accordingly improved as the PSNR is increased in comparison with those obtained by using the conventional frame-based partitioning method. Flexible method can reduce the average number of coding blocks more than the orthogonal method. For example, in both sequences, flexible methods can reduce about one more coding block in comparison with the results of the orthogonal methods. For both orthogonal and flexible methods, optimum approach can reduce slightly more blocks (around 0.3–0.4 blocks in average) than the simplified approach, however, the efficiencies which can be obtained from both approaches are almost the same. Therefore, the simplified approaches are more attractive to use because their complexities are significantly less than the optimum approaches.

V. Conclusions

We propose the SARP methods by modifying the block positions of the conventional frame-based partition. The modifications are performed based on the region-based orthogonal and flexible block positions, with the purpose of reducing the number of coding blocks that partition the arbitrarily shaped object regions. Simulation results show that the performances of the SADCT's combined with the SARP's are superior to that of the SADCT using the conventional frame partition in terms of the bitrate generated at the same PSNR. The coding efficiency of the simplified orthogonal SARP (simplest method) is a little bit lower than that of the optimum flexible SARP which is the most complex. Therefore, the simplified orthogonal SARP method is the most attractive to use because it is very simple to use with minimal complexity and fairly good performance on coding efficiency. For the complexity issue, it is always expected that the encoder needs to be more complex to perform the SARP procedures. The comparison of the complexities in terms of necessary computation power is as follows: simplified orthogonal SARP $<$ simplified flexible SARP $<$ optimum orthogonal SARP $<$ optimum flexible SARP. However, for decoder, there would be two directions to be considered. If we do not send the overheads (vertical, horizontal reference components) from the encoder to decoder, then the amount of the additional complexity required in the decoder is the same as that needed in the encoder. However, if we send the extra bits for overheads, decoder needs not to be complex. For example, 6 b need to be sent in the orthogonal SARP if $N = 8$ and the fixed length coding (FLC) is used. For flexible SARP, 3 b for Y component and $3K$ bits for X components are required to be sent, where K is the number of block rows occupied by the object region.

VI. Remarks

Actually flexible SARP methods that are mentioned in this paper are not optimum procedure because they may not obtain the minimum number of coded blocks if the object regions, which need to be coded, are not so simple as in Fig. 1. At some block rows, more than one segment can exist, so the minimum block number may not be reached if we use "single" horizontal reference component for each block row. For this case, we can allow many horizontal reference components for each block row. Such a loaded algorithm would be too complex for real application. Therefore, we confined our algorithm to use only one reference component in each row and the minimum number of coding blocks can be obtained under such restriction.

References

[1] "MPEG-4 proposal package description—Revision 3," ISO/IEC JTC1/SC29/WG11 MPEG95/N0998, July 1995.

[2] "Report on the *ad hoc* group on the evaluation of tools for non tested functionalities of video submissions," ISO/IEC JTC1/SC29/WG11 MPEG95/0488, Nov. 1995.

[3] M. Hötter, "Object-oriented analysis-synthesis coder based on the model of flexible 2D objects," *Signal Processing: Image Commun.*, vol. 2, no. 4, pp. 409–428, Dec. 1990.

[4] P. Gerken, "Object-based analysis-synthesis coding of image sequences at very low bit rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 4, pp. 228–235, June 1994.

[5] T. Sikora and B. Makai, "Shape-adaptive DCT for generic coding of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 59–62, Feb. 1995.

[6] M. Gilge, T. Engelhardt, and R. Mehlan, "Coding of arbitrarily shaped image segments based on a generalized orthogonal transform," *Signal Processing: Image Commun.*, vol. 1, pp. 153–180, 1989.

A New Rate Control Scheme Using Quadratic Rate Distortion Model

Tihao Chiang and Ya-Qin Zhang

***Abstract*—A new rate control scheme is used to calculate the target bit rate for each frame based on a quadratic formulation of the rate distortion function. The distortion measure is assumed to be the average quantization scale of a frame. The rate distortion function is modeled as a second-order function of the inverse of the distortion measure. We presented a closed form solution for the target bit allocation which includes the MPEG-2 TM5 rate control scheme as a special case. Model parameters are estimated using statistical linear regression analysis. Since the estimation uses the past encoded frames of the same picture prediction type (I, P, B pictures), the proposed approach is a single pass rate control technique. Because of the improved accuracy of the rate distortion function, the fluctuations of the bit counts are significantly reduced by 20–65% in standard deviation of the bit count while the picture quality remains the same. Thus, the buffer requirement is reduced at a small increase in complexity. This technique has bee adopted by the MPEG committee as part of VM5.0 in November 1996.**

***Index Terms*—Linear regression, MPEG-2, MPEG-4, quadratic modeling, rate control.**

I. Introduction

In this paper, we describe a rate control mechanism using a more accurate rate distortion model. The rate control scheme is implemented in both the picture level and macroblock level where a second-order rate-distortion model is used for target bit allocation. We first verify that a second-order quadratic formulation is sufficient to achieve adequate accuracy in modeling the rate distortion curve. Based on a linear regression analysis, a new formula is then derived to yield a smoother bit rate for the coder.

Experiments are performed for three different coders: an MPEG-2 coder, an MPEG-4 Verification Model (VM) coder, and a zerotree entropy (ZTE) encoder [2], [5], [6]. The first experiment is to encode a common intermediate format (CIF) resolution sequence at 1 Mb/s using an MPEG-2 coder. The experimental results show that the

Manuscript received March 10, 1996; revised July 1, 1996. This paper was recommended by Guest Editors Y.-Q. Zhang, F. Pereira, T. Sikora, and C. Reader.

The authors are with the David Sarnoff Research Center, Princeton, NJ 08543-5300 USA.

Publisher Item Identifier S 1051-8215(97)00942-7.

new rate distortion function introduces less bit rate fluctuation as compared to the TM5 rate control scheme [2]. The second experiment is to encode quarter common intermediate format (QCIF) resolution sequence at very low bit rate. Our approach has achieved significant improvement over VM. The third experiment is performed for a wavelet coder using a ZTE coder [6]. This paper is based on our contribution to MPEG-4 in [1] which has been made as a core experiment for MPEG-4 and later adopted as part of VM5.0 [7]. The technique is extensible to a multilayer scalable coder (for example, [4]).

II. Quadratic Rate Distortion Function Modeling

To illustrate the rationale of quadratic rate distortion function modeling, let us assume that the source statistics are Laplacian distributed

$$P(x) = \frac{\alpha}{2} e^{-\alpha|x|} \qquad \text{where } -\infty < x < \infty.$$

If we use a distortion measure $D(x, \overline{x}) = |x - \overline{x}|$, then there is a close form solution for the rate distortion function as derived in [3]

$$R(D) = \ln \left(\frac{1}{\alpha D}\right)$$

where

$$D_{\min} = 0,$$
$$D_{\max} = \frac{1}{\alpha}, \qquad 0 < D < \frac{1}{\alpha}.$$

The rate distortion function is expanded into a Taylor series

$$\begin{aligned} R(D) &= \left(\frac{1}{\alpha D} - 1\right) - \frac{1}{2}\left(\frac{1}{\alpha D} - 1\right)^2 + R_3(D) \\ &= -\frac{3}{2} + \frac{2}{\alpha} D^{-1} - \frac{1}{2\alpha^2} D^{-2} + R_3(D). \end{aligned}$$

Based on the above observation, we present a new model to evaluate the target bit rate before performing the actual encoding process. The new model is formulated in the equation as follows:

$$R = aQ^{-1} + bQ^{-2}.$$

In this formulation, the distortion measure is represented as the average quantization scale of a frame with a specific picture prediction type. Average quantization scale is chosen as the distortion measure because of its computational simplicity. Mean square error (MSE) or mean absolute difference (MAD) can be used and the same formulation is still valid. Using this distortion measure, we try to predict the target bit rate of the picture before the actual encoding process.

The choice of a second order approximation of the rate distortion function is made empirically. The experimental results indicate that the reduction in root MSE (RMSE) of the prediction diminishes if the order of the model is increased beyond two. In Table I, we present a simple example to demonstrate such an observation using the "Flower Garden" sequence with intra coding at several quantization scales. As we can see, the improvement of mean square error in model prediction diminishes as the degree exceeds two. The F-statistics is a measure for aptness of the fit which can be expressed as

$$F = \frac{\dfrac{\sum_i (Y_i - \overline{Y})^2}{k-1}}{\dfrac{\sum_i (Y_i - \hat{Y}_i)^2}{n-k}}$$

TABLE I
Estimation of the Rate Distortion Curve Using Different Order of Regression Models

Order	F ratio	RMSE
1	814.09	104,382
2	2,957.70	39,433
3	3,641.28	29,057

where

- n number of data points;
- k number of parameters to be estimated;
- $\overline{Y}$ mean of all data points;
- $\hat{Y}_i$ estimated value of data point i.

From the above formula, the smaller the residuals of the prediction is, the greater the F ratio becomes. Thus, the greater the F ratio is, the more accurate the model becomes.

III. Rate Control Description

The encoder collects the bit rate and average quantization step for each type of picture at the end of encoding each frame. Then, the model parameters a_1, a_2, a_3 and b_1, b_2, b_3 can be found. In our experiment, we use linear regression analysis and the formula below to find the model parameters a and b:

$$b_k = \frac{n\sum_{i=1}^{n} R_i - \left(\sum_{i=1}^{n} Q_i^{-1}\right)\left(\sum_{i=1}^{n} Q_i R_i\right)}{n\sum_{i=1}^{n} Q_i^{-2} - \left(\sum_{i=1}^{n} Q_i^{-1}\right)^2}$$

$$a_k = \frac{\sum_{i=1}^{n} Q_i R_i - b_k Q_i^{-1}}{n}$$

where

- n number frames observed in the past;
- Q_i, R_i actual encoding average quantization scale and bit count in the past;
- k $1 \cdots 3$.

It should be pointed out that our scheme is not limited by the method of finding the parameters. We can solve the target bit rates T_I, T_P, and T_B based on the following six equations:

$$K_P \times Q_I = K_I \times Q_P$$
$$K_B \times Q_P = K_P \times Q_B$$
$$T_I + N_P \times T_P + N_B \times T_B = R$$
$$T_I = \alpha_1 \times Q_I^{-1} + b_1 \times Q_I^{-2}$$
$$T_P = \alpha_2 \times Q_P^{-1} + b_2 \times Q_P^{-2}$$
$$T_B = \alpha_3 \times Q_B^{-1} + b_3 \times Q_B^{-2}$$

where

- K_I, K_P, K_B constant ratio for I, P, and B pictures.;
- T_I, T_P, T_B target bit rates for I, P, and B pictures;
- Q_I, Q_P, Q_B average quantization step for I, P, and B pictures;
- N_I, N_P, N_B frames to be encoded for I, P, and B pictures;
- R remaining number of bits in the current GOP.

The first two equations make the assumption that there is a constant ratio between the average quantization scales between individual picture types. The constant K_I, K_P, K_B is initialized to 1, 1, and 1.4, respectively. The third equation states that the total number of bits used is equal to the number of bits used in I, P, and B pictures.

TABLE II
PSNR OF THE CIF RESOLUTION SEQUENCE "STEFAN" AT 1 Mb/s

	TM5(	Quadratic	TM5	Quadratic	TM5	Quadratic
Mean PSNR	32.37(Y)	32.41(Y)	37.43(U)	37.46(U)	37.28(V)	37.29(V)
PSNR Standard Deviation	1.50	1.30	1.25	1.20	1.23	1.17

TABLE III
BIT COUNTS OF THE CIF RESOLUTION SEQUENCE "STEFAN" AT 1 Mb/s

	TM5(I)	Quadratic(I)	TM5(P)	Quadratic(P)	TM5(B)	Quadratic(B)
Mean	107,688	106,205	52,885	52,302	17,819	18,271
Bit Count Standard Deviation	21,689	7,646	4,893	3,917	2,161	1,626
Percentage in reduction		65%		20%		25%

The variable R is updated for every group of picture according to the assigned bit rate. The last three equations state that a second-order rate distortion function is assumed with the distortion measure as the average quantization scales of the encoded picture.

In this formulation, the distortion measure Q and the actual picture rate are found after the encoding of each picture. Using the information, the model parameters for the last three equations can be found. Based on the model found in parameter estimation, we can calculate the target bit rate before encoding. We have found a close form solution of the above six equations for the unknowns T_I, T_P, and T_B.

The formulas are different for I, P, and B frames.

To find T_I (Target bit rate for I frame), go to Step A.

To find T_P (Target bit rate for P frame), go to Step B.

To find T_B (Target bit rate for B frame), go to Step C.

Step A:

$$\begin{cases} \alpha = b_1 + N_p \times b_2 \times \dfrac{K_I^2}{K_P^2} + N_B \times b_3 \times \dfrac{K_I^2}{K_B^2} \\ \beta = a_1 + N_P \times a_2 \times \dfrac{K_I}{K_P} + N_B \times a_3 \dfrac{K_I}{K_B} \\ \gamma = -R \\ \delta = \beta^2 - 4 \times \alpha \times \gamma \end{cases}$$

$$\text{if } \delta < 0, \text{then } \begin{cases} Q_I^{-1} = -\dfrac{\gamma}{\beta} \\ T_I = a_1 \times Q_I^{-1} \end{cases}$$

$$\text{else } \begin{cases} Q_I^{-1} = \dfrac{\sqrt{\delta} - \beta}{2 \times \alpha} \\ T_I = a_1 \times Q_I^{-1} + a_2 \times Q_I^{-2}. \end{cases}$$

Step B:

$$\begin{cases} \alpha = N_p \times b_2 + N_B \times b_3 \times \dfrac{K_P^2}{K_B^2} \\ \beta = N_P \times a_2 + N_B \times a_3 \dfrac{K_P}{K_B} \\ \gamma = -R \\ \delta = \beta^2 - 4 \times \alpha \times \gamma \end{cases}$$

$$\text{if } \delta < 0, \text{then } \begin{cases} Q_P^{-1} = -\dfrac{\gamma}{\beta} \\ T_P = a_1 \times Q_P^{-1} \end{cases}$$

$$\text{else } \begin{cases} Q_P^{-1} = \dfrac{\sqrt{\delta} - \beta}{2 \times \alpha} \\ T_P = a_1 \times Q_P^{-1} + a_2 \times Q_P^{-2}. \end{cases}$$

Step C:

$$\begin{cases} \alpha = N_p \times b_2 \times \dfrac{K_B^2}{K_P^2} + N_B \times b_3 \\ \beta = N_P \times a_2 \times \dfrac{K_B}{K_P} + N_B \times a_3 \\ \gamma = -R \\ \delta = \beta^2 - 4 \times \alpha \times \gamma \end{cases}$$

$$\text{if } \delta < 0, \text{then } \begin{cases} Q_B^{-1} = -\dfrac{\gamma}{\beta} \\ T_B = a_1 \times Q_B^{-1} \end{cases}$$

$$\text{else } \begin{cases} Q_B^{-1} = \dfrac{\sqrt{\delta} - \beta}{2 \times \alpha} \\ T_B = a_1 \times Q_B^{-1} + a_2 \times Q_B^{-2}. \end{cases}$$

The above solution can be reduced to TM5 target bit allocation when the formula for $\delta < 0$ is always used [2].

IV. MPEG-2 CODER RESULTS

Fig. 1 shows an MPEG-2 encoding experimental result using the proposed approach for a CIF resolution sequence "Stefan" at 30 frames/s and 1 Mb/s. The macroblock level rate control uses the same technique as the TM5 specification for a fair comparison. The bit rate distribution of our proposed approach is smother. This experiment demonstrates the effectiveness of this method in bit rate regulation. In Tables II, III, and Fig. 1, the peak signal-to-noise ratio (PSNR) for all three components improves while the bit rate fluctuations are reduced. In Table III, the bit count standard deviations are reduced significantly for all pictures. The percentage of reductions are 65, 20, and 25% for I, P, and B pictures, respectively. In summary, the resultant visual picture quality is uniform with less fluctuations in bit counts using the quadratic rate distortion model.

V. MPEG-4 CODER RESULTS

The ISO MPEG-4 video group has defined a VM which is used as a common platform for experimentation [5]. Thus, we compare our rate control with the VM rate control. The VM rate control needs two input parameters which are Q_i (quantization scale of I pictures)

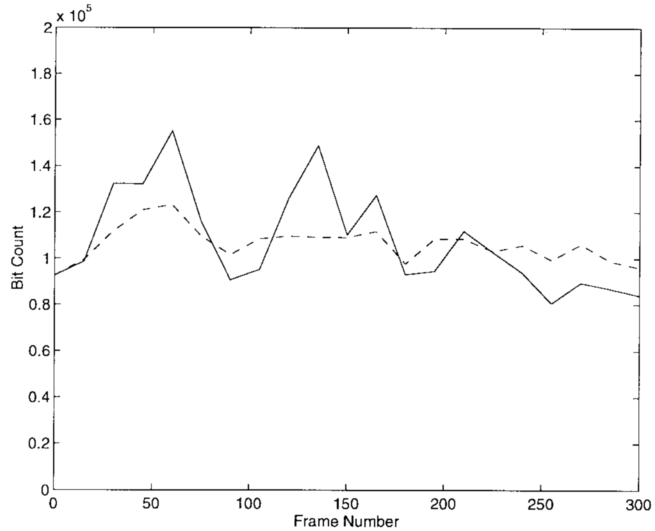

Fig. 1. The bit rate distribution plot for I pictures of the "Stefan" sequence at 1 Mb/s. The solid line is the bit-rate distribution for MPEG-2 TM5 type of target bit-rate allocation while the dotted line is the bit-rate distribution for our new approach.

TABLE IV
PSNR RESULTS USING MPEG-4 VM RATE CONTROL WITH ALL POSSIBLE Q_p AND Q_i COMBINATIONS. THE N/A ENTRY INDICATES THAT THE VM RATE CONTROL SCHEME FAILS TO MEET TARGET RATE. THE SECOND ROW YIELDS THE BEST RESULT

Q_i	Q_p	PSNR(Y)	PSNR(U)	PSNR(V)	Rate
6	2	N/A	N/A	N/A	N/A
*6	4	38.76	40.58	42.35	48.69
6	6	37.73	39.95	41.70	48.59
6	8	37.05	39.53	41.32	48.41
6	10	36.34	39.23	41.17	48.04
8	2	N/A	N/A	N/A	N/A
8	4	38.71	40.65	42.35	48.71
8	6	37.44	39.87	41.59	48.08
8	8	36.48	39.23	41.09	48.60
8	10	35.79	38.94	40.86	48.56
10	2	N/A	N/A	N/A	N/A
10	4	38.70	40.60	42.37	48.70
10	6	37.39	39.87	41.64	48.23
10	8	36.20	39.17	41.03	48.60
10	10	35.32	38.67	40.69	48.68
12	2	N/A	N/A	N/A	N/A
12	4	38.69	40.65	42.40	48.68
12	6	37.33	39.88	41.65	48.15
12	8	36.10	39.22	41.06	48.75
12	10	35.83	38.66	40.75	48.62
14	2	N/A	N/A	N/A	N/A
14	4	38.70	40.65	42.40	48.74
14	6	37.37	39.92	41.66	47.94
14	8	36.08	39.19	41.07	48.67
14	10	34.93	38.61	40.72	47.99

and Q_p (quantization scale of P pictures). Thus, human intervention is necessary and can affect the performance with 3 dB variation in PSNR as shown in Table IV. The VM rate control sometimes cannot exactly meet or can even fail to meet the prescribed rate. In Table V, the quadratic approach always meets the target rate closely. Therefore, the quadratic approach offers an automatic mechanism to meet the target rate and gives significant PSNR improvement.

In Table IV we have tested exhaustive combinations of the VM rate control. The best results of the VM rate control are shown in the

TABLE V
PSNR RESULTS USING THE PROPOSED RATE CONTROL WITH ALL POSSIBLE Q_i COMBINATIONS. THE FIRST ROW YIELDS THE BEST RESULT

Q_i	PSNR(Y)	PSNR(U)	PSNR(V)	Rate
*6	39.68	40.34	42.49	47.99
8	39.33	40.72	42.69	48.07
10	38.56	40.45	42.26	48.08
12	39.68	40.34	42.49	47.99
14	38.56	40.47	42.22	47.99

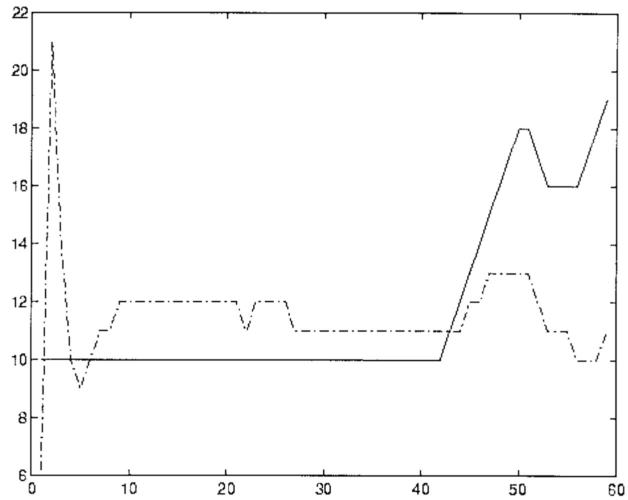

Fig. 2. Quantization scale plot of the "hall monitor" sequence at 48 kb/s. The solid line represents the VM rate control. The dotted line represents our new approach.

second row of Table IV. The quadratic rate control uses the same Q_i as shown in the first row of Table V. The resultant gain is 0.92 dB in PSNR for the Y component. We plot Q_p, bit count, and PSNR in Figs. 2–4. Fig. 2 shows that the quadratic approach quickly stabilizes within the first ten frames and converges to the correct Q_p. The visual evaluation concludes that no significant degradation occurs in the first few frames since the first I-frame is coded at high quality. The overall picture quality is visually superior. Fig. 3 shows less fluctuation in bit count. A more uniform video quality is observed since the PSNR is smoother in Fig. 4.

VI. WAVELET-BASED CODER RESULT

In order to show the versatility of our approach, a variation of the quadratic approach is used for a wavelet-based zerotree coder. The ZTE coder uses quantization scale to perform rate control. The embedded zero tree (EZW) coder uses frame target rate for rate control and meets it exactly. Since our solution yields the target rate and quantization scale simultaneously, an identical rate control scheme can be used for the ZTE coder. The EZW coder does not use quantization scales. Thus, a variation of the original approach is necessary. Since the quantization scale is considered as a distortion measure, we can use the MSE or MAD as the new distortion measure. Therefore, the same formulation can be used for the EZW coder.

In Table VI, we perform a rate control experiment using a ZTE coder [6]. The quantization scale is selected for all frames in this ZTE coder which yield a bit rate 46.56 kb/s. The quadratic rate control scheme tries to meet this rate by varying the quantization step for every frame. Figs. 5 and 6 show that the quadratic approach converges to the desired quantization scale within the first 10 frames

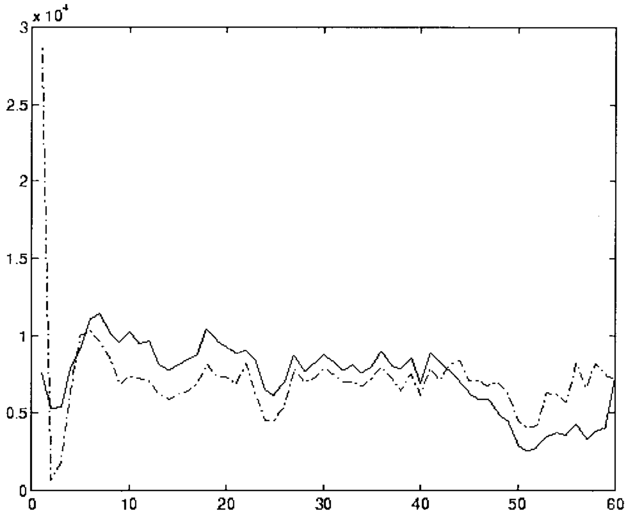

Fig. 3. Bit count plot of the "hall monitor" sequence at 48 kb/s. The solid line represents the VM rate control. The dotted line represents our new approach.

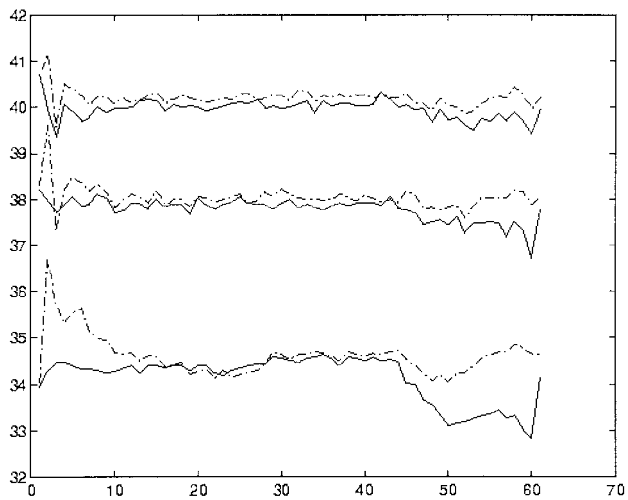

Fig. 4. PSNR plot of the "hall monitor" sequence at 48 kb/s. The solid line represents the VM rate control. The dotted line represents our new approach.

TABLE VI
THE PSNR COMPARISON FOR A ZTE CODER WITH AND WITHOUT RATE CONTROL

	PSNR(Y)	PSNR(U)	PSNR(V)	Rate
ZTE	39.63	38.74	41.23	46.56
ZTE+RC	39.60	38.19	40.64	46.53

despite a poor initial guess. The PSNR is slightly less. However, we demonstrate that the same concept is applicable to the wavelet coders.

VII. CONCLUSIONS

A new rate control scheme is proposed and implemented to address compressed video bit stream transmission in a fixed bandwidth channel. First, a quadratic relationship between the rate and distortion is presented, with the verification that the model provides accurate description of the rate distortion behavior. Then, a more accurate target bit allocation scheme is formulated using the proposed rate-distortion function. To demonstrate the effectiveness of our new model, the rate control scheme is implemented for an MPEG-2 coder at 1 Mb/s, an MPEG-4 coder at 48 kb/s, and a ZTE coder at 46 kb/s.

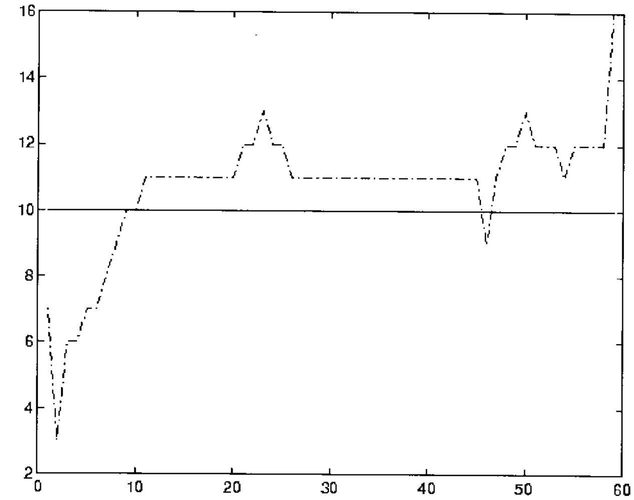

Fig. 5. Quantization scale plot of the "hall monitor" sequence at 46.56 kb/s. The solid line represents the VM rate control. The dotted line represents our new approach.

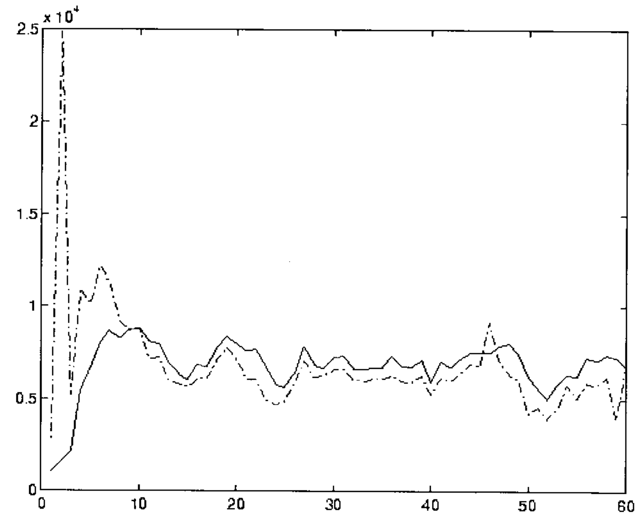

Fig. 6. Bit count plot of the "hall monitor" sequence at 46.56 kb/s. The solid line represents the VM rate control. The dotted line represents our new approach.

The results indicate that less bit rate fluctuation and better PSNR resulted from the proposed algorithm.

REFERENCES

[1] T. Chiang, Y.-Q. Zhang, J. Lee, S. Martucci, H. Peterson, I. Sodagar, R. Suryadevara, and C. Wine, "A rate control scheme using a new rate-distortion model," Dallas, TX: JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 95/0436, Nov. 1995.

[2] ISO-IEC/JTC1/SC29/WG11, "Test model 5," JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio, MPEG 1994.

[3] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding.* New York: McGraw-Hill Electrical Engineering Series, 1979.

[4] T. Chiang and D. Anastassiou, "Hierarchical coding of digital HDTV," *IEEE Commun. Mag.*, May 1994.

[5] Video Group, "MPEG-4 video verification model version 2.0," Firenze, Italy, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 96/1260, Mar. 1996.

[6] S. Martucci, I. Sodager, T. Chiang, and Y.-Q. Zhang, "A zerotree video coder," *IEEE Trans. Circuits Syst. Video Technol.*, this issue, pp. 109–118.

[7] Video Group, "MPEG-4 video verification model version 5.1," Maceio, Brazil, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Asssociated Audio MPEG 96, Nov. 1996