AN ERROR CONCEALMENT ALGORITHM FOR ENTIRE FRAME LOSS IN VIDEO TRANSMISSION

Yu Chen[1*], *Keman Yu*[2], *Jiang Li*[2] *and Shipeng Li*[2]

[1]Department of Electronic Engineering, Tsinghua University
[2]Microsoft Research Asia

ABSTRACT

In video transmission over the Internet, the quality of service is usually not guaranteed. The loss of a packet may result in the loss of a whole video frame. While most of the current error concealment techniques can only deal with the loss of macroblocks, our bi-directional temporal error concealment method can recover a whole missing frame. For each pixel in the lost frame, the proposed algorithm extrapolates two motion vectors from the motion vectors of the previous reconstructed frame and the next frame. The lost pixel is then reconstructed using multi-hypothesis motion compensation. Experiments are carried with an H.263 video codec. The results show that our algorithm obviously outperforms the existing methods on both the PSNR and visual quality. Moreover, our method is efficient in stopping error propagation.

1. INTRODUCTION

With the rapid development of wired and wireless networks, more and more users are seeking video services, including video streaming and video conferencing over the Internet. However, the Internet does not provide guaranteed quality of service (QoS). Traffic congestion usually results in the loss of data packets. In wireless networks, packet losses happen frequently due to multi-path fading, shadowing and noise disturbance of wireless channels. In video transmission, the loss of a packet usually introduces severe distortions to the reconstructed results, because compressed video streams are very sensitive to transmission errors. As most of the current video coding schemes use the spatial-temporal predictive coding to achieve high compression efficiency, an erroneously recovered block may not only lead to errors in the subsequent blocks in the same frame, but also propagate errors to subsequent frames.

In order to control errors in video transmission, lots of error resilience video encoding techniques and decoder error concealment methods have been developed (see [1] for a detailed review). In general, decoder error concealment does not require any change to the encoder. It does not increase the bit rate, or the delay. It can be applied to almost any application, so it is preferable in low bit rate real-time video communications [3].

Temporal error concealment is one of the most important error concealment techniques. It utilizes temporal neighbors, that is, the previous frame or the next frame, to conceal the errors in the current frame. Most of the temporal error concealment methods assume that only a few macroblocks (MB) or slices in a video frame are lost. Typical methods include motion field interpolation (MFI) [2] and its extensions [3]. They estimate the lost motion vectors (MV) from correctly received neighbors in single or multiple reference frames. The method proposed in [4] uses correctly received information surrounding the lost MBs to estimate MVs using an affine transform.

Although temporal error concealment techniques have been extensively studied, they usually cannot handle the loss of a whole video frame. In low bit rate applications, in order to save transmission overhead, one data packet usually carries a whole frame. Consequently, the loss of a packet directly results in the loss of the whole frame. In high bit rate applications, traffic congestion usually leads to a burst of packet losses. If the majority of a frame is missed, the whole frame is hardly able to be recovered by the aforementioned methods.

To recover the missing frames, one straightforward method is to repeat the last received frame with all zero motion vectors, which is called temporal replacement (TR). TR works well at temporal stationary areas but fails at moving areas.

S. Belfiore et al. proposed a multi-frame motion vector averaging method (MMA) [5] that is capable of estimating the entire missing frame. The method exploits motion vectors of a few past frames to estimate the forward motion vector of each pixel in the last received frame, and then projects the last frame onto an estimate of the missing frame. This method usually produces a relatively high Peak Signal-to-Noise Ratio (PSNR) value, but it behaves even worse than TR does under a few conditions, as indicated by the experimental results in the paper. This phenomenon is possible because the farther reference frames usually no longer possess motion vectors similar to the missing frame. In addition, there is not an effective method to determine the number of reference frames.

Motion vector extrapolation (MVE) [6] is another method that can combat the loss of a whole frame. The method first extrapolates motion vectors of macroblocks from the last received frame, and then estimates the overlapped areas between the damaged block and the motion extrapolation macroblocks. It selects the best estimated motion vector, and conceals the corrupted block using general motion compensation. This method overcomes the disadvantage of incorrect macroblock displacement, but the rough motion vectors, of 8×8 pixel size, in the situation without residual information, usually cause block artifacts, especially in large motion scenes.

In this paper we propose a bi-directional temporal error concealment algorithm that can recover loss of a whole frame.

[*]The work presented in this paper was carried out at Microsoft Research Asia.

For each pixel in the lost frame, the proposed algorithm extrapolates two motion vectors from the motion vectors of the previous reconstructed frame and the next frame. It then reconstructs the lost pixel using multi-hypothesis motion compensation.

The rest of the paper is organized as follows. In Section 2 the proposed bi-directional temporal error concealment algorithm is described. Subjective and objective experimental results are shown in Section 3. Finally, we conclude the paper and give future directions in Section 4.

2. BI-DIRECTIONAL TEMPORAL ERROR CONCEALMENT

2.1. Pixel-based motion vector extrapolation

Temporal error concealment methods assume that motion in the video is smooth or continuous. The general approach is to replace the damaged block with the content of the previous frame at the motion-compensated location. The difficulty of this approach is that it relies on the knowledge of motion information that may not be available in all situations, especially when a whole frame is lost. Therefore, techniques to estimate the lost motion vectors have been widely discussed.

Both MVE and MMA contribute to accurately estimating the lost motion information. In MVE, motion vectors are extrapolated from the last received frame. For each 8×8 block in the current frame, its motion vector is determined by the extrapolated macroblock that possesses the largest overlapped area on it. In stationary and slow motion scenes, where the last received frame is highly temporally correlative to the current frame, MVE can yield relatively satisfactory results. However, due to the rough motion vectors, of 8×8 pixel size, and the missing of residual information, this method usually introduces obvious block artifacts. In particular, the method may give erroneously estimated motions in large motion scenes, as illustrated by an example in Figure 1.

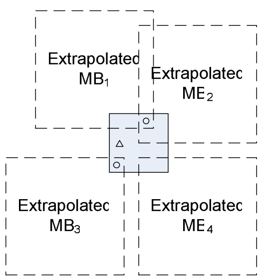


Figure 1: An example of MVE.

In Figure 1, the central square represents an 8×8 block that is lost. It overlaps with four macroblocks that are extrapolated from the last frame. Although the largest overlapped area belongs to MB_2, it is still a small portion of the whole damaged block. Therefore it is improper to use the motion vector of MB_2 to represent the motions of all pixels in the block. Actually, in this case, there is not an extrapolated macroblock that covers the majority of the damaged block. This situation appears frequently in large motion scenes. To achieve more smooth and fine results, pixel-based estimation method is desired.

MMA is a pixel-based temporal error concealment method. It starts from the last received frame. It inversely tracks the motion of each pixel in a few past frames, and then averages the motion vectors in the trace to estimate the forward motion vector of the last received frame. This method can smooth the boundaries of blocks but sometimes it behaves even worse that TR does.

In order to overcome the above shortcomings, we propose a pixel-based MVE method. It differs from block-based MVE in two aspects:

i) For a pixel that is covered by at least one extrapolated MB, such as the circle points in Figure 1, its motion vector is estimated by averaging the motion vectors of all overlapped MBs.

ii) For a pixel that is not covered by any of the extrapolated MBs, such as the triangle point in Figure 1, its motion vector is duplicated from the motion vector of the same pixel in the previous frame. It is worth mentioning that we find that other spatial interpolators, such as mean filter and median filter do not provide better performance, because it is highly probable that the neighbors surrounding the missing one are also empty.

In practice, we find that in stationary or little motion scenes, directly repeating the motion vectors of the previous frame works better than all aforementioned methods. We call a pixel that is covered by more than one extrapolated MBs a multi-covered pixel. In general, the scenes that possess large motion also have a large number of multi-covered pixels. Thus, we add the third rule into our algorithm.

iii) For a frame that has the number of multi-covered pixels smaller than a threshold *T*, its motion vectors are directly duplicated from the corresponding motion vectors of its previous frame. In practice, we find that 2,000 is a suitable value of *T* for QCIF size videos.

If the estimated motion vector is $f = (f_x, f_y)$, each lost pixel $p_f(x, y)$ can be recovered as follows:

$$p_f(x, y) = p_r(x + f_x, y + f_y) \tag{1}$$

where $p_r(x, y)$ refers to pixels in the previous frame.

Although the pixel-based MVE method provides similar performance as block-based MVE does in little motion scenes, it possesses obvious advantage in large motion scenes. The concealment performance of the pixel-based MVE and block-based MVE method is compared frame-by-frame in Figure 2. The value of each point is obtained when the corresponding frame is lost while all the other frames are correctly received.

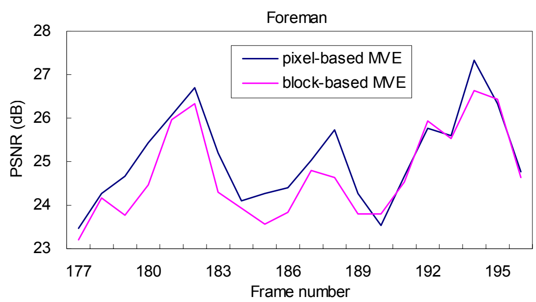


Figure 2: Concealment performance comparison between pixel-based MVE and block-based MVE at 15 fps and 128 kbps.

2.2. Backward estimation

Most of the temporal error concealment techniques utilize only the information of the past frames. Actually, the information of the next frame is usually also available, because we are not aware of the loss of the current frame until we receive the next frame.

Similar to the pixel-based MVE, the backward estimation method extrapolates motion vectors from the next frame. However, the pixel values of the next frame are not available until the current frame is recovered. We solve this problem by compensating each pixel $p_b(x, y)$ of the lost frame on the last reconstructed frame using the backward estimated motion vector (b_x, b_y).

$$p_b(x, y) = p_r(x + b_x, y + b_y) \quad (2)$$

where $p_r(x, y)$ refers to pixels in the last reconstructed frame.

The concealment performance of the forward and backward pixel-based MVE methods is shown in Table 1. The backward method works even better than the forward one.

Table 1: Concealment performance in terms of PSNR (dB)

	Forward	Backward
Foreman	29.19	29.49
Suzie	32.79	33.19
Miss_am	38.65	38.93

2.3. Bi-directional compensation

With the forward and backward methods, we can obtain two estimations of the current frame. It has been found that weighted averaging of multiple candidate concealments can achieve more robust performance [2] [3]. This solution is very similar to multi-hypothesis motion compensation (MHMC) [7].

We propose a more robust error concealment method by combining the forward and backward pixel-based MVE methods. For pixel (x, y) in the lost frame, its value $p(x, y)$ can be estimated as follows:

$$p(x, y) = w \times p_f(x, y) + (1 - w) \times p_b(x, y) \quad (3)$$

where the pixel-based weight $w = w(x, y)$ is used to adjust the weights of the forward and backward methods. In this paper, we set $w(x, y) = 0.5$, namely simply averaging two candidate concealments. A potential piece of work is to design a method that can adaptively adjust multi-hypothesis weights. For example, the weights can be adjusted in terms of the correlativity between the adjacent frames. More weights can be given to the candidate that possesses higher correlation to the lost frame.

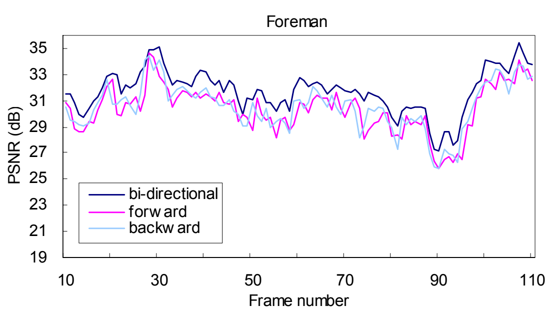


Figure 3: Concealment performance comparison for the QCIF Foreman sequence coded at 15 fps and 128 kbps.

The comparison of the concealment performance of the forward, backward and bi-directional methods is shown in Figure 3. We can see that the bi-directional method is almost consistently better than the other two.

3. EXPERIMENTAL RESULTS

We applied the proposed bi-directional method to an H.263 (TMN5) video codec to examine its effectiveness on concealing the loss of a whole frame and on stopping error propagation. Certainly, this method can be applied to most of the block-based hybrid video coding schemes.

Here, we show the results of three sequences, Miss_am, Suzie and Foreman, which represent small, moderate and large motion scenes respectively. All sequences are in a QCIF size. The frame rate is set to 10 fps. The bit rate is set to 56 kbps for Miss_am and Suzie sequences and 128 kbps for Foreman sequence. Only the first frame is encoded as an I-frame and all the others are encoded as P-frames.

We compare the proposed algorithm with MMA, MVE and motion compensation (MC). In MC, it is assumed that the original motion vectors are correctly received while all the residual information is lost. In the comparison of the performance of temporal error concealment techniques, MC could be regarded as the “upper bound”. In MMA, we find that it is suitable to use only one past frame as reference (called as MMA1) for the Foreman and Suzie sequences, while it is suitable to use two past frames (called as MMA2) for the Miss_am sequence.

As shown in Figure 4 and Table 2, the proposed method is very effective for those scenes that possess large motions, such as Foreman sequence. It outperforms MMA and MVE with 1.85 dB and 1.32 dB on average respectively. On the other hand, for those scenes that have moderate or small motions, such as Suzie and Miss_am sequences, the proposed method still achieves about 0.5 to 0.8 dB improvements to MMA and MVE. Moreover, the PSNR degradation of the proposed method is less than 0.61 dB comparing to the “upper bound”. In summary, the proposed method can provide relatively satisfactory performance in different scenarios with low and high motion activities.

For subjective evaluation, one error-free frame and three frames recovered by MVE, MMA and the proposed algorithm are shown in Figure 5. Image (b) that is recovered by MVE has obvious block artifacts in the face area, while image (c) that is recovered by MMA is smooth in the face area but possesses severe distortions in the background. Obviously, the proposed algorithm provides a more vivid presentation.

Figure 6 illustrates the recovery of PSNR after a loss occurs at the 183rd frame of the Foreman sequence. We can see that the proposed algorithm is more efficient in stopping error propagation than MVE and MMA.

4. CONCLUSIONS

In this paper, we present a bi-directional temporal error concealment algorithm that can recover the loss of a whole frame.

The proposed method exploits the information of both the last reconstructed frame and the next frame to estimate pixel-based motion vectors of the lost frame, and then reconstructs the lost frame using multi-hypothesis motion compensation. We applied the method to an H.263 video codec. Experimental results show that our approach obviously outperforms the existing methods on both PSNR and visual quality. Moreover, it is efficient in stopping error propagation.

Future directions may include designing a method that can adaptively adjust multi-hypothesis weights.

5. REFERENCES

[1] Y. Wang, S. Wenger, J. Wen and A. Katsaggelos, "Error Resilient Video Coding Techniques," *IEEE Signal Processing Magazine,* pp. 61-82, July 2000.

[2] M.E. Al-Mualla, N. Canagarajah and D. R. Bull, "Temporal error concealment using motion field interpolation," *IEE Electronic Letters,* pp. 215-217, Feb. 1999.

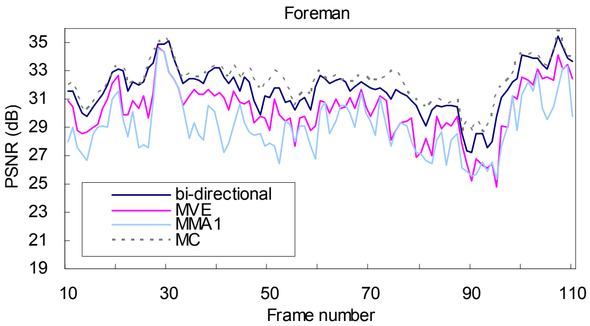
[3] M.E. Al-Mualla, N. Canagarajah and D.R. Bull, "Multiple-reference temporal error concealment," *Proc. IEEE ISCAS 2001*, vol. 5, pp. 149-152.

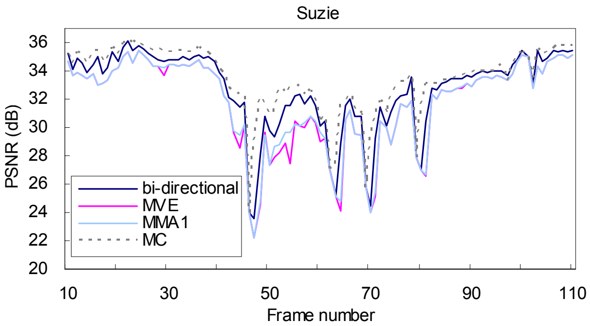
[4] S.H. Lee, D.H. Choi and C.S. Hwang, "Error concealment using affine transform for H.263 coded video transmissions," *Electronics Letters*, 37(4), pp. 218-220, Feb. 2001.

[5] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "An Error Concealment Algorithm for Streaming Video," *Proc. IEEE ICIP 2003*.

[6] Q. Peng, T.W. Yang and C.Q. Zhu, **"**Block-based temporal error concealment for video packet using motion vector extrapolation," *Proc. IEEE Communications, Circuits and Systems and West Sino Expositions*, 2002.

[7] Girod. B, "Efficiency analysis of multi-hypothesis motion-compensated prediction for video coding," *IEEE Transactions on Image Processing,* vol. 9, no. 2, pp.173 – 183, Feb. 2000.





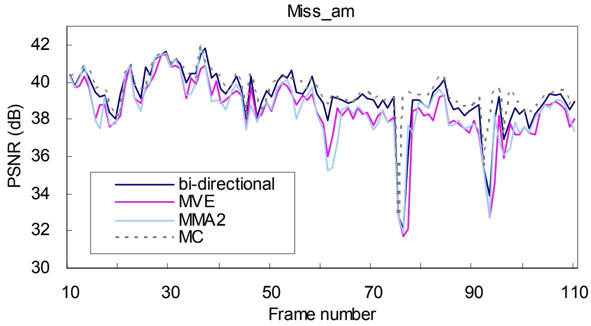


Figure 4: Comparative evaluation of the error concealment methods when each frame is lost.

(a) Error free (33.39 dB) (b) MVE (25.25 dB)

(c) MMA (24.07 dB) (d) Bi-directional (27.27 dB)

Figure 5: Subjective quality of the 180th frame of Foreman

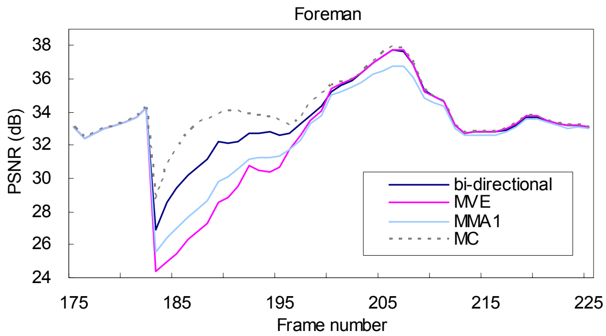


Figure 6: The recovery of PSNR after a lost frame.

Table 2: PSNR comparison on the entire sequences (dB).

	MMA	MVE	Proposed	MC
Foreman	28.54	29.07	30.39	31.01
Suzie	32.79	32.73	33.56	34.17
Miss_am	38.73	38.64	39.28	39.54