Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling

Wei Ding, *Member, IEEE,* and Bede Liu, *Fellow, IEEE*

***Abstract*—For MPEG video coding and recording applications, it is important to select quantization parameters at slice and macroblock levels to produce consistent quality image for a given bit budget. A well-designed rate control strategy can improve overall image quality for video transmission over a constant-bit-rate channel and fulfill editing requirement of video recording, where a certain number of new pictures are encoded to replace consecutive frames on the storage media using, at most, the same number of bits. In this paper, we developed a feedback re-encoding method with a rate-quantization model, which can be adapted to changes in picture activities. The model is used for quantization parameter selection at the frame and slice level. Extra computations needed are modest. Experiments show the accuracy of the model and the effectiveness of the proposed rate control method. A new bit allocation algorithm is then proposed for MPEG video coding.**

I. Introduction

The MPEG video coding standard [1] will have far reaching impact on high-definition television (HDTV), digital video communications, CD-ROM and digital video disk recording, and multimedia computing and networking. For digital video transmitted over a constant bit-rate channel, such as advanced digital TV and HDTV, rate control and quantization are of key importance to visual quality. The objective is to achieve a target bit rate with *consistent* visual quality. The problem can be separated into three parts: 1) to allocate target bits for each picture according to image complexities and buffer fullness for a given channel bit rate, 2) to set reference values of quantization parameters for slices and macroblocks (MB's) to meet the target rate, and 3) to derive the actual quantization parameter for each macroblock from its nominal parameter and the MB activity. In digital video recording and editing [2], it is required that a frame (for intraframe coding only) or a group of pictures (GOP) (for interframe coding) can be replaced by a new frame or GOP and the new frame(s) has to be encoded using, at most, the same number of bits to avoid decoding and re-encoding of the entire sequence. A common problem present in all the above is: how to choose the quantization parameter to meet *exactly* the target bit budget.

The exact bit rate control is also relevant in the following scenario. For a nonrealtime MPEG encoder, such as the IBM Power Visualization System parallel MPEG2 encoder [3], the original sequence is separated in segments, and these segments are encoded in parallel by multiple processors. The encoded bitstreams of the segments are then concatenated together. To comply with the video buffer verifier, each segment is required to be encoded to meet a prespecified bitrate target *exactly.*

The published solutions to this problem suggest the use of previous bit count as a prediction for the current macroblock, frame or sub-GOP [4], [5]. Other approaches use models, estimated off-line from training sequences, to predict the number of bits for the current macroblock under a stationary assumption [6]–[9].

In this paper, we develop a feedback re-encoding method with rate-quantization modeling, which can adapt to changes in picture activities. The model is used for the quantization parameter selection at both frame and slice level in order to distribute bits for uniform quality. One advantage of the approach is that the method is accurate and robust for different sequences and different methods of adaptive quantization. Another advantage is that the algorithm can increase and decrease the quality of an entire picture uniformly. The increased computation in our approach is modest, involving in most cases quantization and run-length entropy coding for only one additional frame. Experimental results are given to show the accuracy of the model and the effectiveness of the proposed rate control method. We then use the model to propose a new bit allocation and rate control algorithm for video coding.

This paper is organized as follows. In Section II, the problem is stated and some previous methods are reviewed. A rate modeling and control algorithm is developed in Section III. Experimental results are presented in Section IV. A new bit allocation and rate control algorithm is then proposed in Section V. And finally conclusions are drawn in Section VI.

II. Rate Control Problem

A. Brief Review of MPEG

In the MPEG syntax [1], an image sequence is separated into GOP's, and each GOP consists of at least one intra-coded picture (I frame), some prediction-coded pictures (P frames) and bidirectionally interpolated pictures (B frames). This is illustrated in Fig. 1. Each picture is divided into 8 $\times$ 8 blocks. A macroblock (MB) is composed of four adjacent luminance

Manuscript received August 23, 1994; revised October 23, 1995. This paper was recommended by Associate Editor C. A. Gonzales. This work was presented in part at SPIE Electronic Imaging-Digital Video Compression: Algorithms and Technologies, San Jose, CA, February 1995. This work was supported in part by a Sony Graduate Fellowship and by an NSF Computing Infrastructure Grant CDA-9216171.

W. Ding was with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA. He is now with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA.

B. Liu is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

Publisher Item Identifier S 1051-8215(96)01324-9.

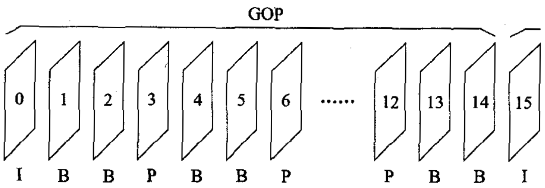


Fig. 1. GOP structure.

blocks and their corresponding chrominance blocks (two such blocks for 4 : 2 : 0 chroma format). The MB is used for motion compensation. A slice is comprised of an arbitrary number of consecutive MB's in the same horizontal row and is the level between a picture and a MB. The MB is the lowest level for the adjustment of the quantizing scale, called MQUANT, which is an integer between one and 31. The block is the unit for DCT, quantization, and run-length entropy coding. The actual quantization step is obtained by multiplying the MQUANT with two prestored quantization matrices, one each for luminance and chrominance.

B. Rate Control

The objective of rate control is to achieve a target bit rate with *consistent* visual quality.[1] In Test Model 5 (TM5) [4], the problem is separated into three steps: 1) to allocate bits to each picture according to image activities and buffer fullness, 2) to set nominal slice quantization parameters to meet the target rate, and 3) to derive MQUANT of each MB from its nominal QP according to its activities to exploit spatial masking effect [6]. The separation of the second and the third steps reduces the complexity of the problem and yet allows adaptation to local activities.

For the third step, MQUANT's can be adjusted from a given reference quantization parameter by adaptive quantization (AQ) in order to adapt to local image content and to exploit spatial masking effect [6]. Several approaches exist [4], [10], e.g., based on MB variance in TM5. We will not consider the problem in this paper.

The first two problems, however, remain relatively open. Optimal bit allocation and quantization, in the sense of a visual quality measure, are difficult to achieve. In order to achieve consistent visual quality, a reasonable alternative is to select identical reference quantization parameters so as to distribute bits within a picture for uniform quality.

One solution to the first step is to allocate bits to I, P, and B frames according to some global complexity measures of previous encoded pictures, such as product of the average MQUANT and the bit count used in TM5 [4]. In [6], the complexity measure for each MB is taken as its spatial mean absolute difference (for I-MB) or its mean absolute prediction difference (for P-MB). In [11], a technique was proposed to utilize temporal masking effect to dynamically choose frame types in a GOP and to allocate bits to each frame. A more involved method is detailed in [7], where prestored bit counts, determined by experiments, are allocated to I, P, and B frames under eight scene complexity classifications. Dynamic programming has also been applied to joint bit allocations of I, P, and B frames [12].

Without accurate rate control, a good bit allocation method cannot be carried out. In this paper, instead of developing the entire algorithm, which probably will adopt and repeat some common techniques, we will focus on the second step, i.e., rate control—to choose the quantization parameters to meet the target number of bits for consistent visual quality. The proposed algorithm can be incorporated into the framework of Test Model 5 [4]. We will then consider the bit allocation problem.

C. Existing Methods

A brute-force exhaustive search is impractical because of the large number of possibilities. A CIF picture, for example, is typically divided into 15 slices, resulting in 15^{31} possibilities to select quantization parameters.

Because of correlations between successive frames, the number of bits to code the previous picture (or sub-GOP) can provide an estimate of the number of bits needed for the current picture (or sub-GOP). In [5], a bit-count estimate of the current picture (or sub-GOP) is taken to be the number of bits of the most recent picture of the same type (or sub-GOP), and is used for adjusting quantization parameters. The prediction method for MB's is used in TM5 [6], [4]. In TM5, before the jth MB is encoded, its quantization parameter is adjusted according to the difference of the actual number of bits and the allocated number of bits of the previous $(j-1)$ MB's in the following way:

$$Q_j^{MB} = \frac{\sum_{k=1}^{j-1} (\mathtt{actual\_bits} - \mathtt{allocation})_k + \mathtt{constant}}{\mathtt{factor}}.$$

Other methods more or less use the same strategy. However, the simple prediction from the previous bit count results in delay of the adjustment of quantization parameters, which will cause uneven image quality either from picture to picture or within each picture. For example, if the lower portion of a picture is more active, this part will need more bits to code than the upper portion. The above method will produce a picture of finely quantized upper part and coarsely quantized lower part, since a normal quantization scale for the upper portion would leave insufficient bits for the lower portion. Even though a moderate QP for the entire picture would be used, if the picture activity distribution had been known beforehand, a moderate QP would be used for the entire picture. The delayed adjustments can also cause uneven quantization from frame to frame.

To provide a better estimate for the current frame, several efforts have been focused on modeling the relationship of bit-count versus quantization. Puri and Aravind [7] made detailed classifications of macroblocks and estimated bit models for each class from training sequences. Such bit models were then stored and used to estimate bit counts of uncoded pictures. Viscito and Gonzales [6] also made MB classifications and used parametric models for different types of MB's to make bit-count estimation. In addition, parametric models were used

[1] We use the term of "consistent" instead of "constant," since it is difficult to characterize visual quality quantitatively.

for the estimation in [8] and [9]. These feed-forward rate modeling methods are based on the implicit assumption that the rate-quantization relationship is stationary. However, such an assumption may not hold for different sequences, not even for different parts of the same sequence, as noted in [7] and [8]. The error in bit estimation due to model mismatch can be quite significant. Adaptive approaches, involving re-estimation from large numbers of frames, will lose the advantage of the feed-forward modeling methods.

III. Rate-Quantization Modeling

In this section, we will develop a re-encoding algorithm to select quantization parameters for a picture with a given bit budget. It is desirable to choose identical reference QP's for all MB's in the picture for uniform visual quality. We will thus first develop a model of bit count versus picture-level reference QP. However, we will find out in later in the section that to match a target rate we have to change the reference QP on a level lower than picture level, e.g., the slice level as used in this paper. As discussed in Section II-B, the actual MQUANT for a MB might be changed from its reference QP. This is done by adaptive quantization to allow adaptation to local image content. Also, an algorithm can use different QP's for I and P MB's. There are several approaches for adaptive quantization [4], [10]. But we do not consider this in this paper, although we use the AQ scheme of TM5 in our simulation.

One advantage of the re-encoding approach is that the estimated model is accurate, since the measurements from actual encoding have already taken into account scene activities and actual use of AQ scheme. The method is thus robust for different sequences. This is opposed to the approaches based on *pre-estimated* bit models. The validity of these models depend on training sequences. The second advantage is that the use of the picture-level QP can increase or decrease the quality of the entire picture uniformly. Yet AQ can take advantage of the spatial masking effect. The third one is that the proposed scheme remains valid for different AQ schemes. We can substitute the AQ method of TM5 for a more complex and better scheme. The proposed algorithm still can perform an accurate control. We will discuss the computation issue later.

A. Global Modeling

The bit count R versus frame quantization scale Q of two sequences, "tennis" and "football," are plotted in Fig. 2 for several I pictures. These two sets of curves are seen to be quite different. For the same sequence, each curve is close to that of the previous I frame, but shifted up or down. However, a simple prediction from the previous frame, for the typical range of $1 \leq Q \leq 10$, will result in a large bit error. These observations also hold for P pictures, although their R-Q curves appear to be more heavily damped than those of I pictures. These show that the R-Q curves are not stationary.

The R-Q curves resemble the rate-distortion curves of Gaussian random variables [13]. This suggests that they may be modeled by the model

$$R = \alpha + \beta \log \frac{1}{Q} \quad (1)$$

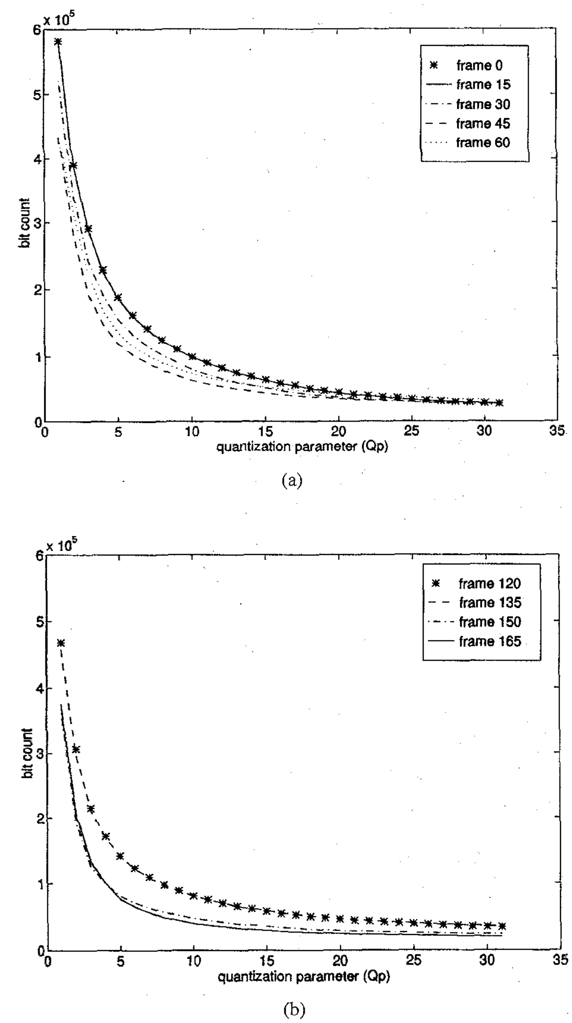


Fig. 2. I-frame R-Q curves of (a) frame 0–60 of "tennis" sequence; (b) frame 120–165 of "football" sequence.

where α and β are parameters. The parameter α accounts for overhead bits. The actual R-Q curves are, however, more heavily damped than (1). One reason is that the quantized DCT coefficients are not just entropy-coded individually, but zigzag ordered and run-length coded. The zero-grouping in the run-length coding really reduces the final bit counts. We thus use a slightly different model

$$R = \alpha + \frac{\beta}{Q^\gamma} \quad (0 < \gamma \leq 2). \quad (2)$$

We fit actual R-Q curves of I pictures of all "tennis" and "football" with various γ values from two sample points $Q = 1$ and $Q = 31$. We also fit P frames with their previous anchor frames (I or P) coded with $Q = 6$.[2] Some typical examples are shown in Fig. 3. We can see that the model (2) fits the actual curves quite well. However, it does not appear that one single model can be used for all sequences. Sometimes the error between the model and the actual curve can be quite

[2] Other quantization parameters give similar results.

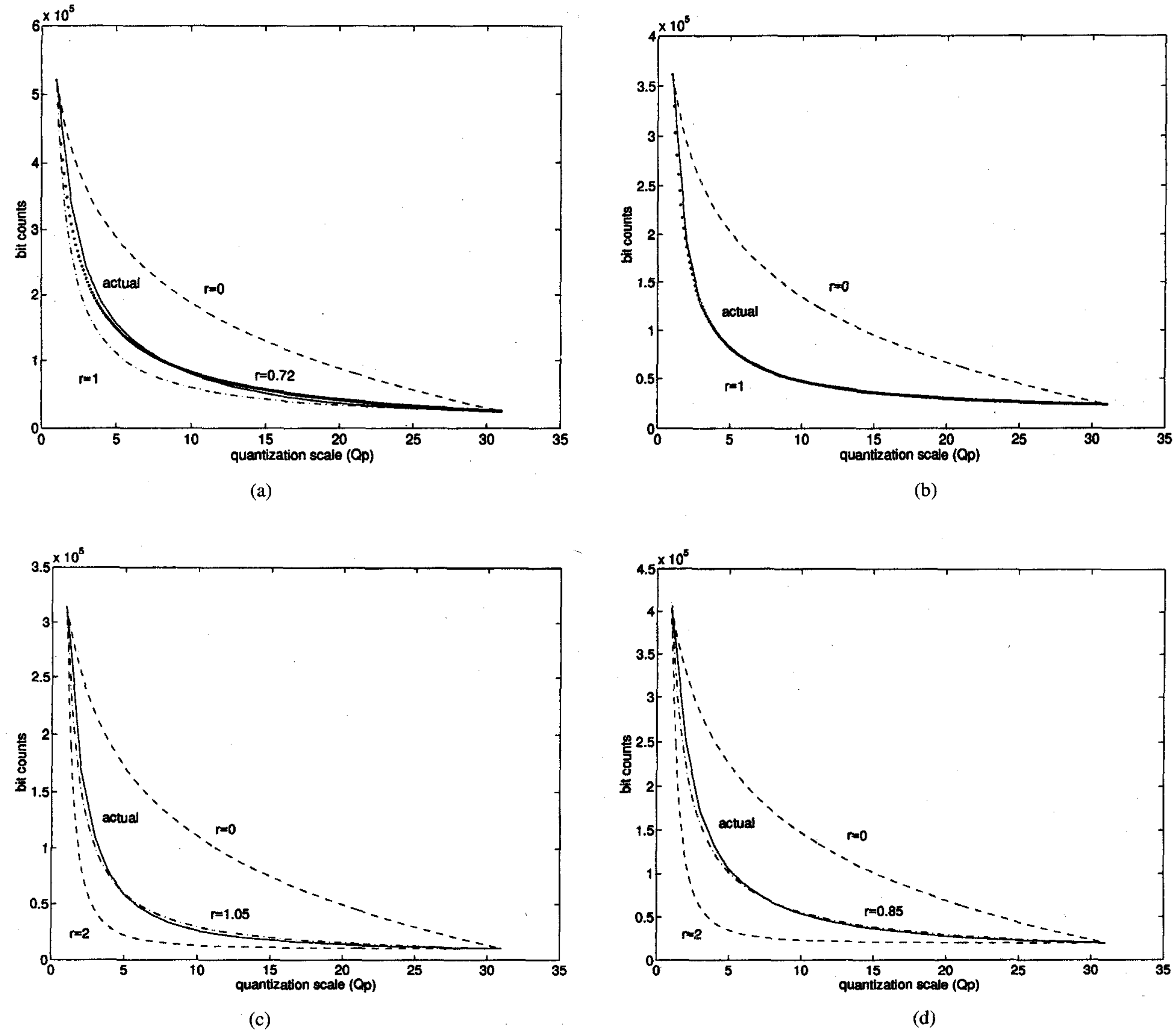


Fig. 3. Global fitting with model (2): (a) I-frame 30 of "tennis"; (b) I-frame 150 of "football"; (c) P-frame 48 of "tennis"; (d) P-frame 138 of "football." The curve of (1) is indicated by $\gamma = 0$.

large, such as for small values of QP in Fig. 3(d). In general, we have found that for both sequences, actual R-Q curves of I frames are bounded between $\gamma = 0.5$ and $\gamma = 1$ and P-frame curves are bounded between $\gamma = 0.5$ and $\gamma = 1.5$.

B. Local Modeling

For different frames and sequences, the values of the three parameters (α, β, γ) are different, and the good global fitting described above is not guaranteed for all frames and sequences. This leads us to consider local fitting. It is more accurate and effective to fit the R-Q curves by the model (2) over a limited range with two parameters (α, β) and with fixed $\gamma = 1$. Fig. 4 shows two examples. We can see that between the two fitting points, $Q = 2$ and 4 for curves C1 or $Q = 5$ and 7 for curve C2, the models fit the actual R-Q curves very well locally.

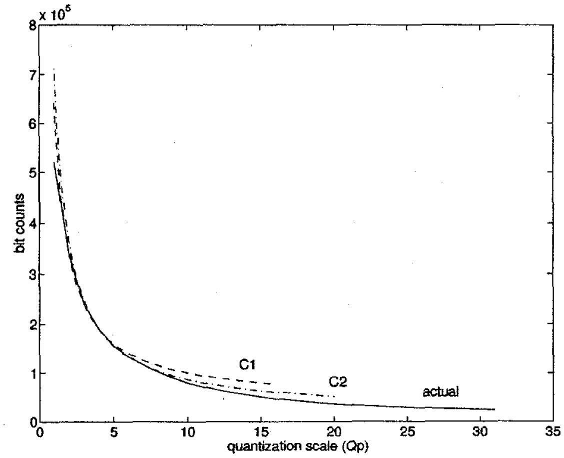
C. Slice Quantization Selection

The quantization parameter can only take on integer values between one and 31. However, the target bit count may not correspond to an integral QP. If the same quantization parameter is applied to the entire picture, the actual bit count can differ significantly from the target bit count, as illustrated in Fig. 5. A sequence coded with integer quantization parameters is shown in Fig. 6. Only I and P pictures are coded and their target bits are set fixed. The large deviation from the target bits can be easily seen.

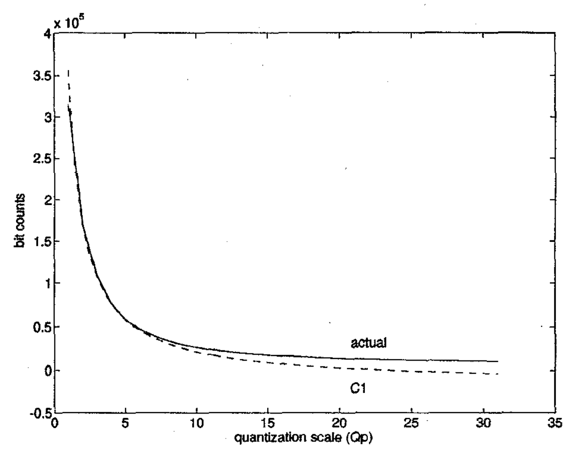
To solve the problem, we will use the extra freedom of slice QP selection. For simplicity of discussion, we use each row of MB's as a slice; other slice structures can be treated similarly. It is assumed that each slice of a picture contributes equally to the total bit count. Suppose that the rate-quantization curve is estimated to be

$$\hat{R} = R(Q).$$

Let the target bit count be R_t. Its corresponding quantization parameter is thus $Q_t = R^{-1}(R_t)$, which, in general, is not an integer. Let $Q_1 = \lfloor Q_t \rfloor$, which is the largest integer not greater than Q_t. The number of bits $R(Q_1)$ or $R(Q_1 + 1)$



(a)



(b)

Fig. 4. Local fitting of "tennis" sequence, with model (2) for $\gamma = 1$. (a) I-frame 30 from $Q_p = 2, 4$ for C1 and from $Q_p = 5, 7$ for C2; (b) P-frame 48 from $Q_p = 2$ and 4.

with all slices quantized by either Q_1 or $(Q_1 + 1)$ may not be close to R_t. The difference $R(Q_1) - R(Q_1 + 1)$ is divided by the total number of slices N_s. Then the target rate R_t can be approximated by a closer rate if

$$N_1 = \left\lfloor \frac{R_t - R(Q_1 + 1)}{R(Q_1) - R(Q_1 + 1)} \cdot N_s + 0.5 \right\rfloor \quad (3)$$

number of slices are quantized with Q_1 and the rest $(N_s - N_1)$ number of slices with $(Q_1 + 1)$.

An example is shown in Fig. 5, where the target R_t is between $R(3)$ and $R(4)$. Suppose there are three slices in a picture. If the three slices are quantized with $Q = 3$ or $Q = 4$, the total number of bits for the picture would be $R(3)$ or $R(4)$, respectively. Both rates are quite different from R_t. However, if two out of the three slices are quantized with $Q = 3$ and the other one with $Q = 4$, the total number of bits is $R(4) + \frac{2}{3}(R(3) - R(4))$, which is much closer to R_t. With more slices, finer division of the R axis will produce more accuracy. Another way to look at the problem is that

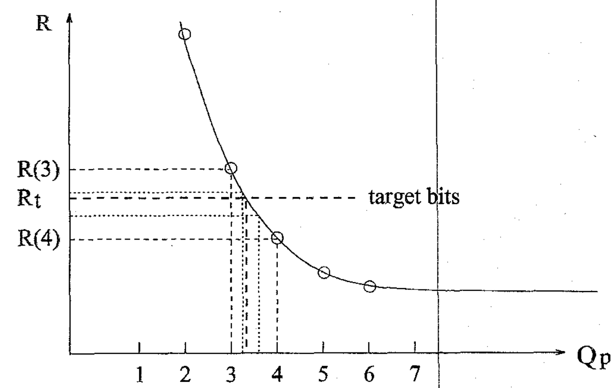


Fig. 5. Illustration of slice Q_p selection with three slices. Equal division at R axis results in unequal division at Q_p axis. For the target number of bits, one slice will be quantized with $Q_p = 4$ and the other two with $Q_p = 3$.

a noninteger frame quantization parameter can be used. It is noted that even though the R-Q model is estimated from integral frame QP's, it is used to determine noninteger frame quantization parameters.

Finally, the assumption of equal contribution from each slice is not always valid. To tackle this, we use a predetermined pattern to mix slices of different QP's to make it appear that quantization parameters are evenly distributed within the entire picture. For example, if five slices out of total 15 slices are to be coded with $Q = 3$ and the other 10 with $Q = 4$, we can mix them in the following pattern: (4, 3, 4, 4, 3, 4, 4, 3, 4, 4, 3, 4, 4, 3, 4).

D. Rate Control Algorithm

We summarize the previous sections into a constrained quantization algorithm for I or P frames:[3]

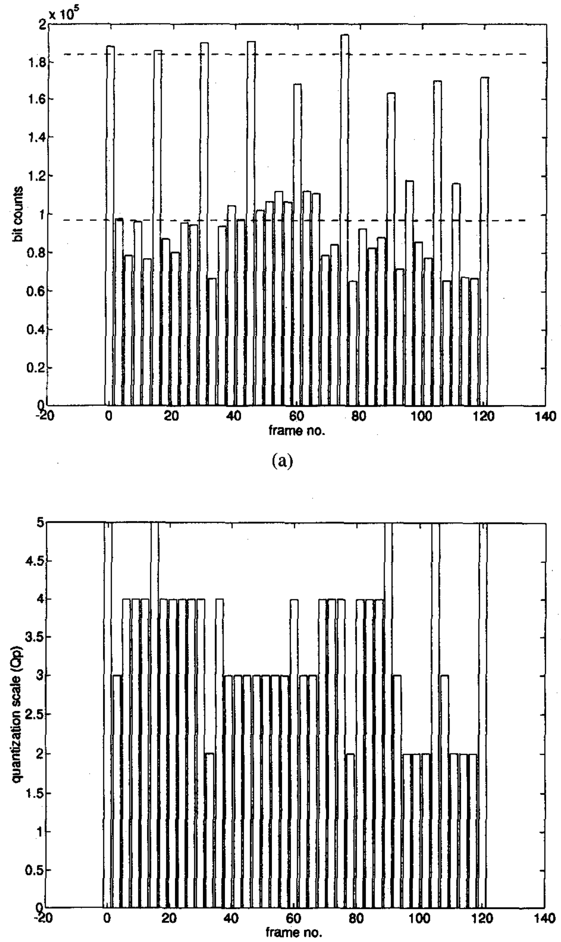
Algorithm 1:

1) Set the initial quantization scale Q_0;
2) Allocate the target R_t bits for the current frame, using algorithms, such as [4] or [7];
3) Perform the normal motion compensation and DCT transform, set $Q_1 = Q_0$, and encode the picture with Q_1 to obtain its bit count R_1;
4) a) If $|R_1 - R_t| < \epsilon$,[4] transmit the coding result, set $Q_0 = \lfloor Q_1 \rfloor$, and go to Step 2 for the next frame;
   b) If $R_1 < R_t - \epsilon$, let $Q_2 = Q_1 - \delta$, or if $R_1 > R_t + \epsilon$, let $Q_2 = Q_1 + \delta$, encode with Q_2 to get R_2;
5) Model the R-Q curve $\hat{R} = R(Q)$ from (Q_1, R_1) and (Q_2, R_2) by Section III-B.
   a) If $Q_t = R^{-1}(R_t)$ is between Q_1 and Q_2, the correct scale is found. Use the technique of Section III-C to quantize and code slices, set $Q_0 = \lfloor Q_t \rfloor$, and go to Step 2 for the next frame;
   b) otherwise, let $Q_1 = Q_2$ and $R_1 = R_2$, go to Step 4.

□

[3]The algorithm is phrased for video coding. It can be straightforwardly applied to video recording. However, extra care has to be taken in order to ensure that the coding bit count does not exceed the target bit count.

[4]ϵ is the deviation tolerance from the target bits and is not a sensitive parameter.



(a)

(b)

Fig. 6. I and P frame coding results with integer frame Q_p of "tennis:" (a) bit counts; (b) frame Q_p. The dashed lines are target numbers of bits of I and P pictures.

We have found that the step size

$$\delta = \begin{cases} 1 & [1,3] \\ 2 \quad Q_1 \in & [4,7] \\ 4 & [8,31] \end{cases}$$

gives good result for MPEG1 and is valid for *q_scale_type* = 0 in MPEG2. The reason is that a more accurate model is needed for a smaller QP, where bit rate has relatively large change.

The above algorithm is designed for I and P frames. B frames can be treated in the same way. However, since B frames usually need fewer bits to code, extra computation efforts for the modeling are not justified. A simple weighted average of quantization scales of its two anchor frames can be used, as was done in [7]. An alternative is to use a usual buffer feedback rate control for B frames to match their bit allocations.

E. Computation

The rate control scheme proposed in this paper requires only a modest amount of extra computation. The additional computation comes from quantization and run-length coding in the modeling step, which are moderate compared with motion compensation and DCT. There are usually no large changes of quantizing scales from one I or P picture to the next picture. In our experiments with two sequences of 120 frames each, there are only one or two occasions needing a second R-Q modeling step. In most cases, only two frames need to be quantized and coded for the model fitting. Furthermore, when sampling points are Q_1 and $Q_2 = (Q_1 + 1)$ and Q_t are between them, the encoding bit streams with Q_1 and Q_2 for modeling can be readily used for the final encoding of slices, no additional encoding is needed. Thus, for most pictures only one frame extra computation of quantization and run-length entropy coding are needed. For a small fraction of the frames, an extra two frames of such computation may be needed.

TABLE I
THE AVERAGE DEVIATIONS FROM THE TARGET NUMBERS OF BITS FOR I AND P FRAMES (IN PERCENT)

(%)	I frame	P frame
tennis	2.38	1.37
football	0.79	0.86
garden	0.70	0.30
average	1.29	0.84

IV. RESULTS

Three CIF sequences, "tennis," "football," and "flower garden" are used to test the proposed method. For a target bitrate of 1.3 Mb/s, fixed numbers of bits are allocated to I and P frames with $R_I = 184328$ bits and $R_P = 97014$ bits respectively. Such a bit allocation scheme is only used for testing purpose. The GOP structure of Fig. 1 is used, and each picture is divided into 15 equal slices. The results presented in Fig. 7 show that the actual numbers of bits by the proposed method are very close to the target number of bits. We calculate the average deviation of actual numbers of bits of I frames from the target R_I by

$$D_I = \frac{1}{N_I} \sum_{k=1}^{N_I} |R_k - R_I| \times 100(\%)$$

where N_I is the number of I frames in the sequence and R_k is the actual number of bits for each I frame. The average deviation for P frames can be calculated similarly. The results for the three sequences are listed in Table I. On average, the deviation of all three sequences is less than 1.5% for the I frames and less than 1.0% for the P frames. A comparison result with integer frame QP is given in Fig. 6, where the deviations from the target number of bits are very large. We would like to point out that even though the model and the method are developed by examining the first two sequences, the experiment with the "flower garden" sequence, which has different statistics from the first two, also exhibits nearly constant bit-rate result around the target number of bits. These tests demonstrate not only that we can produce

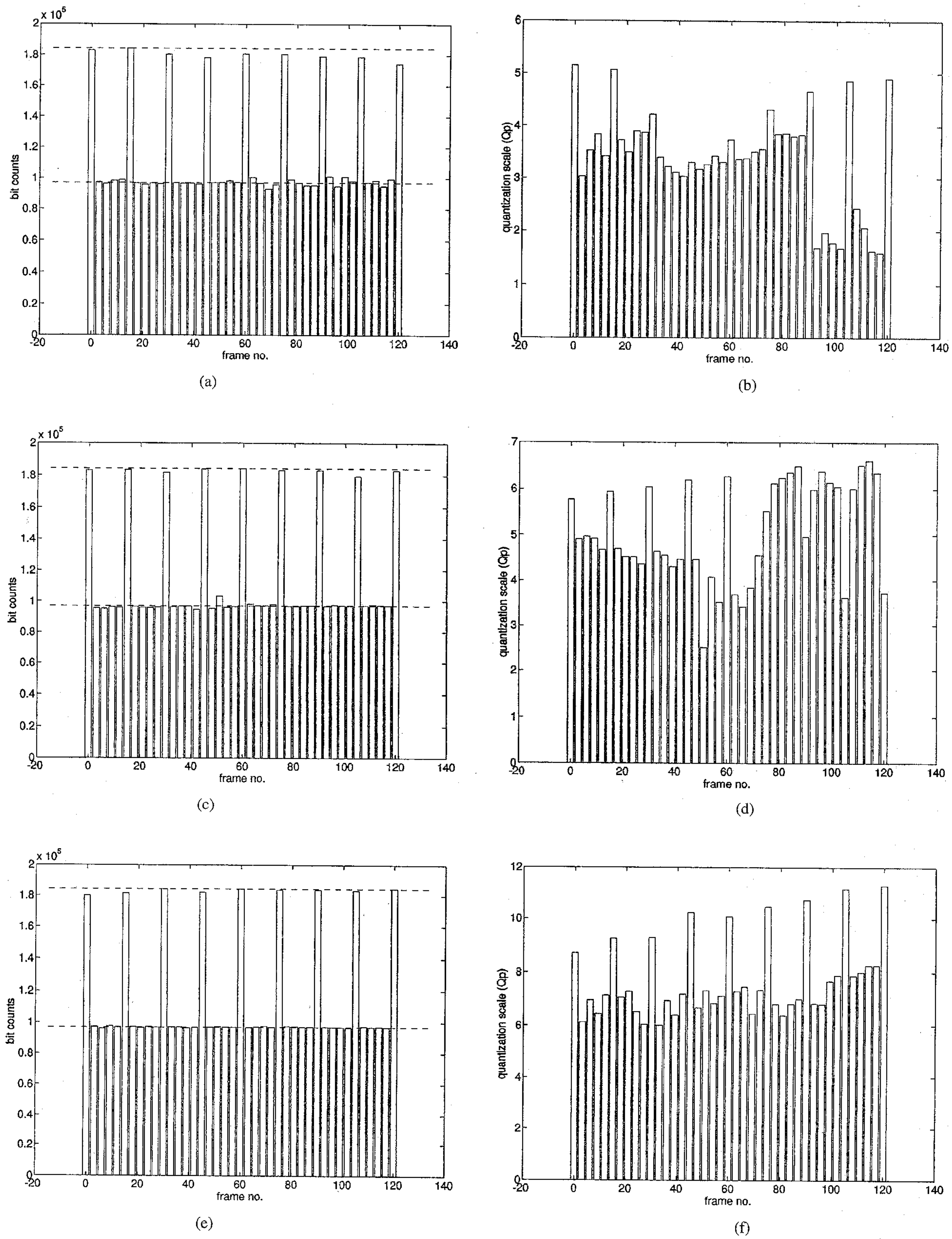


Fig. 7. I and P frame coding results with slice quantization selection: (a) bit counts and (b) noninteger Q_p of "tennis;" (c) bit counts and (d) noninteger Q_p of "football;" (e) bit counts and (f) noninteger Q_p of "flower garden." The dashed lines are target numbers of bits of I and P frames. The noninteger Q_p's are used to select slice quantization parameters.

an almost constant bit-rate stream, which in itself is useful for digital recording applications, but also that we are capable of modeling and controlling the encoding. As we discussed in Section III, we expect the method is robust for other sequences, the re-encoding algorithm estimates model from actual data.

V. A Bit Allocation and Rate Control Algorithm

In this section, we propose a new bit allocation algorithm by modifying the algorithm of TM5 using our rate modeling method. We use complexity measures of current pictures for bit allocation, instead of previous ones as in the TM5. This would allow more accurate bit allocation. The reader is referred to [4] for a detailed description of the TM5 rate control algorithm. The proposed algorithm consists of three steps.

Algorithm 2

Step 1—Bit Allocation: Complexity measures (X_i, X_p, X_b) are first computed for I, P, and B pictures. Suppose that Q_{i0}, Q_{p0}, and Q_{b0} are frame quantization parameters of previous I, P, and B pictures, respectively. Let $Q_i = \lfloor Q_{i0} \rfloor, Q_p = \lfloor Q_{p0} \rfloor$, and $Q_b = \lfloor Q_{b0} \rfloor$. We obtain the corresponding number of bits R_i, R_p, or R_b of the current picture by quantizing and encoding the picture with the appropriate Q_i, Q_p, or Q_b, using the method of Step 3. Their complexity measures are updated as

$$X_i = R_i Q_i, \quad X_p = R_p Q_p, \quad X_b = R_b Q_b.$$

The initial quantization values, for 1.3 Mb/s applications, are set to

$$Q_{i0} = 4, \quad Q_{p0} = 3, \quad Q_{b0} = 6.$$

The exact values are not particularly important. The extra quantizing and encoding computation are shared with Step 2. The target number of bits for the current picture is computed in the same way as in the TM5 [4]

$$T_i = \max\left\{\frac{R}{1 + N_p \dfrac{X_p/K_p}{X_i} + N_b \dfrac{X_b/K_b}{X_i}}, \frac{\text{bit\_rate}}{8 * \text{picture\_rate}}\right\},$$

$$T_p = \max\left\{\frac{R}{N_p + N_b \dfrac{X_b/K_b}{X_p/K_p}}, \frac{\text{bit\_rate}}{8 * \text{picture\_rate}}\right\},$$

$$T_b = \max\left\{\frac{R}{N_b + N_p \dfrac{X_p/K_p}{X_b/K_b}}, \frac{\text{bit\_rate}}{8 * \text{picture\_rate}}\right\}$$

where $K_p = 1.0$ and $K_b = 1.4$ are constants, R is the remaining number of bits for the GOP, and N_p and N_b are the number of P-pictures and B-pictures remaining in the current GOP, including the current picture.

Step 2—Rate Control: Algorithm 1 of Section III-D is used to model and encode the current picture with the target number of bits T_i, T_p, or T_b. Note that the encoding result with Q_i, Q_p, or Q_b computed in Step 1 can be reused here for the model fitting.

Step 3—Adaptive Quantization: The method of either [4] or [10], or any others, can be used. □

The increased computation from the original TM5 algorithm in most cases involves only the quantization and entropy coding of one frame, as discussed in the Section III-E.

The algorithm can be simplified if the B frames are not modeled for the reason stated in Section III-D. Instead, the B pictures are quantized using the quantization parameters of its anchor frames. In this case, Step 1 of Algorithm 2 is modified as follows. The complexity measure for B picture is estimated from the previous B picture as

$$X_b = R_{b0} Q_{b0}$$

where Q_{b0} may not be an integer. Its initial value is set to [4]

$$X_b = 42 * \text{bit\_rate}/115.$$

There is no need to compute the target number of bits for B pictures.

Step 2 of Algorithm 2 is modified as follows. No rate modeling is performed for B pictures. They are quantized by the sum of (noninteger) quantization parameters of its two anchor P frames. For a B picture after an I picture, it is quantized by $2 * Q_p$, where Q_p is the quantization parameter of the anchor P frame immediately after the B picture.

VI. Conclusions

The rate control problem for video coding can be separated into problems of bit allocation, rate control, and adaptive quantization. For the rate control part, selecting the quantization scale at the individual MB level would make the problem intractable in practice. We proposed a re-encoding rate control algorithm using local fitting from models estimated on the fly. The algorithm can adapt to image activity changes and can increase and decrease image quality uniformly. The overhead computation is only modest. The experimental coding results are very close to the target bits for several sequences. The proposed rate control algorithm can also be used for MPEG video editing, where segments of video are required to be replaced for editing purpose with up to the same number of bits. Finally, we modified the bit allocation method of [4], and proposed a new bit allocation and rate control algorithm for video coding.

References

[1] ISO-IEC/JTC1/SC29/WG11/602, "Generic coding of moving pictures and associated audio," MPEG Committee Draft: Recommendation H.262, Nov. 1993.

[2] P. de With and A. Rijckaert, "Digital consumer HDTV recording based on motion-compensated DCT coding of video signals," *Signal Processing: Image Commun.*, vol. 4, pp. 401–420, 1992.

[3] P. Tiwari and E. Viscito, "A parallel MPEG-2 video encoder," in *ICASSP*, May 1996.

[4] ISO-IEC/JTC1/SC29/WG11, "Test Model 5," Draft, Apr. 1993.

[5] C.-T. Chen and A. Wong, "A self-governing rate buffer control strategy for pseudoconstant bit rate video coding," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 50–59, Jan. 1993.

[6] E. Viscito and C. Gonzales, "A video compression algorithm with adaptive bit allocation and quantization," in *SPIE Visual Commun. Image Processing*, vol. 1605-1, pp. 58–72, Nov. 1991.

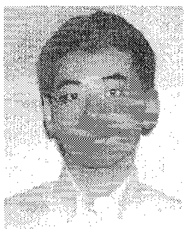
[7] A. Puri and R. Aravind, "Motion-compensated video coding with adaptive perceptual quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 4, pp. 351–361, Dec. 1991.

[8] E. D. Frimout, J. Biemond, and R. L. Lagendijk, "Forward rate control for MPEG recording," in *SPIE Visual Commun. Image Processing*, pp. 184–194, 1993.

[9] S. H. Jang and S. H. Park, "An adaptive rate control algorithm for DPCM/DCT hybrid video codec adopting bi-directional prediction," in *SPIE Visual Commun. Image Processing*, pp. 1237–2094, 1993.

[10] C. A. Gonzales and E. Viscito, "Motion video adaptive quantization in the transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 4, pp. 374–378, Dec. 1991.

[11] J. Lee and B. W. Dickinson, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 513–526, Sep. 1994.
[12] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to MPEG video coders," in *IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. V, 1993, pp. 381–384.
[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.



Wei Ding (S'92–M'95) was born in Beijing, China. He received the B.S.E.E. degree from Beijing University, Beijing, China, in 1989, with the first rank in a class of 70 students, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University in 1992 and 1995, respectively.

He was a research engineer at PECAN Information Technology, Inc., Beijing, China in 1989. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY. His current research interests are in video and image compression and transmission, low bit rate video conferencing and packet video over ATM networks.

Bede Liu (S'55–M'62–F'72) was born in Shanghai, China. He holds the B.S.E.E. degree from the National Taiwan University, and the M.E.E. degree and the D.E.E. degree from the Polytechnic Institute of Brooklyn, NY.

Prior to joining the Princeton UniversityFaculty in 1962, where he is a Professor and Chair of Electrical Engineering, he had been with Bell Laboratories, Allen B. DuMont Laboratory, and Western Electric Company. He has also been a visiting faculty member at the National Taiwan University (Taipei), the University of California (Berkeley), the Imperial College (London), and the Jiao Tong University (Shanghai). His current research interests lie in the image and signal processing.

Dr. Liu served as the Technical Program Chairman for the 1978 IEEE International Symposium on Circuits and Systems and the Géneral Chair of the 1995 IEEE International Conference on Image Processing, He was the President (1982) of the IEEE Circuit and Systems Society, and the IEEE Division I Director (1984, 1985). He was the recipient of an IEEE Centennial Medal, the 1985 Technical Achievement Award of IEEE Acoustics Speech and Signal Processing Society, the 1988 Education Award of the IEEE Circuits and Systems Society, and the Best Paper Award for Video Technology (1994). In 1988, he was named an honorary professor by the Academia Sinica. He edited the research volume *Digital Filters and the Fast Fourier Transform* (DHR, 1975) and coauthored the senior level textbook *Digital Signal Processing: Theory, Design, and Implementation* (Wiley, 1976).