Optimal End-to-end Distortion Estimation for Drift Management 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. The premise of the work is: (i) Decoder drift due to prediction should be controlled but not altogether disallowed; and (ii) Reaping the full benefits of drift management techniques requires accurate estimation of end-to-end distortion. 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. The ROPE method ensures accurate estimation (at the encoder) of the overall decoder distortion, while explicitly accounting for all factors including quantization, drift, error propagation, packet loss, and concealment. The estimate is used to optimize mode and quantizer step-size selection per macro-block in both the base and enhancement layers. Simulations provide evidence, through substantial performance gains, for the importance of combining drift management with the optimal estimation provided by ROPE.

1. Introduction

Scalable coding is a natural paradigm for video transmission over networks, which offers means to mitigate the effects of packet loss [1][2]. In this paradigm, essential information on the video source is encoded into base layer packets and transmitted with high-priority (or high level of protection), while additional information is encoded into one or more enhancement layers and transmitted with lower priority. The design typically provides at least the minimal quality of service of the base layer under poor network conditions, with potential upgrade to the enhanced quality of the enhancement layers under better network conditions. It hence offers a certain degree of error resilience. Another advantage of scalable coding is that it can simultaneously provide multiple levels of service quality, an important feature for applications involving users with differing communication links, e.g., multicast applications.

In the design of a predictive scalable video coding system, an important issue is whether to use information from the enhancement layer for prediction at the base layer [3]. If it is used, a better prediction for the current frame will be provided, and thus the coding gain can be improved. However, if the enhancement information is lost during transmission, then some mismatch between decoder and encoder is expected despite possible error concealment at the decoder. The enhancement layer error will then propagate both temporally and spatially via motion compensated prediction, and cause further degradation in quality. This is the so-called "drift" problem.

Since drift as described above is likely to degrade the base layer video quality, it has been viewed as highly undesirable. Most recent standards favor "no-drift" scalable coding, where the current base layer can only be predicted from previous base layer data [4][5][6]. In particular, neither the H.263 nor the MPEG4 scalable coding syntax supports prediction of base layer from previous enhancement layer data. There is a similar drift management problem in the Fine Granularity Scalability (FGS) framework of MPEG4 as well. The original FGS does not allow the use of enhancement information in prediction so as to avoid the risk of drift. Recently, however, there has been a growing interest in FGS approaches that attempt to optimize a trade-off between some allowed drift and improved compression efficiency [7][8][9][10]. These approaches exploit past enhancement layer information to predict the current frame at the enhancement layers. Nevertheless, the coding mode decisions in [7][9][10] are mainly based on some heuristic criteria. In [8] an RD criterion was employed for mode decision. However, the distortion is not accurately computed, and is instead based on heuristics.

The typical setting, considered in much of the scalable video coding literature, consists of independent channels with differing capacities. Specifically, there is a "base channel" whose capacity allows transmission of the base layer only, and there are better channels whose capacity allows consistent transmission of enhancement layers. From the coding viewpoint this implies that some receivers have only access to the base layer, and others have *always* access to enhancement layers. This fact considerably simplifies the analysis of the prediction since the enhancement decoder always has access to prior enhancement frames for prediction. However, in the work described herein we are concerned with communication through a standard lossy network. We assume a point-to-point transmission where packets are occasionally lost. The scalable setting here is simply a means to packetize the data into packets of differing importance and thereby enable better throughput. This is achieved either by network implementation of packet priority policies, or by unequal protection of packets via retransmission policies or error correction codes. From the coding perspective, this setting implies that the receiver has access to a subset of the packets (most or all of the base layer packets and a smaller subset of the enhancement layer packets) and produces the best reconstruction possible with them. The natural and ultimate objective of the coding scheme is to minimize the end-to-end distortion. Since we are not interested in the base-layer quality per se, there is no motivation to impose preclusion of drift in the prediction. Instead, we allow drift and re-optimize the complete system. It is trivially obvious that the no-drift solution is a special case and will naturally emerge whenever it is optimal. Our line of investigation here is concerned with the crucial importance of the accurate overall distortion estimate in effective management of the drift. We demonstrate our findings within the traditional framework of scalable predictive video coding.

This work is based on the Recursive Optimal per-Pixel Estimate (ROPE) of [11] which is a low complexity tool for optimally (in the mean square error sense) estimating the end-to-end distortion while accounting for all relevant factors including the notorious error propagation phenomenon. A ROPE-based optimal mode selection method is proposed for drift management in SNR scalable video coding. Coding mode selection is performed per macroblock

(MB) so as to optimize the RD criterion. Prediction from the previous enhancement layer frame forms one of the allowable coding modes in both the base and enhancement layers. This coding framework allows prediction from all possible sources and hence enables us to fully exploit all the available information.

The paper is organized as follows. Section 2 provides an overview of the proposed system framework and approach. The specific recursion formulae are derived in Section 3 and adapt ROPE to the base and enhancement layer coding settings. The distortion estimate is embedded within an RD framework in Section 4. Section 5 summarizes the simulation results.

2. Overview of the System Framework and Approach

This work is primarily concerned with the basic SNR scalable video coding system that underlies the H.263 and MPEG standards: The video frame is segmented into MB's. Each MB is encoded using one of the allowed encoding modes. These modes mainly differ in their prediction operation and range from no prediction ("intra" mode) to modes that predict from various reference sources ("inter" modes).

In the system framework we consider here, an inter-mode base-layer MB can be predicted from either the previous base-layer frame or from the previous enhancement-layer frame. Further, an inter-mode enhancement layer MB can be predicted from either the current base-layer frame or from the previous enhancement layer frame. This is in contrast with "no-drift" systems where base-layer prediction from the previous enhancement layer frame is not allowed. Our underlying premise is that the ultimate performance objective is the expected reconstruction quality at the receiver given the packet loss statistics at the base and enhancement layers. In particular, the "base-layer quality" is not important per se but only indirectly as it impacts the expected overall receiver reconstruction given all received packets.

It is generally recognized that coding mode selection per MB is an efficient means to optimize the trade off between coding efficiency and error resilience. For simplicity of interpretation, and without implied loss of generality, let us assume that there is no packet loss at the base layer. Focusing on our framework we observe the following on the effects of mode selection at the base-layer: An intra-mode coded MB will stop error propagation but is most costly in bit rate. If predicted from the base layer, an inter-mode MB does not stop the propagation of an existing error, but will not introduce new errors and consumes fewer bits than an intra-mode MB. An inter-mode MB that is predicted from the enhancement-layer may introduce a new error but is least costly in bits thanks to the better prediction it normally enjoys. It is natural to optimize the tradeoff with respect to the RD criterion as has been recognized for standard video coding (see [12] for a review). The generality of the RD framework allows joint optimization of various compression parameters that impact the distortion and/or the rate. In our scheme, both the coding mode and quantization step size are jointly selected for each MB to minimize the RD cost.

As the rate is easy to calculate precisely at the encoder, the critical difficulty in RD optimization is that of obtaining an accurate estimate of the end-to-end distortion. While the relevant distortion is between the original video and the reconstructed signal at the receiver, it is impossible to calculate at the encoder. It obviously depends on the impact of packet loss events. The best recourse is to estimate the decoder distortion as accurately as possible given available statistical information on the channel, e.g., the packet loss rate. In fact, much research work has been dedicated to this problem in recent years [11][13]. In this paper we adopt the ROPE method [11], which was shown to exactly compute the expected end-to-end distortion. ROPE consists of recursive estimation of the first and second moments of the decoder pixel reconstruction, which determine the expected overall distortion. The estimate is of low complexity, accurate, and takes into account the effects of quantization, packet loss, and the error concealment.

3. End-to-end Distortion Estimation for a Scalable Coder

In this section, we derive the recursion equations needed to accurately estimate the overall decoder distortion at different layers. While the derivation builds on, and is largely similar to the recursions in [6], the major difference is due to allowing drift and accounting for its effects. In other words, we now allow prediction of the current base layer frame from the previous enhancement layer frame.

Assume that all the data of one frame is carried in one packet, and that the packets are independently decodable. 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 (for simplicity) that the packet loss rate is zero. Note that the derivation below can be easily extended to accommodate an arbitrary packet loss rate at the base layer. Note further that the Bernoulli model is not essential to the approach and can be replaced with more complex models.

Let f_n^i denote the original value of pixel i in frame n. Let $\hat{f}_n^i(b)$ and $\hat{f}_n^i(e)$ denote its *encoder* reconstruction at the base and the enhancement layers, respectively. 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. For the encoder, $\tilde{f}_n^i(b)$ and $\tilde{f}_n^i(e)$ are random variables. Assuming the mean square error criterion, the overall expected distortion levels of pixel i, at the base and enhancement layers, are given by

$$d_n^i(b) = E\{(f_n^i - \tilde{f}_n^i(b))^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(b)\} + E\{(\tilde{f}_n^i(b))^2\} \tag{1}$$

$$d_n^i(e) = E\{(f_n^i - \tilde{f}_n^i(e))^2\} = (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i(e)\} + E\{(\tilde{f}_n^i(e))^2\}. \tag{2}$$

It is obvious that the computation of $d_n^i(b)$ and $d_n^i(e)$ requires the first-order and the second-order moments of the corresponding decoder reconstruction random variables. Recursion formulae to compute these moments are derived below.

3.1 Base Layer Recursion

At the base layer there are three available modes: intra-mode, inter-mode predicting from the base layer, and inter-mode predicting from the enhancement layer. Recall that the packet loss rate at the base layer is assumed to be zero.

If pixel i belongs to a MB that is intra-mode coded, then

$$E\{\tilde{f}_n^i(b)\} = \hat{f}_n^i(b)$$
$$E\{(\tilde{f}_n^i(b))^2\} = (\hat{f}_n^i(b))^2 \tag{3}$$

If the MB is inter-mode coded from the base layer, the decoder has access to the quantized residue, $\hat{e}_n^i(b)$, and the motion vector. Let the motion vector be such that pixel i in the current MB is predicted from pixel j in the previous base layer frame. The encoder's prediction is $\hat{f}_{n-1}^j(b)$ and its reconstruction is given by $\hat{f}_n^i(b) = \hat{e}_n^i(b) + \hat{f}_{n-1}^j(b)$. The decoder must use its own prediction $\tilde{f}_{n-1}^j(b)$ and produces the reconstruction $\tilde{f}_n^i(b) = \hat{e}_n^i(b) + \tilde{f}_{n-1}^j(b)$. For pixel i in the MB we obtain:

$$E\{\tilde{f}_n^i(b)\} = \hat{e}_n^i(b) + E\{\tilde{f}_{n-1}^j(b)\}$$
$$E\{(\tilde{f}_n^i(b))^2\} = E\{(\hat{e}_n^i(b) + \tilde{f}_{n-1}^j(b))^2\} = (\hat{e}_n^i(b))^2 + 2\cdot\hat{e}_n^i(b)\cdot E\{\tilde{f}_{n-1}^j(b)\} + E\{(\tilde{f}_{n-1}^j(b))^2\} \tag{4}$$

If the MB is inter-mode coded from the enhancement layer, let the motion vector be such that pixel i in the current MB is predicted from pixel k in the previous enhancement layer frame. Thus we obtain,

$$E\{\tilde{f}_n^i(b)\} = \hat{e}_n^i(b) + E\{\tilde{f}_{n-1}^k(e)\}$$
$$E\{(\tilde{f}_n^i(b))^2\} = E\{(\hat{e}_n^i(b) + \tilde{f}_{n-1}^k(e))^2\} = (\hat{e}_n^i(b))^2 + 2\cdot\hat{e}_n^i(b)\cdot E\{\tilde{f}_{n-1}^k(e)\} + E\{(\tilde{f}_{n-1}^k(e))^2\} \tag{5}$$

3.2 Enhancement Layer Recursion

At the enhancement layer there are also three available modes: intra-mode, inter-mode predicting from the current base layer frame ("upward" mode), and inter-mode predicting from the previous enhancement layer frame ("forward" mode). Recall that the packet loss rate of the enhancement layer is p. We assume (for simplicity) that upward error concealment is used in the decoder: if an enhancement layer MB is lost the MB at the same position in the current base layer is used to replace it.

Let pixel i be in an intra-mode coded MB. If it is lost then its decoder reconstruction is obtained by upward error concealment $\tilde{f}_n^i(b)$. If it is correctly received, then its decoder reconstruction will be $\hat{f}_n^i(e)$. Thus,

$$E\{\tilde{f}_n^i(e)\} = (1-p)\cdot\hat{f}_n^i(e) + p\cdot E\{\tilde{f}_n^i(b)\}$$
$$E\{(\tilde{f}_n^i(e))^2\} = (1-p)\cdot(\hat{f}_n^i(e))^2 + p\cdot E\{(\tilde{f}_n^i(b))^2\} \tag{6}$$

Let pixel i be in an upward-mode coded MB, and let $\hat{e}_n^i(e)$ denote the transmitted residue. Since it is in upward-mode the motion vector is zero, and its prediction from the current base layer in the decoder will be $\tilde{f}_n^i(b)$, which coincides with the decoder upward error concealment value. Thus, we have,

$$E\{\tilde{f}_n^i(e)\} = (1-p)\cdot(\hat{e}_n^i(e) + E\{\tilde{f}_n^i(b)\}) + p\cdot E\{\tilde{f}_n^i(b)\}$$
$$E\{(\tilde{f}_n^i(e))^2\} = (1-p)\cdot E\{(\hat{e}_n^i(e) + \tilde{f}_n^i(b))^2\} + p\cdot E\{(\tilde{f}_n^i(b))^2\}$$
$$= (1-p)\cdot((\hat{e}_n^i(e))^2 + 2\cdot\hat{e}_n^i(e)\cdot E\{\tilde{f}_n^i(b)\} + E\{(\tilde{f}_n^i(b))^2\}) + p\cdot E\{(\tilde{f}_n^i(b))^2\} \tag{7}$$

Let pixel i be in a forward-mode coded MB, and let $\hat{e}_n^i(e)$ denote the transmitted residue and $\tilde{f}_{n-1}^j(e)$ denote its prediction from the previous enhancement layer in the decoder. Its decoder upward error concealment result will be $\tilde{f}_n^i(b)$. Hence we have,

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

We re-emphasize that these recursions are performed per pixel at the encoder in order to calculate the expected overall distortion at the decoder. While for simplicity the recursions have been derived within a two-layer scalable coding setup, they can be extended in a straightforward manner to compute the overall decoder distortion at each layer of a multi-layer scalable video coder.

Note that the estimate is precise for integer-pixel motion estimation. In the half-pixel case, because of the averaging operation involved, the exact computation of the second moment might incur a prohibitive complexity. A similar problem occurs in precise computation of the second moment in the context of the bi-directional mode, where the final prediction is the average of the two available predictions from the base layer and the enhancement layer respectively. For the time being, integer-pixel motion compensation is used and bi-directional mode is not included in mode selection. Variants that involve these options are under current investigation.

4. RD Optimized Mode Selection for Scalable Coding

The distortion estimate provided by ROPE is next employed in 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 bit rate.

The classical rate-distortion formulation of the problem is concerned with joint selection of the coding modes for all MB's so as to minimize the overall distortion, D, subject to a given rate constraint, R. Equivalently, the problem can be recast as an unconstrained minimization of the Lagrangian function, $J = D + \lambda R$, where λ is the Lagrange multiplier [12]. Note that contributions from different individual MB's to this cost are additive and, therefore, the cost J can be independently minimized for each MB.

Obviously, the globally optimal mode selection can only be obtained by jointly optimizing all the possible mode combinations of the base layer MB and the enhancement layer MB, which involves considerable complexity. A simpler and more practical approach is to optimize the coding modes for the base and enhancement layers sequentially. Although it is suboptimal in principle, it is practically implementable and substantial gains can still be achieved as is shown in the results section.

At the base layer, the mode and quantization step size per MB are determined by the simple minimization:

$$\min_{\text{mod}\,e}(J_{MB}(b)) = \min_{\text{mod}\,e}(D_{MB}(b) + \lambda(b)R_{MB}(b)), \quad (9)$$

where the distortion of the MB is the sum of the distortion contributions of all its individual pixels:

$$D_{MB}(b) = \sum_{i \in MB} d_n^i(b). \tag{10}$$

At the enhancement-layer, the prediction mode and quantization step size is selected to minimize

$$\min_{\bmod e}(J_{MB}(e)) = \min_{\bmod e}(D_{MB}(e) + \lambda(e) R_{MB}(e)), \tag{11}$$

where the distortion of the MB is given by:

$$D_{MB}(e) = \sum_{i \in MB} d_n^i(e). \tag{12}$$

Note that we employ ROPE to calculate the distortion per pixel, while the coding mode and quantization step size is selected per MB via (9) and (11). The bit rate is controlled by $\lambda(b)$ and $\lambda(e)$, which are updated frame by frame using the "buffer status" as in [11].

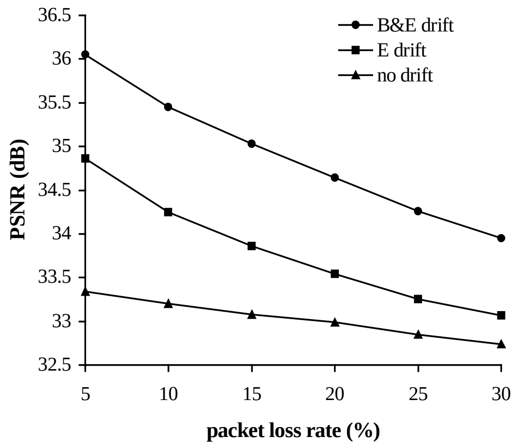
5. Simulation Results

Our simulation system is implemented based on the UBC H.263+ codec with two-layer scalability [14]. In the experiment a small extension is made to the original syntax so that the base layer can be predicted from the previous enhancement layer. We assume that one packet corresponds to one frame. Packet loss patterns are produced by a random number generator and are used to randomly drop packets at the specified packet loss rates. As mentioned in Section 3.2, the enhancement layer error concealment scheme at the decoder consists of simple upward error concealment. In the proposed system, the ROPE-RD algorithm is used for both layers for selection of mode and quantizer parameter.

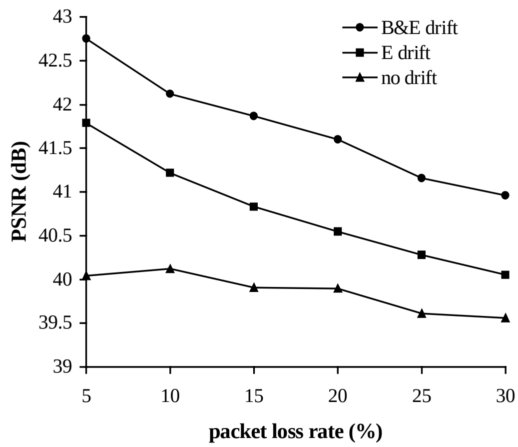
The test is performed on the first 150 frames of the four QCIF sequences: carphone, miss_am, foreman and salesman. A sequence is first encoded into an H.263 bitstream based on the given packet loss rate and bit rates allocated to the base and enhancement layers. The total bit rate used in the following experiment is 300 kb/s with 75 kb/s for the base layer and 225 kb/s for the enhancement layer. The bitstream is then decoded with some packet loss pattern, which is randomly generated given the prescribed packet loss rate. Recall that packet loss is restricted to the enhancement layer. The mean luminance PSNR of each frame is first calculated by averaging over 50 different packet loss patterns. Finally, the mean luminance PSNR of the whole sequence is obtained by averaging over all the frames involved and is used to indicate the system performance.

Fig.1 demonstrates the importance of allowing drift. The competing methods employ the best available estimate, namely ROPE, but differ in whether and to what extent they allow drift. The proposed method allows prediction from the previous enhancement layer for both base and enhancement layer coding. It is identified in the figure as "B&E drift". One reference method disallows all prediction from the enhancement layer and is referred to as "no drift". Note that here there is no error propagation and ROPE itself is of little help. The second reference method applies ROPE to optimize decisions and allows prediction from the previous enhancement layer frame only for encoding the current enhancement layer frame. This method was proposed in [6] and is referred to here as "E drift".

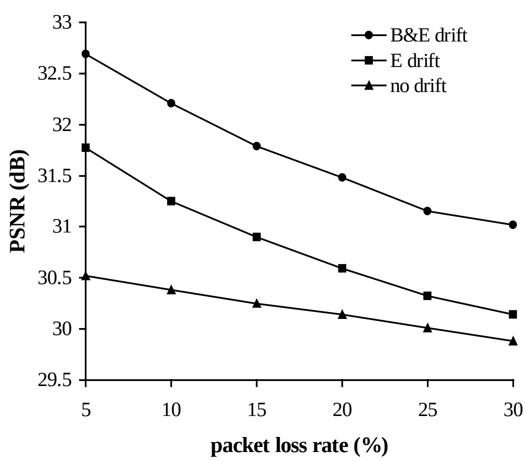
From Fig.1, it is easy to see that for all the sequences and at all packet loss rates the proposed method offers the best PSNR performance. We conclude that when the end-to-end distortion is precisely taken into account, allowing and managing drift is beneficial. For example, at packet loss rate of 5%, the PSNR gains of the proposed ROPE with “B&E drift” over ROPE with “E drift” range from 0.92dB to 2.83dB. The gains of ROPE with “E drift” over “no drift” range from 1.25dB to 2.20dB. Moreover, as it is obvious in the figures, with the packet loss rate increased, the PSNR gain of “B&E drift” over “E drift” is almost unchanged, or even increased in some cases, while the gain of “E drift” over “no drift” diminishes very quickly. This result demonstrates the effectiveness of our “B&E drift” coding scheme as compared to the “E drift” method.



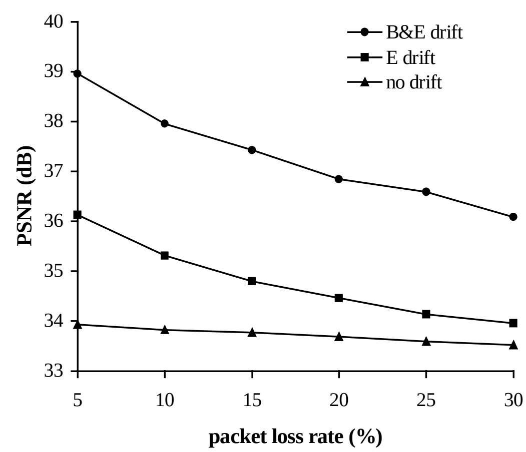
(a) carphone



(b) miss_am



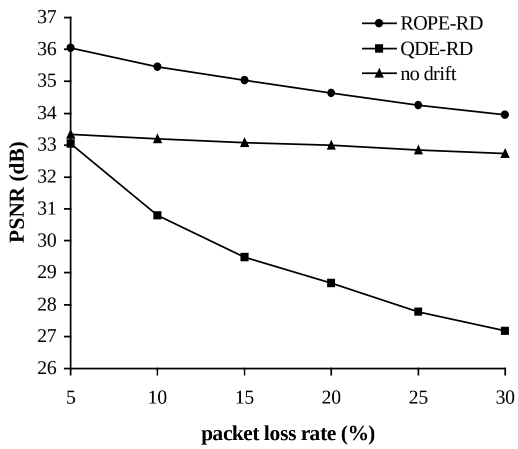
(c) foreman



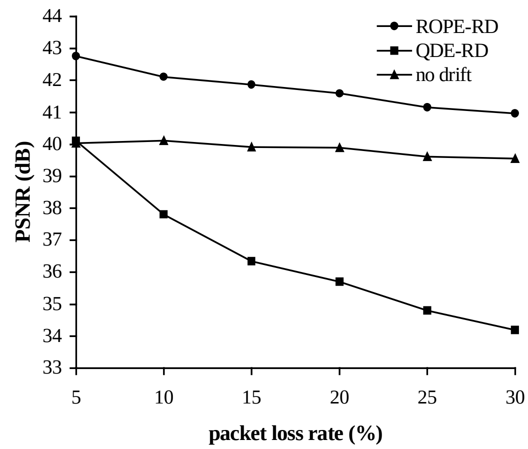
(d) salesman

Fig.1 PSNR vs. packet loss rate. Base layer bit rate: 75 kb/s, enhancement layer bit rate: 225 kb/s. Frame rate: 30 f/s, total number of frames: 150.

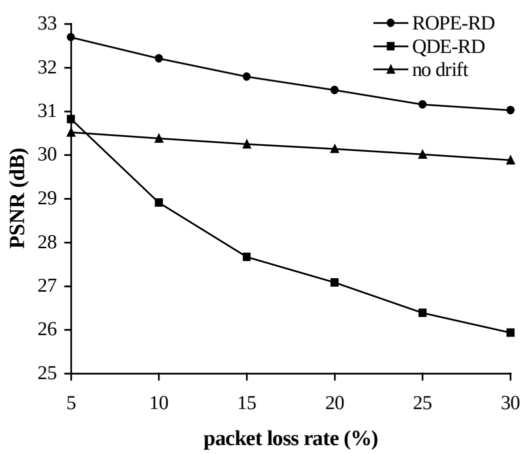
The next experiment focuses on the impact of the accurate estimate on the achievable gains due to drift management. To demonstrate the central role played by ROPE, we compare the proposed ROPE-RD optimized mode selection scheme with another RD optimized mode selection scheme, which does not accurately estimate the end-to-end distortion. The reference method employs the Quantization Distortion Estimate (QDE). QDE estimates the distortion as the quantization distortion. Thus, the packet loss impact is ignored. The comparison result is shown in Fig.2. Here, the QDE-RD method uses the same coding framework as the proposed ROPE-RD method, except that it employs QDE in its RD optimization. For another reference, the "no-drift" method is also included.



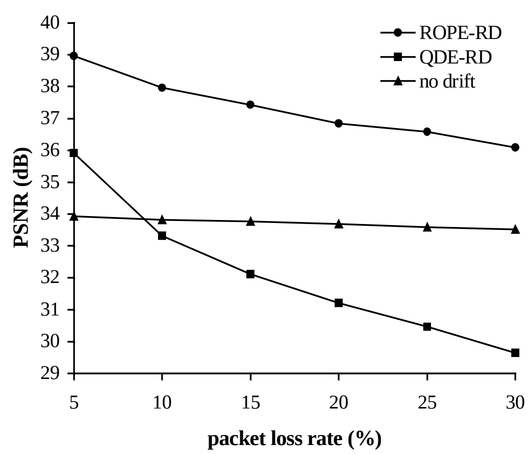
(a) carphone



(b) miss_am



(c) foreman



(d) salesman

Fig.2 PSNR vs. packet loss rate. Base layer bit rate: 75 kb/s, enhancement layer bit rate: 225 kb/s. Frame rate: 30 f/s, total number of frames: 150.

From Fig.2 (a)~(d), it is obvious that the proposed ROPE-RD method always outperforms the competing techniques. Note also that in most cases an inaccurate estimate may even lead to worse performance than the “no drift” method. This demonstrates that the performance of drift management techniques is dependent to a large degree on the quality of the decoder distortion estimate, and hence the significance of employing the proposed ROPE.

6. Conclusion

In the context of point-to-point video transmission over lossy networks, scalable video coding is an important tool for error resilience. In this scenario the ultimate goal is to maximize the reconstructed video quality regardless of the base layer quality itself. There is hence no compelling reason to preclude drift due to prediction at the base layer. If drift is allowed it must be efficiently managed so as to optimize the overall performance. We propose an RD optimized mode selection method for SNR scalable coding that employs ROPE to accurately compute the end-to-end distortion. Simulation results show substantial performance gains. They further provide evidence that the potential benefits of the flexibility due to allowing drift depend critically on the employment of an accurate end-to-end distortion estimate such as ROPE. The implementation of the proposed method requires a small modification of the current H.263 standard, in refining the prediction modes definition.

Reference

[1] M. Ghanbari, "Two-layer coding of video signals for VBR networks," *IEEE J. Select. Areas Commun.*, vol.7, no.5, June 1989, pp. 771-781.

[2] 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.

[3] M. Ghanbari, "Video coding: an introduction to standard codecs," *IEE Press*, 1999.

[4] M. Ghanbari, V. Seferidis, "Efficient H.261-based two-layer video codecs for ATM networks," *IEEE Trans. Circuits Syst. Video Tech.*, vol.5, no.2, Apr. 1995, pp. 171-175.

[5] J. F. Arnold, M. R. Frater, Y. Wang, "Efficient drift-free Signal-to-Noise ratio scalability," *IEEE Trans. Circuits Syst. Video Tech.*, vol.10, no.1, Feb. 2000, pp. 70-82.

[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] A.R. Reibman, L. Bottou, "Managing drift in a DCT-based scalable video encoder, " *IEEE Data Compression Conf.*, Mar. 2001.

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

[9] 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.

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

[11] 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.

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

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

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