Error Concealment-Aware Encoding for Robust Video Transmission

J. Carreira[1,2], P. Assuncao[2], S. Faria[2], E. Ekmekcioglu[1], A. Kondoz[1]
[1]Institute for Digital Technologies, Loughborough University London, UK;
[2]Instituto de Telecomunicações / Instituto Politécnico de Leiria, PT;
j.f.m.carreira@lboro.ac.uk

Abstract—In this paper an error concealment-aware encoding scheme is proposed to improve the quality of decoded video in broadcast environments prone to transmission errors and data loss. The proposed scheme is based on a scalable coding approach where the best error concealment (EC) methods to be used at the decoder are optimally determined at the encoder and signalled to the decoder through SEI messages. Such optimal EC modes are found by simulating transmission losses followed by a lagrangian optimisation of the *signalling rate - EC distortion* cost. A generalised saliency-weighted distortion is used and the residue between coded frames and their EC substitutes is encoded using a rate-controlled enhancement layer. When data loss occurs the decoder uses the signalling information is used at the decoder, in case of data loss, to improve the reconstruction quality. The simulation results show that the proposed method achieves consistent quality gains in comparison with other reference methods and previous works. Using only the EC mode signalling, i.e., without any residue transmitted in the enhancement layer, an average PSNR gain up to 2.95 dB is achieved, while using the full EC-aware scheme, i.e., including residue encoded in the enhancement layer, the proposed scheme outperforms other comparable methods, with PSNR gain up to 3.79 dB.

Index Terms—video broadcasting, error concealment-aware coding, robust video coding.

I. INTRODUCTION

RECENT developments in video compression and broadcasting technologies are progressing towards new Ultra-HD formats (e.g., 4k or 8k resolution), which are also expected to be seamlessly adapted and delivered to mobile devices. The consequent increase in bandwidth demand and multimedia traffic over broadcasting networks also imposes more challenging bounds on the quality and error tolerance of emerging services. The High Efficiency Video Coding (HEVC) [1], as the most recent standard developed by the Joint Collaborative Team on Video Coding (JCT-VC), is foreseen as the main video coding framework to extend current services towards higher resolution formats delivered with better compression efficiency. The coding flexibility and efficiency of HEVC, results from the use of new coding tools and block structures, which enables new block partitions [2], improved prediction modes [3], [4] and new high-level features [5], such as explicit reference picture management and new parameter sets. However, it is known that higher coding efficiency leads to lower robustness to transmission errors and data loss.

The error robustness characteristics of HEVC were analysed in [6] confirming that, in general, the HEVC presents poor error resilience performance, despite its superior coding efficiency in comparison with previous standards. Thus, when HEVC streams incur in transmission losses, this results in significant degradation of both objective and subjective quality [7], [8]. This is mostly due to the strong decoding dependencies imposed by the highly complex prediction modes of HEVC, which are selected according to a rate-distortion (R-D) optimisation criteria, always assuming error-free transmission [9]. A possible approach to encode robust video streams is to increase the error resilience of HEVC by including error-aware mechanisms in the coding process. As a result, the effect of error propagation due to prediction mismatch at the decoding side, will be limited.

The widespread use of HEVC in broadcast and multimedia delivery services, will be based on transport technologies such as RTP [10], MPEG-2 TS [11], or more recent standards as MPEG-DASH [12]. Recent developments also enable combinations of different technologies, e.g, HTTP-based streaming and MPEG-2 TS, to increase the flexibility of over-the-top (OTT) video services [13]. For content delivery over heterogeneous networks, the MPEG Media Transport (MMT) [14] is another possible technology to be used with the HEVC standard. Besides the transport technologies, the overall error robustness of video transmission may also take advantage of the increasing flexibility of multimedia systems to cope with heterogeneous networks, by using auxiliary streams to aid the video decoder to increase the quality of recovered lost slices [15], [16]. Moreover, error resilience techniques may also take advantage of the multi-stream based approach, where each stream allows different levels of robustness to errors [12], [17].

This paper proposes a novel error concealment-aware video encoding scheme to increase the robustness of HEVC streams over transmission channels prone to errors or data loss, such as broadcasting networks. The method aims to improve the quality of recovered lost frames by optimising the efficiency of the error concealment (EC) beforehand at the encoder-side. The proposed approach is based on finding the optimal EC mode, in rate-distortion (R-D) sense, to be used at the decoder side. The process includes data loss simulation and minimisation of a lagrangian cost based on saliency-weighted distortion and signalling rate. Furthermore the residue between the coded frame and its error reconstructed version is also encoded using a rate-controlled enhancement scalable layer.

The remainder of the paper is organised as follows. Section II presents a critical review of related works addressing the

problems of robust coding. Section III describes the proposed EC-aware encoding scheme and Section IV presents and discusses in detail the experimental results along with performance comparison and benchmarking. Finally, Section V concludes the paper.

II. RELATED WORD

In the past, robust transmission has been addressed using different techniques, which can be grouped in four categories: localisation (i.e., to reduce coding dependencies), data partitioning, redundant coding and EC-driven techniques [18]. Different solutions for the problem of spatial and temporal dependencies have been proposed through optimisation, based on the estimated decoded distortion [19]–[21], or by introducing intra-coded CTUs as refresh points based on channel feedback [22]. Still at the encoder side, to reduce mismatched predictions at the decoder, data partitioning was used in [23], dual frame inter-prediction in [24], [25] and reduction of motion vector (MV) dependencies in [26], [27]. Alternatively to increase error resilience, redundant pictures were used in [28], [29] while the use of multiple redundant streams was proposed in [30]. In [31] a visual attention model is used to increase the error robustness of video streams by unequally allocating the error correction codes to those regions with higher impact on visual quality. Finally, EC-driven techniques were developed for the encoder side to ease the reconstruction of missing frames at the decoder [32]–[35]. These approaches have the advantage of not requiring any type of re-transmission which would not be suitable for delay-sensitive applications such as interactive communications with live video broadcasting.

At the decoder-side EC techniques must be applied in order to reduce the effect of mismatched predictions and consequent error propagation in case of frame loss. EC methods based on motion compensation are known to perform better than spatial interpolation, which is not suitable to recover from whole frame losses [36]. In [37], the direct motion copy method presented in [38] was improved by using a recursive difference between the MVs of successive frames, in order to refine the final MV accuracy. In [36], [39] methods to recover missing image regions using a combination of spatial and temporal EC techniques are proposed. Edge estimation and directional interpolation is used in the spatial domain, while in the temporal domain motion refinement with variable block size is used to achieve higher reconstruction quality. In the case of HEVC, in [40] the block partitions of the neighbouring frames were used to assist the motion extrapolation in order to keep the object boundaries smooth. In order to select only the reliable motion information, the residue information can be exploited, as shown in [41]. Overall, the aforementioned methods rely on post-processing without any assistance from the encoder-side, which may lead to inaccurate recovery of the lost information. Thus, in order to achieve higher reconstruction quality further information need to be shared with the decoder to minimise the residual distortion of EC.

In this work we investigate robust coding driven by EC methods with the objective of enhancing the quality of reconstructed video at decoders receiving erroneous streams from the transmission network. The proposed approach combines error resilience tools at the encoder-side with EC at the decoder-side. There are few works addressing this problem that are worth to be mentioned, as they provide relevant insight to the method proposed in this paper. In [32] an estimation of the EC distortion is used at the encoder for slice classification and unequal error protection. Such distortion was also used to optimise motion estimation [20], [33], by including the distortion of the subsequent frame (assuming it is lost and recovered) in the R-D cost optimisation of the current frame. In [42] pixel interlacing is used to divide the frame into different packets in order to enhance the performance of intra-frame EC using linear interpolation. An alternative is to use data hiding 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 [34]. In [43] a low-resolution video is transmitted alongside the main bitstream in order to be used as reference for concealment. The redundant stream is used by the decoder to improve the estimation accuracy of motion information. This method was extended in [44] by using asymmetric coding of the redundant information based on Itti's attention model [45].

Although the above methods facilitate the EC performance either through slice reordering or based on MVs, they always rely on the assumption that a fixed EC method is used at the decoder. The work presented in [35] proposes testing several methods with transmission of extra symbols to indicate the best method to be applied by the decoder, using a region of interest (ROI) to identify the most sensitive image area. Although such an approach achieves reasonably good performance, it relies on the spatial neighbouring information, which may not be available in case of full frame loss. Moreover, it requires transmission of the ROI map to the decoder, which increases the overall bitrate. More recently, a EC signalling methods was proposed for scalable video [46], in order to select the best frame to be used to replace the missing one.

To advance one step further, in this work an EC-aware encoding scheme is proposed to achieve higher performance. This is accomplished by selecting at the encoder, among several EC algorithms, the one that yields the lowest reconstruction distortion at the decoder and also minimises the R-D cost considering the necessary signalling rate. In contrast with previous works [24] that addressed the error resilience problem by optimising the coding dependencies, the approach followed in this work considers both EC performance and the rate required for signalling the best EC mode to the decoder. Thus, the proposed approach outperforms similar previous methods [20], [33] by selecting the best error concealment method without affecting the encoder optimisation. For instance, in comparison with [35], the method devised in this work allows recovery from full frame loss and optimises the signalling overhead without transmitting saliency information. To further improve the quality of recovered frames, an enhancement layer is encoded with the mismatch residue between the transmitted and recovered frames. This is a dynamic optimisation process, which is more efficient than traditional approaches based on
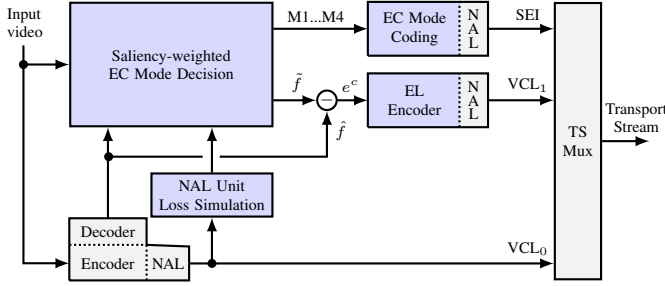
Fig. 1: Architecture of the EC-aware encoding scheme.

redundant pictures [28], [29], which transmit the whole frame as redundancy.

III. PROPOSED ERROR CONCEALMENT-AWARE ENCODING

Figure 1 illustrates the architecture of the proposed encoding scheme. After the video encoder, it includes data loss simulation at the Network Adaptation Layer (NAL) for every slice. Subsequently, error concealment is applied to find the best EC methods that should be used to reconstruct the lost slices in case of transmission errors. Saliency-weighted EC mode decision is used to select the optimal EC mode from the best trade-off between EC distortion, using the original input video (f) as reference, and the number of bits required for signalling. Then the optimal mode is sent to the decoder along with the residue formed between the error-free decoded frame and the EC one, using the scalable extension of the HEVC standard.

For the EC mode decision optimization the missing slices are reconstructed, as they would be in the decoder, but using different EC methods based on motion estimation and extrapolation techniques, to selected the best one at the encoder-side. In this process, each coding tree unit (CTU) is partitioned to form a quadtree structure of EC units (ECU) that are used to test the different EC candidate modes. Then, the best trade-off between ECU partition and signalling is found through R-D optimisation. In order to achieve an efficient use of the signalling overhead a saliency-weighted procedure is embedded in the R-D optimisation based on a spatio-temporal saliency, where more signalling bits are allocated to regions with higher saliency values. Finally, the signalling information is transmitted to the decoder by using the supplemental enhancement information (SEI) NAL unit type [5]. Since in general, the optimal EC mode is still not be able to correctly reconstruct all the missing regions, extra residual information is multiplexed in the stream as a scalable enhancement layer. This residual information contains the difference between the compressed ($\hat{f}_t$) and the EC frames ($\tilde{f}$).

The transmitted information is used to assist the EC operation at the decoder to recover erroneous frames with high accuracy, by using the best EC mode and also residual information to reduce error propagation. The remaining of this section describes in detail the building blocks of the proposed encoding architecture, i.e., the EC candidate modes, the saliency-weighted optimal EC mode decision, signalling and enhancement layer coding.

A. EC candidate modes

A set of four EC candidate modes (M1...M4) is used in the optimisation process that selects one of them as best for each particular ECU. Each EC mode corresponds to a different EC method, as shown in Figure 2, 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 to recover the lost frame, herein referred to as frame f_0 at time instant t_0. This motion field is obtained from the closest available neighbours, i.e., frames f_{-1} and f_{-2}. Then it is extrapolated at the pixel level, i.e., one MV is assigned to each pixel, and coding unit (CU) level, i.e., one MV is assigned to the entire block, at the missing frame instant. This results in two different reconstructions of the missing frame, where the former may increase the accuracy of the pixel extrapolation, but the latter has the advantage of keeping the block structure and generate a smoother reconstruction. The motion field is estimated using an optical flow method [47] based on the Horn and Schunck smoothness condition [48], that forces the estimated field to be regular across neighbouring pixels. Using this method, for each pixel $\mathbf{p}$, the motion field $v(\mathbf{p})$ is obtained by minimising the following energy function,

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

where f_{-1} and f_{-2} are the two previous frames, and Ω defines the sets of pixels of those frames. $|\nabla v(\mathbf{p})|$ is a weighting factor that penalises high variation of the optical flow which effectively disallow discontinuities. This is based on the assumption that neighbouring regions belong to the same object, thus such regions should have similar motion.

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

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

where t_R is the time instant of the reference frame pointed by the original MVs v_{-1}. Then, the MV associated to the block $b_{-1}(x, y)$ at spatial position (x, y) in f_{-1} may be used to recover the block $b_0'(x', y')$ in f_0, at position (x', y') obtained from v_0 components as follows:

$$x' = x - v_{0x} \quad (3)$$
$$y' = y - v_{0y} \quad (4)$$

The fourth EC 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}).
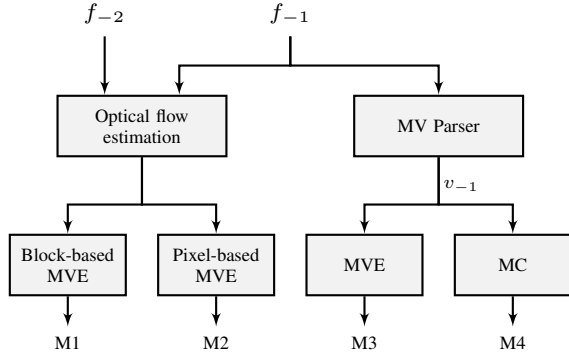
Fig. 2: EC candidates used for optimization.

split: flag to mark quadtree splitting

ecm: error concealment method selected for each ECU

Fig. 3: Payload structure of the SEI NAL units

Although the study carried out in this work only takes into consideration the aforementioned EC methods, the proposed approach is generic and extension to include more EC candidates is straightforward.

B. Optimal saliency-weighted EC mode decision

As described above, in the sub-section III-A, the proposed method includes four EC candidates, which result in different recovered frames for each lost frame (simulated). The lost frame is partitioned into quadtree structures of ECUs, each one forming a set of variable size ECUs. Then, finding the optimal EC method for each ECU is formulated as a rate-distortion cost minimisation. To solve the aforementioned problem, a Lagrangian optimisation is used [49] to achieve the best trade-off between reconstruction quality and signalling overhead. Furthermore, image saliency driven by visual attention modelling is used in the optimisation process to weight the cost of the overhead information of each ECU, according to the visual relevance of its content.

The optimal EC mode $m_{l,i}^*$, for the i_{th} ECU which size is defined by the quadtree partitioning level l, is obtained by minimising the following lagrangian cost:

$$m_{l,i}^* = \operatorname*{arg\,min}_{\substack{l \in \{0...3\} \\ i \in \{1...4^l\}}} \left\{ D(m_{l,i}) + \left(1 + \frac{t_0 - t_I}{T_I}\right) \lambda R(m_{l,i}) \right\}, \quad (5)$$

where $R(m_{l,i})$ represents the number of overhead bits required to encode the EC mode, and $D(m_{l,i})$ is sum of absolute error of the ECU.

In equation 5, besides the parameter λ used to trade-off the cost of distortion and overhead rate in the minimisation process, there is also a weighting factor to control the cost of $R(m_{l,i})$ according to the duration of the error propagation. This is given by the ratio between the temporal distance from the current frame (t_0) to the next refresh frame (t_I) and the period of refresh frames (T_I). This weight varies between $[0; 1]$, achieving the lowest value for intra-coded frames and the maximum for the frame located immediately before an intra-coded one. Moreover, a dynamic value for λ is used to adjust the amount of overhead across the frame, according to the visual relevance of each image region given by a saliency map.
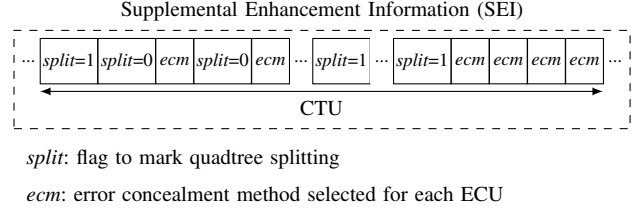
The saliency maps were computed based on the well-known Itti's attention model [45], which applies a hierarchical decomposition based on a set of spatial features and then combines them with center-surround weighting to obtain the saliency map. Since the Itti's model only produces saliency maps for still images, in this work it is further combined with a temporal feature. This is obtained by estimating the optical flow between two consecutive frames, based on the algorithm presented in [48]. The final saliency map is obtained by combining the spatial and temporal components, S_s and S_t respectively, as follows,

$$S = \rho S_t + (1 - \rho) S_s \quad (6)$$

The parameter ρ is used to give higher relevance to the temporal component, since motion is correlated with error propagation, producing longer artefacts and consequently higher impact in visual attention. In this work the temporal component was given weight slightly higher than 50%, i.e. $\rho = 0.6$.

On the one hand, higher values of saliency (i.e. regions with higher visual relevance) should result in lower values of λ, which tend to increase the overhead rate value on those regions, thus reducing EC distortion at the decoder in case of loss. On the other hand, higher values of λ should reduce the overhead, allowing higher EC distortion at the decoder in case of loss. In order to achieve the aforementioned goal the following definition for λ is used:

$$\lambda = \begin{cases} \lambda_0 e^{\frac{\bar{S}-s}{\Delta S}}, & \text{if } \Delta S > 0. \\ \lambda_0, & \text{no saliency used.} \end{cases} \quad (7)$$

where s is the saliency value of the current CTU, $\bar{S}$ is the average saliency value and ΔS is the difference between the maximum and minimum. One should note that the method is not dependent on the existence of a saliency map. If the saliency map is constant or non-existent a fixed value of λ is used, which means that no distinction is made across image regions.

Signalling the optimal EC modes

Figure 3 illustrates the structure of the signalling information for each CTU comprising a quadtree of ECUs. In this structure, one bit is used for the *split* flag and two bits for the *ecm* mode associated to each ECU. The *split* flag indicates whether the ECU is further split into the next level ($split = 1$) or ECU partitioning stops at the current level ($split = 0$) and the corresponding *ecm* modes are inserted. Since no split is necessary at the highest level, then $split = 1$ followed by

the four ecm modes for each ECU. This structure allows dynamic selection of different levels within each frame. Adaptive arithmetic coding [50] is used to efficiently encode both the $split$ and ecm symbols and further reduce the overhead. Note that the SEI information uses much less bitrate than the video signal, thus it is very unlikely to be hit during transmission.

C. Encoding of EC residue - enhancement layer

As described earlier, in order to allow further improvement of recovered slices, besides the best EC mode, the proposed encoding scheme also encodes a controlled amount of EC residue using the scalable extension of HEVC (see Figure 1). The enhancement layer encodes the difference between the coded frame ($\hat{f}$) and its counterpart recovered by using the optimal EC mode ($\tilde{f}$). Note that, the higher the accuracy of the best EC mode, the lower the amount of bits spent in the enhancement layer. At the decoder-side, the reconstruction of a missing frame ($\tilde{f}$) follows three steps:

1) the missing frame is reconstructed using the optimal signalled EC methods for each ECU ($\tilde{f}'$);
2) the enhancement layer residue is decoded (e^c);
3) the recovered frame is obtained as follows:

$$\tilde{f} = \tilde{f}' + e^c. \tag{8}$$

The additional residue received from the enhancement layer is particularly useful to improve the quality of those CTUs where the EC strategies are not able to produce very accurate reconstruction. Moreover, since the enhancement layer only carries residual information the overhead is affordable.

Note that the generated bitstream is compliant with the standard, as the error concealment signalling is encapsulated in SEI NAL units (NALU) while the error concealment residue is carried in NALUs related with the scalable extension. Then, an adapted decoder, capable of decoding both the SEI NALUs and the scalable NALUs takes full benefit from the proposed method. Nevertheless, any standard HEVC decoder is also able to reconstruct the video, because it simply ignores the extra NALUs.

IV. PERFORMANCE EVALUATION

The performance of the proposed EC-aware encoding scheme is evaluated using the HEVC standard as the basis of the encoding framework. This section is organised in the following subsections:

- evaluation of the usage distribution of each EC mode;
- performance evaluation of the EC optimal mode decision (without using the enhancement layer);
- evaluation of the error propagation under single loss events;
- overall performance evaluation of the proposed method using the EC enhancement layer and benchmarking with existing techniques.

The quality obtained with the proposed method is compared with the default method used by reference HEVC, i.e., frame-copy (*Ref*), motion-copy (*MC*) [41], motion vector extrapolation using estimated optical flow [47] (*OptFlow*), and the

TABLE I: Test sequences used in the experiments.

Sequence	SI/TI	Bitrate (Mbps)	Description
Basketball Drill $832 \times 480@50Hz$	33.4 / 14.4	4.50	High motion with several basket ball players
Book Arrival $1024 \times 768@30Hz$	28.4 / 21.7	1.50	Moderate translational motion with two moving persons
BQSquare $416 \times 240@60Hz$	63.2 / 11.5	4.00	Moderate outside motion with moving camera capturing from high point
Four People $1280 \times 720@60Hz$	31.3 / 6.90	2.00	Four people talking of the desk and passing objects to each other
Kimono $1920 \times 1080@24Hz$	23.4 / 32.5	8.00	Capture of a women moving in a forest with moderate motion
Park Scene $1920 \times 1080@24Hz$	31.3 / 11.6	10.00	Moderate motion with cyclists passing across the scene
People on Street $2560 \times 1600@24Hz$	40.0 / 25.4	12.00	With point capture of people moving; high motion and texture complexity
Race Horses $832 \times 480@30Hz$	43.7 / 24.4	8.00	High motion with several horse riders
Tennis $1920 \times 1080@30Hz$	20.3 / 45.3	5.00	High motion with one moving person in the scene

TABLE II: Average Level and usage ratio of each EC mode.

Sequence	Avg. Level	Avg. usage ratio (%)			
		M1	M2	M3	M4
Basketball Drill	0.81	59.61	16.62	14.40	9.37
Book Arrival	0.34	66.77	18.79	7.15	7.29
BQSquare	0.41	15.23	0.64	72.97	11.16
Four People	0.17	85.84	3.22	8.09	2.86
Kimono	0.88	33.76	23.87	26.03	16.35
Park Scene	0.69	39.34	13.58	34.19	12.89
People on Street	1.51	31.17	34.39	16.15	18.29
Race Horses	1.62	34.05	29.89	17.06	19.01
Tennis	1.17	25.27	30.03	23.22	21.48

method proposed in [19], which implements a robust scheme to minimise error propagation. The video quality was evaluated by using two quality metrics: PSNR and a feature similarity index (FSIM) [51] in order to validate the performance of proposed method in a consolidated manner.

The nine video sequences presented in Table I were used in the experiments [52]. These have different types of motion and texture complexity, as evidence by different spatial information (SI) and the temporal information (TI), measured as defined in [53]. The HM reference software, version 16.2 [54] was used with all coding modes enabled, an IDR period of 16 frames and following two recommended prediction structures: Low-Delay (P-frames with 4 references frames) and Random-Access (B-frames with 2 reference frames in each list) [52]. The results were obtained using the same constant bitrate for each test condition. The bitrate used for each sequence is presented in Table I.

A broadcast transmission scenario is simulated by using HEVC NALUs encapsulated into MPEG-2 TS. Each frame is encoded into several slices with a maximum size of 1200 bytes, which are then transmitted as the payload of packetized elementary stream (PES) packets. The PES packets are then split into transport stream (TS) packets. Random loss simulation is carried out using a two-state Markov model. Whenever a TS packet is lost, the entire PES packet is discarded, resulting in the loss of an entire NALU and the corresponding slice. For each test condition, 50 runs were performed and the average PSNR was measured. The filtering

TABLE III: Average quality (PSNR and WPSNR) of the affected frames for Low-Delay configuration.

Sequence	PLR=3%					PLR=10%				
	Ref	*MC*	*OptFlow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*	*Ref*	*MC*	*OptFlow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*
Basketball Drill	28.41 (29.25)	+2.93 (+3.22)	+3.25 (+3.24)	+5.28 (+5.02)	+2.64 (+5.01)	24.64 (22.89)	+2.63 (+3.27)	+3.09 (+3.25)	+4.93 (+5.66)	+4.82 (+5.66)
Book Arrival	30.69 (33.53)	+2.31 (+1.56)	+3.11 (+1.48)	+3.80 (+2.61)	+3.73 (+3.17)	28.29 (27.28)	+2.65 (+2.82)	+1.47 (+2.74)	+4.59 (+4.91)	+4.51 (+5.86)
BQSquare	31.82 (35.39)	+1.97 (+1.59)	+1.23 (+0.90)	+2.21 (+1.75)	+2.19 (+2.29)	28.06 (30.91)	+2.59 (+2.52)	+1.92 (+1.43)	+3.14 (+2.97)	+3.11 (+3.82)
Four People	34.54 (37.03)	+1.46 (+0.75)	+1.87 (+0.73)	+1.91 (+0.96)	+1.89 (+1.14)	33.07 (34.04)	+1.85 (+1.86)	+2.70 (+1.88)	+2.59 (+2.52)	+2.56 (+2.83)
Kimono	32.39 (33.86)	+1.66 (+1.76)	+1.88 (+1.87)	+2.32 (+2.43)	+2.30 (+2.65)	27.98 (29.54)	+0.14 (+0.14)	+0.29 (+0.13)	+0.19 (+0.19)	+0.19 (+0.50)
Park Scene	31.66 (31.76)	+2.28 (+2.89)	+2.15 (+2.39)	+2.96 (+4.03)	+2.93 (+4.27)	27.76 (26.82)	+1.32 (+1.53)	+1.25 (+1.31)	+1.93 (+2.42)	+1.90 (+2.81)
People on Street	26.72 (27.24)	+1.59 (+1.78)	+2.15 (+2.36)	+2.96 (+3.21)	+2.88 (+3.35)	22.53 (22.57)	+0.40 (+0.45)	+1.25 (+0.61)	+0.85 (+0.91)	+0.83 (+1.06)
Race Horses	25.97 (27.81)	+2.55 (+2.73)	+2.59 (+2.57)	+3.80 (+3.98)	+3.69 (+4.29)	21.55 (22.63)	+0.68 (+0.76)	+0.76 (+0.72)	+1.12 (+1.23)	+1.07 (+1.50)
Tennis	28.47 (30.79)	+1.66 (+1.27)	+2.02 (+1.53)	+4.39 (+3.52)	+4.27 (+3.62)	23.77 (24.71)	+1.41 (+1.13)	+1.82 (+1.45)	+4.34 (+3.79)	+4.19 (+3.84)
Average	–	**+2.04 (+1.95)**	**+2.25 (+1.90)**	**+3.29 (+3.06)**	**+2.95 (+3.31)**	–	**+1.52 (+1.61)**	**+1.62 (+1.50)**	**+2.63 (+2.73)**	**+2.58 (+3.10)**

TABLE IV: Average FSIM of the entire video sequence in frame loss conditions (scaled between $[0 - 100]$).

Sequence	PLR=3%					PLR=10%				
	Ref	*MC*	*OptFlow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*	*Ref*	*MC*	*OptFlow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*
Basketball Drill	96.96	+1.13	+1.16	+1.58	+1.59	92.74	+2.38	+2.50	+3.60	+3.65
Book Arrival	98.61	+0.36	+0.41	+0.57	+0.60	96.66	+1.05	+1.21	+1.74	+1.75
BQSquare	98.03	+0.39	+0.28	+0.41	+0.41	96.31	+1.15	+0.81	+1.30	+1.31
Four People	99.41	+0.09	+0.09	+0.11	+0.11	98.97	+0.31	+0.32	+0.39	+0.60
Kimono	98.54	+0.48	+0.52	+0.59	+0.62	95.79	+0.09	+0.12	+0.12	+0.15
Park Scene	98.74	+0.58	+0.55	+0.69	+0.71	96.41	+0.83	+0.79	+1.03	+1.10
People on Street	98.19	+0.67	+0.73	+0.92	+0.94	94.65	+0.45	+0.51	+0.66	+0.70
Race Horses	95.16	+1.78	+1.80	+2.28	+2.29	88.25	+1.18	+1.21	+1.62	+1.68
Tennis	96.64	+0.98	+0.98	+1.69	+1.64	91.02	+2.20	+2.29	+4.27	+4.32
Average	–	**+0.72**	**+0.73**	**+0.98**	**+0.99**	–	**+1.07**	**+1.09**	**+1.64**	**+1.70**

operations were disabled at slices boundaries in order to keep the slices self-contained. The remaining configurations have been kept at their default values as given in the common test conditions [52].

A. EC mode usage distribution

The EC mode usage distribution is an indicator of the usefulness of each EC mode. Table II shows the average *Level* selected for the ECU quadtree and the usage percentage of each EC mode. It is noticeable from the results that the proposed method leads to low average *Level* values (i.e., less than 1), which implicitly results in less signalling information. This also justifies the low overhead shown in Table V.

The results in Table II show that all EC modes are effectively used, which indicates that all of them are found relevant by to recover the missing data. The EC candidate methods based on the estimated motion field (i.e., M1, M2) are selected more often because they provide more accurate representation of the scene's motion than the others (i.e., M3, M4).

TABLE V: Error concealment signalling overhead.

Sequence	Bits used for signalling (%)	
	NoSaliency (Constant λ)	*Saliency (Adaptive λ)*
Basketball Drill	2.88	2.55
Book Arrival	4.03	3.35
BQSquare	0.33	0.24
Four People	3.06	3.18
Kimono	1.75	1.30
Park Scene	1.84	1.37
People on Street	8.39	5.67
Race Horses	1.84	1.21
Tennis	12.15	7.46
Average	**+4.04**	**+2.91 (-28%)**

B. Performance evaluation of optimal EC mode decision

The performance of the proposed saliency-weighted EC mode decision (*Saliency*) was evaluated under loss events and compared with the reference methods. Realistic conditions were used for transmission, by allowing independent losses for any of type of NAL, hence there is no guarantee that the decoder conditions are always equal to those at the encoder, which means that a signalling mismatch may occur. The results obtained without saliency weighting are also shown
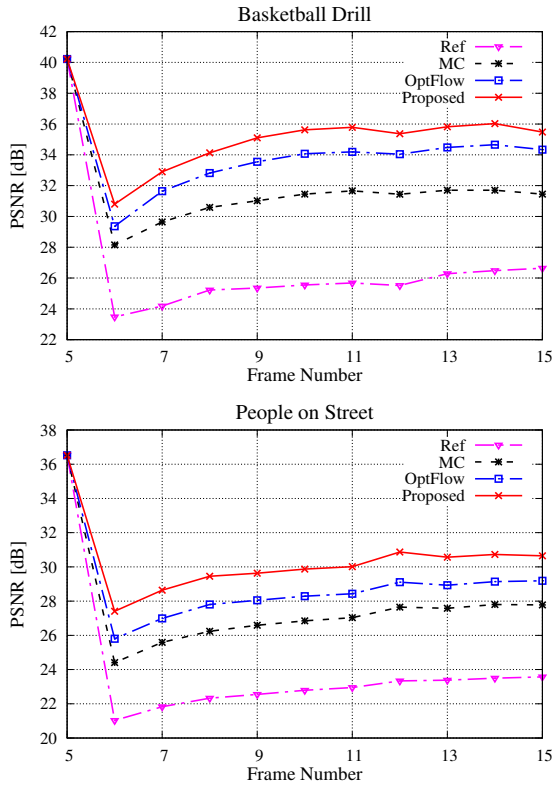
Fig. 4: Error propagation after losing frame #6 for Basketball Drill and People On Street sequences.
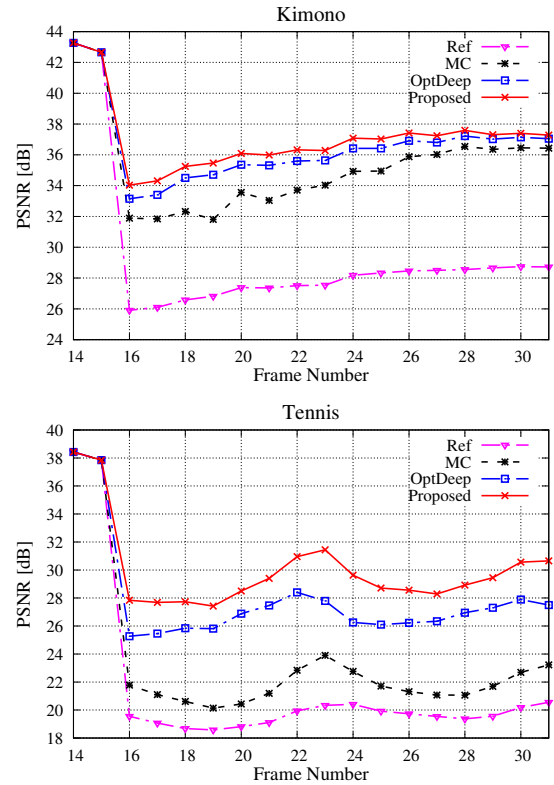

Fig. 5: Error propagation after losing intra-coded frame #16 for Kimono and Tennis sequences.

for comparison (*NoSaliency*). In this case a constant $\lambda_0 = 10$ is used.

Tables III and IV show the average quality under different packet loss ratios (PLRs) for the Low-delay configuration. Table III shows the PSNR and weighted PSNR (WPSNR) between parentheses. The MSE weights used in the WPSNR are defined by the saliency values [55]. Table IV shows the results obtained for the FSIM quality metric. Note that, the absolute value is shown for the reference case (*Ref*), while the differential values are shown for the remaining methods.

From the PSNR results shown in Table III, one may conclude that the proposed method (*Saliency*) consistently outperforms the reference ones for both low and high PLRs. Average PSNR gains up to 2.95 dB for PLR=3% and 2.58 dB for PLR=10% are achieved. It is also worthwhile to notice that higher performance is obtained for sequences with high motion, such as Basketball Drill, which leads to PSNR gains up 4.82 dB Moreover, the results of Table IV confirm the quality improvement achieved by the proposed method, leading to an increase of FSIM up to 4.32, which indicates that the proposed method is able to recover important image features with higher accuracy.

Table V shows the percentage of overhead in the total rate with and without saliency-weighted optimisation, i.e., *NoSaliency* and *Saliency*, respectively. As shown in the Table, the saliency-weighted optimisation reduces the amount of overhead by dynamically assigning unequal number of bits to image regions of different visual importance. These results
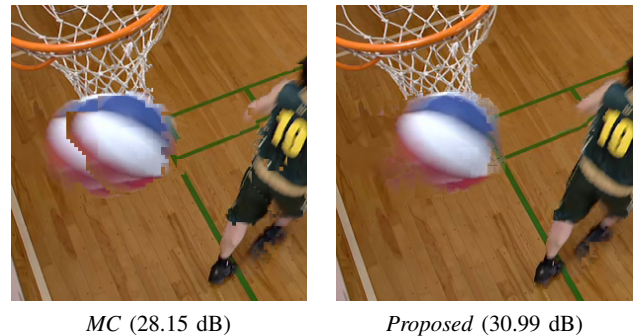

MC (28.15 dB)   *Proposed* (30.99 dB)

Fig. 6: Recovered frame #6 of Basketball Drill

show that using an adaptive value of λ based on the saliency values leads to an average overhead reduction of 28% in comparison with the case where no saliency is used, i.e., constant λ.

It is worth to note that using the saliency-weighted optimisation to reduce the overhead does not compromise the video quality, as can be seen in the results shown in Tables III and IV. The PSNR values for *NoSaliency* and *Saliency* are very close, but in overhead savings are significant for most sequences. The achieved WPSNR and FSIM results further demonstrate the effectiveness of the saliency-based optimisation, because the lower overhead does not affect the most relevant image regions, i.e., those with higher saliency values. Thus, the proposed method using a adaptive λ (i.e., using saliency) is able to not only achieve higher EC quality

TABLE VI: Average PSNR for different methods and different overhead ratios at the same total bitrate.

Sequence	PLR=3%			Proposed		PLR=10%			Proposed	
	Ref	MC	[19]	10%	20%	Ref	MC	[19]	10%	20%
Low-delay configuration										
Basketball Drill	33.31	+2.10	+3.86	+4.62	+4.56	28.22	+2.34	+5.32	+6.26	+6.90
Book Arrival	36.21	+1.33	+2.07	+2.71	+2.99	31.01	+2.48	+4.49	+5.60	+6.31
BQSquare	35.64	+1.47	+1.18	+2.01	+1.77	31.36	+2.50	+2.40	+3.33	+3.55
Four People	39.16	+0.47	-0.65	+0.40	-0.06	37.12	+1.24	-0.99	+1.82	+1.51
Kimono	36.19	+1.41	-0.54	+3.09	+3.47	32.24	+0.12	-1.92	-0.04	-0.21
Park Scene	34.54	+2.06	+2.01	+3.22	+3.15	30.42	+1.22	+1.56	+2.92	+3.64
People on Street	29.48	+1.50	+1.35	+2.64	+3.08	25.05	+0.40	+0.74	+0.97	+2.09
Race Horses	29.08	+2.34	+4.52	+5.35	+6.02	24.19	+0.63	+2.53	+2.75	+3.74
Tennis	32.71	+1.32	+3.13	+3.81	+4.97	27.08	+1.30	+3.51	+4.58	+6.61
Average	–	**+1.55**	**+1.88**	**+3.10**	**+3.33**	–	**+1.36**	**+1.96**	**+3.13**	**+3.79**
Random-access configuration										
Basketball Drill	35.40	+0.17	+0.20	+1.39	+1.53	29.92	+0.14	+1.45	+1.84	+2.42
Book Arrival	37.26	+0.05	+0.72	+1.31	+1.28	32.27	+0.13	+2.58	+2.74	+2.91
BQSquare	34.32	-0.06	+0.51	+1.46	+1.59	28.13	-0.01	+2.74	+2.06	+2.85
Four People	39.92	-0.02	-0.39	+0.29	+0.02	37.81	+0.09	-0.05	+1.46	+1.68
Kimono	36.98	+0.01	-2.44	+0.34	+0.00	33.09	+0.00	-0.03	-1.15	-2.02
Park Scene	35.18	+0.06	-1.27	+1.11	+1.13	30.93	+0.01	+0.73	+0.83	+1.02
People on Street	29.21	+0.03	-0.53	+0.58	+0.95	24.57	+0.00	+1.70	+0.09	+0.35
Race Horses	30.89	+0.09	-0.34	+1.38	+1.42	25.71	+0.01	+1.30	+0.58	+0.71
Tennis	33.86	+0.16	+0.18	+0.98	+1.48	27.74	+0.07	+1.35	+1.17	+2.09
Average	–	**+0.06**	**-0.37**	**+0.98**	**+1.05**	–	**+0.05**	**+1.31**	**+1.07**	**+1.33**

than the reference methods, but also higher quality gains than using a constant λ (i.e., without saliency). This confirms the two-fold superiority of adaptive λ, i.e. not only the total overhead is reduced but also higher quality gains are achieved.

C. Error propagation evaluation

In this section the error propagation is evaluated using single loss events and compared against the reference methods. In this simulation a whole frame loss is enforced in order to assess the impact of error propagation in the worst case scenarios. Experimental evaluations were performed for both intra and inter-coded frame loss. Figure 4 shows the PSNR over a GOP with an IDR period of 16 frames, whole frame #6 is missing in sequences Basketball Drill and People On Street. The figure shows the results obtained by decoding the stream using the proposed method with adaptive λ but without any information sent in the enhancement layer. For comparison, the PSNR obtained from the reference HEVC decoder using *MC* and *Ref* for EC of the lost frame.

Figure 5 shows the PSNR for the same conditions but with a loss event on the intra-coded Frame #16, in sequences Kimono and Tennis. These results reveal that the proposed method is able to improve the reconstruction of inter-coded frames and also outperforms the existing techniques whenever a intra-coded frame is missing.

These results indicate that the proposed method outperforms both reference methods since the time instant when the loss occurs. The lost frame itself is recovered with much higher quality than using the reference EC methods. This is due to the use of the optimal EC mode to recover each ECU, following the SEI information received from the encoder. In this case, the proposed method is able to achieve quality gains up to 5 dB, decreasing the error propagation and increasing the PSNR of affected frames by up to 6 dB.

The visual impact of using the proposed EC-aware encoding can be evaluated by observing the recovered frame (i.e., frame #6) of Basketball Drill as an example. Figure 6 shows that higher visual quality and less reconstruction artefacts are obtained in comparison with the *MC* method. It is worthwhile to note that the proposed method is able to reconstruct the object edges with higher accuracy than the reference *MC*.

D. Overall performance

The overall performance was evaluated using both the optimal EC mode decision and the enhancement layer carrying the EC residue coded at different rates, i.e., 10% and 20% of the total bitrate. As in the previous simulations, the packets were randomly discarded to simulate loss events, at different PLR. In this experiment both layers, i.e., video stream and enhancement layer with the EC residue, may be affected by errors.

Table VI and VII presents the average PSNR and FSIM (respectively) of decoded video for different PLRs and different percentages of total overhead (i.e., signalling rate plus enhancement layer rate). The absolute quality values are shown for the reference case (*Ref*), while the quality differences are presented for the remaining methods. These results show that transmitting the EC residue in the enhancement layer further improves the reconstruction quality obtained when only the optimal EC method is used (see previous section). By only using 10% of the total bit rate, the proposed EC-aware coding scheme is able to outperform the reference HEVC, the *MC* and the method proposed in [19] for both Low-delay and Random-access configurations. This is confirmed by the quality gains in the PSNR and FSIM metrics. Using 20% of the total bitrate the quality gains increase up to 6.90 dB for the Basketball Drill sequence when comparing with the reference HEVC, achieving an average gain of 3.79 dB (2.49 for FSIM) for

TABLE VII: Average FSIM for different methods and different overhead ratios.

Sequence	PLR=3%					PLR=10%				
	Ref	MC	[19]	Proposed 10%	Proposed 20%	Ref	MC	[19]	Proposed 10%	Proposed 20%
				Low-delay configuration						
Basketball Drill	96.96	+1.13	+1.63	+1.83	+1.75	92.74	+2.38	+4.26	+4.44	+4.80
Book Arrival	98.61	+0.36	+0.24	+0.62	+0.69	96.66	+1.05	+1.19	+1.85	+2.15
BQSquare	98.03	+0.39	+0.14	+0.42	+0.34	96.31	+1.15	+0.89	+1.33	+1.36
Four People	99.41	+0.09	-1.20	+0.07	-0.03	98.97	+0.31	-3.35	+0.38	+0.32
Kimono	98.54	+0.48	-1.75	+0.84	+0.94	95.79	+0.09	-4.03	+0.28	+0.31
Park Scene	98.74	+0.58	+0.47	+0.84	+0.83	96.41	+0.83	+0.63	+1.63	+1.98
People on Street	98.19	+0.67	+0.26	+0.90	+1.03	94.65	+0.45	+0.68	+0.82	+1.70
Race Horses	95.16	+1.78	+2.38	+2.77	+3.00	88.25	+1.18	+3.48	+3.42	+4.41
Tennis	96.64	+0.98	+1.43	+1.55	+2.00	91.02	+2.20	+4.17	+3.85	+5.41
Average	**–**	**+0.72**	**+0.40**	**+1.09**	**+1.17**	**–**	**+1.07**	**+0.88**	**+2.00**	**+2.49**
				Random-access configuration						
Basketball Drill	98.07	+0.08	-0.17	+0.29	+0.27	95.11	+0.11	+0.75	+0.88	+1.05
Book Arrival	98.99	+0.00	+0.08	+0.18	+0.16	97.71	+0.03	+0.50	+0.47	+0.46
BQSquare	97.83	-0.02	-0.65	+0.23	+0.26	94.71	+0.01	+0.63	+0.93	+1.24
Four People	99.53	-0.01	-0.13	+0.02	+0.00	99.15	+0.00	-0.38	+0.23	+0.27
Kimono	98.83	+0.00	-1.78	+0.09	+0.07	96.50	+0.00	-0.32	-0.24	-0.50
Park Scene	98.80	+0.01	-0.70	+0.34	+0.32	96.51	+0.00	+0.57	+0.62	+0.69
People on Street	98.27	+0.02	-0.95	+0.10	+0.23	94.90	+0.01	+1.13	+0.00	+0.25
Race Horses	96.77	+0.05	-0.98	+0.39	+0.41	91.65	+0.01	+1.31	+0.36	+0.47
Tennis	97.25	+0.10	-0.23	+0.29	+0.37	92.50	+0.09	+1.19	+0.55	+1.36
Average	**–**	**+0.03**	**-0.61**	**+0.21**	**+0.23**	**–**	**+0.03**	**+0.60**	**+0.42**	**+0.59**

the Low-delay configuration, and 1.33 dB (0.59 for FSIM) for the Random-access.

When comparing the results of Tables VI and VII with the spatial and temporal information of each sequence (SI/TI in Table I) it is noticeable that higher quality gains are achieved for sequences with higher motion information, i.e., Basketball Drill and Tennis sequences. Moreover, for sequences with higher spatial information (e.g., BQSquare and People on Street) considerable quality gains are also achieved. This indicates the superior performance of the proposed method when dealing with complex sequences, both in temporal and spatial levels.

Finally, the results presented in this sub-section confirm the results of Sub-section IV-B, since the use of EC residue is more efficient for those sequences where higher gains are achieved with the proposed EC method.

Visual quality evaluation

To provide an informal subjective quality comparison between the proposed method (using 20% of overhead) and the reference HEVC using *MC* as error concealment, Figure 7 shows a region of the decoded frames #8 and #12 when Frame #6 is lost. The visible artefacts demonstrate how the decoder is able to recover from whole frame loss in both methods, and also the impact of error propagation. Figure 7 shows that the proposed method clearly improves the reconstructed quality in comparison with *MC*.

A closer look at the Race Horses sequence reveals that the horse rider is not correctly reconstructed by the *MC* method (highlighted by the circles and ellipses). As shown in the figure, this not only results in a more accurate reconstruction of the lost frame, but also increases the visual quality of the leading frames. In the Kimono sequence, some artefacts are also noticeable in the *MC* case, but not visible in the proposed method (highlighted with circles). Moreover, the edges of the women's kimono (highlighted with ellipses) are also improved by the proposed method, resulting in a smooth reconstruction without error propagation. From the results discussed above it is clearly observed that the proposed method is able to reconstruct the contours of the moving parts more accurately than traditional EC algorithms.

Comparison with previous works

The results were compared with previous works where fair comparisons could be made. In [35] a method was proposed to improve the recovery quality by signalling the suitable MV to be used for EC. The proposed EC-aware scheme of Figure 1 is able to achieve similar quality gains with less amounts of redundancy achieving a better trade-off between quality and overhead. In comparison with the work described in [46], the proposed method is able to select among more complex EC methods which allows for higher quality gains. Both of the above mentioned methods search for the optimal EC method at the encoder-side, which results in similar complexity as the proposed method.

The methods in [21], [22] both use a feedback channel and intra-refresh to reduce the error propagation. The results presented in [22] revealed an average quality gain of approximately 1.68 dB. Since the proposed EC-aware scheme is able to achieve higher quality without requiring any feedback channel, it presents better performance and functional advantages. An advantage of using a feedback channel is the low extra complexity required to implement robust coding because the encoder receives specific information about transmission errors. This allows straightforward spatio-temporal localization of the lost data without computational effort.

Table VIII shows the average quality differences between the reference method (*Ref*) and the proposed one plus three other previous works. Two different test conditions are shown in the Table: error-free and PLR=10%. Moreover, an efficiency
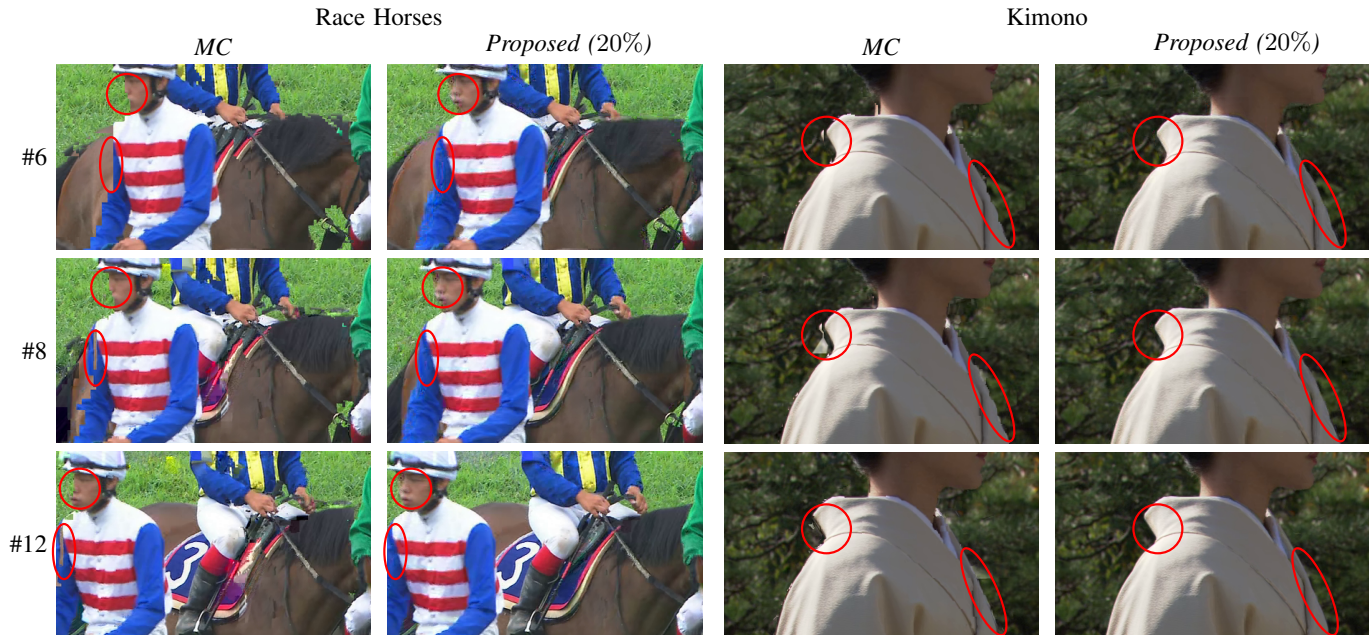
Fig. 7: Crop of the decoded video frame for the Race Horses and Kimono sequences (Loss event at Frame #6)

measure R is defined as the ratio between the quality gains obtained under packet loss and the quality decrease in the error-free case, i.e.,

$$R = \left| \frac{\Delta PSNR(PLR = 10\%)}{\Delta PSNR(Error\text{-}free)} \right|. \qquad (9)$$

The results shown in Table VIII, demonstrate that the effectiveness of the EC-aware coding scheme is higher than all other methods under comparison. In fact, not only the absolute PSNR difference (i.e., PSNR gain) is higher for the same PLR, but also the efficiency of the overhead is also higher. This means that the rate spent to encode the signalling information and the enhancement residue carry more useful information than the other methods. For instance, in comparison with *Ref [25]* the proposed method leads to higher quality degradation in the error-free case. However, for PLR=10% it clearly outperforms *Ref [25]* with a much higher value of R, indicating a far superior efficiency. Results in the table VIII also show that the small loss of coding efficiency in error-free transmission is clearly surpassed by the quality gains under error-prone transmission conditions. The methods in [19], [25] are based on modified predictions, while the method in [29] uses redundant data without any further optimisation. Both of these approaches require low additional computation complexity. However, as observed in the results shown in Table VIII the lower additional complexity also corresponds to lower quality gains in case of transmission errors.

The proposed method uses EC techniques with different levels of computational complexity for each ECU. Since an adaptive approach is used, the average complexity tends to be lower in comparison with other methods that rely on a fixed approach. Thus, less complex video decoding is obtained when using the proposed method when compared with other straight-forward methods that always used the same EC techniques for all CTUs.

TABLE VIII: Comparison against existing methods.

Method	Δ PSNR		R
	Error-free	PLR=10%	
[19]	-0.88	1.64	1.9
[25]	-0.24	1.42	5.9
[29]	-0.37	2.03	5.5
Proposed	-0.35	2.10	6.0

V. CONCLUSIONS

In this paper an EC-aware encoder is proposed by incorporating loss simulation at the encoder-side to enhance the EC efficiency at the decoder. The novelty of this method is to include optimal EC mode decision at the encoder, to be used by the decoder in case of data loss, and also to allow encoding of the EC residue using the enhancement layer of the HEVC standard. The overall results demonstrate that the proposed method is able to efficiently reconstruct lost frames and reduce the error propagation compared with other reference methods. A consistent quality improvement achieved at the cost of low overhead, allows the use of this method in broadcasting video services and applications where packet loss probability should not 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] Y. Yuan, I.-K. Kim, X. Zheng, L. Liu, X. Cao, S. Lee, M.-S. Cheon, T. Lee, Y. He, and J.-H. Park, "Quadtree based nonsquare block structure for inter frame coding in high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1707–1719, Dec. 2012.

[3] J. Lainema, F. Bossen, W.-J. Han, J. 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] I.-K. Kim, S. Lee, M.-S. Cheon, T. Lee, and J. Park, "Coding efficiency improvement of HEVC using asymmetric motion partitioning," in *Broadband Multimedia Systems and Broadcasting (BMSB), 2012 IEEE International Symposium on*, Jun. 2012, pp. 1–4.

[5] 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.

[6] 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)*, Dec. 2012, pp. 785–788.

[7] J. Nightingale, Q. 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.

[8] ——, "Subjective evaluation of the effects of packet loss on HEVC encoded video streams," in *IEEE Third International Conference on Consumer Electronics (ICCE)*, Sep. 2013, pp. 358–359.

[9] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for h.264 video coding over error-prone networks," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 445–454, April 2007.

[10] 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," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 404–412, May 2012.

[11] T. Schierl, M. Hannuksela, Y.-K. 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.

[12] D. Schroeder, A. Ilangovan, M. Reisslein, and E. Steinbach, "Efficient multi-rate video encoding for HEVC-based adaptive HTTP streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, Aug. 2016.

[13] J. Gorostegui, A. Martin, M. Zorrilla, I. Alvaro, and J. Montalban, "Broadcast delivery system for broadband media content," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2017, pp. 1–9.

[14] K. Park, Y. Lim, and D. Y. Suh, "Delivery of ATSC 3.0 Services With MPEG Media Transport Standard Considering Redistribution in MPEG-2 TS Format," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 338–351, Mar. 2016.

[15] J. Wu, B. Cheng, M. Wang, and J. Chen, "Delivering high-frame-rate video to mobile devices in heterogeneous wireless networks," *IEEE Transactions on Communications*, pp. 1–1, 2016.

[16] J. Wu, C. Yuen, M. Wang, and J. Chen, "Content-aware concurrent multipath transfer for high-definition video streaming over heterogeneous wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 3, pp. 710–723, Mar. 2016.

[17] K. Park, N. Kim, and B. D. Lee, "Performance evaluation of the emerging media-transport technologies for the next-generation digital broadcasting systems," *IEEE Access*, vol. 5, pp. 17 597–17 606, 2017.

[18] A. Vetro, J. Xin, and H. Sun, "Error resilience video transcoding for wireless communications," *IEEE Wireless Communications*, vol. 12, no. 4, pp. 14–21, Aug. 2005.

[19] 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.

[20] G. Kulupana, D. S. Talagala, H. K. Arachchi, and A. Fernando, "Error resilience aware motion estimation and mode selection for HEVC video transmission," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2016, pp. 85–86.

[21] H. M. Maung, S. Aramvith, and Y. Miyanaga, "Error resilience aware rate control and mode selection for HEVC video transmission," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2017, pp. 374–375.

[22] ——, "Region-of-interest based error resilient method for HEVC video transmission," in *15th International Symposium on Communications and Information Technologies (ISCIT)*, Oct. 2015, pp. 241–244.

[23] B. Katz, S. Greenberg, N. Yarkoni, N. Blaunstien, and R. Giladi, "New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 308–319, Mar. 2007.

[24] J. Carreira, P. A. Assuncao, S. M. M. de Faria, E. Ekmekcioglu, and A. Kondoz, "A two-stage approach for robust HEVC coding and streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–4, 2017.

[25] D. Liu, L. Wang, C. Li, Y. Hao, and F. Yin, "Error protection with extended dual frame motion compensation," in *IEEE International Symposium on Multimedia (ISM)*, Dec. 2015, pp. 331–334.

[26] 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.

[27] 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 *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct. 2014, pp. 2457–2461.

[28] C. Zhu, Y.-K. Wang, M. Hannuksela, and H. Li, "Error resilient video coding using redundant pictures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 1, pp. 3–14, Jan. 2009.

[29] D. Chen, N. Gadgil, and E. J. Delp, "VPx video coding for lossy transmission channels using error resilience packets," in *2016 Picture Coding Symposium (PCS)*, Dec 2016, pp. 1–5.

[30] P. Correia, P. Assuncao, and V. Silva, "Multiple description of coded video for path diversity streaming adaptation," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 923–935, Jun. 2012.

[31] H. Ha, J. Park, S. Lee, and A. C. Bovik, "Perceptually unequal packet loss protection by weighting saliency and error propagation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 9, pp. 1187–1199, Sep. 2010.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] S. C. Huang and S. Y. Kuo, "Optimization of hybridized error concealment for H.264," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 499–516, Sep. 2008.

[37] J.-T. Chien, G.-L. Li, and M.-J. Chen, "Effective error concealment algorithm of whole frame loss for H.264 video coding standard by recursive motion vector refinement," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 3, pp. 1689–1695, Aug. 2010.

[38] S. K. Bandyopadhyay, Z. Wu, P. Pandit, and J. M. Boyce, "An error concealment scheme for entire frame losses for h.264/avc," in *Sarnoff Symposium, IEEE*, Mar. 2006, pp. 1–4.

[39] M. C. Hwang, J. H. Kim, D. T. Duong, and S. J. Ko, "Hybrid temporal error concealment methods for block-based compressed video transmission," *IEEE Transactions on Broadcasting*, vol. 54, no. 2, pp. 198–207, Jun. 2008.

[40] 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.

[41] 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.

[42] B. Yan, H. Gharavi, and B. Hu, "Pixel interlacing based video transmission for low-complexity intra-frame error concealment," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 253–257, Jun. 2011.

[43] C. Yeo, W. T. Tan, and D. Mukherjee, "Receiver error concealment using acknowledge preview (RECAP) - an approach to resilient video streaming," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Apr. 2009, pp. 785–788.

[44] H. Hadizadeh, I. V. Baji, and G. Cheung, "Video error concealment using a computation-efficient low saliency prior," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 2099–2113, Dec. 2013.

[45] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[46] E. S. Ryu, Y. Ryu, H. J. Roh, J. Kim, and B. G. Lee, "Towards robust UHD video streaming systems using scalable high efficiency video coding," in *International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2015, pp. 1356–1361.

[47] J. S. Pérez, E. Meinhardt-Llopis, and G. Facciolo, "Tv-l1 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.

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

[49] D. Hawkins, "Constrained optimization and lagrange multiplier methods," *Athena Scientific, Belmont*, 1982.

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

[51] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

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

[53] ITU-T, "Recommendation P.910, Subjective video quality assessment methods for multimedia applications," 1998.

[54] JCT-VC, "HM 16.2 reference software," Oct. 2014.

[55] X. Feng, T. Liu, D. Yang, and Y. Wang, "Saliency inspired full-reference quality metrics for packet-loss-impaired video," *IEEE Transactions on Broadcasting*, vol. 57, no. 1, pp. 81–88, Mar. 2011.