A Robust Video Encoding Scheme to Enhance Error Concealment of Intra Frames

J. Carreira[1,2], P. Assuncao[2], S. Faria[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;

joao.f.carreira@ieee.org

Abstract—In this paper a robust encoding scheme is proposed to improve the visual quality of HEVC decoded video when intra frames are lost along the streaming path. For this purpose, the encoding process includes frame loss simulation and subsequent error concealment, to find the most efficient method that should be used by a decoder to recover lost intra frames. In this novel scheme, each image is divided into partitions, which are associated with the error concealment method that achieves the lowest distortion. Then this information is signalled to the decoder through SEI messages in the coded stream. In order to efficiently use the signalling overhead, rate-distortion optimisation is used to achieve the best trade-off between the number of transmitted symbols and distortion of reconstructed frames. Experimental results show the effectiveness of the proposed method to enhance the quality of reconstructed intra frames under different packet loss ratios (PLR). For PLR=10%, the robust coding scheme is able to improve the average PSNR of all frames affected by errors, up to 1.50 dB and 3.44 dB in Low-Delay and Random-Access configurations respectively, at a maximum overhead cost of 0.24%.

I. INTRODUCTION

Recent developments in video compression and networking technologies are contributing to enhance existing services towards new UHD formats. Most of the upcoming applications and services will use the High Efficiency Video Coding (HEVC) standard, which complies with the requirement of increased coding performance, by adopting a new block partition structure and improving the intra and inter coding methods [1]. However, it is known that the higher the coding efficiency, the lower the robustness to transmission errors and data loss. This problem become even worse when intra frames are affected by errors, because these lead to mismatched predictions in all the subsequent dependent frames until a refresh point is reached. This is critical for the perceived quality since the error propagation gives a significant contribution for the overall quality degradation. As more demanding streaming services emerge, such as UHD video streaming, which are expected to provide high quality to the end users, there is a growing interest in methods capable of providing increased HEVC error resiliency by means of robust coding and error concealment.

The problem of robust video streaming has been addressed using error resilience techniques to mitigate the artefacts caused by packet losses. These techniques can be grouped into four categories: localisation (i.e., to reduce coding dependencies), data partitioning, redundant coding and concealment-driven techniques. The problem of spatial and temporal dependencies can be addressed through optimisation, based on the estimated decoded distortion [2], or by reducing the motion vector (MV) coding dependencies [3]. Alternatively, in order to increase error resilience, redundant pictures or multiple redundant streams might be used. Finally, concealment-driven techniques might be adopted at the encoder to ease the reconstruction of missing frames at the decoder [4]–[7].

In this work we investigate concealment-driven techniques to efficiently deal with intra frame loss in communication networks with low overhead. There are few works addressing this problem that are worth to be mentioned, as they provide relevant insights to the method proposed in this paper. In [4] an estimation of the error concealment distortion is used for slice classification and unequal error protection. Moreover, the error concealment distortion may also be taken in consideration to perform motion estimation [5], by including the distortion of the subsequent frame (assuming it is lost and recovered) in the rate-distortion (R-D) cost of the current frame. Alternatively, data hiding can be used at the encoder-side to multiplex extra motion information into DCT coefficients of intra-coded blocks, which is then used at the decoder to recover erroneous neighbouring blocks [6]. Although such methods facilitate the error concealment performed either through slice reordering or based on MVs, they always consider a fixed error concealment approach. Therefore, it is expected that higher performance could be achieved if several error concealment algorithms are considered, since the reconstruction quality can be optimised at the encoder-side. The work presented in [7] proposes testing several methods with transmission of extra symbols to indicate the best method that shall be applied by the decoder, using low overhead (approximately 3%). Although such approach achieves reasonably good performance, it relies on the information available in the neighbouring region, which may not be available in case of full frame loss.

In order to recover intra-coded frames in a video sequence it may be useful to use methods based on motion compensation, which are expected to perform better than spatial concealment methods, especially when large image regions are affected by errors. For instance, in [8], MVs of a previously received frame were extrapolated onto the missing frame, in order to estimate the missing pixels. In [9] the block partitions of the neighbouring frames were used to assist the motion extrapolation, in order to keep the object boundaries smooth. Moreover, residue information may be used to classify the motion information and select only the reliable MVs, as shown in [10]. Despite the goodness of the aforementioned approaches, without any assistance from the encoder-side, the decoder is not able to decide which method leads to the best reconstruction accuracy, which results in sub-optimal reconstruction.

To advance one step further, this paper proposes an error concealment-aware encoding scheme to increase the streaming robustness of HEVC intra frames. This method aims to enhance error concealment in the decoder by optimising it at the encoder-side, where the original signal can be used. It relies on simulation of data loss and test of different error concealment techniques within the encoding loop, in order to identify the best techniques to recover the lost information with the lowest distortion. This information is signalled to the decoder, allowing it to select the best technique
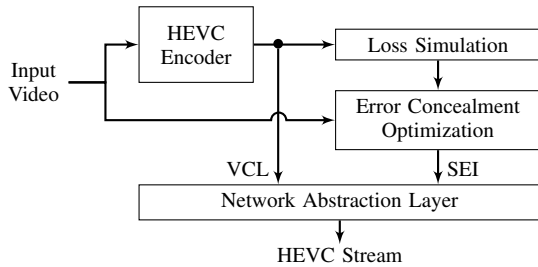
Fig. 1: Functional structure of the proposed encoding scheme.

$Depth$: depth of the coding tree unit quadtree partition

EC: selected concealment candidate for each partition unit

Fig. 2: Symbols transmitted through SEI messages.

whenever errors occur and reducing the error propagation. Finally, in order to keep the signalling overhead affordable, a variable number of symbols are transmitted for each coding tree unit (CTU).

The remainder of the paper is organised as follows. Section II describes the proposed concealment-aware encoding scheme and Section III presents and discusses the experimental results. Finally, Section IV concludes this paper.

II. PROPOSED CONCEALMENT-AWARE ENCODING

Figure 1 illustrates the block diagram of the proposed robust encoding scheme. The error concealment function in the HEVC encoder produces a reconstructed estimate of the lost frames using different methods, and selects the one which recovers the missing data with the lowest distortion. Subsequently, this is signalled to the decoder by multiplexing this information in the bitstream using the supplemental enhancement information (SEI) NAL units [11]. As shown in the figure the encoding process does not depend on the proposed method, therefore both can be applied simultaneous without increasing the processing time and introducing delay.

At the HEVC decoder, such information is used to guide the error concealment operation to recover erroneous frames with high accuracy. The error concealment process combines both motion estimation/compensation and extrapolation techniques to recover lost frames. The CTUs are partitioned to form a quadtree structure (not the same structure used by the video encoder) where each partition is specifically used to test different error concealment methods. Then, the partitioning and signalling of the error concealment methods is achieved through R-D optimisation.

A. Error concealment methods

At the encoder side, four candidate methods for error concealment (M1...M4) are tested for each intra frame that is considered to be lost, herein referred to as frame f_0 at time instant t_0, namely:

- M1, M2: These two methods are based on extrapolation of an estimated motion field;
- M3: This method is based on MV extrapolation of previously received MVs;
- M4: This method uses the co-located motion information from the closest neighbour;

The first two methods M1, M2 rely on an estimated motion field, which is calculated from the closest available neighbours, i.e., frames f_{-1} and f_{-2}, and then extrapolated at the pixel level and coding unit (CU) level to the missing frame instant. This results in two different reconstructions of the missing frame. The motion field is estimated using an optical flow method based on the Horn and Schunck smoothness condition [12], that forces the estimated field to be regular in som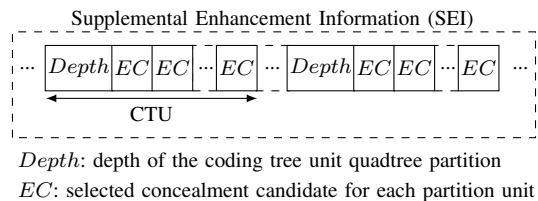e sense. Using this method, for each pixel $\mathbf{p}$, the motion field $v(\mathbf{p})$ is obtained by minimising the following energy function,

$$E(v) = \int_\Omega \left(|f_{-1}(\mathbf{p}) - f_{-2}(\mathbf{p} + v(\mathbf{p}))|^2 + |\nabla v(\mathbf{p})| \right) d\mathbf{p}, \ \mathbf{p} \in \Omega \ (1)$$

where f_{-1} and f_{-2} are two video frames, and Ω defines the sets of pixels of those frames. $|\nabla v(\mathbf{p})|$ is the smoothness condition derived from the assumption that neighbouring regions belong to the same object and, thus, these regions have similar motion. This method has been used because it presents higher performance than block matching algorithms.

The third candidate (M3) recovers the missing intra frame, using motion compensation through the correctly received MVs, of the frame f_{-1} at time instant t_{-1}, according to the following procedure: MVs of the neighbour frame (v_{-1}) are extrapolated to compute a set of MVs for the missing frame (v_0):

$$v_0 = \frac{t_0 - t_{-1}}{t_{-1} - t_R} \times v_{-1}, \qquad (2)$$

where t_R is the time instant of the reference frame pointed by the original motion vectors v_{-1}. Then, the MV associated to the block $b_{f_{-1}}(x, y)$ at position (x, y) in f_{-1} may be used to recover the block $b'_{f_0}(x', y')$ in f_0, at position (x', y') obtained from v_0 components as follows:

$$x' = x - v_{0x} \qquad (3)$$

$$y' = y - v_{0y} \qquad (4)$$

The fourth error concealment method (M4) is obtained by reconstructing the missing frame through motion compensation directly using the co-located vectors in the closest neighbour frame (f_{-1}). Finally, after testing the four methods M1...M4, the candidate technique that achieves the lowest distortion is encoded as a SEI message to be used by the decoder.

B. Error concealment selection and signalling

As described in the previous sub-section, the proposed method includes four error recovery techniques, which result in different estimated frames. The CTU are partitioned into a quadtree structure and for each partition the candidate which results in the lower error is multiplexed in the coded stream. In order to reduce the amount of overhead, larger partition sizes may be used, but lower performance is achieved as the optimum candidate that minimises the distortion may not be found. Therefore, a variable partition size is adopted to obtain the best trade-off between signalling overhead and reconstruction quality. Thus, for each CTU an extra symbol is transmitted to define the partition size to be used. Since partitioning forms a quadtree structure, the extra symbol indicates the maximum depth of the tree, starting with one 64×64 unit ($Depth = 0$) up to sixty four 8×8 units ($Depth = 3$). Figure 2 illustrates the structure of the signalling information for each CTU, where one $Depth$ symbol is transmitted followed by a variable number of EC symbols. These symbols indicate the selected error concealment method that must

be used for each partition. Since four methods are used, this results in an overhead of 2 bit per partition. In order to obtain an optimal $Depth$ value (d^*), the following minimisation is performed:

$$d^* = \arg\min_d \left\{ mse(d) + \lambda \times 4^d \times S \right\}, \qquad (5)$$

where mse is the mean of square differences for each CTU and $S = log_2(4)$ represents the number of bits per symbol. Following the minimisation in (5), a trade-off between distortion and overhead is achieved. In the proposed method $\lambda = 0.5$, in order to obtain a quality level similar to the case of fixed partition size. Note that λ can be controlled to achieve different levels of overhead, but the overhead is not independent from the video content and quantisation parameters.

The use of SEI allows to multiplex this information into the coded bitstream without affecting the syntax of the standard. Moreover, arithmetic coding [13] was used to reduce the overhead. Since the side information uses significantly less bitrate than the video signal, it is less likely to be hit during transmission. Furthermore, the signalling is transmitted using different packets in the transport layer, in order to reduce the probability of being lost alongside of the correspondent video packets. At the decoder-side, signalling bits are decompressed and used to recover the missing frames using the optimum concealment method. Finally, note that the proposed method works for any type of intra coded slice regardless its size, which can be a whole frame or a smaller image segment.

III. Experimental results

The performance of the proposed method (*Prop*) is evaluated in this section against the reference HEVC using motion-copy (*Ref*), which reconstructs the frames using the co-located MV in the closest available frame.

Six well-known video sequences were used in the experiments, namely: Basketball Drill (832×480), Kendo (1024×768), Park Scene (1920×1080), People On Street and Traffic (2560×1600) and Jockey (3840×2160). These test sequences were selected to cover different types of motion and texture complexity, as well as, spatial resolution. The experimental results were obtained using the HEVC reference software, version 16.2. These sequences were encoded using an intra-period of 16 frames and two recommended test configurations: Low-Delay (LD) and Random-Access (RA), corresponding to a GOP size of 4 P-frames and 8 B-frames, respectively [14]. Each NAL unit was limited to a maximum size of 1200 bytes to avoid fragmentation. Therefore, each coded frame was divided into several slice packets. The evaluation of the proposed method is performed only for intra frames to study their impact on error propagation. Moreover, as intra coding units result in higher amount of bits, they also need a higher number of packets. Results using this setup indicate that for an intra-period of 16 frames, approximately 42% of the packets corresponds to intra coded slices. Also, by limiting the proposed method to intra frames, the amount of required overhead is reduced.

Table I shows the overhead introduced by the proposed method and also for fixed size partitions of 32×32 and 16×16, i.e., $Depth = 1$ and $Depth = 2$. While the maximum overhead of these fixed partitions is 1.81% (Traffic) the proposed dynamic partitioning method is able to reduce the overhead to a maximum of 0.24% (Jockey).

Table II shows the average depth selected for CTUs and the usage ratio of each error concealment candidate. It is noticeable from the results that the proposed method leads to low average depth

TABLE I: Percentage of overhead (%)

Sequence	32×32 Depth: 1	16×16 Depth: 2	*Prop* Depth: dynamic
Basketball Drill	0.08	0.27	0.08
Kendo	0.27	1.05	0.19
Park Scene	0.42	1.72	0.05
People On Street	0.06	0.25	0.11
Traffic	0.50	1.81	0.06
Jockey	0.38	1.53	0.24

TABLE II: Usage ratio of each error concealment candidate.

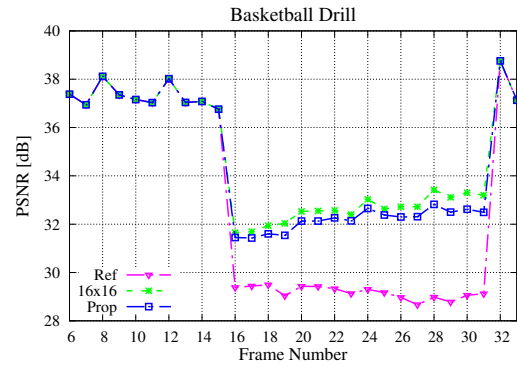| Sequence	Avg. Depth	Avg. usage ratio (%)			
		M1	M2	M3	M4
Basketball Drill	0.60	42.00	27.27	17.93	12.80
Kendo	0.31	38.69	36.63	17.20	7.48
Park Scene	0.25	40.66	31.99	20.46	6.89
Traffic	0.13	49.93	27.32	15.99	6.76
Jockey	0.19	41.38	20.02	22.51	16.09

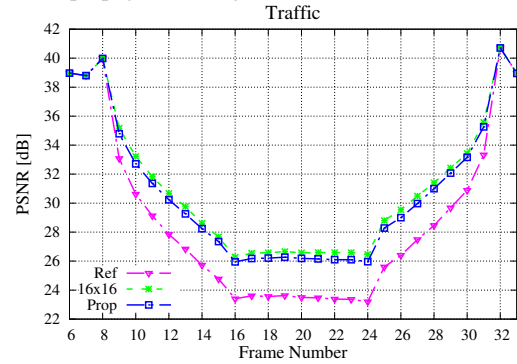Fig. 3: Error propagation using LD for the Basketball Drill sequence.

Fig. 4: Error propagation using RA for the Traffic sequence.

values (i.e., less than 1), which implicitly results in less signalling information. This also justifies the lower overhead obtained for the *Prop* method shown in Table I. These results show that all error concealment candidates are used, which indicates that all of them are found relevant by the proposed method to recover the missing frames. The candidate methods based on the estimated motion field (i.e., M1, M2) are selected more often because they provide more accurate clues about the scene's motion than the others (i.e., M3, M4).

Since a relevant performance indicator is the error propagation over temporally predicted frames, this was evaluated using a single frame loss. Figures 3 and 4 show the PSNR results when frame #16 is missing, for LD and RA encoding configurations, respectively. these Figures show the results for the proposed method (*Prop*), the

TABLE III: Average PSNR and variance of affected frames

Sequence	PLR=3%			PLR=10%		
	Ref	*Prop*	ΔPSNR	*Ref*	*Prop*	ΔPSNR
Low-delay configuration						
Basketball Drill	33.15 (0.42)	34.26 (0.24)	+1.11	32.97 (0.47)	35.90 (0.27)	+2.93
Kendo	34.33 (1.28)	36.57 (1.11)	+2.24	33.84 (1.26)	37.49 (1.05)	+3.65
Park Scene	35.22 (0.12)	35.33 (0.08)	+0.11	34.75 (0.11)	34.96 (0.09)	+0.21
People On Street	34.36 (0.32)	34.98 (0.22)	+0.62	32.17 (0.65)	33.45 (0.50)	+1.28
Traffic	36.77 (0.48)	36.88 (0.57)	+0.11	36.04 (0.30)	36.31 (0.27)	+0.27
Jockey	36.00 (1.14)	37.20 (1.05)	+1.20	35.44 (1.21)	36.74 (1.18)	+1.30
Average	**34.37 (0.73)**	**35.25 (0.60)**	+0.88	**33.64 (0.79)**	**35.14 (0.65)**	+1.50
Random-access configuration						
Basketball Drill	30.22 (8.46)	30.81 (7.66)	+0.59	29.19 (7.60)	29.84 (6.88)	+0.65
Kendo	27.45 (23.9)	32.38 (18.6)	+4.93	25.22 (28.3)	36.98 (22.4)	+11.8
Park Scene	33.91 (1.40)	34.46 (1.07)	+0.55	31.38 (1.85)	32.50 (1.48)	+1.12
People On Street	29.30 (5.12)	31.44 (3.17)	+2.14	24.89 (5.37)	27.69 (4.09)	+2.80
Traffic	35.64 (1.34)	38.12 (1.24)	+2.48	32.85 (1.88)	35.60 (1.40)	+2.75
Jockey	32.35 (12.7)	34.06 (8.60)	+1.71	29.96 (12.5)	32.11 (8.90)	+2.15
Average	**30.89 (5.64)**	**33.03 (2.14)**	+2.14	**28.46 (5.68)**	**31.90 (4.44)**	+3.44

reference method (*Ref*), and also for a fixed partition size (16×16). These results indicate that the proposed method (*Prop*) outperforms the reference one and it is able to achieve quality gains up to 2 dB for LD configuration (see Figure 3), decreasing the error propagation and increasing the quality of the affected frames up to 3 dB. The use of RA configuration leads to higher quality degradation than LD, due the higher temporal distance of the reference frame used for concealment (frame #8 is used to conceal lost frame #16). Note that when Frame # 16 is lost, all frames within the range #9 to #31 (i.e., next I-frame) are affected by error propagation. Nevertheless, the proposed method still achieves similar quality gains (approximately 2 dB). Moreover, for both configurations, even when the overhead is reduced, by moving from fixed partition to dynamic partition size, the proposed method presents only a small reduction in the reconstruction quality. This indicates that using dynamic partitioning allows to achieve a better trade-off between the amount of overhead and the error concealment performance.

Further tests were performed to evaluate the effectiveness of the proposed method under more different loss rates. In the experiments the packets corresponding to intra frames were randomly discarded to simulate loss events, at different packet loss ratios (PLR) rates with an average burst length of 5 packets. For each test condition 25 trials were performed and the average quality of the frames affected by errors (the lost frame and dependent ones) is presented in Table III. The table also shows the PSNR variance value in parentheses and the difference to the PSNR obtained with a reference HEVC decoder The results confirm the superiority of the proposed method, outperforming the reference HEVC for both low and high PLR. Using LD coding configuration, the proposed method is able to achieve quality gains of 0.88 dB for 3% of PLR and 1.50 dB for 10% of PLR. In case of RA, higher quality gains are achieved, leading to an average improvement of 3.44 dB for 10% of PLR. It is also worthwhile to notice that the PSNR variance obtained with the *Prop* method is lower, which indicates lower quality variations in the video segments affected by errors.

The quality gains achieved by the proposed method are similar to those achieved by previous works [6], [7] for the same PLRs. However, these require a higher amount of overhead (approximately 3%), thus the proposed method achieves a better trade-off between quality and overhead. In summary, the proposed technique is able to enhance the performance of the standard decoder by finding the best error concealment method at the encoder and sending such information with a small amount of signalling overhead.

IV. CONCLUSIONS

In this paper an error concealment-aware encoder is proposed by combining loss simulation at the encoder-side to enhance the error concealment efficiency at the decoder. The proposed method is able to improve the reconstruction efficiency of lost intra coded frames and consequently reduce error propagation. The results show that a consistent quality improvement is achieved at the cost of a small overhead, which allows to use this method in networked video services and applications where packet loss probability cannot be ignored.

REFERENCES

[1] G. Sullivan, J. Ohm, W.-J. 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] H. Yang and K. Rose, "Optimizing motion compensated prediction for error resilient video coding," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 108–118, Jan. 2010.

[3] 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.

[4] E. Baccaglini, T. Tillo, and G. Olmo, "Concealment driven smart slice reordering for robust video transmission," in *IEEE International Conference on Multimedia and Expo*, Jun. 2008, pp. 1173–1176.

[5] H. Yang and J. Boyce, "Concealment-aware motion estimation and mode selection for error resilient video coding," in *International Conference on Image Processing (ICIP)*, Oct. 2006, pp. 2229–2232.

[6] S. Chen and H. Leung, "A temporal approach for improving intra-frame concealment performance in h.264/avc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 422–426, Mar. 2009.

[7] J. Y. Pyun, "Error concealment aware streaming video system over packet-based mobile networks," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1705–1713, Nov. 2008.

[8] B. Yan and H. Gharavi, "A hybrid frame concealment algorithm for H.264/AVC," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 98–107, Jan. 2010.

[9] T.-L. Lin, N.-C. Yang, R.-H. Syu, C.-C. Liao, and W.-L. Tsai, "Error concealment algorithm for HEVC coded video using block partition decisions," in *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, Aug. 2013, pp. 1–5.

[10] Y.-L. Chang, Y. Reznik, Z. Chen, and P. Cosman, "Motion compensated error concealment for HEVC based on block-merging and residual energy," in *20th International Packet Video Workshop (PV)*, Dec. 2013, pp. 1–6.

[11] R. Sjoberg, Y. Chen, A. Fujibayashi, M. Hannuksela, J. Samuelsson, T. K. Tan, Y.-K. 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.

[12] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, Aug. 1981.

[13] P. Howard and J. Vitter, "Arithmetic coding for data compression," *Proceedings of the IEEE*, vol. 82, no. 6, pp. 857–865, Jun. 1994.

[14] F. Bossen, "Common test conditions and software reference configurations, document JCTVC-H1100," San Jose, CA, Feb. 2012.