

Error Resilience and Concealment Techniques for High Efficiency Video Coding

João Filipe Monteiro Carreira

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University in London

July 2018

© 2018 by João Carreira

Certificate of Originality
Thesis Access Conditions and Deposit Agreement

Students should consult the guidance notes on the electronic thesis deposit and the access conditions in the University's Code of Practice on Research Degree Programmes

Author……...……...…….

Title……..

I [Please insert name and address], "the Depositor",
would like to deposit [Please insert title of dissertation], hereafter referred to as the "Work", once it has successfully been examined in Loughborough University Institutional Repository

Status of access OPEN / RESTRICTED / CONFIDENTIAL

Moratorium Period……years, ending……../……20…….

Status of access approved by (CAPITALS):……

Supervisor (Signature)……...……...

School of……...……

Author's Declaration *I confirm the following :*

CERTIFICATE OF ORIGINALITY
This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work therein has been submitted to this or any other institution for a degree

NON-EXCLUSIVE RIGHTS
The licence rights granted to Loughborough University Institutional Repository through this agreement are entirely non-exclusive and royalty free. I am free to publish the Work in its present version or future versions elsewhere. I agree that Loughborough University Institutional Repository administrators or any third party with whom Loughborough University Institutional Repository has an agreement to do so may, without changing content, convert the Work to any medium or format for the purpose of future preservation and accessibility.

DEPOSIT IN LOUGHBOROUGH UNIVERSITY INSTITUTIONAL REPOSITORY
I understand that open access work deposited in Loughborough University Institutional Repository will be accessible to a wide variety of people and institutions - including automated agents - via the World Wide Web. An electronic copy of my thesis may also be included in the British Library Electronic Theses On-line System (EThOS). I understand that once the Work is deposited, a citation to the Work will always remain visible. Removal of the Work can be made after discussion with Loughborough University Institutional Repository, who shall make best efforts to ensure removal of the Work from any third party with whom Loughborough University Institutional Repository has an agreement. Restricted or Confidential access material will not be available on the World Wide Web until the moratorium period has expired.

- That I am the author of the Work and have the authority to make this agreement and to hereby give Loughborough University Institutional Repository administrators the right to make available the Work in the way described above.

- That I have exercised reasonable care to ensure that the Work is original, and does not to the best of my knowledge break any UK law or infringe any third party's copyright or other Intellectual Property Right. I have read the University's guidance on third party copyright material in theses.
- The administrators of Loughborough University Institutional Repository do not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of breach of Intellectual Property Rights, or any other right, in the material deposited.

*The statement below shall apply to **ALL** copies:*

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Restricted/confidential work: All access and any copying shall be strictly subject to written permission from the University Dean of School and any external sponsor, if any.

Author's signature……..Date……...……...……

user's declaration: for signature during any Moratorium period (Not Open work): ***I undertake to uphold the above conditions:***			
Date	Name (CAPITALS)	Signature	Address

Anyone who has never
made a mistake has never
tried anything new.

(Albert Einstein)

Acknowledgments

I would like to thank to everyone that help during this research work, and that make it possible to accomplish.

I would like to express my gratitude to my advisers Prof. Pedro Antnio Amado Assuno, Prof. Srgio Manuel Maciel de Faria, Doctor Erhan Ekmekcioglu and Prof. Ahmet Kondoz for their support, guidance and availability that were essential for the proper conduct of this research work.

I would like to thank Fundação para a Ciência e a Tecnologia (FCT), which provided the financial support for this PhD, under the PhD grant SFRH/BD/86368/2012. I am also grateful to the Instituto de Telecomunicacoes and to the Institute for Digital Technologies, which provided the computer resources and laboratory facilities with excellent conditions for my research work.

I would like to thank to all communities around the world that contribute for the development of many free and open-source tools used in this research work, namely the GNU/Linux, KDE software, Ubuntu and Gentoo communities. I also acknowledge the partnership of Luis Lucas in the Calyp software, and his contribution to the Remoterun scripts and the general computer management, which improved our developer capabilities and ease the experimental evaluations.

Finally, a special thanks to my parents, Antnio and Maria, and my sister Marta to whom I owe all that I have become. I also would like to express a special gratitude to Bianca Pires for her support and love during this journey.

Abstract

The increasing diversity of services and the demand for ever higher quality multimedia at wider resolutions, which leads to the growing popularity of Ultra-HD (UHD) format, are pushing technologies forward and opening new research directions. The recent developments in video coding techniques have resulted in a novel standard called High Efficiency Video Coding (HEVC) that is the next breakthrough in multimedia compression after the globally adopted MPEG-4/H.264/AVC standard. However, the significantly improved levels of compression efficiency achieved by the HEVC standard lead to an inherent reduction in the error resilience performance under error prone conditions. It should be aimed at ensuring an adequate level of protection against network and transmission induced impairments (e.g., packet losses), without sacrificing the compression efficiency thanks to the novel block level operations done by HEVC.

My research work takes the HEVC standard to its core and builds new techniques on top of it to enhance its error resilience performance, and develop new processing techniques to improve decoder side error concealment performance. The first scope of this research work is to make an extensive analysis of various error resilience techniques developed for the HEVC standard, but also for the previous video coding standards (most notably H.264/AVC); identify their shortcomings and address those shortcomings. Up to date work has revealed that the Motion Vector (MV) coding technique lacks robustness, which leads to quality degradation due to error propagation in lossy transmission environments. It has also been seen that the temporal dependency of motion information is comparatively higher than that in the H.264/AVC standard. Based on this evidence, a method was developed to make intelligent use of temporal MV candidates during the motion estimation process, in order to decrease the temporal dependency, and improve the error resiliency without penalising the rate-distortion performance. The simulation results have shown that the proposed method improves the error resilience under tested conditions, compared to the reference method that always enables temporal MVs candidates. Another proposed method has exploited the prediction dependency of MVsin order to select the most relevant ones for redundant coding with reduced overhead. Simulation results have shown that the proposed

MV selection method can effectively identify the most relevant motion fields, achieving improved error robustness with a significantly reduced redundancy overhead.

The second scope of this research work is to develop new processing techniques for the concealment of the coding elements lost during the transmission and delivery system, using combined information between the encoder and decoder sides. Error concealment techniques mitigate the detrimental effects of network impairments on the perceptual multimedia quality and the performance these techniques determine the quality of a system directly. The third and the ultimate aim of this research work is to strike a balance between the designed error resilience and error concealment techniques so as to meet a set of QoS/QoE requirements associated with current and future multimedia-based services (e.g., real-time multi-party video conferencing, Ultra-HD video on demand, etc.), such as bandwidth and delay. It is envisaged to produce a set of configuration profiles for chosen services based around the developed tools for HEVC.

Contents

Acknowledgments **v**

Abstract **vii**

Contents **xii**

List of Figures **xiii**

List of Tables **xvii**

List of Abbreviations **xix**

1 Introduction **1**

1.1 Context and motivation 1

1.2 Main objectives and contributions 3

1.3 Outline of this thesis 3

2 The High Efficiency Video Coding Standard **5**

2.1 Introduction 5

2.2 General hybrid video coding 6

2.3 Overview the standard features 7

2.4 High-level syntax 8

2.4.1 Network Abstraction Layer 9

2.4.2 Parameters set 10

2.4.3 Picture types 10

2.4.4 Picture partitioning 13

2.5 Coding structures 16

2.6 Prediction methods 18

2.6.1 Intra-frame 18

2.6.2 Inter-frame 21

2.7 Transform and quantisation 25

2.8 Entropy coding . . . 26
2.9 In-loop filters . . . 26
2.10 Summary . . . 27

3 Robust Video Coding: A Review **29**
3.1 Introduction . . . 29
3.2 Categories of robust video coding techniques . . . 30
3.3 Localisation techniques . . . 31
3.3.1 Error resilient rate-distortion optimisation . . . 31
3.3.2 Intra-refreshing . . . 33
3.3.3 Leaky-predictions . . . 35
3.3.4 Reference picture selection . . . 36
3.4 Data partition and redundant coding . . . 38
3.4.1 Flexible Macroblock Ordering . . . 38
3.4.2 Unequal error protection . . . 40
3.4.3 Redundant pictures . . . 42
3.4.4 Multiple description coding . . . 43
3.5 Error concealment by post-processing . . . 44
3.5.1 Spatial error concealment techniques . . . 44
3.5.2 Temporal error concealment techniques . . . 46
3.6 Concealment-aware error resilience . . . 51
3.7 Summary . . . 52

4 A Method to Increase the Error Robustness of HEVC **55**
4.1 Introduction . . . 55
4.2 Evaluation study of the error robustness of HEVC . . . 56
4.2.1 Motivation and objectives . . . 56
4.2.2 Characterisation of the experiments . . . 57
4.2.3 Evaluation of the impact of the partition units size . . . 59
4.2.4 Evaluation of the impact of intra refreshing . . . 60
4.2.5 Evaluation for using different slice partitioning schemes . . . 62
4.2.6 Motion vector coding impact on error propagation . . . 63
4.2.7 Discussion and conclusions . . . 66
4.3 Dynamic motion vector refreshing . . . 66
4.3.1 Proposed method . . . 67
4.3.2 Performance evaluation . . . 68
4.4 Summary . . . 71

5 A Two-stage Approach for Robust Coding and Streaming 73
5.1 Introduction . . . 73
5.2 Problem formulation . . . 74
5.3 Proposed architecture . . . 75
5.4 Encoding stage: Reference frame selection . . . 76
5.5 Streaming stage: MV selection . . . 81
5.6 Performance evaluation . . . 83
5.6.1 Encoding stage . . . 83
5.6.2 Streaming stage . . . 90
5.7 Summary . . . 92

6 Error Concealment-Aware Coding for Robust Video Transmission 93
6.1 Introduction . . . 93
6.2 Problem formulation . . . 94
6.3 Traditional error concealment algorithms . . . 95
6.3.1 Motion-compensation algorithms . . . 95
6.3.2 Algorithms based on motion estimation . . . 96
6.3.3 Experimental evaluation . . . 99
6.3.4 Discussion and conclusions . . . 100
6.4 Proposed architecture . . . 100
6.5 Error concealment mode decision . . . 102
6.5.1 Error concealment candidates . . . 102
6.5.2 Selection optimization . . . 103
6.5.3 Saliency detection . . . 103
6.5.4 Signalling the optimal EC modes . . . 104
6.5.5 Enhancement-layer concealment residue . . . 105
6.6 Error concealment signalling evaluation . . . 106
6.6.1 Intra-frame EC optimisation . . . 106
6.6.2 Robust video transmission with optimal EC . . . 108
6.7 Overall performance evaluation . . . 115
6.7.1 Visual quality evaluation . . . 117
6.7.2 Comparison with previous works . . . 118
6.8 Summary . . . 119

7 Joint error resilience-concealment for robust video streaming 121
7.1 Introduction . . . 121
7.2 Problem formulation . . . 122
7.3 Performance evaluation under different network condition . . . 122
7.3.1 Test conditions . . . 122

7.3.2 Experimental results . . . 122
7.3.3 Discussion and conclusions . . . 122
7.4 Proposed architecture . . . 122
7.5 Error robustness optimisation . . . 122
7.6 Optimisation method . . . 122
7.6.1 Performance evaluation . . . 122
7.7 Error concealment optimisation based on MV differences . . . 122
7.7.1 Experimental results . . . 122
7.8 Summary . . . 122

8 Conclusion and future work **123**
8.1 Conclusions . . . 123
8.2 Future work . . . 123

A Test video sequences **125**

B Published papers **131**
B.1 Journal publications . . . 131
B.2 Conference publications . . . 131

Bibliography **133**

List of Figures

2.1 Block diagram of a generic video coding system, including the inner blocks of a generic hybrid encoding and decoding scheme. 7
2.2 Generic HEVC standard architecture. 8
2.3 Two bytes Network Abstraction Layer (NAL) unit header in HEVC. . . 9
2.4 Leading and trailing pictures associated with a random-access point. . . 12
2.5 Temporal sub-layer access pictures example. 13
2.6 Examples of the interaction between tiles and slices. 15
2.7 Illustration of the wavefront parallel processing. 16
2.8 Representation of a CTU (64×64) and its partitioning into CUs and PUs (left); corresponding quad-tree with coding order (right) 17
2.9 Partitioning modes of an inter coded Coding Unit (CU) into Prediction Units (PUs) [5]. 18
2.10 Example of HEVC intra prediction modes for a 8×8 block [28]. 19
2.11 Illustration of the MV prediction candidates in the AMVP. 24

3.1 Examples of spatial and temporal error propagation and how it can be reduce using localisation techniques [44]. 32
3.2 Examples of spatial and temporal error propagation and how it can be reduce using localisation techniques [60]. 34
3.3 Example of error propagation with a Region of Interest (ROI) (shaded area indicates error, and the arrows indicate temporal predictions to previous frame) [73]. 36
3.4 Examples of reference picture selection in error resilience video transcoding [78]. 37
3.5 Different types of flexible macroblock reordering [84]. 39
3.6 Architecture of multiple description system [83]. 43
3.7 Lost pixel interpolation from neighbouring error-free pixels. 45
3.8 Illustration of the boundary match algorithm. 47

3.9 Motion Vector Extrapolation (MVE) extrapolation with the corresponding extrapolated macroblock, based on the method presented in [129]. . 48
3.10 Pixel based motion vector extrapolation [131]. 50

4.1 Illustration of the block partition and prediction mode for a region of the Basketball Drill sequence. 59
4.2 Decoded video quality for different intra periods, i.e., distance between intra-frames. 61
4.3 Decoded video quality for different slice partitioning schemes. 62
4.4 Average quality using different number of MV candidates. 63
4.5 Error propagation for Kendo sequence when Frame #6 is lost for the Low-Delay configuration. 64
4.6 Proposed approach to use the temporal motion vector candidate in the HEVC encoder. 67
4.7 Quality obtained with the different approaches tested in the HEVC when the frame #5 is missing for Kendo and BQSquare sequences. 69

5.1 Average usage of the reference frame 1 (r_1) for different sequences. . . . 74
5.2 Functional blocks of the proposed method (*gray* blocks) in the context of the HEVC coding and streaming. 76
5.3 Example of the reference picture selection applied to the frame C based on the checkerboard at Coding Tree Unit (CTU) level (PUs represented by small partitions). 78
5.4 Exponential weight as a function of the global and local deviations . . . 79
5.5 Example of the partitions and references pictures selected using the different encoding methods (the dark gray squares indicate that the MV is using the reference r_1, light gray squares corresponds to the reference r_2 and unfilled regions are intra-coded blocks). 80
5.6 Example of temporal and spatial MV dependencies in HEVC. 81
5.7 Zoom of recovered Frame #10 after a loss event at Frame #9 for Basketball Drill and Race Horses sequence. 85
5.8 Error propagation when errors affect Frame #7 in Kendo sequence with an Instantaneous Decoding Refresh (IDR) period of 32 frames. 85
5.9 Influence of the γ parameter in the average quality gains, comparing with *Ref*, for different packet loss ratios. 88
5.10 Decoded video quality for different amounts of redundancy, selected using a fixed approach and the proposed method. 90

6.1 Representation of the video frame with associated original and extrapolated motion vectors represented at full and dashed arrows, respectively. 96
6.2 Block diagram of the deep matching algorithm [153]. 99
6.3 Architecture of the error concealment-aware encoding scheme. 101
6.4 Error Concealment (EC) candidates used for optimization. 102
6.5 Payload structure of the Supplemental Enhancement Information (SEI) NALs 104
6.6 Error propagation using Low-Delay for the Basketball Drill sequence. . 107
6.7 Error propagation using Random-Access for the Park Scene sequence. . 107
6.8 Error propagation after losing the inter-coded frame #6. 114
6.9 Error propagation after losing intra-coded frame #16. 114
6.10 Recovered frame #6 of Basketball Drill 115
6.11 Crop of the decoded video frame for the Race Horses and Kimono sequences (Loss event at Frame #6) 117

A.1 Basketball Drill, 832 × 480, 50 fps, 240 frames. 125
A.2 Book Arrival, 1024 × 768, 30 fps, 240 frames. 125
A.3 BQSquare, 416 × 240, 60 fps, 240 frames. 126
A.4 Bosphorus, 3840 × 2160, 120 fps, 240 frames. 126
A.5 Cactus, 1920 × 1080, 50 fps, 500 frames. 126
A.6 Four People, 1280 × 720, 60 fps, 240 frames. 126
A.7 Jockey, 3840 × 2160, 120 fps, 240 frames. 127
A.8 Kendo, 1024 × 768, 30 fps, 240 frames. 127
A.9 Kimono, 1920 × 1080, 24 fps, 240 frames. 127
A.10 Park Scene, 1920 × 1080, 24 fps, 240 frames. 128
A.11 People On Street, 2560 × 1600, 24 fps, 150 frames. 128
A.12 Race Horses, 832 × 480, 30 fps, 240 frames. 128
A.13 Traffic, 2560 × 1600, 30 fps, 150 frames. 129

List of Tables

2.1 Picture categories in HEVC standard. 11
2.2 Reference Picture Set example. 23

3.1 Advantages of error robustness techniques by category. 30

4.1 Test sequences used in the experiments. 58
4.2 Average Peak Signal-to-Noise Ratio (PSNR) for different maximum CU size. 60
4.3 Relation between the number of available MV candidates and its usage 64
4.4 PSNR under random packet loss. 65
4.5 Average standard deviation of the decoded video quality (PSNR). . . . 70
4.6 Bjontegaard's average PSNR (dB) differences for all sequences tested at different frame loss conditions. 70

5.1 BD-PSNR in error-free conditions. 84
5.2 Average quality (PSNR): Low-Delay. 86
5.3 Average quality (PSNR): Random-Access. 87
5.4 Relative CPU times of tested methods. 89
5.5 Average quality using redundant MV selected by the proposed approach. 91

6.1 Average quality (PSNR) using different error concealment algorithms . 100
6.2 Percentage of overhead (%) 106
6.3 Average PSNR and variance of affected frames 108
6.4 Average Level and usage ratio of each EC mode. 109
6.5 Average quality (PSNR and WPSNR) of the affected frames for Low-Delay configuration. 110
6.6 Average Feature Similarity Index (FSIM) of the entire video sequence in frame loss conditions (scaled between $[0 - 100]$). 111
6.7 Error concealment signalling overhead. 112

6.8 Average quality under 3% of PLR for different methods and overhead ratios. 115
6.9 Average quality under 10% of PLR. 116
6.10 Comparison against existing methods. 119

List of Abbreviations

AMVP	Advanced Motion Vector Prediction.
BLA	Broken Link Access.
BMA	Boundary Match Algorithms.
CABAC	Context Adaptive Binary Arithmetic Coding.
CB	Coding Block.
CRA	Clean Random Access.
CTB	Coding Tree Block.
CTU	Coding Tree Unit.
CU	Coding Unit.
DCT	Discrete Cosine Transform.
DPB	Decoded Picture Buffer.
DST	Discrete Sine Transform.
ECU	Error Concealment Units.
EC	Error Concealment.
FC	Frame-Copy.
FEC	Forward Error Correction.
FMO	Flexible Macroblock Ordering.
FSIM	Feature Similarity Index.
GOP	Group of Pictures.
H.264/AVC	H.264 - Advanced Video Coding.
HEVC	High Efficiency Video Coding.
HMVE	Hybrid Motion Vector Extrapolation.
IDR	Instantaneous Decoding Refresh.
IEC	International Electrotechnical Commission.
IP	Internet Protocol.
ISO	International Organization for Standardization.
LTE	Long Term Evolution.

MB	Macroblock.
MC	Motion-Copy.
MDC	Multiple Description Coding.
MMCO	Memory Management Control Operation.
MMT	MPEG Media Transport.
MPEG	Moving Picture Experts Group.
MSE	Mean Square Error.
MTU	Maximum Transfer Unit.
MVE	Motion Vector Extrapolation.
MV	Motion Vector.
NAL	Network Abstraction Layer.
PLR	Packet Loss Ratio.
PMVE	Pixel-based Motion Vector Extrapolation.
POC	Picture Order Count.
PPS	Picture Paramenter Set.
PSNR	Peak Signal-to-Noise Ratio.
PU	Prediction Block.
PU	Prediction Unit.
QP	Quantized Parameter.
QoE	Quality of Experience.
R-D	Rate-Distortion.
RADL	Random access decodable leading picture.
RAP	Random-Access Point.
RASL	Random access skipped leading picture.
ROI	Region of Interest.
ROPE	Recursive Optimal Per-Pixel Estimation.
RPSet	Reference Picture Set.
RPS	Reference Picture Selection.
RTP	Real-time Transport Protocol.
SAO	Sample Adaptive Offset.
SEI	Supplemental Enhancement Information.
SI	Spatial Information.
SPS	Sequence Paramenter Set.
SSIM	Structural Similarity Index Metric.
STSA	Step-wise Temporal Sub-layer Access.
TB	Transform Block.
TI	Temporal Information.
TMVP	Temporal Motion Vector Predictor.

TSA	Temporal Sub-layer Access.
TU	Transform Unit.
UEP	Unequal Error protection.
UHD	Ultra-HD.
VCEG	Video Coding Experts Group.
VCL	Video Coding Layer.
VPS	Video paramenter set.
WPP	Wavefront Parallel Processing.
WPSNR	Weighted Peak Signal-to-Noise Ratio.

Chapter 1

Introduction

1.1 Context and motivation

Over the recent years new advances in video compression and networking technologies have been enabling the development of many multimedia applications over the internet, such as video conferencing, video streaming, video publishing, among others. The increasing diversity of services, the demands for high quality multimedia and the growing popularity of Ultra-HD (UHD) formats (*e.g.*, $4k \times 2k$ or $8k \times 4k$ resolution and beyond) are pushing technologies forward and opening new research directions. Moreover, the emerging new video formats with increased resolutions have created stronger needs for higher coding efficiency, beyond the previous standards, i.e., the MPEG-4/H.264 - Advanced Video Coding (H.264/AVC) [1,2]. Such need has been driven not only by higher resolutions but also by stereoscopic and multi-view video formats, which increase the number of viewpoints that need to be captured, coded and delivered simultaneously. Moreover, the traffic generated by mobile multimedia applications [XX] and different video services with ever increasing popularity, e.g., Netflix, impose new challenges to the existing networks. To cope with such demanding requirements a new coding standard was developed by the ITU-T VCEG and the ISO/IEC MPEG standardisation organisations, referred to as H.265 - High Efficiency Video Coding (HEVC) [3, 4].

The HEVC is currently the main video coding framework to extend current services towards higher resolution formats delivered with better compression efficiency. Similar to its predecessor, HEVC is a block-based hybrid video codec using predictive transform and entropy coding. The coding flexibility and efficiency of HEVC, results from the use of new coding tools and block structures, which enables new block partitions [5], improved prediction modes [6, 7] and new high-level features [8], 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. Therefore, the new coding tools in HEVC bring not only high compression efficiency [9], but also some other disadvantages, such as increased complexity [10] and inherent reduction in error robustness.

Earlier studies about the error robustness characteristics of HEVC have shown that in general the use of high efficiency coding tools results in poor error resilience performance, in comparison with previous standards. Thus, when HEVC streams incur in transmission losses, this leads to significant degradation of both objective and subjective quality [11, 12]. This is mostly due to the strong decoding dependencies imposed by the highly complex prediction modes of HEVC, which assume error-free transmission [13]. 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, is be based on transport technologies such as RTP [14], MPEG-2 TS [15], or more recent standards as MPEG-DASH [16]. 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 [17]. For content delivery over heterogeneous networks, the MPEG Media Transport (MMT) [18] is another possible technology to be used with the HEVC standard. Besides the transport technologies, the overall error robustness of video transmission may also benefit from 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 [19, 20]. Moreover, error resilience techniques may also take advantage of the multi-stream based approach, where each stream allows different levels of robustness to errors [16, 21].

In summary, increasing the error robustness of the highly compressed video streams is a relevant research problem whose solutions will be useful to increase the adaptability for transmission diverse network conditions and to achieve robust communications in order to provide acceptable levels of QoE to the end users at all times. The HEVC reference codec is continuously being improved and extended development by multiple proponents from around the world, also addressing the problem of error resilience and robust delivered transmission Although the EC methods have been developed during the past years for H.264/AVC [22] and recently for HEVC [23], a combined and optimised approach between encoder and decoder is yet to be developed, which would improve the users QoE by exploiting shared information between the encoder and the decoder sides.

1.2 Main objectives and contributions

The research objectives of this thesis are focused in robust error resilience for efficient error concealment in highly efficiency video decoders.

Thus, this work investigates novel techniques to increase the video quality of HEVC bitstreams delivered through error-prone networks. The following objectives were pursued in this work:

1. **evaluation study of the error robustness:** the first objective of this research work is to evaluate the robustness against transmission errors of highly efficiently coded video

2. **robust video coding:** At the encoder side, novel techniques to extend the existing error robustness of highly predictive video coding techniques are investigated in order to increase the received video quality over error-prone networks.

3. **robust video decoding:** error concealment techniques are used to increase the error robustness of video decoder and improve the perceived video quality. The paradigm of error concealment are evaluated at the encoder side, as part of the encoding process, in order to optimise the reconstruction process at the decoder side.

1.3 Outline of this thesis

This thesis is organised as follows. The current chapter introduces the context, motivation and objectives of this thesis, as well as the original contributions of this thesis.

The first two chapters review the state-of-the-art concepts and previous work related with this research. Chapter 2 presents an overview of the HEVC standard, focusing on the concepts explored in this work, e.g., high-level features and temporal predictions. Chapter 3 reviews the most important work dealing with robust video transmission also providing relevant background knowledge to this work.

Chapter 4 firstly presents an evaluation study about the error robustness of high efficiently coded video, using the HEVC as the underlying framework for various coding configuration. The study presents the relevant tools that mostly affect the error robustness. Secondly, a new method is proposed to decrease the vulnerability of temporal motion prediction and decrease error propagation, leading to higher video quality in error-prone conditions.

Chapter 5 proposes a novel error resilience approach for high efficiently video codecs. A new method to distribute the use of reference frames is proposed to reduce the

error propagation and improve the EC performance. Moreover, to reduce the effect of mismatch motion predictions (shown in Chapter 4) a method based on selective motion vector redundancies is proposed. This method is able to improve the decoded video quality using a small amount of overhead.

Chapter 6 addresses the problem of robust EC of highly efficiently video streams. Different EC algorithms were taken into consideration in order to develop a concealment-aware coding scheme. The proposed scheme is based on a scalable coding approach where the best EC methods to be used at the decoder are optimally determined at the encoder and signalled to the decoder. A generalised saliency-weighted Rate-Distortion (R-D) optimisation is used and the residue between coded frames and their EC substitutes is encoded using a rate-controlled enhancement layer.

Chapter 7 ... CHECK HERE!

Chapter 8 concludes this dissertation and presents some suggestions for future work.

Appendix A illustrates the original test signals used in simulations throughout the thesis. This includes a partial set of the HEVC test material and other external test material. Appendix B presents a summary of the published papers, describing the contributions of the research work of this thesis.

Chapter 2

The High Efficiency Video Coding Standard

2.1 Introduction

This chapter presents an overview of the H.265 - HEVC [3] standard, focusing on the most relevant features of the encoding algorithm for the work developed in this thesis. The HEVC standard aims to address essentially all existing applications of H.264/AVC [1], extending it in two keys aspects: increased video resolution and increased use of parallel processing architectures. In order to increase the flexibility and follow the increase in the video resolution, the HEVC standard adopts a new block partition structure, as well as, new prediction modes. In this chapter these new techniques introduced are described, especially those which may have a significant impact in the error robustness characteristics of the HEVC.

This chapter is organized as follows. Section 2.2 introduced the functional blocks of a generic video coding system, and Section 2.3 provides the overview of the HEVC standard features. Subsequently, Section 2.4 describes the high-level features which are used to carry relevant information regarding the bitstream, Section 2.5 describes the coding structures used to divide the video frames into smaller partitions and Section 2.6 describes the main predictions techniques, namely intra and inter coding. Finally, a briefly description of the transform and quantisation, entropy coding and in-loop filters is given in Sections 2.7, 2.8 and 2.9, respectively. Section 2.10 concludes this chapter.

2.2 General hybrid video coding

A digital video is a temporal succession of images, also referred to as pictures or frames, which usually presents a significant amount of temporal and spatial correlations. In general, video coding and distribution systems follow a generic set of tools: pre-processing, encoding, storage or transmission, decoding and post-processing. The pre-processing unit is used adapt the raw video signal, in which if the format of raw videos from video cameras is analogue, it will be firstly digitized into RGB images and then converted into YUV representation [24]. In the next processing tool, the video data will be compressed by the respective encoding unit by reducing its redundancy. The encoding process is able to compress the raw video in a high compression ratio and maintains a good trade-off between video quality and bitrate through high efficient algorithms. On the contrary, the output video for displaying can be acquired through decoding and post-processing modules.

Since H.261, the hybrid video coding scheme has been the basic structure for all video coding standards and recommendations of ITU-T VCEG and ISO/IEC MPEG. While the structure has not been changed, the algorithms represented by the building blocks have been refined and the applicable configuration for the algorithms has become more and more flexible over years. The coding scheme is called hybrid as it combines temporal prediction between pictures of the video sequence with transform coding techniques for the prediction error [25] as shown in Figure 2.1 [26].

As aforementioned, the main purpose of video coding is to compress the video by reducing its redundancy. In general, there are two kinds of redundancies existing in video signal: spatial and temporal. On the one hand, the spatial redundancy is exploited through intra frame prediction, of which the mechanism is to find the best matching block for current one from its neighbours. This is due to a high correlation between neighbouring pixels inside a frame. As result, in intra frame prediction a block is reconstructed by the blocks that has already been encoded and stored in the bitstream for output. On the other hand, inter frame prediction is designed to deal with the temporal redundancy existed between consecutive frames such as the frame in the past or in the future. Temporal redundancy is normally referred as the correlation between two adjacent frames since they usually have similar backgrounds or a same object with a relative motion. Thus, the purpose of inter-frame prediction is to search for the best matching block for the current one in the reference frames, ranging from past to future frames. After the best matching candidate for the current block is found by either intra- or inter-frame prediction, only the difference between the original block and the predicted one needs to be encoded. As seen from Figure 2.1, these differences (also called residual signal) will be further processed and converted

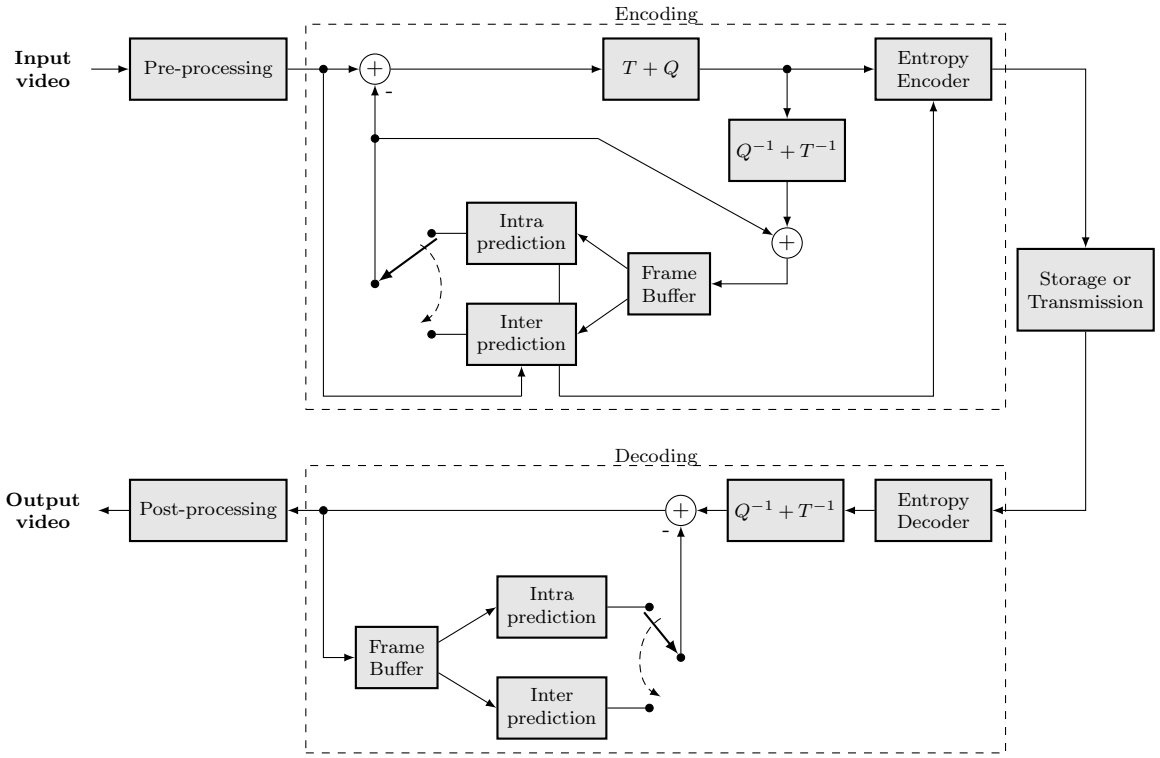


Figure 2.1: Block diagram of a generic video coding system, including the inner blocks of a generic hybrid encoding and decoding scheme.

into the coefficients of transform domain (T). In general, redundant information can be eliminated through transforming. Subsequently, the transformed coefficients are quantized (Q) to reduce the range of the signal and save the number of bits required for signalling. Finally, the quantized transform coefficients together with the prediction information are compressed to form an output bitstream for transmission by the entropy encoder, as shown in Figure 2.1. On contrary, the decoding part performs the inverse operations on the bitstream to reconstruct the video for playing. Subsequently, the aforementioned blocks are described in the context of HEVC standard.

2.3 Overview the standard features

As for all the previous coding standards developed by JCT-VC, in HEVC [4, 27, 28] only the bitstream structure and the coding syntax are standardised. Included are the bitstream constraints, the definition of the semantic meaning of syntax elements and the decoding process, such that every decoder conforming to the standard generates the same output when given a correct bitstream. Such limitation of scope in the standard allows maximal freedom for developers and enables specific optimisations to cope with various applications requirements (*e.g.*, balancing the compression ratio and the coding

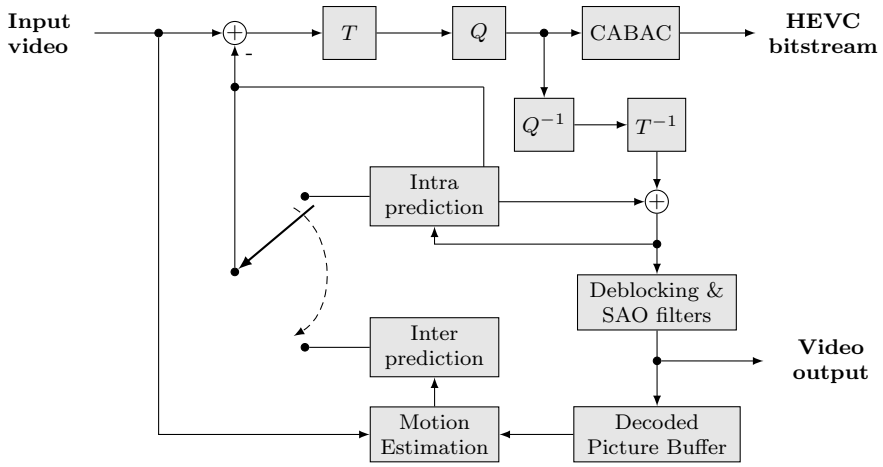


Figure 2.2: Generic HEVC standard architecture.

complexity). Therefore, in order to be considered as an HEVC encoder, the algorithm needs to ensure that the produced bitstream complies with the standard specifications.

Figure 2.2 illustrates the block diagram of the hybrid HEVC encoder. It follows the same overall architecture as its predecessors starting from H.261. The encoder uses both intra- and inter-picture prediction to exploit spatial and temporal redundancies, respectively. Subsequently, the prediction residue transformed by linear spatial transform, which is then quantised, and entropy coded. This allows to reduce the bitstream redundancy to a minimum. The resulting bitstream is combined with control information, which signals the prediction chosen, motion vectors, filtering operations and bitstream headers.

As depicted in Figure 2.2, each picture can be compressed with intra coding algorithms or inter prediction using previously encoded and reconstructed pictures kept in the Decoded Picture Buffer (DPB) as reference. The encoding process follows a block-based approach as in previous standards with the larger units referred to as CTU with a maximum size of 64×64. In order to improve the overall quality of the reconstructed pictures and smooth the coding artefacts, HEVC uses the deblocking and Sample Adaptive Offset (SAO) filters. The encoder has a built-in decoder to generate the reference pictures. In the decoding loop, the residual information is delivered to the inverse quantisation and transform modules. The finally reconstructed samples are obtained by adding the predicted signal and applying the filtering operations.

2.4 High-level syntax

This section describes the high-level components of the HEVC standard that provide signalling information and enable bitstream partitioning. HEVC inherits several high-level features from H.264/AVC, e.g., the NAL units, parameter sets, the use of Picture



Figure 2.3: Two bytes NAL unit header in HEVC.

Order Count (POC), and SEI messages for auxiliary data. Some high-level features that exist in H.264/AVC were not included in HEVC, for example Flexible Macroblock Ordering (FMO), redundant slices, arbitrary slice order, data partitioning, and switching slices. Several new high-level features are introduced, such as the Video paramenter set (VPS), the tiles and wavefront tools for parallel processing, the dependent slices for reduced delay, and a new reference picture management concept. Moreover, new pictures types were introduced new picture types to increase the random-access flexibility, and temporal sub-layer access pictures.

2.4.1 Network Abstraction Layer

Similar to H.264/AVC [1], the HEVC standard used a bitstream based on Network Abstraction Layer (NAL) units, which are useful to deal with lossy network environments, as it is useful for bitstream partitioning and synchronization [29]. The NAL provides the ability to map the Video Coding Layer (VCL) data that represents video frames on various transport layers, such as RTP, MPEG-2 Systems and MPEG DASH. NAL units are classified into two categories: (i) VCL and (ii) non-VCL, as they carry coded video frames or associated data, respectively.

In the HEVC standard two bytes are allocated to describe the content of each NAL unit. These bytes are used to identify which type of NAL is being decoded and are transmitted in a fixed format form without variable length codes. This allow extra error robustness but increases the bit allocation. Figure 2.3 show the structure of the HEVC NAL unit header. The first bit *F*, referred to as *forbiden_bit*, is always zero in order to ensure that a start code can be detected across different protocol standards. This is followed by the NAL unit type which identifies its content, enabling the decoder to access if it is required for a particular decoding purpose, or if it can be discarded. Another syntax element defined by the HEVC standard in the NAL unit header is the TIDP which marks the temporal layer to which the unit belongs. This allows for temporal scalability (layers ranging from 0 to 6) by immediately discarding the NAL unit if it belongs to a layer higher than the desired one. Finally, six bits are reserved for future extensions, preventing the creating of extra NAL units for scalable and multiview extensions as it was done in H.264/AVC.

2.4.2 Parameters set

The encoder uses a set of signalling packets as a robust method to share data with the decoder, referred to as parameter set units. This information was introduced in previous standards to deal with vulnerabilities in the transport of signalling information between the encoder and decoder, which compromised the streaming applications. Within this approach the set of parameters are included in the bitstream and are repeated as many times as required by the application. This structure is inherited from the H.264/AVC standard [29] with modifications to cope with new coding tools. The HEVC includes three sets of parameters, Sequence Paramenter Set (SPS), Picture Paramenter Set (PPS) and the newly introduced Video paramenter set (VPS) [8].

The new set of parameters introduced in the HEVC standard (VPS) contains information related to the different layers of the video, avoiding duplications like in the H.264/AVC and its extensions. VPS aims to provide signalling for every layer, guaranteeing a clean and extendible standard, capable of supporting multiple layers. Moreover, VPS carries information for session negotiation (*e.g.* profile, level and tier). The other sets are the SPS and PPS. SPSs are related to the whole sequence (*i.e.*, should affect all coded slices), and contain information related with the decoder operation point, as well as, flags to control optional tools and scalability control. PPSs carry information that may change for every picture, such as, initial Quantized Parameter (QP), flags for picture related tools and tilling information.

2.4.3 Picture types

The HEVC standard introduces several pictures types as show in Table 2.1. These pictures are divided in the following categories: (i) random-access points, (ii) leading pictures and (iii) temporal sub-layer access pictures [8, 28].

Random-Access Point

The Random-Access Point (RAP) pictures are used to provide points in the bitstream where it is possible to start the decoding process. The standard defines that a conforming bitstream must start with a RAP picture. These frames must belong to temporal sub-layer 0 and be coded using intra coding techniques, i.e., must not using previous decoded frames as reference. One should not that frames may be compressed with solely intra prediction and not be marked as random-access points. It is always possible to decode from the RAP frame and onwards to output any subsequent pictures in output order even if all pictures that precede the RAP in decoding order are discarded from the bitstream.

Table 2.1: Picture categories in HEVC standard.

Random-access points pictures		
IDR	Instantaneous decoding refresh	All leading pictures can be decoded
CRA	Clean random access	Pictures in the display order can be decoded
BLA	Broken link access	Leading pictures cannot be decoded
Leading pictures		
RADL	Random access decodable picture	
RASL	Random access skipped picture	
Temporal sub-layer access pictures		
TSA	Temporal sub-layer access	Allows switching to a given temporal layer or higher
STSA	Step-wise temporal sub-layer access	Allows switching to a given layer only

The IDR frame is an intra coded picture that completely refresh the decoding process and starts a new conforming video stream. This means that neither the IDR nor any other frame following in the decoding order have dependencies to a prior frame. IDR frames are allowed to have leading pictures, but they must be decodable, thus creating a true random-access point where all subsequent frames in the decoding order can be decoder and displayed.

In order to increase the coding efficiency, the HEVC standard introduces a new RAP picture, referred to as Clean Random Access (CRA) picture. Theses frames do not refresh the decoding process and allows leading pictures to depend on frames that precede the CRA picture in the decoding order. This enables a more efficient prediction structure, while maintaining an access point to begin the decoding process. The introduction of CRA frames in HEVC were able to improve the coding efficiency by up to 6% [30].

The HEVC supports bitstream splicing, which means taking a particular bitstream starting from a RAP (both IDR and CRA frames) and introduce it in another bitstream at a random-access instance. Therefore, whenever this is performed for a bitstream starting from a CRA frame, the leading frames associated might not be correctly decoded. Therefore, in order to make splicing operation straightforward the NAL units of the CRA picture are changed to NAL units of type Broken Link Access (BLA). In case of bitstream splicing starting from an IDR frame the leading pictures does not have dependencies to frames prior the RAP, thus converting it to BLA is not required.

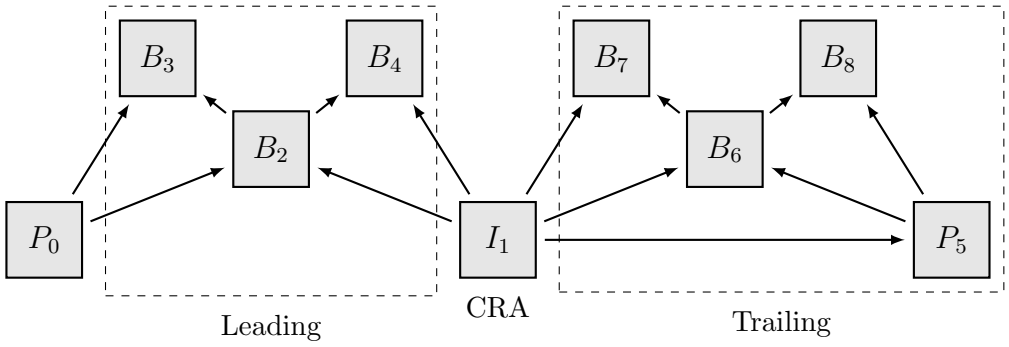


Figure 2.4: Leading and trailing pictures associated with a random-access point.

Leading and trailing pictures

Leading pictures correspond to those frame that follow a particular RAP frame in the decoding order but precedes it in the output order. A trailing picture is a frame that follows a particular RAP frame in both decoding and output order and must not depend on any leading frames nor any trailing frame belonging to a previous RAP. Figure 2.4 shows examples of leading and trailing pictures. In the figure, frames are illustrated in the output order and the number associated with the frame type (i.e., I, P and B) represents the decoding order. In this example frame I_1 corresponds to a CRA picture. Leading and trailing pictures are considered to be associated with the closest previous RAP picture in decoding order, such as frame I_1 in Figure 2.4. Also, all leading frame of a RAP frame must precede in decoding order all trailing frames that are associated with the same RAP. This means that the following order is imposed by the HEVC standard: 1) RAP picture, 2) associated leading pictures and 3) associated trailing pictures.

There is two types of leading pictures as shown in Table 2.1. On the one hand, the Random access decodable leading picture (RADL) pictures only depend on the associated RAP pictures and do not have any prior dependencies to trailing pictures associated with previous RAP frame. These frames can be correctly decoded. On the other hand, Random access skipped leading picture (RASL) picture may have dependencies to a prior trailing picture. Therefore, when a random-access is performed to the associated RAP frame, these frames cannot be correctly decoded, and should be skipped.

Temporal sub-layer switching pictures

A Temporal Sub-layer Access (TSA) picture is a trailing picture that marks a temporal sub-layer switching point. When a decoder is decoding a sub-set of the temporal sub-layers in the bitstream and encounters a TSA picture type of the temporal sub-layer just above the maximum temporal sub-layer it is decoding, it is possible for the decoder

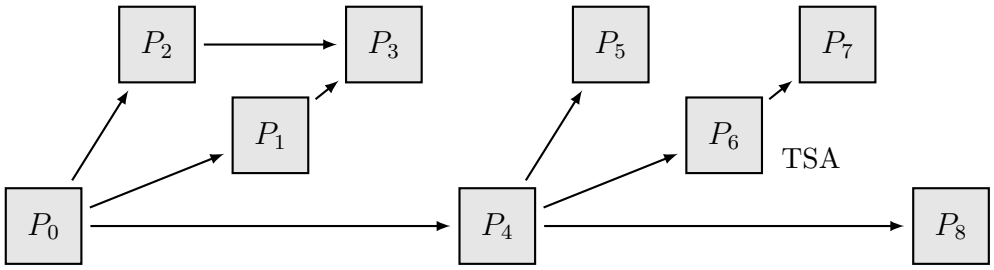


Figure 2.5: Temporal sub-layer access pictures example.

to switch up to and decode any number of additional temporal sub-layers. For example, picture P_6 in Figure 2.5 can use the TSA frame type since it only depends on frame belonging to temporal sub-layer 0 for prediction, as well as, any subsequent frame predicted from the TSA frame P_6. In the example of figure if the decoder was only decoding the temporal sub-layer 0 it can after P_6 starts decoding both sub-layer 0 and 1 or decode all the available three sub-layers. Although theses frame allows higher flexibility in terms of sub-layer switching, they severely constrain the prediction of trailing the TSA frame.

To reduce the constrains in frame prediction due to sub-layer switching, HEVC introduces the Step-wise Temporal Sub-layer Access (STSA) picture. These pictures have similar purpose as the TSA frames but they only guarantee that frames in the same temporal sub-layer as the STSA frame in the decoding order are able to be correctly reconstructed. This is guaranteed by preventing any dependency to frame preceding the STSA to be used as reference in its temporal sub-layer. One example of a STSA frame is given in also given in Figure 2.5 in frame P_2. This frame can be used to switch to the sub-layer 1 as it does not have dependencies to any prior frame in the same sub-layer (P_2 only depends on I_0). However, it cannot be classified as TSA picture since P_3 has dependencies to prior frame in the decoding order to P_1. Summarising, STSA frames can be used to switch to a particular sub-layer as the TSA can be used to switch to any sub-layer.

2.4.4 Picture partitioning

The high-level segmentation of a picture in HEVC is achieved through four techniques: regular slices, dependent slices, tiles and Wavefront Parallel Processing (WPP). Picture partitioning normally serves one or more of the following three purposes:

Error robustness: partitioning the picture into smaller self-contained units in order to gain error robustness by the ability to re-synchronize both the parsing and decoding processes in case of data losses.

Network adaptation: adapt to the network constraint of Maximum Transfer Unit

(MTU) size, found for example in IP networks. Such packetisation scheme restricts the maximum number of payload bits within a slice regardless of the size of the coded frame. To keep each slice within this limit and minimise the packetisation overhead, a variable number of coding units is used for each slice.

Parallel processing: partitioning the coded frame into units which can be processed in parallel. This is achieved by dividing the coding units in such way that they can be encoded and decoded independently of each others.

Slices

The HEVC standard preserves the slice structure present in the H.264/AVC standard. A slice may comprise either the entire frame or a section of it, and all the associated data (*i.e.*, entropy symbols, prediction and residue information) can be independently decoded. Each slice is transported in a different NAL unit. As dependencies across slice boundaries are disabled, each slice can be independently reconstructed, regardless if other slices were lost or incorrectly decoded. Slice partitioning is the only tool that can be used for parallel processing in a virtually identical form as in H.264/AVC. As they are independent from each other, it is straightforward to process multiple slices in parallel, without inter-process communication (except for inter-frame prediction). Although this is a simple approach, it incurs in substantial coding overhead due to the higher number of slice headers, and reduction of causal neighbours for prediction (due to lack of predictions across slice boundaries). Another disadvantage of using slices for parallel processing is that they are also used to meet the MTU size constrains. Therefore, it might be impossible to meet both requirements using only one technique.

Another partitioning tool in HEVC standard is the dependent slices. These slices allow the partitioning of the frame at coding unit boundaries without breaking intra-frame predictions. Moreover, as they have a smaller slice header they are more efficient in terms of signalling overhead than regular slices. Dependent slices are normally used to reduce the end-to-end delay by allowing part of the slice to transmitted while the rest of it are still being processed.

Tiles

A new data partition has been included in the HEVC standard, referred to as Tiles [31]. Tiles enable the partition of a coded frame into a smaller group of CTUs using horizontal and/or vertical boundaries. Figure 2.6 shows two examples of slice partitioning using tiles. In both cases the frame is partitioned into three tiles by using two vertical boundaries. By defining tiles using only boundaries it allows for an efficient signalling with a small overhead. These boundaries are defined in the PPS NAL units.

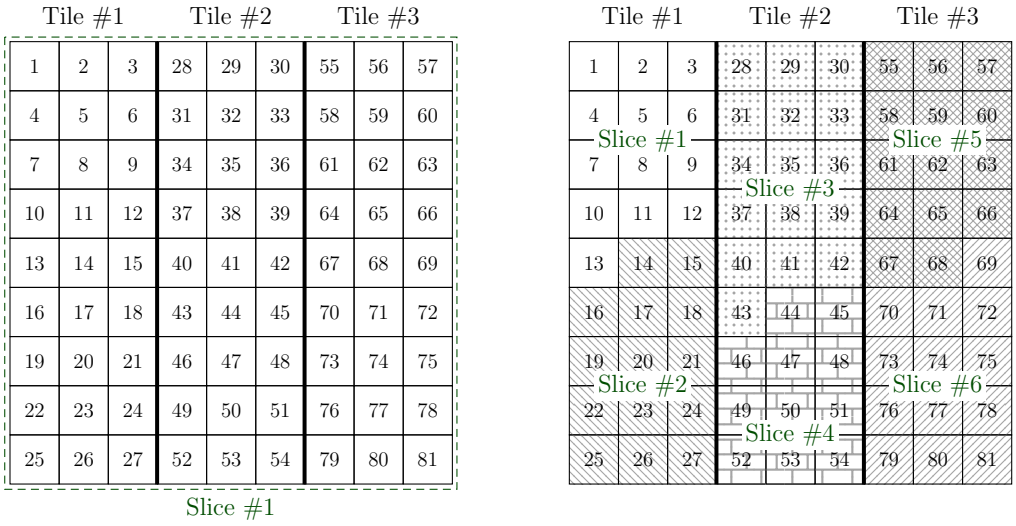


Figure 2.6: Examples of the interaction between tiles and slices.

Although the HEVC standard enables the use of both slices and tiles at the same time, it impose constrains in order to reduce the implementation complexity. One of two cases must be used: (i) all CTUs within a tile must belong to the same slice, or (ii) all CTUs within a slice must belong to the same tile. These cases are illustrated in Figure 2.6 (a) and (b), respectively.

The use of tiles provide several advantages. Specifically, they improve frame partitioning for parallel processing [32] when compared to slices, by reducing the required overhead. Moreover, they also improve the MTU size matching and reduce the line buffer memory (by dividing the frame as in Figure 2.6). Finally, as tiles allow flexible partitioning,they enable the definition of ROIs for asymmetric video coding.

Wavefront Parallel Processing

As mentioned before, the use of either slices or tiles to process each frame using parallel threads require the use of different entropy coding contexts for each slice/tile, in order to make them independently decodable. In consequence the compression efficiency decreases. In order to overcome this issue, the concept of Wavefront Parallel Processing (WPP) was introduced, which enables coding efficient parallel processing [33].

In WPP, each slice is divided into single rows of CTUs which can be processed in parallel. Entropy decoding and prediction are enabled across CTUs of different rows. Therefore, higher compression efficiency can be achieved. In order to re-use the same context in the whole slice each row requires a two CTU delay in order to make sure the required dependencies have already been coded/decoded. Figure 2.7 illustrates

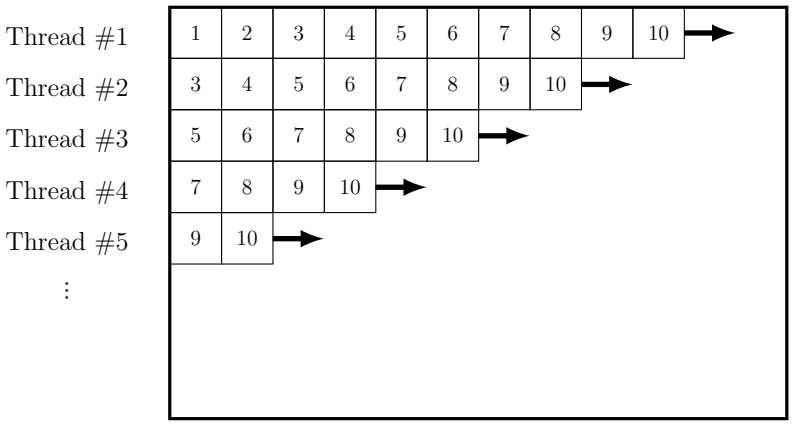


Figure 2.7: Illustration of the wavefront parallel processing.

an example of WPP using at least five different threads. Each square corresponds to a CTU and the numbers corresponds to the time instance in which a CTU may start being processed. Using this staggered start (which appears like a wavefront as represented in the figure), the parallelisation may use as many threads as the number of CTUs rows available in the video frame. Although, the number of threads does not affect the coding efficiency, the required inter-process communication substantially increases.

2.5 Coding structures

Despite the fact that the HEVC standard follows a traditional hybrid video coding architecture, it introduces significant changes in the data structures when compared to the previous standards (e.g., H.264/AVC). In H.264/AVC each frame is divided into units so-called Macroblocks (MBs), which are composed by a 16×16 block of luma samples and two 8×8 blocks of chroma samples, assuming video data in YUV 4:2:0 sub-sampling. The MB is considered the basic processing unit in the H.264/AVC, which define if the containing pixels are compressed with intra or inter prediction techniques. Each MB could also be partitioned into sub-blocks which should follow the same prediction type.

In HEVC each picture is divided into blocks so-called Coding Tree Units (CTUs), which are composed by the luma and chroma Coding Tree Blocks (CTBs) and corresponding signalling [34]. The maximum CTU size is configured by the encoder and signalled to the decoder, and should be one of the following: 64×64, 32×32 and 16×16. The use of coding units with larger size in HEVC was adopted to achieve higher coding efficiency especially in high resolution video. Each CTU can be recursively partitioned into four smaller units, referred to as CUs, until it reaches the minimum unit size, which cannot be inferior to 8×8 pixels. Similar to the maximum CTU size the minimum

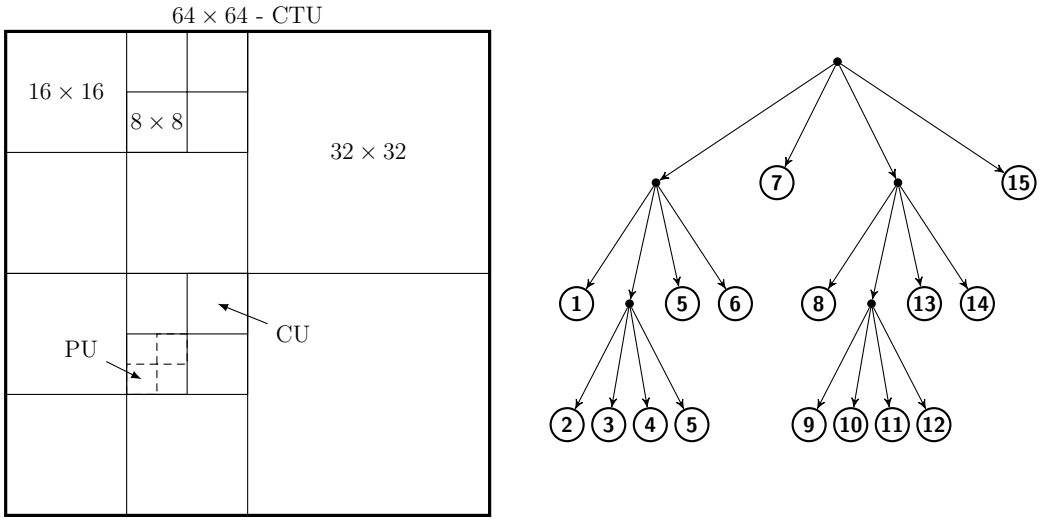


Figure 2.8: Representation of a CTU (64×64) and its partitioning into CUs and PUs (left); corresponding quad-tree with coding order (right)

size is also defined at the encoding time. Figure 2.8 illustrates an example of a CTU partitioning into CUs. The partitioning illustrated in the figure leads to the quad-tree structure represented on the right side with the corresponding coding order.

The CU is the basic processing unit in HEVC for prediction. It is further partitioned into two types of units, specifically PU and Transform Unit (TU), which are used for prediction and transform operations, respectively. Each PU is composed by Prediction Blocks (PUs), and each TU is composed by Transform Blocks (TBs). The decision whether to use intra-frame or inter-frame prediction modes is applied at the CU level. However, depending on the partitioning of the CU into PUs different intra or inter coding modes can be used for each PU.

When the intra prediction mode is signalled for a CU, the size of the PU need to be the same as the Coding Block (CB), except when the smaller CB is selected. In this case further splitting into four PU of smaller size. For example, when the smaller CB size is 8×8, the encoder may decide to divide it into four PUs of 4×4 pixels, each one having its own intra prediction mode. This is signalled using an one bit flag. The use of small PUs is useful for complex regions that required a fine-granularity prediction.

When inter prediction mode is signalled, HEVC defines a higher range of partitioning schemes, allowing the CU to be divided into one, two or four PUs. The splitting into four PUs is only allowed when the CU size is equal to the minimum. Otherwise, the splitting should be carried out at the CU level which leads to the same partition tree. Figure 2.9 illustrates the possible CU partitioning into PUs. One should note that the partition applied to the CU is also applied to divide the CB into PUs. Figure 2.9 (a) shows the symmetric partitions. The $N \times N$ corresponds to the case where the CU is not further divided. Then, the standard defines the horizontal and vertical division

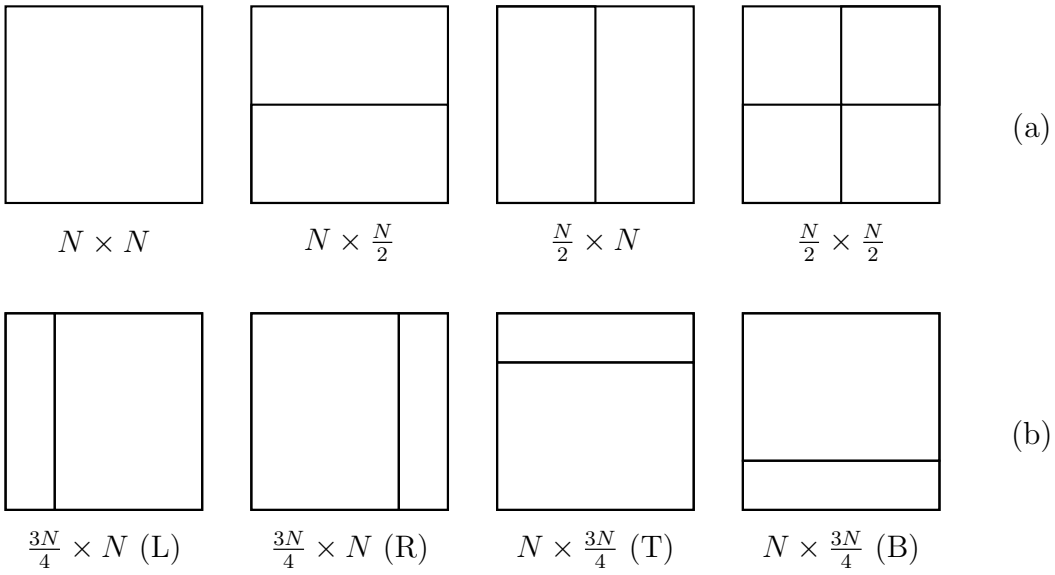


Figure 2.9: Partitioning modes of an inter coded CU into PUs [5].

into two PUs, corresponding to the $\frac{N}{2} \times N$ and $N \times \frac{N}{2}$ cases, respectively. Finally, the $\frac{N}{2} \times \frac{N}{2}$ case result in four PUs. The partitions on Figure 2.9 (b) correspond to the asymmetric partitioning, which is one of the main improvements introduced in HEVC, as it allow for higher flexibility when performing motion estimation and compensation, leading to higher compression efficiency [5]. These modes are only allowed for CUs with 16×16 pixels or larger.

2.6 Prediction methods

Prediction methods are responsible for an important role in image and video coding standards, due to their ability to reduce the signal redundancy based on the previously encoded pixels. This section describes the prediction techniques used in the HEVC standard. These techniques include the directional intra prediction for efficiently reduce spatial redundancy and inter-frame motion compensation for remove the temporal redundancies.

2.6.1 Intra-frame

In previous standards, intra-frame prediction was used to efficiently reduce the spatial redundancy in video frames. Intra predicted blocks are obtained using the previously encoded pixels around the current PU. HEVC adopts three types of intra-frame prediction: directional, Planar and DC [6,35]. Figure 2.10 shows the available intra prediction mode available in HEVC. As shown, the predictions are applied to square blocks only.

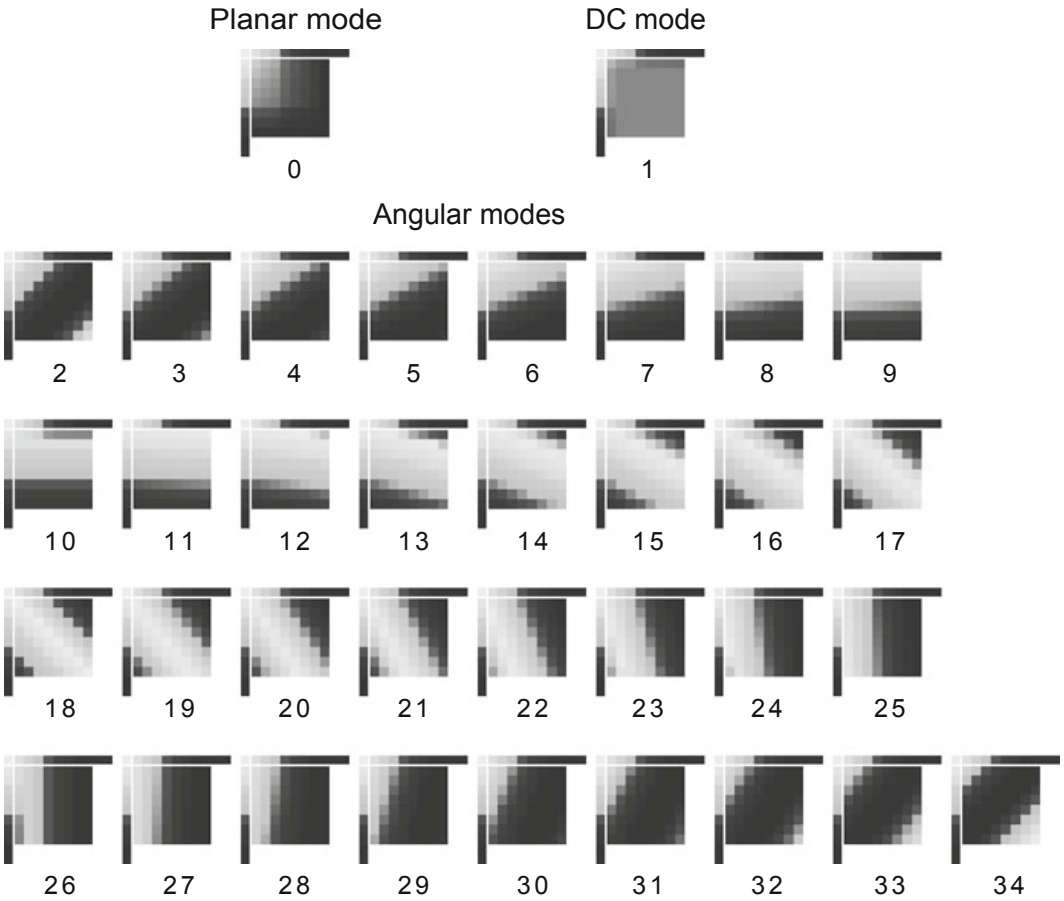


Figure 2.10: Example of HEVC intra prediction modes for a 8×8 block [28].

The example in the figure corresponds to PU of 8×8 pixels, but the standard allows block sizes ranging from 4×4 to 32×32 The encoder tests all the available intra prediction modes and signals the best mode to the decoder. The decoder must generate an identical prediction block, based on the reconstructed pixels.

The reconstructed reference samples used in the prediction process belong to the neighbouring blocks at left-down, left, top-left, top and top-right positions, as can be noticeable in Figure 2.10. Given a PU with size $N \times N$, intra prediction requires a top neighbouring row of $2N$ pixels, a left column of $2N$ pixels plus the top-left neighbouring pixel. Due to unit boundaries, e.g., slice or tile boundaries, neighbouring pixels might not be all available to apply the intra prediction. In addition, when constrained intra prediction is enabled, reference samples belonging to inter-predicted PUs are not used in order to remove error propagation from potentially erroneously reconstructed reference frames. In order to overcome this issue and being able to apply all possible modes, HEVC uses reference sample substitution to replace the unavailable reference samples with the closer available ones, so that all intra prediction modes can be used.

Angular modes

Directional intra prediction was first introduced in the H.264/AVC standard [2], to estimate regions with a structured texture or directional edge. In its process, directional prediction projects the reconstructed samples in the block neighbourhood along a specific direction. The HEVC improves the number of available direction from 8 in H.264/AVC to 33 possible directions [6]. The result of the projections in each possible direction is illustrated in Figure 2.10.

The angular modes are designed to provide a dense coverage near the horizontal and vertical directions. It is not noticeable in the Figure 2.10, but the angular directions are more coarse as it gets closer to the diagonals. This reflects the observed statistical prevalence of the angles and the prediction efficiency.

In order to improve the efficiency of angular prediction the when the reference samples need to be projected a bilinear interpolation is used from the two closest pixels using a 1/32 accuracy.

Planar and DC

In contrast, with the angular mode which cover those regions with strong direction edges, HEVC also implements the Planar and DC modes targeting smooth regions. These modes are useful for smooth regions as they result in a linear or constant prediction.

The DC mode uses the average or the neighbouring reference samples to generate a constant prediction for the current PU. The Planar mode predicts the PU through linear predictions from vertical and horizontal slopes derived from the neighbouring reference samples. As the Planar mode used linear surface for prediction it prevents the discontinuities at PU boundaries. Both of these modes are also illustrated in Figure 2.10.

Sample smoothing

In HEVC sample smoothing is applied at two distinct occasions two improve the overall performance of the prediction accuracy. Firstly, the reference samples are filtered and secondly the PU is filtered after prediction. This improves the previous standard H.264/AVC where only reference samples were filtered.

In HEVC the reference sample smoothing is conditionally applied, based on the PU and intra prediction mode. Similar to H.264/AVC, a three tap $[1\ 2\ 1]/4$ smoothing filter is applied, except for PU sizes of 4×4. The prediction results from Planar mode is also filtered when the PU size is greater than 8×8, and it is not used for the DC mode as it would not be useful.

The second filter in HEVC is the boundary smoothing. This filter aims at removing the discontinuities along block boundaries due to intra prediction. The application of this filter is more exclusive, and it is only applied to the DC and angular mode 10 and 26 (exactly the horizontal and vertical directions, respectively). A two-tap filter $[3\ 1]/4$ fed by the original samples and the adjacent reference sample. For the DC mode this filter is applied to both the top row and the left-most column. In case of the angular mode 10 and 26 only the left-most column and the top row are filtered, respectively. Only the mentioned boundaries are filtered, as they are the most likely to introduce discontinuities. The result of the boundary smoothing can be noticed in Figure 2.10 for the aforementioned modes.

2.6.2 Inter-frame

The efficiency of video coding algorithms consistently relies on inter-frame prediction techniques used to reduce temporal redundancy. The underlying idea of inter-frame prediction is to estimate the current frame from one or more previously encoded reference frames using block-based motion compensated. The most common technique used for motion estimation in HEVC, as well as in H.264/AVC, is the boundary matching algorithm. In the HEVC standard an Advanced Motion Vector Prediction (AMVP) for efficient motion vector signalling, which defines the Merge Mode and differential coding of the motion information.

Motion estimation and compensation

Motion estimation consists in searching for the current block with the highest similarity with the target block to be predicted (PU in HEVC), using block-matching algorithms over the previously encoded frames, i.e. the reference frames. Typically, reference frames correspond to past or future temporal frames. The frame encoding order determines which reference frames are selected and whether future frames are available. Typically, video frame type P normally are predicted from the past as B frame are predicted from frames both in past and future. Block-matching consist on subtracting the current block from one of equal size in the reference frame in a predefined window. The reference block that results in the lowest difference is selected for prediction. In order to signal the selected block to be used for motion compensation the different between the target and the reference block positions are computed to obtain a two-dimensional vector, referred to as MV. This vector is signalled to the decoder, in order to apply the same prediction.

Similar to H.264/AVC the HEVC allows for either one or two MV to be selected to compensate the target PU. This results in uni-predictive and bi-predictive coding,

respectively. HEVC also allows for weighted prediction, which consists in scaling and offset operations the predicted block using determined factors which are explicitly sent to the decoder. The motion vectors are obtained from frames included in the reference frame list. Each frame is identified by an index, therefore to fully describe inter-frame prediction a combination of two parameters is required: MV and reference index.

In recent standards motion compensation has been performed using a variable block size approach. In H.264/AVC block sizes ranging from 4×4 up to $16\times$ pixels were allowed. In HEVC a significant improvement was introduced by extending the block sizes up to 64×64 pixels. As mentioned before it also allows asymmetric rectangular block sizes. By using larger prediction units, HEVC is able to save bits in the signalling of the selected MV, because less MV are required.

In order to increase the efficiency of motion estimation using fractional MVs are used. This has been proven to improve accuracy of motion compensation. To accomplish this, frame interpolation is performed to generated non-integer pixels in the reference frame. Both H.264/AVC and HEVC standards use MVs up to quarter-pixel accuracy for luma component. For chroma samples, the MV accuracy is determined according to the chroma sampling format, which for 4:2:0 results in eighth-pixel accuracy. In contrast, with the two-stage approach used in H.264/AVC, HEVC uses a single separable approach for the interpolation process without intermediate rounding operations. This improves the precision and simplifies the architecture of the fractional interpolation.

The block matching-based motion estimation is one of the most computationally complex procedures in the MV encoder. Normally, the optimal solution for motion estimation is provided by the full-search algorithm, which tests all the possible candidate blocks in the search area. Although the full-search algorithm achieves the optimal performance, it requires the highest computational complexity. Therefore, sub-optimal algorithms have been proposed over the years to reduce the motion estimation complexity and decrease encoding time [36–38]. For example, the HM reference software implements a fast algorithm based on Test Zone Search [36]. Since motion estimation is a non-normative procedure, encoders may implement any alternative search algorithms that suits its needs (i.e., compression ratios or encoding time).

Reference picture management

Another improvement in HEVC is related with the reference picture management scheme. HEVC introduces the Reference Picture Set (RPSet) concept [8], which defines how previously decoded pictures are managed in the DPB. Decoded frame in the DPB are mark grouped in one of the following categories: (i) *short-term reference*, (ii) *long-term reference* and (iii) *unused for reference*. Once a frame is marked as *unused*

Table 2.2: Reference Picture Set example.

Frame	RPSet ({frame, is used})		
I_1	–		
P_5	$\{I_1, 1\}$		
B_6	$\{I_1, 1\}$	$\{P_5, 1\}$	
B_7	$\{I_1, 1\}$	$\{B_6, 1\}$	$\{P_5, 0\}$
B_8	$\{B_6, 1\}$	$\{P_5, 1\}$	

for reference it is no longer used for motion compensation and it will be discarded from the DPB after being displayed/outputted. In H.264/AVC a sliding windows approach was used to implicitly manage the reference picture list. This process discards a given frame whenever the DPB has the number of defined reference frames (which is signalled by the encoder). Moreover, H.264/AVC used an explicitly Memory Management Control Operation (MMCO) to send commands to the decoder and change the normal reference picture management process.

In the HEVC standard instead of a signalling relative changes, the status of the DPB is signalled for every slice. The most fundamental difference with the RPSet concept compared to MMCO/sliding window of H.264/AVC is that for each particular slice a complete set of the reference frames that are used by the current or any subsequent frames are be provided. Thus, a complete set of all frames that must be kept in the DPB (in the present and future) are signalled. This is different from the H.264/AVC scheme where only relative changes to the DPB are signalled. With the RPSet concept, no information from earlier frames in decoding order is needed to maintain the correct status of the DPB. Table 2.2 shows the RPSet for some frames illustrated in Figure 2.4. As expected for the CRA frame (I_2) nothing is signalled. For the remaining frames, two types of signalled are used: frame number and a flag to mark if it is currently used. This is required because any frame not listed in a given RPSet will be discarded and subsequent frames cannot use it as reference.

The RPSet concept in HEVC provides a basic level of error robustness to the reference picture management. In H.264/AVC whenever a reference frame is lost the decoder can look at the value of the *FrameNum* (incremental value associated with each reference frame) to detect the loss event. However, as gaps in *FrameNum* are allowed the decoded might not be able to decide the missing frame was intentional or if it resulted from a loss event in the transmission. Moreover, even if the decoded if able to detect the loss event it might not be able to correctly compute the value the POC associated with the concealed frame. Using the Reference Picture Selection (RPS) concept in the HEVC the decoder is always able to detect loss events and correctly identify the missing frames, preventing it from use wrong frames as reference.

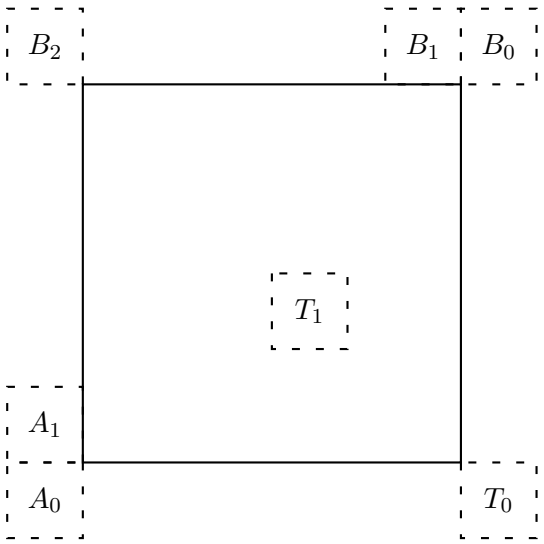


Figure 2.11: Illustration of the MV prediction candidates in the AMVP.

Merge mode

The Merge mode was introduced in HEVC standard, as a new technique to derive motion information from spatial and temporal neighbouring blocks [39]. This technique extends the concepts of the direct and skip modes used in H.264/AVC [2], using a more sophisticated approach with two key differences. Firstly, the Merge Mode allows for more MV candidates. Thus, an index value is included to select one out of several MV candidates. Secondly, it explicitly identifies the reference picture index used for the selected candidate, as well as, the reference picture list. This results in an improved flexibility, compared with the predefined values used in H.264/AVC. The MV candidates allowed in the Merge Mode [39] are illustrated in Figure 2.11. The figure illustrates the spatial candidates' positions (A_0, A_1, B_0, B_1, B_2). Moreover, it also uses a temporal motion vector candidate, derived from the co-located position on the temporal adjacent frame at positions T_0 and T_1. From this MV candidate list the unavailable candidates are not accounted and duplicated candidates (i.e. with the same motion information) are removed. In the Merge Mode, a $C-1$ number of candidates are used, selected from the aforementioned order and one temporal MV candidate is selected to fulfil the total number of MV (C). The value of C is defined at the slice header and a flag controlling the use of temporal MV candidates is sent at the PPS-level. For B-slices, additional MV candidates are generated by choosing two existing candidates according to a predefined order for reference picture list 0 and list 1. A total number of twelve combination are specified in HEVC which are listed in [4]. For P-slices if the number of candidates is less than C zero motion vectors are included to fill the remaining entries.

The Merge Mode allows for the signalling of a MV candidate and the residual

information. In HEVC, the Skip mode, i.e., transmitting only the MV without residual information, is treated as a special case of the Merge Mode. In this case only a skip flag is enabled, and the corresponding merge index is transmitted.

Motion vector prediction

The HEVC also improves the performance of the motion vector coding for those cased where the Merge Mode is not used. HEVC always use a difference encoding approach for efficiently represents the motion information. It considers the same MV candidate of the Merge mode, illustrated in Figure 2.11, to derive the motion vector predictors from which the difference vector is obtained. From these MV candidates only two are selected based on the following:

1. Left candidate: first available from A_0 and A_1.
2. Top candidate: first available from B_0, B_1 and B_2.
3. Temporal candidate: first available from T_0 and T_1.

Therefore, the use of the temporal MV candidate is limited to those cases where the spatial neighbours are not able to provide two valid candidates.

HEVC only allows a much lower number of candidates to be used in the motion vector prediction process, since it can still send a motion vector difference. Moreover, the encoder will perform motion estimation for each candidate, which increase the encoding complexity for every possibility tested. For those cases where the reference index is not the same in the current PU and the neighbouring one, a scaling version of the MV is used.

2.7 Transform and quantisation

Similarly, to the previous standards, HEVC uses a transform followed by a quantisation to compress the residual signal. The residual block is partitioned into multiple square blocks, referred to as Transform Block (TB).

Core transform

A 2D transform of the residual information is obtained by applying 1D transforms in both the horizontal and vertical directions. The core transform matrices are obtained from the Discrete Cosine Transform (DCT) basis functions to define a integer matrix of 32×32 points. Sub-sampled versions of this matrix are used to derive the remaining transform matrix sizes down to 4×4.

A special case is applied for the residual coding of a 4×4 intra coded block. In this case an alternative integer transform based on the Discrete Sine Transform (DST). The motivation for DST is related to the statistics of the residual signal of these blocks, which tend to present higher residue as the distance to the boundary used for prediction (i.e, top or left boundary) increases. Experiments have shown that DST provides up to 1% of bitrate savings for the compression of 4×4 intra predicted blocks, without significantly increase the computational complexity [40].

Quantisation

HEVC standard follows the same principle introduced in H.264/AVC for the quantisation method. After obtaining the residual transform a quantisation is applied using a uniform-reconstruction quantiser scheme based on the QP. The standard supports QP values ranging from 0 to 51, and a increase of 6 in the QP corresponds to doubling the quantisation step. This procedure is the main responsible to introduce the coding errors as it performs many-to-one mapping, which cannot be reversed.

2.8 Entropy coding

In contrast to H.264/AVC standard, which allows two entropy coding methods, HEVC only uses the Context Adaptive Binary Arithmetic Coding (CABAC) [41]. CABAC is an arithmetic coding method that only uses binary symbols, considering different probability models for each symbol. Entropy coding efficiency is closely linked with the context model selected. Therefore, this was carefully designed in HEVC standard, extending the functionality presented in H.264/AVC For example, the HEVC standard exploits the depth of the partition tree or residual transform tree in order to derive the context models for several syntax elements. Regarding transform coefficient coding, CABAC uses a scanning method to firstly organise the coefficients and then transmits the position of the last non-zero transform coefficient. Moreover, it transmits, a significance map, the sign bits and the levels for the transform coefficients. Three coefficient scanning methods are available: the diagonal up-right, the horizontal and the vertical scan. The coefficient scanning is implicitly selected and always performed in 4×4 sub-blocks for all transform block sizes.

2.9 In-loop filters

The HEVC standard uses two filtering procedures over the reconstructed pixels before writing them into the frame buffer, namely the Deblocking Filter and the SAO filter.

The main purpose of deblocking filter is to reduce blocking artefacts caused by the block-based coding tools, operating similarly to the deblocking filter of the H.264/AVC standard. DBF is applied to the samples adjacent to the PU or TU boundaries, that are aligned on a 8×8 grid. Such grid-based restriction reduces the computational complexity and improves parallel-processing implementation which is one of the main goals of HEVC. Three strength levels can be used by filter depending on the coded block characteristics.

The SAO is a non-linear filter, that is adaptively applied to all samples of the image, after the deblocking filter. In its operation, the SAO filter modifies the samples by adding an offset value, extracted from lookup tables transmitted by the encoder. For each CTU, the encoder decides whether to apply the SAO filter or not, and if used, one of the two available filter types is applied, namely the band offset or the edge offset. In the band offset mode, the added offset value depends on the sample amplitude, while the edge offset uses the gradient to classify and derive the offset value.

2.10 Summary

In this chapter we described the main features of the High Efficiency Video Coding (HEVC) standard, giving a special focus on the high-level features, coding structures and newly introduced prediction modes.

At the high-level HEVC introduces new NAL units types, not only for new signalling features but also for the new pictures types introduced. HEVC introduces new picture types to handle different types of random-access strategies, which allow more flexibility and higher coding efficiency, when compared to the random-access in previous standard such H.264/AVC.

Another major improvement comes from the inter-frame prediction. On the one hand new high-level reference picture management was introduced, which increase the error robustness, as it explicitly defines the active reference frame and the DPB status. On the other hand it introduces the AMVP and the Merge mode which efficiently reduce the required bits or motion signalling.

In general, HEVC achieves higher coding performance than its predecessors. Previous studies reported an average bitrate saving of approximately 50% for the same subjective quality. Such amount of bitrate saving is normally related with increased prediction structures with higher dependencies. Such dependencies are more sensible to data loss, which will increase the error propagation to subsequent coding structures. Therefore, it is emergent to evaluate the impact of the transmission errors in the HEVC and develop efficient techniques to cope with data loss and reduce the error propagation.

Chapter 3

Robust Video Coding: A Review

3.1 Introduction

In general, digital communications through networks are prone to transmission errors and data losses. These errors can be handled in different ways, such as retransmission, error correction, and EC [42, 43]. This chapter reviews the main robust video coding techniques that can be applied at the video encoder/decoder to reduce the error propagation and increase the decode video quality. Several techniques have been proposed in the past to deal with the problem of robust video transmission, especially for video coding standard with high temporal dependencies, such as H.264/AVC and more recently HEVC. These techniques normally address the spatio-temporal dependencies in video coders which provide higher coding gains [13] but lead to lower error robustness. In this chapter the error resilience and concealment schemes are divided in different categories depending on they functionality and algorithms used. Although different each approach follows a different techniques they all aim at striking the most efficient balance between coding efficiency and decoded video quality.

This chapter is organised as follows: Section 3.3 describes the localisation techniques, which reduce the temporal and spatial dependencies in coded bitstreams. Section 3.4 provides a overview of methods based on data partition and classification and bitstream redundancy. These techniques ease the introduction of extra information in order to recover from lost data. Section 3.5 covers the techniques used at the decoder-side as post-processing tools to reduce the effect of errors and improve visual quality and smoothness. Section 3.6 provides a description of methods use the EC techniques to provide improved error robustness at the encoder-side. Finally, Section 3.7 concludes this chapter by providing a overall comparison between the different approaches.

3.2 Categories of robust video coding techniques

In order to increase the robustness of video transmission, robust coding mechanisms can be employed in a transmission chain to mitigate the artefacts caused by data loss. These techniques can be divided into several categories depending on the functional characteristics and the main problem they address [44]. In this chapter we divide the robust video coding techniques into four categories:

- Localization techniques;
- Data partitioning and redundant coding;
- Error concealment by post-processing;
- Concealment-aware error resilience.

Table 3.1 provides a summary of the techniques covered in this section and the key advantages associated with each class of techniques. Localisation techniques are able to reduce the spatial and temporal dependencies in the bitstreams [45, 46]. Data partitioning groups the coded data into different categories depending on their relative importance. Redundant coding makes use of redundant blocks of data to correct the transmission errors and achieve robust decoding in error prone conditions. The problem of robust transmission may also be addressed from the decoder side, by using post-processing EC methods to ease the effects of data loss. Finally, concealment-aware techniques might be adopted at the encoder to increase the performance of EC algorithms and increase the reconstruction quality of lost data at the decoder [47–50].

Table 3.1: Advantages of error robustness techniques by category.

Category	**Examples**	**Advantage**
Localization	• End-to-end distortion • Adaptive intra refresh • Reference picture selection • Leaky predictions	Reduces error propagation
Data partitioning and redundancy	• Slice classification • Flexible macroblock order • Redundant pictures • Multiple-description coding	Unequal error protection, transport priority and multi-channel transmission
Error concealment by post-processing	• Linear interpolation • Boundary matching • Motion vector extrapolation	Does not require extra information and adapts to every stream
Concealment-aware	• Error concealment R-D optimisation	Increases the error concealment performance

3.3 Localisation techniques

In general, video coding standard the compression efficiency is achieved by exploiting both the spatial and temporal redundancy of video sources. Since there is high correlation between neighbouring information, predictive coding schemes are very effective and obtain high compression gains exploiting such correlations. The drawback of applying such techniques is the fact that the compressed symbols becomes very dependent on each other's which highly increase the error propagation. Localization techniques break the predictive coding loop in order to reduce the likelihood of errors affect large regions of the video stream, both in the spatial and temporal domains. As expected since localization techniques break the predictive coding, it reduces the efficiency of such methods. This leads to lower compression ratios and increasing bitrate.

This category of error resilience techniques are able to reduce the error propagation at two levels, i.e., temporal and spatial levels as shown in Figure 3.1. As shown in the figure spatial techniques break the dependencies within a single frame, as temporal error propagation can be reduce by selectively breaking the dependencies across different frames.

3.3.1 Error resilient rate-distortion optimisation

A common localization technique addresses the error resilience problem of the rate-distortion optimisation on video encoders. Normally, such optimisation is equivalent to the minimisation of the Lagrangian cost given by:

$$J = D + \lambda R, \tag{3.1}$$

where D is the distortion of all pixels, R the required bitrate, and λ is the Lagrangian multiplier. In [13,51,52] a rate distortion optimisation for error resilience was presented. In order to take into consideration possible network losses a recursive optimal per-pixel estimation (ROPE) of the end-to-end distortion is applied at the encoder-side. Let f_n^i denote the original value of pixel i in frame n, $\hat{f}_n^i$ denote its encoder reconstruction and $tilde f_n^i$ the decoder reconstruction after error concealment if needed. Since data loss may occur in the transmission channel, $tilde f_n^i$ is considered a random variable at the encoder. The expected mean square error of each pixel is given by:

$$\begin{aligned} E\left\{d_n^i\right\} &= E\left\{\left(f_n^i - \tilde{f}_n^i\right)^2\right\} \\ &= (f_n^i)^2 - 2f_n^i E\left\{\tilde{f}_n^i\right\} + E\left\{\left(\tilde{f}_n^i\right)^2\right\} \end{aligned} \tag{3.2}$$

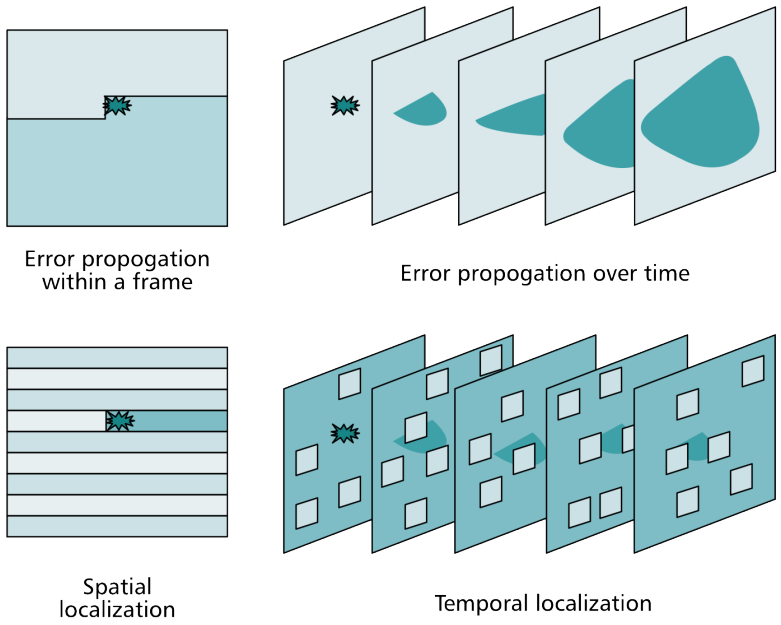


Figure 3.1: Examples of spatial and temporal error propagation and how it can be reduce using localisation techniques [44].

Then to take into consideration the channel losses packet loss probability (p) must be used in the distortion measurement. This leads to the following for intra-coded pixels:

$$\begin{aligned} E\left\{\tilde{f}_n^i\right\} &= (1-p)\,\hat{f}_n^i + pE\left\{\tilde{f}_{n-1}^i\right\} \\ E\left\{\left(\tilde{f}_n^i\right)^2\right\} &= (1-p)\left(\hat{f}_n^i\right) + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\}, \end{aligned} \tag{3.3}$$

and inter-coded pixels:

$$\begin{aligned} E\left\{\tilde{f}_n^i\right\} &= (1-p)\left(\hat{e}_n^i + E\left\{\tilde{f}_{n-1}^j\right\}\right) + pE\left\{\tilde{f}_{n-1}^i\right\} \\ E\left\{\left(\tilde{f}_n^i\right)^2\right\} &= (1-p)\left(\left(\hat{e}_n^i\right)^2 + 2\hat{e}_n^i E\left\{\tilde{f}_{n-1}^j\right\} + E\left\{\left(\tilde{f}_{n-1}^j\right)^2\right\}\right) + pE\left\{\left(\tilde{f}_{n-1}^i\right)^2\right\} \end{aligned} \tag{3.4}$$

where the inter-coded pixel i is predicted from pixel j in the previous frame. The prediction error e_n^i, is quantized to the value $\hat{e}_n^i$, which is transmitted to the decoder along with the motion information. Equations 3.3 and 3.4 allow to measure the distortion taking into consideration the compression error but also the possible of frame loss and consequent EC distortion. This technique was used in [51] to switch between intra and inter coding modes reducing the temporal dependencies, and, thus achieving superior error robustness in video transmission. In [53] this technique was optimized to quarter pixel estimation in H.264/AVC standard. Moreover, in [54] this method is extended for

burst losses by introducing the probability of burst losses into the original end-to-end estimation. This shown to improve the quality under frame loss. In [55] the distortion estimation assumes a temporal EC at the decoder-side. This increases the performance of the proposed as it does not rely of frame-copy which leads to lower reconstruction quality. More advanced EC method were also consider in the study presented in [55].

In order to increase the error robustness of the HEVC standard distortion estimation was also used in the R-D optimisation [45] showing significant quality improvements comparing with the reference HEVC implementation. The method in [56] proposes the estimation of the end-to-end distortion using the Structural Similarity Index Metric (SSIM) [57] due to its trade-off between complexity and efficiency to evaluate the perceptual quality [58].

3.3.2 Intra-refreshing

In order to reduce the spatial dependencies one can introduce refresh points into the bitstream through intra-coded blocks [59–66]. These techniques are fairly efficient to promote the error robustness of bitstreams, which avoid long error propagation at the temporal level. The main feature of these techniques is to increase the number of intra-coded blocks as they have no dependencies to previous frames, however at the expense of higher amount of bits [60]. Therefore, different alternatives have been proposed optimising the number of intra-coded blocks introduced by reducing it the refresh regions. Such refresh regions can be introduced randomly, periodically, based on the motion information or on the end-to-end distortion estimation. Moreover, whenever is possible the encoder can use a feedback channel to decide where to introduce refresh regions [59, 65].

In the H.264/AVC reference software it was introduced a non-normative technique for random intra refreshing which introduces cyclic refresh not in raster scan order. Instead a random pattern is applied. This technique guarantees that all MBs are refreshed at some point, therefore errors do not propagate indefinitely. However, this technique refresh all MB exactly the same number of times without consideration for rate distortion optimisation neither the region containing the MBs. In [67] a rate control scheme is used to decide if a given MB should be refreshed using a intra-coded block. The decision is performed based on the ratio of the Lagrangian cost between the best intra and inter mode. Whenever this ration is lower than a threshold a intra mode is selected. Using this guarantees that intra refreshing only forces an intra prediction mode if its Lagrangian cost is closer to optimal coding method. In [62] this is further extended to adapt the intra refresh algorithm to the network conditions, i.e., the threshold value is obtained from the bitrate and the Packet Loss Ratio (PLR) value. This technique automatically increases the number of intra-coded block as the bitrate

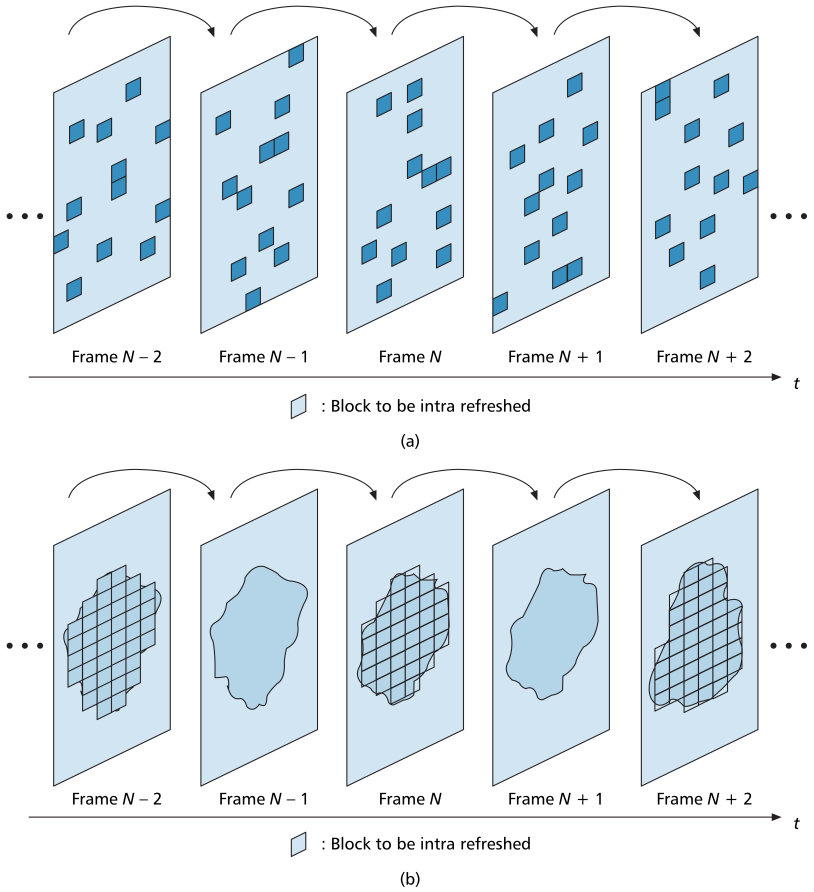


Figure 3.2: Examples of spatial and temporal error propagation and how it can be reduce using localisation techniques [60].

and the packet loss increases. In case of HEVC intra refresh is also being used to reduce error propagation [68]. It was proposed to use joint intra-inter prediction to introduce soft-refresh points in the bitstream. Although such regions are not correctly decoded in case of errors (as happens with intra-refresh), it significantly decreases the bitrate increasing due the amount of intra-coded blocks. When compared with traditional intra-refresh bitrate saving of up to 12% are achieved for a 5% PLR.

A different approach introduces refresh points based on the video contents. In [60] the error resilience of video streams is increased combining rate control and attention based models to define ROIs. A weighting factor is used in the Lagrangian optimization to increase the number of intra blocks areas determined by the attention models. Using the aforementioned method lower refresh points are required to achieve high error resilience in the ROI. A more recent approach was proposed in [69] using a perceptual noticeable difference model to determine an estimation of noticeable distortion. Then, using a visual attention model presented in [XX] intra refresh blocks are introduce in the bitstream. This method is able to improve the results of a previous work [70], especially for high motion sequences. This is due to the fact the previous approach

only used an estimation of the error propagation without accounting for perceptual information.

This was be also exploited for the HEVC standard [71] in combination with a feedback channel. Based on the received information the encoder can introduce intra CUs in the ROI areas. Experimental results reveal an increasing on the achieved video quality when higher number of intra CUs are used. Moreover, applying the proposed method to the ROI exclusively lower bitrate overhead is introduced leading to higher rate distortion gains.

3.3.3 Leaky-predictions

Directly using reconstructed past frames as reference frames for predictive coding, may result in substantial error propagation whenever the reference frames are lost. In order to reduce error propagation one can employ leaky predictions [72, 73], which scales down reconstructed frames. This leads to an exponential decay of error propagation, which leads to higher error robustness. In order to generate such reference frames the following can be applied

$$\check{f}_n^i = \alpha \hat{f}_n^i + (1-\alpha)C, \tag{3.5}$$

where $\check{f}_n^i$ is a reference frame pixel, α is the leak factor, and C is a constant value. The use of leaky predictions has been used in the past for error control especially in scalable video coding [74], where C in Equation (3.5) is replace by a lower layer error-free reconstruction. In case of single layer coding, the use of a constant value has significant impact on the coding efficiency leading to higher bitrate as shown in [75].

In order to overcome the coding efficiency problem weighted prediction can be used with similar advantages. Since H.264/AVC motion compensation can be applied from two individual previously encoded frames and combined together to generate a single prediction for the current frame [76]. In case of leaky prediction instead of applying motion compensation is applied once, but with an interpolated frame obtained as follow:

$$\check{f}_n^i = \alpha \hat{f}_n^i + (1-\alpha)\hat{f}_{n-1}^i. \tag{3.6}$$

Using $\check{f}_n^i$ as reference frame for prediction significantly decrease error propagation if only one of the original coded frames is lost (i.e., $\hat{f}_n^i$ or $\hat{f}_{n-1}^i$).

In [73] in order to increase the error robustness with lower impact on the coding efficiency leaky prediction is confined to a pre-defined ROI. Figure 3.3 shows an example of error propagation within two different regions: ROI region and background. As show in the figure, since the ROI regions have predictions from regions in the background error propagation across regions occurs. Therefore, the method in [73] uses a weighted prediction for the ROI region obtained as follows: 1) a globally optimal

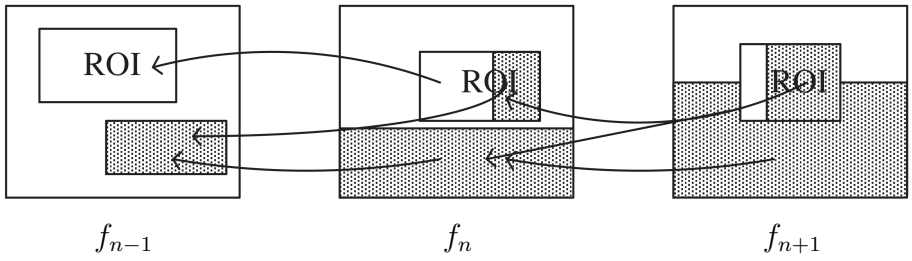


Figure 3.3: Example of error propagation with a ROI (shaded area indicates error, and the arrows indicate temporal predictions to previous frame) [73].

prediction obtained by unrestricted motion estimation, 2) a prediction obtained by restricting motion estimation within the ROI region. This limits the impact on ROI from errors that affect background region. Moreover, since the leaky predictions are confined to a specific regions higher coding efficiency is achieved when compared with other methods. Equation (3.6) results in a finite impulse response based weighting of reference frames, as previously weighted frames are not considered. In [52] an infinitive impulse response based scheme is proposed to increase the decay in error propagation. To accomplish that, in Equation (3.6) the use of the previous encoded frame ($\hat{f}^i_{n-1}$.) is replaced by the previous interpolated frame ($\check{f}^i_n$). The idea of combining multiple predictions was also been used in the HEVC standard by combining it with end-to-end distortion to select the optimal predictions.

3.3.4 Reference picture selection

In order to reduce the temporal dependencies and reduce the error propagation reference picture selection has been widely used in the past [77–80]. In [77], an optimisation algorithm was used to select between short and long term references based on the end-to-end distortion obtained using the Recursive Optimal Per-Pixel Estimation (ROPE) method, in order to increase the prediction distance and reduce error propagation. This method can also be used in combination with feedback channels to improve the references frame selection. In the decoder side, an EC method was later used to take advantage of such prediction structure [81].

A feedback channel was also used in [78], to reduce error propagation at end-user decoders, by dynamically selecting the correctly received frames as references. Figure 3.4 shows an example of error resilience transcoding, where based on network feedback the original temporal predictions (represented by the solid arrows) are transcoded to use the correctly received frame, i.e., prediction represented by the dashed arrow. A more advanced scheme may be considered, where only the regions affected by mismatch predictions are process in the compressed domain, and subsequently transcoded, thus limit the dependencies to lost frames. This was further exploited in [80]. To limit complexity

motion vector concatenation is employed, in order to obtain new inter-predictions to error-free frames. In [82] feedback-based is combined with path diversity to select the reference frames that are more likely to be received at the decoder. In the proposed approach, the feedback is not directly used to select the reference frames. Instead, the reference frames are selected through an end-to-end distortion estimation, taking into account the source distortion and the channel distortion (concealment distortion is also taken into account in channel distortion). In [46] a ROI is defined based on the motion objects is used to limit the reference picture selection based on a feedback channel. Moreover, intra refresh is also used based on the temporal distance to the previous intra-refresh frame.

In [61] an algorithm to introduce intra refresh points at the macroblock level was proposed for the H.264/AVC standard. This method takes into consideration the distortion when no concealment is applied, the distortion when frame-copy concealment is applied and the packet loss conditions, in order to select the refresh blocks. Those refresh blocks are encoded using either intra prediction or inter prediction through a secondary algorithm to select the most reliable reference frame (i.e., frame with less total expected distortion).

In the methods mentioned above the dependencies between the frames are analysed and reduced in order to decrease the error propagation in cases of data loss. Since the existing video coding standards use intense predictive coding, these methods present an efficient approach in highly erroneous conditions. 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).

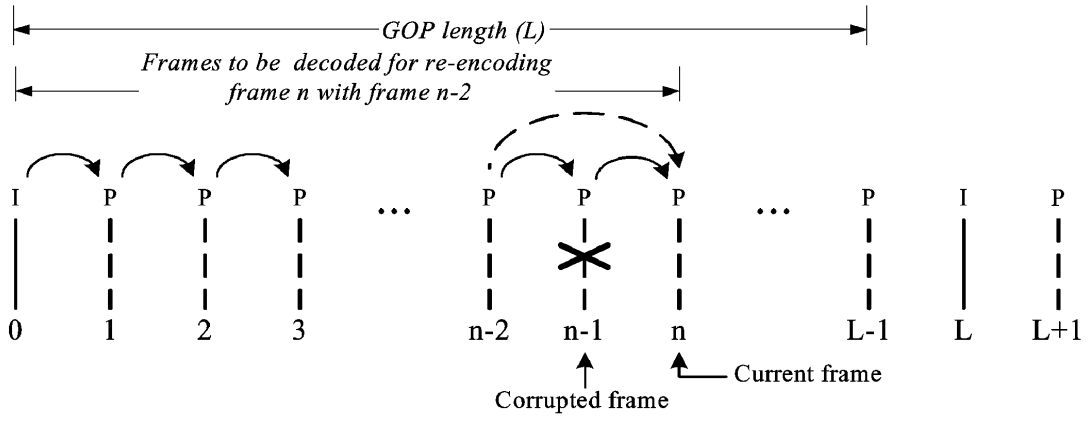


Figure 3.4: Examples of reference picture selection in error resilience video transcoding [78].

3.4 Data partition and redundant coding

In a compressed bitstream different symbols have different relevance in terms of error propagation, therefore the impact on the decoded quality is dependent on lost symbols. Therefore, data partition techniques have been developed to group different parts of the compressed bitstream according to their importance to the decoded quality. Then, some of those groups can be protected more efficiently than others using Unequal Error protection (UEP). These techniques are normally associated with Forward Error Correction (FEC) algorithms. By applying multi-levels of FEC, different levels of error protection are achieved.

In a single channel transmission, the important groups can be protected with stronger channel codes than the remaining groups. In case of multi-channel the important groups can be routed over the reliable channels. Moreover, to achieve higher quality in multi-channel transmission one can use extra information to increase the error resilience of video streams, and be able to decode then in case of errors. Redundancy can be added in an explicit form, e.g., redundant picture, or it can be implicit, e.g., multiple description coding [83].

3.4.1 Flexible Macroblock Ordering

The H.264/AVC [2] introduced an error resilience tool named FMO technique [84]. This technique allows each MB to be assigned freely to a specific slice group (i.e., a set of slices), organized in a mixed-up fashion. The standard supports up to eight slices groups in each picture, and MBs within a group can be assigned to a several slices in the default raster scan order. The case where only one slice groups is used corresponds to the case where the FMO tools is disabled. FMO increases the error resilience by partitioning the bitstream into several groups. Results show that by using FMO it is possible to enhance the subjective and objective quality especially for high packet losses with at cost of an affordable overhead [84].

H.264/AVC FMO support six types of slice groups. Figure 3.5 shows FMO types 0 to 5. In the FMO Type 0 each slice groups has a maximum number of MBs before the next slice group begins. However, the same slice groups can appear more than once in a frame. FMO Type 1 corresponds to scattered slices, where the MBs are assigned to each slice groups with a fixed formula. In 3.5 it is represented the FMO type 1 case with two slice groups resulting in a checkerboard pattern. In the FMO Type 2 slice groups are defined by rectangular shapes defined by the top-left and bottom-right corners. Although the standard allows slice groups to overlap each other, each MB may only belong to one slice group. FMO Type 3, 4 and 5 are special types each allow

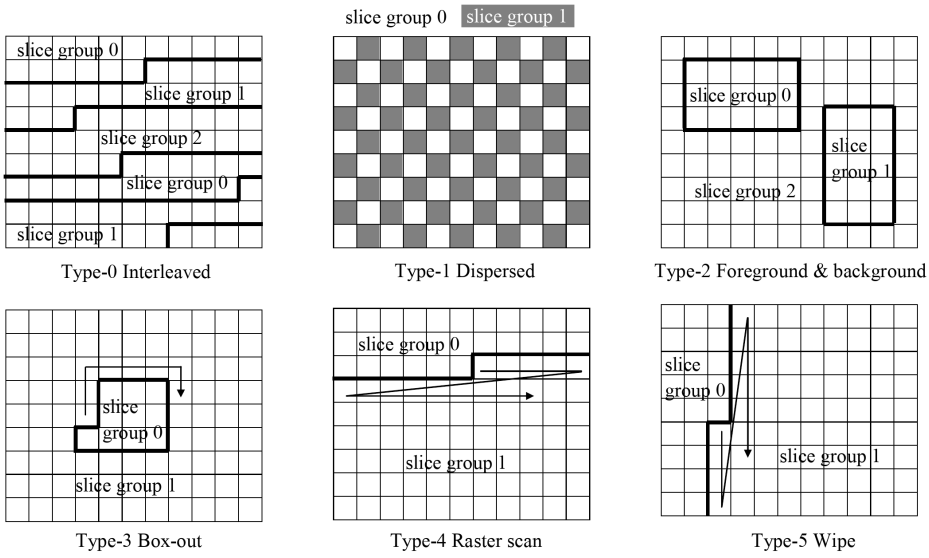


Figure 3.5: Different types of flexible macroblock reordering [84].

the configuration to change on every frame. It only allows two slice groups, in order to keep the signalling overhead affordable. The last FMO type is the most general case, where the slice groups are explicitly defined using high level packets (i.e., signalling NAL units), allow more flexibility at the cost of higher signalling overhead.

In the past FMO has been used widely used to increase the error robustness of video transmissions [85–89] in H.264/AVC streams. In [85] FMO is used to increase the EC performance and thus reduce the error propagation in H.264. The method divides the MBs into the two slice groups based on the EC distortion using a two-pass approach. Then, the slice groups which has more impact of the EC distortion is transmitted using a reliable channel, and it is assumed to be received error-free. Results show that the proposed approach is able to outperform the dispersed mode (FMO type 1). In [86] slice classification is also used based on the weighted sum of the motion vector information and the rate-distortion cost. As in the previous approach, the MBs are ranked based on the importance and distributed in the two groups until the same amount of MBs are presented in the two slice groups. This was further improved in [87] by applying a prediction based on the previous frame to estimate the mapping of the MBs into the slice groups. This reduces the computational complexity of the MB grouping while achieving similar error resilience performance. The method in [88] divides the video into two regions which are then grouped in different slice groups. In order to classify the different MB spatial information is devised from the MB bit allocation, and temporal information is extracted based on the EC distortion. This technique is able to reduce the number of erroneous MB up to 82% achieving quality of up to 3.97%, measured with PSNR, when comparing to the case where FMO is not used. In [89] a method based on motion energy is also used to classify MBs into different groups by . In order

to classify a given MB into a groups its motion energy is compared with the motion energy of its neighbours and the average motion energy of the entire frame. This allows to obtain a dynamic threshold which is then used to determine the MB importance. This method is compared with other UEP error protection schemes showing quality improvements under different PLRs.

In [90] a new FMO type is proposed combining the type 1 and 3 in order to obtain a checkerboard pattern in a box-out scanning order, thus increasing the scattering of lost MBs. Results show that the proposed FMO type is able to outperform the existing types under single and burst loss events. The method described in [91] introduces a new type in order to efficiently divide the video contents in two regions with different levels of importance, based on the FMO Type 6. Then, a spiral protection is applied to scan the MBs grouping then in two groups. Moreover, by proposing a new FMO type they are able to signal several ROI areas achieving lower overhead than using the original FMO Type 6.

3.4.2 Unequal error protection

Video signals normally have different types of content, both at the temporal and spatial level. Therefore, there is the possibility of categorising it into different groups. Moreover, it is well-know that different regions of a video bitstream are affected differently by transmission errors. This motivated the use of Unequal Error protection (UEP) techniques in video streaming, which allows different parts of bitstream to be better protected against errors. This not only ensures that those regions will be delivered with higher quality, but also leads to a reduction of the total overhead required to increase the error robustness of video streams.

UEP has been used for a long time to increase the error robustness of video transmission. A wide variety of unequal loss protection schemes have been developed in the literature to optimize the performance of video communication over networks with packet loss. It has been observed that, within a Group of Pictures (GOP), different frame types (e.g. I, P, or B frames) have different levels of contribution to the reconstructed video quality. This has been exploited for unequal error protection in video transmission as shown in [92, 93].

In [94] an evaluation of layered video was described. In order to increase the error robustness, redundant packets are used which ease the recovering of lost ones. In order to reduce the overhead for error protection each video layer was given a different importance and protected using different levels of redundancy, thus guaranteeing that lower level packets have a higher probability of being decoded without errors. This is a simple approach not dependent on the network capabilities, neither requires a deep knowledge of the video signals. A lightweight model is used to assign packets priorities

to be used by error protection algorithms [95]. This way an UEP can be achieved. The model is based on a weighted sum of four components: (i) a slice type parameter giving higher importance to refresh and reference frames; (ii) a binary flag indicating where a network packet transports video related headers, i.e., NAL or slice header; (iii) a parameter indicating the distance to the packet containing the next slice; (iv) a parameter indicating the distance to the next refresh frame. The encoder-side is able to extract these parameters based on the NAL header, which makes it suitable to low-end encoders or even to prioritise each packet at a streaming level.

However, UEP techniques usually require a deep video analysis which is difficult its integration in every transmission scenario. For instance, one can use the motion information to classify different frames or even different regions within a frame, e.g., MB in H.264/AVC [93]. The motion information is used to determine the motion energy which is given by the L2-norm of the MVs within a determined region. In [96] the motion energy of each frame is computed and compared with neighbouring frames. To achieve this the energy of co-located MBs in subsequent frames are compared, and higher importance is given to frames with higher motion energy differences. This method was compared with equal error protection revealing higher objective and subjective quality revealing the relation between motion information and impact of errors. This work was extended in order to use a two levels unequal protection [97, 98]. Not only MB and frame importance is taken into consideration but also the combined importance of different frames are used to determine the importance of each GOP.

In [99] the transmission distortion is estimated at the packet level, to assess the impact of each packet in the error propagation. Then, different priorities are applied to each packet allow different levels of error protection. This was also be used to schedule the packet transmission through multi path networks [100]. In order to improve the classification of different regions of a video stream a perceptual algorithm can be used. This increases the correlation between the classification and subjective impact of errors. In [101] an method based on a combination of two parameters: a spatial parameter based on a perceptual attention model and a temporal parameters based on the frame position within GOP. The parameter is then used for FEC assignment, and experimental results show a significant quality improvement both used objective metrics and subjective evaluation.

In scalable video coding scheme, the layered structure of the bit stream provides a natural framework for UEP [102–105]. In [103] an algorithm to estimate the overall distortion of reconstructed frame in the decoder is proposed for scalable video. At the encoder-side different enhancement layer truncation, error propagation and EC at taken into account to perform the distortion estimation. The method performs a recursively computation of the total expected distortion in the decoder at the picture-level for each

layer in the prediction hierarchy. This is used to perform a cross-layer optimisation and apply different levels of error protection. The work in [106] proposes a technique for finding the optimal coding rates for coded bit-streams at the encoder-side, in order to optimise the necessary FEC codes at the enhancement layer which aid the decoding of erroneous base-layer video. This not only improves the overall decoded video quality but also prevents that the correctly received enhancement layers are discarded due to missing dependencies.

In the context of HEVC there are also methods to achieve unequal error protection. Since techniques like FMO are not available, different approaches were taken into account [105, 107, 108].The work in [109] proposes a resource allocation scheme based on prioritisation for Long Term Evolution (LTE) Advanced networks [107]. The prioritisation is performed at the slice level (i.e., NAL unit level) based on the inter-frame predictions. Such predictions are analysed and a ranking is given based on the amount of pixels predicted based on the current slice. In [108] a quality degradation model is proposed for video encoded using the HEVC standard. The quality degradation is obtained by simulation error events and measuring the quality using the PSNR metric. From the empirical results a model can be devised to predict the quality under different events. This model is used to control the amount of redundant codes are applied at the network level. The work in [105] proposes the use of UEP for video storage in cloud environments in the context of the scalable extension of the HEVC standard. The proposed method not only takes into consideration the layer information, but also considers the picture type (i.e., I, P or B-frame) to achieve a cross-optimisation.

3.4.3 Redundant pictures

Redundant pictures is a well-known technique adopted in the H.264/AVC standard [110], which allows for different representations of the same picture to be transmitted using different encoding parameters. For example, the primary picture may be coded with fine quantization, while the redundant one with coarse quantisation at lower bitrate. One the one hand if the primary picture is received, the redundant information is discarded. One the other hand if errors occur and the primary picture is lost, the redundant one is used to provide a lower level of reconstructed quality and contain the error propagation.

Zhu *et al.* investigated the paradigm of redundant picture coding in H.264/AVC standard [110]. Different approaches were investigated in order to achieve a mixed method based on both reference picture selection to obtain a clean decoding without error propagation and a hierarchical allocation of the redundant pictures to increase the error robustness. Redundant pictures achieve good performance under error prone conditions, since they introduce secondary pictures that can be decoded when the

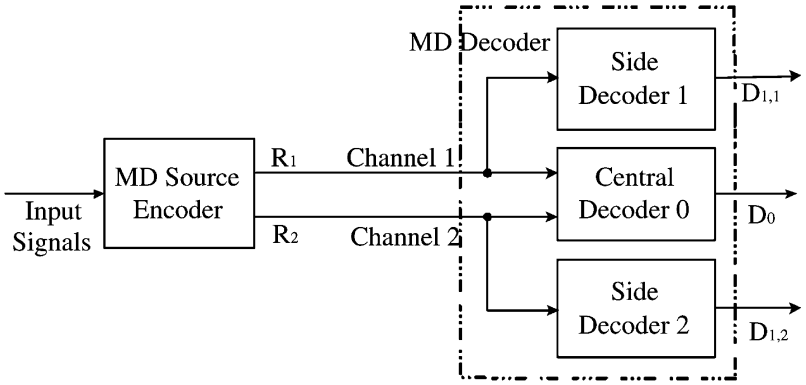


Figure 3.6: Architecture of multiple description system [83].

primary picture is not available.

Moreover, redundant pictures might also work as localisation techniques when different reference frames are selected. However, such approaches generate large amount of redundant data and overheads. In [111] a newer approach was proposed to reduce the amount of information in the redundant pictures of the H.264/AVC standard, by introducing only redundant MVs. This approach increases the overall system performance by presenting an efficient trade-off between the resultant bitrate and the error robustness.

3.4.4 Multiple description coding

In a Multiple Description Coding (MDC) scheme different bitstreams, named descriptions, are encoded such that a acceptable level of quality is achieve with one of the bitstreams isolated [83]. Figure 3.6 illustrates the a basic architecture of encoder and decoder used by the MDC technique. The encoder-side creates two descriptions, i.e., R_1 and R_2, which shall be transmitted using two different channels, reducing the probability of both being lost at the same time. As shown in the figure three situation may happen at the decoder-side: both descriptions are correctly received and decoder using the central decoder, or either one of the descriptions are not received and then one of the side decoders is used. When both bitstreams are received, the central decoder combines the information carried by each description achieving higher quality. In the MDC encoder the redundancy can be controlled by changing the correlation between descriptions. Higher correlation leads to higher redundancy but also higher quality in the output of the side decoders. In contrast with redundant pictures the redundant streams do not carry the same information which means that if both are received the decoded quality is increased.

Different types of multiple description coding [112] can also be used to improve the

quality under transmission errors in multiple path networks [113]. Although these techniques can be very effective in multi-path networks, a significant amount of overhead is required to create the different descriptions.

3.5 Error concealment by post-processing

In general, the Error Concealment (EC) operation is a post processing step in the decoder side and it is the last step to handle errors in the communication chain [114]. Thus, concealment of missing data should take place within the video decoder to minimise perceived effects of lost data. In summary, to enhance the 3D video quality and depth perception, the design of the transmission system must involve the development of adequate EC tools. This section reviews several EC methods based on spatial and temporal redundancies.

3.5.1 Spatial error concealment techniques

A typical approach to conceal missing regions in an image is based on linear interpolation of neighbouring pixels [115,116]. Figure 3.7 shows the estimation process of a lost pixel (lp), within the lost region (R_L), using the four closest pixels (p_i) as reference, and the distance between the lost pixel and the reference ones (d_i). The pixel value (lp) is recovered using a weighted average, as follows:

$$lp = \sum_{i}^{n} \frac{p_i}{d_i} \Bigg/ \sum_{i} \frac{1}{d_i} \,. \tag{3.7}$$

Liner interpolation is a simple yet effective method to recover lost regions in digital images. However, as each lost pixel is recovered by giving the same importance to each reference, the image discontinuities are not taken into account. Thus, some pixels used as reference may belong to distinct regions of the image, without significant correlation with the lost region itself. This leads to a poor estimation of lost pixels, resulting in noticeable artefacts in the reconstructed image. For example, in Figure 3.7 the lost pixel (lp) and the pixel p_4 are separated by an edge. Therefore, p_4 may not be suitable to recover the lost pixel value. So, to ensure a correct concealment of the lost region, the edges and image transitions must be considered, by selecting correct pixels to be used in the interpolation function. To overcome this problem, a directional interpolation can be performed, by recovering the lost pixels based on those within the same region, according to the edge directions.

The directional interpolation improves the performance of the pixel recovery, by estimating the lost edges using either correctly received or previously recovered pixels.

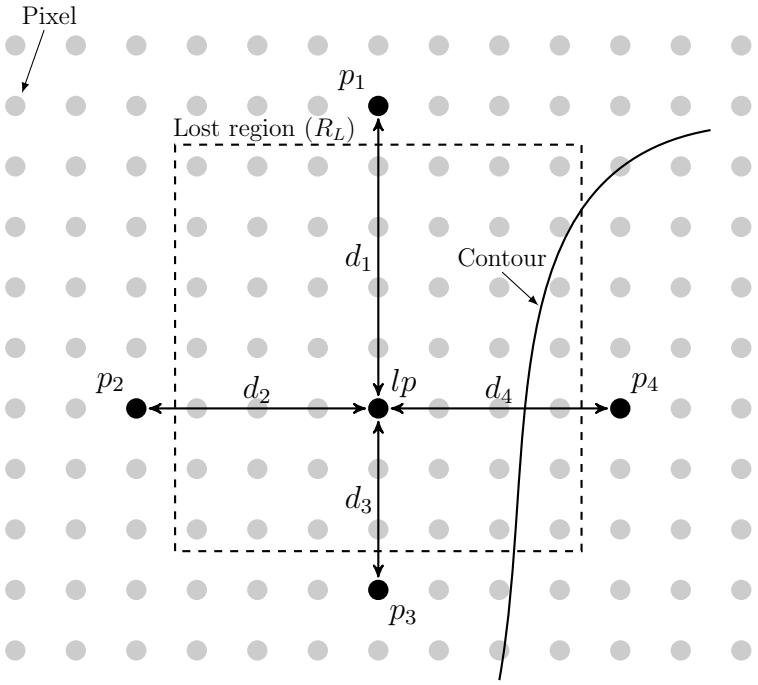


Figure 3.7: Lost pixel interpolation from neighbouring error-free pixels.

In order to obtain the best direction to interpolate, edges are calculated using the image gradient, for example applying the Sobel operator [117]. This operator gives the first-order approximation of the gradient of the image. This is widely used in image processing to locate the image edges, and uses an horizontal and vertical mask given by:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{3.8}$$

The convolution between the image pixels and the Sobel matrixes results in both horizontal (g_x) and vertical (g_y) gradients, from which the magnitude and the angular direction are computed using,

$$\begin{aligned} G &= \sqrt{g_x^2 + g_y^2}, \\ \theta &= tan^{-1}\left(\frac{g_y}{g_x}\right). \end{aligned} \tag{3.9}$$

The Sobel operator is an efficient algorithm to detect edges, as it can reach extreme values at the image edges, so it can be used in pixel interpolation. The usage of this operator for EC was evaluated in [118–122], and it is characterised as a fast and efficient method due to its small kernel [123].

3.5.2 Temporal error concealment techniques

Motion-copy

The simplest way to recover missing frames in the temporal domain is using the Frame-Copy (FC) or the Motion-Copy (MC) method. These two basic concealment methods were proposed for the H.264 reference software in [124]. The FC method consists in repeating (copying) the temporally closest frame in order to fill the gap of the missing frame. The MC method uses the motion information of the closest frame to recover the missing one, thus the missing pixels are recovered through motion compensation. The experimental results showed that the motion-copy method achieves higher quality in the reconstructed frames in most of the cases. However, the complexity increases with the implementation of the MC method, when compared with the FC method.

In [22] a whole frame loss error concealment algorithm was proposed to improve the performance of the MC method, since no refinement is performed in the MVs obtained from the closest decoded frame. Moreover, in this work the MVs between consecutive frames are assumed to have a strong correlation, so the motion information of several preceding frames is taken into account to estimate the lost one. Then, the motion vector differences between the vectors of successive frame pairs are recursively calculated, as well as the next refinement area is determined, until it becomes unchanged. This is shown to improve the performance of the MC algorithm, achieving gains up 0.47 dB (PSNR) for a PLR of 20%.

Boundary match algorithms

The recovery of missing blocks in video frames may rely on the available neighbouring regions, to find the suitable MV to reconstruct the missing pixels. A low complexity approach may use the boundary information of surrounding blocks to search for the motion vector based on a spatial smoothness constraint on boundaries of a lost MB [125]. Figure 3.8 represents the boundary information used to estimate the MV for the lost block from a set of different MVs candidates. The Boundary Match Algorithms (BMA) uses the available MVs provided by the top (T), bottom (B), left (L) and right (R) blocks (see Figure 3.8) to fetch the pixels from the previous decoded frame. Then, the optimal MV is selected according to the sum of differences, between pixels on the internal boundary in the recovered block and the corresponding ones on the external boundary. These pixels are marked in the figure.

In [126], error concealment based on the boundary matching algorithm is improved, by introducing the concept of the interpolated candidate motion vectors. The candidate MVs are calculated with an interpolation of the motion vectors of the neighbouring block in the missing frame and the region around the corresponding block in the pre-

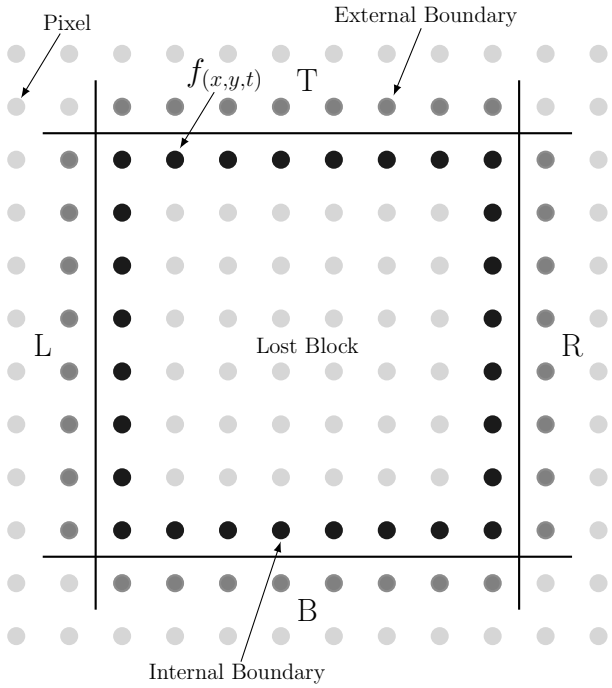


Figure 3.8: Illustration of the boundary match algorithm.

vious frame. The interpolated candidate MVs are created based on the change of magnitude and angle of the top, bottom, left and right neighbouring MVs of the lost block from the previous frame to the current one. Then, these modifications are applied to the MV of the co-located block with the lost one in the previous frame. A more generalised proposal is to apply such transformation not only to the closest neighbours but also to a wider area of blocks. The concept of interpolated candidate MVs can be applied to any method based on BMA, improving its performance, by introducing more accurate candidates.

Another improvement to BMA was obtained with the outer boundary matching algorithm [127], also known as decoder MV estimation algorithm in [128]. In this method the distortion is calculated by the difference of two outer boundaries of reconstructed block, instead of internal and external boundary in BMA. Although BMA and outer BMA have the same design principle of minimising the mismatch distortion, in later the distortion function is measured in the outer boundaries, *i.e.*, external boundaries of the reconstructed block. It was observed in the experimental results that outer BMA has better performance in video recovery [127] than conventional BMA. The performance improvements are explained by the usage of the outer boundary in the distortion functions, which incorporates the edge information within the neighbourhood of the lost block in the decision of the most accurate MV to recover the missing pixels [127]. Moreover, in the method presented in [127] further MV candidates were introduced, a refined local search is performed and multiple boundary layers, *i.e.*, in-

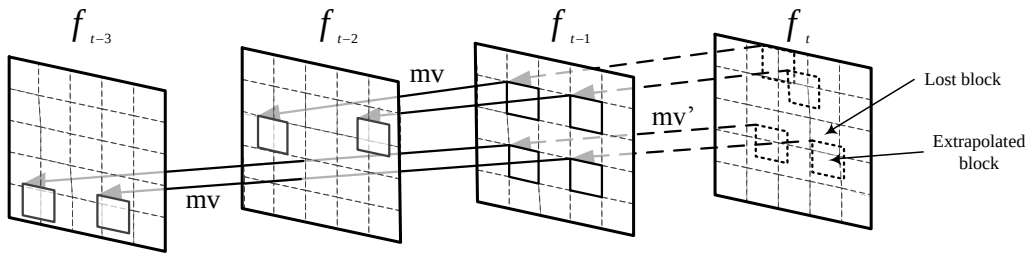


Figure 3.9: MVE extrapolation with the corresponding extrapolated macroblock, based on the method presented in [129].

crease the overlapped region to measure the distortion and to select the final MV, were tested. The experimental results revealed that the usage of a single layer match gives better performance than multiple layers, while selective search gives excellent trade-off between the image recovery and computation complexity

Motion vector extrapolation

As mentioned before, the MC method is a simple way of concealment a missing frame using motion information from previously decoded frames. Although this method can provide an acceptable performance comparing with the FC, as shown in [124], more complex algorithms that use the assumption of motion continuity across consecutive frames are proposed to conceal, not only missing blocks, but also whole frames in compressed video.

MVE [129, 130] methods are a simple but efficient way to reconstruct the MVs of a lost frame, and they are able to achieve good performance. In the method presented in [129], MVs are extrapolated from the last decoded frame (reference frame) onto the missing one. Firstly, each MB in the reference frame is projected to the missing one with its own direction given by the corresponding extrapolated motion vector (mv'). An example of motion vector extrapolation is shown in Figure 3.9, where the vectors of the frame $t-1$ are projected onto the missing frame at instant t. Note that, in the reference frame (f_{t-1}) the motion vectors may point to different frames (f_{t-2} and f_{t-3}), previously decoded, so MV extrapolation must take into account the difference between the original temporal distance and the projected one. Secondly, to improve the accuracy of MVE, the damaged region is divided into 4 blocks (8×8 pixels). Then, according to the matching area between the extrapolated block and the damaged 8×8 region, the best MV is chosen. In order words, the MV for the each block is obtained from the extrapolated block that achieves a higher number of overlapped pixels with the missing 8×8 region. Finally, the reconstructed MVs are used to fetch the missing pixels from the previously decoded frame (f_{t-1}). Although this method is able to overcome the disadvantage of incorrect MB displacement, the choice of the best MV is not very

effective [131]. In [129] the MVE is used to extrapolate the MB of the reference frame. This results in overlapped regions, which are recovered using a weighted average of the multiple pixels candidates. The weight coefficients assume a maximum value in the center of the extrapolated region, and a minimum value at the boundaries.

Since the choice of the best candidate MV relies on the overlapped area of the extrapolated block with the lost region, the method may give erroneously estimated MV in large motion scenes. The example of Figure 3.10 shows a lost 8×8 block (in the centre of the figure) and four possible extrapolated PUs that can be used to estimate lost the motion information. As illustrated in the figure, although the largest overlapped area belongs to the PU_2, the overlapped pixels correspond to a small portion of the whole missing block. Moreover, the majority of the lost block's area was not covered by any extrapolated PU. Thus, motion recovery using only the information available in PU_2 will not properly characterise the motion information for all lost pixels.

To overcome the previous limitation and to achieve smoother and finer results, a Pixel-based Motion Vector Extrapolation (PMVE) algorithm was proposed in [131]. This motion vector extrapolation method is divided into two cases:

1. The pixels in the lost region that are covered by at least one extrapolated block, such as circle points in Figure 3.10, are recovered using the average of the MVs of the overlapped block.

2. Those pixels which are not covered by any extrapolation, such as the triangle point in Figure 3.10, the MV of the co-located pixel in the previous frame is used.

Then, if the estimated MV is (MV_x, MV_y), then the estimation of the missing pixel, $p_m(x, y)$, is obtained by:

$$p_m(x, y) = p_r(x + MV_x, y + MV_y), \tag{3.10}$$

where p_r refers to the pixels in the previous frame. This pixel-based MVE provides similar results than the block-based MVE algorithm in low motion scenarios. Nevertheless, it is able to overcome the previous method in large motion scenes. An extension to the previous approach was considered in [131], by using the MVs of subsequent available frames. Thus, to estimate the lost motion information, forward and backward extrapolations are performed, in order to obtain two candidates to each lost pixel. Then, the lost pixel is estimated using the Equation 3.10, using the MV determined by the average of two MV candidates. The experimental results revealed that the bi-directional PMVE outperforms the isolated use of forward or backward PMVE. Moreover, the bi-directional PMVE achieves gains up to 1.85 dB and 1.32 dB (PSNR) comparing with the work in [132] and MVE [129].

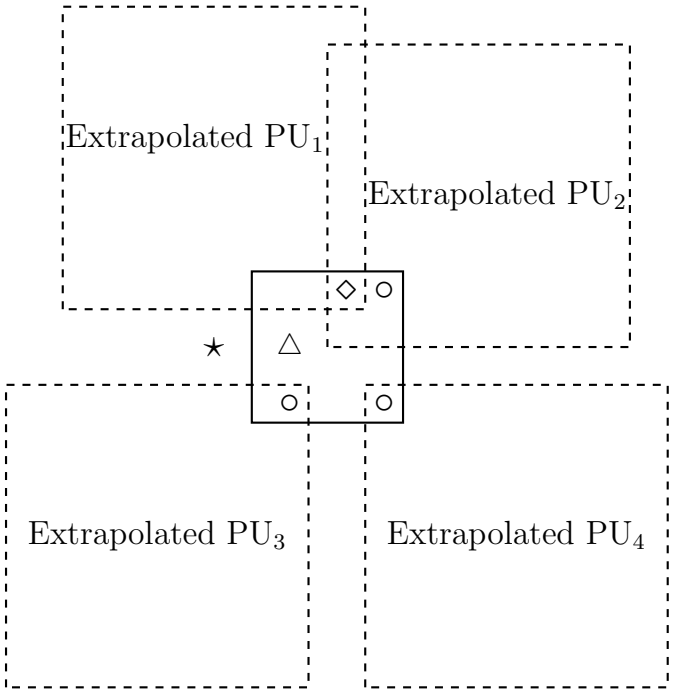


Figure 3.10: Pixel based motion vector extrapolation [131].

In [133] a Hybrid Motion Vector Extrapolation (HMVE) algorithm was proposed, to provide more accuracy to the estimated motion information. The method aims to improve the accuracy of the MVs of the PMVE method, by using the extrapolated vectors at the pixel and block level. Moreover, the proposed method discards the MV that were wrongly extrapolated in order to obtain an accurate MV. In this concealment method, firstly a pixel based motion extrapolation is performed and the pixels is organised into three groups. Group A contains the pixels that are overlapped by at least one extrapolated block (both circle and diamond points in Figure 3.10). Group B includes the pixels that are not covered by any extrapolated block, but belong to a block that is covered, *e.g.*, in Figure 3.10 the triangle point was not covered by any extrapolated MB but it belongs to 8×8 block that is covered by 4 blocks. The remaining pixels are included in group C (star pixels in the figure). Then, through the block-based MVE method described in [129] two new MVs are determined: the first one corresponds to the MV of the most overlapped block with the current one, and second one is obtained by the weighted average value of the MVs corresponding to different overlapped blocks.

The HMVE method aims to find the more accurate MV from the MV set composed described above. To achieve this, the MV set is further refined using the distance between each pair of MVs, in order to discard the wrongly extrapolated MVs. Finally, the pixels of the lost regions (belonging to the groups A and B) are obtained through motion compensation, using the average of the remaining vectors in the MV set. The pixels of the group C, although the MV set is empty, the same process is applied

using the MV of the co-located pixel in the previous frame. This algorithm is able to outperform the PMVE for different packet loss probabilities, as shown by the results in [133]. The method is used to conceal both I (INTRA) and P frames in video sequence compressed with the H.264/AVC, achieving more 1.09 dB (PSNR) in the erroneous frames.

In the case of HEVC, there are also a few works worth to be mentioned. The works presented for the HEVC standard are mainly based on temporal EC approaches. In [23] the block partitions of the neighbouring frames were used to improve the performance motion vector extrapolation. Using this the reconstructed frame keeps the boundaries smooth, leading to higher objective and visual quality. Moreover, in order to select only the reliable motion information, the residue information can be exploited, as shown in [134].

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.

3.6 Concealment-aware error resilience

The last category of techniques for robust video transmission address those algorithms that help the EC process at the decoder-side. Although such techniques might overlap with concepts described above, such as, data partition or redundant information, they have the main goal of improving the EC performance. This group of algorithms combine the error resilience capabilities with EC methods to achieve an improved overall performance.

One of the approaches that can be used to improve the EC performance is to estimate, at the encoding time, the EC distortion introduced by the decoder in case of data loss. In [47] the EC distortion estimation is used for slice classification and UEP. Such distortion is also used to optimise motion estimation [48], by including the distortion of the subsequent frame (assuming it is lost and recovered) in the R-D cost optimisation of the current frame. This was also used in the context of the HEVC to improve its error robustness as shown in [45]. This kinda of approaches take advantage of location techniques and data partitioning to improve the reconstruction quality achieved by the EC algorithms. Normally, they use a single EC method that is implemented both at the encoder and decoder, based on temporal algorithms, e.g., MC and MVE.

Another approach is the use EC motion vectors which are MVs that might be carried by intra-coded units for the purpose of concealing errors. When an error occurs in a given unit, the concealment motion vector from the neighbouring units can be used to

form a prediction from previously received frames. This concept was exploited in [49] by using data-hiding to carry additional motion information in the DCT coefficients. This technique not only improves the reconstruction quality but also shift the algorithm complexity from the decoder to the encoder-side.

A different approach to improve the EC performance was presented in [43]. Pixel interlacing is used to divide the frame into different packets in order to enhance the performance of intra-frame EC using linear interpolation. This technique reduces the probabilities of burst losses, therefore, the neighbouring pixels surrounding the lost regions are available to be used as reference for the concealment operation. 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. In [135] 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 [136] by using asymmetric coding of the redundant information based on Itti's attention model [137].

The work presented in [50,138] proposes testing several methods with transmission of extra symbols to indicate the best method to be applied by the decoder, using 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 [139], in order to select the best frame to be used to replace the missing one.

Summarising, EC-aware error techniques allow the combination of efficient algorithms, e.g., redundant pictures or error concealment-aware motion estimation or embedded information, with algorithms at the decoder which take advantages of such extra information. Such techniques are able to achieve quality improvements in comparison with the application of error recovery approaches a the decoder on its own. Although effective techniques have been proposed in the past, the exploration of dynamic techniques which explore the potential of different error concealment techniques with adaptable decoder have not been fully exploited.

3.7 Summary

This chapter presented an overview of the different error resilience techniques proposed over the last years, based on they key features. Different approaches were described covering techniques for robust video transmission from dependency reduction at the

encoder-side to error concealment by post-processing at the decoder.

The techniques described in this chapter provide relevant insights for the methods proposed in this work. In order to achieve increase the error robustness of the HEVC bitstrems and improved reconstruction quality in case of errors the different algorithms presented before are considered and combined. In Chapter 4 localisation techniques are used to decrease motion vector dependencies in HEVC standard achieving higher quality at cost of small coding efficiency loss. Chapter 5 takes advantage of reference picture selection scheme in combination with motion vector redundancies to improve HEVC streaming. Chapter 6 proposes an error-concealment aware for increase reconstruction quality. Finally, Chapter 7 proposes an end-to-end robust video transmission scheme by combining the error resilience and error concealment based on content, network conditions and application.

Chapter 4

A Method to Increase the Error Robustness of HEVC

4.1 Introduction

The HEVC standard is expected be used by the same range of applications as its predecessors (*e.g.*, H.264/AVC), such as, broadcasting, internet streaming, two-way communications, and storage [15]. More advanced prediction structures, as presented in Section 2, allow bitrate savings, but also to some disadvantages, such as increased complexity [10] and reduced error robustness. Therefore, it is important to evaluate the error robustness of the current standard in order to check for its weaknesses. This can provide relevant insights on future directions on how to improve the standard robustness to transmission losses.

The results provided by this experimental study allowed to improve the HEVC encoder in order to avoid ...

In this chapter an error resilience evaluation of the HEVC is presented in Section 4.2. The evaluation covers different coding configuration to evaluate the individual impact of some of the new coding features introduced, e.g., increased partition size, intra-refreshing, slice partitioning and advanced motion vector prediction. In Section 4.3 a novel method to increase the error robustness of the HEVC MV coding is presented and evaluated against existing techniques. Finally, Section 4.4 concludes this chapter.

4.2 Evaluation study of the error robustness of HEVC

4.2.1 Motivation and objectives

As mentioned before the HEVC standard provide improvements in terms of coding efficiency, however at the cost of more advanced predictions structures, which increases complexity and bitstream dependencies. 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.

A 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 [14], MPEG-2 TS [15], or more recent standards as MPEG-DASH [16] and MPEG Media Transport [18], 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 and streaming technologies. This is the main focus of the study presented in this chapter.

The are some studies presented in the past regarding the error robustness of the HEVC standard which are worth to be mention as they provide relevant insights for the research of carried out in this research work. In [140] the both the coding efficiency and error robustness of the HEVC are evaluated, and compared against the H.264/AVC. Experimental results show that HEVC has reduced error robustness despite its increased coding efficiency. A subjective analysis is presented in [11, 12] revealing that packet loss ratios higher than 3% have a significant effect on the perceived video quality. In both works it is concluded that the robustness of HEVC decreases for high motion sequences. Moreover, in [141] a the vulnerability of MV prediction is presented and its impact of the error robustness is evaluated, showing its weak error resilience in contrast with the coding improvements. Summarising, the decreasing in error robustness 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 optimisation criteria, assuming error-free transmission [13].

This section presents an error robustness evaluation of the HEVC standard. This study focus on the novel techniques introduced extending the generic studies presented in previous works mentioned above. This study may provide relevant insights for future

directions where the error robustness problem should be tackled in the HEVC standard. The following cases studies are covered in this experimental evaluation:

1. different maximum CU sizes to evaluate the impact of higher larger frame partitioning;
2. results for difference distances between intra frames, i.e., evaluate the impact of introducing different amount of refresh points to the overall quality;
3. different slice partitioning schemes by varying the number of bytes per slice;
4. different MV coding schemes using different types of MV candidates, with an analysis at the spatial and temporal level.

4.2.2 Characterisation of the experiments

The experimental setup used in this work to evaluate the error robustness of the HEVC standard used thirteen test sequences, which were chosen to cover different types of motion and texture complexity.

Table 4.1 presents a summary of the main characteristics of the video sequences. Appendix A shows a list of the first frame of each test sequence. In order to evaluation the spatial and temporal complexity of the video sequences the following metrics [142] were used:

- Spatial Information (SI)) based on Sobel filter. Each video frame (f_t) is firstly filtered, then the standard deviation is computed. Finally, the maximum value over time is chosen to represent the SI metric. This process can be represented in the following equation:

$$SI = max_{time}\Big\{ std_{space}\Big(Sobel\,(f_t)\Big)\Big\}. \tag{4.1}$$

- Temporal Information (TI) based upon the motion difference feature, which is the difference between the pixel values at the same location in space but in consecutive time instances. TI is computed as the maximum over time of the standard deviation over space. More motion in adjacent frames will result in higher values of TI. This process can be represented by the following:

$$TI = max_{time}\Big\{ std_{space}\Big(f_t(x,y) - f_{t-1}(x,y)\Big)\Big\}. \tag{4.2}$$

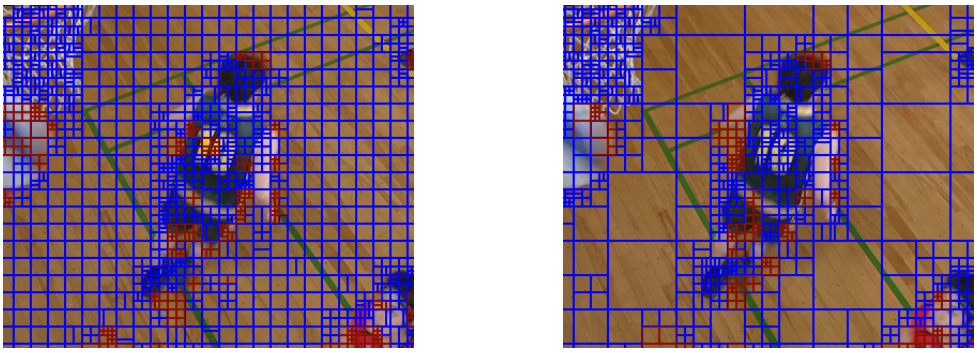
The values of the metrics described above are represented in the Table 4.1 (see second column).

Table 4.1: 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
Bosphorus $3840 \times 2160@120Hz$	13.4 / 3.8	12.00	Boat shipping at low motion with moderate complex background
BQSquare $416 \times 240@60Hz$	63.2 / 11.5	4.00	Moderate outside motion with moving camera capturing from high point
Cactus $1920 \times 1080@50Hz$	30.5 / 11.4	20.00	Several objects with high details with moderate motion
Four People $1280 \times 720@60Hz$	31.3 / 6.90	2.00	Four people talking of the desk and passing objects to each other
Jockey $3840 \times 2160@120Hz$	11.5 / 16.2	12.00	High motion with one horse rider
Kendo $1024 \times 768@30Hz$	19.6 / 16.1	0.60	Moderate motion with two moving persons, and moving camera
Kimono $1920 \times 1080@24Hz$	23.4 / 32.5	8.00	Capture of a women moving in a forest with moderate motion
Kristen and Sara $1280 \times 720@60Hz$	25.6 / 6.30	15.00	Two persons talking to each other 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
Traffic $2560 \times 1600@30Hz$	28.4 / 22.3	6.00	Elevated camera capturing highway with several cars with moderate speed

The experimental results for this study were obtained by using the HM reference software, version 16.2. The sequences were encoded with an IDR period of 16 frames and all the coding modes enabled, following two recommended configurations: Low-Delay (P-frames with 4 references frames) and Random-Access (B-frames with 2 reference frames in each list) [143]. In order to achieve high quality images for all coded sequences (i.e., PSNR around 40 dB) different bitrates were used.

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. 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 for PLR of 1%, 3%, 5%, 10%. For each test condition, 50 runs were performed and the average PSNR was measured. The filtering operations were disabled at slices boundaries in order to keep the slices self-contained. The remaining



(a) 16×16 (b) 64×64

Figure 4.1: Illustration of the block partition and prediction mode for a region of the Basketball Drill sequence.

configurations have been kept at their default values as given in the common test conditions [143].

4.2.3 Evaluation of the impact of the partition units size

The HEVC standard introduced new partitioning schemes to allow coding improvements in high resolution videos. These partition structures, named Coding Unit (CU), are allowed to be as large as 64×64 pixels. In the first set of experiments the impact of the maximum CU size in the decoded video quality is investigated. In this experiment the maximum CU size is set to three different sizes, i.e., 16×16, 32×32 and 64×64. The lower size is targeting the MB size used in the H.264/AVC.

Table 4.2 shows the average PSNR obtained for the three condition tested, for error-free and 4 different packet loss ratios. Results for the error-free scenario shows a consistent quality increasing as the maximum CU size increases, to a maximum of 1.41 dB for the Tennis sequence when comparing the maximum CU size of 64×64 with 16×16. This confirms the principle used by the HEVC standard, that larger coding units lead to higher coding efficiency. In case of data loss, results indicate that larger coding units should be maintained. Even though larger frame partitions lead to larger coding dependencies, it will not affect the robustness to packet loss and error propagation. This is mainly due to the fact that error propagation occurs mostly in complex and moving regions, which normally are coded using smaller partition units.

Figure 4.1 illustrates an example of the frame partitions for a region of the Basketball Drill sequence for two cases: (i) maximum CU size of 64×64 and (i) maximum CU size of 16×16. As shown, in the moving regions, e.g., moving player, the partitions obtained for both cases have a similar pattern, mainly composed of CUs with small size. The differences in the size of the CUs are only noticeable in the background, which is less prone to errors. Therefore, the overall error robustness is not affected by reducing

Table 4.2: Average PSNR for different maximum CU size.

Sequence	Maximum CU size	Error free	Packet loss rate 1%	5%	10%
Basketball Drill	16×16	39.25	37.82	34.66	32.67
	32×32	+0.47	+0.38	+0.66	+0.63
	64×64	+0.65	+0.64	+0.87	+1.01
Book Arrival	16×16	39.85	38.11	33.63	30.60
	32×32	+0.25	+0.19	+0.21	-0.38
	64×64	+0.43	+0.71	+1.06	+0.75
BQSquare	16×16	38.99	38.11	35.04	32.75
	32×32	+0.22	+0.17	+0.33	+0.34
	64×64	+0.31	+0.37	+0.74	+0.88
Cactus	16×16	37.50	37.19	36.72	36.45
	32×32	+0.39	+0.38	+0.41	+0.39
	64×64	+0.50	+0.52	+0.55	+0.48
Four People	16×16	39.54	39.26	38.42	37.54
	32×32	+0.44	+0.43	+0.38	+0.34
	64×64	+0.59	+0.60	+0.59	+0.57
Kendo	16×16	39.76	38.47	33.90	29.57
	32×32	+0.99	+1.04	+0.74	+1.01
	64×64	+1.39	+1.34	+1.00	+1.14
Kimono	16×16	42.42	41.69	40.64	40.13
	32×32	+0.36	+0.39	+0.32	+0.33
	64×64	+0.46	+0.56	+0.47	+0.51
Park Scene	16×16	39.03	38.49	37.64	37.18
	32×32	+0.33	+0.26	+0.38	+0.30
	64×64	+0.44	+0.41	+0.43	+0.38
People On Street	16×16	35.20	34.05	32.66	31.83
	32×32	+0.37	+0.31	+0.34	+0.26
	64×64	+0.70	+0.69	+0.57	+0.57
Race Horses	16×16	39.19	37.92	36.33	35.56
	32×32	+0.51	+0.55	+0.49	+0.47
	64×64	+0.68	+0.81	+0.79	+0.66
Tennis	16×16	39.27	37.91	35.39	34.04
	32×32	+0.95	+0.83	+0.73	+0.58
	64×64	+1.41	+1.33	+1.15	+1.02
Average Difference	**32×32**	**+0.48**	**+0.45**	**+0.45**	**+0.39**
	64×64	**+0.69**	**+0.73**	**+0.75**	**+0.72**

the maximum CU size. One may conclude that the use of larger coding units is able to improve the overall HEVC efficiency in error-free and error prone applications.

4.2.4 Evaluation of the impact of intra refreshing

In this section the relation between the distance of subsequent intra-coded frames, i.e., period of the intra frames, and the quality under errors is evaluated. As the intra-coded frames have no dependencies to prior coded pictures, they introduce a refresh point into the bitstream. This allows the decoded to recover from slice loss and continue a error-free decoding.

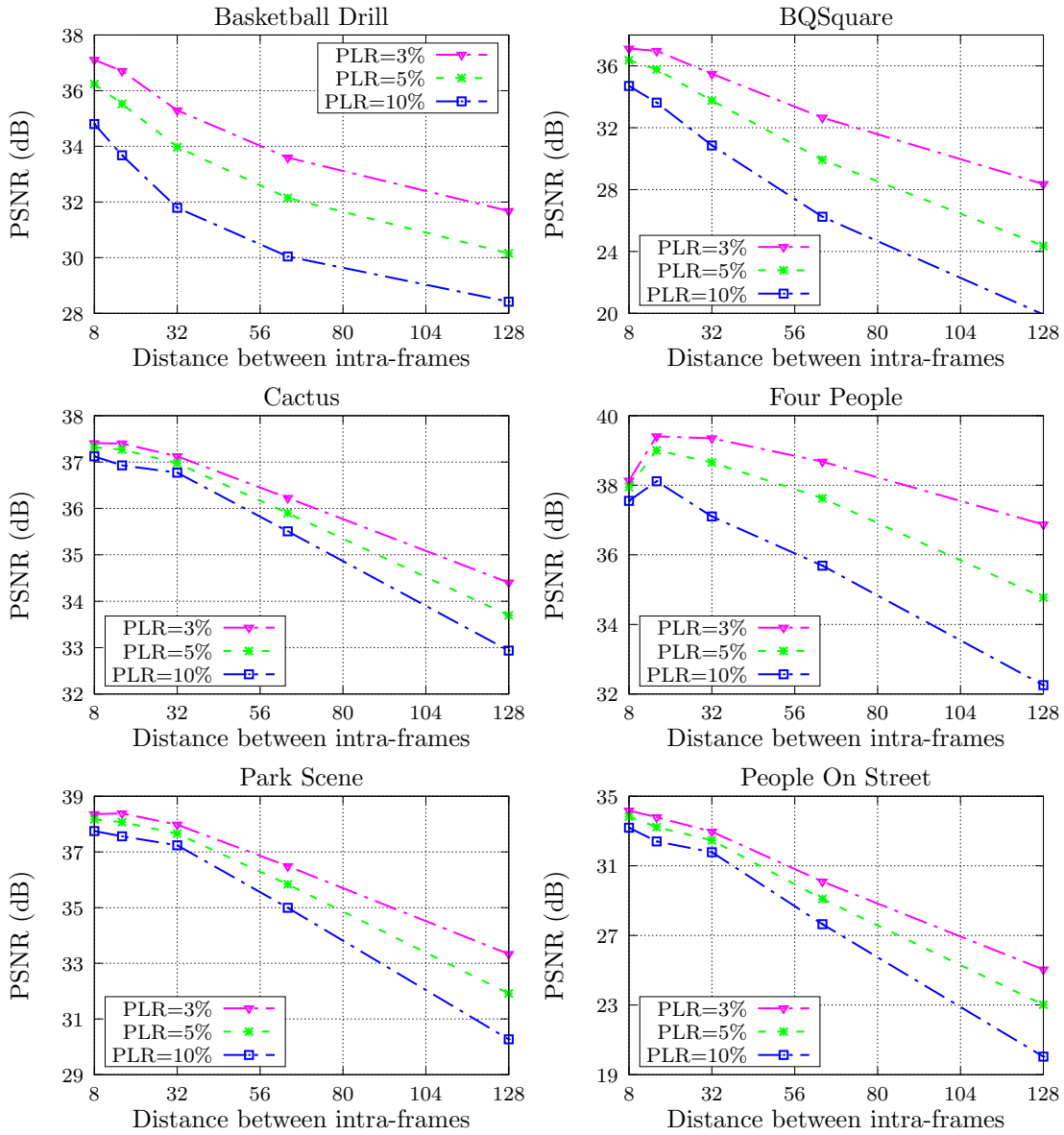


Figure 4.2: Decoded video quality for different intra periods, i.e., distance between intra-frames.

Results on Figure 4.2 shows the average decoded video quality for different distance of intra frames. The figure covers packet loss ratios between 3% and 5%. Results indicate that has the distance between intra frames increases the decoded video quality significantly decreases. This is especially noticeable for an intra period higher than 32 frames, indicating that it is not suitable for video streaming under error-prone conditions.

One should also be noted from the results of Figure 4.2 that using a small distance between intra frames is not always more efficient. Since intra frames require higher amount of bytes, the overall coding efficiency is reduced, resulting in an overall quality decreasing even with a 5% f PLR. This is clear noticeable for the Four People sequence.

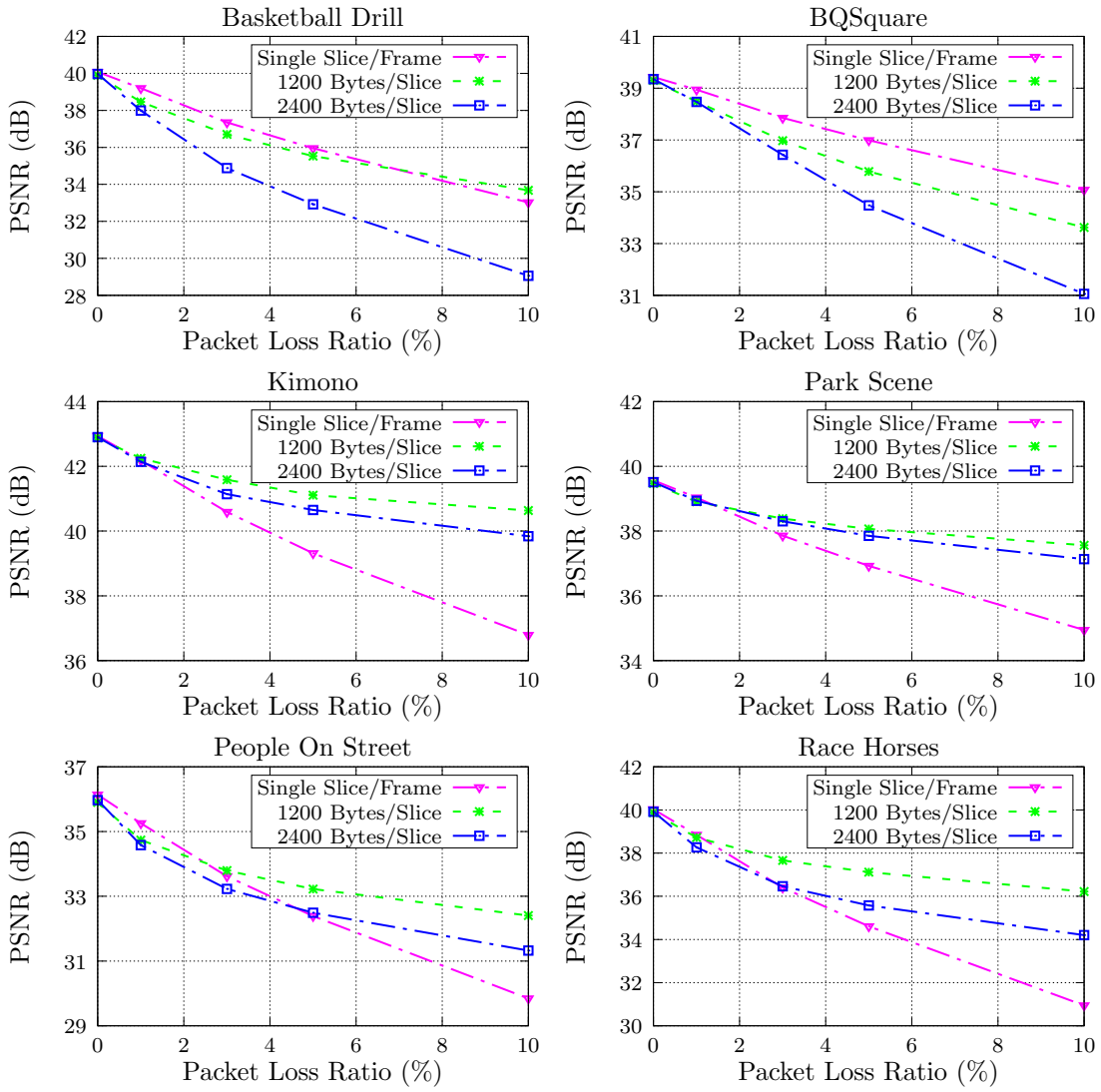


Figure 4.3: Decoded video quality for different slice partitioning schemes.

From the results for Park Scene sequence (and Cactus sequence in Figure **??**) it can be concluded that using sixteen frames between refresh points (i.e., intra frames) achieves an acceptable efficiency under errors, without compromising the coding efficiency.

4.2.5 Evaluation for using different slice partitioning schemes

The partitioning of the CTUs into different slices introduce flexibility into video bitstream, by introducing spatial refresh points and avoiding the error propagation across different CTUs. In this section the amount of bytes per slices is evaluated, since it the most efficient partitioning scheme in the HEVC standard that suits the network requirements of fixed packaging. As a reference for comparison a single slice per frame is also evaluated.

Figure 4.3 shows the average PSNR for three difference slice partitioning schemes.

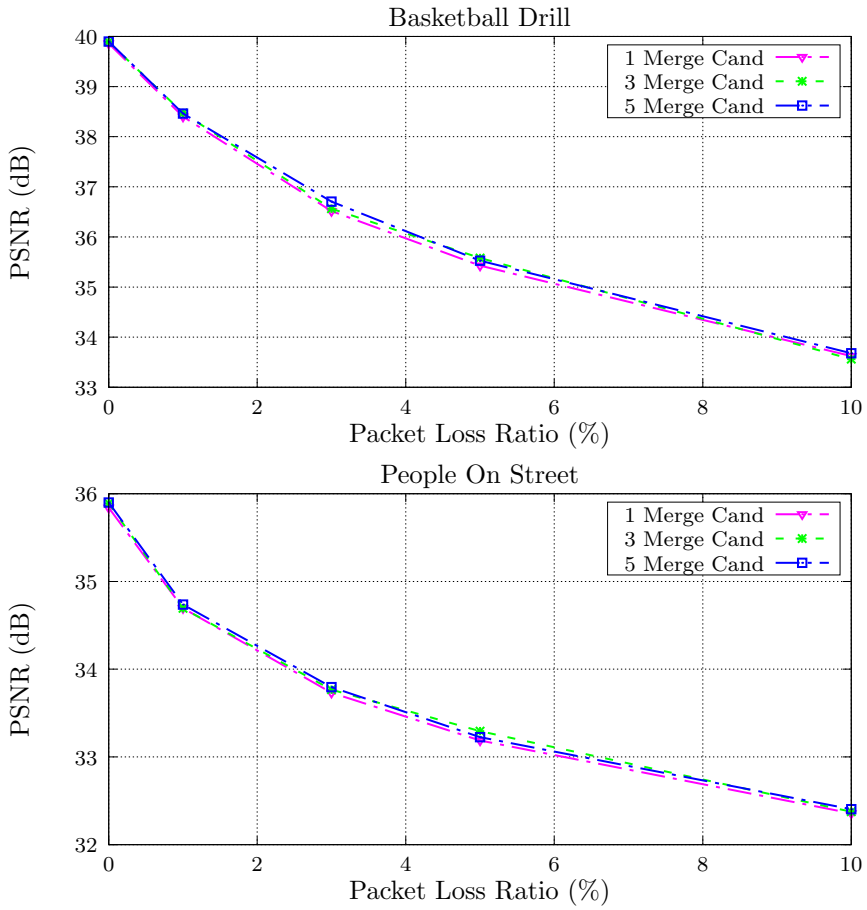


Figure 4.4: Average quality using different number of MV candidates.

Results reveals a clear impact of the slice partitioning on the error robustness of the HEVC, as the quality degradation as the PLR increases varies for each scheme. Firstly it is clearly notice the disadvantage of using a single slice per frame as it leads to the lower quality when comparing the test cases studied. This is due to the lack of refresh points within a frame, resulting in loss events that affects the whole frame which will propagate to the subsequent inter predicted frames. Secondly, comparing slices of different sizes, i.e., comparing the use of 1200 and 2400 bytes per slice, reveals that using small slices leads to higher quality which is able to overcome the inherent degradation on coding efficiency due to higher amount of slice headers.

4.2.6 Motion vector coding impact on error propagation

In HEVC, new motion vector prediction modes are introduced, such as the AMVP and Merge Mode [144]. In this section the the error robustness of the Merge Mode is evaluated, which predicts MVs for the current block from its neighbours, which obviously leads to prediction mismatches at the decoder side in case of packet loss.

Table 4.3: Relation between the number of available MV candidates and its usage

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

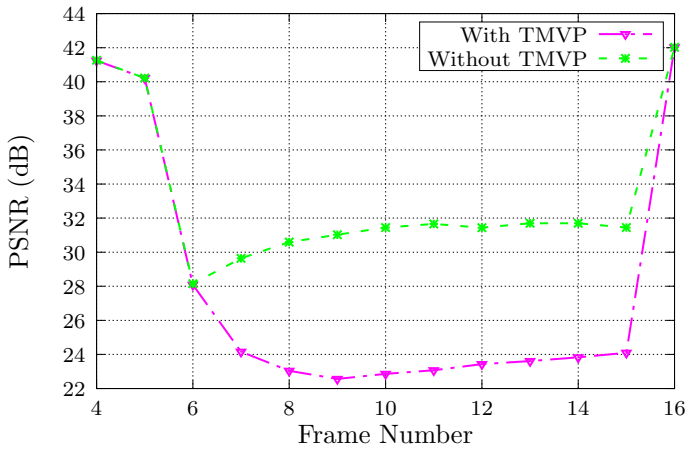


Figure 4.5: Error propagation for Kendo sequence when Frame #6 is lost for the Low-Delay configuration.

Spatial MV candidates

The first set of experiments aimed to evaluate the impact of using several MV candidates in the Merge Mode. As mentioned before, HEVC allows up to five candidates to be used, which may increase the amount of spatial dependencies. In order to check the impact of the MV candidates, different coding configurations were used, by reducing the number of available candidates. Figure 4.4 shows the average quality results for the Basketball Drill sequence under different packet loss conditions. The results show that, besides the amount of MV candidates, the error resilience of the HEVC is not compromised.

The occurrence rate for each spatial MV candidate was analysed in order to check if they are uniformly used. Table 4.3 shows the use of each MV candidate when three and five candidates are available. The results indicate that most predictions use the first candidate and minor differences occur by adding 2 or more MVs. Since the mostly used MVs are the initial ones, changing the number of available candidates does not significantly affects the error resilience. Moreover, the slice partitioning scheme used (*i.e.*, 1200 bytes per slice) adds more refresh points in each frame reducing the spatial error propagation due to mismatch MV predictions.

Table 4.4: PSNR under random packet loss.

Sequence	TMVP	Error Free	Packet loss rate 1%	5%	10%
Basketbal Drill	Enabled	38.69	34.73	27.92	25.41
	Disabled	-0.04	+1.97	+3.93	+3.43
Book Arrival	Enabled	40.62	36.78	30.78	27.29
	Disabled	-0.01	+1.98	+3.32	+3.33
BQSquare	Enabled	39.64	31.99	23.45	20.95
	Disabled	-0.05	+6.14	+10.6	+10.7
Kendo	Enabled	43.31	38.80	29.80	25.45
	Disabled	-0.04	+2.04	+4.51	+4.98
Park Scene	Enabled	36.89	33.78	28.09	25.77
	Disabled	-0.06	+2.33	+5.72	+6.05
Race Horses	Enabled	37.06	29.64	22.92	20.35
	Disabled	-0.07	+3.72	+4.78	+4.19
Tennis	Enabled	41.61	33.74	25.73	22.68
	Disabled	-0.04	+3.10	+4.01	+3.52
Average difference		**-0.04**	**+3.04**	**+5.32**	**+5.17**

Temporal MV candidates

The second set of experiments aimed to find the influence of losing temporal motion predictors on the error resilience of the HEVC standard. In order to perform this test, a single frame is lost, and the error propagation is evaluated for two configurations: with the temporal MV candidate enabled and disabled. Figure 4.5 illustrates the error propagation, when a single frame is lost. For this particular experiment a intra-refreshing period of 32 frames was used to evaluate a long error propagation period. The results show that when the Temporal Motion Vector Predictor (TMVP) mode is enabled, the loss of Frame #6 leads to significant reduction in the reconstruction quality of the subsequent frames. On the contrary, when the temporal MV candidate is disabled, motion information on subsequent frames is not affected by the loss of a single frame, since it is not temporal dependent, achieving significant quality increasing. This is more noticeable in the Low-Delay configuration since more error propagation occurs in P-frames and it is propagated for several frames until an intra frame is decoded.

In order to further evaluate the impact of the temporal MV dependencies, the HEVC streams were subject to random packet losses. Table 4.4 presents the average quality results, measured using the PSNR. As expected, the results indicate that in error free environment, HEVC achieves higher quality by enabling the TMVP candidate, since it presents higher compression efficiency due to more MV predictors in the Merge Mode. This is true for all common test conditions, *i.e.*, both Low Delay and Random Access. However, in the presence of errors, the quality significantly decreases due to low error robustness when temporal MV candidates are used, resulting in an average of 2.65 dB of PSNR increasing when temporal MV predictor is disabled across all

configuration tested. The results also show that the negative impact of using temporal MV dependencies decreases as the number of B-frames increases. The average gain of disabling the temporal MV predictor is 4.51 dB (Low Delay) and 0.78 dB (Random Access). This also confirms the previous results of Figure 4.5. A higher negative impact is obtained when only P-frames are used, since subsequent frames are temporally closer, leading to more accurate temporal MV candidates. Moreover, the results also indicate that for higher loss ratios the negative impact of using the TMVP does not increase as both configuration are subject to high quality degradation.

4.2.7 Discussion and conclusions

The results presented in the previous sub-sections show that HEVC has poor robustness against transmission errors, resulting in significant quality degradation, though dependent on the video content and coding options. As shown coding options as slice partition and intra refreshing period have high correlation with the decoded video quality under errors, which shall be taken into consideration when designing a broadcast application.

One of the main improvements of the HEVC standard is related with the high coding units which does not have an impact of its error robustness, so shall be configured to achieve the highest coding performance. Another improvement in the standard was the Merge Mode, which significantly improves the flexibility in coding and signalling the motion information. One the one hand the spatial MV candidates do not significantly decrease the error robustness. On the other hand, the introduction of temporal MV candidates significantly contribute to increase the temporal dependencies, leading to more severe quality degradation in the presence of slice loss.

Based on these evidences, 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.

4.3 Dynamic motion vector refreshing

The aforementioned results revealed a lack of error resilience in the HEVC standard motion prediction. Based on this evidence, this section proposes a method to make intelligent use of temporal MV candidates during the motion estimation process, in order to decrease the temporal dependency, and improve the error resiliency without penalising the R-D performance.

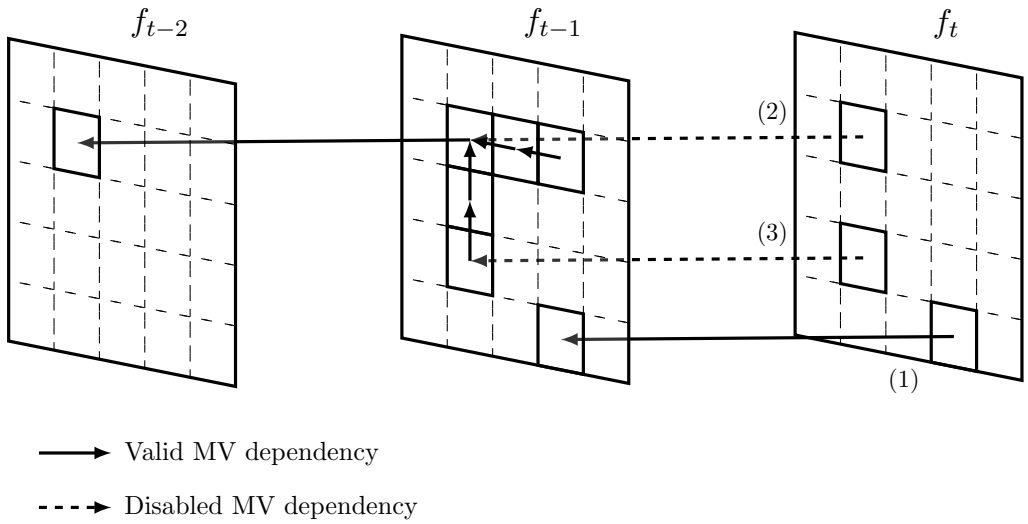


Figure 4.6: Proposed approach to use the temporal motion vector candidate in the HEVC encoder.

4.3.1 Proposed method

In this section a new approach to deal with temporal motion vector dependencies is proposed, aiming at reducing the amount of temporally dependent MVs. This is accomplished by modifying the rate distortion optimisation process in HEVC. Since the number of MVs that can be independently decoded increases, the error recovery percentage is improved after a frame loss event.

In order to overcome in the limitation of the method presented in [141], a new method that breaks the motion information dependencies between frames is proposed, which is used in the Merge Mode. Instead of disabling the temporal motion candidate at the frame level, the proposed method aims to remove the dependencies at the block (CU) level. To achieve this, the temporal dependency of all MV candidates is analysed for each block, in order to decide on the suitable candidates to be used to predict the current MV. Based on the analysis of the temporal dependency, some temporal candidates are marked as unsuitable for the prediction of the current MV, and are not selected in the merge modes. In the proposed approach, a temporal MV candidate is marked as unsuitable based on the following criteria:

1. if it is encoded based on another temporal MV candidate from a previous encoded frame;

2. if it is predicted using a spatial neighbour that was previously encoded using a temporal MV candidate.

Figure 4.6 can be used to explain the proposed method, illustrating different blocks and different MV dependencies represented with arrows. Note that the squares illus-

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

4.3.2 Performance evaluation

In this section the performance of the proposed scheme (*Proposed*) is evaluated against two other approaches: reference HEVC with TMVP enabled and the *TIDR* method as proposed in [141]. To evaluate the proposed method the same experimental setup as presented in Section 4.2 was used.

Error propagation under single loss events

Firstly, the performance of the proposed method was measured under a single loss event. Figure 4.7 shows the error propagation for various tested methods. The results show a significant quality difference between the proposed method, the *TIDR* method, and the reference HEVC. When the temporal MV candidate is disabled (proposed and *TIDR* method) more error robustness can be achieved. The results also show the relevance of using the proposed method, as the *TIDR* is only able to recover from the frame loss after the frame 8 (*i.e.*, the refresh frame).

Quality evaluation under random errors

Secondly, further tests were performed to evaluate the effectiveness of the proposed error resiliency method under realistic packet loss patterns. As in previous experiments, each coded frame is transmitted in one packet. Table 4.5 illustrates the average standard deviation of the PSNR, for three of the sequences tested. The results show that increasing the packet loss rate leads to higher quality deviations, and the impact of errors in the subjective quality experience is higher. However, reducing the temporal

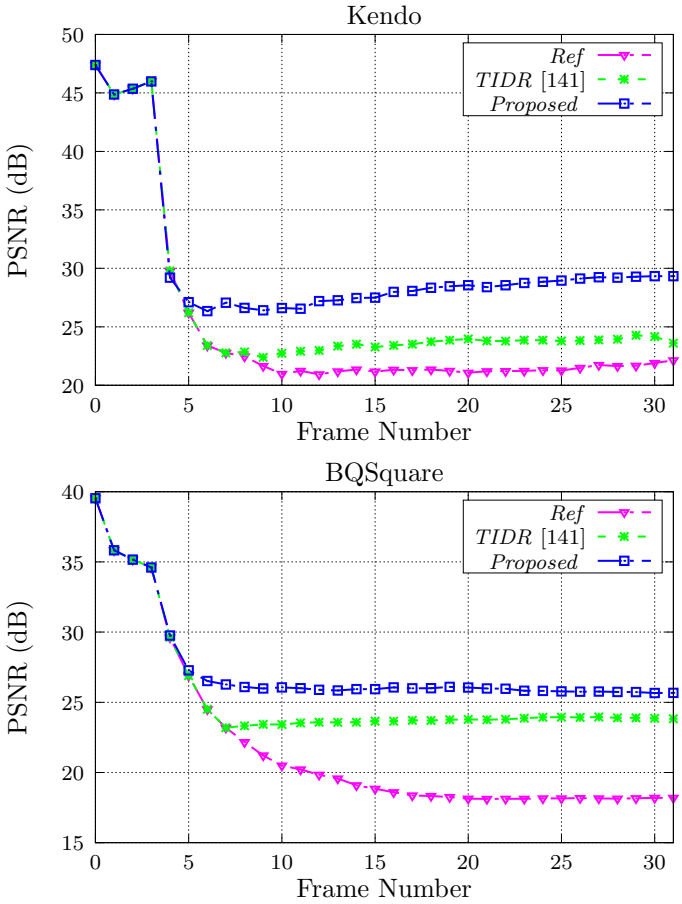


Figure 4.7: Quality obtained with the different approaches tested in the HEVC when the frame #5 is missing for Kendo and BQSquare sequences.

dependencies of the MVs using the proposed method decreases the quality differences variation, achieving the lowest value for all scenarios tested. The proposed method is able to reduce the quality difference variations up to 0.94 dB in Kendo sequence (3% of packet loss) compared to the *TIDR* method.

Table 4.6 shows the Bjontegaard's average PSNR increase [145]. The results are obtained taking as reference the standard HEVC encoder. The results presented in Table 4.6 show that the proposed method presents practically the same rate-distortion performance when compared to *TIDR* method. In general, the average quality, in an error-free scenario, is only decreased by 0.028 dB in comparison to the *TIDR* method and by 0.034 dB in comparison to the reference encoder. However, the proposed method widely outperforms the reference *TIDR* technique [141] in terms of error resiliency performance, achieving in average higher overall quality under error-prone conditions. The effectiveness of the proposed method increases at higher packet loss rates (*e.g.*, 5%), achieving an average gain of 0.573 dB, and a maximum gain of 1 dB (PSNR) for the Book Arrival sequence, when compared to the *TIDR* technique. Higher gains are achieved when compared with the reference HEVC encoder, as shown in Table 4.6.

Table 4.5: Average standard deviation of the decoded video quality (PSNR).

Sequence	**TMVP configuration**	**Packet loss rate**		
		1%	**3%**	**5%**
Basketball Drill	*Ref*	5.09	7.51	7.70
	TIDR [141]	5.00	7.30	7.52
	Proposed	4.78	7.03	7.24
Book Arrival	*Ref*	5.45	8.74	9.08
	TIDR [141]	5.24	8.48	8.88
	Proposed	4.84	7.83	8.26
BQSquare	*Ref*	5.98	8.86	9.14
	TIDR [141]	5.20	7.51	8.01
	Proposed	4.94	7.28	7.68
Kendo	*Ref*	7.56	11.58	12.05
	TIDR [141]	7.25	11.00	11.55
	Proposed	6.62	10.06	10.66
Race Horses	*Ref*	5.74	8.31	8.55
	TIDR [141]	5.47	7.78	8.07
	Proposed	5.29	7.63	7.88
Tennis	*Ref*	6.70	9.55	10.07
	TIDR [141]	6.55	9.25	9.83
	Proposed	6.42	9.11	9.67

Table 4.6: Bjontegaard's average PSNR (dB) differences for all sequences tested at different frame loss conditions.

Sequence	**TMVP configuration**	**Error free**	**1% loss**	**5% loss**
Basketball Drill	*TIDR* [141]	-0.011	0.092	0.365
	Proposed	-0.054	0.218	0.823
Book Arrival	*TIDR* [141]	-0.003	0.099	0.428
	Proposed	-0.012	0.359	1.436
BQSquare	*TIDR* [141]	-0.005	0.418	1.572
	Proposed	-0.032	0.539	1.967
Kendo	*TIDR* [141]	-0.008	0.150	0.618
	Proposed	-0.031	0.380	1.539
Race Horses	*TIDR* [141]	-0.010	0.179	0.689
	Proposed	-0.040	0.236	0.909
Tennis	*TIDR* [141]	-0.003	0.099	0.494
	Proposed	-0.041	0.214	0.932
Average	*TIDR* [141]	-0.006	0.171	0.695
	Proposed	-0.034	0.342	1.339

4.4 Summary

In this chapter, the error resilience performance of the HEVC standard was analysed under error prone conditions. Results revealed that new coding tool introduced in HEVC lead to higher error propagation. This reveals that new techniques should be employed within HEVC encoder and decoder to increase its resilience to errors, achieving higher overall performance under broadcast environments. Therefore, a new approach to deal with a drawback in MV coding was proposed, addressing MV dependencies at the block level. The proposed method selectively disables the temporal MV candidates, in order to break the dependencies across more than one frame. The experimental evaluation shows that the proposed method increases error robustness and outperforms the existing state-of-the-art method and the standard. The results also show that the proposed method is able to provide a good trade-off between the coding performance and error resilience. The results demonstrate that selective usage of temporal MV candidates is an efficient method to enhance error resilience in HEVC.

Chapter 5

A Two-stage Approach for Robust Coding and Streaming

5.1 Introduction

The increased compression ratios achieved by the HEVC standard lead to reduced robustness of coded streams, with increased susceptibility to network errors and consequent video quality degradation, as shown in Chapter 4. Therefore, this chapter proposes a two-stage approach method 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 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.

This chapter is organized as follows. Section 5.2 presents the problem formulation and motivation for the method proposed in this chapter. Section 5.3 provides an overview of the proposed method, describing the functional blocks that compose the two-stage approach for robust coding and streaming. Section 5.4 describes the encoding stage of the proposed method, that is used to selected the optimal reference picture for inter-prediction. Section 5.5 describes the streaming stage used to select a sub-set of MVs to reduce the mismatch motion predictions. Section 5.6 provides a performance evaluation of the proposed method, illustrating the advantages of each stage and how they are combined together to achieve superior error robustness in HEVC standard. Finally, Section 5.7 concludes this chapter.

5.2 Problem formulation

The error robustness characteristics of HEVC analysed in Chapter 4 confirmed that, in general, HEVC presents poor error resilience performance. Thus, HEVC bitstreams subject to transmission losses result in significant degradation of both objective and subjective quality. 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 [13]. Therefore, high compression efficiency is achieved, but resilience to transmission errors is penalised. This compromises the error robustness of video decoders which leads to poor performance under error conditions. Moreover, the use of advanced MV prediction in HEVC [144] 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.

As mentioned before the video decoder may rely of a distributed use of references frames to achieve higher video quality, by experienced lower error propagation and by optimising the error concealment algorithms to extract information from correctly decoded reference frames. In order to evaluate how the reference R-D optimisation presented in the reference HEVC encoder different test sequences were encoded using a configuration with two reference frames, referred to as r_1 and r_2. Figure 5.1 illustrates the average usage ratio of reference r_1 for different sequences. The remaining corresponds to those cases where reference r_2 were selected. Results in Figure 5.1 shows that the reference r_1 is selected in most of the cases when two references are available. This is mainly true for sequences with higher motion. For those cases with lower motion but high spatial complexity, e.g., BQSquare sequence (see TI in Table 4.1), both frames are selected for temporal prediction, with r_1 being used for 58% of cases and r_2 for 42%.

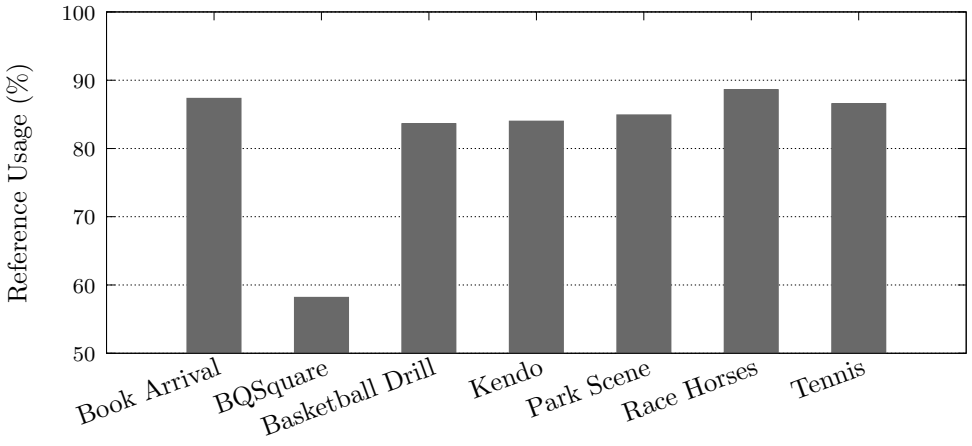


Figure 5.1: Average usage of the reference frame 1 (r_1) for different sequences.

This reveals that even a configuration with two references the encoder will mainly use the temporally closer one as the reference for inter-frame predictions. This will lead to a higher number of mismatch predictions in case of frame loss and consequently to higher error propagation.

As mentioned in Chapter 4, another technique that contributes to increased error propagation is the use of advanced MV prediction in HEVC [144]. This 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, as they propagate errors in the MV decoding to across different time instances.

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. In this chapter a two-fold approach is used to overcome the two problems mentioned above.

5.3 Proposed architecture

The goal of the proposed method is to reduce the error propagation in decoded video in case of frame loss. This is accomplished by constraining inter-frame prediction dependencies, in order to minimise the distortion D_t given by

$$D_t = \tilde{f}_t - \hat{f}_t, \text{when } \tilde{f}_{t-1} \neq \hat{f}_{t-1}, \tag{5.1}$$

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

Figure 5.2 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 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 ranking MVs according to a relevance criterion. Then, a sub-set of these MVs is selected to be transmitted as

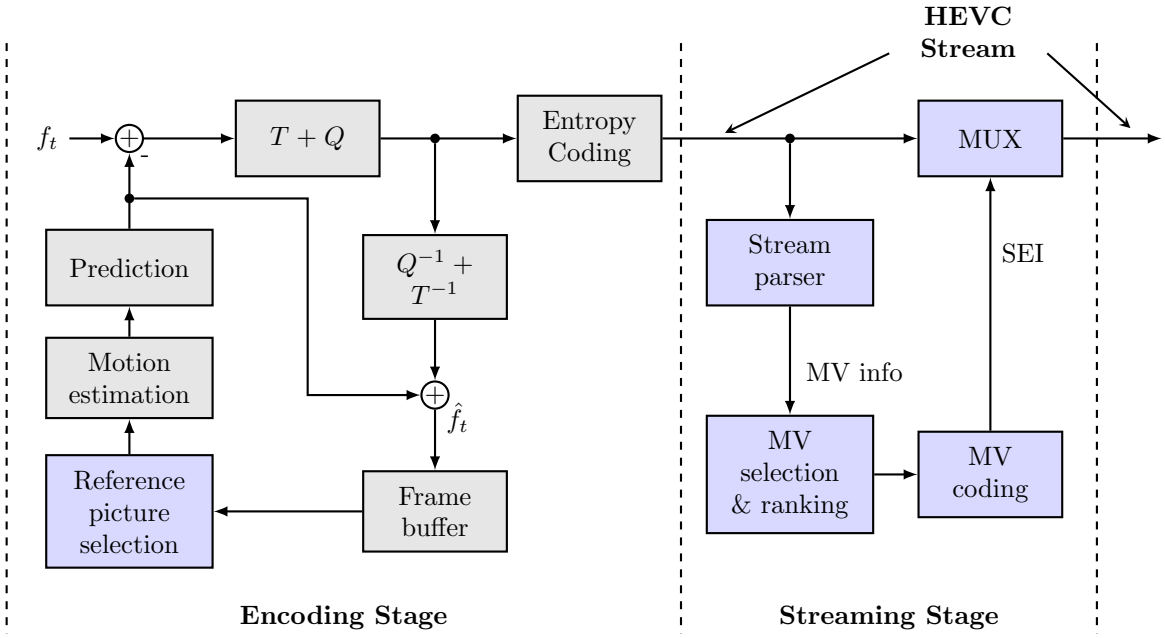


Figure 5.2: Functional blocks of the proposed method (*gray* blocks) in the context of the HEVC coding and streaming.

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 sections.

5.4 Encoding stage: Reference frame selection

The reference picture selection mechanism attempts to uniformly distribute the use of different reference pictures across the 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 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 5.3, 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 5.3. 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 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:

$$\sigma_C = std\Big(\mathbf{N}_C(r_i)\Big), i = 1, 2, \cdots S. \tag{5.2}$$

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

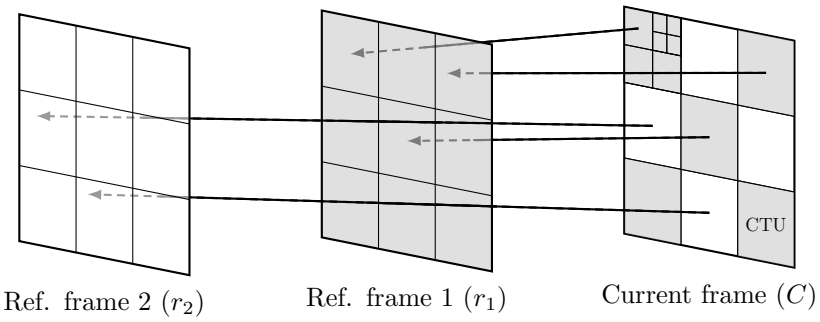


Figure 5.3: Example of the reference picture selection applied to the frame C based on the checkerboard at CTU level (PUs represented by small partitions).

in order to minimise the global standard deviation given by Equation (5.2). Thus, the optimum coding mode ϕ^*, selected from a set of possible coding modes M, is derived based on the following Lagrangian optimisation:

$$\phi^* = arg \min_{\phi \in M} J(\phi) \tag{5.3}$$

$$J(\phi) = \Big(D(\phi) + \lambda \times R(\phi) \Big) \times e^{W(r_\phi)} \tag{5.4}$$

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,

$$W(r_\phi) = \Big(\Delta_G(r_\phi) + \Delta_L(r_\phi) \Big) \times 2^{-T_{ID}} \times \gamma \tag{5.5}$$

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,

$$\Delta_\zeta(r_\phi) = \mathbf{N}_\zeta(r_\phi) - \frac{1}{S} \sum_{i=1}^{S} \mathbf{N}_\zeta(r_i), \;\; \zeta = G, L. \tag{5.6}$$

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

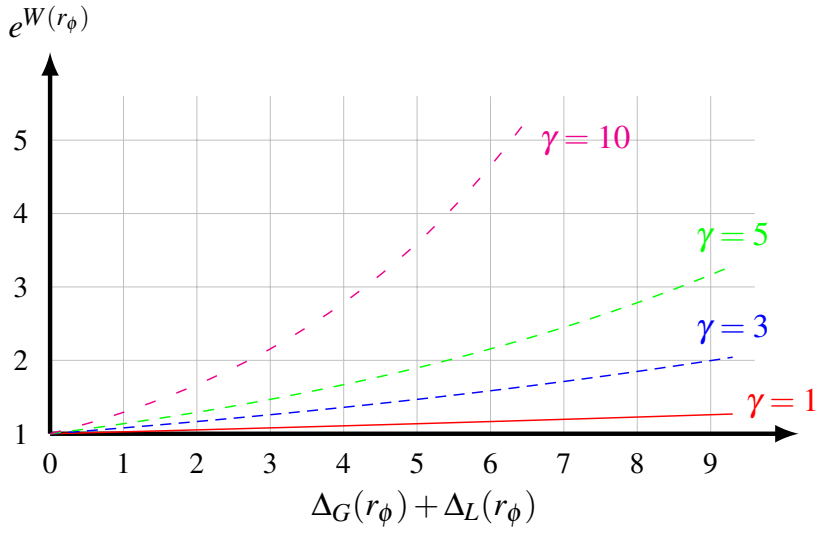


Figure 5.4: Exponential weight as a function of the global and local deviations

(5.6) is used to adjust the penalty weight defined by Equation (5.5) 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 Equation (5.4). Note that an exponential cost increase is used in Equation (5.4) in order to increase the penalty associated to any reference frame that tends to be used much more often than the average.

In Equation (5.5), the γ constant controls the slope of $W(r_\phi)$, leading to higher weights as the value of γ increases. On the one hand 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 5.4 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

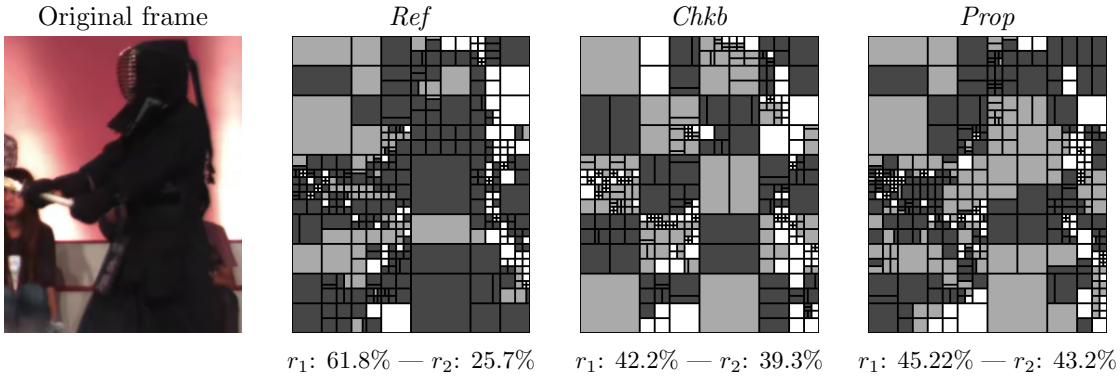


Figure 5.5: Example of the partitions and references pictures selected using the different encoding methods (the dark gray squares indicate that the MV is using the reference r_1, light gray squares corresponds to the reference r_2 and unfilled regions are intra-coded blocks).

50% ($e^{W(r_\phi)} \approx 1.5$).

Figure 5.5 illustrates the distribution of reference frames achieved by the proposed method (*Prop*) across different PUs. This result is compared against a fixed checkerboard approach (*Chkb*) and the reference HEVC (*Ref*). The dark-gray and light-gray squares indicate the reference frame in use, either r_1 or r_2, respectively. The white squares indicate intra-coded PUs. Below each method, the percentages of the image areas predicted by each reference frame are shown. As shown in this figure, the proposed method (*Prop*) is able to distribute the use of the reference frames closely to the *Chkb* method, i.e., r_2 is used as reference for approximately the same amount of PUs as r_1 (see use percentages of the predicted image). However, the predicted regions in the proposed method do not exhibit the same regular pattern as in *Chkb* method.

It is noticeable in the figure that for several CTUs the *Prop* method is able to keep the same partitions and reference frames when compared to the optimal encoding (*Ref*). This can be observed, for example in the small partitions on the left, where the *Prop* method uses the same reference frame as *Ref*.Thus, *Prop* is expected to achieve better coding efficiency than *Chkb*, since more optimal coding modes may be chosen.

Summarising, the underlying principle of Equations (5.4) and (5.5) 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.

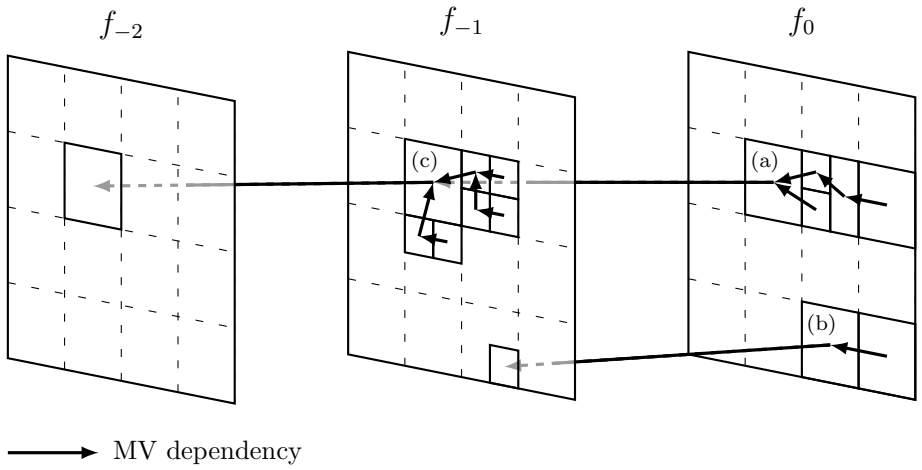


Figure 5.6: Example of temporal and spatial MV dependencies in HEVC.

5.5 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 5.6 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 5.6, 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: $V = \{MV\text{: } MV \textit{ is temporally dependent from others}\}$.

2. the $MVs \in V$ 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 in Equation (5.7).

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 Equation (5.7). Note that any value of n corresponds to a unique subset U.

$$n^* = \arg\max_{n \leq s} \left\{ \sum_{i=1}^{n} A_D(i) - \alpha \frac{t_C - t_I}{T_{IDR}} R_U \right\}, \tag{5.7}$$

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:

$$A_D(i) = \sum_{j=1}^{DV_i} (w_j \times h_j), \;\; i = 1, 2 \cdots s, \tag{5.8}$$

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 5.2 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 SEI NAL units [8]. 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 [146] was used for the redundant MVs, but any other entropy coding method can be used.

5.6 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 Section 5.4 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 Section 5.5, 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 [52], identified as "*Ref [52]*", 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 4.2. The evaluation of the encoding-stage is carried out separately from the streaming-stage.

5.6.1 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 5.4). Further results for a wide range of γ values are also presented.

Table 5.1: BD-PSNR in error-free conditions.

Sequence	**Low-Delay**			**Random-Access**		
	Ref [52]	*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**

Performance under error-free conditions

Table 5.1 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 [52] is tested for the Low-Delay configuration. The results in Table 5.1 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 [52], 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.

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 loss event in the Low-Delay configuration.

Figure 5.7 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 [52], revealing the higher efficiency achieving by distributing the use of the references frames.

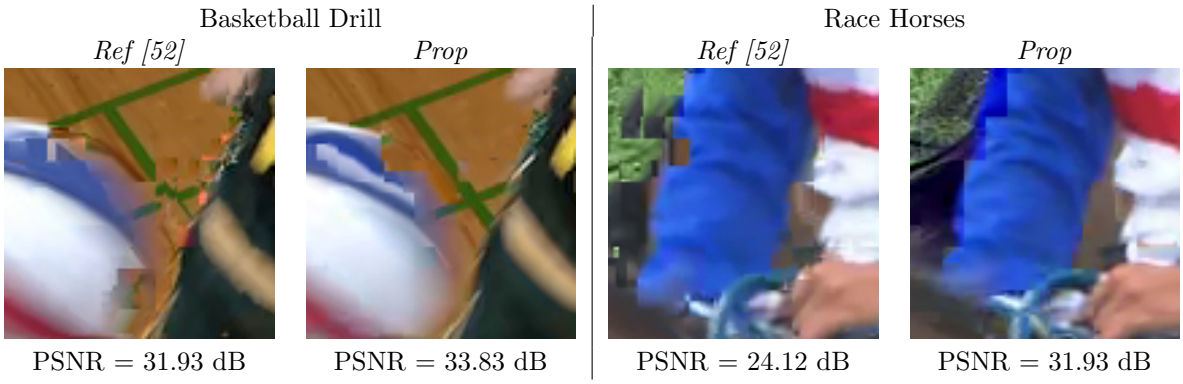


Figure 5.7: Zoom of recovered Frame #10 after a loss event at Frame #9 for Basketball Drill and Race Horses sequence.

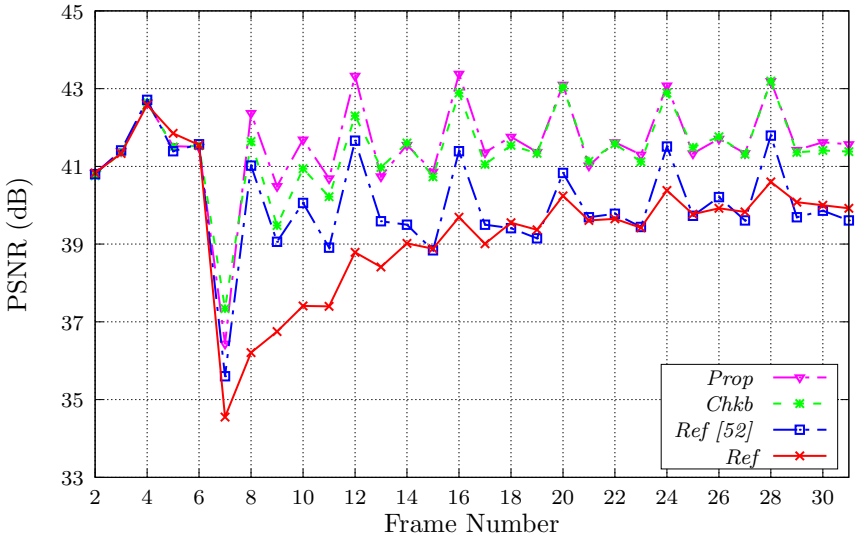


Figure 5.8: Error propagation when errors affect Frame #7 in Kendo sequence with an IDR period of 32 frames.

Figure 5.8 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 5.1), 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

Table 5.2: Average quality (PSNR): Low-Delay.

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 [52]	-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 [52]	-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 [52]	-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 [52]	-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 [52]	-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
Kendo	*Ref*	38.49	30.89	40.45	36.79	33.80	30.35
	Ref [52]	-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 [52]	-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 [52]	+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 [52]	-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 [52]	-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

average quality obtained across all trials for both coding configurations are shown in Table 5.2 and 5.3, 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 5.2), in comparison to the *Ref*, while the fixed checkerboard pattern approach (*Chkb*) only

Table 5.3: Average quality (PSNR): Random-Access.

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
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	+0.16	+0.12	+0.17	+0.51	+0.56	+0.58
	Prop	+1.46	+1.09	+0.65	+0.95	+0.93	+1.04

achieves up to 1.33 dB on average. In [52], 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 [52]*").

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 5.3), which inherently limits the error propagation. In comparison to the refer-

ence 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 4.1, 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.

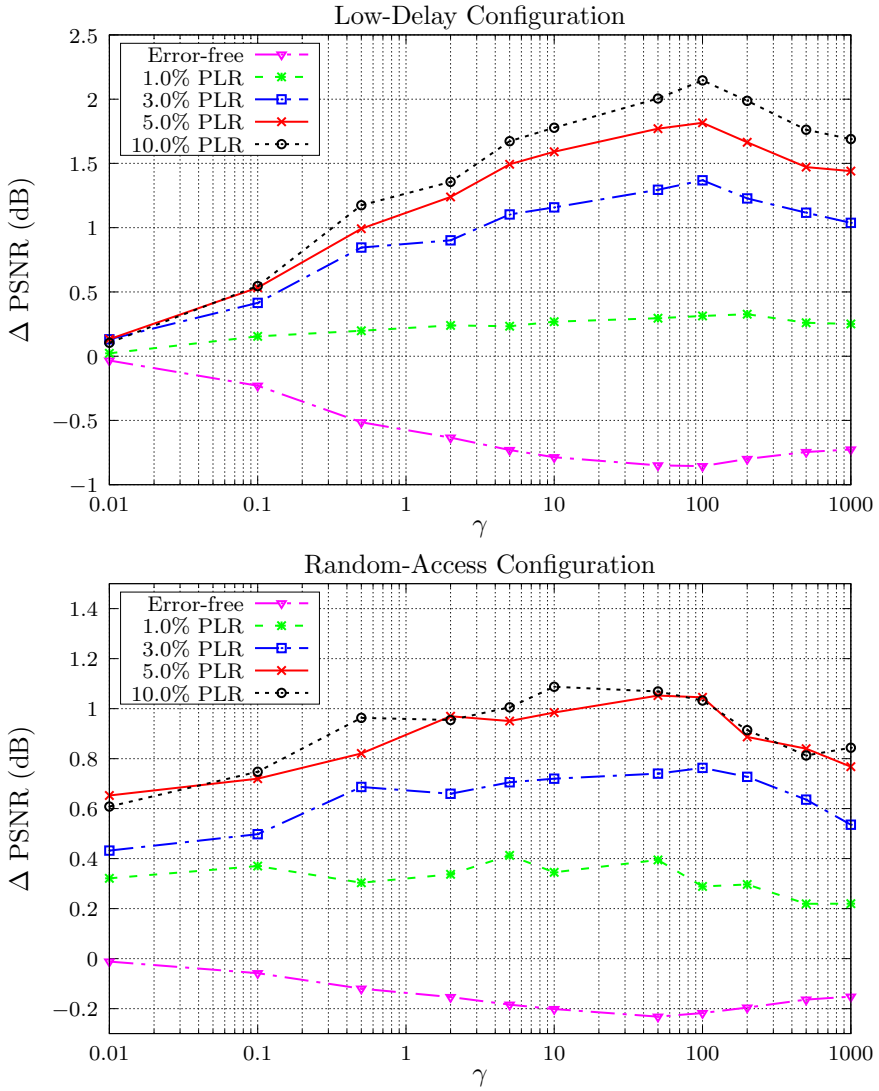


Figure 5.9: Influence of the γ parameter in the average quality gains, comparing with *Ref*, for different packet loss ratios.

Influence of parameter γ

The influence of the γ parameter used in Equation (5.5) was also studied. Figure 5.9 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 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.

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 [52] is shown in Table 5.4.

As shown in this table, the method presented in [52] is the most time consuming because it requires post-processing of the encoded frames in order to interpolate new

Table 5.4: Relative CPU times of tested methods.

Sequence	**Method**		
	Ref [52]	*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%

references. The proposed scheme, using a reference picture selection, *Prop* presents a lower complexity increase than [52], while the fixed checkerboard scheme (*Chkb*) presents a slightly lower complexity than *Prop*.

5.6.2 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 Section 5.5. 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.

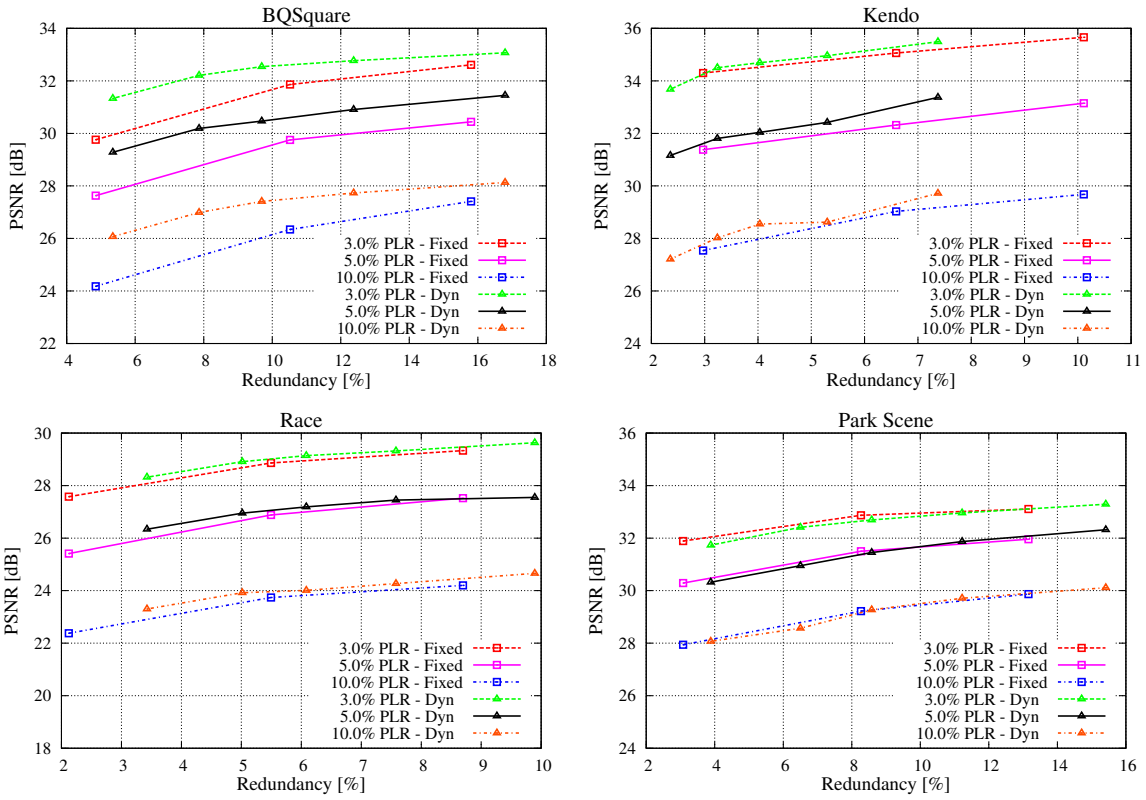


Figure 5.10: Decoded video quality for different amounts of redundancy, selected using a fixed approach and the proposed method.

Table 5.5: Average quality using redundant MV selected by the proposed approach.

Sequence	Method	Overhead ratio (%)	Packet loss rate				
			IDR_{Loss}		IDR_{NoLoss}		
			1%	5%	1%	5%	10%
Basketball Drill	Without MVs	–	30.41	23.31	34.76	28.56	25.89
	Prop-MV (α=1.5)	2.33	+0.68	+0.41	+0.58	+1.60	+1.35
	Prop-MV (α=0.8)	4.52	+0.96	+0.96	+0.95	+2.31	+2.03
	Prop-MV (α=0.6)	5.85	+1.12	+1.16	+0.93	+2.58	+2.14
Book Arrival	Without MVs	–	33.63	25.59	37.47	31.60	28.27
	Prop-MV (α=1.5)	1.81	+0.38	+0.14	+0.02	+0.70	+0.61
	Prop-MV (α=0.8)	2.95	+0.23	+1.15	+0.49	+0.76	+1.06
	Prop-MV (α=0.6)	3.92	+0.44	+1.80	+0.34	+1.53	+1.60
BQSquare	Without MVs	–	28.66	20.22	31.00	23.09	20.21
	Prop-MV (α=1.5)	5.35	+3.04	+3.94	+4.05	+6.19	+5.86
	Prop-MV (α=0.8)	9.69	+3.23	+3.80	+5.03	+7.38	+7.20
	Prop-MV (α=0.6)	12.38	+3.54	+3.93	+5.13	+7.82	+7.52
Kendo	Without MVs	–	36.21	28.38	38.23	29.58	25.52
	Prop-MV (α=1.5)	2.36	+0.78	+1.16	+0.36	+1.58	+1.69
	Prop-MV (α=0.8)	4.03	+0.94	+1.32	+0.79	+2.46	+3.03
	Prop-MV (α=0.6)	5.30	+0.86	+1.08	+0.97	+2.84	+3.10
Park Scene	Without MVs	–	28.60	23.85	33.02	28.26	25.98
	Prop-MV (α=1.5)	3.88	+0.53	+0.72	+1.05	+2.06	+2.09
	Prop-MV (α=0.8)	8.58	+0.79	+0.96	+1.55	+3.19	+3.29
	Prop-MV (α=0.6)	11.22	+0.73	+1.05	+1.58	+3.61	+3.73
Race Horses	Without MVs	–	26.87	21.78	29.67	23.23	20.48
	Prop-MV (α=1.5)	3.43	+2.46	+1.45	+2.28	+3.11	+2.82
	Prop-MV (α=0.8)	6.08	+2.76	+2.23	+2.88	+3.96	+3.53
	Prop-MV (α=0.6)	7.58	+2.93	+2.27	+2.89	+4.22	+3.79
Tennis	Without MVs	–	31.27	24.81	34.33	27.09	23.96
	Prop-MV (α=1.5)	1.63	+0.92	+0.69	+0.87	+1.42	+1.20
	Prop-MV (α=0.8)	3.12	+1.26	+0.76	+1.33	+1.93	+1.73
	Prop-MV (α=0.6)	4.10	+0.97	+1.00	+1.45	+2.17	+2.07
Traffic	Without MVs	–	27.38	22.22	33.92	27.92	24.86
	Prop-MV (α=1.5)	2.09	+0.30	+0.48	+0.95	+1.61	+1.66
	Prop-MV (α=0.8)	5.67	+0.82	+0.95	+1.82	+3.59	+3.67
	Prop-MV (α=0.6)	7.98	+0.96	+0.87	+2.07	+3.89	+4.12

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 5.10 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.

Quality evaluation

Table 5.5 presents the average PSNR of decoded video for different PLR (columns 4 to 8) 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 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.

5.7 Summary

In this chapter 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 contributes 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.

Chapter 6

Error Concealment-Aware Coding for Robust Video Transmission

6.1 Introduction

In this chapter 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 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 scheme is based on a scalable coding approach where the best 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.

This chapter is organized as follows. Section 6.2 presents the problem formulation and motivation for the method proposed in this chapter. In Section 6.3 does an initial evaluation of different error concealment schemes in order to evaluate the relevant ones to be used in the proposed error concealment-aware coding scheme. Section 6.4 provides an overview of the architecture of the proposed method and Section 6.5 describe the proposed method in detail. Section 6.6 presents a discussion of the performance of the proposed EC optimisation and signalling, and Section 6.7 presents an overall performance of the proposed method. Finally, Section 6.8 concludes this chapter.

6.2 Problem formulation

In Chapter 3 described different EC algorithms that can effectively reconstruct lost data, leading to lower error propagation. However, as mentioned before different algorithms might not present the same accuracy for every lost frame or slice. Moreover, selection at the MB/PU level is not straightforward. Approaches in the past proposed the combination of spatial and temporal EC strategies to improve the reconstruction quality [147]. Regarding temporal EC, there are hybrid algorithms to optimise the Motion Vector Extrapolation (MVE) at the pixel and block levels [133, 148]. Regarding the HEVC, existing algorithms also aim to improve the reconstruction accuracy of the motion-based recovery by using the information of correctly received neighbours, e.g., block partitions [23] and prediction residue [134]. Although such methods are able to improve the reconstruction quality and reduce the artefacts, decoder-based decisions are still not fully accurate, as they do not have enough reference information to ease the decision process. Therefore, higher performance can be achieved by performing encoding-side estimations and decisions based on the error-free information. This approach not only improves quality but also reduces the complexity of the video decoders.

In this chapter a 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. Different approaches were proposed in the past to that address the problem of EC at the encoder-side [45, 47, 48]. Moreover, extra information was used to aid the EC operation at the decoder-side, either by sending redundant data [135, 136] or by embedding data into the bitstream [49]. 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 [50, 139] proposes testing several methods with transmission of extra symbols to indicate the best method to be applied by the decoder. Although such approaches achieves reasonably good performance, it relies on the spatial neighbouring information or multi-layer video, which may not be available in all cases.

To advance one step further, in this chapter an EC-aware encoding scheme is proposed to achieve higher flexibility in terms of the error concealment approaches which leads to higher recovered video quality. This is accomplished by selecting at the encoder, among several EC algorithms with different characteristics, the one that yields the lowest reconstruction distortion at the decoder and also minimises the Rate-Distortion (R-D) cost considering the necessary signalling rate. Thus, the pro-

posed approach outperforms similar previous methods [45,48] by selecting the best error concealment method without affecting the encoder optimisation. 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 redundant pictures [110, 149], which transmit the whole frame as redundancy. In contrast with previous work presented in Chapter 5 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.

6.3 Traditional error concealment algorithms

In order to proposed a novel error concealment-aware technique different algorithms were taken into consideration. This section provides a description and evaluation of the error concealment candidates used in the proposed method. Firstly, a detailed description of each approach is provided followed by a performance comparison between each method.

6.3.1 Motion-compensation algorithms

The most common algorithms to recover missing frames used the correctly received MVs from the neighbouring frames. This has been widely used in the past for H.264/AVC standard [124, 129, 130] and also in the HEVC standard [23, 134, 150].

Let's consider the example of Figure 6.1, where a missing frame (f_0) at time instance t_0 is reconstructed based the information present in the neighbouring frames, f_{-1} and f_{-2}. One of the common approaches, i.e., motion-copy based algorithms reconstructs the missing frame through motion compensation directly using the co-located vectors in the closest neighbour frame (f_{-1}). The MVs for the motion compensation (v_0) are obtained by copying the MVs v_{-1} from the same position in f_{-1}, as shown by the example (1) in the figure.

An alternative approach reconstructs 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:

$$v_0 = \frac{t_0 - t_{-1}}{t_{-1} - t_R} \times v_{-1}, \tag{6.1}$$

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:

$$x^{'} = x - v_{0x} \tag{6.2}$$

$$y^{'} = y - v_{0y} \tag{6.3}$$

This is illustrated in Figure 6.1 by the example (2). This method is referred to as Motion Vector Extrapolation (MVE).

Finally, as a base for comparison the default EC implemented in the reference HEVC implementation the FC algorithm is also evaluated. This method consists in replacing the lost frame with a copy of the previous decoded one, which can also be described as a motion compensated reconstruction using only zero MVs.

6.3.2 Algorithms based on motion estimation

In this sub-section describe different method that can be used to estimate a motion-field, which can be used to reconstruct the missing information. The obtained motion-field is used in combination with the MVE method to recover the lost frame.

Optical flow estimation algorithms are used to estimate a 2-D MV, $v = (v_x, v_y)$ for each pixel, such as, the following energy is minimised:

$$E(x,y) = \sum_x \sum_y \left| f_0\Big(x,y\Big) - f_1\Big(x + v_x(x,y), y + v_y(x,y)\Big) \right|^2, \tag{6.4}$$

resulting an unique motion field given by v. Besides using a full search algorithm where all block-based approach is used to search for the best MVs based on the sum of absolute errors, in this work three different approaches were tested to evaluate they effectiveness to be used in EC algorithms, namely:

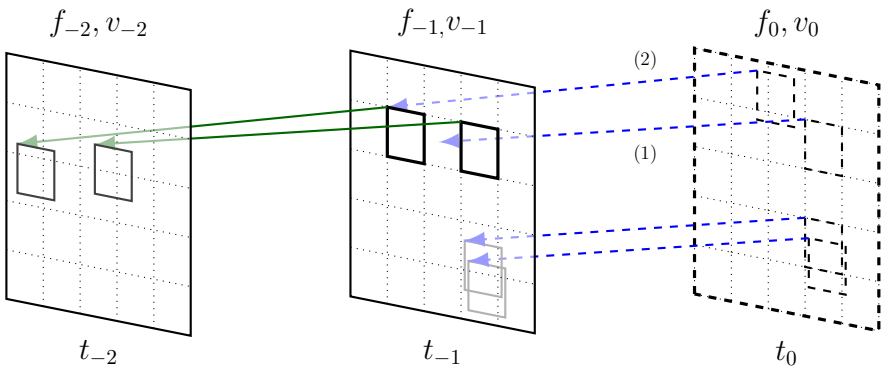


Figure 6.1: Representation of the video frame with associated original and extrapolated motion vectors represented at full and dashed arrows, respectively.

1. a method based on the total variation based on the minimization of the L-1 norm [151];

2. a method which approximates the neighbouring information of a given pixel by quadratic polynomials on both frames [152];

3. a method based on a combination of descriptor matching and large displacement estimation [153].

Motion estimation based on total variation and regularisation (*DualFlow*)

In this method the motion field is estimated based on the minimization of the total variation obtained using the L-1 norm [151,154] and taking into account the Horn and Schunck smoothness condition [155]. This condition forces the estimated field to be regular across neighbouring pixels by penalising deviations in a quadratic way. Using this method, for each pixel $\mathbf{p}$, the motion field $v(\mathbf{p})$ is obtained by minimising the following energy function,

$$E(v) = \sum_{\mathbf{p}\in\Omega} \left(\left| f_{-1}(\mathbf{p}) - f_{-2}(\mathbf{p} + v(\mathbf{p})) \right| + \left| \nabla v(\mathbf{p}) \right| \right) \tag{6.5}$$

where f_{-1} and f_{-2} are the two previous frames at time instances t_{-1} and t_{-2}, as represented in Figure 6.1. In the equation, 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.

Motion estimation based on polynomial expansion (*PolyFlow*)

In this method the motion field is estimated based on a polynomial expansion a neighbourhood of a pixel [152]. In this method only quadratic polynomials are considered, which gives the local signal model, expressed in a local coordinate system as

$$f(\mathbf{x}) \sim \mathbf{x}^T A_1 \mathbf{x} + \mathbf{b}_1^T \mathbf{x} + C_1 \tag{6.6}$$

where $\mathbf{A}$ is a symmetric matrix, $\mathbf{B}$ a vector and C a scalar. The coefficients are estimated from a weighted least squares fit to the signal values in the neighbourhood.

Since the result of polynomial expansion is that each neighbourhood is approximated by a polynomial, we start by analysing what happens if a polynomial undergoes an ideal translation. Let's consider the signal $f_1(\mathbf{x})$ represented by Equation (6.6) and a signal f_2 which corresponds to a displacement of f_1 given by $\mathbf{d}$. This can be expressed

by the following:

$$\begin{aligned} f_2(\mathbf{x}) &= f_1(\mathbf{x}+\mathbf{d}) \\ &= (\mathbf{x}+\mathbf{d})^T A_1(\mathbf{x}+\mathbf{d}) + \mathbf{b}_1^T(\mathbf{x}+\mathbf{d}) + C_1 \\ &= \mathbf{x}^T A_2 \mathbf{x} + \mathbf{b}_2^T \mathbf{x} + C_2, \end{aligned} \tag{6.7}$$

which results in the following relations,

$$\begin{aligned} \mathbf{A}_2 &= A_1 \\ \mathbf{b}_2 &= \mathbf{b} + 2\mathbf{A}_1\mathbf{d}, \\ C_2 &= \mathbf{d}^T\mathbf{A}_1\mathbf{d} + \mathbf{b}_1\mathbf{d} + C_1. \end{aligned} \tag{6.8}$$

This allows us to find the displacement $\mathbf{d}$ for any signal dimensions by solving the above equations. Detailed implementations of the polynomial expansion and solving can be found in [156, 157].

Large displacement optical flow with deep matching (*DeepFlow*)

In this method the motion field is estimated based on the large displacement optical flow presented in [158], which extends the total variation methods, as the one expressed by Equation (6.5), including descriptor matching. The correspondences between descriptions are used to apply a coarse-to-fine wrapping, thus avoiding local minimum values. The matching algorithm is built upon a multi-stage architecture with six layers with different patches sizes, interleaving convolutions and max-pooling [153]. The combination of both approaches leads towards a better matching.

In [153] to improve the estimated motion field accuracy, the following extension to the method proposed in [158] were introduced: (i) description matching was used in the minimisation problem; (ii) a normalisation parameter was introduced to down-height the impact of regions with high spatial changes; (iii) different weights are used for each layers of optimisation. Figure 6.2 shows the blocks that are included in the estimation process. As shown in the figure a multi-size optimisation is applied by using different patches sizes. This is used to estimate an initial motion field, referred to as v' .

Subsequently, the optical flow is obtained by minimising an energy function, which can be described by the following:

$$E(v) = \sum_{\mathbf{p}\in\Omega} \Big(E_D(\mathbf{p}) + \alpha E_S(\mathbf{p}) + \beta E_M(\mathbf{p}) \Big), \tag{6.9}$$

where E_D is the data term, E_S is a smoothing term and E_M is the matching term. The data term is obtained from the difference between matching pixels in both video

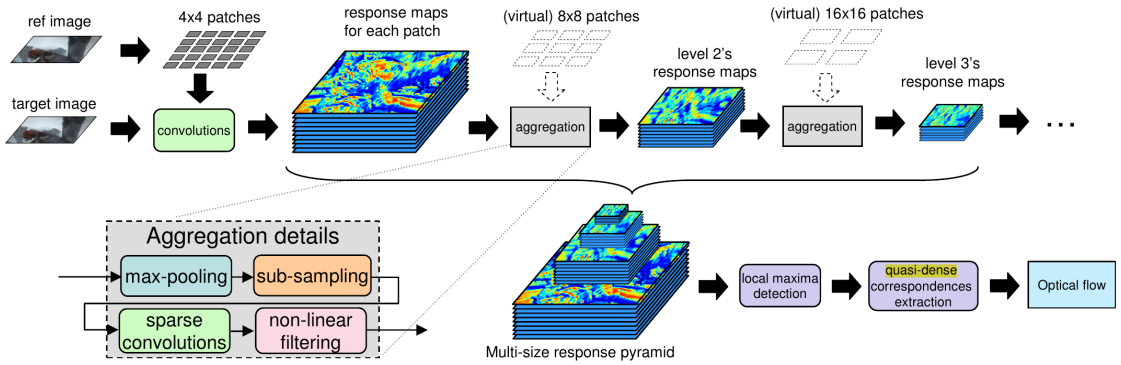


Figure 6.2: Block diagram of the deep matching algorithm [153].

frames. The smoothness term is a robust penalisation of the gradient of the flow as used in the Equation (6.5), i.e., $E_S(\mathbf{p}) = |\nabla v(\mathbf{p})|$. Finally, the matching term (E_M) is used to encourage the flow estimation to be similar to the precomputed motion field (v').

6.3.3 Experimental evaluation

In this section a performance evaluation of the different EC schemes previously described is carried out. This evaluation aims at finding the suitable methods to be used for efficiently reconstruct the missing frames in case of data loss. In this evaluation the default EC method of the reference HEVC implementation is used as a reference for comparison. Then, two method based on the correctly received MVs are tested, i.e., MC and MVE. Moreover, the four different approaches based on a estimated motion field are compared: (i) full search based on the minimisation of the sum of absolute errors (*FullSearch*), (ii) optical flow based a duality between total variation and smoothness term (*DualFlow*), (iii) optical flow based on polynomial approximations (*PolyFlow*) and (iv) optical flow based on deep matching (*DeepFlow*).

Table 6.1 shows the average quality results obtained using the PSNR metric for different methods. The absolute value is shown for the frame-copy method (*FC*), and differential values are shown for the remaining methods. Results reveal that the methods *MC* and *MVE* are able to overcome *FC* which does not use any MVs, achieving average quality gains of up to 1.41 dB and 1.46 dB, respectively. Moreover, by extrapolating the MVs from the correctly receiving frame to the lost time instance (*MVE*) is able to outperform simple motion-copy (*MC*).

Results for EC methods based motion estimation (see columns 5 to 8 of Table 6.1) reveal a overall superior performance when comparing with methods which rely of correctly received MVs. Results for *FullSearch* method reveals that only minimising the sum of absolute errors between two consecutive frames might not always be directly related with objects' motions. Thus, this may results in a inaccurate motion estimation

Table 6.1: Average quality (PSNR) using different error concealment algorithms

Sequence	*FC*	*MC*	*MVE*	*Full Search*	*Dual Flow*	*Poly Flow*	*Deep Flow*
Basketball Drill	28.22	+2.43	+2.56	+2.61	+2.71	+3.08	+3.33
Book Arrival	31.01	+2.50	+2.49	+2.31	+2.81	+3.05	+3.39
BQSquare	31.36	+2.28	+2.32	+0.43	+1.38	+1.93	+1.88
Cactus	31.93	+0.91	+0.96	+0.77	+0.96	+1.05	+1.08
Four People	37.12	+1.24	+1.26	+0.33	+1.38	+1.44	+1.44
Kendo	31.57	+2.34	+2.36	+2.04	+2.80	+2.78	+3.05
Kimono	32.24	+0.12	+0.12	+0.10	+0.13	+0.13	+0.14
Park Scene	30.42	+1.30	+1.38	+1.21	+1.23	+1.50	+1.57
People on Street	25.05	+0.40	+0.43	+0.36	+0.56	+0.61	+0.61
Race Horses	24.19	+0.68	+0.67	+0.66	+0.66	+0.76	+0.81
Tennis	27.08	+1.34	+1.56	+1.17	+1.65	+0.74	+2.37
Average	–	**+1.41**	**+1.46**	**+1.09**	**+1.48**	**+1.55**	**+1.79**

to be used for missing frames reconstruction. Finally, results from Table 6.1 reveal that the *DeepFlow* algorithm leads to the most accurate motion field to be used for EC, in comparison with the other methods. It is able to achieve an average quality gain of 1.79 dB, and up to 3.39 dB for Book Arrival sequence, where *PolyFlow* only achieves 3.05 dB.

6.3.4 Discussion and conclusions

Summarising, the analysis of the previous results indicates that all methods cope with lost frames and manage to recover the missing data with acceptable quality. Nevertheless, it is noticeable that the recovering lost frames with an estimated motion field leads to higher quality, especially using the *DeepFlow* method. However, as the other methods tested are also able to achieve acceptable video quality they shall not be neglected. Thus, in order to achieve higher reconstruction quality different approaches shall be considered.

6.4 Proposed architecture

This section presents an overview of the proposed method architecture describing the functional blocks presented providing an introduction for the subsequent detailed descriptions.

Figure 6.3 illustrates the architecture of the proposed encoding scheme. After the video encoder, it includes data loss simulation at the NAL for every slice. Subsequently, error concealment is applied to find the best Error Concealment (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

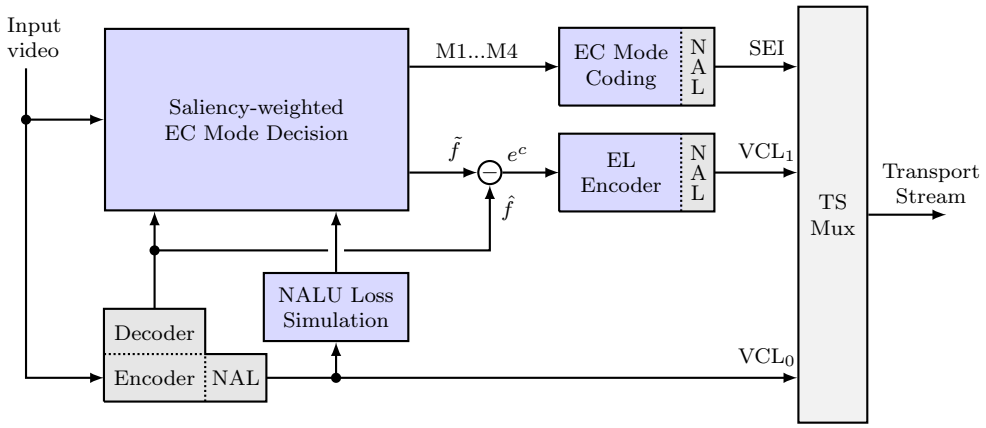


Figure 6.3: Architecture of the error concealment-aware encoding scheme.

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 error concealed 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 CTU is partitioned to form a quadtree structure of Error Concealment 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 SEI NAL unit type [8]. 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.

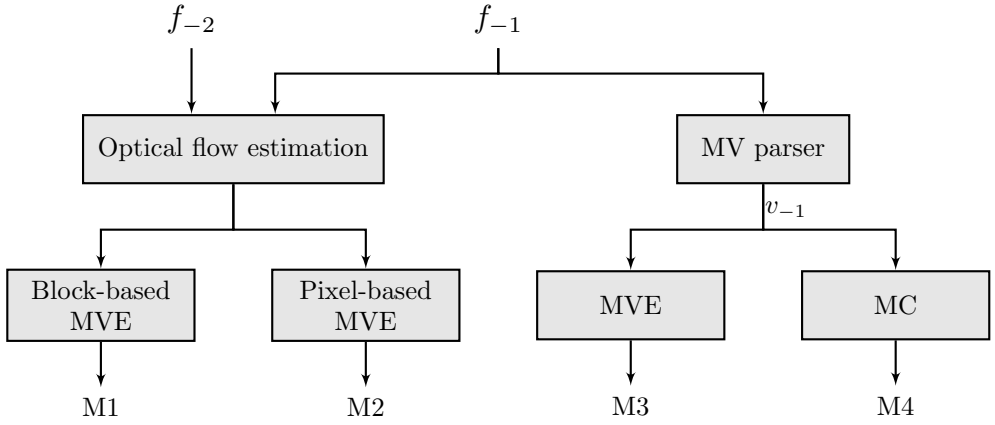


Figure 6.4: EC candidates used for optimization.

6.5 Error concealment mode decision

In this section the saliency-weighted EC mode decision algorithm is described. Firstly, the list of EC candidates from the previously described methods are enumerated. Then the R-D optimisation is proposed based on a saliency detection to decrease the overall bitrate without compromising the reconstruction quality. Moreover, the signalling coding strategy is also described. Finally, the introducing of EC residue as auxiliary information to improve the reconstruction quality is described.

6.5.1 Error concealment candidates

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 6.4, 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, using the method *DeepFlow* [153], to recover the lost frame. Then, it is extrapolated at the pixel level, i.e., one MV is assigned to each pixel, and 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 method M3 is based on a traditional Motion Vector Extrapolation (MVE), and finally M4 reconstructs the missing frames using a Motion-Copy (MC) algorithm.

6.5.2 Selection optimization

As described above, in the sub-section 6.3, 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 ECU, 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 [159] 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:

$$m^*_{l,i} = \underset{\substack{l\in\{0\ldots3\}\\ i\in\{1\ldots4^l\}}}{\arg\min} \left\{ D(m_{l,i}) + \left(1 + \frac{t_0 - t_I}{T_I}\right) \lambda R(m_{l,i}) \right\}, \tag{6.10}$$

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 6.10, 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.

6.5.3 Saliency detection

The saliency maps were computed based on the well-known Itti's attention model [137], 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

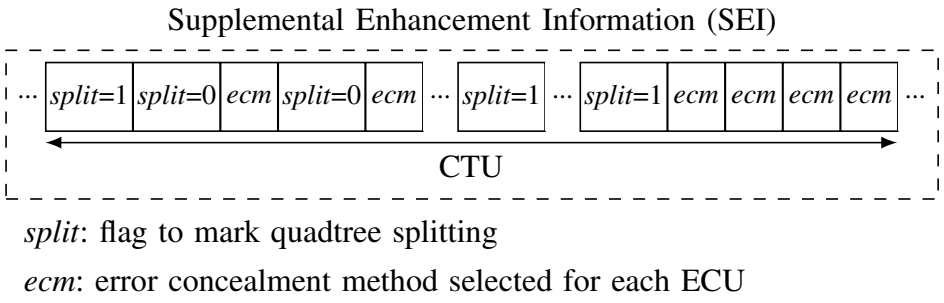


Figure 6.5: Payload structure of the SEI NALs

combined with a temporal feature. This is obtained by estimating the optical flow between two consecutive frames, based on the algorithm presented in [155]. The final saliency map is obtained by combining the spatial and temporal components, S_s and S_t respectively, as follows,

$$S = \rho S_t + (1 - \rho) S_s \tag{6.11}$$

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:

$$\lambda = \begin{cases} \lambda_0 e^{\frac{\bar{S}-s}{\Delta S}}, & \text{if } \Delta S > 0. \\ \lambda_0, & \text{no saliency used.} \end{cases} \tag{6.12}$$

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.

6.5.4 Signalling the optimal EC modes

Figure 6.5 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 [146] 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.

6.5.5 Enhancement-layer concealment residue

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 6.3). 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:

$$\tilde{f} = \tilde{f}' + e^c. \tag{6.13}$$

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 NALs 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.

Table 6.2: Percentage of overhead (%)

Sequence	32×32 **Level: 1**	16×16 **Level: 2**	*Proposed* **Level: dynamic**
Basketball Drill	0.08	0.27	0.08
Kendo	0.27	1.05	0.19
Park Scene	0.42	1.72	0.05
People On Street	0.06	0.25	0.11
Traffic	0.50	1.81	0.06

6.6 Error concealment signalling evaluation

6.6.1 Intra-frame EC optimisation

The evaluation of the proposed method is performed only for intra frames to study their impact on error propagation. Moreover, as intra coding units result in higher amount of bits, they also need a higher number of packets. Results using this setup indicate that for an intra-period of 16 frames, approximately 42% of the packets corresponds to intra coded slices. Also, by limiting the proposed method to intra frames, the amount of required overhead is reduced.

Overhead analysis

Table 6.2 shows the overhead introduced by the proposed method and also for fixed size partitions of 32×32 and 16×16, i.e., $Depth = 1$ and $Depth = 2$. While the maximum overhead of these fixed partitions is 1.81% (Traffic) the proposed dynamic partitioning method is able to reduce the overhead to a maximum of 0.19% (Kendo).

Quality evaluation

Since a relevant performance indicator is the error propagation over temporally predicted frames, this was evaluated using a single frame loss. Figures 6.6 and 6.7 show the PSNR results when frame #16 is missing, for Low-Delay and Random-Access encoding configurations, respectively. these Figures show the results for the proposed method (*Proposed*), the motion-copy method (*MC*), and also for a fixed partition size (16×16). These results indicate that the proposed method (*Proposed*) outperforms the reference one and it is able to achieve quality gains up to 2 dB for Low-Delay configuration (see Figure 6.6), decreasing the error propagation and increasing the quality of the affected frames up to 3 dB. The use of Random-Access configuration leads to higher quality degradation than LD, due the higher temporal distance of the reference frame used for concealment (frame #8 is used to conceal lost frame #16). Note that when Frame # 16 is lost, all frames within the range #9 to #31 (i.e., next I-frame) are

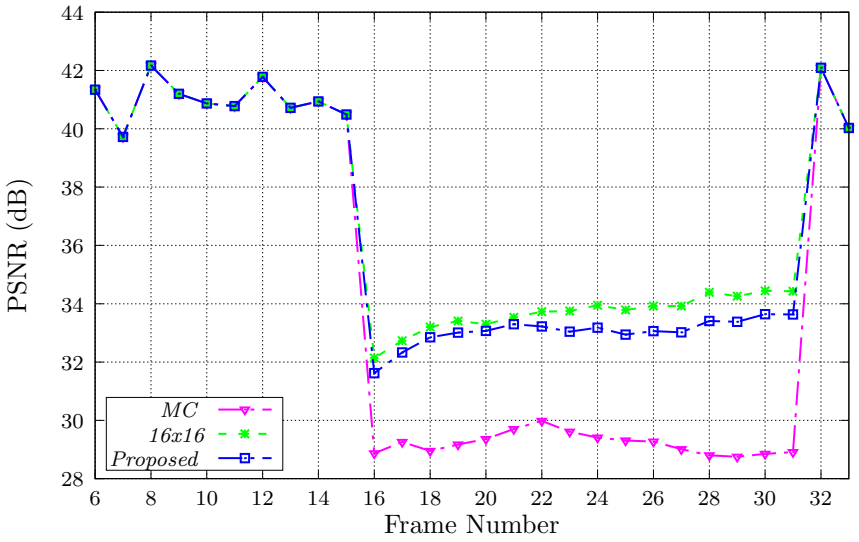


Figure 6.6: Error propagation using Low-Delay for the Basketball Drill sequence.

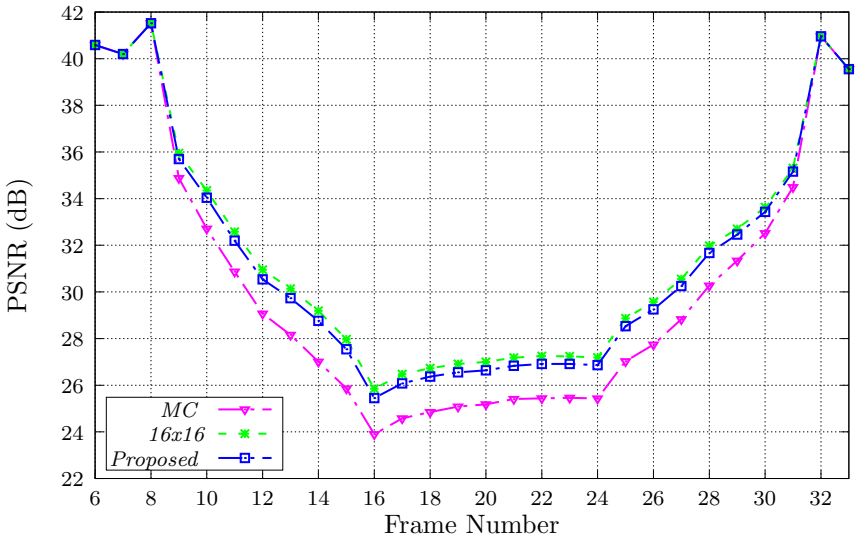


Figure 6.7: Error propagation using Random-Access for the Park Scene sequence.

affected by error propagation. Nevertheless, the proposed method still achieves similar quality gains (approximately 2 dB). Moreover, for both configurations, even when the overhead is reduced, by moving from fixed partition to dynamic partition size, the proposed method presents only a small reduction in the reconstruction quality. This indicates that using dynamic partitioning allows to achieve a better trade-off between the amount of overhead and the error concealment performance.

Further tests were performed to evaluate the effectiveness of the proposed method under more different loss rates. In the experiments the packets corresponding to intra frames were randomly discarded to simulate loss events, at different PLRs rates with an average burst length of 5 packets. For each test condition 25 trials were performed and the average quality of the frames affected by errors (the lost frame and dependent ones)

Table 6.3: Average PSNR and variance of affected frames

Sequence	**PLR=3%**			**PLR=10%**		
	MC	*Proposed*	ΔPSNR	*MC*	*Proposed*	ΔPSNR
Low-delay configuration						
Basketball Drill	33.15 (0.42)	34.26 (0.24)	**+1.11**	32.97 (0.47)	35.90 (0.27)	**+2.93**
Kendo	34.33 (1.28)	36.57 (1.11)	**+2.24**	33.84 (1.26)	37.49 (1.05)	**+3.65**
Park Scene	35.22 (0.12)	35.33 (0.08)	**+0.11**	34.75 (0.11)	34.96 (0.09)	**+0.21**
People On Street	34.36 (0.32)	34.98 (0.22)	**+0.62**	32.17 (0.65)	33.45 (0.50)	**+1.28**
Traffic	36.77 (0.48)	36.88 (0.57)	**+0.11**	36.04 (0.30)	36.31 (0.27)	**+0.27**
Jockey	36.00 (1.14)	37.20 (1.05)	**+1.20**	35.44 (1.21)	36.74 (1.18)	**+1.30**
Average	**34.37 (0.73)**	**35.25 (0.60)**	+0.88	**33.64 (0.79)**	**35.14 (0.65)**	+1.50

is presented in Table 6.3. The table also shows the PSNR variance value in parentheses and the difference to the PSNR obtained with a reference HEVC decoder The results confirm the superiority of the proposed method, outperforming the reference HEVC for both low and high PLR. Using Low-Delay coding configuration, the proposed method is able to achieve quality gains of 0.88 dB for 3% of PLR and 1.50 dB for 10% of PLR. In case of RA, higher quality gains are achieved, leading to an average improvement of 3.44 dB for 10% of PLR. It is also worthwhile to notice that the PSNR variance obtained with the *Proposed* method is lower, which indicates lower quality variations in the video segments affected by errors.

In summary, the proposed technique is able to enhance the performance of the standard decoder by finding the best error concealment method at the encoder and sending such information with a small amount of signalling overhead.

6.6.2 Robust video transmission with optimal EC

The performance of the proposed EC-aware optimisation 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;

EC mode usage distribution

The EC mode usage distribution is an indicator of the usefulness of each EC mode. Table 6.4 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 6.7.

The results in Table 6.4 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).

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. The video quality was evaluated by using two quality metrics: PSNR and a FSIM [160] in order to validate the performance of proposed method in a consolidated manner. Realistic conditions were used for transmission, by allowing independent losses for any of type of NALs, 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 for comparison (*NoSaliency*). In this case a constant $\lambda_0 = 10$ is used.

Tables 6.5 and 6.6 show the average quality under different PLRs for the Low-delay configuration. Table 6.5 shows the PSNR and weighted PSNR (WPSNR) between parentheses. The Mean Square Error (MSE) weights used in the WPSNR are defined by the saliency values [161]. Table 6.6 shows the results obtained for the FSIM quality metric. Note that, the absolute value is shown for the reference case (*Ref*), while the

Table 6.4: 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

Table 6.5: Average quality (PSNR and WPSNR) of the affected frames for Low-Delay configuration.

Sequence	**PLR=3%**					**PLR=10%**				
	Ref	*MC*	*Deep Flow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*	*Ref*	*MC*	*Deep Flow*	*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 6.6: Average FSIM of the entire video sequence in frame loss conditions (scaled between $[0 - 100]$).

Sequence	**PLR=3%**					**PLR=10%**				
	Ref	*MC*	*Deep Flow*	*NoSaliency (Constant λ)*	*Saliency (Adaptive λ)*	*Ref*	*MC*	*Deep Flow*	*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**

Table 6.7: 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%)**

differential values are shown for the remaining methods.

From the PSNR results shown in Table 6.5, 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 6.6 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 6.7 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 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 6.5 and 6.6. 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 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.

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 6.8 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 6.9 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 6.10 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*.

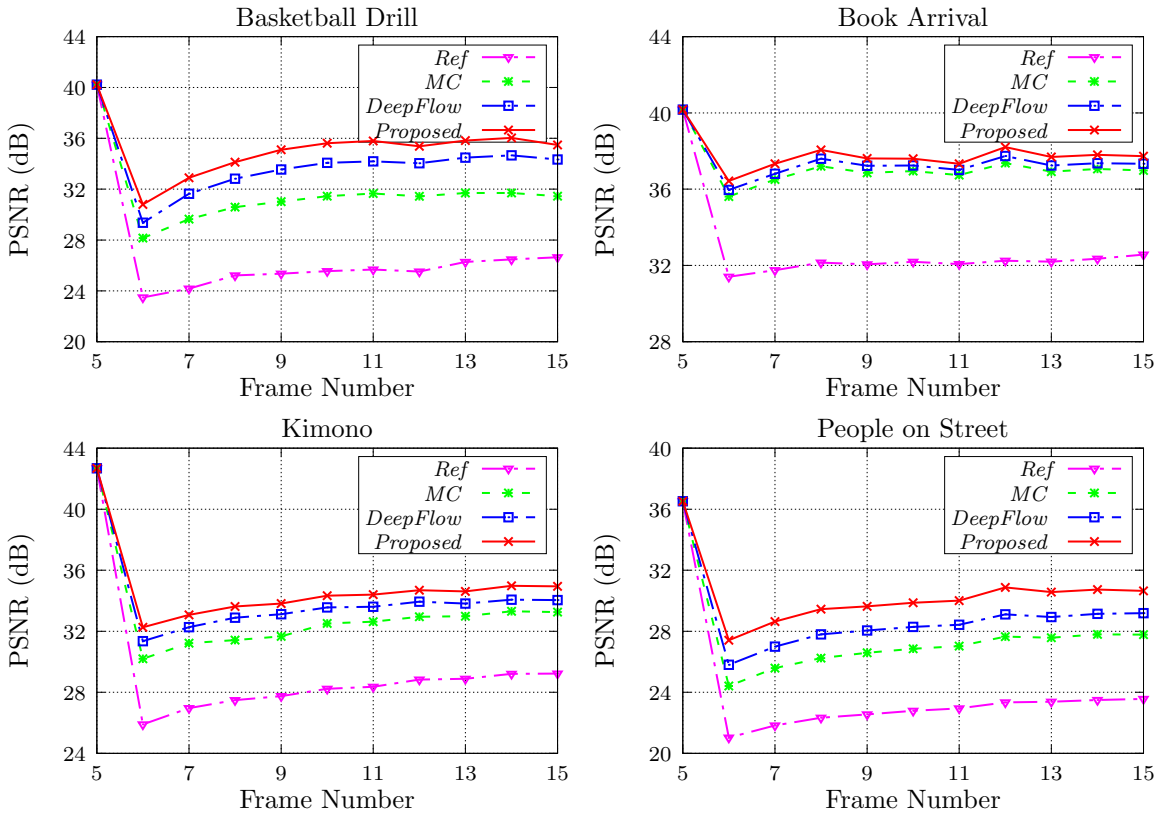


Figure 6.8: Error propagation after losing the inter-coded frame #6.

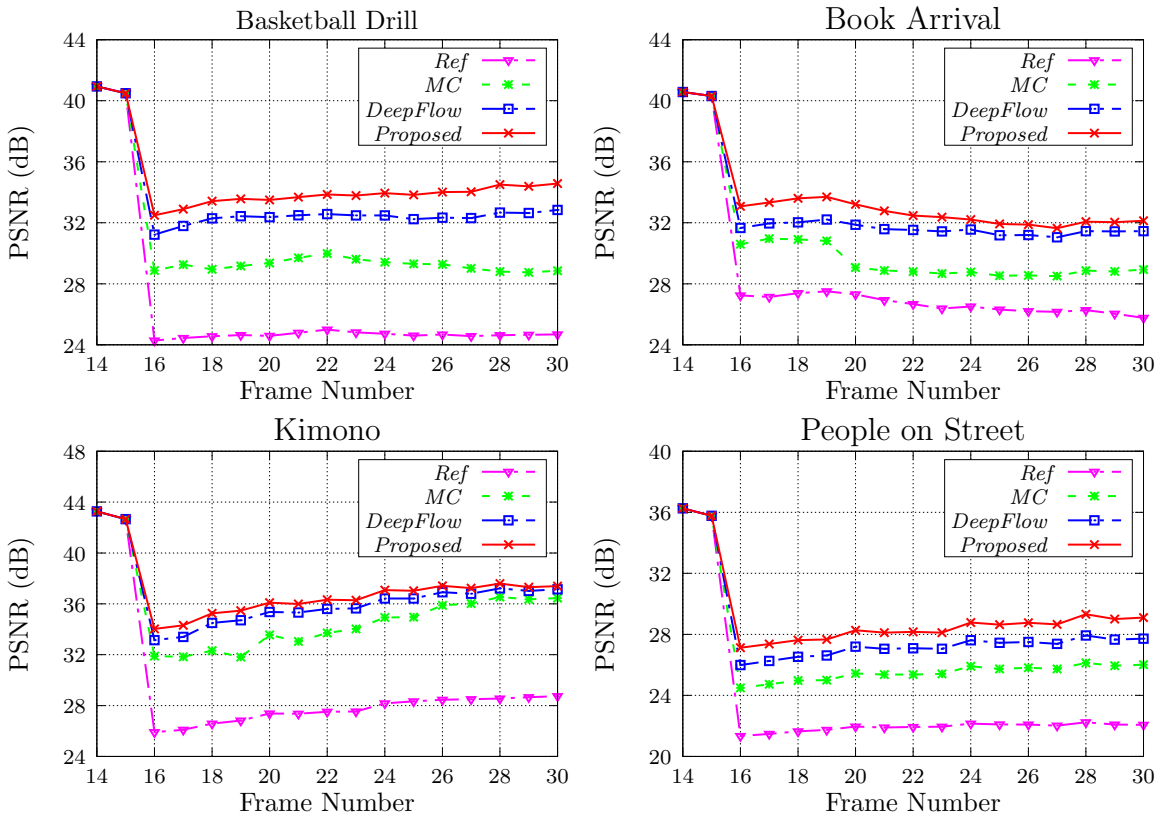
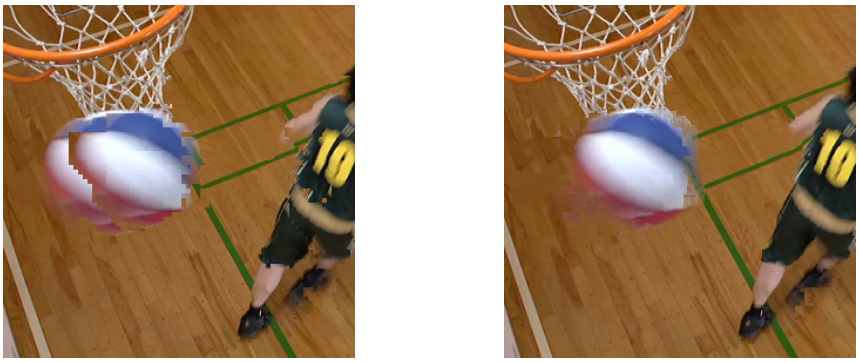


Figure 6.9: Error propagation after losing intra-coded frame #16.



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

Figure 6.10: Recovered frame #6 of Basketball Drill

Table 6.8: Average quality under 3% of PLR for different methods and overhead ratios.

Sequence	*Ref*	*MC*	*Ref [52]*	*Proposed* 10%	*Proposed* 20%
	Low-delay configuration				
Basketball Drill	33.31	+2.10	+3.86	+4.62	+4.56
Book Arrival	36.21	+1.33	+2.07	+2.71	+2.99
BQSquare	35.64	+1.47	+1.18	+2.01	+1.77
Four People	39.16	+0.47	-0.65	+0.40	-0.06
Kimono	36.19	+1.41	-0.54	+3.09	+3.47
Park Scene	34.54	+2.06	+2.01	+3.22	+3.15
People	29.48	+1.50	+1.35	+2.64	+3.08
Race Horses	29.08	+2.34	+4.52	+5.35	+6.02
Tennis	32.71	+1.32	+3.13	+3.81	+4.97
Average	–	**+1.55**	**+1.88**	**+3.10**	**+3.33**
	Random-access configuration				
Basketball Drill	35.40	+0.17	+0.20	+1.39	+1.53
Book Arrival	37.26	+0.05	+0.72	+1.31	+1.28
BQSquare	34.32	-0.06	+0.51	+1.46	+1.59
Four People	39.92	+0.02	-0.39	+0.29	+0.02
Kimono	36.98	+0.01	-2.44	+0.34	+0.00
Park Scene	35.18	+0.06	-1.27	+1.11	+1.13
People on Street	29.21	+0.03	-0.53	+0.58	+0.95
Race Horses	30.89	+0.09	-0.34	+1.38	+1.42
Tennis	33.86	+0.16	+0.18	+0.98	+1.48
Average	–	**+0.06**	**-0.37**	**+0.98**	**+1.05**

6.7 Overall performance evaluation

As in the previous simulations, the packets were randomly discarded to simulate loss events, at different PLRs. The PSNR obtained with the proposed method is compared with the default method used by reference HEVC , i.e., frame-copy (*Ref*), motion-copy (*MC*) [134], and the method proposed in [52], which implements a robust scheme to minimise error propagation. The following results were obtained using the same constant bitrate for each test condition.

Table 6.9: Average quality under 10% of PLR.

Sequence	*Ref*	*MC*	*Ref [52]*	*Proposed* 10%	*Proposed* 20%
	Low-delay configuration				
Basketball Drill	28.22	+2.34	+5.32	+6.26	+6.90
Book Arrival	31.01	+2.48	+4.49	+5.60	+6.31
BQSquare	31.36	+2.50	+2.40	+3.33	+3.55
Four People	37.12	+1.24	-0.99	+1.82	+1.51
Kimono	32.24	+0.12	-1.92	-0.04	-0.21
Park Scene	30.42	+1.22	+1.56	+2.92	+3.64
People	25.05	+0.40	+0.74	+0.97	+2.09
Race Horses	24.19	+0.63	+2.53	+2.75	+3.74
Tennis	27.08	+1.30	+3.51	+4.58	+6.61
Average	–	**+1.36**	**+1.96**	**+3.13**	**+3.79**
	Random-access configuration				
Basketball Drill	29.92	+0.14	+1.45	+1.84	+2.42
Book Arrival	32.27	+0.13	+2.58	+2.74	+2.91
BQSquare	28.13	-0.01	+2.74	+2.06	+2.85
Four People	37.81	+0.09	-0.05	+1.46	+1.68
Kimono	33.09	+0.00	-0.03	-1.15	-2.02
Park Scene	30.93	+0.01	+0.73	+0.83	+1.02
People on Street	24.57	+0.00	+1.70	+0.09	+0.35
Race Horses	25.71	+0.01	+1.30	+0.58	+0.71
Tennis	27.74	+0.07	+1.35	+1.17	+2.09
Average	–	**+0.05**	**+1.31**	**+1.07**	**+1.33**

Tables 6.8 and 6.9 present the average PSNR of decoded video for different PLRs and different percentages of total overhead (i.e., signalling rate plus enhancement layer rate). The absolute PSNR value is shown for the reference case, while the PSNR difference is 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 [52] for both Low-delay and Random-access configurations. 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 for the Low-delay configuration (1.33 dB for the Random-access).

When comparing the results of Tables 6.8 and 6.9 with the spatial and temporal information of each sequence (SI/TI in Table 4.1) 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

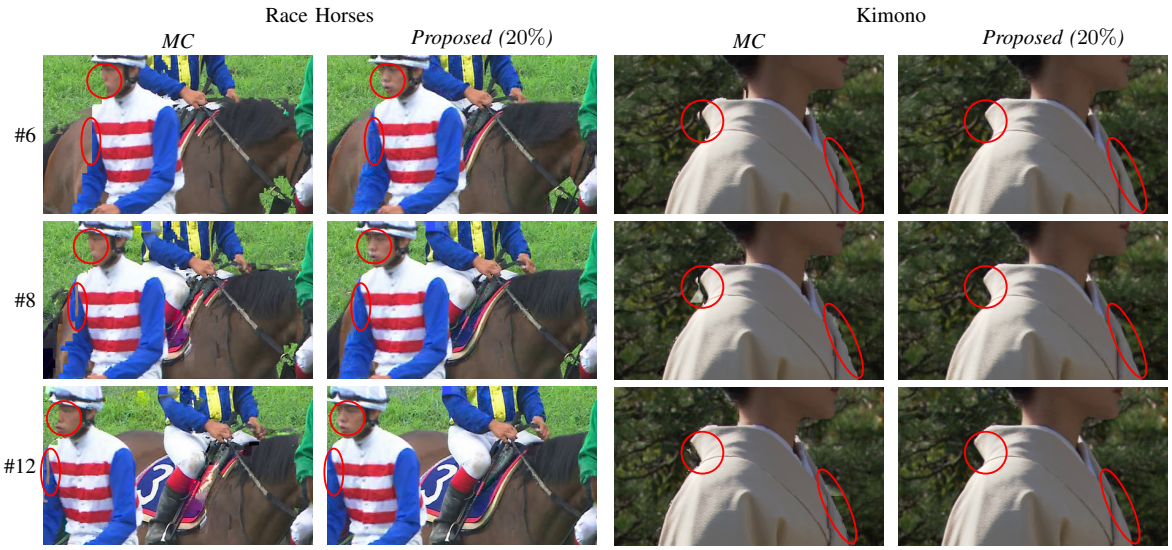


Figure 6.11: Crop of the decoded video frame for the Race Horses and Kimono sequences (Loss event at Frame #6)

sequences, both in temporal and spatial levels.

6.7.1 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 6.11 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 6.11 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.

6.7.2 Comparison with previous works

The results were compared with previous works where fair comparisons could be made. In [50] 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 6.3 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 [139], 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 [46, 71] both use a feedback channel and intra-refresh to reduce the error propagation. The results presented in [71] 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 6.10 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 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.,

$$R = \left| \frac{\Delta PSNR(\text{PLR} = 10\%)}{\Delta PSNR(Error\text{-}free)} \right|. \tag{6.14}$$

The results shown in Table 6.10, 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 [162]* the proposed method leads to higher quality degradation in the error-free case. However, for PLR=10% it clearly outperforms *Ref [162]* with a much higher value of R, indicating a far superior efficiency. Results in the table 6.10 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 [52, 162] are based on modified predictions, while the method in [149] uses redundant data without any further optimisation. Both of

Table 6.10: Comparison against existing methods.

Method	**Δ PSNR**		**R**
	Error-free	PLR=10%	
Ref [52]	-0.88	1.64	1.9
Ref [162]	-0.24	1.42	5.9
Ref [149]	-0.37	2.03	5.5
Proposed	-0.35	2.10	6.0

these approaches require low additional computation complexity. However, as observed in the results shown in Table 6.10 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 straightforward methods that always used the same EC techniques for all CTUs.

6.8 Summary

In this chapter an error concealment-aware encoder is proposed by incorporating loss simulation at the encoder-side to enhance the error concealment efficiency at the decoder. The novelty of this method is to include optimal error concealment mode decision at the encoder, to be used by the decoder in case of data loss, and also to allow encoding of the error concealment 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 is achieved at the cost of low overhead, which allows to use this method in broadcasting video services and applications where packet loss probability should not be ignored.

Chapter 7

Joint error resilience-concealment for robust video streaming

7.1 Introduction

In this chapter the error resilience and error concealment approaches proposed in the previous chapters, namely Chapter 5 and 6, are combined to achieved an higher decoder video quality across different network conditions and applications. This chapter presents a combined evaluation of different approaches and proposes a method to efficiently select the optimal one in terms of reconstructed quality.

This chapter is organised as follows. Section 7.2 presents the problem formulation and motivation for the method proposed in this chapter. ... Section 7.8 concludes this chapter.

7.2 Problem formulation

7.3 Performance evaluation under different network condition

7.3.1 Test conditions

7.3.2 Experimental results

7.3.3 Discussion and conclusions

7.4 Proposed architecture

7.5 Error robustness optimisation

7.6 Optimisation method

7.6.1 Performance evaluation

7.7 Error concealment optimisation based on MV differences

7.7.1 Experimental results

7.8 Summary

Chapter 8

Conclusion and future work

This chapter concludes this thesis, presenting some conclusions about this research work and some future research perspectives in the area of video transmission, error resilience and error concealment.

8.1 Conclusions

8.2 Future work

Appendix A

Test video sequences

This appendix illustrates the original test signals used in simulations throughout the thesis. This includes a partial set of the HEVC test material and other external test material.



Figure A.1: Basketball Drill, 832 × 480, 50 fps, 240 frames.



Figure A.2: Book Arrival, 1024 × 768, 30 fps, 240 frames.



Figure A.3: BQSquare, 416 × 240, 60 fps, 240 frames.



Figure A.4: Bosphorus, 3840 × 2160, 120 fps, 240 frames.



Figure A.5: Cactus, 1920 × 1080, 50 fps, 500 frames.



Figure A.6: Four People, 1280 × 720, 60 fps, 240 frames.



Figure A.7: Jockey, 3840 × 2160, 120 fps, 240 frames.



Figure A.8: Kendo, 1024 × 768, 30 fps, 240 frames.



Figure A.9: Kimono, 1920 × 1080, 24 fps, 240 frames.



Figure A.10: Park Scene, 1920 × 1080, 24 fps, 240 frames.



Figure A.11: People On Street, 2560 × 1600, 24 fps, 150 frames.



Figure A.12: Race Horses, 832 × 480, 30 fps, 240 frames.



Figure A.13: Traffic, 2560 × 1600, 30 fps, 150 frames.

Appendix B

Published papers

This appendix presents a list of the published papers that resulted from the research work done during this thesis proposal. Submitted papers are also presented.

B.1 Journal publications

1. J. Carreira, P. Assuncao, S. 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*, Apr. 2017.

2. J. Carreira, P. Assuncao, S. Faria, E. Ekmekcioglu, and A. Kondoz, "Error Concealment-Aware Encoding for Robust Video Transmission," *IEEE Transactions on Broadcasting*, **(Submitted)**.

B.2 Conference publications

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

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

3. J. Carreira, S. Faria, P. Assuncao, E. Ekmekcioglu, and A. Kondoz, "Error resilience analysis of motion vector prediction in HEVC," in *Conference on Telecom-*

munications (Conftele), Oct. 2015, pp. 1–4.

4. J. Carreira, P. Assuncao, S. Faria, E. Ekmekcioglu, A. Kondoz, and H.Lim, "Reference picture selection using checkerboard pattern for resilient video coding," in *IEEE International Conference on Visual Communications and Image Processing (VCIP)*, Dec. 2015, pp. 1–5.

5. J. Carreira, P. Assuncao, S. Faria, E. Ekmekcioglu, and A. Kondoz, "A robust video encoding scheme to enhance error concealment of intra frames," in *IEEE International Conference on Circuits and Systems (ISCAS)*, May 2017, pp. 1–5.

Bibliography

[1] I. J. ITU-T, "Advanced video coding for generic audiovisual services. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)," London, UK, May 2003.

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

[3] I. J. ITU-T, "High efficiency video coding. itu-t recommendation h.265 and iso/iec 23008-2," Feb. 2018.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[10] G. Correa, P. Assuncao, L. Agostini, and L. S. da 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.

[11] 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.

[12] ——, "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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] R. Hoffman, *Data Compression in Digital Systems*, 1st ed. London, UK, UK: Chapman & Hall, Ltd., 1997.

[25] B. Bing, *Video Coding Fundamentals*. Wiley-Blackwell, 2015, ch. 2, pp. 29–82.

[26] L. F. R. Lucas, E. A. B. da Silva, S. M. M. de Faria, N. M. M. Rodrigues, and C. L. Pagliari, *Efficient Predictive Algorithms for Image Compression*. Springer Publishing Company, 2017.

[27] M. Pourazad, C. Doutre, M. Azimi, and P. Nasiopoulos, "Hevc: The new gold standard for video compression: How does hevc compare with H.264/AVC?" *IEEE Consumer Electronics Magazine*, vol. 1, no. 3, pp. 36–46, Jul. 2012.

[28] V. Sze, M. Budagavi, and G. J. Sullivan, *High Efficiency Video Coding (HEVC): Algorithms and Architectures*. Springer Publishing Company, 2014.

[29] S. Wenger, "H.264/AVC over IP," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 645–656, Jul. 2003.

[30] A. Fujibayashi and T. Tan, "Random access support for hevc, document JCTVC-D234," Daegu, Korea, Jan. 2011.

[31] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth, and M. Zhou, "An overview of tiles in HEVC," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 6, pp. 969–977, Dec. 2013.

[32] R. Rodrguez-Snchez and E. S. Quintana-Ort, "Tiles-and wpp-based hevc decoding on asymmetric multi-core processors," in *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, Apr. 2017, pp. 299–302.

[33] K. Chen, J. Sun, Y. Duan, and Z. Guo, "A novel wavefront-based high parallel solution for HEVC encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 181–194, Jan. 2016.

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

[35] T. Nguyen and D. Marpe, "Performance analysis of hevc-based intra coding for still image compression," in *Picture Coding Symposium (PCS)*, May 2012, pp. 233–236.

[36] N. Purnachand, L. N. Alves, and A. Navarro, "Improvements to TZ search motion estimation algorithm for multiview video coding," in *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Apr. 2012, pp. 388–391.

[37] T. K. Lee, Y. L. Chan, and W. C. Siu, "Adaptive search range for HEVC motion estimation based on depth information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2216–2230, Oct. 2017.

[38] S. H. Yang, J. Z. Jiang, and H. J. Yang, "Fast motion estimation for HEVC with directional search," *Electronics Letters*, vol. 50, no. 9, pp. 673–675, Apr. 2014.

[39] J.-L. Lin, Y.-W. Chen, Y.-W. Huang, and S.-M. Lei, "Motion vector coding in the hevc standard," *IEEE Journal of Selected Topics in Signal Processing*, 2013.

[40] A. Saxena and F. C. Fernandes, "DCT/DST-based transform coding for intra prediction in image/video coding," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3974–3981, Oct. 2013.

[41] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[42] Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, no. 4, pp. 61–82, Jul. 2000.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[52] 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.

[53] ——, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 7, pp. 845–856, Jul. 2007.

[54] Y. Liao and J. D. Gibson, "Enhanced error resilience of video communications for burst losses using an extended ROPE algorithm," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Apr. 2009, pp. 1853–1856.

[55] Q. Cheng and D. Agrafiotis, "End to end video distortion estimation with advanced error concealment considerations," in *IEEE Visual Communications and Image Processing Conference*, Dec. 2014, pp. 303–306.

[56] Q. Peng, L. Zhang, X. Wu, and Q. Wang, "Modeling of SSIM-based end-to-end distortion for error-resilient video coding," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 45, 2014.

[57] Z. Wang, L. Lu, and A. Bovik, "Video quality assessment using structural distortion measurement," in *IEEE International Conference on Image Processing*, vol. 3, 2002, pp. 5–68.

[58] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[59] B. Girod and N. Farber, "Feedback-based error control for mobile video transmission," *Proceedings of the IEEE*, vol. 87, no. 10, pp. 1707–1723, Oct. 1999.

[60] Q. Chen, Z. Chen, X. Gu, and C. Wang, "Attention-based adaptive intra refresh for error-prone video transmission," *IEEE Communications Magazine*, vol. 45, no. 1, pp. 52–60, Jan. 2007.

[61] Y. Zhou, W. Xu, and Y. Chen, "A network-aware error-resilient video coding using adaptive intra and reference selection refresh," in *International Symposium on Computer Network and Multimedia Technology (CNMT)*, Jan. 2009, pp. 1–4.

[62] P. Nunes, L. D. Soares, and F. Pereira, "Automatic and adaptive network-aware macroblock intra refresh for error-resilient H.264/AVC video coding," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 3073–3076.

[63] S. Moiron and M. Ghanbari, "Enhanced error resiliency for video with cyclic intra-refresh lines," in *2011 18th IEEE International Conference on Image Processing*, Sep. 2011, pp. 3229–3232.

[64] A. R. dela Cruz and R. D. Cajote, "Error resilient joint source-channel adaptive intra-refresh rate for wireless video transmission," in *19th International Conference on Digital Signal Processing*, Aug. 2014, pp. 509–514.

[65] Y. R. Zhou, G. Q. Li, and S. S. Ning, "A new feedback-based intra refresh method for robust video coding," in *International Conference on Computer Science and Applications (CSA)*, Nov. 2015, pp. 218–221.

[66] Z. Wang, Z. Hou, J. Xiao, R. Wang, and R. Zhu, "A hybrid intra refresh and reference selection scheme for mobile video coding," in *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 357–362.

[67] P. Nunes, L. D. Soares, and F. Pereira, "Error resilient macroblock rate control for H.264/AVC video coding," in *15th IEEE International Conference on Image Processing*, Oct. 2008, pp. 2132–2135.

[68] B. Li, T. Nanjundaswamy, and K. Rose, "An error-resilient video coding framework with soft reset and end-to-end distortion optimization," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 1910–1914.

[69] V. A. Nguyen, Z. Chen, and Y. P. Tan, "Perceptually optimized error resilient transcoding using attention-based intra refresh," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2010, pp. 4217–4220.

[70] H.-J. Chiou, Y.-R. Lee, and C.-W. Lin, "Content-aware error-resilient transcoding using prioritized intra-refresh for video streaming," *Journal of Visual Communication and Image Representation*, vol. 16, pp. 563–588, Aug. 2005.

[71] H. M. Maung, S. Aramvith, and Y. Miyanaga, "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.

[72] U. Celikcan and E. Tuncel, "Bimodal leaky prediction for error resilient video streaming," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, Nov. 2009, pp. 583–587.

[73] Q. Chen, X. Yang, L. Song, and W. Zhang, "Robust video region-of-interest coding based on leaky prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 9, pp. 1389–1394, Sept 2009.

[74] H.-C. Huang, C.-N. Wang, and T. Chiang, "A robust fine granularity scalability using trellis-based predictive leak," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 6, pp. 372–385, Jun. 2002.

[75] D. Connor, "Techniques for reducing the visibility of transmission errors in digitally encoded video signals," *IEEE Transactions on Communications*, vol. 21, no. 6, pp. 695–706, Jun. 1973.

[76] J. M. Boyce, "Weighted prediction in the H.264/MPEG AVC video coding standard," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, vol. 3, May 2004, pp. III–789–92 Vol.3.

[77] 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.

[78] 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.

[79] W. Tu and E. Steinbach, "Proxy-based reference picture selection for error resilient conversational video in mobile networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 2, pp. 151–164, Feb. 2009.

[80] 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.

[81] 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.

[82] 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.

[83] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[84] P. Lambert, W. D. Neve, Y. Dhondt, and R. V. de Walle, "Flexible macroblock ordering in H.264/AVC," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 358–375, 2006, introduction: Special Issue on emerging H.264/AVC video coding standard.

[85] S. K. Im and A. J. Pearmain, "Error resilient video coding with priority data classification using H.264 flexible macroblock ordering," *IET Image Processing*, vol. 1, no. 2, pp. 197–204, June 2007.

[86] 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.

[87] ——, "A new error resilience scheme based on FMO and error concealment in H.264/AVC," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 1057–1060.

[88] J. Panyavaraporn and R. D. Cajote, "Flexible macroblock ordering based on region of interest for H.264/AVC wireless video transmission," in *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, Apr. 2012, pp. 384–387.

[89] H. D. Pham and S. Vafi, "Motion-energy-based unequal error protection for H.264/AVC video bitstreams," *Signal, Image and Video Processing*, vol. 9, no. 8, pp. 1759–1766, Nov. 2015.

[90] 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.

[91] Y. Wang, M. O'Neill, and F. Kurugollu, "New FMO type to flag ROI in H.264/AVC," in *5th European Workshop on Visual Information Processing (EUVIP)*, Dec. 2014, pp. 1–5.

[92] X. Yang, C. Zhu, Z. G. Li, X. Lin, and N. Ling, "An unequal packet loss resilience scheme for video over the internet," *IEEE Transactions on Multimedia*, vol. 7, no. 4, pp. 753–765, Aug. 2005.

[93] Q. Qu, Y. Pei, and J. W. Modestino, "An adaptive motion-based unequal error protection approach for real-time video transport over wireless IP networks," *IEEE Transactions on Multimedia*, vol. 8, no. 5, pp. 1033–1044, Oct. 2006.

[94] T. Zhang and Y. Xu, "Unequal packet loss protection for layered video transmission," *IEEE Transactions on Broadcasting*, vol. 45, no. 2, pp. 243–252, Jun. 1999.

[95] P. Perez and N. Garcia, "Lightweight multimedia packet prioritization model for unequal error protection," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 132–138, Feb. 2011.

[96] H. D. Pham and S. Vafi, "An adaptive unequal error protection based on motion energy of H.264/AVC video frames," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2013, pp. 4594–4599.

[97] H. D. Pham and V. Sina, "Unequal error protection of H.264/AVC video bitstreams based on data partitioning and motion information of slices," in *IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, Aug. 2012, pp. 634–639.

[98] H. D. Pham and S. Vafi, "Unequal error protection of H.264/AVC video bitstreams based on the motion energy estimation for group of pictures," in *7th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec. 2013, pp. 1–6.

[99] 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.

[100] Y. Zhang, S. Qin, and Z. He, "Fine-granularity transmission distortion modelling for video packet scheduling over mesh networks," *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 1–12, Jan. 2010.

[101] 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.

[102] H. Ha and C. Yim, "Layer-weighted unequal error protection for scalable video coding extension of H.264/AVC," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 736–744, May 2008.

[103] E. Maani and A. K. Katsaggelos, "Unequal error protection for robust streaming of scalable video over packet lossy networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp. 407–416, Mar. 2010.

[104] J. Jose and S. M. Sameer, "A new unequal error protection technique for scalable video transmission over multimedia wireless networks," in *IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, Feb. 2015, pp. 1–5.

[105] X. Song, X. Peng, J. Xu, G. Shi, and F. Wu, "Unequal error protection for scalable video storage in the cloud," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 699–710, Mar. 2018.

[106] Y. Huo, M. El-Hajjar, R. G. Maunder, and L. Hanzo, "Layered wireless video relying on minimum-distortion inter-layer fec coding," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 697–710, Apr. 2014.

[107] R. Perera, A. Fernando, H. K. Arachchi, and M. A. Imran, "Adaptive modulation and coding based error resilience for transmission of compressed video," in *International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug. 2015, pp. 1127–1132.

[108] P. Tovstogan and H. F. Hsiao, "Video streaming optimization using degradation estimation with unequal error protection," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 1–4.

[109] R. Perera, A. Fernando, T. Mallikarachchi, H. K. Arachchi, and M. Pourazad, "Qoe aware resource allocation for video communications over LTE based mobile networks," in *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, Aug. 2014, pp. 63–69.

[110] 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.

[111] 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.

[112] 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.

[113] 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.

[114] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[115] P. Salama, N. Shroff, E. Coyle, and E. Delp, "Error concealment techniques for encoded video streams," in *International Conference on Image Processing*, vol. 1, Oct. 1995, pp. 9–12.

[116] S. Valente, C. Dufour, F. Groliere, and D. Snook, "An efficient error concealment implementation for MPEG-4 video streams," *IEEE Transactions on Consumer Electronics*, vol. 47, no. 3, pp. 568–578, Aug. 2001.

[117] J. W. Woods, *Multidimensional Signal, Image, and Video Processing and Coding*, 2nd ed. Orlando, FL, USA: Academic Press, Inc., 2012.

[118] R. Bernardini, L. Celetto, G. Gennari, M. Petrani, and R. Rinaldo, "Error concealment of INTRA coded video frames," in *International Workshop on Image Analysis for Multimedia Interactive Services*, Apr. 2010, pp. 1–4.

[119] W. Kim, J. Koo, and J. Jeong, "Fine directional interpolation for spatial error concealment," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 3, pp. 1050–1056, Aug. 2006.

[120] S.-C. Hsia, "An edge-oriented spatial interpolation for consecutive block error concealment," *IEEE Signal Processing Letters*, vol. 11, no. 6, pp. 577–580, Jun. 2004.

[121] J. Chen, J. Liu, X. Wang, and G. Chen, "Modified edge-oriented spatial interpolation for consecutive blocks error concealment," in *IEEE International Conference on Image Processing,*, vol. 3, Sep. 2005, pp. 904–913.

[122] H. Gharavi and S. Gao, "Spatial interpolation algorithm for error concealment," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2008, pp. 1153–1156.

[123] H. Senel, "Gradient estimation using wide support operators," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 867–878, Apr. 2009.

[124] 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.

[125] Y. Chen, Y. Hu, O. Au, H. Li, and C. W. Chen, "Video error concealment using spatio-temporal boundary matching and partial differential equation," *IEEE Transactions on Multimedia*, vol. 10, no. 1, pp. 2–15, Jan. 2008.

[126] S. Garg and S. Merchant, "Interpolated candidate motion vectors for boundary matching error concealment technique in video," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 53, no. 10, pp. 1039–1043, Oct. 2006.

[127] T. Thaipanich, P.-H. Wu, and C.-C. Kuo, "Low-complexity video error concealment for mobile applications using OBMA," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 753–761, May 2008.

[128] J. Zhang, J. Arnold, and M. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 659–665, Jun. 2000.

[129] Q. Peng, T. Yang, and C. Zhu, "Block-based temporal error concealment for video packet using motion vector extrapolation," in *IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions*, vol. 1, Jul. 2002, pp. 10–14.

[130] H. Liu, W. Li, and O. Issa, "New algorithm for motion vector extrapolation for concealing entire frame loss in the H.264 video receiver," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2011, pp. 525–526.

[131] Y. Chen, K. Yu, J. Li, and S. Li, "An error concealment algorithm for entire frame loss in video transmission," in *Picture Coding Symposium (PCS)*, 2004.

[132] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "An error concealment algorithm for streaming video," in *International Conference on Image Processing*, vol. 3, Sep. 2003, pp. 49–52.

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] 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.

[138] J. Y. Pyun and H. J. Choi, "Error concealment aware error resilient video coding over wireless burst-packet-loss network," in *5th IEEE Consumer Communications and Networking Conference*, Jan. 2008, pp. 824–828.

[139] 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.

[140] 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.

[141] 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.

[142] ITU-T, "Recommendation P.910, Subjective video quality assessment methods for multimedia applications," Apr. 2008.

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

[144] 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.

[145] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves, VCEG contribution VCEG-M33," Austin, Apr. 2001.

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

[147] 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.

[148] 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.

[149] 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.

[150] K. H. Choi and D. H. Kim, "A downhill simplex approach for HEVC error concealment in wireless IP networks," in *IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2016, pp. 143–146.

[151] J. S. Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.

[152] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 363–370.

[153] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Deepflow: Large displacement optical flow with deep matching," in *IEEE Intenational Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013.

[154] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Pattern Recognition*, F. A. Hamprecht, C. Schnörr, and B. Jähne, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 214–223.

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

[156] H. Knutsson and C.-F. Westin, "Normalized and differential convolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1993, pp. 515–523.

[157] G. Farnebäck, "Polynomial expansion for orientation and motion estimation," Ph.D. dissertation, Linköping University Electronic Press, 2002.

[158] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

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

[160] 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.

[161] 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.

[162] 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.