Temporal concealment of packet-loss related distortions in video based on Structural Alignment

Ajit S. Bopardikar, Odd Inge Hillestad and Andrew Perkis

Centre for Quantifiable Quality of Service in Communication Systems,
Norwegian University of Science and Technology,
Trondheim, Norway.
e-mail: {ajit, hillesta, andrew}@q2s.ntnu.no

Abstract

Concealment of packet-loss related distortions in multimedia communications assumes great importance with the emergence of packet switched communication using the IP protocol. This paper proposes a novel temporal algorithm to conceal packet-loss related distortion in video transmission. It is based on the insight that the spatial region around the area in the previous frame that yields the best concealment will in general be highly correlated with the corresponding area around the region to be concealed in the present frame. The proposed structural alignment algorithm (SAA) consistently yields better concealment and exhibits a superior ability to align edges. We compare performance of the proposed algorithm with that of a well known side matching algorithm (SMA). Finally, we note the strengths and the weaknesses of SAA and SMA and propose a combined concealment algorithm (CCA) that uses the strengths of both, SAA and SMA for more effective concealment. With their superior concealment ability, the SAA and CAA can find application in real-time concealment of packet-loss related artifacts in streaming video.

1 Introduction

The simplicity and flexibility of packet switched communication using the IP protocol has played an important role in its emergence as the method of choice for multimedia delivery. However, multimedia over IP and wireless networks faces many challenges due to network variability and lack of service guarantees with respect to available bandwidth and delay jitter. These result in packet-loss or, for streaming services, in packets being delivered after they are required. While playout buffers and protocols such as automatic repeat request (ARQ) try to mitigate the effects of delay jitters and lost packets to some extent, packet loss continues to be an important source of distortion for multimedia delivery.

Clearly, the effect of such losses on video depends on how the video stream has been coded, how it has been mapped into flows and packetized. Typically, for block-based video compression schemes such as the ISO/IEC and ITU video coding standards (e.g. MPEG-1/2/4, H-261/2/3/4), consecutive macroblocks (MBs) in a frame are transmitted together as a *slice* in a single packet. Thus, for many of these systems, packet-loss results in slices of a frame being lost.

Several methods have been proposed to minimize the effects of damaged and lost packets [1]. Source-based techniques for error robustness and recovery include forward error correction (FEC), multiple description coding, joint source and channel coding and, robust entropy coding [1, 2]. *Error concealment* (EC) by post processing methods, on the other hand try to minimize the perceptual effects of lost or damaged packets at the decoder. Apart from improving the perceptual quality of a given frame, EC methods attempt to minimize the errors that propagate through the decoder to subsequently decoded frames. EC methods can be classified into spatial and temporal error concealment methods.

Spatial EC methods utilize information from the neighboring MBs in a given frame to predict the information in lost packets. Algorithms proposed for spatial EC include bilinear interpolation-based methods in spatial and transform domain [1, 3, 4], projection on to convex sets (POCS) based [1] and the split-match algorithm[5]. While spatial interpolation methods are often simple, they can cause blurring.

Temporal EC methods [1, 3, 4, 5, 6, 7, 8, 9, 10, 11] on the other hand use appropriate sections from the previous frame (assumed to be correctly concealed or error-free) for concealment in the present frame. They make use of the high temporal correlation that typically exists between consecutive frames in video can improve concealment quality.

In this paper, we present novel temporal EC algorithms based on the insight that the spatial region around the area in the previous frame that yields the best concealment will in general be highly correlated with the corresponding area around the region to be concealed in the present frame. This results in algorithms that give improved edge and feature alignment. Given that the human visual system is sensitive to edges and structures[12], such algorithms are capable of producing superior concealment.

The paper is organized as follows. Section 2 briefly describes the structure of video streams and relevant aspects of video packetization using the MPEG-2 video bitstream as an example. Section 3 gives an overview of existing algorithms for temporal EC and describes the side matching algorithm (SMA). A novel temporal algorithm based on matching structural information called the structural alignment algorithm (SAA) is presented in Section 4 and its performance is compared to the SMA algorithm, two-line search as described in [11].Section 5 proposes an algorithm that is based on using the strengths of the SAA and SMA and presents concealment results for this algorithm. Finally, Section 6 presents the conclusions.

2 Video Stream Structures

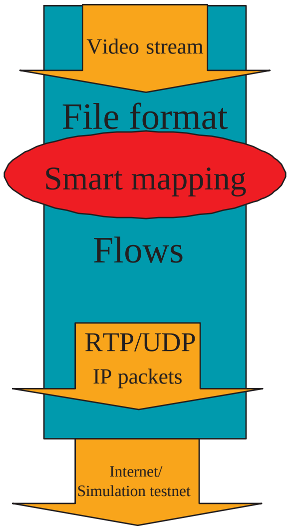


Fig. 1. General flowgraph for video stream packetization

Figure 1 shows a general flowgraph for video stream packetization. First, video frames are coded to produce a video stream wrapped in a particular file format depending on the compression scheme used. This video stream is divided into *flows* that are geared toward effective packetization and transportation by means of a *mapping*. Packetization can be carried out using for example, the RTP/UDP protocol. Such a scenario can represent a generalized version of packetization or can be envisaged in a test bed framework to study the performance of video transmission over packet-switched networks. As a specific example, we consider MPEG-2 [13] which we use for simulations in this paper. The basic unit for compression in MPEG-2 is the macroblock which is composed of four adjacent 8×8 pixels luminance blocks which form the 16×16 luminance macroblock and two 8×8 chrominance blocks. In general MPEG-2 video frames are coded on a macroblock-by-macroblock basis. For all the coding modes, namely, the I, P and B modes, consecutive macroblocks in a frame are grouped into several *slices*. The pictures (frames) between two successive I-frames constitute a group of pictures (GOP). For more information in MPEG-2 data structures see [13]

The syntax of the bitstream is shown in Figure 2 [14]. The bitstream of a frame is composed of the bitstreams of its constituent slices and a picture header. A slice bitstream likewise, is composed of bitstreams of its macroblocks and a slice header. Finally, a macroblock bitstream is preceeded by a header which

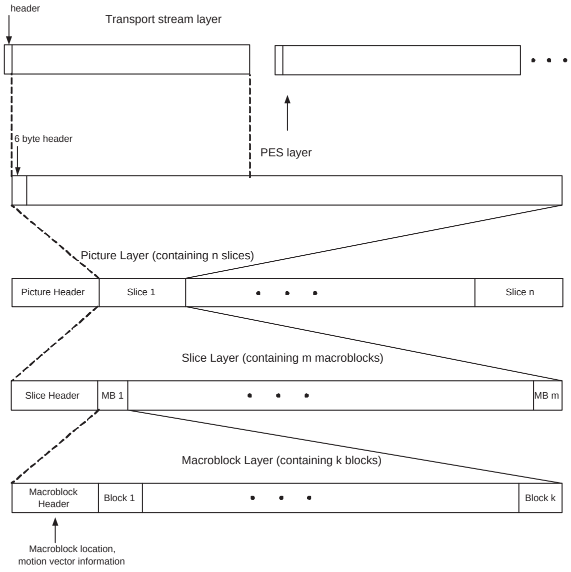


Fig. 2. MPEG-2 video bit-stream.

contains information about the macroblock location in the frame and motion vectors if required. Note that in a slice, while the motion vector associated with the first macroblock is coded absolutely, the rest are coded differentially.

For packetization purpose, the bit stream is divided into variable length flows called packetized elementary streams (PES) by a mapping operation. For example, one slice could be mapped into each PES as illustrated in Figure 2. PES packets are then broken into transport stream (TS) packets for transmission. This packetization strategy means each packet can be decoded independently, irrespective of errors in decoding previous packets. At the same time, loss of a packet means loss of spatial and motion vector information for the corresponding part of the frame. EC methods attempt to conceal the result of such losses that affect the overall video quality.

3 Previous Work

Figure 6(b) where the dark region represents information lost due to packet-loss, shows an example of packet-loss related artifacts. The simplest temporal EC algorithm that can be used to conceal such artifacts is *temporal replacement* (TR). TR consists of replacing lost and damaged macroblocks in the present frame with macroblocks from the same spatial location in the previous, error-free frame. However, in presence of significant motion, TR can give rise to visible discontinuity at slice boundaries corresponding to replaced macroblocks. In addition, edges in image content may also not be well-aligned.

One way to improve on TR is to make use of motion vector information for the lost macroblock to choose the appropriate macroblock-sized area from the previous, undamaged frame for concealment [1]. While this gives considerably better results, motion information is often lost along with the macroblock data. In such cases, motion vectors can be estimated from those corresponding to the macroblocks neighboring the area to be concealed. The boundary matching algorithm (BMA) [6] uses such motion vectors to effect concealment. First it assembles a set of candidate motion vectors and then chooses a best fit from among the macroblocks in the previous frame associated with them. The best fit macroblock is the one that minimizes the sum of mean squared error between its boundaries and the boundaries adjacent to them from the top, bottom and left macroblocks around the area to be concealed. For example, Chen et. al. [8] use the SAD in a full search iterative method to determine the best concealment.

An improvement over BMA-related algorithms were algorithms based on side matching [5, 3, 9, 10, 11]. Tsekeridou et. al. [5] proposed an algorithm that chose a candidate MB area from the previous frame for concealment for which the SAD between the top and bottom MBs of the MB to be concealed and the corresponding areas around the candidate MB area was minimum. Zhang et. al. [11] proposed a side matching algorithm (SMA) that compared 2 to 8 layers of pixels around the area to be concealed to the corresponding areas around the candidate macroblock. An example of such an area, 2 pixels thick, is shown in Figure 3.

Other algorithms along —similar lines were presented in [3, 10] while in [9] the authors proposed minimizing a weighed side matching algorithm. These matching algorithms in general give excellent concealment results but often fail to align strong edges. In the next section, we present an algorithm that is capable of doing the latter.

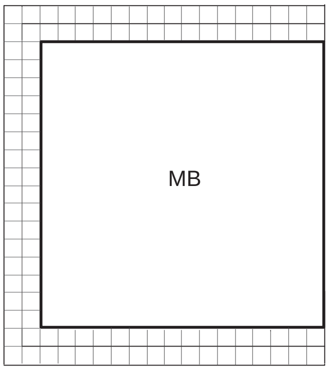


Fig. 3. An example of the area around an MB used for comparison in SMA.

4 Proposed error concealment algorithm

In this section, we present a novel EC algorithm called the structural alignment algorithm (SAA) based on the high degree of correlation that can exist between spatial region around the best-fit (that provides the best concealment) candidate macroblock in the previous frame and the corresponding region around the area to be concealed in the present frame. It is a refinement of the algorithm presented in [15].

Let F_0 denote the present frame to which EC has to be applied and let F_{-1} be the previous frame that we assume to be error free. MBs are assumed to have $n \times n$ pixels. The MB to be concealed is denoted by M_{0C} and the MBs surrounding it are M_{0T},M_{0TL}, M_{0L}, M_{0BL} and M_{0B}, respectively. A candidate MB in F_{-1} is denoted by M_{-1C} and the MBs surrounding it are M_{-1T},M_{-1TL}, M_{-1L}, M_{-1BL} and M_{-1B}, respectively. This is illustrated in Figure 4. The algorithm involves the following steps for each candidate MB, M_{-1C}:

1. First insert M_{-1C} between M_{0T},M_{0TL}, M_{0L}, M_{0BL} and M_{0B} to form a composite block, M_{0Comp}. Similarly, denote the composite block made up of M_{-1T},M_{-1TL}, M_{-1C}, M_{-1B}, M_{-1BL} and M_{-1L} by M_{-1Comp}. These are shown in Figure 5. Next, apply the 3×3 Sobel mask operators

$$H_x = \frac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

   and $H_y = H_x^T$, where ^{T} represents the transpose, to M_{-1Comp} and M_{0Comp} to get gradient maps in the x and the y directions, respectively. The horizontal and vertical gradient maps for M_{0Comp} are denoted by X_{0Comp} and Y_{0Comp}, respectively. Similarly, X_{-1Comp} and Y_{-1Comp} denote the horizontal and vertical gradient maps for M_{-1Comp}. The overall gradient magnitude for the two composite blocks is then computed as

$$G_{0Comp}(i,j) = |X_{0Comp}(i,j)| + |Y_{0Comp}(i,j)| \tag{1}$$

$$G_{-1Comp}(i,j) = |X_{-1Comp}(i,j)| + |Y_{-1Comp}(i,j)| \tag{2}$$

2. Compute

$$S_T = \sum_{i=n}^{2n-1} |G_{0Comp}(n-1,i) - G_{-1Comp}(n-1,i)|, \tag{3}$$

$$S_B = \sum_{i=n}^{2n-1} |G_{0Comp}(2n,i) - G_{-1Comp}(2n,i)|, \tag{4}$$

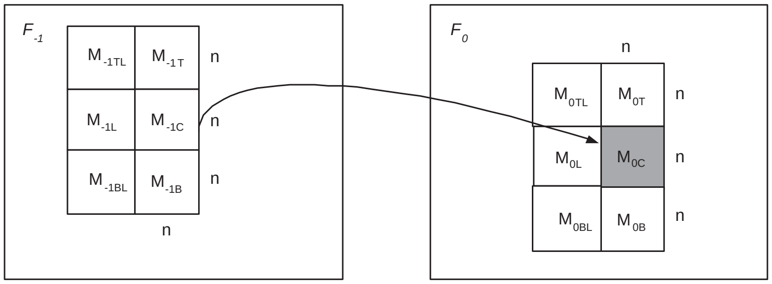


Fig. 4. Frame F_{-1} is assumed to be error-free and is used to conceal macroblock M_{0C} in frame F_0.

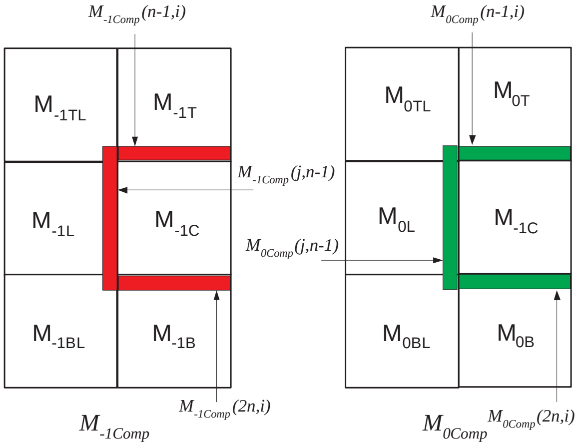


Fig. 5. Composite MBs M_{-1Comp} and M_{0Comp} in frames F_{-1} and F_0, respectively.

$$S_L = \sum_{i=n-1}^{2n} |G_{0Comp}(i, n-1) - G_{-1Comp}(i, n-1)|, \tag{5}$$

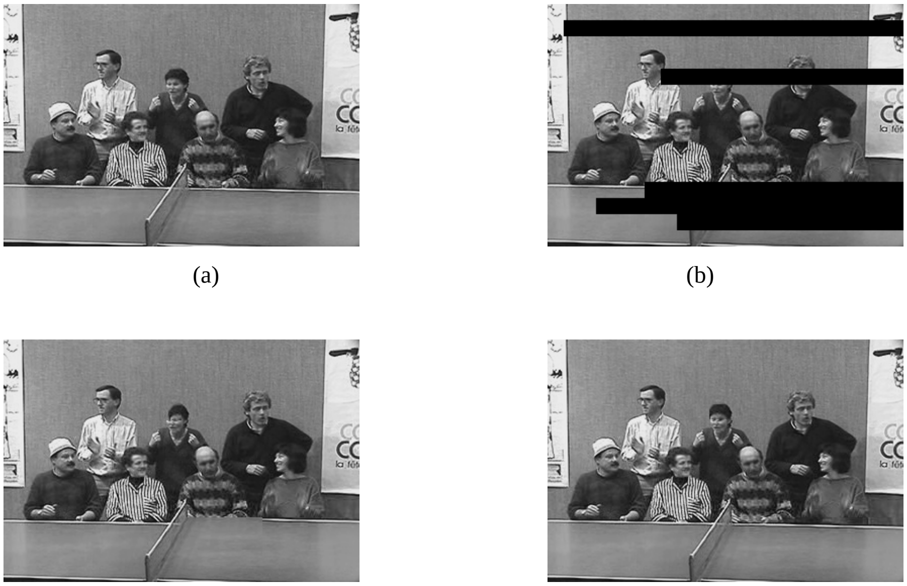
Note that $G_{0Comp}(n-1, i)$, $G_{0Comp}(2n, i)$, $G_{0Comp}(i, n-1)$ are the gradients corresponding to the last row of M_{0T}, first row of M_{0B} and last column of M_{0L}, respectively when M_{-1C} is used for concealment. Similarly,
$G_{-1Comp}(n-1, i)$, $G_{-1Comp}(2n, i)$ and $G_{0Comp}(i, n-1)$ are the gradients of the last row of M_{-1T}, first row of M_{-1B} and last column of M_{-1L}, respectively. They are represented by the shaded regions in Fig. 5.

3. Finally calculate $S = S_T + S_B + S_L$.

Error concealment is effected by inserting between M_{0T}, M_{0B} and M_{0L}, the candidate MB which gives minimum S. If there is more than one MB for which the minimum S is obtained then the one closest in distance to the concealment location is chosen.

S is computed using the available MBs neighboring the MB to be concealed. For example, for a MB in middle of the top slice of the frame, S will be computed using the bottom, bottom-left and the left MBs which are the only ones available. Note that the Sobel edge operators need not be applied to the entire composite block. Because, the gradient is only required for rows $n-1$ and $2n$ and column $n-1$ for M_{-1Comp} and M_{0Comp}, it is computed for only these rows and columns for both the composite blocks.

SAA is based on the assumption that the MBs surrounding M_{0C}, the lost MB in M_{0Comp}, will exhibit a strong correlation to their counterparts in M_{-1Comp}, corresponding to the best-fitting M_{-1C} in frame F_{-1}. The Sobel operation applied to the two composite blocks will tend to emphasize important structural information contained in the above rows. Inserting a poor fitting M_{-1C} will produce $G_{0Comp}(n-1, i)$, $G_{0Comp}(2n, i)$ and $G_{0Comp}(j, n-1)$ that are dissimilar to $G_{-1Comp}(n-1, i)$, $G_{-1Comp}(2n, i)$ and $G_{-1Comp}(j, n-1)$, respectively, because the Sobel edge operators will emphasize the discontinuities at the MB boundaries. This will give a higher S. On the other hand a candidate MB that fits better (provides better concealment) will result in greater correlation of structural details between both pairs of gradients and therefore a lower S. Because the comparison is between spatially equivalent rows around the concealment area and the candidate MBs, greater alignment between salient structures such as edges can be achieved.



(a) (b)

(c) (d)

Fig. 6. Error concealment performance of different algorithms for frame 418 in the 'table-tennis' sequence. (a) Original/error-free, (b) with packet-loss artifacts (PSNR=11.70), concealed with:(c) two-line search SMA (PSNR = 27.04)and (d) SAA (PSNR=33.74).

Thus, the SAA differs from the SMA in that it attempts to emphasize structural information to effect better concealment.

Finally, the candidate MBs in F_{-1} can be chosen based on a set of motion vectors belonging to MBs in F_0 around M_{0C}. Alternatively, they can be determined by performing full motion search around an area in F_{-1} spatially corresponding to M_{0C} in F_0.

4.1 Comparison with SMA

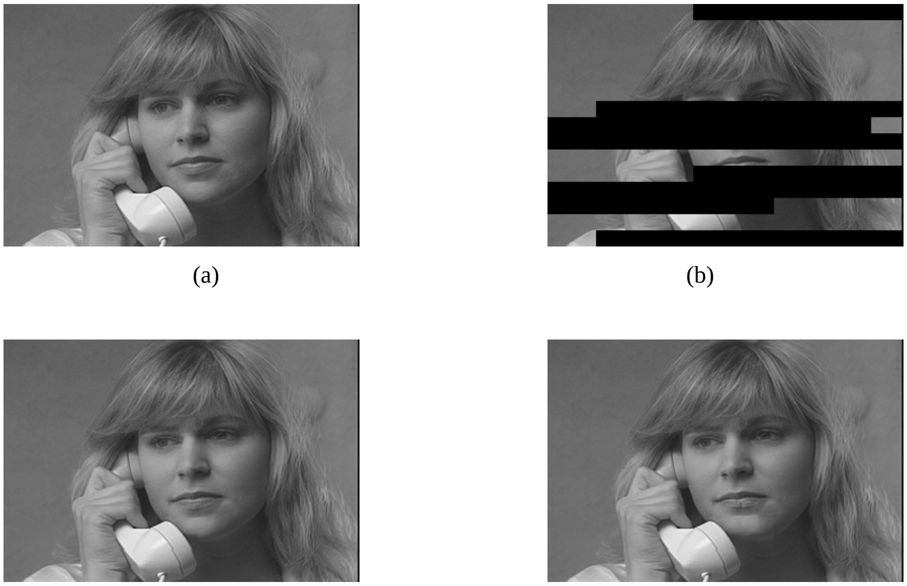
As mentioned above, the SAA is capable of effecting superior edge alignment. An example of this is given in Figure 6 which shows frame 418 of the "table-tennis" sequence in the SIF format (352×240) with a frame rate of 30 fps and a duration of 15 secs. It was MPEG-2 coded at 1.5 Mbps and packet-loss was simulated using the packet-loss generation software from NTT DoCoMo. Full motion search is used to determine the best-fitting MB for concealment. These specifications hold true for all other sequences used to generate results for this paper.

The result is compared with concealment performed by two-line search SMA [11] which is considered for comparison in our simulations because of its low computational complexity and good performance.

As can be seen from Figure 6 the SAA concealed frame exhibits superior concealment and edge alignment as compared to the SMA. On the other hand, SMA performs better in matching lower frequency variations and textured areas of the frame. This can be seen in Figure 7 for frame 120 of the sequence "susie" (note the concealment around the models chin in Figure 7(d)). The next section proposes an algorithm based on using the strengths of both the SMA and the SAA algorithm.

5 A Combined concealment algorithm

In this section, we propose an algorithm that makes use of the strengths of the SMA and SAA. For this purpose, we use the standard deviations, σ_T, σ_B and σ_L of $M_{0Comp}(n-1,i),\ i=n,\ldots 2n-1$, $M_{0Comp}(2n,i),\ i=n,\ldots 2n-1$ and $M_{0Comp}(j,n-1),\ j=n-1,\ldots 2n$, respectively which are the shaded rows and columns of M_{0Comp} in Figure 5. A high standard deviation indicates high activity such as a very busy texture or a strong edge. Thus, it is important to give such a boundary more importance in concealment. At the same time, the presence of a strong edge suggests the use of the SAA could result in better concealment. Based on these ideas, the combined concealment algorithm (CCA) involves the following steps:



(a) (b)

(c) (d)

Fig. 7. Error concealment performance of different algorithms for frame 120 in the 'Susie' sequence. (a) Original/error-free, (b) with packet-loss artifacts (PSNR=10.435), concealed with:(c) two-line search SMA (PSNR = 34.25)and (d) SAA (PSNR=34.17).

(a) (b)

Fig. 8. Error concealment performance of CCA for frame 120 in the 'Susie' sequence. (a) $\tau = 25$ **(PSNR= 34.8) (b)** $\tau = 50$ **(PSNR= 34.81).**

1. Compute σ_T, σ_B and σ_L.
2. If $\sigma_m = \max(\sigma_T, \sigma_B, \sigma_L) > \tau$ use the two-line search SAA for concealment, else use SMA.
3. $S = (\sigma_T * S_T + \sigma_B * S_B + \sigma_L * S_L)/(\sigma_T + \sigma_B + \sigma_L)$. Where S_T, S_B and S_L represent the sum of absolute difference in the case of SMA. For the case of SAA they represent differences (3), (4) and (5) plus similar differences for rows $n-2$, $2n+1$ and column $n-2$. Thus, we extend SAA to compare two rows and columns around the MB to be concealed and the candidate MB.

As before, the candidate MB that yields the minimum S is used for concealment. The parameter τ determines the threshold that determines the number of times each algorithm is used for EC. It can be chosen based on image content ot can be present to a particular value. Figure 8 presents concealment results for two values of τ namely, 25 and 50 for frame 120 of the sequence "susie". The CCA exhibits an improvement in the PSNR and improved concealment (again, note the concealment around the models chin and the phones mouthpiece). Table 1 gives PSNR values for various frames from different sequences where packet-loss related artifacts are concealed using CAA with $\tau = 25$ and 50 under consideration. Overall, we can see that the CAA consistently gives a superior PSNR performance.

Table 1. PSNR (dB) value for frames from three test sequences with packet-loss (PL) related distortions concealed using the SMA and CCA with parameters, 25 and 50.

Sequence	Fr. no.	PL (dB)	SMA (dB)	CCA (25) (dB)	CCA (50) (dB)
Susi	40	20.37	41.45	43.47	43.47
	67	17.53	37.82	43.53	43.67
Mobile	72	21.93	33.33	33.5	33.04
	139	13.19	30.17	30.0	30.17
	203	14.03	39.07	39.07	39.07
Table-tennis	143	10.41	36.00	36.24	36.22
	346	13.92	32.15	36.78	36.07
	418	11.70	26.22	32.77	33.37

6 Conclusions

In this paper, we presented algorithms for concealment of distortions due to packet loss in video transmission. The proposed concealment algorithms, SAA and the combined concealment algorithm from Section 5 make use of the high spatial correlation between the region around the best-fit candidate MB-sized area from the previous frame and the area around the MB to be concealed. In our simulations, we have seen that this algorithm consistently gives better concealment and superior edge and structural alignment compared to the two-line search SMA algorithm we compare them to. These algorithms can thus be useful tools in a real-time streaming video scenario to conceal packet-loss related distortions.

References

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

[2] A. Perkis, "On the importance of error resilience in visual communications over noisy channels," *Circuits, Systems and Signal Processing; Special issue on Multimedia Communications* , vol. 20, no. 3, pp. 415–445, 2001.

[3] M. C. Hong, H. Schwab, L. P. Kondi, and A. K. Katsaggelos, "Error concelament algorithms for compressed video," *Signal Processing: Image Communication* , vol. 14, no. 6-8, pp. 473–492, May 1999.

[4] H. Sun, J. Zdepski, W. Kwok, and D. Raychaudhuri, "Error concelament algorithms for robust decoding of MPEG compressed video," *Signal Processing: Image Communcation* , vol. 10, no. 4, September 1997.

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

[6] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneaously received motion vectors," in *IEEE International Conference on Acoustics, Speech and Signal Processing* , Minneapolis, MN, USA, May 1993, pp. V–417 – V420.

[7] L. Atzori, F. G. B. De Natale, and C. Perra, "A spatio-temporal concelament techniue using boundary matching algorithm and mesh based warping (BMA-MBW)," *IEEE Transactions on Multimedia*, vol. 7, no. 3, pp. 560–563, June 1997.

[8] M. J. Chen, C. S. Chen, and M. C Chi, "Recursive block-matching principle for error concealment algorithm.," in *International Symposium on Circuits and Systems* , May 2003, pp. II–528 – II–531.

[9] E. T. Kim, S. J Choi, and H. M. Kim, "Weighted boundary matching algorithm for error concealment in the MPEG-2 video bitstream," *Signal Processing*, vol. 73, no. 3, pp. 291–295, January 1999.

[10] W. Y. Kung, C. S. Kim, and C. C. Jay Kuo, "A dynamic error concealment for video transmission over noisy channles," in *IEEE GLOBECOM2002*, Taipei, Taiwan, November 2002.

[11] J. Zhang, J. F. Arnold, and M. R. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Transactions on Circuits and Systems for Video Technology* , vol. 10, no. 4, pp. 659–665, June 2000.

[12] Z. Wang, A. C. Bovik, , H. R. Sheik, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on Image Processing* , vol. 13, no. 1, January 2004.

[13] K. R. Rao and J. J. Hwang, *Techniques and Standards for Image, Video and Audio Coding.* , Prentice-Hall, Englewood Cliffs, NJ, USA, 1996.

[14] J. F. Arnold, M. R. Frater, and J. Zhang, "Error resilience in MPEG-2 video coding standard for cell based networks-a review," *Signal Processing: Image Communication* , vol. 14, no. 6-8, pp. 607–633, May 1999.

[15] A. S. Bopardikar, O. I. Hillestad, and A. Perkis, "A temporal error concealment algorithm based on structural alignment for packet video.," in *International Conference on Signal Processing and Communications (SPCOM)* , Bangalore, India., December 2004.