

HYBRID SPATIAL AND TEMPORAL ERROR CONCEALMENT FOR DISTRIBUTED VIDEO CODING

Shuiming Ye, Mourad Ouaret, Frederic Dufaux, Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

Distributed Video Coding (DVC) is based on a new paradigm in coding, which has received many interests recently. This paper proposes a hybrid spatial and temporal error concealment scheme to conceal errors in Wyner-Ziv (WZ) frames. We first use a spatial concealment based on edge directed filter. This step is exploited to improve the performance of subsequent temporal concealment. An enhanced temporal concealment based on motion compensated temporal interpolation is proposed, including motion vector refinement and smoothing, optimal compensation mode selection, and a new matching criterion for motion estimation. Simulation results show that the objective qualities as well as the perceptual qualities of the corrupted sequences are significantly improved by the hybrid error concealment, outperforming both spatial and temporal concealments alone.

1. INTRODUCTION

Recently, the rapid growth of Internet and wireless communications has led to increased interest for robust transmission of compressed video. However, transmission errors may severely impact video quality as compressed data is very sensitive to these errors [1]. Thus, error control techniques are necessary for efficient video transmission over error prone channels.

Error Concealment (EC) consists in estimating or interpolating corrupted data at the decoder from the correctly received information. It can improve the quality of decoded video corrupted by transmission errors, without any additional payload on the encoder or channel. EC can be classified into three categories: temporal concealment, spatial concealment, and hybrid spatial and temporal concealment [2]. Temporal EC usually leads to better results when compared to spatial concealment, given the typically high temporal correlation in video. However, for video with scene changes or very high or irregular motion, spatial EC is preferred. Some attempts have been made to combine both spatial and temporal EC to improve performance [2, 3, 4]. These schemes use some mode selection methods to decide whether to use spatial or temporal EC. For example, temporal activity (measured as the prediction error in the surrounding blocks) and spatial activity (measured as the variance of the same surrounding blocks) are used to decide which concealment mode to use [3]. In general, however, these methods have achieved very limited success, mostly due to the simple mode selection mechanisms at the decoder to merge the results from both spatial and temporal EC.

Distributed Video Coding (DVC) is based on a new paradigm in coding. As it allows for very low complexity encoding, it is well-suited for applications such as wireless low-power video

surveillance, multimedia sensor networks, wireless PC cameras, and mobile camera phones. Another appealing property of DVC is its good resilience to transmission errors. A thorough analysis of its performance in the presence of transmission errors has been presented in [5], showing its good error resilience properties.

In DVC, the WZ bits are used to improve the quality of an initial estimate of the decoded WZ frame, commonly referred to as Side Information (SI) and generated from some conventionally coded key frames. Therefore, transmission errors in WZ bits result in different distortions from that in conventional video codecs. Per consequent, error concealment schemes proposed for conventional video coding cannot be directly applied to conceal errors in DVC. On the other hand, in the literature, few error concealment schemes can be found for WZ frames.

In this paper, we propose a new hybrid spatial and temporal concealment scheme for WZ frames in DVC. It uses the error-concealed results from spatial EC to improve the performance of the temporal EC, instead of simply switching between spatial and temporal EC. Spatial EC based on the edge directed filter [6] is firstly applied to the corrupted blocks, and the results are exploited to improve the temporal EC. An enhanced temporal EC is then applied, which includes motion vector refinement and smoothing, optimal compensation mode selection based on the spatial error-concealed results, and a new matching criterion for motion estimation. In other words, the temporal EC is not only based on the key frames, but also on the WZ bits already decoded. This constitutes one of the main novelties of the proposed approach. Experimental results show that the performance of the proposed hybrid scheme is superior to spatial EC and temporal EC alone.

2. DVC ARCHITECTURE

Without loss of generality, in this paper, we consider more specifically the DVC architecture from [7], but without the feedback channel. As shown in Figure 1, a video sequence is divided into key frames and WZ frames. Hereafter, we consider a Group of Pictures (GOP) size of 2, namely the odd and even frames are key frames and WZ frames, respectively. Key frames are conventionally encoded using H.264/AVC intra coding. Conversely, for WZ frames, a DCT transform is firstly applied, and the resulting transform coefficients undergo quantization. The quantized values are then split into bitplanes which are turbo encoded. At the decoder, SI approximating the WZ frames is generated by Motion Compensated Temporal Interpolation (MCTI) of the decoded key frames. The SI is used in the turbo decoder, along with WZ parity bits, in order to reconstruct the decoded video sequence.

The feedback channel is not a desirable feature for real-time video application scenarios, since it increases transmission delay.

Note that encoder rate control has been previously proposed [8]. In this paper, we suppose that the encoder performs ideal rate control, i.e. the number of requests for each bitplane for the error free case is determined a priori and used for decoding the corresponding corrupted bitstream.

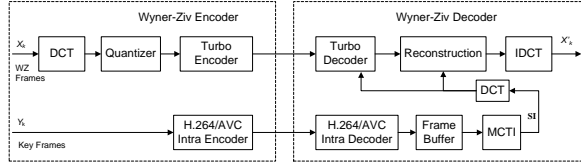


Figure 1: DVC architecture.

3. PROPOSED ERROR CONCEALMENT FOR DVC

The proposed hybrid error concealment scheme is illustrated in Figure 2. The locations of corrupted blocks are firstly detected. In this paper, we assume that the error locations are detected at the decoder, as often presumed in error concealment literature, which can be done at transport level, or based on syntax and watermarking [1]. Spatial EC is then applied to obtain a partially error-concealed frame. This frame is much closer to the error free frame than the corrupted frame. The partially error-concealed frame is used for motion vector refinement, smoothing, and optimal compensation mode selection, to obtain an estimate of the motion vector of the corrupted block. Motion compensation is finally used to obtain the reconstructed block.

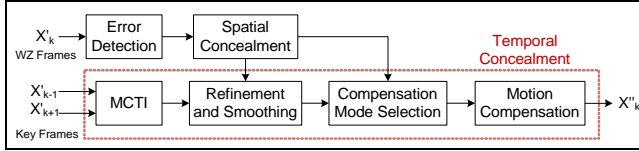


Figure 2: Proposed spatio-temporal error concealment.

3.1. Spatial concealment based on edge directed filter

In DVC, the decoded WZ frames are based on the SI generated by MCTI of the key frames. SI is then used by the turbo decoder, along with the WZ bits, to obtain the decoded WZ frame. The transmission errors in WZ bits tend to cause noises around edges in the corrupted WZ frames (as shown in Figure 7a). It is different from traditional video coding schemes such as H.264/AVC, where errors result in loss of whole blocks.

In this paper, because of the error patterns generated by WZ corrupted frames, an edge directed filter is constructed to remove the noises around edges caused by errors in the WZ bits. Anisotropic diffusion techniques have been widely used in image processing for their efficiency at smoothing noisy images while preserving sharp edges. We adopt the anisotropic diffusion as a direction diffusion operation, and use the diffusion method for spatial error concealment as in [6].

3.2. Enhanced temporal error concealment

An enhanced temporal EC is proposed to exploit spatio-temporal correlations. The approach is based on MCTI and motion vector filtering as proposed in [9]. One of the key novelties is that the partially error-concealed frame is used to improve the temporal EC, unlike [9] where MCTI is based on the previous and next key frames. Indeed, the reconstructed frame by spatial concealment contains additional information about the current frame carried by the correctly received WZ bits. Therefore, by using the partially

error-concealed frame resulting from spatial EC, the spatio-temporal correlations between this frame and the reference key frames can be better exploited. Hence the performance of the temporal EC is improved.

3.2.1. Matching criterion

To exploit the spatio-temporal correlations based on the partially error-concealed frame, a new matching criterion is used to evaluate the error in motion estimation, based on mean absolute difference (*MAD*) and boundary absolute difference (*BAD*) [10]. *MAD* often leads to false motion vectors in the presence of changes in pixel intensity. Conversely, *BAD* is not efficient when local variation is large. Therefore, both are combined to construct the new matching distortion (D_{ST}) between two frames F_1 and F_2 , which is defined as

$$D_{ST}(F_1, F_2, MV) = \alpha BAD(F_1, F_2, MV) + (1 - \alpha) MAD(F_1, F_2, MV) \quad (1)$$

where α is a weighting factor, set to 0.3 empirically in this paper, and MV the estimated motion vector.

3.2.2. Motion vector refinement and smoothing

The spatio-temporal correlations between the partially error-concealed frames and the key frames are exploited to refine and smooth the estimated motion vectors. The motion vectors are re-estimated using the new matching criterion and partially error-concealed frame by bi-directional motion estimation, and are then filtered using motion vector spatial smoothing. This process generates a new estimation of the motion vector for the block to be concealed using information which is not only from reference key frames, as in a classical MCTI [9], but also from the partially error-concealed frame. Therefore, as more information is exploited, it produces better estimates.

It is observed that the motion vectors have sometimes low spatial coherence. A spatial motion smoothing filter is therefore used as in [9], but with the new matching criterion and the partially error-concealed frame. A weighted median vector filter is used to maintain the motion field spatial coherence. More precisely, it looks, at each block, for candidate motion vectors at neighboring blocks. This filter is adjusted by a set of weights controlling the filter smoothing strength. The new matching criterion is used to calculate the weights of the blocks for each candidate motion vector (calculated between key frames and the partially error-concealed frame). The weighted median vector filter is defined as

$$MV_F = \arg \min_{MV_i} \sum_{j=1}^{Num} w_j \|MV_i - MV_j\|_L, i \in [1, Num] \quad (2)$$

where MV_1, \dots, MV_{Num} are the motion vectors of the corresponding nearest neighboring blocks. MV_F is the motion vector output of the weighted vector median filter, which is chosen in order to minimize the sum of distances to the other $Num-1$ vectors. w_1, \dots, w_{Num} correspond to a set of weights.

The weights are calculated based on the new matching criterion and the partially error-concealed frame

$$w_j = \frac{D_{ST}(F_{n-1}, F_n, MV_c) + D_{ST}(F_{n+1}, F_n, MV_c)}{D_{ST}(F_{n-1}, F_n, MV_j) + D_{ST}(F_{n+1}, F_n, MV_j)} \quad (3)$$

where D_{ST} represents the matching measure as defined in Eq. (1). MV_c is the re-estimated vector for the block to be concealed.

3.2.3. Optimal motion compensation mode selection

The objective of the proposed optimal motion compensation mode selection is to generate an optimal motion compensated estimate that is more accurate than MCTI. In the DVC codec in [9], while

bi-directional prediction is shown to be effective, it is limited to motion-compensated average of the previous and the next key frames.

Based on the partially error-concealed frame, the most similar block to the corrupted block could be selected from a number of sources: the previous frame, the next frame, the bi-directional motion-compensated average of the previous and the next frame. In other words, the block in the reconstructed frame is reconstructed using only one block from the previous key frame in backward mode, from the next key frame in forward mode, and from both frames in bi-directional mode. This mode selection further improves the performance of temporal EC, especially when the motion between the frames is significant but not symmetric. It differs from the classical MCTI using bi-directional motion compensated average of the previous and the next key frames in the sense that the partially error-concealed frame is used to decide how to perform motion compensated prediction.

Based on the estimated motion vectors and the interpolation modes, motion compensation is applied to generate the reconstructed blocks as the result of temporal concealment.

4. RESULTS AND DISCUSSIONS

In simulations, only luminance data is coded, and a communication channel, characterized by the error pattern files provided in [11] with different packet loss ratios (PLR), is used. Test sequences in QCIF format at 15 fps are corrupted with packet loss rates of 3%, 5%, 10%, and 20%. For the various test conditions in terms of PLR and quantization parameters (QP), the average PSNR is measured. Results are obtained by averaging over ten runs using different error patterns. In this paper, if the bit error probability of the decoded bitplane is higher than 10^{-3} , the decoder uses the corresponding bitplane of the side information, which improves the quality of the decoded frame. The header of the WZ bitstream, which contains critical information such as frame size, quantization parameters, and intra period, is assumed to be correctly received.

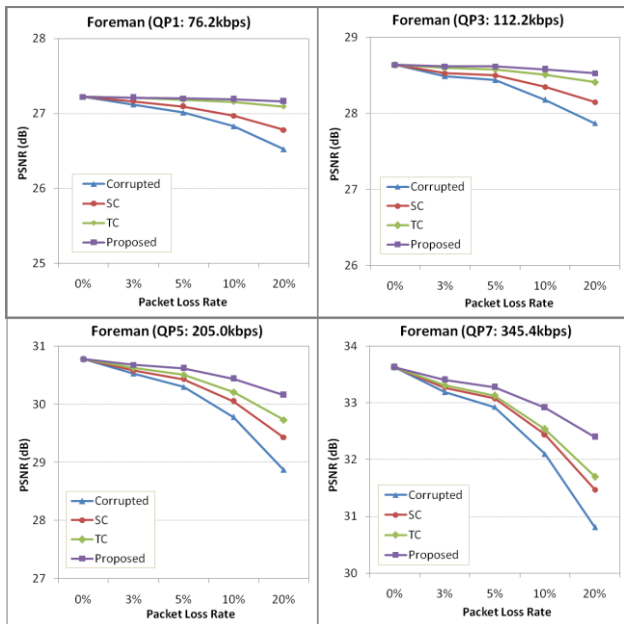


Figure 3: PSNR performance for Foreman (only WZ frames are corrupted).

To evaluate the performance of the proposed method, first only errors in WZ frames are simulated. Figure 3 shows the PSNR values of error-concealed results at several quantization indexes for Foreman. The results of the proposed concealment method are compared to those of spatial concealment (SC, based on edge directed filter [6]), and temporal concealment (TC, based on MCTI [9]). As shown in Figure 3, the proposed error concealment method achieves better objective qualities than both SC and TC, for all packet loss ratios and quantization indexes. The improvements of the proposed method can be considered as quite good when compared to the SC and TC only. From the results, we can also observe that, although TC generally performs better than SC, the performance of SC can be very close to TC when the video quality is good (for high QP or low PLR).

Figure 4 shows results for Soccer (Figure 4a) and Hallmonitor (Figure 4b). For sequences with large and complex motion such as Soccer, the temporal EC performance can become worse than that of spatial EC (Figure 4a). For sequences with simple motion such as Hallmonitor, the performance of the proposed EC becomes close to TC (Figure 4b).

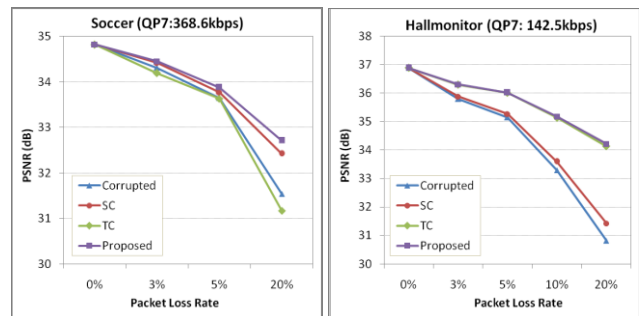


Figure 4: PSNR performance for Soccer and Hallmonitor (only WZ frames are corrupted).

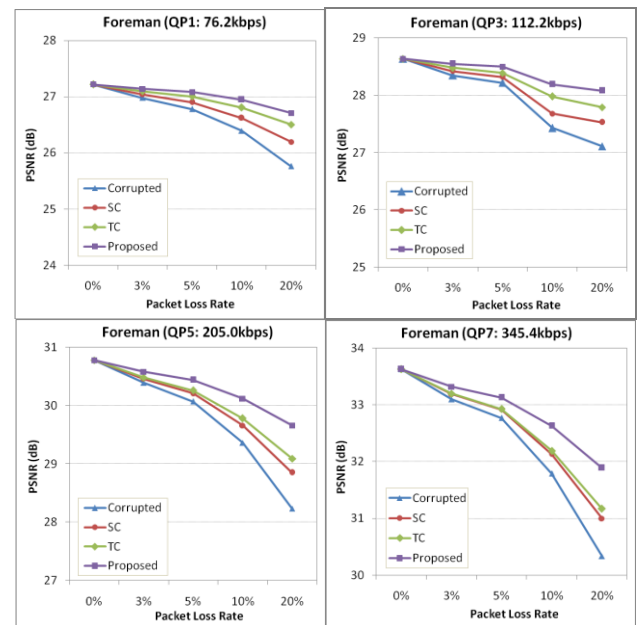


Figure 5: PSNR performance for Foreman (both key and WZ frames are corrupted).

To evaluate the performance of the proposed method in a more realistic scenario, errors in both key and WZ frames are also simulated. The spatial EC defined in H.264/AVC is used to conceal errors in key frames. Figure 5 shows the PSNR values of error-concealed results at several quantization indexes for Foreman. As shown in Figure 5, errors in key frames further decrease PSNR values for all EC methods, but the proposed error concealment method leads to even larger improvements when compared to SC and TC only. The reason may be that the errors in key frames will also cause distortions in WZ frame, and these distortions can be better concealed by the proposed method.

Figure 6 shows the PSNR values of all frames of Foreman (QP7, PLR 20%). The improvement by the proposed method is quite important, with gains up to 8.0 dB. The average improvement for the whole sequence is 2.0 dB, significantly outperforming SC or TC. Table 1 shows processing time comparisons for results in Figure 6 (Windows XP, 2.6 GHz CPU, 2.0 G RAM). Processing time consumed by the proposed scheme is longer than both SC and TC, since both SC and TC are applied in the proposed scheme.

Figure 7 shows the visual results of frame 56 (QP7, PLR 20%). Distortions are still visible near the edges in the frames concealed by SC (Figure 7b) and TC (Figure 7c). The visual quality obtained with the proposed method (Figure 7d) is clearly much better.

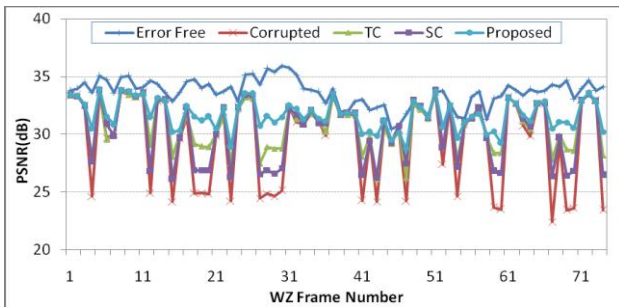


Figure 6: PSNR of error-concealed frames for Foreman.

Table 1: Comparison of processing time (QP7, PLR 20%)

Running Time	SC	TC	Proposed
Second/frame	3.47	0.47	3.93

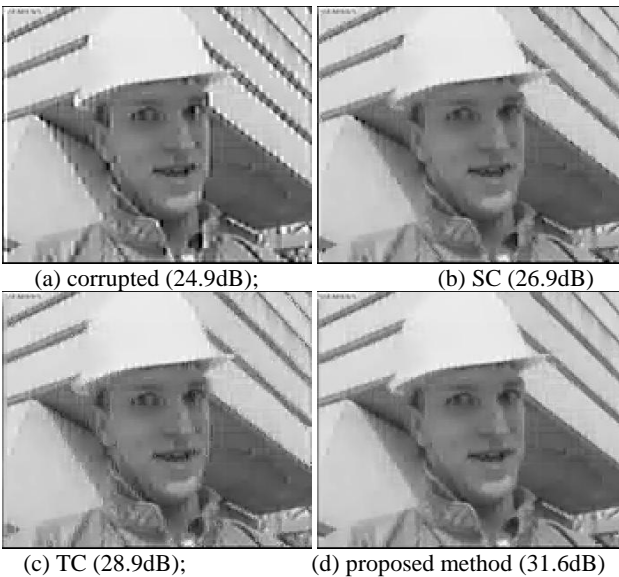


Figure 7: Error-concealed results for Foreman

5. CONCLUSIONS

A new error concealment scheme for WZ frames in DVC is proposed. An edge directed filter is used for spatial concealment to remove annoying artifacts concentrated on edges in the corrupted WZ frame. The use of the partially error-concealed frame from spatial concealment improves the performance of the temporal EC by motion vector refinement, smoothing, and optimal compensation mode selection. Simulation results show the proposed method can bring significant improvements in terms of both objective as well as visual quality.

ACKNOWLEDGEMENT

This work was partially supported by the European Network of Excellence **VISNET2** IST Contract 1-038398, funded under the European Commission IST 6th Framework Program.

REFERENCES

- [1] Y. Wang and Q. Zhu, "Error Control and Concealment for Video Communication: A Review", *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998, pp.974-997.
- [2] O. Hadar, M. Huber, R. Huber, and S. Greenberg, "New Hybrid Error Concealment for Digital Compressed Video", *EURASIP Journal on Applied Signal Processing*, Vol. 2005, No. 12, 2005, pp. 1821-1833.
- [3] D. Agrafiotis, D. R. Bull, T.K. Chiew, P. Ferre, and A. Nix, "Enhanced Error Concealment for Video Transmission over WLANs", *International Workshop on Image Analysis for Multimedia Interactive Services*, Switzerland, Apr. 2005.
- [4] C. Yim, W. Kim, and H. Lim, "Hybrid Error Concealment Method for H.264 Video Transmission over Wireless Networks", *International Conference on Wireless Communications and Mobile Computing*, Hawaii, USA, Aug. 2007.
- [5] J. Pedro, L. Soares, C. Brites, J. Ascenso, F. Pereira, C. Bandeirinha, S. Ye, F. Dufaux, and T. Ebrahimi, "Studying Error Resilience Performance for a Feedback Channel Based Transform Domain Wyner-Ziv Video Codec", *Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.
- [6] S. Ye, Q. Sun, and E.C. Chang, "Edge Directed Filter based Error Concealment for Wavelet-based Images", *IEEE International Conference on Image Processing*, Singapore, 2004.
- [7] C. Brites, J. Ascenso, and F. Pereira, "Improving Transform Domain Wyner-Ziv Video Coding Performance", *International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.
- [8] C. Brites and F. Pereira, "Encoder Rate Control for Transform Domain Wyner-Ziv Video Coding", *International Conference on Image Processing*, San Antonio, USA, Sept. 2007.
- [9] J. Ascenso, C. Brites, and F. Pereira, "Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding", *EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Slovak, July 2005.
- [10] Y. Chen, O. C. Au, C.-W. Ho, and J. Zhou, "Spatio-Temporal Boundary Matching Algorithm for Temporal Error Concealment", *IEEE International Symposium on Circuits and Systems*, Greece, May, 2006.
- [11] S. Wenger, "Proposed Error Patterns for Internet Experiments", Doc. VCEG Q15-I-16R1, VCEG Meeting, Red Bank, NJ, USA, Oct. 1999.