SPATIO-TEMPORAL VIDEO ERROR CONCEALMENT WITH PERCEPTUALLY OPTIMIZED MODE SELECTION

S. Belfiore, M. Grangetto, E. Magli, G. Olmo

CERCOM - Center for Multimedia Radio Communications
Dipartimento di Elettronica - Politecnico di Torino
Corso Duca degli Abruzzi 24 - 10129 Torino - Italy
Ph.: +39-011-5644195 - Fax: +39-011-5644099
`grangetto(magli,olmo)@polito.it`
`belfiore@mail.tlc.polito.it`
URL: `www.cercom.polito.it`

ABSTRACT

We propose a spatio-temporal error concealment algorithm for video transmission in an error-prone environment. The proposed technique employs motion vector estimation, edge-preserving interpolation, and texture analysis/synthesis. It has two main advantages with respect to existing methods, namely i) it aims at optimizing the visual quality of the restored video, and not only PSNR, and ii) it employs an automatic mode selection algorithm in order to decide, on a macroblock basis, whether to use the spatial restoration, the temporal one, or a combination thereof. The algorithm has been applied to H.26L video, providing satisfactory performance over a large set of operating conditions.

1. INTRODUCTION AND STATE-OF-THE-ART

A considerable effort is recently being spent in the development of effective techniques for the transmission of compressed video sequences over networks potentially subject to packet losses. In these applications, one has to cope with the fact that the network may be subject to congestion, thus causing some packets to be unusable at the receiver side because of excessive delay. In order to provide a minimum degree of quality even in case of packet losses, it is necessary that the decoder be equipped with error concealment tools, in such a way as to estimate the missing data from the received ones, and to optimize the end-to-end quality of service.

As for video sequences, both spatial and temporal error concealment algorithms have been proposed. Algorithms of the temporal class typically require less computational resources. Proposed techniques include, amongst others, projection onto convex sets [1] and boundary matching. Temporal concealment may fail in the presence of scene changes, fast motion, rotation and deformation of objects [2]. Therefore, a practical algorithm should also include a spatial interpolator to recover missing macroblocks (MB), e.g. based on a DCT domain formulation [3] or on projection onto convex sets [4]. Spatio-temporal techniques have also been proposed, using boundary matching [2] or Markov random fields (MRF) [5, 6, 7].

In most spatio-temporal error concealment algorithms, once a corrupted MB or slice has been detected, one has to decide whether to apply spatial or temporal restoration. For example, in [8] the spatial part is only used in the first I frame of the video sequence, or in case of scene changes. In [5, 9] the spatial algorithm is applied to all MBs in I-frames, while the temporal one to all MBs in P-frames. In [6] it is assumed that the coding modes are known; thus, the spatial algorithm is applied to all intra-coded MBs, and the temporal one to all inter-coded ones. It must be noticed that spatial and temporal concealment provide quite different visual quality and PSNR. If the motion vectors (MV) can be estimated with a satisfactory degree of accuracy, temporal concealment usually yields better results, since the replaced MB is sharp and fairly similar to the original. Conversely, if one employs inaccurate MVs, discontinuities are likely to appear. Spatial interpolation usually provides blurred estimates of the missing MB; however, it does not introduce discontinuities, and is hence preferable to an inaccurate temporal estimate. Moreover, one should be cautious in the interpretation of the coding mode of a MB or frame. Intra frames or MBs do not necessarily mean that the lost MB cannot be easily predicted from the reference frame; as an example, several encoders force a certain percentage of intra frames or MBs to improve robustness to bit errors or packet losses.

In this paper we propose a spatio-temporal error concealment algorithm, based on MV estimation and edge-preserving MAP estimation based on a MRF model. We propose to use gradient-based boundary matching (GBM) directional information to set the parameters of the temporal and spatial concealment algorithms, independently of the original MB and frame coding modes, in order to achieve optimal reconstruction as for PSNR and visual quality.

2. PROPOSED ALGORITHM

The proposed error concealment algorithm is sketched in Fig. 1. A buffer stores the reference frame for prediction. First, MVs are re-estimated by searching for the best matching MB in the reference frame, using classical boundary matching. Then, the restoration mode is selected based on the value of the GBM metric. Three possibile modes are supported, namely i) fully temporal reconstruction, ii) combined spatio-temporal reconstruction, and iii) fully spatial reconstruction. Details of the temporal and spatial algorithms, along with their combination, are given in the following.

0-7803-7663-3/03/$17.00 ©2003 IEEE

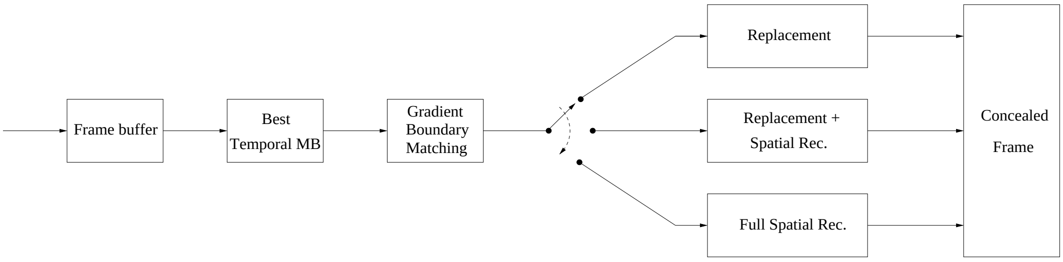


Fig. 1. Block scheme of the proposed algorithm

2.1. Temporal restoration

The MV estimation is based on the algorithm in [6] (see the reference paper for more details). Given the 16x16 lost MB, four 8x8 blocks are considered, i.e. the ones on the top and left side of the MB. For each of these blocks, motion estimation with respect to the reference frame is performed, thus yielding four candidate MVs. Among the four corresponding MBs, the one that exhibits the best boundary matching with the surrounding MBs is selected as replacement MB.

2.2. Spatial restoration

Spatial restoration is performed using the coarse-to-fine MRF interpolator in [7], improved with texture synthesis in order to optimize visual appearance, as in [10]. The algorithm first finds a smooth estimate of the missing MB by means of an averaging filter. It then performs edge-directional interpolation based on the MRF model in [6], first at the MB level, and then at the quarter-MB level [7]. The MRF interpolator converges after a certain number of iterations (typically 70 for spatial restoration).

Finally, the algorithm analyzes texture in the neighboring MBs and, if a texture template compatible with the current reconstruction is found, a new texture is synthesized and merged with it. In this paper we do not focus on implementation details, but rather on mode selection; the interested reader can find more information in [7, 10].

2.3. Mode selection

The mode selection rule is based on the value of the new GBM metric. This metric is designed so as to penalize directional edge discontinuity between the replacement MB corresponding to the estimated MB, and the surrounding part of the frame, since this kind of discontinuity negatively impacts on PSNR and especially visual quality.

In the GBM, we consider the pixels on the internal upper and lower boundaries of the missing MB, replaced by the MB in the reference frame pointed at by the estimated MV; we do not use the left and right boundaries, since they can be missing if a complete slice is lost. For each pixel of these boundaries, we consider a clique c and a complement c' as in Fig. 2, and eight directions between the current pixel and each component of its clique. For each direction, we define a counter c_m, $m = 1, \ldots, 8$ for each pixel of the clique and its complement, which is used to measure edge strength and direction for the current pixel; unlike [6], we use this information as a block matching metric.

In particular, we consider a window of 5x5 pixels centered in the current boundary pixel; for each pixel of the window, we compute the gradient magnitude G and phase Θ. Upon detection of an edge passing through the clique or its complement, we increment by G the counter related to the edge direction.

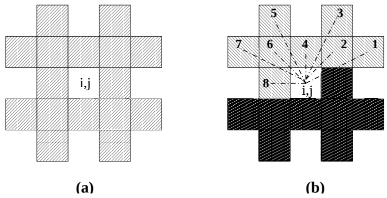


Fig. 2. (a) a pixel and its 16-pel clique; (b) a pixel, its clique c and the complement c' (dark area), and the eight directions.

The information on the counters is used to evaluate the degree of matching of the replenishment block. A pixel of the boundary is labeled as "correctly aligned" (i.e. with good directional boundary match) if, in at least five of the eight directions, the corresponding counters c_m of the clique and its complement both exceed a threshold ξ, or they are both below it; otherwise, a bad match is declared. The threshold ξ is selected automatically as a function of the standard deviation of the current frame, as $\xi = \frac{\sigma^2}{2}$.

Based on the labels assigned by the GBM metric, mode selection is performed as follows; notice that the rule does not depend on the frame or MB coding modes, this latter ones likely having been lost along with the DCT coefficients of the MB.

- If most boundary pixels of a MB (>90%) exhibit a good match, this means that the MV estimate is accurate. As a consequence, temporal replacement is performed using the MB pointed at by the estimated MV. In this case, temporal replacement is preferable because it provides a sharper and hence more pleasant reconstruction.
- If not enough pixels (<75%) exhibit a good match, the MV estimate is likely inaccurate. Thus, temporal replacement would provide a reconstruction with discontinuities. In this case fully spatial restoration is performed as described in Sect. 2.2.
- If the percentage of pixels with good match is between 75% and 90%, an intermediate solution is adopted. The MB pointed at by the estimated MV is used, not as replacement, but as initial estimate for the spatial reconstruction method. Fewer iterations of the MRF interpolator are required (say 4), whose main purpose is to smooth out discontinuities across the MB boundary. Thus, the obtained reconstruction is sharper than the fully spatial interpolation, but reasonably well connected to the neighboring MBs.

3. EXPERIMENTAL RESULTS

The proposed algorithm with GBM mode selection (labeled as GBM-MS in the following) has been tested using H.26L [11] as coding framework (JM 2 codec, ver. 2.1). In the following we report the results obtained on two QCIF video sequences, namely *foreman* and *carphone*; the first one has been coded at 188 kbps with a temporal resolution of 30 fps, while the second one has been coded at 90 kbps with a temporal resolution of 10 fps; slices are formed so as to contain 11 MBs. The proposed algorithm has been compared with two other mode selection rules. The first one applies the temporal algorithm to all inter-coded MBs, and the full spatial reconstruction to all intra-coded ones, as in [6] (notice that this is possible only if the coding modes are correctly received); in the following it is labeled as MB-MS. The second one applies the temporal algorithm to predictive frames, and the full spatial reconstruction to intra-coded frames [8, 9]; in the following it is referred to as F-MS.

We report the results achieved by the proposed algorithm and the other two methods concealing an intra-coded frame (Fig. 3 and 4) and an inter-coded frame (Fig. 5 and 6) from the two sequences. We assume a loss of about 20% of the image (i.e. two slices) due to network congestion.

Fig. 3. An intra-coded frame from the *carphone* sequence: **(a)** original; reconstruction using **(b)** F-MS, **(c)** MB-MS, **(d)** GBM-MS

Let us consider the *carphone* sequence, and particularly the I-frame shown in Fig. 3; in this case the two MB-MS and F-MS methods apply a fully spatial technique that can hardly recover the lost information, because only the above and below MBs provide uncorrupted information. The GBM-MS algorithm is capable of exploiting the temporal redundancy also if the frame is intra-coded, and is able to improve the temporal reconstruction if it is not smoothly connected to the surrounding area, by integrating it with a spatial interpolation or by replacing it with a full-spatial reconstruction.

In Fig. 5 the recovery of Inter-coded MBs is shown: in this case the F-MS method reconstructs all the MBs by using the temporal algorithm, thus generating some artifacts. Moreover, the

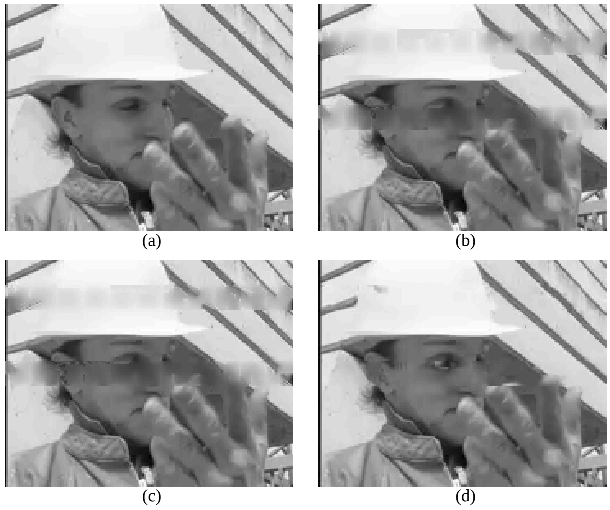


Fig. 4. An intra-coded frame from the *foreman* sequence: **(a)** original; reconstruction using **(b)** F-MS, **(c)** MB-MS, **(d)** GBM-MS

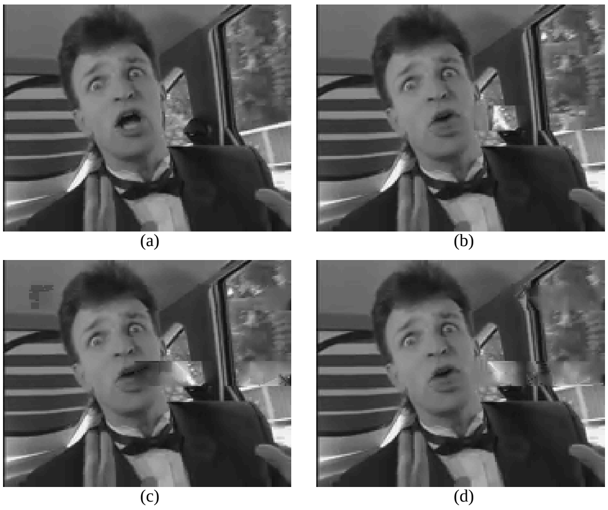
MB-MS also produces some artifacts due to an inappropriate use of the spatial reconstruction, and to an inaccurate estimation of MVs. Conversely, the GBM algorithm provides a more pleasant reconstruction by avoiding the discontinuities caused by the mismatch between the missing MB and its replacement.

Analogous considerations can be made for the *foreman* sequence (Fig. 4 and 6). As can be seen in Fig. 4, the reconstruction quality of the intra-frame is significantly improved by exploiting temporal information. Let us then consider the P-frame shown in Fig. 6: due to the large amount of motion of this part of the sequence, the MB-MS and the F-MS methods, which mainly exploit the temporal redundancy, produce reconstructions affected by remarkable artifacts. On the contrary, the GBM method attempts to smooth the discontinuities produced by the fast camera movement, thus providing a better reconstruction.

As can be seen in Tab. 1, the PSNR values obtained by the three methods confirm these considerations, showing that the proposed method is able to yield reconstructions that are more similar to the original image than those provided by the other methods. These values have also been compared with those obtained by the built-in error concealment algorithm of H.26L [12]. As can be seen, in the intra frames our proposed technique outperforms the built-in algorithm, though at the expenses of a higher complexity. The results in the P frames are not comparable, and are reported only for completeness, since the built-in algorithm does employ the MVs of available neighboring MBs, while the proposed one does not.

4. CONCLUSIONS

In this paper we have proposed a spatio-temporal error concealment algorithm that aims at optimizing both visual quality and PSNR. We have shown that, by means of mode selection, a careful combination of the spatial and temporal interpolators can provide largely improved results. Application to H.26L video has shown that the proposed algorithm provides satisfactory performance with respect to other strategies.



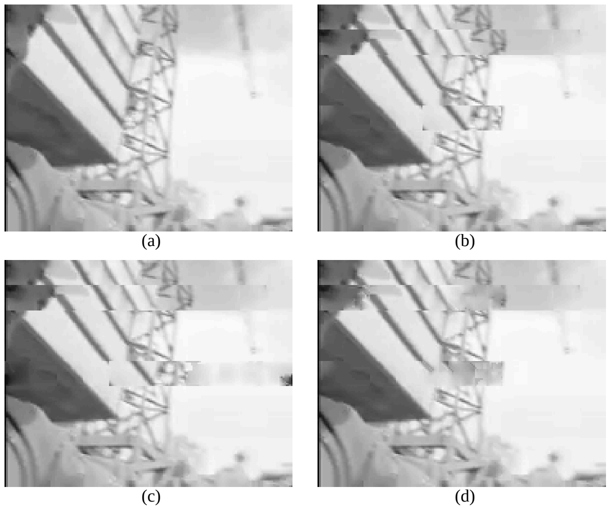


Fig. 5. An inter-coded frame from the *carphone* sequence: **(a)** original; reconstruction using **(b)** F-MS, **(c)** MB-MS, **(d)** GBM-MS

Fig. 6. An inter-coded frame from the *foreman* sequence: **(a)** original; reconstruction using **(b)** F-MS, **(c)** MB-MS, **(d)** GBM-MS

Table 1. PSNR (dB) comparison for the *carphone* and *foreman* video sequences (20% packet loss rate)

Method	**I-frame**	**P-frame**
Carphone		
F-MS	26.29	25.56
MB-MS	26.29	25.14
GBM-MS	28.52	26.17
H.26L	27.74	27.90
Foreman		
F-MS	26.70	25.18
MB-MS	26.68	24.48
GBM-MS	32.15	26.23
H.26L	25.71	36.71

5. REFERENCES

[1] G.S. Yu, M.M.K. Liu, and M.W. Marcellin, "POCS-based error concealment for packet video using multiframe overlap information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 422–434, Aug. 1998.

[2] L. Atzori, F.G.B. De Natale, and C. Perra, "A spatio-temporal concealment technique using boundary matching algorithm and mesh-based warping (BMA-MBW)," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 326–338, Sept. 2001.

[3] Z. Alkachouch and M.G. Bellanger, "DCT-based spatial domain interpolation of blocks in images," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 729–732, Apr. 2000.

[4] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 470–477, Apr. 1995.

[5] P. Salama, N.B. Schroff, and E.J. Delp, "Error concealment in MPEG video streams over ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1129–1144, June 2000.

[6] S. Shirani, F. Kossentini, and R. Ward, "A concealment method for video communications in an error-prone environment," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1122–1128, June 2000.

[7] S. Belfiore, L. Crisà, M. Grangetto, E. Magli, and G. Olmo, "Robust and edge-preserving video error concealment by coarse-to-fine block replenishment," in *Proc. of ICASSP - IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[8] S. Tsekeridou and I. Pitas, "MPEG-2 error concealment based on block-matching principles," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 646–658, June 2000.

[9] L. Wai-Man and A.R. Reibman, "An error concealment algorithm for images subject to channel errors," *IEEE Transactions on Image Processing*, vol. 4, no. 5, pp. 533–542, May 1995.

[10] S. Belfiore, L. Crisà, M. Grangetto, E. Magli, and G. Olmo, "An edge and texture preserving algorithm for error concealment in video sequences," in *Tyrrhenian International Workshop on Digital Communications (IWDC)*, 2002.

[11] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Document JVT-B118r2, "Working Draft Number 2, Revision 2 (WD-2)"*, Mar. 2002.

[12] W. Ye-Kui, M.M. Hannuksela, V. Varsa, A. Houruwranta, and M. Gabbouj, "The error concealment feature in the H.26L test model," in *Proc. of ICIP - IEEE International Conference on Image Processing*, 2002.