Distortion Variation Minimization in low-bit-rate Video Communication

Sang-Hyun Park, *Member, KIMICS*

***Abstract*—**A real-time frame-layer rate control algorithm with a token bucket traffic shaper is proposed for distortion variation minimization. The proposed rate control method uses a non-iterative optimization method for low computational complexity, and performs bit allocation at the frame level to minimize the average distortion over an entire sequence as well as variations in distortion between frames. The proposed algorithm does not produce time delay from encoding, and is suitable for real-time low-complexity video encoder. Experimental results indicate that the proposed control method provides better visual and PSNR performances than the existing rate control method.

***Index Terms*—**Constant video quality, Distortion variation minimization, Rate control, Video streaming.

I. INTRODUCTION

Video communication over the Internet has become popular due to the fast development of networking and compression technologies. The digital video compression technique plays an important role in development of an audiovisual communication system. To transmit compressed video efficiently over the Internet, we should consider both the underlying video content and channel conditions, and develop an effective rate control scheme accordingly.

Current Internet Protocol (IP) provides a service that can be characterized as a best-effort service. The best-effort model is sufficient for non-real-time video or non-real-time data communication applications like web browsing, file transfer, remote login, and electronic mail. In order to provide QoS guarantee to real-time applications, the IETF has defined the integrated services (IntServ) and the differentiated services (DiffServ). The IETF has specified a token bucket algorithm as a traffic policing mechanism in the IntServ [1]. In the case of the DiffServ, the token bucket algorithm is also one of the most commonly used methods in edge routers. In the DiffServ network, a service subscriber first sets up a service profile with an Internet service provider regarding the desired type of service. At the ingress of the DiffServ network, edge routers classify all packets passing through them into several predefined service classes, and mark the packets with different drop precedence according to the subscriber's service profile using the token bucket algorithm [2,3]. Therefore, this work focuses on an effective rate control for low bit rate video over the Internet where video traffic is policed by the token bucket shaper.

Many frame-layer rate control schemes have been proposed in [4-9]. In general, these schemes can be thought of as having a frame and a macroblock layer. The frame-layer rate control assigns a target number of bits to each video frame, and, at a given frame, the block-layer rate control selects the block-quantization parameters to achieve the target number of bits for the frame [5-6]. Some frame-layer rate control approaches use simple formulas, but these simple methods generally do not achieve the target number of bits accurately [7]. Other approaches use various rate-distortion strategies to assign a target number of bits to each frame [8-9]. However, since they usually use either an iteration method for optimal bit allocation or a pre-analysis method on a group of frames before encoding, they produce time delay of high computational complexity. In addition, they are optimized for the constant bit-rate (CBR) networks, therefore they cannot be applied to the QoS-guaranteed networks where traffic flows are regulated by the token bucket.

In this paper, we propose a real-time frame-layer rate control method with the token bucket policing for low bit rate video compression standard. In order to achieve accurate rate control, a new frame-layer rate-distortion (R-D) model is derived. Furthermore, we use a non-iterative method with low computational complexity for real-time rate control. It is seen that the proposed rate control algorithm does not produce time delay from encoding.

The paper is organized as follows. In the next section, the proposed frame-layer rate control scheme is presented. Section 3 presents and discusses the experimental results. Finally, our conclusions are given in Section 4.

II. PROPOSED RATE CONTROL METHOD

For frame-layer rate control, the rate and the distortion of the current frame are first estimated as a function of the average quantization parameter (QP). Then, the

---

Manuscript received February 12, 2007.

Sang-Hyun Park, Department of Multimedia Engineering, Sunchon National University, 315 Maegok, Sunchon, Jeonnam 540-742, Korea. (email: shark@scnu.ac.kr)

target distortion for the current frame is determined according to the parameters of the token bucket. Finally, the optimization problem with the bit budget constraint is solved.

A. Estimation of Rate and Distortion

To estimate the rate and the distortion of each frame, an empirical data-based frame-layer R-D model is employed using the quadratic rate model and the affine distortion model [10] with respect to the average QP in a frame, which is given by

$$\hat{R}(\overline{q}_i) = (a \cdot \overline{q}_i^{\,-1} + b \cdot \overline{q}_i^{\,-2}) \cdot MAD(\hat{f}_{ref}, f_{cur}), \tag{1}$$

$$\hat{D}(\overline{q}_i) = a' \cdot \overline{q}_i + b', \tag{2}$$

where a, b, a', and b' are the model coefficients, $\overline{q}_i$ is the average QP of all macroblocks in the *i*th frame, and $\hat{R}(\overline{q}_i)$ and $\hat{D}(\overline{q}_i)$ are the estimated rate and the estimated distortion of the *i*th frame, respectively. Here, $\hat{f}_{ref}$ is the reconstructed reference frame at the previous time instant, $\hat{f}_{cur}$ is the uncompressed image at the current time instant, and $MAD(\cdot,\cdot)$ is the mean of absolute difference between two frames. The model coefficients are determined by using the linear regression analysis and the formula considering of the previous encoding results.

To determine the model coefficients, the previous approaches used a jumping window method for the analysis of scene characteristics [11]. However, the jumping window scheme requires a pre-analysis process for analyzing a group of frames before encoding, and thus causes an additional delay. A sliding window method is introduced that can analyze scene characteristics without time delay. Fig. 1 shows the sliding window method where the window moves one frame at a time to determine the coefficients for each frame. P_i is the current sliding window consisting of frames $\{f_{i-L+1}, f_{i-L+2}, \cdots, f_{i-1}, f_i\}$, where L is the number of frames within the sliding window.

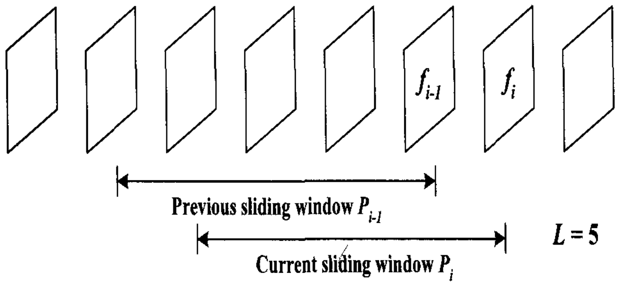


Fig. 1 Concept of the proposed sliding window method.

B. Determination of the Target Distortion

At the encoding time of the *i*th frame, the target distortion D_i^T is determined considering buffer overflow and underflow. In the proposed scheme, rate allocation is performed accordingly to achieve the constant video quality since token bucket policing allows the output rate to vary depending on the state of the token bucket.

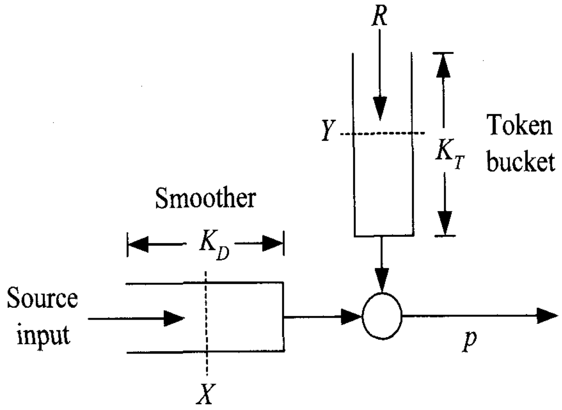


Fig. 2 Token bucket traffic shaper.

To manage the token bucket, we use a virtual buffer random variable method. Fig. 2 shows the token bucket shaper where K_D is the smoothing buffer size, K_T is the token bucket size, p is the peak rate, and R is the token generation rate which is equivalent to the channel rate of the CBR network. Two random variables X and Y represent the occupancy of the smoothing buffer and the token bucket, respectively. Note that the token bucket shaper has following properties: The smoothing buffer can be occupied only if the token bucket is empty. Conversely, the token bucket can be occupied only if the smoothing buffer is empty. We now define the following single virtual buffer random variable.

$$\widetilde{W} \equiv X - Y + K_T \tag{3}$$

Then, $\widetilde{W}$ has the following properties:

$$\begin{aligned} X = 0, \quad \widetilde{W} = K_T - Y, \quad \mathit{if}\ 0 \le \widetilde{W} \le K_T \\ Y = 0, \quad \widetilde{W} = X + K_T, \quad \mathit{if}\ K_T < \widetilde{W} < K \end{aligned} \tag{4}$$

where $K = K_D + K_T$.

To preserve smooth video quality, we define the target distortion D_i^T at the encoding time of the *i*th frame as

$$D_i^T = \begin{cases} \alpha \cdot D_{i-1}^T, & \mathit{if}\ \widetilde{W} > \alpha' \cdot K \\ \beta \cdot D_{i-1}^T, & \mathit{if}\ \widetilde{W} < \beta' \cdot K \\ D_{i-1}^T, & \mathit{otherwise} \end{cases} \tag{5}$$

where α' and β' are parameters used for the safety margin to avoid the buffer overflow and underflow, respectively. To prevent the buffer overflow, D_i^T has a greater value than the previous target distortion. That is, the parameter α is greater than one. On the other hand, to prevent the buffer underflow, D_i^T has a smaller value than the previous target distortion. That is, the parameter β is smaller than one. If $\widetilde{W}$ is not in the safety margin, D_i^T preserves the same target distortion value of the previous frame. In our experiment, α, β, α', and β' are set to 1.1, 0.9, 0.9 and 0.1, respectively.

C. Optimization Process

We consider a new formulation of frame-layer rate control based on the R-D model as follows: Determine $\bar{q}_i$, $i = 1, 2, \ldots, L$ to minimize

$$\sum_{i=1}^{L} | \hat{D}_i(\bar{q}_i) - D_i^T | \quad \text{subject to } \beta' \cdot K \le \widetilde{W} \le \alpha' \cdot K. \tag{6}$$

This formulation minimizes the average distortion variation over the sliding window. Since the encoder buffer allows limited bit rate fluctuation without buffer overflow or underflow, the optimization for the minimum distortion variation is a local optimization and is formulated as

$$\bar{q}_i^* = \arg\min_{\bar{q}_i} | \hat{D}_i(\bar{q}_i) - D_i^T | \quad \text{subject to } \beta' \cdot K \le \widetilde{W} \le \alpha' \cdot K. \tag{7}$$

The proposed rate control algorithm does not produce encoding time delay. However, a negligible performance loss due to its intrinsic sub-optimality is inevitable in our design. Using the R-D curves, we can determine $\bar{q}_i^*$, therefore we do not have to encode the current frame to determine the distortion.

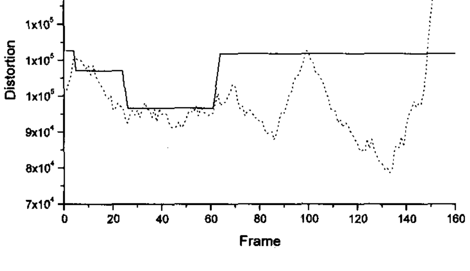


Fig. 3 Comparison between the actual distortion and the target distortion: QCIF Carphone.

In the case of H.263+ standard, what we finally need is not $\bar{q}_i^*$, but $\hat{R}(\bar{q}_i^*)$ which is the target bit budget for the *i*th frame. Once the target bit budget $\hat{R}(\bar{q}_i^*)$ is allocated to the current frame using the aforementioned frame-layer rate control, the TMN8 macroblock layer rate control algorithm allocates the bit budget to each macroblock.

III. EXPERIMENTAL RESULTS AND DISCUSSION

For low bit rate video coding, the delay is a bottleneck since packets arriving later will be discarded resulting in part of the bandwidth being wasted. The encoder buffer is used to monitor the corresponding delay of the generated bit stream to ensure that buffer does not overflow or underflow. Utilizing the encoder buffer, limited bit rate fluctuation is allowed while buffer overflow or underflow does not occur. In this experiment, the smoothing buffer size is set to $5 \cdot R / F$, where F is frame rate in frames per second. In addition, the token bucket size is also set to $5 \cdot R / F$.

To evaluate the performance of the proposed method, the H.263+ video codec is employed, and the proposed method is compared with the MINVAR approach [12]. In the MINVAR approach, the target quantization parameter is determined as

$$\begin{aligned} & V_n(\bar{q}_n) = | D_n(\bar{q}_n) - D_{n-1}(\bar{q}_{n-1}) |, \\ & \bar{q}_n^* = \arg\min_{\bar{q}_n} V_n(\bar{q}_n) \\ & \text{subject to } R_n^l \le \hat{R}(\bar{q}_n^*) \le R_n^u \end{aligned} \tag{8}$$

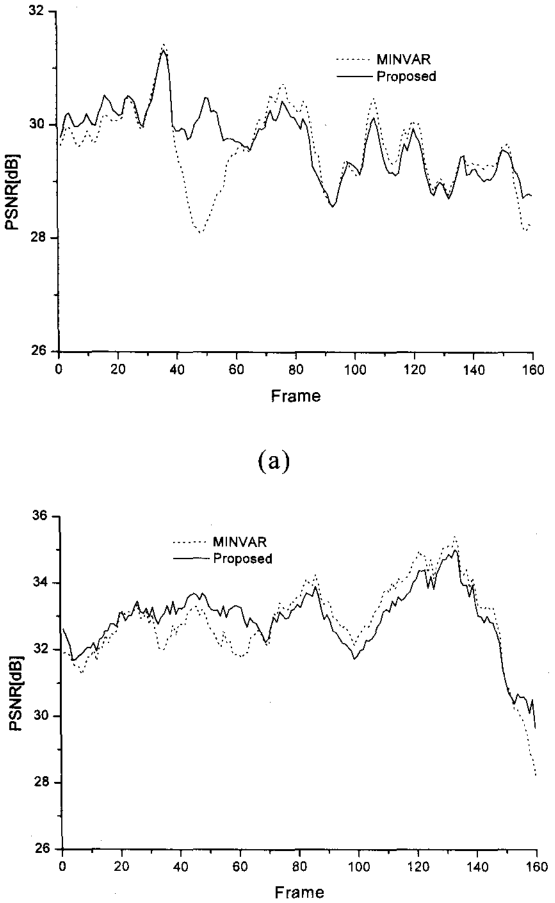
where R_n^l and R_n^u are the lower and upper bounds corresponding to the target bit rate obtained from the buffer feedback.

Extensive experimental testing and comparison were performed on several sequences with different characteristics: Foreman, Carphone, and Silent. These sequences are in QCIF format (176×144) and the frame rate F is 30 fps. The token generation rate R and the size of the sliding window L are set to 64kbps and 12, respectively.

We first present simulation results to validate the proposed method to determine the target distortion D_i^T in (5). Fig. 3 shows the actual distortion and the target distortion for each frame of the Carphone sequence. As shown in this figure, the proposed method works reasonably well. Note that D_i^T preserves the same value for some period since D_i^T has the same value of D_{i-1}^T while $\widetilde{W}$ is less than $\alpha' \cdot K$ and greater than $\beta' \cdot K$. It is also seen that D_i is properly regulated by D_i^T.

Table 1 Performance comparison of the proposed algorithm with MINVAR

Test sequence	Rate control method	Average of PSNR	σ of PSNR	Distortion Variation
Foreman	MINVAR	29.486	0.860	0.172
	Proposed	29.608	0.681	0.139
Carphone	MINVAR	30.951	2.499	0.197
	Proposed	30.977	2.276	0.183
Silent	MINVAR	32.702	0.742	0.140
	Proposed	32.682	0.530	0.110



(a)

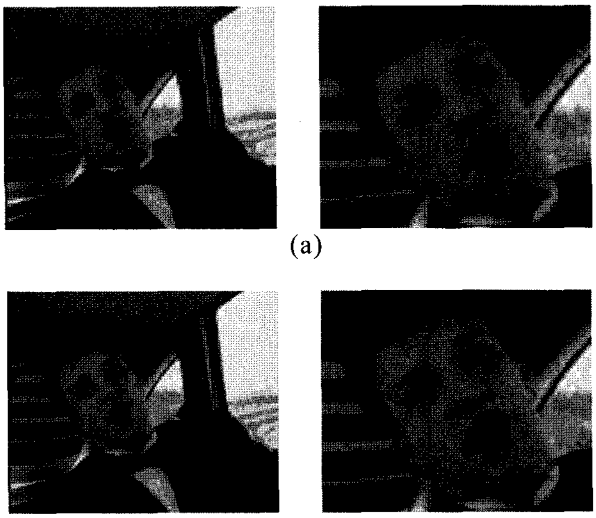
(b)

Fig. 4 PSNR comparison with a token generation rate at 64kbps. (a) QCIF Foreman and (b) QCIF Carphone sequences.

Performance was mainly evaluated by visual judgement since there is no standard measure currently available to evaluate subjective quality. In addition, as an objective measurement of the distance between an original image $f(x,y)$ and its reconstructed image $g(x,y)$, peak-signal-to-noise ratio (PSNR) is used. For a $m \times n$ image with [0,255] gray-level range, PSNR can be defined as

$$PSNR = 10\log_{10}\frac{m\times n\times 255^2}{\sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(f(x,y)-g(x,y))^2}.$$

The performance of the proposed frame-layer rate control scheme is compared with that of MINVAR. For the performance comparison for three test sequences, we show the average PSNR value, the standard deviation (σ) of PSNR, and the distortion variation [12] in Table 1.



(a)

(b)

Fig. 5 Visual quality comparison between proposed algorithm and MINVAR scheme on Carphone sequence. (a) MINVAR, (b) Zoomed image of (a), (c) Proposed algorithm, and (d) Zoomed image of (c)

Note that when compared with MINVAR method the proposed frame rate control algorithm can not only improve the average PSNR value, but also reduce the variation of PSNR. The PSNR plots associated with the Foreman and the Carphone sequences as a function of the frame number are shown in Fig. 4 (a) and (b), respectively. Note that the proposed frame rate control scheme can reduce the quality degradation and the quality variation better than MINVAR method.

Visual comparison of the proposed algorithm with the MINVAR method is also provided in Fig. 5. To make a comparison of the subjective quality more clear, zoomed images are also presented. It is observed in Fig. 5 that edges are well preserved by the proposed technique. Note that the proposed algorithm performs better than MINVAR approach. Similar results were also observed on the other test images. Experimental results indicate that the proposed algorithm is a useful alternative to MINVAR in terms of both the PSNR and subjective quality.

IV. CONCLUSIONS

In this paper, we presented a new real-time frame-layer rate control for low bit rate video over the networks where flows are regulated by a token bucket. In order to reduce the quality fluctuation, the sliding window method has been utilized. In addition, we introduced a frame-layer rate control scheme to minimize the average distortion over an entire sequences as well as variations in distortion between frames with the target distortion method. Since the proposed technique uses fast convergence method and does not require pre-analysis, it is suitable to real-time low-complexity video coding. The proposed algorithm has been tested on several sequences, and found to provide better visual and PSNR performances than the existing MINVAR rate control algorithm.

The propose frame-layer rate control method can be effectively used for the transmission of the low bit rate video over the next generation Internet, such as the DiffServ and the IntServ, where the video traffic is policed by the token bucket traffic shaper.

ACKNOWLEDGMENT

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2006-331-D00550).

REFERENCES

[1] C. Metz, "IP QoS: Traveling in first class on the Internet," *IEEE Internet Computing*, vol 3, pp. 84-88, Mar. 1999.

[2] H. Su and M. Atiquzzaman, "Comprehensive performance model of differentiated service with token bucket marker," *IEE Proceedings-Communications.*, vol. 150, no. 5, pp. 347-353, Oct. 2003.

[3] E. C. Park and C. H. Choi, "Adaptive Token Bucket Algorithm for Fair Bandwidth Allocation in DiffServ Networks," *in Proc. IEEE Globecom'03*, pp. 3176-3180, 2003.

[4] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, "Rate-distortion optimized mode for very low bit rate video coding and emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Tech.*, vol. 6, no. 2, pp. 182-190, Apr. 1996.

[5] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Tech.*, pp. 172-185, Feb. 1999.

[6] ITU-T, *Video codec test model, near term, version 8 (TMN8),* H.263 AdHoc Group. Jun. 1997.

[7] MPEG-2, *Test Model 5*, ISO/IEC/JTC1/SC29 WG11, Apr. 1993.

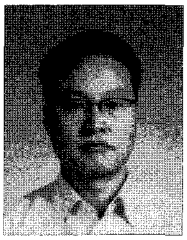
[8] L. J. Lin and A. Ortega, "Bit rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Tech.*, pp. 446-459, Aug. 1998.

[9] Y. Cheng and C. C. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Trans. Circuits Syst. Video Tech.*, pp. 696-702, Aug. 1997.

[10] T. Chiang and Y. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Tech.*, pp. 246-250, Feb. 1997.

[11] Y. Yokoyama and S. Nogaki, "A rate control method with pre-analysis for real-time MPEG-2 video coding," *in Proc. Of Int. Conf. Image Processing*, pp. 514-517, Oct. 2001.

[12] Z. Chen and K. N. Ngan, "Distortion variation minimization in real-time video coding," *Signal Processing: Image Communication*, pp. 273-279, Apr. 2006.



Sang-Hyun Park
Received B.S., M.S., and Ph.D degrees in Electronics Engineering at Korea University in 1995, 1997, and 2002, respectively. In 2004, he joined the Department of Multimedia Engineering at Sunchon National University where he is currently an Assistant Professor. From 2001 to 2003, he was with Hyundai Networks where he was involved in research and development on multimedia communication equipments. His major areas of interest include video signal processing and multimedia communications.