A Sequence-Based Rate Control Framework for Consistent Quality Real-Time Video

Bo Xie and Wenjun Zeng, *Senior Member, IEEE*

***Abstract*—Most model-based rate control solutions have the generally questionable assumption that video sequence is stationary. In addition, they often suffer from the fundamental problem of model parameter misestimation. In this paper, we propose a *sequence-based* frame-level bit allocation framework employing a *rate-complexity* model that has the capability of tracking the nonstationary characteristics in the video source without look-ahead encoding. In addition, a new *nonlinear* model parameter estimation approach is proposed to overcome the existing problems in previous model parameter estimation schemes where quantization parameter (QP) is determined to achieve the allocated bits for a frame. Furthermore, a general concept of bit allocation guarantee is discussed and its importance is highlighted. The proposed rate control solution can achieve smoother video quality with less quality flicker and motion jerkiness. Both a complete solution where requantization is employed to guarantee the achievement of the allocated bits, and a simplified solution without requantization are studied. Experimental results show that they both provide significantly better performance, in terms of average peak-signal-to-noise ratio and quality smoothness, than the MPEG-4 Annex L frame-level rate control solution.**

***Index Terms*—Bit allocation, model parameter estimation, rate control, real-time encoding, smooth quality, stationarity assumption.**

I. Introduction

RATE control is a central piece for standard video codecs (e.g., MPEG-1, MPEG-2, MPEG-4, and H.26x) to achieve *consistently* good quality for the whole video sequence under the channel bandwidth and delay/buffer constraints. In general, rate control includes two parts, one is bit allocation and the other is bit allocation achievement, i.e., quantization parameter (QP) determination for achieving the allocated bits for the current frame or macroblock (MB) accurately.

For bit allocation, MPEG-2 TM5 [1] is a benchmark, i.e., group of pictures (GOP)-based bit allocation in which the GOP size is fixed, and constant bit allocation among GOPs and among frames of the same coding type is used. Many subsequent solutions, for example, MPEG-4 Annex L rate control [2] and ITU H.263+ TMN8 [3] rate control, follow the same spirit, albeit with some improvements. The implicit assumption of this class of bit allocation is that video sequence is stationary and all GOPs have similar characteristics, which is usually questionable. To overcome some of the problems of this GOP-based bit allocation, work has been done in the following three directions.

First, most work focuses on improving the bit allocation *within a GOP*. Some bit allocation schemes [4], [5] allocate more bits to high complexity frames and less bits to low complexity frames within a GOP. Others do rate-distortion (R-D) optimized bit allocation, with or without considering the temporal dependency [6]–[10], or with different optimization criteria (e.g., minimizing the average distortion, or minimizing the distortion variation) [23], [24]. Note that all these bit allocation solutions are restricted to fixed size GOPs, and among GOPs, constant bit allocation is used. As a result, quality fluctuation in a larger scale (e.g., among scenes) is not addressed. Second, since fixed-size GOPs do not match the scene structure in a multiple-scene sequence, dynamic GOP [11]–[14] has been proposed to determine the GOP structure and frame characteristics using some look-ahead techniques. Dynamic GOP solves the scene change problem, but does not address the problem of bit allocation among different GOPs (fixed/dynamic size). Third, GOP-based bit allocation, if not done wisely, is hard to achieve consistent quality across the whole sequence. Scene complexity was taken into account in [4], [5] in an effort to solve the problem of bit allocation among scenes (or GOPs) in a sequence. The authors proposed a perceptual rate-quantization (R-Q) model to improve the bit allocation and combined this model with the R-Q model used in QP determination. However, the scheme used only previous data to estimate the model parameters and determine the R and Q of the current scene or frame, without using the actual coding complexity of the current scene/frame, which may cause some problems because the model parameters estimated based on the previous data may not truly reflect the statistics of the current scene/frame.

For bit allocation achievement, i.e., QP determination, most previous work focuses on developing different kinds of R-Q models, for example, logarithmic [16], [17], power [15], [18], spline [19], polynomial (including linear and quadratic) [20]–[23], exponential [25] models, etc. Yang *et al.* [26] proposed a more complex model that combines a logarithmic and a quadratic model. Recently, He *et al.* [27], [28] proposed ρ-domain source modeling where ρ is defined as the percentage of zero quantized discrete cosine transform (DCT) coefficients. They found linear R-ρ relationship and established a ρ to Q mapping, based on which a more accurate R-Q relationship is built. Note that a potential problem with model-based approaches for QP determination is that there is no guarantee that a model will always accurately characterize the encoder's

Manuscript received November 22, 2003; revised May 4, 2004. This research was supported in part by a grant from the University of Missouri System Research Board, and in part by the National Science Foundation (NSF) under Grant CNS-0423386. Part of this work was presented at *IEEE ICIP'2000*, Rochester, NY. This paper was recommended by Associate Editor K. Aizawa.

B. Xie is with PacketVideo Corporation, San Diego, CA 92121 USA (e-mail: xie@packetvideo.com).

W. Zeng is with the Department of Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA (e-mail: zengw@missouri.edu).

Digital Object Identifier 10.1109/TCSVT.2005.856911

behavior. An exception is Ding and Liu's work [15] where re-encoding to guarantee the achievement of frame-level bit allocation was considered. Other work, for example, operational R-D optimized rate control [8]–[10], does not have the bit allocation achievement problem, but requires high complexity.

In summary, most existing rate control solutions may suffer from two fundamental problems, i.e., stationarity assumption that results in constant bit allocation among GOPs/scenes, and model parameter misestimation. In general, the variability of video sequences has not been sufficiently addressed in these bit allocations. These prior solutions, although working fine for some standard test sequences that are usually simple and more stationary, typically have problems dealing with more complex sequences. We believe that, to achieve consistently good video quality for the whole sequence, *sequence-based*, instead of GOP-based, bit allocation is necessary. The key of this sequence-based bit allocation is to efficiently track the variation of the characteristics across the video sequence.

The model parameter misestimation problem exists in both bit allocation and bit allocation achievement. In general, there are three kinds of problems in model parameter estimation: 1) Stationarity assumption. In particular, only historical data, not the current data, is used for parameter estimation. 2) Linear estimation methods, especially the well-known mean-squared error (MSE)-based linear regression approach, are used in parameter estimation. This linear method may not work well for nonstationary parameter estimation. 3) Data outlier may exist to prevent accurate parameter estimation. This problem becomes more prominent in two-pass rate control solutions where the second pass rate control usually makes use of the first-pass data, assuming that it is reliable.

For model-based QP determination, there exist cases for any model to fail. We refer to this as *model mismatch*. Model mismatch may potentially lead to big discrepancy between the actual encoded bit count and the target (allocated) bit count, and thus may impose penalty for the bit allocation of the future frames. For example, some frames may consume much more bits than expected (allocated), resulting in bit shortage for the future frames. In addition, from buffer management point of view, model mismatch will increase the chance of buffer overflow and underflow, resulting in more *unexpected* frame dropping.

In principle, operational R-D optimization using dynamic programming can be applied to the *whole* sequence to solve all the above problems. However, this is computationally intractable and the delay is usually unacceptable in a real-time system. Our goal is thus to develop a generic rate control solution [30] for *real-time* encoding that intends to solve the above mentioned problems and achieve more consistent video quality across the entire sequence for a large variety of sequences with different characteristics. We propose a fast *sequence-based* bit allocation solution employing a *rate-complexity* model that has certain capability of tracking the variability of nonstationary characteristics in the video source *without* resorting to two-pass or look-ahead encoding. We also provide a solid mechanism to achieve the allocated bits for each frame. This includes a nonlinear approach for improving the R-Q model parameter estimation and an efficient way to do requantization *if necessary*.

We realize that the above-mentioned problems exist in both frame-level rate control and MB-level rate control, but are more prominent in frame-level rate control. In this paper, we will focus on building the framework for addressing the problems in frame-level rate control. The framework may potentially be extended or combined with some previous MB-level rate control mechanisms to further improve the performance, which is a topic of future research.

For comparison, we choose the frame-level rate control in MPEG-4 Annex L [2] as the main reference, as this rate control solution is a typical example that may exhibit many of the problems mentioned above. We will show that both the proposed *complete* solution where requantization is invoked for some frames to guarantee the achievement of the allocated bits, and the *simplified* solution where no requantization is applied, can provide significantly better performance, in terms of average PSNR (peak-signal-to-noise ratio) and quality smoothness, than the MPEG-4 Annex L frame-level rate control solution.

This paper is organized as follows. In Section II, we present our proposed real-time *sequence-based* frame-level bit allocation solution. In Section III, we address the bit allocation achievement issue, and discuss a general concept of bit allocation guarantee. We then propose a new nonlinear model parameter estimation method to improve the accuracy of QP estimation. Finally, we compare our solution with the MPEG-4 Annex L frame-level rate control and show the simulation results in Section IV.

II. Sequence-Based Bit Allocation Framework for Real-Time Video

A. Problem Formulation

The goal of our *sequence-based* bit allocation is to maximize the overall video quality, at the same time minimize/reduce the distortion variation *across* the whole sequence within the constraints of the given target bit rate, frame rate and delay requirement [in terms of video buffering verifier (VBV) buffer size], with only *single-pass* real-time encoding. Note that *single-pass* here refers to only one pass through the sequence without look-ahead, as commonly understood. As will be seen later, one of our solutions includes possible requantization of the *current* frame. The problem is formulated as follow.

Let $\boldsymbol{Q} = \{1, 2, \cdots, 31\}$ be the set of quantization parameter values in standard video codecs, and let $\boldsymbol{B_s}$ be the given maximum buffer size. Find $\boldsymbol{q}^* = \{q_1^*, q_2^*, \cdots, q_{N_{\text{seq}}}^*\}^T$, with $q_i^* \in \boldsymbol{Q}$ for $i = 1, 2, \cdots, N_{\text{seq}}$, where N_{seq} is the total number of frames in the sequence, such that

$$\boldsymbol{q}^* = \arg \min_{\mathbf{q} \in Q^{N_{\text{seq}}}} \frac{1}{N_{\text{seq}}} \sum_{i=1}^{N_{\text{seq}}} \left| d_i(q_i) - \overline{d(\boldsymbol{q})} \right|$$

subject to the constraints

$$0 \le B(i, \boldsymbol{q}) \le \boldsymbol{B_s}, \quad i = 1, 2, \cdots, N_{\text{seq}} \tag{1}$$

where $d_i(q_i)$ is the MSE-based distortion for the ith frame with the quantization parameter q_i, $\overline{d(\boldsymbol{q})}$ is the average distortion for all the frames in a sequence, i.e.,

$\overline{d(\boldsymbol{q})} = (1/N_{\rm seq})\sum_{i=1}^{N_{\rm seq}} d_i(q_i)$, $B(i,\boldsymbol{q})$ is the VBV buffer fullness at the ith frame interval, and is calculated as

$$B(i,\boldsymbol{q}) = max\left(B(i-1,\boldsymbol{q}) + r_i(q_i) - C, 0\right)$$
$$i = 1, 2, \cdots, N_{\rm seq}$$

with

$$B(0,\boldsymbol{q}) = \text{initial buffer fullness} \tag{2}$$

where $r_i(q_i)$ is the actual bit count for the ith frame encoded with q_i, $C =$bit rate/frame rate. Note that the buffer constraints in (1) guarantee that there is no buffer overflow and underflow at the decoder. When the total size of the compressed bitstream is sufficiently large with respect to the VBV buffer size, the buffer constraints also guarantee the convergence of the actual encoded bit rate to the target bit rate.[1]

In the *single-pass* real-time encoding scenario, a global optimal solution to (1) is impossible, due to the unavailability of future frames. Nevertheless, at a specific encoding time, it is possible to develop a greedy suboptimal solution based on all *available* information of the frames that have already been encoded, as will be shown in the following. In particular, instead of explicitly minimizing the distortion variation or the average distortion across the frames, we address this problem by exploring a *global* bit allocation model we developed that aims to characterize the relationship between the appropriate amount of allocated bits and a coding complexity measure of a frame.

B. Global Bit Allocation Model

The basic idea of our proposed sequence-based frame-level bit allocation solution is to develop an effective and quantifiable way to allocate more bits to high complexity scenes/frames (such as scene change frames), and less bits to low complexity scenes/frames, such that consistently good quality across the sequence can be achieved. We will refer to the *coding complexity* of a frame as the number of bits required to encode such frame to achieve *similar* quality as other frames. The goal is to develop a good model that would tell the right amount of bits to be allocated based on an appropriate coding complexity measure of the frames.

Traditionally, the variance of the DCT coefficients of the motion compensated difference frames σ^2 is often used to measure the coding complexity of a frame, based on the Laplacian distribution assumption [6]. For example, for *independent* video sources, or video frame encoding without considering the frame dependency, Ribas and Lei [6], [7] used Lagrangian multiplier optimization to calculate the optimal bit allocation *within a coding unit* (e.g., a GOP).

Frame-level bit allocation

$$T_j^* = \left(T - AN\sum_{m=1}^{M} C_m\right)\frac{S_j}{\sum\limits_{m=1}^{M} S_m} + ANC_j \tag{3}$$

where $S_j = \sum_{i=1}^{N}\sigma_{i,j}^2$, which can be interpreted as the energy of the jth frame, and $\sigma_{i,j}^2$ is the variance of the ith MB in the jth frame. T_j^* and T, respectively, are the "optimal" target bit count for the jth frame and the total bit budget for the corresponding GOP, M is the number of frames in the GOP, and A, N, and C_j are the number of pixels for each MB, the number of MBs in each frame, and the average rate (in bits/pixel) to encode the motion vectors, header and other syntax for the jth frame, respectively. This optimal bit allocation for *independent* sources/frames *within a coding unit* suggests a linear relationship between the target allocated bits and the coding complexity σ^2.

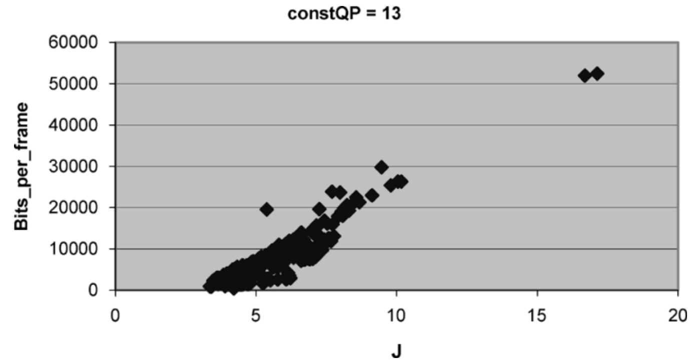


Fig. 1. R-J results for the movie sequence "Hangingup" with constant $\mathrm{QP}(\lambda = 2.3 * \mathrm{QP})$. "Hangingup": 320 × 224, 300 frames, lots of scene change and high motion.

We have studied an alternative coding complexity measure that accounts for both motion vector coding and residue coding, i.e.,

$$J = mdev + \lambda \cdot \overline{R_{\rm motion}} \tag{4}$$

where $mdev$ is referred to as the mean deviation, and is defined as $mdev = (1/N)\sum_{i=1}^{N} mdev_i$, $mdev_i = (1/A)\sum_{j=1}^{A}|x_{j,i} - \overline{x_i}|$, $x_{j,i}$ and $\overline{x_i}$ are the intensity of the j-th residue pixel of the i-th MB and the average residue intensity of the i-th MB, respectively;$\overline{R_{\rm motion}}$ is defined as $\overline{R_{\rm motion}} = (1/N)R_{\rm motion}$, where $R_{\rm motion}$ is the total coding bits spent on the motion vectors of a frame, λ is a scaling factor to control the relative contribution from the motion information and the residue. If $\lambda = 0$, then J measure becomes very similar to σ^2.

It is desirable that the appropriate rate versus J (R-J) relationship can be analyzed to facilitate bit allocation. In general, the goal of the bit allocation is to achieve small distortion variation across the sequence while minimizing the overall average distortion. It is well known that the asymptotically optimal bit allocation for coding *independent* sources at high bit rates will result in constant distortion across the sources [31]. In fact, many experiments have indicated that using constant QP for the entire video sequence typically results in good performance (in terms of both average PSNR and consistent quality). Therefore, we build the rate-complexity model by analyzing the actual data generated by coding using constant QPs across frames of the whole sequence. Fig. 1 shows the R-J curve with constant QP equal to 13 for the entire video sequence, where the scaling factor λ is chosen as $\lambda = k \cdot \mathrm{QP}$, which is similar to that used for rate-distortion-based motion estimation in [29]. In the experiment, we found that when λ is within the range of $[1.8 * \mathrm{QP}, 3 * \mathrm{QP}]$, the R-J data shows some linear relationship.

The measures of J or σ^2, however, have some inherent problems. Bit allocation based on linear R-J or R-σ^2 relationship

[1]The VBV buffer size signifies the maximum possible difference between the actual size and the target size of the compressed bitstream. The actual bit *rate* will thus converge to the target bit rate asymptotically.

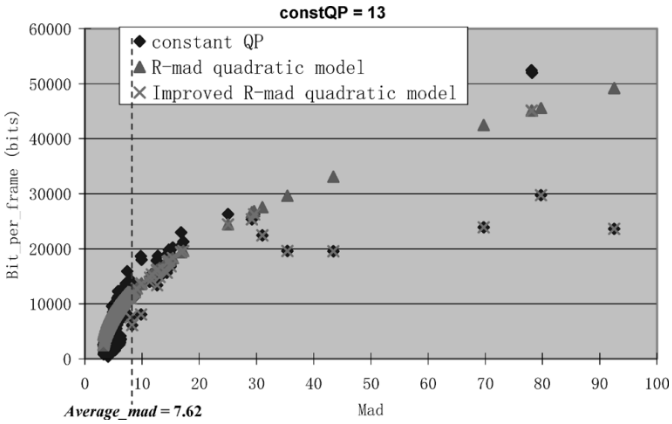


Fig. 2. R-MAD modeling results for the movie sequence "Hangingup" with constant $\mathrm{QP} = 13$, and $K = 5200$.

is more suitable for the scenario where each frame is *independently* coded because they are derived without addressing the dependency issue in video coding. These measures, therefore, may not effectively differentiate I-frames from P/B-frames, resulting in overall nonoptimal performance. In addition, J or σ^2 does not account for the coding complexity incurred by coding the DC coefficients of I-frames or I-blocks, which will further penalize I-frames. In fact, we found in our experiments that in many cases, J or σ^2 of the I-frames or scene change frames are very close to those of the low-motion P-frames, resulting in low quality for I or scene change frames. Therefore, these two coding complexity measures are not sufficient to achieve our goal of achieving consistently good quality using sequence-based bit allocation.

We, therefore, propose to use mean absolute difference (MAD) as an alternative for the coding complexity measure. For I-frames or intra-MBs, we use the absolute value of each *original* pixel in calculating the MAD, while the absolute value of the residue/difference pixel is used for inter-coded MBs. This is to differentiate intra-coded MBs from inter-coded MBs, thus further differentiate I or scene change frames from P- or B-frames.

Fig. 2 shows the actual R-MAD data when constant QP is used for coding the entire sequence. To characterize the R-MAD relationship, we choose the following model:

$$R = K * (\mathrm{MAD} - \alpha)^{\beta}, \quad (\mathrm{MAD} \geq \alpha \geq 0, \beta > 0). \tag{5}$$

where K and β are some constants that control the slope of the R-MAD curve, while α is an offset that aims to account for the quantization effect (i.e., when MAD is small, the quantized DCT coefficients may become zero, resulting in zero coding bits.). In our experiments, we use 0.5 for β, which corresponds to a quadratic model. The constant K can be estimated from the actual data.

We can see that the proposed quadratic model basically follows the trend of the actual curve, especially for frames with low to mid MAD. For frames with high MAD (which typically correspond to I or scene change frames), there is an apparent discrepancy between the actual number of bits and the number of bits derived from the model, i.e., the proposed quadratic model tends to prioritize the I-frames or scene change frames, which may not be too bad from the coding dependency point of view. Nevertheless, to make sure that the bit allocation model does not over-prioritize those frames with large MAD, we further impose a constraint on the bit allocation, i.e., the QP value derived from the bits allocated by the R-MAD model for a frame that has larger MAD than the average MAD of all previously encoded frames should not be smaller than the average QP of all previously encoded frames. This constraint basically suggests that if a frame has a MAD that is larger than the average MAD, and its QP derived from the quadratic R-MAD model is smaller than the average QP, then we simply use the estimated average QP to encode such a frame. The quadratic R-MAD model with such additional constraint will be referred to as the *improved quadratic* R-MAD *model*. Fig. 2 shows the results of the improved quadratic R-MAD model where for frames with a MAD lower than the average MAD, the quadratic model is used; while for most of the frames that have a MAD that is larger than the average MAD, it resorts to using the constant (or average) QP, because the derived QPs from the R-MAD model (over-allocated bits) would be smaller than the average QP. More details will be discussed in Section III-C. Note that a more complex model such as a higher order polynomial model or a hybrid model may also be used to better approximate the actual curve, with the cost of extra complexity. Considering the real-time constraint we intend to address in this paper, we leave this investigation for future work.

It should be pointed out that, in terms of bit allocation modeling, the proposed improved R-MAD model and R-J model both approximate the constant QP solution reasonably well. Note that the assumption here is that the constant QP solution generally provides consistent reconstructed image quality across the entire video sequence. The main difference lies in the different properties of the two complexity measures. As mentioned above, MAD measure is able to differentiate I-frames from P/B-frames. Thus, the bit allocation for I-frames and P/B-frames can be effectively differentiated as well. However, the J measure fails to achieve this. As a result, the bit variation shown in the R-J curve (i.e., different R's given the same J) in Fig. 1 may represent the variation of the bit allocation for *different types* of frames (e.g., I-frames or P/B-frames), as opposed to the variation of the bit allocation for the *same type* of frames as shown in the R-MAD curves in Fig. 2. The inability to differentiate scene change frames (or I-frames) from less complex frames becomes a critical problem in R-J modeling.

Note that to match the actual data well, we introduce the shifting factor α in the model. In general, this shifting factor should be very small. In the implementation, we use the following model for bit allocation

$$R = K * \sqrt{\mathrm{MAD}}. \tag{6}$$

C. Target Bit Calculation

As discussed above, without being constrained to the GOP structure, our goal is to allocate bits in a way such that consistently good video quality across the whole sequence will be achieved. To that end, based on the experimental data (e.g., in Fig. 2), we conclude that there exists an appropriate R-MAD relationship characterized by the model we introduced in (6) for the whole sequence, given the target bit rate, frame rate, buffer constraint and other encoder settings, which we refer to as a *specific coding scenario*. For each specific coding scenario, there will be a fixed sequence-based K in our model. This fixed K for a specific coding scenario, however, can not be accurately determined until the whole sequence is available or encoded. This is similar to (3) where the slope of the linear $T_j - S_j$ relationship can not be completely determined until all S_m's are available. Within the constraint of one-pass real-time encoding, at any encoding time instance, the best thing we can do is to estimate the K value based on all frames that have already been encoded, for example, using the least MSE-based estimation.

Specifically, based on (6), we propose to use the following bit allocation

$$T_n = C * \sqrt{\frac{\mathrm{MAD}_n}{\overline{\mathrm{MAD}}_{n-1}}}. \tag{7}$$

where C is the VBV buffer output rate, i.e., $C =$bit rate/frame rate, T_n is the target bit allocated for the current frame, MAD_n is the MAD of the current frame, and $\overline{\mathrm{MAD}}_{n-1}$ is the average MAD of all the previously encoded frames. This effectively estimates the K value as $C/\sqrt{\overline{\mathrm{MAD}}_{n-1}}$. This estimate is not the least MSE solution, but is simpler to compute.

In (7), we choose C instead of the average bits per frame of all the previous encoded frames. The reasons are two folds. First, the average bit count per frame of the previous encoded frames should converge asymptotically to C, e.g., when the buffer size is negligible as compared to the total size of the already compressed bitstream. Second, C can help correct the bit allocation "error" and prevent "error propagation" resulted from some bad bit allocation, which is likely to happen, for example, to the frames at the beginning of the sequence; while using the average bit count per frame of the previous encoded frames may result in bit allocation error propagated to the following frames.

In (7), $\mathrm{MAD}_n/\overline{\mathrm{MAD}}_{n-1}$ indicates how complex the current frame is, compared with all the available encoded frames. For scene change frames or high complexity scenes/frames, $\mathrm{MAD}_n$is usually larger than the average complexity $\overline{\mathrm{MAD}}_{n-1}$, thus a larger number of bits will be allocated to such frames. While for low complexity frames/scenes, a smaller number of bits will be allocated. Since $\mathrm{MAD}_n/\overline{\mathrm{MAD}}_{n-1}$ can effectively track the complexity variation of each frame across the whole sequence, we feel that scene change detection and the bit allocation for the scene change frames can be easily addressed. In fact, no separate data pre-analysis for scene change detection is necessary. In addition, since MAD is capable of differentiating the coding complexity of I-, P-, and B-frames, bit allocations for I-, P-, and B-frames in the sequence are determined in a *unified* way.

For the proposed bit allocation solution, the stationarity assumption is no longer necessary, as opposed to the GOP-based bit allocation. The only factor that may affect the performance of the proposed solution is the closeness of each instantaneously estimated value of K (or equivalently $\overline{\mathrm{MAD}}_{n-1}$) to the K (or $\overline{\mathrm{MAD}}$) value that is estimated based on the whole sequence. Typically the estimate is getting better and better as more and more frames have been encoded. In addition, if the scene complexity fluctuates across the sequence, then it is likely we will have a good instantaneously estimated K value (therefore good rate control) after the first couple of scenes have been encoded. The worst case is that the scenes get increasingly more complex (or simpler) throughout the whole sequence. Then it is generally very difficult to achieve consistent quality across the whole sequence for any *single pass* encoding solution where no information about the future frames is available. Note also that the coding complexity of the *current* frame, i.e., current MAD, is used in our solution; while in most existing GOP-based bit allocation schemes, only previous source data has been used, usually based on the assumption that the previous source data and the current source data follow similar statistics, which is not true in general. This signifies a significant advantage of the proposed scheme over some of the previous schemes, especially for scene change frames where the stationary assumption is violated.

After this sequence-based bit allocation, the existing discrepancy between the actual bit count and the target bit count needs to be taken into account. Thus, one minor adjustment for the current frame is made

$$T_n = T_n - \Delta_n, \quad \Delta_n = \sum_{i<n}(R_i - T_i) \tag{8}$$

where R_i and T_i are the actual bit count and the target bit count for the previous frames. Note that Δ_n represents the cumulative difference between the actual bit count and the target bit count.

Up to this point, the target bit count for the current frame is calculated from the consistently good quality point of view. Next, we have to consider the buffer constraint to avoid buffer overflow or underflow, as is done in all previous rate-control techniques.

$$\begin{aligned}
&\texttt{If } (\mathrm{VBV_}\texttt{fullness} + T_n - C > B_s)\\
&\qquad T_n = B_s - \mathrm{VBV_}\texttt{fullness} + C;\\
&\qquad\qquad /^{*}\texttt{avoid buffer overflow}^{*}/\\
&\texttt{Else if } (\mathrm{VBV_}\texttt{fullness} + T_n - C < 0)\\
&\qquad T_n = C - \mathrm{VBV_}\texttt{fullness};\\
&\qquad\qquad /^{*}\texttt{avoid buffer underflow}^{*}/
\end{aligned} \tag{9}$$

where VBV_fullness shows the buffer usage, and B_s is the buffer size.

We have derived the target bit for the current frame that meets both the goal of achieving consistently good quality and the buffer constraint without being restricted to the GOP structure. As pointed out previously, meeting the buffer constraint will automatically guarantee the convergence of the actual bit rate to the target bit rate, as long as the buffer size is negligible with respect to the total size of the compressed sequence, which is usually the case. For variable bit rate (VBR) coding with a very large (or unlimited size) buffer, to guarantee this convergence, we can impose an appropriate "virtual" buffer size constraint in the rate control process.

III. Quantization Parameter Determination

A. Bit Allocation Guarantee

After target bit calculation, quantization parameter will be determined. An important concept in quantization parameter determination is *bit allocation guarantee*, i.e., the achievement of the allocated target bit count should be enforced. Without bit allocation guarantee, a good bit allocation scheme will not produce expected performance. In practice, given a target bit budget allocated for a frame, typically a R-Q model is used to determine an appropriate QP for such target bits. Therefore, an important step for bit allocation guarantee is to develop an accurate R-Q model, as evidenced in many prior works in the literature. We will propose an improved R-Q model using a *nonlinear* approach in Section III-B. The concept of bit allocation guarantee we discuss here, however, is more generic in the sense that it should be able to handle the cases where the R-Q model fails.

The benefit of bit allocation guarantee is further highlighted in the following. In traditional model-based QP determination, such as MPEG-4 Annex L, since there is no bit allocation guarantee, there may exist "error propagation" in the bit allocation process (i.e., bit overuse for one frame will impose penalty for the future frames). As a result, the role of VBV buffer turns out to be *passively* absorbing the bit discrepancy between the allocated bit and the actual bit, which is not easily predictable. In order to reduce the risk of buffer overflow and underflow, the bit allocation has to be conservative, e.g., using constant bit allocation. On the other hand, suppose we can achieve 100% bit allocation guarantee, then the VBV buffer can be used to accommodate bit allocation *variation* among frames. In other words, we can choose a more aggressive and better bit allocation scheme that makes sufficient use of the flexibility provided by the VBV buffer to improve the encoding quality. In this sense, bit allocation guarantee serves as the basis for any effective variable rate bit allocation scheme, including our proposed sequence-based bit allocation solution.

To address the model failure issue, we extend the idea in Ding and Liu's work [15], and perform requantization (only if necessary) to achieve the bit allocation guarantee. In particular, if $|\mathrm{Actualbit} - \mathrm{Targetbit}|/\mathrm{Targetbit} > \mathrm{Threshold}$, where $Actual\ bit$ is the actual bit count resulted from operational encoding, then we will do QP re-adjustment and entropy coding for the current frame; otherwise the normal encoding process follows. The difference between our scheme and Ding and Liu's work lies in the initial QP determination. In their work, two initial QPs are required, and at least two requantizations (three candidate QPs) for their proposed R-Q model are needed, which is not efficient. In our QP determination scheme, any model-based initial QP calculation could be used, and if the model used is accurate enough, then requantization is not necessary. We will show in Section IV that the simple quadratic R-Q model and the nonlinear parameter estimation algorithm we propose in the next sub-section are often sufficiently accurate to achieve the allocated bit without requantization. Therefore, the requantization process in our scheme only serves as the safeguard to guarantee bit allocation achievement, and is invoked *only if necessary*.

B. New Non-Linear Approach for Improving R-Q Model Parameter Estimation

To achieve bit allocation guarantee, besides addressing the model failure problem, there is a constant need for improving the R-Q model accuracy. Section I lists a number of such works. However, we realize that there is another fundamental issue in most of the previous works that may greatly affect the accuracy of the QP determination, i.e., model parameter estimation. As pointed out in Section I, there might be three problems in previous QP determination solutions: stationarity assumption (e.g., only previous data was used in parameter estimation), linear estimation method, and data outlier. In this section, we propose a generic approach to solve these problems in a coherent way.

The basic idea is borrowed from the approximation theory. A more accurate approximation to a *general* function can be achieved when the area where the approximation is applied becomes smaller, regardless of the actual form of the approximation/model. Therefore, the idea is to find the smallest area around the reference point (e.g., data of the current frame) to apply an approximation/model such that the most accurate estimation can be achieved.

In the following, we will apply this general idea to the simple quadratic R-Q model $R/\mathrm{MAD} = X/Q^2$ where X is the sole parameter), which is similar to the model used in MPEG-4 Annex L, i.e., $R/\mathrm{MAD} = (X1/Q) + (X2/Q^2)$ (where $X1$, $X2$ are two parameters), to solve the QP determination problem. This R-Q model is different from the R-MAD model in (6) which addresses a completely different problem, i.e., the problem of frame-level bit allocation to achieve consistent quality.

1) Initial QP Determination: Due to potential requantization, there *may* be several coding iterations with different QPs and different actual number of encoded bits during the encoding of *each* previous frame. So the basic idea is to pick up a particular coding iteration for certain frame in the past (restricted by a window), which has the most similar statistical characteristics as the current frame. Then we use the data of that particular coding iteration of that particular previous frame to estimate the parameter (X) of the R-Q model to be used for calculating the current QP. We use two metrics, MAD and $\mathrm{Actualbit}$, to describe the statistical similarity between two frames.

Specifically, suppose we have the historical data set $\boldsymbol{X} = \{(A(i,j), Q(i,j), \mathrm{MAD}(i))\}_{i=n-1,j=1}^{i=n-M,j=k_i}$, where $A(i,j)$ and $Q(i,j)$ are, respectively, the actual number of bits and the QP used for the jth coding iteration of the ith frame in the past; $\mathrm{MAD}(i)$ is the MAD for the ith frame in the past; M is the window size that specifies the number of frames in the past that will be considered; K_i is the total number of coding iterations of the ith frame; n represents the current frame. For the current frame, the data we can obtain before quantization and entropy coding are target bits $T(n)$ and $\mathrm{MAD}(n)$. Note that if no requantization is employed, e.g., in the *Simplified Solution* referred to in Section IV, K_i will always be 1.

We first search the historical data set to get

$$p^* = \arg \min_{n-1 \le i \le n-M} \left(|\mathrm{MAD}(i) - \mathrm{MAD}(n)| \right), \quad n-1 \le p^* \le n-M, \quad (10)$$

Then

$$(p^*, q^*) = \arg \min_{1\le j\le K_{p^*}} \left(|A(p^*, j) - T(n)|\right), \quad 1 \le q^* \le K_{p^*}. \tag{11}$$

Then by applying the quadratic model $R/\mathrm{MAD} = x/Q^2$, we can get the current QP

$$\mathrm{QP}(n) = \mathrm{QP}(p^*, q^*) * \sqrt{\frac{\frac{A(p^*,q^*)}{\mathrm{MAD}(p^*)}}{\frac{T(n)}{\mathrm{MAD}(n)}}}. \tag{12}$$

In summary, based on the above search [i.e., (10) and (11)], the smallest area around the current point/frame for the polynomial approximation is found. As a result, accurate parameter estimation is achieved. Note that, with such *nonlinear* approach, the stationarity assumption is no longer needed and data outlier will be eliminated in general. On the contrary, the traditional MSE estimator is subject to the average effect of all the data in the data set $\boldsymbol{X}$, relying on the stationarity assumption. In addition, any data outlier would easily ruin the approximation. In Section IV, we will show that the proposed nonlinear parameter estimation algorithm is much more efficient than the MSE estimation used in MPEG-4 Annex L.

2) QP Re-Adjustment: The above idea can be extended to the requantization process of the current frame. As will be shown in Section IV, the initial QP calculated using the proposed nonlinear parameter estimation scheme is accurate for most of the frames. As a result, on average, only less than 30% frames may need requantization. If QP readjustment is necessary for the current frame, we will make use of the *actual* number of bits associated with the $\mathrm{QP}(s)$ used in the previous quantization iterations of the current frame. We then use the quadratic R-Q model to calculate the updated QP for the new quantization iteration. In this case, the above two metrics reduce to one, i.e., $\mathrm{Actualbit}$. In particular

$$\begin{aligned}\mathrm{QP\_new} &= \mathrm{QP\_prev} * \sqrt{\frac{\frac{\mathrm{Actual\_bit\_prev}}{\mathrm{MAD}}}{\frac{\mathrm{Target\_bit}}{\mathrm{MAD}}}} \\ &= \mathrm{QP\_prev} * \sqrt{\frac{\mathrm{Actual\_bit\_prev}}{\mathrm{Target\_bit}}}. \end{aligned} \tag{13}$$

Since the actual encoding data ($\mathrm{QP\_prev}$ and $\mathrm{Actual\_bit\_prev}$) of the current frame is used in the parameter estimation, the QP estimation is typically very accurate, and typically only one extra requantization is sufficient to meet the target bit requirement, as will be shown in Section IV. Since the quadratic model sometimes may not be sufficiently accurate, the quantization iteration based on the above equation alone may not guarantee 100% convergence of the actual bit count to the target bit count. Therefore, in our implementation, we combine (13) with the following variant of the bisection algorithm to achieve the final convergence. The basic idea is, during each iteration, we first calculate the QP of the current iteration using (13). If the calculated QP is within an appropriate QP range, then we use this QP to do the encoding; otherwise we take the median of the current QP range as the new QP. At the end of each iteration, the QP range will be updated. The pseudocode for this process is shown below.

We define the following two integer variables $\mathrm{QP\_left}$ and $\mathrm{QP\_right}$ to determine the QP range ($\mathrm{QP\_left}$, $\mathrm{QP\_right}$) where final QP should be located.

1) Initialization: $\mathrm{QP\_left} = 1$; $\mathrm{QP\_right} = 31$;
2) Determine the current QP:

```
First calculate the current QP based
on (13)
If (QP <= QP_left||QP >= QP_right) {
    /* QP is outside QP range, do bi-
section */
    QP = (QP_left + QP_right)/2;
}
Encode the frame using the current QP;
```

3) Update $\mathrm{QP\_left}$ and $\mathrm{QP\_right}$:

```
for the current QP,
if (|Actualbit − Targetbit|/Targetbit > Threshold)
{
  /*current QP not accurate enough*/
  if (Actualbit > Targetbit)
    QP_left = QP(current iteration);
  else
    QP_right = QP(current iteration);
  Go to Step 2).
}
else
  Stop the process
```

C. QP Final Sanity Check

After the QP determination (with or without the above QP re-adjustment process), we propose to do the following sanity check for QP, as discussed in Section II-B. Specifically, for frames with higher complexity (i.e., $\mathrm{MAD}_n > \overline{\mathrm{MAD}}_{n-1}$), if the calculated QP is smaller than the average QP of all previously encoded frames, then the current QP will be clipped to the average QP, i.e., if

$$(\mathrm{MAD}_n > \overline{\mathrm{MAD}}_{n-1} \;\&\&\; \mathrm{QP_n} < \overline{\mathrm{QP}}_{n-1})$$

and

$$\mathrm{QP}_n = \overline{\mathrm{QP}}_{n-1} \tag{14}$$

where QP_n is the QP of the current frame determined by the above models, $\overline{\mathrm{QP}}_{n-1}$ is the average QP of all previously encoded frames.

As discussed in Section II-B, the main reason for doing this sanity check is to avoid the potential negative impact of the large discrepancy between our proposed quadratic bit allocation model (6) and the actual data, for frames with larger MAD, as shown in Fig. 2. Intuitively, (14) is a good constraint and would help prevent over-prioritizing frames with high MAD and achieve consistent quality across the sequence.

After final QP determination, we will do the actual encoding and update the parameters Δ_n [as shown in (8)] and $\mathrm{VBV\_fullness}$ in our rate control. In addition, we need to check if frame dropping is needed to avoid buffer overflow after the current buffer status has been updated. We use the same procedure as used in MPEG-4 Annex L

```
If (VBV_fullness >= Bs){ /* Bs is the
VBV buffer size*/
  Drop the current frame;
  Update VBV_fullness accordingly;
  }
Else {
  while (VBV_fullness > Bs * 80%)
  {

    Drop the next frame;
    Update VBV_fullness accordingly;   (15)
  }
}
```

IV. Simulation Results

We did three sets of simulations to evaluate the performance of the proposed rate control framework. First we verify the accuracy of the proposed nonlinear R-Q model parameter estimation for QP determination, considering that this is an important and independent component in our solution. Then we compare the overall performance between the proposed rate control framework and the MPEG-4 Annex L frame-level rate control solution. Finally we compare the effectiveness of two coding complexity measures we proposed in Section II, i.e., J and MAD, for bit allocation modeling within our framework, to further justify our choice of using MAD as the coding complexity measure.

Note that we are more focusing on low bit rate applications (e.g., 2.5 G or 3 G wireless applications) where the bit rates and frame rates are relatively low. In such application scenarios, frame dropping will have a bigger impact on the perceived video playback quality, e.g., it is easier to notice the motion jerkiness when some frames are dropped. It is also easier to notice the quality fluctuation in such low bit rate application scenarios. Furthermore, the test sequences we used include a lot of complex sequences that are generally more difficult to handle than most of the standard test sequences. This is to address the diverse applications of MPEG-4, and get a better understanding of to what extent the proposed solution, and in general, a one-pass rate control solution, can work effectively in a practical system.

A. Performance Comparison of Two R-Q Models for QP Determination

We compare two approaches. One is to use the quadratic R-Q model in MPEG-4 Annex L (where Least MSE estimation is used) for initial QP determination, and then use the traditional bisection search algorithm (over the QP range of [1], [31]) for QP re-adjustment. The other approach is the one we proposed in Section III. We define the R-Q model accuracy as the probability of the event $\{|A - T|/T < \mathrm{Threshold}\}$, i.e., $P\{|A - T|/T < \mathrm{Threshold}\}$, where A and T are the actual bit count and the target bit count of a frame, respectively. To ensure a fair comparison, we need to maintain similar bit allocation for both models. This is achieved by using the *same* bit allocation scheme and by using the requantization process to ensure bit allocation achievement. Two variables are defined: $\mathrm{re\_quantized\_frames}$ represents the total number of frames that require requantization to meet the criterion $\{|A - T|/T < \mathrm{Threshold}\}$ (which signifies a model failure); and $\mathrm{re\_quantized\_times}$ records the total number of requantization iterations. Thus, $P\{|A - T|/T < \mathrm{Threshold}\}$ can be approximated as $1 - \mathrm{re\_quantized\_frames}/\mathrm{total\_encoded\_frames}$. The variable $\mathrm{re\_quantized\_times}$ signifies the extra computational complexity introduced by requantization.

In the simulation, we used constant bit allocation among frames and set $\mathrm{Threshold}$ to be 30%. We chose QCIF (176 $\times$ 144) "Foreman" (400 frames, medium motion, one scene change), "Sea World" (600 frames, high motion, a couple of scene changes), "Glasgow" (750 frames, lots of scene cuts) and "Charlies Angels" (1000 frames, action movie, lots of scene changes and high motion) as the test sequences.

The simulation results are shown in Table I and Table II. Table II is a summary of Table I. From Table II, we can see that our proposed R-Q modeling approach improves the model accuracy rate by 29.5% up to 56%, or on average 40.7%, over the MPEG-4 Annex L quadratic R-Q model. Note that this is consistent with the frame dropping comparison to be shown in the next sub-section. Since our proposed R-Q model is much more accurate than the MPEG-4 Annex L quadratic R-Q model, it significantly reduces the occasions of frame dropping, even *without* requantization, as shown in Section IV-B. From the computation point of view, our approach can save 60.7% up to 226.2%, or on average 139% requantization effort. When compared to an encoder without using any requantization, our approach with requantization incurs, on average, only 30% extra quantization. For most of the frames that need requantization, only one extra requantization is sufficient. Note that the quantization and entropy-coding modules do not contribute significantly to the overall video encoding complexity. Therefore, real time encoding can be achieved in our approach, even when requantization is employed.

B. Comparison of the Overall Encoder Performance

We applied the proposed rate control solution to MPEG-4 simple profile encoder [2] and compared it with the MPEG-4 Annex L frame-level rate control.

For the proposed solution, we developed two versions. One is the *Complete Solution*, which includes the proposed sequence-based bit allocation, and the new R-Q model parameter estimation for QP determination with requantization incorporated. The other is the *Simplified Solution*, which is the same as the *Complete Solution* except that requantization is NOT used. This aims to evaluate the validity and accuracy of the proposed R-Q model and its impact on the encoder performance.

The basic coding settings are: no B-frames, combined mode with no error resilience (no data partitioning and resync marker); 10-s I-frame refresh time; four motion vectors in motion estimation with 16×16 motion search range; and 0.5 s VBV buffer size. We again use the same set of test sequences as used in Section IV-A. For MPEG-4 Annex L, we set the initial QP for the first I-frame to 10, and the allocated bits for

TABLE I
COMPARISONS BETWEEN TWO R-Q MODELS ON MODEL FAILURE RATE AND RE-QUANTIZATION COMPLEXITY. (1ST LINE: ANNEX L QUADRATIC MODEL; 2ND LINE: THE PROPOSED MODEL). MODEL FAILURE RATE IS DEFINED AS THE RATIO OF $Re\_quantized\_frames$ TO THE *TOTAL ENCODED FRAMES*

Sequence	Bitrate(bps)@ Framerate(fps)	*Re_quantized_frames* (Model failure rate, Total encoded frames)	*Re_quantized_times* (percentage)
Foreman	48k@7.5	80 (80.0%, 100) 24 (24.0%, 100)	215 (215.0%) 25 (25.0%)
	64k@10	88 (65.7%, 134) 40 (29.9%, 134)	259 (193.3%) 44 (32.8%)
	128k@15	119 (59.5%, 200) 60 (30.0%, 200)	418 (209.0%) 61(30.5%)
Sea World	48k@7.5	101 (67.3%, 150) 27(18.0%, 150)	270 (180.0%) 29 (19.3%)
	64k@10	124 (62.0%, 200) 16 (8.0%, 200)	289 (144.5%) 18 (9.0%)
	128k@15	141 (47.0%, 300) 34 (11.3%, 300)	218 (72.7%) 36 (12.0%)
Glasgow	48k@7.5	140 (74.5%, 188) 63 (33.5%, 188)	296 (157.4%) 77 (41.0%)
	64k@10	194 (77.6%, 250) 88 (35.2%, 250)	439 (175.6%) 109 (43.6%)
	128k@15	281 (74.9%, 375) 150 (40.0%, 375)	769 (273.7%) 178 (47.5%)
Charles Angeles	48k@7.5	157 (62.8%, 250) 70 (28.0%, 250)	323 (129.2%) 86 (34.4%)
	64k@10	227 (68.0%, 334) 84 (25.1%, 334)	478 (143.1%) 108 (32.3%)
	128k@15	318 (63.6%, 500) 154 (30.8%, 500)	705 (141.0%) 192 (38.4%)
On Average (%)		67% 26%	169% 30%

TABLE II
R-Q MODEL ACCURACY IMPROVEMENT AND RE-QUANTIZATION SAVING OF THE PROPOSED MODEL, AS COMPARED TO THE ANNEX L QUADRATIC MODEL

Sequence	Bitrate(bps)@ Framerate(fps)	Model accuracy Improvement	Re-quantization Saving (%)
Foreman	48k@7.5	+56%	190%
	64k@10	+35.8%	160.5%
	128k@15	+29.5%	178.5%
Sea World	48k@7.5	+49.3%	160.7%
	64k@10	+54%	135.5%
	128k@15	+35.7%	60.7%
Glasgow	48k@7.5	+41%	116.4%
	64k@10	+42.4%	132%
	128k@15	+34.9%	226.2%
Charles Angeles	48k@7.5	+34.8%	94.8%
	64k@10	+42.9%	110.8%
	128k@15	+32.8%	102.6%
On Average		+40.7%	139%

Note: Model accuracy improvement = (Proposed model accuracy) − (Annex L quadratic model accuracy). Requantization saving = (Requantized_times (Annex L quadratic model) − Requantized_times (proposed model))/ Total encoded frames.

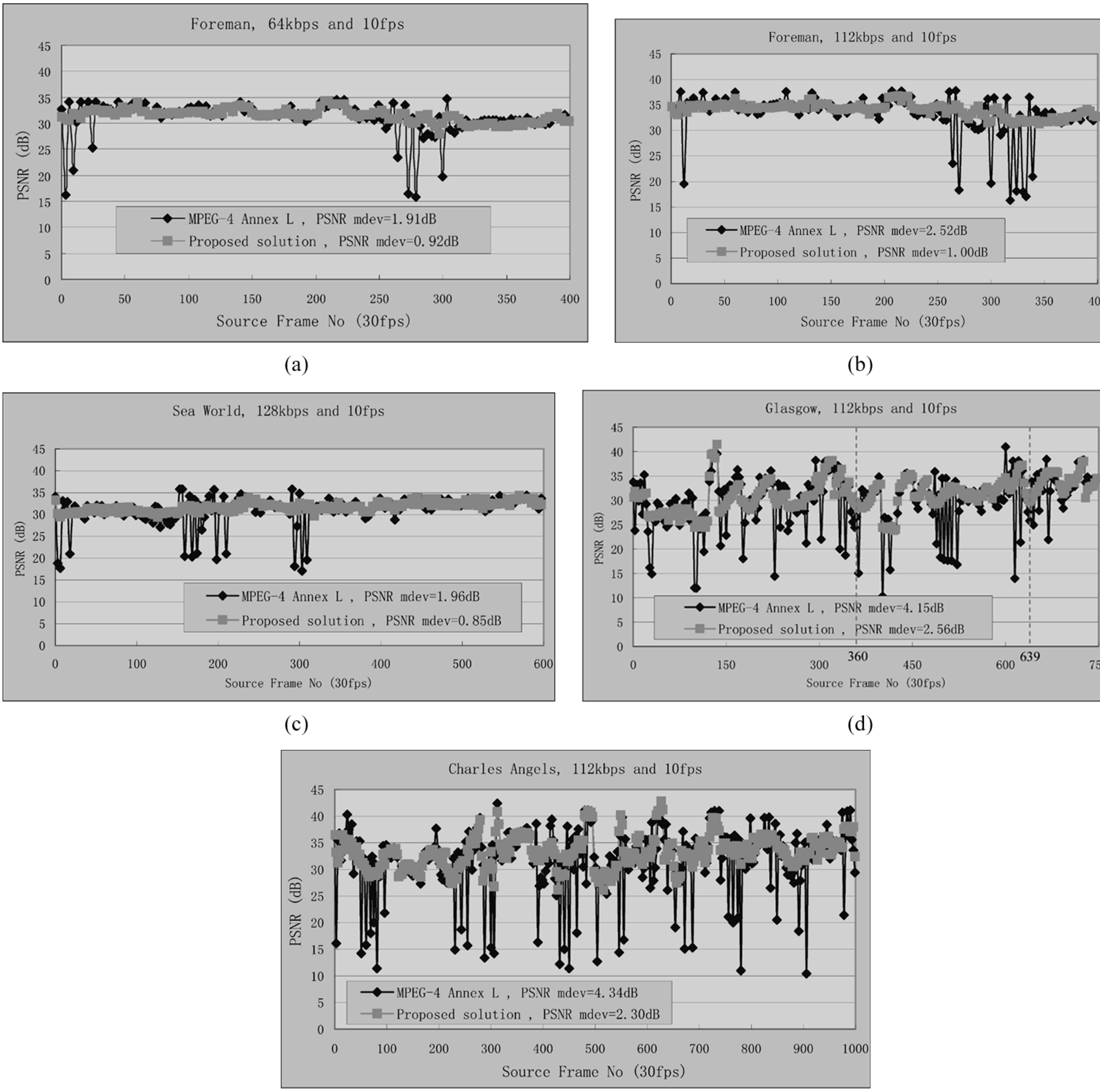
other I-frames are three times as many as the up-to-date average allocated bits of P-frames. In our *Complete Solution*, the first I-frame bit allocation is chosen as 40% of the available VBV buffer (i.e., $\mathrm{VBV\ buffer\ size}/2 * 40\%$). For our *Simplified Solution*, the same initial QP as used in MPEG-4 Annex L is used for the first I-frame, for fair comparison. The reason of

TABLE III
COMPARISONS OF FRAME DROPPING AND PSNR (LUMINANCE ONLY), BETWEEN MPEG-4 ANNEX L FRAME-LEVEL RATE CONTROL, AND THE PROPOSED RATE CONTROL SOLUTIONS (1ST LINE: ANNEX L; 2ND LINE: PROPOSED *COMPLETE SOLUTION*; 3RD LINE: PROPOSED *SIMPLIFIED SOLUTION*; 4TH LINE: PROPOSED BIT ALLOCATION WITH LINEAR R$-\rho$ MODEL)

Sequence (bitrate, framerate)	Actual Bitrate (kb/s)	Actual framerate (f/s)	# of dropped frames (Percentage)	PSNR (db)	PSNR Gain (db)
Foreman (48kb/s, 7.5fps)	48.30 45.94 46.46 47.36	6.90 7.50 7.35 7.27	8 (8.00%) 0 (0) 2 (2.00%) 3 (3.00%)	30.30 30.89 30.74 30.61	0 +0.59 +0.44 +0.31
Foreman (64kb/s, 10fps)	64.56 62.44 64.20 63.29	9.53 10.00 9.75 9.90	7 (5.25%) 0 (0) 4 (3.00%) 2 (1.5%)	31.09 31.51 31.29 31.37	0 +0.42 +0.20 +0.28
Foreman (112kb/s, 10fps)	112.07 110.13 111.22 110.54	9.38 10.00 10.00 9.97	9 (6.75%) 0 (0) 0 (0) 1 (0.75%)	33.20 34.03 34.05 33.96	0 +0.83 +0.85 +0.76
Sea World (48kb/s, 7.5fps)	47.85 47.61 47.45 47.46	6.45 7.50 7.35 7.40	21 (14.00%) 0 (0) 3 (2.00%) 2 (1.33%)	27.63 28.88 28.60 28.71	0 +1.25 +0.97 +1.08
Sea World (64kb/s, 10fps)	63.54 63.28 63.22 63.27	9.20 10.00 9.90 9.90	16 (8.00%) 0 (0) 2 (1.00%) 2 (1.00%)	28.51 29.23 29.13 29.17	0 +0.72 +0.62 +0.66
Sea World (128kb/s, 10fps)	128.57 126.50 125.97 125.59	9.45 10.00 9.95 9.95	11 (5.50%) 0 (0) 1 (0.50%) 1 (0.50%)	31.22 31.77 31.68 31.68	0 +0.55 +0.46 +0.46
Glasgow (48kb/s, 10fps)	48.24 47.63 46.20 47.58	7.64 9.92 9.52 9.80	59 (23.60%) 2 (0.80%) 12 (4.80%) 5 (2.00%)	25.81 27.20 26.85 27.04	0 +1.39 +1.04 +1.23
Glasgow (64kb/s, 10fps)	64.28 63.47 62.26 63.41	8.32 10.00 9.56 9.88	42 (16.80%) 0 (0) 11 (4.40%) 3 (1.20%)	27.64 28.54 28.07 28.39	0 +0.90 +0.43 +0.75
Glasgow (112kb/s, 10fps)	112.34 111.05 109.19 111.07	8.68 10.00 9.48 9.92	33 (13.20%) 0 (0) 13 (5.20%) 2 (0.8%)	29.98 31.21 30.65 31.09	0 +1.23 +0.67 +1.11
Charles Angels (64kb/s, 10fps)	64.09 63.81 62.68 63.91	8.49 9.99 9.81 9.90	50 (15.00%) 1 (0.30%) 7 (2.10%) 4 (1.20%)	28.94 30.30 29.96 30.34	0 +1.36 +1.02 +1.40
Charles Angels (112kb/s, 10fps)	112.52 111.73 108.71 111.73	8.91 10.00 9.78 9.99	36 (10.80%) 0 (0) 8 (2.40%) 1 (0.30%)	31.73 33.23 32.67 33.23	0 +1.50 +0.94 +1.50
On Average			11.5% 0.1% 2.5% 1.23%		0 0.98 0.70 0.86

doing bit allocation for the first I-frame slightly differently in the *Complete Solution* is that a fixed initial QP cannot adapt to different bit rates or frame rates. The results are summarized as follows. Note that for fair comparison, we used the same buffer overflow threshold of 80% of buffer size [see (15)] for all the above solutions.

Table III shows the overall performance comparison among the MPEG-4 Annex L frame-level solution, and the proposed *Complete and Simplified* solutions. The proposed *Complete Solution* can achieve much less frame dropping and higher frame rate than MPEG-4 Annex L frame-level solution, for example, on average, 6.7% less frame dropping for "Foreman," 9.2% less frame dropping for "Sea World," 17.6% less frame dropping for "Glasgow," and 12.8% less frame dropping for "Charlies Angels." In most cases, the proposed *Complete Solution* introduces no frame dropping, which results in temporally smoother video. In terms of PSNR performance, the *Complete Solution* provides an average of 0.98 dB gain, over the MPEG-4 Annex L frame-level solution. The proposed *Simplified Solution* is also much better than the MPEG-4 Annex L frame-level solution. It achieves, on average, 9.0% less frame dropping, and an average of 0.70 dB PSNR gain. The improvement is more significant for sequences with lots of scene changes (e.g., "Glasgow" and "Charlies Angels"). When compared with the *Complete Solu-*



(a) (b) (c) (d) (e)

Fig. 3. (a)–(e) PSNR variation for different sequences with smaller (0.5 s) VBV buffer. $\mathrm{PSNR\ mdev} = (1/N_{\mathrm{seq}}) \sum_{i=1}^{N_{\mathrm{seq}}} |PSNR_i - \overline{PSNR}|$, where $\overline{PSNR}$ is the average PSNR value.

tion, on average, the *Simplified Solution* introduces 2.4% more frame dropping and incurs about 0.28 dB PSNR loss. Therefore, our *Simplified Solution* achieves close performance as our *Complete Solution*. On the other hand, we also notice that our *Simplified Solution* incurs bigger loss for more complex sequences (e.g., "Glasgow" and "Charlies Angels"), which highlights the importance of bit allocation guarantee. Table III also shows the results of the approach by combining our proposed frame-level bit allocation scheme with the linear $R-\rho$ model for *frame-level* QP determination as proposed in [28]. The performance of this approach is somewhere between our *Complete Solution* and our *Simplified Solution*. We make a few notes here. First, in terms of complexity, the linear $R-\rho$ model for frame-level QP determination needs at least two to three quantization processes for each frame [28]. A fast implementation using two quantization processes per frame is proposed in [32]. Our *complete solution* needs only 30% requantization and entropy coding on average, which appears to be less complex than the linear $R-\rho$ model approach proposed in [28]. Second, since our QP determination approach can determine the QP prior to performing DCT, it can be combined with any fast encoding algorithms (e.g., [33]) where pretransform decisions can be made to save a lot of the computation of DCT, quantization/inverse quantization and IDCT. On the other hand, the linear $R-\rho$ model approach

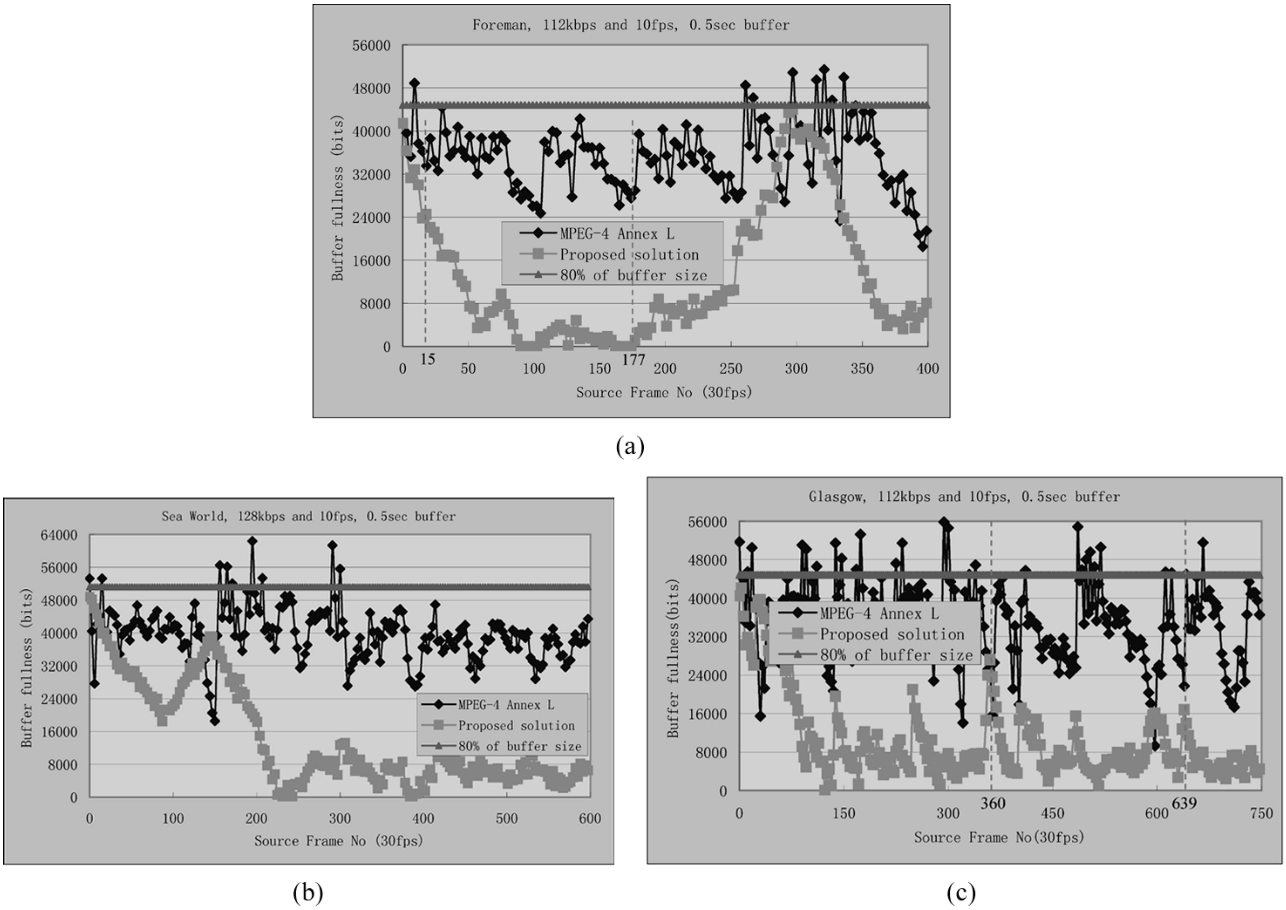


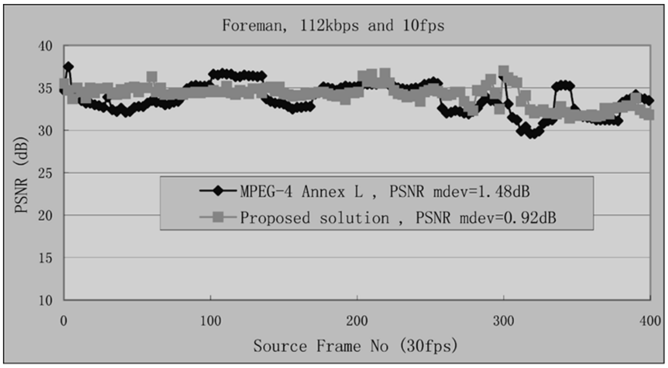
Fig. 4. (a)–(c) Comparison of the encoder VBV buffer fullness for different sequences with different bit rates at the frame rate of 10 fps.

relies on the availability of the distribution of the DCT coefficients, therefore, may not allow fast encoding desirable in the real-time encoding scenario. Third, it appears that, with slightly increased complexity, the linear R$-\rho$ model is more robust and accurate than the quadratic R-Q model used in MPEG-4 Annex L rate control. However, our proposed nonlinear model parameter estimation approach is generic. It can be applied to the linear $R-\rho$ model as well to provide further improvement.

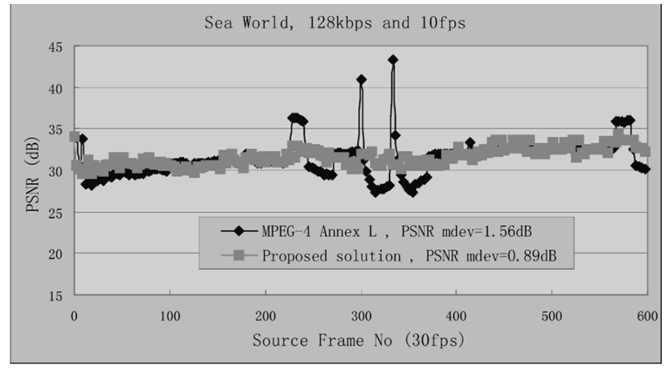
In the following simulations and comparisons, the term "the proposed solution" always refers to our proposed *Complete Solution* to show the potentially big advantage over the MPEG-4 Annex L frame-level rate control solution. Fig. 3 shows the PSNR curves for different sequences with different bit rates and frame rates. The PSNR curves are generated in conformance with the decision made in MPEG-4 rate control testing, where it was decided that when a frame was skipped, the previous encoded frame should be used in the PSNR calculation because the decoder displays the previous encoded frame instead [6]. The same policy has also been specified in ITU document [34] for video performance evaluation. While it is debatable how well this measure can reflect the temporal visual artifact introduced by frame dropping, it is used here just to provide a very rough quantitative comparison of the *motion* video quality. The readers are encouraged to take these numbers cautiously using their own judgment. With such PSNR calculation, the skipped frames can be easily identified, e.g., the points with lower than 20–25 dB PSNR values.

All these curves show that our proposed solution can achieve much less PSNR variation than MPEG-4 Annex L frame-level rate control, which leads to more consistent visual quality across the sequence. This can also be seen from the PSNR *mean absolute deviation* values shown in the figures. For example, for more complex sequences such as "Glasgow" and "Charlies Angels", due to many dropped frames, MPEG-4 Annex L frame-level rate control introduces huge PSNR fluctuation; while with our proposed solution, only one frame for "Charlies Angels", and none for "Glasgow" were dropped. For simpler sequences such as "Foreman" and "Sea World", we can see that, even only the PSNR values of the *encoded* frames are taken into account, the proposed solution still delivers much more consistent quality than MPEG-4 Annex L frame-level rate control.

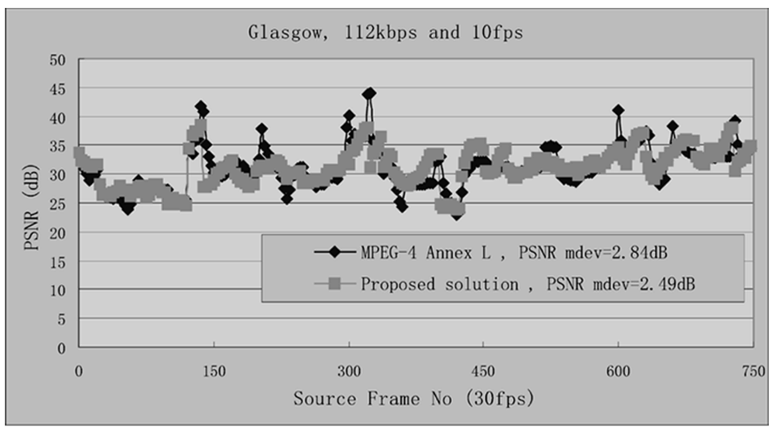
Fig. 4 shows the difference between these two solutions in terms of the encoder VBV buffer usage. In all these figures, the maximum value of the vertical axis represents the buffer size. The initial buffer fullness is chosen as half of the buffer size. Note that the curves show the buffer fullness after the buffer underflow and overflow control as described in (15), i.e., the buffer status after potentially dropping the frames, e.g., when the buffer fullness exceeds 80% of the buffer size, the *next* frame(s) gets dropped. From these curves, we can see that, for MPEG-4 Annex L, the buffer fullness is usually relatively flat with some small to mid *local* fluctuation, suggesting that it does not take full advantage of the flexibility provided by the buffer to accommodate scene variation; while for the proposed solution, there



(a)



(b)



(c)

Fig. 5. (a)–(c) PSNR variation for different sequences with larger (2 s) VBV buffer.

is usually large fluctuation across a larger scale (from scene to scene) with smaller local (within a scene) variation. This big difference is attributed to the two totally different approaches. Basically, with better rate-complexity modeling and better R-Q modeling, our approach tends to make full use of the buffer to track and adapt to the bit allocation variation demanded by the content variability to achieve smoother quality across scenes; while MPEG-4 Annex L frame-level rate control fails to do so. As a result, from the buffer usage point of view, MPEG-4 Annex L is much more conservative. Note that all these buffer fullness fluctuations are within the VBV buffer constraint, imposed by the delay requirement. In other words, the proposed solution tends to take full advantage of the available buffer without violating the delay constraints.

For example, Fig. 4(a) shows the result for the "Forman" sequence where there is a high motion and scene change segment from the 250th frame to the 350th frame. From the bit allocation point of view, to achieve consistent quality, more bits (than the average target bits per frame) need to be allocated to these high motion and scene change frames. As a result, the buffer fullness corresponding to this segment tends to be high. Actually, in order to allocate more bits to this segment without violating the buffer constraint, the bit allocation model should be capable of "saving" some bits in advance. In Fig. 4(a), the curve shows that from the 15th frame to the 177th frame where the scene is quite stationary, the buffer usage is below the equilibrium line (i.e., initial buffer fullness line), which means fewer bits (than the average target bits per frame) have been allocated. The "saved" bits instead are later allocated to the subsequent high complexity frames. All these are achieved by our proposed bit allocation model shown in (7). On the contrary, for MPEG-4 Annex L, the buffer usage overall tends to be flat for the whole sequence, which indicates that the bit allocation and buffer usage are not efficient.

Fig. 4(b) shows the results for the "Sea World" sequence where the first 100 frames belong to a low motion scene and the next 100 frames belong to a high motion scene. Note that from the PSNR curves shown in Fig. 3(c), MPEG-4 Annex L frame-level rate control incurs a lot of frame dropping in the high motion scene (from the 100th frame to the 200th frame), while the proposed rate control solution can successfully avoid any frame dropping. To avoid the frame dropping and achieve decent quality for the high motion scene, our rate control solution is able to spend fewer bits in the first scene (which has low motion) and thus leave more room to accommodate the second scene with high motion, as indicated in Fig. 4(b). On the contrary, MPEG-4 Annex L frame-level rate control failed to do so. Fig. 4(b) indicates that it spends more bits on the first scene (where the buffer fullness goes up), leaving little room to accommodate the bit/complexity variation of the second scene, which causes a lot of frame dropping. In addition, MPEG-4 Annex L does not guarantee the achievement of the allocated bits [as indicated by the more severe *local* buffer fullness variation in Fig. 4(b)], which further contributes to frame dropping.

TABLE IV
COMPARISONS BETWEEN R-MAD AND R-J MODEL. (1ST LINE: R-MAD MODEL; 2ND LINE: R-J MODEL)

Sequence (bitrate, framerate)	Actual Bitrate (kb/s)	Actual framerate (f/s)	Dropped frames (Percentage)	PSNR (db)	PSNR Gain (db)
Glasgow (64kb/s, 10fps)	61.56 61.52	10.00 10.00	0(0) 0(0)	28.60 28.48	0 -0.12
Charles Angels (64kb/s, 10fps)	62.27 62.77	10.00 9.33	0(0) 22(6.6%)	29.46 28.90	0 -0.56
Charles Angels (112kb/s, 10fps)	108.98 109.10	10.00 9.57	0(0) 14(4.20%)	32.03 31.77	0 -0.26
Hangingup (256kbps, 30fps)	258.40 271.21	30.00 29.50	0(0) 42(7.0%)	31.81 31.59	0 -0.22

Fig. 4(c) shows the cases for the sequence “Glasgow” that has multiple scenes. The buffer fullness of our proposed solution shows the trend of the content complexity variation. For example, from both the PSNR curve (Fig. 3(d)) and the buffer fullness curve [Fig. 4(c)], we can see that for the 360th frame which is a scene change frame, the buffer fullness gradient/change (an indication of the amount of allocated bits) in our solution is much larger than that in MPEG-4 Annex L, and the corresponding PSNR value is 29.2 dB, much higher than the PSNR value (26.6 dB) when MPEG-4 Annex L is used. Note that this frame is *not* dropped in both cases. Another scene change frame is the 639th frame, where our solution provides a PSNR of 29.7 dB, as compared to 25.8 dB provided by MPEG-4 Annex L.

To further compare the visual quality variation, we relax the buffer constraint and increase the VBV buffer size from 0.5 to 2 s, since we believe a 2.5-s VBV buffer is too small to accommodate large bit variation to achieve more consistent quality for our test sequences, most of which are very complex. To focus on the spatial image quality, Fig. 5 shows only the PSNR values of the *encoded* frames for different sequences. In fact, it should be pointed out that there is no frame dropping for both solutions in all these cases. We can see that the proposed solution can achieve much more consistent visual quality than the MPEG-4 Annex L frame-level solution, especially on the scene change frames or high complexity frames, thus significantly reducing the annoying flickering effect. The PSNR *mean absolute deviation* values of the proposed solution are much smaller than those in MPEG-4 Annex L solution. Fig. 5(b) shows that for the “Sea World” sequence, our proposed solution basically achieves constant quality. Fig. 5(c) shows that our proposed solution provides larger quality variation for “Glasgow” sequence than other sequences, though it is better than MPEG-4 Annex L frame-level rate control in terms of quality consistency. We believe this is because this sequence is too complex for a *one-pass* encoding solution to achieve truly constant quality across the whole sequence, a limitation imposed by the real-time encoding constraint.

In summary, the proposed solution not only achieves overall decent PSNR gain but also delivers temporally smoother video with more consistent visual quality across the sequence. In particular, it takes care of scene change and scene complexity variation very well and makes more efficient use of the VBV buffer.

C. Performance Comparison Between Two Proposed Bit Allocation Models

In Section I, we discuss several coding complexity measures for bit allocation modeling. Here we present some simulation results to compare the MAD measure with the J measure. In order to better show the effect of bit allocation models on encoding quality, we used a 2.5-s VBV buffer, instead of a 0.5-s VBV buffer.

For the R-J model, we use the linear model, i.e., $R = K \cdot J$. Similarly, we use the following bit allocation

$$T_n = C \cdot \frac{J_n}{\overline{J}_{n-1}},$$

where $C = \mathrm{Bitrate}/\mathrm{Framerate}$, J_n and $\overline{J}_{n-1}$ are the current J and the average J of all previous encoded frames, and $K = C/\overline{J}_{n-1}$. All the remaining modules such as the R-Q modeling and requantization remain the same.

We used almost the same coding settings as presented in the above sub-section, except that the following test sequences are used: “Charlies Angels” (QCIF, a different segment of 1000 frames, more difficult to encode), “Hangingup” ($320 \times 224, 600$ frames, lots of scene changes, different kinds of motion and fade in) and “Glasgow.”

Table IV shows the comparison of the overall performance. For these test scenarios, R-MAD shows better PSNR performance than R-J. In one case, R-MAD model provides 0.56-dB gain over R-J model. Note that in most cases shown in Table IV, R-J model uses slightly more bits than R-MAD model. Typically, R-J model introduces more frame dropping than R-MAD model. This is because for frames with low to medium complexity which account for the majority of the frames in the entire sequence, the bit allocation based on the *linear* R-J model has larger variation than the bit allocation based on the *quadratic* R-MAD model. This has been verified by analyzing the data in Figs. 1 and 2. As a result, when using the R-J model, the VBV buffer tends to build up or deplete more quickly, and thus increase the chances of buffer overflow or underflow, when compared to the R-MAD model. Note that for some other simpler sequences, we observed that the PSNR performance of R-J tends to be close to that of R-MAD.

As mentioned in Section I, one of our main motivations to use R-MAD instead of R-J is that J, similar to σ^2 the variance of a frame), does not differentiate I frames or scene change frames from nonscene-change P frames very well. Therefore, for scene

Fig. 6. 274th frame, left side: using R-MAD model; right side: using R-J model.

change frames, R-J model does not necessarily allocate sufficient bits to them. On the contrary, R-MAD takes care of scene change very well. In the simulation, we observed in one test case that most scene change frames have bad quality when R-J model is used. Fig. 6 shows a side-by-side comparison of captured scene change frames for the "Hangingup" sequence encoded at 256 kbps and 30 fps. The limitation of the J measure is apparent. Note that typically people tend to judge the video quality by the worst case.

V. Conclusion

In this paper, we have proposed a new rate control framework to achieve more consistent quality across the whole video sequence for *one-pass* real-time encoding. To our best knowledge, our work is among the firsts, if not *the* firsts, to address the issue of achieving consistent quality *across various scenes* in the context of *one-pass* real-time encoding. There are two major contributions in this work. First, we propose a novel *sequence-based* (as opposed to GOP-based) bit allocation model to track the nonstationary characteristics in the video source without look-ahead encoding. By doing so, the stationarity assumption typically made in most existing one-pass rate control solutions is no longer needed, scene change detection and the bit allocation for I-, P-, and B-frames can be addressed in a coherent way, and better handling of various scenes with different complexity is achieved. Second, we propose a solid mechanism to achieve the bit allocation, once determined. Specifically, a new *nonlinear* approach for model parameter estimation is proposed to solve the problems in some existing R-Q model parameter estimation schemes. We also highlight and demonstrate the importance of the notion of bit allocation achievement guarantee. We show that the proposed rate control solution can produce significantly better PSNR performance (in terms of both average value and consistency across scenes) as well as temporarily smoother video with less quality flicker and motion jerkiness than MPEG-4 Annex L frame-level rate control. The proposed rate control solution is robust against various sequences, bit rates and frame rates, and has been used in commercial products.

Our experiments also demonstrate that it is a very challenging problem to achieve consistent quality across scenes for some *highly complex* video sequences using one-pass real-time encoding, especially when the delay (or VBV buffer size) constraint is stringent. Our work provides some insight and some initial and promising results. We hope it will stimulate more work along this direction. Future work may include further improving the sequence-based bit allocation model (e.g., including motion information in the model), extending the proposed nonlinear parameter estimation approach to other R-Q models, and investigating the extensibility of the proposed frame-level bit allocation model to MB-level bit allocation.

Acknowledgment

The authors would like to thank J. Lan from the University of Missouri-Columbia for providing a fast implementation of the linear R-ρ model approach for QP determination. The authors also thank the anonymous reviewers for many constructive comments that have helped improve the paper significantly.

References

[1] *Test Model 5*, ISO-IEC/JTC1/SC29/WG11/N0400, Apr. 1993.

[2] *Coding of Moving Pictures and Associated Audio*, ISO/IEC 14 496-2, Oct. 1998.

[3] Video codec test model, near-term, TMN8, in ITU Study Group 16, Video Coding Experts Group, Portland, OR, Jun. 1997.

[4] N. Mohawnian, R. Rajagopalan, and C. A. Gonzales, "Single-pass constant- and variable-bit-rate MPEG-2 video compression," *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 489–509, Jul. 1999.

[5] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, "Two-pass MPEG-2 variable-bit-rate encoding," *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 471–488, Jul. 1999.

[6] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[7] J. R.-Corbera and S. Lei, Extension of TMN8 rate control to B frames and enhancement layer, in ITU SG 16, Video Coding Experts Group, Document Q15-C-19r1, Eibsee, Germany, Dec. 1997.

[8] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 9, pp. 1445–1453, Sep. 1988.

[9] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 533–544, Sep. 1995.

[10] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 26–40, Jan. 1994.

[11] J. Lee and B. Dickinson, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 513–526, Sep. 1994.

[12] B. Yoneyama, H. Yanagihara, and Y. Nakajima, "One-pass VBR MPEG encoder using scene-adaptive dynamic GOP structure," in *Proc. Int. Conf. Consumer Electronics*, Jun. 2001, pp. 174–175.

[13] F.-C. Jeng, C.-T. Chen, and C.-C. Huang, "Single-pass variable bit-rate control for digital video coding," U.S. patent 6 192 075, Aug. 21, 1997.

[14] D. Legall, A. Wells, and K. Uz, "Variable bit rate encoding," U.S. Patent 5 929 916, Dec. 26, 1995.

[15] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 12–20, Feb. 1996.

[16] E. D. Frimout, J. Biemond, and R. L. Lagendik, "Forward rate control for MPEG recording," in *Proc. SPIE Visual Commun. Image Process.*, Cambridge, MA, Nov. 1993, pp. 184–194.

[17] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application—Part I: Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 287–298, Apr. 1997.

[18] B. Tao, H. A. Peterson, and B. W. Dickinson, "A rate-quantization model for MPEG encoders," in *Proc. 1997 Int. Conf. Image Processing*, vol. I, Santa Barbara, CA, Oct. 1997, pp. 338–341.

[19] L.-J. Lin, A. Ortega, and C.-C. J. Kuo, "Rate control using spline-interpolated R-D characteristics," in *Proc. VCIP 1996*, Orlando, FL, Mar. 1996, pp. 111–122.

[20] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 246–250, Feb. 1997.

[21] A. Yan and M. L. Liou, “Adaptive predictive rate control algorithm for MPEG videos by rate quantization method,” in *Proc. Picture Coding Symp.*, Berlin, Germany, Sep. 1997, pp. 619–624.
[22] J. I. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, “Rate control and bit allocation for MPEG-4,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1243–1258, Dec. 1999.
[23] L.-J. Lin and A. Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 446–459, Aug. 1998.
[24] Y. Yang and S. Hemami, “Rate control for VBR over ATM: simplification and implementation,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 9, pp. 1045–1058, Sep. 2001.
[25] J.-F. Bai, C. Feng, Q.-M. Liao, X.-G. Lin, and X.-H. Zhuang, “Accurate estimation of R-D characteristics for rate control in real-time video encoding,” in *Proc. ICASSP*, vol. 3, Salt Lake City, UT, May 2001, pp. 1381–1384.
[26] K. H. Yang, A. Jacquin, and N. S. Jayant, “A normalized rate-distortion model for H.263-compatible codecs and its application to quantizer selection,” in *Proc. Int. Conf. Image Processing*, vol. II, Santa Barbara, CA, Oct. 1997, pp. 41–44.
[27] Z. He, Y. Kim, and S. Mitra, “Low-delay rate control for DCT video coding via ρ-domain source modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.
[28] Z. He and S. K. Mitra, “A unified rate-distortion analysis framework for transform coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1221–1236, Dec. 2001.
[29] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
[30] B. Xie and W. Zeng, “Sequence-based rate control for constant quality video,” in *Proc. 2002 Int. Conf. Image Processing*, Rochester, NY, Sep. 2002, pp. 77–80.
[31] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
[32] J. Lan, X. Zhuang, and W. Zeng, “Single-pass frame-level constant distortion bit allocation for smooth video quality,” in *IEEE Int. Conf. Multimedia and Expo*, Jun. 2004, pp. 387–390.
[33] I.-M. Pao and M.-T. Sun, “Modeling DCT coefficients for fast video encoding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.
[34] Common conditions for video performance evaluation in H.324/M error-prone systems, in ITU Study Group 16, Video Coding Experts Group, Document Q15-G-46, Feb. 1999.

Bo Xie received the B.S. and M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1993 and 1997, respectively.

At Tsinghua University, he developed a PC-based Internet Videophone software system. In 1999, he joined IBM China Development Laboratory as a Software Engineer. Since May 2000, he has been with PacketVideo Corporation, San Diego, CA. His research interests include video coding, especially on encoding quality improvement and performance optimization and video streaming.

Wenjun Zeng (S'94–M'97–SM'03) received the B.E. degree from Tsinghua University, Beijing, China, in 1990, the M.S. degree from the University of Notre Dame, Notre Dame, IL, in 1993, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1997, all in electrical engineering.

He has been an Associate Professor with the Computer Science Department of University of Missouri, Columbia, since August 2003. He worked at Matsushita Information Technology Laboratory, Panasonic Technologies Inc., Princeton, in the summer of 1995, and at Multimedia Communication Laboratory, Bell Laboratories, Murray Hill, NJ, in the summer of 1996. From 1997 to 2000, he was with Sharp Laboratories of America, Camas, WA. He was with PacketVideo Corporation, San Diego, CA, from December 2000 to August 2003, where he was leading research and development projects on wireless multimedia streaming, encoder quality optimization, and digital rights management. His current research interests include multimedia communications and networking, content and network security, and wireless multimedia. He has been an active contributor to the JPEG 2000 image coding standard and the MPEG4 IPMP Extension standard, where four of his proposals have been adopted into the standards. He has been awarded 11 patents.

Dr. Zeng has served as a Special Issue Guest Editor, Special Session, and Panel Session Organizer, and Technical Program Committee Member for several IEEE international journals and conferences. He was the Lead Guest Editor of the IEEE TRANSACTIONS ON MULTIMEDIA Special Issue on Streaming Media published in April 2004. He is an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA, and was the Technical Program Co-Chair of the *Multimedia Communications and Home Networking Symposium*, *2005 IEEE International Conference on Communications*, and the Chair of the *Workshop on Digital Rights Management Impact on Consumer Communications*, *2005 IEEE Consumer Communications and Networking Conference*.