THE ERROR CONCEALMENT FEATURE IN THE H.26L TEST MODEL

Ye-Kui Wang[1], *Miska M. Hannuksela*[2], *Viktor Varsa*[3],
Ari Hourunranta[2], *and Moncef Gabbouj*[4]

[1]Tampere International Center for Signal Processing (TICSP), Tampere, Finland
[2]Nokia Mobile Software, Tampere, Finland
[3]Nokia Research Center, Irving, TX, USA
[4]Tampere University of Technology, Finland

ABSTRACT

This paper presents the error concealment (EC) feature implemented by the authors in the test model of the draft ITU-T video coding standard H.26L. The selected EC algorithms are based on weighted pixel value averaging for INTRA pictures and boundary-matching-based motion vector recovery for INTER pictures. The specific concealment strategy and some special methods, including handling of B-pictures, multiple reference frames and entire frame losses, are described. Both subjective and objective results are given based on simulations under Internet conditions. The feature was adopted and is now included in the latest H.26L reference software TML-9.0.

1. INTRODUCTION

The H.26L project [1] aims at developing a new ITU-T video coding standard. The project started in January 1998 and is to be finalized by the end of 2002. Compared to ITU-T Recommendation H.263, H.26L aims at better compression efficiency, more flexible network adaptation, and enhanced error robustness among other things. The test model reference software, which is called "Test Model Long Term (TML)", was established to include the adopted algorithms and to verify new proposals.

In order to have a realistic view of H.26L's operation in error-prone environment, basic error resilience tools had to be implemented in the TML reference software. One of these basic tools was decoder-side error concealment (EC). As EC falls outside the scope of the standard itself, we proposed non-normative EC algorithms for the TML software. These were adopted and included in the latest reference software TML-9.0 [2]. Any new error-robust coding scheme proposal for H.26L should be compared against the H.26L test model equipped with this feature.

In section 2 we present the general concealment strategy. The selected EC algorithms for the different types of encoded pictures are described in section 3. Section 4 discusses the EC implementation issues for multiple reference frames, entire frame losses, and B-pictures. Simulation results and conclusion are addressed in sections 5 and 6, respectively.

2. THE ERROR CONCEALMENT STRATEGY

It is assumed that erroneous or incomplete slices are not decoded but discarded before decoding, therefore no integrity checking or bit-error detection is performed. All correctly received slices of a picture are decoded first, and then the lost slices are concealed according to the presented algorithms. In practice, a record is kept in a macroblock (MB) based status map of the frame. The status of an MB in the status map is "Correctly received" whenever the slice where the MB resides is available for decoding, and "Lost" otherwise. After the frame is decoded, if the status map contains "Lost" MBs, concealment is initiated.

Given the slice structure and the MB-based status map of a frame, the proposed concealment algorithms are MB-based. The missing frame area (pixels) covered by MBs marked as "Lost" in the status map are concealed MB-by-MB (16x16 Y pixels, 8x8 U, V pixels). After an MB has been concealed, it is marked in the status map as "Concealed". Not only the "Correctly received" but also the "Concealed" MBs are treated as reliable neighbors in the concealment process whenever no "Correctly received" immediate neighbor of a "Lost" MB exists. In such cases, an unsuccessfully concealed MB can result in propagation of this concealment mistake to several neighboring concealed MBs. Therefore, the order in which "Lost" MBs are concealed is important. The processing starts with MB columns at the frame boundaries and then moves inwards column-by-column. This processing order helps to prevent a typical concealment mistake that is made in the usually "difficult" (discontinuous motion areas, large coded prediction error) center part of the frame from propagating to the "easy" (continuous motion area, similar motion over several frames) parts of the frame.

Fig. 1 shows a snapshot of the status map during the concealment phase where already concealed MBs have the status of "Concealed", and the currently processed (being concealed) MB is marked as "Current MB".

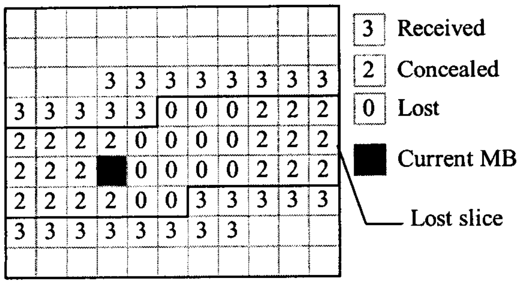

Fig. 1. MB status map at the decoder

3. THE SELECTED EC ALGORITHMS

3.1. INTRA Frame Concealment

Lost areas in INTRA frames have to be concealed spatially as preceding frames may not resemble the INTRA frame. The proposed spatial concealment algorithm is based on weighted pixel averaging [3]. The weight used for averaging is simply the inverse distance between the source and destination pixels. Only "Correctly received" neighboring MBs are used for concealment if at least two such MBs are available. Otherwise, neighboring "Concealed" MBs are also used in the averaging operation.

3.2. INTER Frame Concealment

Instead of directly operating in the pixel domain a more efficient approach is to try to "guess" the motion in the missing pixel area (MB) by some prediction schemes from available motion information of spatial or temporal neighbors. This "guessed" motion vector (MV) is then used for motion compensation using the reference frame. The copied pixel values give the final reconstructed pixel values for concealment, and no additional pixel domain operations are used [4].

The motion activity of the correctly received slices of the current picture is investigated first. If the average MV is smaller than a pre-defined threshold (currently ¼ pixels for each MV component), all lost slices are concealed by copying from the spatially corresponding positions in the reference frame. Otherwise, motion-compensated error concealment is used, and the MVs of the lost MBs are predicted as described in the following paragraphs.

The motion of a "Lost" MB is predicted from a spatial neighbor MB's motion relying on the statistical observation, that motion of spatially neighboring frame areas is highly correlated. For example, in a frame area covered by a moving foreground scene object, the MV field is continuous, which means that it is easy to predict.

The MV of a "Lost" MB is predicted from one of the neighbor MBs (or 8x8 blocks). This approach assumes that the MV of one of the neighbor MBs (or blocks) models the motion in the current MB well. It was found in previous experiments, that median or averaging over all neighbors' MVs did not give better results. For simplicity, in the current implementation the smallest neighbor block size that is considered separately as prediction is set to 8x8 Y pixels. The motion of any 8x8 block is calculated as the average of the motion of the spatially corresponding 4x4 or other shaped (e.g. 4x8) blocks.

The decision of which neighbor's MV to use as prediction for the current MB is made based on the smoothness of the concealed (reconstructed) image. During this trial procedure the concealment pixel values are calculated using the MV of each candidate (motion compensated pixel values). The MV that results in the smallest luminance change across block boundaries when the block is inserted into its place in the frame is selected. (see Fig. 2). The zero MV case is always considered and this copy concealment (copy pixel values from the spatially corresponding MB in the reference frame) is evaluated similarly as other MV candidates.

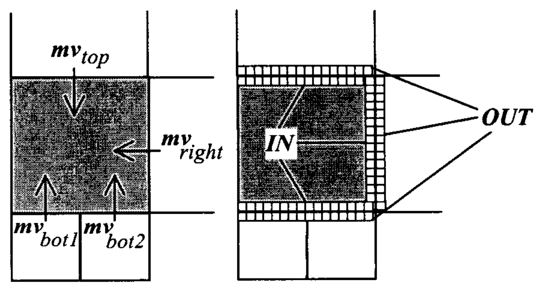

Fig. 2. Selecting the motion vector for prediction

The winning prediction MV is the one which minimizes the side match distortion d_{sm}, which is the sum of absolute Y pixel value differences of the *IN*-block and neighboring *OUT*-block pixels at the boundaries of the current block, as shown in Eqn. (1):

$$\min_{dir\in\{top,bot,left,right\}} \arg\left\langle d_{sm} = \frac{1}{N}\sum_{j=1}^{N}\left|\hat{Y}(\mathbf{mv}^{dir})_j^{IN} - Y_j^{OUT}\right|\right\rangle \qquad (1)$$

where $\hat{Y}(\mathbf{mv}^{dir})_j^{IN}$ is the j-th concealed Y value in the *IN*-blocks using $\mathbf{mv}^{dir}$ to predict the MV, and Y_j^{OUT} is the j-th reconstructed Y value in the *OUT*-blocks, and N is the total number of the calculated boundary pixels.

When "Correctly received" neighbor MBs exist, the side match distortion is calculated only over them. Otherwise, "Concealed" neighbor MBs are included in the calculation.

4. ADDITIONAL IMPLEMENTATION ISSUES

4.1. Handling of Multiple Reference Frames

When multiple references are used, the reference frame of the candidate MV is used as the reference frame for the current MB. That is, when calculating the side match distortion, the IN-block pixels come from the reference frame of the candidate MV.

4.2. Handling of Entire Frame Losses

Since multiple reference frames might be used in H.26L, and in the reference software a simple sliding window buffer model is used, a picture is referred using its index in the buffer. Consequently, when entire frames are lost, the reference buffer needs to be adjusted. Otherwise, the subsequent received frames would use wrong reference frames. To solve this problem, the reference picture ID (the ID which counts the number of frames that are used as reference frames, therefore B-frames are not counted) is used to infer whether and how many entire frames are lost, and the picture indices in the sliding window buffer are shifted appropriately.

4.3. B-frame Concealment

For B-frames, a simple MV prediction scheme according to the prediction mode of the candidate neighbor MB is used as follows: If the prediction mode of the candidate MB is

1. forward prediction mode, use the forward MV as the prediction, in the same way as for P frames.
2. backward prediction mode, use the backward MV as the prediction.
3. bi-directional prediction mode, use the forward MV as the prediction, and discard the backward MV.
4. direct prediction mode, use the backward MV as the prediction.

Note that 1) Each MV, whether forward or backward, has its own reference frame. 2) An Intra coded block is not used as a motion prediction candidate.

5. SIMULATIONS

The simulations were carried out using H.26L TML-8.4 [5] under the H.323/Internet conditions [6] of ITU-T Video Coding Experts Group. The proposed algorithms were compared to unmodified TML-8.4, which concealed errors by copying the spatially corresponding macroblocks from the previous frame.

The PSNR of each decoded video sequence was calculated for 10 runs of the packet loss pattern files [7]. The next run in the simulation continued from where the previous run ended. The average PSNR of the 10 runs was calculated. PSNR was calculated between each and every frame of the source sequence (at full frame rate) and the corresponding reconstructed frame, including repetitions of previous frames in place of skipped and lost frames. Since no rate control strategy was implemented in the TML software, we acquired the desired bit rates by simply adjusting the quantization parameter (QP). QP was fixed for all the frames (including the I-frames).

The slice interleaving packetization mechanism was used. It is based on interleaving even and odd numbered GOB-shaped slices, and arises from two design considerations. First, since the packetization overhead for the IP/UDP/RTP headers in the Internet is 40 bytes per packet, reasonably big packets have to be used. Second, to allow effective error concealment, consecutive slices should not be placed in the same packet. This results in two packets per picture and enables reasonable concealment of missing macroblocks if only one of the two packets is lost. A maximum payload size of 1400 bytes was assumed, and a new packet was established every time one of the packets in the initial pair for the same picture was about to exceed 1400 bytes. The packet header overhead was subtracted from the target bitrate to obtain the available video bitrate.

The simulation results with and without the EC feature are shown in table 1. To stop infinite error propagation, one GOB for each frame is forced to be INTRA coded. Fig. 3 shows a few decoded pictures from the Foreman sequence.

From the objective results, we can see that the EC feature significantly improves the performance of the original H.26L decoder in high-motion sequences (Foreman, Irene), and performs similarly or slightly better in low-motion sequences (Hall, Paris). As can be seen from Fig. 3, the subjective quality is greatly improved.

6. CONCLUSION

The error concealment feature in the H.26L test model is presented in this paper. The feature provides H.26L with a basic error resiliency for the decoder, and could, therefore, be used as a basis for comparison for any new error robust coding proposal. Simulation results show important improvement in both objective and subjective quality. Future work may focus on more sophisticated error concealment techniques to further improve performance.

7. REFERENCES

[1] T. Wiegand (editor), "H.26L Test Model Long-Term Number 9 (TML-9) draft0", ITU-T Video Coding Experts Group, document VCEG-N83d1, December 2001.

[2] H.26L Test Model Software, version 9.0, December 2001.

[3] P. Salama, N. B. Shroff, and E. J. Delp, "Error Concealment in Encoded Video Streams", Chapter 7, in A. K. Katsaggelos and N. P. Galatsanos (editors), *"Signal Recovery Techniques for Image and Video Compression and Transmission"*, Kluwer Academic Publishers, 1998.

[4] W. -M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in Proc. ICASSP'93, Minneapolis, Apr. 1993, pp. V417–V420.

[5] H.26L Test Model Software, version 8.4, August 2001.

[6] S. Wenger, "Common conditions for wire-line, low delay IP/UDP/RTP packet loss resilient testing", ITU-T Video Coding Experts Group, document VCEG-N79r1, September 2001.

[7] S. Wenger, "Error patterns for Internet video experiments", ITU-T SG16 document Q15-I-16-R1, October 1999.

Table 1. Average luminance PSNR (unit: dB) results obtained by the TML codec without and with the EC feature. The frame rates for Foreman, Hall, Irene and Paris are 7.5, 10, 30 and 15 frames per second, respectively.

Foreman 64 kbps	Packet Loss Rate				Hall 32 kbps	Packet Loss Rate			
	3%	5%	10%	20%		3%	5%	10%	20%
No EC	23.24	22.35	21.04	19.56	No EC	27.89	27.75	27.44	26.77
EC	23.77	22.97	21.91	20.41	EC	27.88	27.75	27.44	26.77
Irene 384 kbps	Packet Loss Rate (%)				Paris 144 kbps	Packet Loss Rate (%)			
	3%	5%	10%	20%		3%	5%	10%	20%
No EC	34.06	32.33	30.17	27.23	No EC	25.99	25.53	24.80	23.55
EC	34.74	33.32	31.26	28.66	EC	26.01	25.58	24.91	23.84

Fig. 3. Subjective image quality comparison: the top three are the decoded pictures without error concealment, and the bottom three are the corresponding ones with error concealment, all under 10% packet loss rate.