MULTIPLE DESCRIPTION VIDEO CODING USING ADAPTIVE TEMPORAL SUB-SAMPLING

[1]Huihui Bai, [2]Yao Zhao and [3]Ce Zhu

[1,2] Beijing Jiaotong University, China, [3]NanyangTechnological University, Singapore

ABSTRACT

Multiple description coding (MDC) is a promising alternative for robust transmission of information over non-prioritized and unpredictable networks. Especially, MDC has emerged as an attractive approach for video applications where retransmission is unacceptable or infeasible. In this paper, an effective MD video codec is designed based on pre- and post-processing of video sequences, without any modification to the source or channel codec. Considering different motion information of inter-frame, adaptive temporal sub-sampling is employed to the original video data. As a result, adaptive redundancy can make a better tradeoff between the reconstruction quality and compression efficiency. The experimental results exhibit better performance of the proposed scheme than other schemes.

1. INTRODUCTION

Network congestion and delay sensibility pose great challenges for multimedia communication system design. This creates a need for coding approaches combining high compression efficiency and robustness. Multiple description (MD) coding has emerged as an attractive framework for robust transmission over unreliable channels. It can efficiently combat packet loss without any retransmission thus satisfying the demand of real time services and relieving the network congestion [1].

Multiple description coding encodes the source message into several bit streams (descriptions) carrying different information which then can be transmitted over the channels. If only one channel works, the descriptions can be individually decoded to sufficiently guarantee a minimum fidelity in the reconstruction at the receiver. However, when more channels work, the descriptions from the channels can be combined to yield a higher fidelity reconstruction. In the past years, the MDC of still image has attracted a lot of attention, and several algorithms have been proposed [2-5].

MDC is especially promising for video coding due to the very stringent delay requirement in many video applications [6]. Recently, the MD version of sampling [7] is a popular technique for the design of MD video coding. In [7], an MDC scheme is presented with pre- and post-processing stages. Redundancy is added by padding various numbers of zeros in one dimension DCT domain of each frame and multiple descriptions are generated by sub-sampling zero padded frames. It is showed in [7] that the 1D approach performs much better than 2D padding techniques in [8], at a much lower computational complexity. However, in [7] through zero padding inside the frames, only the correlation of intra-frame is considered to improve side distortion and the temporal correlation of inter-frame is neglected completely.

In this paper, we attempt to design a more effective MD video coder based on pre-/post- processing and sub-sampling. In view of the different motion information between frames, the redundancy is added adaptively to make a better tradeoff between the reconstructed quality and the compression efficiency. Consequently, the abrupt motion may result in worse side distortion so more redundancy is needed to guarantee the correlation between frames while less or no redundancy is needed if the motion of inter-frame is enough smooth.

The rest of this paper is organized as follows. In Section 2, an overview of the proposed MD video coding scheme is given. In the following sub-section 2.1, the pre-processing stage is presented in detail, and the post-processing stage is presented in sub-section 2.2. The performance of the proposed video coding scheme is analyzed and examined against other MD coders in Section 3. We conclude the paper in Section 4.

2. OVERVIEW OF THE PROPOSED SCHEME

Fig. 1 illustrates our proposed scheme for MD video coding. In the pre-processing stage, the original video sequence is up-sampled to generate a new-length video with adaptively redundant frames. And then by means of odd and even frames the new video sequence is divided into two descriptions, which can be compressed by any standard video codec. Here, the latest video coding standard H.264 is employed for a suitable comparison. In the post-processing, the decoded video stream is down-sampled to reconstruct original-length video and error concealment method is used to estimate lost frames. The details of pre- and post-processing are showed in the following sub-sections.

1-4244-1017-7/07/$25.00 ©2007 IEEE

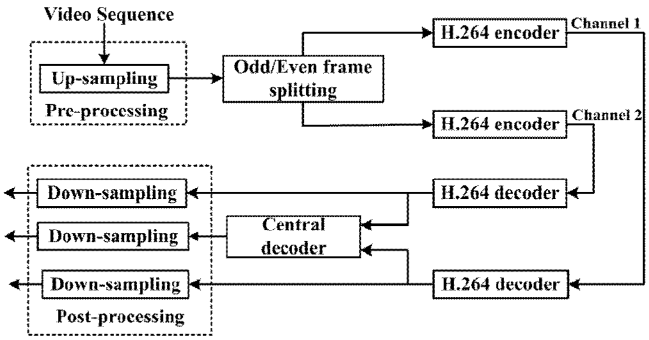

Fig.1 Block diagram of the proposed scheme

2.1. The pre-processing stage

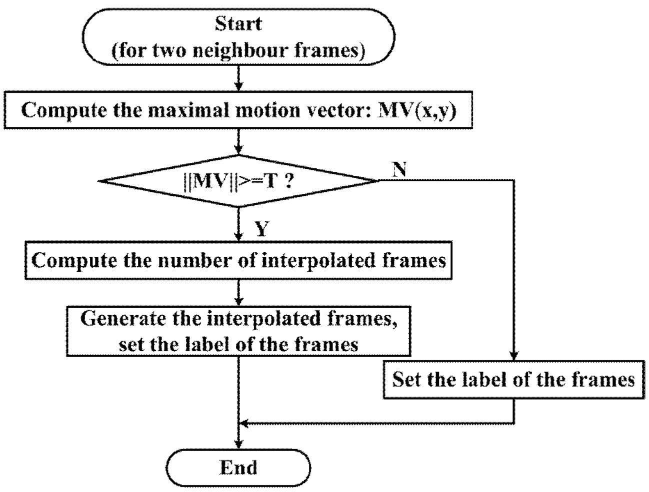

Fig.2 The flow chart of up-sampling

In terms of the principle of MDC, higher quality of side decoded video will result from more correlations in descriptions for better error concealment is available, but more redundancy introduced will bring about lower efficiency to the central decoder. Obviously, it is a better solution that the redundancy added can make a tradeoff between the reconstructed quality and the compression efficiency. As a result, in the pre-processing stage, the up-sampling with changeable rate is employed to introduce the adaptive redundancy. Since different motion appearance of inter-frames will affect the quality of side decoded video, the rate of up-sampling is various according to the motion information between any two neighboring frames. The obvious motion between the frames can result in more interpolated frames and higher rate of up-sampling. On the other hand, if the motion information of the inter-frame is very smooth, fewer or no redundant frames are needed. Such method of up-sampling mainly aims to generate descriptions with smooth motion which makes better evaluation of lost frames available at the side decoder.

In Fig.2 the flow chart of up-sampling is shown step by step. For any two neighboring frames, all the motion vectors for each macro block are computed and the maximum can be obtained. Here it is denoted by $\|MV\|$ and $\|MV\| = \sqrt{x^2 + y^2}$ ((x, y) is the coordinates of the maximal motion vector). Then the maximal motion vector is compared with the threshold T. If $\|MV\| < T$, the motion between the two frames is considered smooth so no frames need to be interpolated. Otherwise, redundant frames are needed. In view of the balance of two channels, even frames are interpolated to maintain equal frame number of two descriptions. Consequently, the number of interpolated frames is computed by $2\times\lceil(\|MV\|/T-1)/2\rceil$. In the end the redundant frames can be generated using the general algorithm of motion compensated interpolation, such as [9]. Additionally, the label with one bit ('1' or '0') is set for each frame to distinguish the original frame and interpolated one, then duplicated and transmitted on two channels. Here, we assume label=1 represents the original frame and label=0 is the interpolated one.

2.2. The post-processing stage

In the post-processing stage, two situations for decoding exist, that is, the design of central or side decoder.

Since the two descriptions are generated by odd and even means, at the central decoder, the video streams from H.264 decoder can be interleaved and realigned in the same way to achieve the redundant video. According to the labels ('1' or '0'), the interpolated frames can be deleted, so the redundant video can be down-sampled to obtain the final reconstruction which has the same frame number with the original video sequences.

If only one channel works, the side decoder is employed and four possibilities should be taken into account.

1) If the current label is '1' but its following label is '0', the represented frame is just the reconstructed one.

2) If the current label is '0' but its following label is '1', the represented frame is the interpolated frame and it can be regarded as the reconstructed one.

3) If the current label is '0' and its following label is also '0', the continuous frames represented by '0' should be merged to a reconstructed frame.

4) If the current label is '1', and its following label is also '1', a new frame should be interpolated between the two frames denoted by '1'.

In Fig.3, a simple example illustrates the pre- and post-processing. The original video sequence has 10 frames denoted by frame 1 to frame 10. After up-sampling, the redundant video has 18 frames. From the figure, we can see even frames are interpolated adaptively, such as two frames interpolated between frame 1 and frame 2, four frames interpolated between frame 4 and frame 5. After splitting by odd and even means, the generated descriptions are denoted by video on channel 1 and video on channel 2 and the labels are '101001101' and '011001101' respectively. When only channel 1 works, the reconstruction from side decoder 1 is achieved like the figure. The two interpolated frames between frame 3 and frame 5 will be merged into a new reconstructed one while a new frame is interpolated between frame 5 and frame 7 to evaluate the lost frame 6. On the other hand, if two channels work, the lossless video may be obtained without the processing by H.264 codec.

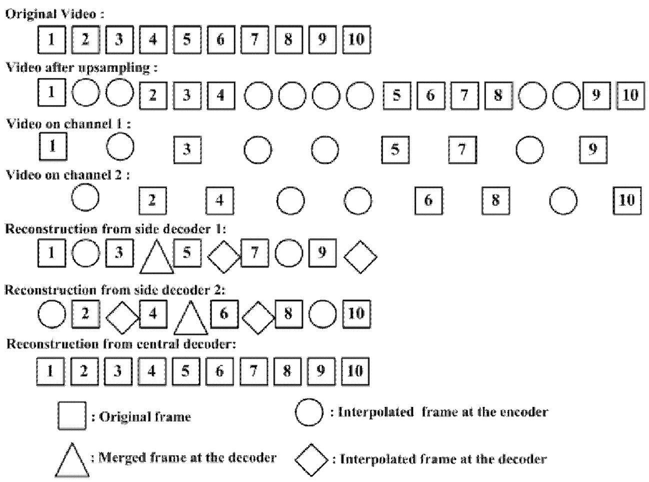

Fig.3 A simple example for the pre- and post-processing

3. EXPERIMENTAL RESULTS

Here, there are mainly two experiments taken into account to present the efficiency of up-sampling in temporal domain. The first one is shown the better performance of the proposed scheme than the conventional scheme without up-sampling in the pre-processing stage. In the second experiment, the advantage of the proposed scheme is illuminated compared with the up-sampling in the spatial domain.

The conventional scheme is similar with Fig.1 but without up-sampling block. The standard test video "coastguard.qcif " is used with 30 frames per second. For a fair comparison, the same mode and parameters are chosen in H.264 encoder and decoder. Additionally, we also employ the same method of motion compensated interpolation.

In Fig.4, at the almost same central distortion, the side distortion of the proposed scheme has 0.5 dB better than the conventional scheme. However, this is just a global comparison for the average of the whole video. In fact, some individual frames may achieve more advantages over the conventional schemes. Fig.5 shows the central and side PSNR of each frame at the bit rate of 138kbps achieved by the proposed and conventional schemes. From Fig.5 (a), the central PSNR of the two compared schemes can be found the almost same performance. However, in Fig.5 (b), the proposed scheme can perform obvious improvement in the side distortion around the frame number 70. From Fig.6 to Fig.9, the side reconstructed frames by the two compared schemes are presented to illustrate the efficiency of the proposed scheme.

To present the advantage over the scheme using up-sampling in spatial domain, the proposed scheme is compared with [7] in Fig.10. For the fair comparison, the same code mode of H.264 is employed in the two compared scheme for the test video "foreman.qcif". From the Fig.10, we can see better rate and central/side distortion performance achieved by the proposed scheme, especially at the lower bit rate.

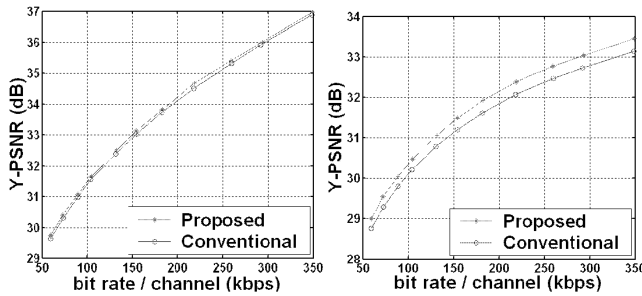

Fig.4 (a) Fig.4 (b)

Fig.4 Comparison for the test video "coastguard.qcif".
(a) Rate/central distortion performance
(b) Rate/side distortion performance

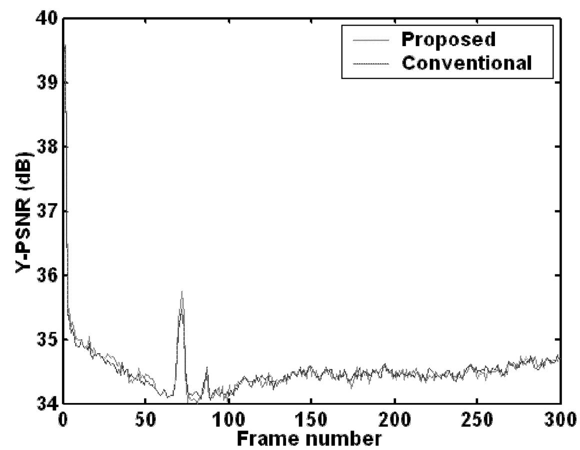

Fig.5 (a)

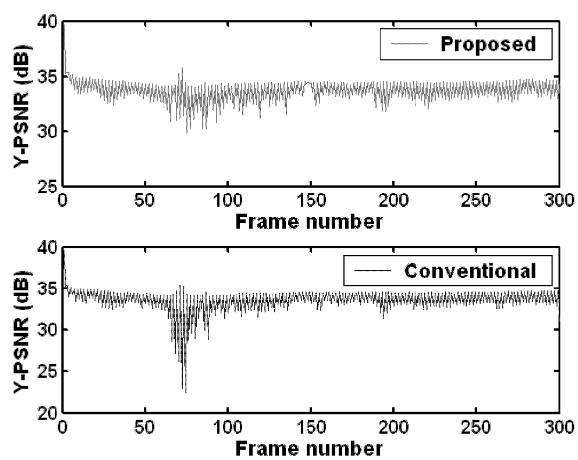

Fig.5 (b)

Fig.5 Comparison for the test video "coastguard.qcif".
(a) Central PSNR values for each frame
(b) Side PSNR values for each frame

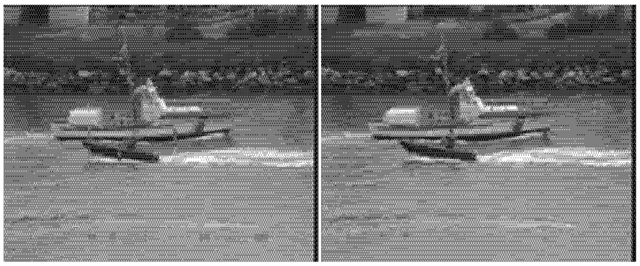

Fig.6 (a) Fig.6 (b)

Fig.6 (a) the conventional scheme (Frame no.:70, Side PSNR: 26.267dB)
Fig.6 (b) the proposed scheme (Frame no.:70, Side PSNR: 31.352dB)

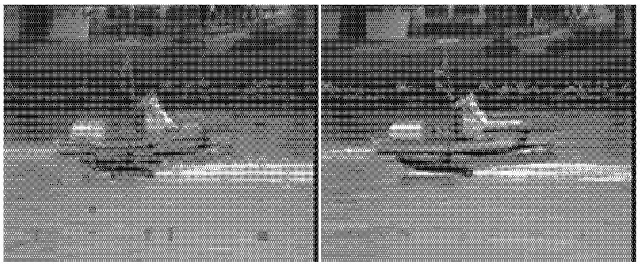

Fig.7 (a) Fig.7 (b)

Fig.7 (a) the conventional scheme (Frame no.:72, Side PSNR: 22.999dB)
Fig.7 (b) the proposed scheme (Frame no.:72, Side PSNR: 31.681dB)

Fig.8 (a) Fig.8 (b)

Fig. 8 (a) the conventional scheme (Frame no.:74, Side PSNR: 22.435dB)
Fig.8 (b) the proposed scheme (Frame no.:74, Side PSNR: 26.698dB)

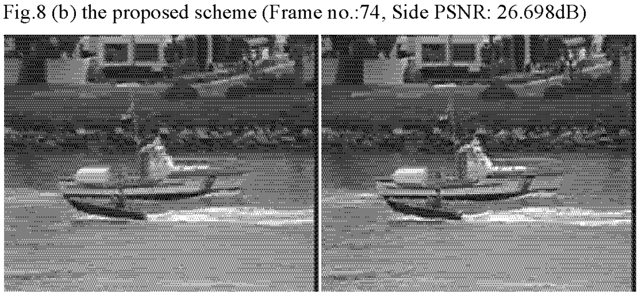

Fig.9 (a) Fig.9 (b)

Fig.9 (a) the conventional scheme (Frame no.:76, Side PSNR: 28.702dB)
Fig.9 (b) the proposed scheme (Frame no.76, Side PSNR: 31.075dB)

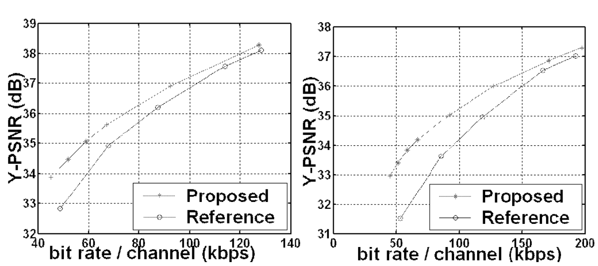

Fig.10 (a) Fig.10 (b)

Fig.10 Comparison for the test video "foreman.qcif".
(a) Rate/central distortion performance (b) Rate/side distortion performance

4. CONCLUSION

An MD video coding scheme based on the pre- and post-processing has been developed in the paper, without any modification to the source or channel codec. In view of the temporal redundancy and actual motion information, adaptive motion compensated interpolation has been accommodated in the proposed system to achieve better tradeoff between the reconstructed quality and the compression efficiency. As a result, the proposed MD video coding scheme has demonstrated superior rate-distortion performance to the MD video coder in [7].

ACKNOWLEGEMENT

This work was supported in part by NSFC(No. 90604032, No. 60373028), SRFDP，NCET and Specialized Research Foundation of BJTU, 973 program (No.2006CB303104).

REFERENCES

[1] V.K. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp.74–93, Sept. 2001.

[2] S.D. Servetto, K. Ramchandran, V.A. Vaishampayan and K. Nahrstedt, "Multiple Description Wavelet Based Image Coding," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp.813–826, May 2000.

[3] T. Tillo and G. Olmo, "A Novel Multiple Description Coding Scheme Compatible with the JPEG2000 Decoder," *IEEE Signal Processing Letters*, vol.11, no.11, pp. 908-911, Nov. 2004.

[4] C. Tian and S.S. Hemami, "A New Class of Multiple Description Scalar Quantizer and Its Application to Image Coding," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp.329–332, Apr. 2005.

[5] H. Bai, Y. Zhao, C. Zhu, "Multiple Description Shifted Lattice Vector Quantization for Progressive Wavelet Image Coding," *IEEE International Conference on Image Processing*, Atlanta, pp.797-800, 2006.

[6] Y. Wang, A.R.Reibman, S.Lin, "Multiple Description Coding for Video Delivery," *Proceedings of the IEEE*, vol.93, no.1, pp57-70, Jan. 2005.

[7] D. Wang, N. Canagarajah, D. Redmill and D. Bull," Multiple Description Video Coding Based on Zero Padding," *Proceedings of the 2004 International Symposium on Circuits and Systems 2004*, ISCAS '04, vol. 2, pp.205-208, May 2004

[8] M. Gallant, S.Shirani and F. Kossentini, " Standard-compliant Multiple Description Video Coding," *IEEE International Conference on Image Processing*, Thessaloniki, Oct. 2001.

[9] L.Wei, Y.Zhao and A.Wang,"Improved Side-information in Distributed Source Coding," *International conference on Innovative Computing, Information and Control*, Beijing, vol.2,pp.209-212, Aug. 2006.

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