Scalable Rate Control for MPEG-4 Video

Hung-Ju Lee, *Member, IEEE*, Tihao Chiang, *Senior Member, IEEE*, and Ya-Qin Zhang, *Fellow, IEEE*

***Abstract*—This paper presents a scalable rate control (SRC) scheme based on a more accurate second-order rate-distortion model. A sliding-window method for data selection is used to mitigate the impact of a scene change. The data points for updating a model are adaptively selected such that the statistical behavior is improved. For video object (VO) shape coding, we use an adaptive threshold method to remove shape-coding artifacts for MPEG-4 applications. A dynamic bit allocation among VOs is implemented according to the coding complexities for each VO.**

SRC achieves more accurate bit allocation with low latency and limited buffer size. In a single framework, SRC offers multiple layers of controls for objects, frames, and macroblocks (MBs). At MB level, SRC provides finer bit rate and buffer control. At multiple VO level, SRC offers superior VO presentation for multimedia applications. The proposed SRC scheme has been adopted as part of the International Standard of the emerging ISO MPEG-4 standard [1], [2].

***Index Terms*—Bit allocation, MPEG, multiple video object, rate control, rate-distortion optimization, shape coding control.**

I. Introduction

THE MAIN challenge in designing a multimedia application over communication networks is how to deliver multimedia streams to users with minimal replay jitters. In general, a network-based multimedia system can be conceptually viewed as a layer-structure system, which consists of application layer on the top, compression layer, transport layer, and transmission layer, as shown in Fig. 1. To diminish the impact on the video quality due to the delay jitter and available network resources (e.g., bandwidth and buffers), traffic shaping and scalable rate control (SRC) are qualified candidates at two different system levels. Traffic shaping is a transport-layer approach, while SRC is a compression-layer approach, which is the focus of this paper. The basic concept behind the traffic-shaping approach is that before the encoded video bitstreams are injected into the network for transmission, the traffic pattern is already shaped with the desired characteristics, such as maximal delay bounds and peak instantaneous rate [3]–[6]. Therefore, all the system components along the network path from the sender to the receiver can be configured to meet the quality of service (QoS) as desired by allocating the appropriate resources *a priori*. On the other hand, SRC approach is a compression-layer technique where the source video sequence is compressed according to the application's requirement and available network resource; e.g., 10 frames per second (fps) playback rate and 500-ms maximal accumulated delay. In this paper, our attention focuses on the issue of SRC that arises from efficient management of network bandwidth with sufficient video quality to support current multimedia applications.

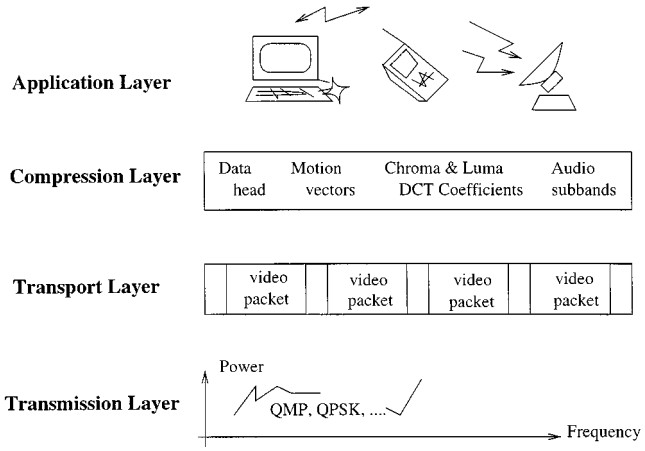

Fig. 1. Layer structure of a network-based multimedia system.

In the development of an SRC scheme, we need to consider a common feature of employing an inter-frame coding between two consecutive video frames in several widely used video compression schemes such as MPEG-1, MPEG-2, and H.263. Although the inter-frame coding scheme exploits the similarity usually found in encoding two consecutive video frames, and achieves significantly coding efficiency, the output with variable-length video bitstream is not well suited for a fixed-rate communication channel. To better utilize network resources and transmit coded video bitstream as accurately as possible, the network parameters and encoding parameters should be jointly considered, and their relationship should be modeled accurately. Technically speaking, rate control is a decision-making process where the desired encoding rate for a source video can be met accurately by properly setting a sequence of quantization parameters (QP). To cope with various requirements of different coding environments and applications, a rate-control scheme needs to provide sufficient flexibility and scalability. For example, multimedia applications are categorized into two groups, which are variable-bit-rate (VBR) application and constant-bit-rate (CBR) application. For VBR applications, rate control attempts to achieve the optimum quality for a given target rate. In CBR and real-time applications, a rate-control scheme must satisfy low-latency and buffer constraints. In addition, the rate-control scheme has to be applicable to a variety of sequences and bit rates. Thus, a rate-control scheme must be scalable for various bit rates (e.g., 10 kbits/s to 1 Mbits/s), various spatial resolutions (e.g.,

Manuscript received September 15, 1998; revised February 8, 2000. This paper was recommended by Associate Editor C. W. Chen.

H.-J. Lee is with Multimedia Technology Laboratory, Sarnoff Corporation, Princeton, NJ 08543 USA (e-mail: hjlee@sarnoff.com).

T. Chiang is with National Chiao Tung University, Taipei, Taiwan, R.O.C. (e-mail: tchiang@cc.nctu.edu.tw).

Y.-Q. Zhang was with Multimedia Technology Laboratory, Sarnoff Corporation, Princeton, NJ 08543 USA. He is now with Microsoft Research, Beijing 100080, China (e-mail: yzhang@microsoft.com).

Publisher Item Identifier S 1051-8215(00)07555-8.

1051–8215/00$10.00 © 2000 IEEE

Peloton - EX1003, Page 1
IPR2023-01239 (Peloton v. NEC)

QCIF to CCIR-601), various temporal resolutions (e.g., 7.5–30 fps), various coders (e.g., DCT and wavelet), and various granularities of video object (VO) (e.g., single VO to multiple VOs, frame-layer to macroblock (MB)-layer).

In developing a rate-control technique, there are two widely used approaches: 1) an analytical model-based approach and 2) an operational rate-distortion (R-D) based approach. In the model-based approach, various distribution and characteristics of signal source models with associated quantizers are considered. Based on the selected model, a closed-form solution is derived using optimization theory. Such a theoretical optimization solution cannot be implemented easily because there is only a finite discrete set of quantizers and the source signal model varies spatially. Alternatively, an operational R-D based approach is used in practical coding environment. For example, to minimize the overall coding distortion subject to a total bit budget constraint, lots of techniques based on dynamic programming or Lagragian multiplier for optimization solutions have been developed [7]–[12]. These methods share the similar concept of data pre-analysis. By analyzing the R-D characteristics of future frames, the bit-allocation strategy is determined afterwards. The Lagragian multiplier is a well-known technique for optimal bit allocation in image and video coding, but with an assumption that the source consists of statistically independent components. Thus, an inter-frame based coding may not find the Lagragian multiplier approach applicable because of the temporal dependency. Although Ramchandran [9] takes frame dependencies into account in bit-rate control, its potentially high complexity with increasing the operating R-D points make it unsuitable for the applications requiring interactivity or low encoding delay. In [13], [14], Ding investigated a joint encoder and channel rate-control scheme for VBR video over ATM networks and claimed that the rate-control scheme has to balance both issues of consistent video quality in the encoder side and bitstream smoothness for statistical multiplexing gain in the network side. Tao *et al.* [15] proposed a parametric R-D model for MPEG encoders, especially for the picture-level rate control. Based on the bit-rate *mquant* model, the desired *mquant* is calculated and used for encoding every MB by combining with appropriate quantization matrix entry in a picture. A normalized parametric R-D model based approach [16] has been also developed for H.263-compatible video codecs. By providing good approximations of all 31 rate-distortion relations, the authors claim that the proposed model offers an efficient and less memory requirement approach to approximate the rate and distortion characteristics for all QPs. Recently, Vetro and Sun [17], [18] and Ribas-Corbera and Lei [19], [20] also proposed rate-control techniques for MPEG-4 object-level and MB-level video coding, respectively. However, most of the aforementioned techniques only focus on a single coding environment, either frame level, object level, or macro level. None of these techniques demonstrates its applicability to MPEG-4 video coding including the above three coding granularities simultaneously.

In this paper, based on a newly revised quadratic R-D model, our SRC proposes a single framework which is designed to meet both VBR without delay constraints and CBR with low-latency and buffer constraints. With this scalable framework based on a new R-D model and several new concepts [21] in our scheme, not only more accurate bit-rate control with buffer regulation is achieved, but scalability is also preserved for all test video sequences in various applications. For example, in the object-based video coding of the emerging ISO MPEG-4 international standard, it is very important to appropriately allocate bits among different VOs. In allocating bits among VOs, video contents and coding complexity must be considered. Otherwise, over- and under-runing of the bit budget can occur. In general, a rate-control scheme should spend more bits in the VO of user interests (e.g., foreground VO) than in other areas (e.g., background VO). Without employing proper bit allocation, for example, the background VO could have excellent quality, while the foreground VO could suffer from lots of annoying distortion, even though bits are evenly distributed. By considering video contents and coding complexity in our quadratic R-D model, our rate-control scheme with joint buffer control can dynamically and appropriately allocate the bits among VOs to meet the overall bit-rate requirement with uniform video quality.

Another unique merit of MPEG-4 VO-based coding from the other video coding standards like MPEG-1, MPEG-2, and H.263 is that an encoder can separately encode any VO from the rest and transmit these individual elementary bitstreams independently. In this category of applications, our proposed joint-buffer rate-control scheme seems imposing fairly restrictive conditions that each VO shares the same buffer management. As a matter of fact, the proposed rate-control scheme can quite easily handle this scenario. A simple way is to consider each individual VO as a “frame” so that each VO can operate on its own (i.e., separate buffer control and separate R-D model and so on). In this case, the frame-level rate control can be directly applied with minor modifications (e.g., initial setting of the buffer fullness). Note that since there is no joint buffer control in this case, before an encoder starts encoding each VO, the application needs to specify the buffer condition for each VO.

With its precision of the R-D model and ease of implementation, our rate-control scheme with the following new concepts and techniques has been adopted as part of the rate-control scheme in the International Standard of the emerging ISO MPEG-4 standard:

1) a more accurate second-order R-D model for the target bit-rate estimation;
2) a sliding-window method for smoothing the impact of scene change;
3) an adaptive selection criterion of data points for better model updating process;
4) an adaptive threshold shape control for better use of bit budget;
5) a dynamically bit-rate allocation among VOs with different coding complexities.

The proposed rate-control scheme provides a scalable solution, meaning that our rate-control technique offers a general framework for multiple layers of controls for objects, frames, and MBs in various coding contexts.

The rest of the paper is organized as follows. Section II reviews the theoretical foundation of the proposed rate control, and characterizes new features for generalizing and improving

Peloton - EX1003, Page 2
IPR2023-01239 (Peloton v. NEC)

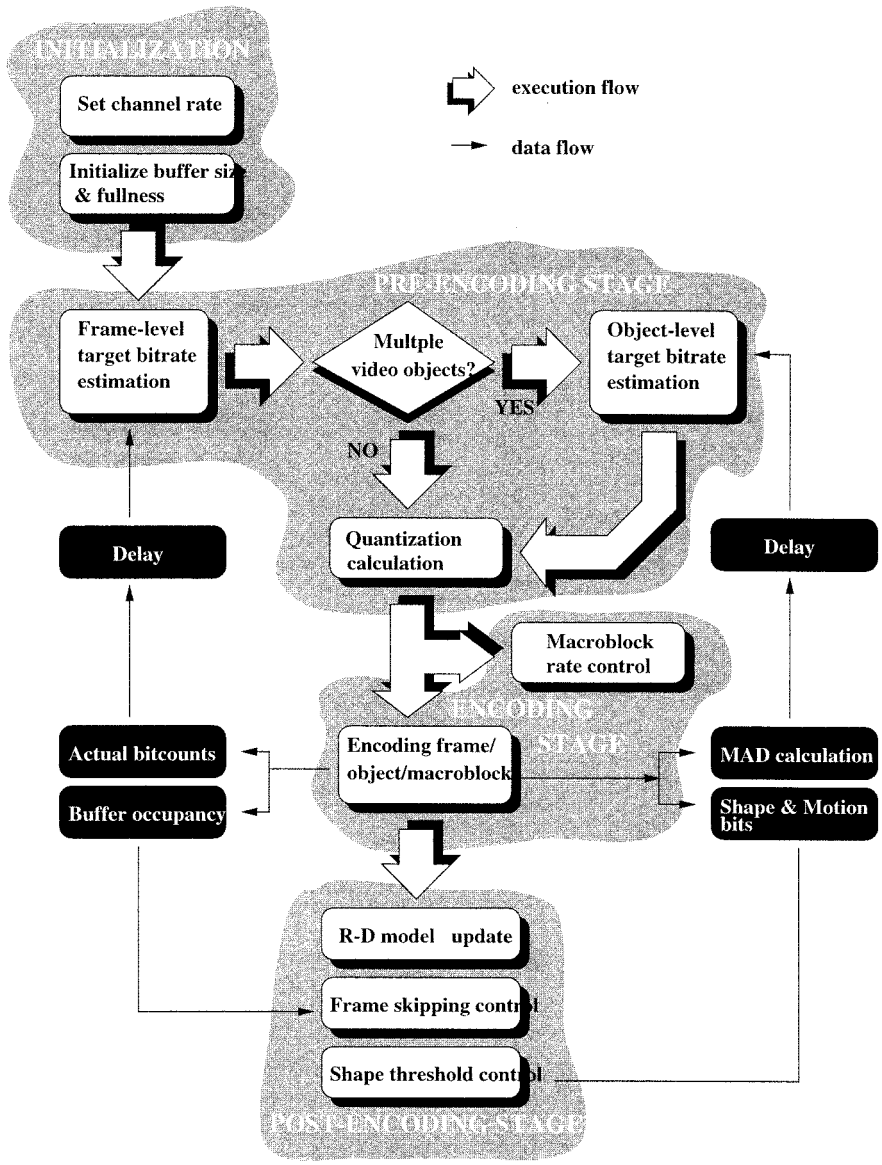

Fig. 2. Block diagram of the proposed SRC.

the R-D models. In Section III, detailed descriptions of our generalized rate control are presented. Some fundamental research problems for rate-control scheme in various coding granularities are also addressed. In Section IV, extensive experiments are conducted to evaluate the performance of the scheme. This paper concludes with Section V.

II. Framework of SRC

In this section, we describe the framework of the proposed SRC scheme which provides an integrated approach with three different coding granularities, including frame level, object level, and MB level. The theoretical foundation behind the proposed rate-control scheme is based on the R-D model, where the distortion is measured in terms of quantization parameter. The block diagram of the proposed SRC is depicted in Fig. 2, where the proposed rate control consists of four stages: 1) initialization stage; 2) pre-encoding stage; 3) encoding stage; and 4) post-encoding stage.

A. Scalable Quadratic Rate Distortion Model

To illustrate the rationale of quadratic R-D function modeling, we summarize the result derived in [22], [23]. Assuming that the source statistics are Laplacian distributed

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

the distortion measure is defined as $D(x, \tilde{x}) = |x - \tilde{x}|$, then there is a closed-form solution for the R-D functions as derived in [24]

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

where

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

Peloton - EX1003, Page 3
IPR2023-01239 (Peloton v. NEC)

The R-D function is expanded into a Taylor series

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

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

$$R_i = a_1 \times Q_i^{-1} + a_2 \times Q_i^{-2}.$$

Although the above R-D model provides the theoretical foundation of the rate-control scheme, the major drawback is its lack of considering the following two factors.

1) The R-D model is not scalable with video contents. The model was developed based on the assumption that each video frame has similar coding complexity, resulting in similar video quality for each frame by properly target bit-rate estimation. By introducing an index for video coding complexity such as mean absolute difference (MAD), the R-D model becomes scalable.
2) The R-D model does not exclude the bit counts used for coding the overhead including video/frame syntax, motion vectors and shape information. Although the bit count used for motion vectors, for example, is video content dependent, its variation in bit count is relatively smaller, compared to the texture information.

To enhance our R-D model with more accuracy, a simple prediction is used to predict those bit counts using the last coded frame as a reference. These bits used for nontexture information are considered as constant numbers irrespective of its distortion and excluded from the target bit-rate estimation. To accurately estimate the target bit rate with scalability, the original quadratic R-D formula is modified by introducing two new parameters: MAD and nontexture overhead (H).

$$\frac{R_i - H_i}{M_i} = a_1 \times Q_i^{-1} + a_2 \times Q_i^{-2}$$

where

- R_i total number of bits used for encoding the current frame i;
- H_i denotes the bits used for header, motion vectors, and shape information,
- M_i MAD, computed using motion-compensated residual for the luminance component (i.e., Y component);
- Q_i quantization level used for the current frame i;
- a_1, a_2 first- and the second-order coefficients.

To solve the target bit rate, we assume the video is encoded first as an I frame, and subsequently P frames. The scheme has been extended to variable GOP structure and B frames as well

$$R = T_i + N_p \times T_p$$
$$\frac{T_p - H_p}{M_p} = a_1 \times Q_p^{-1} + a_2 \times Q_p^{-2}$$

where

- R total bit budget;
- T_i bits budget used for the first I frame;
- N_p number of P frames;
- T_p the bit budget used for all P frames, and
- H_p bit budget used for nontexture information.

Then the T_p and Q_p can be obtained based on the technique [23]. Let $X_{n\times 2} = [1,\ 1/Q_p(i)]$ and $Y_{n\times 1} = [Q_p(i) \times T_p(i)]$, where $i = 1, 2, \cdots, n$, and n is the number of selected data samples. Then

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y.$$

Based on these two model parameters a_1 and a_2, the quantization level Q_p and target bit rate T_p can be computed before encoding the next frame.

B. Initialization Stage

In the initialization stage, the major tasks the encoder has to complete with respect to the rate control include:

1) initializing the buffer size based on latency requirement;
2) subtracting the bit counts of the first I-frame from total bit counts;
3) initializing the buffer fullness in the middle level.

Without loss of generality, we assume that the video sequence is encoded first as an I frame, and subsequently P frames. In this stage, the encoder encodes the first I-frame using an initial QP value specified as an input parameter. Then the remaining available bits for encoding the subsequent P frames can be calculated as

$$\mathcal{R}_t = \mathcal{T} \cdot R - \mathcal{I} \tag{1}$$

where

- $\mathcal{R}_t$ remaining available bits for encoding subsequent P frames at the coding time instant t (e.g., initially $t = 0$);
- $\mathcal{T}$ duration of the video sequence in the unit of seconds (e.g, 10 s);
- R itrate for the sequence in the unit of bits per second (e.g., 10 kbits/s);
- $\mathcal{I}$ number of bits used for the first I frame.

Thus, the channel output rate is $\mathcal{R}_0/N_0$, where N_0 is the number of P frames in the sequence or in a GOP. The buffer size is set based on the latency requirement specified by the user. The default buffer size is $R \cdot 0.5$ (i.e., the maximum accumulated delay is 500 ms.). The target buffer fullness is set at the middle level of the buffer (i.e., $R \cdot 0.25$) is to have more rooms to effectively offset the potential buffer overflow or underflow caused by suddenly increasing or decreasing bit count in a segment of video sequence. It is clear that higher target buffer level may cause the buffer overflow, while lower target buffer level may cause the buffer underflow. Although stuffing useless bits is a straightforward way to handle it, the waste of network resource, e.g., bandwidth and network buffer is its major disadvantage.

Peloton - EX1003, Page 4
IPR2023-01239 (Peloton v. NEC)

C. Pre-Encoding Stage

In the pre-encoding stage, the tasks of the rate-control scheme include: 1) target bits estimation,; 2) further adjustment of the target bits based on the buffer status for each VO; and 3) quantization parameter calculation. The target bit count is estimated in the following phases including: 1) frame-level bit rate; 2) object level if desired; and 3) MB-level bit-rate estimation if desired. At the frame level, the target bit count for a P frame at time $(t+1)$, R_{t+1}, is estimated as

$$R_{t+1} = \frac{\mathcal{R}_t}{N_t} \times (1-S) + A_t \times S$$

where

$\mathcal{R}_t$ remaining bit counts at time t;
N_t emaining number of P frames at time t;
A_t actual bits used for the P frame at time t (i.e., the previous P frame).

Note that S is the weighting factor to determine impact from the previous frame on the target bit estimation of the current frame. It can be determined adaptively or set as a constant number. The default value of S is 0.05 in our experiments.

To get a better target bit-rate estimation, we need to consider buffer fullness. Hence, the target bit-rate estimation can be further adjusted with the following equation:

$$R_t = (B_t + 2 \times (\mathcal{B} - B_t))/(2 \times B_t + (\mathcal{B} - B_t)) \times R_t \quad (2)$$

where B_t is the current buffer fullness at time t and $\mathcal{B}$ is the buffer size. The adjustment in the above equation aims to keep the buffer fullness in the middle level to reduce any chance of buffer overflow or underflow. To achieve constant video quality for each video frame or VO, the encoder must allocate a minimum number of bits, which is denoted as R/F, where R and F are application's bit rate and frame rate of the source video, respectively. That is

$$R_t = \max\{R/F, R_t\}. \quad (3)$$

Then, the final adjustment is made to predict the impact of R_t on the future buffer fullness. A safety margin, denoted as M, of the buffer, which is pre-set before encoding, is enforced to avoid the potential buffer overflow and buffer underflow. To avoid buffer overflow, if $(R_t + B_t > (1-M) \cdot \mathcal{B})$, then the target bit rate is decreased and becomes

$$R_t = R_t + B_t - (1-M) \cdot \mathcal{B}.$$

To avoid buffer underflow, on the other hand, if $(R_t + B_t - C < M \cdot \mathcal{B})$, then the target bit rate is increased and becomes

$$R_t = C - R_t - B_t + M \cdot \mathcal{B}$$

where $C(= \mathcal{R}_0/N_0)$ is the channel output rate.

D. Encoding Stage

In the encoding stage, the major tasks that the encoder has to complete include:

1) encoding the video frame (object) and recording all actual bit rate;
2) activating the MB-layer rate control if desired.

In the encoding stage, if either the frame- or object-level rate control is activated, the encoder compresses each video frame or VO using QP as computed in the pre-encoding stage. However, some low-delay applications may require strict buffer regulations, less accumulated delay, and better spatially perceptual quality. An MB-level rate control is necessary. However, an MB-level rate control is costly at low rate since there is additional overhead if quantization parameter is changed frequently within a frame. For instance, in the MPEG-4, the MB type requires one to three more bits to indicate the existence of the differential quantization parameter (i.e., dquant). Furthermore, two bits need to be sent for dquant. In the worse case, for the same prediction mode, additional 5 bits need to be transmitted in order to change QP. In the case of encoding at 10 kbits/s, 7.5 fps, QCIF resolution, the overhead can be as high as $99 * 5 * 7.5 = 3.7$ kbits/s. If only 33 MBs are encoded, the overhead $33 * 5 * 7.5 = 1.2$ kbits/s. Thus, there will be about 10% loss in compression efficiency at low bit-rate encoding. At high bit rate, the overhead bit count is less significant than the residual bit count. The details of the MB-level rate control will be presented in the next section.

E. Post-Encoding Stage

In the post-encoding stages, the encoder needs to complete the following tasks: 1) updating the correspondent quadratic R-D model for the entire frame or an individual VO; 2) performing the shape-threshold control to balance the bit usage between shape information and texture information; and 3) performing the frame-skipping control to prevent the potential buffer overflow and/or underflow.

1) R-D Model Update: After the encoding stage, the encoder has to update each VO's respective R-D model based on the following formula, for example, for VO i:

$$\frac{V_t^i - H_t^i}{M_t^i} = \frac{X_1^i}{Q_t^i} + \frac{X_2^i}{(Q_t^i)^2} \quad (4)$$

where V_t^i is the actual bit count used for object i, and H_t^i is the overhead bit count used for syntax, motion and shape coding. Note that in the case of MB rate control, the quantization parameter is defined as the average of all encoded MBs. To make our R-D model more accurate to reflect the video contents, the R-D model update process is consisted of the following three steps. The motivation and technical details of these three steps will be described in the next section.

Step 1—Selection of Data Points: Those data points selected by the encoder are used to update the R-D model. The quality and quantity of the data set are critical to accuracy of the model. With respect to the quantity of the data set, generally speaking, more data points are likely to yield a more accurate model at the expense of higher complexity. With respect to the quality of the data set, some objective indices, such as MAD or SAD, of the current data point can be used. By considering these factors, a sliding-window based data selection mechanism is developed, and its details will be presented in the next section. Although inserting an I frame is an more effective way to handle scene change, its higher complexity, e.g., resuming the buffer control

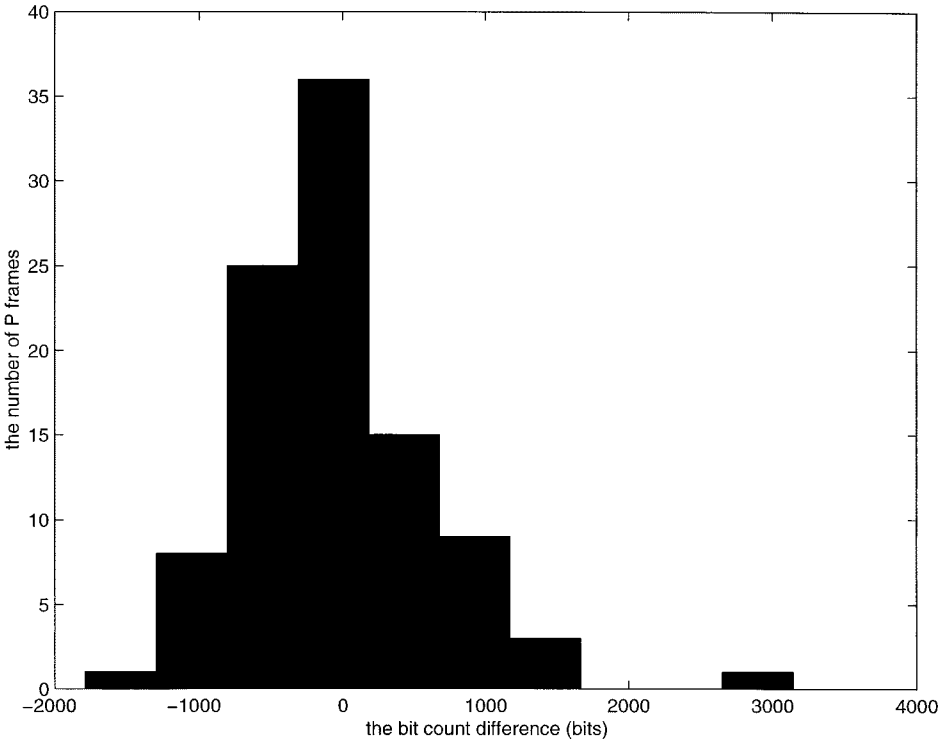

Fig. 3. Distribution of the bit-count difference.

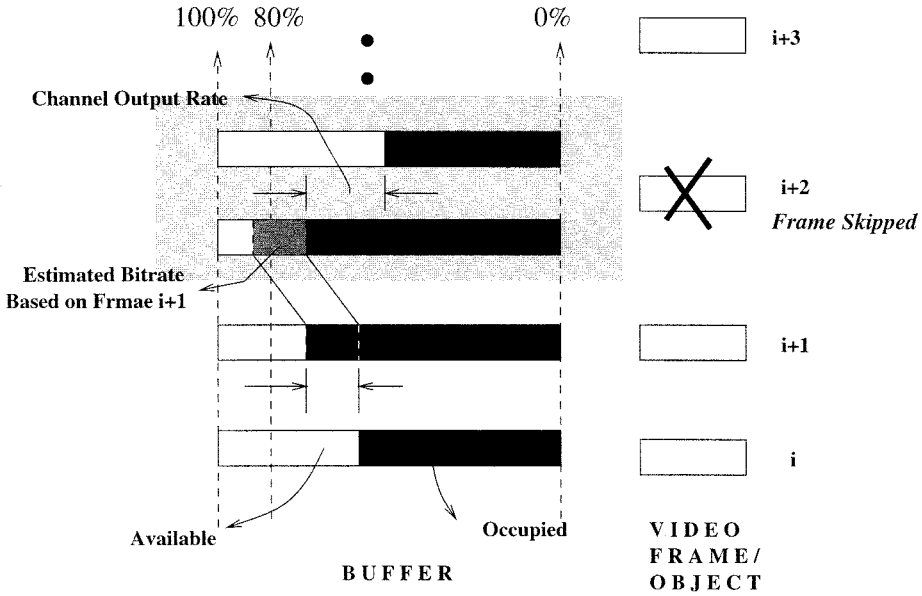

Fig. 4. Scenario of frame-skipping control.

to prevent buffer overflow, is its major drawback. Note that since our rate control is a GOP basis, the insertion of I-frame can be viewed as the resume of another GOP round, thus our rate control still applicable.

Step 2—Calculation of the Model Parameters X_1 and X_2: Based on these two values, the theoretical target bit rate can be calculated for each data point within the sliding window obtained in step 1. For those selected data points, the encoder collects quantization levels and actual bit-rate statistics. Using a linear regression technique, the two model parameters X_1 and X_2 can be obtained as

$$X_2 = \frac{n\sum_{i=1}^{n} R_i - \left(\sum_{i=1}^{n} Q_i^{-1}\right)\left(\sum_{i=1}^{n} Q_i R_i\right)}{n\sum_{i=1}^{n} Q_i^{-2} - \left(\sum_{i=1}^{n} Q_i^{-1}\right)^2} \tag{5}$$

Peloton - EX1003, Page 6
IPR2023-01239 (Peloton v. NEC)

$$X_1 = \frac{\sum_{i=1}^{n} Q_i R_i - X_2 Q_i^{-1}}{n} \tag{6}$$

where n is the number of selected past frames, Q_i and R_i are the actual average quantization levels, and actual bit counts in the past, respectively.

Step 3—Removal of the Outliers from the Data Set: After the new model parameters X_1 and X_2 are derived, the encoder performs further refinement step to remove some bad data points. Then the actual bit rate for each past frame is known and the target bit rate is recalculated based on the new model parameters obtained in Step 2, the actual bit rate R_i, and the average quantization level Q_i. By applying the refinement process, more representative data points are selected, and the final model parameters can be derived based on these new data points. The detail of this process will be presented in the next section.

2) Shape-Threshold Control: Next, we discuss the issue of shape-threshold control. The issue arises from the fact that only a limited bit budget is available for object-based video coding. In the object-based video coding, the development of an efficient and effective way to allocate limited bit budget used for coding shape and texture information becomes a very important research problem. In MPEG-4, there are two ways to control the bit count used for the shape information: 1) size conversion process and 2) shape-threshold setting. The size conversion process adopted by MPEG-4 is used to reduce the amount of shape information for rate control. The size conversion can be carried out on an MB basis [25], [26]. The shape-threshold setting, on the other hand, is a controllable parameter, and is carried out on an object basis. The threshold value could be either a constant or a variable. In this paper, an adaptive shape-threshold control is developed with its details presented in the next section.

3) Frame-Skipping Control: The objective of the frame-skipping control is to prevent buffer overflow. Once the encoder predicts that encoding the next frame would cause buffer overflow, the encoder skips the encoding of the next frame. The buffer occupancy will be decreased by the channel transmission rate at the expense of lower frame rate. Although the frame skipping is an effective way to prevent buffer overflow, the overall perceptual quality will be reduced significantly, especially in the case of continuous frame skipping. To fight the problem with continuous frame skipping, a frame-skipping mechanism is proposed and will be presented in details in the next section.

III. New Features of the Proposed SRC

In this section, we introduce several new concepts with its associated technical merits in our rate-control scheme and elaborate those techniques leading to a more accurate and scalable target bit-rate estimation.

In our frame-level rate control, to make the quadratic R-D model more accurate and scalable for different coding requirements, the following new concepts and mechanisms are introduced:

1) sliding-window data-point selection;
2) statistical removal of data outliers;
3) predictive frame-skipping control.

In object-based rate control, we would like to address the following research problem: 1) calculating the target bit rate among the VOs; 2) balancing the bit budget between the shape information and the texture information without introducing noticeable distortion; and 3) encoding the VOs with proper temporal resolution so that the quality of the composite video frame is sufficient.

To solve these problems, we propose the following solutions based on our experimental results and some theoretical implications.

1) dynamic target bit-rate distribution among VOs;
2) adaptive shape-threshold control

A. Sliding-Window Data-Point Selection

The sliding-window mechanism is used to adaptively smooth the impact of a scene change for certain number of frames in updating the R-D model. If the complexity changes significantly (i.e., high motion scenes), a smaller window with more recent data points is used. By using such a mechanism, the encoder can intelligently select those representative data points for R-D model updates. The selection of data points is based on the ratio of "scene change" between the current frame (object) and the previous encoded frame (object). Note that the sliding-window mechanism with data-point selection is to diminish the effect of partially deals with scene change, not completely solve the abrupt scene change.

To quantify the amount of scene change, various indices, such as MAD or SAD, or their combinations can be used. A sophisticated weighting factor can be considered, too. To make our rate-control scheme easy to implement with lower complexity, only MAD is used as an index to quantify the amount of scene change. If a segment of video contains a trend of higher motion scene (i.e., increasing coding complexity), then a smaller number of data points with recent data are selected. On the other hand, a larger number of data points with historic data are selected for a lower motion scene of video contents. Algorithmically, if $\text{MAD}_{t-1} > \text{MAD}_t$, where t is the time instant of coding, the size of the sliding window is $(\text{MAD}_t/\text{MAD}_{t-1} \cdot \text{MAX_SLIDING_WINDOW})$, otherwise it is $(\text{MAD}_{t-1}/\text{MAD}_t \cdot \text{MAX_SLIDING_WINDOW})$, where MAX_SLIDING_WINDOW is a preset constant (e.g., 20 in our experiments).

B. Statistical Removal of Data Outliers

The R-D model is further enhanced and calibrated by employing a statistical method of rejecting some erroneous data points. Those erroneous data points are defined, in the statistical sense, as the data points whose prediction errors between the actual bit rate and the target bit rate is larger than k standard deviations (e.g., as a rule of thumb, k is set to one in our experiments). In updating the R-D model after encoding a frame, the removal of the outliers from the data set can diminish the impact of bad data points on the model update process. In other words, after new model parameters X_1 and X_2 are derived, the encoder performs this refinement step to remove some erroneous data points. Since the actual bit rate for each past frame is known,

Peloton - EX1003, Page 7
IPR2023-01239 (Peloton v. NEC)

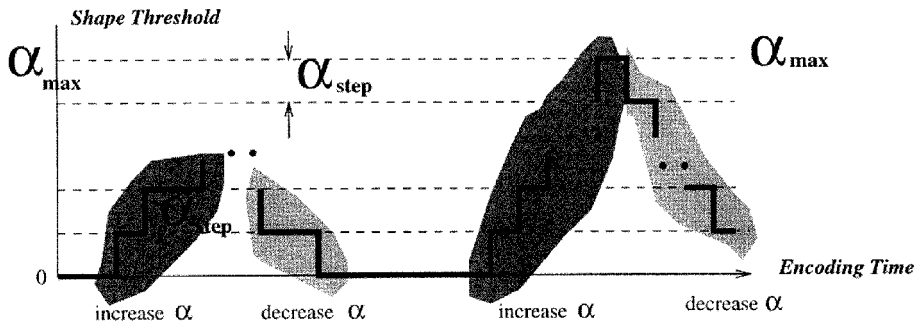

Fig. 5. Scenario of shape-threshold settings.

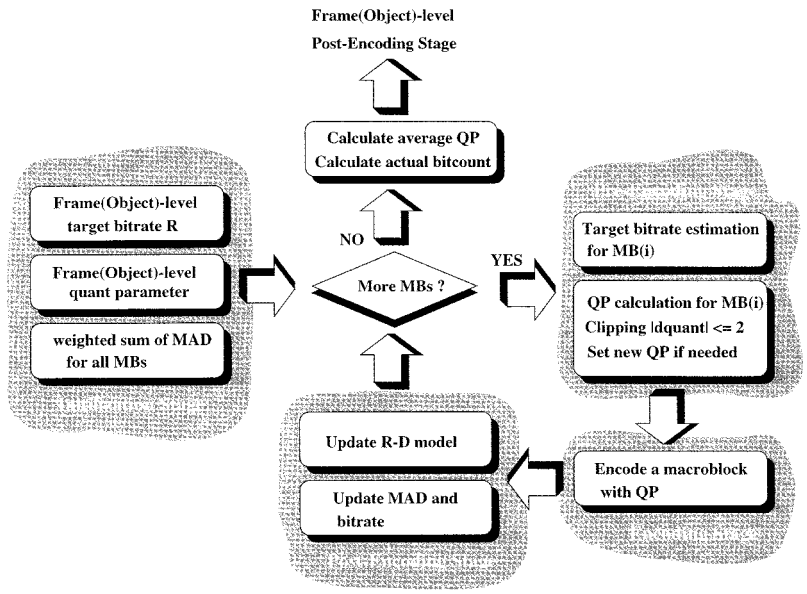

Fig. 6. Block diagram of the proposed MB-level rate control.

then the target bit rate is recalculated based on the new model parameters. By applying this statistical technique, more representative data are selected, and the final model parameters can be derived based on new data. Note that, to avoid the removal of all data points, the latest data point is always selected in the data set.

C. Predictive Frame-Skipping Control

Ideally, if the input bit rate to the buffer is equal to the channel output rate, then the buffer fullness will keep its middle level, which is the target buffer level. Otherwise, it changes its fullness upwards or downwards which may potentially cause buffer overflow or underflow, respectively. To effectively prevent the buffer overflow without losing overall video quality due to continuous frame skipping, we proposed a predictive frame-skipping control. By setting a safety threshold, if an encoder predicts to have a potential buffer overflow, then it skips next frame (object) and subtracts the buffer fullness from the channel output rate. The safety threshold is selected based on the distribution of bit-count difference of a predicted frame (object) from the average bit count of coded frame (object) in a video sequence or GOP. For example, a *containership* video sequence with 300 qcif frames is coded as the first I-frame and subsequent P-frame coding pattern at 10 fps and 24 kbits/s. Its distribution of the bit-count difference is depicted in Fig. 3, where the X-axis denotes the bit-count difference of a coded P frame from the average coding bit count in a sequence or GOP, while the Y-axis denotes the number of P frames. In this example, the average bit count for a P-frame is calculated as (24 kbits $*10$ sec -1 kbit)/100 frames $= 2.39$ kbits, given 1 kbit is used for coding the first I frame. If the target buffer level is in the middle level (i.e., 12 kbits), then the encoder has a upper or lower half of 6/2.39 frames for buffering any sudden bit count increasing, or decreasing, respectively. Considering the buffer-overflow case, if a threshold value is properly set, then there is only a slight chance of overflowing the buffer. In this case, if the threshold is set as 80% of the buffer size (i.e., 4.8 kbits), the consecutive frame skipping occurs very rarely from the statstical viewpoint. This empirical assumption by setting a proper threshold demonstrates although our predictive frame-skipping scheme cannot guarantee no consecutive frame skipping occurring; however, there exist fairly high confidence that consecutive frame skipping rarely occurs.

The proposed frame-skipping mechanism is described as follows. Before encoding the next frame, the encoder first examines the current buffer occupancy and the estimated target bit rate for the next frame. If the current buffer fullness plus the estimated frame bits for the next frame is larger than some pre-de-

Peloton - EX1003, Page 8
IPR2023-01239 (Peloton v. NEC)

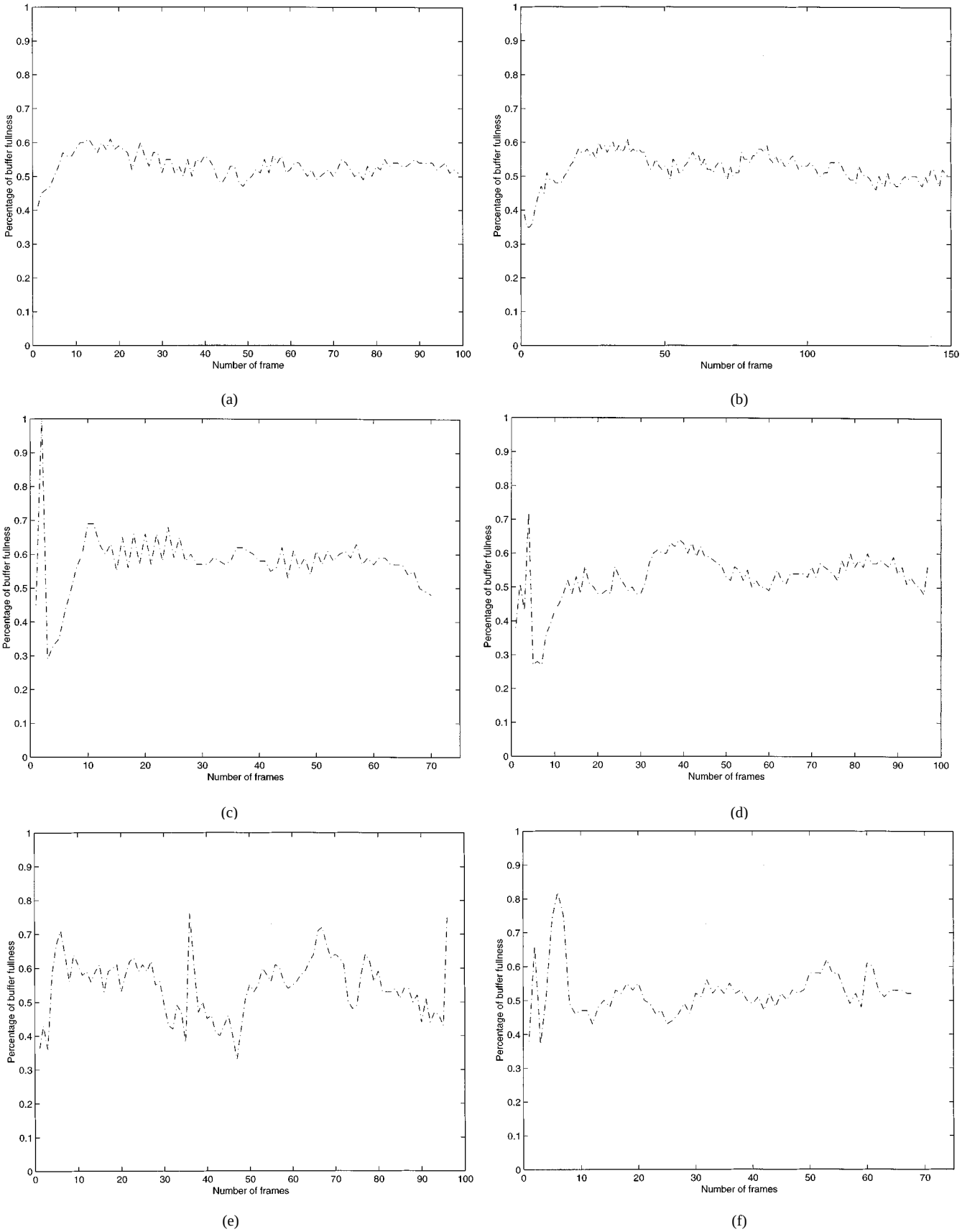

Fig. 7. Frame-level rate control: the buffer occupancy for various test conditions. SVORC: (a) $silent_qcif$ at 10 fps and 24 kbits/s; (b) $hall_qcif$ at 7.5 fps and 10 kbits/s; (c) mad_qcif at 10 fps and 24 kbits/s; (d) $madl_qcif$ at 7.5 fps and 10 kbits/s; (e) $news_cif$ at 15 fps and 112 kbits/s. (f) $news_cif$ at 7.5 fps and 48 kbits/s.

TABLE I
VBR EXPERIMENTAL RESULTS OF THE FRAME-LEVEL RATE CONTROL

Video sequence	target frame rate (fps)	actual frame rate (fps)	target bit rate (Kbps)	actual rate (Kbps)	PSNR_Y (dB)
coastguard(qcif)	10	10	48	48.00	29.16
coastguard(qcif)	15	15	112	112.08	31.68
container(qcif)	10	10	24	23.97	32.37
container(qcif)	7.5	7.5	10	9.99	29.77
hall(qcif)	7.5	7.5	10	10.02	30.44
mad(qcif)	10	10	24	24.02	35.03
mad(qcif)	7.5	7.5	10	10.20	32.11
news(cif)	15	15	112	112.17	33.80
news(cif)	7.5	7.5	48	48.01	31.30

TABLE II
CBR EXPERIMENTAL RESULTS OF THE FRAME-LEVEL RATE CONTROL

Video sequence	target frame rate (fps)	actual frame rate (fps)	traget bit rate (Kbps)	actual bit rate (Kbps)	PSNR_Y (dB)
coastguard(qcif,48k)	10	9.8	48	47.90	29.32
coastguard(qcif,112k)	15	15	112	111.93	31.72
container(qcif,24k)	10	9.9	24	23.96	32.58
container(qcif,10k)	7.5	7.2	10	10.00	29.96
hall(qcif,10k)	7.5	6.5	10	10.01	31.06
mad(qcif,24k)	10	9.6	24	23.85	35.25
mad(qcif,10k)	7.5	6.0	10	10.15	32.87
news(cif,112k)	15	15	112	112.11	33.96
news(cif,48k)	7.5	7.3	48	47.63	31.61
silent(qcif,24k)	10	9.6	24	24.04	31.01

termined threshold, called the safety margin, for instance 80% of the buffer size, the next frame will be skipped. Note that the safety margin is used to reduce continuous frame skipping and can be adaptively changed based on the content, or a pre-determined constant for simplicity. An example, as shown in Fig. 4, demonstrates our frame-skipping control. Before encoding the frame $i+2$, the encoder uses the actual bit count used for the frame $i+1$ to predict the next frame $i+2$. Then the buffer occupancy will exceed 80% of the buffer size, resulting in skipping the frame $i+2$ and decreasing the buffer occupancy by the channel output rate. The frame-skipping condition can be formulated as follows.

WHILE ($buffer_fullness + actual_bitcounts_for_last_frame - channel_output_rate \geq buffer_size *$ skip_margin) {
  Skip the next frame;
$buffer_fullness = buffer_fullness - channel_output_rate$

D. Dynamic Target Bit-Rate Distribution Among VOs

To estimate the target bit rate for each VO, a straightforward way is to allocate an equal number of bits to each VO without considering its complexity and perceptual importance. However, this simple scheme suffers several serious problems. For example, the background VO may have bits left unused, while the foreground VO requires more. Therefore, we propose a new bit-allocation method to adaptively adjust the bit budget for each VO. Based on the coding complexity and perceptual importance, the distribution of the bit budget is proportional to the square of MAD of a VO, which is obtained empirically [27]. As long as the target bit rate is known at the frame level, say R_t, the target bit rate for VO i at time t, V_t^i, is

$$V_t^i = K_t^i \cdot R_t \tag{7}$$

where $K_t^i = (\mathrm{MAD}_t^i \cdot \mathrm{MAD}_t^i)/(\sum_{k=i}^{n} \mathrm{MAD}_t^i \times \mathrm{MAD}_t^i)$ and n are the number of VOs in the coding frame at time t.

In addition to the distribution of bits among VOs in the spatial domain, we also need consider the composition of VOs in the temporal domain. Since each VO has different coding complexity (e.g., low or high motion), to obtain better coding efficiency, it is fairly straightforward to encode the VOs at different frame rates, then the decoder can display the composite video. However, our experimental results show that the significant quality deterioration is experienced in the "gluing" boundary of VOs. Thus, encoding the VOs at the same frame rate is a better alternative to yield better video quality.

Peloton - EX1003, Page 10
IPR2023-01239 (Peloton v. NEC)

TABLE III
EXPERIMENTAL RESULTS OF MULTIPLE VO'S RATE CONTROL

Video sequence	video object	frame rate	bit rate	PSNR_Y	bitrate distribution				
					syntax	motion	shape	texture	video object
akiyo	VO0	100/100	24.06k	36.83	9%	19%	18%	48%	32%
(qcif,24k)	VO1			31.91					68%
coastguard	VO0	95/100	48.22k	28.87	8%	29%	21%	39%	30%
(qcif,48k)	VO1			25.83					32%
	VO2			27.30					18%
	VO3			26.46					20%
container	VO0	79/100	24.28k	31.45	21%	22%	21%	27%	24%
(qcif,24k)	VO1			24.09					24%
	VO2			27.58					10%
	VO3			27.33					15%
	VO4			29.64					13%
	VO5			28.75					14%
akiyo	VO0	149/150	63.97k	40.41	5%	10%	11%	72%	23%
(qcif,64k)	VO1			36.79					77%
container	VO0	150/150	64.21k	31.86	15%	13%	14%	54%	15%
(qcif,64k)	VO1			26.45					22%
	VO2			33.87					12%
	VO3			30.47					11%
	VO4			33.25					8%
	VO5			42.04					31%
news	VO0	147/150	24.06k	39.00	3%	21%	11%	61%	14%
(cif,192k)	VO1			34.19					49%
	VO2			32.94					33%
	VO3			44.33					5%

E. Adaptive Shape-Threshold Control

To avoid using excessive bits for motion and shape information instead of texture and to balance the bit usage without introducing noticeable distortion, the threshold values can be set adaptively based on the previous coding information. The proposed adaptive threshold shape control is described as follows.

For a VO, if the actual bit count used for nontexture information (e.g., shape information) exceeds its estimated bit budget, the encoder will increase the threshold value to reduce the bit count for nontexture coding at the expense of shape accuracy of a VO. On the other hand, if the above condition does not hold, then the encoder decreases the threshold value to get a better video shape accuracy. A scenario of the threshold setting by the proposed adaptive shape control mechanism is depicted in Fig. 5. Initially, the threshold, α_i, for a VO i is set to zero, then it is increased by α_{step} if the target bit rate V_i is larger than H_i $(= syntax_i + motion_i + shape_i$ bits used for object i in the previous frame), otherwise it is decreased by α_{step}. To maintain the accuracy of the video shape to a certain degree, (i.e, to avoid a negative threshold or a large α), the α_i is bounded between 0 and $\alpha_{\max}$. The shape-threshold control mechanism is described as follows:

IF $(V_i \leq syntax_i + motion_i + shape_i)$ **THEN**
$\alpha_i = \max(\alpha_{\max}, \alpha_i + \alpha_{\text{step}})$
ELSE
$\alpha_i = \min(0, \alpha_i - \alpha_{\text{step}})$

F. MB-Level Rate Control

The block diagram of the proposed MB rate control [28] is depicted in Fig. 6. Once the MB-level rate control is activated to provide more strict buffer regulations and higher bit-rate encoding, the following steps are taken. Based on the target bit rate and the quantization level calculated in frame (object) level, the encoder starts to perform the encoding for every MB. The model is now updated on an MB basis, instead of frame (object) basis. Since our rate-control scheme is scalable with the MAD, the flow of the MB rate control is the same as the frame-level rate control with a smaller update step: 1) target bit-rate calculation; 2) quantization parameter calculation; and 3) update the R-D model. Assuming that the following information are available before encoding the first MB of a frame (object): weighting factor w_i for MB i and MAD m_i for MB i. The specific technique in calculating the weighting factor can be selected at the designer's choice [29]. Thus, the MB-level rate control consists of the following steps:

Peloton - EX1003, Page 11
IPR2023-01239 (Peloton v. NEC)

1) Target Bit-Rate Calculation r^i for an MB i: Given R_t for the entire frame or V_t^j for VO j, then

$$r^i = \frac{m^i \cdot w^i}{\sum_{k=1}^{n} m^i \cdot w^k} \cdot R_t \qquad (\text{or } V_t^j) \tag{8}$$

where n is the number of MBs in a frame (object).

2) Quantization Parameter Calculation: Solve QP using m_i and r_i. Note that m_i is not considered in the modeling.

3) Update R-D Model: In the first step, the formula can be derived from the case of equal weighting among MBs. Assuming that weighting is 1 for all MBs, then the problem is formulated as follows:

$$\frac{r_i}{m_i} = \frac{x_1}{q_i} + \frac{x_2}{q_i^2}$$
$$\sum_i r^i = R_t \qquad (\text{or } V_t^i).$$

Since all MBs are equally important, every MB is quantized with the same QP. Thus

$$q_i = \cdots = q_n$$

implies

$$\frac{r_i}{m_i} = \cdots = \frac{r_n}{m_n}.$$

Solving with a total target rate of R_t, we have

$$r_i = \frac{m_i}{\sum_{k=1}^{n} m_k} \cdot R_t.$$

The complete MB rate-control scheme is presented as follows:

```
/* starting encoding a frame (VO) */
calculate T_f and Q, where T_f is the
  target bit count for this frame and
H_q is a threshold for eliminating in-
  significant MBs with low MADs.
IF (m_i < H_q) THEN
  m_i = 0
compute the weighted sum of all existing
  MAD as
mw_f = Σ m_i · w_i .
/* start encoding each MB */
FOR (i = 0; i < N; i++) DO
  quantization parameter calculation
  IF (Q < Q_last) THEN
    Q = max(Q_last − 2, Q)
  ELSE
    Q = min(Q_last + 2, Q)
  encode the MB i
  update the R-D model
  Q_last = Q
  mw_f −= m_i · w_i
  T_f −= T_actual
  IF ( MB i is skipped) THEN
    N_q++
    H_q += m_i
END
H_q = H_q/N_q;
```

IV. Experimental Results

To evaluate the performance of the proposed rate-control scheme, extensive experiments for three different coding granularities with various test conditions are conducted. The performance results of the proposed SRC scheme for the frame level, object level, and MB level are reported in the following sections.

A. Frame-Level Rate Control

For the frame-level rate control, the experiments are conducted in two cases of VBR without any buffer restriction and contant bit-rate CBR with the limited buffer size $(0.5 * target_bit_rate)$, i.e., the maximum accumulated latency is set to be 500 ms, as required by MPEG-4. Tables I and II demonstrate that in VBR and CBR both cases, our frame-level rate control achieves its accuracy in target frame rate and target bit rate. In addition, in CBR case, on average, less than 1% variation of target bit rate and around 3% variation of the target frame rate are experienced. Fig. 7 illustrates that during the encoding process, the buffer occupancy is around 50% of the buffer size with a small variation, and it also shows that with our SRC algorithm, buffer overflow and/or underflow is improbable. The results demonstrate that our SRC scheme can achieve very accurate target bit-rate allocation, and satisfy the buffer with low latency (i.e., $\leq$ 500 ms) for all test sequences. In addition, there are only a few video frames which are skipped in order to meet the buffer constraints, implying that the target frame rate is also obtained.

B. Object-Level Rate Control

For multiple VO rate control, the performance results are reported in Table III. The top three rows report the experiment results obtained in lower bit-rate coding context (e.g., 24–48 kbits/s), comparing to the results reported in the bottom three rows with higher bit-rate coding context (e.g., 64–192 kbits/s.) By examining the results of the number of encoded frames and the target bit-rate allocation, only small variations with target frame rate and target bit rate are experienced. As aforementioned, the allocation of target bit rate among VOs is an important research issue. As reported in the last column of Table III, with our target bit-rate allocation based on the MAD^2, our rate-control scheme demonstrates its capability of automatically and appropriately allocating the target bit rate among VOs. For example, in coding the *Akiyo* video sequence at 10 fps and 25 kbits/s, most of the frame (object)-level bit budget (i.e., 68% of the texture bit budget) is used to encode the VO with the higher coding complexity (e.g., foreground VO), while only 32% of the bit budget is used to encode the VO with lesser coding complexity (e.g., background VO). With its appropriate target bit distribution, each VO has a very similar visual quality for both low bit rate and high bit-rate coding contexts, even

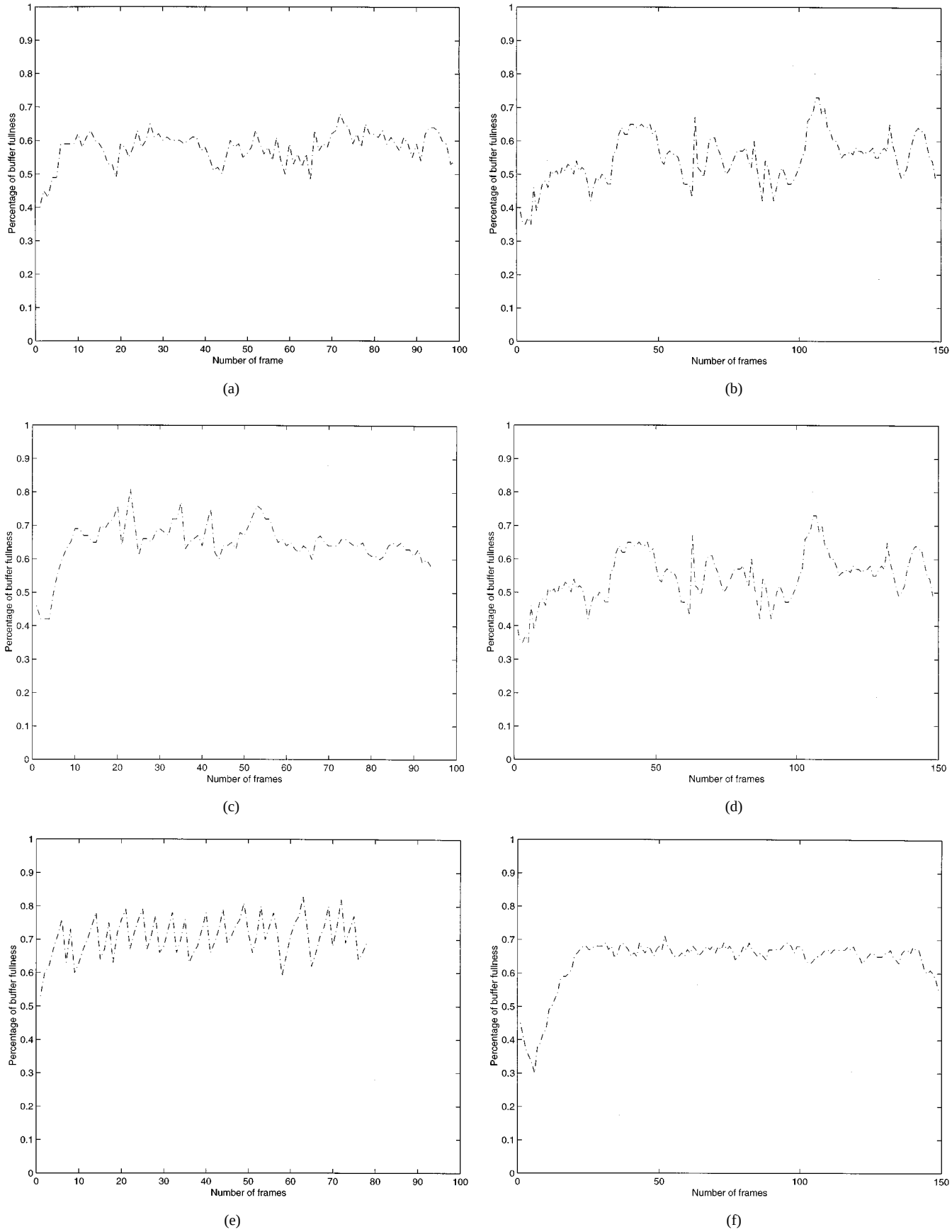

Fig. 8. Object-level rate control: the buffer occupancy for various test conditions. MVORC: (a) $akiyo_q cif$ at 10 fps and 24 kbits/s; (b) $akiyo_qcif$ at 15 fps and 64 kbits/s; (c) $coastguard_q cif$ at 10 fps and 48 kbits/s; (d) $news_cif$ at 15 fps and 192 kbits/s. (e) $container_q cif$ at 10 fps and 24 kbits/s. (f) $container_qcif$ at 15 fps and 64 kbits/s.

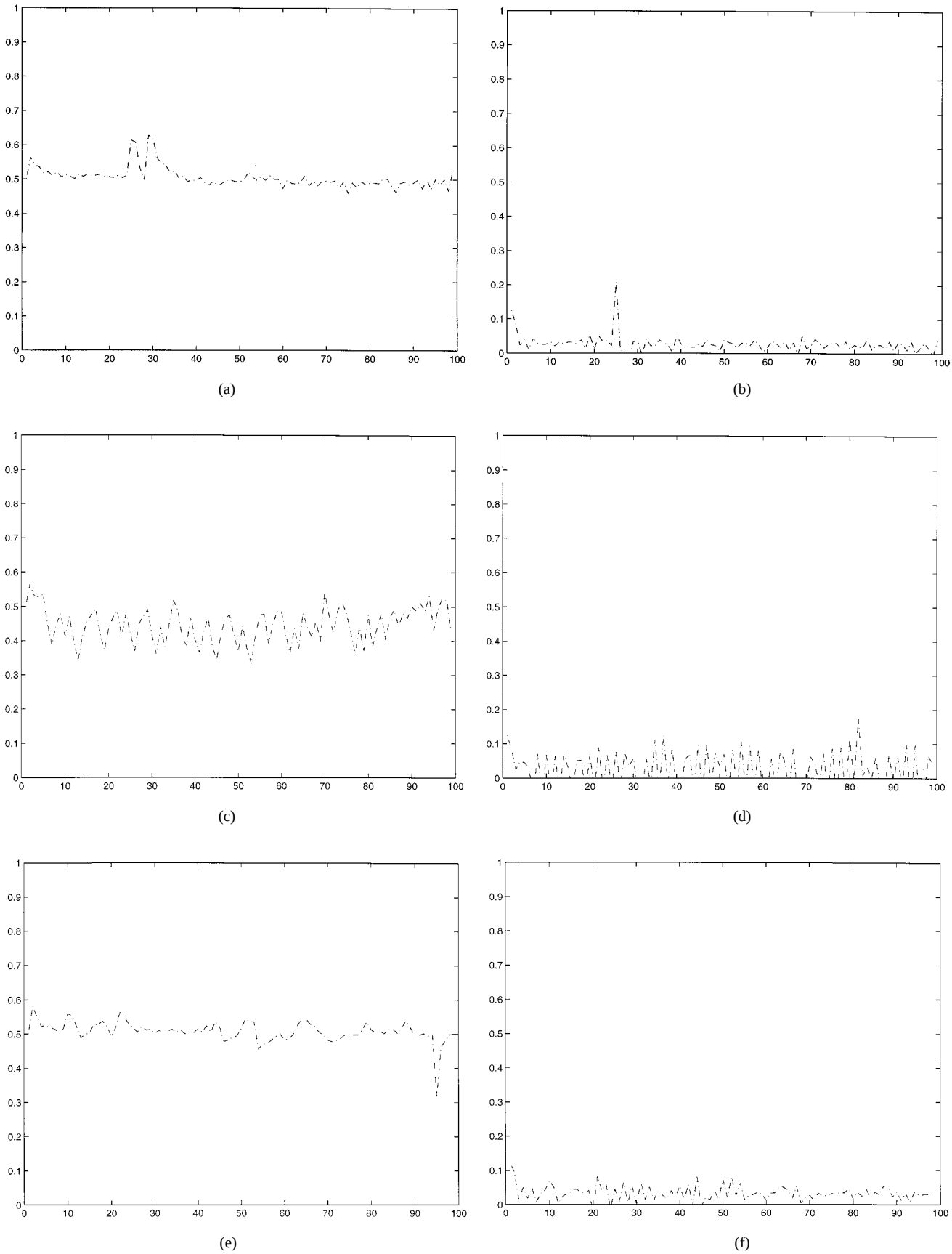

Fig. 9. MB-level rate control: the buffer occupancy for various test conditions under the maximum accumulated delays are 500 ms (left column) and 250 ms (right column). (a), (b): MBRC $coastguard_qcif$ at 10 fps and 24 kbits/s. (c), (d): MBRC $container_qcif$ at 10 fps and 24 kbits/s. (e), (f): mad_qcif at 10 fps and 24 kbits/s.

Peloton - EX1003, Page 14
IPR2023-01239 (Peloton v. NEC)

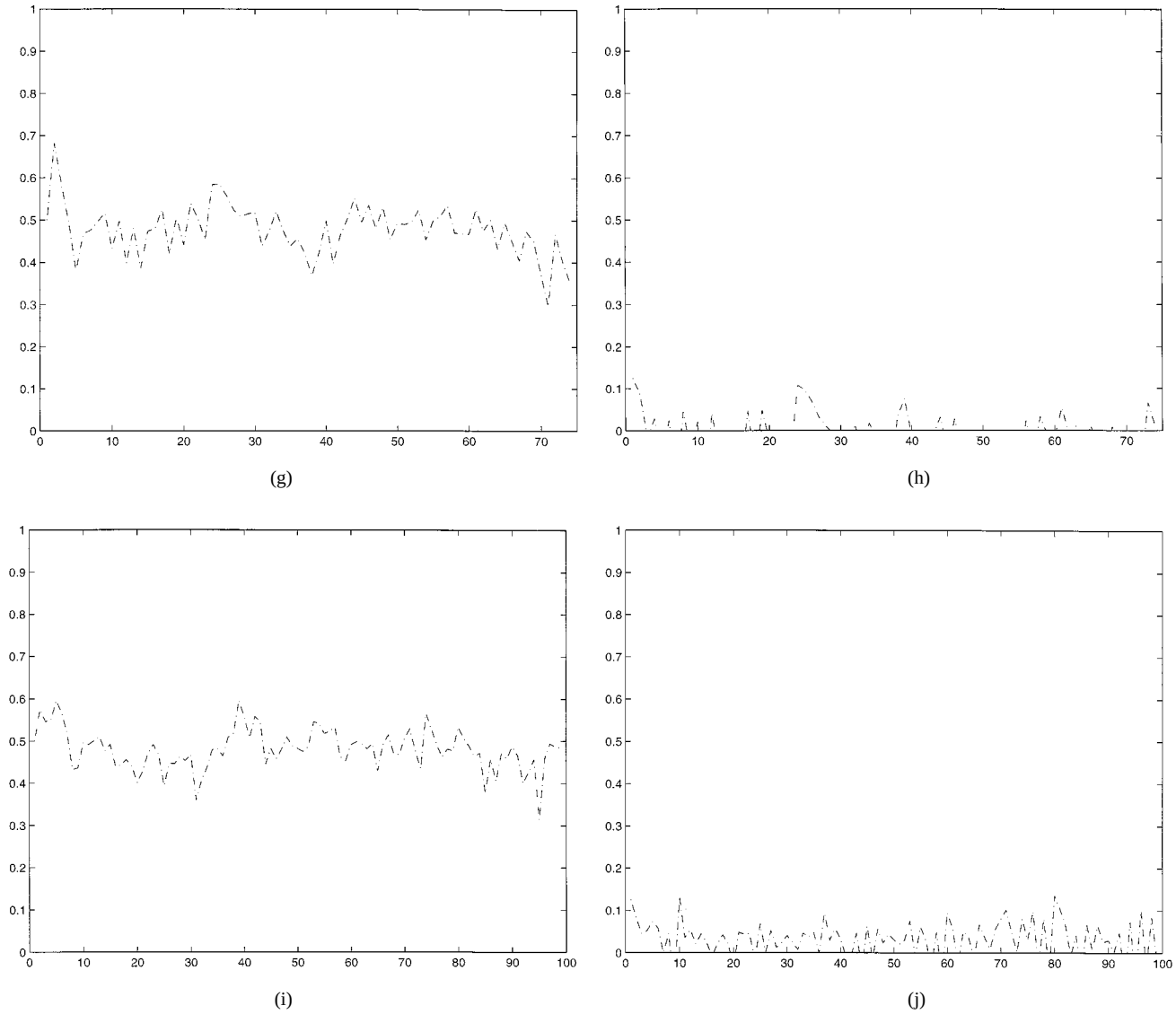

Fig. 9. *(Continued).* MB-level rate control: the buffer occupancy for various test conditions under the maximum accumulated delays are 500 ms (left column) and 250 ms (right column). (g), (h): $news_cif$ at 7.5 fps and 192 kbits/s. (i), (j): $silent_qcif$ at 10 fps and 24 kbits/s. (i), (j): $silent_qcif$ at 10 fps and 24 kbits/s.

TABLE IV
EXPERIMENTAL RESULTS OF MB-LAYER RATE CONTROL GIVEN THE MAXIMUM ACCUMULATED DELAYS ARE 500 AND 125 MS

video sequence	max. delay=500ms			max. delay=125ms		
	frame rate	bitrate	PSNR_Y	frame rate	bitrate	PSNR_Y
coast(qcif,48k)	100/100	47.98k	29.04	99/100	47.78k	29.09
coast(qcif,112k)	150/150	112.00k	31.62	150/150	111.82k	31.60
container(qcif,24k)	100/100	24.13k	32.55	100/100	23.81k	32.56
container(qcif,10k)	75/75	99.77k	30.05	74/75	9.91k	30.06
hall(qcif,10k)	75/75	9.95k	30.62	75/75	9.95k	30.51
mad(qcif,24k)	100/100	24.02k	35.01	100/100	23.90k	34.96
mad(qcif,10k)	73/75	10.13k	32.11	70/75	9.92k	32.27
silent(qcif,24k)	100/100	24.04k	30.74	100/100	23.95k	30.77
news(cif,112k)	150/150	111.75k	34.09	150/150	111.61k	34.29
news(cif,48k)	75/75	48.26k	31.64	75/75	47.58k	31.62

though there exist fairly large dB differences up to 5, 3, 7, 3, 12, and 11 dB between objects in the respective tests. It is obvious that the background object has pretty much static with very little difference from its previous frame. In this case, the peak signal-to-noise-ratio (PSNR) value is not a sufficient indicator for visual quality. Our experimental results show that equal or similar PSNR values in this case create more coding artifacts in background objects and in a composite frame. So, even the PSNR values are fairly different among VOs; however, their composite video frame shows a better visual quality.

Next the buffer occupancy is examined, and plotted in Fig. 8. In our experiments, a single joint buffer control is used and the maximum accumulated delay is set to be 500 ms for the encoding process. As shown in Fig. 8, the buffer occupancy is maintained at around 50%–60% of the buffer size with a small variation, as expected, even for the case that there are up to six VOs (e.g., container video sequence). It also demonstrates that with our rate-control scheme, the buffer overflow or underflow does not occur for all test conditions of high- and low-bit rate coding contexts.

C. MB-Level Rate Control

Table IV reports the results of MB layer rate control under two maximum accumulated latencies which are set to be 500 and 250 ms, respectively. Compared to the frame-level rate control, the MB-level rate control generates more accurate target bit estimation, resulting in the higher frame rate, and lower latency delay at the expense of slightly quality loss. For the case of the maximum accumulated delay equal to 250 ms, a larger—but within the tolerance—fluctuation of buffer occupancy is experienced, as expected, since a smaller buffering is used to smooth out the bit rate fluctuation of encoding process (see Fig. 9). However, under this strict test condition, no buffer overflow occurs, proving the robustness and stability of our MB rate-control scheme. The results show that our MB-level rate control can be applied for the various applications with higher bit rate and strict buffer regulations.

V. Summary and Conclusions

In this paper, we proposed a single, integrated and SRC scheme for different coding granularity including the frame level, the object level, and the MB level. Our scheme not only provides a framework for the very low bit-rate control, but also comes up with several new techniques to improve the accuracy of the R-D model. Those new methods or concepts are listed below:

1) a more accurate R-D model which is scalable with MAD;
2) a dynamically bit-rate allocation among VOs with various coding complexities;
3) a sliding-window mechanism for smoothing the impact of scene change;
4) an adaptive selection criterion of data points for the R-D model update process;
5) an adaptive threshold setting for rate reduction in shape coding;
6) an effective frame-skipping control for the prevention of the potential buffer-overflow problem.

The proposed SRC has shown the following advantages: 1) low latency and the limited buffer constrains are satisfied for CBR applications; 2) the VBR quality is maintained; 3) both the target bit rate and the target frame rate are obtained within a negligible error; and 4) easy extension to the multiple VOs and MB layer. The proposed SRC has been adopted in the International Standard of the emerging MPEG4 standard.

References

[1] *Coding of Moving Pictures and Associated Audio ISO/IEC 14496-2*, MPEG-4 Committee Draft ISO/IEC JTC1/SC29/WG11, Oct. 1997.

[2] *Coding of Moving Pictures and Associated Audio ISO/IEC 14496-2*, MPEG-4 Committee Draft ISO/IEC JTC1/SC29/WG11, Oct. 1998.

[3] H. Zhang and D. Ferrari, "Rate-controlled static priority queueing," in *Proc. IEEE INFOCOM'93*, Apr. 1993, pp. 227–236.

[4] D. Ferrari and D. Verma, "A scheme for real-time channel establishment in wide-area networks," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 368–379, Apr. 1990.

[5] L. Zhang, "Virtual clock: A new traffic control algorithm for packet switching networks," in *Proc. ACM SIGCOMM'90*, Sept. 1990, pp. 19–29.

[6] S. J. Golestani, "A stop-and-go queueing framework for congestion management," in *Proc. ACM SIGCOMM'90*, Sept. 1990, pp. 8–18.

[7] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1445–1453, 1988.

[8] S. W. Wu and A. Gersho, "Rate-constrained optimal block-adaptive coding for digital tape recoding of HDTV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, Mar. 1991.

[9] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit Allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sept. 1994.

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

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

[12] L. J. Lin and A. Ortega, "Bitrate control using Piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 446–459, Aug. 1998.

[13] W. Ding, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 12–20, Feb. 1996.

[14] ——, "Joint encoder and channel rate control of VBR video over ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, Aug. 1997.

[15] B. Tao, H. A. Peterson, and B. W. Dickinson, "A Rate-Quantization model for MPEG Encoders," in *Proc. 1997 Int. Conf. Image Processing*, Oct. 1997, pp. 338–341.

[16] 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. 1997 Int. Conf. Image Processing*, Oct. 1997, pp. 41–44.

[17] *Joint rate control for multiple-VO*, MPEG96/M1890 ISO/IEC JTC/SC29/WG11, Apr. 1997.

[18] A. Vetro, H. F. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 186–199, Feb. 1999.

[19] J. Ribas-Corbera and S. M. Lei, "Contribution to rate control Q2 experiment: A quantizer control tool for achieving target bitrates accurately," Coding of Moving Pictures and Associated Audio MPEG96/M1812 ISO/IEC JTC/SC29/WG11, Sevilla, Spain, Feb. 1997.

[20] ——, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 172–185, Feb. 1999.

[21] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for very low bitrate video," in *Proc. 1997 Int. Conf. Image Processing*, vol. II, Oct. 1997, pp. 768–771.

[22] T. Chiang, "A rate control scheme using a new rate-distortion model," JCoding of Moving Pictures and Associated Audio MPEG95/0436 TC1/SC29/WG11, Dallas, TX, Nov. 1995.

[23] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using a new rate-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 246–250, Feb. 1997.

Peloton - EX1003, Page 16
IPR2023-01239 (Peloton v. NEC)

[24] A. Viteribi and J. Omura, "A new rate control scheme using a new rate-distortion model," in *Principle of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
[25] *Coding of Moving Pictures and Associated Audio MPEG96*, MPEG-4 video verification model V5.0 ISO/IEC JTC1/SC29/WG11, Nov. 1996.
[26] *Coding of Moving Pictures and Associated Audio MPEG96*, MPEG-4 video verification model V8.0 ISO/IEC JTC1/SC29/WG11, July 1997.
[27] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Multiple-VO Rate Control," ISO/IEC JTC/SC29/WG11 MPEG97/M2554, July 1997.
[28] T. Chiang, H.-J. Lee, and Y.-Q. Zhang, "Macroblock Layer Rate Control," ISO/IEC JTC/SC29/WG11 MPEG97/M2555, July 1997.
[29] S. Ryoo, J. Shin, and E. Jang, "Rate control tool: Based on human visual sensitivity (HVS) for low bitrate coding," ISO/IEC JTC/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG96/M0566, Munich, Germany, Jan. 1996.

Hung-Ju Lee (M'96) received the B.S. degree from Tatung Institute of Technology, Taipei, Taiwan, in 1987, and the M.S. and Ph.D. degrees from Texas A&M University, College Station, TX, in 1993 and 1996, respectively, all in computer science.

In 1996, he joined Sarnoff Corporation, Princeton, NJ, as a Member of Technical Staff. He actively participates in ISO's MPEG digital video standardization process, focusing particularly on wavelet-based visual texture coding and scalable rate control for MPEG-4 video. He received Sarnoff Technical Achievement Award in 1998 for his contributions on the development of MPEG-4 rate control. His current research interests include image and video coding, and network resource management for multimedia applications.

Tihao Chiang (S'90–M'95–SM'99) was born in Cha-Yi, Taiwan, R.O.C., in 1965. He received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1987, the M.S. degree from Columbia University in 1991, and the Ph.D. degree from Columbia University, New York, in 1995, all in electrical engineering.

In 1995, he joined David Sarnoff Research Center, Princeton, NJ, as a Member of Technical Staff. He was later promoted to Technology Leader and then Program Manager. While at Sarnoff, he led a team of world-class researchers and developed an optimized MPEG-2 software encoder. For his work in the encoder and MPEG-4 areas, he received two Sarnoff Achievement Awards and three Sarnoff team awards. In September 1999, he joined the faculty at National Chiao-Tung University, Taiwan, R.O.C. Since 1992, he has actively participated in ISO's MPEG digital video-coding standardization process, with particular focus on the scalability/compatibility issue. He is currently the co-chair for encoder optimization on the MPEG-4 committee, has made more than 40 contributions to the MPEG committee over the past eight years, and has co-authored the rate-control technology that was adopted as part of the MPEG-4 International Standards in 1998. He holds two U.S. patents and more than 10 pending patents. He published over 20 technical journal and conference papers in the field of video and signal processing. His main research interests include compatible/scalable video compression, stereoscopic video coding, and motion estimation.

Ya-Qin Zhang (F'97) was born in Taiyuan, China, in 1966. He received the B.S. and M.S. degrees from the University of Science and Technology of China (USTC) in 1983 and 1985, respectively, and the Ph.D. degree from George Washington University, Washington, DC, in 1989, all in electrical engineering.

In 1999, he joined Microsoft Research, Beijing, China, as the Assistant Managing Director. Previously, he was the Director of the Multimedia Technology Laboratory, Sarnoff Corporation, Princeton, NJ. During 1989–1994, he was with GTE Laboratories Inc., Waltham, MA, and Contel Technology Center, Washington, DC. He has authored and co-authored over 150 refereed papers and 30 U.S. patents in digital video, Internet, multimedia, wireless, and satellite communications. Many of the technologies he and his team developed have become the basis for start-up ventures, commercial products, and international standards. His recent focus has included MPEG2/DTV, MPEG4/VLBR, and the Internet. He has been an active contributor to the ISO/MPEG and ITU standardization efforts in digital video and multimedia.

Dr. Zhang served as the Editor-In-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from July 1997 to July 1999. He serves on the Editorial boards of seven other professional journals and over a dozen conference committees.