ON THE EFFECTS OF ENCODER-DECODER CONCEALMENT MISMATCH ON VIDEO DISTORTION ESTIMATION

Fabio De Vito[§] *and Juan Carlos De Martin*[*]

[§]Dipartimento di Automatica e Informatica/[*]IEIIT-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [fabio.devito|demartin]@polito.it

ABSTRACT

Many advanced video transmission techniques rely on per-packet distortion estimates. To compute reliable estimates, however, the decoder inner workings, including the concealment module used in case of packet losses, should be fully known at the encoder. This paper explores the effects on video distortion estimation of encoder-side erroneous assumptions about the concealment technique used by the decoder. Several concealment techniques, roughly representative of the main families of concealment algorithms (i.e., spatial, temporal and hybrid), have been implemented and then distortion has been evaluated at the encoder using an analysis-by-synthesis approach for all possible combinations of encoder-decoder concealment pairs. The results for nine, widely known, test video sequences show that as long as the encoder concealment algorithm belongs to the same family of the decoder one, the effect of concealment mismatch on distortion estimation is quite small. The results have also been validated by measuring the effects of suboptimal distortion estimation on packet classification for video transmission on a 2-class DiffServ IP network. Simulations results show that, for intra-family concealment mismatch, packet misclassifications affect only 1–5 % of the packets, yielding, on average, perceptually variations of only about 0.2 dB PSNR.

1. INTRODUCTION

The increasing importance of multimedia communications over IP networks has generated, in recent years, a large number of proposals aimed at improving the perceptual quality experienced by end users. IP networks, in fact, do not guarantee the service quality, therefore, techniques to contain the effects of packet losses, delay and jiter have to be implemented at the application layer.

Several state-of-the-art media encoders implement resilience tools, including, e.g., resynchronization markers, forward error correction, packet classification, and layer coding [1]. Such tools are used to prevent errors, to stop their propagation or to concentrate losses in given low-importance regions (i.e., packets). At the decoder side, an error concealment module is typically included to lower the distortion caused by packet losses.

Determining the perceptual importance on each individual packet has become, in recent years, a prerequisite for several multimedia communications techniques. A family of techniques, for instance, has been recently proposed for audio and video transmission over Differentiated Services (DiffServ) IP networks [2], which support several *classes* of traffic with varying quality of service; several works [3, 4, 5] showed that perception-based packet classification can considerably increase the quality of the received video.

Video coding standards, however, usually do not define the concealment algorithm, which may then vary even among compliant implementations of the same standard. The non-standardization of the concealment module affects the estimation of the perceptual importance of video packets. Reliable estimation of the perceptual importance of multimedia packets, in fact, depends on the specific error concealment technique employed at the decoder. When the concealment used at the encoder does not match the actual decoder–side concealment algorithm, distortion estimates are affected by an error. This paper aims at estimating such error by sistematically studying the effects of encoder-decoder concealment mismatch on distortion estimation.

The paper is organized as follows: technical background is presented in Section 2. The concealment techniques implemented for this work are described in Section 3, while per-packet distortions and misclassification results are reported in Section 4. Section 5 shows perceptual results for the case of video transmission over DiffServ networks. Conclusions are drawn in Section 6.

2. BACKGROUND

In the case of video coding, a large number of error concealment algorithms —see, e.g.,[6, 7, 1, 8, 9]— have been proposed. The techniques can be grouped –with some degree of approximation, which can be excused considering that this paper does not address concealment per se— in three main families: *spatial* algorithms, that interpolate the missing information using surrounding data within the same frame. *Temporal* concealments techniques, which mask errors using information from a previously decoded frame, either selecting MBs according to motion vectors of neighbouring macroblocks or simply replicating the pixels in the same position of the lost ones. *Mixed* concealment approaches, which are a combination of spatial and temporal approaches; the most popular uses a spatial concealment on the I-frame and temporal concealment on P- and B- frames. Mixed schemes often combine the good performance of temporal approaches with the absence of error propagation to following GOPs due to the self-concealment of the I-frames. Recently, a distinction between *first* and *second* generation algorithms has been proposed [10]; second generation concealment techniques are based on the training of a model for the selection of concealing information.

An example of usage of per–packet distortion estimation is multimedia communications over DiffServ networks. In that case, an accurate identification of the packets that should experience lower delays or lower packet loss rates is key in delivering high perceptual quality to the end users. In a multimedia stream, in fact, not all the portions of the compressed bitstream have the same importance. Encoder-side packet classification depends on the concealment used at the decoder; if this is not known, it potentially leads to packet misclassification, and, consequently, to lower perceptual quality. The problem of the encoder-decoder concealment mismatch and its effects on distortion estimation, although mentioned in [11], has not been extensively studied so far.

3. CONCEALMENT TECHNIQUES

The concealment algorithms studied in this paper are first generation ones. We implemented several techniques within the H.264 reference code [12], overriding the concealment already present within the reference decoder [13]; in this work we use ten different algorithms, defined as follows:

- spatial algorithms:
  - sp_1 copy of the uppermost neighboring MB, if available;
  - sp_2 copy of the leftmost neighboring MB, if available;
  - sp_3 for each 4x4 block, average color of the three upper-left 4x4 neighboring blocks;
  - sp_4 for each MB, average color of the three upper-left MBs;
- temporal algorithms:
  - te_1 copy of the MBs in the same position of the lost ones, in the previous I- or P-frame;
  - te_2 copy of the MBs pointed by the average of surrounding MVs;
  - te_3 predict MVs as an extension of the previous P-frame MVs;
- mixed algorithms: obtained as sp_3 on the I-frame and one of the temporal algorithms on remaining frames; mix_1, mix_2 and mix_3 use respectively te_1, te_2 and te_3.

The above techniques have been chosen to cover a wide spectrum of approaches, although the list is by no means intended to be exhaustive; our focus is mainly centered on the behavior of the different algorithmic families.

4. PER-PACKET DISTORTION AND MISCLASSIFICATION

The distortion introduced by a packet loss (assumed to be isolated) is measured by the *Mean Square Error* (MSE) between the correctly decoded sequence and the corrupted one. To lower the computational complexity of this measure, in this work future frame distortion is estimated using a statistical model of future distortion as described in [14]. Table 1 shows the average per-packet MSE values for two of the nine test sequences analyzed for this work; similar results have been obtained for the remaning seven test sequences.

The per-packet MSEs of the spatial algorithms is at least one order of magnitude higher than the other two families, while spatial and mixed approaches show closer values. In all cases, intra-family distortions are very close to each other. As a consequence, we expect, for the case of transmission of DiffServ networks, that the number of misclassified packets for intra-family mismatch should be significantly lower than misclassifications between algorithms belonging to different families. The results shown in Table 2 confirm our expectations.

The percentage of misclassified packets, in fact, when encoder and decoder concealments belong to the same family is always below 6%, while it is never lower than 13% if the algorithms belong to different families. Similar results have been obtained also for premium bandwidths of 10% and 30%.

The marking patterns for a particular sequence, i.e., precisely which packets are assigned to the premium class, strongly

Table 1. Average MSE values for different concealment algorithms and sequences.

Sequence name	Concealment name	Average per-pcket MSE
foreman	sp_1	6509.7
	sp_2	8179.7
	sp_3	6373.1
	sp_4	6372.7
	te_1	260.9
	te_2	173.7
	te_3	187.8
	mix_1	312.2
	mix_2	235.0
	mix_3	247.0
tempete	sp_1	1101.3
	sp_2	1440.8
	sp_3	1007.0
	sp_4	988.4
	te_1	71.5
	te_2	42.4
	te_3	53.7
	mix_1	94.7
	mix_2	68.0
	mix_3	78.6

depend on the concealment family used, with minor differences among algorithms within the same family. The quality of the video obtained at the decoder side is then the result of the correctness of the marking pattern generated by the encoder, which tries to concentrate losses in low-importance regions, as well as the result of the actual performance of the decoder–side concealment algorithm.

5. DIFFSERV TRANSMISSION RESULTS

Nine, widely known test sequences have been encoded and transmitted over a simulated 2-class (best effort and premium) DiffServ network, with 20% of premium bandwidth. Each sequence has been marked according to all of the proposed algorithms, transmitted and then decoded using all of the available decoder concealments, for an aggregate of one hundred pairs of encoder and decoder algorithms for each of the nine sequences. Table 3 shows the PSNR values obtained with three decoders and all the encoders, for sequences *foreman* (high motion), *mobile* (medium motion) and *news* (slow motion). PSNR values are computed with respect to the original uncompressed sequence.

Concealments belonging to the same family show similar PSNR results. Encoders and sequences not shown, due to space constraints, in Table 3 exhibit the same behavior.

Results are affected by three factors: the correctness of the importance estimation, the masking capability of the specific decoder concealment employed, and, in case of mix_x family, the lack of inter-GOP error propagation. The best

Table 2. Average percentage of misclassified packets for couples of algorithms belonging to the same family and to different families; two classes, 20% premium bandwidth.

Sequence name	Misclassified packets (%) same family	different families
foreman	3.352	14.683
tempete	4.332	13.558
mobile	5.238	14.603
news	2.354	16.150
akiyo	1.420	26.237
silent	3.408	14.747
sean	1.715	21.822
paris	2.822	14.771
table	3.772	14.291

performance is most of the time achieved by matching pairs of concealments, since in that case the predicted importance of a packet is the closest possible to the real impact experienced at the decoder side. If the encoder matches at least the family of the decoder algorithm, performance is only very slightly affected, while PSNR degrades much more markedly —from more than half a dB to several dB's— if families do not match. Results also show that, at least at the considered packet loss rate, decoder–side temporal concealment techniques deliver better performance than the other two families, whatever the encoding algorithm is.

6. CONCLUSIONS

In this paper we addressed the problem of the mismatch between the concealment implemented at the decoder and the one used for distortion estimation at encoder side. We implemented several error concealment algorithms both at encoder and decoder side, and performed a study of the per-packet MSE values for nine, widely known test video sequences.

Packet classification results for different encoder-side algorithms show that the percentage of misclassified packets is very low for concealment algorithms belonging to the same family, and quite high in case of different families. This behavior was confirmed by network simulationa; we demonstrated that in almost all cases the perfect matching between the two concealments ensures best perceptual performance. Moreover, it is sufficient to just match the family of the decoder algorithm, while marked performance degradations (up to several dB's PSNR) are observed when the concealments belong to different families.

Finally, as a side result, it was shown that the temporal concealment algorithms studied in this work behave sensibly better than the mixed and spatial approaches.

Table 3. PSNR with respect to the original uncompressed sequence as a function of the concealment algorithms; DiffServ network with 20% premium bandwidth and 10% packet loss rate.

Sequence name	Encoder	PSNR (dB) Decoder sp_3	te_2	mix_2
foreman	sp_1	30.51	31.05	31.11
	sp_2	30.34	31.19	31.24
	sp_3	30.51	31.05	31.11
	sp_4	30.51	31.05	31.11
	te_1	29.70	31.35	30.35
	te_2	29.19	31.33	29.77
	te_3	29.75	31.35	30.39
	mix_1	30.43	31.19	31.49
	mix_2	30.60	31.19	31.58
	mix_3	30.63	31.21	31.62
mobile	sp_1	22.30	25.20	22.51
	sp_2	22.03	25.82	22.42
	sp_3	21.89	25.08	22.10
	sp_4	22.05	25.09	22.32
	te_1	22.12	26.25	22.91
	te_2	22.20	26.23	22.91
	te_3	22.16	25.99	22.63
	mix_1	22.69	25.66	23.30
	mix_2	23.54	26.04	24.23
	mix_3	23.35	25.85	24.02
news	sp_1	27.65	34.30	28.24
	sp_2	27.69	34.64	28.37
	sp_3	27.54	34.49	28.06
	sp_4	27.54	34.49	28.06
	te_1	27.64	34.98	28.52
	te_2	27.66	35.00	28.43
	te_3	27.40	34.95	28.17
	mix_1	25.29	33.19	26.37
	mix_2	25.26	33.12	26.35
	mix_3	25.28	32.90	26.33

7. REFERENCES

[1] Y. Wang, S. Wenger, J. Wen, and A.K. Katsaggelos, "Real-time video communications over unreliable networks," in *IEEE Signal Processing Magazine*, July 2000, pp. 61–82.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," *RFC 2475*, December 1998.

[3] J. Shin, J. Kim, and C.-C. J. Kuo, "Quality-of-service mapping mechanism for packet video in differentiated services network," *IEEE Transactions on Multimedia*, vol. 3, no. 2, pp. 219–231, June 2001.

[4] F. De Vito, L. Farinetti, J.C. De Martin, "Perceptual classification of MPEG video for Differentiated-Services communications," in *Proc. IEEE Int. Conf. on Multimedia & Expo*, Lausanne, Switzerland, August 2002, vol. 1, pp. 141–144.

[5] F. D'Agostino and E. Masala and L. Farinetti and J.C. De Martin, "A simulative study of analysis-by-synthesis perceptual video classification and transmission over diffserv IP networks," in *Proceedings of ICC*, Anchorage, Alaska, May 2003, vol. 1, pp. 572–576.

[6] P. Salama, N. Shroff, and E. Delp, "Error concealment in encoded video streams," in *Signal Recovery Techniques for Image and Video Compression and Transmission*. 1998, N. P. Galatsanos and A. K. Katsaggelos, Kluwer Academic Publishers, Boston.

[7] S. Cen and P. Cosman, "Comparison of error concealment strategies for MPEG video," in *IEEE Wireless Communications and Networking*, New Orleans, September 1999, pp. 329–333.

[8] Y. Wang and Q. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[9] B.W. Wah, X. Su, and D. Lin, "A survey of error-concealment schemes for real-time audio and video transmissions over the internet," in *Proceedings of the 2000 International Conference on Microelectronic Systems Education*. 2000, p. 17, IEEE Computer Society.

[10] T.P. Chen and T. Chen, "Second-generation error concealment for video transport over error-prone channels," *Wireless Communications and Mobile Computing*, vol. 2, pp. 607–624, October 2002.

[11] G. Cote, S. Shirani, and F. Kossentini, "Optimal model selection and synchronization for robust video communications over error-prone networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952–965, June 2000.

[12] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "DRAFT H.264 Standard ISO/IEC 14496-10 (MPEG-4 Part 10)," June 2003.

[13] Y. Wang, M.M. Hannuksela, V. Varsa, A. Hourunranta, and M. Gabbouj, "The error concealment frature in the h.26l test model," in *Proc. IEEE Int. Conf. on Image Processing*, September 2002, vol. 2, pp. 729–732.

[14] F. De Vito, D. Quaglia, and J.C. De Martin, "Model-based distortion estimation for perceptual classification of video packets," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, September 2004, vol. 1, pp. 79–82.