Error resilience analysis of motion vector prediction in HEVC

J. Carreira[1,2], S. Faria[2], P. Assuncao[2], E. Ekmekcioglu[1], A. Kondoz[1]
[1]Institute for Digital Technologies, Loughborough University in London, UK;
[2]Instituto de Telecomunicações / Instituto Politécnico de Leiria, PT;
{*jcarreira,sergio.faria, amado*}*@co.it.pt*

***Abstract*—The High Efficiency Video Coding standard achieves a high level of compression efficiency when compared to previous standards. As a counter effect, its error resilience is decreased and performance under error prone conditions is greatly affected. In this paper the error resilience of HEVC is evaluated, focusing on the new motion vector prediction technique referred to as Merge Mode, by analysing the impact of using different number of motion vector (MV) candidates and temporal MV dependencies. Since spatial dependencies are not the most relevant for error resilience, the focus of this performance study is on the temporal dependency of motion information due to its greater impact on error propagation. The results indicate that quality gains can be achieved by disabling the temporal motion vector candidate (TMVP), without compromising the coding efficiency. In our experiments average quality gains up to 5.17 dB (PSNR) can be achieved when TMVP is disabled for a 10% of loss ratio.**

***Index Terms*—HEVC, error resilience, motion vector prediction, merge mode, temporal dependencies.**

I. INTRODUCTION

The increasing diversity of multimedia applications and services, and the emergence of Ultra-HD formats (*e.g.*, $4k \times 2k$ or $8k \times 4k$ resolution) are reinforcing the requirements for coding efficiency superior to H.264/AVC capabilities. This need is even stronger when higher resolutions are needed for stereo and multi-view video. Moreover, the increasing amount of traffic generated by mobile multimedia applications demanding for better quality and higher resolutions are imposing new challenges to the existing networks. To match more challenging requirements, the High Efficiency Video Coding (HEVC) [1] was the latest standard developed by the Joint Collaborative Team on Video Coding (JCT-VC). The HEVC is essentially aimed at existing applications based on H.264/AVC, and extending these in two keys aspects: increased video resolution and use of parallel processing architectures. To increase coding flexibility and efficiency, the HEVC standard adopts a new block partition structure, enabling block sizes up to 64×64 [2]. The standard also improves the intra [3] and inter [4] coding, and includes new high-level features [5], such as, explicit reference picture management with improved error resilience, and a new parameter sets.

More efficient prediction structures allow higher bitrate saving, despite other disadvantages, such as complexity increasing [6] and error robustness decreasing. Although complexity may not pose significant problems due to the rapid evolution of hardware technologies, error robustness is strongly affected under packet loss conditions. The HEVC standard enables complex prediction modes for both pixels and MV, which contribute for achieving higher compression ratios. The introduction of the so called Merge Mode in the standard highly increases the dependencies between MVs at the spatial and temporal level. Since more dependencies are introduced, in the presence of transmission errors or data loss, error propagation is increased through coding units for several frames. To devise efficient methods for robust coding, it is necessary to first evaluate the impact of packet losses on the error resilience performance of such modes.

In this paper the error resilience of the emerging HEVC standard is analysed, considering random packet losses at various rates. This study extends previous ones by focusing in the HEVC Merge Mode, which is a new coding technique introduced in this standard. As mentioned before, the Merge Mode allows for spatial and temporal MV predictions. Therefore, our study covers both types of prediction separately. Finally, some approaches will be discussed to cope with the underlying problems presented along the paper.

The remainder of the paper is organised as follows. Section II describes previous studies addressing error resilience of the HEVC standard and several techniques to improve it. Section III describes the Merge Mode method to improve motion vector coding in HEVC. Section IV evaluates the error resilience performance of the HEVC standard focusing on the motion vector coding modes. Finally, Section V concludes this paper.

II. RELATED WORK

In this section, we aim to provide an overview of previous work related with error resilience of the HEVC standard. HEVC streams are used to deliver compressed video across different network technologies, therefore, there are some related works worth to be mentioned. A system integration of HEVC with existing technologies was presented in [7], showing that HEVC is suitable for use with well-known existing techniques, such as RTP, MPEG-2 TS and MPEG-DASH. Both scalability and parallel processing features are allowed to be used within these technologies. An end-to-end framework

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT), PhD Grant SFRH/BD/86368/2012) and Project UID/EEA/50008/2013.

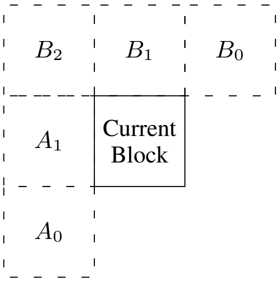

Fig. 1. Spatial motion vector predictors (MVP) used in the Merge Mode of HEVC.

for HEVC streaming based on RTP was proposed in [8], where the perceived video quality was analysed at different bitrates, providing relevant insights for video streaming. The results show that HEVC is sensitive to bandwidth reduction which decrease the quality up to 4 dB (10% of reduction).

The error robustness of HEVC is further analysed in [9], and compared against the H.264/AVC. Experimental results show that HEVC has reduced error robustness despite its increased coding efficiency. A subjective analysis is presented in [10] revealing that packet loss ratios higher than 3% have a significant effect on the perceived video quality. In both works it is concluded that the robustness of HEVC decreases for high motion sequences. In [11], the vulnerability of MV prediction is evaluated, showing its weak error resilience in contrast with the coding improvements. To overcome this vulnerability, the temporal motion vector predictor is disabled at constant frame intervals. In [12] authors extended the idea to the block level achieving a more robust MV prediction, without compromising the error resilience. To improve the error resilience of HEVC, the work described in [13] use redundant motion vectors to reduce the prediction mismatch of motion vectors at the decoder. To achieve reduced redundancy ratios the spatial dependencies are analysed in order to select a sub-set of MVs to transmit as auxiliary information. Although previous studies covered the performance of the HEVC standard under error conditions, a study covering the quality reduction due to the spatial and temporal MV predictors has not been presented.

III. MOTION VECTOR PREDICTION IN HEVC

In HEVC, new motion vector prediction modes are introduced, such as the advanced motion vector prediction (AMVP) and Merge Mode [4]. This work is primarily focused on the error robustness of the Merge Mode, which predicts MVs for the current block from its neighbours, which obviously leads to prediction mismatches at the decoder side in case of packet loss. The Merge Mode is based on the same concept of the H.264/AVC skip mode, with two main differences. Firstly, the Merge Mode allows for more motion vector candidates, where an index is added to select one MV out of several candidates. Secondly, it explicitly identifies the reference picture index used for the selected candidate, as well as the reference picture list. This increases the flexibility, when compared to the predefined values used by H.264/AVC. The MV candidates allowed in the Merge Mode and their positions are illustrated in Figure 1. Moreover, a temporal motion vector candidate is also used, derived from the co-located position on the temporal adjacent frame.

TABLE I
CHARACTERIZATION OF THE TEST SEQUENCES.

Sequence	Resolution	Description
Basketball Drill	832 × 480@50	High motion with several basket ball players
Book Arrival	1024 × 768@30	Low translational motion with two moving persons
BQSquare	416 × 240@60	Moderate outside motion with moving camera capturing from high point
Kendo	1024 × 768@30	Moderate motion with two moving persons, and moving camera
Race Horses	832 × 480@30	Moderate motion with several horse riders
Park Scene	1920 × 1080@24	Moderate motion with cyclists passing across the scene
People On Street	2560 × 1600@24	With point capture of people moving; high motion and texture complexity
Tennis	1920 × 1080@24	High motion with one moving person in the scene

Compared to the H.264/AVC standard, HEVC allows more spatial candidates, and also introduces the temporal MV candidate. Although the increased flexibility, the introduction of the temporal candidate in the set of possible MVs leads to temporal dependencies between subsequent frames. This may lead to higher error propagation and inherent quality degradation in the presence of packet loss. Thus, it is important to evaluate the impact of losing the motion information in HEVC.

IV. ERROR RESILIENCE OF HEVC MERGE MODE

In this section, the error resilience of the HEVC standard is analysed. This analysis is focused on the use of the temporal motion vector prediction introduced in the motion compensated modes of HEVC.

Seven well-known video sequences with 240 frames each and different resolutions were used in the experiments. Table I presents a summary of the sequences' characteristics. These test sequences were selected to cover different types of motion and texture complexity. The experimental results were obtained using the HEVC reference software, version 16.2. The sequences were encoded using an IDR period of 32 frames and the recommended test configurations: Lowdelay (GOP size of 4 P-frames) and Random Access (GOP size of 8 B-frames) [14]. In order to simulate a more realistic transmission environment, each NAL unit was defined with a maximum size of 1200 bytes to avoid fragmentation. Therefore, each coded frame was divided into several slices packets. In the experiments, random packet loss was simulated using a two-state Markov model. At the decoder side, in order to recover the missing frames, motion vector extrapolation is performed from the closest decoded frame. This recover the missing motion field, which is used to reconstruct the missing frame using motion compensation.

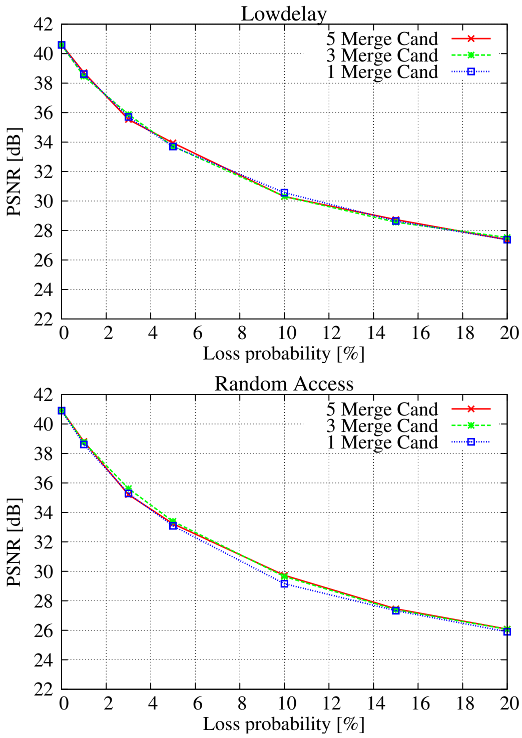

Fig. 2. Average PSNR for Basketball Drill encoded at $2Mbps$.

TABLE II
RELATION BETWEEN THE NUMBER OF AVAILABLE MV CANDIDATES AND ITS USAGE

Max. number of candidates	**Candidate selected (ratio)**				
	1	2	3	4	5
3 MVs	66.17%	22.79%	11.04%	-	-
5 MVs	64.54%	22.24%	7.38%	3.37%	2.47%

A. *Spatial MV candidates*

The first set of experiments aimed to evaluate the impact of using several MV candidates in the Merge Mode. As mentioned before, HEVC allows up to five candidates to be used, which may increase the amount of spatial dependencies. In order to check the impact of the MV candidates, different coding configurations were used, by reducing the number of available candidates. Figure 2 shows the average quality results for the Basketball Drill sequence under different packet loss conditions. The results show that, besides the amount of MV candidates, the error resilience of the HEVC is not compromised.

The occurrence rate for each spatial MV candidate was analysed in order to check if they are uniformly used. Table II shows the use of each MV candidate when three and five candidates are available. The results indicate that most predictions use the first candidate and minor differences occur by adding 2 or more MVs. Since the mostly used MVs are the initial ones, changing the amount of available candidates does not significantly affects the error resilience. Moreover, the slice partitioning scheme used (*i.e.*, 1200 bytes per slice) adds more refresh points in each frame reducing the spatial error propagation due to mismatch MV predictions.

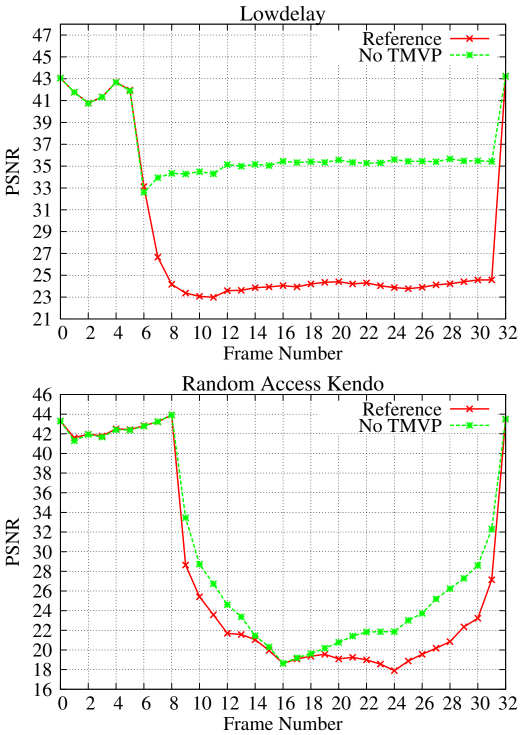

Fig. 3. Error propagation for Kendo encoded at $1Mbps$ (Frame #6 is lost in the Lowdelay configuration and Frame #16 is lost in the Random Access configuration).

B. *Temporal MV candidates*

The second set of experiments aimed to find the influence of losing temporal motion predictors on the error resilience of the HEVC standard. In order to perform this test, a single frame is lost and the error propagation is evaluated for two configurations: with the temporal MV candidate enabled (Reference) and disabled (No TMVP). Figure 3 illustrates the error propagation, when a single frame is lost. The results show that when the TMVP mode is enabled (Reference encoding), the loss of Frame #6 in the Lowdelay leads to significant reduction in the reconstruction quality of the subsequent frames. On the contrary, when the temporal MV candidate is disabled (No TMVP), motion information on subsequent frames is not affected by the loss of a single frame, since it is not temporal dependent, achieving significant quality increasing. This is more noticeable in the Lowdelay configuration, since more error propagation occurs in P-frames and it is propagated for several GOPs until an I-frame is decoded.

In order to further evaluate the impact of the temporal MV dependencies, the HEVC streams were subject to random packet losses. Table III presents the average quality results, measured using the PSNR. As expected, the results indicate that in error free environment, HEVC achieves higher quality by enabling the TMVP candidate, since it presents higher compression efficiency due to more MV predictors in the Merge Mode. This is true for all common test conditions, *i.e.*, both Low Delay and Random Access. However, in the presence of errors, the quality significantly decreases due to

TABLE III
PSNR UNDER RANDOM PACKET LOSS.

Sequence	**TMVP**	**Error Free**	**Packet loss rate**		
			1%	**5%**	**10%**
Lowdelay configuration					
Bask. Drill	Enabled	38.69	34.73	27.92	25.41
	Disabled	-0.04	+1.97	+3.93	+3.43
Book Arrival	Enabled	40.62	36.78	30.78	27.29
	Disabled	-0.01	+1.98	+3.32	+3.33
BQSquare	Enabled	39.64	31.99	23.45	20.95
	Disabled	-0.05	+6.14	+10.6	+10.7
Kendo	Enabled	43.31	38.80	29.80	25.45
	Disabled	-0.04	+2.04	+4.51	+4.98
Park Scene	Enabled	36.89	33.78	28.09	25.77
	Disabled	-0.06	+2.33	+5.72	+6.05
Race	Enabled	37.06	29.64	22.92	20.35
	Disabled	-0.07	+3.72	+4.78	+4.19
Tennis	Enabled	41.61	33.74	25.73	22.68
	Disabled	-0.04	+3.10	+4.01	+3.52
Average difference		**-0.044**	**+3.044**	**+5.317**	**+5.170**
Random Access configuration					
Bask. Drill	Enabled	39.48	36.84	30.50	27.15
	Disabled	-0.03	+0.25	+0.71	+0.61
Book Arrival	Enabled	40.94	38.38	32.43	28.55
	Disabled	-0.01	+0.38	+0.61	+1.00
BQSquare	Enabled	41.47	35.86	26.77	23.13
	Disabled	-0.08	+0.74	+1.58	+1.7
Kendo	Enabled	43.90	40.08	29.77	25.71
	Disabled	-0.04	-0.05	+0.55	+0.45
Park Scene	Enabled	37.90	35.26	30.51	27.74
	Disabled	-0.06	+0.67	+1.29	+1.34
Race	Enabled	37.43	32.36	25.35	22.47
	Disabled	-0.05	+0.61	+1.08	+1.00
Tennis	Enabled	41.78	36.33	28.80	25.49
	Disabled	-0.04	+0.68	+0.74	+0.67
Average difference		**-0.044**	**+0.469**	**+0.937**	**+0.967**

low error robustness when temporal MV candidates are used, resulting in an average of 2.65 dB of PSNR increasing when temporal MV predictor is disabled across all configuration tested. The results also show that the negative impact of using temporal MV dependencies decreases as the number of B-frames increases. The average gain of disabling the temporal MV predictor is 4.51 dB (Low Delay) and 0.78 dB (Random Access). This also confirms the previous results of Figure 3. A higher negative impact is obtained when only P-frames are used, since subsequent frames are temporally closer, leading to more accurate temporal MV candidates. Moreover, the results also indicate that for higher loss ratios the negative impact of using the TMVP does not increase as both configuration are subject to high quality degradation.

Summarising, the temporal MV candidate included in the Merge Mode improves the flexibility of the HEVC encoder, especially when no neighbour candidates are available (*e.g.*, corner positions of the frame). However, as discussed above, this increases the temporal dependencies, which lead to higher quality degradation when a frame loss occurs.

V. DISCUSSION AND CONCLUSIONS

The results shown in this paper indicate that the Merge Mode used in the HEVC standard decreases its error resilience due to the fact that more dependencies are introduced. This dependencies are not only at the pixel level, but also at the motion information level leading to incorrect MV decoding. Therefore, it is relevant to develop robust coding techniques to make HEVC bitstreams less prone to errors. In previous standards, redundant information was introduced to increase the resilience of the video transmissions. In the HEVC standard these techniques should also focus in MV predictions, decreasing the probability of incorrect decoding of MVs in case of packet loss. The analysis in this paper provide relevant insights to the robustness of HEVC motion vector prediction, which should be taken into account to achieve efficient robust video transmission.

REFERENCES

[1] G.J. Sullivan, J. Ohm, Woo-Jin Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] Il-Koo Kim, Junghye Min, T. Lee, Woo-Jin Han, and JeongHoon Park, “Block partitioning structure in the HEVC standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1697–1706, Dec. 2012.

[3] J. Lainema, F. Bossen, Woo-Jin Han, Junghye Min, and K. Ugur, “Intra coding of the HEVC standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1792–1801, 2012.

[4] Jian-Liang Lin, Yi-Wen Chen, Yu-Pao Tsai, Yu-Wen Huang, and Shawmin Lei, “Motion vector coding techniques for HEVC,” in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2011, pp. 1–6.

[5] R. Sjoberg, Ying Chen, A. Fujibayashi, M.M. Hannuksela, J. Samuelsson, Thiow Keng Tan, Ye-Kui Wang, and S. Wenger, “Overview of HEVC high-level syntax and reference picture management,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1858–1870, Dec. 2012.

[6] G. Correa, P. Assuncao, L. Agostini, and L.A. da Silva Cruz, “Performance and computational complexity assessment of high-efficiency video encoders,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1899–1909, Dec. 2012.

[7] T. Schierl, M.M. Hannuksela, Ye-Kui Wang, and S. Wenger, “System layer integration of high efficiency video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1871–1884, Dec. 2012.

[8] J. Nightingale, Qi Wang, and C. Grecos, “HEVStream: a framework for streaming and evaluation of high efficiency video coding (HEVC) content in loss-prone networks,” *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 404–412, May 2012.

[9] B. Oztas, M.T. Pourazad, P. Nasiopoulos, and V.C.M. Leung, “A study on the HEVC performance over lossy networks,” in *19th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Dec. 2012, pp. 785–788.

[10] J. Nightingale, Qi Wang, C. Grecos, and S. Goma, “The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming,” *IEEE Transactions on Consumer Electronics*, vol. 60, no. 2, pp. 242–250, May 2014.

[11] Bin Li, Jizheng Xu, and Houqiang Li, “Parsing robustness in high efficiency video coding - analysis and improvement,” in *IEEE Visual Communications and Image Processing*, Sept. 2011, pp. 1–4.

[12] J. Carreira, V. De Silva, E. Ekmekcioglu, A. Kondoz, P. Assuncao, and S. Faria, “Dynamic motion vector refreshing for enhanced error resilience in HEVC,” in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 281–285.

[13] J. Carreira, E. Ekmekcioglu, A. Kondoz, P. Assuncao, S. Faria, and V. De Silva, “Selective motion vector redundancies for improved error resilience in HEVC,” in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2014, pp. 2457–2461.

[14] F. Bossen, “Common test conditions and software reference configurations, document JCTVC-H1100,” Feb. 2012.