OPEN ACCESS

University of
BRISTOL

Link to published version (if available):
10.1109/ISCAS.2005.1465188

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

S FRAME DESIGN FOR MULTIPLE DESCRIPTION VIDEO CODING

D. Wang[†], N. Canagarajah and D. Bull
Centre for Communications Research,
University of Bristol, Bristol, BS8 1UB, UK
Tel: +44 (0)117 9545125, Fax: +44 (0)117 9545206
[†]Email: Dong.Wang@bristol.ac.uk

Abstract—**Multiple description (MD) video coding generates several descriptions so that any subset of descriptions can reconstruct video, which provides much error resilience. But most of current MD video coding schemes are for two descriptions and for "on-off" channels, which is not suitable for packet-loss networks. This paper proposes a scheme to enhance the error resilience of traditional MD video coding in such environment, by periodically inserting S frames, a kind of switching frame, in the video stream to make good description recover the 'bad' description, with very small redundancy. This proves to perform well in packet lossy networks especially lower packet loss rate.**

I. INTRODUCTION

Video transmission over lossy network is a challenging problem. In video compression, due to predictive coding, any bit loss may cause great quality degradation. Multiple description coding is one approach to address this problem, where several sub bit streams called descriptions are generated from source video. Each description can reconstruct video of acceptable quality and all the descriptions together can reconstruct higher quality video.

Unlike layered video coding techniques, each description generated by MDC can independently be decoded and reconstructed to acceptable quality. This can give a graceful degradation of received video with loss, while avoiding catastrophic failure of layered coding due to loss of base layer.

An MDC system consists of two kinds of decoders, as shown in Fig. 1. One is the central decoder which is used when all the descriptions are received, and the other is side decoder which just uses one or a subset of descriptions to reconstruct video of acceptable quality.

More correlations in descriptions will result in higher quality of side decoded video. At the same time central decoder must perform with lower efficiency because more redundancy is introduced. Extensive research on MDC to increase the efficiency has been conducted.

MDC based on Scalar Quantization is developed in [1] to divide a signal by two coarser quantizers, and it's applied to predictive video coding in [2]. Output of each quantizer is the approximation of single description. Any one description can use its coarse data to generate a basic video and both of them can be combined to reconstruct higher quality video. Another approach on image coding is addressed in [3] and [4] using pairwise correlating transforms to transform a vector of DCT coefficients into another vector of correlated components, which introduces additional redundancy between components. This was used in motion compensated video coding [5]. Another simple way of generating MDC is that through pre- and post- processing, as in [6]. Redundancy is introduced by padding zeros in frequency domain. The source video frames are transformed using DCT. Certain number of zeros are padded in frequency domain, and after inverse transform, the video is sub-sampled into two descriptions. The two descriptions are independently coded at the encoder. In [7], video sequence is divided into two by means of odd and even frames and different concealment methods are used to estimate lost frames. In [8] odd and even frames compose two descriptions, which is similar to [7], but three MC loops are maintained. It performs well on ideal MDC environments and packet lossy network. A restriction is that it can only use previous two frames with constant weights of two motion vectors. Overlapping technique is used on motion vectors in [9] to achieve more accurate prediction of lost data.

Many of them contains only two descriptions and are mainly for the "on-off" channels, under the assumption that multiple independent channels are either error-free or temporarily down. In this environment, it can perform very well and the decoded quality is the side results generated from just one description. But if the channel is packet lossy or has burst error, each description may not be good but are not totally useless, hence the results will not be good as expected. Traditional error concealments or error concealments between two descriptions can be used, but they both cannot make the descriptions 'communicate' well with their qualities. For example with two descriptions, if we recover lost packets by copying the contents of the other description, there must be some degradations of quality for each copying. And with packet losses in both descriptions, we need to cross-copy one description to the other, which makes the quality decrease very fast. Schemes being able to recover the lost packet very good, which re-synchronizes the bit stream, will be more useful for such environments.

We propose a scheme to enhance its capability in packet loss network, by inserting S frames periodically in the video stream. Although S frame cannot reconstruct exactly the same as original, it can almost make video stream synchronized and recovered from errors. It is especially good for small burst errors. Since the descriptions have correlations between each other, encoding some S frames does not bring much redundancy to the stream. Experiments show that it can make the stream with fairly average quality by recovering 'bad' description.

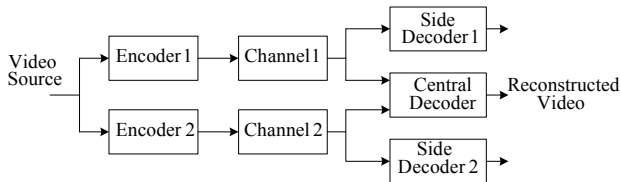The rest of this paper is organized as follows. In Section 2 our

Figure. 1. Example of MDC

Figure. 2. S frame scheme

scheme is described. Section 3 gives the results and analysis of experiments. Conclusions are presented in section 4.

II. DESCRIPTION OF THE PROPOSED S FRAME DESGIN

The problem of the MDC for "on-off" channels is the difficulties in recovering bad description. If there are several packet loss in one description which results in bad quality, this bad quality will remain until next Intra frame or macroblocks occur. But actually intra frame will not be so frequent because of very low efficiency. Since there are much correlations between descriptions, we introduce the S frame between them to synchronize them to maintain fairly good quality. The basic structure is as Fig.2. The specified frames in the stream are encoded by using corresponding frames in the other description as the reference, and this encoded frame is called S frame. We call positions of these specified S frame as S frame position. S frame is originally used for stream switching. Through S frame, one bitstream can be switched to another, such as different bit rate. It is used here for synchronization. It is obvious that we have one stream of S frame per description. And we call it S frame period from one S frame to the adjacent one. When decoding, if we found one description is worse than the other, we can use S frame at the S frame position to recover the frames in bad description. The frame at the position can be recovered directly. And if there is multiple reference frame coding, the previous frames can be backward recovered or concealed. Thus, the bad description can be recovered at each S frame position to the similar quality as the best one.

Although S frame coding is not like intra frame coding, which is independently coded and highly resilient to losses, the results can benefit much from it if the two channels are not in the same error statistics. If one description suffers error greatly, while the other is not so bad or contains no error, we can recover this description by the other, and we will get the quality at least in the level similar as the good one. This makes the bad description re-synchronized, which means previous error are eliminated from this S frame position. Thus the incompatibility with packet loss network is mitigated.

There are two conflicting things for this scheme. One is number of S frames inserted. The other is the redundancy. More number of S frames will provides better quality, but increase redundancy. It is lucky that for MDC the S frame is not so costly. For several MD schemes, encoding one S frame is just half of or at most near cost of encoding one normal frame. If the S frame period is not too short, the redundancy will not be so much. In the
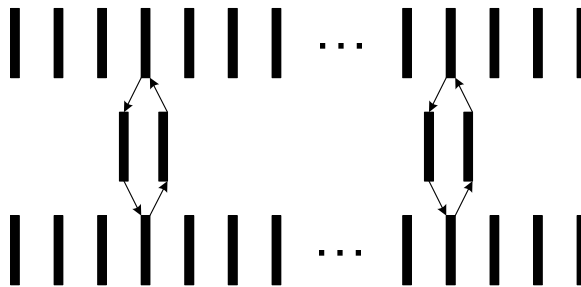
next section, the experiments are based on S frame period being 20. the redundancy is just around 2.5% adding to the MD scheme used.

Another similar technique is SP frame which is used in current H.264 standard. It is also a kind of switching frame. The difference is that SP frame reconstruct exactly the same frame as that in the stream by primary SP frame, a more complicated design of the frame in the stream, and the frame outside the stream, called secondary SP frame. Although SP frame has its great advantage that exact same reconstruction can be got, it is not chosen as our scheme. There are mainly three reasons. SP frame is much more complicated than S frame, and two quantizers are used. Moreover, since exact reconstruction is needed, SP frame has less efficiency than S frame, especially for this application. Two description is similar but different. Encoding S frame is very efficient, while encoding SP frame will keep every coefficient coded which make it very costly. The last reason is that we focus on packet loss network. It means each description may have loss. At the S/SP frame position, no matter how accuracy SP coding makes it, the recovered frame will not be the same as lost one. This makes SP design completely meaningless.

III. RESULTS AND ANALYSIS

We examine the performance of our proposed S frame scheme. We choose [6] as the base MDC system. This MDC system is based on latest video coding standard H.264. Two descriptions are generated at the encoder side, and they are merged together if correct or one description reconstructs video if only one is received. Fixed frame rate (30frames/second) and constant quantizer step size are used for each slice in all frames of sequence. No B frame is used. Entropy coding is CAVLC. We encode one packet per frame, which means packet loss rate is equal to frame loss rate.

Fig. 3 is an example of results of our proposed scheme. 'foreman' QCIF sequence is used and QP is set to 30. We assume one packet loss means one lost frame. For the lost frame, we copy previous frame as the error concealment, to simplify the experiment. We insert S frames every 20 frames. There are two lines representing two descriptions. We can see that the beginning PSNR is around 35-36dB. In the first period, there is no loss for description 1, but description 2 has several loss which makes PSNR dropping to 26dB. At the S frame position it is recovered by description 1. The PSNR is recovered back to the similar level as description 1. In the second period, both descriptions have losses. But it can be seen that description 2 has lower PSNR. At the S frame position, this worse description is recovered by description
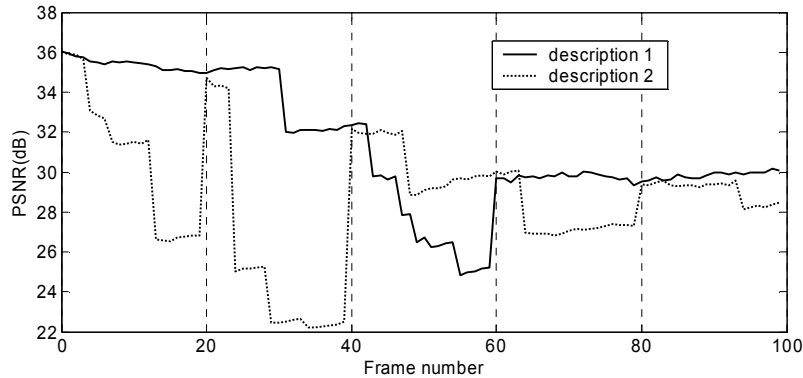
2720

Figure.3.   Results with example packet loss.

1, and the PSNR is around over 32dB. In the third period, the description 1 become worse and at S frame position it is much worse than description 2 and it is recovered by description 2. In the last two periods, description 1 is always better and it recovers description 2. From this example, it can be observed that in each period, the PSNR can always be kept similar as the best description.

In the practical environment, the quality degradations of each description is unknown. At the decoder side, we only have number of lost packet. Sometimes number of lost packets is similar between two descriptions. Due to property of video sequence, the importance of each frame or packet is different. This means equal number of lost frames may results in different PSNR. In this case, it's hard to say which description is better then hard to decide whether we should use this S frame or not. If we use S frame, the recovered frame may be worse than just error concealment, even if number of lost packets are the same. Here we introduce an additional parameter, cost of frames (COF) to solve this problem. This is calculated based on video source to let decoder know the importance of frames. We divide it into 4 levels, which take 2bits for each frame. It can be heavily protected by FEC or any other means and the additional bits can be discarded because of very small bits spent.

When both descriptions have loss in one S frame period, a decision will be made for which description is recovered using S frame, i.e. to decide which description is better. We make it through error value E_i (i=0,1) for each description.

$$E_i = \sum_{j=n-S}^{n-1} e_{ij}$$

$$e_{ij} = \begin{cases} COF_{ij} & , no\ burst\ error \\ \sum_{k=m}^{j} COF_{ik} & , with\ burst\ error \end{cases}$$

e_{ij} is the error value of j_{th} frame of description i. n is the frame number of current S frame position, and S is S frame period. When there is no burst errors, e_{ij} is the value of COF_{ij}. Otherwise, where the degradation is much greater, it is the sum of COF since beginning of burst error, denoted by m. If $E_i > E_j$ then description i is recovered by description j using S frame.

It should be noted that the decision method varies depending on the error concealment method. And it can be improved to be more accurate to estimate the better description.

Fig. 4 is the simulation results of our proposed scheme. Experiments with various qualities are done and they all perform better. For each loss rate, we run experiments 100 times. We evaluate the results for balanced channels which have the same loss rate, and the unbalanced channels with different loss rate in two channels. Fig. 4 (a) to (c) show the balanced channels for different encoding quality. It can be seen that with S frame the average PSNR is improved by 0.4-1.2dB, and the improvements are higher with lower loss rate. For the unbalanced channels, we fix the loss rate of one channel to 3% and vary the other loss rate in the fig.4 (d) and (e). The improvements are always higher than 1dB. The last figure show the PSNR improvements for each simulation. They vary depending on the detailed loss statistics. Sometimes there will be near 5dB better for some conditions. In several simulations, S frame makes it worse than without S frame because the error values in these simulations are not estimated precisely.

With S frame, the result of every simulation keeps similar, not like without S frame which sometimes causes very bad results, hence some simulations has 4-6dB improvements. This is useful for applications which need nearly steady quality. The redundancy is very small as mentioned above, which is just around 3% based on MDC stream. It is higher for lower bitrate, but still acceptable.

More S frames will bring more benefits, but with more redundancy.   It is a balance between efficiency and redundancy.

IV.   CONCLUSIONS

In this paper we introduced an approach based on S frame, to be used on the traditional MD video coding, which is designed for 'on-off' channels. With very small amount of redundancy, it can recover 'bad' description by good one using S frame. It is shown through simulations that it performs very well especially for lower packet loss rate, and efficiently mitigates the incompatibility of these kinds of MDC in packet-loss network.

REFERENCES

[1] V. Vaishampayan, "Design of multiple description scalar quantizers," IEEE Trans. On information Theory, vol. 39, no.3, pp.821-834, May 1993.
[2] V. Vaishampayan and S. John, "Interframe balanced-multiple-description video compression," Packet Video '99, New York, NY, USA, Apr. 1999.
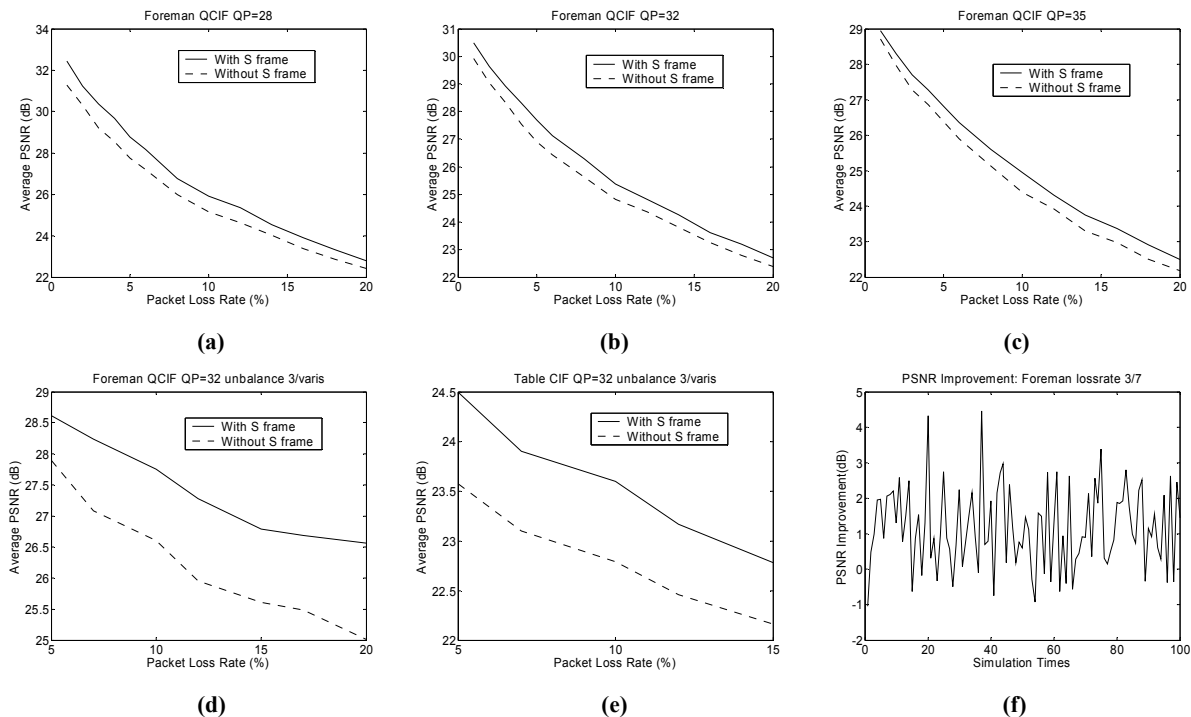[3] M. Orchard, Y. Wang, V. Vaishampayan, and A. Reibman,

Figure. 4. Results with various qualities and loss rates: (a), (b) and (c) are for the channels with the same loss rate. They have different bit rate. (d) and (e) are for the unbalanced channels in which different loss rates is set for them. (f) shows the improvement of each of 100 simulations.

TABLE 1. REDUNDANCY OF ENCODING WITH S FRAMES

Quantizer	Foreman QCIF		Table CIF	
	Bitrate	*Redundancy*	*Bitrate*	*Redundancy*
QP=28	159.06 kbits/s	2.39%	1075.35 kbits/s	2.07%
QP=32	95.90 kbits/s	3.22%	512.03 kbits/s	2.94%
QP=35	65.70 kbits/s	3.51%	315.00 kbits/s	3.43%

"Redundancy rate distortion analysis of multiple description coding using pairwise correlating transforms," Proc. IEEE Int. Conf. Image Proc, Santa Barbara, CA, USA, Oct. 1997.

[4] Y. Wang, M. Orchard, and A. Reibman, "Optimal pairwise correlating transforms for multiple description coding,", Proc. IEEE Int. Conf. Image Proc, Chicago, Illinois, USA, Oct. 1998.

[5] A. Reibman, H. Jafakhani, Y. Wang, and M. Orchard, "Multiple description coding for video using motion compensated prediction," Proc. IEEE Int. Conf. Image Proc, Kobe, Japan, Oct. 1999.

[6] M. Karczewicz, R. Kurceren, "The SP- and SI-Frames Design for H.264/AVC," IEEE Trans. On Circuits and Systems for Video Technology, Vol. 13, No.7, July 2003.

[7] D. Wang, N. Canagarajah, D. Redmill, D. Bull, "Multiple Descriptoin Video Coding Based on Zero Padding," Proc. IEEE Int. Symposium on Circuits and Systems, Vancouver, Canada, May 2004.

[8] J. G. Apostolopoulos, "Error-resilient video compression through the use of multiple states," Proc. IEEE Int. Conf. Image Proc, Vancouver, CA, USA, Sept. 2000.

[9] Y. Wang, S. Lin, "Error-Resilient Video Coding Using Multiple Description Motion Compensation,", IEEE Trans. On Circuits and Systems for Video Technology, Vol.12, No.6, June 2002.

[10] C. –S. Kim and S. –U. Lee, "Multiple description motion coding algorithm for robust video transmission," IEEE Int. Symp. on Circuits and Syst., Geneva, Switzerland, May 2000.

[11] H.264 standard, JVT-G050, 7th meeting, Pattaya, Thailand, 7-14 March, 2003

[12] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, pp. 560- 576, July 2003.