Rate Control for Low-Bit-Rate Video via Variable-Encoding Frame Rates

Hwangjun Song and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*—A novel rate-control algorithm with a variable-encoding frame-rate method is proposed in this work for low-bit-rate video coding. Most existing rate-control algorithms for low-bit-rate video focus on bit allocation at the macroblock level under a constant frame-rate assumption. The proposed rate-control algorithm is able to adjust the encoding frame rate at the expense of tolerable time-delay. The new rate-control algorithm attempts to achieve a good balance between spatial quality and temporal quality to enhance the overall human perceptual quality at low bit rates. It is demonstrated that the rate-control algorithm achieves higher coding efficiency at low bit rates, with a low additional computational cost. The proposed variable-encoding frame-rate method is compatible with the bit-stream structure of H.263+.**

I. Introduction

DIGITAL video-coding techniques have advanced rapidly in recent years. International standards such as MPEG-1, MPEG-2 [1], H.261, and H.263 have been established to accommodate different application needs. New standards such as H.263+/++ [2], H.26L, and MPEG-4 are also under development to achieve more functionalities. Among various video-coding standards, we focus on the rate-control problem for a near-term enhancement of H.263, known as H.263+ in this paper. H.263+ is an emerging low-bit-rate video compression standard that provides various advanced options [2]. Core ingredients of H.263+ include block-based motion compensation and block-based DCT coding. Rate control plays a critical role in the video encoder. It regulates the coded bit stream to satisfy certain given conditions, on the one hand, and enhances the quality of coded video, on the other hand. However, the rate-control algorithm is often not standardized, since it can be independent of the decoder structure.

Depending on channel conditions and characteristics of the storage medium, many MPEG rate-control algorithms have been proposed. Among the three types of frames in MPEG-1/2 and H.263+, i.e., intra (I)-, predictive (P)-, and bidirectional (B)-frames, I-frames are required to control the accumulated mismatch error and scene change. MPEG 1 and 2 can provide high spatial/temporal quality at high bit rates. Generally speaking, I-frames are encoded with high spatial quality. The spatial quality of predictive frames is degraded gradually after I-frames. This phenomenon is more obvious for fast-moving video. Thus, I-frames are required to refresh the accumulated mismatch error. H.263+ recommends the macroblock-based update, where I-frames can be employed for significant scene change to reduce the output bit rate.

The spatial and temporal compression artifacts of coded video have recently been studied intensively. Blocking, ringing, and texture-deviation artifacts are often observed in low-bit-rate video as spatial quality degradations. As to temporal visual degradation, few research results are available. Flickering (or blinking) and motion jerkiness are the two major artifacts often observed. The flickering artifact is caused by the fluctuation of spatial image quality between adjacent frames, while motion jerkiness occurs when there is an abrupt change of the coding frame rate or when the frame rate goes below a certain threshold required to generate smooth motion. Despite the fact that the change of the peak signal-to-noise ratio (PSNR) does not correspond to flickering completely, we observe that the flickering effect can be reduced by keeping the image quality of each frame almost constant. Since the measure of the flickering effect and motion smoothness is a very complicated and challenging problem, the subjective visual test is always needed to evaluate the performance of a video-coding scheme.

It is clear that the coding error of I-frames can be propagated to subsequent predictive frames, such as P- and B-frames. However, the coding of I frames demands much higher bit rates than P- and B-frames, since motion compensation is not employed. As the bandwidth becomes narrower, the bit budget for I-frames is a growing burden for the whole coded bit stream. One important difference between MPEG and H.263+ rate-control algorithms is in the length of group of pictures (GOP). Note that even though the GOP term is not employed in H.263, a definition similar to that given by MPEG can be adopted. That is, it includes a leading I frame and all its following frames before the appearance of the next I frame. In MPEG, one GOP consists of one I-frame and several predictive B- and P-frames. The GOP structure is repeated periodically, so that it is often used as a basic rate-control unit. In H.263+, we have to reduce the number of I-frames due to the low-bit-rate constraint. It implies that the frame number between adjacent I-frames in H.263+ should be very large. Consequently, GOP is not suitable to be used as a basic rate-control unit due to the high computational complexity and long time-delay. This explains the reason why existing MPEG-1 and MPEG-2 frame-level bit-allocation algorithms cannot be straightforwardly extended to H.263+.

Manuscript received April 24, 1998; revised September 30, 1999. This work was supported by the Integrated Media Systems Center (a National Science Foundation Engineering Research Center), by the Annenberg Center for Communication, University of Southern California, and by the the California Trade and Commerce Agency. This paper was recommended by Associate Editor C. W. Chen.

H. Song is with the School of Electronic and Electrical Engineering, Hongik University, Seoul, Korea (e-mail: hwangjun@wow.hongik.ac.kr).

C.-C. J. Kuo is with the Integrated Multimedia Systems Center and the Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: cckuo@sipi.usc.edu).

Publisher Item Identifier S 1051-8215(01)03012-9.

It is worthwhile to point out that rate-control algorithms, in terms of bit allocation at the macroblock level, have been studied for H.263. In [3], Wiegand *et al.* applied rate-distortion (R-D) theory for the optimal selection of modes in the H.263 video encoder. The Viterbi algorithm was employed to find the optimal path through a trellis. This approach was extended by Mukherjee and Mitra to combine the optimal macroblock mode selection and the macroblock quantization step adaptation in [4]. Another practical rate-control algorithm was considered by Ribas-Corbera and Lei [5], where rate and distortion models of macroblocks were derived, and a rate-control algorithm based on derived models was studied. Their main contribution is the macroblock-level bit-allocation to keep the output bit-stream rate almost constant with small fluctuation for the low time-delay and jitter under constant bit rate (CBR) channel. Even though a study of the frame-level bit allocation was also performed in [5], results on this part were still preliminary. Since the temporal quality of coded video was not considered explicitly in [5], the resulting quality can be degraded. Buffer constraints are also important in rate control since the buffer underflow and overflow problems at the encoder and the decoder can degrade video quality. The general buffer-constraint problem was examined by Reibman and Berger [6] and Hsu *et al.* [7].

We examine the rate-control algorithm for H.263+ with very large spacing between I-frames in this research. The proposed rate-control algorithm attempts to achieve a good balance between spatial quality and temporal quality to enhance the overall human perceptual quality by adopting a variable-encoding frame rate. We cannot support high spatial and temporal video quality at low bit rates. Up until now, only spatial quality degradation has been considered under a fixed-encoding frame rate in most work. The variable-encoding frame rate control can improve the visual perceptual quality by pursuing an efficient tradeoff between spatial and temporal qualities. Furthermore, more flexible and robust rate control is needed under time-varying communication channels such as the Internet and the mobile channels. Under these environments, variable-encoding frame-rate control can provide a satisfactory solution [8].

Usually, the I- and P-frames are considered separately for the H.263 rate-control problem. In this work, the I-frame is encoded by using the current H.263 I-frame coding scheme at a predetermined bit rate, which highly depends on the time delay requirement. The H.26L Evaluation Delay Model User Guide recommends that the bit rate for the I-frame must not be greater than one second worth of bit transmission at the assumed channel bit rate. Our focus here is the determination of an appropriate frame rate and the bit allocation at the frame layer. The proposed variable-encoding frame rate method is compatible with the bit-stream structure of H.263+.

This paper is organized as follows. We briefly review the rate-control problem and formulate the problem of our interest in Section II. Then, the proposed variable-encoding frame rate-control and the bit-allocation algorithm are described in Section III. Experimental results are given in Section IV. Finally, concluding remarks are provided in Section V.

II. Problem Formulation

Generally speaking, rate control for a video codec can be done at two levels: the frame and the macroblock levels. Most effort on H.263 rate control has been spent on the macroblock-level bit allocation. Efficient macroblock-level bit-allocation algorithms have been proposed in [3]–[5], [9]–[12]. However, one cannot treat the temporal quality of video effectively with the macroblock-level bit allocation only. In this work, we examine the frame-level bit allocation. Two approaches have been considered in the frame-level bit allocation for long image sequences, i.e., dependent [13], [14] and independent frame coding schemes [15], [16]. Since the GOP structure of MPEG is relatively short and there are quite a few I frames available over hundreds of frames, dependency among frames is less important. The frame dependency becomes more important as the frame number between two adjacent I-frames increases in the context of H.263. Thus, the dependent frame coding scheme is adopted here.

The dependent frame coding without buffer constraints for MPEG rate control can be described as follows. Determine $q_i, i = 1, 2, \ldots, N$ to minimize

$$\sum_{i=1}^{N} d_i(q_1, q_2, \ldots, q_i), \quad \text{subject to } \sum_{i=1}^{N} r_i(q_1, q_2, \ldots, q_i) \le B \tag{1}$$

where N is the frame number of a GOP, B is the given bit budget for a GOP, and $d_i(q_1, q_2, \ldots, q_i)$ and $r_i(q_1, q_2, \ldots, q_i)$ are the distortion measure and allocated bit rates for the ith frame.

By using the Lagrange multiplier method, we can define a penalty function by combining the cost function and the constraint for minimization, i.e.,

$$\min_{q_1, q_2, \ldots, q_N} \sum_{i=1}^{N} J_i(q_1, q_2, \ldots, q_i) \tag{2}$$

where

$$J_i(q_1, q_2, \ldots, q_N) = d_i(q_1, q_2, \ldots, q_N) + \lambda r_i(q_1, q_2, \ldots, q_i)$$

and where λ is the Lagrange multiplier which serves as a weighting factor for the constraint. Many methods have been proposed to solve the above optimization problem. Well-known examples include the Viterbi algorithm [16], [13] and the gradient method [15], [14]. The Viterbi algorithm is basically a trellis-based dynamic programming procedure. Although it can guarantee the optimal solution, it is usually too complicated to be used in practice. As to gradient-like search algorithms, its computational complexity increases exponentially with the number of frames in one GOP. Also, it is unavoidable to have the time-delay of one GOP due to the above formulation. Since one GOP delay is very significant in H.263, the formulation in (1) has to be modified. In the following, we reformulate the above dependent frame coding problem to reduce the computational complexity and time-delay and to make the solution more tractable with two simplifying assumptions.

First, the I and P frames are often treated separately in developing the H.263 rate-control algorithm. By assuming that the bit rate for the I frame is determined by the maximum buffer size, time delay and image quality requirements, we can have a simpler problem, i.e., to determine $q_i, i = 1, 2, 3, \ldots, N-1$ to minimize

$$d_I + \sum_{i=1}^{N-1} d_i(d_I, q_1, q_2, \ldots, q_i),$$
$$\text{subject to } r_I + \sum_{i=1}^{N-1} r_i(r_I, q_1, q_2, \ldots, q_i) \le B \qquad (3)$$

where d_I and r_I are the distortion measure and the bit rate for the I-frame, respectively. They are considered to be known quantities when one attempts to solve (3) for simplicity.

Second, we have to look for a basic unit for rate control which is different from the GOP structure used in MPEG. Although the PB-frame mode and the improved PB-frame mode are supported as advanced modes, the standard frame structure of H.263+ is $IPPPPPP\ldots PP$. Based on the reason given above, we can exclude the I frame and modify the GOP definition to be all P-frames between two adjacent I-frames. As the frame number of GOP becomes longer, the quality improvement due to rate control is more substantial but at the expense of a larger computational complexity and longer time delay. To achieve a balance, we divide one GOP into several sub-GOPs and use each sub-GOP as the basic rate-control unit. At the same time, dependency among sub-GOPs is considered to compensate the performance degradation caused by the division of GOP.

We develop an variable-encoding frame rate control and bit-allocation algorithm that preserves the quality of P frames constant as much as possible along the time in Section III.

III. Frame-Rate Control and Bit Allocation

As stated earlier, the objective of rate control presented in this work is to keep the quality of P frames nearly constant or degrade very slowly. Since each P-frame is used as the reference frame for the following P-frame, quality degradation propagates to later frames when a P-frame is degraded severely. Thus, the quality of P-frames has to be kept in a tolerable range to reduce error propagation, and the dependent frame coding framework [13] has to be considered. The proposed rate-control algorithm consists of two parts: control of the encoding frame rate and bit allocation at the frame level.

It is difficult to support both good spatial and temporal quality at very low bit rates. An encoding frame-rate-control scheme is proposed for a tradeoff of spatial/temporal quality. The proposed scheme aims at the reduction of temporal degradation in terms of motion jerkiness perceived by human beings. Since the R-D characteristics of predictive frames are related to the amount of motion involved, we will predict the R-D relation roughly based on the motion information existing in video and develop an encoding frame rate-control scheme accordingly in Section III-A. For H.263+ implementation, the information of the variable-encoding frame rate can be stored by temporal reference (TR) and temporal reference for B-frames (TRB) in the header and the optional *custom picture clock frequency code* [2].

TABLE I
CODING WITH EVEN-INDEXED FRAME NUMBER

Frame rate	Encoded frame No.
1/1	1,2,3,4,5,6,7,8,9,10,11,12
1/2	2,4,6,8,10,12
1/3	3,6,9,12
1/4	4,8,12
1/6	6,12
1/12	6

TABLE II
CODING WITH ODD-INDEXED FRAME NUMBER

Frame rate	Encoded frame No.
1/1	1,2,3,4,5,6,7,8,9,10,11,12
1/2	1,3,5,7,9,11
1/3	1,4,7,10
1/4	1,5,9
1/6	1,7

After the encoding frame rate is specified, the second issue is bit allocation at the frame level. Since the computational complexity to determine the optimal frame-level bit allocation increases exponentially with the length of GOP and results in longer time delay, it is very difficult to get the optimal bit allocation for one GOP in practice. Instead of getting the optimal solution for one GOP, we seek a local optimal solution within a sub-GOP to reduce the computational complexity and time delay and, in the meanwhile, dependency among sub-GOPs will be also handled with a low additional computational complexity. The length of sub-GOP has a direct consequence on time delay. In our design, each sub-GOP consists of 12 frames, which is the least common multiplier of 2, 3, 4, and 6. With a frame rate of 30 frames per second (fps), this implies a time delay of 400 ms. We also assume that one GOP is made up of $12M$ frames, where M is the number of sub-GOPs within one GOP. In the experiment, M is chosen to be 8. This procedure of frame-level bit allocation is described in Section III-B. It is worthwhile to point out that the problem of macroblock-level bit allocation is not our concern and several existing methods can be employed [3]–[5], [9]–[12].

A. Encoding Frame-Rate Control

By encoding frame-rate control, we can avoid or reduce the sudden frame skipping in existing rate-control algorithms, which degrades motion smoothness disastrously. Two problems have to be addressed for frame rate control. They are: 1) when the frame rate should be changed and 2) how to change the encoding frame-rate to preserve motion smoothness.

We assume that the video camera generates 30 fps. It is well known that human eyes are sensitive to abrupt (temporal) interval change between adjacent frames, where unsmooth motion is perceived more obviously. Thus, to handle the second problem, we have to choose the encoded frame positions properly to reduce the degradation due to motion unsmoothness. The relation between the encoding frame-rate and the encoded frame position adopted in this work is given in Tables I and II. Usu-

TABLE III
CALCULATION OF $\hat{D}_e$ BASED ON HOD ($\omega_h = 3$) AND THE RESULTING FRAME RATE FOR EACH SUB-GOP FOR SALESMAN WHEN $T = 0.03$

No. of sub-GOP	Slope of approx. line	Last HOD	$m(D_h)$	$\hat{D}_e$	No. of encoded frames
1	0.0063	0.062	0.039	0.0804	3 (initial value)
2	-0.0061	0.026	0.045	0.0080	2
3	0.00049	0.019	0.018	0.0202	3
4	0.00030	0.021	0.022	0.0220	3
5	0.00486	0.055	0.037	0.0696	3
6	-0.0063	0.018	0.067	-0.0008	2
7	0.00063	0.009	0.005	0.0109	3
8					3

ally, if the even-indexed (or odd-indexed) frames are chosen in the current sub-GOP, the even-indexed (or odd-indexed) frames will also be chosen in the next sub-GOP. The only exception is that if the $1/12$ frame rate in Table I is chosen for the current sub-GOP, then the 1/6 frame rate in Table II must be used for the next sub-GOP to form an alternating pattern. Thus, the encoding frame-rate will not change abruptly. As a result, the degradation in motion smoothness will not be noticeable or at least not obvious visually. Our scheme can support a frame rate ranging from 3.7 fps (which corresponds to three coded frames for 24 frames) to 30 fps. Furthermore, to ensure a smooth encoding frame-rate change, it is also required that the encoding frame-rate can only move up or down one level at one time. For example, if the frame rate in the current sub-GOP is $1/3$, then the allowed frame rates in the next sub-GOP are $1/2, 1/3$ and $1/4$. By imposing the frame-rate change pattern and the encoded frame position in the sub-GOP, we can provide a smooth transition in the location of selected coding frames.

In principle, the R-D curve of the current sub-GOP can be used to estimate the appropriate encoding frame-rate for the next sub-GOP. Many R-D models have been proposed to speed up the rate-control algorithm, including the MPEG test model, the statistical model, the exponential model, and the approximating spline model [14]. The R-D curve provides a good estimate of the encoding frame-rate if the model is accurate. However, an accurate R-D model requires a large amount of computation. In this work, two simplified approaches are adopted. One is based on the histogram of difference image (HOD) while the other uses the ratio between the rate and quality of frames. Both methods can be used to detect motion change in video with a low computational cost.

Histogram-based methods [17] such as the difference of histograms (DOH), the block histogram difference (BH), and the histogram of difference image (HOD) can be used to detect motion activities in video. Since the required bit rates for predictive frames are also related to the motion activity (i.e., higher bit rates for faster motion), histogram based methods can be used to estimate the proper encoding frame-rate as a short cut. Besides histogram-based methods, other methods [17] such as block variance difference (BV) and motion compensation error (MCE) can also be used. MCE is too expensive computationally. DOH and BH perform better in detecting global changes rather than local motion, while HOD is more sensitive to local motion. HOD is examined below.

TABLE IV
RESULTING VARIABLE ENCODING FRAME RATES FOR THE FOREMAN SEQUENCE

No. of sub-GOP	No. of encoded frames
1	3 (initial value)
2	3
3	2
4	3
5	3
6	2
7	3
8	2

TABLE V
VALUES OF THE WEIGHTING FACTOR λ_m USED TO DETERMINE THE QUANTIZATION PARAMETER IN EACH SUB-GOP FOR ALL SEQUENCES

No. of sub-GOP(m)	λ_m
1	0.001
2	0.002
3	0.002
4	0.01
5	0.01
6	0.01
7	0.1
8	1

One way to measure the difference of two frames f_n and f_m can be defined as

$$D_h(f_n, f_m) = \frac{\sum_{i>|\mathrm{TH}_0|} \mathrm{hod}(i)}{N_{\mathrm{pixel}}} \tag{4}$$

where

- i index of the quantization bin;
- $\mathrm{hod}(i)$ histogram of the difference image;
- TH_0 threshold value for detecting the closeness of the position to zero;
- N_{pixel} number of pixels.

Once all HOD values for consecutive frames in the current sub-GOP are calculated, the estimated HOD value $\hat{D}_e$ for the following sub-GOP can be computed by

$$\hat{D}_e = D_h + \omega_h a_h \tag{5}$$

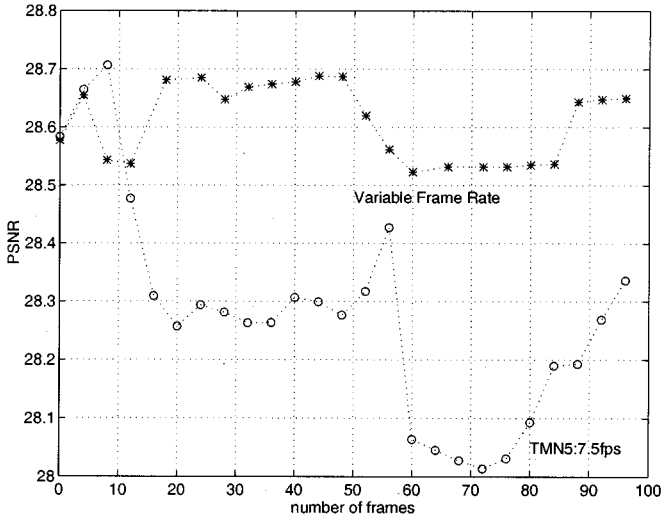

Fig. 1. PSNR performance comparison between TMN5 and the proposed rate-control algorithm for the Y component for Salesman with TMN5.

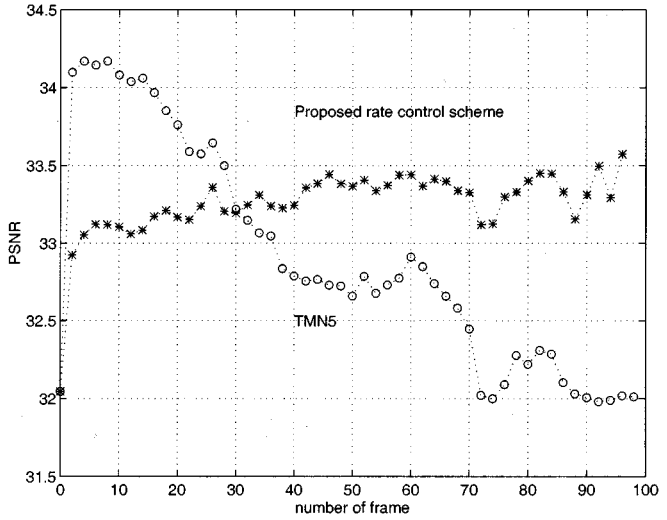

Fig. 2. PSNR performance comparison between TMN5 and the proposed rate-control algorithm for the Y component for Akiyo with TMN5.

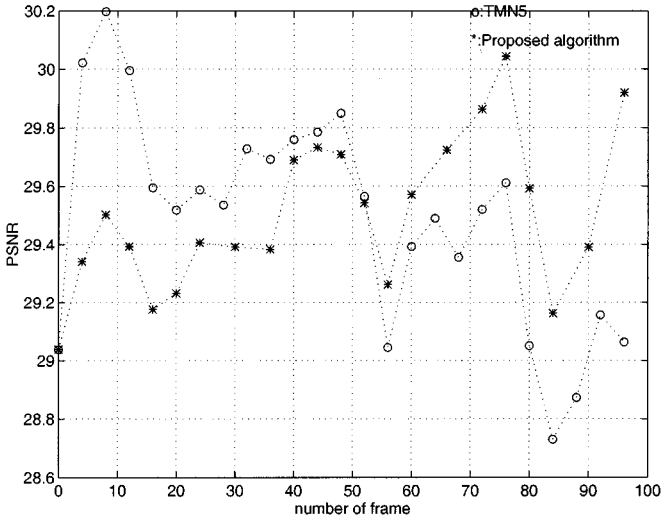

Fig. 3. PSNR performance comparison between TMN5 and the proposed rate-control algorithm for the Y component for Foreman with TMN5.

where

- D_h HOD value between the two last encoded frames in the current sub-GOP;
- ω_h weighting factor;
- a_h slope of the approximating line which minimizes the mean square error of HODs in the current sub-GOP.

Actually, D_h provides the latest motion change information and a_h corresponds to the current motion change trend. Based on them, we can estimate the future motion change by using (5).

The second approach to estimate the motion activity is to consider the ratio between the rate and the distortion function, i.e.,

$$D_r = \frac{r}{d_{\text{psnr}}} \tag{6}$$

where r is the consumed bit rate and d_{psnr} is the PSNR of the encoded frame of the current sub-GOP. Since a higher bit budget is required for faster motion in predictive frames, faster motion in video leads to a larger ratio. As a result, the ratio can also be used to determine the supportable encoding frame-rates. Similar to the histogram-based case, we adopt the estimate for the following sub-GOP:

$$\hat{D}_e = D_r + \omega_r a_r \tag{7}$$

where

- D_r R-D ratio value of the last two encoded frames in the current sub-GOP;
- ω_r weighting factor;
- a_r slope of the approximating line which minimizes the mean square error of ratio coefficients in the current sub-GOP.

Let us use $m(D_h)$ to denote the mean of all HOD values of frames in the current sub-GOP. Then, we can adjust the encoding frame-rate for the next sub-GOP based on the difference of $\hat{D}_e$ and $m(D_h)$. That is

1) if $\delta \geq T$, the encoding frame-rate is decreased by one level;
2) if $\delta \leq -T$, the encoding frame-rate is increased by one level;
3) if $|\delta| < T$, the encoding frame-rate remains the same.

where $\delta = \hat{D}_e - m(D_h)$ and the threshold value T is chosen to be the averaged HOD over the first sub-GOP. By changing D_h to D_r, we can obtain the decision rule in terms of the ratio between the rate and the distortion function as well.

It is interesting to point out that both a_h and a_r are related to motion change in video. The positive value means that the motion in video becomes faster while the negative value means that the motion in video becomes slower. Also, a larger value of $|a_h|$ or $|a_r|$ implies a larger motion change. Parameters ω_h, ω_r, and T work as weighting factors for controlling the tradeoff between temporal and spatial quality. As T decreases and ω_h and ω_r increase, the spatial quality is more emphasized than the motion smoothness, and vice versa.

B. Bit Allocation

After determining the proper frame rate and the selected frame location, we consider bit allocation among these frames.

TABLE VI
COMPARISON WITH TMN5 (SALESMAN). ENCODED FRAME NUMBER: TMN5 (25 FRAMES) AND PROPOSED ALGORITHM (23 FRAMES)

Rate control method	Average of PSNR	Standard deviation of PSNR
TMN 5	28.28	0.187
Proposed rate control(HOD)	28.64	0.081
Proposed rate control(Ratio)	28.61	0.065

TABLE VII
COMPARISON WITH TMN5 (AKIYO). ENCODED FRAME NUMBER: TMN5 (50 FRAMES) AND PROPOSED ALGORITHM (49 FRAMES)

Rate control method	Average PSNR	Standard deviation of PSNR
TMN 5	32.90	0.725
Proposed rate control	33.26	0.223

TABLE VIII
COMPARISON WITH TMN5 (FOREMAN). ENCODED FRAME NUMBER: TMN5 (25 FRAMES) AND PROPOSED ALGORITHM (22 FRAMES)

Rate control method	Average PSNR	Standard deviation of PSNR
TMN 5	29.49	0.373
Proposed rate control	29.50	0.262

The bit budget and image quality have to be examined simultaneously. By adopting the dependent frame coding scheme, the optimal frame-level bit allocation of a GOP based on (3) can be formulated as follows.

Determine $\vec{q} = (q_1, q_2, \ldots, q_{N_P})$ to minimize

$$D(\vec{q}) + \omega_q E(\vec{q}),$$
$$\text{subject to } \sum_{i=1}^{N_P} r_i(q_1, q_2, \ldots, q_i) \le B_{\text{gop}} \tag{8}$$

where

N_P encoded P-frame number;
B_{gop} total bit budget for a GOP;
ω_q weighting factor for abrupt quality change and flickering;

and

$$D(\vec{q}) = \frac{1}{N_P}\sum_{i=1}^{N_P} d_i(q_1, q_2, \ldots, q_i)$$
$$E(\vec{q}) = \frac{1}{N_P}\sum_{i=1}^{N_P}(d_i(q_1, q_2, \ldots, q_i) - d_{i-1}(q_1, q_2, \ldots, q_{i-1}))^2.$$

Ramchandran *et al.* [13] proposed a bit-allocation algorithm which determined a set of optimal quantization parameters for all encoded frames by the Viterbi algorithm for MPEG video. It primarily serves as a benchmark rather than a practical method due to the high computational complexity and long time delay involved. In the case of H.263 video, it is also impractical to get the optimal frame-level bit allocation for a GOP by using the Lagrange multiplier method because of the long length of GOP. Even though approximating R-D models can be used to reduce the computational complexity substantially, they are still not efficient enough to tackle the optimal frame-level bit-allocation problem for one GOP. To reduce the computational complexity and time delay, we simplify the optimization problem by employing the dependent frame coding scheme within all sub-GOPs and considering dependency among sub-GOPs. Then, we are led to the following simplified problem.

Determine $\vec{q_m}, m = 1, 2, \ldots, M$ to minimize

$$\sum_{m=1}^{M}(D_m(\vec{q_m}) + \omega_q E_m(\vec{q_m})),$$
$$\text{subject to } \sum_{m=1}^{M} r_m(\vec{q_m}) \le B_{\text{subgop}} M \tag{9}$$

where $\vec{q_m} = (q_{m,1}, q_{m,2}, \ldots, q_{m,N_m})$ is the quantization parameter vector for the mth sub-GOP, and

N_m encoded frame number of the mth sub-GOP;
$r_m(\vec{q_m})$ assigned number of bits for the mth sub-GOP;
M number of sub-GOPs in a GOP;
N_{subgop} itotal frame number of a sub-GOP;
N_{gop} total frame number of a GOP;

and

$$D_m(\vec{q_m}) = \frac{1}{N_m}\sum_{i=1}^{N_m} d_i(q_1, q_2, \ldots, q_i),$$
$$E_m(\vec{q_m}) = \frac{1}{N_m}\sum_{i=1}^{N_m}(d_i(q_1, q_2, \ldots, q_i) - d_{i-1}(q_1, q_2, \ldots, q_i, q_{i-1}))^2. \tag{10}$$

It is clear that we have the following relationship:

$$\sum_{m=1}^{M} N_m = N_P, \quad B_{\text{subgop}} = \frac{N_{\text{subgop}}}{N_{\text{gop}}} \cdot B_{\text{gop}}.$$

Note that $E(\vec{q})$ in (8) and $E_m(\vec{q_m})$ in (9) are inserted to reduce the flickering effect and to avoid abrupt quality change. By adjusting the weighting factor ω_q, the abrupt quality change and

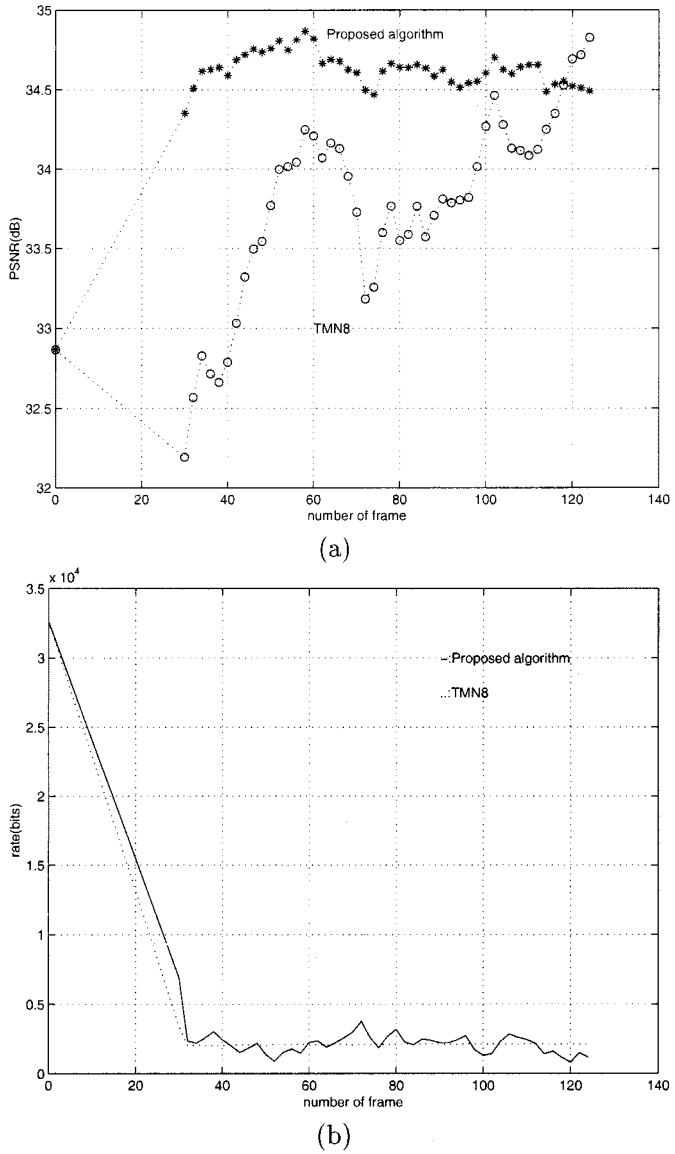

Fig. 4. PSNR and rate performance comparison between TMN8 and the proposed rate-control algorithm with the same frame skipping after I-frame coding for the Y component. (a) PSNR plot for Akiyo. (b) Rate plot for Akiyo.

the flickering effect are controllable. Now, the Lagrange multiplier method can be used to solve the optimization problem with constraints. The penalty function for the mth sub-GOP, which is derived from the bit budget constraint in (9), can be written as

$$P_m(\vec{q_m}) = \sum_{i=1}^{m} r_i(\vec{q_i}) - m \cdot B_{\text{subgop}}. \tag{11}$$

By combining (9) and (11), we can define a new penalty function for the mth sub-GOP as

$$\Phi_m(\vec{q_m}, \lambda_m) = J_m(\vec{q_m}) + \lambda_m \max\{0, P_m(\vec{q_m})\}, \quad \text{for } m = 1, 2, \ldots, M \tag{12}$$

where

$$J_m(\vec{q_m}) = D_m(\vec{q_m}) + \omega_q E_m(\vec{q_m}).$$

In our implementation, a gradient search method was used to find the optimal solution of (12). Since $\Phi_m(\vec{q_m}, \lambda_m)$ is a convex function [14], the optimal solution in each sub-GOP can be found easily. It is possible to insert an additional item into the penalty function to handle the buffer constraint [14], [13], [6].

The dependency of sub-GOPs is taken into account based on the following facts.

1) As shown in (11), the sum of bit rates for the 1st, 2nd , ... and mth sub-GOPs is used as the penalty function so that a GOP satisfies the constraint of the total bit budget more smoothly.
2) The dependency among sub-GOPs is considered by controlling weighting factors $\{\lambda_m, m = 1, 2, \ldots, M\}$ of the Lagrange multiplier method in (12).
3) d_0 in (10) is the error of the last encoded frame in the previous sub-GOP.

To compensate the dependency among sub-GOPs properly, weighting factors for sub-GOPs are set to satisfy the following equation:

$$\lambda_i \le \lambda_j, \quad \text{if } i \le j. \tag{13}$$

We argue that (13) is reasonable in this new context based on the monotonicity property given in [13]. By the assumption that each sub-GOP is independent, the propagation effect from the preceding sub-GOPs to the succeeding sub-GOPs is neglected. To compensate this effect, we require that all values of λ satisfy (13). Note that λ_i is related to buffer underflow, buffer overflow and rate fluctuation. Several adaptive control algorithms of λ were proposed in [3], [8], [18], [19].

Actually, we need a frame interpolation technique to achieve 30 fps. Three techniques are found in the literature [20]: the simple frame repetition (or intrafiltering), the frame averaging and the motion compensated frame interpolation [21]. In our implementation reported in Section IV, the simple frame repetition technique is adopted.

IV. Experimental Results

In this research, the conventional PSNR quality measure is reported for the comparison purpose. The difference of PSNR values of adjacent frames is used to measure quality change (or flickering). However, we also attempt to comment on the subjective quality evaluation of the coded video whenever it is appropriate.

Experimental results are reported in this section with four test image sequences. They are: "Salesman," "Akiyo," "Silent Voice," and "Foreman" of the CIF format. In the experiment, the average target bit rates are set to 32 kbits/s for Akiyo, Silent Voice, and Salesman, and 90 kbits/s for Foreman. Each I-frame is encoded by using the H.263 I-frame coding scheme with about 32–35 kbits for all cases. The proposed encoding frame-rate control algorithm is compared with TMN5 [22] and TMN8 [23]. The length of GOP (N) is 96 frames and the length of sub-GOP (M) is 12 frames in this experiment.

To estimate the supportable encoding frame-rate of the following sub-GOP, both HOD and the R-D ratio are calculated in the current sub-GOP. Based on these data, the encoding frame-rate for the next sub-GOP is predicted. The new frame

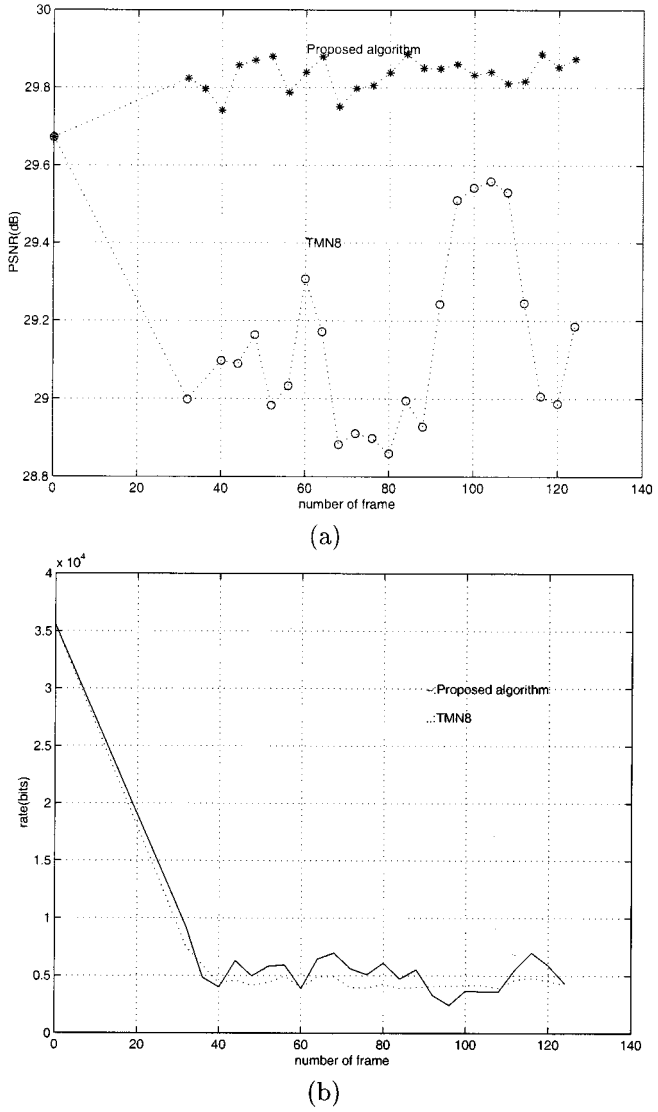

Fig. 5. PSNR and rate performance comparison between TMN8 and the proposed rate-control algorithm with the same frame skipping after I-frame coding for the Y component. (a) PSNR plot for Silent Voice. (b) Rate plot for Silent Voice.

rate is basically determined by the motion activity in the underlying video and the initial frame rate for the first sub-GOP. The threshold value TH_0 in (4) is set to 32, and the initial encoding frame-rates for all test videos are set to 3. It is observed that the two methods give almost the same results. Thus, we choose the HOD method for all sequences due to its simplicity. The encoding frame-rate results for Salesman with the moderate motion change among test videos are shown in Table III. It is observed that the encoding frame-rate for Foreman, which has a faster motion change than Salesman, fluctuates more frequently as shown in Table IV while the encoding frame-rate for Akiyo remains about the same since the associated motion is relatively slow and uniform. These results imply that the faster the motion change in a video clip, the more frequently the encoding frame-rates change. Parameter T in Section III-A is empirically adjusted to reduce the motion artifact.

Based on the above discussion, we expect a more dynamic encoding frame-rate change to be observed for video with a faster motion change such as movies and TV sport programs. Since digital movies and TV are most encoded with MPEG-2, it is out of the scope of this paper. Furthermore, if we choose 4 and 6 as the initial encoding frame-rates for the first sub-GOP of Salesman, then the encoded frame rates for sub-GOPs become $\{4,3,4,4,4,3,4,4\}$ and $\{6,4,6,6,6,4,6,6\}$, respectively. The same results are observed for all test videos. In these cases, the motion smoothness degradation is not observable by using the proposed encoding frame-rate-control scheme and the intra-filtering interpolation scheme with a subjective test.

Quantization parameters can be determined based on the computed encoding frame-rate with the encoding frame-rate-control scheme. The new cost function $\Phi_m(\vec{q_m}, \lambda_m)$ in (12) is minimized by adjusting quantization parameters for encoded frames in the sub-GOP while λ_m takes the dependency among the sub-GOPs into consideration. The gradient search algorithm is used for the solution of (12). To reduce the abrupt change in quality, ω_q in (9) is set to 2. In the experiment, we employ λ_m as shown in Table V.

By combining the proposed encoding frame-rate-control scheme and the frame-level bit-allocation scheme, we obtain the final rate-control results. These results are compared with those coded by TMN5 [22] and TMN8 [23] in Tables VI–VIII and Figs. 1–3. We have the following observations. As shown in Tables VI–VIII, the proposed rate-control scheme increases the averaged PSNR by 0.3–0.4 dB while it reduces the standard deviation of PSNR by 40%–70% for Salesman and Akiyo. For Foreman, the averaged PSNR is improved and the standard deviation of PSNR is reduced about 30%. Although the standard deviation of PSNR is not an exact measure of the flickering effect, it is fair to say that the flickering effect can be reduced by the smaller standard deviation of PSNR for the slow-moving *head & shoulder* video. In this experiment, we observed a significant flickering effect in the neck-tie area of Salesman coded by the TMN5 version of Telenor. For the Foreman sequence, the face is clearer, especially in the parts of eyes and mouth, and he blocking artifact of the background is reduced. For the Akiyo sequence, the subjective quality is not improved as obviously as that for Salesman and Foreman. However, with some attention, one can still see the face and the background more clearly. To conclude, we can reduce the flickering effect subjectively and objectively by using the proposed rate-control algorithm.

We also compare the proposed rate-control algorithm with TMN8. The frame skipping is required to reduce the time-delay caused by I-frame coding. The results are shown in Figs. 4–6. The proposed algorithm can keep video quality almost constant by increasing the output bit-stream rate fluctuation while TMN8 can only get almost constant bit rates by sacrificing video quality as shown in the figure. For Silent Voice, the 36th frame in Fig. 5(a) is skipped after the first P-frame coding. In this case, motion smoothness is seriously degraded. However, the proposed rate-control algorithm does not degrade the temporal quality. Actually, the generated output bit-stream rate attributes are determined by the given channel conditions and characteristics. The statistical data of Figs. 4–6 are summarized in Table IX. It is observed that the proposed algorithm can reduce the spatial quality change at the cost of the increased output bit-rate fluctuation.

TABLE IX
COMPARISON WITH TMN8 (STATISTICAL PSNR (dB) AND RATE (BITS) INFORMATION)

Sequence	Rate control	Avg PSNR	STD of PSNR	Avg Rate	STD Rate
Akiyo	TMN 8	33.78	0.589	2124	190
	Proposed alg.	34.63	0.103	2205	923
Silent Voice	TMN 8	29.14	0.224	4450	742
	Proposed alg.	29.83	0.040	5020	1484
Foreman	TMN 8	30.96	0.719	12015	862
	Proposed alg.	30.71	0.412	11957	4743

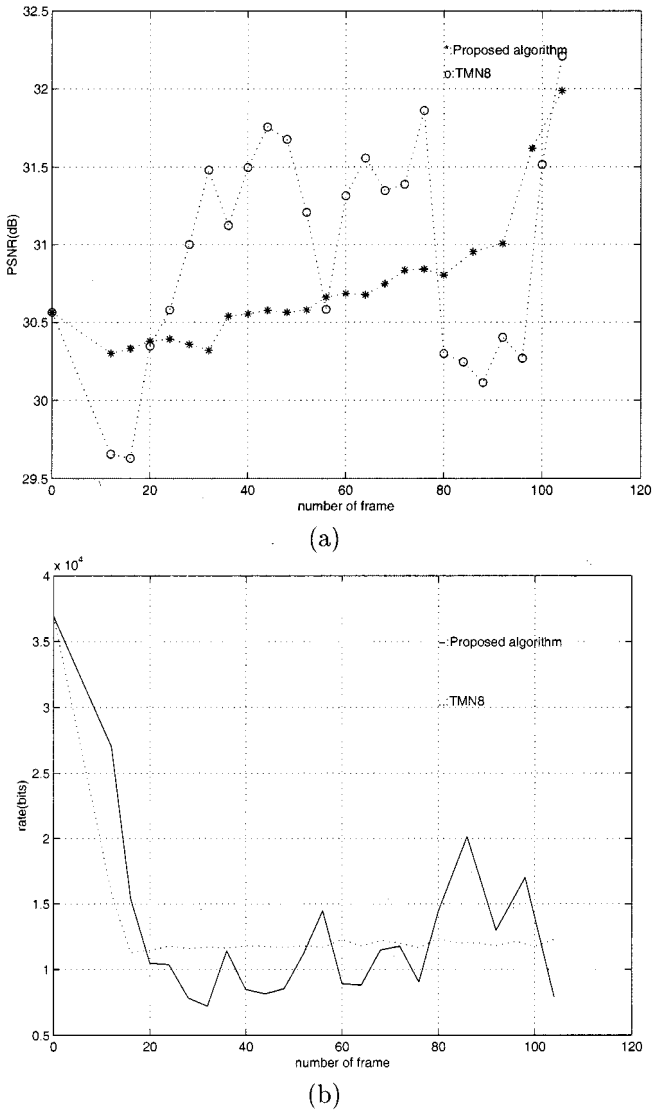

Fig. 6. PSNR and rate performance comparison between TMN8 and the proposed rate-control algorithm with the same frame skipping after I-frame coding for the Y component. (a) PSNR plot for Foreman. (b) Rate plot for Foreman.

V. CONCLUSIONS AND FUTURE WORK

A new rate-control algorithm, which consists of frame-rate control and bit allocation, was proposed in this work. There are several significant advantages with the new rate-control algorithm. Since the frame rate is treated as a control variable, we can determine a better tradeoff between the spatial and temporal quality to improve the overall perceptual quality at low bit rates. With the new frame-rate-control scheme, motion smoothness is better preserved. The abrupt frame skipping phenomenon, which degrades motion smoothness, can be reduced. Efficient frame-level bit allocation was also presented. With this scheme, we can determine a good balance between the computational complexity and the R-D performance with a tolerable time-delay constraint. It was demonstrated that the proposed frame-level bit-allocation algorithm can reduce the flickering effect subjectively and objectively in terms of the PSNR value.

There are still problems to be solved along this research direction to make the quality of low-bit-rate video higher. As mentioned before, the proposed rate-control algorithm requires 400-ms time delay. It is desirable to reduce time delay in real-time interactive video transmission applications. Even though some ingredients of the rate-control algorithm described in this work are expected to be applicable in low time-delay video coding as well, we have to examine a different tradeoff with the new constraint. Another interesting problem is to determine an efficient motion-compensated frame interpolation scheme at the decoder end to allow a constant frame rate display with higher quality reconstructed frames.

REFERENCES

[1] *Generic Coding of Moving Pictures and Associated Audio: (MPEG-2)*, MPEG (JTC1/SC29/WG11) and E. G. on ATM Video Coding (ITU-T SG15), Mar. 1994.
[2] *Video Coding for Low Bitrate Communication*, ITU-T Recommendation H.263 Version 2, Jan. 1998.
[3] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 182–190, Apr. 1996.
[4] D. Mukherjee and S. K. Mitra, "Combined mode selection and macroblock step adaptation for H.263 video encoder," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 37–40, Oct. 1997.
[5] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay video communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 172–185, Feb. 1999.
[6] A. R. Reibman and A. W. Berger, "Traffic descriptors for VBR video teleconferencing," *IEEE Trans. Networking*, vol. 3, pp. 329–339, June 1995.
[7] C. Y. Hsu, A. Ortega, and A. R. Reibman, "Joint selection of source and channel rate for VBR video transmission under ATM policing constraints," *IEEE J. Select. Areas Commun.*, vol. 15, Aug. 1997.
[8] H. Song, J. Kim, and C. C. J. Kuo, "Real-time encoding frame rate control for H.263+ video over the Internet," *Signal Processing: Image Commun.*, vol. 15, no. 1–2, pp. 127–148, 1999.
[9] 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," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 41–44, Oct. 1997.
[10] K. T. Ng, S. C. Chan, and T. S. Ng, "Buffer control algorithm for low bit rate video compression," *Proc. IEEE Int. Conf. Image Processing*, Sept. 1996.

[11] K. Oehler and J. L. Webb, "Macroblock quantizer selection for H.263 video coding," *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 365–368, Oct. 1997.
[12] G. Schuster and A. Katsaggelos, "Fast and efficient mode and quantizer selection in the rate and distortion sense for H.263," in *Proc. SPIE Visual Communication and Image Processing Conf.*, 1996.
[13] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with application to multiresolution and MPEG video coder," *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sept. 1994.
[14] L. J. Lin, A. Ortega, and C. C. J. Kuo, "Rate control using spline interpolated R-D characteristics," in *Proc. SPIE Visual Communication and Image Processing Conf.*, Mar. 1996, pp. 111–122.
[15] 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, Sept. 1988.
[16] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation," *IEEE Trans. Image Processing*, vol. 3, pp. 26–40, Jan. 1994.
[17] J. Lee and B. W. Dickinson, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans. Image Processing*, vol. 3, pp. 513–526, Sept. 1994.
[18] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled multiplier method," *IEEE Trans. Image Processing*, vol. 3, pp. 546–558, Sept. 1994.
[19] J. J. Chen and D. W. Lin, "Optimal bit allocation for coding of video signals over ATM networks," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1002–1015, Aug. 1997.
[20] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[21] T. Kuo and C.-C. J. Kuo, "Motion-compensated interpolation for low-bit-rate video quality enhancement," in *Proc. SPIE Int. Symp. Optical Science, Engineering and Instrumentation*, May 1998.
[22] T. Research, *TMN (H.263) encoder/decoder, version 2.0, tmn (h.263) codec*, June 1996.
[23] "H.263+ Encoder/Decoder," Image Processing Lab, Univ. British Columbia, Canada, Feb. 1998. TMN(H.263) codec.

Hwangjun Song received the B.S. and M.S. degrees from the Department of Control and Instrumentation (EE), Seoul National University, Seoul, Korea, in 1990 and 1992, respectively, and the Ph.D. degree in electrical engineering systems from the University of Southern California, Los Angeles, in 1999.

During 1992, he was a Research Engineer with LG Industrial Laboratories, Korea. From 1995 to 1999, he was a Research Assistant with the Signal and Image Processing Institute and the Integrated Media Systems Center, University of Southern California at Los Angeles. He was with Sejong University, Seoul, Korea, during the spring semester 2000. Since the full semester 2000, he has been with Hongik University, Seoul, Korea. His research interests include multimedia signal processing and communication, image/video compression, digital signal processing, and network protocols necessary to implement a functional.

C.-C. Jay Kuo (S'86–M'87–SM'92–F'99) received the B.S. degree from the National Taiwan University, Taipei, in 1980, and the M.S. and Ph.D. degrees from Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He was a Computational and Applied Mathematics (CAM) Research Assistant Professor in the Department of Mathematics, University of California, Los Angeles, from October 1987 to December 1988. Since January 1989, he has been with the Department of Electrical Engineering Systems and the Signal and Image Processing Institute, University of Southern California, Los Angeles, where he currently has a joint appointment as Professor of Electrical Engineering and Mathematics. His research interests are in the areas of digital signal and image processing, audio and video coding, wavelet theory and applications, multimedia technologies, and Internet and wireless communications. He has authored more than 380 technical publications in international conferences and journals.

Dr. Kuo is a member of SIAM and ACM, and a Fellow of SPIE. He is the Editor-in-Chief for the *Journal of Visual Communication and Image Representation*, and served as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING during 1995–1998 and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1995–1997. He received the National Science Foundation Young Investigator Award and Presidential Faculty Fellow Award in 1992 and 1993, respectively.