SOURCE-CHANNEL PREDICTION IN ERROR RESILIENT VIDEO CODING

Hua Yang and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560

ABSTRACT

Motion compensated prediction in conventional video coders is based on the encoder reconstruction of the previous frame where motion is estimated so as to minimize the encoder prediction error. Such design paradigms optimize the prediction at the encoder and ignore the effects of packet loss. In this paper, we propose a new prediction framework, which uses the expected decoder reconstruction of past frames for prediction. Moreover, the motion estimation criterion is accordingly modified to minimize the expected decoder prediction error. The approach builds on the recursive optimal per-pixel estimate (ROPE), which ensures accurate estimation of the expected values of decoder quantities. The proposed prediction scheme achieves better error resilience performance than its conventional counterpart.

1. INTRODUCTION

A widely recognized critical concern with virtually all video-networking applications is that of efficient error resilience to adequately mitigate the impact of packet loss. A large variety of error control and concealment approaches have been proposed so far (see recent review in [1].)

Since the damage due to packet loss is greatly exacerbated by error propagation, many error resilience techniques are focused on the prediction mechanism. For example, the H.263 standard divides a picture into non-overlapping spatial regions, i.e. slices, and limits spatial and temporal prediction within each slice. Video redundancy coding (VRC) suppresses the effect of temporal error propagation by partitioning input video frames into several independently predicted groups, the so-called threads [2]. When an error occurs in one frame, only one thread will be affected so that the video signal can still be well reproduced with other correctly received threads. Some other approaches involve multiple reference frames, where one appropriate frame is selected out of several previously coded frames for prediction, subject to prediction accuracy and error resilience considerations [3][4]. While different in many respects, these methods have one feature in common: they assume the same underlying prediction, which uses past encoder-reconstructed frames for prediction, and performs motion estimation under the minimum prediction error criterion. The approach we pursue in this work is to modify conventional motion compensated prediction so as to improve or optimize the error resilience performance of the overall system.

The approach here is a source-channel coding approach. Traditionally source and channel coding were separated, as motivated by Shannon's celebrated theorem. However, at finite complexity and delay separate design is in general suboptimal. In fact, most encoder-based error resilient methods embody this realization by attempting to account for packet loss effects during encoding [1].

Continuing along this line of thought, we propose a new prediction scheme in this paper. It takes into account the packet loss effect and uses the expected *decoder* reconstruction for prediction. This is the optimal prediction in the sense of minimum end-to-end mean squared error (MSE) distortion as will be shown in the next section. We also correspondingly modify the motion estimation criterion, which becomes minimum *expected decoder* prediction error.

The success of a method that minimizes decoder prediction error and distortion crucially depends on the accuracy of estimates of expected decoder quantities. So far, much research effort has been devoted to end-to-end distortion estimation (e.g., [5][6]). Herein, we build on the ROPE approach of [6], which optimally estimates the decoder reconstruction and distortion while taking into account virtually all relevant factors including quantization, packet loss, and error concealment. The optimal estimate is derived from the first and second order moments of the decoder reconstruction, which are recursively calculated for each pixel. ROPE was employed to optimize encoding decisions, most notably mode selection. This work expands ROPE's utility to enhance the prediction module. Simulation results show that the resulting source-channel prediction scheme achieves consistently better overall performance than conventional schemes.

The paper is organized as follows. In Section 2, MSE optimality of source-channel prediction is shown. Section 3 provides the details of the proposed motion estimation criterion. Simulation results are summarized in Section 4.

0-7803-7965-9/03/$17.00 ©2003 IEEE

2. PREDICTION BASED ON THE DECODER RECONSTRUCTION ESTIMATE

In video-over-Internet applications, the overall distortion is primarily due to two factors: quantization at the encoder and packet loss in the channel. Conventional prediction schemes use the encoder reconstruction for prediction, and hence only account for quantization noise. In order to improve error resilience we propose a new prediction scheme that employs the *expected decoder* reconstruction for prediction, and takes into account the impact of both quantization and packet loss.

Conventional prediction: $$\bar{f}_n^i = \hat{f}_{n-1}^j \tag{1}$$

Source-channel prediction: $$\bar{f}_n^i = E\{\tilde{f}_{n-1}^j\} \tag{2}$$

Prediction residue: $$res_n^i = f_n^i - \bar{f}_n^i \tag{3}$$

Here, f_n^i and $\bar{f}_n^i$ denote the original and predicted values of pixel i in frame n, respectively, while $\hat{f}_{n-1}^j$ and $\tilde{f}_{n-1}^j$ denote the encoder and decoder reconstruction values for pixel j in frame n-1, which is employed to predict pixel i in frame n (given the motion vector). The quantity res_n^i is the prediction error to be quantized. Note that due to packet loss the decoder reconstruction is viewed as a random variable at the encoder.

Next we show that the proposed prediction in (2) is the optimal prediction that minimizes the expected prediction error at the decoder and, in fact, the overall end-to-end distortion (prior to quantization of the residual). It is necessary to emphasize that we are still making the common assumption that predictor decisions are made prior to quantization. This means that the residue we consider is the prediction residue *prior to quantization*, not the *quantized* residue.

The estimate of MSE end-to-end distortion is given by:

$$\begin{aligned} d_n^i &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\} \end{aligned} \tag{4}$$

As already mentioned, we apply ROPE to accurately estimate the two expected values, namely, the first and second moments in (4). Since there is no prediction in Intra macro-block (MB) coding, we are only interested in Inter mode MB's. For simplicity but without implied loss of generality: (a) We model the channel as a Bernoulli process with packet loss rate p. (b) We assume that data of one frame are carried in one packet. Hence, the pixel loss rate equals the packet loss rate. (c) We assume that to conceal a lost frame it is simply replaced by the reconstructed frame. As in [6] the moments are computed recursively by

$$\begin{aligned} E\{\tilde{f}_n^i\} &= (1-p)\cdot(res_n^i + E\{\tilde{f}_{n-1}^j\}) + p\cdot E\{\tilde{f}_{n-1}^i\} \\ E\{(\tilde{f}_n^i)^2\} &= (1-p)\cdot E\{(res_n^i + \tilde{f}_{n-1}^j)^2\} + p\cdot E\{(\tilde{f}_{n-1}^i)^2\} \end{aligned} \tag{5}$$

Substituting (5) and (3) into (4), we can get:

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

It is easy to see that for a given motion vector (j fixed) distortion of (6) is minimized by (2), which proves the optimality of the proposed source-channel predictor.

3. MODIFIED MOTION ESTIMATION CRITERION

In the conventional scheme, the motion estimation criterion is to minimize the encoder prediction error:

$$\min_{mv} \sum_{i\in MB} \left(f_n^i - \bar{f}_n^i\right)^2 = \min_{mv} \sum_{i\in MB} \left(f_n^i - \hat{f}_{n-1}^{i+mv}\right)^2 . \tag{7}$$

Here, mv denotes the motion vector assigned to the specific MB. Obviously, as we replace the conventional predictor of (1) with the source-channel predictor proposed in (2), we must reconsider and accordingly modify the motion estimation criterion of (7).

It turns out that there is more subtlety to this problem than may initially be expected. One may consider two different motion estimation criteria:

Criterion I: $$\min_{mv} \sum_{i\in MB} \left(f_n^i - E\{\tilde{f}_{n-1}^{i+mv}\}\right)^2 , \tag{8}$$

Criterion II: $$\min_{mv} \sum_{i\in MB} E\{\left(f_n^i - \tilde{f}_{n-1}^{i+mv}\right)^2\} . \tag{9}$$

Note that Criterion I (8) is a natural choice for criterion as it considers the difference between the original pixels and the actual prediction employed at the encoder (which was appropriately modified to reflect the expected reconstruction at the decoder). However, Criterion II (9) computes an expectation over the actual prediction error *at the decoder* and explicitly accounts for the fact that the decoder's predictor is random. In other words, although the best predictor to use is given in (2) and indeed appears in (8), the best motion vector is the one that minimizes (9) which is exactly the expected squared prediction error at the decoder. Clearly, Criterion II is theoretically superior to Criterion I, and is hence adopted in our proposed scheme. Simulation results to support this analysis are included for completeness in the next section.

To further illustrate the difference between the two criteria, we re-express (9) in terms of distortions:

$$\min_{mv}\sum_{i\in MB}E\left\{\left(f_n^i-\tilde{f}_{n-1}^{i+mv}\right)^2\right\}$$
$$=\min_{mv}\sum_{i\in MB}\left[\left(f_n^i-E\{\tilde{f}_{n-1}^{i+mv}\}\right)^2+\left(E\{(f_{n-1}^{i+mv})^2\}-E^2\{\tilde{f}_{n-1}^{i+mv}\}\right)\right] \quad (10)$$
$$=\min_{mv}\left[(1-p)\cdot D_R+D_D\right]$$

where

$$D_R=\sum_{i\in MB}\left(f_n^i-\bar{f}_n^i\right)^2 \quad (11)$$

$$D_D=\sum_{i\in MB}E\left\{\left(f_n^i-\tilde{f}_n^i\right)^2\right\}\ . \quad (12)$$

Bearing the R-D perspective in mind, we denote the squared prediction residue at the encoder by D_R as it directly impacts bit rate, and we denote the end-to-end distortion by D_D. We can see from (10) that Criterion I only considers D_R while Criterion II considers the properly weighted impacts of D_D and D_R.

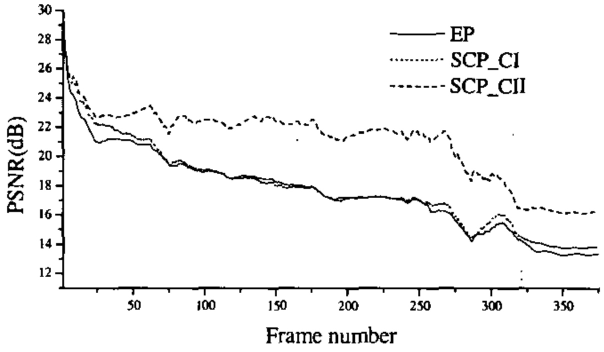
4. SIMULATION RESULTS

Our simulation system is based on the UBC H.263+ codec [7]. A sequence is encoded into an H.263 bitstream given the packet loss rate and total bit rate. The bitstream is then undergoes a packet loss pattern that is randomly generated with the prescribed packet loss rate. System performance is measured by the average luminance PSNR. In the experiment, we averaged over 50 different packet loss patterns.

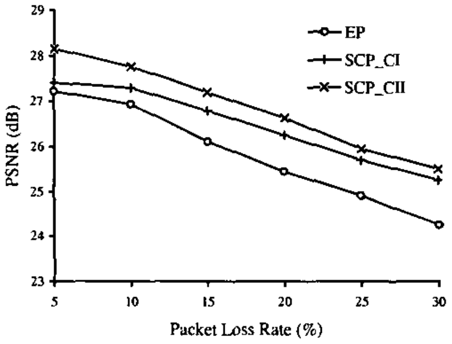
In the following figures, the conventional encoder-based prediction scheme is labeled as "EP". Our proposed source-channel prediction scheme [minimizing criterion II of (9) for motion estimation] is denoted by "SCP_CII". For completeness we also include source-channel prediction whose motion estimation minimizes Criterion I of (8), and denote it "SCP_CI".

We test the above three methods under three different scenarios, namely no Intra updating, periodic Intra updating and R-D optimized Intra updating. In the "no Intra updating" case, all the MB's are coded in Inter mode. Periodic Intra updating means that an MB is coded in Intra mode once per $1/p$ frames, where p is the packet loss rate. R-D optimized Intra updating performs the coding mode selection for each MB with the R-D criterion. While only the results for the QCIF sequence "foreman" are provided here, similar results were obtained in experiments with other sequences including "carphone", "salesman" and "miss_am".

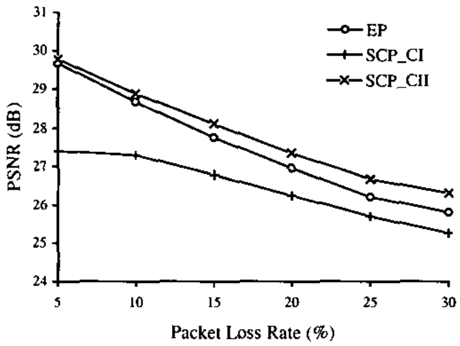
From the figures, it is evident that in all the scenarios and at all packet loss rates the proposed "SCP_CII" method offers the best performance. Note that in simulations, performance improvement not only shows in PSNR gains, but also can be perceptually observed on the actual decoder-reconstructed video signals. This result demonstrates that accounting for packet loss at the prediction stage, by estimating the decoder reconstruction, enhances the error resilience performance of the system.



(a) No Intra updating



(b) Periodic Intra updating



(c) R-D optimized Intra updating

Fig.1 PSNR performance comparison. "foreman", QCIF, 30f/s. (a) No Intra updating: 375 frames, 300kb/s, p=10%. (b) Periodic Intra updating: 1st 150 frames, 200kb/s. (c) R-D optimized Intra updating: 1st 150 frames, 200kb/s.

Another observation is that when Intra updating is efficiently performed, smaller gains are achieved by "SCP_CII" over "EP". For example, as we switch from periodic Intra updating to R-D optimized Intra updating, the maximal performance gain decreases from 1.3dB to 0.5dB. This reflects the fact that the gains depend on how much "damage" was caused by packet loss and was unaccounted for in conventional prediction. Clearly, effective mode-selection combats the effects of error propagation and thereby also reduces the relative gains of SCP_CII. Nevertheless, as shown in Fig. 1 (c) the proposed scheme consistently outperforms the conventional scheme.

We also note that "SCP_CI" always underperforms "SCP_CII", and in R-D optimized Intra updating case it performs much worse than the other two methods, especially at low packet loss rates. This result supports the analysis in Section 3.

5. CONCLUSION

The novelty of our work primarily lies in the proposal of further enhancement of error resilience via fundamental modification of the conventional prediction structure. Rather than predict based on the encoder-reconstructed signal, we propose to predict based on the expected decoder reconstruction. Such an approach hinges on accurate estimation of decoder quantities and we hence build on the ROPE approach for recursive decoder reconstruction estimation. In spite of the loss in source coding gain due to the lower source prediction quality of the proposed source-channel predictor, our scheme always achieves better overall R-D tradeoff than the conventional scheme. Moreover, in conjunction with the proposed source-channel predictor, it is necessary to modify the motion estimation criterion. Rather than identify the motion vector that minimizes the (directly computable) prediction error of the proposed predictor, it is advantageous to select the motion vector that minimizes the expected prediction error to be experienced at the decoder. This subtle point is shown to be of considerable significance to the performance of the system.

REFERENCES

[1] Y. Wang, Q. Zhu, "Error Control and concealment for video communication: A review," *Proc. of IEEE,* vol.86, no.5, May 1998, pp. 974-997.

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

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

[4] T. Wiegand, N. Farber, K. Stuhlmuller, B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Select. Areas Commun.*, vol.18, no.6, June 2000, pp. 1050-1062.

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

[6] 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.

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