New H.266/VVC Based Multiple Description Coding for Robust Video Transmission over Error-Prone Networks

Dinh Trieu Duong
Faculty of Electronics and Telecommunications (FET)
University of Engineering and Technology (UET) - Vietnam National University, Hanoi (VNU)
Hanoi, Vietnam
duongdt@vnu.edu.vn

***Abstract*—In this paper, we propose a novel multiple description coding (MDC) method to operate at network edges for robust video transmission. The proposed MDC method, named VVC-MDC offers benefits of both the new H.266 Versatile video coding (H.266/VVC) and Distributed video coding (DVC) standards, which can provide not only higher performance compared to the traditional MDC methods but also effective scheme for the error resilience. At the encoder, the proposed VVC-MDC coder encode the source video sequence into two descriptions including odd and even subsequences and then transmit these descriptions to the receiver. At the receiver, our proposed MDC decoder is designed using a novel Wyner-Ziv (WZ) coding introduced in the DVC to provide a high image quality for the video sequence. Experimental results show that the proposed method can achieve a wide range of tradeoffs between coding efficiency and error resilience, and provide much better PSNR performance than other conventional MDC methods.**

Keywords—Multiple description coding (MDC), H.266 Versatile video coding, H.266/VVC, Distributed video coding (DVC)

I. Introduction

Recently, multiple description coding (MDC) has emerged as a promising approach to enhance the error resilience of a video delivery system. It can effectively combat packet loss without retransmission thus satisfying the demand of real-time services and relieving the network congestion [1].

In MDC, the source video is encoded into two (or more) correlated descriptions, which are then individually packetized and sent through either the same or separate physical channels. At the receiver, if both the descriptions are correctly received, the decoder provides a high-quality reconstruction of the source data. On the other hand, if one of the descriptions is lost, the decoder estimates it from the other description, and then provides a lower but acceptable video quality reconstruction.

Several methods have been proposed for the MDC technique [2]-[8]. One of the most popular MDC methods is the scalar quantization based MDC [2], which is applied to the MDC coders in [3], [4]. However these methods focus on stand-alone MDC codecs then they are not compatible with standards like H.264/AVC or H.265/HEVC [5]. To address this, Indoonundon *et al.* [6] proposed another MDC method based on the H.264/AVC named FMO-MDC. In the FMO-MDC method, the flexible macroblock ordering (FMO) scheme is combined with the H.264/AVC based MDC coder to enhance the performance of error concealment for the lost description. In [7], Xiang *et al.* introduced a 2-D layered multiple description coding (2DL-MDC) for efficient error resilience while preserving compatibility with the H.264 Scalable video coding (H.264/SVC) standard. Majid *et al.* [8] proposed a MDC coder which splits the input video sequence into even and odd subsequences and then encodes these subsequences using H.265 High efficiency video coding (H.265/HEVC). These methods can provide an effective error resilient coding solution for the MDC codec. However, it is the fact that the best available coding standard recently is not any more H.264/AVC or H.265/HEVC but rather the H.266 Versatile video coding (H.266/VVC) [9], [10].

MDC has also been investigated for non-standard video coding algorithms such as in [11]- [12], where MDC is combined with distributed video coding (DVC) approaches. Generally, there are two main approaches to the DVC design: the DVC Stanford [13] and the DVC Berkeley [14] solutions. Milani *et al.* [12] presented an effective DVC based MDC approach, named Multiple description distributed video coder (MD-DVC) that encoded the input video signal and created different descriptions multiplexing primary and redundant video packets. The proposed MD-DVC can provide a good redundancy tuning mechanism and overcome the limitations posed by the conventional predictive video codecs. However, this coder is also conceived as a stand-alone MDC codec, and thus, the descriptions generated by the MD-DVC coder are not compatible with video standards, e.g. H.264/AVC or H.265/HEVC.

In this paper, we propose a novel multiple description coding (MDC) method to operate at network edges for increasing robustness of video streaming. The proposed MDC method offers benefits of both the new H.266/VVC and Distributed video coding (DVC) standards. At the encoder, the proposed VVC-MDC coder encode the source video sequence into two descriptions including odd and even subsequences and then transmit these descriptions to the receiver. At the receiver, our proposed MDC decoder is designed using a novel Wyner-Ziv (WZ) coding scheme introduced in the DVC to provide a high image quality for the video sequence, even if one of the descriptions is lost during the transmission. Unlike the conventional MDC methods, the redundant data in our proposed MDC can be effectively controlled based on the WZ coding scheme.

The rest of the paper is organized as follows. Section II describes the proposed method in detail. Experimental results

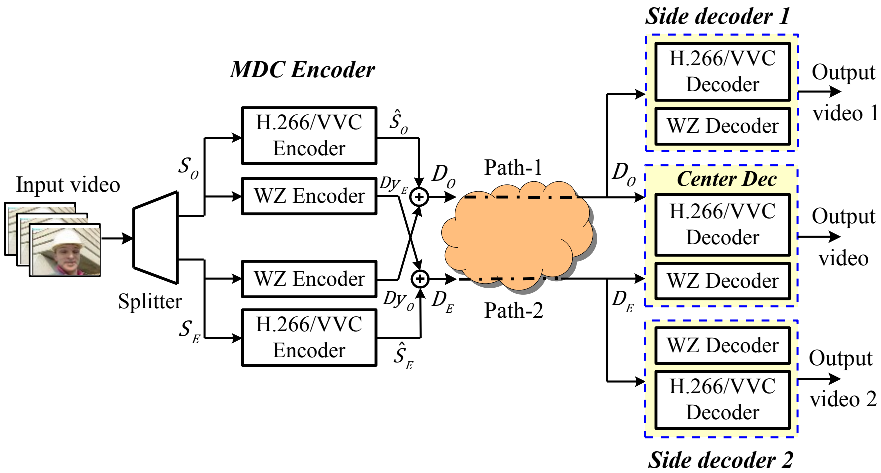

Fig. 1. The proposed VVC-MDC method.

are discussed in Section III. Finally, Section IV concludes this paper.

II. Proposed H.266/VVC Based Multiple Description Coding (VVC-MDC)

Fig. 1 shows the general framework of the proposed VVC-MDC method. In Fig. 1, the input video sequence is separated into two parts: the odd and even subsequences including the odd and even frame indexes of the input sequence, respectively. These subsequences are then encoded using H.266/VVC and WZ coding to obtain the odd and even compressed and syndrome bitstreams, $\hat{S}_i$ and Dy_i $(i = O, E)$, respectively, which are then encapsulated into two corresponding descriptions named D_O and D_E to transmit to the receiver.

At the receiver, the proposed MDC includes two types of decoders, namely central and side decoders. The central decoder is utilized when all descriptions are correctly received as shown in Fig. 1. Otherwise, when only one description is available and correctly received, it is decoded using the corresponding side decoder to obtain the reconstructed video sequence.

A. *Proposed VVC-MDC Encoder*

As shown in Fig. 2, at the proposed MDC encoder, instead of using the conventional video coding standards like H.264/AVC or H.265/HEVC, we utilize H.266/VVC which provides several advanced video coding techniques to encode the odd and even video frames [9]. This make our proposed MDC can not only satisfy the requirement of fully standard compatible codec but also can provide an effective solution to improve the coding efficiency for the proposed MDC coder. In addition, though the codec itself is not the core novelty of this paper, our proposed VVC-MDC codec is the first MDC codec in literature employing H.266/VVC coding.

Compared to the H.264/AVC and H.265/HEVC, H.266/VVC standard is designed from the ground up to be both efficient and versatile to address today's media needs. H.266/VVC is also the evolution of H.265/HEVC codec: With the same perceptual quality, H.266/VVC can offer up to 50% compression efficiency than HEVC and support a wide range of resolutions from 4K to 16K as well as 360° videos

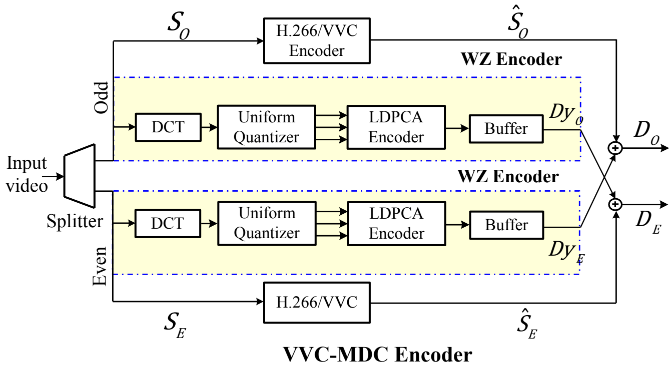

Fig. 2. Proposed VVC-MDC Encoder

[9].

Let S_O and S_E denote the odd and even subsequences, respectively. As shown in Fig. 2, at the encoder, both S_o and S_E are independently encoded using H.266/VVC to achieve two encoded bitstreams, $\hat{S}_o$ and $\hat{S}_E$, respectively. $\hat{S}_o$ and $\hat{S}_E$ are then encapsulated into two corresponding descriptions named D_O and D_E to transmit to the receiver.

Thought based on the H.266/VVC standard, the proposed MDC encoder can achieve high performance for the description coding, it would also be suffered from the predictive mismatch and predictive error propagation, which are general problems in most conventional standard compatible MDC coder [5]. To solve these problems, in the proposed MDC encoder, we employ a novel concept, namely WZ coding introduced in the DVC technique [13] to encode the descriptions, S_O and S_E. As shown in Fig. 2, together with the H.266/VVC coding, the odd and even subsequences S_o and S_E, are also transformed using Discrete cosine transform (DCT).

The quantized coefficients are then encoded using entropy (biplane per biplane) and LDPCA coding. LDPCA code is described in [15] as an efficient way of using low-density parity-check (LDPC) code for a rate adaptive scheme. An LDPCA encoder consists of an LDPC syndrome-former concatenated with an accumulator as shown in Fig. 3. In our proposed MDC encoder, for each bit plane, syndrome bits, Dy_O and Dy_E , are created using the LDPC code and accumulated modulo 2 to produce the accumulated syndrome.

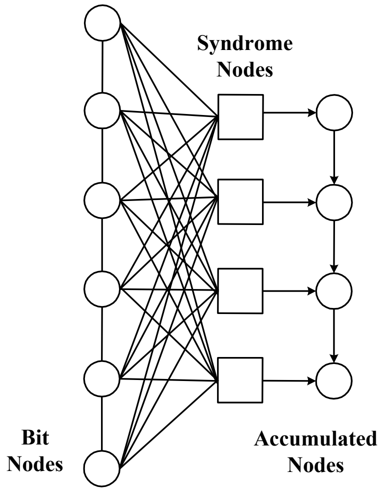

Fig. 3. LDPCA code

It is noted that in our MDC method, to improve the coding efficiency for the MDC coder, only a minimum rate of accumulated syndromes Dy_O and Dy_E is estimated, and then put into two descriptions, D_O and D_E, to send to the MDC decoder. The remaining syndrome bits, $\breve{D}y_O$ and $\breve{D}y_E$ are stored in the encoder buffer to be sent later depend on the channel feedbacks.

After encoding, two descriptions, D_O and D_E are transmitted over two distinct paths, Ph_O and Ph_E, of a path diversity system to the MDC decoder as shown in Fig. 2.

B. *Proposed VVC-MDC Decoder*

1) MDC Center decoder: At the MDC central decoder, both descriptions D_O and D_E, which are correctly received without errors are decoded by using H.266/VVC. In this case, D_O and D_E are jointly decoded, thus leading to a higher reconstruction quality for the reconstructed frames.

Compared to the conventional single description coding like H.265/HEVC or H.266/VVC, at the same image quality, the coding efficiency of the center decoder is decreased since the additional data, Dy_O and Dy_E, received at the center decoder in this case is not the decoded video data but the redundant data. However, the cost of these redundant data is acceptable because these data are essencial for the error resilient scheme provided for the proposed MDC coder.

2) MDC Side decoder: When only one description, D_O or D_E, is available and correctly received at the receiver, it is decoded using the corresponding MDC side decoders as shown in Fig. 4.

Without loss of generality, it is assumed that D_O is transmitted to the decoder over the path-1 and D_O is lost due to the transmission errors. In this case, the side decoder 2 is employed not only to decode the correctly received description, D_E, but also to interpolate for the lost description, D_O, to provide an acceptable quality for the entire video sequence. Since D_O is not available, the side decoder 2 need to employ the correlation between D_E and D_O to obtain the interpolated description $\widetilde{D}_O$ for D_O.

It is worth noticing that the interpolated quality of $\widetilde{D}_O$ plays an important role for improving the total rate-distortion performance of the proposed MDC coder. The higher image

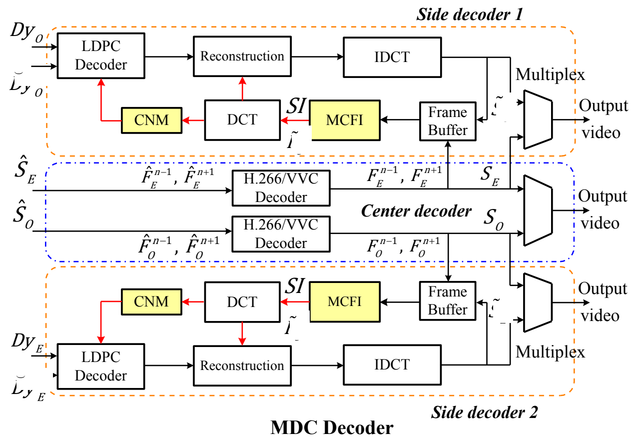

Fig. 4. Proposed MDC Decoder

quality gained for $\widetilde{D}_O$, the smaller amount of redundant data required for Sy_O, and then the higher coding efficiency can be achieved for the MDC coder [5]. In this work, we propose to use an algorithm named Motion compensated frame interpolation (MCFI) which can effectively employ the high correlations between D_E and D_O to obtain a good image quality for $\widetilde{D}_O$. More details on the MCFI algorithm are described in the following subsection.

a) Motion compensated frame interpolation (MCFI):

The main concept of MCFI is introduced in [16] and it has been successfully applied to many applications [17]. In this work, based on the high correlation between odd and even frames included in D_O and D_E, the MCFI algorithm is employed to obtain the interpolated frames for D_O.

Specifically, let F^n denote the nth frame in the original input sequence, F^{n-1} and F^{n+1} be the previous and next frames of F^n, respectively. The input video sequence is split into S_E and S_O as shown in Fig. 1. Thus, after splitting and H.266/VVC encoding, the frames F^{n-1}, F^{n+1}, and F^n become $\hat{F}_E^{n-1}$, $\hat{F}_E^{n+1}$, and $\hat{F}_O^n$, respectively, where $\hat{F}_E^{n-1}$, $\hat{F}_E^{n+1}$ are located in $\hat{S}_E$, and $\hat{F}_O^n$ is located in $\hat{S}_O$. $\hat{S}_o$ and $\hat{S}_E$ are then encapsulated into D_O and D_E to transmit to the receiver as explained in the previous section.

At the receiver, when D_O is lost due to the transmission errors, $\hat{S}_O$ and thus $\hat{F}_O^n$ are lost also. In contrast, D_E is correctly received, then $\hat{F}_E^{n-1}$ and $\hat{F}_E^{n+1}$ can be correctly decoded to obtain F_E^{n-1} and F_E^{n+1}, respectively as shown in Fig. 4.

In the MCFI algorithm, the high temporal correlation between successive decoded frames, F_E^{n-1} and F_E^{n+1}, are employed to obtain the interpolated frame $\tilde{F}_O^n$ for $\hat{F}_O^n$. Specifically, let $\boldsymbol{v}(\boldsymbol{x})$ be the 2D motion vector of the pixel $\boldsymbol{x}$, $\boldsymbol{v}(\boldsymbol{x})$ is estimated in the motion estimation between F_E^{n-1} and F_E^{n+1}, where F_E^{n-1} is referred to as the reference frame of F_E^{n+1}. In other words, $F_E^{n+1}(\boldsymbol{x})$ is the predicted pixel of $F_E^{n-1}(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x}))$ in the forward motion estimation process. Then,

$$F_E^{n+1}(\boldsymbol{x}) = F_E^{n-1}(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x}))$$

In the forward direction, along the motion trajectory passing through $\tilde{F}_O^n$ from F_E^{n-1} to F_E^{n+1} as shown in Fig. 5, we can approximate $\tilde{F}_O^n(\boldsymbol{x})$ as

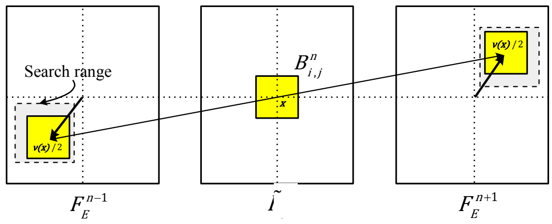

Fig. 5. Bi-directional MCFI scheme

$$\tilde{F}_O^n(\boldsymbol{x}) = F_E^{n-1}(\boldsymbol{x} - \boldsymbol{v}(\boldsymbol{x})/2) \quad (1)$$

And, in the backward direction:

$$\tilde{F}_O^n(\boldsymbol{x}) = F_E^{n+1}(\boldsymbol{x} + \boldsymbol{v}(\boldsymbol{x})/2) \quad (2)$$

Then, $\tilde{F}_O^n$ can be interpolated using the Bi-directional motion compensation as follows:

$$\tilde{F}_O^n(\boldsymbol{x}) = \frac{1}{2}[F_E^{n-1}(\boldsymbol{x} - \boldsymbol{v}(\boldsymbol{x})/2) + F_E^{n+1}(\boldsymbol{x} + \boldsymbol{v}(\boldsymbol{x})/2)]. \quad (3)$$

b) Side decoding with Side information (SI): At the side decoder, $\tilde{F}_O^n$ can be used as a simply replacement for the lost frame F_O^n as in the other conventional frame error concealments. However, it can be seen that, these approaches can only provide an acceptable prediction image if the motion vectors between F_E^{n-1} and F_E^{n+1} are highly correlated. Otherwise, the prediction image and so the quality of MDC codecs can be severely degraded due to the effect of annoying artifacts observed at the block region boundaries. To solve the problem, in our proposed MDC, we utilize $\tilde{F}_O^n$ as the side information (SI) frame only, $\tilde{F}_O^n = F_{SI}^n$, based on which the proposed MDC side decoder processes F_{SI}^n further to achieve higher image quality for the reconstructed lost frame, F_O^n.

There are several researches have been introduced to model the correlation between $\mathrm{F}_{\mathrm{SI}}^{\mathrm{n}}$ and $\mathrm{F}_{\mathrm{O}}^{\mathrm{n}}$. In [18], Brites *et al*. has shown that the SI frame F_{SI}^n can be considered as the noise version of the frame F_O^n, and the residual data which is the different between $\mathrm{F}_{\mathrm{SI}}^{\mathrm{n}}$ and $\mathrm{F}_{\mathrm{O}}^{\mathrm{n}}$ (in both the pixel and transform domains) can be modelled as the correlation noise model (CNM) that follows the Laplacian distribution. Thus, in our works, given SI frame F_{SI}^n, the LDPCA decoder is designed to iteratively request more syndrome bits $\breve{D}y_E$ to correct the mismatch between F_{SI}^n and F_O^n. In addition, at the sender, the MDC encoder replies to each request by sending additional syndrome bits, which combined with the previously sent ones, until they are sufficient for successful decoding F_O^n.

After LDPCA decoding, the decoded frames are inverted using the invert quantization and invert DCT transform to obtain the reconstructed description $\widehat{D}_O$ which is then combined with the reconstructed description $\widehat{D}_E$ to obtain a full resolution for the output video sequence as shown in Fig. 4.

Similarly, when the description D_E is lost, all the approaches mentioned above can be utilized again in the side decoder 1 to provide a faithful image quality for the reconstructed description.

III. Experimental Results

Several experiments have been performed to illustrate the effectiveness of the proposed VVC-MDC method. The experiment results are reported for several video sequences using VTM reference software [19]of the H.266/VVC standard.

In these experiments, two descriptions D_O and D_E are generated and simultaneously transmitted over the path-1 and path-2, respectively, to the receiver. At the receiver, both center and side decoders are employed to provide faithful image quality for the decoded descriptions, even if one description is lost due the transmission errors.

First, we compare the PSNR performance of the proposed method with that of the FMO-MDC method introduced in [6] and the conventional H.266/VVC single description coding (SDC) [10]. In the FMO-MDC method, the flexible macroblock ordering (FMO) scheme is combined with the H.264/AVC based MDC coder to enhance the performance of error concealment for the corrupted description. For the conventional H.266/VVC SDC, the encoded stream is transmitted over one single path, and the PLR of this path is set to p_S.

Fig. 6 shows the PSNR performance of the proposed MDC, the conventional H.266/VVC, and the FMO-MDC methods corresponding to a wide range of encoding bitrates. As seen in Fig. 6, in the error-free condition, the PSNR performance obtained in the center decoder of the proposed method is about 0.6dB lower than that of the conventional H.266/VVC SDC method.

In the case of error-free where both descriptions are correctly received at the decoder, these redundant data might result in the degradation on the RD performance of the proposed method. However, in cases of lossy packet networks where the encoded descriptions are suffered from the transmission errors, the proposed method can provide much higher PSNR performance than conventional methods. As seen in Figs. 6 and 7, with the PLRs of channels are equal to 5% ($p_1 = p_2 = 0.05$), at the bitrate of 2.0Mbps, the proposed VVC-MDC can provide 5.8dB better performance

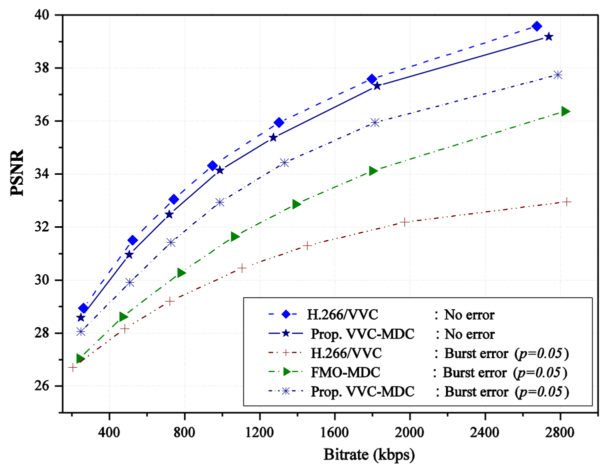

Fig. 6. PSNR performance for *Coastguard* sequence when PLR=5%

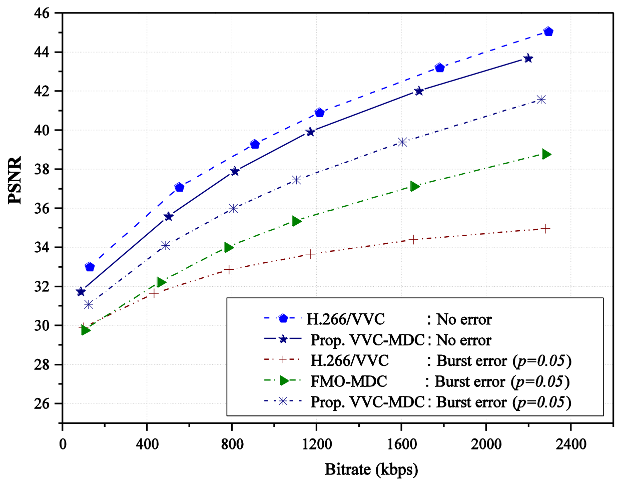

Fig. 7. PSNR performance for *Foreman* sequence when PLR=5%

than the conventional H.266/VVC SDC. And, with the same amount of redundancy data required, the performance of FMO-MDC method is lower than that of the proposed VVC-MDC at all values of bitrates.

TABLE I. PSNR PERFORMANCES ON TEST VIDEO SEQUENCES AND DIFFERENT PLR (P1, P2)

(dB)

Sequence	*(p1,p2)*	H.266/VVC	FMO-MDC	Prop. VVC-MDC
Foreman	(0.01, 0.05)	33.87	36.13	**38.05**
Coastguard	(0.05, 0.05)	31.48	33.21	**35.14**
Hall	(0.10, 0.05)	34.47	36.33	**38.51**

Table I shows more details on the average PSNR performance of the conventional and proposed methods performed on different video test sequence, PLRs and QPs. As shown in Table I, the proposed MDC method always provides higher PSNR performance than the H.266/VVC and FMO-MDC methods. For example, the proposed algorithm provides up to 3.66 dB and 1.93 dB gains as compared with the H.266/VVC and FMO-MDC methods, respectively, for the *Coastguard* sequence when QP=25 and PLR= 5%.

IV. Conclusion

In this paper, we have proposed a novel multiple description coding (MDC) method which offers benefits of both the new H.266/VVC and DVC standards. The proposed VVC-MDC coder encodes the source video sequence into two descriptions including odd and even subsequences and then transmit these descriptions to the receiver. At the receiver, our proposed MDC decoder is designed using a novel Wyner-Ziv (WZ) coding introduced in the DVC to provide a high image quality for the video sequence. Experimental results show that the proposed method can effectively provide higher PSNR and error resilience performance than other conventional MDC methods.

References

[1] V.K. Goyal, "Multiple description coding: compression meets the network," IEEE Signal Process. Mag., vol. 5, no. 18, p. 74–93 , 2001.

[2] V. A. Vaishampayan, "Design of multiple description scalar quantizers," IEEE Trans. Inform. Theory, vol. 39, no. 3, p. 821–834, 1993.

[3] O. Crave, B. P. Popescu, and C. Guillemot, "Robust Video Coding Based on Multiple Description Scalar Quantization With Side Information," IEEE Trans. Circuits Syst. Video Techn., vol. 20, no. 6, pp. 769 - 779, June 2010.

[4] T. Guionnet, C. Guillemot, and S. Pateux, "Embedded multiple description coding for progressive image transmission over unreliable channels," Proc. Int. Conf. Image Process., vol. 1, no. 1, pp. 94-97, Oct. 2001.

[5] Y. Wang, A.R. Reibman, and S. Lin, "Multiple description coding for video delivery," Proceedings of the IEEE, vol. 93, no. 1, pp. 57-70, Jan. 2005.

[6] D. Indoonundon, T.P Fowdur, and K.M.S Soyjaudah, "Enhanced H.264 Transmission with Multiple Description Coding, Prioritised Concealment and FMO," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 9, no. 2, pp. 81-90, June 2017.

[7] W. Xiang, C. Zhu, C. K. Siew, Y. Xu, and M. Liu, "Forward error correction-based 2-D layered multiple description coding for error-resilient H.264 SVC video transmission," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 12, pp. 1730-1738, Dec. 2009.

[8] M. Majid, M. Owais, and S. M. Anwar, "Visual saliency based redundancy allocation in HEVC compatible multiple description video coding," Multimedia Tools and Applications, vol. 77, p. 20955–20977, Dec. 2017.

[9] ITU (2018-04-27), "Beyond HEVC: Versatile Video Coding project starts strongly in Joint Video Experts Team," ITU News. Retrieved 2019-01-21.

[10] B. Bross, J. Chen, S. Liu, Y.-K. Wang, "Versatile Video Coding (Draft 8)," doc. JVET-Q2001 of ITU-T/ISO/IEC Joint Video Exploration Team (JVET), Brussels, 17th meeting, 2020.

[11] M. Wu, A. Vetro, and C. Wen, "Multiple-description image coding with distributed source coding and side information," Multimedia Syst. Applicat., vol. 5600, no. 7, p. 120–127, Dec. 2004.

[12] S. Milani and G. Calvagno, "Multiple Description Distributed Video Coding Using Redundant Slices and Lossy Syndromes," IEEE Signal Processing Letters, vol. 17, no. 1, pp. 51-54, Jan. 2010.

[13] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed Video Coding," Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, 2005.

[14] R. Puriand and K. Ramchandran, "PRISM:a new robust video coding architecture based on distributed compression principles," Proceedings of the 40th Allerton Conference Communication, Control and Computing, 2002.

[15] R. G. Gallager, "Low-density parity-check codes," IRE Transactions on Information Theory, vol. 8, no. 1, pp. 21-28, Jan. 1962.

[16] B.-T. Choi, S.-H. Lee, and S.-J. Ko, "New frame rate up-conversion using bi-directional motion estimation," IEEE Trans. Consum. Electron., vol. 46, no. 3, p. 603–609, Aug. 2000.

[17] B.-D. Choi, J.-W. Han, C.-S. Kim, and S.-J. Ko, "Motion-Compensated Frame Interpolation Using Bilateral Motion Estimation and Adaptive Overlapped Block Motion Compensation," IEEE Trans. Circuits Syst. Video Techn., vol. 17, no. 4, pp. 407 - 416, Apr. 2007 .

[18] C. Brites, F. Pereira, "Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner–Ziv Video Coding," IEEE Trans. Circuits Syst. Video Techn., vol. 18, no. 9, pp. 1177 - 1190, Sept. 2008.

[19] F. Bossen, X. Li, and K. Sühring, "Guidelines for H.266/VVC reference software development," JVET-N1003 for Geneva meeting (MPEG number m48033), 2019.