Delay Constrained Spatio-temporal Video Rate Control for Time-Varying Rate Channels

Myeong-jin Lee[1] and Dong-jun Lee[2]

[1] Dept. of Electrical Engineering, Kyungsung University, Busan, 608-736, Korea
mjlee@ieee.org
[2] School of Electronics, Telecomm. and Computer Engineering, Hankuk Aviation
University, Gyeonggi, 412-791, Korea

Abstract. In this paper, we propose a delay constrained spatio-temporal video rate control method for lossy channels where the effective channel rate available to the video encoder is time-varying. Target bit-rate constraint for encoding is derived, which guarantees in-time delivery of video frames. By using empirically obtained rate-quantization and distortion-quantization relations of video, the distortions of skipped and coded frames in near future can be calculated in real-time. For the window expanding from the current to the firstly coded frame including the skipped frames in between, the number of frames to skip from the current and the quantization parameter(QP) for the firstly coded frame are decided in the direction to minimize the average distortion of frames in the window. From the simulation results, the proposed method is shown to enhance the average PSNR performance compared to TMN8 with some increase in the number of skipped frames and less number of delay violations.

1 Introduction

Most recently, with the increasing demand of video services over the Internet and the wireless networks, adaptive video transmission over the lossy channels has been the main focus of the research. Because there exists inevitable packet losses and bit errors in the channels, video encoders adopt error protection, recovery, and concealment mechanisms, which may require overhead at the expense of some quality degradation. Packet-loss and bit-error ratios generally vary over time, and they cause the effective channel rate available to the video encoder to be time-varying. Thus, video encoders should adjust encoding parameters under the end-to-end delay constraint continuously sensing the time-varying characteristics of channels.

Rate control plays an important role in the video encoder, which may have great effect on the channel adaptability and the perceived quality. There have been many research works on the rate control under the fixed frame rate or spatial resolution. However, because the coding complexity of video source and the effective channel rate available to the encoder are time-varying, the overall distortion of video cannot be minimized by controlling just one encoding parameter, i.e. a QP or a frame rate.

Recently, rate control algorithms[5, 6, 7] considered both temporal and spatial qualities jointly. However, additional buffering delay for pre-analysis of video source or the method of dynamic programming for optimal solving[6, 7] would not be applicable to real-time applications such as video phones and video conferences. Though the spatio-temporal optimization problem was simplified in [5] by using the explicit distortion models for coded and skipped frames, the distortion for skipped frames was not accurate enough and the end-to-end delay constraint was not considered.

In this paper, we propose a delay constrained spatio-temporal video rate control method which enables video encoders to efficiently adapt to the time-varying effective channel rate. In section 2, we discuss the delay constraint in video transmission systems and derive a constraint on the target bit-rate for encoding. In section 3, by using empirically obtained rate-quantization and quantization-distortion relations of video, the distortion models for skipped and coded frames in near future are proposed. In section 4 and 5, a problem is formulated and a real-time algorithm is presented for delay constrained spatio-temporal video rate control. For the window expanding from the current to the firstly coded frame including the skipped frames in between, the number of frames to skip from the current and the QP for the firstly coded frame are decided in the direction to minimize the average distortion of frames in the window. Simulation results and conclusion are presented in section 6 and 7, respectively.

2   Delay Constraint in Video Transmission Systems

For lossy channels, as shown in Fig. 1(a), joint source and channel coding is generally used to minimize the overall distortion by controlling source and channel coding parameters[3]. The effective channel rate available to the video encoder is time-varying because the rate allocation for the channel coding is done based on the time-varying characteristics of bit errors or packet losses. In this paper, we focus on the spatio-temporal video rate control method which can adapt to the time-varying effective channel rate. Joint optimization of source and channel coding parameters is not considered and left for further study.

Fig. 1(b) shows the video transmission system considered for spatio-temporal rate control. The encoder buffer is served with the effective channel rate mentioned above. We do not directly consider the time-varying characteristics of the channel, but only the time-varying effective channel rate available to the encoder.

Then, the encoder buffer occupancy is given by

$$B^e(j) = \max\left\{B^e(j-1) + e(j, q_j) - s(j), 0\right\},\qquad(1)$$

where $B^e(j)$, $e(j, q_j)$, and $s(j)$ are the encoder buffer occupancy, the generated bit-rate for the jth frame with the QP q_j, and the effective channel rate. The encoder buffer size is assumed to be sufficiently large.

The actual encoder buffer service rate is given by

$$\tilde{s}(j) = \min\left\{s(j), B^e(j-1) + e(j, q_j)\right\}.\qquad(2)$$

(a) video transmission system for lossy channels

(b) video transmission system: source encoder's side view
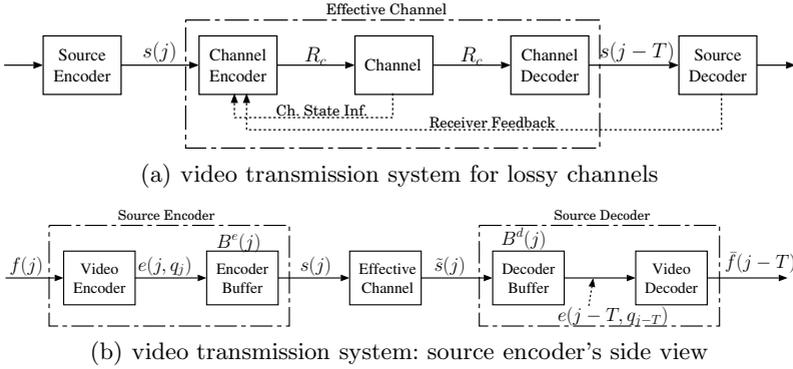
Fig. 1. Video communication system

Then, the decoder buffer occupancy is given by

$$B^d(j) = \begin{cases} \sum_{i=1}^{j} \tilde{s}(i) - \sum_{i=1}^{j-T} e(i, q_i), & \text{if } j \geq T \\ \sum_{i=1}^{j} \tilde{s}(i), & \text{if } j < T \end{cases} \tag{3}$$

where T is the end-to-end delay between the input instances to the encoder buffer and to the decoder. For $j \geq T$, the decoder buffer occupancy can be modified as

$$B^d(j) = \sum_{i=j-T+1}^{j} \tilde{s}(i) - B^e(j - T), \tag{4}$$

where the encoder buffer occupancy is also represented by $B^e(j) = \sum_{i=1}^{j} \{e(i, q_i) - \tilde{s}(i)\}$ using Eq. 1 and 2. By applying the decoder buffer underflow condition $B^d(j) \geq 0$ to Eq. 4, we obtain the condition of $B^e(j) \leq \sum_{i=j+1}^{j+T} \tilde{s}(i)$. For buffered data, because the encoder transmits data in its maximum channel rate $s(j)$, we obtain $\tilde{s}(j) = s(j)$. Finally, the decoder buffer underflow constraint is given by

$$B^e(j) \leq \sum_{i=j+1}^{j+T} s(i), \tag{5}$$

for $j \geq T$. Eq. 5 means that the last frame $e(j, q_j)$ entering the encoder buffer during the jth frame period should leave the encoder buffer no later than the $(j + T)^{th}$ frame time.

For the case of the bounded network delay, the time left for the encoded data is decreased by the amount of the maximum transfer delay. Then, by combining Eq. 1 and 5, the constraint on the target bit-rate for encoding is given by

$$e(j, q_j) \leq \sum_{i=j}^{j+T_e} s(i) - B^e(j - 1). \tag{6}$$

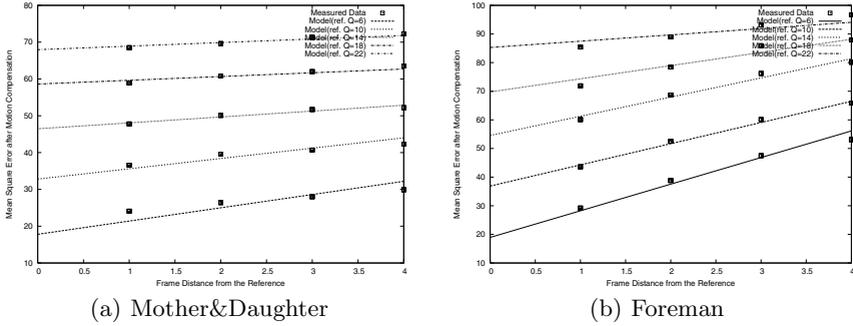where T_e is the time limit of video frames allowed to stay in the encoder buffer.

Fig. 2. Variance of the residual image after motion compensation($\sigma^2_{MC,j}$). The 100th frame is selected as the reference. QP's of 6, 10, 14, 18, and 22 are used for the reference.

3   Distortion Models for Coded and Skipped Frames

3.1   Distortion Model for Coded Frames

The distortion of the coded frame generally depends on the QP of the reference and the distance from it. Generally, QP's vary over macroblocks in a frame depending on rate control algorithms. But, it is difficult to find distortion models using the average QP because the distortion or generated bit-rate may also differ for the same average QP. Thus, in this paper, all the macroblocks in a frame are assumed to be quantized with the same QP. The distortion generally increases as the QP increases. As the distance from the reference increases, the variance of the residual image after motion compensation also increases. It is because the prediction efficiency gets lower as the distance from the reference increases.

The variance of the residual image depends on the coded distortion of the reference and the distance from it. Thus, the variance of the residual image for the jth frame, which is predicted from the j_c^{th} frame, could be modeled as

$$\sigma^2_{MC,j} = D_c(q_{j_c}) + \alpha \cdot (j - j_c), \qquad (7)$$

where $D_c(q_{j_c})$ and α are the coded distortion of the reference and a constant parameter. An index j_c and a QP q_{j_c} are used for the reference frame.

Fig. 2 shows the relation of $\sigma^2_{MC,j}$ with the coded distortion of the reference and the distance from it. Though the proposed model is simple, it fits well with the measured data. The model can be used to predict the variance of the residual image without motion compensation and the result can be used to estimate the coded distortion of future frames without encoding.

To decide whether to encode or to skip a frame, the distortion model for coded frames should be investigated. There have been so many studies on the rate distortion relation for video coding. For real-time video transmission systems such as videophones, it is needed to estimate the coded distortion of incoming frames and to decide coding parameters fast, i.e. QP's, frame skip, and etc., in the
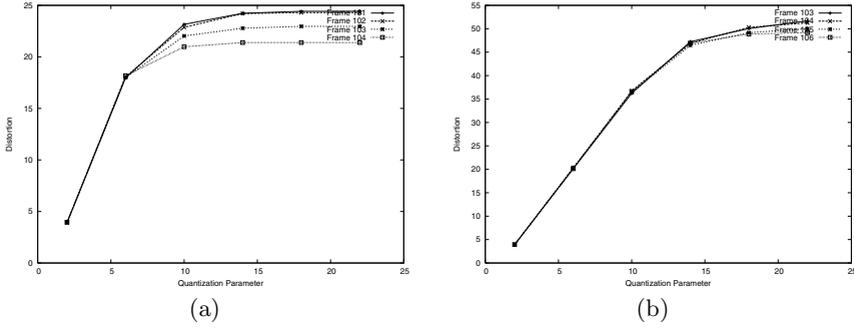
Fig. 3. Distortion of coded frames over QP's. *Mother&Daughter*. (a) Distance from the ref.=1, QP of the ref.=6 (b) Distance from the ref.=3, QP of the ref.=14.

direction to maximize the decoded video quality. Also, for wireless applications, the computational cost should be kept low for the longer battery life of mobile devices. Thus, it is needed to estimate the coded distortion fast enough with less computation.

Fig. 3 shows the distortion of coded frames over QP's for different combinations of reference distortions and the distances from the refence. While the coded distortion is linear for QP's below a threshold, it saturates to the variance of the residual image for QP's above the threshold. The threshold value generally depends on the QP of the reference.

Then, the distortion of coded frames can be estimated by

$$D_c(q_j) = \min\{c_j \cdot q_j, \sigma_{MC,j}^2\}, \tag{8}$$

where q_j and c_j are the QP and the rate-distortion parameter for the jth frame, respectively. The effect of the distance on the distortion is already considered in the variance model of the residual image.

For bit-rate estimation, the quadratic rate-quantization model[4] is used.

$$e(q_j) = a_j \cdot q_j^{-1} + b_j \cdot q_j^{-2} \tag{9}$$

where a_j and b_j are the model parameters.

For real-time applications, it is hard to get two rate-quantization points for parameter calculation. Though two rate-quantization points from adjacent frames could be used for the calculation, the difference between the QP's is not large enough for the model to cover the wider range of QP's. Also, changes in the characteristics of video source sometimes makes it difficult to get correct parameters.

By assuming a virtual rate-quantization point, q_v and $e(q_v)$, the model parameters can be calculated by using a single set of the average QP and the generated bit-rate of the recently coded frame. For real-time applications, the virtual QP q_v is set to a value outside the real range of QP's and the corresponding bit-rate $e(q_v)$ is set to a value less than the possible minimum bit-rate of frames.
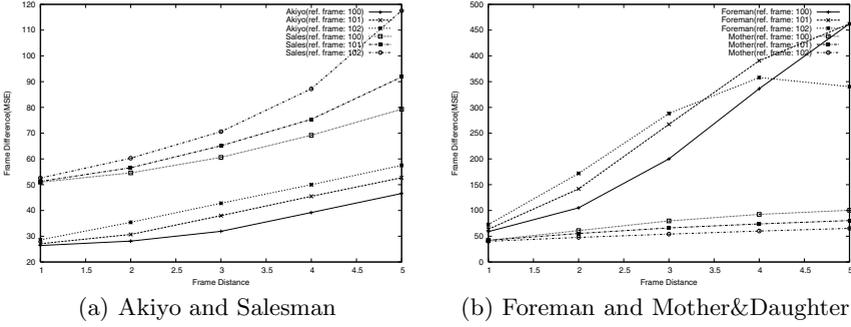
(a) Akiyo and Salesman   (b) Foreman and Mother&Daughter

Fig. 4. MSE of the source frame difference. Reference frame(j_c): $100 \sim 102^{th}$ frame.

3.2 Distortion Model for Skipped Frames

For skipped frames, if there is no frame interpolation or no other quality en-
hancing techniques used, the recently decoded frame is shown at their decoding
times. Vetro analyzed the distortion for future skipped frames in [5] and it is
given by

$$D_s(j_c, j) = D_c(q_{j_c}) + E\{\Delta^2 z_{j_c, j}\} \tag{10}$$

where j_c and j are the indexes of the recently coded frame and the future skipped
frame, respectively. $\Delta z_{j_c, j}$ represents the source frame difference between the j_c^{th}
and the jth frames.

To calculate the source frame difference, it is needed to predict the future
skipped frame correctly. In [5], Vetro assumed that all pixels in the skipped frame
have corresponding motion vectors and the motion vectors between frames are
linear. Based on the assumption, the pixels in the skipped frame is predicted
from the recently coded frame. However, it cannot be applied to real applica-
tions because the motion is not linear and the motion vector range is sometimes
confined to provide error resilience. Also, the computational cost is high because
the gradient and the motion vector should be calculated for all pixels.

Fig. 4 shows the mean square error(MSE) of the source frame difference with
respect to the different frame distances. For neighboring frames, we argue that
it is not the position of the reference frame, but the frame distance to greatly
affect the MSE of the source frame difference. Thus, the source frame difference
between the lastly coded frame and the future skipped frame can be predicted
using the difference between the previous frames of the same distance. Then, the
distortion for future skipped frames can be given by

$$D_s(j_c, j) = D_c(q_{j_c}) + f_d(j - j_c) \tag{11}$$

where $f_d(i)$ is the MSE of the frame difference between frames of distance i.
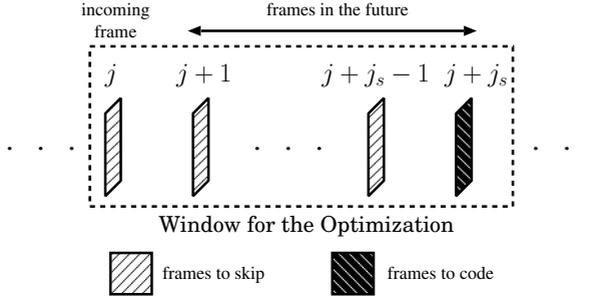
Fig. 5. Window for the optimization of average distortion

4   Problem Formulation for Delay Constrained Spatio-temporal Video Rate Control

Delay constrained spatio-temporal video rate control can be defined as the optimization problem of the average distortion under the delay constraint. The optimization is performed for the frame window shown in Fig. 5, which consists of successive skipped frames from the current and the next coded frame. Then, the problem can be formulated as follows.

$$\min_{j_s, q_{j+j_s}} D_{avg}(j_s, q_{j+j_s}) \tag{12}$$

$$s.t. \ \ e(j + j_s, q_{j+j_s}) \leq \sum_{i=j}^{j+j_s+T_e} s(i) - B^e(j-1)$$

where j_s is the number of frames to skip from the current frame j.

An encoder distortion which includes the distortion for coded and skipped frames can be defined as follows.

$$D_{enc}(j) = \begin{cases} D_c(q_j), & \text{for encoded frames} \\ D_s(j_c, j), & \text{for skipped frames.} \end{cases} \tag{13}$$

Then, the average distortion over the frame window is given by

$$D_{avg}(j_s, q_{j+j_s}) = \frac{1}{j_s + 1} \sum_{i=j}^{j+j_s} D_{enc}(i). \tag{14}$$

Under the assumption of stationary video source, by combining Eq. 10, 13, and 14, the average distortion can be calculated as follows.

$$D_{avg}(j_s, q_{j+j_s}) = D_c(q_{j_c}) + \frac{D_c(q_{j+j_s}) + \sum_{i=j}^{j+j_s-1} f_d(i - j_c)}{j_s + 1} \tag{15}$$

5 Delay Constrained Spatio-temporal Video Rate Control Algorithm

5.1 Frame-Level Rate Control

Given the ranges of the number of frames to skip and the QP, the average distortion is calculated using the distortion models in Section 3. The rate controller determines a set of parameters and the corresponding target bit-rate which minimizes the average distortion.

Because the proposed distortion and rate-quantization models are based on the previous results of encoding and frame analysis, it is needed to update the model parameters frequently. Also, due to the time-varying characteristics of video source, the decided number of frames to skip(j_s^{min}) would not be accurate enough. Thus, only the current frame is skipped if j_s^{min} is larger than zero. Otherwise, it is encoded with the QP $q_{j_s}^{min}$. The frame-level rate control is performed in frame-by-frame basis.

5.2 Macroblock-Level Rate Control

If the current frame is encoded, the rate controller decides a QP for each macroblock. Basically, the QP decided in the proposed frame-level rate control is used as a lower bound for the quantization. Considering the time-varying characteristics of video, it is needed to regulate the bit-rate to meet the target. For this purpose, the QP in TMN8(QP_i^{TMN8}) is used as a reference and it is further adjusted considering the encoder buffer occupancy.

The encoder buffer occupancy is updated in macroblock level with the drain rate of the average bit-rate for a macroblock. For the jth frame, the encoder buffer occupancy after encoding the ith macroblock is given by

$$B_{MB}^e(i) = \max\{0, B_{MB}^e(i-1) + e_{MB}(i) - \frac{s(j)}{N}\} \tag{16}$$

where $e_{MB}(i)$ and N are the generated bits for the ith macroblock and the number of macroblocks in a frame, respectively, and $B_{MB}^e(0) = B^e(j-1)$.

The QP for the ith macroblock is given by

$$QP_i = \begin{cases} \max\{q_{j+j_s}^{min}, \min\{QP_i^{TMN8}, q_{j+j_s}^{min} + q_\Delta\}\} & ,\text{if } \frac{B_{MB}^e(i-1)}{s(j)} > (T_e - 2) \\ QP_i^{TMN8} & ,\text{otherwise} \end{cases} \tag{17}$$

where q_Δ is 2, 4, or 6 for the cases of $(T_e - 2) < \frac{B_{MB}^e(i-1)}{s(j)} < (T_e - 1.5)$, $(T_e - 1.5) < \frac{B_{MB}^e(i-1)}{s(j)} < (T_e - 1)$, or $\frac{B_{MB}^e(i-1)}{s(j)} > (T_e - 1)$, respectively.

Algorithm. Delay Constrained Spatio-temporal Video Rate Control

- *Step 0:* **Initialization**
    - $B^e(0) = 0, j_c = 0, j_s^{min} = 0, q_{j+j_s}^{min} = 0, a_j = a_0, b_j = b_0, c_j = c_0$

- *Step 1:* **Frame-Level Rate Control**
  - Given ranges of $j_s(j_s = 0, \ldots, j_{max})$ and $q_{j+j_s}(|q_{j+j_s} - q_{j_c}| \leq \Delta q)$
    * calculate the estimated bit-rate $e(j + j_s, q_{j+j_s})$ using Eq. 9
    * if $e(q_{j+j_s})$ satisfies the delay constraint and the average distortion(D_{avg}) is less than D_{avg}^{min}, update the parameter set of $\{j_s^{min}, q_{j+j_s}^{min}, D_{avg}^{min}\}$
  - If j_s^{min} is larger than 1, skip current frame and go to *Step 3*
  - else, go to *Step 2*
- *Step 2:* **Macroblock-Level Rate Control**
  - Update the encoder buffer occupancy using Eq. 16 in MB-by-MB basis
  - Quantization parameter decision for each macroblock
    * Calculate new QP_i^{TMN8} for the ith MB according to TMN8
    * Adjust the QP according to the buffer occupancy(Eq. 17)
  - Encode the ith macroblock with QP_i
  - After encoding all the macroblocks in current frame, $j_c = j$ and goto *Step 3*
- *Step 3:* **Parameter Update**
  - Update the encoder buffer occupancy $B^e(j)$
  - Update encoding parameters and distortion model parameters: a_j, b_j, c_j, q_{j_c}
  - Frame index increment: $j = j + 1$
  - go to *Step 1*

6   Simulation Results

6.1   Simulation Environment

In this section, we present some experimental results that demonstrate the performance of the proposed spatio-temporal video rate control algorithm. The simulations are based on the H.263+ codec[1]. We used three video sequences(*Akiyo, Salesman, Mother&Daughter*), all in QCIF format with frame rate of 30 fps. To evaluate the performance for different channel rates and delay constraints, we consider various combinations of channel rates($s(j) \in \{32, 64, var\}$ kbps) and encoder buffering limits($T_e \in \{2, 3, 4, 5\}$). To show how the proposed algorithm responds to time-varying channel status, the channel rate in Fig. 7 is used for simulation and will be refered as '*var*'. The time-varying bit error ratio in wireless channels with constant bit-rate(CBR) may cause the effective channel rate available to the encoder to vary over time as in Fig.7.

The performance of the proposed algorithm is compared with that of the TMN8[2] rate control algorithm. For the performance comparison of the algorithms, we used two measures of video quality. The first is the average PSNR to evaluate the spatial quality of video. For skipped frames, the PSNR is calculated by considering the recently decoded frame as the decoded frame. The PSNR of the skipped frames may be increased if the frame interpolation is applied. The second is the number of decoder buffer underflows to check how many frames have arrived at the decoder in time. The maximum deviation of the QP(q_Δ) is set to 3 to prevent abrupt quality variation. The larger the values, the less number of skipped frames, the larger quality fluctuation, and the larger probability of decoder underflow are expected.
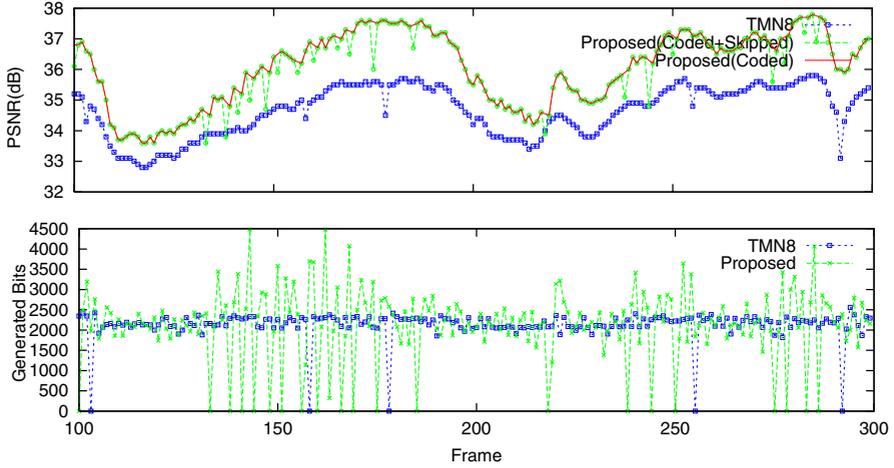
Fig. 6. Performance of the proposed algorithm. *Salesman*($T_e = 3$, $s(j) = 64$kbps).

6.2   Performance for the Constant Channel Rates

Fig. 6 plots the encoding results for *Salesman*. For the skipped frames, the generated bit is zero. The proposed algorithm shows enhanced PSNR performance with more frame skips over the sequence. There is somewhat larger PSNR fluctuation in the proposed, which is caused from the relative low PSNR of the skipped frames.

Table 1 shows the performance for *Akiyo* for various combinations of channel rates and the encoder buffering limits. As the channel rate or the encoder buffering limit increases, the average PSNR increases in the proposed algorithm. Because the TMN8 does not consider the end-to-end delay constraint for rate control, the average PSNR of TMN8 does not change for different encoder buffering limits. The number of skipped frames of the proposed is larger than TMN8. The number of decoder buffer underflows in TMN8 is generally larger than that in the proposed.

Compared with the TMN8, the proposed algorithm shows enhanced performance in the average PSNR and the number of decoder buffer underflows for different combinations of channel rates and encoder buffering limits. It is because the proposed algorithm determines the frame skip and the QP based on the distortion models for the coded and the skipped frames in the direction to minimize the average distortion within a window. The saved bits in the skipped frames are allocated to the next coded frames, which decreases the distortion of them. Then, the distortions of the following coded and skipped frames are recursively decreased because the distortion of the coded and skipped frames depends on that of the reference as in Eq. 7, 8, 9. The PSNR variation due to the skipped frames is somewhat large in the proposed algorithm. However, the PSNR of the skipped frames is generally larger than that of the corresponding frames in TMN8. Thus, we argue that the perceived quality of the proposed
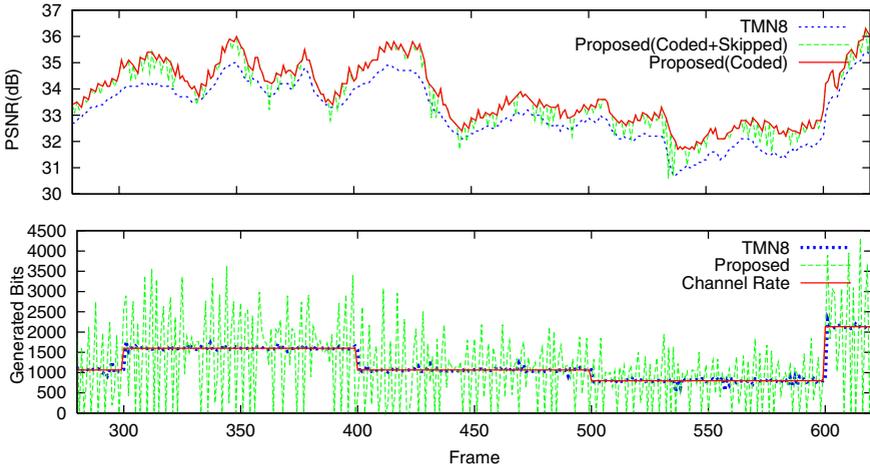
Fig. 7. Performance for the time-varying channel rate. *Mother&Daughter*($T_e = 3$).

Table 1. Performance of the proposed rate control method. *Akiyo*, 300 frames.

$s(j)$	T_e	Proposed			TMN8		
		skip	PSNR(dB)	dec. under.	skip	PSNR(dB)	dec. under.
32	2	65	35.4	0	7	35.5	7
32	3	63	36.3	0	7	35.5	6
32	4	67	36.2	0	7	35.5	5
64	2	52	38.6	0	4	39.0	4
64	3	40	39.7	0	4	39.0	2
64	4	45	39.8	0	4	39.0	1
var	2	61	37.9	0	4	38.3	7
var	3	48	39.0	0	4	38.3	3
var	4	51	39.1	1	4	38.3	1

algorithm is better than that of the TMN8. Also, the PSNR variation can be decreased by interpolating the skipped frames from the neighboring frames.

6.3   Performance for the Time-Varying Channel Rate

Fig. 7 shows the performance of the proposed algorithm for the time-varying effective channel rate. There are 0.4 dB enhancement over the TMN8 in the average PSNR for 900 frames. The number of the delay violations are 1 and 2 in the proposed and the TMN8 algorithms, respectively. There exist some delay violations in the proposed algorithm due to the abrupt changes in the channel rate. Because the number of delay violations depends on both the rate control method and the channel statistics, it is impossible to completely remove them. Thus, it is needed to jointly consider the source and the channel coding

to minimize the overall distortion by source coding and channel errors, which is left for further study.

7   Conclusion

In this paper, we proposed a delay constrained spatio-temporal video rate control method. Target bit-rate constraint for encoding is derived, which guarantees in-time delivery of video frames. By using empirically obtained rate-quantization and quantization-distortion relations of video, the distortion models for skipped and coded frames in near future are proposed. The number of frames to skip from the current and the QP for the firstly coded frame are decided in the direction to minimize the average distortion of the frames. From the simulation results, it is shown that the proposed algorithm enhances the average PSNR performance compared to TMN8 with some increase in the number of skipped frames and less number of delay violations. The results can be utilized for video codecs based on motion-compensation and DCT, e.g. H.26X and MPEG-4, to adapt to the time-varying channel such as wireless networks and the Internet where the effective channel rate available to the encoder changes over time. For further works, it is needed to jointly control the spatio-temporal parameters and the code rate in channel codecs, which requires the video packetization, distortion and error propagation model for channel errors, and etc.

Acknowledgements

References

1. ITU-T Recommendation H.263, Version 2, 1998.
2. ITU-T/SG15, Video Codec Test Model, Near-Term, Version 8(TMN8), Portland, June 1997.
3. K. Stuhlmuller, N. Faber, M. Link, and B. Girod, *Analysis of Video Transmission over Lossy Channels*, *IEEE J. Select. Areas Commun.*, Vol. 18, No. 6, pp. 1012-1032, 2000.
4. T. Chiang and Y.-Q. Zhang, *A New Rate Control Scheme Using Quadratic Rate Distortion Model*, *IEEE Trans. Circuit Syst. Video Technol.*, Vol. 7, No. 1, pp. 246-250, 1997.
5. A. Vetro, Y. Wang, and H. Sun, *Estimating Distortion of Coded and Non-Coded Frames for Frameskip-Optimized Video Coding*, *IEEE ICME*, pp. 541-544, 2001.
6. S. Liu and C.-C. J. Kuo, *Joint Temporal-Spatial Bit Allocation for Video Coding with Dependency*, *IEEE Trans. Circuit Syst. Video Technol.*, Vol. 15, No. 1, pp. 15-26, 2005.
7. R. C. Reed and J. S. Lim, *Optimal Multidimensional Bit-Rate Control for Video Communication*, *IEEE Trans. Image Processing*, Vol. 11, No. 8, pp. 873-885, 2002.