DYNAMIC MOTION VECTOR REFRESHING FOR ENHANCED ERROR RESILIENCE IN HEVC

J. Carreira[12*], *V. De Silva*[4], *E. Ekmekcioglu*[1], *A. Kondoz*[3], *P. Assuncao*[2], *S. Faria*[2]

[1]I-Lab, University of Surrey, UK;
[2]Instituto de Telecomunicações, Polytechnic Institute of Leiria/ESTG PT;
[3]Institute for Digital Technologies, Loughborough University in London, UK;
[4]Apical Ltd, Leicestershire, UK;
*j.f.carreira@surrey.ac.uk

ABSTRACT

The high level of compression efficiency achieved by HEVC coding techniques decreases the error resilience performance under error prone conditions. This paper addresses the error resiliency of the HEVC standard, focusing on the new motion estimation tools. It is shown that the temporal dependency of motion information is comparatively higher than that in the H.264/AVC standard, causing an increase in the error propagation. Based on this evidence, this paper proposes a method to make intelligent use of temporal motion vector (MV) candidates during the motion estimation process, in order to decrease the temporal dependency, and improve the error resiliency without penalising the rate-distortion performance. The simulation results show that the proposed method improves the error resilience under tested conditions by increasing the video quality by up to 1.7 dB in average, compared to the reference method that always enables temporal MV candidates.

***Index Terms*—** HEVC, video coding, error resilience.

1. INTRODUCTION

The growing popularity of HD video and the emergence of Ultra-HD formats (*e.g.*, 4k or 8k resolution) are creating strong needs for coding efficiency beyond to H.264/AVC capabilities. Thus, a new coding standard was developed by the Joint Collaborative Team on Video Coding (JCT-VC), referred to as High Efficiency Video Coding (HEVC) [1]. The HEVC aims to address essentially all existing applications of H.264/AVC, extending it in two key aspects: increased video resolution and increased use of parallel processing architectures. In order to increase the flexibility and follow the demand for increased video resolution, the HEVC standard adopts a new block partition structure, enabling block sizes of up to 64×64 [2]. The standard also improves the intra and inter [3] coding techniques, and includes new high-level features.

New coding tools bring not only compression efficiency, but also other disadvantages, such as increased complexity [4] and lower error robustness. The HEVC standard enables the motion vectors (MVs) of the current frame to be predicted not only using the surrounding MVs, but also MVs from the temporally adjacent frame. The introduction of the temporal motion vector predictor (TMVP) in the HEVC design leads to increased dependencies between MVs of different frames. Since the motion vectors in the reference frame will be used as the predictors, the loss of one frame will affect the decoding of the MVs in subsequent slices which use the lost frame as reference. Hence, the error propagation rate increases.

Related with HEVC errors resilience and transmission that some works worth to be mentioned. A system integration of HEVC with existing technologies, such as RTP and MPEG-2, was presented in [5], evaluating the compatibility of the new scalable features of the standard with existing techniques. An end-to-end framework for HEVC streaming was proposed in [6], where the perceived video quality was analysed at different bitrates providing relevant insights for video streaming. The error robustness of HEVC was also analysed in [7], and compared against the H.264/AVC. The results had shown that HEVC has reduced error robustness despite its increased coding efficiency. In [8], the vulnerability of the MV prediction was evaluated under different packet loss conditions, showing the lack of error robustness in the HEVC standard. To overcome such vulnerability in HEVC, a method is proposed to limit the TMVP at frame level. Although previous studies [7, 8] covered the performance of the HEVC standard under error conditions, a study covering the quality reduction due to the temporal MV predictors and its influence when compared with the H.264/AVC has not been done yet.

In this paper the error robustness of the emerging HEVC standard is firstly analysed, considering random packet losses at various rates. This study is focused on the impact on error

This work was supported by the Fundação para a Ciência e a Tecnologia, Portugal (under the PhD grant SFRH/BD/86368/2012), and the ROMEO project (grant number: 287896), which was funded by the EC FP7 ICT collaborative research programme.

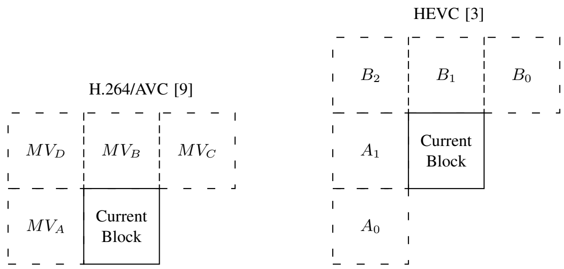

Fig. 1: Motion vector predictor (MVP) in the two state-of-the-art video coding standards.

resilience of MV loss and usage of the TMVP. The quality degradation influenced by the temporal dependencies is also analysed. The performance is also compared against that of H.264/AVC under error prone conditions. To improve the error resilience, a new approach to use the temporal MV predictor is proposed, which reduces the temporal dependencies at the block level, enhancing the error recovery.

The rest of the paper is organized as follows. Section 2 briefly compares the existing SKIP mode in H.264/AVC and the new inter coding modes included in the HEVC standard. Then, in sub-section 2.2 the error resilience evaluation of the HEVC bitstreams under single and random loss conditions is presented. The proposed method to reduce the temporal dependencies and improve the error resilience is described and evaluated in Section 3 and 4, respectively. Finally, the conclusions are drawn in Section 5.

2. ANALYSIS OF THE HEVC ERROR RESILIENCE

In this section the error resilience of the HEVC standard is analysed. The analysis is focused on the usage of the temporal motion vector predictor introduced in the motion compensated modes of HEVC. A brief comparison to the SKIP mode in H.264/AVC and the Merge Mode in HEVC is drawn in the first sub-section. Then, an error resilience evaluation is presented in sub-section 2.2.

2.1. Comparison of the SKIP mode in H.264/AVC and Merge Mode in HEVC

In the H.264/AVC standard along different motion compensated coding techniques [9], the so-called SKIP mode was introduced. In this mode, neither quantized residual information, nor the MV is transmitted. This mode is useful for large areas with no change or constant motion. The MV for the SKIP mode is derived based on the median value of the following three neighbours: left, above and above right, as shown in Figure 1 (a) by MV_A, MV_B and MV_C, respectively. Whenever the MV_C is not available, the MV_D is used instead.

Table 1: Description of the sequences used in the experiments.

Sequence	Resolution	Description
Basketball Drill	832×480 50 fps	High motion with several basket ball players
Book Arrival	1024×768 30 fps	Low translational motion with two moving persons
BQSquare	416×240 60 fps	Moderate outside motion with moving camera capturing from high point
Kendo	1024×768 30 fps	Moderate motion with two moving persons, and moving camera
Tennis	1920×1080 24 fps	High motion with one moving person in the scene

In HEVC, a new method to derive the motion information from previously encoded MV is included, referred to as Merge Mode [3]. Although this mode is based on the concept of the skip mode of the H.264/AVC, it allows more MV candidates, leading to improved flexibility. The MV candidates allowed in the Merge Mode [3] are illustrated in Figure 1 (b). The figure illustrates the spatial candidates' positions (A_0, A_1, B_0, B_1, B_2). Moreover a temporal motion vector candidate is also used, derived from the co-located position on the temporal adjacent frame.

In contrast with the H.264/AVC standard, the HEVC standard allows the usage of a temporal MV predictor, as well as, more spatial candidates to predict the current motion vector. Despite introducing more flexibility, the dependencies between the MVs might increase, making the motion information more prone to errors. Especially the introduction of the temporal candidate in the set of possible MVs that increases the dependencies between subsequent frames, which may lead to higher quality degradation in presence of errors, when compared with H.264/AVC.

2.2. Error resilience performance

In this sub-section, the error resilience of the HEVC standard is evaluated under different network conditions, using different encoding configurations. Moreover, it is also compared against the H.264/AVC standard. Five well-known video sequences with 240 frames and different resolutions were used in the tests. Table 1 presents a summary of the sequences' features. The test sequences were selected to cover different types of motion and texture complexity. To test the error resilience, version 11.0 of HM, and version 18.3 of the JM reference software were used. The sequences were encoded using different bitrates, an IDR period of 32 frames and a GOP size of 1 (*i.e.*, I-P-P...) using one reference frame. Similar configurations were used in both H.264/AVC and HEVC cases. For different bitrates were used covering from 500 $kbps$ to 3.0 $Mbps$. In the experiments each frame was transmitted within one packet and random packet loss was simulated using a two-states Markov model. The sequences

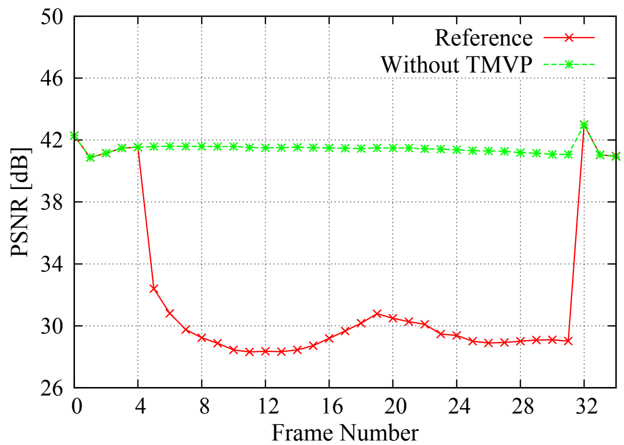

Fig. 2: Error propagation for the Book Arrival sequence encoded at $2Mbps$ (Frame #4 is lost).

Table 2: Bjontegaard's average PSNR (dB) differences between H.264/AVC and the HEVC reference and without TMVP (H.264/AVC as reference).

Sequence	HEVC configuration	Error free	1% loss	5% loss
Basketball Drill	Reference	1.685	0.149	-0.341
	Without TMVP	1.572	0.978	0.875
Book Arrival	Reference	0.876	-0.804	-1.252
	Without TMVP	0.846	0.309	0.357
BQSquare	Reference	1.168	-2.312	-3.785
	Without TMVP	1.130	0.709	0.616
Kendo	Reference	1.361	-0.927	-1.808
	Without TMVP	1.318	0.716	0.620

were decoded 50 times and the average quality was measured.

The first set of experiments aimed to find the influence of losing the motion information on the error resilience of the HEVC standard. To achieve this goal, the motion information was removed for a given frame, in order to be unavailable to decode the subsequent ones (hypothetical scenario). This means that the reconstruction is not affected by errors, but error propagation still occurs due to mismatch prediction of the MVs in the subsequent frames. Figure 2 illustrates the error propagation, when errors affect Frame #4. Note that the concealment of the lost temporal candidates is not performed and the decoded behaves as it is not available. The results show that when the TMVP mode is enabled (reference encoding), the loss of motion information in Frame #4 leads to significant reduction in the reconstruction quality of the subsequent frames. This is true even for sequences with low motion activity as per Table 1. On the contrary, when the temporal MV candidate is disabled (without TMVP), the sequence is not affected by the loss of motion vector information in Frame #4, since the original frame is used in the concealment process and the motion information is not temporal dependent, thus always correctly decoded.

In order to further evaluate the error resilience, HEVC and H.264/AVC bitstreams were subjected to random packet losses. In this experiment each coded frame was transmitted in one packet. Table 2 presents the Bjontegaard's average PSNR increase for the HEVC encoded bitstreams comparing with the H.264/AVC. The results are obtained taking as reference the H.264/AVC. In the HEVC two configuration were tested, *e.g.*, with the temporal MV candidate enabled (Reference) and disabled (Without TMVP). As expected, the results indicate that in a error free scenario, HEVC achieves higher quality, since it presents higher compression efficiency comparing with H.264/AVC. However, in case of errors the quality of HEVC significantly decreases due the lack of robustness of the temporal MV candidates, resulting in up to 4.3 dB of PSNR reduction in the sequence BQSquare at 5% of packet loss. Thus, the reference HEVC achieves lower quality than H.264/AVC encoded streams. However, the error robustness of HEVC increases when TMVP is disabled, outperforming the H.264/AVC standard.

Summarising, the temporal MV candidate included in the Merge Mode improves the flexibility of the HEVC encoder, especially when no neighbour candidates are available. However, as discussed above, this increases the temporal dependencies, which lead to higher quality degradation when a frame loss occurs. Therefore, it is necessary to develop an algorithm for reducing the temporal dependencies in order to achieve higher error robustness without fully disabling the temporal MV candidates.

3. PROPOSED DYNAMIC MOTION VECTOR REFRESHING FOR HEVC

In this section a new approach to deal with temporal motion vector dependencies is proposed, aiming at reducing the amount of temporally dependent MVs. This is accomplished by modifying the rate distortion optimisation process in HEVC. Thus, since the number of MVs that can be independently decoded without previous temporal references increases, the error recovery is improved after a frame loss event.

Regarding this work, it is worth to mention a previous approach to reduce the temporal dependencies in HEVC bitstreams [8]. The number of dependencies are reduced by coding the MVs of certain frames without inter-frame dependencies, thus the motion information of those frames have no temporal relation with the previous ones. Therefore the video decoding is not affected by errors in previous frames. In [8] temporal information decoding refresh (TIDR) pictures were introduced, and included every 8 frames. Although these frames are inter coded, they do not use temporal MV candidates as predictors for other MVs. Although the error robustness is improved by using refreshing points in the bitstream, error propagation still exists in between the refresh picture boundaries.

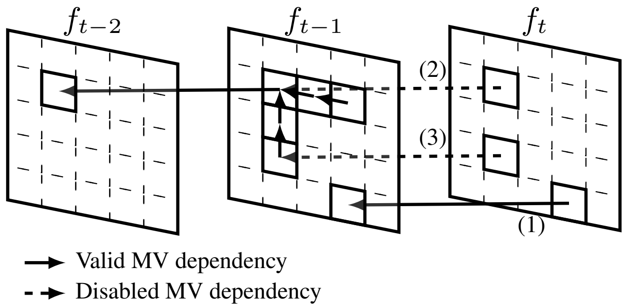

Fig. 3: Proposed approach to use the temporal motion vector candidate in the HEVC encoder.

In order to overcome this limitation, a new method that breaks the motion information dependencies between frames is proposed, to be used in the merge and AMVP modes. Instead of disabling the temporal motion candidate at the frame level, the proposed method aims to remove the dependencies at the block (CU) level. To achieve this, the temporal dependency of all MV candidates is analysed for each block, in order to decide on the suitable candidates to be used to predict the current MV. Based on the analysis of the temporal dependency, some temporal candidates are marked as unsuitable for the prediction of the current MV, and are not included in the candidates list. In the proposed approach, a given temporal MV candidate is marked as unsuitable based on the following criteria:

1. it was encoded based on another temporal MV candidate from a previous encoded frame;
2. it was predicted using a spatial neighbour that was previously encoded using a temporal MV candidate.

Figure 3 can be used to explain the proposed method, illustrating different blocks and different MV dependencies represented with arrows. Note that the squares illustrate a generic block size, and the representation of the temporal dependencies is simplified, and in some cases may not correspond to the exact co-located block in the previous frame. In frame f_{t-1} there are several MVs encoded using spatial predictions and one MV encoded using a temporal prediction (arrow pointing to f_{t-2}). In frame f_t, the MV corresponding to the block (1) may use the temporal MV candidate, since the co-located block in f_{t-1} is not temporally dependent. However, the MV prediction marked with the arrow (2) is not allowed in the proposed scheme in order to avoid propagation from temporal dependencies. Moreover, the block corresponding to the arrow (3) cannot use the temporal MV candidate since it already depends on a MV that was previously encoded using the temporal candidate. The proposed method reduces the temporal dependencies of the MVs at the block level, providing improved error robustness to the encoded streams. Using this approach, the dependencies are selectively removed, improving the error recovery, without fully disabling the temporal MV candidates for a given frame.

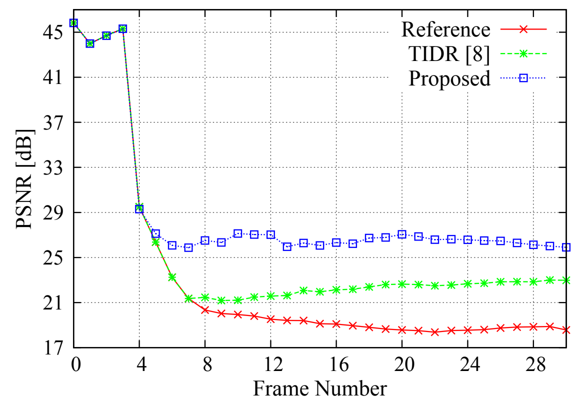

Fig. 4: Quality obtained with the different approaches tested in the HEVC when the Frame #4 is missing for the BQSquare sequence @ 1.0 $Mbps$.

Table 3: Average standard deviation of the decoded video quality (PSNR) under different packet loss conditions.

Sequence	**TMVP configuration**	**1% loss**	**3% loss**	**5% loss**
Kendo	Reference	7.56	11.58	12.05
	TIDR [8]	7.25	11.00	11.55
	Proposed	6.62	10.06	10.66
Race Horses	Reference	5.74	8.31	8.55
	TIDR [8]	5.47	7.78	8.07
	Proposed	5.29	7.63	7.88
Tennis	Reference	6.70	9.55	10.07
	TIDR [8]	6.55	9.25	9.83
	Proposed	6.42	9.11	9.67

4. PERFORMANCE EVALUATION

In this section the performance of the proposed scheme is evaluated against two other approaches: reference HEVC with TMVP enabled and the TIDR method as proposed in [8]. To evaluate the proposed method, the sequences described in Table 1 were used, with the same configuration as in Section 2.

Firstly, the performance of the proposed method was measured under a single loss event. Figure 4 shows the error propagation for various tested methods. The results show a significant quality difference between the proposed method, the TIDR method, and the reference HEVC. When the temporal MV candidate is disabled (proposed and TIDR method) more error robustness can be achieved. The results also show the relevance of using the proposed method, as the TIDR is only able to recover from the frame loss after the Frame #8 (*i.e.*, the refresh frame).

Secondly, further tests were performed to evaluate the effectiveness of the proposed error resiliency method under realistic packet loss patterns. As in previous experiments, each coded frame is transmitted in one packet. Table 3 illustrates the average standard deviation of the PSNR, for three of the

Table 4: Bjontegaard's average PSNR (dB) differences for all sequences tested at different frame loss conditions.

Sequence	TMVP configuration	Error free	1% loss	5% loss
Basketball Drill	TIDR [8]	-0.011	0.092	0.365
	Proposed	-0.054	0.218	0.823
Book Arrival	TIDR [8]	-0.003	0.099	0.428
	Proposed	-0.012	0.359	1.436
BQSquare	TIDR [8]	-0.005	0.418	1.572
	Proposed	-0.032	0.539	1.967
Kendo	TIDR [8]	-0.008	0.150	0.618
	Proposed	-0.031	0.380	1.539
Tennis	TIDR [8]	-0.003	0.099	0.494
	Proposed	-0.041	0.214	0.932
Average	TIDR [8]	-0.006	0.171	0.695
	Proposed	-0.034	0.342	1.339

sequences tested. The results show that increasing the packet loss rate leads to higher quality deviations, and the impact of errors in the subjective quality experience is higher. However, reducing the temporal dependencies of the MVs using the proposed method decreases the quality differences variation, achieving the lowest value for all scenarios tested. The proposed method is able to reduce the quality difference variations up to 0.94 dB in Kendo sequence (3% of packet loss) compared to the TIDR method.

Table 4 shows the Bjontegaard's average PSNR increase. The results are obtained taking as reference the standard HEVC encoder. The results presented in Table 4 show that the proposed method presents practically the same rate-distortion performance when compared to TIDR method. In general, the average quality, in an error-free scenario, is only decreased by 0.028 dB in comparison to the TIDR method and by 0.034 dB in comparison to the reference encoder. However, the proposed method widely outperforms the reference TIDR technique [8] in terms of error resiliency performance, achieving in average higher overall quality under error-prone conditions. The effectiveness of the proposed method increases at higher packet loss rates (*e.g.*, 5%), achieving an average gain of 0.573 dB, and a maximum gain of 1 dB (PSNR) for the Book Arrival sequence, when compared to the TIDR technique. Higher gains are achieved when compared with the reference HEVC encoder, as shown in Table 4.

5. CONCLUSION

In this paper, the error resilience performance of the HEVC standard was analysed under error prone conditions. A new approach to deal with such drawback was proposed, addressing MV dependency at the block level. The proposed method selectively disables the temporal MV candidates, in order to break the dependencies across more than one frame. The experimental evaluation shows that the proposed method increases error robustness and outperforms an existing state-of-the-art method and the current standard. The results also show that the proposed method is able to provide a good trade-off between the coding performance and resilience against errors. Moreover, it is able to achieve quality improvements of up to 1.69 dB under packet loss conditions. In summary, this paper demonstrates that selective usage of temporal MV candidates is an efficient method to enhance error resilience in HEVC.

REFERENCES

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

[2] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1697–1706, 2012.

[3] J.-L. Lin, Y.-W. Chen, Y.-P. Tsai, Y.-W. Huang, and S. Lei, "Motion vector coding techniques for HEVC," in *IEEE 13th International Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2011, pp. 1–6.

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

[5] T. Schierl, M. Hannuksela, Y.-K. Wang, and S. Wenger, "System layer integration of high efficiency video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1871–1884, 2012.

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

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

[8] B. Li, J. Xu, and H. Li, "Parsing robustness in high efficiency video coding - analysis and improvement," in *IEEE Visual Communications and Image Processing*, Sep. 2011, pp. 1–4.

[9] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, Jul. 2003.