EFFICIENT ERROR RECOVERY FOR MULTIPLE DESCRIPTION VIDEO CODING

Guanjun Zhang and Robert L. Stevenson

Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556, USA
Email: {gzhang, rls}@nd.edu

ABSTRACT

Multiple Description Coding (MDC) and Multihypothesis Motion-Compensated Prediction(MHMCP) are two promising error resilient video coding techniques. We propose an enhanced error resilient video coding scheme based on Multiple State Video Coding(MSVC). It uses multihypothesis motion prediction in addition to normal motion-compensated video compression steps. Yielded additional motion vectors between frames of two descriptions can help efficiently and quickly estimate lost content in one description when the other one is received correctly. Negligible redundancy is added to originally coded MSVC bitstreams while complicated motion search is avoided at decoder when there is a need to estimate lost content. Simulation results show that the proposed scheme has better error recovery performance than previously used approaches in most situations, especially when multiple and complex motions appear.

1. INTRODUCTION

Stopping or controlling error propagation due to motion compensation is a desirable feature of compressed video streams over error-prone networks. Numerous source coding approaches have been reported and adopted in international standards. A common idea is to diversify the source of motion references, which greatly limits the effect of error propagation to the following pictures. Two successful examples are MDC and MHMCP.

MHMCP is the generalization of the B-frame structure widely used in video coding. It improves coding gain [1] with increased number of hypothesis. Its error resilient effect has also been studied [2]. An optimal predictor will find a balance between coding gain and error resilience capability.

MDC provides intrinsic robustness and significant performance improvement over conventional error protections in lossy packet networks when a back channel is not available or retransmission delay is intolerable. MDC matches heterogeneous network constructions by taking advantage of the transmission diversity of networks [3]. Its basic assumption is that the possibility of simultaneous errors in multiple channels is smaller than that in one channel. Multiple descriptions can mutually be enhanced when correctly received, or efficiently recovered from erasure errors: the more descriptions received, the higher the quality.

Various MD approaches have been applied to generate equally important video descriptions from one input source. We can roughly group them into two catalogs: spatial approaches and temporal approaches, according to how final reconstruction is enhanced by the received descriptions. Spatial domain methods use information from different descriptions in the same time slot. These include Multiple Description Scalar Quantization (MDSQ) [4], overcomplete transforms such as the Multiple Description Transform Coding (MDTC) [5], and polyphase spatial down-sampling multiple description coding [6]. In temporal enhancement MDC approaches, reconstruction of current frames is improved by information in adjacent frames from different descriptions. This type approaches include Multiple Description Motion Compensation (MDMC) [7] and MSVC [8]. In both MDMC and MSVC, an input sequence is grouped and coded into an odd frame subsequence and an even frame subsequence. In MSVC, errors affect frames in one subsequence, while in MDMC, errors are strictly constrained in the local frames with the price of higher redundancy.

In MSVC, the assumed situation is loss of whole frames, and several error concealment approaches were used to address this problem. Erasure of a whole frame can be recovered from adjacent two, three or even more correctly received frames in both directions. Search of a dense motion field from the closest (say, an odd frame lost) even frames is applied to estimate the motions of every pixel, and then the lost frame is interpolated. It yields more accurate estimations than other simple algorithms when multiple motions appear [9]. But in most application scenarios, such as real time video applications, decoders can not afford the time or computation for complex motion search. Further more, the applied motion estimation at decoders is not efficient for complex motions such as rotation and zooming [10].

0-7803-8554-3/04/$20.00 ©2004 IEEE.

We argue that temporal correlation between two descriptions in MSVC is not fully extracted to promote error concealment. In this paper, we present an improved MSVC scheme featured with fast and efficient lost frame estimation. It involves multihypothesis motion prediction at the encoder and simple reconstruction at the decoder with small amount of additional block motion information. In section 2, we will introduce and discuss the proposed coding scheme. Simulation results will be presented and evaluated in section 3. Conclusions are given in section 4.

2. PROPOSED CODING SCHEME

In MSVC, an original input sequence is simply divided into one odd frame subsequence and one even frame sequence. The two subsequences are independently compressed using conventional hybrid motion-compensated video coding approach, as shown in Figure 1(a). The two descriptions enhance temporal resolution of final decoded streams. When erasure errors happen to one frame, it can be approximated from nearby correctly decoded frames. As shown in Figure 1(b), for example, if P_3 is lost, the temporally nearby even frames P_0, P_2, and P_4 can be used to estimate it. Estimation algorithms without motion search are simple but have poor performance, such as replacing the lost odd frame with previous odd frame. Complicated motion search is time and computation consuming.

The error recovery approaches used in MSVC rely on accurate motion information to estimate lost frames, especially when there are multiple motions. If we can extract proper motion information at the encoder and transmit it along with coded bitstreams, we can avoid complex motion search at the decoder while yielding the same or even more accurate prediction.

Figure 1(c) explains the proposed coding scheme. The input sequence is set apart as odd and even subsequences as in MSVC. Take the odd subsequence for example. In addition to the normal encoding processing for an odd frame to generate the bitstream, motion prediction is applied to its nearby even frames. For each block, we use more than one additional motion vectors from different reference frames, respectively. Multiple reference prediction has the advantage of more accurate prediction, and is more suitable for error concealment. The additional motion information of the odd frame is transmitted along with the even reference frames. Since two descriptions are independently decodable, when erasure errors happen to the odd frames, the decoder can use the corresponding motion information in the even description to quickly estimate the lost content. For the whole sequence, we apply the same process to every frame of two subsequences and embed these motion vectors into corresponding reference frames. A simple version of the scheme using only two nearest reference pictures is shown in Figure 1(d). Thus every coded frame in one description contains additional motion vectors of two frames in the other stream.

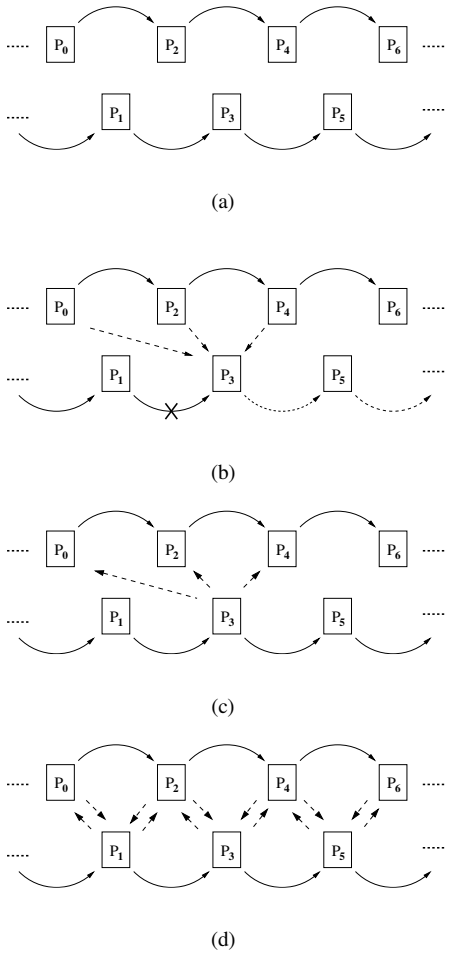


Fig. 1. a) Original MSVC encoding; b) error recovery of MSVC decoding; c) proposed encoding in MSVC for one frame; d) proposed simple encoding for two subsequences.

These additional motion vectors can be computed for each pixel or each block. The more dense the motion field is, the more accurate the approximation is, and more redundancy is introduced. In this paper, we use 16×16 size blocks. For every block, motion vectors are searched using multiple reference frames. More accurate prediction will be yielded if more hypothesis are used, and again, more redundancy to the bitstreams. Linear combination of predictions in each reference frame is used as the prediction of the block. If we assume that all frames in a sequence have the same size, and every frame use the same block size, then no block position information is needed for the decoder. We can put all these motion vectors at the end of a frame and organize them in a fixed order. After predictive coding

and proper entropy coding, the additional motion information will add little burden to the original MSVC bitstreams. Since the number of motion vectors for each frame is constant in the same video, the higher the bitrate and quality, the lower the redundancy ratio. Thus, for most high quality video applications, the added bitrate is negligible.

If part of or whole frame is lost, the decoder fills desired areas with predictions from its reference frames using corresponding motion vectors. Although no prediction errors are transmitted, the proposed scheme can yield better effect than most straight forward postprocessing error recover approaches because of highly efficient multihypothesis motion-prediction. The motion information is not used when decoded frame is error free. How to make use of the motion information when there is no error can be a future research topic.

3. SIMULATION RESULTS

Our algorithm is implemented on an MPEG-2 compatible codec. All test sequences are coded at a rate of 25 f/s, yielding 12.5 f/s for each description. For simplicity, only one previous and one future frame are used as additional references, as in Figure 1(d). Bidirectional motion search is applied to every 16×16 macroblock in a frame. Only I and P type frames are used for an appropriate comparison with the original methods in MSVC. We set the encoder to generate constant bitrate streams, which makes compressed video quality vary slightly due to picture content. It does not affect the purpose and conclusions of the test.

The proposed method is compared to several estimation approaches previously used in MSVC [9]. Efficiency of each scheme is tested on sequence football and garden (both SIF format, 352×240), respectively. The sequence football contains complex and combined motions, and the sequence garden has large global motions. For simplicity, we assume that the odd sequence lost one whole frame when errors happen, while all even frames are correctly decoded. This is equal to an 80 ms duration burst error or 3 to 4 packet loss. Decoders use different approaches to estimate the lost odd frames from both even frames around it. In our test results, the proposed method is marked as "MVpred". Recovery by averaging previous and future even frames is marked as "average even". "inplaceMC" represents the approach that the decoder motion-compensates the lost frame using prediction errors in the future even frame and replace the motion vectors by $1/2$ of those between previous and future even frames. "MCinterp" indicates a motion-compensated interpolations in which the decoder predicts a dense motion field for every pixel using a basic version of [10] through two nearest even frames. Correctly decoded odd frames are also presented as "original decoded" to indicate how good the reconstructed picture quality is.

Each test sequences is first encoded at the same bitrate and same quality: $15.84KB/f$ for every P-frame in football; $10.56KB/f$ for every P-frame in garden. In the proposed scheme "MVpred", additional motion vectors are compressed using predictive coding, run-length coding, and arithmetic coding, and the bits are added to the end of each frame. The average number of additional bits is around $450Bytes/f$ in football and $275Bytes/f$ in garden. This results in approximately 3% bitrate increment in both test sequences, compared to the original MSVC coded video.

Figure 2 and 3 demonstrate the efficiency of the above recovery methods for every odd frames of sequences football and garden, respectively. In football, the average PSNR of MVpred is 29.3dB, which is 3.1dB higher than MCinterp, and the curve of MVpred is also flatter than the curves of the other three methods. In garden, this difference is 1.6dB. Averaging nearest even frames only yields good results when there is little motion in the scene, such as the frames from No.50 to No.70 in football. The proposed MVpred produces the best estimation at most time instance thus is less influenced by channel loss.

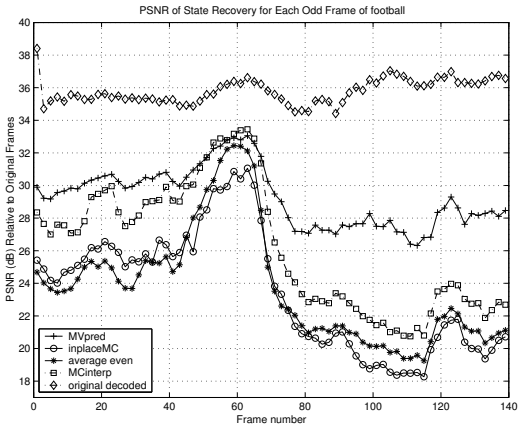


Fig. 2. Error recovery for odd frames in sequence football.

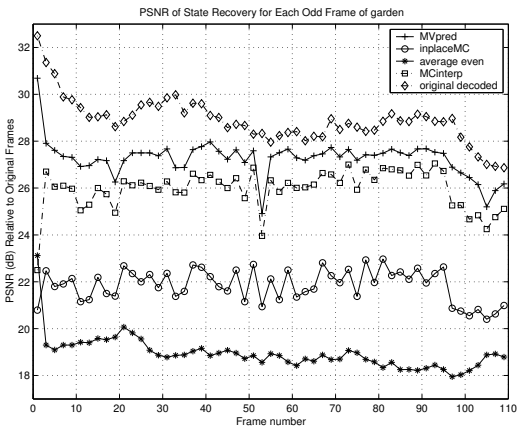
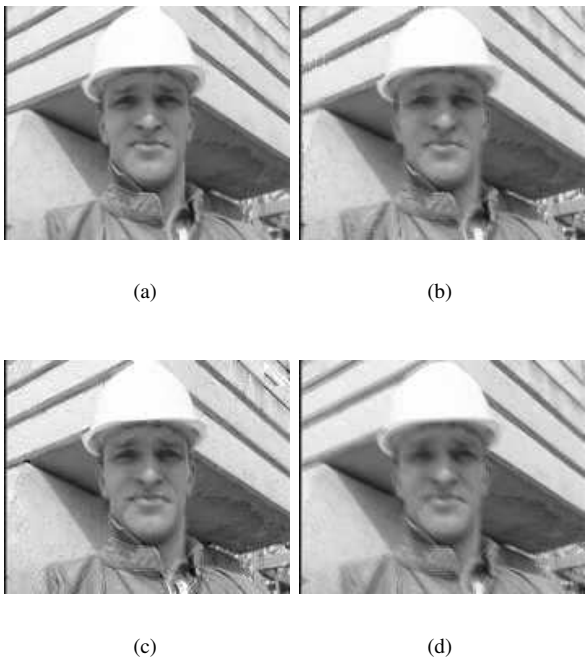


Fig. 3. Error recovery for odd frames in sequence garden.

MVpred only needs one plus and two bit shifting operations per pixel, while MCinterp needs much more computation in the decoder. For every block, MCinterp detects three motion candidates by using a 2D cosine window, FFT, normalization of frequency domain coefficients, and IFFT. The motion vector of every pixel is compared and selected from the candidates. Then temporal and spatial interpolations are applied pixel by pixel. The required operations of the whole processing of MCinterp is over 30 times more than that of MVpred.

Visual effect of different methods is shown in Figure 4. The sample pictures are from the fifth frame of sequence foreman (QCIF format, 176×144) predicted by different methods. Our proposed prediction method still yields the best result most of time. The additional bitrate for foreman is also around 3% ($110Bytes/f$ compared to $3.17KB/f$).



(a) (b)

(c) (d)

Fig. 4. Lost frame (No.5 in foreman) recovered by a) MVpred, 35.1dB; b) MCinterp, 34.4dB; c) inplaceMC, 29.9dB ; d) average even, 30.1dB.

4. CONCLUSION

We present an improved error resilience multiple description video coding scheme based on Multiple State Video Coding. Multihypothesis motion predictions are performed in addition to normal coding predictions at the encoder to extract useful motion information for fast error recovery at the decoder. The additional motion vectors of each odd frame are compressed and transmitted in the even subsequence and vice versa.

Simulation results demonstrate that the proposed scheme is robust to channel loss. It estimates lost frames more accurate than other previous used approaches in most cases, especially when multiple and complex motions appear. Applying this coding scheme can avoid complex motion search at decoder while adding minor redundancy to the bitstreams.

5. REFERENCES

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

[2] S. Lin and Y. Wang, "Error resilience property of multihypothesis motion-compensated prediction," *IEEE International Conference on Image Processing (ICIP)*, pp. 545–548, Oct. 2002.

[3] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Singal Processing Magazine*, pp. 74 – 93, Sep. 2001.

[4] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Info. Theory*, vol. 39, pp. 821 – 834, May 1993.

[5] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Info. Theory*, vol. 47, no. 6, pp. 2199 – 2224, Sep. 2001.

[6] N. Franchi, M. Fumafalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over ip," *Paket Video 2003, Nantes, France*, April 2003.

[7] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Tran. on Cir. Sys. for Video Tech.*, vol. 12, no. 6, pp. 438–452, June 2002.

[8] J. G. Apostolopoulos, "Error-resilient video compression via multiple state streams," *International Workshop on Very Low Bitrate Video Coding(VLBV99), Kyoto, Japan*, Oct. 1999.

[9] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," *Proc. of Visual Communications and Image Processing(VCIP) 2001*, vol. 4310, pp. 392–409, January 2001.

[10] G. Thomas, "Television motion measurement for datv and other applications," *Research Department Report, British Broadcasting Corporation*, September 1987.