Science in China Series F: Information Sciences

www.scichina.com
info.scichina.com
www.springerlink.com

Robust multiple description distributed video coding using optimized zero-padding

WANG AnHong[1,2†], ZHAO Yao[1] & BAI HuiHui[1]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China;
[2] Taiyuan University of Science and Technology, Taiyuan 030024, China

Distributed video coding (DVC) arouses high interests due to its property of low-complexity encoding. This paper proposes a robust multiple description DVC (MDDVC) under the constraint of low-complexity encoding. In MDDVC, zeros are padded to each frame and the resulting big-size video is divided into multiple descriptions. Then, each description is compressed by a hybrid DVC (HDVC) codec and transmitted over different channel. When one channel does not work, the lost HDVC description is estimated by the received from other channel, which guarantees the robustness of the system; MDDVC moves the complex motion estimation totally to the decoder so it features low-complexity encoding. In the pre-processing, an optimized zero-padding is also proposed to improve the performance. Experimental results exhibit that the proposed MDDVC scheme achieves better rate-distortion performance and robustness than the referenced especially when packet-loss rate is high.

distributed video coding, robustness, multiple description coding, zero-padding

1 Introduction

The increasing demand for friendly up-link communication of low-power video captures has generated a lot of research interests in developing video codec of low-complexity encoding. As a new video coding framework, DVC[1] also called Wyner-Ziv video coding gives strong support to low-complexity encoding. DVC is built on 1970's two information theories, the Slepian-Wolf and Wyner-Ziv theories which state that almost the same rate-distortion performances may be achieved when the correlation of source is exploited at the decoder only. Based on these theories, DVC[1−3] shifts the motion estimation which exploits correlation of frames to the decoder. Since motion estimation is a high computational burden, DVC achieves low-complexity encoding compared with the conventional video coding algorithms, such as H.26x and MPEG series which exploit motion estimation at the encoder.

On the other hand, robust DVC methods are necessary especially when the video of low-power

Received May 20, 2008; accepted November 10, 2008
doi: 10.1007/s11432-009-0037-5
†Corresponding author (email: wah_ty@yahoo.com.cn)
Supported in part by the National Natural Science Foundation of China (Grant Nos. 60776794 and 90604032), the National Basic Research Program of China (973 program) (Grant No. 2006CB303104), the National High-tech Research & Development Program of China (863 program) (Grant No. 2007AA01Z175), Program for Changjiang Scholars and Innovative Research Team in University (Grant No. IRT0707), and Specialized Research Foundation of BJTU, Youth Science Technology Research Foundation of Shanxi Province (Grant No. 2008021020)

captures is transmitted over wireless network. DVC itself takes on inherent robustness because of the error-correcting channel decoding algorithm adopted. Ref. [2] proves the robustness of DVC through the experiments. But this robustness is achieved at the cost of compression efficiency. DVC assumes a correlation channel between the source to be encoded and its side information available at the decoder. In the view of principle of DVC, the compression efficiency comes from the correlation of side information, i.e., the more correlation means the higher compression efficiency, and vice versa. While in the case of high packet-loss rate, the correlation of side information becomes low, so the compression efficiency is strongly limited. In some other works related to robustness, Wyner-Ziv video coding is used as forward error correction (FEC) to protect the video transmission. For example, Girod et al.[1,4] provided a systematic lossy error protection (SLEP) based on Wyner-Ziv coding. The scheme is two-layer scalable in the sense of having one base layer with MPEG encoder and the corresponding Wyner-Ziv bits as the enhancement layer. However, SLEP scheme still applies motion estimation in its MPEG encoder so the property of low-complexity encoding cannot be guaranteed to some degree. And, error propagation in the MPEG-encoded stream may negatively impact the quality of the side information and degrade the robustness of the system especially when the packet-loss rate is high. To improve the robustness of SLEP, Crave[5] addressed a distributed multiple description coding, while still, the high-complexity encoding is preserved due to the motion compensation temporal filtering at the encoder.

MDC has emerged as an attractive framework for robust transmission over unreliable channels. MDC encodes the source message into several bit streams (called descriptions) carrying different but correlated information which then can be transmitted over the different channels. If some channels do not work, the lost descriptions can be estimated by the correlated descriptions received. MDC gives guarantee for robustness especially when the pack-loss rate is high due to its structure. Then, the robustness problem of DVC is addressed here by combining the multiple description coding. In this paper, we attempt to design a robust multiple description DVC under the constraints of low-complexity encoding.

Recently, the multiple description (MD) version of pre-/post-processing is a popular technique because it does not require the modification to the source codec. It is proven that MDC with pre-/post-processing is much better than directly sub-sampling the original frames. For instance, in ref. [6], the redundancy frames are added in the pre-processing based on the temporal correlation estimated by motion estimation. But the complexity of encoder is higher. In ref. [7], redundancy is added firstly by padding various numbers of zeros in DCT domain to each frame. Especially, 1-D DCT zero-padding is attractive for it is an easy but efficient way. Thus in this paper, we attempt to design the robust MDDVC based on pre-/post-processing with simple 1-D DCT zero-padding scheme. Typically, our scheme includes the following steps: first, in the pre-process stage, proper zeros are padded to each frame and a new sequence with big size is created. Next, MDDVC splits the sequence into two descriptions with each description compressed independently by HDVC algorithm of low-complexity encoding. At the decoder, the recovered video with good or acceptable quality is obtained by central or side decoder after post-processing stage. Here, note that, the compressed HDVC descriptions are transmitted over different channels; when one channel does not work, the lost HDVC description can be estimated by the received from other channels so the robustness is guaranteed even in the case of high packet-loss rate. Meanwhile, the proposed MDDVC avoids the complex motion estimation and its encoding computation is similar to the intra-frame coding, thus it has low-complexity encoding. In addition, althought Fan[8] and Wu et al.[9] have done some benefit work for MDC based on Slepian-Wolf to achieve better robustness, there is much work to do in robust DVC combining MDC under the constrain of low-complexity encoding.

2 Overview of the proposed MDDVC

Figure 1 illustrates our proposed scheme for MDDVC. In the pre-processing stage, the original video sequence is pre-processed to generate a new big-size video sequence using optimized 1-D DCT zero-padding. Then by means of interleaving-spliting, the new sequence is divided into two descriptions, which can be compressed by any DVC scheme. Here, a proposed HDVC scheme is employed for its efficient compression performance and low-complexity encoding. At the decoder, firstly, the central decoder merges the two decoded descriptions in case of no loss, or the side decoder uses error concealment method to estimate the lost description and create the concealed big-size video in loss case; and then in the post-processing, the inversed 1-D DCT zero padding is implemented to recover the pixels. The details of HDVC, pre-processing and post-processing are shown in the following two sections respectively.

3 HDVC scheme used in MDDVC

To each description, an independent DVC codec is used to its big-size sequence. In the view of low-complexity encoding and high rate-distortion performance, a new HDVC codec combining residual DVC[10] and Slepian-Wolf set partition in hierarchical trees (SW-SPIHT) algorithm[11] is proposed. The codec including HDVC encoder and decoder is shown in Figure 2.

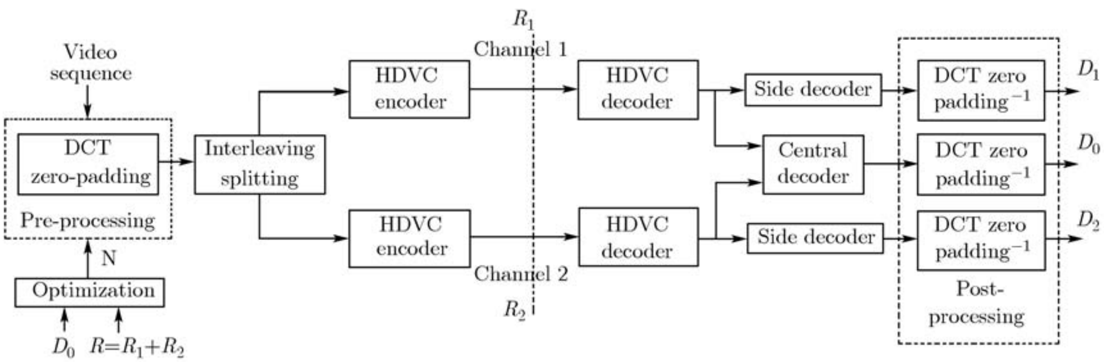

Figure 1 Framework of MDDVC proposed.

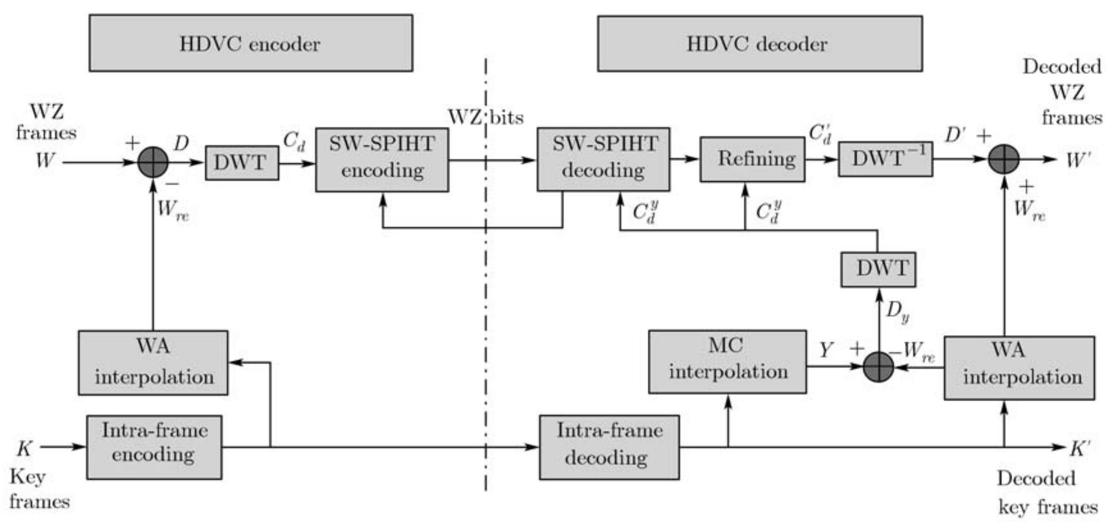

Figure 2 HDVC framework used in the MDDVC.

3.1 HDVC encoder

In each description, the frames are classified as the key frames and Wyner-Ziv frames according to the coding method adopted. To the key frames, the intra-frame encoding of H.264/AVC is used. While to the other frames, the Wyner-Ziv ones, the following processes are implemented.

Firstly, the weighted average (WA) interpolation is implemented to generate the referenced frame W_{re}

$$W_{re} = \alpha K_j^{'} + \beta K_{j+1}^{'}, \tag{1}$$

where $K_j^{'}$ and $K_{j+1}^{'}$ are the previous and next decoded key frames; $\alpha = 1 - l/g$, $\beta = 1 - \alpha$, where l is the distance between current frame and the previous key frame in the same group of picture (GOP), which has g frames totally.

Secondly, for each Wyner-Ziv frame W, the residual frame $D = W - W_{re}$ is decomposed by discrete wavelet transform (DWT) and the wavelet coefficient C_d is encoded by SW-SPIHT encoder. SW-SPIHT means that Slepian-Wolf codec is used to compress the SPIHT information of C_d, including the wavelet tree distribution information, the significant information, the sign information and the refinement information. The details of SW-SPIHT can be seen in ref. [12], where it is noted as "SPIHT with SI". The output of SW-SPIHT encoder forming the so-called Wyner-Ziv bits is sent to the decoder. The rate-adaptive low density parity coding accumulation (LDPCA)[13] with feedback is used for Slepian-Wolf codec. LDPCA with 396 nodes is adopted in our work. In the case of the sequence length less than 396, the padding-zero is applied.

Here, the encoding of HDVC mainly consists of DWT with SPIHT, LDPCA encoding, etc., which means that the complexity is similar to the intra-frame encoding so the low-complexity encoding is preserved. But note that, the simultaneous exploitation of temporal and spatial correlations by the residual coding and DWT with SPIHT helps HDVC achieve better rate-distortion performance than DCT-domain DVC, which can be seen in the experimental results in section 5. In addition, SPIHT-based HDVC can be easily extended to the scalable system.

3.2 HDVC decoder

In HDVC decoding, the key frames are recovered by the intra-frame H.264/AVC decoder. While to the Wyner-Ziv ones, the following processes are implemented.

Firstly, the decoded key frames are used to generate high quality side information based on the bidirectional motion compensation interpolation as in formula (2).

$$\begin{aligned} Y(x,y) = {} & \alpha \times K_j'(x + \beta \times dx_f, y + \beta \times dy_f) \\ & + \beta \times K_{j+1}'(x - \alpha \times dx_f, y - \alpha \times dy_f) \\ & + \alpha \times K_j'(x - \beta \times dx_b, y - \beta \times dy_b) \\ & + \beta \times K_{j+1}'(x + \alpha \times dx_b, \\ & \qquad y + \alpha \times dy_b), \end{aligned} \tag{2}$$

where (x, y) is the coordinates of the interpolated frame; $[dx_b, dy_b]$ and $[dx_f, dy_f]$ are the backward and forward motion vectors between the decoded key frames respectively, which may be obtained by the half-pixel motion estimation similar to ref. [3]; α, β and l are similar to that in formula (1).

Secondly, DWT is implemented to the difference $D_y = Y - W_{re}$ and the resulting coefficient C_d^y is used as the side information for SW-SPIHT decoding. The refinement module is to recover the wavelet coefficients with the side information

$$C_d' = \begin{cases} V_{\max}, & \text{if } C_d^y > V_{\max}, \\ C_d^y, & \text{if } C_d^y \in (V_{\min}, V_{\max}), \\ V_{\min}, & \text{if } C_d^y < V_{\min}, \end{cases} \tag{3}$$

where C_d' is the final wavelet coefficient value after refinement, $V_{\max}$ and $V_{\min}$ are the possible maximal and minimal value of wavelet coefficient if SPIHT decoding is implemented to all bit-planes.

Thirdly, after the inversed DWT, the recovered difference D' is achieved. Finally, the referenced W_{re} is added to restore the original pixels.

4 Pre-/post-processing scheme with optimized zero-padding in MDDVC

MDC with pre-/post-processing is much better than directly sub-sampling the original frames. Especially, pre-/post-processing with DCT zero-padding technique is based on the theory of zero padding which shows padding zeros in frequency domain will result in the interpolation in time domain[7]. The interpolated frame is easy to be compressed because the correlation between pixels becomes strong although it increases the frame size to be coded[7]. The concrete process of pre-/post-processing with DCT zero-padding is shown in the following.

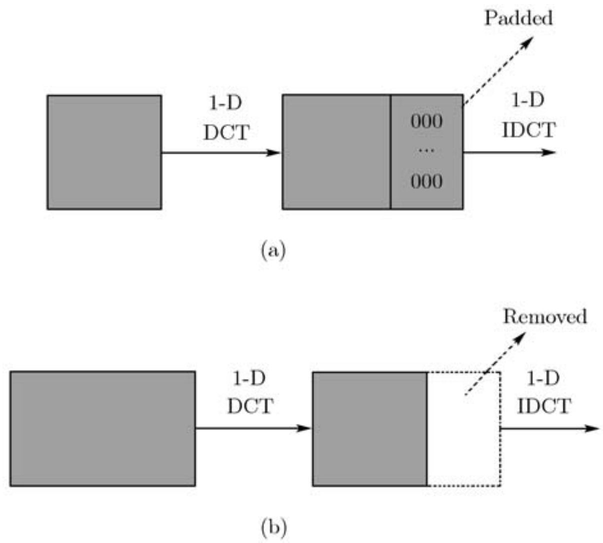

Figure 3 To illustrate (a) 1-D DCT zero-padding and (b) its reversed process.

(i) Pre-processing. 1-D DCT zero-padding, which has lower complexity and better rate-performance than 2-D DCT zero-padding, is illustrated in Figure 3(a). Firstly, each frame is transformed using 1-D DCT on each row, then padded with zeros horizontally. After 1-D IDCT on new-sized row, we obtain the enlarged frame which is sub-sampled into two sub-frames by interleaving and independently coded by HDVC codec explained above.

(ii) Post-processing. After MDDVC central decoder or side decoder, two decoded sub-frames are obtained. Then the two sub-frames are merged to get the big size frames. Next, the inversed 1-D DCT padding illustrated in Figure 3(b) is implemented to recover the original pixels, i.e., the big size frame is firstly transformed using 1-D DCT on each row, then padded zeros are removed and IDCT is implemented to get the frame needed.

(iii) Optimized zero-padding. In the MDDVC with zero-padding, the number of zero padded will affect the correlation as well as the rate-distortion performance of side and central decoder. Generally, when more zeros are padded, correlation between two descriptions will be higher producing better estimation and better quality from side decoder, while the central quality drops with the increasing of zeros. That is, more zero-padding takes benefit to the compression in two aspects, one is the lost description is estimated accurately, the other is that the frames can be compressed easily due to the increased correlation between adjacent pixels. While on the other hand, more zero-padding makes the size of frame increase, which maybe require more bits to represent it, and also, when no loss happens, the more zero-padding means more redundancy added so the central quality will decrease. The aforementioned analysis requires an optimization for the number of zeros padded.

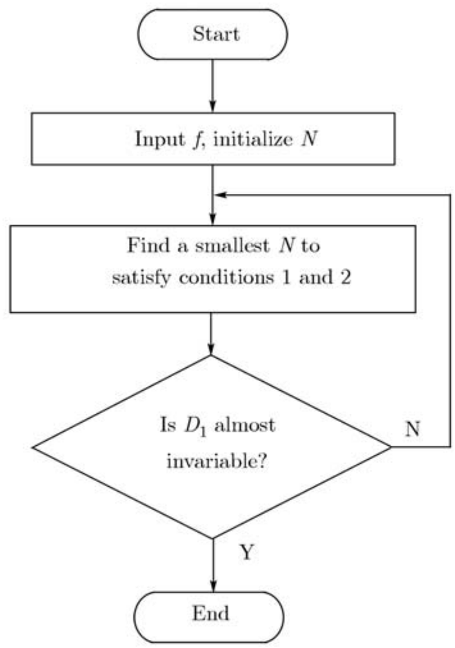

Figure 4 Zero-padding optimization process.

Let $D_0(f,N)$ and $D_1(f,N)$ (or $D_2(f,N)$) denote the mean squared errors (MSE) from the central and side decoder for the input image f, respectively, given the number of zero padded is N. Let $R(f,N)$ be the bit rate for two descriptions, while $R_1(f,N)$ and $R_2(f,N)$ be the bit rates for the two balanced descriptions 1 and 2, respectively. Our goal is to find the optimal parameter N in solving the following optimization problem:

$$\min_N D_1(f,N), \tag{4}$$

subject to

condition 1:

$$R(f,N) = 2R_1(f,N) = 2R_2(f,N) \leqslant R_{\text{budget}}; \quad (5)$$

condition 2:

$$D_0(f,N) \leqslant D_{\text{budget}}, \quad (6)$$

where R_{budget} is the available total bit rate to encode two descriptions and D_{budget} is the maximum distortion acceptable for central decoder reconstruction. The encoding optimization module in Figure 1 is based on the above function. With the constrain on the total bit rate and the central distortion, N is adjusted accordingly to minimize the side distortion.

The optimization for the problem is carried out in an iterative way. The basic algorithm shown in Figure 4 is to make use of the monotonicity of R and D as the function of N. After initialization a smallest N is searched to minimize D_1 subject to condition 1 and condition 2.

5 Experimental results and analyses

Here, there are mainly three groups of experiments taken into account to present the efficiency of MDDVC proposed. They are performance comparison of different coding methods, performance comparison with optimized zero-padding and robustness comparison over packet-loss channel. In all of these experiments, the luminance components of standard sequences Foreman and Hall in QCIF@15 fps (frame per second) are tested. The total bit rate of two channels is accounted.

5.1 Comparison of different DVC methods

The performance of single description DVC (SD DVC) used will affect the performance of MDDVC proposed, so firstly we compare our proposed SD HDVC with other SD DVCs referenced, including DVC[3] and residual DVC[10]. The GOP varies in 2, 4 and 8 to show the performance comparison.

Comparison curves in Figure 5 show that the proposed SD HDVC obtains up to 3 dB improvement for Hall sequence. There is even up to 3.9 dB improvement over residual DVC in ref. [10]. These improvements originate from the fact that the hybrid scheme can make use of some temporal and spatial correlations simultaneously at the encoder. For Foreman sequence, the improvement can also be achieved though it is not as valuable as the Hall sequence due to the low temporal correlation.

Then, we plot the performance of the proposed MDDVC in Figures 6 and 7, where the MD channels are assumed to ideal, in which one channel is either intact or totally lost, i.e., the information loss rate is 0 or 50%. The results show that the number of zeros added will affect the results of MDDVC, and even show that more zeros padded are not always helpful for the side decoder of MDDVC, such as in Hall sequence with 176 zeros. So an optimized zero-padding is necessary for introducing the proper number of zeros in the pre-processing stage.

5.2 Comparison of optimized zero-padding

This experiment is to test the efficiency of the op-

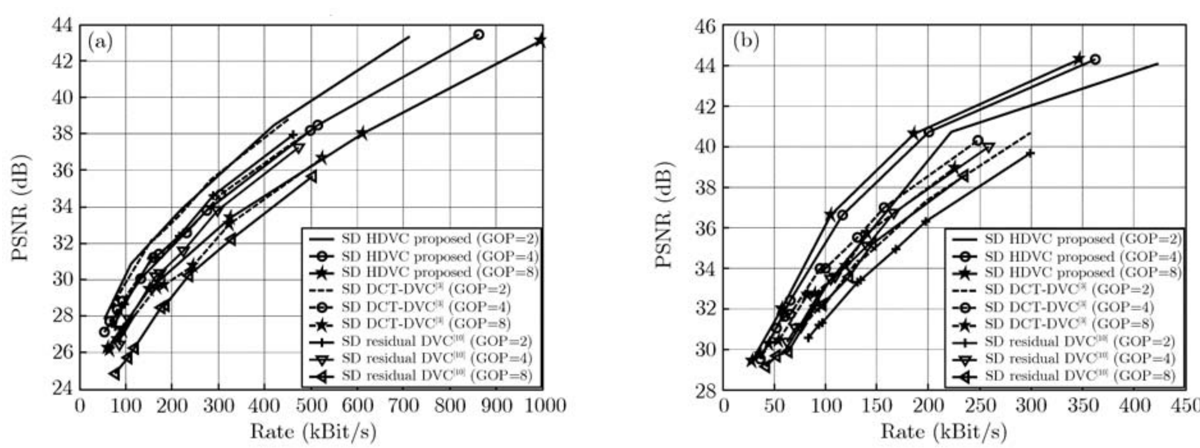

Figure 5 Comparison of different SD DVC. (a) Foreman sequence; (b) Hall sequence.

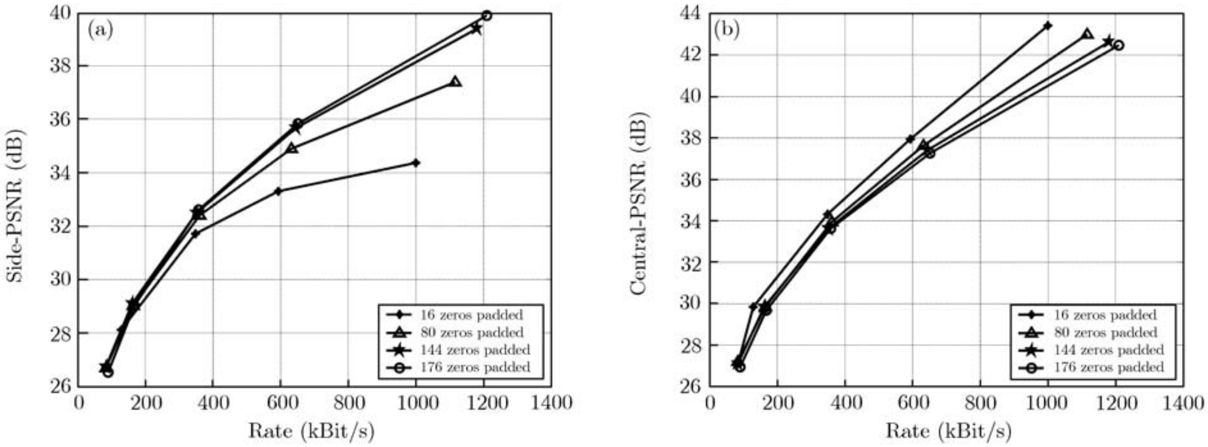

Figure 6 Performance of different zero-padding for Foreman sequence. (a) Side rate-distortion; (b) central rate-distortion.

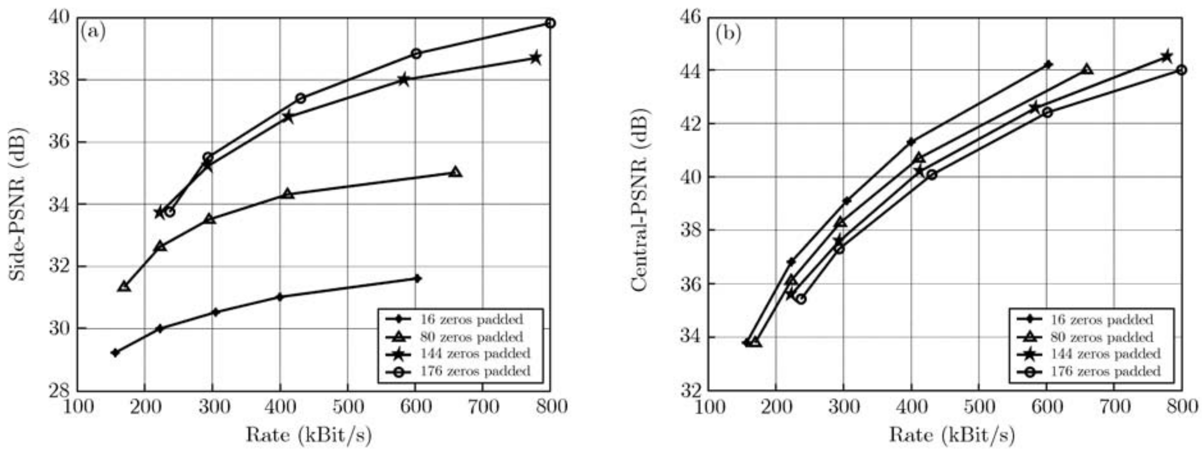

Figure 7 Performance of different zero-padding for Hall sequence. (a) Side rate-distortion; (b) center rate-distortion.

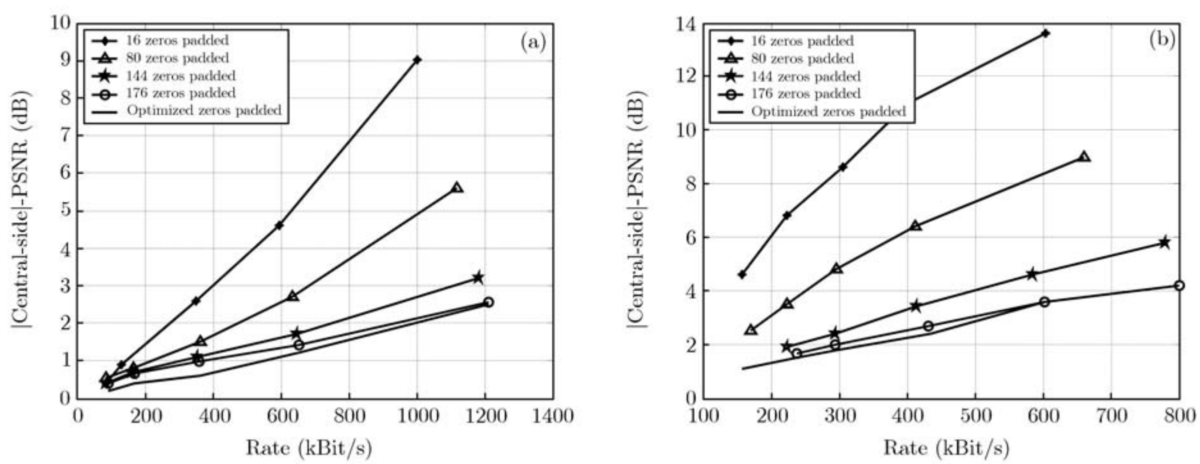

Figure 8 Comparison of optimized zero-padding. (a) Foreman sequence; (b) Hall sequence.

timized zero-padding in MDDVC proposed. The comparison is based on the fact that the difference between the central and side quality illustrates the optimization performance of MDDVC when the same central quality and same bit rate are constrained during changing the number of zeros padded, that is, the smaller the difference is, the better the optimization is, and vice verse.

The optimization is implemented to each frame with the searching range variable from 0 to 176 zeros padded. The performance comparison is shown in Figure 8. Obviously, the optimized zero-padding achieves the smallest difference between the central and side quality due to its tradeoff consideration of side and central quality. And it also shows that the less zeros should be padded in the case of low bit rate.

5.3 Robustness comparison over packet-loss channel

The two above experiments test the performance when half of the information is lost, while in practice, the loss is not so serious, so this experiment will divide the frames into packets and test the performance when partial packets are lost.

This experiment compares two methods: MDDVC proposed in the paper, SLEP based on ref. [1] but with H.264 and SW-SPIHT codec instead of the MPEG and SW-DCT. In case of SLEP, each frame is divided into two sub-frames by interleaving and each sub-frame is compressed and packed independently. The packet loss pattern is shown in Figure 9, which assumes that the two sub-frames in the same frame are not lost simultaneously. At the decoder, the lost sub-frames are estimated by average interpolation, and then the estimated values are corrected by the SW-SPIHT bitstream. In the case of MDDVC, it is just each big-size frame that is divided into two sub-frames and packed independently. While, the packet loss pattern is the same as that of SLEP mentioned above. GOP is set to 2, 4 and 8 for SLEP with “I-B-B-B” order and for HDVC with “K-W-W-W”.

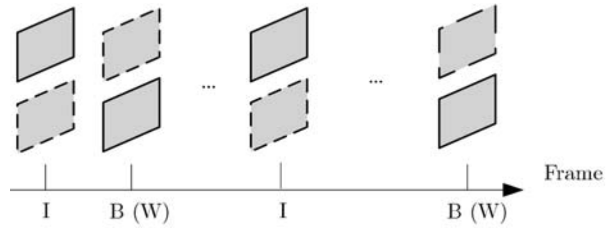

Figure 9 Packet loss pattern.

Figure 10 shows the results of reconstruction quality with different packet-loss rates when total bit rate is 1100 kBit/s for Foreman and 610 kBit/s for Hall. It is clear that our scheme has evidently better robustness than SLEP when loss rate is high. While, unlike SLEP, MDDVC does not performance better with the increasing of GOP especially for the high motion sequence Foreman. This is because the side information becomes bad in big GOP, which affects the performance of SDDVC and MDDVC. While note that, SLEP performs good when the GOP value is big at the cost of the increasing of encoding computation.

6 Conclusion

A new robust MDDVC scheme combining the distributed video coding (DVC) and multiple description coding (MDC) is proposed. The encoding of

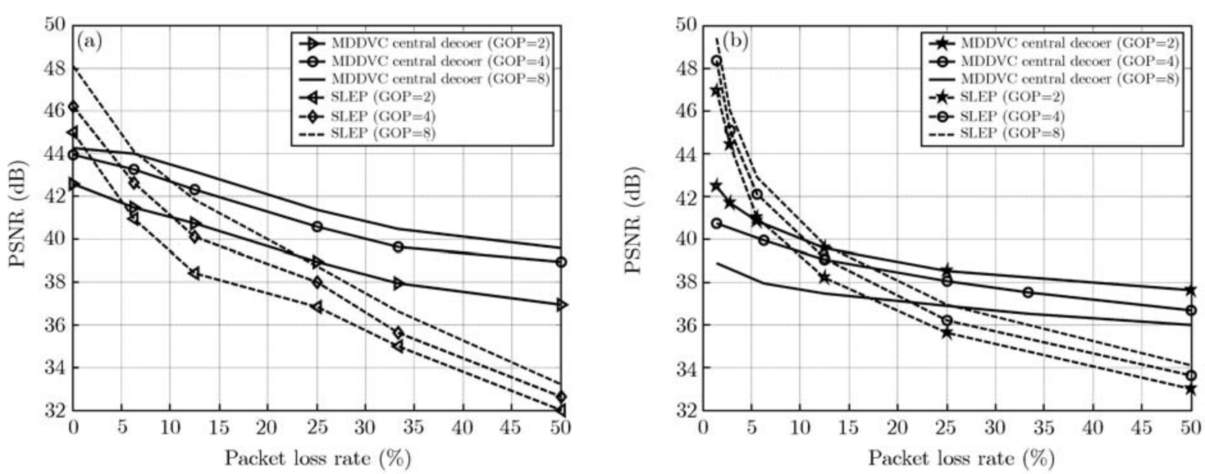

Figure 10 Robustness comparison. (a) Hall sequence; (b)Foreman sequence.

MDDVC mainly consists of 1-D DCT zero-padding, DWT, SPIHT, and LDPCA encoding, etc. Thus, the proposed scheme avoids the complex motion estimation and preserves the low-complexity encoding similar to the intra-frame coding. MDDVC guarantees its robustness in view of structure, that is, when one channel does not work, the lost HDVC description is efficiently estimated by the received from other channel. The proposed scheme is promising in the video communications of low-power devices over network, such as wireless sensor network, mobile camera, and so on.

1 Girod B, Aaron A, Rane S, et al. Distributed video coding. Proceed IEEE, 2005, 93(1): 71–83

2 Puri R, Ramchandran K. PRISM: an uplink-friendly multimedia coding paradigm. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing, 2003, 4: 856–859

3 Artigas X, Ascenso J, Dalai M, et al. The DISCOVER codec: architecture, techniques and evaluation. In: Proceedings of Picture Coding Symposium, 2007. 1950–1953

4 Rane S, Aaron A, Girod B. Systematic lossy forward error protection for error-resilient digital video broadcasting——a Wyner-Ziv coding approach. In: Proceedings of International Conference on Image Processing, 2004. 609–612

5 Crave O, Guillemot C, Pesquet-Popescu B, et al. Distributed temporal multiple description coding for robust video transmission. EURASIP J Wireless Commun Netw, 2008. Article ID: 183536

6 Bai H, Zhao Y, Zhu C. Multiple description video coding using adaptive temporal subsampling. In: Proceedings of International Conference on Multimedia & Expo, 2007. 1331–1334

7 Wang D, Canagarajah N, Redmill D, et al. Multiple description video coding based on zero padding. In: Proceedings of International Symposium on Circuits and Systems, 2004. 205–208

8 Fan Y H, Wang J, Sun J, et al. A novel multiple description video codec based on Slepian-Wolf coding. In: Proceedings of IEEE Data Compression Conference, 2008. 515

9 Wu M, Vetro A, Chen C W. Multiple description image coding with distributed source coding and side information. In: Proceedings of SPIE Multimedia System and Application, 2004, VII: 3–10

10 Aaron A, Varodayan D, Girod B. Wyner-Ziv residual coding of video. In: Proceedings of Picture Coding Symposium, 2006. 2–7

11 Tang C, Cheung N, Ortega A, et al. Efficient inter-band prediction and wavelet based compression for hyerspectral imagery: a distributed source coding approach. In: Proceedings of IEEE Data Compression Conference, 2005. 437–446

12 Wang A, Zhao Y. Efficient wavelet zero-tree video coding based on Wyner-Ziv coding and lattice vector quantization. In: Proceedings of International Conference on Innovative Computing, Information and Control, 2006. 245–258

13 Varodayan D, Aaron A, Girod B. Rate-adaptive distributed source coding using low-density parity-check codes. In: Proceedings of Asilomar Conference on Signals, Systems and Computers, 2005. 1–8