To appear in *IEEE Inter. Conf. Multimedia and Expo*, Baltimore, MD, July, 2003

SPATIAL-TEMPORAL ERROR CONCEALMENT WITH SIDE INFORMATION FOR STANDARD VIDEO CODECS

Wenjun Zeng[1]
PacketVideo Corporation
E-mail: wzeng@packetvideo.com

ABSTRACT

Error concealment (EC) is important for transmitting video over error prone networks such as the Internet or wireless networks. Different EC strategies have their own advantages for different scenarios. However, it is generally difficult for the decoder to figure out which strategy works the best for a specific case. This paper proposes to use data embedding to convey the necessary *high level* side information in a standard compliant way to help improve the decoder's EC performance. We show that the EC "mode" information (i.e.., whether spatial EC or temporal EC should be used) is critical for a spatial-temporal EC approach.

1 INTRODUCTION

Compressed digital media is very vulnerable to imperfect communication channels such as the Internet or wireless networks. The problem of recovering lost data in compressed video is thus of importance. Among others, error concealment (EC) techniques take advantage of the spatial and/or temporal correlation of video without incurring much overhead and delay.

Different EC strategies have their pros and cons for different scenarios [1]. An adaptive approach that aims to exploit the advantages of different EC strategies is thus desirable. However, it is generally difficult for the *decoder* to figure out which strategy works the best for a specific case. On the other hand, the relative performance of different EC strategies for a specific case can be easily evaluated at the *encoder* side, thanks to the availability of all original video data. In this paper, we propose to use data embedding/hiding to convey the necessary *high level* side information, as opposed to simply some redundant information, in a standard compliant way to help improve the decoder's error concealment performance. We will show that the EC "mode" information, i.e., which strategy should be used, is critical for a spatial-temporal EC approach, and thus can be embedded into the compressed bitstream to facilitate the decoder's EC operation.

This paper is organized as follows. Section 2 briefly reviews some of the general signal-processing based EC techniques. The general concept of exploiting data hiding to facilitate error concealment is discussed in Section 3. Section 4 proposes to use data hiding to convey the high level EC "mode" information for a spatial-temporal EC approach. Simulation results are presented in Section 5. Section 6 draws the conclusion.

2 SPATIAL TEMPORAL ERROR CONCEALMENT

In this section, we briefly review some general approaches for post-processing based error concealment. The post-processing based EC basically exploits the correlation between the damaged block and its spatially and temporally adjacent blocks. It relies on the fact that generally there still exists some redundancy in the compressed bitstream.

<u>Motion-compensated temporal prediction</u>

A very simple scheme for temporal error concealment is to just copy the macroblock (MB) at the same spatial location in the previous frame to conceal the lost MB. If motion vector (MV) is not damaged, then the motion-compensated prediction from the reference frame can be used to conceal the damaged block. This can often produce good results. The problem, however, is that MV and coding-mode (intra/inter) may not be available when a packet is lost. One can try to estimate the lost MV, but the accuracy of the estimation may affect the results significantly. Even with the correct MV, the decoder may not know how good the prediction is. A good measure to evaluate the estimation is the spatial smoothness around the block boundary. In [2], it was proposed to choose one from a set of candidate MVs available from neighboring MBs that minimizes the boundary variation between the concealed MB and its neighboring correctly received MBs. The complexity of this approach is about 3-4 operations per missing pixel. It, however, may not be applicable in some cases (e.g. scene change frames, or Intra MBs with large irregular motion).

<u>Spatial/frequency interpolation</u>

The second technique is to interpolate the lost region from its spatially neighboring available pixels or coefficients. This technique could be very simple, but a general problem is that it may blur the edges, and may not be able to recover very complex structures. With some extra complexity, an edge-based directional interpolation can usually provide much better results than bilinear interpolation [3][4]. For example, a simple scheme was proposed in [3] where the edge direction is estimated by evaluating the pixel-wise difference between two sets of projection data on the block boundary. Directional interpolation

[1] The author would like to thank Gene Wen for some valuable discussion.

along the estimated edge direction is then applied to recover the lost block. The computation overhead is less than 2 operations per missing pixel.

Combined spatial-temporal error concealment

There have been some attempts to combine both spatial and temporal error concealments to improve the performance. For example, the smoothness constraint can be applied to both temporally and spatially neighboring pixels to construct a maximally smooth missing block [5]. A typical problem with such an approach is the loss of high frequency information.

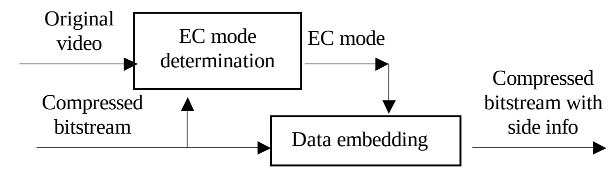
3 DATA HIDING FOR ERROR CONCEALMENT

Data hiding provides an efficient way to convey side information that can be used to achieve better performance and additional functionalities for many different applications, e.g., in multimedia communication.

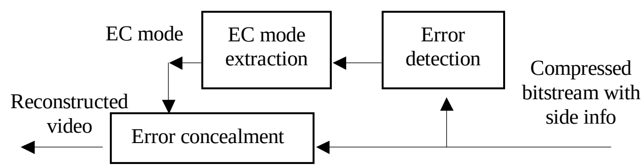
To our best knowledge, we are one of the firsts to suggest using data hiding for error concealment/resiliency purpose *in a standard compliant way*. In [6], we proposed to use data hiding to convey a rate-distortion based significance score associated with an image block that is used to signify the effectiveness that the associated block can be interpolated based on its surrounding information. The embedded significance score can later be extracted to assist real-time dynamic rate shaping by intelligently dropping some selective blocks for transmitting compressed video over time-varying channels. It has been shown in [6] that the bit-overhead introduced by data hiding can be made very low. In particular, it is possible to achieve fractional overhead-bit per embedded information bit. For example, it has been shown [6] that by embedding one bit to each 8x8 block using an *odd-even scheme*, it incurs no bit overhead and no perceptual degradation of the video quality.

In a recent work [7], Yin et al proposed to use data hiding to facilitate spatial EC by embedding edge directional information of a block into a companion block prior to the transmission. This edge information can then be extracted at the receiver side for directional filtering to recover the lost blocks, thus reducing the computational load (for inferring the edge directional information) at the decoder. Similar idea has also been used to embed some important redundant information such as the motion vector information in the compressed bitstream to facilitate motion-compensated error concealment at the decoder [8].

In general, data hiding may potentially introduce bit overhead or video quality degradation. Therefore, it is important to identify and extract highly summarized information that is critical to the performance improvement at the decoder. This will not only improve the EC performance, but also avoid the potential negative impact on the bit rate or video quality. Specifically, in this paper, we propose to use data hiding to facilitate the employment of spatial or temporal error concealment at the decoder. As discussed in Section 1, it is generally very difficult for the decoder to figure out which one performs better than the other one for a particular MB, due to lack of information about the original MB. But the encoder is capable of knowing exactly which one performs better. This EC "mode" information can thus be determined at the encoder and then embedded into the compressed bitstream as side information so that the decoder can take advantage of it when necessary.



(a) encoder side



(b) decoder side

Fig. 1: The general diagram of the proposed spatial-temporal error concealment scheme.

4 SPATIAL-TEMPORAL ERROR CONCEALMENT WITH SIDE INFORMATION

In this section, we describe a spatial-temporal error concealment scheme that utilizes some important side information embedded in the standard compressed bitstream to improve the EC performance. The basic idea is to determine the EC mode information at the encoder side, which is then embedded into the compressed bitstream and can be later extracted as necessary at the decoder side to facilitate the EC operation. Since typically more information is available at the encoder side, this side information is more accurate than what can be determined at the decoder side. It also reduces the computation load of the decoder.

Fig. 1 shows the general diagram. Specifically, at the encoder side, we will compute the error concealed MBs using both spatial and temporal error concealment separately. Using the original video frame available at the encoder as the benchmark, we then identify which one reconstructs a better MB in terms of mean square error (MSE), with respect to the original MB. This one bit information, referred to as "EC mode" in the paper, is then embedded into the next row of MBs (GOB) using the even/odd approach as described below. At the decoder side, when a MB or a GOB is lost, this EC mode information can be extracted from the neighboring correctly received GOBs, and the appropriate EC operation (spatial or temporal) will be applied.

For each lost MB, we will assume all other MBs in the same GOB are also lost. Therefore only MBs in the previous GOB or the next GOB will be used in error concealment of the current

MB. The specific spatial error concealment and temporal error concealment schemes to be used are described as follows.

Spatial error concealment

We use spatial directional interpolation along the estimated edge direction for spatial error concealment. The edge direction is estimated as one of three directions (135, 90, 45 degree) that minimizes the pixel-wise difference between two sets of projection data locating on, respectively, the top and bottom GOB boundaries associated with each missing MB [3]. For example, for 45^o direction, one set of projection data consists of the 16 boundary pixels in the MB above, and another 16 in the neighboring MB to the northeast. The other set of projection data consists of the 16 boundary pixels in the neighboring MB to the southwest, and another 16 in the MB below.

Temporal error concealment

We first estimate the MV by choosing one that minimizes the block boundary variation [2] based on a set of candidates, including MVs of the top three, and the bottom three neighboring MBs, the MV of the MB located at the same spatial location in the previous frame, average of the above MVs, and zero MV. Motion-compensated error concealment using the estimated MV is then applied.

Data embedding method

The even-odd embedding scheme is referred to as hiding one bit to one 8x8 block by forcing the sum of the quantized AC levels to be odd or even. To minimize the potential impact on the perceptual quality, one can exploit the visual masking effect by choosing the largest AC coefficient to change the level value (by one unit, if necessary) [6]. We will embed the EC mode bits of the even GOBs to the coded 8x8 blocks of the odd GOBs, and vice versa. For example, for a QCIF size frame, only 99 EC mode bits need to be embedded for each frame. In the rare case that there are less than 99 coded 8x8 blocks in a frame, we will embed the side information sequentially until all coded 8x8 blocks have been used. The rest of the EC mode bits will not be embedded, and temporal EC will be used for their associated lost MBs.

5 SIMULATION RESULTS

Simulation has been done to test the effectiveness of the proposed spatial-temporal error concealment scheme. The experimental setting is as follow. We consider a worse case scenario and assume that every other GOB is lost, i.e., a packet loss rate of 50%. This is consistent with the RTP packetization scheme recommended in [10][9] for H.263 compressed video, where all even GOBs of a frame is packed into one RTP packet, and all odd GOBs into another. When reconstructing a frame, the reference frame is assumed to be correctly received, i.e., temporal error propagation effect is not considered. This is to isolate the effect of temporal error propagation, and to signify the gain the proposed scheme can achieve over other prior solutions on a frame-by-frame basis. Our experiment shows that the bit overhead and perceptual quality degradation introduced by the data hiding process is negligible. For example, in one case, with data hiding, the average PSNR value is reduced by 0.08 dB while the bit rate is also reduced by 0.002 bpp.

We compare the following six error concealment schemes:

- Use_ZeroMV: where the MB in the same spatial location in the previous frame is copied
- Use_EstimatedMV: where temporal error concealment based on estimated MV is used alone
- SP/TP_w/o_SideInfo: estimate the coding mode (inter/intra) using the method described in [1], and then use temporal error concealment for inter MBs and spatial error concealment for intra MBs
- SP/TP_w_SideInfo: use temporal or spatial error concealment based on the "EC mode" bits extracted from the bitstream at the decoder
- Use_OriMV_SP4intra: assume the original MVs are available, and are used for motion-compensated EC. Spatial error concealment is used for intra MB
- Use_OriMV_0MV4intra: same as above, except the MB in the same spatial location in the previous frame is copied to conceal a lost intra MB

Fig. 2 shows the PSNR results of the reconstructed video frames for the movie sequence "hangingup" (size 320x224). The average PSNR values for the cases of SP/TP_w_SideInfo, SP/TP_w/o_SideInfo, Use_EstimatedMV, and Use_ZeroMV are 30.2 dB, 28.2 dB, 26.9 dB and 25.8 dB, respectively. It is seen that the side information helps to significantly improve the EC performance, when compared to the SP/TP_w/o_SideInfo scheme. The PSNR gain is up to 4-5 dB for some frames. The Use_ZeroMV approach results in very poor performance for this complex sequence. For many cases such as the scene change or high motion frames, spatial error concealment is very important.

Fig. 3 shows that with the EC mode information, even with 50% packet loss, the decoder can reconstruct a decent frame. The EC mode bit conveys some information that is very difficult, if not impossible, to infer at the decoder side.

Fig. 4 shows that, if we assume that the original MVs are available for error concealment (for example, they can also be embedded in the bitstream redundantly [8]), it provides much better results than using the estimated MVs for this particular test sequence that has high and irregular motion. However, with only one bit EC mode information embedded, we can achieve close performance as using the original MVs. Note that the average PSNR values for the cases of SP/TP_w_SideInfo, SP/TP_w/o_SideInfo, Use_OriMV_SP4intra, and Use_OriMV_0MV4intra are 30.2 dB, 28.2 dB, 30.4 dB and 27.7 dB, respectively.

6 CONCLUSION

In this paper, we discuss the application of data hiding to achieve improved performance for error concealment. We point out the importance of identifying some high level information as side information, as opposed to redundant information, in such an approach. We show that the EC mode, i.e., using spatial EC or temporal EC, is very important side information. Experiments show that with such side information embedded in the standard compressed bitstream, significant improvement can be achieved in recovering the lost packets.

7 REFERENCES

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

[2] W. Lam, A. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," *Proc. ICASSP*, pp. 417-420, 1993

[3] K. Jung, J. Chang, and C. Lee, "Error concealment technique using projection data for block-based image coding," *Proc. SPIE Conf. Visual Communications and Image Proc*, vol. 2308, pp. 1466-1476, 1994.

[4] W. Zeng and B. Liu, "Geometric-structure-based error concealment with novel applications in block-based low bit rate coding," *IEEE Trans. on Cir. and Sys. for Video Technology*, pp. 648-665, June 1999.

[5] Y. Wang and Q. Zhu, "Signal loss recovery in DCT-based image and video codecs," *Proc. SPIE Conf. Visual Communications and Image Proc*, vol. 1605, pp. 667-678, Nov. 1991.

[6] W. Zeng, *Resilient video transmission and multimedia database applications*, Ph.D Dissertation, Electrical Engineering Dept., Princeton University, June 1997.

[7] P. Yin, H. Yu, and B. Liu, "Error concealment using information hiding," *Proc. IEEE ICASSP*, 2001.

[8] P. Yin, M. Wu and B. Liu, "A robust error resilient approach for MPEG video transmission over internet," *Proc. VCIP02*.

[9] C. Bormann et al, "RTP payload format for the 1998 version of ITU-T Rec. H.263 video (H.263+)," IETF RFC2429, Oct. 1998.

[10] S. Wenger and G. Cote, "Using RFC2429 and H.263+ at low to medium bit-rates for low-latency applications," Packet Video Workshop, 1999.

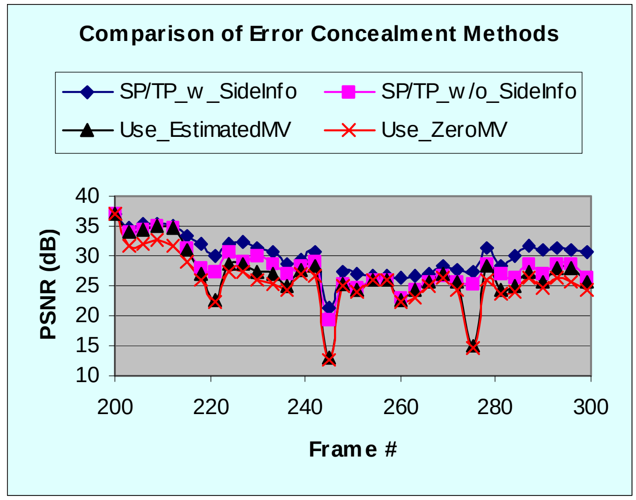


Fig. 2: PSNR comparison of different EC schemes for the "hangingup" sequence.

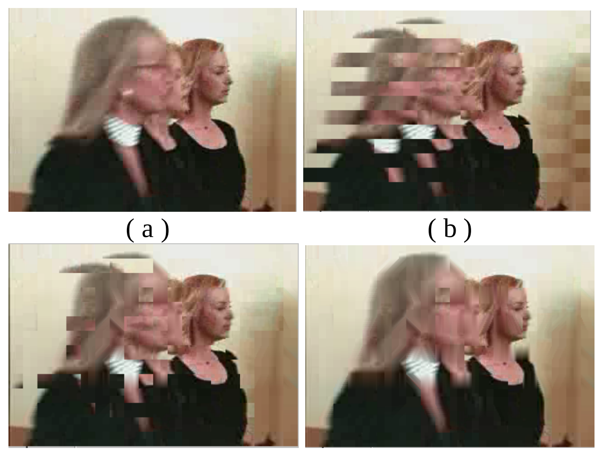


Fig. 3: Reconstructed frames using different EC schemes for the "hangingup" sequence, assuming every other GOB lost. (a): error free; (b): Use_estimatedMV; (c): SP/TP_w/o_SideInfo; (d):SP/TP_w_SideInfo

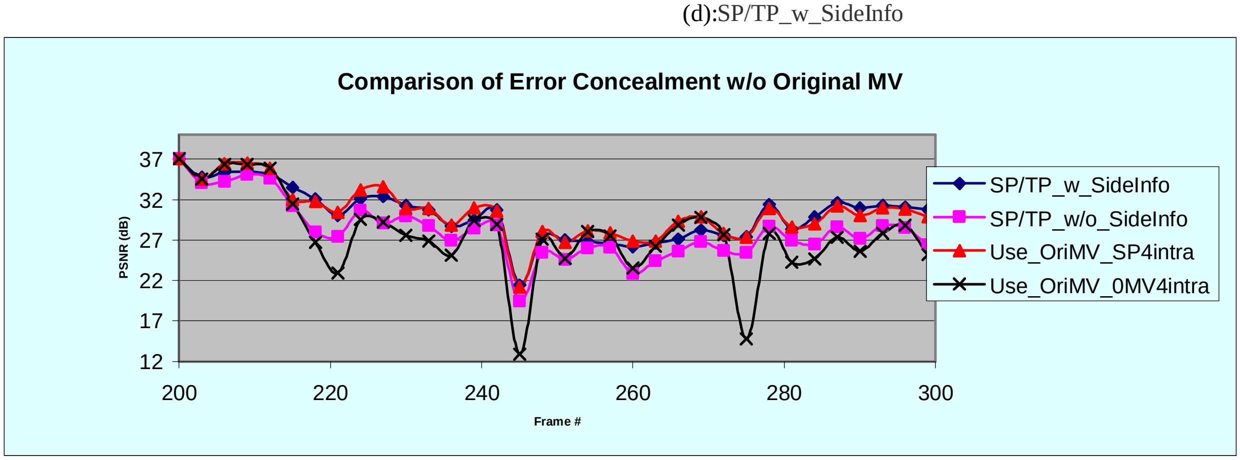


Fig. 4: Comparison of different EC schemes with or without original MV information.