Frame-Layer Constant-Quality Rate Control of Regions of Interest for Multiple Encoders With Single Video Source

Ping-Hao Wu and Homer H. Chen, *Fellow, IEEE*

***Abstract*—In this work, we develop a constant-quality rate control algorithm for a surveillance system which consists of one "base encoder" that encodes a down-sampled full-view version of the input video sequence, and one "region of interest (ROI) encoder" that encodes the region of interest of the input video at the original resolution. Exploiting the inter-relationship between these two independent encoders, the algorithm allocates the bits for the ROI encoder according to the distortion obtained from the corresponding region in the base encoder. Simulation results show that the proposed algorithm can achieve significant reduction in the image quality variation. Compared to the rate control algorithm in JM 8.4, the overall quality is improved and the bit rate is saved.**

***Index Terms*—Bit allocation, H.264/AVC, rate control, surveillance, video coding.**

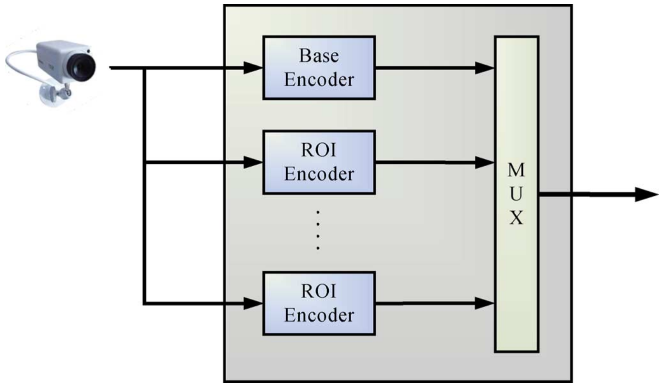


Fig. 1. A high-resolution camera and the base and ROI encoders.

I. Introduction

FUELED by the rapid development of international digital video coding standards [1]–[6], various visual information processing and communication systems have been developed. Rate control plays an important role in any video communication system. It deals with control mechanisms for determining the date rate of compressed video so that successful delivery of video streams and best visual quality can be achieved.

According to the bit rate characteristics of the compressed video, rate control can be either constant bit rate (CBR) or variable bit rate (VBR). CBR video has been widely adopted in many digital video applications, like digital TV broadcast or video conferencing, that are subject to the constraint imposed by constant channel bandwidth. However, due to the non-stationary nature of video signals, it is almost impossible to achieve constant video quality with CBR encoding. On the other hand, VBR encoding is able to provide constant video quality.

Many constant-quality rate control algorithms have been proposed [9]–[16], [22], [23]. Adaptive algorithms such as the one proposed in [9] vary the quantization step size according to the properties of an image sequence. However, they cannot guarantee to meet the constraint on storage size. Two-pass algorithms [11]–[13] generate constant-quality video in the second-pass of the encoding process according to the information obtained from the first-pass encoding. Such algorithms perform effectively, but the computational complexity and the two-pass nature make them unsuitable for real-time applications. To solve this problem, several single-pass constant-quality rate control algorithms have been developed [14]–[16], [22], [23]. In these algorithms, the quantization parameter for a frame is selected according to the statistics gathered from previously encoded frames. The information of the current frame is not available since it is not encoded yet, which is not the case for the problem considered here.

In this paper, we consider an application scenario where multiple video encoders are employed in a surveillance system to encode the full view as well as the regions of interest (ROI) of the scene and the input videos of these video encoders come from a single camera, as shown in Fig. 1. Since the video encoders share a common video source, the inter-relationship between these encoders can be exploited to improve the rate control. With the information of the current frame and the previous frames obtained during the encoding process of the base encoder, the output bit rate and the video quality of the ROI encoders can be better controlled. Constant output video quality can be achieved without pre-analysis of the whole video sequence while information of the current frame is available. With the proposed algorithm, video quality fluctuates less, and the

Manuscript received July 25, 2006; revised January 31, 2007. This work was supported in part by the National Science Council of Taiwan, R.O.C., under contracts NSC 94-2752-E-002-006-PAE and NSC 94-2219-1-002-016 and in part by the Industrial Technology Research Institute under contract T1-94081-10. This paper was recommended by Associate Editor M. Comer.

P.-H. Wu was with the Graduate Institute of Communication Engineering, National Taiwan University, Taipei 10617, Taiwan, R.O.C. He is now with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA.

H. H. Chen is with the Department of Electrical Engineering, Graduate Institute of Communication Engineering, and Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei 10617, Taiwan, R.O.C. (e-mail: homer@cc.ee.ntu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2007.899605

1051-8215/$25.00 © 2007 IEEE

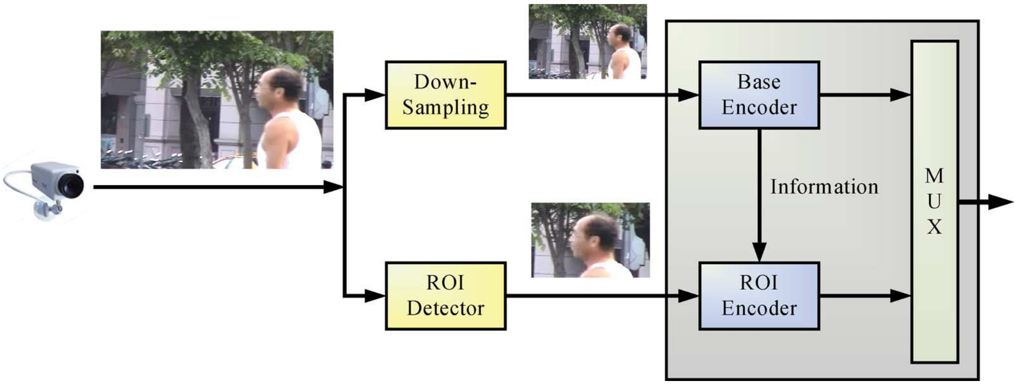


Fig. 2. Video surveillance system with one single video source, one down-sampler, and one ROI detector.

overall performance is also significantly improved in terms of average peak-signal-to-noise-ratio (PSNR) and output bit rate.

This paper is organized as follows. Section II describes the video surveillance system that consists of one single camera and multiple video encoders. The proposed rate control algorithm is described in Section III. Section IV shows the simulation results, followed by a conclusion in Section V.

II. Multiple Encoders With Single Video Source

With the high compression efficiency brought by the digital video coding standards, wide deployment of video surveillance systems is coming of age. For security reasons, some regions of the video, called the ROI, are often required to be shown in more details than other regions. Usually, lower resolution is sufficient for the full-view video, while the ROI video needs higher resolution so that the details can be clearly seen. However, conventional surveillance systems are unable to provide both full-view and ROI videos at the same time to meet different resolution requirements.

Though multiple cameras can solve the problem, the cost is an issue. Besides, the synchronization problem between different cameras is not trivial. To avoid these problems, the system considered in the paper uses one high-definition camera to capture the video. The full-view video is obtained by down-sampling the captured video, while the ROI videos are encoded at the original resolution.

For simplicity, only one ROI video is considered in this work, as shown in Fig. 2. The system consists of two video encoders: one base encoder that encodes the full-view video at a low resolution and one ROI encoder that encodes the ROI video at the original resolution. In this system, the high-resolution frame is down-sampled horizontally and vertically by a factor of two prior to being fed into the base encoder. The ROI region is extracted from the original frame and then encoded by the ROI encoder. The next frame is processed after the current frame is encoded.

Note that the ROI contains the same content as its corresponding region in the base sequence. The only difference between them is the resolution. Therefore, certain correlations between the ROI sequence and its corresponding region in the base sequence can be expected. By exploiting such correlations, the rate control of the ROI sequence can be improved, as described in the following section. The work described in [17] is related to ours in that it adaptively skips non-ROI area during bit allocation to improve overall subjective quality.

The choice of separate base and ROI encoders in the coding framework described in this paper is the result of a tradeoff between cost and performance. It is well recognized that conventional surveillance systems fail to offer images at the level of quality needed for critical missions such as the crime investigation of a bank robbery. With a high-resolution camera, images with detailed content can be captured. However, the requirement for large transmission bandwidth and storage capacity becomes a problem. Although the problem may be resolved in the long run when bandwidth and storage cost is dropped, the application scheme described in this work represents a plausible solution that is applicable now and in the future.

III. Proposed Rate Control Algorithm for the Region of Interest

Because the ROI sequence and its corresponding region in the base sequence have similar content but at different resolutions, information obtained through the encoding of the base video can be used to improve the encoding of the ROI video, which is different from the case of a single video encoder. Our previous work described in [18] shows that the mode information obtained from the base sequence can be used to predict the mode in the ROI sequence. In this paper, the statistics of the base sequence are used to better control the bit rate and the quality of the ROI sequence.

As suggested by the name, ROI are often needed to be shown in more details. By employing a constant-quality encoding algorithm, each frame of the ROI sequence can be encoded at the same quality level, avoiding unnecessarily high PSNR of the low-activity frames and saving more bits for the high-activity frames.

Since constant video quality is not needed for the base video, the base video in the surveillance system is encoded with a CBR rate control algorithm which has relatively low computations compared to VBR algorithms. The statistics of the base encoder collected during the encoding are used to estimate how many bits are needed for a frame in the ROI video to achieve constant quality.

The proposed algorithm can be divided into five parts: initial quantization parameter (QP) determination, mean absolute difference (MAD) prediction, remaining bits estimation, target bit allocation, and QP determination. We describe each of them in the following subsections.

A. Initial QP Determination

We start with the determination of the initial quantization parameter. In one-pass encoding systems, statistics of the video sequence cannot be obtained, which makes the initial QP determination difficult. In MPEG-2 TM5 [8], a fixed initial QP at the beginning of the encoding is selected, whereas in JM 8.4 [7], which is the reference software of H.264, the initial QP is selected among several predefined QPs

$$\mathrm{QP}_0 = \begin{cases} 35, & \mathrm{bpp} \le l_1 \\ 25, & l_1 < \mathrm{bpp} \le l_2 \\ 20, & l_2 < \mathrm{bpp} \le l_3 \\ 10, & \mathrm{bpp} > l_3 \end{cases} \tag{1}$$

where $(l_1,\ l_2,\ l_3) = (0.1,\ 0.3,\ 0.6)$ is recommended for QCIF size video, $(l_1,\ l_2,\ l_3) = (0.2,\ 0.6,\ 1.2)$ is recommended for CIF size video, and $(l_1,\ l_2,\ l_3) = (0.2,\ 1.4,\ 2.4)$ for video with frame size larger than CIF. In (1)

$$\mathrm{bpp} = \frac{R_t}{f \times N_{\mathrm{pixel}}} \tag{2}$$

where R_t denote the target bit rate and N_{pixel} is the total number of pixels in a frame. For example, $N_{\mathrm{pixel}} = 352 \times 288 \times 1.5$ for a YUV 4:2:0 sequence of CIF size.

This approach maps the average bits per pixel into four possible initial quantization parameters. The rate control algorithm is thus inflexible because it considers only four candidates.

To solve the problem, an analytical model with a logarithm form is proposed in [20]

$$\mathrm{QP}_0 = a_1 \times \log(\mathrm{bpp}) + a_2 \tag{3}$$

where QP_0 denotes the initial QP.

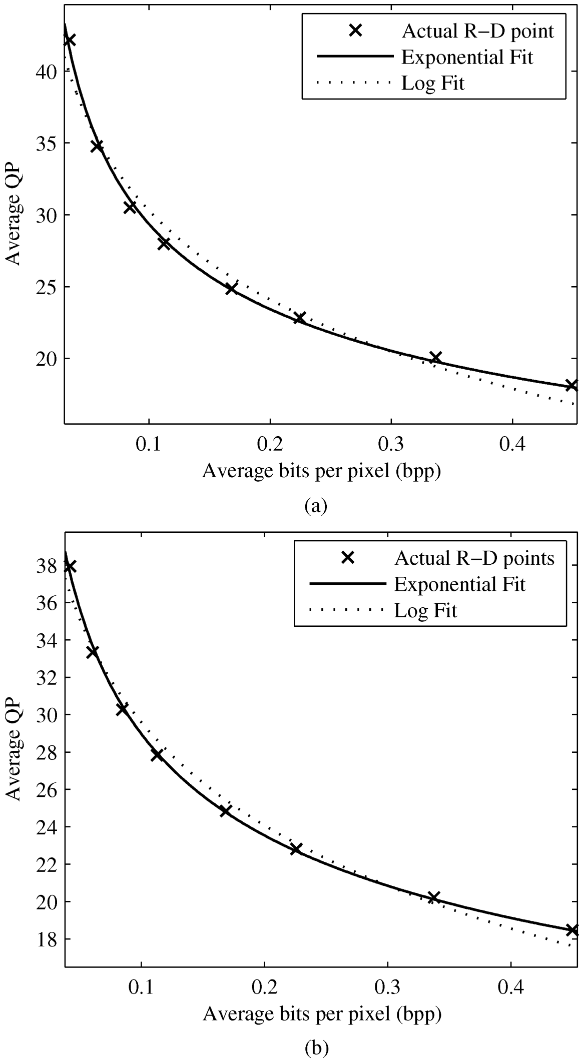
In our experiments, we find that the model

$$\mathrm{QP}_0 = a \times \mathrm{bpp}^b \tag{4}$$

is more accurate than the logarithm model. Typical values for a and b are 14 and -0.32, respectively, for CIF size sequences according to our experiments. Specifically, this model is empirically determined by data fitting using 12 CIF sequences that are encoded using JM 8.4 [7] at 8 different bit rates, and the average QP versus average bits per pixel is plotted. Fig. 3(a) is the case when the sequence contains no B-frame, and Fig. 3(b) is the case of two B-frames between two I- or P-frames. From the figures we can see that (4) is a much better approximation to the actual data than (3). In this way, the initial QP is not restricted to a few candidates any more. An appropriate initial QP can be selected according to the encoding setting.

B. Mean Absolute Difference (MAD) Prediction

H.264 adopts the rate-distortion optimization (RDO) to optimally select the mode and motion vectors, which makes the task of rate control for H.264 much harder than the previous standards. The reason is that the quantization parameter is involved in both rate control and RDO. RDO needs the QP to calculate the Lagrange multiplier λ and the number of bits required to encode the frame or macroblock. Different QPs may result in different motion vectors and modes, which implies that the QP must be determined first by the rate control. However, the MAD between the original and the motion-compensated block, which is needed to calculate QP before RDO [19], is only available after performing the RDO. Consequently, a chicken and egg dilemma arises.



(a)

(b)

Fig. 3. Average QP-bpp curve of 12 CIF sequences encoded at 8 different bit rates, (a) $\mathrm{N} = 30$ and $\mathrm{M} = 1$, (b) $\mathrm{N} = 30$ and $\mathrm{M} = 3$, where N is the GOP size and M the sub-GOP size.

In the rate control of H.264 [7], this chicken and egg dilemma is solved by linearly predicting the MAD of the current frame from the MAD of the previous frame

$$\mathrm{MAD}_{cb} = a_1 \times \mathrm{MAD}_{pb} + a_2 \tag{5}$$

where MAD_{cb} denotes the predicted MAD of the current frame, MAD_{pb} denotes the actual MAD of the previous frame, and a_1 and a_2 are model parameters that are updated through linear regression by using the encoding statistics of the previous frame.

Obviously, if scene change occurs, this approach would fail. To prevent the failure, we avoid using the previous ROI-frame and propose to predict the MAD of the current frame in the ROI sequence from the corresponding region in the base sequence. The MAD prediction by linear regression becomes

$$\mathrm{MAD}_{\mathrm{curr,ROI}} = a_1 \times \mathrm{MAD}_{\mathrm{curr,base}} + a_2 \tag{6}$$

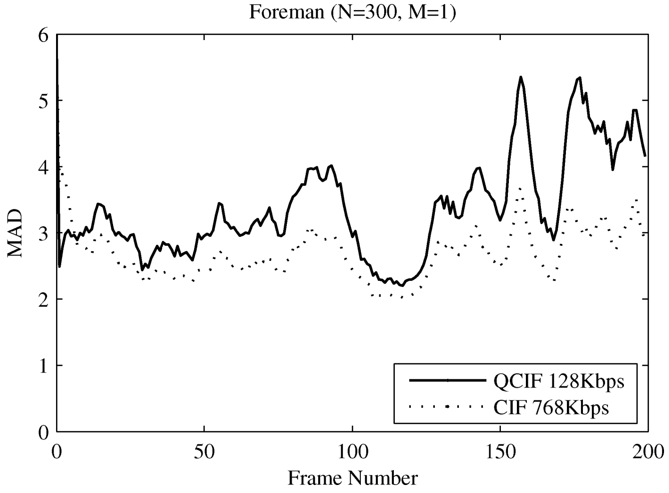


Fig. 4. MAD curves of *Foreman*.

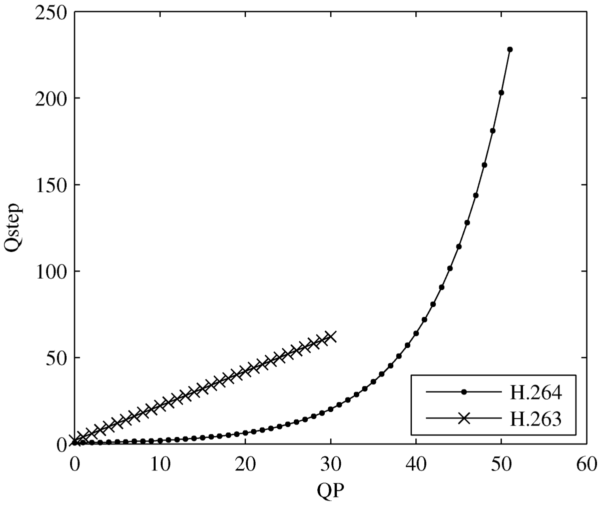


Fig. 5. Quantization step size versus quantization parameter.

where $\mathrm{MAD}_{\mathrm{curr,ROI}}$ denotes the MAD of the current frame in the ROI sequence and $\mathrm{MAD}_{\mathrm{curr,base}}$ denotes the MAD of the corresponding region in the base sequence.

Fig. 4 shows the MAD curves of the CIF size and the QCIF size *Foreman* encoded at 76 and 128 kbps, respectively, using JM 8.4. Only the first frame is set as the intra frame. From the figure, we can see that the MAD curves of these two sequences, which contain similar content at different resolutions, are very similar in trend.

C. Remaining Bits Estimation

Remaining bits are allocated to each picture type according to the corresponding complexity measure. In the proposed algorithm, the complexity measure of MPEG-2 TM5 [8] is adopted

$$X^t = b^t \times Q^t, \quad t \in I, P, B \tag{7}$$

where X denotes the complexity, b the actual number of bits, Q the quantization parameter, and t the frame type. However, unlike previous standards where the relationship between QP and the quantization step size (Q_{step}) is linear,[1] H.264 adopts an exponential relationship

$$Q_{\mathrm{step,H.264}} = 2^{(\mathrm{QP}-4)/6} \tag{8}$$

as shown in Fig. 5. We use the quantization step size instead of the quantization parameter for Q^t in (7) to determine the complexity of the picture since the quantization step size is the true value used to quantize the discrete cosine transform (DCT) coefficients while the quantization parameter indirectly indicates the step size.

The remaining bits are allocated according to the formula

$$T^t_{\mathrm{rem}} = \frac{T_{\mathrm{total}} \times X^t_{\mathrm{avg}} N^t_{\mathrm{rem}}}{X^I_{\mathrm{avg}} N^I_{\mathrm{rem}} + X^P_{\mathrm{avg}} N^P_{\mathrm{rem}} + X^B_{\mathrm{avg}} N^B_{\mathrm{rem}}}, \quad t \in I, P, B \tag{9}$$

where T^t_{rem} is the number of remaining bits for frame type t, T_{total} is the total number of remaining bits, X^t_{avg} is the average complexity for frame type t, and N^t_{rem} is the number of remaining frames of type t.

The average complexity X^t_{avg} is updated using the average number of bits b^t_{avg} and the average quantization step size $Q_{\mathrm{step,avg}}$ of previously encoded frames of type t

$$\begin{aligned} X^t_{\mathrm{avg}} &= b^t_{\mathrm{avg}} \times Q^t_{\mathrm{step,avg}} \\ &= \left(\frac{1}{W^t_X} \sum_{i=N^t_c - W^t_X + 1}^{N^t_c} b^t(i) \right) \times \left(\frac{1}{W^t_X} \sum_{i=N^t_c - W^t_X + 1}^{N^t_c} Q^t_{\mathrm{step}}(i) \right) \end{aligned} \tag{10}$$

where $b^t(i)$ and $Q^t_{\mathrm{step}}(i)$ are the actual number of bits and the average quantization step size for the ith type-t frame, respectively, and N^t_c denotes the number of coded frames of type t. The window length W^t_X is computed by

$$W^t_X = \min\{N^t_c, \theta_X\} \tag{11}$$

where θ_X is the upper bound of the window length.

D. Target Bit Allocation

Because the content is the same but at different resolutions, it can be expected that the ROI sequence and its corresponding region in the base sequence have similar picture quality trends. As an illustration, Fig. 6 shows the PSNR curves of the CIF size and the QCIF size *Table Tennis* sequences encoded at target bit rate 768 and 128 kbps, respectively. Different bitrate settings produce similar results. We can see that the PSNR curves of these two sequences are very similar except at the first GOP. Based on this observation, we use the distortion obtained from the base sequence to allocate bits for the ROI video.

To allocate bits for the current frame so that constant quality can be achieved, a frame complexity measure is defined according to the distortion of the corresponding region in the base sequence. Let $D^t_{\mathrm{base}}(i)$ denote the distortion of the ith type-t

[1] Taking H.263 for example, the quantization step size is computed by $Q_{\mathrm{step},H.263} = 2 \times \mathrm{QP}$.

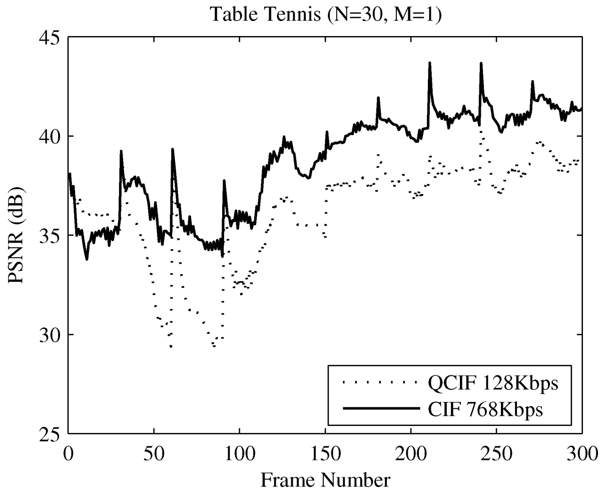


Fig. 6. PSNR curves of *Table Tennis*.

frame obtained from the corresponding region of the base sequence. Then the frame complexity S_{base}^{t} is defined as

$$S_{\text{base}}^{t} = \frac{\ln\left(D_{\text{base}}^{t}\left(N_c^t+1\right)\right)}{\ln\left(\bar{D}_{\text{base}}^{t}\right)} \tag{12}$$

and

$$\bar{D}_{\text{base}}^{t} = \frac{1}{W_D^t}\sum_{i=N_c^t-W_D^t+2}^{N_c^t+1} D_{\text{base}}^{t}(i) \tag{13}$$

where $D_{\text{base}}^{t}\left(N_c^t+1\right)$ is the distortion of the current frame and $\bar{D}_{\text{base}}^{t}$ denotes the average distortion of the previous frames obtained from the corresponding region in the base sequence. The distortion is computed using the MAD between the original and the reconstructed pixels (luminance only)

$$D_{\text{base}}^{t}(i) = \sum_{x,y}\left|p_{\text{org}}^{t}(x,y) - p_{\text{rec}}^{t}(x,y)\right| \tag{14}$$

where $p_{\text{org}}^{t}(x,y)$ denotes the pixel (x,y) in the original frame and $p_{\text{rec}}^{t}(x,y)$ the pixel (x,y) in the reconstructed frame. W_D^t in (13) is the window length, which is computed by

$$W_D^t = \min\left\{N_c^t,\ \theta_D\right\} \tag{15}$$

where θ_X is the upper bound of the window length.

With the frame complexity measure defined in (12), the number of bits allocated to the current frame is then computed according to the average bits actually produced

$$T^t = S_{\text{base}}^{t} \times \frac{1}{W_D^t}\sum_{i=N_c^t-W_D^t+1}^{N_c} b^t(i) \tag{16}$$

where T^t represents the bits needed for the current frame to achieve the average quality of previous W_D^t frames. If the current distortion of the corresponding region in the base sequence is larger than the average distortion of previous frames in the base sequence, S_{base}^{t} would be larger than 1, and thus more bits are allocated to the current frame so that constant quality can be achieved.

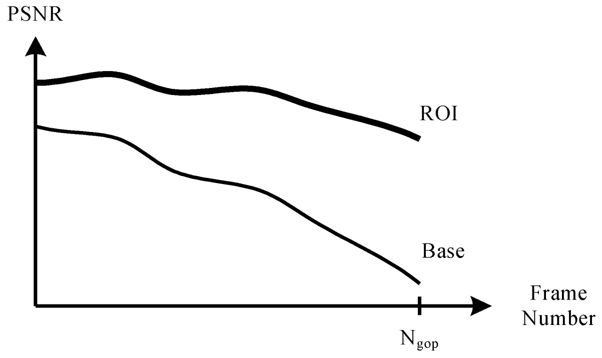


Fig. 7. PSNR curves of the ROI and the base sequence at different quality levels.

However, in this way, the remaining number of bits is not taken into consideration, and the target bit rate constraint may not be satisfied. In storage applications, this may cause unnecessary waste of the storage space; even worse, the size of the entire bitstream may exceed the total available storage space. To handle the target bit rate constraint, the target number of bits is then modified as

$$\begin{aligned} T^t &= S_{\text{base}}^{t} \times \text{Average remaining bits} \\ &= S_{\text{base}}^{t} \times \left(\frac{T_{\text{rem}}^{t}}{N_{\text{rem}}^{t}}\right). \end{aligned} \tag{17}$$

Because the rates of the remaining bits and the used bits both should be close to the target bit rate, it is reasonable to replace the average used bits with the average remaining bits. This way, the target bit rate constraint can be achieved more nicely.

This approach is based on the assumption that the quality trends of the corresponding regions between the base sequence and the ROI sequence are similar, which is reasonable since they represent the same scene at different resolutions. However, this does not take the speed of quality change into consideration. If the base sequence and the ROI sequence are encoded to be at different quality levels, one of them may change faster in terms of PSNR than the other one does. As illustrated in Fig. 7, the PSNR of the base sequence drops much faster than that of the ROI sequence. This may make the mechanism described by (17) to allocate too many bits to the current frame.

To solve this problem, we propose to monitor the distortion variation of more than one previous frame in the ROI sequence. Similar to (5), the distortion of the current frame in the ROI sequence can be linearly predicted by exploiting the locally stationary property

$$\tilde{D}_{\text{ROI}}(i) = c_1 \times D_{\text{ROI}}(i-1) + c_2 \tag{18}$$

where $D_{\text{ROI}}(i-1)$ is the distortion of frame $(i-1)$, $\tilde{D}_{\text{ROI}}(i)$ is the linear prediction of the distortion of frame i, and c_1 and c_2 are the model parameters that are updated by linear regression after the current frame is coded. Then, similar to S_{base}^{t}, the adjustment factor S_{ROI}^{t} for the target bits is calculated as follows:

$$S_{\text{ROI}}^{t} = \frac{\ln\left(\tilde{D}_{\text{ROI}}^{t}\left(N_c^t+1\right)\right)}{\ln\left(\bar{D}_{\text{ROI}}^{t}\right)} \tag{19}$$

and

$$\bar{D}^t_{\mathrm{ROI}} = \frac{1}{W^t_D} \sum_{i=N^t_c - W^t_D + 1}^{N^t_c} D^t_{\mathrm{base}}(i) \tag{20}$$

where $\bar{D}^t_{\mathrm{ROI}}$ denotes the average distortion of W^t_D previous frames. If the predicted distortion of the current frame is larger than the average distortion of previously encoded frames, more bits are allocated to the current frame, and vice versa.

The frame complexity measure for the current frame is adjusted by the factor calculated in (19)

$$S^t_{\mathrm{current}} = S^t_{\mathrm{ROI}} \times S^t_{\mathrm{base}}. \tag{21}$$

To avoid allocating too many or too few bits, S^t_{current} is further bounded

$$S^t_{\mathrm{current}} = \min\left\{\theta_1,\ \max\left\{\theta_2,\ S^t_{\mathrm{current}}\right\}\right\} \tag{22}$$

where θ_1 and θ_2 are the upper and lower bound of the frame complexity measure, respectively.

Finally, the target number of bits for the current frame is computed by (17), except that S^t_{base} is replaced with the adjusted and bounded frame complexity S^t_{current}. That is

$$T^t = S^t_{\mathrm{current}} \times \left(\frac{T^t_{\mathrm{rem}}}{N^t_{\mathrm{rem}}}\right). \tag{23}$$

E. QP Determination

We now discuss how the quantization parameters of I- and P-frames are determined. For simplicity, the quantization parameter for B-frames is determined in the same way as the one in JM 8.4.

1) QPs of I-Frames: In the rate control algorithm [19] adopted in H.264 JM, the quantization parameters for I-frames are determined by the average of quantization parameters for all P-frames in the previous GOP and are bounded within $\mathrm{QP}^I_{\mathrm{prev}} \pm 2$, where $\mathrm{QP}^I_{\mathrm{prev}}$ is the quantization parameter of the previous I-frame. Since the scene may have already changed after one GOP, it is unnecessary and unrealistic to put restriction on the difference of quantization parameters between the two I-frames of successive GOPs.

Furthermore, the quantization parameter for I-frame in [19] is usually set too small. Although an I-frame with better quality can reduce the bits needed to encode the following frames at the same quality level, it may overuse its bit budget if the quantization parameter is set too small, causing quality degradation for the following inter frames. Therefore, an appropriate quantization parameter for I-frame is needed to ensure that the inter frames of this GOP have a reference with sufficient quality and to avoid spending too many bits on this reference frame.

Based on these observations, the quantization parameter for I-frame is modified from [19] as follows:

$$Q_{\mathrm{step},i}(1) = \frac{\mathrm{SumPQ]_{step}}\,(\mathrm{i}-1)}{N_P(i-1)} \tag{24}$$

$$\mathrm{QP}_i(1) = Q_{\mathrm{step}} to \mathrm{QP}\left(Q_{\mathrm{step},i}(1)\right) - \min\left\{1, \frac{N_{i-1}}{15}\right\} \tag{25}$$

where $Q_{\mathrm{step},i}(1)$ denotes the quantization step size of the first frame (an I-frame) in the ith GOP, $\mathrm{SumPQ}_{\mathrm{step}}(i-1)$ is the sum of quantization step sizes of all P-frames in the previous GOP, and Q_{step} to $\mathrm{QP}(\cdot)$ denotes the function that converts a quantization step size to a quantization parameter. The result of the division operation in (25) is truncated to integer.

Note that, compared to the original scheme [19], the quantization step sizes instead of quantization parameters are used to determine the quantization parameter for the current I-frame. Note also that the maximum adjustment is 1, as shown in the second term on the right-hand side of (25), instead of 2. This prevents the I-frame from overusing its bit budget while still providing reference frames with sufficient quality. Furthermore, smaller adjustment helps to maintain uniform picture quality.

2) QPs of P-Frames: The quantization parameter for P-frames is determined according to three different quantization parameters, $Q_{\mathrm{distortion}}$, Q_{CBR}, and Q_{constant}, which are computed with different considerations and are described in detail in the following.

In [22], a one-pass VBR control algorithm based on the average coding complexity is proposed

$$Q = \frac{X_{\mathrm{avg}}}{R_t} \tag{26}$$

where Q denotes the quantization parameter, R_t denotes the target bit rate, and X_{avg} is the average complexity of all coded frame calculated by (7). Instead of using the target bit rate R_t in (26), Song *et al.* [23] proposed to use the average of the remaining number of bits. That is

$$Q = \frac{X_{\mathrm{avg}}}{R_{\mathrm{avg}}}. \tag{27}$$

To produce constant video quality, it is usually needed to pre-analyze the whole video sequence or to encode the video several times. By using the long-term average of the coding complexities, these one-pass VBR rate control algorithms achieve nearly uniform picture quality without pre-analysis. Note that uniform quality can be achieved by keeping constant quantization parameter. Because X_{avg} is close to the actual average complexity of the whole sequence and R_{avg} is close to the target bit rate, the resulting Q approaches to a constant which is the actual average quantization parameter. Base on this concept, Q_{constant} is determined by

$$Q_{\mathrm{constant}} = Q_{\mathrm{step\ to}}\ \mathrm{QP}\left(\frac{X^P_{\mathrm{avg}}}{\frac{T^P_{\mathrm{rem}}}{N^P_{\mathrm{rem}}}}\right). \tag{28}$$

Note that, unlike [22] and [23], X^P_{avg} in the proposed algorithm is calculated using the quantization step sizes rather than the quantization parameters. The resulting quantization step size, which would be close to the actual average step size of the whole sequence, is mapped to a quantization parameter by the function Q_{step} to $\mathrm{QP}(\cdot)$.

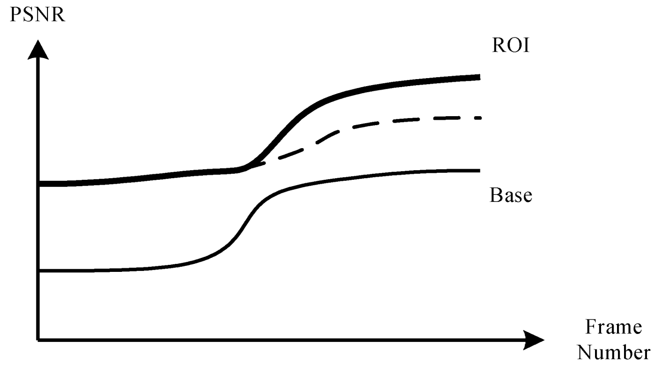


Fig. 8. Transition from a high-activity scene to a low-activity scene.

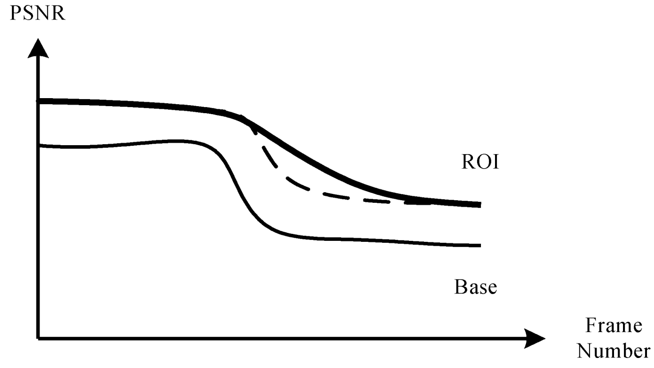


Fig. 9. Transition from a low-activity scene to a high-activity scene.

The quantization parameter for the current frame is computed by

$$Q_{\text{current}} = \min\left(Q_{\text{distortion}},\ Q_{\text{constant}},\ Q_{\text{CBR}}\right) \tag{29}$$

where Q_{CBR} is the quantization parameter computed according to the CBR rate control algorithm described in [19], and the distortion-based quantization parameter $Q_{\text{distortion}}$ is computed by the quadratic model described in [21] with the target bits allocated according to the distortion of the base sequence and the ROI sequence.

The three quantization parameters have different purposes. $Q_{\text{distortion}}$ is calculated for the purpose of ensuring the current frame to have the same quality as the average quality of previous frames, and Q_{constant} is for producing almost constant quantization parameter to eliminate quality fluctuation.

Consider the case of a transition from a scene with high activity (low PSNR) to a scene with low activity (high PSNR) as shown in Fig. 8. The solid line represents the PSNR curve of the base sequence, the dashed line and the bold line, represent the PSNR curves of the ROI sequence obtained by a CBR rate control algorithm and the proposed algorithm, respectively. Recall that the proposed algorithm aims at producing constant-quality video. However, when the PSNR of the current frame in the base sequence is higher than the average PSNR of the previous frames, the allocated bits calculated by (17) would be unnecessarily fewer than the channel can provide. Furthermore, the adjustment factor introduced in (19), which takes advantages of the local stationary property, prevents the distortion of the ROI video from changing too drastically. As a result, the quality level would not rise as the activity becomes lower.

To solve this problem, we introduce the quantization parameter of the CBR rate control algorithm, Q_{CBR}, into the algorithm. If the scene is changing from high activity to low activity, the quantization parameter selection mechanism will choose Q_{CBR} instead of other values so that the output quality level can match the changing of the scene activity.

On the other hand, consider the case shown in Fig. 9 where there is a transition from a scene with low activity (high PSNR) to a scene with high activity (low PSNR). If the ROI video is encoded with a CBR rate control algorithm, the resulting curve is shown by the dashed line. Because of the constant channel bandwidth constraint, a sudden drop of the quality arises. To solve this problem, the quantization parameter Q_{constant} is introduced to prevent the PSNR from dropping suddenly. The resulting PSNR would change smoothly to the new quality level as indicated by the bold line in Fig. 9.

Based on the selected quantization parameter, RDO motion estimation and mode decision can then be performed.

IV. Simulation Results

To evaluate its performance, the proposed rate control algorithm is implemented on the H.264 reference software JM 8.4 [7] which serves as the test benchmark. The output video quality and the quality variation of the proposed algorithm are compared with those of JM 8.4.

To simulate the surveillance system that consists of multiple video encoders with single video source, two sets of sequences, QCIF size and CIF size, are used to represent the base sequences and the ROI sequences, respectively. The CIF size sequences represent the input sequences at the original resolution, while the QCIF size sequences represent the down-sampled version of the corresponding regions in the input video sequences.

There are eight sequences used in the simulations: *Container*, *Foreman*, *Mobile & Calendar*, *Mother & Daughter*, *Salesman*, *Silent*, *Stefan*, and *Table Tennis*. They all consist of 300 frames at 30 fps. The base sequences are encoded by JM 8.4, and after coding the current frame in the base sequence, the reconstructed distortion and the MAD are input to the ROI encoder for MAD prediction and bit allocation.

Two sets of simulations are performed. The first one is baseline profile simulation, where no B-frame is involved. The second one is main profile simulation with an [IBBPBBP…] GOP structure. Both sets of simulations have a GOP size of 30. The ROI encoder is configured to have five reference frames, a search range of 16, and quarter-pel motion vector resolution. Rate-distortion optimization and the Hadamard transform are turned on. In the baseline profile simulation, CAVLC is used for symbol coding, while CABAC is used in the main profile simulation. The parameters θ_X, θ_D, θ_1, and θ_2 in the proposed algorithm are set to 30, 40, 1.2, and 0.8, respectively.

The performance of rate control is measured in terms of the average output bit rate, the average PSNR, and the variance of PSNR for the whole sequence. Note that the PSNR values are measured on the luminance component only. Table I shows the encoding results of JM 8.4 and the proposed algorithm in the

TABLE I
PERFORMANCE COMPARISON FOR [IPPP…] GOP STRUCTURE

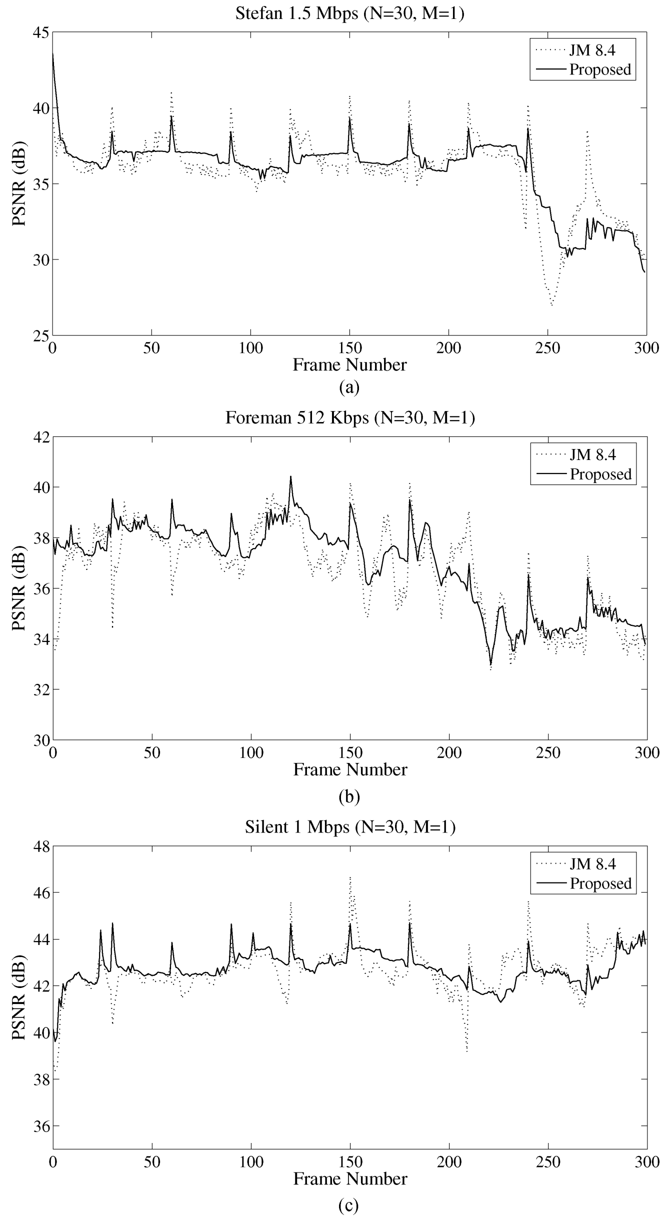
Sequence	Algorithm	Bit-Rate (Kbps)		PSNR (dB)	
		Target	Actual	Average	Variance
Container	JM 8.4	512	512.37	38.655	1.371
	Proposed	512	512.05	38.845	0.282
	Gain	-	0.32	0.190	1.089
Foreman	JM 8.4	512	513.31	36.631	3.190
	Proposed	512	512.92	36.914	2.786
	Gain	-	0.39	0.283	0.404
Mother & Daughter	JM 8.4	512	513..03	43.607	2.955
	Proposed	512	512.77	43.842	1.000
	Gain	-	0.26	0.236	1.954
Table Tennis	JM 8.4	512	512.35	36.658	5.665
	Proposed	512	512.28	36.527	4.166
	Gain	-	0.07	-0.132	1.499
Silent	JM 8.4	1000	999.82	42.662	0.984
	Proposed	1000	1000.09	42.742	0.488
	Gain	-	-0.27	0.080	0.496
Mobile & Calendars	JM 8.4	1500	1501.68	33.181	1.862
	Proposed	1500	1499.94	33.748	1.859
	Gain	-	1.74	0.567	0.003
Salesman	JM 8.4	1500	1499.76	44.279	1.776
	Proposed	1500	1499.22	44.434	0.942
	Gain	-	0.54	0.155	0.834
Stefan	JM 8.4	1500	1500.98	35.755	5.744
	Proposed	1500	1499.84	35.889	4.897
	Gain	-	1.14	0.134	0.847

TABLE II
PERFORMANCE COMPARISON FOR [IPPP…] GOP STRUCTURE

Sequence	Algorithm	Bit-Rate (Kbps)		PSNR (dB)	
		Target	Actual	Average	Variance
Container	JM 8.4	512	513.65	40.016	5.135
	Proposed	512	512.06	40.381	0.632
	Gain	-	1.59	0.365	4.502
Foreman	JM 8.4	512	512.58	37.531	6.088
	Proposed	512	513.00	37.648	4.123
	Gain	-	-0.42	0.111	1.965
Mother & Daughter	JM 8.4	512	512.67	43.327	8.378
	Proposed	512	512.31	44.390	1.558
	Gain	-	0.36	1.063	6.820
Table Tennis	JM 8.4	512	512.64	37.272	7.159
	Proposed	512	512.03	37.278	4.954
	Gain	-	0.61	0.006	2.205
Silent	JM 8.4	1000	1002.23	43.787	2.806
	Proposed	1000	999.34	43.794	1.661
	Gain	-	2.89	0.007	1.145
Mobile & Calendars	JM 8.4	1500	1501.26	35.286	4.536
	Proposed	1500	1498.60	35.195	3.624
	Gain	-	2.66	-0.091	0.912
Salesman	JM 8.4	1500	1501.97	45.390	4.273
	Proposed	1500	1497.08	45.351	1.821
	Gain	-	4.89	-0.039	2.452
Stefan	JM 8.4	1500	1501.20	35.684	11.969
	Proposed	1500	1502.31	35.880	8.802
	Gain	-	-1.11	0.196	3.167

baseline profile simulation. From the values of the PSNR variance, we can see that the quality variation between frames for our algorithm is considerably less than the JM 8.4 rate control algorithm. In addition, the proposed rate control algorithm saves about 0.52 kbps in bit rate, while the overall quality is on average improved by 0.2 dB. All together, the proposed algorithm can achieve significant overall-quality improvement, less quality fluctuation, and bit rate saving. The encoding results of the main profile simulation shown in Table II also show that the quality variation between frames for our algorithm is reduced significantly. Besides, it gains an average of 0.2 dB in PSNR and 1.43 kbps in bit rate over the rate control algorithm in JM 8.4.

Several sequences are selected to demonstrate the transient behavior. The plots of PSNR versus frame number of sequence

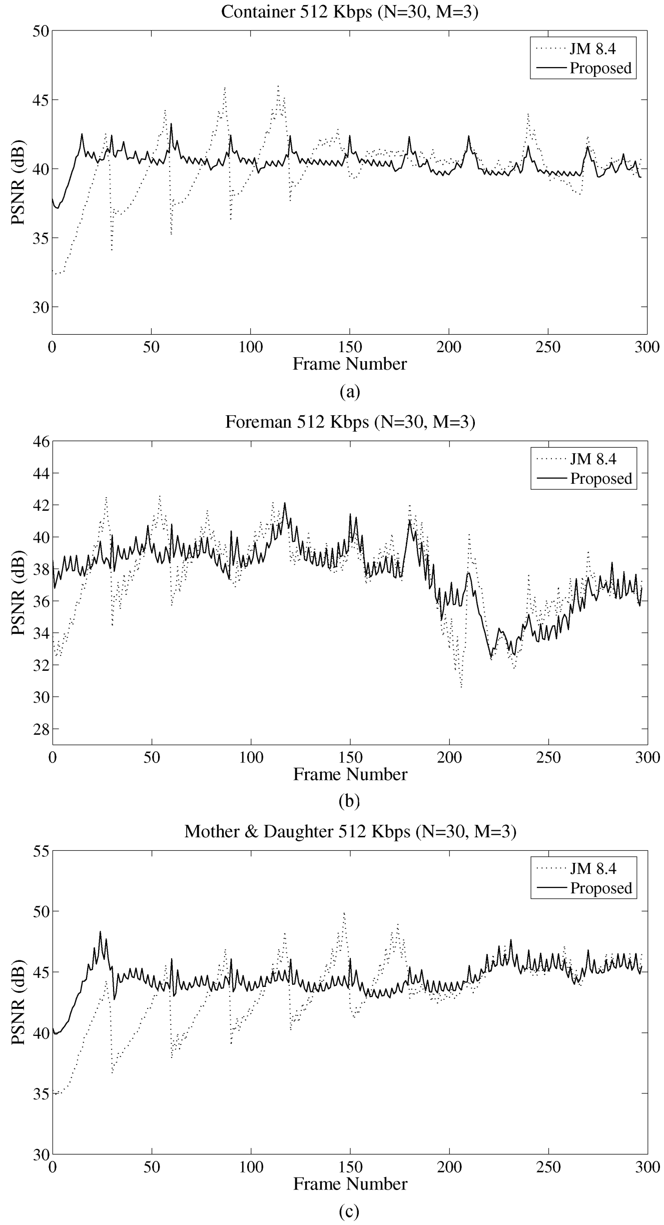


(a)

(b)

(c)

Fig. 10. PSNR curves of JM 8.4 and the proposed algorithm for [IPPP…] GOP structure.



(a)

(b)

(c)

Fig. 11. PSNR curves of JM 8.4 and the proposed algorithm for [IBBP…] GOP structure.

Stefan, *Foreman*, and *Silent* with [IPPP…] GOP structure are shown in Fig. 10. The dashed line represents the PSNR curve of the JM 8.4 rate control algorithm while the solid line represents that of the proposed algorithm. We can see that quality variation of the proposed algorithm is much smaller than that of the JM 8.4. Similar behaviors are also observed in the second set of simulations. The plots of PSNR versus frame number of sequence *Container*, *Foreman*, and *Mother & Daughter* with [IBBP…] GOP structure are shown in Fig. 11.

Since constant-quality rate control cannot guarantee constant output bit rate, it is necessary to ensure that buffer overflow does not occur. Fig. 12 shows the virtual buffer fullness. As can be seen, constant video quality achieved by the proposed algorithm comes at the cost of higher buffer level. However, frame dropping would not occur since the buffer fullness of the proposed algorithm is within the buffer size constraint.

V. Conclusion

A constant-quality rate control algorithm for multiple video encoders with single video source has been presented in this paper. With the additional information obtained from the base encoder, the proposed algorithm can calculate the bits that are needed to encode the current frame of the ROI sequence at the same quality level as the previous frames. The performance of the proposed algorithm is evaluated by comparing with the JM 8.4 rate control algorithm. The simulation results show that the proposed algorithm significantly outperforms JM 8.4 by achieving an average of 38% reduction in the PSNR variation, 0.2-dB gain in the overall PSNR, and 0.97 kbps reduction in the overall bit rate. The proposed algorithm operates in the frame layer. However, it can be applied to macroblock-layer QP adaptation as well.

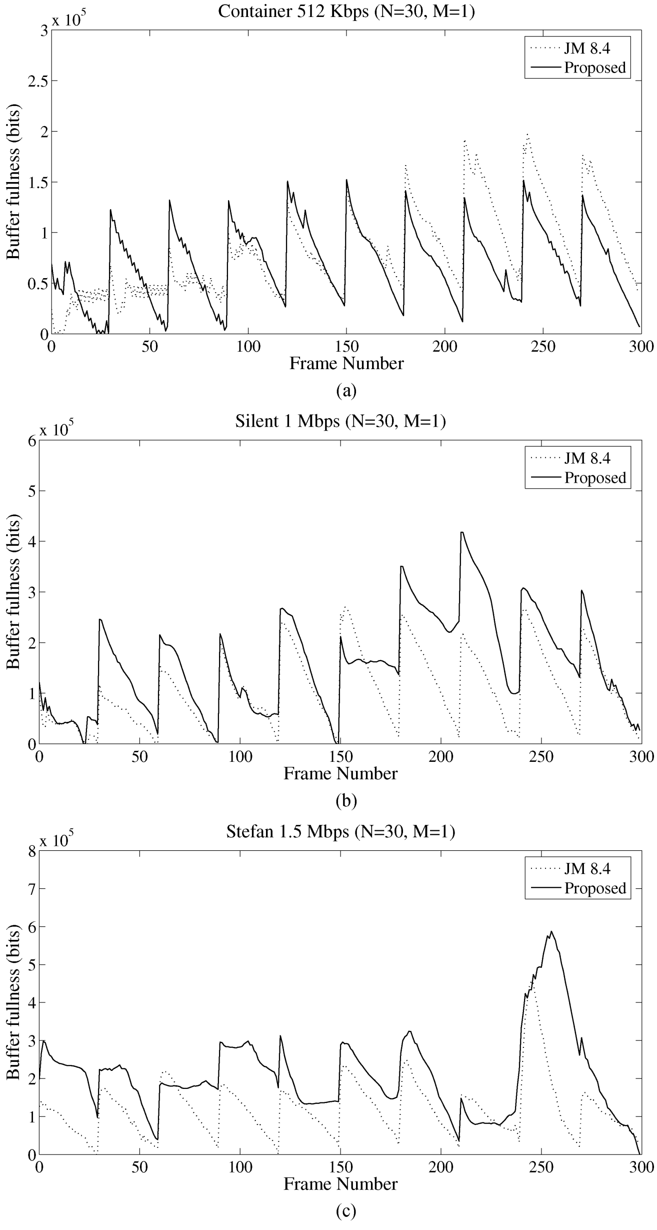


Fig. 12. Buffer fullness curves of JM 8.4 and the proposed algorithm for [IPPP…] GOP structure.

In this work, we explore the inter-relationship between two independent encoders that encode the same content at different image resolutions and exploit such inter-relationship to better control the quality of the ROI encoder. To extend the proposed rate control algorithm to a multilayer video coding scheme such as SVC [24]–[27] with inter-layer prediction, the R-D characteristic of the residual signal in each layer has to be studied. An accurate R-D model for the residual signal is needed. Then the bits can be allocated according to the new frame complexity measure calculated based on the new R-D model.

REFERENCES

[1] *Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to About 1.5 Mbit/s-Part 2: Video*, ISO/IEC 11172-2, 1993.

[2] *Information Technology-Generic Coding of Moving Pictures and Associated Audio Information-Part 2: Video*, ISO/IEC 13818-2, 1995.

[3] *Information Technology-Coding of Audio/Visual Objects*, ISO/IEC 14496-2, 1999.

[4] *Video Codec for Audio Visual Services at $p\times$ 64 Kbits/s*, ITU-T Rec. H.261, 1990.

[5] *Video Coding for Low Bit Rate Communication*, ITU-T Rec. H.263, Jan. 1998.

[6] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Rec. H.264|ISO/IEC 14496-10 AVC, 2003.

[7] Joint Model Reference Software Version 8.4. JVT of ISO/IEC MPEG and ITU-T VCEG.

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

[9] M. R. Pickering and J. F. Arnold, "A perceptually efficient VBR rate control algorithm," *IEEE Trans. Image Process.*, vol. 3, pp. 527–532, Sep. 1994.

[10] T. V. Lakshman, A. Ortega, and A. R. Reibman, "VBR video: Tradeoffs and potentials," *Proc. IEEE*, vol. 86, no. 5, pp. 952–972, May 1998.

[11] Y. Yu, J. Zhou, Y. Wang, and C. W. Chen, "A novel two pass VBR coding algorithm for fixed-size storage application," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 345–356, Mar. 2001.

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

[13] H. B. Yin, X. Z. Fang, L. Chen, and J. Hou, "A practical consistent-quality two-pass VBR video coding algorithm for digital storage application," *IEEE Trans. Consum. Electron.*, vol. 50, no. 4, pp. 1142–1150, Nov. 2004.

[14] D. Bagni, B. Biffi, and R. Ramalho, "A constant-quality, single-pass VBR control for DVD recorders," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 653–662, Aug. 2003.

[15] A. Jagmohan and K. Ratakonda, "MPEG-4 one-pass VBR rate control for digital storage," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 447–452, 2003.

[16] S. Kondo and H. Fukuda, "A real-time variable bit rate MPEG2 video coding method for digital storage media," *IEEE Trans. Consum. Electron.*, vol. 43, no. 3, pp. 537–543, Aug. 1997.

[17] Y. J. Liang, H. Wang, and L. El-Maleh, "Design and implementation of content-adaptive background skipping for wireless video," in *Proc. IEEE Int. Symp. Circuits Syst.*, Kos, Greece, May 2006, pp. 2865–2868.

[18] Y.-L. Lin, S.-F. Lin, Y.-F. Hsu, and H. H. Chen, "Improving the coding of regions of interest," in *IEEE Int. Symp. Circuits Syst.*, Kos, Greece, May 2006, pp. 4313–4316.

[19] Z. G. Li *et al.*, "Adaptive rate control with HRD consideration," presented at the 8th JVT Meeting, Geneva, Switzerland, May 2003, JVT-H014.

[20] J. Yang, Q. Dai, W. Xu, and R. Ding, "A rate control algorithm for MPEG-2 to H.264 real-time transcoding," in *Proc. SPIE Vis.Commun. Image Process.*, Beijing, China, Jul. 2005, pp. 1995–2003.

[21] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[22] Y. Yokoyama and Y. Ooi, "A scene-adaptive one-pass variable bit rate video coding method for storage media," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1999, vol. 3, pp. 827–831.

[23] B. C. Song and K. W. Chun, "A one-pass variable bit rate video coding for storage media," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 689–692, Aug. 2003.

[24] H. Schwarz, D. Marpe, and T. Wiegand, *Subband Extension of H.264/AVC,* 2004, ISO/IEC JTC1/SC29/WG11 MPEG04/M10569/S03.

[25] J.-R. Ohm, *Complexity and Delay Analysis of MCTF Interframe Wavelet Structures,* 2002, ISO/IEC JTC1/WG11 Doc. M8520.

[26] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," in *Proc. PCS*, Apr. 2003, pp. 59–62.

[27] L. Xu, S. Ma, D. Zhao, and W. Gao, "Rate control for scalable video model," in *Proc. SPIE Vis. Commun. Image Process.*, Jul. 2005, vol. 5960, pp. 525–534.

Ping-Hao Wu received the B.S. degree in electrical engineering and the M.S. degree in communication engineering from National Taiwan University, Taiwan, R.O.C., in 2004 and 2006, respectively.

He is currently a graduate student in the Department of Electrical Engineering, University of Southern California, Los Angeles. His research interests include digital signal processing and video coding.

Homer H. Chen (S'83–M'86–SM'01–F'03) received the Ph.D. degree in electrical and computer engineering from University of Illinois at Urbana-Champaign.

Since August 2003, he has been with the College of Electrical Engineering and Computer Science, National Taiwan University, Taiwan, R.O.C., where he is Irving T. Ho Chair Professor. Prior to that, he had held various research and development management and engineering positions in leading US companies including AT&T Bell Labs, Rockwell Science Center, iVast, and Digital Island over a period of 17 years. He was a US delegate of the ISO and ITU standards committees and contributed to the development of many new interactive multimedia technologies that are now part of the MPEG-4 and JPEG-2000 standards. His research interests lie in the broad area of multimedia processing and communications.

Dr. Chen is an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He served as Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING from 1992 to 1994, Guest Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 1999, and Editorial Board Member for *Pattern Recognition* from 1989 to 1999.