Multiple Description Video Coding Using Inter- and Intra-Description Correlation at Macro Block Level*

Huihui BAI††, Nonmember, Mengmeng ZHANG†††, Anhong WANG††††, Members, Meiqin LIU†, and Yao ZHAO††, Nonmembers

SUMMARY A novel standard-compliant multiple description (MD) video codec is proposed in this paper, which aims to achieve effective redundancy allocation using inter- and intra-description correlation. The inter-description correlation at macro block (MB) level is applied to produce side information of different modes which is helpful for better side decoding quality. Furthermore, the intra-description correlation at MB level is exploited to design the adaptive skip mode for higher compression efficiency. The experimental results exhibit a better rate of side and central distortion performance compared with other relevant MDC schemes.

key words: video coding, multiple description coding, correlation

1. Introduction

In recent years with the explosion of the Internet, video transmission has become increasingly popular in the world and will continue to flourish in the future. However, network congestion, random bit errors and packet losses may cause substantial quality degradation of the compressed video data, which impose tremendous challenge on video communications. Due to delay sensibility, the retransmission of corrupted data may be impossible in some specific applications such as real time video transmission or video-conference. Therefore, this creates a urgent need to develop the compression technology combining high compression efficiency and robust transmission.

Multiple description coding (MDC) has emerged as a promising technology for robust transmission over error-prone channels, which has been attracting more and more researchers [1]. MDC are proposed based on the assumption that multiple channels exist between the source and destination and it is impossible for all channels to fail at the same time. Therefore, using MDC the source can generate multiple bit streams (descriptions) with equal priority at the encoder, which then can be transmitted over multiple channels. At the decoder, each description can be decoded independently to produce the minimum fidelity which is measured by side distortion. With an increase in the received descriptions, the reconstructed quality can be enhanced further. In a simple architecture of two channels, the distortion generated by two received descriptions is called central distortion [2]. Here, we mainly focus on MD design for two channels.

The basic idea of MDC is how to allocate the essential redundancy of the descriptions for better tradeoff between encoding efficiency and robust transmission. The classical MDC versions use quantizers and transform. Based on the principle of MD scalar quantizer [3], an MD scheme for video coding is proposed in [4] while MD correlation transform is also employed to design motion compensated MD video coding [5]. Although the above methods have shown good performance, they are incompatible with widely-used standard codecs, such as H.26x and MPEG series. To solve this problem, in [6] MDC is designed to introduce redundancy into descriptions through utilizing the advanced video coding tools and features provided in H.264/AVC. Furthermore, many approaches are proposed for a certain standard codec such as H.264/AVC [7]–[9]. Especially, in view of better performance, in [8] and [9] the MDC schemes for H.264/AVC are presented at slice and macro block (MB) level, respectively.

To overcome the limitation of a given standard codec, in this paper we attempt to design a novel MD video codec with generalized compatibility. In this scheme, the redundancy from inter- and intra-descriptions is effectively allocated to gain better tradeoff between compression efficiency and robust transmission. At the encoder for simplicity odd/even frame splitting can be performed firstly to generate the video sub-sequence as the original descriptions, which can ensure fully compatibility with the current standard source codec such as H.26x or MPEG series. Then the correlation of inter-description at MB level is applied to produce side information of different modes. At the decoder the side information is helpful to estimate the loss information for better side decoding quality. Furthermore, in view of higher compression efficiency, the flexible skip mode is employed to adapt the MB level correlation of intra-description.

The rest of this paper is organized as follows. In Sect. 2, an overview of the proposed MD video coding scheme is presented step by step. In Sect. 3, the performance of the proposed scheme is examined. We conclude the paper in Sect. 4.
2. Overview of the Proposed Scheme

In the conventional MD video coding scheme, an original video sequence can be split directly into odd and even frames as two descriptions without any pre-processing. Although such method is easy to perform, temporal correlation of the original video sequences will be compromised to lead to poor estimation of lost frames in side reconstruction, especially for the frames containing irregular motion. Figures 1 shows a simple example, where irregular motion of a block (indicated by the black circle) exhibits from frame \( k - 1 \) to frame \( k + 1 \). Suppose a simple method, such as motion-compensated interpolation (MCI) [10] based on piecewise uniform motion assumption is used to estimate the lost frame \( k \) in side reconstruction. The reconstructed frame may suffer a substantial distortion loss due to the wrong displacement of the block. To address this problem, a preliminary scheme based on pre-/post-processing in [11] is proposed, where some redundant frames are inserted to regulate motion change for post-processing based on MCI. For the example in Fig. 1, if an interpolated frame is inserted between frame \( k \) and frame \( k + 1 \) in the pre-processing, the frame can be well estimated with the frame \( k - 1 \) and the inserted one. However, more inserted frames may be needed for some more complex irregular motion, which could produce much redundancy. Therefore, in [11] the redundancy allocation may be not enough effective at frame level. In this paper, the redundant information will be taken into account at MB level. Figure 2 illustrates the proposed scheme and a step-by-step recipe is explained as follows.

Step 1. Frame splitting

For simplicity, the original video sequence can be split into odd and even frames, which is fully compatible with the current standard video codec, such as H.26x and MPEG series. Here, the two sub-sequences, that is, odd frames and even frames can be regarded as the original descriptions. Furthermore, the original video sequences can be sub-sampling in temporal domain to generate multiple descriptions. Here, we focus on the two descriptions because the proposed method can be extend to multiple descriptions easily.

Step 2. Side information generated

From Fig. 1, it can be seen that if only odd or even frames can be received the side reconstruction may suffer a substantial distortion loss due to the wrong estimation for the block displacement. Therefore, for better estimation side information will be generated using MB level inter-description correlation. Here take an example of odd frames. Our task is to produce some useful side information from even frames for the side decoder.

Firstly, odd frames can be utilized to estimate the even frames according to MCI, shown in Fig. 3. The process is similar with the estimation at the side decoder. The MCI method will be presented in Step 5. Each MB of the reconstructed frames in Fig. 3 will be compared with the same position MB in the real even frames. Thus we can calculate the correlation coefficients of each pair MB using the Eq. (1).

$$\rho_{\text{inter}}(B_k, B'_k) = \frac{\text{Cov}(B_k, B'_k)}{\sqrt{D(B_k)D(B'_k)}}$$

where \( B_k \) is the current MB in even frame \( k \) and \( B'_k \) is the compared MB in the estimated frames \( k \). The covariance of \( B_k \) and \( B'_k \) is \( \text{Cov}(B_k, B'_k) \) and their variances can be denoted by \( D(B_k) \) and \( D(B'_k) \), respectively. Next, according to the MB level correlation of inter-descriptions, three modes of redundancy allocation will be designed as follows.

If \( \rho_{\text{inter}} > T_1 \) and \( T_1 = \frac{1}{N} \sum_{k=1}^{N} \rho_{\text{inter}}(B_k, B'_k) \), the correlation is considered enough to normally yield good estimation for the lost MB of even frames in side reconstruction. Here, \( T_1 \in [0, 1] \). \( N \) is the number of MB in the current video. In this case, no extra redundancy needs to be inserted, which can be regarded as Mode 1.

If \( T_2 < \rho_{\text{inter}}(B_k, B'_k) \leq T_1 \) and \( T_2 < T_1 \), the correlation between the pair of MB is worse which may suffer from the irregular motion. In this case, it is difficult for the method of MCI to estimate the accurate motion vector. Therefore, the real motion vector for this MB of even frames can be regarded as the essential side information, which is Mode 2.

It is noted that such motion vector can be searched in the
channels. According to a lot of experiments on the standard information will also be transmitted over the corresponding residuals are also needed as the side information. In this case, besides the motion vectors like Mode 0, the corresponding residuals are also needed as the side information. This can be considered as Mode 3. The side information will also be transmitted over the corresponding channels. According to a lot of experiments on the standard video sequences, if the correlation coefficient is smaller than 0.97, some large irregular motion may occur and more side information is needed for good estimation. In this case, besides the motion vectors like Mode 0, the corresponding residuals are also needed as the side information. Therefore, the compression performance will not drop effectively. In this section, besides the motion vectors like Mode 0, the corresponding residuals are also needed as the side information. This can be considered as Mode 3. The side information will also be transmitted over the corresponding channels. According to a lot of experiments on the standard video sequences, if the correlation coefficient is smaller than 0.97, some large irregular motion may occur and more side information is needed for good estimation. In this case, besides the motion vectors like Mode 0, the corresponding residuals are also needed as the side information. Therefore, the compression performance will not drop effectively.

Step 3. Adaptive skip mode

In Step 2, the introduced redundancy may impact the compression efficiency. Therefore, the redundancy within each description should be removed effectively. In this paper, a flexible skip mode is designed which can make good use of the MB level correlation of intra-descriptions. The correlation of intra-description (denoted by \( \rho_{\text{intra}} \)) is dependent on the temporal motion-compensated correlation within the same description. Therefore, \( \rho_{\text{intra}} \) between the current MB \( B_k \) in frame \( k \) and its forward motion-compensated MB \( B'_{k-1} \) in frame \( k - 1 \) can be computed by

\[
\rho_{\text{intra}}(B_k, B'_{k-1}) = \frac{\text{Cov}(B_k, B'_{k-1})}{\sqrt{\text{Var}(B_k)\text{Var}(B'_{k-1})}}
\] (2)

where \( B'_{k-1} \) is the motion-compensated MB of \( B_k \) in frame \( k - 1 \). Then if \( \rho_{\text{intra}}(B_k, B'_{k-1}) > T_3 \) and \( T_3 = \frac{1}{N} \sum_{k=1}^{N} \rho_{\text{intra}}(B_k, B'_{k-1}) \), then the MB \( B_k \) can be encoded using skip mode. As a result, the flexible skip mode can adapt to the correlation within the same descriptions. Compared with the uniform period of skip mode, adaptive skip mode can keep up with the temporal correlation between frames, so that better error concealment can be achieved at the decoder if information loss occurs in descriptions.

Step 4. Standard encoder

Each video sub-sequence (odd or even frames) can be encoded to bit streams using current standard codec. Here, H.264 encoder is chosen and obviously the proposed scheme is compatible with the generalized standard codec. In addition, the residuals generated in Mode 3 also can be processed using intra-coding of H.264 for compression.

Step 5. Reconstruction at the decoder

Here two cases for decoding will be considered, that is the design of central decoder and side decoder. If both channels work, the received two bit rate streams will be decoded to generate two video sub-sequences which then can be interleaved for central reconstruction. Due to the MB level correlation of intra-descriptions based, adaptive skip mode will counteract some effect from the redundant side information. Therefore, the compression performance will not drop seriously with increasing redundancy, which is exhibited in Sect. 3.

If only one channel works, the side decoder is designed to estimate the lost information according to the different modes. In the case of Mode 1, the lost frames can be reconstructed using the method of MCI directly. The widely-used method of MCI based on the piecewise uniform motion assumption is performed by bi-directional motion estimation, which may produce overlapped pixels and holes in the reconstructed frame. We denote by \( f \) the estimated frame between frame \( f_k \) and frame \( f_{k+1} \) and by \( MV(\tilde{p}) \) the motion vector for the pixel location \( \tilde{p} \). To avoid the holes in the estimated frame, we can compute a preliminary reconstruction as background.

\[
f(\tilde{p}) = \frac{1}{2}(f_k(\tilde{p}) + f_{k+1}(\tilde{p}))
\] (3)

Furthermore, the forward and backward motion compensation can be performed for frame \( f_k \) and \( f_{k+1} \), respectively. To solve the overlapped problem of MCI, the mean values of overlapped pixels are adopted for motion compensation. Then the preliminary background may be replaced by the MCI-based reconstruction according to

\[
f(\tilde{p}) = \frac{1}{2}(f_k(\tilde{p} - MV(\tilde{p})) + f_{k+1}(\tilde{p} + MV(\tilde{p})))
\] (4)

In Mode 2, the encoded MVs can be used as accurate motion vectors at the decoder, which are denoted as \( MV_f(\tilde{p}) \) from forward prediction and \( MV_b(\tilde{p}) \) from backward prediction, respectively. Therefore, the Eq. (4) can be refined as follows.

\[
f(\tilde{p}) = \frac{1}{2}(f_k(\tilde{p} - MV_f(\tilde{p})) + f_{k+1}(\tilde{p} + MV_b(\tilde{p})))
\] (5)

Furthermore, in Mode 3 accurate residuals can also be applied to refine the reconstructed quality of MCI.

3. Experimental Results

Two standard video sequences “Mobile.qcif” and “Paris.cif” are used to test our scheme against others. To make a fair comparison, the same experimental setup is applied for all the compared schemes. The same parameters are chosen in H.264 encoder and decoder [12]. Additionally, we also employ the same MCI method for the estimation of lost frames. It is noted that the total bit rate is the sum of two descriptions with the labels and the side distortion is the mean PSNR value from two side decoders.

Figure 4 shows the side and central distortion of the proposed scheme against other relevant schemes for the test video “Mobile.qcif” at the total bit rate from 100 kbps to 800 kbps. For the conventional method without any preprocessing, since the temporal correlation may be preserved by the direct frame splitting, it has shown the worst rate side-distortion performance in Fig.4(a). Compared with the conventional one, the redundancy inserted in [11] can improve the side reconstruction to some extent, but much redundant information at frame level can lead to lower rate central-distortion performance shown in [11] can improve the side reconstruction to some extent, but much redundant information at frame level can lead to lower rate central-distortion performance shown in [11].
Fig. 4 (b). From the figures, at the same bit rate the proposed scheme can consistently perform better than the other scheme in both central and side distortion. This is just a comparison for the average PSNR values of the whole video. Actually, some individual frames in the proposed scheme may achieve more improvements. Figure 5 shows the side PSNR of each frame (from the 200th to 300th frame) at the total bit rate 400 kbps, from which it can be seen that for either channel the side reconstruction of the proposed scheme has substantial improvement compared with the reference scheme [11] and the maximal improvement is about 10 dB at the 290th frame.

Furthermore, in Fig. 6 the test video “Paris.cif” is utilized for examine the proposed scheme at the total bit rate from 400 kbps to 1800 kbps. From the results, we can find out that the proposed temporal sampling scheme outperforms the other schemes in both central and side distortion simultaneously at the same bit rate over a wide range from 400 kbps to 1800 kbps, with both improvements of 0.5–1 dB in side distortion and 0.5–1.7 dB in central distortion.

4. Conclusion

An MD video coding scheme based on effective redundancy allocation has been developed in this paper, without any modification to the standard codec. In view of the MB level correlation of inter-descriptions, a different mode of redundancy is designed for better side reconstruction. Furthermore, adaptive skip mode has been accommodated based on intra-description correlation to enhance the compression efficiency. As a result, the proposed MD system has also demonstrated superior rate-distortion performance to the conventional and reference schemes. Due to the generalized compatibility with the standard codec, the proposed scheme may be a better choice in the practical applications.

References

[1] Y. Wang, A.R. Reibman, and S. Lin, “Multiple description coding for video delivery,” Proc. IEEE, vol.93, no.1, pp.57–69, 2005.
[2] V.K. Goyal, “Multiple description coding: compression meets the network,” IEEE Signal Processing Magazine, vol.18, no.5, pp.74–93, 2001.
[3] V. Vaishampayan, “Design of multiple description scalar quantizers,” IEEE Trans. Inf. Theory, vol.39, no.3, pp.821–834, 1993.
[4] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, “Multiple description coding for video using motion compensated prediction,” Proc. IEEE International Conference on Image Processing (ICIP ’99), pp.837–841, 1999.
[5] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, “Multiple description coding for video using motion compensated prediction,” Proc. IEEE International Conference on Image Processing (ICIP ’99), pp.837–841, 1999.
[6] N. Conci and F.G.B. De Natale, “Multiple description video coding using coefficients ordering and interpolation,” Signal Process. Image Commun., vol.22, no.3, pp.252–265, 2007.
[7] Y. Wang, A.R. Reibman, and S. Lin, “Multiple description coding for video delivery,” Proc. IEEE, vol.93, no.1, pp.57–69, 2005.
[8] V.K. Goyal, “Multiple description coding: compression meets the network,” IEEE Signal Processing Magazine, vol.18, no.5, pp.74–93, 2001.
[9] V. Vaishampayan, “Design of multiple description scalar quantizers,” IEEE Trans. Inf. Theory, vol.39, no.3, pp.821–834, 1993.
[10] V.A. Vaishampayan and S. John, “Balanced interframe multiple description video compression,” Proc. IEEE International Conference on Image Processing (ICIP ’99), pp.812–816, 1999.
[11] A.R. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, “Multiple description coding for video using motion compensated prediction,” Proc. IEEE International Conference on Image Processing (ICIP ’99), pp.837–841, 1999.
[12] N. Conci and F.G.B. De Natale, “Multiple description video coding using coefficients ordering and interpolation,” Signal Process. Image Commun., vol.22, no.3, pp.252–265, 2007.