A Highly Effective Error Concealment Method for Whole Frame Loss

Chen Zhao[1], Siwei Ma[1],

[1]Institute of Digital Media
Peking University
Beijing, China
{czhao, swma, wgao}@jdl.ac.cn

Jian Zhang[2], Wen Gao[1]

[2]Department of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
zhangjian@jdl.ac.cn

Abstract—**When videos are transmitted over the Internet, packet missing is inevitable and an entire frame may get lost. However, most of the literature on error concealment problems can only deal with block loss. For the case of frame loss, they usually fail to achieve satisfactory results. In this paper, we propose a highly effective frame concealment method, which gives amazingly good recovery quality. We base on the theory of multiple hypotheses and devise an adaptive integration scheme to make full use of each hypothesis' strength. Different from the existing methods which mostly rely on motion vectors of previous frames, we fully exploit the correlation between consecutive frames. A novel idea for generating multiple estimates of the lost frame is adopted. Experimental results demonstrate that the proposed algorithm significantly outperforms the state-of-the-art error concealment methods for whole frame loss in both subjective and objective quality.**

I. INTRODUCTION

Owing to the fast development of digital technology, the demand for video transmission is increasing rapidly. However, the communication channel is not yet reliable and packet loss may frequently occur. Plus the temporal predictive coding used in most coding standards, errors may propagate easily to succeeding frames. When this happens, the decoded video sequences will be severely distorted, which is intolerable for the users.

In order to alleviate the corruption of video frames and control the propagation of errors, error concealment (EC) techniques are developed to minimize the distortion at the decoder side. So far, many works [1-10] have been proposed. However, most of these methods assume that only a few blocks in a video frame are missing, and they are incompetent when a whole frame is encapsulated into a single packet and all its data are lost during transmission.

One straightforward method to recover a lost frame is to simply repeat the earlier decoded frame (frame copy or FC) [15]. Another method called motion copy (MC) [15] uses the motion vectors (MVs) of the previous frame and restores the current missing frame through motion compensation. Both methods are simple but could only work well for low-motion areas.

Some other algorithms seek to generate a better motion vector field (MVF), like block-based motion vector extrapolation (BMVE) [12]. BMVE determines the best estimated MV of each corrupted block according to its overlapped areas with the motion extrapolated macroblocks. [13] puts forward a similar method, which extends BMVE to the pixel level (PMVE). Merging the two schemes, [14] proposes a hybrid motion vector extrapolation (HMVE) method. Although these methods may work not bad in some scenarios, holes and overlaps areas, which may not match well with MVs of previous frames, would seriously affect the global video quality.

All of the algorithms above resort to MVs of previous frames, which could only provide limited information for frame recovery. In this paper, other than the MV information, we make full use of the correlations between consecutive frames to build multiple more reasonable estimates of the lost frame. The frame concealment algorithm is on the basis of the theory of multiple hypotheses [18, 19] and proposes a novel method to adaptively integrate these estimates.

The rest of the paper is organized as follows. In Section II, we give the details of the proposed highly effective error concealment algorithm. Extensive experimental results are reported in Section III, and conclusions are drawn in Section IV.

II. MULTI-HYPOTHESIS BASED FRAME CONCEALMENT

Based on the multi-hypothesis theory, we build multiple estimated pictures for the lost frame. This makes sense in that one estimate may well reflect the true data of the lost frame in some areas and may disastrously distort it in other areas. Adaptive integration helps enhance the influence of accurately estimated areas and attenuate that of wrongly estimated areas. The process could be mathematically formulated as

$$\hat{f}_n = \sum_k w_k \odot f_n^k \qquad (1)$$

where $\hat{f}_n$ represents the reconstructed estimate of the missing frame f_n, and $f_n^k, k = 1, 2 \cdots S$ is one estimated frame, in which S refers to the total number of estimates. w_k is a weight for each hypothesis, characterizing the reliability of the estimate f_n^k, and $\odot$ stands for the element-wise product of two vectors. Here w_k adaptively adjusts the contribution of each estimate to the final recovery of the missing frame f_n. Remodeling the multi-hypothesis framework on a pixel-by-pixel basis, we get

$$\hat{f}_n(x, y) = \sum_k w_k(x, y) * f_n^k(x, y) \qquad (2)$$

where $\hat{f}_n(x, y)$ and $f_n^k(x, y)$ refer to pixels at the location of (x, y) in the reconstructed frame and the kth estimated frame, respectively.

According to (2), in order to reconstruct the frame f_n, we need to develop an approach to achieve several predictions of f_n and to design an appropriate weight model to well exhibit the reliability of each prediction. In the following sections, our solutions to the two aspects will be presented: motion vector field derivation and adaptive weight model.

A. Motion Vector Field Derivation

To better estimate the lost frame, the correlations between consecutive frames are taken into consideration. We assume that objects move in a constant speed along a straight line. Thus, as depicted in Fig. 1, block B_n in the missing frame f_n and its referenced block B_{n-1} in frame f_{n-1} share a motion vector, which refers B_{n-1} to B_{n-2} in frame f_{n-2} as well. The better B_{n-2} matches B_{n-1}, the more accurate the motion vector should be. The matching degree is measured by the sum of absolute differences (SAD) between corresponding pixels in the pair of blocks B_{n-2} and B_{n-1}, which is denoted by D_i in the following equation.

$$mv = \arg \min_{(mv_{ix}, mv_{iy})} D_i \qquad (3)$$

Here, (mv_{ix}, mv_{iy}) represents all the candidate motion vectors within the searching window and mv is the estimated optimal motion vector for block B_n. Hence, the optimal motion vector could be chosen according to (3).

In this way, the motion vector for block B_n is estimated, and covering all the blocks in the missing frame f_n, one MVF could be generated. In order to obtain multiple MVFs, we repeat the process above using various block sizes. We name one repetition a step. Specifically, 32×32, 16×16, 8×8, 4×4, 2×2 to 1×1 block sizes are adopted and six MVFs are accordingly derived (Note that the 1×1 block is down to the pixel level).
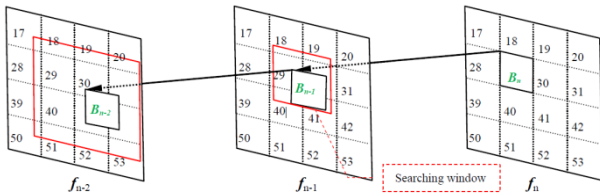

Figure 1.   Optimal Motion Vector Determination

To ensure the reliability of each estimate, we regulate the initial motion vectors for each step, which indicate the centers of the searching windows. For the kth step, the initial motion vector mvp^k of block B_n is defined as the median value among the motion vectors of its neighboring blocks and the motion vector of B_n obtained in the k-1th step. This can be formulated as follows

$$mvp^k = \mathrm{median}\left\{mv^{k-1}, mv_{topleft}^k, mv_{top}^k, mv_{topright}^k, mv_{left}^k\right\} \quad (4)$$

where mv^{k-1} refers to the motion vector estimated from the k-1th step, if the current step is not the first. And for the first step of MVF derivation, which is performed with block size 32×32, mv^{k-1} is the motion vector of the co-located block in the previous frame f_{n-1}. The other four motion vectors denote the MVs of the top-left, top, top-right and left adjacent blocks achieved in the current step.

With the estimated six MVFs, six hypotheses of the lost frame f_n could be acquired through motion compensated prediction, which is shown below.

$$f_n^k(x, y) = f_{n-1}(x + mv_x^k, y + mv_y^k) \qquad (5)$$

where the referenced pixel of $f_n^k(x, y)$ in the previous frame f_{n-1} is denoted by $f_{n-1}(x + mv_x^k, y + mv_y^k)$, in which (mv_x^k, mv_y^k) is the motion vector achieved in the kth step of motion estimation.

B. Adaptive Weight Model

Considering that the weight is responsible of trading off the devotion of each hypothesis to the reconstructed frame, it should reflect the reliability of the hypothesis. At a pixel level, this means that the more reliable the estimated pixel value $f_n^k(x, y)$ is, the larger its weight $w_k(x, y)$ should be. Intuitively, its reliability could be measured by the difference between the actual value of the pixel $f_n(x, y)$ in frame f_n and the estimated value $f_n^k(x, y)$. However, the frame f_n is already lost so that in no way could the real value of $f_n(x, y)$ be obtained. Helpfully, the preceding frames f_{n-2} and f_{n-1} are available and differences between the corresponding pixels in the two frames could be calculated as a substitute.

Therefore, we assume that the reliability of the estimated pixel value $f_n^k(x, y)$ could be approximately modeled by the variation between the referenced pixels $f_{n-2}(x + 2mv_x^k, y + 2mv_y^k)$ and $f_{n-1}(x + mv_x^k, y + mv_y^k)$ in frames f_{n-2} and f_{n-1} along the motion trajectory. Then we have

$$r^k(x, y) = \frac{1}{\left| f_{n-2}(x + 2mv_x^k, y + 2mv_y^k) - f_{n-1}(x + mv_x^k, y + mv_y^k) \right|} \qquad (6)$$

in which $r^k(x, y)$ represents the reliability of the estimated pixel value $f_n^k(x, y)$.

The model above simply considers the difference between two referenced pixels, which may lead to improper approximation when there is a fluctuation for either of the two

values. Hence, we extend the pair of referenced pixels into two patches centered on them, and by modifying (6) we get a new model as following

$$r^k(x, y) = \frac{1}{S^k(x, y)} \qquad (7)$$

where

$$S^k(x, y) = \sum_{-M \leq j_x, j_y \leq M} \left| f_{n-2}(x + 2mv_x^k + j_x, y + 2mv_y^k + j_y) \right. \qquad (8)$$
$$\left. - f_{n-1}(x + mv_x^k + j_x, y + mv_y^k + j_y) \right|$$

In the equation, j_x and j_y refer to the offset of the pixels in the referenced patches and M is the size of the patches.

Now by normalizing the reliability, we can get the expression of the weight $w_k(x, y)$ for the estimated pixel value $f_n^k(x, y)$.

$$w_k(x, y) = \Psi\left(r^k(x, y)\right) \qquad (9)$$

where Ψ represents the normalization factor, which makes the weight $w_k(x, y)$ range from 0 to 1.

III. EXPERIMENTAL RESULTS

The proposed algorithm is simulated on the H.264/AVC reference software JM10.0. Due to the platform-independent feature of error concealment, the proposed method can be easily transplanted to other versions of the codec and achieve an equally good performance. Four test sequences "Bus", "Mobile", "Coastguard" and "Tempete" are chosen to evaluate the performance of the algorithm. All of them are in CIF size with the frame rate of 30 frames per second. The period of I frame reset is 15 and the number of reference frames is 1. A constant QP of 24 is maintained for all the frames.

We compare the performance of the proposed algorithm with that of BMVE, PMVE and HMVE. In this simulation, one frame is dropped in each group of pictures (GOP) and the dropped frames will be concealed with the four methods. We use the peak signal-to-noise ratio (PSNR) as the objective measurement, which is computed using the original uncompressed video as reference. The PSNR values of each method for the four sequences are plotted in Fig. 2. It is obvious that the curves of the proposed algorithm are above the other ones in all the cases.

For clearer comparison, Table 1 presents the average PSNR performances over all the erroneous frames. As shown in this table as well as in Fig 2, the proposed algorithm yields higher PSNR performances than all the other methods. For sequences with high motion like "Bus", the proposed algorithm is able to provide up to 6.10 dB, 3.06 dB and 1.43 dB better PSNR performances than BMVE, PMVE and HMVE respectively. For the worst case, for sequences that have low motion like "Coastguard", it still outperforms BMVE, PMVE and HMVE by 2.28 dB, 0.67 dB and 0.50 dB.
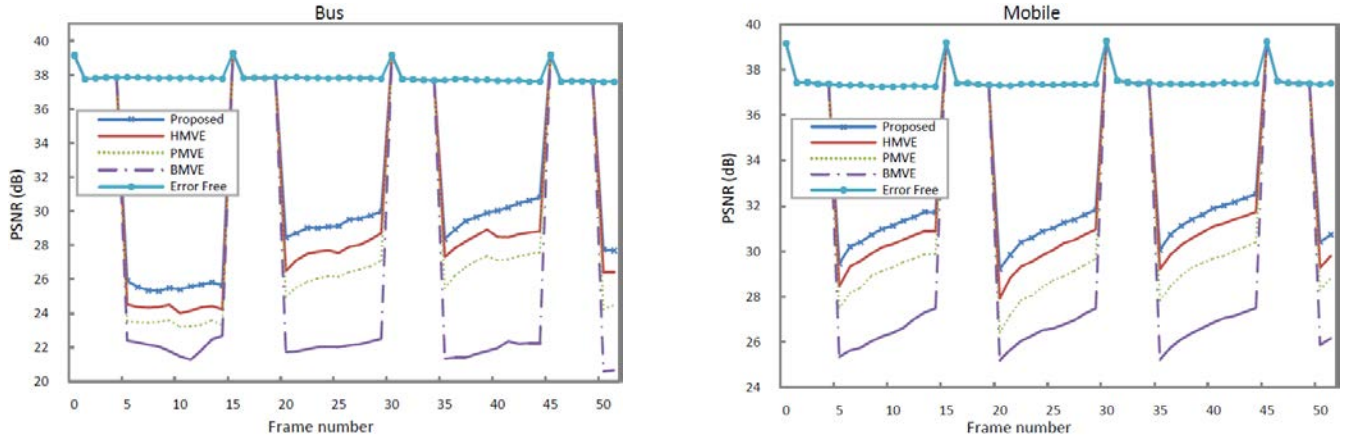

Figure 2. PSNR comparison versus frame number for the two sequences: for each sequence, four frames (the 5th, 20th, 35th, 50th) are dropped and concealed with the four methods (Notice that the sequence starts with the 0th frame, not the 1st frame as used in reference [14])

TABLE I.   COMPARISON OF THE AVERAGE PSNR PERFORMANCE

Sequence	PSNR (dB)				Gain (dB)		
	BMVE [12]	PMVE [13]	HMVE [14]	Proposed	Over BMVE	Over PMVE	Over HMVE
Bus	21.51	24.55	26.18	**27.61**	6.10	3.06	1.43
Mobile	25.41	27.50	28.72	**29.79**	4.38	2.29	1.07
Coastguard	29.21	30.82	30.99	**31.49**	2.28	0.67	0.50
Tempete	27.23	26.56	27.17	**27.88**	0.65	1.32	0.71
Average	25.84	27.36	28.27	**29.19**	3.35	1.83	0.92

For subjective evaluation, one error-free frame and four recovered frames reconstructed by BMVE, PMVE, HMVE and the proposed algorithm are demonstrated in Fig. 3. The 35th frame of "Bus" is chosen. We can observe that the proposed algorithm almost perfectly recovered the fences area, keeping the balusters straight and undivided, compared with BMVE and PMVE. While HMVE destroys the smoothness of the bus top, our method has well preserved it.
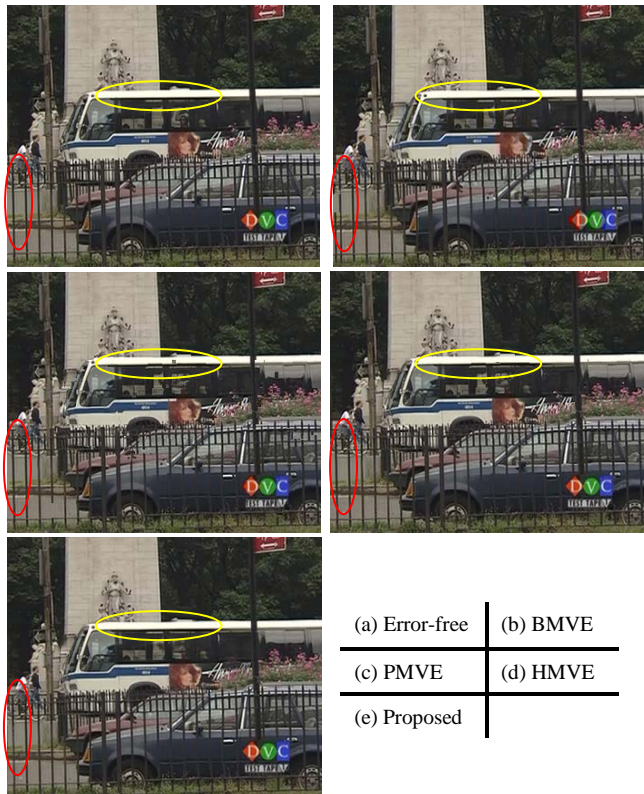

(a) Error-free	(b) BMVE
(c) PMVE	(d) HMVE
(e) Proposed	

Figure 3.   Reconstructed 35th frames of "Bus" CIF sequence.

IV.   CONCLUSIONS

In this paper, we propose a novel frame concealment algorithm, which can achieve superior recovery results to the state-of-art methods. It reconstructs the lost frame via an adaptive integration of its multiple hypotheses. With a discreetly designed weight model, the advantage of each hypothesis is well taken. Besides, the correlations between neighboring frames are observed, so that more information is employed to reconstruct an estimate of the lost frame. The experimental results demonstrate that the proposed algorithm has better performance than the existing methods both on PSNR and visual quality.

As for the time complexity, although multiple motion estimations is a bit time-consuming, it can be easily mitigated by means of parallel processing, such as SIMD implementation and multi-thread programming.

REFERENCES

[1] W. Kumwilaisak and C. J. Kuo, "Spatial error concealment with sequence-aligned texture modeling and adaptive directional recovery," *Journal of Visual Communication and Image Representation*, vol. 22, no. 2, pp. 164–177, Feb. 2011.

[2] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical mpeg-2 video codec," in *Proc. Int. Conf. Commun.(ICC)*, pp. 1778–1783, 1995.

[3] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projection onto convex set," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 470–477, Apr. 1995.

[4] R. F. Zhang, Y. H. Zhou and X. D. Huang, "Content-adaptive spatial error concealment for video communication," *IEEE Trans. Consumer Electron.*, vol. 50, no. 1, pp. 335–341, Feb. 2004.

[5] G. Zhai, J. Cai, W. Lin, X. Yang and W. Zhang, "Image Error-concealment via Block-based Bilateral Filtering," in *Proc. IEEE Int. Conf. Multimedia and Expo (ICME)*, pp. 621–624, 2008.

[6] Y. B. Zhang, X. G. Xiang, S. W. Ma, D. B. Zhao, and W. Gao, "Auto Regressive Model and Weighted Least Squares based Packet Video Error Concealment, " in *Proc. IEEE Data Compression Conference (DCC)*, pp. 455-464, 2010.

[7] D. Nguyen, M. Dao, and Trac Tran, "Video Error Concealment using Sparse Recovery and Local Dictionaries," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, pp. 1125-1128, 2011.

[8] J. Zhou, B. Yan, and H. Gharavi, "Efficient Motion Vector Interpolation for Error Concealment of H.264/AVC," *IEEE Trans. Broadcasting*, vol. 57, no. 1, pp. 75-80, 2011.

[9] M. E. Al-Mualla, C. N. Canagarajah, and D. R .Bull, "Multiple-reference temporal error concealment," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, pp. 149-152, 2001.

[10] W.-Y. Kunag, C.-S. Kim, and C.-C. J. Kuo, "A dynamic error concealment for video transmission over noisy channels," in *Proc. IEEE Global Telecommunications Conference*, pp. 1769-1773, 2002.

[11] H. Liu, R. Xiong, S. Ma, D. Zhao and W. Gao, "Bayesian frame interpolation by fusing multiple motion-compensated prediction frames," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 1197–1200, 2011.

[12] Q. Peng, T. Yang and C. Zhu, "Block-based temporal error concealment for video packet using motion vector extrapolation," in *Proc. IEEE Int. Conf. Communications, Circuits and Syst. and West Sino Exp.*, pp. 10–14, 2002.

[13] Y. Chen, K. Yu, J. Li and S. Li, "An error concealment algorithm for entire frame loss in video transmission," in *Proc. IEEE Picture Coding Symp. (PCS)*, 2004.

[14] B. Yan and H. Gharavi, "A hybrid frame concealment algorithm for H.264/AVC," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 98–107, Jan. 2010.

[15] Z. Wu and J. M. Boyce, "An Error Concealment Scheme for Entire Frame Losses Based on H.264," in *Proc. IEEE Int. Symp. Circuits and Syst. (ISCAS)*, 2006.

[16] J. Chien, G. Li and M. Chen, "Effective Error Concealment Algorithm of Whole Frame Loss for H.264 Video Coding Standard by Recursive Motion Vector Refinement," *IEEE Trans. Consumer Electronics.*, vol. 56, no. 3, pp. 1689–1695, Aug. 2010.

[17] S.K. Bandyopadhyay, Z.Y. Wu, P.Pandit, and J. Boyce, "Frame loss error concealment for H.264/AVC," Joint Video Team of ISO/IEC MPEG &ITU-T VCEG, JVT-P072, Jul. 2005.

[18] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 173-183, Feb. 2000.

[19] T. Wiegand, X. Zhang, and B. Girod, "Long term memory motion–compensated prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 70-84, Jan. 1999.