ENCODER AND DECODER OPTIMIZATION FOR SOURCE-CHANNEL PREDICTION IN ERROR RESILIENT VIDEO TRANSMISSION

Hua Yang and Kenneth Rose

Department of Electrical and Computer Engineering
University of California Santa Barbara, CA, 93106-9560, USA
hua, rose@ece.ucsb.edu

ABSTRACT

Motion-compensated prediction that accounts for loss in the channel is achieved by the source-channel prediction (SCP) method, which is based on the expected decoder reconstruction of past frames (rather than their encoder reconstruction). The decoder reconstruction is estimated by exploiting the recursive optimal per-pixel estimate (ROPE), which explicitly accounts for the quantization distortion, channel loss, error propagation, as well as the decoder operation, and achieves improved error resilience. We take this paradigm further by noting that the decoder can, in turn, be re-optimized to match the modification introduced to the encoder for SCP. Simulation results demonstrate substantial performance gains over conventional decoding. We then examine the benefits of re-optimizing the encoder for the newly matched decoder, and then re-optimizing the decoder, etc., and note that further incremental gains are minor. Hence, one complete round of SCP optimization offers significant gains, but multiple re-optimization iterations may not be cost-effective.

1. INTRODUCTION

Packet loss is inevitable in many video transmission settings including in particular those involving the Internet or various wireless networks. How to effectively mitigate its impact is a critical concern for video-over-network applications. A large variety of error control and concealment approaches have already been proposed [1], including some that focus on modifying the standard motion compensated prediction mechanism, such as video redundancy coding [2] and multiple reference prediction [3][4]. These methods, nevertheless, assume the conventional prediction paradigm, where the motion compensated prediction of the current frame is derived from the *encoder* reconstruction of past frames.

In [5], we proposed the Source-Channel Prediction (SCP) scheme, where the prediction is computed from the *expected decoder* reconstruction of past frames. In contradistinction with the conventional mechanism, SCP effectively takes into account the impact of both source coding distortion, and transmission loss, and thus improves the overall performance in lossy channel environments [5]. The performance of SCP critically depends on the accuracy of the end-to-end distortion estimate, for which purpose we adapted the recursive optimal per-pixel estimate (ROPE) technique [6].

The starting point for this paper is the observation that while SCP offers improvement via modification to the encoder, it leaves the decoder unchanged. However, an optimal decoder must account for modifications to the encoder. In particular, even when all the data are correctly received at the decoder, the mismatch between the conventional decoder and the SCP encoder will lead to additional distortion, and may hence compromise the overall system performance. In this work we match the decoder to the SCP encoder so that no additional distortion is incurred during error free transmission. We note in passing that further decoder optimization is possible to jointly match the SCP encoder and channel error effects. This is the subject of ongoing research. Decoder matching as reported here is shown by simulation to offer significant performance gains under lossy transmission conditions.

On the other hand, recall that for accurate estimation of end-to-end quantities of relevance to SCP, the encoder has to take into account the exact decoding procedure. Hence, the latest modification of the decoder calls, in turn, for a corresponding modification of the original SCP encoder, which may entail further modification of the decoder, and so on. Clearly, in order to make the most of SCP, one may need to iterate encoder and decoder design to convergence.

This paper is concerned with the SCP-based codec design. We provide explicit derivation of the “one round” and “two rounds” of SCP codec design. As our main result we show that SCP provides substantial performance gains after one complete design iteration (whose complexity is modest). We complement this result with the observation (based on two rounds of SCP re-optimization) that additional iterations may not be cost effective as they involve a consid-

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

erable increase in the number of end-to-end quantities to estimate, yet provide only minor additional gains.

The rest of the paper is organized as follows. Section 2 reviews the SCP encoder design. In Section 3, a decoding algorithm is proposed to match the SCP encoder. The procedure for second-round optimization of the SCP codec is briefly outlined in Section 4. Simulation results and conclusions are provided in Section 5 and Section 6, respectively.

2. SOURCE-CHANNEL PREDICTION AT THE ENCODER

In [5] we proposed a source-channel prediction scheme to enhance the error resilience of video transmission. Let f_n^i, $\hat{f}_n^i$ and $\tilde{f}_n^i$ denote the values of pixel i in frame n, in the original, encoder reconstruction, and decoder reconstruction sequences, respectively. Let mv denote the motion vector, and let res_n^i be the prediction error before quantization.

- Conventional prediction:

$$\min_{mv} \sum_{i \in MB} (f_n^i - \hat{f}_{n-1}^{i+mv})^2 \tag{1}$$

$$res_n^i = f_n^i - \hat{f}_{n-1}^{i+mv} \tag{2}$$

- Source-channel prediction:

$$\min_{mv} \sum_{i \in MB} E\{(f_n^i - \tilde{f}_{n-1}^{i+mv})^2\} \tag{3}$$

$$res_n^i = f_n^i - E\{\tilde{f}_{n-1}^{i+mv}\}, \tag{4}$$

where the expectation is over the channel statistics. Note that for convenience we allow some abuse of notation where the minimization of (1) and (3) is over all mv referring to all the possible motion vectors, but elsewhere (throughout the paper) mv denotes the optimal motion vector that achieves the minimum.

The conventional scheme uses the encoder reconstruction for motion compensated prediction. Hence, only the source coding quantization distortion is considered. In contrast, SCP considers the *decoder* reconstruction, and hence takes into account the impact of both the quantization and transmission loss. As shown in [5], SCP is the optimal prediction in the sense of minimum end-to-end mean squared error (MSE) distortion.

The end-to-end quantities required to compute (3) and (4) are determined by the first and second order moments of the decoder reconstructions which can be estimated by the ROPE method [6] as follows:

$$\begin{aligned} E\{\tilde{f}_n^i\} &= (1-p)\cdot(\widehat{res}_n^i + E\{\tilde{f}_{n-1}^{i+mv}\}) \\ &\quad + p\cdot E\{\tilde{f}_{n-1}^i\} \end{aligned} \tag{5}$$

$$\begin{aligned} E\{(\tilde{f}_n^i)^2\} &= (1-p)\cdot E\{(\widehat{res}_n^i + \tilde{f}_{n-1}^{i+mv})^2\} \\ &\quad + p\cdot E\{(\tilde{f}_{n-1}^i)^2\} \end{aligned} \tag{6}$$

where p is the packet loss rate in the channel. $\widehat{res}_n^i$ denotes the quantized prediction error, which is part of the transmitted data.

Note that in (5) and (6) a conventional decoder is assumed, which can be explicitly described as follows:

- If $\widehat{res}_n^i$ and mv are correctly received:

$$\tilde{f}_n^i = \tilde{f}_{n-1}^{i+mv} + \widehat{res}_n^i \tag{7}$$

- If $\widehat{res}_n^i$ and mv are lost:

$$\tilde{f}_n^i = \tilde{f}_{n-1}^i \tag{8}$$

3. THE SCP MATCHED DECODER

So far, SCP has only been introduced at the encoder, while decoding is conducted in the conventional way as in (7) and (8). In fact, the SCP encoder is optimized for the conventional decoder. However, having modified the encoder, we must revisit the decoder and consider its re-optimization to match the new encoder.

To optimize the decoder, we first identify the mismatch under the simpler case of error-free transmission. Specifically, as the conventional decoder does not account for the actual SCP scheme adopted at the encoder, the received prediction error data will be misinterpreted as being generated from (2), but not (4). Therefore, even when all the data are correctly received at the decoder, the mismatch of the conventional decoder with the SCP encoder results in additional reconstruction error. We propose a simple decoding algorithm as shown below, which completely eliminates the mismatch. Its efficacy is well supported by the significant performance gains observed in the simulations summarized in Section 5.

Assume that p is also available at the decoder.

- If $\widehat{res}_n^i$ and mv are correctly received:

$$\tilde{f}_n^i = E\{\tilde{f}_{n-1}^{i+mv}\}_d + \widehat{res}_n^i \tag{9}$$

$$\begin{aligned} E\{\tilde{f}_n^i\}_d &= (1-p)\cdot(E\{\tilde{f}_{n-1}^{i+mv}\}_d + \widehat{res}_n^i) \\ &\quad + p\cdot E\{\tilde{f}_{n-1}^i\}_d \end{aligned} \tag{10}$$

- If $\widehat{res}_n^i$ and mv are lost:

$$\tilde{f}_n^i = \tilde{f}_{n-1}^i \tag{11}$$

$$E\{\tilde{f}_n^i\}_d = (1-p)\cdot X_1 + p\cdot E\{\tilde{f}_{n-1}^i\}_d \tag{12}$$

For error concealment: $X_1 = \tilde{f}_{n-1}^i$.

Here, $E\{\tilde{f}_{n-1}^{i+mv}\}_d$ is the quantity calculated by the decoder to emulate or track the SCP estimate $E\{\tilde{f}_{n-1}^{i+mv}\}$, which is used by the encoder.

4. THE TWICE RE-OPTIMIZED SCP CODEC

Once the optimized decoding algorithm takes effect at the decoder, a new mismatch arises at the encoder. The SCP encoder must account for the exact decoding process so as to accurately estimate end-to-end quantities such as the decoder reconstruction. So far, the conventional decoder has been assumed for such estimation as in (5) and (6), which is inconsistent with the matched SCP decoder. To resolve this mismatch, the SCP encoder should be re-modified accordingly, followed by re-modification of the decoder, and so on. In principle, one may expect that, to achieve optimal performance, the SCP encoder and decoder design would be iterated until convergence.

Herein, we examine two-round optimization of the SCP codec. We first redesign the encoder. It is easy to see that (3) and (4) still hold, except that we need to re-derive the estimates therein. Note that the decoder's tracking estimate $E\{\tilde{f}_{n-1}^{i+mv}\}_d$ is a random variable for the encoder. Based on (9) and (11), we obtain:

$$\begin{aligned} E\{\tilde{f}_n^i\} &= (1-p)\cdot(E\{E\{\tilde{f}_{n-1}^{i+mv}\}_d\} + \widehat{res}_n^i) \\ &\quad + p\cdot E\{\tilde{f}_{n-1}^i\} \qquad (13) \\ E\{(\tilde{f}_n^i)^2\} &= (1-p)\cdot E\{(E\{\tilde{f}_{n-1}^{i+mv}\}_d + \widehat{res}_n^i)^2\} \\ &\quad + p\cdot E\{(\tilde{f}_{n-1}^i)^2\} \qquad (14) \end{aligned}$$

In this case, two new end-to-end quantities are involved, $E\{E\{\tilde{f}_{n-1}^{i+mv}\}_d\}$ and $E\{(E\{\tilde{f}_{n-1}^{i+mv}\}_d)^2\}$, whose recursive estimation formula can be similarly derived from (10) and (12). (The latter 2nd order term explicitly shows up once (14) is expanded. For space considerations we omit their estimation formula here.) It is important to note that cross-correlation terms arise in computing $E\{(E\{\tilde{f}_{n-1}^{i+mv}\}_d)^2\}$. In practice, to estimate them with tractable complexity, the estimation accuracy will be slightly and inevitably compromised (see [7] for analysis and practical approach to this problem in the context of plain ROPE).

The twice re-optimized SCP decoder can be similarly constructed in a straightforward manner (omitted here), as well as further rounds of SCP codec re-optimization. One should expect additional iterations of SCP re-optimization to involve more end-to-end quantities, and hence higher complexity. Fortunately, simulation results suggest that most of the available gains are captured by one round of codec re-optimization, while further iterations offer diminishing returns that appear not to justify the complexity incurred.

5. SIMULATION RESULTS

Our simulation adopts the UBC H.263+ codec. System performance is measured by the average luminance PSNR over the results of 50 different packet loss patterns. We examine the performance of five different codec solutions: conventional encoder with conventional decoder, i.e., sorce-based prediction referred to as "SP codec"; SCP encoder with conventional decoder, referred to as "SCP1 enc"; the outcome of a complete round of SCP codec optimization (SCP encoder with matched decoder) is denoted "SCP1 codec"; the 2nd round SCP encoder to match the 1st round re-optimized SCP decoder, is denoted "SCP2 enc"; the outcome of two complete rounds of SCP codec re-optimization is referred to as "SCP2 codec". Note that to estimate the cross-correlation terms appearing in the 2nd round optimization of the encoder, we use the Model II method proposed in [7].

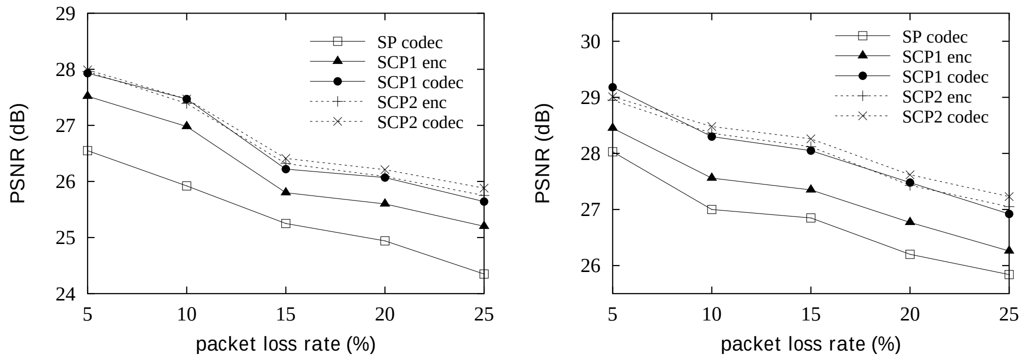
Fig. 1 shows the results under the scenario of periodic Intra updating, where an MB is coded in Intra mode once per $1/p$ frames. The results in Fig. 2 are obtained with RD optimal Intra updating, where the Inter/Intra mode selection of each MB is optimized by the ROPE technique of [6]. It is worth noting that while we show results for two specific test sequences, similar results were obtained with several other sequences.

It is observed that the "SCP1 codec" consistently outperforms both the conventional codec and "SCP1 enc" with significant gains. For example, for $p = 15\%$ in Fig. 1 (a), (b), Fig. 2 (a) and (b), the gains of "SCP1 codec" over "SP codec" are 0.97dB, 1.20dB, 0.84dB and 0.50dB, respectively, while the gains of "SCP1 codec" over "SCP1 enc" are 0.42dB, 0.70dB, 0.40dB and 0.47dB, respectively. The results provide experimental evidence for the benefit of accounting for channel loss statistics in motion-compensated prediction, and for re-optimizing the decoder to match the encoder revision.

Another observation is that, in all cases, the performance gain of "SCP2 codec" over "SCP1 codec" is relatively minor. From this, we expect that multiple rounds of SCP codec re-optimization provide diminishing returns in terms of performance gains and are usually not cost-effective. Also note that sometimes "SCP2 enc" or "SCP2 codec" performs worse than "SCP1 codec", e.g., in Fig. 2 (b). We attribute this to the approximation implicit in the estimation of the cross-correlation terms, as mentioned in Section 4, which also impacts the performance gains due to the additional SCP re-optimizations.

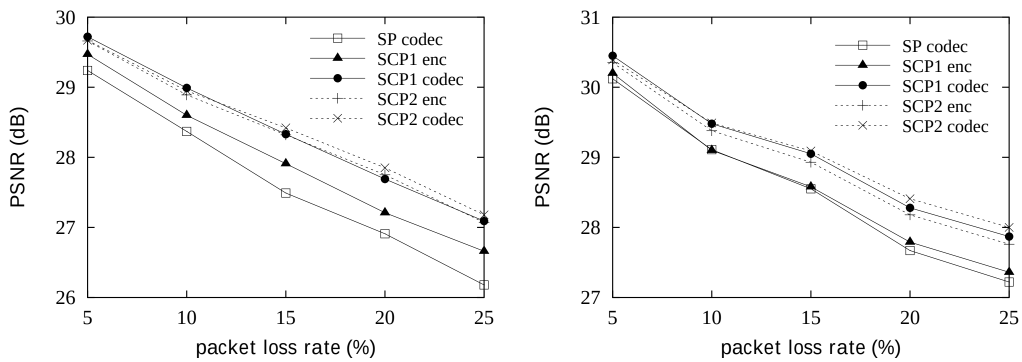
6. CONCLUSIONS

This paper builds on our prior work on source-channel prediction. First, we identify the decoder mismatch problem, and propose a procedure to re-optimize the SCP decoder to match the original SCP encoder. Experimental evidence supports the substantial performance gains due to this extension. We then consider the question of multiple rounds of SCP codec re-optimization. In particular we examine the "twice re-optimized" SCP codec. The minor additional



(a) News (b) Carphone

Fig. 1. Performance comparison with periodic Intra updating. QCIF, 10f/s, 50kb/s.



(a) News (b) Carphone

Fig. 2. Performance comparison with RD optimized Intra updating. QCIF, 10f/s, 50kb/s

gains observed in experiments suggest that additional rounds of SCP optimization may not be cost-effective. Research in progress is focused on investigation of the potential of optimizing the SCP decoder while accounting for both the SCP encoder in use *and* all effects of transmission loss including error propagation. Preliminary results suggest that substantial gains can be recouped by such a comprehensive approach.

7. REFERENCES

[1] Y. Wang and Q. F. Zhu, “Error control and concealment for video communication: a review,” *Proc. of the IEEE*, vol. 86, p. 974-97, May 1998.

[2] S. Wenger, “Video redundancy coding in H.263+,” *Proc. AVSPN*, Aberdeen, U.K., Sept. 1997.

[3] M. Budagavi and J. Gibson, “Error propagation in motion compensated video over wireless channels,” *Proc. of IEEE ICIP*, p. 89-92, Oct. 1997.

[4] T. Wiegand, N. Farber, K. Stuhlmuller and B. Girod, “Error-resilient video transmission using long-term memory motion-compensated prediction,” *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, , p. 1050-62, June 2000.

[5] H. Yang and K. Rose, “Source-channel prediction in error resilient video coding,” *Proc. of IEEE ICME 2003*, Aug. 2003.

[6] R. Zhang, S. L. Regunathan, and K. Rose, “Video coding with optimal intra/inter mode switching for packet loss resilience,” *IEEE Journal of Selected Areas in Communications, special issue on Error-Resilient Image and Video Transmission*, vol. 18, no. 6, p. 966-76, June 2000.

[7] H. Yang and K. Rose, “Recursive end-to-end distortion estimation with model-based cross-correlation approximation,” *Proc. of IEEE ICIP 2003*, Sept. 2003.