Distributed Coding Based Multiple Descriptions for Robust Video Transmission over Error-Prone Networks

Dinh Trieu Duong
University of Engineering and Technology (UET) - Vietnam National University, Hanoi (VNUH)
G2 Building, Faculty of Electronics and Telecommunications
144 Xuan Thuy Rd., Hanoi, Vietnam
+84 024 37547 9338
duongdt@vnu.edu.vn

ABSTRACT
In this paper, we propose a novel multiple description coding (MDC) method to enhance the robustness of video transmission over error-prone networks. The proposed MDC method provides benefits of both distributed video coding (DVC) and multiple description coding techniques, which can offer not only higher performance compared to the conventional MDC methods but also effective scheme for the error resilience. In the proposed MDC method, the input video sequence is split into odd and even group of pictures (GOPs) subsequences, which are independently encoded using the new H.265/High efficiency video coding (H.265/HEVC) based DVC technique. Though the codec itself is not the core novelty of this paper, our proposed codec is the first MDC codec in literature employing H.265/HEVC based DVC approaches, thus all results presented in this paper are new. Experimental results show that the proposed method can achieve a wide range of tradeoffs between coding efficiency and error resilience, and provide much better peak signal-to-noise ratio (PSNR) performance than other conventional MDC methods.

Keywords
Multiple description coding (MDC), H.265 High efficiency video coding, H.265/HEVC, Distributed video coding (DVC).

1. INTRODUCTION
Recently, multiple description coding (MDC) has emerged as an attractive framework for robust video transmission over packet lossy networks [1]. In MDC, the source video data 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 [2].

Several methods have been proposed for the MDC technique [3]-[4]. In [3], Tillo *et al.* proposed a MDC method that utilized the slice coding tool available in H.264/AVC to create two balanced descriptions. These descriptions are then encoded and analyzed at the MDC encoder to optimally select the amount of redundancies inserted in each slice. Majid *et al.* [4] 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 are standard compatible MDC methods and they can provide an effective error resilient coding solution for the proposed MDC codecs. However, it is the fact that the use of standard compatible MDC codecs like H.264/AVC or H.265/HEVC based MDC often leads to predictive mismatch and predictive error propagation. At the decoder, the prediction signal may differ from the one used by the encoder. That means, it may not always be possible to reconstruct the video properly due to this predictive mismatch. To solve the mismatch problem, the DVC based MDC method can be considered as an efficient approach for robust video transmission and avoid predictive mismatch of predictive coding [2].

In the DVC architecture, the input video is separated into two parts: the key and WZ frames. The key frames are encoded using the conventional Intra coding and the Wyner–Ziv (WZ) frames are coded using WZ coding scheme. One of the most popular methods introduced for the DVC based MDC is multiple description using scalar quantization (MDSQ) method [5], which is applied to the MDC coders in [6], [7]. Milani *et al.* in [8] 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. This method can provide a good redundancy tuning mechanism and overcome the limitations posed by the conventional predictive video codecs. However, when there is no packet loss, the WZ data is not exploited by the MD-DVC, which we aim at avoiding in the approach described here. In [9], X. Ou *et al.* introduced the multiple descriptions Wyner–Ziv video coding (MDWZVC) which utilized the Stanford DVC architecture with the DISCOVER codec [10]. The DISCOVER codec is based on the H.264/AVC Intra coding for the key frame. However, it has been shown in [11] that the best available coding standard recently is not any more H.264/AVC but rather the H.265/HEVC Intra coding.

In this paper, we propose a novel MDC method to enhance the robustness of video transmission over lossy packet networks. The proposed MDC method provides benefits of both H.265/HEVC and DVC techniques, which can offer not only higher performance compared to the conventional MDC methods but also effective scheme for the error resilience. In the proposed MDC method, the input video sequence is structured into groups of pictures (GOPs) containing the Key and Wyner-Ziv (WZ) frames, which are encoded using the Intra and WZ coding schemes, respectively. Though the codec itself is not the core novelty of this paper, our proposed codec is the first MDC codec in literature employing DVC and H.265/HEVC based DVC approaches, thus all results presented in this paper are new. Experimental results show that the proposed method can achieve a wide range of tradeoffs between coding efficiency and error resilience, and provide much better peak signal-to-noise ratio (PSNR) performance than other conventional MDC methods.

The rest of the paper is organized as follows. Section 2 describes the proposed method in detail. Experimental results are discussed in Section 3. Finally, Section 4 concludes this paper.

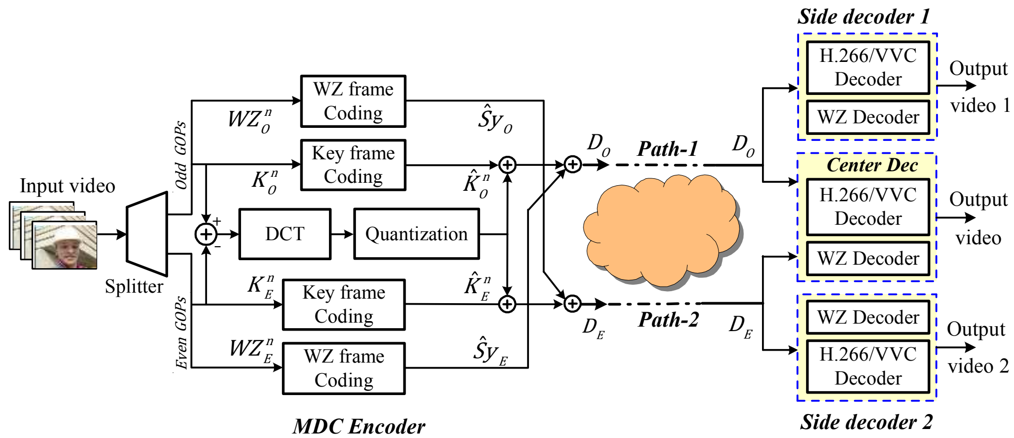

Figure 1. General video streaming framework of the proposed MDC method.

2. PROPOSED MULTIPLE DESCRIPTION CODING

2.1 Proposed MDC Encoder

Figure 1 shows a video streaming framework of the proposed MDC method. In Figure 1, the input video sequence is separated into two parts: the odd and even subsequences including the odd and even indexes of GOPs, respectively. In order to improve the coding efficiency and provide a robustness error resilience scheme for the proposed MDC codec, each GOP is configured to have only one key frame and one WZ frame as shown in Figure 2.

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. Otherwise, when only one description is available and correctly received, it is decoded using the corresponding side decoder to obtain the reconstructed video sequence.

Figure 3 shows the architecture of our proposed MDC encoder. As shown in Figure 3, for each GOP, the key and WZ frames are encoded using H.265/HEVC Intra and WZ coding scheme, respectively. In our method, key frames are periodically inserted in each GOP with the GOP size of 2 ($N_{GOP} = 2$).

It can be seen in Figure 3 that the proposed MDC encoder also employs the Stanford DVC approach, specifically, the DISCOVER codec which is described in [10]. However, instead of using the conventional H.264/AVC Intra coding as in the DISCOVER, our proposed method employs H.265/HEVC standard [11] which offers several advanced coding techniques to effectively improve the coding efficiency for the proposed MDC coder.

Let K_i^n and WZ_i^n denote the original key and WZ frames in the *n*th GOP_i $(i = O, E)$, respectively. Figure 2 illustrates more details on the positions of K_i^n and WZ_i^n in each GOP_i^n configured from the input video sequence. At the encoder, the key frames K_O^n and K_E^n are encoded using H.265/HEVC Intra coding to obtain the compressed bitstreams, $\widehat{K}_O^n$ and $\widehat{K}_E^n$, respectively. Let ΔD_k be the difference between K_O^n and K_E^n, ΔD_K^n is defined as

$$\Delta D_K^n = K_O^n - K_E^n \quad (1)$$

Then, in the proposed MDC, ΔD_K^n is transformed using the DCT transform and then the resulting DCT coefficients are quantized to obtain $\Delta\widehat{D}_K^n$ as shown in Figure 3. $\widehat{K}_O^n$, $\widehat{K}_E^n$, and $\Delta\widehat{D}_K^n$ are then encapsulated into descriptions D_O and D_E to transmit to the receiver.

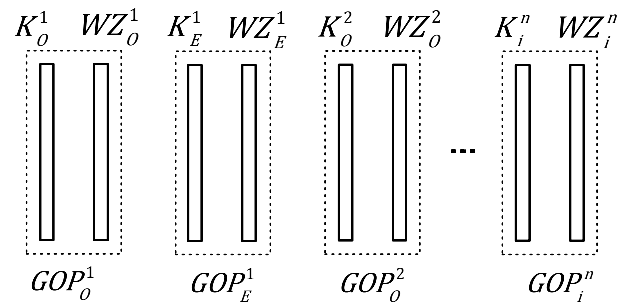

Figure 2. Structure of key and WZ frames in each GOP.

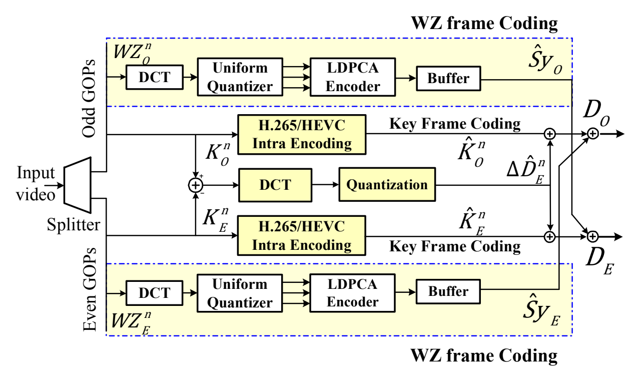

Figure 3. Proposed MDC Encoder.

For WZ frame coding, an integer 4x4 block-based DCT is applied prior to quantization. The quantized values are then split into bitplanes which go through a Low density parity check code (LDPC) encoder. LDPC code is described in [12] as an efficient way of using low-density parity-check (LDPC) code for a rate adaptive scheme. An LDPC encoder consists of an LDPC syndrome-former concatenated with an accumulator as shown in Figure 4. In our proposed MDC encoder, for each bit plane, syndrome bits, $\hat{S}y_O$ and $\hat{S}y_E$, are created using the LDPC code and accumulated modulo 2 to produce the accumulated syndrome.

It is noted that in our MDC method, to improve the coding efficiency for the MDC coder, only a minimum rate of accumulated syndromes

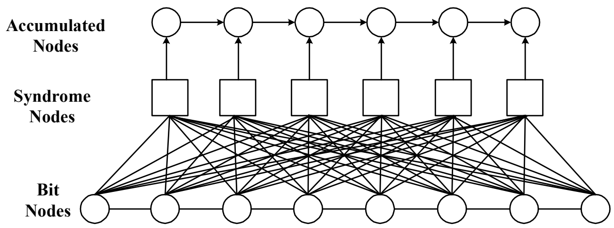

Figure 4. LDPC coder

$\hat{S}y_O$ and $\hat{S}y_E$ is estimated, and then put into two descriptions, D_O and D_E, to send to the MDC decoder. The remaining syndrome bits, $\breve{S}y_O$ and $\breve{S}y_E$ 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 Figure 1.

2.2 Proposed MDC Decoder

2.2.1 MDC Center decoder

At the reciever, *w*hen both the descriptions D_O and D_E are correctly received without errors, the proposed MDC center decoder is employed to decode D_O and D_E. As shown in Figure 5, the received key and WZ frames included in D_O and D_E are decoded using H.265/HEVC Intra and WZ decoding, respectively. In this case, D_O and D_E are jointly decoded, thus leading to a higher reconstruction quality for the reconstructed frames.

2.2.1.1 Motion Compensated Frame Interpolation (MCFI)

In the MCFI algorithm, the high temporal correlation between successive decoded frames, K_O^n and K_E^n, are employed to obtain the interpolated SI frame WZ_{SI}^n for WZ_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 K_O^n and K_E^n, where K_O^n is referred to as the reference frame of K_E^n as shown in Figure 6. In other words, $K_O^n(\boldsymbol{x})$ is the predicted pixel of $K_E^n(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x}))$ in the forward motion estimation process. Then,

$$K_O^n(\boldsymbol{x}) = K_E^n(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x})) \quad (2)$$

It is generally assumed that motion vectors between the consecutive frames in a video sequence are analogous to each other because the movement of objects in real video sequences tends to change smoothly [13]. Thus, in the forward direction, along the motion trajectory passing through WZ_{SI}^n from K_O^n to K_E^n as shown in Figure 6, we can approximate $WZ_{SI}^n(\boldsymbol{x})$ as

$$WZ_{SI}^n(\boldsymbol{x}) = K_O^n(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x})/2) \quad (3)$$

And, in the backward direction:

$$WZ_{SI}^n(\boldsymbol{x}) = K_E^n(\boldsymbol{x}+\boldsymbol{v}(\boldsymbol{x})/2) \quad (4)$$

Then, WZ_{SI}^n can be interpolated using the Bi-directional motion compensation as follows:

$$WZ_{SI}^n(\boldsymbol{x}) = \frac{1}{2}[K_O^n(\boldsymbol{x}-\boldsymbol{v}(\boldsymbol{x})/2) + K_E^n(\boldsymbol{x}+\boldsymbol{v}(\boldsymbol{x})/2)]. \quad (5)$$

2.2.1.2 WZ Decoding with Side Information (SI)

In our proposed MDC, we utilize WZ_{SI}^n as the side information (SI) frame only, $\widetilde{WZ}_O^n = WZ_{SI}^n$, based on which the proposed MDC side decoder processes WZ_{SI}^n further to achieve higher image quality for the reconstructed frame, WZ_O^n.

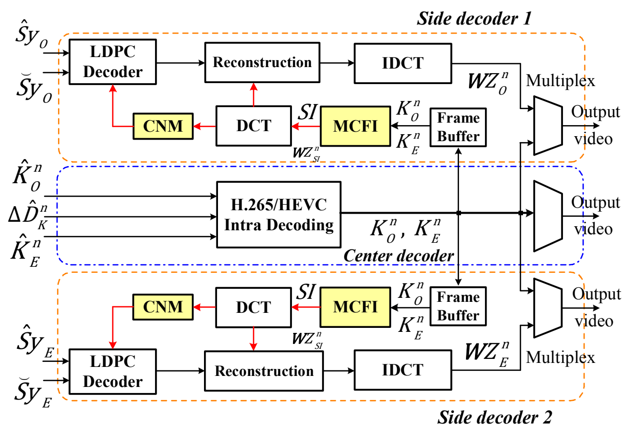

Figure 5. Proposed MDC Decoder

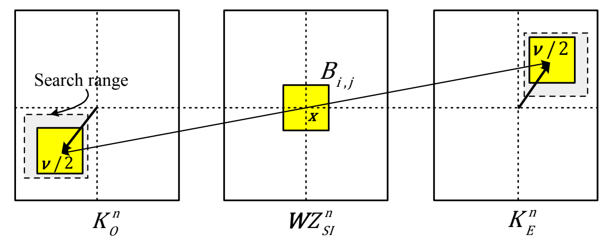

Figure 6. Bi-directional MCFI scheme

There are several researches have been introduced to model the correlation between WZ_{SI}^n and WZ_O^n. In [14], Brites *et al*. has shown that the SI frame WZ_{SI}^n can be considered as the noise version of the frame WZ_O^n, and the residual data which is the different between WZ_{SI}^n and WZ_O^n (in both the pixel and frequency domains) can be modelled as the correlation noise model (CNM) that follows the Laplacian distribution. Thus, in our works, given the SI frame WZ_{SI}^n, CNM of WZ_{SI}^n, and the received syndrome bits Sy_O, the LDPC decoder is designed to iteratively request more syndrome bits $\breve{S}y_O$ to correct the mismatch between WZ_{SI}^n and WZ_O^n. In addition, at the sender, the MDC encoder replies to each request by sending additional syndrome bits, which is combined with the previously sent ones, until they are sufficient for successful decoding WZ_O^n.

Similarly, we can reconstruct the original WZ frames WZ_E^n. Then, after decoding, both the key frame K_i^n and WZ frame WZ_i^n $(i = O, E)$ belong to D_O and D_E, are successfully decoded at the proposed MDC center decoder. D_O and D_E are then combined to provide a full resolution for the output video sequence.

2.2.2 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 Figure 5.

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 contrast, the description D_E which includes the key frame K_E^n, key frame difference ΔD_K^n, and syndrome data $\hat{S}y_O$ is correctly received at the side decoder. It is worth noticing that in the case of center decoding where both the descriptions D_O and D_E are correctly received, the additional data $\Delta\hat{D}_K^n$ is not the decoded video data but the redundant data. However, the cost of these redundant data is acceptable because these data are essential

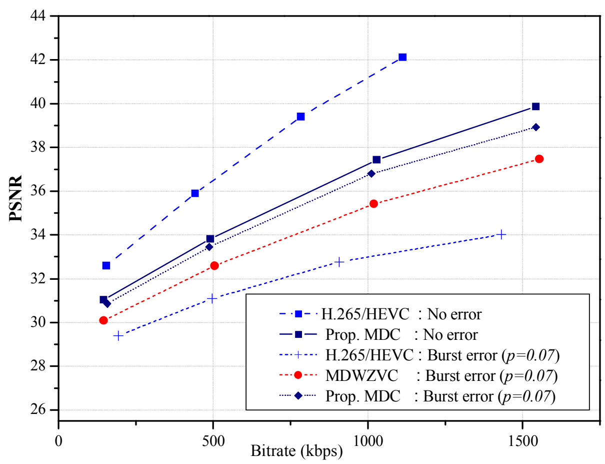

Figure 7. Rate-distortion performance for *Foreman* sequence when PLR=7%

to recover D_O or D_E when one of these descriptions is lost during transmissions.

Specifically, in our proposed side decoder, when D_O is lost, that means, K_O^n and $\hat{S}y_E$ included in D_O are lost also. However, as in (1), K_O^n can be recovered from K_E^n as follows:

$$K_O^n = K_E^n + \Delta D_K^n \quad (6)$$

Based on the recovered key frame K_O^n, we can also recover the corresponding WZ frame WZ_O^n as explained in the proposed MDC center decoder above.

3. EXPERIMENTAL RESULTS

Several experiments have been performed to illustrate the effectiveness of the proposed MDC method. The experiment results are reported for several video sequences using HM 16.2 reference software [15] of the H.265/HEVC standard. The test sequences including *Foreman*, *Soccer*, and *Hall* are in YUV 4:2:0 format with CIF (352x288) resolution.

3.1 Packet loss pattern simulation

In our tests, we use the two-state Markov channel model based on the simulation data of modulation and coding scheme 2 (MCS-2) in 3GPP TR 26.904 [16] to simulate proper packet loss patterns along each path. For video transmission over path diversity system, we set target PLRs of path-1 and path-2 as p_1 and p_2, respectively. The values of parameters which are characterized for the two-state

3.2 Performance comparison

In order to illustrate the effectiveness of the proposed MDC method, we compare the PSNR performance of the proposed method with that of the *multiple descriptions Wyner–Ziv video coding* (MDWZVC) method introduced in [9] and the conventional H.265/HEVC single description coding (SDC) in [11]. In the MDWZVC method, WZ frames are sub-sampled into four parallel low resolution to generate four descriptions, which are encoded and transmitted to the decoder independently. At the receiver, the MDWZVC method utilizes a successive refined SI algorithm, to exploit both temporal and spatial correlations between successive video frames.

Figure 7 shows the PSNR performance of the proposed MDC, the

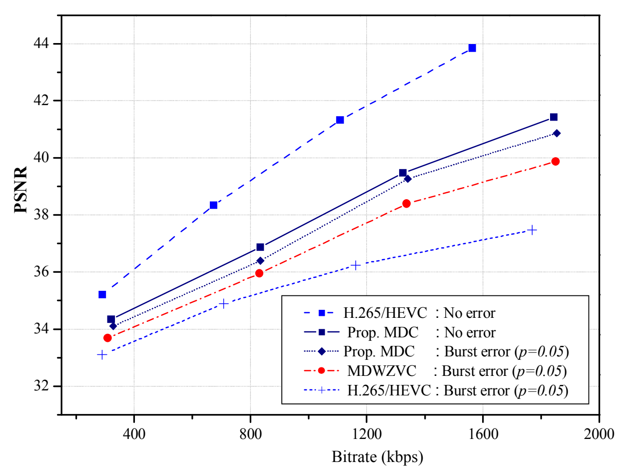

Figure 8. Rate-distortion performance for *Hall* sequence when PLR=5%

Table 1. Comparisons of the average PSNRs performed on different test video sequences and different PLR (p1, p2)

(dB)

Seq.	*(p1,p2)*	H.265 - SDC	MDWZVC	Prop. MDC
Foreman	(0.01, 0.07)	33.07	35.23	**37.31**
Soccer	(0.05, 0.07)	31.02	32.33	**34.14**
Hall	(0.05, 0.05)	36.12	37.08	**37.81**

conventional H.265/HEVC SDC, and the MDWZVC methods corresponding to a wide range of encoding bitrates. As seen in Figure 7, in the error-free condition, the PSNR performance of the the conventional H.265/HEVC SDC method is higher than that of the proposed MDC method. The reason lies in the fact that in H.265/HEVC SDC, many recursively searches have been performed to find the optimal motion estimation and Intra/Inter prediction modes. On every possible coding tree units (CTUs) sizes, a complex rate distortion optimization (RDO) is performed among up to 35 prediction mode candidates to screen out the best prediction mode. These lead to a high coding efficiency achieved for the H.265/HEVC SDC. However, this method is also faced with big challenges of very high coding complexity and low capability of error resilience against transmission errors. In contrast, the proposed MDC have proved to be significantly more efficient as compared to the SDC for robust video transmissions. As shown in Figure 7, in cases of lossy packet networks where the encoded descriptions are suffered from the transmission errors, the proposed method outperforms the conventional methods by a large margin of performance. For example, with the PLRs of channels are equal to 7% ($p_1 = p_2 = 0.07$), at the bitrate of 1.0Mbps, the proposed MDC provides up to 4.2dB better performance than the conventional H.265/HEVC SDC.

The PSNR performance obtained in the MDWZVC method is higher than that obtained for the H.265/HEVC SDC method. However, with the same amount of redundancy data required, the MDWZVC method yields worse performances than the proposed MDC method at all values of bitrates as shown in Figures 7 and 8.

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 consistently provides better performance than the H.265/HEVC and MDWZVC methods. For example, the proposed algorithm provides up to 3.12 dB and 1.81 dB gains as compared with the H.265/HEVC and MDWZVC methods, respectively, for the *Soccer* sequence when QP=25 and PLR= 7%.

4. CONCLUSION

In this paper, we have proposed a novel multiple description coding (MDC) method to enhance the robustness of video transmission. The proposed MDC method provides benefits of both distributed video coding (DVC) and multiple description coding techniques, which can offer not only higher performance compared to the conventional MDC methods but also effective scheme for the error resilience. In the proposed MDC method, the input video sequence is split into odd and even group of pictures (GOPs) subsequences, which are independently encoded using the new H.265/High efficiency video coding (H.265/HEVC) based DVC technique. Experimental results show that the proposed method can achieve a wide range of tradeoffs between coding efficiency and error resilience, and provide much better peak signal-to-noise ratio (PSNR) performance than other conventional MDC methods.

5. REFERENCES

[1] Wang, Y., Wenger, S., Wen, J., and Katsaggelos, A. K. 2000. Error resilient video coding techniques. *IEEE Signal Process. Mag.*, 17, 1 (Jul. 2000), 61-82. DOI= 10.1109/79.855913.

[2] Wang, Y., Reibman, A.R., and Lin, S. 2005. Multiple description coding for video delivery. *Proceedings of the IEEE*, 93, 1 (Jan. 2005), 57–70. DOI= 10.1109/JPROC.2004.839618.

[3] Tillo, T., Grangetto, M., and Olmo, G. 2008. Redundant slice optimal allocation for H.264 multiple description coding. *IEEE Trans. Circuits Syst. Video Technol.*, 18, 1 (Jan. 2008), 59–70. DOI= 10.1109/TCSVT.2007.913751.

[4] Majid, M., Owais, M., and Anwar, S. M. 2017. Visual saliency based redundancy allocation in HEVC compatible multiple description video coding. *Multimedia Tools and Applications*, 77 (Dec. 2017), 20955–20977. DOI= https://doi.org/10.1007/s11042-017-5499-7.

[5] Vaishampayan, V. A. 1993. Design of multiple description scalar quantizers. *IEEE Trans. Inform. Theory*, 39, 3 (1993), 821–834. DOI= 10.1109/18.256491.

[6] Crave, O., Popescu, B. P., and Guillemot, C. 2010. Robust Video Coding Based on Multiple Description Scalar Quantization with Side Information. *IEEE Trans. Circuits Syst. Video Techn.*, 20, 6 (Jun. 2010), 769 - 779. DOI= 10.1109/TCSVT.2010.2045805.

[7] Guionnet, T., Guillemot, C., and Pateux, S. 2001. Embedded multiple description coding for progressive image transmission over unreliable channels. *Proc. Int. Conf. Image Process.*, 1, 1 (Oct. 2001), 94-9. DOI= 10.1109/ICIP.2001.958961.

[8] Milani, S. and Calvagno, G. 2010. Multiple Description Distributed Video Coding Using Redundant Slices and Lossy Syndromes. *IEEE Signal Processing Letters*, 17, 1 (Jan. 2010), 51-54. DOI= 10.1109/LSP.2009.2031748

[9] Ou, X., Qing, L., He, X., Yu, Y., Yu, Y. 2014. Multiple descriptions Wyner–Ziv video coding. *International Journal for Light and Electron Optics*. 125, 5 (Mar. 2014), 1650–1656. DOI= 10.1016/j.ijleo.2013.09.065.

[10] Artigas, X., Ascenso, J., Dalai, M., Klomp, S., Kubasov, D., and Ouaret, M. 2007. The discover codec: architecture, techniques and evaluation. *Proceedings of Picture Coding Symposium (PCS '07)* (Nov. 2007). DOI= 10.1.1.131.1402.

[11] Sullivan, G. J., Ohm, J.-R., Han, W.-J., and Wiegand, T. 2012. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits Syst. Video Technol.*, 22, 12 (Dec. 2012), 1649-1668. DOI= 10.1109/TCSVT.2012.2221191.

[12] Gallager, R. G. 1962. Low-density parity-check codes. *Transactions on Information Theory*, 8, 1 (Jan. 1962), 21 - 28. DOI= 10.1109/TIT.1962.1057683

[13] Choi, B.-T., Lee, S.-H., and Ko, S.-J. 2000. New frame rate up-conversion using bi-directional motion estimation. *IEEE Trans. Consum. Electron.*, 46, 3 (Aug. 2000), 603–609. DOI= 10.1109/30.883418.

[14] Brites, C. and Pereira, F. 2008. Correlation Noise Modeling for Efficient Pixel and Transform Domain Wyner–Ziv Video Coding. *IEEE Trans. Circuits Syst. Video Techn.*, 18, 9 (Sept. 2008), 1177 - 1190. DOI= 10.1109/TCSVT.2008.924107.

[15] Frank, B., David, F., Karl, S., and Suhring, K. 2014. HEVC reference software HM16.2 reference manual. 2014.

[16] 3GPP TR 26.904. Study on improved video coding support, ETSI TR 126904 v. 10.0, *Universal Mobile Telecommunications System (UMTS)*.

Table 2. Authors' background

Author Name	Title*	Research Field	Personal website
Dinh Trieu Duong	Associate Professor	Image/Video Processing and Communication	https://sites.google.com/site/xiemhoang/3-research-team