16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

VIDEO PACKET LOSS CONCEALMENT DETECTION BASED ON IMAGE CONTENT

Dor Shabtay, Nissan Raviv, and Yair Moshe

Signal and Image Processing Laboratory (SIPL), Department of Electrical Engineering, Technion - IIT
Technion City, 32000, Haifa, Israel
email: yair@sipl.technion.ac.il
web: sipl.technion.ac.il

ABSTRACT

Compressed video transmitted over unreliable channels such as wireless networks or the Internet may suffer from packet loss. A packet loss leads to image impairment that may cause significant degradation in image quality. In most practical systems, packet loss is detected at the transport layer and a decoder error concealment post-processor tries to mitigate the effect of lost packets. This helps to improve image quality but could still leave some noticeable impairments in the video. In some applications such as no-reference video quality evaluation, detection of concealment impairments is desired. If only video coding layer information is available, concealment has to be detected based on image content. In this paper, a novel technique for detecting impairments in a video sequence caused by packet loss after its concealment is proposed. The detection is based on image content only. Simulation results show good temporal and spatial concealment detection accuracy.

1. INTRODUCTION

Video compression is widely employed in many applications due to the need for bit rate reduction for transmission or storage. The ITU H.261/3/4 and ISO MPEG-1/2/4 are well-known video compression standards that share the same hybrid DPCM-transform structure. These algorithms compress the video by removing redundancies using a 2D transform, temporal and possibly also spatial prediction, and a variable length coder. However, as the output of the encoder contains less redundant information, it is more sensitive to channel errors. In the context of a hybrid DPCM-transform codec structure, variable length coding may lead to a loss of synchronization at the decoder, which can cause error propagation. Since a packet typically contains information of a few dozen macroblocks, its loss could cause an impairment that spans several consecutive macroblock rows. Moreover, temporal prediction by motion compensation may cause error propagation to consecutive Inter (temporally predicted) frames, until the next Intra frame.

Packet loss is typically detected at the transport layer or at the video syntax layer. Then, a decoder post-processing algorithm tries to mitigate the effect of lost packets by concealing the corrupted macroblocks. Concealment fills in the corrupted region by utilizing redundancies in the received video. Packet loss concealment helps to improve image quality but could still leave some noticeable impairments in the video.

(a) (b)

Figure 1: Error concealment impairments: (a) Temporal concealment – Foreman CIF, frame #8. (b) Spatial concealment – Tempete CIF, frame #16.

In some applications, detection of concealment impairments is desired. One example is no-reference video quality evaluation, e.g. [1][5]. Although image quality could be significantly degraded by packet loss, most modern no-reference video quality evaluation measures do not consider packet loss concealment impairments. Error concealment detection is the first step in estimating subjective concealment impairment severity. If only video coding layer information is available, concealment has to be detected based on image content. This could be the case if, for example, a video sequence has been coded, transmitted with errors, decoded with appropriate error concealment, and saved to disk. At a later time, a detection of concealment impairments in the video sequence is desired.

There are essentially two approaches for video packet loss error concealment – temporal concealment and spatial concealment. For temporal concealment, missing motion vectors are interpolated and damaged regions are filled in by applying motion compensation. The simplest and most commonly used technique in this category is just selecting the zero motion-vector. Temporal error concealment does not work well when the video sequence contains unsmooth moving objects or in the case of a scene change. An example of an impairment caused by temporal concealment is shown in Figure 1a.

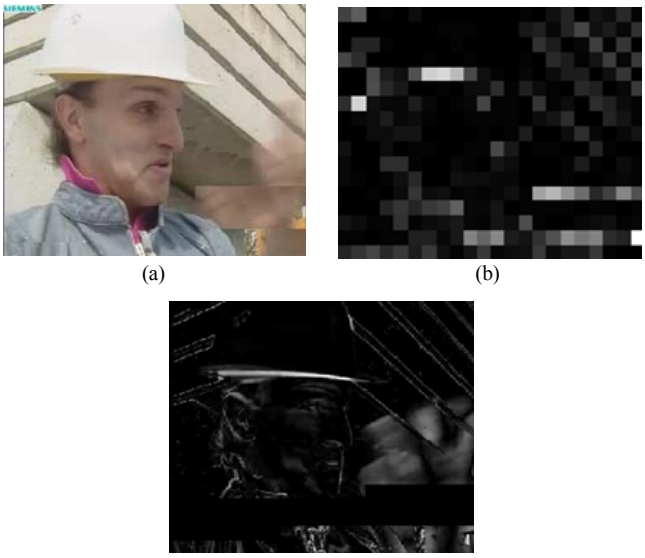
For spatial concealment, spatially neighboring pixels are used to conceal the corrupted macroblock. A corrupted macroblock is usually interpolated using boundary pixels

surrounding it, so the concealed macroblock satisfies a smoothness property. One of the simplest and most commonly used techniques in this category is bilinear interpolation. Spatial concealment works well only on smooth regions, but it could be the only plausible technique if motion information is unavailable or in the case of a scene change. An example of an impairment caused by spatial concealment is shown in Figure 1b.

Only few previous works for packet loss concealment detection are available. Temporal concealment detection based on image content and a model for its subjective impairment severity is first described in [1]. Detection is based on the assumption that sharp edges in a frame are rarely aligned with macroblock boundaries. Based on this assumption, strong edges at both the upper boundary and the bottom boundary of a macroblock row indicate a concealed macroblocks row. The intersample difference both at the horizontal boundary of each macroblock row and inside that macroblock row, are carefully checked. If certain conditions are satisfied, that edge point is counted for the overall metric. The same detection technique with minor changes is also used in [2][3][4], possibly combined with bitstream information. In [3] it is reported that when used alone, this pixel-based technique results in high false alarm rates. A different subjective impairment severity model is proposed in [5] but with a similar technique used in the detection process. Spatial discontinuity is calculated as an average of the absolute difference of the luminance values at the boundary between correctly decoded regions and regions to which error-concealment has been applied. All these works do not consider impairments that span more than a single macroblock row and do not consider spatial error concealment.

Some previous works for (unconcealed) packet loss detection could also be relevant for concealment detection. Almost all these works use the transport layer or the video syntax layer; few also use image content for detection. Algorithms for detecting transmission errors in JPEG images [6], H.261 video sequences [7], and MPEG-2 video sequences [8][9] were proposed in the past. First, error existence is determined by checking a set of syntax-based conditions. Then, the precise location of the error is located using a set of spatial error detection measures. According to the assumption that an unimpaired region holds similar measurement values for neighboring macroblocks, each macroblock with exceptional measurement values is regarded as erroneous. In [10] a DCT coefficient-based error detection technique is suggested. The properties of the DCT coefficients of each macroblock are utilized to calculate the average luminance values and to detect the presence of edges. Erroneous macroblocks are detected by monitoring the variation of average energy from macroblock to macroblock.

The works described in [6][7][8][9][10] share two drawbacks that make them inappropriate for solving the aforementioned problem. First, they deal only with scattered macroblock errors and do not consider packet loss. Second, these works do not assume that corrupted regions have been concealed. Our attempt to explicitly adapt the measures presented in these works for packet loss concealment detection yielded unsatisfactory results.



(a) (b)

(c)

Figure 2: Temporal concealment measures for Foreman CIF. (a) Frame #16 with a corrupted region. (b) The average of intersample difference across the macroblock upper boundary (AIDB-U) is large at the upper macroblocks of the corrupted region and at the macroblocks below the corrupted region - larger values are indicated by lighter colors. (c) The absolute difference between consecutive frames is negligible at the corrupted region.

In this paper, a novel algorithm for packet loss concealment detection is presented. Both spatial and temporal concealment are detected based only on luminance image content. In spite of the technical feasibility of detecting concealment impairments propagation in many cases, error concealment is detected only in the first corrupted frame. This decision is motivated by a future complementary algorithm that may evaluate impairment severity, including its propagation to following frames. The proposed algorithm was tested with spatial bilinear interpolation concealment and temporal zero motion vector concealment but the techniques used for detection are not limited to these specific concealment techniques. Each packet is assumed to contain between 10 and 100 macroblocks so impairment may span several macroblock rows.

The rest of the paper is organized as follows. The proposed algorithms for temporal concealment detection and spatial concealment detection are described in Section 2 and Section 3, respectively. Results are given is Section 4 and conclusions are drawn in Section 5.

2. TEMPORAL CONCEALMENT DETECTION

Temporal error concealment fills in missing macroblocks by copying regions from the previous frame. In case of an unsmooth or fast movement or in case of a scene change, an unnatural discontinuity might result across the boundaries of the concealed region. This undesirable effect can be observed in Figure 1a. Measures that will be used for detecting such a discontinuity are the average of intersample

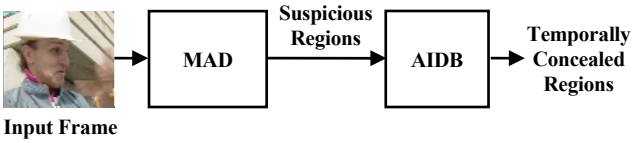


Figure 3: Temporal concealment detection.

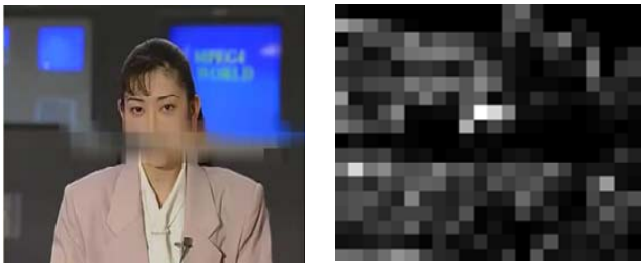
difference across the macroblock upper boundary (AIDB-U) and the average of intersample difference across the macroblock lower boundary (AIDB-L). The intersample difference across the macroblock right boundary and the intersample difference across the macroblock left boundary are not used due to the fact that consecutive macroblocks are lost. Thus, corrupted region's right and left boundaries are short and contain less distinguishing information. For a macroblock of size $N \times N$, the AIDB-U is computed as:

$$AIDBU = \frac{\sum_{x=0}^{N-1} |I(x,0) - I(x,-1)|}{N}$$

where $I(x,y)$ is the intensity value at location (x,y) in a local coordinate system. The AIDB-L is computed in a similar way. Values of AIDB-U are shown in Figure 2b. It can be observed that large AIDB-U values appear at the upper macroblocks of the corrupted region and at the macroblocks below the corrupted region. Similarly, large AIDB-L values typically appear at lower macroblocks of the corrupted region and at macroblocks above the corrupted region.

The most commonly used temporal concealment technique is to copy corresponding macroblocks from the previous frame (zero motion-vector). As a result, the difference between consecutive frames is negligible at macroblocks concealed using this technique. The difference is not necessarily zero at these regions due to quantization during possible reencoding of the video sequence. This distinguishing property of temporally concealed regions can be observed in Figure 2c and can be quantified by computing the mean absolute difference (MAD) with the previous frame for each macroblock.

Figure 3 illustrates the proposed algorithm for temporal error concealment detection based on these two characteristics of concealed regions – large AIDB-U and AIDB-L values across the upper and lower horizontal boundaries together with negligible MAD values of the concealed macroblocks. The MAD is computed for every macroblock and the current frame is searched for contiguous sets of macroblocks with negligible MAD values. The AIDB-U and AIDB-L are computed for the upper and lower boundaries of these suspicious regions. Any contiguous set of macroblocks with large AIDB-U and large AIDB-L values is detected as temporally concealed. In order to enable robustness to varied images properties and to possible requantization, thresholds for detecting negligible MAD and for detecting large AIDB-U and large AIDB-L values are adaptively computed. These thresholds are computed relatively to the mean of the corresponding measure values in the frame.



(a) (b)

Figure 4: Spatial concealment macroblock variance measure for Akiyo CIF (a) Frame #1 with a corrupted region. (b) The macroblock variance is small at the corrupted region - larger values are indicated by lighter colors.

(a) (b)

Figure 5: Spatial concealment macroblock average vertical gradient (AVG) measure for Paris CIF. (a) Frame #46 with a corrupted region. (b) The vertical gradient is small at the corrupted region - larger values are indicated by lighter colors.

3. SPATIAL CONCEALMENT DETECTION

Spatial packet loss concealment could also cause noticeable impairments. A corrupted macroblock is usually interpolated using boundary pixels surrounding it so the concealed macroblock satisfies a smoothness property. Packet loss causes consecutive macroblocks to be corrupted. The concealment algorithm conceals the impairment one macroblock at a time in some order, fulfilling the smoothness property only with neighboring uncorrupted macroblocks and possibly also with neighboring macroblocks that have already been concealed. As a result, more upper and lower neighboring macroblocks than right and left neighboring macroblocks are available. This results in an effect that may be described as 'vertical bars' – a region that is vertically smooth and contains horizontal discontinuities. Horizontal discontinuities are prevalent but not limited to macroblock right and left boundaries. The 'vertical bars' effect can be observed in Figure 1b.

Three measures are used for spatial concealment detection. Two measures facilitate the detection of smooth macroblocks - macroblock variance and the macroblock average vertical gradient (AVG). The third measure facilitates the detection of horizontal discontinuities using vertical edge detection in the macroblock's horizontal gradient (VEHG).

Spatial concealment creates spatially smooth macroblocks. Thus, the variance of the pixel values in such a

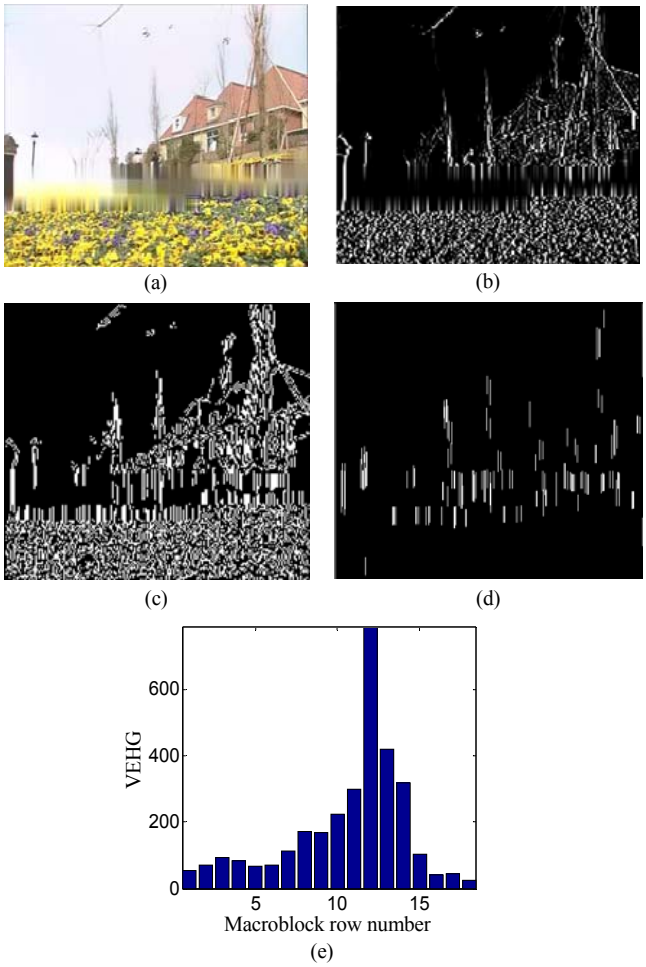


Figure 6: Spatial concealment vertical edge detection in the macroblock's horizontal gradient (VEHG) measure for Flower CIF. (a) Frame #136 with a corrupted region. (b) Horizontal gradient contains vertical edges at the corrupted region. (c) Canny edge detection in the horizontal gradient image emphasizes edges. (d) Morphological image opening leaves only vertical edged that are at least of the height of a macroblock. (e) Corrupted macroblock rows exhibit large VEHG values.

macroblock is typically small. This can be observed in Figure 4.

The macroblock average vertical gradient (AVG) quantifies vertical macroblock smoothness. The vertical derivative of the frame is computed using a simple $[-1\ 0\ 1]^T$ mask operation. Then, vertical derivative values in every macroblock are averaged. AVG values in spatially concealed macroblocks are typically small, as can be observed in Figure 5.

Horizontal discontinuities are discovered using vertical edge detection in the macroblock's horizontal gradient (VEHG). First, the horizontal derivative of the frame is computed using a simple $[-1\ 0\ 1]$ mask operation. The resulting vertical gradient image contains vertical edges at the corrupted regions, but these edges may be discontiguous or not strong enough (Figure 6b). Thus, the Canny edge detector is used on the vertical gradient image. The Canny edge detector was chosen due to its non-maximal suppression and hysteresis stages that connect and emphasize weak

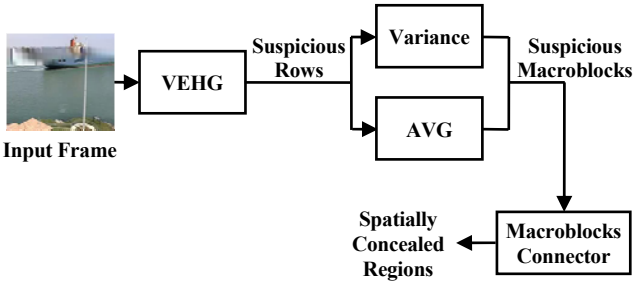


Figure 7: Spatial concealment detection.

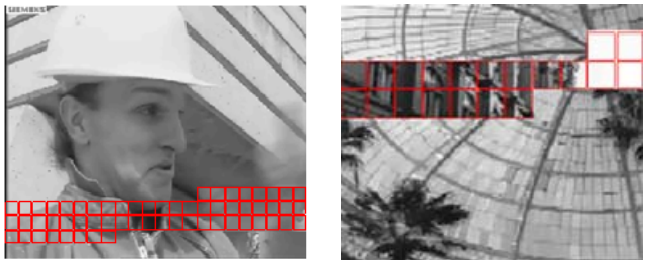
edges. The result is a binary image of (vertical and non-vertical) edges (Figure 6c). Then, a morphological image opening is used with a structuring element that is a vertical line, 16 pixels in length. This morphological image opening operation leaves only vertical edges that are at least of the height of a macroblock (Figure 6d). Finally, the number of vertical edges in each macroblock row is summed to result in the VEHG of that macroblock row (Figure 6e).

Figure 7 illustrates the proposed algorithm for spatial error concealment detection based on these three characteristics of concealed regions – large VEHG, small variance, and small AVG values of concealed macroblocks. The VEHG is computed for every macroblock row and macroblock rows with large VEHG values are considered as suspicious. The variance and AVG are then computed for each macroblock in a suspicious macroblock row. If both variance and AVG values are small, the macroblock is considered as suspicious. The resulting suspicious macroblocks may be discontiguous. Thus, a final stage rechecks with less strict thresholds macroblocks in suspicious rows that are between suspicious macroblocks in a sequential scan order. Then, only contiguous sets of suspicious macroblocks are considered as spatially concealed regions. As for temporal concealment detection, all thresholds are adaptively computed in order to enable robustness to varied images properties and to possible requantization.

4. RESULTS

To evaluate the effectiveness of the proposed method, 30 standard video sequences in QCIF (176x144) and CIF (352x288) resolution were used. Some video sequences were encoded using an H.263 baseline encoder and the rest using an H.264/AVC baseline encoder. The video sequences were reencoded after concealment. Output bitrates were between 100Kbps and 1Mbps. Packet loss was randomly simulated using a varied ratio in the range 0.5%-1%. The length of each packet was also randomly simulated to contain between 10 and 100 consecutive macroblocks so impairment may span several macroblock rows. Packet loss distribution was i.i.d. Zero motion-vector was used for temporal concealment and bilinear interpolation was used for spatial concealment. The algorithm had no prior knowledge of the concealment type so both temporal concealment detection and spatial concealment detection was performed for each sequence.

Examples of detection results can be observed in Figure 8 and in Figure 9.



(a) (b)

Figure 8: Temporal concealment detection examples. (a) Foreman CIF, frame #158. (b) Glasgow QCIF, frame #25.

(a) (b)

Figure 9: Spatial concealment detection examples. (a) Tempete CIF, frame #16. (b) Akiyo CIF, frame #1.

Evaluation of the results was performed both per packet loss and per macroblock. For per packet loss evaluation, a correct concealment detection is a detection of a true concealed region, maybe with missing or with extra macroblocks. In case of error propagation to following frames, a correct detection is considered a detection of the corrupted region in the first corrupted frame. The precision (fraction of the detections that are correct) and recall (fraction of the concealments that are successfully detected) measures were used. For per macroblock evaluation, detection precision and detection recall were evaluated for every macroblock separately (here again, in case of error propagation a correct detection is only in the first corrupted frame).

The results are summarized in Table 1. For both spatial and temporal concealment detection, precision and recall are high. Temporal concealment is more easily detected due to the severe impairments it may cause. The tradeoff between precision and recall can be controlled by changing the global thresholds for the different measures used.

5. CONCLUSIONS

In this paper, a novel technique for packet loss concealment detection is described. This technique is able to detect both spatially and temporally concealed macroblocks based only on image content. Such a technique could be useful as part of a no-reference video quality evaluation measure. The proposed algorithm can handle video sequences with varied image properties and is robust to reencoding. Simulation results show precision of 0.92 and recall of 0.93 for temporal concealment detection and precision of 0.89 and recall of 0.90 for spatial concealment (measured per packet loss event).

Per packet loss:	Precision	Recall
Temporal	0.92	0.93
Spatial	0.89	0.90
Per macroblock:		
Temporal	0.90	0.91
Spatial	0.86	0.71

Table 1: Packet loss results per packet loss and per macroblock.

ACKNOWLEDGMENT

The authors would like to thank the head of the Signal and Image Processing Laboratory (SIPL), Prof. David Malah, and the chief engineer of SIPL, Nimrod Peleg, for their support and valuable comments.

The work described in this paper has been initiated and supported by RADVISION. The authors would like to thank their contact persons at RADVISION – Yair Wiener and Tamar Barzuza.

REFERENCES

[1] R. V. Babu et al., "No-Reference Metrics for Video Streaming Applications," in *Proc. Int. Workshop on Packet Video*, Dec. 2004.

[2] H. Rui, C. Li, and S. Qiu, "Evaluation of Packet Loss Impairment on Streaming Video," *Journal of Zhejiang University SCIENCE B*, vol. 7 (1), pp. 131-136, Jan. 2006.

[3] A. R. Reibman and D. Poole, "Characterizing Packet-loss Impairments in Compressed Video," in *Proc. ICIP 2007*, San Antonio, TX, USA, Oct. 2007, vol. 5, pp. V-77-V-80.

[4] A. R. Reibman and D. Poole, "Predicting Packet-loss Visibility Using Scene Characteristics," in *Proc. Packet Video 2007*, Lausanne, Switzerland, Nov. 2007, pp. 308-317.

[5] T. Yamada, Y. Miyamoto, and M. Serizawa, "No-Reference Video Quality Estimation Based on Error Concealment Effectiveness," in *Proc. Packet Video 2007*, Lausanne, Switzerland, Nov. 2007, pp. 288-293.

[6] Y. H. Han and J. J. Leou, "Detection and Correction of Transmission Errors in JPEG Images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8 (2), pp. 221-231, April 1998.

[7] W. J. Chu and J. J. Leou, "Detection and Concealment of Transmission Errors in H.261 Images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8 (1), pp. 74-84, Feb. 1998.

[8] H. C. Shyu and J. J. Leou, "Detection and Concealment of Transmission Errors in MPEG-2 Images - A Genetic Algorithm Approach," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9 (6), pp. 937-948, Sep. 1999.

[9] J. Cao and Z. Wang, "Analysis and Detection of Transmission Errors for MPEG-2 Video Signal," in *Proc. APCCAS 2000*, Tianjin, China, Dec. 2000, pp. 105-108.

[10] K. Bhattacharyya and H. S. Jamadagni, "DCT Coefficient-Based Error Detection Technique for Compressed Video Stream," in *Proc. ICME 2000*, New York, NY, USA, July-Aug. 2000, vol. 3, pp. 1483-1486.