Multiple description video coding for real-time applications using HEVC

Trung Hieu Le[1] Marc Antonini[1] Marc Lambert[2] Karima Alioua[2]

[1] Laboratoire I3S - Université Côte d'Azur et CNRS, UMR 7271, Sophia Antipolis, France

[2]Lextan SAS, Gemenos, France

{thle, am}@i3s.unice.fr

Résumé

Remote control vehicles require the transmission of large amounts of data, and video is one of the most important sources for the driver. To ensure reliable video transmission, the encoded video stream is transmitted simultaneously over multiple channels. However, this solution incurs a high transmission cost. To address this issue, it is necessary to use more efficient video encoding methods that can make the video stream robust to noise. Moreover it should have a less complexity to adapt to the real time requirement. In this paper, we propose a low-complexity, low-latency 2-channel Multiple Description Coding (MDC) solution with an adaptive Instantaneous Decoder Refresh (IDR) frame period, which is compatible with the HEVC standard which adaptive redundancy adjustment. This method shows a better resistance to high packet loss rates with lower complexity.

Mots clefs

Multiple Description coding, HEVC, noisy channel, Error Correction, low latency

1 Introduction

Remote driving requires continuous video data transmission to allow the driver to perceive the environment. The vehicle's video is transmitted to the driver through a wireless channel, and the latency and frame quality must be within limits to guarantee safety. However, wireless networks are more susceptible to noise than wired networks and are thus less reliable. Forward Error Correction (FEC) has been implemented in the transport layer to improve the resilience of video transmission to losses, but it is designed to be robust only up to a certain limit of packet error rate and requires high complexity computation[1]. Furthermore, the wireless channel's characteristics are dynamic and vary over time, making it difficult to estimate the channel state accurately for FEC.

To mitigate interference in real-time video transmission, one approach is to use two separate wireless channels to send the same video sequence as a backup. However, this can result in a waste of bandwidth since the same information is being transmitted twice. Multiple Description Coding (MDC) can be used as a solution to address this issue. In the case of MDC for two channels, the MDC encoder produces two different descriptions, S1 and S2, of the video with bit rates R1 and R2, respectively, from the original video source. These two descriptions are then transmitted through two independent channels by two transmitters. If only one description is available at the MDC decoder, either S1 or S2, the side decoder will be used to produce the video sequence with distortion D1 or D2, respectively. However, if both descriptions are available at the MDC decoder, the central decoder merges them to construct the central reconstruction by removing redundant information and retaining the primary one, resulting in higher video quality with a smaller central distortion D0.

MDC has been studied in the past and can be classified into three categories based on how the redundancy information is added to each description. There have been many studies on MDC in the spatial domain, most of which are compatible with the H.264 encoding standard. In [2], Tilo *et* al proposed a method of MDC compatible with the H.264 standard based on adjusting the redundancy level of the different slices. However, this method required to calculate all the GOP to find the optimum bit allocation. In [3], the author used the Multiple-State Video Coding with Redundant Pictures method by shifting the I and P frames of Group of Picture (GOP) between the two descriptions. In [4], the authors proposed an H.264 MDC scheme based on block permutation and DCT coefficients splitting. However, These methods didn't include the redundancy control thus it cannot adapt in the redundancy level based on network condition.

However, recent studies have attempted to include an MDC scheme in the HEVC standard, which has higher performance and is more resilient to packet erasure, as demonstrated in [5]. Many of these studies, such as [6, 7], are based on visual saliency and transform domain splitting. However, these methods may not be suitable for real-time applications due to their use of Bidirectional Motion Compensation (B-Frame). In [8], authors propose a frame rate variation MDC scheme adapted for remote control vehicles, but this scheme does not exploit the temporal dependency between frames, resulting in lower compression performance.

In this article, a spatial domain multiple description codec based on the HEVC standard with Rate Distortion Optimization (RDO) at the CTU level for each frame independently is proposed. The scheme uses unidirectional temporal prediction, which allows for a reduction in transmission costs and meets real-time delay requirements. The proposed solution is presented in the first part of the article, and the second part shows its performance under a packet erasure channel.

2 Proposed Method

We propose a balanced MDC with two descriptions as described in Figure 1. Assume that the packet error distribution of both channels is independently and identically distributed (i.i.d). Therefore, for each frame of the sequence, the expected distortion at the decoder is expressed as :

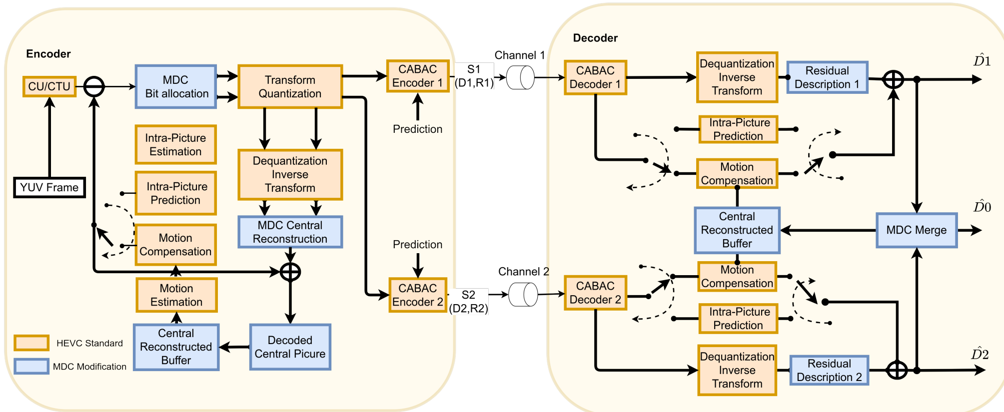

FIGURE 1 – *Proposed coding scheme : First, the frame will be split in CTU/CU structure optimized for the given target bitrate. Based on this structure and the residual produced by this step. The MDC bit allocation will allocate the QP value for each CTU using algorithm 1. Then, these QP values will feed the Transform and Quantization blocks at encoder side. After that the central reconstruction will be constructed by discarding the coarse quality CTU, this central reconstruction is stored in the Central Reconstructed Buffer. Finally MC process is performed on the central reconstruction.*

$$D_e = (1-p_e)\sum_{j=1}^{2} D_{p,j} + p_e(1-p_e)\sum_{j=1}^{2} D_{r,j} + p_e^2 D_{error} \quad (1)$$

Here, $D_{p,j}$ and $D_{r,j}$ are the total distortion of the principal CTUs and the redundant CTUs, respectively, in the same description j of the same frame. We define the following relationship :

$$D_{p,j} + D_{r,j} = \sum_{i=1}^{N} d_{i,j} \quad \forall j \in 1,2 \quad (2)$$

In this equation, $d_{i,j}$ is the quantization distortion of a CTU$_{i,j}$, where i is the CTU index in a frame, j is the description and N is the total number of CTUs in each frame. pe is the probability of packet error. The term D_{error} is the distortion when the two descriptions are lost simultaneously. Thus D_{error} is a constant and can be omitted from the cost function. The problem is to find the set of $QP_{i,j}$ to use for each CTU$_{i,j}$ in a frame which minimizes the expected distortion D_e under a frame target bit rate R_t. The optimal MDC rate-constrained optimization problem is then given by :

$$\begin{aligned} \min_{QP_{i,j}} \quad & D_e \qquad \forall j \in 1,2 \quad \forall i \in 1,...,N \\ \text{s.t.} \quad & R_j = \frac{R_t}{2}, \\ & QP_{min} \leq QP_{i,j} \leq QP_{max} \end{aligned}$$

This problem can be solved using the standard Lagrangian approach and minimizing the following cost function :

$$J_{\lambda_1,\lambda_2}(R_1,R_2) = D_e + \sum_{j=1}^{2} \lambda_j (R_j - R_t/2) \quad (3)$$

As the two descriptions are independent from each other, we can therefore establish :

$$J_{\lambda_1,\lambda_2}(R_1,R_2) = J_{\lambda_1}(R_1) + J_{\lambda_2}(R_2) \quad (4)$$

where $J_{\lambda_j}(R_j)$ contains only the terms of the corresponding description j and is given by :

$$J_{\lambda_j}(R_j) = (1-p_e)D_{p,j} + p_e(1-p_e)D_{r,j} + \lambda_j(R_j - R_t/2) \quad (5)$$

The solution to the optimization-constrained problem is given by the first-order conditions, leading to :

$$\begin{cases} \dfrac{\partial J_{\lambda_j}(R_j)}{\partial \lambda_j} = 0 \qquad \forall j \in \{1,2\} & (6) \\ \dfrac{\partial J_{\lambda_j(R_j)}}{\partial R_{i,j}} = 0 \qquad \forall j \in \{1,2\} \quad \forall i \in \{1,..,N\} & (7) \end{cases}$$

Due to the nonlinearity of the scalar quantizer in HEVC, minimizing the cost function is not straightforward. As a result, the rate-distortion function $d_{i,j}(R_{i,j})$ is not continuous and therefore not differentiable. To address this challenge, we approximate the rate-distortion relationship using the exponential function, given by :

$$d_{i,j}(R_{i,j}) = a_i e^{b_i R_{i,j}} \quad \forall i,j \quad (8)$$

where the parameters a_i and b_i are estimated using linear regression in each CTU of the residual frame produced by a pre-encoding process before encoding the two descriptions. Then, we can solve the problem using Algorithm 1, where $R_{i_{min}}$ and $R_{i_{max}}$ are given by QP_{max} and QP_{min} repectively for each CTU$_{i,j}$. To prevent error propagation, the Instantaneous Decoder Refresh (IDR) frame allows the encoder to send an intra-frame signal to the decoder, clearing the Central Reconstructed Buffer. All frames can then be decoded from this IDR frame. Therefore, the encoder needs to select the appropriate amount of IDR frames to achieve the best coding quality concerning the channel noise. The study [9] has shown that the optimal IDR frame period under i.i.d. packet error distribution is given by :

Algorithm 1 Minimizing cost function (5)

Initialize $R_t, \lambda_{MAX}, \lambda_{MIN}, \epsilon, R_{i_{max}}, R_{i_{min}}$
while $|R_j - \frac{R_t}{2}| > \epsilon$ **do**
  $\lambda_j \leftarrow \frac{\lambda_{MAX}+\lambda_{MIN}}{2}$
  $\{R^*_{i,j}\} \leftarrow minimize(J_{\lambda_j}(R_j))$
  $\{R^*_{i,j}\} \leftarrow clip(R_{i_{max}}, R_{i_{min}}, R_{i,j})$
  if $R_j - \frac{R_t}{2} < 0$ **then**
    $\lambda_{MAX} \leftarrow \frac{\lambda_j+\lambda_{MIN}}{2}$
  else
    $\lambda_{MIN} \leftarrow \frac{\lambda_j+\lambda_{MAX}}{2}$
  end if
end while

$$T_{IDR} = \frac{1}{p_e} \quad (9)$$

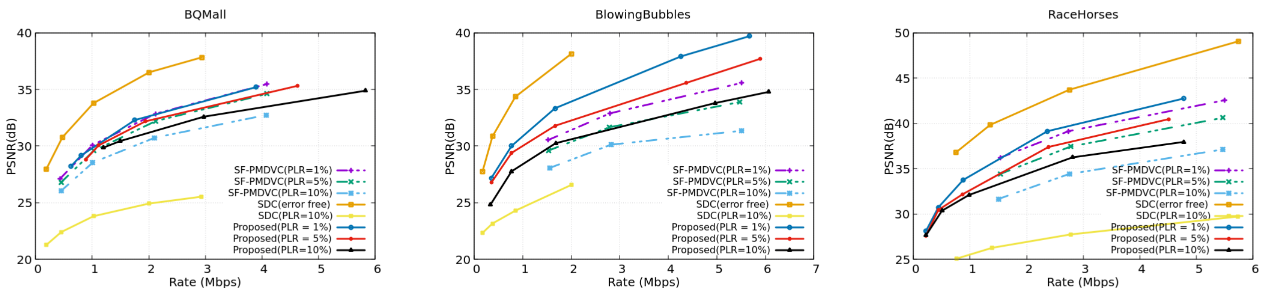

FIGURE 2 – *Comparison of the average Rate-Distortion with the method SF-PMDVC [7] under packet erasure : The experiment was conducted for three different packet loss rates : 0.1, 0.05, and 0.01. Each compressed stream was simulated three times over with different packet lost rate, and the average PSNR was computed. Our proposed MDC used LD-P configuration for encoding, while the referenced method used RA. SDC stands for Single Description Coding*

To decode the erroneous bitstream, if the principal CTU is lost, its redundant version will replace it. If two versions of the CTU are lost, the basic error concealment, which consists of replacing the block with the previous one, is applied.

3 Experimental Result

In this section, we evaluate the performance of our framework under a packet erasure channel. As mentioned earlier, the solution is implemented inside the HM codec [10]. To simulate the transmission, we use HEVC compressed streams with varying packet loss rates. Figure 2 shows that our proposed solution, which employs the LD-P configuration, outperforms the SF-PMDVC method with Random Access (RA) configuration for high packet error rates and high-motion sequences like BQ-Mall, RaceHorses, and BlowingBubbles, while having a lower complexity encoding profile. Therefore, our proposed method is better suited for real-time applications.

4 Conclusion

In this study, we proposed a spatial-based multiple description encoding bit allocation and decoding solution that is adapted to HEVC standard. Our proposed MD coding scheme includes a bit allocation that distributes the redundancy between descriptions by adjusting the QP value for each CTU within a frame based on the channel characteristics and an IDR adaptation. This solution meets the requirements of low latency and good compression performance, making it suitable for use in remote control vehicles.

In our perspective, a more robust error handling mechanism with finer grain error detection at the CU level will enhance the performance of the system. Various methods such as those discussed in [11, 12], could be employed to improve the decoding performance. Additionally, a scheme of optimization with error mismatch propagation model should improve the performance of the system.

Appendix

This is the summary of the article that has been submitted for the IEEE ICIP 2023 conference and is currently awaiting the review process.

Références

[1] C. Soldani, G. Leduc, F. Verdicchio, et A. Munteanu. Multiple description coding versus transport layer fec for resilient video transmission. Dans *International Conference on Digital Telecommunications (ICDT'06)*, pages 20–20, 2006.

[2] Tammam Tillo, Marco Grangetto, et Gabriella Olmo. Redundant slice optimal allocation for H.264 multiple description coding. *IEEE Trans. Circuits Syst. Video Technol.*, 18(1) :59–70, 2008.

[3] Ivana Radulovic, Pascal Frossard, Ye-Kui Wang, Miska M. Hannuksela, et Antti Hallapuro. Multiple description video coding with H.264/AVC redundant pictures. *IEEE Trans. Circuits Syst. Video Technol.*, 20(1) :144–148, 2010.

[4] Chia-Wei Hsiao et Wen-Jiin Tsai. Hybrid multiple description coding based on H.264. *IEEE Trans. Circuits Syst. Video Technol.*, 20(1) :76–87, 2010.

[5] Kostas E. Psannis. HEVC in wireless environments. *J. Real Time Image Process.*, 12(2) :509–516, 2016.

[6] Muhammad Majid, Muhammad Owais, et Syed Muhammad Anwar. Visual saliency based redundancy allocation in HEVC compatible multiple description video coding. *Multim. Tools Appl.*, 77(16) :20955–20977, 2018.

[7] Feifeng Wang, Jing Chen, Huanqiang Zeng, et Canhui Cai. Spatial-frequency HEVC multiple description video coding with adaptive perceptual redundancy allocation. *J. Vis. Commun. Image Represent.*, 88 :103614, 2022.

[8] Mohamed Aymen Labiod, Mohamed Gharbi, François-Xavier Coudoux, Patrick Corlay, et Noureddine Doghmane. Cross-layer scheme for low latency multiple description video streaming over vehicular ad-hoc networks (vanets). *AEU - International Journal of Electronics and Communications*, 104 :23–34, 2019.

[9] Guy Côté et Faouzi Kossentini. Optimal intra coding of blocks for robust video communication over the internet. *Signal Process. Image Commun.*, 15(1-2) :25–34, 1999.

[10] High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Encoder Description Update 10 | MPEG.

[11] Trung Hieu Le, Marc Antonini, Marc Lambert, et Karima Alioua. Codage vidéo à description multiple basé sur hevc pour le pilotage de véhicules semi-autonomes. Dans *GRETSI*, 2022.

[12] Marie Andrée Agostini, Marc Antonini, et Michel Kieffer. Map estimation of multiple description encoded video transmitted over noisy channels. Dans *ICIP*, pages 3069–3072. IEEE, 2009.