SWITCHED ERROR CONCEALMENT AND ROBUST CODING DECISIONS IN SCALABLE VIDEO CODING

Rui Zhang, Shankar L. Regunathan and Kenneth Rose

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

Abstract

This work introduces two complementary techniques to improve the packet loss resilience of scalable video coding systems. First, a "switch per-pixel" error concealment (SPEC) scheme is proposed, which allows the decoder to exploit information from both the current base layer and previous enhancement-layer frame for the reconstruction of missing enhancement-layer blocks. Based on the packet loss history and the quantized base-layer data, the algorithm switches per pixel *between the two information sources. SPEC is shown to consistently outperform standard concealment methods. The second main contribution is concerned with encoder decision optimization. Enhancement layer prediction modes are selected so as to minimize the overall decoder reconstruction distortion, which is due to quantization, packet loss and error propagation. The distortion computation uses a recursive optimal per-pixel estimate (ROPE) to accurately account for the effects of error concealment as well as spatial and temporal error propagation. Simulation results show that ROPE-based mode selection substantially outperforms conventional prediction mode selection schemes. Finally, the combination of SPEC at the decoder and ROPE-based mode selection at the encoder is shown to achieve significant additional performance gains.*

1. INTRODUCTION

Scalable coding, in conjunction with unequal error protection, is an important tool for video transmission over packet-switched networks. Reliable communication of base-layer packets by giving them higher priority, or error protection, ensures a coarse quality of reconstruction which is acceptable as the worst-case scenario [1] [2]. However, transmission can be guaranteed for only a small fraction of the total number of packets and, hence, a relatively low base layer rate. On the other hand, the enhancement layer is allocated larger rate, but its packets are subject to packet loss. Further, decoder reconstruction that exploits previous enhancement-layer frames may suffer additional degradation due to error propagation. In summary, a minimal amount of error-free information is available from the base layer, while the enhancement layer contains a substantial amount of additional (albeit often unreliable) information.

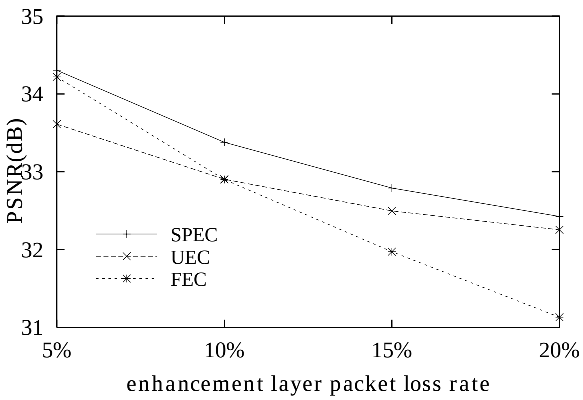
This partition of information between the layers forms an obstacle to the design of two major components of the scalable video coding system: (i) At the decoder, the error concealment module for missing enhancement-layer blocks; (ii) At the encoder, the prediction mode selector for encoding enhancement-layer blocks. Note that both modules are mainly concerned with the estimation of enhancement-layer blocks. Conventional algorithms for either error concealment or prediction mode selection do not make full use of the available information for this estimate [3] [4]. The premise of this work is that the above suboptimal use of information is the main obstacle on the way to major performance enhancement in scalable video coding. Our focus is therefore on deriving algorithms for concealment and prediction mode selection that efficiently exploit information provided by the base and enhancement layers.

In section 2, we derive an error concealment scheme that switches between the two information sources *per pixel* based on the packet loss history and the base layer quantized residual. A strategy for prediction mode selection per enhancement layer macroblock (MB) is derived in section 3, which explicitly computes and minimizes the expected overall distortion at the decoder.

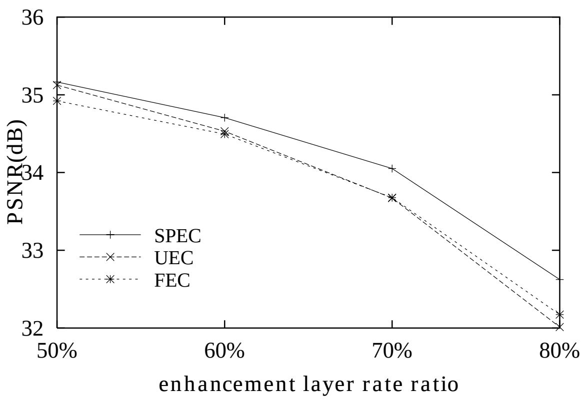
2. "SWITCH PER-PIXEL" ERROR CONCEALMENT

Error concealment is used by the decoder to reconstruct blocks that have been affected by packet loss. The standard technique for enhancement-layer error concealment is to use the current base layer data [3]. However, this ignores information from prior enhancement-layer frames, and is efficient only when the quality of the base layer is moderate-

This work is supported in part by the NSF under grant MIP-9707764, the University of California MICRO Program, Conexant Systems, Inc., Fujitsu Laboratories of America, Inc., Lernout & Hauspie Speech Products, Lucent Technologies, Inc., and Qualcomm, Inc.



(a)



(b)

Fig. 1. Comparison of error concealment schemes. Methods: SPEC(proposed), UEC [3], FEC [4]. Sequence *carphone*, frame rate = 30fps, total bit rate = 200kbps. (a) PSNR vs. enhancement-layer packet loss rate; Enhancement-layer rate = 75% of total rate. (b) PSNR vs. enhancement-layer rate (as percentage of total rate); Enhancement-layer packet loss rate = 10%

ly inferior to that of the enhancement layer. On the other hand, error concealment using motion-compensated prior enhancement-layer frames [4] ignores base-layer data and is ineffective during rapid scene changes. It also promotes propagation of errors due to packet loss.

We propose to combine the two sources of information by means of a simple switch operation at the pixel level. To reconstruct each missing pixel, we switch between the base-layer pixel or the corresponding (motion compensated) pixel in the previous enhancement-layer frame. The decision depends on the base-layer residue and the enhancement-layer error propagation history. Note that the enhancement-layer pixel in previous frame is normally useless if it has been corrupted or if there is a scene change (indicated by a significant level base-layer residue). The Switch per-Pixel Error Concealment (SPEC) algorithm is summarized as follows: For each pixel in a missing block:

- If the current base-layer residue at this location is zero, *and* if the corresponding enhancement-layer packet in the previous frame had been received correctly
  - Copy the corresponding enhancement-layer pixel in the previous frame
- Otherwise
  - Copy the corresponding base-layer pixel

For the simulations, we modified the UBC H.263+ decoder [5]. Each packet contains only one GOB. Two standard error concealment schemes are used as reference: (i) Upward Error Concealment (UEC) which uses base-layer data [3], and (ii) Forward Error Concealment (FEC) which uses previous enhancement-layer data [4]. We present the results for the QCIF sequences “carphone”, “foreman”, “mother&daughter” and the CIF sequence “LTS”. The PSNR of luminance reconstruction is averaged over 30 different channel realizations.

Figure 1 (a) presents PSNR versus packet loss rate and Figure 1 (b) depicts PSNR versus enhancement layer bit rates (as a fraction of total bit rate). Table 1 summarizes the performance on the different sequences. Note that the relative performance of UEC and FEC depends on the packet loss rate and the enhancement layer bit rate. SPEC, however, consistently outperforms the other two methods.

Table 1. Comparison of error concealment methods. (Enhancement layer rate = 75% of total rate, enhancement layer packet loss rate = 10%, frame rate = 30fps, total bit rate: 200kbps for QCIF, 1000kbps for CIF)

Sequence	SPEC	UEC	FEC
Carphone	33.38dB	32.90dB	32.90dB
Foreman	30.74dB	30.46dB	30.27dB
Mother&Daughter	33.94dB	33.57dB	33.50dB
LTS	32.17dB	31.78dB	32.09dB

3. PREDICTION MODE SELECTION

Mode selection is a standard compatible tool to achieve resilience to packet loss. To our knowledge, there has been little work on mode selection for scalable coding, unlike the case of mode selection for single-layer video coding (see [6] [7]). In H.263+ [8], the enhancement layer has a choice of three prediction modes for each MB: upward prediction from the base layer, forward prediction from the previous enhancement-layer frame, and bi-directional prediction using both.

Recall that the base-layer reconstruction is error free. Thus, the use of upward prediction limits error propagation at the enhancement layer, and is more effective during scene changes. However, this may imply reduced compression efficiency. On the other hand, forward prediction is superior if the reproduced quality of the enhancement layer is substantially higher than that of the base layer. Our objective is to derive a mode selection algorithm that optimizes the overall rate-distortion performance.

The key step is the estimation of the overall decoder distortion which accounts for quantization, packet loss, and the error concealment scheme. However, this task is complicated by two factors. Spatial error propagation beyond MB boundaries can only be accurately accounted for by computing the distortion *per pixel*. Further, the contributions of quantization and packet loss to the overall distortion are not additive. To accurately compute the distortion, we extend the Recursive Optimal per-Pixel Estimate (ROPE) which had been originally derived for non-scalable coding [6].

We assume that the group of blocks (GOB) is carried in a separate 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. 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 enhancement layer respectively. As the base layer is loss free, the base-layer reconstruction at the decoder is $\hat{f}_n^i(b)$. However, the enhancement-layer reconstruction at the decoder, $\tilde{f}_n^i(e)$, can be different from its reconstruction at the encoder due to packet loss and error propagation. For the encoder, $\tilde{f}_n^i(e)$ is a random variable. Assuming mean square error distortion, the overall expected distortion for this pixel, at the enhancement layer, is given by

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

The computation of d_n^i requires the first and second moments of the corresponding random variable $\tilde{f}_n^i(e)$. We develop recursion formulae to compute these two moments.

Let the prediction value at the encoder side be $\hat{g}_n^i(e)$, and that of the decoder side be $\tilde{g}_n^i(e)$. Let the motion vector of the MB associate pixel i with pixel j in the previous frame. The prediction, at the encoder and decoder, corresponding to the three prediction modes are given by:

- upward prediction:

$$\hat{g}_n^i(e) = \tilde{g}_n^i(e) = \hat{f}_n^i(b). \quad (2)$$

- forward prediction:

$$\begin{aligned} \hat{g}_n^i(e) &= \hat{f}_{n-1}^j(e), \\ \tilde{g}_n^i(e) &= \tilde{f}_{n-1}^j(e). \end{aligned} \quad (3)$$

- bi-directional prediction:

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

Note that $\hat{g}_n^i(e)$ and $\tilde{g}_n^i(e)$ may be different. The prediction error of the current pixel is given by $\hat{e}_n^i(e) = \hat{f}_n^i(e) - \hat{g}_n^i(e)$. If the packet carrying prediction error is received correctly, the decoder reconstruction is given by $\hat{e}_n^i(e) + \tilde{g}_n^i(e)$. The probability of this event is $(1-p)$. Error propagation can occur if $\tilde{g}_n^i(e)$ is different from $\hat{g}_n^i(e)$. If the packet is lost, error concealment is used. The probability of this event is p. Let the reconstruction in this case be denoted by $c_n^i(e)$. If upward error concealment is used, $c_n^i(e) = \hat{f}_n^i(b)$. For the case of forward error concealment, $c_n^i(e) = \tilde{f}_{n-1}^j(e)$. When SPEC is used, if $\hat{e}_n^i(b) \neq 0$, we have $c_n^i(e) = \hat{f}_n^i(b)$. If $\hat{e}_n^i(b) = 0$, we have $c_n^i(e) = \tilde{f}_{n-1}^j(e)$ with probability $(1-p)$, and $c_n^i(e) = \hat{f}_n^i(b)$ with probability p.

We thus have the following recursion functions for the expected moments of $\tilde{f}_n^i(e)$:

$$\begin{aligned} E\{\tilde{f}_n^i(e)\} &= (1-p)(\hat{e}_n^i(e) + E\{\tilde{g}_n^i(e)\}) \\ &+ pE\{c_n^i(e)\}, \\ E\{(\tilde{f}_n^i(e))^2\} &= (1-p)E\{(\hat{e}_n^i(e) + \tilde{g}_n^i(e))^2\} \\ &+ pE\{(c_n^i(e))^2\}. \end{aligned} \quad (5)$$

An overall rate-distortion framework is used for mode selection. For each MB, the prediction mode and quantization step size are chosen to minimize $\min_{\text{mode}}(D_{\text{MB}} + \lambda R_{\text{MB}})$ where $D_{\text{MB}} = \sum_{i \in \text{MB}} d_n^i$, and R_{MB} is the rate for the MB. Note that while the distortion is calculated *per pixel*, the prediction mode and quantization step size are selected *per MB*.

We demonstrated the power of ROPE-based prediction mode selection in conjunction with conventional upward error concealment (UEC) in [9]. Table 2 shows that similar gains can be achieved when ROPE-based prediction mode selection is used in conjunction with SPEC. We compare ROPE with two standard approaches for prediction mode selection in the enhancement layer. The first method considers only quantization distortion for prediction mode selection (QDE), and the second mode always uses the upward prediction (UP). For comparison, we also give the results when UEC is used as the error concealment in Table 3. The complementary nature of performance gains due to improved prediction mode selection (ROPE) and better error concealment (SPEC) is seen by comparing the results in the two Tables. Thus ROPE-SPEC can achieve up to 1.5dB in PSNR gains over the standard coding algorithm which uses upward prediction and upward concealment (UP-UEC). This is shown in Figure 2, where the proposed overall system outperforms the best of the other two by at least 1.0dB over the entire range of packet loss rate.

Table 2. Comparison of prediction mode selection methods in conjunction with SPEC. (Enhancement layer rate = 75% of total rate, enhancement layer packet loss rate = 10%, frame rate = 30fps, total bit rate: 200kbps for QCIF, 1000kbps for CIF)

Sequence	ROPE-SPEC	UP-SPEC	QDE-SPEC
Carphone	34.85dB	33.49dB	34.14dB
Foreman	31.99dB	30.46dB	31.10dB
Mother&Daughter	35.63dB	34.35dB	34.91dB
LTS	33.00dB	31.74dB	32.30dB

Table 3. Comparison of prediction mode selection methods in conjunction with UEC. (Same conditions as in Table 2)

Sequence	ROPE-UEC	UP-UEC	QDE-UEC
Carphone	34.41dB	33.39dB	33.52dB
Foreman	31.46dB	30.32dB	30.47dB
Mother&Daughter	35.17dB	34.26dB	34.29dB
LTS	32.59dB	31.62dB	31.71dB

4. CONCLUSION

We propose two complementary techniques to improve the robustness of scalable video coding. The switch per-pixel error concealment (SPEC) technique is an enhancement of the decoder. It exploits all the available sources of information on lost blocks and takes into account the loss history in their reconstruction. Simulation results demonstrate significant gains over traditional methods. An optimal prediction mode selection strategy is proposed for the encoder that accurately computes and minimizes the total expected distortion of decoder reconstruction. Experiments show substantial and consistent gains over existing techniques. The integration of ROPE mode selection with SPEC gives further gains.

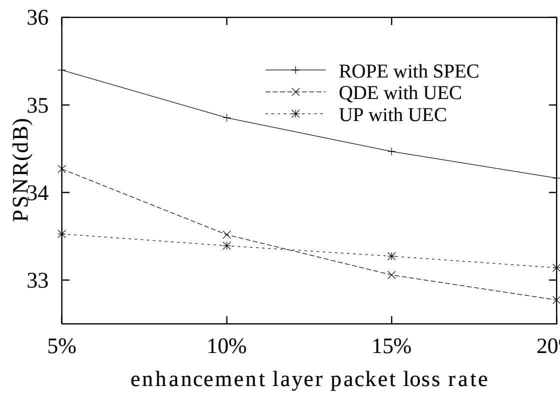


Fig. 2. PSNR vs. enhancement-layer packet loss rate. Methods: ROPE with SPEC(proposed), QDE with UEC, UP with UEC. Sequence *carphone*, frame rate = 30fps, total bit rate = 200kbps, enhancement layer rate = %75 of total rate.

5. REFERENCES

[1] R. Aravind, M. R. Civanlar and A. R. Reibman, “Packet loss resilience of MPEG-2 scalable video coding algorithms,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.6, no. 5, Oct. 1996, pp. 426-435.

[2] J. Villasenor, Y. Zhang, and J. Wen, “Robust video coding algorithms and systems,” *Proceedings of the IEEE*, vol.87, no.10, Oct. 1999, pp. 1724-33.

[3] S. Aign and K. Fazel, “Temporal & spatial error concealment techniques for hierarchical MPEG-2 video codec,” *1995 IEEE International Conference on Communications*, pp. 18-22, June 1995, Seattle WA, USA.

[4] M. Ghanbari and V. Seferidis, “Cell-loss concealment in ATM video codecs,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, no. 3, June 1993, pp. 238-247.

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

[6] R. Zhang, S. L. Regunathan and K. Rose, “Video coding with optimal intra/inter mode switching for packet loss resilience”, to appear in *IEEE Journal of Selected Areas in Communications*, special issue on Error-Resilient Image and Video Transmission.

[7] G. Cote and F. Kossentini, “Optimal intra coding of blocks for robust video communication over the Internet,” *Signal Processing: Image Communication*, vol.15, (no.1-2), Sept. 1999, pp. 25-34.

[8] ITU-T Recommendation H.263, “Video coding for low bit rate communication,” 1998.

[9] R. Zhang, S. L. Regunathan and K. Rose, “Scalable Video Coding with Robust Mode Selection”, *Packet Video Workshop*, Cagliari, Italy, May 2000.