A Novel Long-Span Traffic Predictor for Real-Time VBR Videos via ρ-Domain Rate Model

Chu-Chuan Lee and Pao-Chi Chang, *Member, IEEE*

***Abstract*— To predict the traffic of frames that may be far from the current frame, this letter extends the use of the ρ-domain rate model from macroblock-based source rate control to frame-based long-span traffic prediction. Moreover, this work enhances the linearity and convergence speed of ρ-domain frame-based rate function by adding a parameter that is the number of non-zero motion vectors. Simulation results reveal that the proposed predictor can significantly lower the prediction error compared with two conventional LMS methods. More importantly, the process of the proposed predictor is unique but simple for different video contents and prediction spans.**

***Index Terms*— Real-time videos, traffic prediction, source rate control.**

I. Introduction

VARIABLE Bit Rate (VBR) video is extremely vulnerable against packet losses due to the decoding error propagation. From the viewpoint of resource management, an online traffic smoothing mechanism can effectively increase the bandwidth utilization and thus reduce video packet losses, if sufficient and accurate video traffic prediction results are timely sent to the mechanism. To predict the real-time video traffic, Adas [1] proposed a Least Mean Square (LMS) predictor for forecasting bandwidth demands of future I-, P- and B-frames. To simplify the Rate-Distortion analysis, He and Mitra [2] developed a ρ-domain source rate control scheme for video coding systems, where ρ denotes as the percentage of zeros among quantized coefficients. To increase the bandwidth utilization of networks, Sen *et al*. [3] proposed an online traffic smoothing scheme for delivering real-time streaming videos. From [3] it is obvious that a traffic predictor which only predicts the traffic of the next frame is insufficient in time for efficient online smoothing operations. Besides, regarding the operations of LMS, the step size μ is a very important parameter to the prediction accuracy [1]. However, the optimal μ depends on video contents. In contrast to pre-stored videos, it is difficult to determine timely an optimal μ for a real-time video application. Therefore, this letter proposes a frame-based Long-Span Predictor (ρ-LSP) for real-time videos over networks, where the *long-span* means that predicted frames may be far from the current frame in time. ρ-LSP effectively extends the use of ρ-domain rate model [2] from the macroblock (MB) based source rate control to frame-based traffic prediction. Moreover, this work enhances the linearity and convergence speed of the extended model by adding a parameter that is the number of non-zero motion vectors. The details of ρ-LSP are described as follows.

Manuscript received April 27, 2004. The associate editor coordinating the review of this letter and approving it for publication was Prof. Iakovos Venieris. This work was supported in part by National Science Council under Contract No. NSC93-2213-E-008-015.

C. C. Lee is with the Department of Electrical Engineering, National Central University, Jongli City, Taiwan, R.O.C. (e-mail: chuchu@cht.com.tw).

P. C. Chang is with the Department of Electrical Engineering, National Central University, Jongli City, Taiwan, R.O.C. (e-mail: pcchang@ee.ncu.edu.tw).

Digital Object Identifier 10.1109/LCOMM.2005.03032.

II. ρ - Domain Frame-Based Rate Function

For the sake of explanation, the encoded *j*-th α-frame of the *i*-th GOP (Group of Pictures) of a scene is denoted as F_{ij}^{α}, where α represents one of three video frame types, e.g., I or P or B. Considering the F_{ij}^{α} with encoded bits R_{ij}^{α}, the number of non-zero motion vectors is denoted as M_{ij}^{α} and the number of non-zero quantized coefficients is denoted as Q_{ij}^{α}. Since the coding property of each frame type is different, ρ-LSP executes independent yet identical prediction operations for I-, P- and B-frames.

From [2], involving in the N_c MBs already encoded in the current frame, a linear relationship with an estimated slope θ exists between the bits used to encode these N_c MBs and the number of non-zero quantized coefficients in these N_c MBs. However, when this study directly extends the above linear function from correlated spatial MBs within a frame to correlated temporal inter-frames within a scene, the number of non-zero quantized coefficients of a frame with low rate is usually small. The phenomenon is particularly obvious at B-frames or low picture complexity. In such a situation, the linearity of the extended frame-based rate function may be vulnerable due to insufficient non-zero quantized coefficients. Therefore, this work adds a parameter that is the number of non-zero motion vectors to the extended frame-based rate function. Meanwhile, the modified estimation function of θ for frame prediction is given by

$$\hat{\theta}_{ij}^{\alpha} = \frac{\sum_{h=1}^{i-1}\sum_{m=1}^{C^{\alpha}} R_{hm}^{\alpha} + \sum_{m=1}^{j} R_{im}^{\alpha}}{\sum_{h=1}^{i-1}\sum_{m=1}^{C^{\alpha}} \bar{\rho}_{hm}^{\alpha} + \sum_{m=1}^{j} \bar{\rho}_{im}^{\alpha}}, \quad j \le C^{\alpha},\ i > 1 \qquad (1)$$

where C^{α} is the number of α-frames in a GOP. As mentioned in (1), the slope $\hat{\theta}_{ij}^{\alpha}$ estimated at the time instant of F_{ij}^{α} is the cumulated bits used to encode the $[(i-1)\cdot C^{\alpha}+j]$ α-frames divided by the cumulated $\bar{\rho}_{ij}^{\alpha}$ in the $[(i-1)\cdot C^{\alpha}+j]$ encoded α-frames. Note that $\bar{\rho}_{ij}^{\alpha}$ represents the summation of M_{ij}^{α} and Q_{ij}^{α}, which is different from the definition of ρ_c or ρ in [2] where only the zeros among quantized coefficients are considered. In the initial stage, Eq. (1) is simplified to $\sum_{m=1}^{j} R_{1m}^{\alpha} / \sum_{m=1}^{j} \bar{\rho}_{1m}^{\alpha}$ when i = 1.

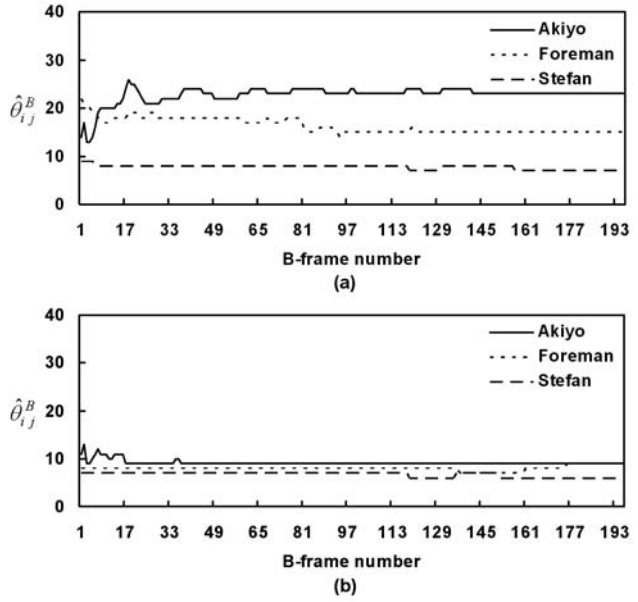

Fig. 1. The $\hat{\theta}^B_{ij}$ value estimated at each B-frame: (a) with Q^B_{ij} only; (b) with the summation of M^B_{ij} and Q^B_{ij}.

Fig. 1 presents the effect of adding M^B_{ij} to the estimation of $\hat{\theta}^B_{ij}$ by three video test segments with different picture complexities. These segments are encoded in CIF format, where a GOP consists of 30 frames with the IBBP pattern. To simplify the notation of horizontal axis, the B-frames are re-numbered successively. Comparing Figs. 1(a) and 1(b) reveals that adding M^B_{ij} can effectively increase the linearity and convergence speed of $\hat{\theta}^B_{ij}$, especially for videos with low complexity such as *Akiyo*.

III. Long-Span Frame Prediction

To predict the traffic of frames that are far from the current frame, we propose the ρ-LSP that consists of two prediction algorithms. The first algorithm considers that the current frame F^α_{ij} and the predicted frame $\hat{F}^\alpha_{i^*(j+w)}$ are at the same GOP, such that $(j+w) \le C^\alpha$ and $i^* = i$. The second algorithm considers that the current frame and the predicted frame are at different GOPs, such that $(j+w) > C^\alpha$ and $i^* > i$. Note that both frames have the same type and the w represents the prediction span. Given the estimated $\hat{\theta}^\alpha_{ij}$, ρ-LSP can predict the bits $\hat{R}^\alpha_{i^*(j+w)}$ of $\hat{F}^\alpha_{i^*(j+w)}$ by

$$\hat{R}^\alpha_{i^*(j+w)} = \hat{\theta}^\alpha_{ij} \cdot \bar{\rho}^\alpha_{ij} \cdot \chi \tag{2}$$

where χ is a weighting factor to respond to the possible influence of the distance between F^α_{ij} and $\hat{F}^\alpha_{i^*(j+w)}$ to $\hat{R}^\alpha_{i^*(j+w)}$, as expressed by

$$\chi = \begin{cases} \dfrac{\sum_{h=1}^{i-1} (\hat{\theta}^\alpha_{h\ (j+w)} \cdot \beta^k)}{\sum_{h=1}^{i-1} (\hat{\theta}^\alpha_{h\ j} \cdot \beta^k)}, & j+w \le C^\alpha, i > 1 \\ \dfrac{\sum_{h=2}^{i-1} (\hat{\theta}^\alpha_{h\ ((j+w)\,\%\,C^\alpha)} \cdot \beta^k)}{\sum_{h=1}^{i-2} (\hat{\theta}^\alpha_{h\ j} \cdot \beta^{k-1})}, & j+w > C^\alpha, i > 2 \end{cases} \tag{3}$$

TABLE I
Prediction performance of LMS, A-LMS, and ρ-LSP on the whole video sequence

D	Video source	LMS μ=0.006	LMS μ=opt.	A-LMS μ=0.006	A-LMS μ=opt.	ρ-LSP no M^α_{ij}	ρ-LSP add M^α_{ij}
1	Video-1	0.343	0.107	0.252	0.074	0.059	0.054
	Video-2	0.300	0.118	0.202	0.076	0.062	0.052
3	Video-1	0.349	0.115	0.257	0.082	0.064	0.058
	Video-2	0.305	0.125	0.207	0.084	0.068	0.057
12	Video-1	0.422	0.412	0.385	0.405	0.110	0.100
	Video-2	0.353	0.319	0.258	0.299	0.115	0.100
30	Video-1	0.634	>100	0.584	2.580	0.154	0.143
	Video-2	0.499	>100	0.410	3.070	0.161	0.144
45	Video-1	0.844	>100	>100	>100	0.218	0.205
	Video-2	0.619	>100	>100	>100	0.238	0.215

where $k = i - h - 1$ and $(j+w)\,\%\,C^\alpha$ is the remainder of dividing $(j+w)$ by C^α. The β with power k is a weighting factor to react to the correlation between GOPs. As mentioned in (3), considering the $(i-1)$ GOPs already encoded in the current scene, there exists $(i-1)$ or $(i-2)$ pairs of α-frames having the same location context as the pair of F^α_{ij} and $\hat{F}^\alpha_{i^*(j+w)}$, where the number of pairs depends on the value of $(j+w)$. In each pair mentioned above, there exists a difference of the estimated slopes between the two frames. Therefore, χ is determined by averaging the differences given by these $(i-1)$ or $(i-2)$ pairs. Note that the cumulation in the numerator of (3) begins from the second GOP of the current scene when $(j+w) > C^\alpha$, for responding to possible influences of crossing GOP boundary on prediction results. In the initial stage, χ is set to 1 when $[(i=1) \cap ((j+w) \le C^\alpha)]$or $[(i \le 2) \cap ((j+w) > C^\alpha)]$.

IV. Performance Evaluation

Two video sequences, *Video*-1 and *Video*-2, which are constructed separately by concatenating sixteen standard test segments with different orders, are used. The sixteen segments are *Table*, *Container*, *Mobile*, *Salesman*, *Coastguard*, *Paris*, *Singer*, *Claire*, *Akiyo*, *Stefan*, *Children*, *Dancer*, *Silent*, *Foreman*, *Tempete* and *Hall monitor*. Both sequences have 5160 frames in CIF format, where a GOP consists of 30 frames with IBBP pattern. We utilize an expression, i.e., PE = $\sum_n e^2(n) / \sum_n x^2(n)$, to represent the degree of Prediction Error (PE), where $e(n)$ represents the prediction error and $x(n)$ represents the actual encoded bits of frame *n*. Two conventional methods, the LMS predictor [1] and the Adaptive LMS (A-LMS) predictor [4], are compared with ρ-LSP because they provide similar capabilities of long span frame prediction. For comparison, the optimal step sizes of LMS and A-LMS for *Video*-1 and *Video*-2 are determined herein in the range of $0.001 \le \mu \le 2$, respectively. The order of LMS and A-LMS is set to five and the value of β in (3) is set to one.

To simplify the presentation of Table I, we translate the prediction spans for I-, P-, and B-frames, namely w_I, w_P, and w_B, into a single distance parameter *D*, where $w_I = \lceil D/30 \rceil$, $w_P = \lceil 9D/30 \rceil$, and $w_B = \lceil 20D/30 \rceil$. During the simulation, this work repeats the prediction process with the span w_α whenever an α-frame is encoded. From the PE results of

LMS and A-LMS in Table I, the use of $\mu = 0.006$, which is optimal for the "*Star Wars*" in [4], results in larger prediction error than that of the optimal step sizes determined herein for *Video*-1 and *Video*-2. This reveals that the optimal μ for LMS or A-LMS obviously depends on video contents and affects the performance. Besides, the prediction operation using LMS or A-LMS may diverge at large D even the optimal step sizes for *Video*-1 and *Video*-2 are found at D = 1. In other words, the decision of the optimal μ depends on the prediction distance. In contrast, the simulation results reveal that ρ-LSP can significantly reduce the prediction error whether *Video*-1 or *Video*-2 is used, especially at large D. The added parameter M_{ij}^{α}, that is the number of non-zero motion vectors, can further increase the prediction accuracy based on the results of the last two columns in Table I. More importantly, the process of ρ-LSP is unique but simple for different video contents and prediction spans in which no search of optimal parameter is required.

V. Conclusion

The proposed ρ-LSP effectively extends the use of ρ-domain source rate control to long-span frame prediction. Given the estimated slope and coding information of the current frame, ρ-LSP can easily predict the bits of a future frame with an arbitrary span w. More importantly, the improvement of prediction accuracy using ρ-LSP is significant and the process of ρ-LSP is unique but simple for different video contents and prediction spans.

References

[1] A. M. Adas, "Using adaptive linear prediction to support real-time VBR video under RCBR network service model," *IEEE/ACM Trans. Networking*, vol. 6, pp. 635-644, Oct. 1998.

[2] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 840-849, Oct. 2002.

[3] S. Sen, J. L. Rexford, J. K. Dey, J. F. Kurose, and D. F. Towsley, "Online smoothing of variable-bit-rate streaming video," *IEEE Trans. Multimedia*, vol. 2, pp. 37-48, Mar. 2000.

[4] S. J. Yoo, "Efficient traffic prediction scheme for real-time VBR MPEG video transmission over high-speed networks," *IEEE Trans. Broadcast.*, vol. 48, pp. 10-18, Mar. 2002.