DRIFT MANAGEMENT AND ADAPTIVE BIT RATE ALLOCATION IN SCALABLE VIDEO CODING

Hua Yang, Rui Zhang and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

ABSTRACT

This paper is concerned with performance optimization of point-to-point communication of scalable video over lossy networks. A variant of the recursive optimal per-pixel estimate (ROPE) is developed and embedded within a rate-distortion (RD) optimized mode selection scheme for SNR scalable video coding. The system allows predicting the current base layer frame from past enhancement layer frame data. Drift management and adaptive bit rate allocation are both RD optimized within this framework. The ROPE method ensures accurate estimation (at the encoder) of the overall decoder distortion, which is critical to the substantial performance gains achieved. A low complexity layer sequential optimization scheme is proposed as well, which approximates the more complex joint optimization scheme, while maintaining most of the performance gains.

1. INTRODUCTION

Scalable coding is a natural paradigm for video transmission over lossy networks, which offers means to mitigate the effects of packet loss [1]. In designing a predictive scalable video coding system, an important issue is whether to use the enhancement layer information for prediction [2]. Its use enables better prediction and hence improves the coding gain. However, if the enhancement layer information is lost during transmission, the incurred mismatch between decoder and encoder will trigger error propagation via prediction and thus degrade the reconstructed video quality. This is the so-called “drift” problem. Drift is usually viewed as highly undesirable and most recent standards, such as H.263 and MPEG4, favor “no-drift” scalable coding. Recently, however, there has been a growing interest in approaches that attempt to optimize the trade-off between some allowed drift and improved compression ef£ciency [3][4][5][6].

The typical communication setting, considered in much of the traditional scalable video coding literature, consists of independent channels with differing capacities. To cater to all these various channels, a coarse but acceptable base layer video quality is necessary, and bit rates of the base and enhancement layers are determined by the channel capacities. However, in the work described herein we are concerned with point-to-point communication through a standard lossy network, which means that only one channel is considered. The scalable setting here is simply a means to packetize the data into packets of differing importance and thereby enable better throughput. Obviously, in this setting there is no need to impose preclusion of drift in prediction or to pre-specify the bit rates of different layers prior to encoding. Instead, we allow drift and adaptively allocate bit rate to the layers per frame. Our line of investigation here is primarily motivated by: (i) optimization of the error resilience performance achievable by scalable coding; (ii) the crucial importance of accurate overall distortion estimation for effective drift management and bit rate allocation.

The proposed approach takes as starting point the basic macroblock (MB)-based SNR scalable video coding system. Our coding framework allows utilization of enhancement layer information for prediction at both the base and enhancement layers, which offers improved prediction but also entails greater risks of damage due to packet loss and drift. Drift management and adaptive bit rate allocation are implemented in conjunction with RD optimized coding mode selection for each MB. In RD optimization, the critical dif£culty is to obtain an accurate estimate of the end-to-end distortion. Much research work has been dedicated to this problem in recent years (see, e.g., [7][8]). This work builds on and extends the ROPE method of [8], which provides an accurate estimate by taking into account the effects of quantization, packet loss, and error concealment.

The paper is organized as follows. Section 2 re-derives the recursion formulae of ROPE while allowing for unrestricted prediction from the enhancement layer. The resultant distortion estimate is embedded within an RD framework in Section 3. Section 4 summarizes the simulation results.

This work is supported in part by the NSF under grants EIA-9986057 and EIA-0080134, the University of California MICRO Program, Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, Inc. and Qualcomm, Inc.

2. END-TO-END DISTORTION ESTIMATION FOR A SCALABLE CODER

For simplicity but without implied loss of generality, we make the following assumptions. The data of one frame is carried in one packet. Thus, the pixel loss rate equals the packet loss rate. We model the channel as a Bernoulli process with packet loss rate p for the enhancement layer. For the base layer, we assume that the packet loss rate is zero.

Assuming the mean squared error criterion, the overall expected distortion levels of pixel i in frame n, at the base and enhancement layers, are given by

$$\begin{aligned} d_n^i(b) &= E\{(f_n^i - \tilde{f}_n^i(b))^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(b)\} + E\{(\tilde{f}_n^i(b))^2\} \quad (1) \end{aligned}$$

$$\begin{aligned} d_n^i(e) &= E\{(f_n^i - \tilde{f}_n^i(e))^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(e)\} + E\{(\tilde{f}_n^i(e))^2\}. \quad (2) \end{aligned}$$

Here f_n^i is the original pixel value. The reconstructed values at the *decoder*, possibly after error concealment, are denoted by $\tilde{f}_n^i(b)$ and $\tilde{f}_n^i(e)$, respectively, which are random variables for the encoder. For future use we also let $\hat{f}_n^i(b)$ and $\hat{f}_n^i(e)$ denote the *encoder* reconstruction at the base and the enhancement layers, respectively. Recursion formulae to compute the £rst and second-order moments of the decoder reconstruction variables, which determine the expected distortion in (1) and (2), are derived below.

2.1. Base Layer Recursion

At the base layer there are three available coding modes: Intra-mode, Inter-mode prediction from the previous base layer frame ("Inter $B \rightarrow B$"), and Inter-mode prediction from the previous enhancement layer frame ("Inter $E \rightarrow B$"). Note that there is no packet loss in the base layer. Thus, we have:

- Intra mode:

$$\begin{aligned} E\{\tilde{f}_n^i(b)\} &= \hat{f}_n^i(b) \\ E\{(\tilde{f}_n^i(b))^2\} &= (\hat{f}_n^i(b))^2 \quad (3) \end{aligned}$$

- Inter $B \rightarrow B$ mode:

$$\begin{aligned} E\{\tilde{f}_n^i(b)\} &= \hat{e}_n^i(b) + E\{\tilde{f}_{n-1}^j(b)\} \\ E\{(\tilde{f}_n^i(b))^2\} &= E\{(\hat{e}_n^i(b) + \tilde{f}_{n-1}^j(b))^2\} \quad (4) \end{aligned}$$

- Inter $E \rightarrow B$ mode:

$$\begin{aligned} E\{\tilde{f}_n^i(b)\} &= \hat{e}_n^i(b) + E\{\tilde{f}_{n-1}^k(e)\} \\ E\{(\tilde{f}_n^i(b))^2\} &= E\{(\hat{e}_n^i(b) + \tilde{f}_{n-1}^k(e))^2\}. \quad (5) \end{aligned}$$

Here $\hat{e}_n^i(b)$ denotes the quantized residue. Due to motion compensation, pixel i in the current MB is predicted from pixel j in the previous base layer frame or from pixel k in the previous enhancement layer frame.

2.2. Enhancement Layer Recursion

The three enhancement layer modes are: Intra-mode, Inter-mode prediction from the current base layer frame ("Upward"), and Inter-mode prediction from the previous enhancement layer frame ("Inter $E \rightarrow E$"). In the case of packet loss in the enhancement layer, we use base layer information at the same position for fallback. In other words, the upward error concealment is used.

- Intra mode:

$$\begin{aligned} E\{\tilde{f}_n^i(e)\} &= (1-p)\cdot \hat{f}_n^i(e) + p \cdot E\{\tilde{f}_n^i(b)\} \\ E\{(\tilde{f}_n^i(e))^2\} &= (1-p)\cdot (\hat{f}_n^i(e))^2 \\ &\quad +p \cdot E\{(\tilde{f}_n^i(b))^2\} \quad (6) \end{aligned}$$

- Upward mode:

$$\begin{aligned} E\{\tilde{f}_n^i(e)\} &= (1-p)\cdot (\hat{e}_n^i(e) + E\{\tilde{f}_n^i(b)\}) \\ &\quad +p \cdot E\{\tilde{f}_n^i(b)\} \\ E\{(\tilde{f}_n^i(e))^2\} &= (1-p)\cdot E\{(\hat{e}_n^i(e) + \tilde{f}_n^i(b))^2\} \\ &\quad +p \cdot E\{(\tilde{f}_n^i(b))^2\}. \quad (7) \end{aligned}$$

- Inter $E \rightarrow E$ mode:

$$\begin{aligned} E\{\tilde{f}_n^i(e)\} &= (1-p)\cdot (\hat{e}_n^i(e) + E\{\tilde{f}_{n-1}^j(e)\}) \\ &\quad +p \cdot E\{\tilde{f}_n^i(b)\} \\ E\{(\tilde{f}_n^i(e))^2\} &= (1-p)\cdot E\{(\hat{e}_n^i(e) + \tilde{f}_{n-1}^j(e))^2\} \\ &\quad +p \cdot E\{(\tilde{f}_n^i(b))^2\}. \quad (8) \end{aligned}$$

3. RD OPTIMIZED MODE SELECTION FOR SCALABLE CODING

The distortion estimate provided by ROPE is then incorporated into an RD framework to select the coding mode and quantization step size for each MB, in order to minimize the overall decoder distortion for the given total bit rate.

The RD optimization problem is typically recast as an unconstrained minimization of the Lagrangian function, $J = D + \lambda R$, where λ is the Lagrange multiplier [9]. Note that contributions from different individual MB's to this cost are additive. Therefore, J can be independently minimized for each MB.

3.1. Joint Optimization

Obviously, the globally optimal mode selection can only be obtained by jointly optimizing over all the possible mode combinations of the base layer MB and the enhancement layer MB, which however involves a non-trivial complexity cost. In joint optimization, the coding mode and quantization step size per MB are determined by

$$\min_{mode} \{J_{MB}(e)\} = \min_{mode} \{D_{MB}(e) + \lambda \cdot (R_{MB}(e) + \gamma \cdot R_{MB}(b))\}. \quad (9)$$

Here, $R_{MB}(e)$ and $R_{MB}(b)$ are the bit rates consumed by the enhancement layer and the base layer, respectively. The estimated enhancement layer distortion of the MB is denoted by $D_{MB}(e)$ and can be expressed as:

$$D_{MB}(e) = \sum_{i \in MB} d_n^i(e), \quad (10)$$

where $d_n^i(e)$ is calculated via ROPE. Moreover, γ is a coefficient whose purpose is to account for the possible bit rate increase incurred by the protection of the base layer, such as error correcting codes (ECC) or re-transmission. In the case of simple retransmission, $\gamma = 2$. The total bit rate is controlled by λ, which is updated frame by frame using the "buffer status" as in [8].

3.2. Sequential Optimization

Joint optimization involves a substantial increase in complexity, which grows exponentially with the number of layers. We next propose a low complexity variant, namely, sequential optimization, which optimizes the layers sequentially.

- For the base layer:

$$\min_{mode} \{J_{MB}(b)\} = \min_{mode} \{D_{MB}(b) + \lambda \cdot (\gamma \cdot R_{MB}(b))\} \quad (11)$$

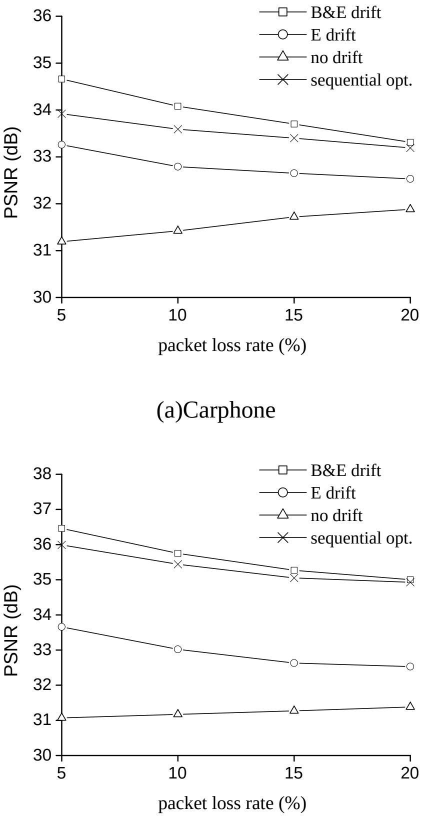
- For the enhancement layer:

$$\min_{mode} \{J_{MB}(e)\} = \min_{mode} \{D_{MB}(e) + \lambda \cdot (\gamma \cdot R_{MB}(e))\} \quad (12)$$

Note that $R_{MB}(b)$ is determined by (11). It is therefore omitted in (12).

4. SIMULATION RESULTS

Our simulation system is based on the UBC H.263+ codec with two-layer scalability [10]. A sequence is first encoded into an H.263 bitstream given the packet loss rate and total bit rate. The bitstream is decoded after undergoing packet loss, whose pattern is randomly generated at the prescribed packet loss rate. System performance is quantified by the mean luminance PSNR of the sequence. In experiments, we use 150 frames and 50 different packet loss patterns. We assume that simple retransmission is used with $\gamma = 2$. Tests were performed on the QCIF sequences Carphone and Salesman.



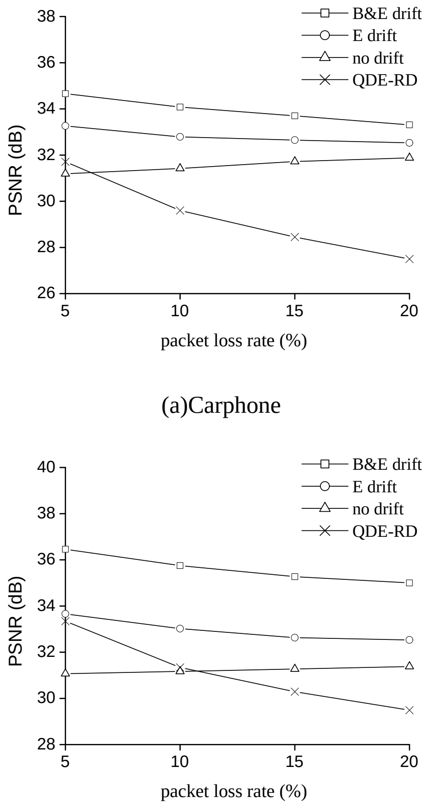
(a)Carphone

(b)Salesman

Fig. 1. PSNR vs. packet loss rate. Frame rate: 30fps. Total bit rate: 300kbps.

Fig. 1 shows the importance of allowing drift. Our proposed joint optimization method is identified as "B&E drift" as it allows prediction from the previous enhancement layer, for both current base and enhancement layer frames. In the similar way, we label the other two reference methods as "E drift" and "no drift". Note that all the three methods employ joint optimization. The proposed sequential optimization scheme is also included as "sequential opt.". It is identical to "B&E drift" in all respects, except that it optimizes the layers sequentially.

From Fig. 1, it is easy to see that for both sequences and at all packet loss rates the proposed joint optimization method offers the best PSNR performance. We conclude that allowing and managing drift is beneficial as long as the end-to-end distortion is accurately estimated and taken into account. In particular, for both sequences and at all packet loss rates, the PSNR gains of the proposed "B&E drift" over "E drift" range from 0.78dB to 2.80dB. The gains of "E drift" over "no drift" range from 0.65dB to 2.59dB. We further observe that the proposed sequential optimiza-



(a)Carphone

(b)Salesman

Fig. 2. PSNR vs. packet loss rate. Frame rate: 30fps. Total bit rate: 300kbps.

tion scheme also consistently outperforms the two reference methods as well. It clearly captures much of the gain of the joint optimization scheme, while the complexity ratio (measured in encoding time) is approximately 1:13. Nevertheless, joint optimization does provide non-trivial gains over sequential optimization at low packet loss rates, e.g., 5%.

Fig. 2 shows the importance of accurate distortion estimation in effective drift management and adaptive bit rate allocation. The first three methods in Fig. 2 are the same as in Fig. 1, all of which employ ROPE-RD optimization. The reference method “QDE-RD” employs the Quantization Distortion Estimate (QDE) for RD joint optimization, but identical in all other respects to the proposed “B&E drift” method. QDE estimates the decoder distortion simply as the quantization distortion. Thus, the packet loss impact is ignored. From Fig. 2, it is obvious that the proposed ROPE-RD method always largely outperforms the QDE-RD method. Moreover, at high packet loss rates, the inaccurate estimate of the QDE method results in worse performance than the “no drift” ROPE-RD method. This demonstrates that the performance of drift management is critically dependent on the quality of the encoder’s end-to-end distortion estimate. It hence substantiates the importance of the proposed ROPE approach.

5. CONCLUSION

In the context of point-to-point scalable video transmission over lossy networks: (i) Decoder drift due to packet loss and prediction should be controlled but not altogether disallowed; (ii) Bit rates of different layers should be adaptively allocated per frame; and (iii) Reaping the full benefits of drift management and adaptive bit rate allocation requires accurate estimation of end-to-end distortion.

6. REFERENCES

[1] R. Aravind, M. R. Civanlar, A. R. Reibman, “Packet loss resilience of MPEG-2 Scalable Video Coding Algorithms,” *IEEE Trans. Circuits Syst. Video Tech.* , vol.6, no.5, Oct. 1996, pp. 426–435.

[2] M. Ghanbari, “Video coding: an introduction to standard codecs,” *IEE Press*, 1999.

[3] A. R. Reibman, L. Bottou, A. Basso, “DCT-based scalable video coding with drift,” *IEEE ICIP 2001*, Oct. 2001.

[4] X. Sun, F. Wu, S. Li, W. Gao, Y.-Q. Zhang, “Macroblock-based progressive fine granularity scalable (PFGS) video coding with flexible temporal-SNR scalabilities,” *IEEE ICIP 2001*, Oct. 2001.

[5] W.-S. Peng, Y.-K. Chen, “Mode-adaptive fine granularity scalability,” *IEEE ICIP 2001*, Oct. 2001.

[6] S. Regunathan, R. Zhang, K. Rose, “Scalable video coding with robust mode selection,” *Signal Processing: Image Communication*, vol.16, no.8, May 2001, pp. 725–732.

[7] G. Cote, F. Kossentini, “Optimal intra coding of blocks for robust video communication over the Internet,” *Image Commun.*, Sept. 1999, pp. 25–34.

[8] R. Zhang, S. L. Regunathan and K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” *IEEE J. Select. Areas Commun.*, vol.18, no.6, June 2000, pp. 966–976.

[9] A. Ortega, K. Ramchandran, “Rate-distortion methods for image and video compression,” *IEEE Signal Processing Mag.*, vol.15, Nov. 1998, pp. 23–50.

[10] H.263+ codec, http://spmg.ece.ubc.ca/.