SCIENCE CHINA
Information Sciences

Multiple description video coding using correlation optimized temporal sampling

HuiHui BAI[1,*], MengMeng ZHANG[2], AnHong WANG[3] and Yao ZHAO[1]

Citation: SCIENCE CHINA Information Sciences **57**, 052113 (2014); doi: 10.1007/s11432-013-4861-2

View online: https://engine.scichina.com/doi/10.1007/s11432-013-4861-2

View Table of Contents: https://engine.scichina.com/publisher/scp/journal/SCIS/57/5

Published by the Science China Press

SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

May 2014, Vol. 57 052113:1–052113:10
doi: 10.1007/s11432-013-4861-2

Multiple description video coding using correlation optimized temporal sampling

BAI HuiHui[1]*, ZHANG MengMeng[2], WANG AnHong[3] & ZHAO Yao[1]

[1] *Institute of Information Science, Beijing Jiaotong University, Beijing* 100044*, China;*
[2] *North China University of Technology, Beijing* 100144*, China;*
[3] *Taiyuan University of Science and Technology, Taiyuan* 030024*, China*

Received July 5, 2013; accepted September 13, 2013; published online October 30, 2013

Abstract We propose a novel multiple description video coding scheme based on pre- and post-processing of video sequences without modifying the actual coding process itself, thus making it compatible with the current standard coding. In pre-processing, adaptive modes of temporal sampling are employed to regulate the motion change between frames before performing odd/even frame splitting, which facilitates good estimation of lost frames in post-processing after side decoding. Significantly, given a central distortion, rate-distortion optimization with Lagrangian formulation is applied in pre-processing to achieve a good tradeoff between bit rate and side distortion. Furthermore, to simplify Lagrangian optimization, a new optimization criterion is proposed based on the correlation of intra- and inter-descriptions. In the experiments, the proposed scheme achieves better performance compared with other tested multiple description coding (MDC) schemes in both an on-off MDC environment and lossy packet networks.

Keywords video coding, multiple description coding, temporal sampling, rate-distortion optimization, correlation optimization

Citation Bai H H, Zhang M M, Wang A H, et al. Multiple description video coding using correlation optimized temporal sampling. Sci China Inf Sci, 2014, 57: 052113(10), doi: 10.1007/s11432-013-4861-2

1 Introduction

Multiple description coding (MDC) is a promising framework for solving compressed video transmission over non-prioritized and unreliable networks [1]. The original video signal can be split into multiple bit streams (descriptions) using a multiple description (MD) encoder. These multiple descriptions can then be transmitted over multiple channels. Since there is a very small probability that all channels will fail at the same time, only one description is needed by the MD decoder to reconstruct the video with acceptable quality (measured by side distortion) and more descriptions can produce the video with better quality (measured by central distortion) [2]. Here, we focus mainly on an MD design for two channels.

Over the past few years, several methods for MDC have been presented, such as the MD scalar quantizer [3], MD lattice vector quantizer [4,5], MD based on pairwise correlating transforms [6] and MD based on FEC [7]. Although all these methods have claimed to achieve good performance, they are difficult to

*Corresponding author (email: hhbai@bjtu.edu.cn)

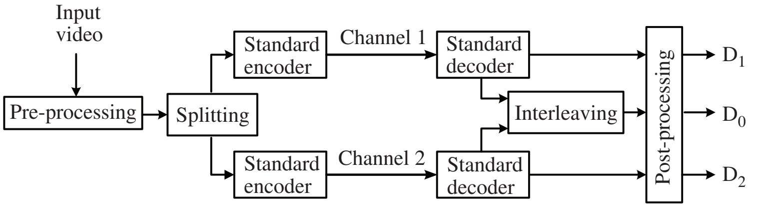


Figure 1 Block diagram of a typical MD video coding with pre- and post-processing.

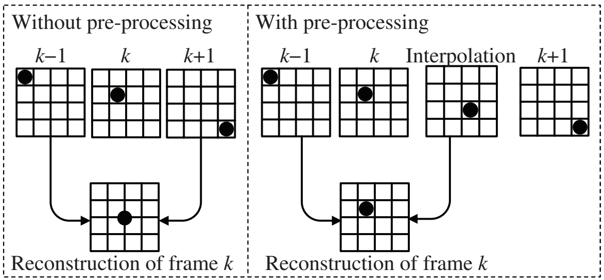


Figure 2 Side reconstruction without/with pre-processing.

apply in practical applications because the specially designed MD encoders are not compatible with the widely-used standard codecs, such as H.26x and MPEG-x. The authors in [8] set out to generate two descriptions using standard encoders, but the compression efficiency needs to be improved.

To address this problem, pre- and post-processing based MDC may be a good choice. In Figure 1, a video sequence is first processed and then it is split into two sub-sequences and passed through any standard encoder. At the receiver, post-processing is performed for error concealment. Depending on the pre-processing, there are two types of sampling based MDC, namely, spatial sampling [9,10] and temporal sampling [11,12]. In the pre-processing reported in [9], zero padding is proposed to add spatial redundancy to each original frame and then spatial sampling can be applied to produce two descriptions. Furthermore, in [10] optimized zero-padding improves the performance of multiple description coding. Coversely, a temporal sampling version is proposed in [11]. Redundant frames are up-sampled in the temporal domain to regulate motion change so that smooth motion in the video can be utilized in post-processing to estimate lost frames. Furthermore, a preliminary MDC scheme based on frame duplication and interpolation is proposed in [12] to improve the rate-distortion performance reported in [11].

In this paper we attempt to exploit the intra- and inter-description correlation to design an optimized MD video coder based on pre- and post-processing. In pre-processing, using the appropriate mode of frame duplication/interpolation for temporal sampling, motion change between frames can be regulated before performing odd/even frame splitting, which facilitates good estimation of lost frames in post-processing after side decoding. In addition, to avoid the complexity of rate-distortion optimization with the Lagrangian formulation, we propose a new optimization criterion based on intra- and inter-description correlation to perform mode selection for temporal sampling.

2 Proposed MDC scheme

2.1 Pre-processing

In the conventional MDC scheme based on temporal sampling, odd and even frames can be regarded as two descriptions without any pre-processing. Although such a method is easy to perform, temporal correlation is compromised, which may lead to poor estimation of lost frames in side reconstruction, especially for frames containing irregular motion. In Figure 2, a simple example shows that pre-processing has a significant impact on side reconstruction, where irregular motion of an object (indicated by the black circle) is exhibited between frame $k-1$ and frame $k+1$. With the simple odd/even frame splitting

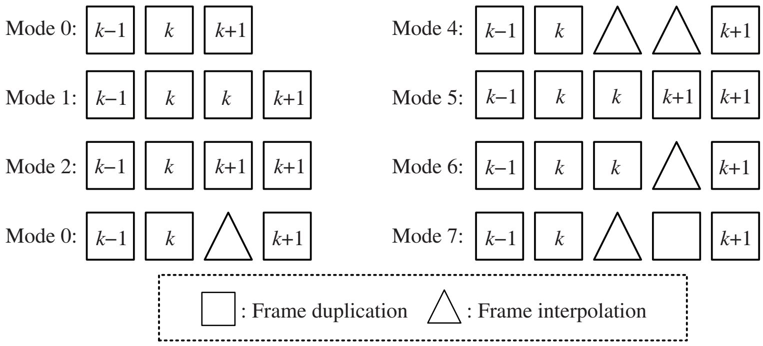


Figure 3 Eight modes of frame duplication/interpolation.

scheme, frame k must be estimated in side reconstruction. Assume that a simple method, such as motion-compensated interpolation (MCI) [12], is applied to estimate the lost frame k in side reconstruction. The reconstructed frame may suffer substantial distortion owing to the wrong displacement of the object as shown in Figure 2. Here, if an interpolated frame is inserted between frame k and frame $k+1$ during the pre-processing, the frame can be better estimated using the frame $k-1$ and the inserted frame, as shown in Figure 2. However, for more complex irregular motion, additional interpolated frames may be needed, which could produce a great deal of redundancy. In this case frame duplication may be a better option. In [12] we proposed a preliminary scheme to address the problem, where one interpolated frame or two duplicated frames were used to regulate motion change for post-processing based on MCI.

Compared with the two simple modes of redundancy allocation in [12], this paper presents an improved scheme using eight different modes, distinguished by how many (one or two) and how the redundant frames are inserted. To realize adaptive temporal sampling, two issues need to be addressed; the first relates to checking motion regularity while the other involves determining how to perform frame duplication or interpolation in the event of irregular motion.

Here we adopt a simple yet effective method for checking motion regularity. The motion between adjacent frames is determined using the magnitude of the maximal motion vectors in the frames. For any two adjacent frames, the motion vector for each macroblock is computed and the maximal motion vector $(\mathrm{MV}_x, \mathrm{MV}_y)$ can be found based on the magnitude of $\|\mathrm{MV}\| = \sqrt{\mathrm{MV}_x^2 + \mathrm{MV}_y^2}$. A greater difference between the two maximum motion vector magnitudes in adjacent frames means that irregular motion exists, whereas a smaller difference implies more uniform motion. Here we consider at most two redundant frames inserted between two original frames, which can normally handle the non-uniform motion problem. Accordingly, two thresholds, T_1 and T_2 are used to determine the extent of irregular motion. Note that T_1 is adaptively determined using the mean value of the variety of motion, that is, for an original video with N frames,

$$T_1 = \frac{1}{N-2} \sum_{k=1}^{N-2} \left| \|\mathrm{MV}\|_{k+1} - \|\mathrm{MV}\|_k \right|, \tag{1}$$

and T_2 is chosen as $T_2 = 2T_1$ for simplicity. Since T_1 is evaluated using all the frames in the video sequence, the proposed scheme cannot be used in real-time applications. More suitable applications thereof can be found in video-on-demand (VOD) systems.

Depending on whether one or two redundant frames are inserted, different combinations of frame duplication and frame interpolation are used. For a better estimation of frame k as shown in Figure 2, eight possible modes, shown in Figure 3, have been designed.

If $\left| \|\mathrm{MV}\|_k - \|\mathrm{MV}\|_{k+1} \right| < T_1$, the motion is considered to be smooth enough for the MCI based post-processing to obtain a good reconstruction of the lost odd/even frames in the side decoder. In this case, which we call mode 0, no redundant frames need to be inserted.

If $T_1 \leqslant \left| \|\mathrm{MV}\|_k - \|\mathrm{MV}\|_{k+1} \right| < T_2$, there is a small degree of irregular motion and a single additional frame is inserted to regulate the motion using one of three modes shown in Figure 3, that is, duplication of one frame (mode 1 or mode 2) or interpolation of one frame (mode 3).

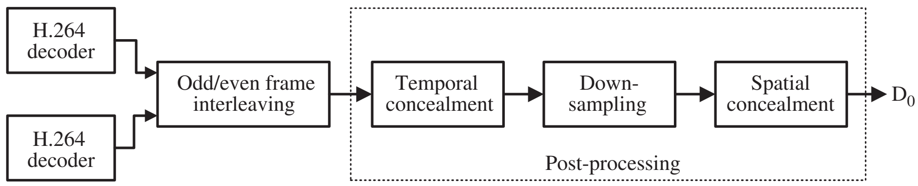


Figure 4 Block diagram of central decoder for lossy packet networks.

If $\left|\|\mathrm{MV}\|_k - \|\mathrm{MV}\|_{k+1}\right| \geqslant T_2$, there may be some larger irregular motion and two extra frames are needed to regulate the motion, corresponding to the four modes shown in Figure 3; that is, interpolation of two frames (mode 4), duplication of two frames (mode 5), duplication of one frame followed by interpolation of one frame (mode 6) or interpolation of one frame followed by duplication of one frame (mode 7).

It should be noted that a label with two bits is needed for each frame to denote the frame style (original, duplicated or interpolated). Interpolated frames can be generated using the conventional method for MCI presented in [12]. Although extra frames generated by interpolation or duplication during pre-processing are helpful for side reconstruction, they increase the size of the original video, which in turn leads to a higher coding bit rate. Therefore, for a good tradeoff of rate and side distortion, the interpolation/duplication modes during pre-processing should be selected based on rate distortion optimization, which is discussed in detail in Section 3.

2.2 Post-processing

Here different decoder designs are needed to satisfy the different environments for MDC.

1) Decoder design for an on-off channel environment: Based on the pre-processing described above, we discuss two aspects of post-processing for video reconstruction: central reconstruction and side reconstruction. For central reconstruction, two video sub-sequences can initially be organized after standard decoding. Then according to the labels, redundant frames with the same labels are exploited by averaging both representations to obtain a better reconstruction. If only one channel works, the positions of the lost frames can be obtained from the labels. Finally, we use motion-compensated interpolation (MCI) to estimate the lost frames [12].

2) Decoder design for lossy packet networks: Since packet losses occur in each of the descriptions, only a central decoder should be designed in post-processing. This differs from the approach in an on-off MDC environment.

The generated descriptions have temporal correlations between frames and spatial correlations within each frame. Therefore, in view of exploring these correlations, joint temporal-spatial error concealment is proposed as show in Figure 4. After odd/even frame interleaving, temporal concealment is used to reconstruct the lost packets. Since redundant frames exist, more motion information can be preserved than in the conventional scheme, which in turn results in better reconstruction quality. Here, two types of lost information need to be estimated for a damaged macroblock, specifically the motion vector and the pixel values. Note that at the encoder packets in each frame are organized like a checkerboard, that is, a macroblock is not grouped in the same packet as its surrounding ones. Owing 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 neighboring macroblocks. Then, we can use the macroblock in the previous frame pointed to by the generated motion vector for motion compensation. After temporal concealment, the redundant frames are down-sampled.

Nevertheless, with an increased packet loss rate, there are still errors that are difficult to conceal using only motion compensation. For example, owing to the greater packet loss rate, not only the current macroblock, but also the motion-compensated ones in the previous frames may be damaged. To address this problem, spatial concealment is employed to improve the effect of temporal concealment.

3 Optimization of mode selection

3.1 Correlation optimization

As mentioned in Section 2, different combination modes of frame duplication/interpolation have been designed to regulate motion change in the original video for better side reconstruction. However, the bit rate increases owing to the extra frames inserted. Mode optimization aims to select the mode that achieves,

$$\min(D_S), \quad \text{subject to } R \leqslant R_T, D_C \leqslant D_T, \tag{2}$$

where R_T is the target (total) bit rate for two descriptions and D_T is the maximum acceptable distortion for the central reconstruction. D_S is the side distortion with respect to rate R and central distortion D_C. The optimization problem can be solved as a Lagrangian formulation by

$$\min(J), \quad \text{with } J = D_S + \lambda_1 R + \lambda_2 D_C. \tag{3}$$

Here, the Lagrangian cost function J is minimized for particular values of the Lagrangian multipliers λ_1 and λ_2. Each solution of (3) for a given pair of λ_1 and λ_2 corresponds to an optimal solution of (2) for particular values of R_T and D_T. To solve the optimization problem, the Lagrangian multipliers, λ_1 and λ_2, must be determined, which is a difficult problem in multiple description coding. Furthermore, given λ_1 and λ_2, a triple (D_S, R, D_C) needs to be computed for each mode selection. This requires actual time-consuming encoding and decoding for each mode.

To simplify mode optimization, we consider a new optimization criterion based on the intra- and inter-description correlation, instead of rate distortion optimization of the triple (D_S, R, D_C). First, we give the formulas for intra- and inter-description correlation. The intra-description correlation coefficient (denoted by ρ_{intra}) is dependent on the temporal motion-compensated correlation within the odd or even description. Therefore, ρ_{intra} between the current frame f_k and its preceding frame f_{k-1} can be computed as

$$\rho_{\text{intra}}(f_k, f_{k-1}) = \frac{\text{Cov}(f_k, f'_{k-1})}{\sqrt{D(f_k)D(f'_{k-1})}}. \tag{4}$$

Here f'_{k-1} is the motion-compensated frame for f_k. The covariance of f_{k-1} and f'_{k-1} is $\text{Cov}(f_k, f'_{k-1})$ and their variances can be denoted by $D(f_k)$ and $D(f'_{k-1})$, respectively. The inter-description correlation coefficient (denoted by ρ_{inter}) can be computed as the correlation between the estimated frame f^*_k and the original frame f_k. Note that here f^*_k is generated using the same MCI method as in post-processing. In the same way, we obtain

$$\rho_{\text{inter}}(f_k, f^*_k) = \frac{\text{Cov}(f_k, f^*_k)}{\sqrt{D(f_k)D(f^*_k)}}. \tag{5}$$

Next we explain the corresponding relationship between $(\rho_{\text{inter}}, \rho_{\text{intra}})$ and (D_S, R, D_C). According to the formula for ρ_{intra}, a higher ρ_{intra} normally leads to higher compression, which results in a lower bit rate for a given central distortion (or smaller central distortion for a given bit rate). Consequently, bit rate and central distortion can be approximately translated as the intra-description correlation. According to (5), since ρ_{inter} is computed from the estimated and original frames, a higher ρ_{inter} is useful for estimating lost frames for a better side reconstruction. Therefore, we consider side distortion as the inter-description correlation. With the two correlation coefficients, we can convert the optimization given in (3) to

$$\max(J'), \quad \text{with } J' = \rho_{\text{inter}} + \lambda \cdot \rho_{\text{intra}}, \tag{6}$$

where, as mentioned above, ρ_{inter} and ρ_{intra} are used to measure, respectively, the side distortion D_S, and the bit rate R and central distortion D_C. Then the optimized mode of frame duplication/interpolation will be selected according to (6).

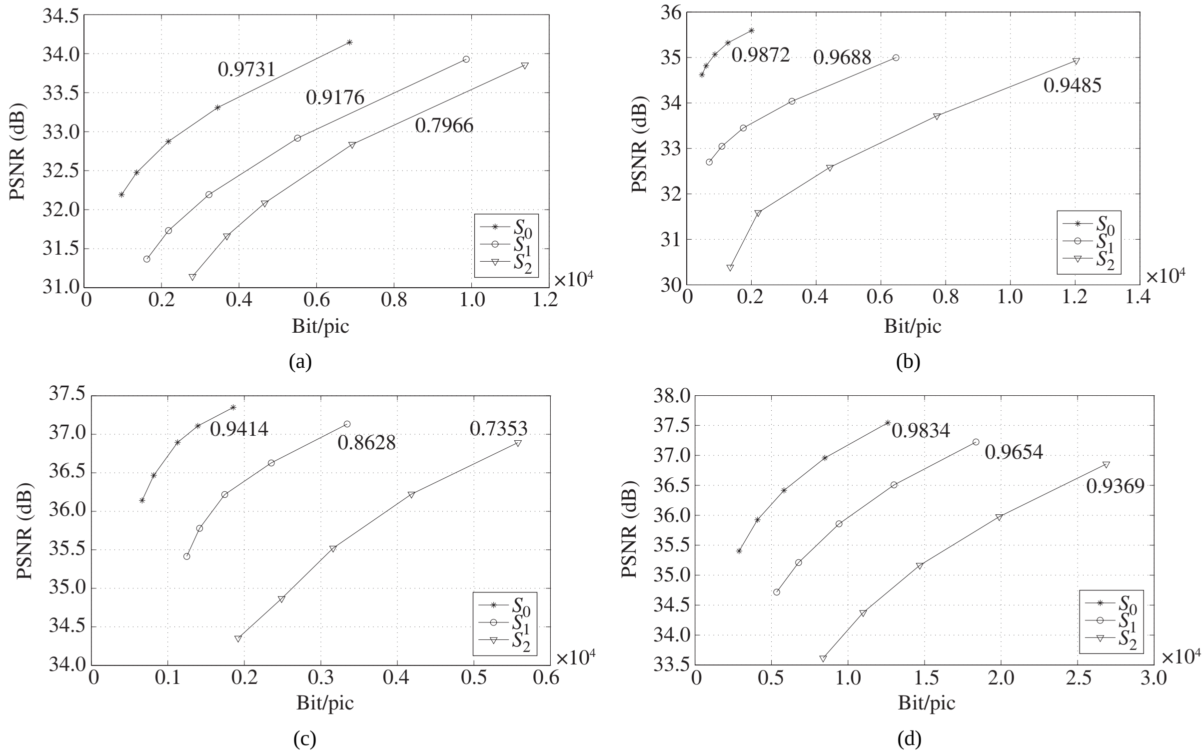


Figure 5 Rate distortion performance using different intra-description correlations. (a) Mobile.qcif; (b) Coastguard.qcif; (c) Foreman.qcif; (d) Paris.cif.

3.2 Relationship between $(\rho_{\text{intra}}, \rho_{\text{inter}})$ and (R, D_C, D_S)

To substantiate the corresponding relationship between $(\rho_{\text{inter}}, \rho_{\text{intra}})$ and (D_S, R, D_C), four standard video sequences were chosen for the experiments, namely, *Mobile.qcif*, *Coastguard.qcif*, *Foreman.qcif* and *Paris.cif*.

First, we carried out an experiment to substantiate the relationship between ρ_{intra} and R (or D_C). For each video sequence (denoted by S_0), two corresponding down-sampled sub-sequences (denoted by S_1 and S_2) were constructed, for example, $S_0 = \{f_1, f_2, \ldots, f_k\}$, $S_1 = \{f_1, f_3, \ldots, f_{2m-1}\}$ and $S_2 = \{f_1, f_5, \ldots, f_{4n-3}\}$. It should be noted that k, m and n are positive integers with $2m-1 < k$ and $4n-3 < k$. Then, the same frame, e.g., f_5 , existing in S_0, S_1, and S_2 was tested for its rate distortion performance and intra-description correlation. For a fair comparison, the frames preceding f_5, that is, f_4, f_3, and f_1 in S_0, S_1, and S_2 respectively, were encoded as I frames and f_5 as a P frame with the same parameters as in the H.264 encoder [13]. Furthermore, the intra-description correlation coefficients of f_5 in S_0, S_1, and S_2 were computed using $\rho_{\text{intra}}(f_5, f_4)$, $\rho_{\text{intra}}(f_5, f_3)$, and $\rho_{\text{intra}}(f_5, f_1)$, respectively. For each standard video sequence, Figure 5 shows that the best rate distortion performance of f_5 is achieved in S_0 with better performance in S_1 than in S_2. At the same time, it was found that $\rho_{\text{intra}}(f_5, f_4) > \rho_{\text{intra}}(f_5, f_3) > \rho_{\text{intra}}(f_5, f_1)$. From the above experiments, the changing trend in intra-description correlation was found to be analogous with that of rate distortion performance. Consequently, it is feasible that bit rate R and central distortion D_C can more or less be considered ras to ρ_{intra}.

In view of the relationship between ρ_{inter} and D_S, corresponding experiments were designed as follows. For any lost frame f_k, a reconstructed frame f_k^* can be estimated using the MCI method. Then we can compute the reconstructed quality for f_k and the inter-description correlation $\rho_{\text{inter}}(f_k, f_k^*)$. Here, assuming that the lost frame is f_5, f_5^* can be estimated using a prior and subsequent frame, for example, f_4 and f_6 in S_0, f_3 and f_7 in S_1, f_1 and f_9 in S_2. In Table 1, it can be seen that as the inter-description correlation coefficient increases, the reconstructed quality improves for standard video sequences. As a result, it is reasonable that ρ_{inter} can be used as a measure of the reconstructed quality of the lost frame, that is, side distortion D_S.

Table 1 The relationship between ρ_{inter} and D_S

Video sequence	(forward, back)	ρ_{inter}	Recon. (dB)
Mobile.qcif	(f_4, f_6)	0.9955	33.1019
	(f_3, f_7)	0.9730	25.4625
	(f_1, f_9)	0.9554	23.3396
Coastguard.qcif	(f_4, f_6)	0.9975	37.8429
	(f_3, f_7)	0.9931	33.5001
	(f_1, f_9)	0.9847	30.0360
Foreman.qcif	(f_4, f_6)	0.9931	32.3692
	(f_3, f_7)	0.9794	27.6490
	(f_1, f_9)	0.9239	22.0435
Paris.cif	(f_4, f_6)	0.9957	34.1520
	(f_3, f_7)	0.9888	30.0151
	(f_1, f_9)	0.9744	26.4454

Table 2 Test sequences

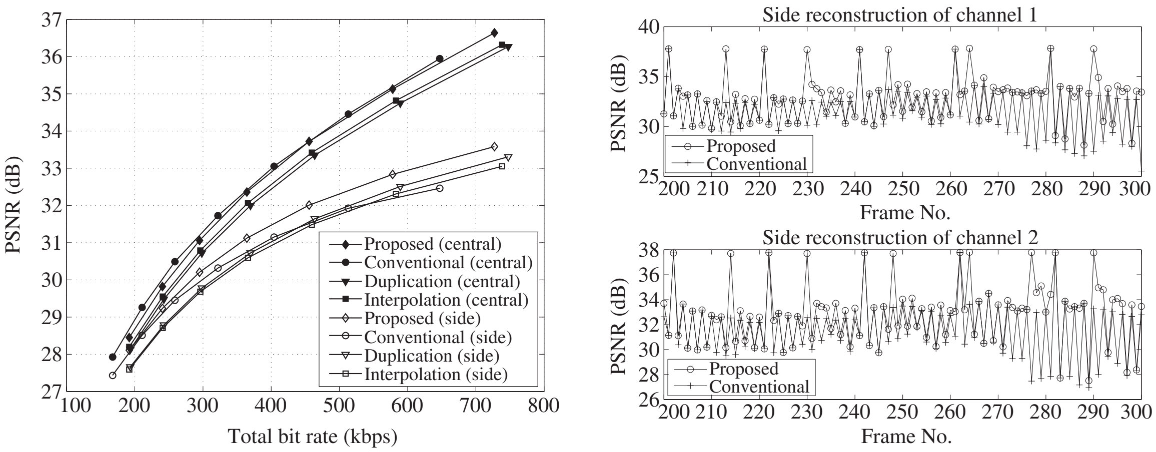
Sequence name	Format	Resolution	Input frames (fps)	Redundant frames	Interpolated frames (%)	Duplicated frames (%)
Mobile	QCIF	176 × 144	300@30	62	40.32	59.68
Paris	CIF	352 × 288	300@30	74	55.41	44.59
Crew	4CIF	704 × 576	100@30	24	12.50	87.50
Soccer	4CIF	704 × 576	100@30	25	12.00	88.00
City	720P	1280 × 720	100@60	23	21.74	78.26

4 Experimental results

Next, the performance of the proposed MDC scheme is discussed for on-off channels and lossy packet networks. Table 2 lists some standard video sequences in the experiments, and also summarizes the percentage of interpolated/duplicated frames used as redundant frames in the different sequences. In this paper, we focus on the comparison of our proposed temporal sampling scheme and other temporal or spatial sampling schemes since all of these are compatible with the standard video codec. Furthermore, a performance comparison between the proposed scheme and the video standard H.264/AVC [13] is also shown when packet loss has occurred in the network. To make a fair comparison, the same experimental setup is applied to all the compared schemes. The same parameters are used for the H.264 encoder and decoder with version JM10.2 of the software. For simplicity the H.264 coding structure in all the compared schemes is IPPP... without B frames. Additionally, we also employed the same MCI method for the estimation of lost frames. It should be noted that the total bit rate is the sum of the two descriptions with their labels and the side distortion is the average PSNR value from two side decoders.

4.1 Performance in on-off MDC channels

Figure 6(a) shows the central and side distortion of the proposed scheme compared with three other temporal sampling schemes for the test video *Mobile.qcif* with the total bit rate between 100 kbps and 800 kbps. From the results, the conventional scheme without pre-processing has a slight advantage over the other temporal sampling schemes in terms of central distortion. However, at the same bit rate the proposed scheme consistently performs better than the conventional scheme in terms of side distortion. This figure shows a comparison of the average PSNR values only for the whole video. In fact, some individual frames in the proposed scheme showed even greater improvement. Figure 6(b) shows the side PSNR for each frame (from the 200th to 300th frame) with a total bit rate of 400 kbps achieved by the proposed scheme and the scheme without pre-processing. It can be seen that for either channel the side reconstruction by the proposed scheme is substantially better than that of the conventional scheme with



(a) (b)

Figure 6 Performance of the proposed scheme vs. other temporal sampling in on-off MDC channels. (a) Rate-central/side distortion performance; (b) side reconstruction for each frame.

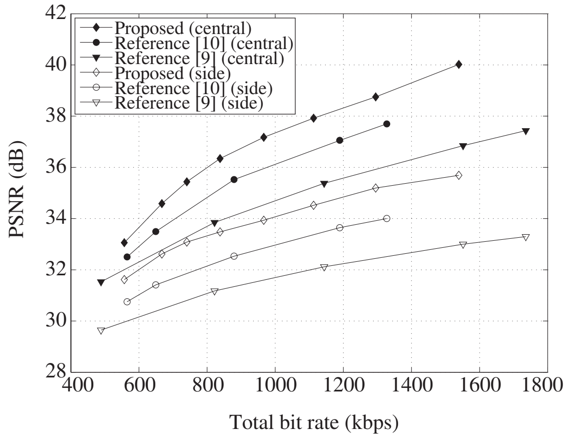


Figure 7 Performance of the proposed scheme vs. spatial sampling in on-off MDC channels.

a maximal improvement of about 10 dB for the 290th frame. Although in Figure 6(b) the frame-by-frame quality fluctuates by nearly 3 dB, the PSNR of most of the frames is good at 32–35 dB, which may reduce the influence of quality fluctuations. Furthermore, from Figure 6(a) we can see that the proposed scheme achieves better performance than those schemes using only frame duplication or interpolation, which may be attributed to the optimized mode selection of frame duplication/interpolation.

Figure 7 shows a comparison of the proposed scheme, optimized spatial sampling [10] and conventional spatial sampling [9] for the test video *Paris.cif* with the total bit rate between 400 kbps and 1800 kbps. From the results, we can see that our proposed temporal sampling scheme outperforms the other spatial sampling schemes for both central and side distortion simultaneously at the same bit rate ranging between 400 kbps and 1800 kbps, with improvement in both side and central distortion.

4.2 Performance in lossy packet networks

Next, we discuss the performance of the proposed MDC scheme in lossy packet networks. For fair comparison, the organization of packets is checkerboard-like in both cases. In Figure 8(a), for the test video *Mobile.qcif*, the performance of the proposed and conventional temporal sampling schemes is compared in lossy packet networks at the same total bit rate of 400 kbps. According to Figure 8(a), the proposed scheme achieves better performance than the conventional one. Improvements between 0.3 dB and 0.5 dB are obtained. Note that with a higher packet loss rate, the proposed scheme performs better than the conventional one since the efficient redundant information is helpful for estimating of lost information at the central decoder.

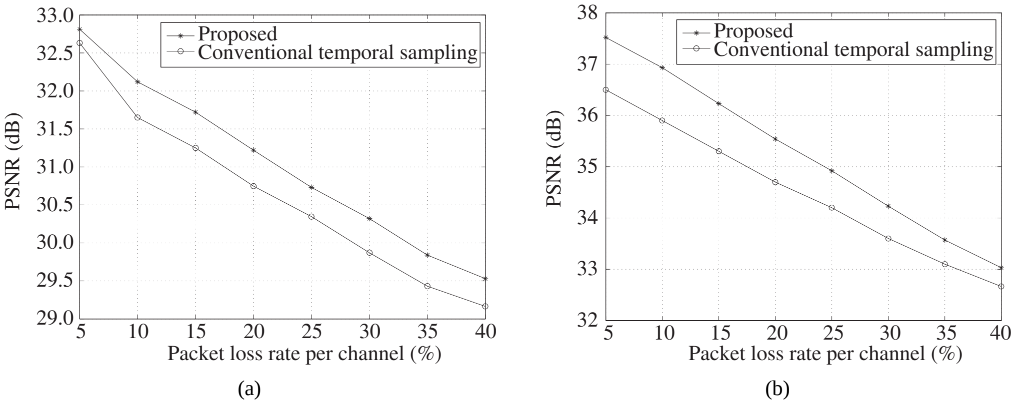


Figure 8 Performance comparison in lossy packet networks. (a) Mobile (QCIF); (b) Paris (CIF).

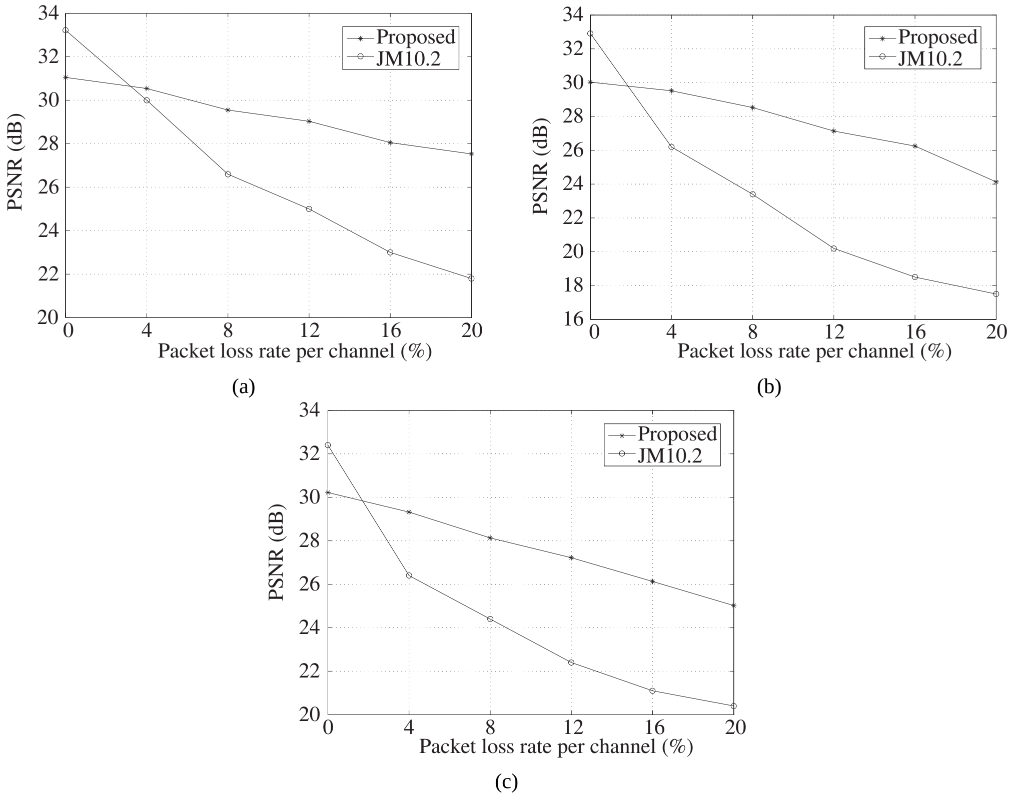


Figure 9 Performance comparison between the proposed system and H.264/AVC. (a) Crew (4CIF); (b) Soccer (4CIF); (c) City (720P).

Furthermore, Figure 8(b) shows a comparison of the performance of the proposed and optimized spatial sampling schemes in lossy packet networks at the same total bit rate of 1100 kbps for the test video *Paris.cif*. It can be seen in this figure that the proposed scheme shows a performance improvement of between 0.4 dB and 1dB compared with that of optimized spatial sampling [10]. Note that for a lower packet loss rate, the improvement is greater. The reason for this is that at a higher packet loss rate, spatial correlation becomes more significant than at a lower packet loss rate. Therefore, in this paper, joint temporal and spatial error concealment has been adopted for better reconstruction in lossy packet networks.

Next, we compared the performance of the proposed MDC scheme and the video standard H.264 with version JM10.2, as shown in Figure 9. For the three test sequences, the total bit rate was 8.1 Mbps, 7.9 Mbps and 28.6 Mbps, respectively, with the packet loss rate varying between 0% and 20%. From the figures, it can be seen that with no packet loss, the performance of JM10.2 surpasses the proposed scheme with the largest gaps between the two schemes being about 2 dB for the sequence *Crew* and *City* and 3 dB for the sequence *Soccer*. This is because the motion in the sequence *Soccer* is very irregular and more redundancy is inserted. However, when packet loss occurs in the network, the proposed scheme degrades gracefully while JM10.2 shows a sharp decline. Here, the largest gap by which the proposed scheme surpasses JM10.2 is about 5 dB for the sequence *Crew* and *City* and 6 dB for the sequence *Soccer*. This is because the irregular motion in *Soccer* may lead to worse estimation of the lost information.

5 Conclusion

An MD video coding scheme based on pre- and post-processing was presented in this paper, without any modification to the standard codec. In view of the motion change characteristic between frames, an optimized mode of temporal sampling during pre-processing can be accommodated based on the intra- and inter-description correlation for better side reconstruction. Furthermore, we showed that the proposed MD system has superior rate-distortion performance compared with the spatial sampling based scheme [9,10]. Owing to its perfect compatibility with the standard codec, the proposed MD video scheme may be a better choice in practical applications.

Acknowledgements

This work was supported in part by National Basic Research Program of China (973 Program) (Grant No. 2012CB316400), National Natural Science Foundation of China (Grant Nos. 60903066, 61073142, 61103113, 61272051, 61272262), Beijing Natural Science Foundation (Grant No. 4102049), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No. 20090009120006), Fundamental Research Funds for the Central Universities(Grant Nos. 2011JBM214, 2012JBZ012), and Jiangsu Provincial Natural Science Foundation (BK2011455). The authors would like to thank Wu Feng and Sun XiaoYan at Microsoft Asia Research (MSRA) for their valuable comments that greatly improved this paper.

References

1 Wang Y, Reibman A R, Lin S. Multiple description coding for video delivery. Proc IEEE, 2005, 93: 57–69
2 Goyal V K. Multiple description coding: compression meets the network. IEEE Signal Process Mag, 2001, 18: 74–93
3 Lin C, Zhao Y, Zhu C. Two-stage diversity-based multiple description image coding. IEEE Signal Process Lett, 2008, 15: 837–840
4 Bai H, Zhu C, Zhao Y. Optimized multiple description lattice vector quantization for wavelet image coding. IEEE Trans Circ Syst Video Technol, 2007, 17: 912–917
5 Liu M, Zhu C. M-description lattice vector quantization: index assignment and analysis. IEEE Trans Signal Process, 2009, 57: 2258–2274
6 Reibman A R, Jafarkhani H, Wang Y, et al. Multiple-description video coding using motion-compensated temporal prediction. IEEE Trans Circ Syst Video Technol, 2002, 12: 193–204
7 Chang S, Cosman P C, Milstein L B. Performance analysis of channel symmetric FEC-Based multiple description coding for OFDM networks. IEEE Trans Image Process, 2011, 20: 1061–1076
8 Zhao A, Yu Y, Sun G, et al. Multiple description video coding based on residuum compensation. Sci China Ser F-Inf Sci, 2009, 52: 1439–1445
9 Wang D, Canagarajah N, Redmill D, et al. Multiple description video coding based on zero padding. In: International Symposium on Circuits and Systems, Vancouver, 2004. 205–208
10 Wang A, Zhao Y, Bai H. Robust multiple description distributed video coding using optimized zero-padding. Sci China Ser F-Inf Sci, 2009, 52: 206–214
11 Bai H, Zhao Y, Zhu C. Multiple description video coding using adaptive temporal sub-sampling. In: IEEE International Conference on Multimedia and Expo, Beijing, 2007. 1331–1334
12 Bai H, Zhao Y, Zhu C, et al. Multiple description video coding using joint frame duplication/interpolation. Comput Inf, 2010, 29: 1267–1282
13 H.264 standard. JVT-G050 7th meeting, Pattaya, 2003