# Optimal Estimation for Error Concealment 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 is concerned with scalable video coding and, specifically, with error concealment for the mitigation of damage caused by packet loss at the enhancement layer. We propose a frequency-domain error concealment scheme that makes full use of the information available at the decoder from both the current base layer, and prior enhancement layer frames. It employs a statistical model for the evolution of transform coefficients from frame to frame and implements the optimal estimate of the reconstructed coefficient. Moreover, the ideas underlying the proposed approach allow detection and correction of blocks that suffer from significant error propagation from past losses. Substantial PSNR gains are observed in the experiments.*

## 1. Introduction

Scalable coding is an important tool to enable efficient transmission of video over packet-switched networks. In a scalable coder, essential information about the source is transmitted in the base layer, and can be independently decoded to obtain a coarse quality of reconstruction. Supplementary information is transmitted in higher enhancement layers, which, when combined with base layer information, improves the reconstruction quality at the decoder.

In conjunction with unequal error protection, scalable coding provides robustness for the video transmission in packet-switched networks. The base-layer packets are given higher priority in error protection or network transportation [1]. Thus, reliable communication of base-layer packets can be reasonably guaranteed, and this ensures a minimum quality of reconstruction.

However, it is important to note that transmission can be guaranteed for only a small fraction of the total number of packets. This implies a relatively low base-layer rate and, consequently, coarse quantization in the base layer. On the other hand, the enhancement layer is allocated a higher 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. Error concealment for the missing blocks in the enhancement layer at the decoder is thus necessary to improve the performance.

Conventional algorithms for error concealment at the enhancement layer have not made efficient use of all the information available at the decoder. In particular, known error concealment methods make exclusive use of either current base-layer data [2] or previous enhancement-layer data [3]. A method for adaptive concealment via a neighborhood matching criterion was proposed, but only for the case of fixed step size quantization of the enhancement layer [4]. Recently, algorithms that switch between the two sources of information (current base-layer and previous enhancement-layer data) have been proposed. Such switching is done in the spatial domain at either the block level [5] or the pixel level [6].

The objective of this work is to derive an algorithm for optimal concealment of the enhancement-layer blocks whose packets were lost. We propose a frequency domain error concealment scheme which makes full use of all the information available at the decoder. It employs a statistical model for the evolution of transform coefficients from frame to frame, and implements the optimal estimate of the reconstructed coefficient given both current base-layer and previous enhancement-layer data. Moreover, this frequency domain error concealment (FDEC) scheme naturally enables, and is complemented by, post-processing of correctly received blocks to detect and mitigate error propagation from past losses. Substantial PSNR gains were obtained in the experiments.

The paper is organized as follows. In section 2, we derive the frequency domain error concealment (FDEC) algorithm. We present the simulation results to demonstrate its performance in section 3. Concluding remarks are provided in section 4.

## 2. Frequency Domain Error Concealment by Optimal Estimation

For simplicity we specialize to two-layer SNR scalable coding, where the base layer and enhancement layer have the same spatial and temporal resolution. We also assume that base-layer packets are always received at the decoder, and losses are confined to enhancement-layer packets. In this case, error concealment is used by the decoder to reconstruct blocks whose enhancement-layer packets have not arrived at the decoder.

In the standard approach to SNR scalable video coding, the video frame is segmented into macroblocks (MBs), the coding mode of each MB determines the type of prediction to be applied. At the base layer, the block can be encoded without prediction (intra-mode), or using motion compensated prediction (inter-mode). The enhancement-layer error concealment of blocks that have been intra-coded at the base layer is straightforward and is usually performed using its corresponding base-layer reconstruction. The concealment of blocks that have been inter-coded is more complicated since there are multiple sources of information about such blocks. In particular, one can use either current base-layer data or previous enhancement-layer data. We define optimal error concealment as the best estimate of the block, given the information received about it from the base-layer packet, and the information available from the enhancement-layer reconstruction of previous frames. It is important to note that while the base-layer information is error free, information extracted from prior enhancement-layer frames may have been altered due to packet loss and concealment.

The proposed optimal error concealment technique builds on earlier work of the optimal predictor design for scalable video coding [7]. Unlike conventional error concealment schemes that operate in the spatial domain [2]-[6], we propose to implement optimal error concealment in the discrete cosine transform (DCT) domain, i.e., to directly reconstruct the DCT coefficients of the missing block. The frequency domain offers two important advantages. First, the quantization interval of each DCT coefficient in the base layer is readily available in the frequency domain. Thus, the possible range of the DCT signal is easily determined. Second, the DCT coefficients are almost uncorrelated, and the estimate may be computed independently for each coefficient at negligible loss of optimality.

### 2.1 The optimal estimation algorithm

To jointly exploit information from current base-layer data and previous enhancement-layer data, we employ a statistical model for the evolution of DCT coefficients. The evolution of a DCT coefficient in time ("from frame to frame") may be modeled by a first-order Markov process

$$x_{i,n} = \rho x_{i',n-1} + z_{i,n}, \qquad (1)$$

where $x_{i,n}$ is the $i$th DCT coefficient in frame $n$, and $x_{i',n-1}$ is the DCT coefficent in the previous frame that was mapped to it by motion compensation. It is reasonably assumed that $z_{i,n}$ is zero-mean, stationary, and independent of $x_{i',n-1}$.

Let us assume that due to a packet loss at the enhancement layer a block needs to be reconstructed. We first use the information provided by the (error-free) base layer that determines via its quantization interval a corresponding interval in which the original DCT coefficient value must lie:

$$x_{i,n} \in (a_{i,n}, b_{i,n}). \qquad (2)$$

Further, let $\tilde{x}^e_{i',n-1}$ be the decoder-reconstruction (which may be corrupted and hence differ from the encoder reconstruction) of the corresponding DCT coefficient in the previous enhancement layer frame. Taking into account all the available information, we employ the best estimate to reconstruct the lost DCT coefficient:

$$\tilde{x}^e_{i,n} = E\{x_{i,n} | \tilde{x}^e_{i',n-1}, x_{i,n} \in (a_{i,n}, b_{i,n})\}. \qquad (3)$$

Note how the estimate accounts for prior enhancement-layer information, while maintaining consistency with the quantization interval specified by the (error-free) current base-layer reconstruction.

The estimate in (3) may be conveniently approximated by the reconstruction

$$\tilde{x}^e_{i,n} = \rho \tilde{x}^e_{i',n-1} + E\{z_{i,n} | z_{i,n} \in (c_{i,n}, d_{i,n})\}, \qquad (4)$$

where,

$$\begin{aligned} c_{i,n} &= a_{i,n} - \rho \tilde{x}^e_{i',n-1}, \\ d_{i,n} &= b_{i,n} - \rho \tilde{x}^e_{i',n-1}. \end{aligned} \qquad (5)$$

To evaluate this expectation we employ an appropriate probabilistic model for $z_{i,n}$. It is well known that the marginal density function of the DCT coefficient is approximated by a Laplacian distribution. If $x_{i,n}$ is a Markov-Laplace process, then the density of $z_{i,n}$ is [8]:

$$p_{z_{i,n}}(z) = \rho^2 \delta(z) + \frac{(1-\rho^2)}{2} \alpha \exp\{\frac{-|z|}{\alpha}\}. \qquad (6)$$

The parameters $\rho$ and $\alpha$ can be estimated from a training set. We found that $\rho \approx 1$ for "low and intermediate frequency" DCT coefficients. The optimal estimate computation consists of calculating the centroid of the quantization interval (specified by the base layer) with respect to the density of (6), and is therefore of moderate complexity.

### 2.2 Post-processing to mitigate error propagation

In addition to concealing blocks whose packets are missing, the proposed estimation technique is beneficial in reconstructing blocks whose packets have arrived at the decoder. Even though the quantized prediction error is correctly received, the previous-frame block used for prediction might be corrupted due to prior packet losses. The damage due to error propagation can be mitigated by employing the optimal estimation algorithm as a post-processing step. The post-processing is carried out as follows:

- Reconstructed DCT coefficient $\tilde{x}^e_{i,n}$ is first calculated by the usual reconstruction procedure for received blocks.
- If $\tilde{x}^e_{i,n} \in (a_{i,n}, b_{i,n})$, the reconstructed DCT coefficient is consistent with the corresponding base layer quantizer interval. No post-processing is performed.
- Otherwise, $\tilde{x}^e_{i,n}$ lies outside its respective interval due to error propagation. Hence it is reconstructed through optimal estimation according to (3).

Simulation results show that this post-processing step yields additional significant gains in performance.

## 3. Simulation Results

For the simulations, we employed a modified version of the UBC H.263+ decoder [9]. The RTP payload format [10] is assumed for packetization and each packet contains a single GOB. A random packet loss generator is used to drop packets at a specified loss rate. The QCIF sequences “carphone”, “foreman”, “mother&daughter” and the CIF sequence “LTS” are compressed by the original UBC H.263+ encoder. The PSNR of luminance reconstruction is computed for the sequences and averaged over 30 different channel realizations (with different packet loss patterns).

The proposed scheme, frequency domain error concealment (FDEC), is compared with two standard error concealment schemes: (i) Upward Error Concealment (UEC) which uses base-layer data [2], and (ii) Forward Error concealment (FEC) which uses the motion-compensated previous enhancement-layer data [3]. As additional reference we use the the Switched per-Pixel Error Concealment (SPEC) we recently proposed [6]. SPEC performs error concealment in the pixel domain. To reconstruct each missing pixel, either the base-layer pixel or the corresponding previous enhancement-layer pixel is chosen according to the base-layer information and the enhancement-layer loss history.

Figure 1 presents reconstructed PSNR versus packet loss rate of enhancement layer for the sequence “carphone”.

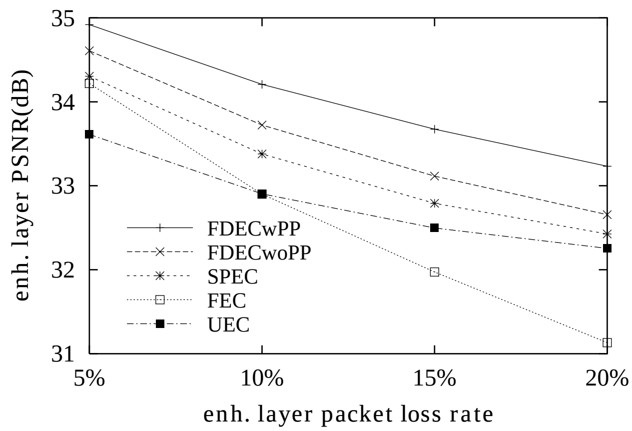


**Figure 1. PSNR vs. enhancement layer packet loss rate for concealment methods: FDEC(proposed), SPEC [6], FEC [3], UEC [2]. Sequence “carphone”, frame rate=30fps, total bit rate=200kbps, enhancement layer rate=75% of total rate.**

Figure 2 depicts PSNR versus enhancement layer bit rates (as a fraction of total bit rate). Figure 3 gives the frame-by-frame evolution of the reconstruction PSNR. Table 1 presents the PSNR of reconstructed video for the four video sequences.

The results strongly support the claim that, while the relative performance of UEC and FEC depends on the packet loss rate and the enhancement-layer bit rate, FDEC yields consistent and substantial gains over them. It also outperforms the SPEC algorithm that operates in the spatial domain. The gains due to post-processing are illustrated in Figure 1 where we compare the performance of FDEC algorithms with and without post-processing of correctly received blocks - (FDECwPP) and (FDECwoPP), respectively. As results in Figure 2 indicate that the gains from FDEC increase as the enhancement layer is allocated a larger portion of the total bit rate.

## 4. Conclusion

We propose a frequency domain error concealment method for blocks whose enhancement-layer packets have not arrived at the decoder. The method exploits all the information available to the decoder within an estimation-theoretic framework. Furthermore, the scheme enables post-processing of correctly received blocks to mitigate the effects of error propagation. Simulation results demonstrate that this algorithm significantly improves the robustness of scalable video coding to packet loss.

**Table 1. Performance 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 sequences, 1000kbps for CIF sequence.**

| Sequence | FDEC | SPEC | UEC | FEC |
|---|---|---|---|---|
| Carphone | 34.21dB | 33.38dB | 32.90dB | 32.90dB |
| Foreman | 31.32dB | 30.74dB | 30.46dB | 30.27dB |
| Mother&Daughter | 35.16dB | 33.94dB | 33.57dB | 33.50dB |
| LTS | 33.19dB | 32.17dB | 31.78dB | 32.09dB |

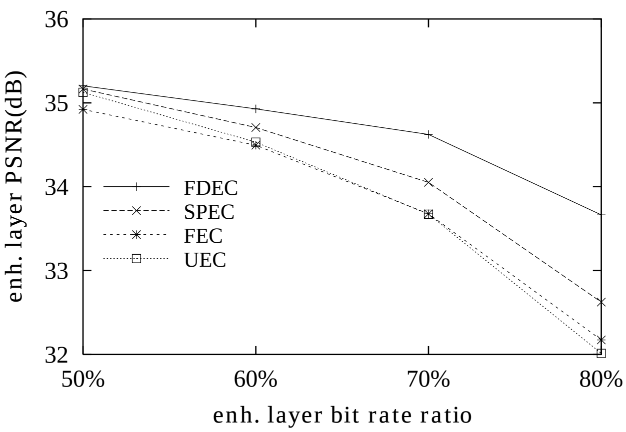


**Figure 2. PSNR vs. enhancement layer rate (as percentage of total rate) for concealment methods: FDEC(proposed), SPEC [6], FEC [3], UEC [2]. Sequence "carphone", frame rate=30fps, total bit rate=200kbps, enhancement layer packet loss rate=10%.**

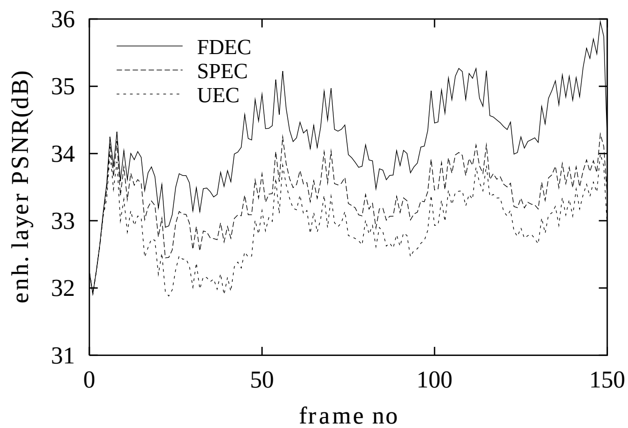


**Figure 3. Frame-by-Frame PSNR of concealment methods: FDEC, SPEC [6], UEC [2]. Sequence "carphone", frame rate=30fps, total bit rate=200kbps, enhancement layer packet loss rate=10%, enhancement layer rate=75% of total rate.**

## References


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

[2] 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.

[3] 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, pp. 238-247, June 1993.

[4] C. Hahm and J. Kim, "An adaptive error concealment in SNR scalable system," *Proceedings of the SPIE, Visual Communications and Image Processing '95*, vol. 2501, pp. 1380-7, May 24-26, 1995, Taipei, Taiwan.

[5] M. Gallant and F. Kssentini, "Rate-distortion optimal joint source/channel coding for robust and efficient low bit rate packet video communications," *International Conference on Image Processing, ICIP00*, Sept. 2000, Vancouver, Canada.

[6] R. Zhang, S. L. Regunathan and K. Rose, "Switched Error Concealment and Robust Coding Decisions in Scalable Video Coding," *International Conference on Image Processing, ICIP00*, Sept. 2000, Vancouver, Canada.

[7] K. Rose and S. L. Regunathan, "Toward Optimal Scalability in Predictive Video Coding," *International*

*Conference of Image Processing, ICIP97*, vol. 3, pp. 432-435, Oct. 1997, Santa Barbara, CA, USA.

[8] N. Farvardin and J. W. Modestino, “Rate-distortion performance of DPCM schemes for autoregressive sources,” *IEEE Trans. on Information Theory*, vol. 31, pp. 402-18, May. 1985.

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

[10] “RTP Payload Format for the 1998 Version of ITU-T Rec. H.263 Video (H.263+),” Internet Draft, RFC2429, http://www.faqs.org/rfcs/rfc2429.html.