ROBUST AND EFFICIENT SCALABLE VIDEO CODING WITH LEAKY PREDICTION

Sangeun Han and Bernd Girod

Information Systems Laboratory
Stanford University, Stanford CA 94305
{sehan,bgirod}@stanford.edu

ABSTRACT

This paper presents a robust scalable video coding scheme with leaky prediction, suitable for time-varying error-prone channels, such as the Internet or wireless channels. The proposed scheme shows how leaky prediction can be utilized for efficient motion-compensated video coding, as well as for robust video transmission over packet erasure networks. We show that leaky prediction with a temporal prediction coefficient of less than one increases the robustness in the presence of errors significantly, while introducing only a small loss in coding efficiency in the error-free case. The proposed scalable video coding scheme is combined with a prioritized ARQ scheme in which higher priority is assigned to more important layers. Our simulation results demonstrate graceful degradation with an increasing error rate over wireless links.

1. INTRODUCTION

Scalable video codecs, generating bitstreams decodable at different bitrates, have been proposed for heterogeneous networks, where users must cope with different receiver capabilities and changing transmission conditions, such as available bitrate and packet loss rate. The MPEG-4 standard supports Fine-Granular-Scalability (FGS) [1]; the forthcoming H.26L standard also incorporates scalability [2]. Scalable codecs have been shown to outperform single-layer codecs for Internet video-on-demand or multicast [3]. Additionally, scalable video streams can benefit networks that support priority mechanisms such as DiffServ [4].

While providing rate adaptability, scalable coding is less efficient than non-scalable video coding optimized for a particular transmission bitrate. This coding inefficiency is mainly due to the reduced exploitation of the temporal dependency between adjacent frames. Even if high quality reference frames are available at the encoder, the base layer codec does not use them for prediction because the enhancement layer bitstream is not guaranteed to arrive error-free at the decoder over error-prone networks. Recently, techniques that include a certain amount of enhancement layer information into the prediction loop have been proposed to improve the coding efficiency of FGS [5][6]. However, to minimize the effect of drift introduced by prediction mismatch, higher computational effort is required at the decoder.

In this paper we propose a novel scheme for improving motion-compensation in scalable coding by introducing leakage into the prediction loops *not* in the base layer. The proposed structure partially employs enhancement layer information in the prediction loop, considerably improving the overall coding efficiency while preserving bitrate scalability. In addition, leaky prediction limits error propagation in the less protected enhancement layers.

The remainder of this paper is organized as follows. In Section 2, we describe the proposed scalable coding scheme with leaky prediction. We then describe methods of mitigating the prediction mismatch for channel errors in Section 3. In Section 4, we present our simulation results, which demonstrate the efficiency and robustness of the proposed scalable video coding scheme.

2. SCALABLE VIDEO CODEC

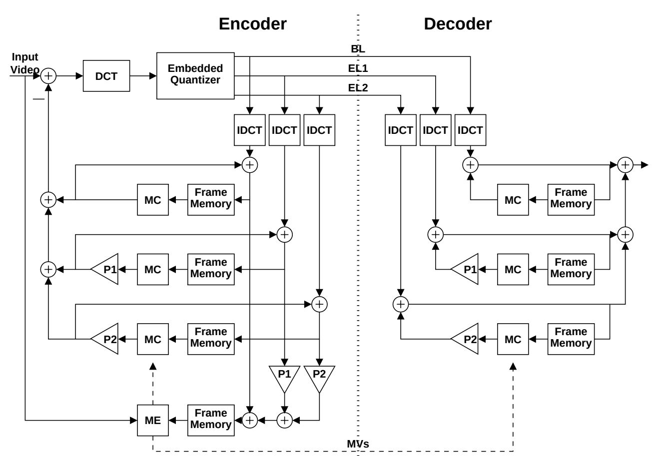


Fig. 1. Structure of the proposed scalable video codec

2.1. Encoder

The proposed system, depicted in Figure 1, uses the motion estimation and compensation algorithm of the H.26L codec TML8.7 [7] which performs a block-based motion estimation with variable block sizes and 1/4-pel accuracy. Multiple reference frames can be also used. Instead of scalar quantization, we use an embedded multi-stage quantizer to generate a layered representation of DCT coefficients of the displaced frame difference (DFD).

The key advantage of our scheme is that prediction is based on the base layer *and* the enhancement layers. We refer to this as **Inter Layer Prediction**. Typically, the base layer is encoded at a lower bitrate to facilitate its reliable delivery. Thus, the motion-compensated prediction of the next frame using only the base layer is imprecise. Higher quality reference frames, a weighted sum of reconstructed pictures in each layer, provide more accurate motion-compensated prediction for the base layer to improve coding efficiency.

While our scheme benefits from Inter Layer Prediction, it also makes use of **Intra Layer Prediction**, in which motion compensation is applied within a single quality layer. This results in prediction mismatch in the base layer when the transmission bitrate drops and cannot support all the layers. Figure 2 illustrates the coding dependencies for the 2-layer case.

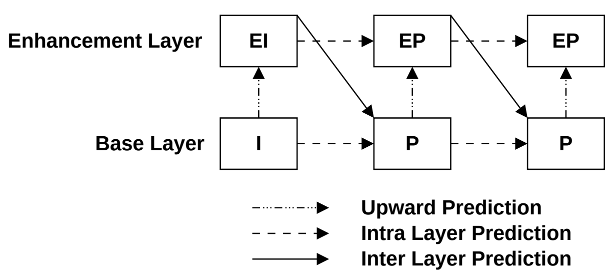


Fig. 2. Coding dependencies for a two-layer encoder

For our work, we use a simple embedded quantizer; the index in each layer is independently entropy coded. A better embedded coding structure, such as EZW, SPIHT, or bitplane coding of DCT combined with efficient entropy coding, can replace the quantizer to improve the performance. A fixed quantization parameter is used for the entire sequence, and rate control is not implemented.

2.2. Decoder

At the decoder, shown in Figure 1, each layer has an *independent* motion compensation loop. As shown in the Figure, we add leaky prediction for the enhancement layers, by introducing gain factors P_1 and P_2, both less than 1, in the DPCM feedback loop. Side information specifying the prediction mode and base layer motion vectors is *reused* to decode enhancement layer pictures, however. Simulation results in Section 4.2 will show that the propagation of channel errors to subsequent frames is reduced at the cost of less prediction gain by this leaky prediction.

3. MITIGATION OF ERROR PROPAGATION

At the encoder, the residual error of motion compensation remains after Inter Layer Prediction. At the decoder, each layer creates a reconstructed picture by combining its Intra Layer Prediction with the portion of the dequantized bitstream corresponding to the layer. Because the final reconstructed picture is the sum of the pictures from each layer, the amount of image distortion introduced for a given error event, e.g., a packet loss in the enhancement layer, will propagate to successive frames and remain visible for a long period. We discuss two approaches to mitigate the effect of error propagation in this section.

3.1. Leaky Prediction

Leaky prediction is a well-known technique to increase error robustness by trading off coding efficiency [8]. In [9] a similar optimization problem has been investigated for a DPCM system with a predictor optimized with respect to the characteristic of a particular noisy channel. Since the leakage introduced by spatial filtering in a motion-compensated predictor is not strong enough for error resilience [10], we have added additional gain factors P_i : $P_i < 1$ in the enhancement layers to severely attenuate the prediction signal. The effectiveness of leaky prediction is shown in Figure 3 for a 2-layer system. After the loss of one GOB in the enhancement layer, the loss in picture quality, $\Delta PSNR$, decreases faster with a smaller prediction coefficient.

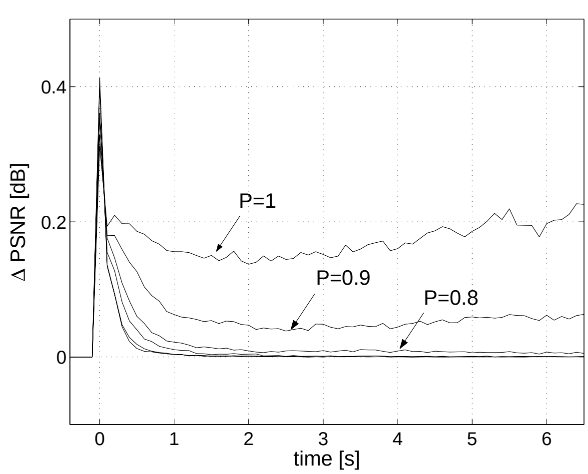


Fig. 3. Loss recovery in PSNR with leaky prediction

3.2. Intra Layer Update

Intra layer update is the extreme case of leaky prediction, corresponding to leaky prediction coefficients of zero in each enhancement layer; i.e., the prediction is based only on low quality reference frames in the base layer, and the enhancement layers are coded using Upward Prediction. The periodic intra layer update removes the prediction mismatch in the base layer due to Inter Layer Prediction of the proposed scheme. This is analogous to the regular insertion of I-frames in non-scalable coding to stop interframe error propagation.

4. SIMULATION RESULTS

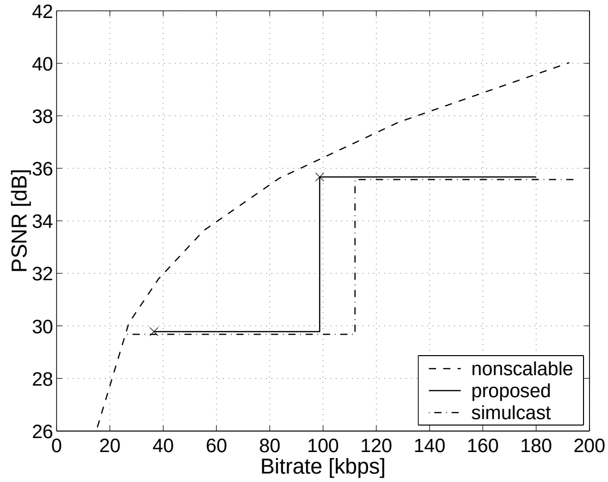
We present two experiments performed with the proposed scalable video codec. In the first simulation, we compare the rate-distortion (RD) performance of our scheme to that of non-scalable coding and multiplexed bitstream, running independent encoders in parallel (*simulcast*). To demonstrate the robustness of our proposed scheme, in the second part of our experiments we show how the error propagation will be reduced by leaky prediction. We have used the test sequence *Foreman* (QCIF, 30 Hz, 300 frames) at a constant frame rate of 10 Hz for all simulations.

4.1. Rate-Distortion Performance

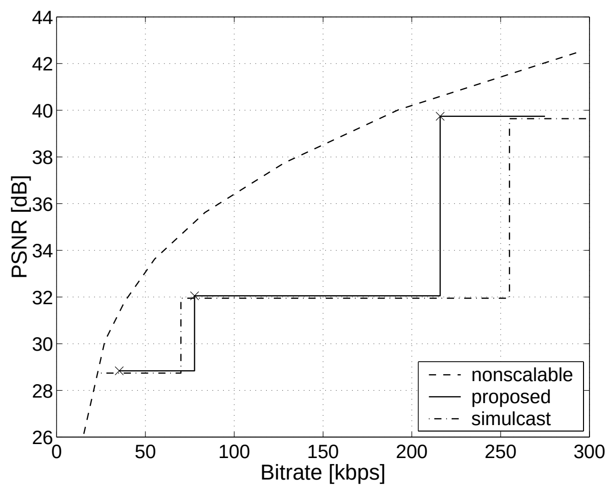
We encode the test sequence with a given quantization parameter and only one I-frame at the beginning of the whole sequence. As a reference system, we use the non-scalable H.26L TML8.7 [7]. In our proposed scheme, the RD performance of the base layer is affected by the enhancement layer because the higher quality reference frames including the enhancement layer are used for prediction. While no leakage, i.e., $P_1 = P_2 = 1$ in Figure 1, certainly gives the best RD performance for the enhancement layer, it causes a drift in the base layer that is intrinsic to the system, resulting in the base layer picture quality degrading over time. The base layer performance in Figure 4 is obtained by the two approaches discussed in Section 3 with leaky prediction coefficients $P_1 = P_2 = 0.9$, eliminating the downward drift in base layer picture quality. The overall bitrate is only slightly inferior to the non-scalable coding method of H.26L. In addition, the proposed scheme offers a 10-15% bitrate reduction compared to simulcast.

4.2. Performance with Packet Loss

To evaluate the performance of our scheme in a packet loss scenario, we simulate streaming video from a server on the Internet to a user's mobile over a 3G UMTS radio link. Once the video sequence is encoded, we packetize it into a RTP bitstream and use the software provided in [11] for



(a) 2-layer



(b) 3-layer

Fig. 4. Comparison between simulcast and the proposed scalable coding

framing at the PDCP/PPP level. We apply bit-error patterns to the RLC/RLP frames using the error pattern files provided with the software, but varying starting positions in the error files. We assume that RTP packets are either received correctly or dropped completely.

For non-scalable coding, I-frames are inserted every 10 frames for error robustness, which increases the overall bitrate to that of the scalable coded bitstream. We have three RTP packets for each frame. If the decoder receives an error indication, the corresponding GOBs from the previous frame are simply copied to the current frame buffer (*previous frame concealment*). For scalable coding, on the other hand, the base layer is packed into one packet, and the enhancement layer is divided into two packets. In this case no error concealment technique is used. However, the base layer receives higher protection through priority retransmission with explicit marking, resulting in the fact that the base layer is guaranteed to transmit without any loss.

In addition to the notable RD performance, Figure 5 shows that our proposed scheme shows further advantages in error-prone networks. We can see significant gains compared to non-scalable coding. In our scheme the quality does not decrease dramatically as the packet loss rate increases. In non-scalable coding, however, the picture quality degrades severely at high packet loss rates, even though I-frames are periodically inserted. If the I-frame packet happens to be lost, the next 9 P-frames suffer from error propagation. Note that the proposed coding structure achieves data partitioning as a side effect and protects the motion information in the base layer for the decoding of all layers. Moreover, with leaky prediction, another 1 dB gain is obtained for the higher packet loss rate.

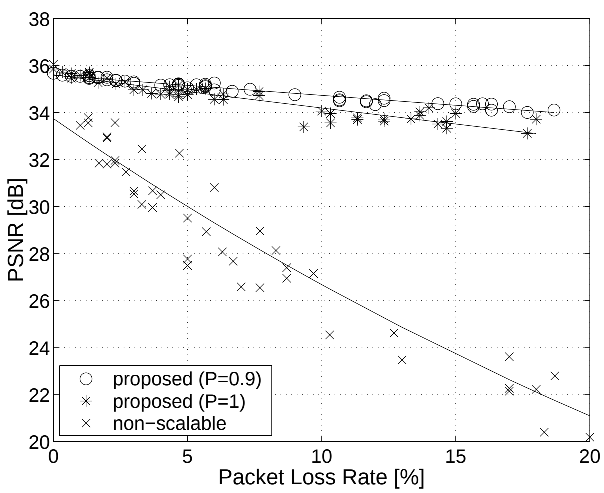


Fig. 5. Performance of different schemes under lossy transmission

5. CONCLUSIONS

We have presented a new scalable video coding method, which utilizes higher quality reference frames for prediction in order to improve coding efficiency. The overall rate-distortion performance is close to that of non-scalable coding, and our scheme provides several quality levels at a bitrate significantly lower than simulcast. Error propagation due to the prediction mismatch is effectively controlled by leaky prediction in the enhancement layers and Intra Layer Update. Simulation results show that the leakage built into the enhancement layers helps to combat channel errors. We report significant gains of up to 1 dB in PSNR compared to scalable coding without leaky prediction and 10 dB to non-scalable bitstream for higher packet loss rates.

6. REFERENCES

[1] ISO/IEC 14496-2, "MPEG-4 video FGS v. 4.0," N3317, Noordwijkerhout, NL, March 2000.

[2] G. S. Greenbaum, "Remarks on the H.26L project: streaming video requirements for next generation video compression standards," ITU-T SG16/Q15, Q15-G-11, Monterey, CA, February 1999.

[3] B. Girod, N. Färber, and U. Horn, "Scalable codec architectures for Internet Video-on-Demand," in *Proceedings of 1997 Asilomar Conference on Signals and Systems*, Pacific Grove, CA, USA, November 1997.

[4] J. Kimura, F. Tobagi, J. Pulido, and P. Emstad, "Perceived quality and bandwidth characterization of layered MPEG-2 video encoding," in *Proceedings of the SPIE International Symposium on Voice, Video and Data Communications*, Boston, Mass, September 1999.

[5] M. van der Schaar and H. Radha, "Motion-Compensation Fine-Granular-Scalability (MC-FGS) for wireless multimedia," in *Proceedings of IEEE Symposium on Multimedia Signal Processing (Special Session on Mobile Multimedia Communications)*, October 2001.

[6] F. Wu, S. Li, and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 332–344, March 2001.

[7] G. Bjontegaard, *H.26L Test Model Long Term Number 8 (TML-8) draft 0*, ITU-T Standardization Sector, June 2001.

[8] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[9] K. Chang and R. Donaldson, "Analysis, optimization, and sensitivity study of differential PCM systems operating on noisy communication channels," *IEEE Transactions on Communications*, vol. COM-20, no. 3, pp. 338–350, June 1972.

[10] N. Färber, *Feedback-Based Error Control for Robust Video Transmission*, Ph.D. thesis, University Erlangen-Nuremberg, Germany, 2000.

[11] V. Varsa, M. Karczewicz, G. Roth, R. Sjöberg, G. Liebl, and T. Stockhammer, "Common test conditions for RTP/IP pver 3GPP/3GPP2 - software and amendments," ITU-T SG16/Q6, VCEG-N37, Santa Barbara, CA, September 2001.