# Improved H.263+ Rate Control via Variable Frame Rate Adjustment and Hybrid I-frame Coding

정회원 송 황 준*

Hwangjun Song* ***Regular Member***

ABSTRACT

A novel rate control algorithm consisting of two major components, i.e. a variable encoding frame rate method and a hybrid DCT/wavelet I-frame coding scheme, 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. Furthermore, an R-D optimized hybrid DCT/wavelet scheme is used for effective I-frame coding. 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 variable frame rate method and the hybrid I-frame coding scheme are compatible with the bitstream structure of H.263+.

## I. INTRODUCTION

Digital video coding techniques have advanced rapidly in recent years. International standards such as MPEG-1, MPEG-2, H.261 and H.263 have been established to accommodate different application needs. New standards such as H.263+/++, H.26L, MPEG-4 and MPEG-7 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. 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 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 and H.263+, i.e. intra (I), predictive (P), and bidirectional (B) frames, I-frames are required to control the accumulated mismatch error. It is clear that the coding error of I-frames can be propagated to subsequent predictive frames such as P-frames 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

* Dept. of Software Engineering, Sejong University, Seoul, Korea. Email: hwangjun@sejong.ac.kr
논문번호 : 99339-0823, 접수일자 : 1999년 8월 23일

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 there has been little research work about the frame level bit allocation for H.263+. Almost existing H.263 rate control algorithms focus on macroblock level bit allocation. However, we cannot treat the temporal quality of video efficiently with only macroblock level bit allocation.

We examine the rate control algorithm for H.263+ with very long predictive 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. The hybrid DCT/wavelet I-frame coding scheme, which is more efficient at low bit rates, is also studied. Usually, I-frame and P-frames are considered separately for the rate control problem of the H.263 family. However, given the important role of I-frames, we feel that it is essential to include the I-frame coding scheme as an integral part of the solution. The variable frame rate method and the hybrid I-frame coding scheme are compatible with the bitstream structure of H.263+.

## Ⅱ. PROBLEM FORMULATION

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. 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 $q_i$ $i=1,2,3,\cdots,N-1$ to minimize

$$d_I + \sum_{i=1}^{N-1} d_i(q_1, q_2, \cdots, q_i),$$

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

where $d_I$ and $r_I$ are the distortion measure and the bit rate for the I-frame, $d_i(q_1, q_2, \cdots, q_i)$ and $r_i(q_1, q_2, \cdots, q_i)$ are the distortion measure and the allocated bit rate for the $i$th frame, respectively, and $B$ is the given bit budget for a GOP. We address two parts of the rate control problem in the following two sections. First, we develop a frame rate control and bit allocation algorithm that preserves the quality of P frames to be constant as long as possible. To reduce the computational complexity and encoding time-delay, GOP is divided into several sub-GOPs. Second, a hybrid DCT/wavelet I-frame coding scheme is proposed, which is used to reduce the distortion $d_I$ of the I-frame with the given bit budget $r_I$

## Ⅲ. PROPOSED RATE CONTROL

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 [1] 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 rate-distortion characteristics of predictive frames are related to the amount of motion involved, we will predict the rate-distortion relation roughly based on the motion information existing in video, and develop an encoding frame rate control scheme accordingly in Section 3.1. In this algorithm, each sub-GOP consists of 12 frames.

### 3.1 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 frames per second. 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 1 and 2. Usually, 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 1 is chosen for the current sub-GOP, then the 1/6 frame rate in Table 2 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 3 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 rate-distortion (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 [2]. 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.

Table 1. 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 2. 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 |

Histogram-based methods [3] 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 [3] such as BV (block variance difference) and MCE (motion compensation error) 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.

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

$$D_h(f_n, f_m) = \frac{\sum_{i>\text{left}|TH_0\text{right}|} hod(i)}{N_{pixel}}, \quad (1)$$

where $i$ be the index of the quantization bin, $hod(i)$ is the histogram of the difference image, $TH_0$ is the threshold value for detecting the closeness of the position to zero, and $N_{pixel}$ is the number of pixels. Once all HOD values for consecutive frames in the current sub-GOP are calculated, the estimated HOD value $\widehat{D_e}$ for the following sub-GOP can be computed by

$$\widehat{D_e} = D_h + \omega_h a_h, \quad (2)$$

where $D_h$ is the HOD value between the two last encoded frames in the current sub-GOP, $\omega_h$ is a weighting factor and $a_h$ is the 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 (2)

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

$$D_r = \frac{r}{d_{psnr}}, \quad (3)$$

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:

$$\widehat{D_e} = D_r + \omega_r a_r, \quad (4)$$

where $D_r$ is the rate-distortion ratio value of the last two encoded frames in the current sub-GOP, $\omega_r$ is a weighting factor, and a_r$ is the 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 $\widehat{D_e}$ and $m(D_h)$. That is,

- *If* $\delta \geq T$, the encoding frame rate is decreased by one level.
- *If* $\delta \leq -T$, the encoding frame rate is increased by one level.
- *If* $|\delta| < T$, the encoding frame rate remains the same.

where $\delta = \widehat{D_e} - m(D_h)$ and the threshold value $T$ is chosen to be the averaged *HOD* over the current 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 $\omega_h$, $\omega_r$ and $T$

work as weighting factors for controlling the tradeoff between temporal and spatial quality. As $T$ decreases and $\omega_h$ and $\omega_r$ increase, the spatial quality is more emphasized than the motion smoothness, vice versa.

### 3.2 Bit Allocation

After determining the proper frame rate and the selected frame location, we consider bit allocation among these frames. 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 can be formulated as follows.

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

$$D(\vec{q}) + \omega_q E(\vec{q}),$$

$$\text{subject to } \sum_{i=1}^{N_P} r_i(q_1, q_2, \ldots, q_i) \le B_{gop}, \tag{5}$$

where $N_P$ is the encoded P-frame number and $B_{gop}$ is total bit budget for a GOP, $\omega_q$ is the 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.* [1] 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

$$\sum_{m=1}^{M} (D_m(\vec{q_m}) + \omega_q E_m(\vec{q_m})),$$

$$\text{subject to } \sum_{m=1}^{M} r_m(\vec{pm}) \le B_{subgop} M, \tag{6}$$

where $\vec{q_m} = (q_{m,1}, q_{m,2}, \ldots, q_{m,N_m})$ is the quantization parameter vector for the $m$th sub-GOP, $N_m$ is the encoded frame number of the $m$th sub-GOP, $r_m(\vec{q_m})$ is the assigned number of bits for the $m$th sub-GOP, $M$ is the number of sub-GOPs in a GOP, $N_{subgop}$ is the total frame number of a sub-GOP and $N_{gop}$ is the total frame number of a GOP, and

$$D_m(\vec{q_m}) = \frac{1}{N_m} \sum_{i=1}^{N_m} d_i(q_1, q_2, \ldots, q_i),$$

$$E_m(\vec{q_m}) =$$

$$\frac{1}{N_m} \sum_{i=1}^{N_m} (d_i(q_1, q_2, \ldots, q_i) - d_{i-1}(q_1, q_2, \ldots, q_i, q_{i-1}))^2. \tag{7}$$

It is clear that we have the following relationship

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

Note that $E(\vec{q})$ in (5) and $E_m(\vec{q_m})$ in (6) are inserted to reduce the flickering effect and to avoid abrupt quality change. By adjusting the weighting factor $\omega_q$, abrupt quality change and 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 $m$th sub-GOP, which is derived from the bit budget constraint in (6), can be written as

$$P_m(\vec{q_m}) = \min\{\sum_{i=1}^{m} r_i(\vec{q_i}) - m \cdot B_{subgop}, 0\}, \tag{8}$$

By combining (6) and (8), we can define a new penalty function for the $m$th sub-GOP as

$$\Phi_{m}(\overrightarrow{q_m}, \lambda_m) = J_m(\overrightarrow{q_m}) + \lambda_m \max\{0, P_{m(\text{vbc})}\}, \quad \text{for } m = 1, 2, \ldots, M, \tag{9}$$

where

$$J_{m}(\overrightarrow{q_m}) = D_{m}(\overrightarrow{q_m}) + \omega_q E_m(\overrightarrow{q_m}).$$

In our implementation, a gradient search method was used to find the optimal solution of (9). Since $\Phi_m(\overrightarrow{q_m}, \lambda_m)$ is a convex function [2], 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[2], [1], [4].

The following facts are taken into account to improve the video quality and satisfy the bit constraint.

- As shown in (8), the sum of bit rates for the 1st, 2nd, ··· and $m$th sub-GOPs is used as the penalty function so that a GOP satisfies the constraint of the total bit budget more smoothly.
- $d_0$ in (7) is the error of the last encoded frame in the previous sub-GOP.
- We adapt the Lagrange multiplier for each sub-GOP to meet the given total bit budget as follows.

$$\lambda_{m+1} = \lambda_m + \Delta\lambda_m, \tag{10}$$

$$\Delta\lambda_m = \frac{\sum_{i=1}^{m} r_i(\overrightarrow{q_i})}{m \cdot B_{\text{subgop}}} - 1 \tag{11}$$

Several adaptive control algorithms of $\lambda$ were proposed in [5], [6], [7], [8].

The maximum required computational complexity for (5) and (6) are $31^N$ and $\sum_{i=1}^{M} 31^{N_i}$ respectively where $N = \sum_{i=1}^{M} N_i$ and $N_i = \dim(\overrightarrow{q_i})$.

## Ⅳ. PROPOSED H.263+ I-FRAME CODING SCHEME

Generally speaking, there are two major approaches to I-frame (or still image) coding based on the block DCT and the wavelet transforms, respectively. JPEG is a popular image coding standard using the block DCT transform. Recently, coders based on the multi-resolution wavelet transform such as EZW [9], LZC [10] and SPIHT [11] were proposed. These schemes provide a better performance over the DCT-based schemes. However, it is difficult to apply the above schemes directly to the I-frame coding of H.263+, since they are totally different from DCT-based schemes.

Besides the multiresolution wavelet approach, model-based UTQ (uniform threshold quantizer) is another popular approach. UTQ is employed for subband signals since it approximates the optimum entropy-constrained scalar quantizer (ECSQ) for the generalized Gaussian distribution without complicated ECSQ. The characteristics of UTQ is fully determined by the step size and the deadzone size. A fast wavelet-based compression technique with model-based UTQ was proposed by LoPresto *et. al.* [12] for still image compression. In his work, each wavelet coefficient is treated as a random variable with the Laplacian distribution. The standard deviation of the wavelet coefficient to be encoded is estimated by using wavelet coefficients coded earlier. Then, the optimal UTQ for each wavelet coefficient is determined based on the estimated standard deviation. This scheme reduces the side information for quantizers. However, iterative calculation is required for the estimation of the standard deviation of each wavelet coefficient. Thus, it is not efficient for I-frame coding.

Silva and Ghanbari [13] used a hybrid DCT/wavelet approach to encode a high resolution still image contained by video. Even though their scheme satisfies the bit stream structure of H.261, it is not efficient to use block DCT for high frequency subbands. Various wavelet based I-frame coding schemes were proposed to replace the DCT based I-frame coding in MPEG-4. A hybrid DCT/wavelet transform [14] was also

proposed in MPEG-4. However, these wavelet coders are not compatible with the H.263+ bitstream syntax.

An efficient I-frame coding scheme at low bit rates is studied in this section. More specifically, we examine a hybrid DCT/wavelet I-frame coding scheme with UTQ, which provides an enhanced performance over a pure DCT-based H.263+ codec. The compressed bit stream of our proposed scheme is still compatible with the syntax of H.263+.

### 4.1 Coding Algorithm

The block diagram of the proposed hybrid DCT /wavelet coding scheme is depicted in Figure 3. Each I-frame is decomposed into four subbands. The DCT-based H.263+ I-frame coding can still be applied to the LL subband without loss of efficiency. However, images in the three high frequency subbands look like ghost images since only high frequency residual signals exist. We adopt a hybrid DCT/wavelet transform, and apply a fast R-D optimized bit allocation scheme to each subband. The high frequency subbands information can be transmitted with the supplementary option of H.263+. Since the I frame is only decomposed to four subbands, this approach can also be called the hybrid DCT/subband coding scheme. We adopt the term "wavelet" here, since there is a trend to use wavelet coding to contain all subband coding schemes.

The image in the LL subband is very similar to the original I-frame image with a quarter size of the original one. We can apply the conventional block DCT transform, which is the standard I-frame coding scheme of H.263+, without loss of efficiency. In H.263+, the I-frame is divided into GOBs, each GOB is divided into macroblocks (16 × 16) and each macroblock is divided into blocks (8 × 8). For each block, the 8 × 8 DCT transform is applied. Unless the optional advanced intra coding mode or the modified quantization mode is used, the same quantizer is used for all coefficients within a macroblock except the first one of INTRA blocks. The quantization step size $q_L$ ranges from 1 to 31. The DCT-transformed and quantized coefficients are entropy encoded. To apply H.263+ I-frame coding scheme to the LL subband, we quantize the wavelet coefficients in the LL subband and convert them into a range between 0 and 255. We do not need the side information for the LL subband quantizer, except for the quantization step size. This is justifiable due to the lowpass filtering characteristics of the wavelet transform.

The uniform threshold quantizer (UTQ) and the adaptive arithmetic coder are used for the coding of wavelet coefficients in the three high frequency subbands. UTQ is similar to the scalar quantizer except that the quantization step size in a specific region called the deadzone can be different. It is employed here since it approximates the optimum entropy-constrained scalar quantizer (ECSQ) for the generalized Gaussian distribution without computing optimum ECSQ explicitly. In this work, we choose a step size of $\Delta$ and the interval [-1.5 $\Delta$, 1.5 $\Delta$] as the deadzone. Reconstruction levels are selected by using the centroid condition under the Laplacian distribution assumption. Let us consider UTQ with $2L+1$ quantization levels, i.e. one for the deadzone, $L$ above the deadzone and $L$ below the deadzone. Under the zero-mean and symmetric distribution assumption, it is clear that the reconstruction level for the deadzone is zero (i.e. $q_{0=0}$). After some calculation, reconstruction level $q_i$ for $i$=1,2, ···, $L$ can be obtained as

$$q_i(\Delta)=\frac{1}{2p_i}(0.5+i)\Delta e^{-\lambda(0.5+i)\Delta} - \frac{1}{2p_i}(1.5+i)\Delta e^{-\lambda(1.5+i)\Delta}+\frac{1}{\lambda}, \qquad (12)$$

where $p_i$ denotes the probability of being in the $i$th quantization bin and $\lambda=\frac{\sqrt{2}}{\sigma}$, and where $\sigma$ is the standard deviation as estimated in (15) and $q_{-i}=-q_i$. For the three high frequency subbands, only the luminance components (Y) of the wavelet coefficients are encoded. If the step size information $\Delta$ is available, we can reconstruct the quantized original data.

### 4.2 Subband Bit Allocation

In order to perform the optimal bit allocation, we have to study the rate-distortion models for the LL and high frequency subbands. The rate-distortion analysis of the LL subband can be performed based on two models parameterized by the quantization step size $q_L$, i.e. one for the rate and the other for the distortion. We adopt the quadratic rate model, which is the extended version of the one used in the MPEG-2 test R-D model. The quadratic rate model can be written as

$$R_L(q_L) = aq_L^{-1} + bq_L^{-2}. \tag{13}$$

A linear regression method is used to determine coefficients $a$ and $b$ [15]. The distortion measure is represented as the averaged quantization scale of a frame to reduce the computational complexity in [15]. Here, we adopt the affine distortion model *w.r.t.* quantization parameter $q_L$, i.e.

$$D_L(q_L) = cq_L + d, \qquad 1 \le q_L \le 31, \tag{14}$$

where coefficients $c$ and $d$ are determined by minimizing MSE between observed and estimated data.

For the three high frequency subbands, it is often assumed that wavelet coefficients are in the generalized Gaussian distribution of the form

$$p(x) = \frac{\nu\eta(\nu,\sigma)}{\Gamma(1/\nu)} \exp\{-(\eta(\nu,\sigma)\,|\,x\,|\,)^{\nu}\},$$

where $\sigma$ is the standard deviation, $\nu$ is a shape parameter,

$$\eta(\nu,\sigma) = \sigma^{-1}\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}, \text{ and } \Gamma(z) = \int_0^{\infty} \exp(-t)t^{z-1}dt.$$

For $\nu = 1$, it becomes the well known Laplacian distribution. In the following derivation, we simply consider the Laplacian distribution, which yields a reasonably good result. Note that the Laplacian density function is determined completely by its standard deviation $\sigma$. A practical variance estimation algorithm was considered in [16]. That is, it can be computed as

$$\sigma^2 = \frac{1}{N_{pixel}} \sum_{i=1}^{N_{pixel}} c_i^2. \tag{15}$$

where $c_i$ is wavelet coefficient and $N_{pixel}$ is the number of wavelet coefficients in a subband.

For a Laplacian distributed source, we can calculate the rate and distortion in terms of quantization parameter $\Delta$ based on the estimated variances of high frequency subbands. Even though the computational complexity required by this approach is low, the accuracy of the resulting model is not good enough. To get a better model, we use the affine distortion model and the quadratic rate model parameterized by $\Delta$ as the case in the LL subband. That is, we have

$$\begin{aligned} R_H(\Delta) = {} & p_0(\Delta)\log(p_0(\Delta)) \\ & + 2\sum_{i=1}^{L} p_i(\Delta)\log(p_i(\Delta)) \end{aligned} \tag{16}$$

where $p_i$ denotes the probability of being in the $i$th quantization bin. Equation (16) is used to find a sufficient number of control points in the bit rate range of our interest and then use them to determine model coefficients $a$, $b$ $c$ and $d$ in terms of $\Delta$. By combining the rate and the distortion models, we can derive the R-D curve for high frequency subbands as plotted in Figure 2. Based on this R-D model, we can exploit existing fast algorithms for bit allocation.

For bit allocation, we consider an iterative procedure to determine the optimal bit budget allocated to four subbands denoted by $R_1$, $R_2$, $R_3$ and $R_4$. The basic idea of the iterative optimization algorithm is that when the optimal bit allocation is achieved, the four subband should operate with the same slope, which corresponds to the same Lagrangian multiplier, at their respective R-D curves. A fast bit allocation algorithm is described below.

1. Choose the initial Lagrange multiplier parameter $\lambda$ and apply it to all four subbands.
2. Based on models specified in (13) and (14), we find the quantization step size $q_1$ (or $q_L$)

for the LL subband that satisfies

$$\min_{q_1} D_1(q_1) + \lambda R_1(q_1).$$

Also, based on models specified in (14), (13) and (16), we can determine $\Delta_i$ that yields

$$\min_{\Delta_i} D_i + \lambda R_i, \quad \text{for} \quad i = 2, 3, 4.$$

3. Let R denote the given total bit budget for the whole I frame. If

$$\sum_{i=1}^{4} R_i - R > \varepsilon_1$$

we go back to Step 2 with a larger value of $\lambda$. If

$$R - \sum_{i=1}^{4} R_i > \varepsilon_2,$$

we go back to Step 2 with a smaller value of $\lambda$. Thresholds $\varepsilon_1$ and $\varepsilon_2$ are two small positive numbers. If the desired bit rate $R$ and the actual bit rate are close enough, we can proceed to the next step.

4. In Steps 2 and 3, all quantization steps $q_1$ and $\Delta_i$ are adjustable. In this step, we consider to finetune the bit rate by modifying $\Delta_i$ only. Let $q_1^*$ be the quantization parameter in the LL subband determined in the 3rd step. If

$$R - \left( R_1(q_1^*) + \sum_{i=2}^{4} R_i \right) > \varepsilon_3,$$

we repeat Step 4 with a smaller value of $\lambda$. If

$$\left( R_1(q_1^*) + \sum_{i=2}^{4} R_i \right) - R > \varepsilon_4,$$

we repeat Step 4 with a larger value of $\lambda$. If the desired bit budget and the sum of allocated bit budgets are close, the bit allocation task is completed.

After bit allocation, we can perform the quantization operation in each subband with the selected quantization step sizes $q_1$, $\Delta_2$, $\Delta_3$ and $\Delta_4$.

## V. Experimental Results

In this research, the conventional PSNR (peak signal-to-noise ratio) 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 target bit rates are set to 32 kbps for Akiyo, Silent Voice and Salesman, and 90 kbps for Foreman. Each I-frame is encoded with about 32-35 kbits for all cases. The frame rate control algorithm is compared with TMN5 [17] and TMN8 [18] in Section 5.1 while the hybrid I-frame coding scheme is compared with H.263+ I-frame coding scheme in Section 5.2. The length of GOP ($N$) is 96 frames and the length of sub-GOP ($M$) is 12 frames in this experiment.

### 5.1 Frame Rate Control and Bit Allocation

To estimate the supportable encoding frame rate of the following sub-GOP, both HOD and the rate-distortion ratio are calculated in the current sub-GOP. Based on these data, the encoding frame rate for the next sub-GOP is predicted. The encoding frame rate results for Salesman are shown in Table 3 and 4. The two methods give almost the same results. This observation is typical. Thus, we choose the HOD method for all sequences due to its simplicity.

The encoding frame rate for Akiyo remains about the same since the associated motion is relatively uniform and slow. The result of Foreman is more interesting as shown in Table 5. The threshold value $TH_0$ in (1) is set to 32. 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

Table 3. Calculation of $\widehat{D_e}$ 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 | $m(D_h)$ | $\widehat{D_e}$ | No. of encoded frames |
|---|---|---|---|---|---|
| 1 | 0.006290 | 0.061553 | 0.039358 | 0.0804 | 3 (initial value) |
| 2 | -0.006145 | 0.026407 | 0.044843 | 0.0080 | 2 |
| 3 | 0.000492 | 0.018742 | 0.018255 | 0.0202 | 3 |
| 4 | 0.000300 | 0.021100 | 0.022385 | 0.0220 | 3 |
| 5 | 0.004858 | 0.054983 | 0.036632 | 0.0696 | 3 |
| 6 | -0.006279 | 0.018012 | 0.066848 | -0.0008 | 2 |
| 7 | 0.000631 | 0.008967 | 0.005337 | 0.0109 | 3 |
| 8 | | | | | |

Table 4. Calculation of $\widehat{D_e}$ based on HOD ( $\omega_r$=2) and the resulting frame rate for each sub-GOP for Salesman when $T$=5

| No. of sub-GOP | Slope of approx. line | Last coeff. | $m(D_r)$ | $\widehat{D_e}$ | No. of encoded frames |
|---|---|---|---|---|---|
| 1 | 1.980963 | 26.4220 | 16.8983 | 30.3839 | 3 (initial value) |
| 2 | -3.719367 | 11.9224 | 23.0805 | 4.4837 | 2 |
| 3 | 0.134012 | 10.9854 | 10.7915 | 11.2534 | 3 |
| 4 | 0.487675 | 13.5254 | 11.6786 | 14.5008 | 3 |
| 5 | 1.817487 | 22.0173 | 13.4961 | 25.6523 | 3 |
| 6 | -1.109783 | 12.9331 | 16.2625 | 10.7135 | 2 |
| 7 | 0.017300 | 9.14930 | 9.09740 | 9.1837 | 3 |
| 8 | | | | | |

Table 5. The resulting variable 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 | 3 |
| 7 | 3 |
| 8 | 2 |

cost function $\Phi_m(\overrightarrow{q_m}, \lambda_m)$ in (9) is minimized by adjusting quantization parameters for encoded frames in the sub-GOP while $\lambda_m$ takes the dependency among the sub-GOPs into consideration. The gradient search algorithm is used for the solution of (9). To reduce the abrupt change in quality, $\omega_q$ in (6) is set to 2.

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 [17] and TMN8 [18] in Tables 6~8 and Fig. 7. We have the following observations. As shown in Tables 6~8, 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, specially in the parts of eyes and mouth. The 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 in two different situations. First, the same frame skipping as TMN8 is used after I-frame coding. The frame skipping is required to reduce the time-delay caused by the I-frame. Second, the proposed rate control algorithm without frame skipping after the I-frame is compared with TMN8. The second case may include the video playback at low bit rates. The results are shown in Fig. 8. The proposed algorithm can keep video quality almost constant by increasing the output bitstream 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. 8 (c) 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 bitstream rate attributes are determined by the given channel conditions and characteristics. The statistical data of Fig. 8 are summarized in Table 9. It is observed that the proposed algorithm can reduce the spatial quality change at the cost of the increased output bit rate fluctuation.

Table 6. 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 | 28.2796 | 0.1870 |
| proposed rate control(HOD, TMN5 I) | 28.6410 | 0.0813 |
| proposed rate control(Ratio, TMN5 I) | 28.6103 | 0.0649 |
| proposed rate control & hybrid I | 28.7653 | 0.1832 |

Table 7. 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 | 32.8986 | 0.7249 |
| proposed rate control(TMN5 I) | 33.2571 | 0.2229 |
| proposed rate control & hybrid I | 33.5880 | 0.1250 |

Table 8 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 | 29.4862 | 0.3732 |
| proposed rate control(TMN5 I) | 29.5028 | 0.2624 |
| proposed rate control & hybrid I | 31.7480 | 0.6521 |

Table 9. 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 | 33.7822 | 0.5893 | 2.124k | 189.9 |
| | proposed alg. | 34.6265 | 0.1028 | 2.205k | 922.9 |
| Silent | TMN 8 | 29.1359 | 0.2240 | 4.450k | 742.0 |
| | proposed alg. | 29.8345 | 0.0400 | 5.020k | 1483.9 |
| Foreman | TMN 8 | 30.9605 | 0.7188 | 12.015k | 861.8 |
| | proposed alg. | 30.7136 | 0.4119 | 11.957k | 4743.0 |

### 5.2 Hybrid I-frame Coding

In this experiment, the biorthogonal 9-7 tap spline filter is used for the wavelet transform. Control points are required to model the R-D curves of subbands. For the LL subband, {3,5,8,13,21,31} is used for $q_1$. The set of points { $1.3\sigma$, $1.6\sigma$, $2\sigma$, $2.5\sigma$, $3\sigma$} is employed as control points for high frequency subbands because we are interested in very low bit rates in high frequency subbands. These control points are determined by (16). The PSNR plots of I-frames of CIF Salesman, Akiyo and Foreman are shown in Figs. 4, 5 and 6. In (a),(c) and (e) of Figs. 7 and 8, we see that PSNR has improved greatly at low bit rates. In addition, we can see that the PSNR plots of the following P-frame has increased in (b), (d) and (f) when the I frame is encoded at about 32 kbits. Since the $i$th P-frame is used as a reference frame for the $(i+1)$th P-frame, the proposed hybrid I-frame coding scheme improves all frames of a GOP. The proposed hybrid I-frame coding scheme improves the quality in terms of the objective measure (PSNR). The blocking artifact is very obvious at low bit rate in the DCT block-based coding scheme. The proposed coding scheme reduces the blocking artifact significantly. Even though the PSNR values of two schemes are almost the same, the subjective quality of the proposed scheme is better than DCT block-based scheme. When the hybrid DCT/wavelet I-frame coding scheme is combined with the proposed rate control algorithm, the results are shown in (b), (d) and (e) of Fig. 7 and Tables 6-8. Actually, we use the TMN5 I-frame coding scheme for the LL subband coding for the new hybrid I-frame coding. It is observed that the averaged PSNR of test sequences is improved by about 0.5-1.3 dB.

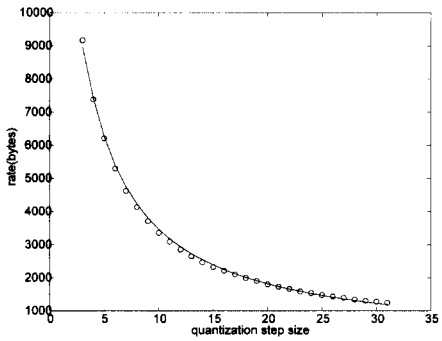
## Ⅵ. Conclusions and Future Work

A new rate control algorithm, which consists of frame rate control and bit allocation and hybrid DCT/wavelet I-frame coding, 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 trade-off 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 rate-distortion 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. Given the important role of the I frame at low bit rates, effective I-frame coding is essential. The proposed hybrid DCT/wavelet I-frame coding scheme can reduce the blocking artifact to enhance both objective and subjective quality substantially. However, a post-processing task is required at the decoder to implement the hybrid I-frame coding scheme.

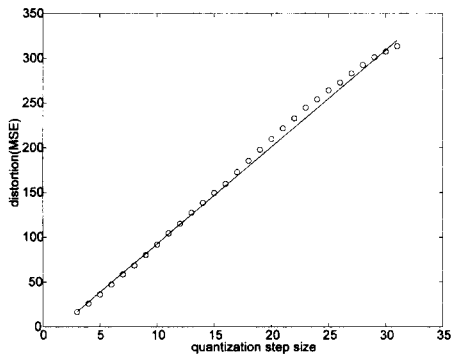
There are still problems to be solved along this research direction to make the quality of low bit rate video higher. As mentioned before, the proposed rate control algorithm requires $400ms$ time delay. It is desirable to reduce time delay in real-time interactive video transmission applications. Even 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 trade-off 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.

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 $400ms$ 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 trade-off 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.

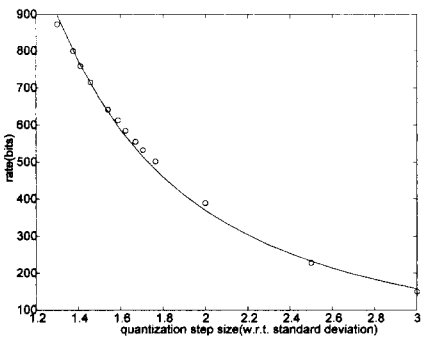



(a)

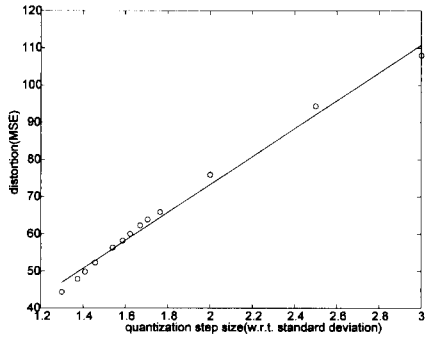



(b)

Figure 1. (a) The rate and (b) the distortion models parameterized by the quantization step size for the LL subband for the Salesman sequence.




(a)




(d)

Figure 2. (a) The rate and (b) the distortion models parameterized by the quantization step size for the LH subband for the Salesman sequence.

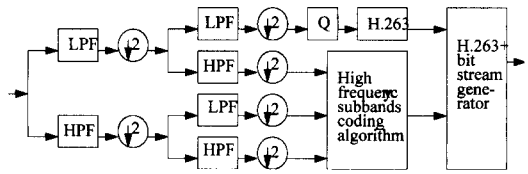
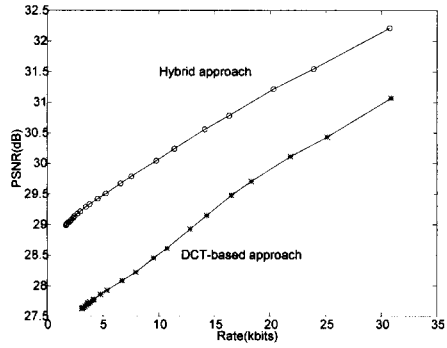
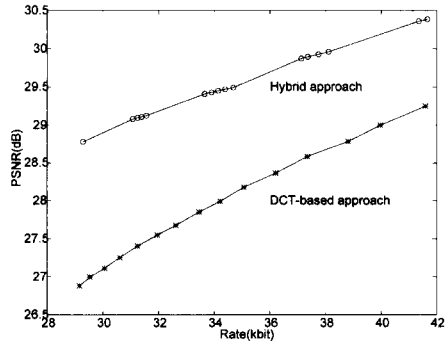



Figure 3. Diagram of the proposed hybrid DCT/wavelet I-frame coding for H.263+.




(a)




(b)

Figure 4. Performance comparison for the Y component of the CIF Salesman sequence: (a) the R-D plot of I-frame and (b) the R-D plot of the following P-frame.

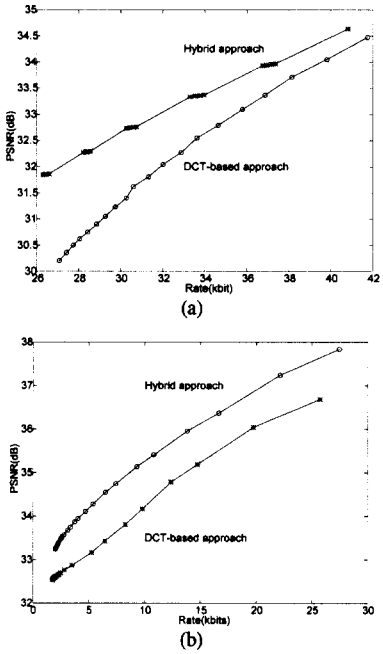



Figure 5. Performance comparison for the Y component of the CIF Akiyo sequence: (a) the R-D plot of I-frame and (b) the R-D plot of the following P-frame.

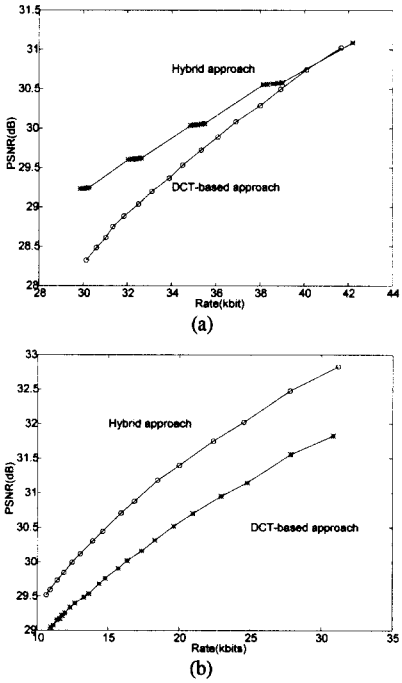
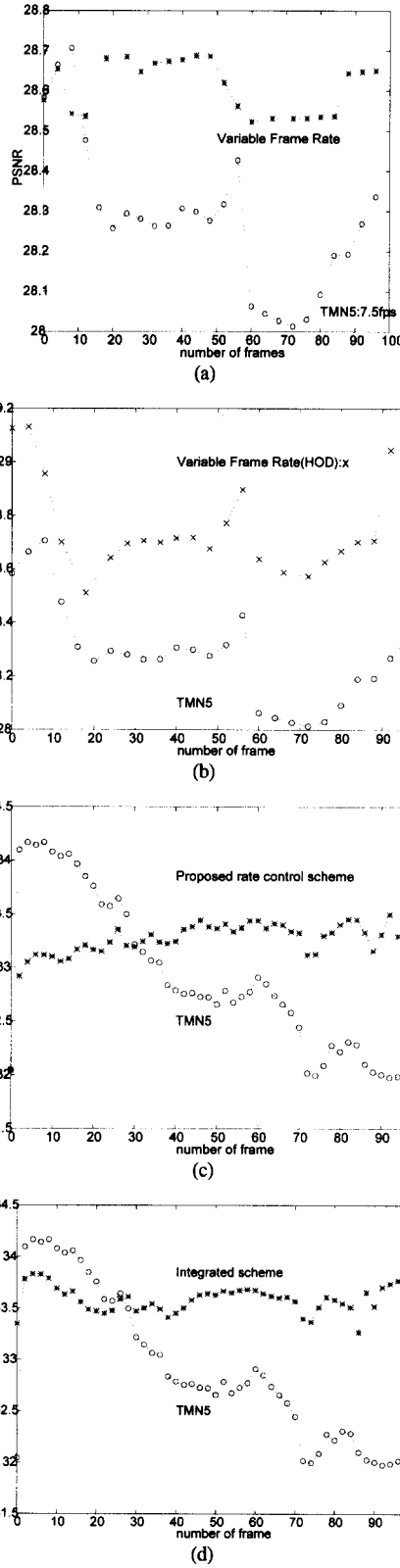
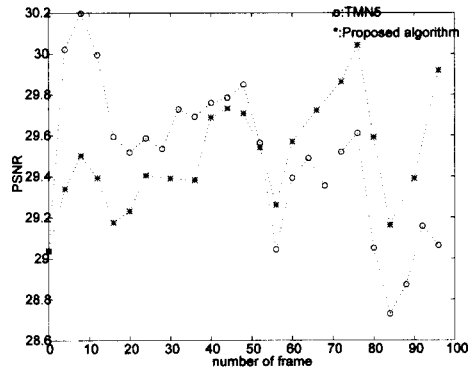
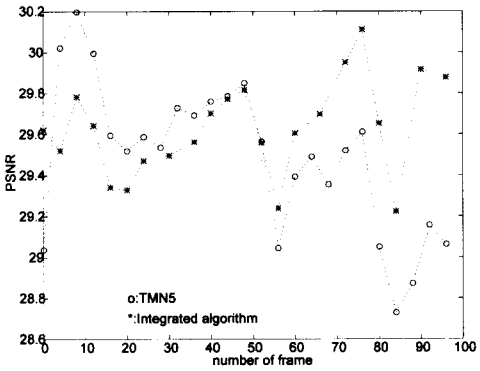



Figure 6. Performance comparison for the Y component for the CIF Foreman sequence: (a) the R-D plot of I-frame and (b) the R-D plot of the following P-frame, where $\circ$ denotes observed data and the solid line denotes the calculated model.





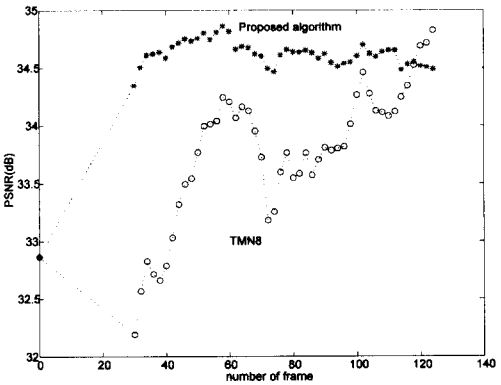

(e)



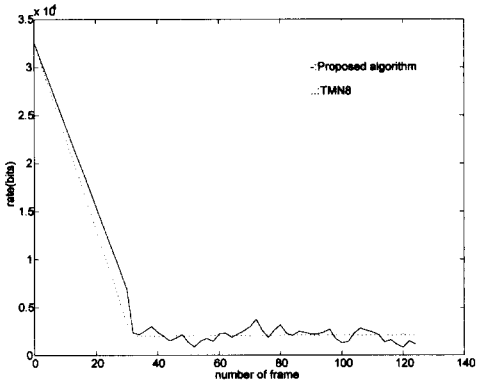

(f)

Figure 7. The PSNR performance comparison between TMN5 and the proposed rate control algorithm for the Y component:

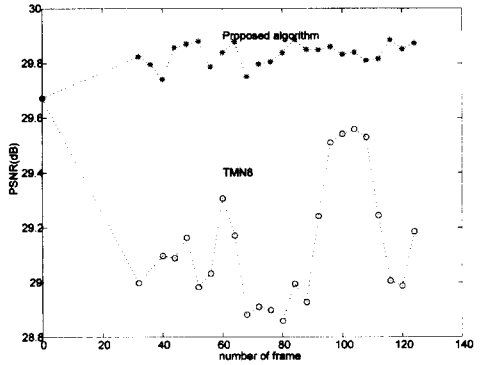
(a) Salesman with TMN5,
(b) Salesman with hybrid,
(c) Akiyo with TMN5,
(d) Akiyo with hybrid,
(e) Foreman with TMN5, and
(f) Foreman with hybrid I.



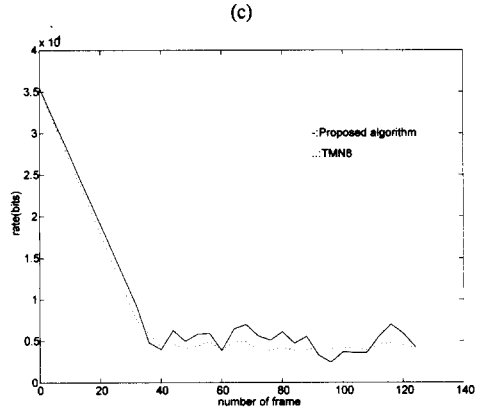

(a)



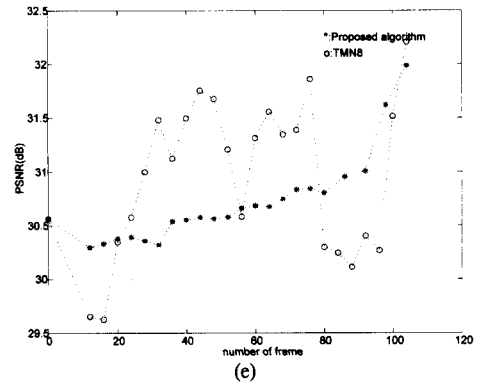

(b)



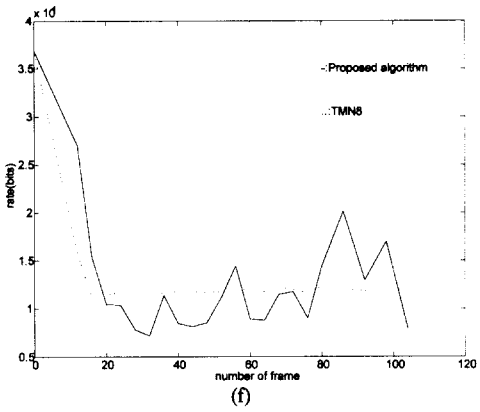

(c)




(d)




(e)



(f)

Figure 8. The 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) the PSNR plot for Akiyo,
(b) the rate plot for Akiyo,
(c) the PSNR plot for Silent Voice,
(d) the rate plot for Silent Voice,
(e) the PSNR plot for Foreman, and
(f) the rate plot for Foreman.

## References


[1] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with application to multiresolution and MPEG video coder," *IEEE Trans. on Image Preocessing*, vol. 3, pp. 533-545, Sept. 1994.

[2] L.J.Lin, A Ortega, and C. C. J. Kuo, "Rate control using spline interpolated R-D characteristics," in *Proc. of SPIE Visual Communication and Image Processing*, pp. 111-122, Mar. 1996.

[3] J. Lee and B. W. Dickinson, "Temporally adaptive motion interpolation exploiting temporal masking in visual perception," *IEEE Trans, on Image Processing*, vol 3, pp. 513-526, Sept. 1994.

[4] A. R. Reibman and A. W. Berger, "Traffic descriptors for VBR video teleconferencing" *IEEE/ACM Trans. on Networks*, vol. 3, pp. 329-339, June 1995

[5] 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. on Circuits and Systems for Video Technology*, vol. 6, pp. 182-190, Apr. 1996.

[6] 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 Communition*, vol 15, Sept 1999.

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

[8] J. J. Chen and D. W. Lin, "Optimal bit allocation for coding of video signals over ATM networks," *IEEE Journal on Selected Areas in Communication*, Vol. 15, pp. 1002-1015, Aug. 1997.

[9] J. M. Shapiro "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3455-3462, Dec. 1993.

[10] D. Tauman and A. Zakhor, "Multirate 3-d subband coding of video," *IEEE Trans. on Image Processing*, vol. 3, pp. 572-588, Sept. 1994.

[11] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, pp. 243-250, June 1996.

[12] S. M. LoPresto, K. Ramchandram, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantixation framework," in *Proc. of SPIE Visual Communication and Image Processing*, Jan. 1997.

[13] E.A.B.D. Sliva and M. Ghanbari, "A hybrid subband-DCT coder for transmission of high resolution still pictures at 64 kbits," *Journal of Visual Communication and Image Representation*, vol. 6, pp. 164-177, June 1995.

[14] Moving Picture Expert Group Video Group "Report of core experiment t1: wavelet coding of intra frames," Draft International Standard MPEG96/M1869, ISO/IEC JTC1/SC29/WG11, Feb. 1997.

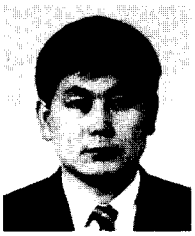
[15] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 246-250, Sept. 1997

[16] J. Li, P.-Y. Cheng, and C.-C. J. Kuo, "Image compression using fast rate-distortion optimized wavelet packet transform," *IEEE Trans. on Circuits and Systems for Video Technology,* submitted 1996.

[17] T. Research, "TMN(H.263) encoder/decoder, version 2.0," tmn(h.263) codec, June 1996.

[18] Image Processing Lab, University of British Columbia, "H.263+ encoder/decoder," TMN (H.263) codec, Feb 1998.

송 황 준(Hwangjun Song)



1990.2 서울대학교 공과대학 제어계측학과 졸업 (공학사)

1992.2 서울대학교 공과대학원 제어계측학과 졸업 (공학석사)

1999.5 EE-Systems, University of Southern California, Los Angeles, USA.(공학 박사)

1999.5~1999. 12 SIPI&IMSC(sponsored by NSF), USC. 연구원

2000.3~현재 세종대학교 전자정보공학대학 소프트웨어학과 전임강사

<주관심 분야> Multimedia communication/signal processing, Internet video, Packet video, Network protocols Image/video compression