Computing and Informatics, Vol. 29, 2010, 1267–1282

MULTIPLE DESCRIPTION VIDEO CODING USING JOINT FRAME DUPLICATION/INTERPOLATION

Huihui Bai, Yao Zhao

Institute of Information Science
Beijing Jiaotong University
100044 Beijing, China
e-mail: `luckybhh@gmail.com`

Ce Zhu

School of Electrical and Electronic Engineering
Nanyang Technology University
Singapore, Singapore

Anhong Wang

Taiyuan University of Science and Technology
030024 Taiyuan, China

Manuscript received 30 October 2008; revised 3 July 2009
Communicated by Vladimír Britaňák

Abstract. Multiple description coding (MDC) is a promising alternative to combatting information loss without any retransmission. In this paper, an effective MD video codec is designed based on temporal pre- and post-processing of video sequences without modifying the actual coding process itself, which makes it compatible with the current standard source or channel codec. For ease of post-processing, motion-compensated interpolation (MCI) based on piecewise uniform motion assumption is adopted to estimate the lost frame in side decoding. Accordingly, to match the post-processing, in the pre-processing joint frame duplication/interpolation is first applied to the original video data before performing odd/even frame splitting, which attempts to make the motion variety in the generated descriptions piecewise uniformly thus achieving better side reconstructed quality based on

MCI. The experimental results exhibit better performance of the proposed scheme than some other tested schemes, in both the on-off channel environment and packet loss network.

Keywords: Video coding, multiple description coding, frame duplication, frame interpolation

1 INTRODUCTION

With the explosion of the Internet, video transmission has become increasingly popular in the recent years and will continue to flourish in the future. However, network congestion and delay sensibility impose tremendous challenge on video communications. Due to network congestion, random bit errors and packet losses may cause substantial quality degradation of the compressed video sequence. In the case of real-time video application, delay sensibility has made the retransmission of corrupted data impossible. Therefore, this creates a need for coding approaches combining high compression efficiency and robustness. Multiple description (MD) coding is an attractive framework to solve this problem. It can efficiently combat information loss without any retransmission, thus satisfying the demand of real-time services and relieving the network congestion [1].

MD coding (MDC) encodes the video message into several bit streams (descriptions) carrying different information which can be transmitted over channels. The descriptions can be individually decoded to reconstruct video with a minimum fidelity which is measured by side distortion. However, when more channels work, the descriptions from the channels can be combined to yield a higher fidelity reconstruction. In a simple architecture of two channels, the distortion generated by two received descriptions is called central distortion. Here, we mainly focus on MD design for two channels.

During the past years, the MD versions of quantizers and transform are the two popular methods for the design of MD image/video codec. Based on the principle of MD scalar quantizer [2], an MD scheme for video coding is proposed in [3] while MD correlation transform is also employed to design motion compensated MD video coding [4]. Although the above methods have shown good performance, they are incompatible with widely-used standard codecs, such as H.26x and MPEG-x.

To address the limitation, subsampling-based MD is developed with pre- and post-processing shown in Figure 1. A video sequence is firstly processed and then it can be split into two subsequences going through any standard encoder. At the receiver, post-processing is used for error concealment. In [7], a spatial subsampling-based MD video coder is developed as an extension of MD image coding [6] with a frame by frame approach. In the pre-processing, redundancy is added by padding a number of zeros in vertical direction in the DCT domain of each frame. After the inverse DCT, two descriptions are generated by spatially subsampling the smooth and enlarged frames. It is shown in [7] that the 1-D padding approach performs better than 2-D method in [5] with a lower computational complexity. Using the zero padding within each individual frame, spatial correlation of intra-frame is mainly exploited and therefore, in the post-processing the weighted mean value of the adjacent pixels in the same frame is used to estimate the lost pixel.

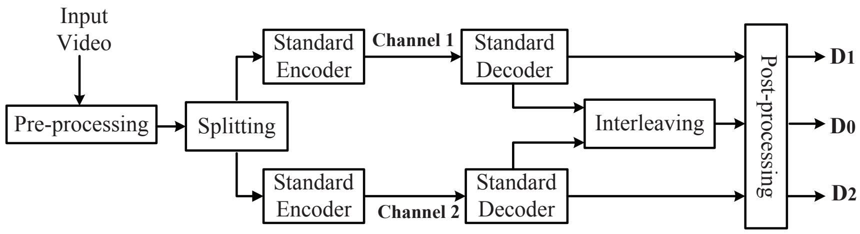


Fig. 1. Block diagram of a typical MD video coding with pre-/post-processing

However, in terms of error concealment, spatial interpolation tends to be less effective than temporal interpolation especially for low-bit-rate video application [8] because considering a video sequence is mainly featured with temporal correlation. Furthermore, due to spatial subsampling, temporal correlation will also be compromised, because it is also highly dependent on spatial resolution of each frame (That is why fractional motion estimation can improve motion compensation quality). Compared to the approach in [7], we attempt to exploit temporal correlation to design a more effective MD video coder based on pre- and post-processing. In order to obtain better side reconstructed quality, the pre-processing should be designed to match the temporal error concealment in the post-processing. Here, a widely-used method of motion-compensated interpolation (MCI) is adopted in the post-processing based on piecewise uniform motion assumption. Accordingly, in the pre-processing, joint frame duplication/interpolation is first applied to the original video data before performing odd/even frame splitting, which attempts to make the generated descriptions satisfying the piecewise uniform motion assumption for MCI.

The rest of this paper is organized as follows. In Section 2, the proposed MD video coding scheme is presented including pre- and post-processing. In Section 3, the performance of the proposed scheme is examined against some other relevant existing MD coders. We conclude the paper in Section 4.

2 THE PROPOSED MD SCHEME

2.1 The Pre-Processing

As mentioned above, the pre-processing should be designed to match the post-processing based on the piecewise uniform motion assumption. In Figure 2, for example, in the original video the motion speed of the macroblock (indicated by the black circle) between frame k and frame $k+1$ is changed compared with the motion between frame $k-1$ and frame k. In the reconstructed video, due to the piecewise uniform motion assumption for MCI, the motion speed between neighboring frames will be considered unvaried. Therefore, when frame k is reconstructed by frame $k-1$ and frame $k+1$ in odd/even subsequence generated by splitting, obvious distortion will occur in the reconstructed frame k shown in Figure 2. In view of the motion variety in the original video, we can interpolate one frame for motion adjustment shown in Figure 3. Then, the motion from frame $k-1$ to frame $k+1$ is smooth enough to reconstruct frame k accurately. Therefore, the pre-processing is necessary for side reconstructed quality based on MCI, especially when large motion variety occurs.

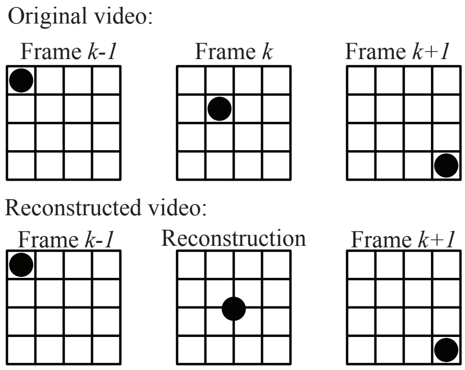


Fig. 2. MCI for frame k without pre-processing

In a simple example shown in Figure 2 the motion variety is only one unit shift in horizontal and vertical direction. If more significant motion change occurs, more interpolated frames should be needed to regulate the motion uniformly, which means too much redundancy is increased. In this case, it may be more effective to have only two corresponding frames containing large motion variety to duplicate and transmit on both channels. As a result, an adaptive pre-processing is developed by the combination of frame duplication and interpolation to adjust the motion speed uniformly in the original video. Here, for simplicity only two modes of joint frame duplication/interpolation are taken into account, that is, two frames duplicated and one frame interpolated. They are shown in Figure 4. Although more combined modes of frame duplication/interpolation may lead to better side distortion, the computational complexity will also be increased at the same time.

The flowchart of pre-processing is shown in Figure 5. For any two adjacent frames, the motion vector for each macroblock is computed and the maximal motion vector (MV_x, MV_y) can be obtained. We calculate the magnitude using $\|MV\| = \sqrt{MV_x^2 + MV_y^2}$. Here, the motion variety is measured by the difference of $\|MV\|_{k+1}$

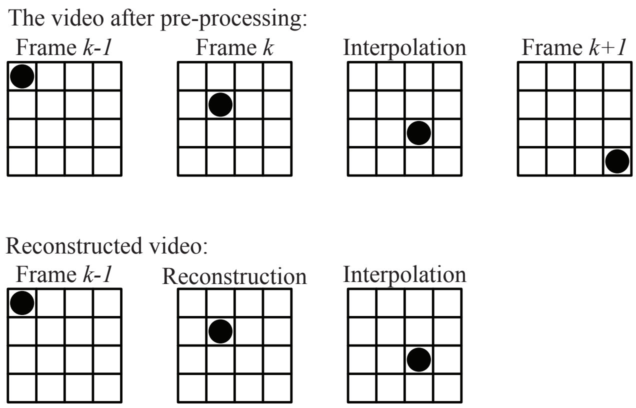


Fig. 3. MCI for frame k with pre-processing

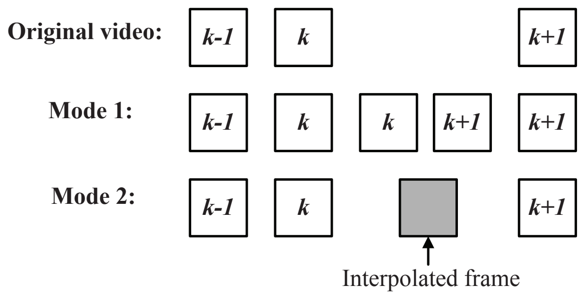


Fig. 4. The modes of joint frame duplication/interpolation

and $\|MV\|_k$. If $\left|\|MV\|_{k+1} - \|MV\|_k\right| \geq T_1$, significant motion variety is considered and two duplicated frames are inserted, which is mode 1. At the same time, if $T_2 \leq \left|\|MV\|_{k+1} - \|MV\|_k\right| \leq T_1$, smaller but not negligible motion variety occurs and one interpolated frame is needed, which is mode 2. Additionally, no redundant frames are needed for uniform motion speed, which can be labelled as mode 0.

It is noted that the threshold T_2 is set empirically using the mean value of all the motion variety, that is, for an original video with N frames,

$$T_2 = \frac{1}{N-2} \sum_{k=1}^{N-2} \left|\|MV\|_{k+1} - \|MV\|_k\right|, \tag{1}$$

and $T_1 = 2T_2$. Furthermore, the interpolated frames can be generated using the conventional MCI method, which is introduced in subsection 2.2. At the end, the label with two bits is set for each frame to distinguish the three modes adop-

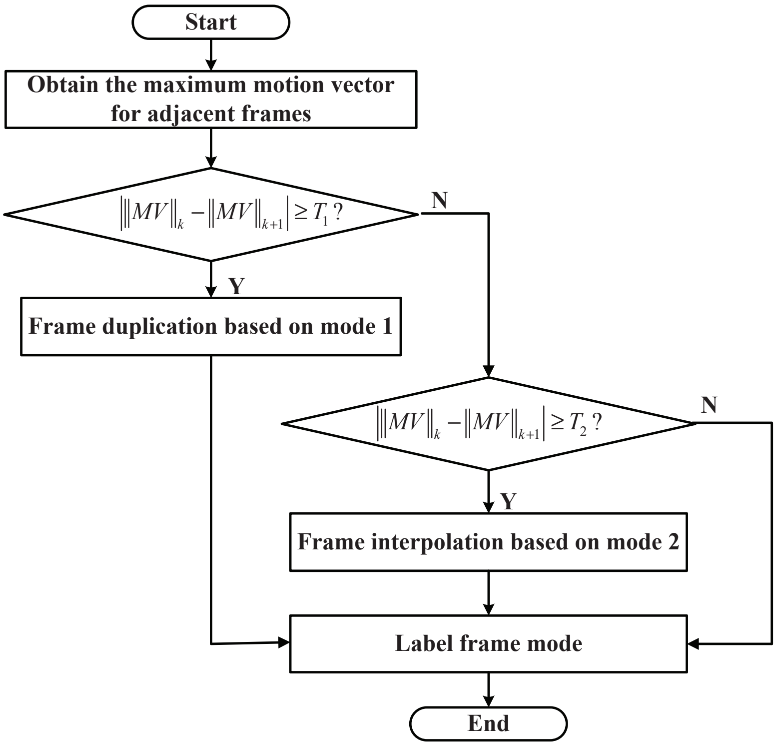


Fig. 5. The flowchart of pre-processing

ted.

Although frame duplication/interpolation in the pre-processing is helpful to regulate the motion change between frames for better side reconstruction, at the same time the inserted frames also increase the size of original video, which may impact the bit rate. However, due to frame duplication/interpolation, the temporal correlation in each generated description of the proposed scheme is better than the conventional scheme without pre-processing and spatial-based subsampling scheme. The temporal correlation coefficient ρ between the current frame f_k and its previous frame f_{k-1} can be computed by

$$\rho(f_k, f_{k-1}) = \frac{\sum\sum\left(f_k - \overline{f_k}\right)\left(f'_{k-1} - \overline{f'_{k-1}}\right)}{\sqrt{\left(\sum\sum\left(f_k - \overline{f_k}\right)^2\right)\left(\sum\sum\left(f'_{k-1} - \overline{f'_{k-1}}\right)^2\right)}}, \tag{2}$$

where f'_{k-1} is the reconstructed frame by f_{k-1} after motion estimation and compensation. $\overline{f_k}$ is the mean of f_k and $\overline{f'_{k-1}}$ is the mean of f'_{k-1}. It is shown in Figure 6 that the proposed scheme has improved both the minimum and the mean value of temporal correlation. Higher temporal correlation in one description means easy compression which will counteract some effect from the size of the redundant frames. Therefore, the compression performance will not drop seriously with increasing frame number, which is exhibited in Section 3.

2.2 The Post-Processing

There are two environments for MDC. One is the on-off channel environment and the other is packet loss network. In the on-off channel environment, at the decoder only one intact description is received or both descriptions are received accurately. In packet loss network, packet losses occur in each description, and both descriptions have to be used at the decoder.

1. **Decoder Design for the On-off Channel Environment** In the on-off channel environment, two cases for decoding should be taken into account, that is, the design of central decoder and side decoder.

   Since the two descriptions are generated by odd and even means, at the central decoder, the two video subsequences after standard decoding can be interleaved firstly. Then according to the labels, the inserted frames will be removed to obtain the central reconstruction.

   If only one channel works, the side decoder is employed to estimate the lost information according to the labels. The widely-used method of MCI based on the piecewise uniform motion assumption is performed by the bi-directional motion estimation, which may produce overlapped pixels and holes in the estimated frame.

   For convenience, some notations are defined as follows. Let f denote the interpolated frame between frame f_k and frame f_{k+1}, and let $MV(\vec{p})$ denote the motion vector between the two reference frames at the pixel position $\vec{p}$.

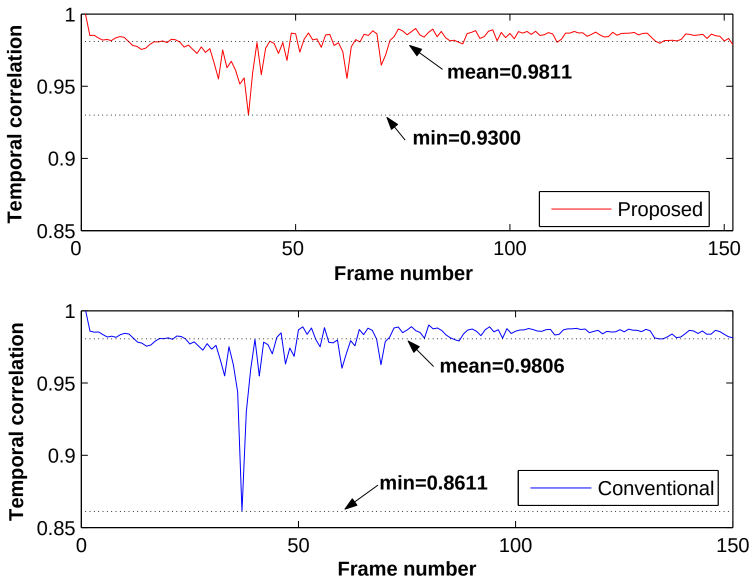
   To avoid the holes in the interpolated frame, we can compute the preliminary reconstruction as background information using

$$f = \frac{1}{2}\left(f_k(\vec{p}) + f_{k+1}(\vec{p})\right). \tag{3}$$

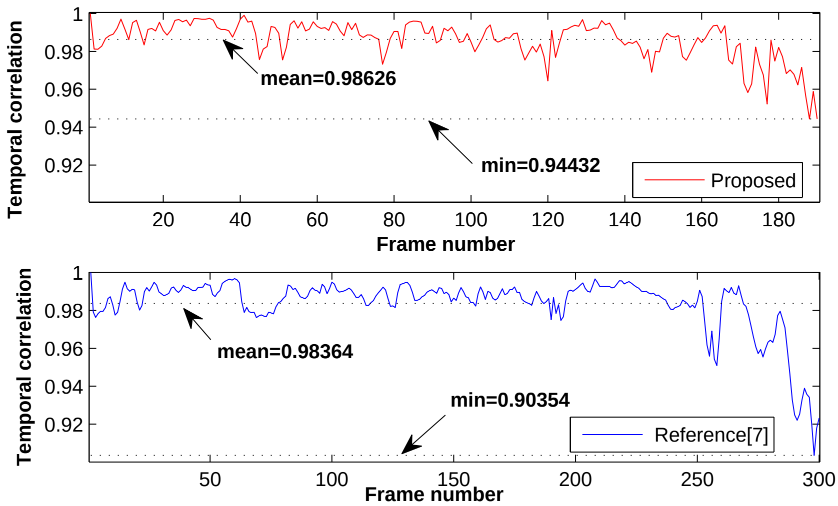
   Then, the preliminary reconstruction will be replaced by MCI according to

$$f = \frac{1}{2}\left(f_k\left(\vec{p} - \frac{1}{2}MV(\vec{p})\right) + f_{k+1}\left(\vec{p} + \frac{1}{2}MV(\vec{p})\right)\right). \tag{4}$$

   Furthermore, to avoid overlapped problem of MCI, the mean value of overlapped pixels can be adopted for the final reconstruction.



(a)



(b)

Fig. 6. Temporal correlation in each description: a) the proposed scheme vs. the conventional scheme for "coastguard.qcif", b) the proposed scheme vs. the reference scheme [7] for "foreman.qcif"

2. Decoder Design for Packet Loss Network In packet loss network, due to both descriptions received suffering from packet losses, only central decoder should be designed. Here, two types of lost information need to be estimated for a damaged macroblock, specifically the motion vector and the pixel values. It is noted that at the encoder packets in each frame are organized like a checkerboard, that is, any macroblock and its surrounding ones are not grouped in the same packet.

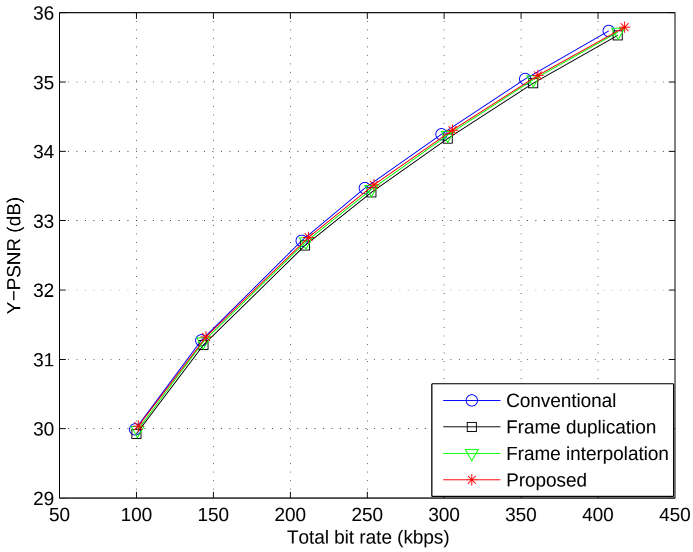
After standard decoding, the two generated video subsequences are interleaved by odd and even means firstly to produce a video with redundant frames. Due to the packet organization of a checkerboard, the motion vector of the lost macroblock is first obtained using the mean of the motion vectors from its spatially adjacent macroblocks. Then, we can use the macroblock in the previous frame pointed by the generated motion vector for motion compensation. Furthermore, the method of MCI is also employed to improve the reconstructed quality of lost macroblock, especially when at higher packet loss rate spatially adjacent macroblocks tend to be also damaged. In the end, according to the labels, the inserted frames will be removed to obtain the central reconstruction.

3 EXPERIMENTAL RESULTS AND COMPARISONS

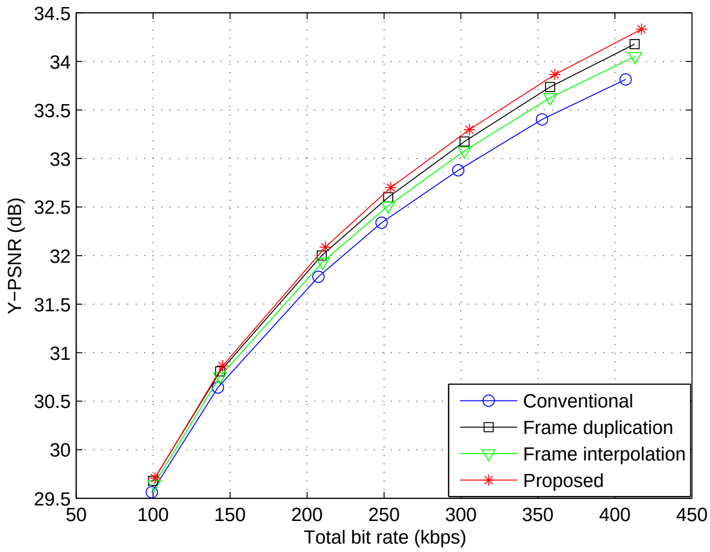
Three experiments are considered to present the efficiency of proposed joint frame duplication/interpolation. The first one is shown in the on-off channel environment demonstrating better rate/side distortion performance of the proposed scheme than the conventional scheme without pre-processing and the schemes with pure frame duplication or interpolation. Comparison of the proposed and conventional scheme in packet loss network is presented in the second experiment. In the last experiment, the advantage of the proposed scheme is exhibited compared with the spatial subsampling schemes ([5] and [7]).

The standard test video "coastguard.qcif" is used with 30 frames per second. For a fair comparison, the same parameters are chosen in H.264 encoder and decoder [9]. Additionally, we also employ the same MCI method for error concealment.

Figure 7 shows the central and side distortion of the proposed scheme against other existing schemes at the bit rate from 50 kbps to 450 kbps. Here, the total bit rate is attained from two descriptions including their labels; the side distortion is the average PSNR value from two side decoder. From the figures, in the on-off channel environment, at the same bit rate and the almost same central distortion, the proposed scheme can consistently perform better than the conventional scheme and the schemes with pure frame duplication or interpolation in side distortion. This is just a comparison for the average PSNR values of the whole video. In fact, some individual frames in the proposed scheme may achieve more advantages over the conventional schemes. Figure 8 shows the side PSNR of each frame at the bit rate of 210 kbps achieved by the proposed and conventional schemes. It can be

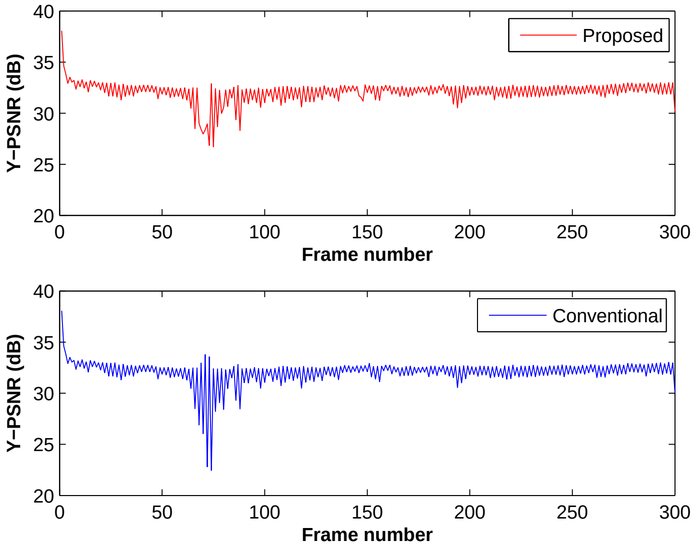


(a)

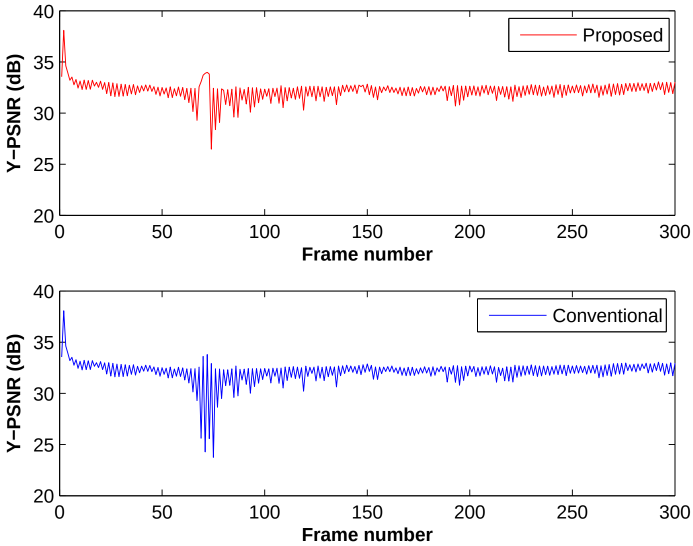


(b)

Fig. 7. The proposed scheme vs. conventional one and pure frame duplication or interpolation in on-off MD environment: a) rate/central distortion performance, b) rate/side distortion performance (test video: coastguard.qcif)



(a)



(b)

Fig. 8. Side distortion for each frame: a) channel 1, b) channel 2 (test video: coastguard.qcif)

seen that for either channel the side distortion of the proposed scheme has obvious improvement around the frame number 70.

In Figure 9, the performance of the proposed and conventional schemes is compared in packet loss network at the same total bit rate 210 kbps. For a fair comparison, the organization of packets is a checkerboard type for both cases. In Figure 9, the proposed scheme has achieved better performance than the conventional one. The improvements are obtained from 0.5 dB to 1.2 dB. Note that for low packet loss rate, the improvement is greater. With higher packet loss rate, both the proposed scheme and the conventional one cannot do well in error concealment due to so many packets lost.

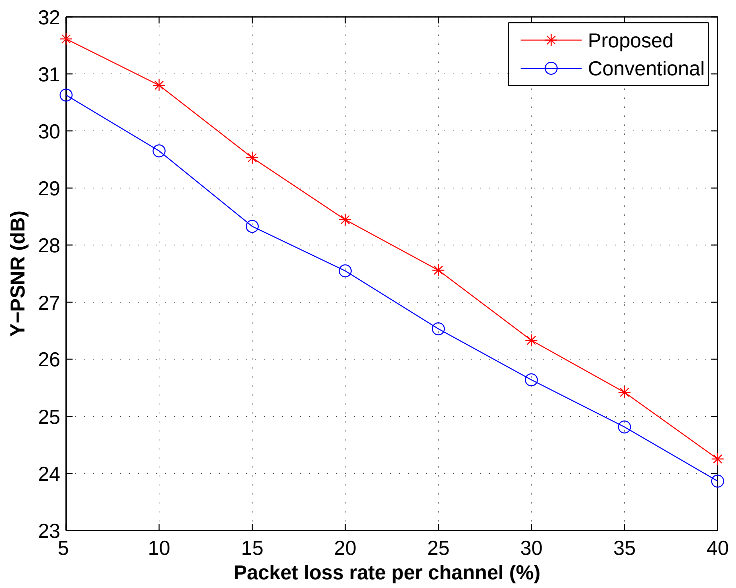
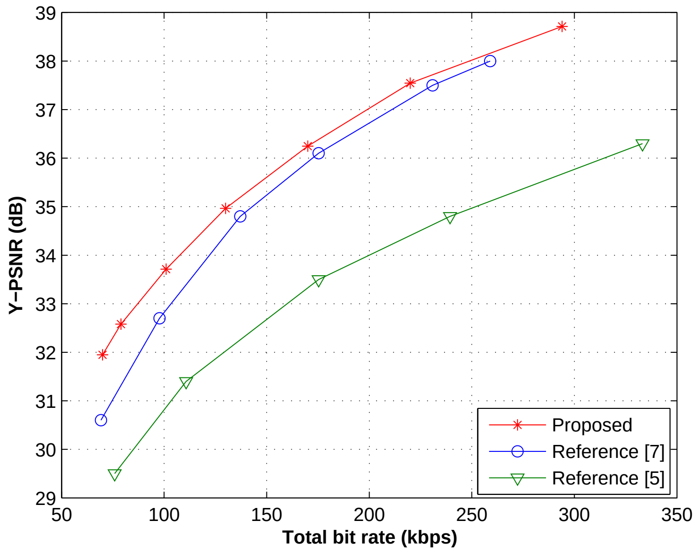
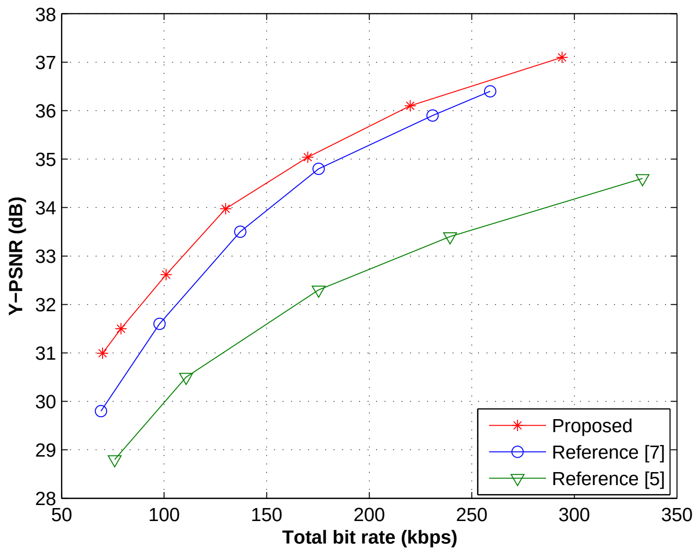


Fig. 9. The proposed scheme vs. the conventional one in packet loss network (test video: coastguard.qcif)

To present the advantage over the schemes using spatial subsampling, the proposed scheme is compared with [5] and [7] in Figure 10. For the fair comparison, the same code parameters of H.264 are also employed in the three compared schemes for the test video "foreman.qcif". From the Figure 10, we can find out better rate and central/side distortion performance achieved by the proposed scheme, especially at the lower bit rate.



(a)



(b)

Fig. 10. The proposed scheme vs. reference ones in on-off MD environment: a) rate/central distortion performance, b) rate/side distortion performance (test video: foreman.qcif)

4 CONCLUSION

The MD video coding scheme based on the pre- and post-processing has been developed in the paper without any modification of the source or channel codec. In view of the motion variety characteristic between frames, joint mode of frame duplication/interpolation in the pre-processing has been accommodated to match the post-processing for better side distortion in on-off channel environment and better robustness in packet loss network compared with the conventional MD video coding. Furthermore, the proposed MD system has also demonstrated superior rate-distortion performance to spatial subsampling-based schemes. Due to the perfect compatibility with the standard source and channel codec, the proposed MD video scheme may be a better choice in the practical applications.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China (No. 60903066, No. 60972085, No. 60776794, No. 90604032, No. 61073142), Beijing Natural Science Foundation (No. 4102049), New Teacher Foundation of State Education Ministry (No. 20090009120006), Beijing Lab of Intelligent Information Technology in Beijing Institute of Technology, 863 program (No. 2007AA01Z175), Sino-Singapore JRP (No. 2010DFA11010), Fundamental Research Funds for the Central Universities (No. 2009JBZ006-3) Research Foundation of BJTU (2008RC-027), Youth Science Technology Research Foundation of Shanxi Province (No. 2008-021020) and Doctor Startup Foundation of TYUST (No. 20092011).

REFERENCES

[1] Goyal, V. K.: Multiple Description Coding: Compression Meets the Network. IEEE Signal Processing Magazine, Vol. 18, 2001, No. 5, pp. 74–93.

[2] Vaishampayan, V.: Design of Multiple Description Scalar Quantizer. IEEE Trans. on Information Theory, Vol. 39, 1993, No. 3, pp. 834–921.

[3] Vaishampayan, V.—John, S.: Balanced Interframe Multiple Description Video Compression. International Conference on Image Processing, Kobe, Vol. 3, October 24–28, 1999, pp. 812–816.

[4] Reibman, A. R.—Jafarkhani, H.—Wang, Y.—Orchard, M. T.—Puri, R.: Multiple Description Coding for Video Using Motion Compensated Prediction. International Conference on Image Processing, Kobe, Vol. 3, October 24–28, 1999, pp. 812–816.

[5] Gallant, M.—Shirani, S.—Kossentini, F: Standard-Compliant Multiple Description Video Coding. International Conference on Image Processing, Tessaloniki, Vol. 1, October 7–10, 2001, pp. 946–949.

[6] Olmo, G.—Tillo, T.: Directional Multiple Description Scheme for Still Images. International Conference on Electronics, Circuits and Systems, Sharjah, Vol. 2, December 14–17, 2003, pp. 886–889.

[7] Wang, D.—Canagarajah, N.—Redmill, D.—Bull, D.: Multiple Description Video Coding Based on Zero Padding. International Symposium on Circuits and Systems, Vancouver, Vol. 2, May 23–26, 2004, pp. 205-208.

[8] Wang, Y.—Ostermann, J.—Zhang, Y.: Video Processing and Communication. Prentice Hall, NJ, 2002.

[9] H.264 standard. JVT-G050. 7$^{\text{th}}$ meeting, Pattaya, Thailand, March 7–14, 2003.

Huihui Bai received the Ph. D. degrees in signal and information processing from Beijing Jiaotong University (BJTU), Beijing, China, in 2008. She is currently a lectuer in Institute of Information Science in BJTU. She has been engaged in R & D work in video coding technologies and standards such as multiple description video coding (MDC) and distributed video coding (DVC). She is leading or participating in several research projects such as 973 Program, 863 Program, National Natural Science Foundation of China, Natural Science Foundation of Beijing, New Teacher Foundation of State Education Ministry, and Research Foundation of BJTU.

Yao Zhao received the Doctor degree in signal and information processing from Beijing Jiaotong University (BJTU), Beijing, China in 1996. He worked as a Post-Doctor in Delft University of Technology in Netherlands from 2001 to 2002. He is currently a Professor and Director in Institute of Information Science in BJTU. His research interests include image and video coding, digital watermark and content based video retrieval. He has published more than 130 papers on IEEE Transactions on CSVT, Electronics Letters, Signal Processing: Image Communication and so on. He has finished 25 national projects sponsored by NSFC, 973 program and 863 program.

Ce Zhu received the Ph. D. degrees from Southeast University, Nanjing, China in 1994 in electronic engineering. He is currently an Associate Professor in the School of Electrical & Electronic Engineering, Nanyang Technological University. As a Principle Investigator, He has successfully completed a few research and industrial projects on image/video coding and processing. He has published about 60 papers in refereed journals and conferences, and filed three patents (1 granted and 2 pending). He is currently a member of the Technical Committee on Multimedia Systems and Applications, IEEE Circuits and Systems Society, and a senior member of IEEE.

Anhong Wang received the Ph. D. degree in signal and information processing from Beijing Jiaotong University (BJTU), Beijing, China, in 2009. She became an Associate Professor with TYUST in 2005 and became a Professor in 2009. Her research interest includes image/video coding and compressed image/video sensor. She has published more than 30 papers in international journals and conferences. Now she is leading or participating in several national research projects from 863 Program, National Science Foundation of China and Youth Science Technology Research Foundation of Shanxi Province.