_1} \log_2 \frac{\sigma_1^2}{D_1}, & i = 1 \\ \frac{1}{\beta_i} \log_2 \frac{\alpha_i D_{i-1} + \sigma_i^2}{D_i} & i = 2, 3, \cdots, N. \end{cases} \tag{30}$$

TABLE II
AVERAGE PSNR FOR VIDEO CODERS AT 1.152 MB/S

	Flower	Mobile	Tennis	Cheer
MPEG TM5 (dB)	24.29	22.97	30.04	25.42
Wavelet - TM5 rate (dB)	24.79	23.45	31.22	25.93
Wavelet - fixed rate (dB)	24.85	23.54	31.03	25.93
Wavelet - proposed (dB)	25.06	23.62	31.51	26.10

TABLE III
AVERAGE PSNR FOR VIDEO CODERS AT 1.152 MB/S FOR SCENE CHANGE IN TENNIS

	Tennis (60-80)	Tennis (91-100)
MPEG TM5 (dB)	31.33	30.14
Wavelet - TM5 rate (dB)	33.11	32.27
Wavelet - fixed rate (dB)	32.39	30.50
Wavelet - proposed (dB)	33.72	33.18

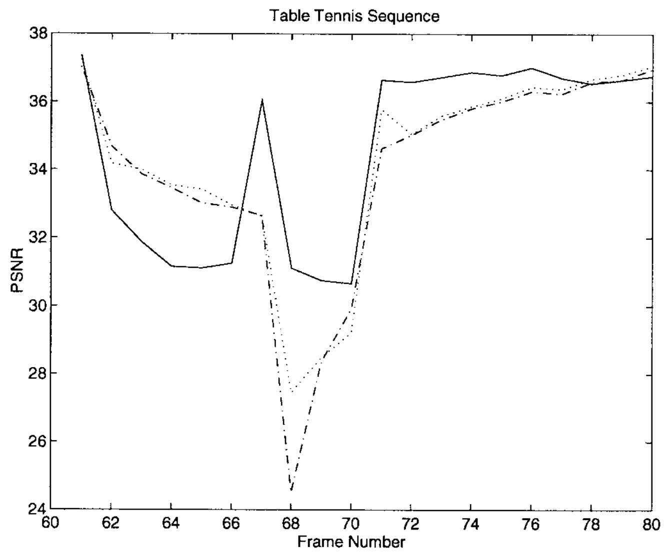
In the implementation of the proposed rate control scheme (30), we need the variance σ_i^2, the coding efficiency parameter β_i, and the frame dependency parameter α_i. The variance σ_1^2 for I frame is the mean square error of the wavelet decomposition. The variance σ_i^2 for $1 < i \leq N$ is the mean square error of the motion prediction residue with respect to the original frame. Note that in order to calculate the allocated rate R_i for frame i, we need the parameter σ_j^2, β_j, and α_j for frames $j = i+1, \cdots, N$. In applications such as CD-ROM or DVD, we can afford the delay of one GOP, thus we can prescan the whole GOP, determine parameter σ_i^2, β_i, and α_i for every frame, and calculated the allocated bit rate for each frame with (30). However, in real time applications, a GOP delay is not affordable. In those applications, we assume parameter σ_j^2, β_j, and α_j for frames $j = i+1, \cdots, N$ to be equal to that of the corresponding frames in the previous GOP. This strategy is also used in the experiment of Section IV. In (30), K is an adjustable control parameter so that the bit rate constraint in (2) can be met. It is adjusted through simple bisection iteration consisting of the following three steps.

1) Select two initial parameters K_1 and K_2 so that bit rates obtained from (30) satisfy $R_{\text{GOP}}(K_1) > R_{\text{GOP}} > R_{\text{GOP}}(K_2)$.
2) Let $K_t = (K_1 + K_2)/2$. If $R_{\text{GOP}}(K_t) > R_{\text{GOP}}$, K_1 is replaced by K_t, otherwise K_2 is replaced by K_t.
3) Repeat Step 2 until $|R_{\text{GOP}}(K_t) - R_{\text{GOP}}| < \epsilon$, where ϵ is a nonnegative number close to zero. (It is chosen to be 10^{-6} in the experiments in Section IV.)

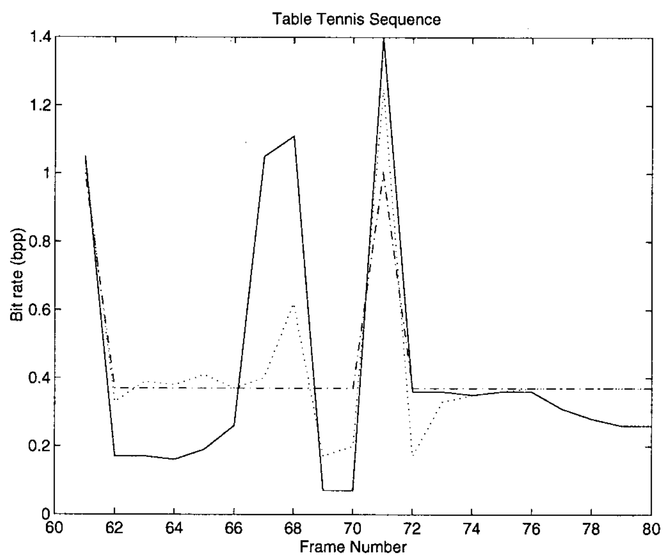
Note that each iteration of the bisection algorithm only involves the evaluation of D_i by (17), (20), and (29) and R_i by (30) with frame parameters α_i, β_i, and σ_i^2. The complexity of the algorithm is $O(N)$. Thus, the proposed rate control scheme is computationally efficient.

IV. EXPERIMENTAL RESULTS

Four test image sequences are used in the experiments: Flower, Mobile, Tennis, and Cheer. The sequences are of the CIF format with frame size 352 × 240 and 30 frames/s. Each sequence runs for 5 s, or 150 frames. The structure of each GOP is chosen to be one I frame followed by nine P frames so that there are three GOP's in one second. The target bit rate is 1.152 Mb/s. To demonstrate the performance of the proposed rate-control scheme, we compare it with two other rate control strategies. As a reference, we also show the experimental results of MPEG TM5. Our first comparison rate control strategy is to let the allocated bit rate for each frame follow that of MPEG TM5. We denote this coder as wavelet coder with TM5 rate.



(a)



(b)

Fig. 3. The (a) PSNR values and (b) coding bit rates in the neighborhood of scene change (frame 60–80 of Tennis) for an embedded wavelet coder using the TM5 rate (dotted line), the fixed rate assignment (dash dotted line), and the proposed rate control scheme (solid line).

Such rate control strategy cannot be used in real coder, as the TM5 rate will not be available. Nevertheless, it shows the improvement of the wavelet coder over DCT-based MPEG TM5. The second rate control strategy is to assign a fixed number of bits for each I and P frame. We call this coder a wavelet coder with fixed rate. It is a first-cut rate control strategy for embedded coder. In this experiment, the exact number of bits allocated for I and P frames are obtained by averaging the number of coding bits for I and P frames in MPEG TM5.

We compare MPEG TM5, the embedded wavelet coders with TM5 rate, fixed rate, and the proposed rate control strategy and show average peak signal-to-noise ratio (PSNR) performance in Table II. The embedded wavelet coder with TM5 rate offers an average PSNR improvement of 0.5–1.2 dB over MPEG TM5, where the gain comes from the better residue image coder. The proposed rate control scheme gives an average 0.1–0.4 dB gain over the embedded wavelet coder with TM5 rate and fixed rate assignment. It is also observed that the proposed rate control scheme performs especially well around scene change. To demonstrate this, we focus on the PSNR and bit rate variation in the neighborhood of a scene change (Frames 60–80, and Frames 91–100 of Tennis), where the scene change occurs at the sixty-eighth and the ninety-seventh frame, respectively. We show the average PSNR for MPEG TM5 and the three embedded wavelet based coders with different rate control strategies in Table III. We also plot the PSNR and bit rates as functions of the frame number for frames 60–80 of Tennis in Fig. 3. The wavelet coders with the TM5, fixed, and the proposed rate control strategies are depicted with the dotted, dashed dotted, and solid lines, respectively. The proposed rate control scheme is about 0.6–0.9 dB better than the wavelet coder with TM5 rate and 1.3–2.7 dB better than wavelet coder with fixed rate. By examining the allocated rate around scene change in Fig. 3(b), we find that the bit rate assigned to the frame immediately after the scene change is not large enough for the wavelet coder with the TM5 rate and the fixed-rate assignment. This causes the quality of the scene change frame to be very poor and also affects the quality of frames thereafter as shown in Fig. 3(a). Note that the bit rate for each GOP of the proposed wavelet video coder is exactly the same as the allocated bit rate $R_{\rm GOP}$ for one GOP. Such a property provides an additional advantage in video editing since we can modify or replace a whole GOP without affecting other GOP's.

V. Conclusions

In this research, we proposed a new rate control scheme for a wavelet video coder. It not only met the CBR transmission constraint but also minimized the coding distortion. In deriving the proposed rate control scheme, we investigated the rate–distortion performance of the coder and the relationship between the reference and predictive frames. We showed that the variance of the motion-compensated residue depended linearly on the coding distortion of the reference frame. By incorporating these observations, we solved the optimal bit allocation problem by using the Lagrangian method and demonstrated that the allocated bit rate for each frame R_i could be calculated as a function of several image dependent parameters such as the variance of each frame (σ_i^2), the frame dependency parameter (α_i) between the reference and the predictive frame, and the coding efficiency parameter (β_i). This scheme had a relatively low computational complexity. Experimental results demonstrated that the proposed rate control schemes achieved a better PSNR performance and performed especially well around the scene change.

References

[1] MPEG-2, Test Model 5 (TM5) Doc. ISO/IEC JTC1/SC29/WG11/93-225b. Test Model Editing Committee, Apr. 1993.

[2] D. W. Lin, M.-H. Wang, and J.-J. Chen, "Optimal delayed-coding of video sequences subject to a buffer-size constraint," in *Proc. SPIE Visual Communications and Image Processing'93,* Cambridge, MA, Nov. 1993, pp. 223–234.

[3] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation," *IEEE Trans. Image Processing,* vol. 3, pp. 26–40, Jan. 1994.

[4] L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "A gradient-based rate control algorithm with applications to MPEG video," in *Proc. ICIP'95,* Washington, DC, 1995, vol. III, pp. 392–395.

[5] E. Frimout, J. Biemond, and R. L. Lagendijk, "Forward rate control for MPEG recording," in *Proc. SPIE Visual Communications and Image Processing'93,* Cambridge, MA, Nov. 1993, pp. 184–194.

[6] J.-J. Chen and H. M. Hang, "A transform video coder source model and its application," in *Proc. ICIP'94,* Austin, TX, 1994, vol. II, pp. 962–966.

[7] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled Lagrange multiplier method," *IEEE Trans. Image Processing,* vol. 3, pp. 546–558, Sept. 1994.

[8] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Image Processing*, vol. 5, pp. 12–20, Feb. 1996.
[9] S.-W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, pp. 100–112, Mar. 1991.
[10] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
[11] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, pp. 572–588, Sept. 1994.
[12] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996.
[13] K. Ramchandran, Z. Xiong, K. Asai, and M. Vetterli, "Adaptive transforms for image coding using spatially-varying wavelet packets," *IEEE Trans. Image Processing*, 1995, submitted.
[14] J. Li, P.-Y. Cheng, and C.-C. J. Kuo, "On the improvements of embedded zerotree wavelet (EZW) coding," in *SPIE: Visual Communication and Image Processing'95*, Taipei, Taiwan, May 1995, vol. 2501, pp. 1490–1501.
[15] J. Li, J. Li, and C.-C. J. Kuo, "An embedded DCT approach to progressive image compression," in *IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 1996, pp. I:201–I:205.
[16] Z. Xiong, O. Guleryuz, and M. T. Orchard, "A DCT-based embedded image coder," *IEEE Signal Processing Lett.*, vol. 3, pp. 289–290, Nov. 1996.

Low-Complexity Block-Based Motion Estimation via One-Bit Transforms

Balas Natarajan, Vasudev Bhaskaran, and Konstantinos Konstantinides

***Abstract*—We present an algorithm and a hardware architecture for block-based motion estimation that involves transforming video sequences from a multibit to a one-bit/pixel representation and then applying conventional motion estimation search strategies. This results in substantial reductions in arithmetic and hardware complexity and reduced power consumption, while maintaining good compression performance. Experimental results and a custom hardware design using a linear array of processing elements are also presented.**

***Index Terms*—Architectures, CPU performance, instructions, motion estimation, multimedia, video compression standards.**

I. Introduction

Digital video is typically stored and transmitted in compressed form conforming to the MPEG standards for motion sequences [1]. These standards utilize block-based motion estimation as a technique for exploiting the temporal redundancy in a sequence of images, thereby achieving increased compression. The simplest abstraction of the motion estimation problem is as follows. Given two blocks of pixels, a source block of size $b \times b$ and a search window larger than the source block, find the $b \times b$ subblock in the search window that is closest to the source block.

Manuscript received September 30, 1996; revised January 31, 1997. This paper was recommended by Guest Editors B. Sheu, C.-Y. Wu, H.-D. Lin, and M. Ghanbari.

The authors are with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA.

Publisher Item Identifier S 1051-8215(97)05878-3.

The distance between two blocks can be measured by a number of different metrics [2], and typically the l_1 metric (mean absolute deviation) is used. Using this metric and a search strategy, we can evaluate candidate subblocks of the search window to find the subblock that is closest to the source block. The search strategy may be exhaustive search, evaluating each one of the candidate blocks from the search window and selecting the one that is closest in appearance to the source block. Or we may employ faster but approximate strategies, such as logarithmic search [1], to find a subblock that is close in appearance to the source block but is not necessarily the closest.

Whatever the search strategy, evaluating the l_1 metric on pixels of full intensity resolution is computationally expensive. To overcome this obstacle, we propose to transform the current and reference frame to frames of binary-valued pixels. We then apply one of the conventional search strategies to these frames. The l_1 metric then amounts to computing the exclusive-or of a sequence of bits and adding up the number of ones in the result. This can result in substantial savings in software implementations as well as reduced complexity and power consumption in hardware implementations. Our experiments show that a careful choice of the one-bit transform can realize these gains with a small sacrifice in compression efficiency.

Previously, a one-bit modification of the l_1 metric was proposed in [3], and we will compare our approach to theirs later in this paper. Recently and independently, Feng *et al.* [4] proposed a one-bit transform similar to ours, but exploited it as a preprocessing step to exhaustive search with the l_1 metric. Their approach differs from ours on three counts. 1) They use the block mean as the threshold. However, we have found that the block mean does not offer the best results in our experiments. 2) The complexity of their strategy is roughly six times that of ours. 3) Their strategy is adaptive and not suited for simple hardware implementation at low power consumption. In [5], Mizuki *et al.* describe a binary block matching architecture where block matching is performed on the binary edge maps of the current and the reference frames. They also present a custom hardware implementation that includes circuitry for edge detection and a two-dimensional (2-D) array of elementary processors, where the number of elementary processors is equal to the number of candidate blocks for full-search motion estimation. Compared to conventional block matching schemes, they estimate that binary block matching for motion estimation reduces the silicon area required by a factor of five.

In Section II, we establish the preliminaries and define the problem; In Section III, we give details of the proposed one-bit transform; in Section IV, we present a custom architecture for the one-bit motion estimation strategy, and in Section V, we present experimental results from applying our technique to sample video sequences.

II. Preliminaries

Let s denote the source block of $b \times b$ pixels, with $s_{i,j}$ being the pixel at row i and column j. Similarly, let w denote the search window with $w_{i,j}$ being the pixel at row i and column j. The subblock of w at position x, y is denoted by $w^{x,y}$, and is the block of $b \times b$ pixels $w_{x+i,y+j}$, for $i = 1, 2, \cdots, b$, $j = 1, 2, \cdots, b$.

The distance between two blocks u and v can be measured in many metrics, but typically the mean absolute deviation is used. The mean absolute deviation or l_1 metric is given by

$$\|u, v\|_1 = \frac{1}{b^2} \sum_{i,j} |u_{i,j} - v_{i,j}|. \tag{1}$$