A Two-stage Approach for Robust HEVC Coding and Streaming

J. Carreira, P. A. Assuncao, S. M. M. de Faria, E. Ekmekcioglu and A. Kondoz

Abstract—The increased compression ratios achieved by the High Efficiency Video Coding (HEVC) standard lead to reduced robustness of coded streams, with increased susceptibility to network errors and consequent video quality degradation. This paper proposes a method based on a two-stage approach to improve the error robustness of HEVC streaming, by reducing temporal error propagation in case of frame loss. The prediction mismatch that occurs at the decoder after frame loss is reduced through the following two stages: (i) at the encoding stage, the reference pictures are dynamically selected based on constraining conditions and Lagrangian optimisation, which distributes the use of reference pictures, by reducing the number of prediction units (PUs) that depend on a single reference; (ii) at the streaming stage, a motion vector (MV) prioritisation algorithm, based on spatial dependencies, selects an optimal sub-set of MVs to be transmitted, redundantly, as side information to reduce mismatched MV predictions at the decoder. The simulation results show that the proposed method significantly reduces the effect of temporal error propagation. Compared to the reference HEVC, the proposed reference picture selection method is able to improve the video quality at low packet loss rates (e.g., 1%) using the same bitrate, achieving quality gains up to 2.3 dB for 10% of packet loss ratio. It is shown, for instance, that the redundant MVs are able to boost the performance achieving quality gains of 3 dB when compared to the reference HEVC, at the cost using 4% increase in total bitrate.

Index Terms—HEVC, robust coding, reference picture selection, redundant MV, error propagation.

I. INTRODUCTION

THE current diversity of multimedia applications and services, and the emergence of Ultra-HD formats (e.g., 4k or 8k resolution) are continuously reinforcing the need for efficient video coding. Moreover, the increasing amount of mobile multimedia traffic demands for more bandwidth, and the use of higher resolutions impose more challenging bounds on quality and error tolerance. The High Efficiency Video Coding (HEVC) [1] is the most recent standard developed by the Joint Collaborative Team on Video Coding (JCT-VC), essentially aiming to extend previous standards towards increased video resolutions and use of parallel processing architectures. The higher coding flexibility and efficiency of

J. Carreira is with Institute for Digital Technologies, Loughborough University London, 3 Lesney Avenue, Here East, Queen Elizabeth Olympic Park, London E15 2GZ, UK and Instituto de Telecomunicações, 2411-901 Leiria, Portugal, (e-mail: j.f.m.carreira@lboro.ac.uk).

P. A. Assuncao and S. M. M. de Faria are with Instituto de Telecomunicações, and Instituto Politécnico de Leiria, 2411-901 Leiria, Portugal (e-mail: {amado,sergio.faria}@co.it.pt).

E. Ekmekcioglu and A. Kondoz are with Institute for Digital Technologies, Loughborough University London, 3 Lesney Avenue, Here East, Queen Elizabeth Olympic Park, London E15 2GZ, UK (e-mail: {e.ekmekcioglu,a.kondoz}@lboro.ac.uk, tel: +442038051325).

HEVC results from the use of new block partition structures, enabling flexible block partitions [2], improved prediction modes [3], [4] and new high-level features [5], such as explicit reference picture management and new parameter sets.

More efficient prediction structures allow higher bitrate savings, but also to some disadvantages, such as increased complexity [6] and reduced error robustness. Although the complexity may not pose significant problems due to the rapid evolution of hardware technologies, reduced error robustness strongly affects quality under packet loss conditions. Another critical aspect related to the performance of robust video streaming methods is the error detection accuracy. Since HEVC streaming is expected to use known transport technologies, such as RTP [7], MPEG-2 TS [8], or more recent standards as MPEG-DASH [9] and MPEG Media Transport [10], their error detection capabilities have significant impact on the overall system performance. In general, the error detection operation at the transport layers is decoupled from the video coding layer. The encoding side implements the robust coding methods, while the efficient recovery of lost video data is implemented in the decoder. Between them, an inter-layer signalling ensures that the decoder is given the necessary information to identify the lost slices/frames in the video data. Therefore, stream robustness can be investigated on its own, assuming any possible form of error detection.

The error robustness characteristics of HEVC analysed in [11], [12] confirmed that, in general, HEVC presents poor error resilience performance, despite its superior coding efficiency. Thus, HEVC bitstreams subject to transmission losses result in significant degradation of both objective and subjective quality [12], [13]. 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, assuming error-free transmission [14]. Therefore, high compression efficiency is achieved, but resilience to transmission errors is penalised. Moreover, the use of advanced motion vector (MV) prediction in HEVC [15] also contributes to increase the spatial and temporal dependencies to a great extent. Although the spatial MV candidates are used more often, temporally predicted MVs are responsible for a greater impact on error propagation.

In general, increased error robustness in video streaming can be jointly achieved at two different stages: (i) while encoding live video, by selecting robust coding modes and (ii) after encoding, by adding robust features to the compressed streams. While in the former this is accomplished by optimising coding parameters and decisions to increase robustness, in the latter some type of compressed domain processing must be used to

increase error robustness of pre-encoded streams, e.g., error resilient transcoding. Inevitably, in both cases this is achieved at the expense of some loss in coding efficiency, therefore the challenge is to find an optimum trade-off between robustness and bitrate overhead.

This paper addresses the problem of HEVC robust streaming by combining the two stages previously referred. Besides the novelty of the methods devised for each stage, a distinct characteristic of the proposed scheme relies on its ability to cope with increased robustness requirements of both live video and pre-encoded streams. At the encoding stage, an optimal reference picture selection method is proposed to reduce error propagation in case of frame loss, by distributing the temporal dependencies across different frames. A Lagrangian R-D cost function is devised to select the optimal reference frame for each prediction unit. At the streaming stage, the MVs are parsed from the coded stream to optimally select a small set of them, based on a priority criterion derived from MV spatial dependencies. Then the optimal set of MVs are transmitted as side information to substitute predicted MVs, whenever their references cannot be decoded due to data loss.

The remainder of the paper is organised as follows: Section II provides an overview of the previous studies addressing error-prone video transmission and related work on methods to improve error robustness; Section III describes motion vector prediction in HEVC and evaluates its error resilience performance; Section IV describes the proposed method, and Section V covers the discussion of the experimental results. Finally Section VI concludes the paper.

II. RELATED WORK

The error robustness of HEVC streaming over diverse transport protocols has been addressed in the literature, where relevant insight can be found regarding various aspects that influence the quality delivered to end users. For instance, previous works have shown that HEVC streams are highly sensitive to bandwidth fluctuations, which may be responsible for decreasing the quality (i.e., PSNR) by up to 4 dB for just 10% of bandwidth reduction [8], [7]. This is not only due to the intrinsic nature of predictive coded video, but it is also related to the error detection capabilities of transport and encapsulation protocols, which may have significant impact on the overall performance [16], [17]. In this regard, the performance of error detection is known to be also highly dependent on the syntax elements affected by errors [18]. In the past, different approaches have been used for this purpose. For instance, in [19] a content-based approach was proposed to detect errors based on the spatial and temporal similarities in video sequences, while in [20] extra information is multiplexed into the video stream to detect errors and ease the recovery of wrongly decoded regions. In this paper, error detection is assumed to be handled by the lower layers of the video communication system, such that the decoder is aware of the losses occurring in the video signal, including their spatial and temporal locations.

The error resilience techniques, normally used in robust coding and streaming to mitigate the artefacts caused by packet losses, can be grouped into four categories: localisation, data partitioning, redundant coding and concealment-driven techniques [21]. In [22], a feedback-based transcoder is used to reduce error propagation at end-user decoders, by dynamically selecting the correctly received frames as references. This was further exploited in [23] to limit the complexity, by using MV concatenation in motion re-estimation. In [24] feedback-based is combined with path diversity to select the reference frames that are more likely to be received at the decoder. Although these closed-loop techniques may offer a reliable control of the decoder distortion, the feedback required from the receiver side and multipath networking are not always available neither feasible in real-time and unidirectional channels (e.g. broadcast).

A R-D optimisation method was presented in [25], based on the recursive optimal per-pixel distortion estimation (ROPE). In [26] the transmission distortion is estimated at the encoder-side to assign different priorities to each packet, allowing different levels of error protection. In [27], an optimisation algorithm was used to select between short and long term references, in order to increase the prediction distance and reduce error propagation. In the decoder side, an error concealment method was later used to take advantage of such prediction structure [28]. The use of leaky prediction is also able to achieve an exponential decay of error propagation, but coding efficiency is significantly penalised [29]. Although these methods are able to efficiently predict the decoder distortion in case of packet losses, they rely upon either interpolated reference frames or long-term references, which might dramatically decrease the coding performance.

The robustness of transmitted streams may also be increased through bitstream partitioning and unequal error protection. This is done in [30], by assigning different importance to each macroblock and then using the Flexible Macroblock Ordering (FMO) mechanism of H.264/AVC to encode and transmit the more important slices with higher levels of error protection. Different types of multiple description coding [31] can also be used to improve the quality under transmission errors in multiple path networks [32]. Although these techniques can be very effective in multi-path networks, a significant amount of overhead is required to create the different descriptions.

The vulnerability of MV prediction in HEVC was evaluated in [33], showing its weak error resilience in contrast to the coding improvements. As a solution to overcome such vulnerability, the temporal motion vector predictor can be disabled at constant frame intervals. In [34] this idea was extended to the prediction unit (PU) level achieving a more robust MV prediction, without compromising error resilience. Redundant pictures were used in [35] to embed redundant MVs without encoding any residual information, in order to increase error robustness with low penalty in coding efficiency. While this method may be more efficient than using standard redundant pictures [36], since all MVs are encoded, it may still lead to high redundancy without always being worthwhile in terms of increased robustness. This approach was further improved in [37], where the authors proposed a method based on selective motion information, rather than the entire motion field. In the scope of HEVC, combining R-D efficiency and

robust coding with error concealment of missing frames, as proposed in this paper, is still an open issue, without well established solutions.

The problem of robust transmission may also be addressed from the decoder side, by using coding methods that facilitate error concealment at the decoder [38], [39]. 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 presented in [41], where the authors concluded that a higher residue implies less accurate MVs. In general, these methods rely on post-processing of past frames, which may lead to incorrect recovery of motion information, because in general motion is neither constant nor purely translational. Therefore, simple error concealment techniques might not be capable of mitigating all the effects of transmission errors, and specific coding techniques are necessary in order to improve the robustness of compressed video streams in the presence of data loss.

This paper advances one step further beyond previous works, by proposing an integrated two-stage robust streaming approach to address the problem of error resilience in HEVC. The encoding stage is based on a reference picture selection scheme that uses a R-D optimised reference picture assignment, without requiring transcoding or network feedback as in previous methods [22]–[24]. Therefore, this method does not depend on the network characteristics (e.g., feedback channel) and does not increase complexity to estimate the end-to-end distortion. Moreover, as the proposed scheme does not use long-term neither interpolated reference frames, the impact in the coding efficiency is lower. The streaming stage reduces the impact of mismatched MV temporal dependencies on the error robustness of HEVC, by selecting a number of MVs for redundant transmission in the post-processing stage. This paper extends the authors' previous work [37], by finding the optimal set of MVs for each frame, rather than using a fixed pre-defined amount of MVs. Moreover, it further improves the performance of [34], which followed a different approach based on constrained inter-coding modes. In comparison to other similar approach that uses the whole motion field [35], the proposed method has the advantage of using a much smaller amount of redundant MVs.

III. ANALYSIS OF HEVC ERROR ROBUSTNESS

To highlight the impact of transmission errors in HEVC and the importance of error propagation, an experimental study about the error robustness of HEVC is presented in this section. Special emphasis is given to the use of temporal motion vector prediction due to its relevance for this work. A brief description of the Merge Mode in HEVC is first described and then the evaluation study and results are presented in Section III-A.

In HEVC, the Advanced Motion Vector Prediction (AMVP) and Merge Mode [15] are based on MV predictions, which obviously leads to mismatch errors at the decoder whenever a packet loss affects motion data. Figure 1 illustrates the spatial MV candidates allowed in the Merge Mode and their
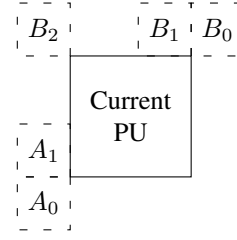

Fig. 1. Spatial MV predictor positions used in the Merge Mode of HEVC.

relative positions, where each dashed square corresponds to one prediction unit (PU) from which a MV candidate is derived, for possible prediction of the MV in the current PU. These candidates are checked in the following order: A_1, B_1, B_0, A_0, B_2. Moreover, a temporal MV candidate (TMVP) is also used, derived from the co-located position in the temporally adjacent frame. By adding a temporal candidate in the set of possible MVs creates temporal dependencies between consecutive frames. This leads to higher error propagation and inherent quality degradation in the presence of packet loss, as demonstrated by the experimental study presented in the next section. These results provide relevant insights on how HEVC error resilience may be improved.

A. Evaluation of error robustness

This section presents an evaluation study about the error robustness of HEVC using different sequences and coding configurations. Table I presents a summary of the main characteristics of the video sequences used in this work. Each sequence comprises 240 frames and the whole set covers

TABLE I
TEST SEQUENCES USED IN THE EXPERIMENTS.

Sequence	Resolution	SI/TI	Description
Basketball Drill	832 × 480 50 fps	33.4 / 14.4	High motion with several basketball players
Book Arrival	1024 × 768 30 fps	28.4 / 21.7	Moderate translational motion with two moving persons
Bosphorus	3840 × 2160 120 fps	13.4 / 3.8	Boat shipping at low motion with moderate complex background
BQSquare	416 × 240 60 fps	63.2 / 11.5	Moderate outside motion with moving camera capturing from high point
Cactus	1920 × 1080 50 fps	30.5 / 11.4	Several objects with high details with moderate motion
Jockey	3840 × 2160 120 fps	11.5 / 16.2	High motion with one horse rider
Kendo	1024 × 768 30 fps	19.6 / 16.1	Moderate motion with two moving persons, and moving camera
Park Scene	1920 × 1080 24 fps	31.3 / 11.6	Moderate motion with cyclists passing across the scene
People On Street	2560 × 1600 24 fps	40.0 / 25.4	With point capture of people moving; high motion and texture complexity
Race Horses	832 × 480 30 fps	43.7 / 24.4	High motion with several horse riders
Tennis	1920 × 1080 30 fps	20.3 / 45.3	High motion with one moving person in the scene
Traffic	2560 × 1600 30 fps	35.5 / 25.4	Elevated camera capturing highway with several cars with moderate speed

TABLE II
AVERAGE PSNR FOR DIFFERENT HEVC CONFIGURATIONS

| Sequence | Configuration | Low-Delay | | | | | Random-Access | | | | |
| | | IDR_{Loss} | | IDR_{NoLoss} | | | IDR_{Loss} | | IDR_{NoLoss} | | |
		1%	3%	1%	3%	5%	1%	3%	1%	3%	5%
Basketball Drill	Ref [42]	33.77	28.24	34.44	30.58	28.10	34.84	29.80	36.66	33.12	30.64
	Max_CU_Size_16	+0.36	+0.45	+0.54	+0.06	+0.37	+0.11	-0.13	-0.18	-0.28	+0.04
	Single_Merge_Cand	-0.65	-0.78	-0.42	-0.71	-0.57	+0.12	+0.16	+0.18	-0.09	+0.03
	No_TMVP	+2.16	+3.41	+2.08	+3.12	+3.60	+0.16	+0.41	+0.24	+0.42	+0.41
Book Arrival	Ref [42]	36.28	30.48	37.24	33.04	30.57	37.67	30.65	38.26	35.28	32.41
	Max_CU_Size_16	+0.14	+0.22	-0.23	+0.05	+0.43	-0.70	+0.08	-0.37	-0.40	+0.04
	Single_Merge_Cand	-0.55	-0.98	-0.15	-0.32	-0.96	-0.63	+0.28	-0.20	-0.25	+0.05
	No_TMVP	+1.61	+2.77	+1.29	+2.78	+3.06	-0.04	+0.64	+0.45	+0.62	+0.60
BQSquare	Ref [42]	30.57	23.71	31.39	25.85	23.36	33.64	25.99	37.03	31.66	28.65
	Max_CU_Size_16	+0.64	-0.27	+0.72	-0.46	-0.32	+0.22	-0.20	+0.06	-0.18	-1.45
	Single_Merge_Cand	-0.17	+0.06	-0.13	-0.24	-0.35	-0.14	-0.03	-0.24	-0.23	-1.36
	No_TMVP	+7.19	+10.1	+6.81	+10.2	+11.2	+0.90	+1.39	+0.82	+1.79	+1.79
Kendo	Ref [42]	37.11	30.89	38.66	33.16	29.61	37.75	29.75	39.41	33.94	30.37
	Max_CU_Size_16	+0.39	-0.15	-0.99	-0.20	-0.56	-0.85	-0.19	-0.57	-0.39	-0.84
	Single_Merge_Cand	-0.19	-0.87	-0.23	+0.16	+0.33	-0.52	-0.04	-0.07	-0.09	-0.53
	No_TMVP	+2.24	+4.08	+1.79	+3.63	+4.19	+0.20	+0.74	+0.37	+0.75	+0.36
Park Scene	Ref [42]	32.15	28.20	33.36	30.07	28.29	33.44	29.46	35.70	33.12	31.33
	Max_CU_Size_16	+0.42	+0.03	-0.05	-0.22	-0.51	-0.01	+0.08	-0.23	-0.44	-0.78
	Single_Merge_Cand	-0.19	-0.38	-0.16	-0.25	-0.73	-0.02	+0.01	+0.10	-0.58	-0.63
	No_TMVP	+2.82	+4.99	+2.70	+4.76	+5.35	+0.94	+1.27	+0.48	+0.90	+1.15
Race Horses	Ref [42]	28.60	23.48	29.39	24.78	22.79	31.07	26.02	32.46	28.26	25.74
	Max_CU_Size_16	+0.10	-0.19	+0.04	+0.04	-0.44	+0.01	-0.25	-0.18	-0.32	-0.19
	Single_Merge_Cand	-0.13	-0.39	-0.36	-0.24	-0.45	-0.08	-0.12	+0.02	-0.27	-0.04
	No_TMVP	+3.62	+4.78	+3.61	+4.63	+4.70	+0.93	+0.92	+0.76	+0.88	+1.10
Tennis	Ref [42]	32.53	26.88	33.94	28.18	25.86	35.13	29.64	36.52	31.48	29.01
	Max_CU_Size_16	-0.79	-0.66	-1.23	-0.29	-0.49	-0.53	-0.26	-0.82	-0.37	-0.61
	Single_Merge_Cand	-0.06	-0.63	-0.39	-0.13	-0.30	+0.01	+0.31	-0.09	+0.14	-0.13
	No_TMVP	+2.95	+3.37	+2.80	+3.65	+3.68	+0.81	+0.74	+0.41	+0.79	+0.60
Traffic	Ref [42]	21.95	20.09	34.16	29.90	28.12	32.46	28.03	35.73	31.77	29.61
	Max_CU_Size_16	+5.19	+3.75	-0.77	+0.61	+0.19	+0.03	-0.02	-0.14	-0.26	-0.08
	Single_Merge_Cand	+9.19	+6.06	-0.68	-0.60	-0.98	-0.01	-0.04	+0.04	-0.05	+0.03
	No_TMVP	+13.6	+12.8	+3.80	+6.37	+6.75	+1.33	+1.13	+0.90	+1.43	+1.59

different types of motion and texture complexity. This table also shows the spatial information (SI) and the temporal information (TI), as defined in [43]. The experimental results for this study were obtained by using the HM reference software, version 16.2. The sequences were encoded with all coding modes enabled 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) [42]. These prediction structures cover the use of P-frames and hierarchical B-frames [44], which are commonly used in HEVC streaming [45]. To simulate a broadcast transmission an IDR period of 32 frames was used and the filtering 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 [42]. In order to achieve high quality images for all coded sequences (i.e., PSNR around 40 dB) different bitrates were used. At the decoder side, error concealment is based on motion compensation from previously decoded frames. The motion field of a missing frame is obtained applying MV extrapolation from the two closest neighbours. Afterwards, the final MVs are selected based on the residual energy of their original prediction [41]. To simulate the lossy transmission environment, each frame is encoded into several slices, each one packetized into one NAL unit, which are then transmitted as the payload of independent packets. Two test cases were

considered for simulation: (i) in IDR_{Loss} all frames may be equally affected by packet loss, including IDR frames; (ii) in IDR_{NoLoss} IDR frames are assumed to be prioritised in the network and delivered error-free, or using some combined approach as recently proposed in [46]. In our experiments a whole packet is lost whenever an error occurs, originating a lost slice. Random packet loss was simulated using a two-state Markov model with the following average burst length 1.24, 1.47, 1.83 and 2.05, and a maximum burst length of 2, 3, 5 and 7, for packet loss ratios (PLR) of 1%, 3%, 5%, 10%, respectively.

Four different configurations were used, each one corresponding to specific coding conditions, using different CU sizes and different number of Merge Mode candidates: *Ref [42]*, *Max_CU_Size_16*, *Single_Merge_Cand* and *No_TMVP*. Table II shows the average quality (PSNR) obtained for these four configurations under different PLRs. The reference for comparison (*Ref [42]*) uses a maximum CU size of 64 and 5 spatial MV candidates for the Merge mode, following the recommended test configuration [42]. The following three coding modes were tested and compared with the reference to evaluate the impact of different coding tools in error robustness: (i) reducing the maximum CU size to 16, similar to the H.264/AVC macroblocks (*Max_CU_Size_16*), (ii) using only one MV candidate in the Merge Mode (*Single_Merge_Cand*) and (iii) disabling the temporal MV

candidate: (*No_TMVP*). Table II shows the absolute value of PSNR for the *Ref [42]* case, while for the other configurations the PSNR difference is presented. A positive difference means that better quality is achieved for a different coding configuration with the same PLR.

The results across all configurations and sequences show that the hierarchical coding structure used in Random-Access is more efficient in mitigating the effect of errors than the Low-Delay structure. This is mainly because non-reference B-frames do not propagate errors. Comparing the results of IDR_{Loss} and IDR_{NoLoss}, one can further observe that IDR losses result in worst quality due longer error propagation, as expected.

The results for individual configurations show that reducing either the maximum CU size (*Max_CU_Size_16*) or the number of MV candidates (*Single_Merge_Cand*) do not significantly improve error robustness, because the prediction modes are the same. This leads to the conclusion that inter-frame dependencies play a dominant role in quality degradation. It is interesting to note that the quality gains between IDR_{Loss} and IDR_{NoLoss} is roughly the same for each sequence, configuration and PLR.

When the TMVP is disabled (*No_TMVP*), the results show significant improvements in the reconstructed video quality for different PLR. This confirms that the error propagation decreases when motion information is not encoded with temporal dependencies. Therefore, quality degradation is more prominent in Low-Delay rather than Random-Access. This is because errors are not self-contained within each group of pictures (GOP) and also, since frames are temporally closer, temporal MV predictions are more frequently used. For example in Table II, BQSquare sequence presents the highest PSNR gain when using the *No_TMVP* mode. This is because the smooth translational motion of BQSquare is more friendly for temporal prediction.

These results show that HEVC has poor robustness against transmission errors, resulting in significant quality degradation, though dependent on the video content and coding options. Despite the fact that temporal MV candidates improve coding efficiency, they also significantly contribute to increase the temporal dependencies, leading to more severe quality degradation in the presence of frame loss. Based on this evidence, the next section presents an efficient method to improve the error robustness of HEVC based on the underlying principle of reducing the mismatch in MV predictions caused by temporal dependencies.

IV. PROPOSED METHOD

In this section a two-stage approach for robust HEVC coding and streaming is proposed. The following sub-section presents an overview of the proposed method, and then the stages of the proposed method are outlined.

A. Overview

The goal of the proposed method is to reduce the error propagation in decoded video in case of frame loss. This is
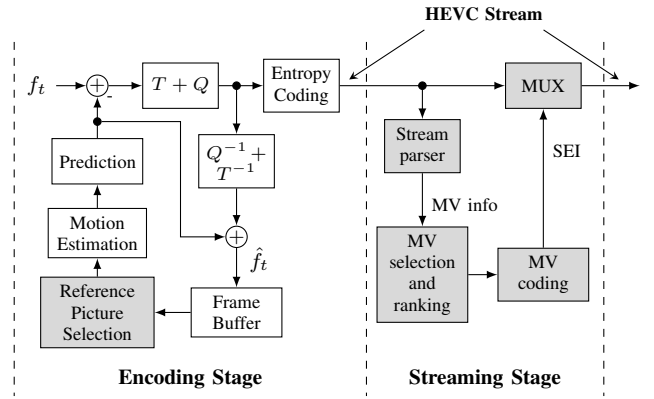

Fig. 2. Functional blocks of the proposed method (gray blocks) in the context of the HEVC coding and streaming.

accomplished by constraining inter-frame prediction dependencies, in order to minimise the distortion D_t given by

$$D_t = \tilde{f}_t - \hat{f}_t, \text{when } \tilde{f}_{t-1} \neq \hat{f}_{t-1}, \quad (1)$$

where $\hat{f}_t$ and $\tilde{f}_t$ are the encoded and decoded frames, respectively, at instant t.

Figure 2 illustrates the block diagram of the two-stage method proposed in this work. At the encoding stage, a reference frame selection mechanism is used for each prediction unit (PU) within a coding tree unit (CTU) forcing the use of different reference frames in adjacent CTUs. The decision criterion contributes to reduce the amount of predictions from a single reference frame, and forces to interleave different reference frames along all encoded PUs, based on a R-D optimised criterion. At the streaming stage, the coded stream is parsed to extract the MV information, which allows to rank MVs according to a relevance criterion. Then, a sub-set of these MVs is selected to be transmitted as side information. The amount of selected MVs is not fixed and can be optimised for different content and end-network characteristics. Note that the streaming stage is an independent and complementary subsystem to the encoder. Thus it is able to operate at network nodes, including the streaming server on pre-encoded streams.

In case of frame loss these two stages have the common objective of helping the decoder to recover from mismatched predictions at the expense of a small increase in bitrate. This is equivalent to reducing the error propagation by using two techniques: (i) reducing the mismatch of inter-frame predictions, by encoding each PU from a different reference frame and (ii) reducing the mismatch of MV predictions by sending the most important MVs as side information. These are the two stages of the proposed method described in the following sub-sections.

B. Encoding stage: Reference frame selection

The reference picture selection mechanism attempts to uniformly distribute the use of different reference pictures across the coding tree units (CTU) of a coded video frame. In case of frame loss, the error propagation in the decoder is reduced because the number of CTUs not predicted from
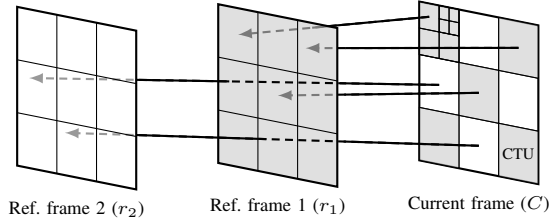
Fig. 3. Example of the reference picture selection applied to the frame C based on the checkerboard at CTU level (PUs represented by small partitions).

the lost frame decreases and thus error concealment benefits from more data correctly decoded. Note that standard compliance is maintained. The underlying principle of this method is to minimise the probability of using the same reference picture for neighbouring CTUs, which tends to result in a checkerboard structure, where a different reference frame is used for predicting every other CTU within the current frame. For instance, in a coding structure with two reference frames, as shown in Figure 3, the reference frames r_1 and r_2 are used to predict the same number of CTUs in current frame (C). If each reference frame is used to predict half of the pixels in the current frame, when any of these references is lost, the subsequent predicted frame is only partially distorted by mismatched predictions. This is because only half of its predictions depend on the missing reference frame, which limits the overall error propagation. Moreover, since the motion information of the predicted frame is encoded with reference not only to the missing frame but also to the past ones, those MVs that cross over a missing reference frame, pointing to the previous ones, are still useful for reconstructing correct MVs and also for error concealment.

In the limit, a straightforward implementation of the previous example leads to a simple interleaved uniform distribution of the reference frames over all CTUs in the current frame, forcing the checkerboard structure shown in Figure 3. In the case of two reference frames, 50% of the CTUs would use one reference frame, while the other 50% would use the other reference. However, such solution would not take into consideration coding efficiency, since the selection process of the reference frame would not be R-D optimised for each prediction unit (PU) within a CTU.

Therefore, a more efficient approach to improve error robustness was devised to dynamically choose the coding mode for each PU, that ensures a balanced use of all possible reference frames $F = \{r_1, r_2, ... r_S\}$ across the CTUs of the current frame C, also optimising the R-D cost. This leads to the overall goal of finding the optimum R-D coding mode for each PU, as the one which tends to minimise the standard deviation of the reference frames use count $\mathbf{N}_C(r_i)$, considering all encoded PUs of the current frame C, as follows:

$$\sigma_C = std\Big(\mathbf{N}_C(r_i)\Big), i = 1, 2, \cdots S. \qquad (2)$$

In this case σ_C measures the deviation from the checkerboard solution, where $\sigma_C = 0$, since all possible reference frames $r_i \in F$ are uniformly used throughout the PUs of C.

To attain the goal defined above, the best coding mode for each PU is chosen through a Lagrangian R-D optimisation that tends to minimise σ_C. This is done by penalising (or benefiting) the R-D cost of those coding modes which intend to use reference frames that have already been used more (or less) frequently than their average use in previous PUs of the current frame. For each PU, this mechanism has the effect of reducing (or increasing) the probability of using each possible reference frame $r_i \in F$, according to the number of previous PUs that have used each one. Such approach allows to select the reference frames, by independently processing each PU in order to minimise the global standard deviation given by (2). Thus, the optimum coding mode ϕ^*, selected from a set of possible coding modes M, is derived based on the following Lagrangian optimisation:

$$\phi^* = arg \min_{\phi \in M} J(\phi) \qquad (3)$$

$$J(\phi) = \Big(D(\phi) + \lambda \times R(\phi)\Big) \times e^{W(r_\phi)} \qquad (4)$$

where $D(\phi)$ and $R(\phi)$ are the distortion and rate, respectively, associated with the coding mode ϕ. The exponential weighting factor $e^{W(r_\phi)}$ is used to penalise (or benefit) the Lagrangian cost of a given coding mode ϕ associated to reference frame r_ϕ. $W(r_\phi)$ is obtained according to the global and local deviation between the number of times that r_ϕ was used and the average use of all $r_i \in F$ across the PUs of current frame C, as follows,

$$W(r_\phi) = \Big(\Delta_G(r_\phi) + \Delta_L(r_\phi)\Big) \times 2^{-T_{ID}} \times \gamma \qquad (5)$$

where T_{ID} is given by the temporal hierarchy in HEVC, and γ is a parameter to control the slope of $W(r_\phi)$. The global and local deviations, $\Delta_G(r_\phi)$ and $\Delta_L(r_\phi)$, respectively, are computed as follows,

$$\Delta_\zeta(r_\phi) = \mathbf{N}_\zeta(r_\phi) - \frac{1}{S} \sum_{i=1}^{S} \mathbf{N}_\zeta(r_i), \ \ \zeta = G, L. \qquad (6)$$

For $\zeta = G$, function $\Delta_G(r_\phi)$ measures the difference between the number of times that reference frame r_ϕ was used and the global average use of all $r_i \in F$ in the PUs of C that were encoded before the current PU. $\mathbf{N}_G(r_i)$ is the number of times that each reference frame r_i was used across all previously encoded PUs. For $\zeta = L$, function $\Delta_L(r_\phi)$ only considers the three top-left neighbouring CTUs for counting the number of times each reference frame was used, which is expressed by $\mathbf{N}_L(r_i)$. Overall, Equation (6) is used to adjust the penalty weight defined by (5) in order to achieve an approximately uniform use of all possible reference frames. This corresponds to the general objective of minimising the values of $\Delta_\zeta(r_\phi)$. The optimum is reached when all reference frames are used to encode the same amount of PUs (i.e., $\Delta_G(r_\phi) = \Delta_L(r_\phi) = 0$), and the original Lagrangian cost is not affected, i.e., $e^{W(r_\phi)} = 1$ in (4). Note that an exponential cost increase is used in (4) in order to increase the penalty associated to any reference frame that tends to be used much more often than the average.

In (5), the γ constant controls the slope of $W(r_\phi)$, leading to higher weights as the value of γ increases. On the one hand
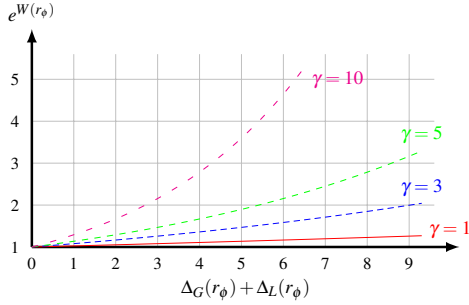
Fig. 4. Exponential weight as a function of the global and local deviations

Fig. 5. Example of temporal and spatial MV dependencies in HEVC.

a very high value for γ leads to the checkerboard pattern in the reference picture usage, as previously described. On the other hand, a low value of γ reduces the importance of the choice of the reference frames in the Lagrangian cost. Therefore, the impact of the proposed method on the coding efficiency can be controlled through γ. Another way of controlling the impact of the proposed method is the use of the T_{ID}. This allows using lower weights in higher temporal layers, since such frames cannot be used as references for others with lower temporal ID, thus having less impact on error propagation. Therefore, there is a reduction of the impact of the proposed method when hierarchical coding configuration is adopted without compromising the error robustness.

As an example, Figure 4 illustrates the exponential weight used in the proposed method for different values of γ and $T_{ID} = 0$. The horizontal axis represents the sum of $\Delta_G(r_\phi)$ plus $\Delta_L(r_\phi)$. For instance, for $\gamma = 5$, when a reference frame r_ϕ is used to predict 3 more CTUs ($\Delta_G(r_\phi) + \Delta_L(r_\phi) = 3$) than the other reference frames (all PUs within those CTUs use the same reference), then the Lagrangian cost is increased by 50% ($e^{W(r_\phi)} \approx 1.5$).

As mentioned before, the underlying principle of (4) and (5) is that the cost of using a given reference frame increases with the number of PUs using it for prediction. However, the reference frames are not hard-selected without taking into account the R-D cost. In general, it is more difficult to reach a uniform use of reference frames in complex sequences than simpler ones. Note that standard compatibility is maintained, since the proposed optimisation method does not require any syntax change. Moreover, as the proposed approach is applied at the Lagrangian computation already present in the HEVC optimisation, the complexity of the encoder is not significantly affected.

C. Streaming stage: MV selection

Since HEVC uses differential and predictive MV coding, whenever a MV is lost the subsequent ones will also be affected, until a refresh point is reached to break the dependency chain. This increases the error propagation whenever the temporal MV candidate is used in the Merge Mode. To increase the number of possible refresh points and improve error robustness, a reduced number of MVs are selected to be transmitted as side information. The selection of these redundant MVs 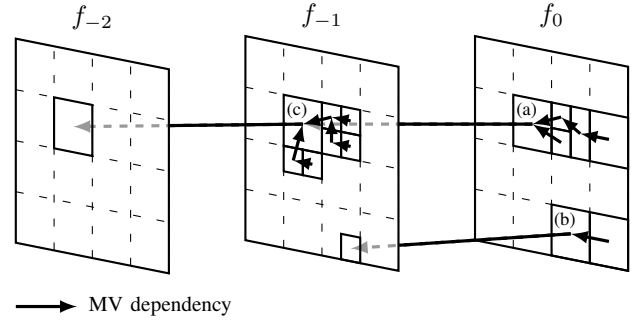is based on a trade-off that attempts to maximise the image area covered by such MVs using the lowest number of bits to encode them.

Figure 5 illustrates how MVs have different importance in error propagation. In this figure, each block corresponds to a PU, which has an associated MV. The arrows point to the PU containing the MV used as predictor. In the current frame f_0, two PUs have temporally dependent MVs. While the MV of PU (a) is used to predict a total of four other MVs, the MV of PU (b) is only used to predict one single MV. Thus, in case of data loss, the MV of PU (a) has more impact on error propagation and quality degradation than the MV of PU (b), because of its higher number of dependent MVs. For the this reason, in the MV selection process, the proposed method assigns higher importance to the MV of PU (a) than PU (b). Moreover in Figure 5, it can be observed that although the number of MVs predicted from the MVs of PU (c) is higher than PU (a) (6 and 4 MVs, respectively), the image area dependent from the former is smaller that the latter. This is due to the variable size of PUs. Therefore, the image area covered by all dependent PUs is separately considered in the MV selection process.

Based on the evidence mentioned above, the best MVs to be encoded redundantly as side information for each frame are selected through the following procedure:

1) the MVs with temporal dependencies are firstly selected as the most important set: $V = \{MV: MV \text{ is temporally dependent from others}\}$.
2) the $MV \in V$ are ranked according to the number of spatial dependencies associated to each one.
3) the optimal sub-set $U \subset V$ is found as the best trade-off between the image area covered by the dependent PUs and bitrate overhead, i.e., the one that maximises the cost function (7).

Considering V the ordered set of MVs encoded for the current frame, i.e., $V = \{MV_1, MV_2, \cdots MV_s\}$ and $U \subset V$ always starting in the first vector, i.e., $U = \{MV_1, MV_2, \cdots MV_n\}$ with $n \leq s$, the best set of redundant MVs (U^*) is selected based on the optimisation procedure given by (7). Note that any value of n corresponds to a unique subset U.

$$n^* = \arg\max_{n \leq s} \left\{ \sum_{i=1}^{n} A_D(i) - \alpha \frac{t_I - t_c}{T_{IDR}} R_U \right\}, \quad (7)$$

where R_U is the amount of bits required to encode the n MVs of U and $A_D(i)$ is the total area of the PUs whose MVs are dependently encoded from MV_i:

$$A_D(i) = \sum_{j=1}^{DV_i} (w_j \times h_j), \quad i = 1, 2 \cdots s, \tag{8}$$

where DV_i is the number of MVs dependent from MV_i and w_j and h_j are the width and height of the corresponding PU. The weighting factor used to control the amount of redundancy R_U is used to increase the cost of the overhead rate as the duration of error propagation decreases. It is given by the ratio between the distance from the current frame (t_c) to the next IDR (t_I) and the IDR period (T_{IDR}), combined with the α parameter. Therefore, for higher values of α, less redundant MVs are selected.

As pointed out before, this streaming stage may be implemented at any point of the network, to increase robustness without fully decoding the stream. Note that the parsing operation represented in Figure 2 does not decode the video frames, introducing only one frame delay because the MV selection is performed on a frame-by-frame basis.

To maintain compatibility with the HEVC standard, the redundant motion information is transmitted using the supplemental enhancement information (SEI) NAL units [5]. Thus, this information can be multiplexed into the coded bitstream without affecting the compressed video. The subset of selected MVs is independently encoded, ensuring that no reference information is needed in the decoder to properly decode the redundant MVs. In this work arithmetic coding [47] was used for the redundant MVs, but any other entropy coding method can be used.

V. Performance Evaluation

In this section the robustness of the proposed method is evaluated and discussed for two cases: (i) only using the encoding stage and (ii) using the streaming stage in addition to the encoding stage. In the former case, identified as "*Prop*", only the dynamic reference picture selection method described in Sub-section IV-B is evaluated. In the latter case, identified as "*Prop-MV*", besides the previous method of reference picture selection, redundant MVs are encoded as described in Sub-section IV-C, resulting in a combination of both.

As references for comparison, three other methods are used: (i) a fixed checkerboard approach (*Chkb*), (ii) the reference picture generation method proposed in [25], identified as "*Ref [25]*", and (iii) long-term reference frames (*Long*). The *Long* method uses only the key frames as reference, i.e., in the Random-Access configuration only one out of eight frames is used as reference frame, allowing a more robust transmission, as the reference frames are less likely to be hit by errors.

The experimental setup used in this section is the same as the one previously described in Section III. The evaluation of the encoding-stage is carried out separately from the streaming-stage.

TABLE III
BD-PSNR IN ERROR-FREE CONDITIONS.

Sequence	Low-Delay			Random-Access		
	Ref [25]	*Chkb*	*Prop*	*Long*	*Chkb*	*Prop*
Basketball Drill	-1.14	-1.15	-0.51	-0.85	-0.29	-0.22
Kendo	-2.10	-1.39	-0.99	-0.91	-0.39	-0.28
Park Scene	-1.79	-1.38	-0.81	-0.68	-0.38	-0.28
Traffic	-1.91	-1.21	-0.71	-0.55	-0.36	-0.21
Bosphorus	-0.97	-0.95	-1.10	-0.64	-0.20	-0.20
Average	**-1.58**	**-1.22**	**-0.82**	**-0.73**	**-0.32**	**-0.24**

A. Performance of the encoding stage

The performance of the encoding stage is evaluated on its own in order to consider only the improvements achieved by the dynamic reference frame selection method, when reference frames were reconstructed with distortion (e.g. as a result of non-perfect error concealment). Since the encoding stage does not deal with errors in temporal MV predictions (this is handled by the streaming-stage of the proposed method), in this evaluation the temporal MV predictions were not used. This is necessary to avoid masking the results with the errors caused by wrong MV predictions.

The performance of the proposed encoding stage (*Prop*) was evaluated using $\gamma = 5$. This allows to increase the Lagrangian cost by about 50%, when a reference frame is used for three or more CTUs in comparison to the usage of the other ones (see Figure 4). Further results for a wide range of γ values are also presented.

1) Performance under error-free conditions: Table III shows the Bjontegaard's Delta PSNR (BD-PSNR) values compared to the reference HEVC encoder for different coding configurations in error-free decoding. Four QPs were used from the common test conditions: 22, 27, 32, 37. The aim is to evaluate the R-D penalty incurred by the *Prop* method in comparison with *Chkb*, as well as the other reference methods. The use of long-term reference pictures (*Long*) is only tested for the case of Random-Access, where key frames are available, while method [25] is tested for the Low-Delay configuration. The results in Table III indicate a small loss of coding efficiency when the proposed reference picture selection approach is used under error-free conditions. As expected, this is due to the sub-optimal R-D optimisation of the proposed method, because the cost function also includes the cost associated to the choice of the reference frame, in addition to the rate and distortion.

While the fixed checkerboard approach leads to an average quality reduction of 1.22 dB for the Low-Delay (0.32 dB for Random-Access), the proposed method only loses 0.82 dB and 0.24 dB in the Low-Delay and Random-Access configurations, respectively. In comparison to *Long* and the method in [25], the proposed method is still more efficient, consistently achieving better results. Overall, the loss in the quality compared to the reference HEVC is considered acceptable, given the increase in robustness obtained in lossy transmission, as discussed in the following sections.

2) Quality evaluation under error-prone conditions: The robustness achieved by the proposed method was evaluated by comparison of the error propagation resulting from a single
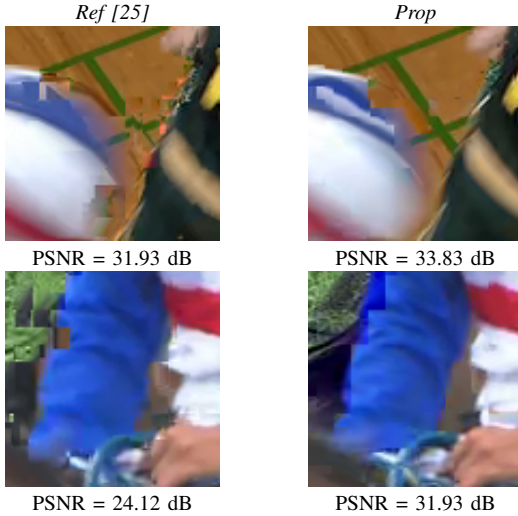
Fig. 6. Zoom of recovered Frame #10 after a loss event at Frame #9 for Basketball Drill and Race Horses sequence.
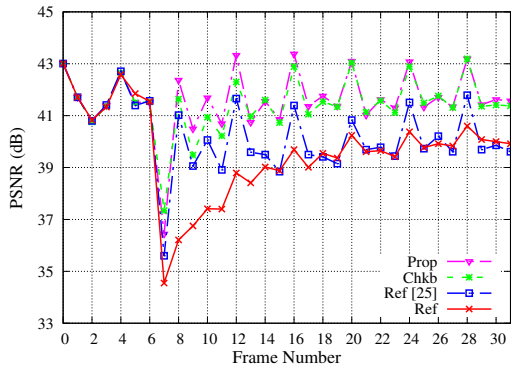

Fig. 7. Error propagation when errors affect Frame #7 in Kendo sequence.

loss event in the Low-Delay configuration.

Figure 6 shows the visual impact of error propagation on a region of the decoded Frame #10 when Frame #9 is affected by errors, for Basketball Drill and Race Horses sequences (the PSNR corresponds to the entire frame). As shown in this figure, the proposed method (*Prop*) is able to reduce the impact of mismatch decoding in comparison with the method in [25], revealing the higher efficiency achieving by distributing the use of the references frames.

Figure 7 shows the PSNR over a GOP with an IDR period of 32 frames, encoded at the same constant bitrate. This figure shows that all methods are able to outperform the reference HEVC (*Ref*), gaining approximately 3 dB in reconstruction quality. Since the proposed method is more efficient than the others for error-free transmission (see Table III), and it also provides better error robustness, it can be considered as an efficient solution to better deal with transmission errors. The similar quality levels achieved by *Chkb* and *Prop* reveal that dynamic selection of the reference frames is more robust because a better trade-off between coding efficiency and error resilience is attained. Overall, the method proposed for the encoding stage is able to increase the error robustness of HEVC without severely compromising the coding efficiency.

Further tests were run to evaluate the effectiveness of the proposed method under various packet loss rates. For each test condition, 50 trials were performed and the average quality obtained across all trials for both coding configurations are shown in Table IV and V, for the Low-Delay and Random-Access, respectively. In both tables the absolute PSNR value is shown for the reference HEVC case (*Ref*), while the PSNR difference is presented for the other cases. Results show that the proposed method is able to outperform other methods, by increasing the robustness of HEVC. An average gain up to 1.47 dB is obtained for the Low-Delay configuration (see Table IV), in comparison to the *Ref*, while the fixed checkerboard pattern approach (*Chkb*) only achieves up to 1.33 dB on average. In [25], the use of interpolated references in the encoding loop reveals to reduce error propagation, at a cost of high degradation of the coding efficiency in the error-free environment, thus the overall performance is lower (see results for "*Ref [25]*").

The proposed method is also able to outperform the HEVC standard when an hierarchical coding structure is used (see results for Random-Access configuration in Table V), which inherently limits the error propagation. In comparison to the reference HEVC, quality improvements up to 1.42 dB at PLR=10% are obtained (Kendo), while the average gain across all test sequences is 1 dB. When comparing results for both cases, it is noticeable that the proposed method is able to outperform the reference methods both when IDR frame loss is allowed or not (case 1 and 2, respectively). As the PLR increases the gains of the proposed method decreases when comparing with the case where the IDR frames delivered error-free (i.e., case 1), due to the higher error propagation of the missing IDR frame. Taking into account the above results and the spatio-temporal information shown in Table I, one can observe that the performance of the proposed method is lower for sequences with high spatial details and low temporal activity (see results for BQSquare and Park Scene), which indicates higher effectiveness for video sequences with higher motion activity.

3) Influence of parameter γ: The influence of the γ parameter used in Equation (5) was also studied. Figure 8 shows the average quality gains in comparison to the reference HEVC (*Ref*) across all test sequences, when using different values of γ in the proposed method (*Prop*). For the error-free case, the results in this figure show that increasing γ leads to a decrease in the coding efficiency, especially for the Low-Delay configuration. The reason is that higher values of γ lead to higher values of $W(r_\phi)$, therefore the reference picture selection is more constrained by the exponential weights. Since T_{ID} is constant for all frames in the Low-Delay, higher weights are used for all frames (T_{ID} is used to reduce the weights in case of hierarchical coding) and the encoder is not able to select the best reference frame in terms of R-D optimisation. In the case where packet loss occurs, higher values of γ lead to higher error resilience. Since higher weights used in the R-D cost, the encoder is constrained to distribute the use of the reference frames, reducing the error propagation. For very high values (e.g., $\gamma = 1000$) the encoder is forced to uniformly select the reference frames, resulting in a fixed

<div align="center">

TABLE IV
AVERAGE QUALITY (PSNR): LOW-DELAY.

</div>

Sequence	Method	Packet loss rate					
		IDR_{Loss}		IDR_{NoLoss}			
		1%	5%	1%	3%	5%	10%
Basketball Drill	*Ref*	32.25	24.43	36.52	33.70	31.70	28.51
	Ref [25]	-1.05	+0.50	-1.95	+0.12	+1.32	+2.90
	Chkb	-0.39	+0.60	-0.09	+0.69	+0.97	+1.34
	Prop	+0.08	+0.86	+0.45	+1.26	+1.54	+1.78
Book Arrival	*Ref*	34.47	26.95	38.53	35.82	33.63	30.67
	Ref [25]	-1.84	-0.72	-1.76	-0.64	-0.05	-0.04
	Chkb	-0.08	+0.81	+0.49	+1.33	+1.66	+2.15
	Prop	-0.37	+0.56	+0.35	+1.17	+1.41	+1.55
Cactus	*Ref*	34.37	31.94	37.04	35.67	34.51	32.65
	Ref [25]	-0.13	+0.20	+0.31	+0.46	+0.35	+0.83
	Chkb	-0.34	-0.12	+0.02	+0.37	+0.71	+1.00
	Prop	-0.23	+0.04	+0.06	+0.21	+0.43	+0.54
Bosphorus	*Ref*	31.79	24.00	42.19	41.31	40.56	38.74
	Ref [25]	-0.09	-1.04	-0.34	-0.15	+0.09	+0.67
	Chkb	+0.16	-0.49	-0.38	-0.13	+0.13	+0.72
	Prop	+0.23	-0.31	-0.47	-0.18	+0.07	+0.69
BQSquare	*Ref*	32.47	25.11	38.20	36.08	34.57	31.93
	Ref [25]	-4.42	-3.79	-4.42	-3.79	-3.73	-3.74
	Chkb	-0.38	+0.63	-0.12	+0.29	+0.44	+0.67
	Prop	+0.74	+1.37	-0.14	+0.66	+0.88	+1.35
Jockey	*Ref*	33.52	26.57	37.91	33.23	30.75	27.53
	Ref [25]	+1.29	-0.03	+2.10	+3.21	+2.84	+1.00
	Chkb	+2.16	+0.19	+1.47	+2.03	+2.20	+2.05
	Prop	+2.17	+0.23	+1.68	+2.41	+2.30	+2.16
Kendo	*Ref*	38.49	30.89	40.45	36.79	33.80	30.35
	Ref [25]	-1.82	+0.07	-1.82	+0.10	+1.34	+1.30
	Chkb	+0.09	+1.16	+0.24	+1.38	+2.27	+2.15
	Prop	+0.51	+1.33	+0.19	+1.34	+2.26	+2.26
Park Scene	*Ref*	29.81	25.29	36.06	34.83	33.64	31.63
	Ref [25]	-2.29	-1.63	-3.49	-3.40	-3.34	-3.45
	Chkb	-0.05	+0.10	-0.69	-0.20	+0.22	+0.71
	Prop	-0.08	-0.04	-0.50	-0.05	+0.29	+0.86
People on Street	*Ref*	26.07	21.36	32.42	29.61	27.85	25.00
	Ref [25]	+0.26	+0.82	-0.66	+0.42	+0.83	+1.26
	Chkb	+1.05	+0.43	+0.43	+0.95	+1.13	+1.38
	Prop	+0.94	+0.34	+0.58	+1.02	+1.17	+1.42
Race Horses	*Ref*	29.34	23.88	33.00	29.41	27.49	24.55
	Ref [25]	-0.10	+1.25	+0.60	+2.71	+3.31	+3.88
	Chkb	+0.69	+0.64	+0.46	+1.35	+1.47	+1.64
	Prop	+1.11	+0.84	+0.79	+1.80	+1.94	+2.11
Tennis	*Ref*	32.50	25.69	36.50	31.90	29.67	26.15
	Ref [25]	-1.04	+0.12	-1.74	-0.72	-0.08	+1.71
	Chkb	+0.48	+0.45	+0.69	+1.19	+0.81	+1.17
	Prop	+1.15	+1.17	+1.20	+2.03	+1.95	+2.12
Traffic	*Ref*	28.63	23.44	37.92	36.18	34.93	32.51
	Ref [25]	-0.13	+0.19	-0.73	-0.09	+0.12	+0.71
	Chkb	+0.16	+0.10	-0.43	+0.13	+0.51	+1.00
	Prop	-0.21	-0.08	-0.25	+0.28	+0.45	+0.76
Average	*Ref [25]*	**-0.95**	**-0.34**	**-1.16**	**-0.15**	**+0.25**	**+0.59**
	Chkb	**+0.30**	**+0.38**	**+0.17**	**+0.78**	**+1.04**	**+1.33**
	Prop	**+0.50**	**+0.53**	**+0.33**	**+1.00**	**+1.22**	**+1.47**

<div align="center">

TABLE V
AVERAGE QUALITY (PSNR): RANDOM-ACCESS.

</div>

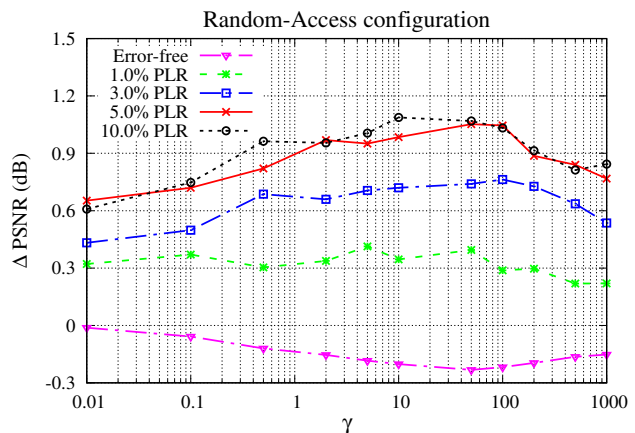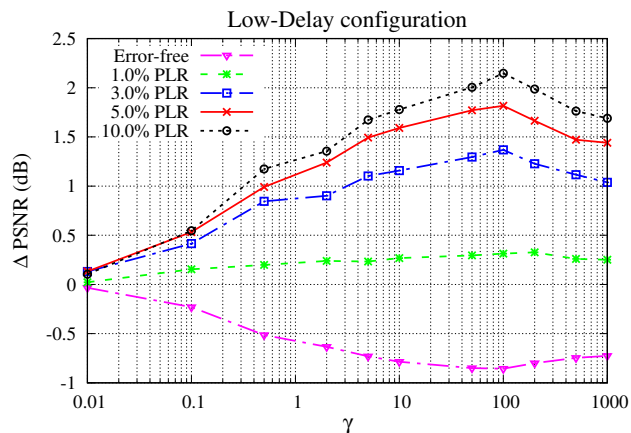| Sequence	Method	Packet loss rate					
		IDR_{Loss}		IDR_{NoLoss}			
		1%	5%	1%	3%	5%	10%
Basketball Drill	*Ref*	31.36	25.07	36.90	33.54	31.05	27.88
	Long	+0.09	-0.12	-0.02	+0.31	+0.57	+0.54
	Chkb	+0.12	+0.20	+0.15	+0.46	+0.60	+0.69
	Prop	+0.39	+0.37	+0.26	+0.57	+0.79	+0.87
Book Arrival	*Ref*	33.39	25.48	38.71	35.90	33.01	29.53
	Long	-0.14	+1.01	+0.18	+0.38	+0.96	+0.94
	Chkb	+0.30	+0.10	+0.07	+0.06	+0.31	+0.70
	Prop	+0.26	+1.16	+0.34	+0.28	+0.67	+0.97
Cactus	*Ref*	34.58	32.36	36.92	35.08	33.92	32.12
	Long	-0.32	-0.10	-0.34	-0.32	-0.30	-0.30
	Chkb	+0.15	+0.22	+0.21	+0.36	+0.42	+0.53
	Prop	+0.19	+0.13	+0.19	+0.34	+0.46	+0.50
Bosphorus	*Ref*	32.43	24.86	41.64	39.70	38.28	36.05
	Long	+0.09	-0.87	-0.60	-0.62	-0.61	-0.66
	Chkb	-0.22	+0.37	+0.19	+0.43	+0.55	+0.63
	Prop	+0.21	+0.49	+0.18	+0.47	+0.59	+0.69
BQSquare	*Ref*	33.85	23.63	37.85	33.45	30.44	27.10
	Long	-0.24	+0.46	-0.75	-0.06	-0.04	-0.41
	Chkb	+0.32	+0.62	+0.66	+0.65	+1.06	+1.05
	Prop	+0.42	+0.31	+0.50	+1.11	+0.93	+0.90
Jockey	*Ref*	33.86	26.52	36.80	31.82	29.27	26.21
	Long	+1.22	+0.45	+0.70	+0.50	+0.25	+0.05
	Chkb	+1.00	+0.79	+0.74	+1.08	+1.12	+0.96
	Prop	+1.08	+0.86	+0.85	+1.30	+1.16	+1.13
Kendo	*Ref*	36.81	28.11	39.78	34.69	30.73	26.62
	Long	+1.04	+0.97	+0.59	+1.43	+1.44	+0.80
	Chkb	+0.55	+1.09	+0.24	+0.62	+1.07	+0.98
	Prop	+0.84	+1.65	+0.37	+0.95	+1.53	+1.42
Park Scene	*Ref*	30.03	25.37	36.74	32.20	29.60	26.35
	Long	-0.15	-0.27	-0.56	-0.44	-0.39	-0.12
	Chkb	+0.24	+0.16	+0.17	+0.51	+0.56	+0.58
	Prop	+0.36	+0.31	+0.65	+0.95	+0.93	+1.04
People on Street	*Ref*	26.04	20.91	30.86	27.21	25.17	22.39
	Long	+0.26	+0.14	-0.22	-0.27	-0.30	-0.27
	Chkb	+0.56	+0.81	+0.56	+0.64	+0.68	+0.68
	Prop	+0.74	+0.83	+0.76	+0.87	+0.96	+0.93
Race Horses	*Ref*	30.43	24.11	33.22	29.14	26.84	23.93
	Long	-0.17	-0.20	+0.09	+0.50	+0.56	+0.34
	Chkb	+0.50	+0.43	+0.29	+0.69	+0.83	+0.89
	Prop	+0.50	+0.38	+0.54	+1.08	+1.10	+1.08
Tennis	*Ref*	31.89	25.89	36.74	32.20	29.60	26.35
	Long	+1.12	+0.58	+0.09	+1.55	+1.61	+1.25
	Chkb	+0.12	+0.17	+0.51	+0.56	+0.58	
	Prop	+1.46	+1.09	+0.65	+0.95	+0.93	+1.04
Traffic	*Ref*	28.60	23.47	37.29	34.52	32.54	29.60
	Long	-0.09	-0.31	-0.37	-0.32	-0.34	-0.24
	Chkb	-0.05	-0.07	+0.16	+0.54	+0.63	+0.96
	Prop	-0.09	+0.06	+0.41	+0.61	+0.76	+0.95
Average	*Long*	**+0.23**	**+0.15**	**-0.10**	**+0.22**	**+0.28**	**+0.16**
	Chkb	**+0.30**	**+0.40**	**+0.29**	**+0.53**	**+0.69**	**+0.78**
	Prop	**+0.53**	**+0.64**	**+0.43**	**+0.75**	**+0.87**	**+0.94**

Fig. 8. Influence of the γ parameter in the average quality gains, comparing with *Ref*, for different packet loss ratios.

TABLE VI
RELATIVE CPU TIMES OF TESTED METHODS.

Sequence	Method		
	Ref [25]	Chkb	Prop
Basketball Drill	+4.19%	+2.13%	+2.93%
Kendo	+4.96%	+1.83%	+2.05%
Park Scene	+4.98%	+3.90%	+4.04%
Traffic	+8.65%	+2.28%	+3.05%

checkerboard pattern. Overall, using the γ parameter in the proposed method, the impact of the reference frame, selected by the R-D optimisation, can be controlled, and thus a better trade-off between higher error resilience and loss of coding efficiency can be achieved.

4) Complexity overhead: The implementation of the proposed method has slightly increased the complexity of the standard non-robust encoder. Therefore, the computational complexity has also been considered as a performance metric in this research, which is measured as the encoding time in a controlled hardware platform. The reference used for comparison with other methods is the HEVC reference software encoding time, running on the same platform with the same coding configuration. The relative complexity increase of the robust encoder using *Prop*, *Chkb* and the method presented in [25] is shown in Table VI.

As shown in this table, the method presented in [25] is the most time consuming because it requires post-processing of the encoded frames in order to interpolate new references. The proposed scheme, using a reference picture selection, *Prop* presents a lower complexity increase than [25], while the fixed checkerboard scheme (*Chkb*) presents a slightly lower complexity than *Prop*.

B. Streaming stage

In this section the coding performance that results from using redundant MV (streaming stage) in combination with the reference picture selection scheme (encoding stage) is presented. In the experimental tests the temporal MV candidates are enabled and redundant MVs are added into the stream, as described in Sub-section IV-C. This is referred to as "*Prop-MV*", and compared to cases where a fixed amount of redundant MVs per frame are transmitted (i.e., the most important 10%, 30% and 50% $MV \in V$). Different α values are used for different redundancy ratios. All streams were encoded with the same total bitrate, including the redundant bits, in order to make a fair quality comparison.

1) Dynamic MV selection performance: In the case where variable redundancy is allowed, MVs are selected using the proposed dynamic method (Dyn) and a fixed approach (Fixed). Figure 9 illustrates the relation between the quality obtained and the amount of redundancy. The results indicate that for approximately the same amount of redundant bits, dynamic selection of MVs lead to higher performance in all test sequences. For instance, the proposed method is able to gain up to 1.5 dB for the BQSquare sequence. Moreover, the results are consistent for different PLR.
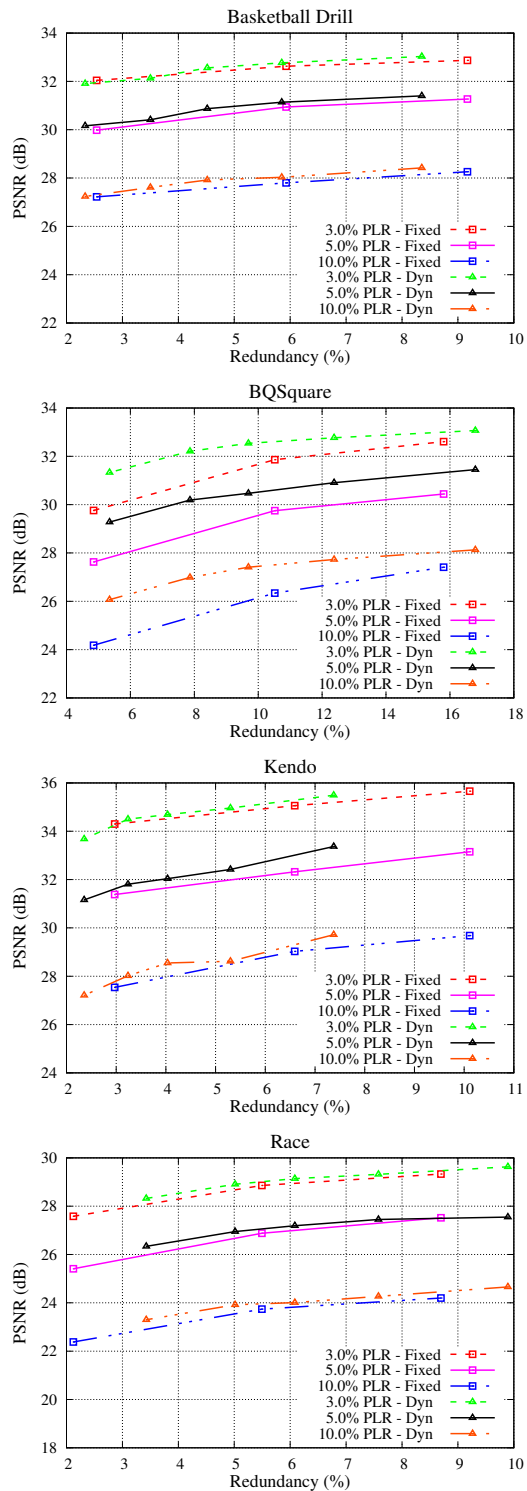

Fig. 9. Decoded video quality for different amounts of redundancy, selected using a fixed approach and the proposed method.

2) Quality evaluation: Table VII presents the average PSNR of decoded video for different PLR (columns 4 to 6) and for various percentages of MV redundancy (third column). It can be confirmed from the results that, for each sequence, as α decreases more redundant MVs are used, resulting in higher objective image quality for all tested sequences. This indicates

TABLE VII
AVERAGE QUALITY (PSNR) USING REDUNDANT MV SELECTED USING THE PROPOSED APPROACH

Sequence	Method	Redundancy ratio (%)	IDR_{Loss}		IDR_{NoLoss}			
			1%	5%	1%	3%	5%	10%
Basketball Drill	Without MVs	−	30.41	23.31	34.76	30.87	28.56	25.89
	Prop-MV (α=1.5)	2.33	+0.68	+0.41	+0.58	+1.04	+1.60	+1.35
	Prop-MV (α=0.8)	4.52	+0.96	+0.96	+0.95	+1.69	+2.31	+2.03
	Prop-MV (α=0.6)	5.85	+1.12	+1.16	+0.93	+1.90	+2.58	+2.14
Book Arrival	Without MVs	−	33.63	25.59	37.47	33.53	31.60	28.27
	Prop-MV (α=1.5)	1.81	+0.38	+0.14	+0.02	+0.49	+0.70	+0.61
	Prop-MV (α=0.8)	2.95	+0.23	+1.15	+0.49	+0.85	+0.76	+1.06
	Prop-MV (α=0.6)	3.92	+0.44	+1.80	+0.34	+1.22	+1.53	+1.60
Bosphorus	Without MVs	−	40.58	38.86	41.32	40.48	40.14	39.19
	Prop-MV (α=1.5)	1.46	+0.54	+1.21	+0.27	+0.63	+0.62	+0.75
	Prop-MV (α=0.8)	3.05	+0.73	+1.63	+0.31	+0.80	+0.85	+1.09
	Prop-MV (α=0.6)	3.94	+0.75	+1.70	+0.33	+0.81	+0.92	+1.17
BQSquare	Without MVs	−	28.66	20.22	31.00	25.40	23.09	20.21
	Prop-MV (α=1.5)	5.35	+3.04	+3.94	+4.05	+5.93	+6.19	+5.86
	Prop-MV (α=0.8)	9.69	+3.23	+3.80	+5.03	+7.14	+7.38	+7.20
	Prop-MV (α=0.6)	12.38	+3.54	+3.93	+5.13	+7.37	+7.82	+7.52
Kendo	Without MVs	−	36.21	28.38	38.23	32.85	29.58	25.52
	Prop-MV (α=1.5)	2.36	+0.78	+1.16	+0.36	+0.83	+1.58	+1.69
	Prop-MV (α=0.8)	4.03	+0.94	+1.32	+0.79	+1.84	+2.46	+3.03
	Prop-MV (α=0.6)	5.30	+0.86	+1.08	+0.97	+2.11	+2.84	+3.10
Park Scene	Without MVs	−	28.60	23.85	33.02	29.70	28.26	25.98
	Prop-MV (α=1.5)	3.88	+0.53	+0.72	+1.05	+2.03	+2.06	+2.09
	Prop-MV (α=0.8)	8.58	+0.79	+0.96	+1.55	+2.99	+3.19	+3.29
	Prop-MV (α=0.6)	11.22	+0.73	+1.05	+1.58	+3.26	+3.61	+3.73
Race Horses	Without MVs	−	26.87	21.78	29.67	25.19	23.23	20.48
	Prop-MV (α=1.5)	3.43	+2.46	+1.45	+2.28	+3.13	+3.11	+2.82
	Prop-MV (α=0.8)	6.08	+2.76	+2.23	+2.88	+3.95	+3.96	+3.53
	Prop-MV (α=0.6)	7.58	+2.93	+2.27	+2.89	+4.13	+4.22	+3.79
Tennis	Without MVs	−	31.27	24.81	34.33	29.35	27.09	23.96
	Prop-MV (α=1.5)	1.63	+0.92	+0.69	+0.87	+1.35	+1.42	+1.20
	Prop-MV (α=0.8)	3.12	+1.26	+0.76	+1.33	+1.95	+1.93	+1.73
	Prop-MV (α=0.6)	4.10	+0.97	+1.00	+1.45	+2.07	+2.17	+2.07
Traffic	Without MVs	−	27.38	22.22	33.92	29.86	27.92	24.86
	Prop-MV (α=1.5)	2.09	+0.30	+0.48	+0.95	+1.84	+1.61	+1.66
	Prop-MV (α=0.8)	5.67	+0.82	+0.95	+1.82	+3.15	+3.59	+3.67
	Prop-MV (α=0.6)	7.98	+0.96	+0.87	+2.07	+3.47	+3.89	+4.12

that redundant MVs are able to minimise the error propagation. This is accomplished with an acceptable increase in the bitrate, since only a sub-set of MVs is transmitted. The use of redundant MVs reveals better results for PLR as low as 1%, achieving a maximum gain of 3.93 dB for IDR_{Loss}(7.82 dB for IDR_{NoLoss}) for the BQSquare sequence at PLR=5%.

In summary, the proposed method is able to select the most relevant motion MV information and when combined with a reference picture selection scheme the problem of mismatch MV predictions is mitigated. Moreover, the dynamic approach devised to select the best MVs also contribute to increase the overall performance. Thus, using the proposed two-stage approach to recover temporal predictions consistently improves the error robustness of HEVC.

VI. CONCLUSION

In this paper a two-stage approach is proposed to increase the error robustness of HEVC streaming and reduce the error propagation in case of packet loss. A constrained coding approach was devised to select reference frames and dynamically distribute temporal dependencies. This is jointly used with a controlled amount of side information in coded streams, comprising a small set of the most relevant MVs, to minimise MV mismatch at the decoder in the presence of frame loss. As can be concluded from the results, this method contri Step 1: View and Respond to Decisibutes to reduce the temporal dependencies with consistent gains for different PLR and coding configurations, only incurring in a small drop of coding efficiency. Overall, the proposed approach is an effective method for coping with video transmission over error-prone networks.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] ——, "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.

[14] 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, Apr. 2007.

[15] 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.

[16] "ITU-T Rec. H.222.0 and ISO/IEC 13818-1:2015 information technology - generic coding of moving pictures and associated audio information: Systems," Jul. 2015.

[17] "ISO/IEC 23008-1:2014 Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 1: MPEG media transport (MMT)," Jun. 2014.

[18] B. Yan and K. W. Ng, "An efficient error detection technique for MPEG-4 video streams," in *First IEEE Consumer Communications and Networking Conference*, Jan. 2004, pp. 442–446.

[19] G. L. Wu and S. Y. Chien, "Spatial-temporal error detection scheme for video transmission over noisy channels," in *Ninth IEEE International Symposium on Multimedia (ISM)*, Dec. 2007, pp. 78–85.

[20] K. L. Hung and C. H. Tsai, "Image error detection and error concealment technique based on interleaving prediction and direction information hiding," in *First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)*, Sep. 2010, pp. 371–376.

[21] 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.

[22] Y.-L. Chan, H.-K. Cheung, and W.-C. Siu, "Compressed-domain techniques for error-resilient video transcoding using RPS," *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 357–370, Feb. 2009.

[23] H. Hadizadeh and I. Bajic, "Rate-distortion optimized pixel-based motion vector concatenation for reference picture selection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 8, pp. 1139–1151, Aug. 2011.

[24] Y. J. Liang, E. Setton, and B. Girod, "Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection," in *IEEE Workshop on Multimedia Signal Processing*, Dec. 2002, pp. 420–423.

[25] 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.

[26] Y. Zhang, S. Qin, B. Li, and Z. He, "Rate-distortion optimized unequal loss protection for video transmission over packet erasure channels," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1390 – 1404, 2013.

[27] A. Leontaris and P. Cosman, "Video compression for lossy packet networks with mode switching and a dual-frame buffer," *Image Processing, IEEE Transactions on*, vol. 13, no. 7, pp. 885–897, Jul. 2004.

[28] V. Chellappa, P. Cosman, and G. Voelker, "Error concealment for dual frame video coding with uneven quality," in *Proceedings of Data Compression Conference*, Mar. 2005, pp. 319–328.

[29] H. Kim, P. Cosman, and L. Milstein, "Motion-compensated scalable video transmission over MIMO wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 116–127, Jan. 2013.

[30] K. Tan and A. Pearmain, "An improved FMO slice grouping method for error resilience in H.264/AVC," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 1442–1445.

[31] W.-J. Tsai and H.-Y. You, "Multiple description video coding based on hierarchical b pictures using unequal redundancy," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 2, pp. 309–320, Feb. 2012.

[32] 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.

[33] 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.

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

[35] M. Dissanayake, C. T. E. R. Hewage, S. Worrall, W. A. C. Fernando, and A. Kondoz, "Redundant motion vectors for improved error resilience in H.264/AVC coded video," in *IEEE International Conference on Multimedia and Expo*, Jun. 2008, pp. 25–28.

[36] 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.

[37] J. Carreira, E. Ekmekcioglu, A. Kondoz, 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.

[38] 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.

[39] 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.

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

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

[44] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[45] G. Kokkonis, K. E. Psannis, M. Roumeliotis, and Y. Ishibashi, "Efficient algorithm for transferring a real-time HEVC stream with haptic data through the internet," *Journal of Real-Time Image Processing*, vol. 12, no. 2, pp. 343–355, Aug. 2016.

[46] 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.

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

João F. M. Carreira (M'11) was born in Portugal in 1989. He received the Engineering and M.Sc. degrees in Electrical Engineering from the Escola Superior de Tecnologia e Gestão of Instituto Politécnico de Leiria, Portugal, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree at Institute for Digital Technologies of Loughborough University London, United Kingdom, in collaboration with the Instituto de Telecomunicações, Portugal.

He has been a research fellow with the Instituto de Telecomunicações, Portugal, since 2008. His research interests include 2D/3D image and video processing, robust video streaming over constrained networks and subjective quality evaluation.

Pedro A. Assunção (M'98, SM'14) obtained the degree of Licenciado and Master of Science in Electrical Engineering from the University of Coimbra in 1988 and 1993, respectively. In 1998 he received the PhD in Electronic Systems Engineering from the University of Essex, United Kingdom. He is currently coordinator professor of electronic and telecommunications at Politécnico de Leiria and he is also Senior Researcher at Instituto de Telecomunicações, Portugal in the field of Networks and Multimedia. Since 1999, he was vice-president of the School of Technology, Head of the Department of Electrical and Computer Engineering, President of the Pedagogical Board and also President of the Scientific Board. He has been teaching Electronics, Analogue and Digital Communications, and Multimedia Communications, in undergraduate and MSc courses of electrical engineering. He has been active as researcher and supervisor of MSc and PhD students, and he is co-author/author of more than one hundred publications in international conferences, journals, book chapters, books and four US patents. He has been a reviewer for several scientific conferences and journals published by the IEEE, Elsevier, Springer. He was the Chair of COST Action 3D-ConTourNet (IC1105) and General Chair of 3DTV-Con 2015 and he is Senior Member member of the Institute of Electrical and Electronics Engineering (IEEE). His current research interests include high efficiency coding and processing of 3D video, Light Field and UHD panoramic video, Quality of Experience (QoE) in multimedia communications systems, complexity of High Efficiency Video Codecs, visual attention modelling and applications.

Sérgio M. M. de Faria (M'93, SM'08) was born in Portugal in 1965. He received the Engineering degree in Electrical Engineering from Universidade de Coimbra, Portugal, in 1988, the M.Sc. degree in Electrical Engineering from Universidade de Coimbra, Portugal, in 1992, and the Ph.D. degree in Electronics and Telecommunications from the University of Essex, England, in 1996. In 2014 received the title of Agregado by Instituto Superior Técnico, University of Lisbon.

He is Professor with the Department of Electrical Engineering, in Escola Superior de Tecnologia e Gestão of Instituto Politécnico de Leiria, Portugal, since 1990. He has collaborated in master courses with Faculty of Science and Technology and with Faculty of Economy of Universidade de Coimbra, Portugal. He is an Auditor with A3ES organization for Electrical and Electronic Engineering courses in Portugal. He is a Senior Researcher with Instituto de Telecomunicações.

His research interests include 2D/3D image and video processing and coding, motion representation, and medical imaging. In this field, he has published 1 book, edited 2 books and authored 7 book chapters, 23 journal papers, and 117 referred conference papers. He has been participating and he is responsible for several, national and international (EU), funded projects.

He is an Area Editor of Signal Processing: Image Communication. He has been a Scientific and Program Committee Member of many international conferences. He is a reviewer for several international scientific journals and conferences (IEEE, IET and EURASIP). He is a Senior Member of the IEEE.

Erhan Ekmekcioglu received his PhD degree from the University of Surrey, UK, in 2010, and worked there as a post-doc researcher until 2014. Since 2014 he has been working as a research associate in the Institute for Digital Technologies at Loughborough University London, a post graduate teaching, research and enterprise institute. His research interests include video processing and compression, machine learning applications on image and video, video streaming over constrained networks, mobile video communication, Quality of Experience, immersive and 3D media systems, augmented reality, and virtual reality. He has participated in a number of European Union funded collaborative research projects and conducted research mainly on multi-view video acquisition, compression, distribution, and Quality of Experience. He is the author/co-author of several research articles, book chapters, a book on 3D-TV systems, and the guest editor of several special issues published by the IEEE Multimedia Communications Technical Committee. He has been a regular reviewer of prestigious IEEE Transactions including Circuits and Systems for Video Technology, Multimedia, and Image Processing. Since January 2017 he has been co-chairing the 3D rendering, processing and communications Interest Group (3DIG) of the IEEE Multimedia Communications Technical Committee.

Ahmet Kondoz (M'91, SM'11) received his PhD degree in 1987 from University of Surrey, UK. From 1986 to 1988, he was employed as a research fellow in the communication systems research group. He became a lecturer in 1988, reader in 1995, and in 1996 he was promoted to professor in multimedia communication systems. He was the founding head of I-LAB, a multi-disciplinary multimedia communication systems research lab at the University of Surrey. Since January 2014, he has been appointed as the founding Director of the Institute for Digital Technologies, at Loughborough University London, a post graduate teaching, research and enterprise institute. He is also serving as the Associate Dean for Research at Loughborough University London.

His research interests include digital signal processing and coding, fixed and mobile multimedia communication systems, 3D immersive media applications for the future Internet systems, smart systems such as autonomous vehicles and assistive technologies, big data analytics and visualisation and related cyber security systems. He has over 400 publications, including six books, several book chapters, and seven patents, and graduated more than 75 PhD students. He has been a consultant for major wireless media industries and has been acting as an advisor for various international governmental departments, research councils and patent attorneys.

Dr. Kondoz has been involved with several European Commission FP6 & FP7 research and development projects, such as NEWCOM, e-SENSE, SUIT, VISNET, MUSCADE, etc. involving leading universities, research institutes and industrial organisations across Europe. He coordinated FP6 VISNET II NoE, FP7 DIOMEDES STREP and ROMEO IP projects, involving many leading organisations across Europe which deals with the hybrid delivery of high quality 3D immersive media to remote collaborating users including those with mobile terminals. He co-chaired the European networked media advisory task force, and contributed to the Future Media and 3D Internet activities to support the European Commission in the FP7 programmes.