
ADAPTIVE RESOLUTION VIDEO CODING

Brett Hosking



Centre of Doctoral Training in Communications
Visual Information Laboratory

PhD
Thesis
Electronic & Electrical Engineering
University of Bristol

Sponsored by:

BAE SYSTEMS

EPSRC

A thesis submitted to the University of Bristol in accordance with the requirements for
the degree of Doctor of Philosophy in the Faculty of Engineering

September 2016

ABSTRACT

Video coding and compression is a necessary process for the transmission or storage of digital video. Over time the increasing demand for higher spatial resolutions, faster frame rates and greater dynamic range has led to the need for more intelligent and efficient video coding algorithms. However, many applications fail to meet these demands due to low bandwidth restrictions – high levels of coding distortion can render a video transmission useless by failing to provide enough detail from the original captured scene.

Previous work has shown that some degree of this distortion can be alleviated by coding at lower spatial resolutions. Both coding and resampling introduces distortion and therefore best results can be obtained by minimising the distortion produced by these two processes. True video quality is assessed over the duration of the sequence; high variability of picture distortion can alter the perceived level of coding efficiency. A well coded sequence is therefore one that maintains a high but also consistent level of quality, and also a well regulated bitrate that prevents skipping of frames due to the over fullness of the channel buffer.

The main problem addressed in this thesis is the high bitrate requirement of the intra-coded Instantaneous Decoder Refresh (IDR) pictures featuring in all video coding standards. Frequent coding of IDR pictures is essential for error resilience in order to prevent the occurrence of error propagation. However, as each one consumes a huge portion of the available bitrate, the quality of future coded frames is hindered by high levels of compression.

The work presented in this thesis demonstrates a novel coding technique to increase the efficiency of IDR pictures by coding at lower spatial resolutions. This is achieved while enabling inter-coded pictures to be coded efficiently at the original High Resolution (HR). This differs from existing methods as resampling is performed within the codec and not as pre- and post-processing operations. This means that the proposed coding technique can not be applied to standard decoders but the spatial resolution of intra-coded pictures can be adapted independently according to content and the available bandwidth to provide a greater level of rate-distortion performance. Results show that the proposed method increases rate-distortion performance up to 11.5% within the range 100kbps-5Mbps for HD sequences.

ACKNOWLEDGMENTS

I would first like express my appreciation to everyone involved in the Centre of Doctoral Training in Communications. I am very grateful and proud to have been given the opportunity to undertake my PhD with all the expertise, experience and support that the centre provides. This appreciation extends to the EPSRC and all the industrial partners, especially BAE Systems who funded my own research.

I also have great appreciation for my many supervisors that have provided me with guidance throughout my PhD. Dr Dimitris Agrafiotis, Dr Nick Easton, Dr Angela Doufexi, Professor David Bull and Dr Paul Hill, although some of these people were only officially my supervisors for a relatively short time, I am thankful for their invaluable input and insights that they shared with me during that time. In particular I would like to thank Dr Dimitris Agrafiotis without whom this thesis would not have been possible. His guidance, encouragement, patience and enthusiasm in my work have been greatly appreciated. He also provided me with lots of good ideas, many of which helped shaped my thesis into what it is. Special thanks also goes to Dr Nick Easton at BAE Systems who made regular visits to Bristol all the way from Chelmsford to attend meetings, discuss the work and offer advice – I really appreciate it.

I am indebted to my many friends and colleagues in the department for providing a stimulating and fun working environment. I am especially grateful to Dr Umar Ozgunalp and Nicolas Epaminonda who initially welcomed me into the Visual Information Laboratory and became my reliable *coffee buddies*.

I wish to thank my family for all their support. My parents for always encouraging me to pursue my own career path (and not forcing me to become a dairy farmer). My brother, Scott, for believing in my abilities and encouraging me to aim high. Also my sister, Selina – thanks for giving me a place to crash in Bristol when my funding ran out.

DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author

SIGNED:

DATE:.....

PUBLICATIONS

B. Hosking, D. Agrafiotis, D. Bull, and N. Easton. Spatial resampling of IDR frames for low bitrate video coding with HEVC. IS&T/SPIE Electronic Imaging, (2015)

B. Hosking, D. Agrafiotis, D. Bull, and N. Easton. An adaptive resolution rate control method for Intra Coding in HEVC. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2016).

B. Hosking, D. Agrafiotis, D. Bull, and N. Easton. Enhancement of intra-coded pictures for greater coding efficiency. Picture Coding Symposium (PCS), (2016) - Paper Accepted

B. Hosking, D. Agrafiotis, D. Bull, and N. Easton. Adaptive Resolution of Intra-Pictures for Bandwidth Constrained Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, (2016) - Submitted

TABLE OF CONTENTS

	Page
List of Tables	xiii
List of Figures	xiv
1 Introduction	1
1.1 Motivation	1
1.1.1 Project Proposal	2
1.1.2 Problem Statement	2
1.1.3 Aims and Objectives	4
1.2 Thesis Outline	4
1.3 Contributions	6
2 Video Coding and Sampling	7
2.1 Sampling Fundamentals	8
2.1.1 Downsampling	10
2.1.2 Upsampling	11
2.2 Colour Space	11
2.2.1 RGB	11
2.2.2 YCrCb	11
2.2.3 Conversion	13
2.3 Video Compression	13
2.3.1 Transforms	13
2.3.2 Quantisation	15
2.3.3 Inverse Quantisation and Transformation	17
2.3.4 Run-Length Coding & Entropy Coding	17
2.3.5 Prediction	18
2.4 H.265 - High Efficiency Video Coding	21
2.4.1 Coding Units	22
2.4.2 Prediction	24
2.4.3 Performance Comparisons	25

TABLE OF CONTENTS

2.5 Scalable Video Extensions	26
2.5.1 Spatial Scalability	27
2.6 Test Sequences	28
2.7 Prior Art	30
2.8 Performance Evaluation	33
2.8.1 Peak Signal-to-Noise Ratio	33
2.8.2 Structural Similarity Index Metric	33
2.8.3 Best Rate-Distortion Curves	33
2.8.4 BD-PSNR	34
2.9 Chapter Summary	34
3 Spatial Resampling of IDR Pictures	37
3.1 Performance of Resampled Intra Coded Pictures	38
3.1.1 Intra Coding at Low Spatial Resolutions	39
3.1.2 Analysis of Intra Coded Results	42
3.2 Visual Comparisons of Intra Coded Pictures	44
3.3 Methodology and Proof-of-Concept	45
3.3.1 Proposed Coding Technique	45
3.3.2 Experimental Results	46
3.4 Chapter Summary	52
4 Adaptive Resolution R-Q Model for Intra-Coded Pictures	55
4.1 Complexity Measures	56
4.1.1 Average Gradient Per Pixel	58
4.1.2 Edge Point Ratio	60
4.2 Single Resolution R-Q Model	62
4.2.1 Initial QP Selection	64
4.2.2 Practical Single Resolution Algorithm	65
4.2.3 Proposed Modified Updating Procedure	65
4.3 Adaptive Resolution R-Q Model	67
4.3.1 Best Coded Picture Selection Problem	69
4.3.2 Best Coded Picture Selection Method	72
4.3.3 Computational Complexity	73
4.4 Experimental Results	75
4.5 Chapter Summary	77
5 Techniques for Spatial Resampling of IDR Pictures	79
5.1 Spatially Invariant Filtering	80
5.1.1 Nearest Neighbour	81

TABLE OF CONTENTS

5.1.2	Bilinear	82
5.1.3	Bicubic	83
5.1.4	SHVC Sampling	83
5.2	Comparison of Sampling Techniques	87
5.2.1	Objective Quality Analysis	87
5.3	Spatially Variant Filtering	90
5.3.1	Joint Bilateral Upsampling	91
5.4	Super Resolution & Enhanced Sampling Techniques	93
5.4.1	Iterative Back Projection	93
5.4.2	Interpolation Dependent Image Downsampling	96
5.4.3	Cross-Scale Self-Similarity	98
5.5	Comparison of Enhanced Sampling Techniques	101
5.5.1	Objective Comparison of IBP and IDID	102
5.5.2	Visual Comparison of IBP and IDID	102
5.5.3	Region Objective Quality Performance of IDID and IBP	104
5.5.4	IBP vs IDID for Enhancement of Intra Coded Sequences	107
5.6	Intra Coding Performance of IDID Resampled Sequences	111
5.7	Chapter Summary	115
6	Adaptive Resolution Coding	117
6.1	Rate Control Modifications	118
6.1.1	The Rate Control Problem	118
6.1.2	Adapting the Bit-Budget	120
6.1.3	Performance After Bit-Budget Modifications	122
6.2	Adaptive Resolution of IDR Pictures	124
6.2.1	Performance Factors	124
6.3	Experimental Results	125
6.3.1	Average Rate-Disotrtion Performance	127
6.3.2	Temporal Variation of Quality & Bitrate	128
6.3.3	Percentage of Selected Resolutions	135
6.4	Chapter Summary	138
7	Mixed Resolution Enhancement Coding	141
7.1	Enhancement of Resampled Intra-Coded Pictures	143
7.2	Experimental Results	145
7.2.1	Test Data and Coding Parameters	145
7.2.2	Average Bit Savings	146
7.2.3	Regional Quality Evaluation	148
7.3	Mixed Resolution Region-of-Interest Coding	150

TABLE OF CONTENTS

7.4 Chapter Summary	152
8 Conclusions & Further Work	155
8.1 Conclusions & Discussions	155
8.2 Further Work	158
8.2.1 Adaptive Rate Control Algorithm for Varying Video Content	158
8.2.2 Spatially Invariant IDID	159
8.2.3 Adaptive Resolution of Intra-Coded Data within Inter-Coded Pictures . .	159
8.2.4 Evaluation of the Adaptive Resolution R-Q model for Inter-Coded Video .	159
8.2.5 More Effective Models for Picture Complexity	160
Bibliography	161
Appendices	169

LIST OF TABLES

TABLE	Page
2.1 Spatial resolution formats for video	9
2.2 The result of <i>Run-Length</i> coding for the given example	17
2.3 Comparison of video coding standards for streaming applications	25
2.4 Comparison of video coding standards for entertainment applications	26
3.1 List of spatial resolutions and corresponding scale factors	42
3.2 Average bitrate savings for intra-coded and resampled sequences	43
3.3 Statistical results of temporal variation of quality - InToTree 1000kbps	48
4.1 Average coding time for different coding configurations	74
4.2 Average reduction of computation time when coding at lower spatial resolutions . . .	75
4.3 Adaptive Resolution R-Q model results	76
5.1 SHVC downsampling coefficients - low range	84
5.2 SHVC downsampling coefficients - high range	84
5.3 SHVC Luma coefficients for interpolation	85
5.4 SHVC Chroma coefficients for interpolation	85
5.5 Objective quality comparison between Bicubic and SHVC sampling - Group A	88
5.6 Objective quality comparison between Bicubic and SHVC sampling - Group B	89
5.7 Objective quality comparison between Bicubic and SHVC sampling - Group C	89
5.8 Objective quality comparison between Bicubic and SHVC sampling - average results .	90
5.9 Average bitrate savings for each resampling method compared to standard HEVC . .	110
5.10 Average bitrate savings for IDID resampled and coded sequences	111
6.1 PSNR statistics for different reductions of the IDR bit-budget	120
6.2 BD-PSNR average savings for different reductions of the IDR bit-budget	122
6.3 BD-PSNR average savings using the Adaptive Resolution coding technique	125
7.1 Average PSNR difference of intra-coded techniques compared to standard HEVC . .	147
7.2 Coded bits and corresponding PSNR for different EL QPs	148

LIST OF FIGURES

FIGURE	Page
1.1 Example of the problem that coding at low bitrates produces	3
2.1 Illustration of spatial and temporal sampling	9
2.2 Downsampling theory	10
2.3 Upsampling theory	11
2.4 YCrCb colour space sampling	12
2.5 Surface plot of DCT coefficients example	15
2.6 Method of reordering coefficients in a <i>zig-zag</i> fashion	18
2.7 Inter prediction by exploiting temporal redundancies	19
2.8 Basic GOP structure commonly used in inter-coding configurations	20
2.9 Block diagram of the H.265/HEVC standard	21
2.10 Example of CTU partitioning into smaller CUs	22
2.11 Illustration of the data contained within a Coding Tree Unit (CTU)	22
2.12 Illustration of the data contained within a Coding Unit (CU)	23
2.13 Prediction Block (PB) Configurations	23
2.14 Transform Block (TB) Configurations	24
2.15 Directional modes for intra prediction in H.265/HEVC	25
2.16 Inter-layer prediction in SHVC with two spatial layers	26
2.17 Inter-layer prediction in SHVC with two spatial layers and an IDR picture	27
2.18 Spatial and temporal information indexing for test sequences	28
2.19 Illustration of producing best rate-distortion curve	34
3.1 Rate-distortion - BlueSky intra-coded and resampled using Bicubic	38
3.2 Rate-distortion - InToTree intra-coded and resampled using Bicubic	39
3.3 Rate-distortion - OldTownCross intra-coded and resampled using Bicubic	39
3.4 Rate-distortion - Station2 intra-coded and resampled using Bicubic	40
3.5 Rate-distortion - Tennis intra-coded and resampled using Bicubic	40
3.6 Rate-distortion - ParkJoy intra-coded and resampled using Bicubic	41
3.7 Rate-distortion - Jockey intra-coded and resampled using Bicubic	41
3.8 Rate-distortion - ReadySetGo intra-coded and resampled using Bicubic	42

3.9	Visual comparison of an intra-coded picture from the sequence Tennis	43
3.10	Visual comparison of an intra-coded picture from the sequence Jockey	44
3.11	Block diagram of the proposed technique	45
3.12	Rate control performance - InToTree 1000kbps	47
3.13	Comparison of temporal variation of quality - InToTree 1000kbps	48
3.14	Comparison of bitrate variation between standard and the proposed coding methods .	49
3.15	Rate-distortion - InToTree comparison between standard and proposed method . . .	49
3.16	Rate-distortion - OldTownCross comparison between standard and proposed method	50
3.17	Rate-distortion - Station2 comparison between standard and proposed method . . .	50
3.18	Rate-distortion - ParkJoy comparison between standard and proposed method . . .	51
3.19	Rate-distortion - Tennis comparison between standard and proposed method . . .	51
4.1	Variation of complexity (actual and measured) over time	57
4.2	Correlation between BPP and GPP for sequence ParkJoy	59
4.3	R^2 correlation over varying QP for complexity measure GPP	60
4.4	R^2 correlation over varying QP for complexity measures EPR and GPP	61
4.5	Optimal EPR threshold over varying QP	62
4.6	Modelling rate over Q_{step} for standard intra-coding	63
4.7	Linear models for the initial prediction of the best QP given a bit-budget	64
4.8	Illustrative rate-distortion curve	69
4.9	The Adaptive Resolution problem - possible picture selection scenarios	70
4.10	The Adaptive Resolution problem with known data point	71
4.11	The Adaptive Resolution problem with rate-distortion models	72
4.12	The Adaptive Resolution selection method	73
4.13	Average computational complexity of intra-coded pictures at different spatial resolutions	74
4.14	Bitrate and PSNR variation for each R-Q model	77
5.1	<i>Lena</i> test image (RGB)	80
5.2	<i>Lena</i> test image (luminance channel)	81
5.3	Nearest Neighbour interpolated image	81
5.4	Bilinear interpolated image	82
5.5	Bicubic interpolated image	83
5.6	Interpolated image using SHVC sampling functions	86
5.7	Thumbnails of still images used for comparison of sampling techniques	87
5.8	JBU multiplication of the spatial and range filters	91
5.9	Scaled vector graphic of image <i>Lena</i>	92
5.10	Block diagram of the IBP upsampling technique	94
5.11	PSNR for each iteration of the IBP technique	95
5.12	Block diagram of the IDID downsampling technique	96

LIST OF FIGURES

5.13	PSNR for each iteration of the IDID technique	97
5.14	Average PSNR plotted for each iteration of IBP and IDID	102
5.15	Sampling comparisons ($\times 2.5$) using the test image from the sequence <i>PlasmaFree</i> . . .	102
5.16	Sampling comparisons ($\times 2.5$) using the test image from the sequence <i>TreeWills</i>	103
5.17	Sampling comparisons ($\times 1.5$) using the test image from the sequence <i>TreeWills</i>	104
5.18	Local PSNR colourmap for the test image from sequence <i>TreeWills</i> scaled by $\times 2.5$. . .	105
5.19	Local PSNR colourmap for the test image from sequence <i>TreeWills</i> scaled by $\times 1.5$. . .	105
5.20	Local PSNR improvement of enhanced sampling techniques when scaled by $\times 2.5$. . .	106
5.21	Local PSNR improvement of enhanced sampling techniques when scaled by $\times 1.5$. . .	106
5.22	Direct comparison of local PSNR for IBP and IDID	107
5.23	Block diagram of the IBP upsampling technique within a video codec	108
5.24	Block diagram of the IDID downsampling technique within a video codec	109
5.25	Rate-distortion curves showing performance of resampling techniques when coding .	110
5.26	IDID rate-distortion curves for each sequence and scale factor	112
5.27	Best IDID rate-distortion curve for each sequence given each scale factor	113
6.1	The effect of reducing the IDR bit-budget	119
6.2	Comparison of PSNR variation for different reductions of the IDR bit-budget	120
6.3	Coding structure of an inter-coded picture with different reductions of the IDR bit-budget	121
6.4	Rate-distortion performance with different reductions of the IDR bit-budget	123
6.5	Rate-distortion performance of the Adaptive Resolution coding technique	126
6.6	Illustration of high temporal correlation within the sequence <i>Squirrel</i>	128
6.7	Temporal variation of performance for sequence <i>Squirrel</i> . Target bitrate: 250kbps .	129
6.8	Temporal variation of performance for sequence <i>Squirrel</i> . Target bitrate: 500kbps .	131
6.9	Temporal variation of performance for sequence <i>Squirrel</i> . Target bitrate: 1000kbps .	132
6.10	Temporal variation of performance for sequence <i>Squirrel</i> . Target bitrate: 4000kbps .	134
6.11	Percentage of each spatial resolution selected per sequence	136
7.1	Illustration of Mixed Resolution coding	142
7.2	Block diagram of Mixed Resolution Enhancement Coding	143
7.3	Enhancement of an intra-coded picture using inter-layer prediction	144
7.4	Test images and corresponding picture complexity	145
7.5	Block based local PSNR colour map representing a resampled intra-coded picture . .	148
7.6	Blocks based local PSNR difference with the addition of an EL	149
7.7	ROI mapping and block based local PSNR difference by applying ROI approach	151
7.8	Visual comparisons between standard HEVC and Mixed Resolution ROI coding	152

ACRONYMS

AI All Intra

AMVP Advanced Motion Vector Prediction

ASP Advanced Simple Profile

AVC Advanced Video Coding

BL Base Layer

BLA Broken Link Access

BPP Bits per Pixel

BVI Bristol Vision Institute

CB Coding Block

CBR Constant Bitrate

CCITT International Telegraph and Telephone Consultative Committee

CD Compact Disk

CHC Conversational High Compression

CIF Common Intermediate Format

CRA Clean Random Access

CTB Coding Tree Block

CTU Coding Tree Unit

CU Coding Unit

DCT Discrete Cosine Transform

DVD Digital Versatile Disk

LIST OF FIGURES

EL Enhancement Layer

EPR Edge Point Ratio

GOP Group of Pictures

GPP Gradient per Pixel

HD High Definition

HEVC High Efficiency Video Coding

HLP High Latency Profile

HP High Profile

HR High Resolution

HVS Human Visual System

IBP Iterative Back Projection

IDCT Inverse Discrete Cosine Transform

IDID Interpolation Dependent Image Downsampling

IDR Instantaneous Decoding Refresh

IRAP Intra Random Access Point

ISDN Integrated Services Digital Network

ITU-T International Telecommunications Union - Telecommunication Standardisation Section

JBU Joint Bilateral Upsampling

JCT-VC Joint Collaborative Team on Video Coding

JPEG Joint Photographic Experts Group

LCU Largest Coding Unit

LD Low Delay

LR Low Resolution

MISR Multi-Image Super Resolution

MP Main Profile

MPEG Moving Picture Experts Group

MSE Mean Squared Error

NAL Network Abstraction Layer

NN Nearest Neighbour

PB Prediction Block

PU Prediction Unit

PSNR Peak Signal-to-Noise-Ratio

QP Quantisation Parameter

QCIF Quarter - Common Intermediate Format

RA Random Access

RDO Rate Distortion Optimisation

ROI Region of Interest

RPB Reference Picture Buffer

R-Q Rate-Quantisation

RSS Residual Sum of Squares

SAD Sum of Absolute Difference

SD Standard Definition

SHVC Scalable HEVC

SI Spatial Information

SISR Single-Image Super Resolution

SNR Signal-to-Noise Ratio

SPS Sequence Parameter Set

SR Super Resolution

SSE Sum of Squared Error

SSIM Structural Similarity Index Metric

LIST OF FIGURES

SVC Scalable Video Coding

TB Transform Blocks

TI Temporal Information

UAV Unmanned Aerial Vehicle

UHD Ultra High Definition

VCEG Video Coding Experts Group

VLC Variable Length Coding

INTRODUCTION

Over the last few decades wireless technology has flourished and has re-defined the way in which we communicate on a global scale. It has also enabled us to transfer data in ways not previously possible with wired technology, such as over a mobile network and even retrieving images from other planets.

The first transatlantic radio communication was made by Guglielmo Marconi on the 12th December 1901 [46]. The signal was transmitted in Morse code from Poldhu, a small area situated in southern Cornwall on the Lizard Peninsula, United Kingdom, and was received at Signal Hill in Newfoundland, Canada. Since then, many advancements in wireless technology have been achieved leading to many new technological possibilities. As a result there is an increasing demand for research in these areas.

1.1 Motivation

Video compression and coding is a necessary part of the multimedia services used for storage and transmission of visual data. An uncompressed Standard Definition (SD) video (720×576 at 25Hz and 8 bits/pixel) would equate to 248.832 Mbps. Compact Disk (CD) and Digital Versatile Disk (DVD) have about 700MB and 4.7GB capacity, respectively. Thus, a CD would contain less than three seconds and a DVD would contain less than twenty seconds of uncompressed SD video data. Therefore, in order for DVDs to become a viable format for digital video it was necessary to first compress the data. As a result of compression, the DVD format has become a world wide success.

Further enhancements to compression algorithms and optical disk technology have lead to the switch to High Definition (HD) video and also the emerging Ultra High Definition (UHD) video.

For many applications higher orders of compression are required when transmitting over a wireless channel. Like an optical disk, the frequency spectrum has a finite capacity. Over the last few decades the mobile industry has grown exponentially. In addition to voice calls, data transfer such as video requires bandwidth. In order to accommodate the vast number of users accessing a network at any given time various methods have been devised to improve bandwidth efficiency. Bandwidth efficiency is determined by the rate at which a communications system is capable of transmitting information given a fixed bandwidth. It is especially important when bandwidth is limited and lossless compression algorithms are incapable of reducing the data bitrate to a level below or equal to the channel bitrate. In such circumstances more intelligent solutions are required to maximise bandwidth efficiency.

1.1.1 Project Proposal

The motivation behind this work is to find solutions to the problems that arise within aerial-to-ground video communications. The proposed scenario involves the operation of an Unmanned Aerial Vehicle (UAV), also commonly known as a *Drone*, to survey the ground from an aerial view. It is required that the captured video is transmitted to a Command Centre where it can be analysed and enable decisions to be made regarding the further operation of the UAV in real-time. Due to severely restricted bandwidths, video bitrates will need to be limited. For surveillance systems it is imperative that sufficient information can be gained from the captured scene without high levels of latency and a large number of errors. Under such conditions video content produced by current systems is usually limited to a low spatial resolution and a low level of visual quality. The proposed project aims to find solutions to enable more important information from the original scene to be coded and transmitted at low bitrates. This effectively means providing greater coding efficiencies than the current standards and possibly regional prioritisation of image content. Ideally, HD or UHD video would be captured and transmitted to provide the necessary level of detail. The target bitrate range is between 100kbps-5Mbps for HD sequences and between 100kbps-9Mbps for UHD sequences. The level of visual quality required is not explicitly defined as results will be highly affected by video content. However, the proposed solution should reduce bitrate requirements for a wide range of quality levels, ideally for an average PSNR up to 40dB.

1.1.2 Problem Statement

Wireless communication links may be subject to errors due to a low Signal-to-Noise-Ratio (SNR) or interference from other communication systems. Video codecs implement tools to reduce the impact of these errors within the video bitstream but this comes at the cost of reduced compres-

sion efficiency. The latest video coding standard, High Efficiency Video Coding (HEVC), provides significantly higher compression efficiencies at HD and UHD resolutions compared to previous standards [49]. Despite increased coding performance, bandwidth restricted video applications often fail to achieve an acceptable and useful level of video quality, warranting the need for additional coding and processing.

Instantaneous Decoder Refresh (IDR), Clean Random Access (CRA) and Broken Link Access (BLA) are all forms of Intra Random Access Point (IRAP) pictures. IRAP pictures provide points in the bitstream where decoding can be initialised and therefore must remain independent from any other picture. Intra-coded pictures are necessary for all video coding applications; a HEVC compliant bitstream must at the very least start with an IRAP picture. IDR pictures are a special case as they refresh the decoding process thus preventing themselves and future decoded pictures having any dependency on previously decoded pictures. Frequent coding of IDR pictures can provide a more error resilient and robust video bitstream at the expense of reduced compression efficiency [23]. At low bitrates this results in a low average level of video quality that is also highly variant over time due the high compression of the remaining inter-coded pictures contained within the Group of Pictures (GOP). The reason for this is illustrated in Figure (1.1) which shows how the encoder allocates bits and corresponding Quantisation Parameter (QP) values over five GOPs of a HD sequence. In this example, the IDR pictures consume more than ten orders of magnitude more coded data compared to the inter-coded B-pictures and as a result a large portion of the information contained in later frames is lost. There is therefore a trade-off, which becomes more of an issue at low bitrates; IDR pictures should be of high quality to provide a good reference for future coded frames but this can consume the majority of the available bits allocated to the GOP resulting in high quantisation of the remaining frames in order to meet the target bitrate, which may reduce the overall quality.

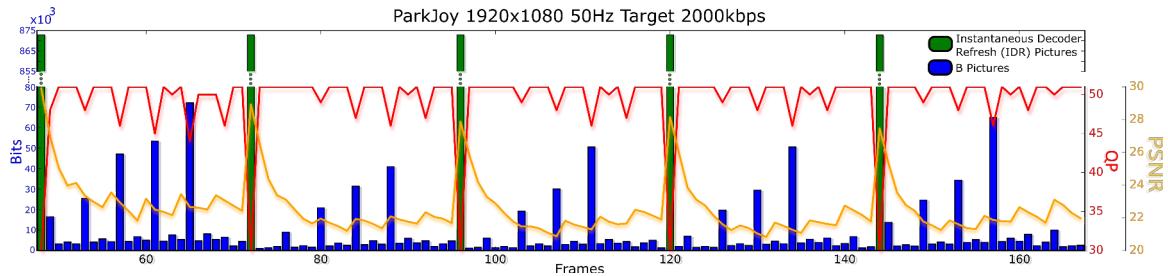


Figure 1.1: Variation of the number of coded bits, and corresponding QP and PSNR for sequence *ParkJoy* 1920×1080 coded using the Low Delay configuration (50Hz Target 2000kbps Intra Period 24). At low bitrates, IDR pictures consume a large portion of the available number of bits resulting in high quantisation of the remaining frames in the Group of Pictures (GOP)

1.1.3 Aims and Objectives

Previous work has shown that spatial resampling of video sequences, or even single images, can provide better compression performance than simply coding at the original High Resolution (HR) [7, 12, 13, 42, 48, 59, 74]. The term *Resampling* is used throughout this thesis to indicate an image that has been downsampled and then upsampled back to the original resolution. In the aforementioned work, coding is applied at the Low Resolution (LR) to reduce coding distortion. In [7], Bruckstein et al. show that at low bitrates, an image downsampled, compressed using JPEG and later interpolated, produces better results than an image compressed at the original HR. Wu et al. demonstrate that coding oversampled frames is not only a waste of resources but can also be counterproductive to image quality given a tight bit-budget [74]. Dong et al. proposed resampling the entire video sequence after determining the optimal scale factor by minimising the overall distortion caused by downsampling and coding [12, 13]. In [48], Nguyen et al. proposed a method of adapting the scale factor and the quantisation step size according to spatial content. More in-depth detail on the prior state-of-the-art is given in Chapter 2.

This work looks at methods of improving the compression efficiency of intra-coded pictures to alleviate some of the distortion introduced at low bitrates when quantisation is high. An error resilient and low latency coding structure is assumed to account for the requirements of an aerial-to-ground communications link transmitting video captured from a UAV. To this end, the Low Delay configuration is used for all inter-coded results utilising a frequent intra period with intra frames coded as IDR pictures. In order to improve rate-distortion performance, the resolution of IDR pictures is adapted according to the amount of bandwidth available and the rate-distortion properties that coding at each spatial resolution provides. This work focuses on resampling IDR pictures only as inter-coded pictures are already well compressed, especially in HEVC. This means that the majority of the frames will be coded at the original HR. This differs from previous work where the entire sequence is resampled and coded at a lower spatial resolution and therefore is no longer a true HD or UHD sequence. As content can vary over time, the overall aim is to produce a system that determines the best scale factor and QP for each IDR picture independently and therefore maximise rate-distortion performance.

1.2 Thesis Outline

There are various approaches to tackling the problem of enabling effective transmission of video under non-ideal channel conditions. A great deal of research is carried out at the Physical Layer to increase spectral efficiency and provide greater bandwidths per channel, thus avoiding many of the issues that arise when coding video at low bitrates. This thesis aims to tackle these issues directly by assuming that only bandwidth constrained communications links are available.

Chapter 2 gives an overview of video coding and compression. The aim is to provide the reader with knowledge of the relevant principles required for understanding the work presented in later chapters.

Chapter 3 describes the initial version of the proposed technique for providing greater coding efficiency at low bitrates. Results provided in this chapter confirm that coding IDR pictures at lower spatial resolutions enables greater rate-distortion performance. Work presented in this chapter was published in [23].

Chapter 4 builds upon the work presented in Chapter 3 by demonstrating an adaptive resolution rate control method for intra-coding. This work enables adaptation of spatial resolution for each intra-coded picture. Rate-distortion performance is therefore further improved by minimising the combine coding and resampling distortion for each frame independently. Work presented in this chapter was published in [25].

Chapter 5 contains an analysis of image sampling techniques. Previous chapters utilise Bicubic to generate the necessary LR pictures and also for reconstructing the HR pictures after coding. Applying better resampling techniques will reduce the overall distortion of intra-coded pictures and also provide greater quality reference pictures when applying inter-prediction. Some of the work presented in this chapter has been submitted as part of a journal in [24].

Chapter 6 combines elements from the previous three chapters to demonstrate the potential rate-distortion gain that the proposed method provides. Unlike the work presented in Chapter 3, the resolution of IDR pictures is adapted using the method described in Chapter 4. Also, the IDID sampling technique is used as described in Chapter 5. This chapter contains a large portion of the work that is submitted within the journal [24].

Chapter 7 presents a novel coding approach that enables resolution to be adapted for different regions of an intra-coded picture. Content may change throughout an image, as well as temporarily throughout a video sequence. For this reason, a Mixed Resolution coding approach is described and initial results are provided. This framework also lends itself to Region of Interest (ROI) coding as areas of high priority can be coded at a higher spatial resolution to maintain the fidelity of high frequency content such as textures and edges. The work presented in this chapter has been accepted for publication [26].

Chapter 8 contains conclusions regarding the effectiveness of the proposed Adaptive Resolution coding technique for low bitrate video coding, and also the other contributions provided in

this thesis. Additionally, details of further work are given describing areas of work that would provide further benefit and complement the work already achieved.

1.3 Contributions

The main contributions can be summarised as follows:

- Demonstration that coding intra pictures at lower spatial resolutions and upsampling after decoding can increase rate-distortion performance at low bitrates. This is shown to be true for different sampling filters and techniques. An Interpolation Dependent Image Downsampling (IDID) technique is presented, based on Iterative Back Projection (IBP), that is shown to provide the greatest rate-distortion performance.
- The focus of this thesis is the proposed novel coding method for which IRAP pictures such as IDR pictures are resampled, before and after coding, to provide greater levels of efficiency without resampling inter-coded pictures. This enables adaptation of the spatial resolution for varying picture content, unlike previous methods that apply a fixed scale factor for the entire sequence.
- An Adaptive Resolution R-Q model is presented that enables accurate prediction of the best QP for each scaled picture. The best coded and reconstructed picture is then selected given a set of defined criteria.
- Demonstration that inter-layer prediction can increase rate-distortion performance of intra-coded pictures. Scalable video coding extensions, such as SHVC, have much of the functionality required but they are not used for this purpose. The work presented in this thesis combines Adaptive Resolution coding with inter-layer prediction to demonstrate that further gains in rate-distortion performance are possible.

VIDEO CODING AND SAMPLING

Video coding and compression is a necessary process for the transmission or storage of digital video. RAW video may be captured and used by professionals who wish to perform pre-processing techniques but ultimately all video will be coded before commercial use. Video coding standards are formed as a compilation of algorithms, some providing lossless compression but greater compression efficiencies can be achieved using lossy compression. Research on video compression has often received inspiration from the field of Cognitive Psychology as information that can not be perceived by the Human Visual System (HVS) is effectively redundant and can be removed without considerable loss of quality

The first video coding standard was termed H.120 and was produced by the International Telecommunications Union - Telecommunication Standardisation Section (ITU-T) in 1984 and revised in 1988, although at the time the ITU-T were known as the International Telegraph and Telephone Consultative Committee (CCITT). After implementation it was realised that the codec was incapable of producing video quality to an acceptable level and as a result the ITU-T began working on the next standard in succession through the Video Coding Experts Group (VCEG). H.261 was the first of the H.26x family of standards, published in 1990. The original design goal was for transmission over Integrated Services Digital Networks (ISDNs) with supported frame size formats CIF and QCIF using 4:2:0 sampling (see Section 2.1). H.261 was very successful but over time new applications demanded increased coding efficiency for larger spatial resolutions.

From the success of the image compression standard JPEG, produced by the Joint Photographic Experts Group, and the H.261 video coding standard, the Moving Picture Experts Group (MPEG) formed in 1988 and commenced work on producing a new standard supporting higher coding efficiencies. MPEG-1 was later published in 1993. Success of the standard was hindered by its coding limitations; MPEG-1 supports just one profile, one sampling configuration and lacks interlaced video coding functionality. To address these problems, MPEG-2 was published in 1996 and became highly successful. Digital television and optical disk technology adopted MPEG-2 as its video coding format. The video part of MPEG-2 was jointly developed by the VCEG and the MPEG, thus it also became known as H.262/MPEG-2 Part 2.

H.263 and MPEG-4 Part-2 were published in the late 1990's by the VCEG and MPEG, respectively, to address the need of newer and better technology. In 2001 the two groups combined their efforts and formed the Joint Collaborative Team on Video Coding (JCT-VC). By 2003 the JCT-VC published the H.264/MPEG-4 Advanced Video Coding (AVC) standard. At the time of writing, H.264/AVC is the most widely used video coding standard for HD video technology. However, H.265/*High Efficiency Video Coding* (HEVC) was released in January 2013 which provides even greater compression efficiencies at HD and UHD resolutions, and in time will replace AVC.

2.1 Sampling Fundamentals

To produce a digital representation of a natural scene over time, a camera is required to sample in both the spatial and temporal domain. A natural scene is temporally and spatially continuous and therefore sampling produces a corresponding discrete signal. The accuracy of the captured digital video is dependent on the sampling rate, in both space and time, and the bitrate in terms of the number of bits used to represent each pixel – also known as the dynamic range. Generally it is important to capture the scene with a high sampling rate and high bitrate to ensure high quality. This is because it is not possible to improve the accuracy of the digital video after it has been captured. Sampling can also be performed on a discrete signal, such as when capturing video from a computer generated scene or an already existing digital video. In such cases, applying a sampling rate greater than the number of pre-existing samples will result in oversampling, see Chapter 5 for further discussion on sampling rates and sampling techniques. Figure 2.1 provides an illustration of how sampling is performed.

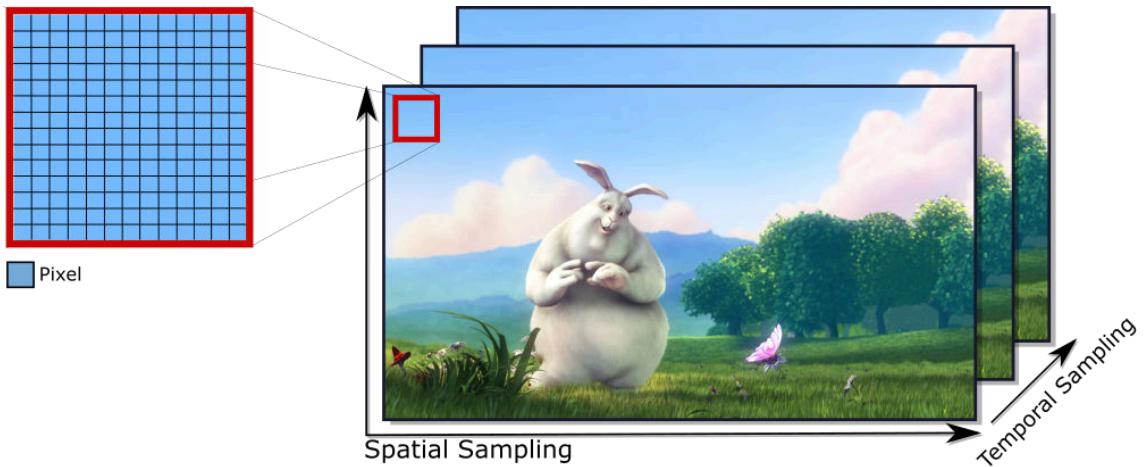


Figure 2.1: Spatial and temporal sampling is required when capturing a natural or computer generated scene. Each sample in the spatial domain is called a *Pixel* and each sample in the temporal domain is called a *Frame*, which is composed of many pixels

Spatial sampling is performed when capturing both a video or a single image. A single captured image is known as a frame within a video sequence. As Figure (2.1) illustrates, each frame is constructed of many spatial samples known as pixels. In the case of colour images, each pixel will have 3 components in order to produce representative luminance and chrominance intensities – see Section 2.2 for more details on Colour Space. Increasing the spatial sampling rate increases the number of pixels within a given area, resulting in greater spatial resolution. Table (2.1) provides details of resolutions used for standard image formats.

Table 2.1: Spatial resolution formats for video

Format	Resolution
Sub-QCIF	128×96
QCIF	176×144
CIF	352×288
4CIF	704×576
720p	1280×720
1080p/HD	1920×1080
4K/UHD	3840×2160

Temporal sampling is applied during video capture only. A frame is sampled spatially during one instance of time. Temporal sampling therefore applies this process by capturing images as time progresses within the scene. The amount of images captured by the camera for any given amount of time is determined by the temporal sampling rate, also known as the frame rate. Displaying these images in a linear fashion can give the illusion of movement – which is where the American word *Movie* comes from, it is an abbreviation of *Moving Pictures*. Generally a higher

frame rate can increase the perceived smoothness of motion within the scene. The necessary frame rate for consistently smooth motion is therefore determined by the rate of change within the scene and also the limitations of the HVS. Typically frame rates of 25Hz - 30Hz are used, however, some recent films have been mastered at 50Hz and this is expected to increase as compression and communications technology improves. As the frame rates reduces, motion appears unnatural and *jerk*. There is therefore a trade off between ensuring perceived smooth motion and the required data rates for higher frame rates. The work presented in this thesis involves resampling already discrete 2-dimensional signals at different sampling rates. As initially defined in Chapter 1, the term *resampling* is used to describe an image that is downsampled and upsampled. The process and implications of these operations are provided in Sections 2.1.1 and 2.1.2.

2.1.1 Downsampling

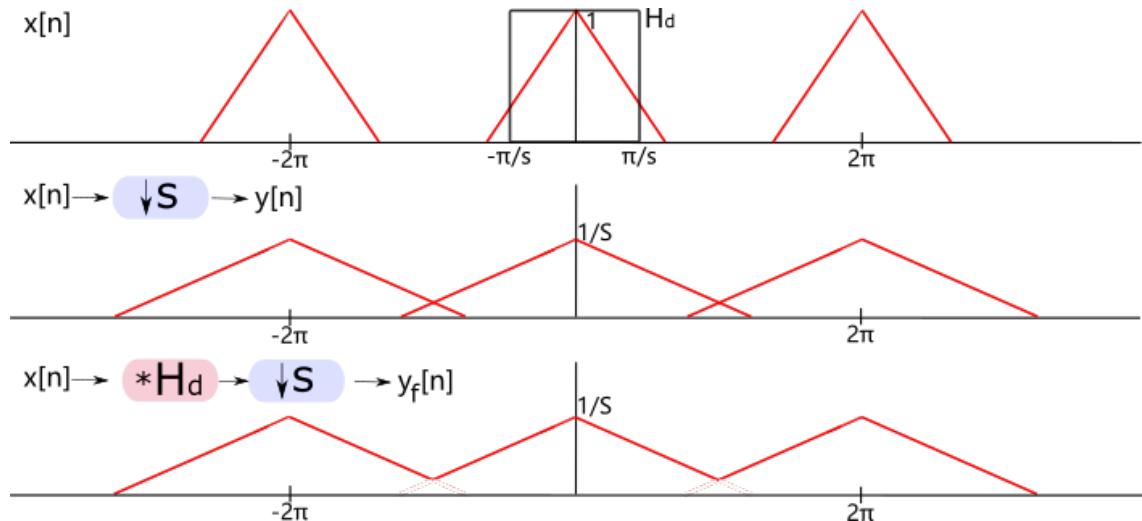


Figure 2.2: Representation of the downsampling process in the frequency domain.

Downsampling effectively reduces the amount of information contained within the original HR image. In the normalised frequency domain, downsampling stretches the spectral extent of the original signal within the fundamental frequency range, as illustrated in Figure (2.2). Low pass filtering is applied prior to reducing the sampling rate in order to minimise overlapping spectral copies. Downsampling by a large factor can lead to significant loss of high frequency content, such as sharp edges and textures, as the normalised cut-off frequency of the filter applied needs to be further restricted if aliasing is to be reduced.

2.1.2 Upsampling

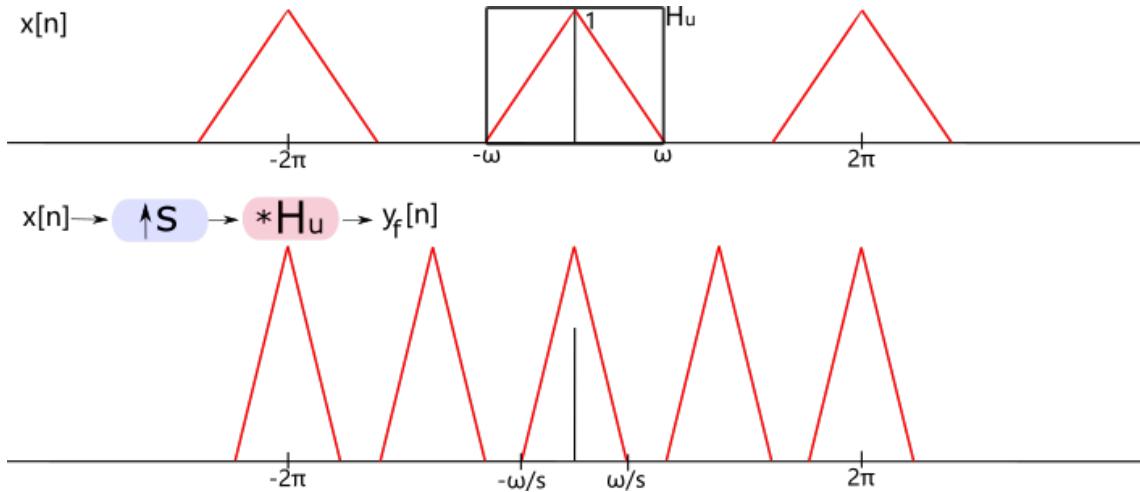


Figure 2.3: Representation of the upsampling process in the frequency domain.

Upsampling follows the reverse process of downsampling by applying a low pass filter after increasing the sampling rate. This effectively means that the new signal is initially composed of zeros and the low pass filter is applied to remove the spectral copies during the interpolation process to calculate the missing sample values. As a result of the higher sampling rate, the spectral extent of the original signal compresses in the normalised frequency domain, as illustrated in Figure (2.3).

2.2 Colour Space

A monochrome image requires just one value to represent the luminance of each pixel. For colour images, three values are required, a combination of all three values can describe the luminance and chrominance of each pixel. How these values are represented forms what is known as the *Colour Space*.

2.2.1 RGB

A combination of the three primary colours of light (Red, Green and Blue) can be used to create any other colour. The RGB colour space requires one value for each colour component to represent the intensity of that colour at each pixel. Therefore each component in the RGB colour space is equally important, thus each colour channel is stored at the same resolution.

2.2.2 YCrCb

The HVS is less sensitive to changes in chrominance than it is to changes in luminance. RGB is therefore inefficient as chrominance and luminance are stored at the same resolution. By

exploiting properties of the HVS it is possible to represent colour more efficiently. YCrCb separates the luminance from the chrominance information, thus enabling resampling of the chrominance components at a lower sampling rate. The result may be visually similar to that of RGB but with fewer bits required for storage or transmission. Despite separating the luminance (Y) information it is still possible to represent a pixel with only three values. The summation of Cr, Cb and Cg is a constant value and so one can be calculated with knowledge of the other two.

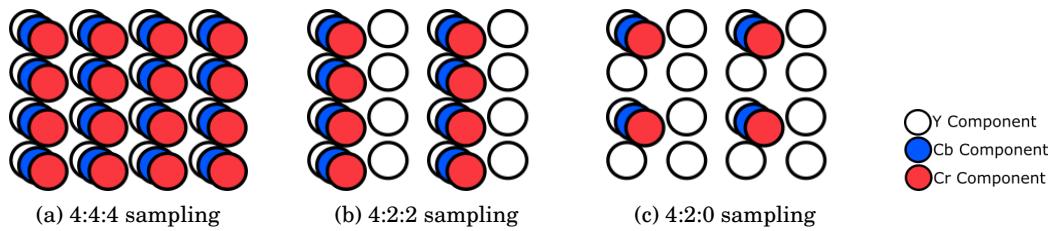


Figure 2.4: YCrCb colour space sampling. The YCrCb colour space enables downsampling of the chrominance information while preserving the luminance information

Figure (2.4) shows how three common sampling patterns are constructed. 4:4:4 sampling contains the same number of chrominance samples as luminance samples, thus the full resolution of colour is maintained. The three numbers represent the number of samples for each channel component in the horizontal direction – Y, Cr and Cb. Therefore, for every 4 Y samples there are 4 Cr samples and 4 Cb samples. 4:2:2 sampling reduces the number of chrominance samples in the horizontal direction by half. Therefore, for every 4 Y samples there are 2 Cr samples and 2 Cb samples. For most viewers there is no difference in visual quality between the 4:4:4 and 4:2:2 sampling patterns and so 4:2:2 is highly favourable due to the reduced number of required bits. The 4:2:0 pattern does not follow the same structure as the previously mentioned sampling patterns; 4:2:0 sampling reduces the number of chrominance samples in the horizontal and vertical direction by half. It is highly popular for applications and devices with low requirements and specifications as it requires half the number of bits than the full resolution 4:4:4 sampling pattern.

Downsampling of chrominance samples is therefore a form of image and video compression. It exploits the limitations of the HVS by removing redundant information not visible to the human eye. As a result visual quality can be maintained while the number of bits required to represent each frame is reduced. This form of compression is therefore *lossy*; when reconstructing the image/video it is not possible to reproduce the full resolution of colour.

2.2.3 Conversion

The video file format encoded using $YCrCb$ is commonly known as YUV , where Y is the luminance channel and U and V are the chrominance channels. To convert between RGB and YUV the following formulas can be used:

$$(2.1) \quad Y = (0.257R) + (0.504G) + (0.098B) + 16,$$

$$U = -(0.148R) - (0.291G) + (0.439B) + 128,$$

$$V = (0.439R) - (0.368G) - (0.071B) + 128.$$

$$(2.2) \quad R = 1.164(Y - 16) + 1.596(V - 128),$$

$$G = 1.164(Y - 16) - 0.813(V - 128) - 0.391(U - 128),$$

$$B = 1.164(Y - 16) + 2.018.(U - 128).$$

2.3 Video Compression

As mentioned in Chapter 1 it is necessary to compress a video before storage or transmission in order to efficiently utilise channel capacity and maximise bandwidth efficiency. Section 2.2 described an elegant method of downsampling chrominance samples, while maintaining visual quality, in order to reduce the number of bits required to represent an image or video. In this section standard video compression techniques, which form the basis of all video standards, will be described and analysed.

2.3.1 Transforms

Section 2.3.5 describes the prediction process, at the front end of the encoder. After prediction and motion compensation, transform coding is used to convert the residual data or source data from the spatial domain into the transform domain. The transform operation alone is a lossless process but enables the effective use of quantisation, which is lossy, as discussed in Section 2.3.2.

A popular transform used in earlier standards, such as JPEG, MPEG-2 Video, and MPEG-4 Visual, is the Discrete Cosine Transform (DCT). The advantage of using the DCT for video compression is that it decorrelates the spatial data allowing effective use of quantisation and also it is easily implemented into hardware and software. The DCT can be applied by:

$$(2.3) \quad \mathbf{C} = \mathbf{A}\mathbf{X}\mathbf{A}^T,$$

where \mathbf{C} is the DCT coefficients matrix, \mathbf{A} is the DCT transform matrix and \mathbf{X} is the $N \times N$ input block. The matrix \mathbf{A} for the k^{th} row and the n^{th} column, can be given as:

$$(2.4) \quad \mathbf{A}_{k,n} = \sqrt{\frac{2}{N}} \epsilon_k \cos\left(\frac{\pi(2n+1)k}{2N}\right),$$

where

$$(2.5) \quad \epsilon_k = \begin{cases} \sqrt{\frac{1}{2}} & \text{for } k = 0, \\ 1 & \text{otherwise.} \end{cases}$$

To demonstrate the operation of the DCT an 8×8 block of pixels is given as an example:

$$(2.6) \quad \mathbf{X} = \begin{bmatrix} 157 & 176 & 177 & 174 & 190 & 186 & 187 & 183 \\ 166 & 177 & 180 & 183 & 187 & 188 & 181 & 167 \\ 177 & 181 & 188 & 189 & 194 & 185 & 160 & 95 \\ 181 & 184 & 189 & 193 & 185 & 143 & 81 & 74 \\ 188 & 192 & 192 & 179 & 133 & 75 & 77 & 76 \\ 193 & 195 & 175 & 108 & 67 & 61 & 74 & 77 \\ 200 & 169 & 93 & 53 & 69 & 57 & 70 & 72 \\ 163 & 74 & 52 & 53 & 59 & 65 & 61 & 65 \end{bmatrix},$$

$$(2.7) \quad \mathbf{A} = \begin{bmatrix} 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 \\ 0.4904 & 0.4157 & 0.2778 & 0.0975 & -0.0975 & -0.2778 & -0.4157 & -0.4904 \\ 0.4619 & 0.1913 & -0.1913 & -0.4619 & -0.4619 & -0.1913 & 0.1913 & 0.4619 \\ 0.4157 & -0.0975 & -0.4904 & -0.2778 & 0.2778 & 0.4904 & 0.0975 & -0.4157 \\ 0.3536 & -0.3536 & -0.3536 & 0.3536 & 0.3536 & -0.3536 & -0.3536 & 0.3536 \\ 0.2778 & -0.4904 & 0.0975 & 0.4157 & -0.4157 & -0.0975 & 0.4904 & -0.2778 \\ 0.1913 & -0.4619 & 0.4619 & -0.1913 & -0.1913 & 0.4619 & -0.4619 & 0.1913 \\ 0.0975 & -0.2778 & 0.4157 & -0.4904 & 0.4904 & -0.4157 & 0.2778 & -0.0975 \end{bmatrix},$$

$$(2.8) \quad \mathbf{C} = \begin{bmatrix} 1111.9 & 203.7 & 5.2 & 20.5 & 6 & -1.8 & -4.2 & 10.1 \\ 285.5 & -112.3 & -111.7 & -12.5 & -23.2 & -5.7 & -5.8 & -3.4 \\ -59.5 & -118.1 & 77.7 & 39.7 & 5.2 & 1.7 & 3.4 & 2.1 \\ 16.7 & 36.5 & 48 & -60.8 & -15 & -12.2 & -8.9 & 3.1 \\ -20.9 & -8.2 & -37.3 & -5.3 & 34.9 & 1.8 & 8.8 & -2.9 \\ 2.1 & 15.9 & 2.3 & 23.6 & -6.5 & -21.5 & -3.6 & 1.1 \\ -4.2 & -3.2 & -8.1 & -2.1 & -17.2 & 16.2 & 7.5 & 2.7 \\ 6.6 & 1.9 & 1.3 & 6.6 & -2 & 6.1 & -17.1 & 6.7 \end{bmatrix}.$$

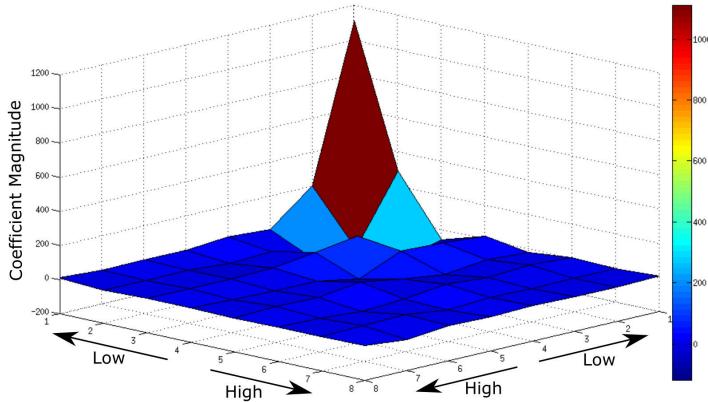


Figure 2.5: DCT coefficients of input block corresponding to the matrix **C** 2.8

From **C** in 2.8 and Figure (2.5) it can be seen that a larger amount of energy is conjugated in the upper left corner of the DCT block. The energy of the coefficients generally reduces from left to right and top to bottom. The upper left value is known as the DC coefficient and is the magnitude of the lowest frequency component. Coefficients representing lower frequencies tend to have more energy while, conversely, coefficients representing higher frequencies have less energy. In this form removal of these higher frequencies can be efficiently achieved.

2.3.2 Quantisation

After applying the DCT the spatial frequencies become decorrelated, thus enabling quantisation of the less significant coefficients. The quantisation process is lossy; data removed can not be re-obtained at the decoder. The DCT matrix **C** is divided by a quantisation matrix and rounded to the nearest integer. From Figure (2.5) it can be seen that many of the high frequency coefficients are close to zero. After quantisation these will likely be set to zero, thus reducing the number of significant bits required to represent this block. An example of an 8×8 quantisation block, as used in the JPEG standard, is given by matrix **Q** 2.9.

$$(2.9) \quad \mathbf{Q} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}.$$

Matrix \mathbf{Q} can be applied to the DCT block \mathbf{C} by:

$$(2.10) \quad \mathbf{CQ}_{i,j} = \text{round}\left(\frac{\mathbf{C}_{i,j}}{\mathbf{Q}_{i,j}}\right) \text{ for } i,j = 0,1,\dots,7.$$

The resulting quantised DCT matrix is given as:

$$(2.11) \quad \mathbf{CQ} = \begin{bmatrix} 69 & 19 & 1 & 1 & 0 & 0 & 0 & 0 \\ 24 & -9 & -8 & -1 & -1 & 0 & 0 & 0 \\ -4 & -7 & 4 & 1 & 0 & 0 & 0 & 0 \\ 1 & 3 & 3 & -3 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

When coding a video, the encoder assigns a QP for each frame which determines how much high frequency information is removed by applying the corresponding quantisation matrix. An example of this was illustrated in Figure (1.1) and it can be seen that the selected QP can have a huge impact of the level of quality and the bitrate produced. Although the transform and quantisation processes are not a major focus of research within this thesis, the proposed solution aims to reduce the required QP for a given bit-budget and therefore reduces the amount of high frequency information lost due to quantisation. As highlighted in Chapter 1, this is achieved by downsampling intra-coded pictures prior to coding. The downsampling process will also remove high frequency information, as described in Section 2.1.1, and therefore each process will introduce distortion which needs to be balanced in order to provide the best overall performance. Further details regarding how this is achieved are given in Chapter 3.

2.3.3 Inverse Quantisation and Transformation

Reconstructing the image block follows the reverse process from Sections 2.3.2 and 2.3.1. First, the quantisation block \mathbf{Q} 2.9 is applied to the quantised matrix \mathbf{CQ} to form the reconstructed DCT matrix \mathbf{C} , by:

$$(2.12) \quad \mathbf{C}_{i,j} = \text{round}(\mathbf{CQ}_{i,j} \times \mathbf{Q}_{i,j}) \text{ for } i,j = 0, 1, \dots, 7.$$

The result is given as:

$$(2.13) \quad \mathbf{C} = \begin{bmatrix} 1104 & 209 & 10 & 16 & 0 & 0 & 0 & 0 \\ 288 & -108 & -112 & -19 & -26 & 0 & 0 & 0 \\ -56 & -119 & 88 & 29 & 0 & 0 & 0 & 0 \\ 14 & 39 & 48 & -72 & 0 & 0 & 0 & 0 \\ -18 & 0 & -37 & 0 & 68 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The Inverse DCT (IDCT) can be obtained by:

$$(2.14) \quad \mathbf{X} = \mathbf{A}^T \mathbf{CA},$$

where \mathbf{X} is the reconstructed input block, \mathbf{A} is the DCT transform matrix and \mathbf{C} is the DCT coefficient matrix after inverse quantisation.

2.3.4 Run-Length Coding & Entropy Coding

Table 2.2: Run-Length Coding applied to matrix \mathbf{CQ} 2.11. In (R,L), R is the number of zeros within the array prior to the significant value L

(0,69),(0,19),(0,24),(0,24),(0,-4),(0,-9),(0,1),(0,1),(0,-8),(0,-7),(0,1),(0,-1),(0,3),(0,4),(0,-1),(2,-1),
(0,1),(0,3),(4,-1),(0,-3),(14,1)

As can be seen from matrix \mathbf{CQ} 2.11, after quantisation only a few significant coefficients remain, clustered around the upper left corner, leaving the rest equal to zero. This produces a large amount of redundancy within the data. This data can be reduced further using *lossless* coding methods. Reordering the coefficients by performing a zigzag scan, starting with the DC component, will result in a 1D array with most of the significant coefficients at the beginning and

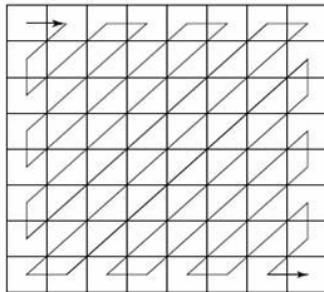


Figure 2.6: Zig-zag recording of coefficients to produce a vector of coefficients with the most significant values first

a large number of zeros towards the end. This process is illustrated in Figure (2.6). The reordered 1D array will contain a large number of zeros that are stored consecutively, one after another. Run-length coding can then be performed to reduce the number of bits required to represent this array. The *run* applies to the number of consecutive zeros within the array before a significant value is given. The *length* applies to this significant value.

Table (2.2) gives the result of *run-length* coding, after zigzag recording, for the data in matrix **CQ 2.11**. The remaining coefficients are zero. Entropy coding can further compress the data by utilising the statistical redundancies within the output of the *run-length* encoder. *Huffman* coding was once a popular form of Variable Length Coding (VLC). Huffman coding works by applying a short codeword to frequently occurring *run-length* pairs and long codewords to less frequently occurring *run-length* pairs.

2.3.5 Prediction

Predictive coding relies on spatial and temporal redundancies within a video sequence. The correlation between neighbouring pixels and the correlation between pixels between frames is usually high within a video of a captured natural scene.

2.3.5.1 Inter Prediction

Figure (2.7) shows an example of how motion compensation can exploit temporal correlation to increase redundancy between frames. The residual frame is the difference between the current frame and the next, or the current frame and a previously coded reference frame. It is therefore representative of the motion occurred during the two instances of time at which the frames were sampled and so it can be said that increasing the frame rate could potentially decrease the amount of residual information between two frames. It would be highly inefficient to transmit a video by encoding each frame independently. As shown in Figure (2.7b), there can be high tempo-



Figure 2.7: Inter prediction by exploiting temporal redundancies. The motion compensated residual image is shifted to account for camera pan

ral redundancy within a video scene and this redundancy can be exploited by only transmitting residual data. The decoder is then able to reconstruct the current frame by applying the residual to a previously coded reference frame stored in a buffer. Typically a residual frame will contain less information than any standard frame, thus this process increases bandwidth efficiency. If a video sequence contains *scene changes* then this may not be true; residual data from one scene to the next will likely contain a large amount of information. During a scene change it would be beneficial to encode and transmit the new frame independently using only intra prediction, this will be discussed in Section 2.3.5.2.

Motion estimation and compensation can be used to reduce the amount of information within each residual frame but requires transmission of additional motion vectors. Inter prediction must be applied to frames contained at both the encoder and decoder to ensure an identical prediction at the decoder. Therefore, the encoder can only make predictions on frames that have already been coded and transmitted. Prediction is applied on a block by block basis. The size of the block directly affects the performance of motion compensation, the amount of computations required by the encoder and the number of bits needed for transmission. Smaller block sizes provide better predictions in detailed areas with complex motion but at the cost of more bits to code the

additional motion vectors. As it is unlikely that features within the scene will propagate over a large area, a search window can be defined to reduce the number of computations that would usually be required to search the entire frame.

Various metrics have been used to compare coding blocks within the reference frame and the current frame in order to calculate motion vectors. These include: Mean Squared Error (MSE), Sum of Squared Error (SSE) [34], and *Sum of Absolute Difference* (SAD). SAD has commonly been used for motion estimation as it contains the least amount computations but still provides good performance.

$$(2.15) \quad MSE = \frac{1}{N \times M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (X_{ij} - Y_{ij})^2.$$

$$(2.16) \quad SSE = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (X_{ij} - Y_{ij})^2.$$

$$(2.17) \quad SAD = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} |(X_{ij} - Y_{ij})|.$$

where N and M are the dimensions of the blocks X and Y .

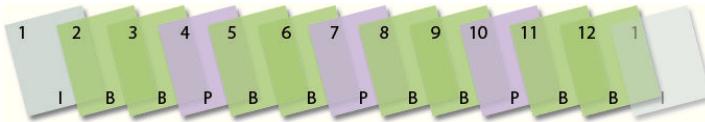


Figure 2.8: Basic GOP structure commonly used in inter-coding configurations

An error in a reference picture may result in poor prediction of the current picture. A succession of inter-predicted pictures can therefore lead to the propagation of errors. As a result, pictures are usually encoded within a group, known as a Group of Pictures (GOP). Figure (2.8) shows an example of a basic GOP structure commonly used in earlier standards such as MPEG-1 and MPEG-2 [55]. The first coded picture within this GOP is an intra-coded picture, also known as an *I-picture*, which does not apply temporal prediction. Once encoded and transmitted, this picture can be used as a reference. *P-pictures* (Predicted) are inter-predicted from previous *I* or *P* pictures, thus *P-pictures* can also be used for reference. Each *B-picture* (Bidirectionally Predicted) is predicted from *I* and/or *P* pictures either preceding or succeeding its position in the GOP. It is therefore required that the decoder receives and decodes all the necessary *P-pictures* prior to receiving a *B-picture*. *B-pictures* are generally more efficient than *P-pictures* as they enable prediction from reference pictures sampled at instances of time closer to the picture to be encoded. However, this comes at the cost of increased delay and larger buffer requirements. This GOP example provides a good compromise between coding efficiency and complexity.

2.3.5.2 Intra Prediction

Intra coding exploits the spatial redundancies within a picture. The current block can be predicted from previously encoded and reconstructed blocks. A residual is initially calculated and then a prediction mode is selected based on the one that minimises the difference between the reference block and the current block to be encoded. Given raster-scan ordering, prediction can be performed from any block positioned to the left and/or above the current block. Intra-coding will be described in more detail in Section 2.4.2.

2.4 H.265 - High Efficiency Video Coding

In July 2010, many new proposals were given to the JCT-VC for a new video coding standard. In October, later that year, the new standard H.265/HEVC was accepted. HEVC is the latest video coding standard, although it is yet to surpass AVC in terms of deployment. The need for the new standard became apparent from AVC's low coding efficiency at high resolutions [19]. This inefficiency is partly due to the limitation of the maximum 16×16 luma block size. The *High Efficiency Video Coding (HEVC) text specification draft 8* was released 20th July 2012 [65] and the first edition of the standard was finalised in January 2013.

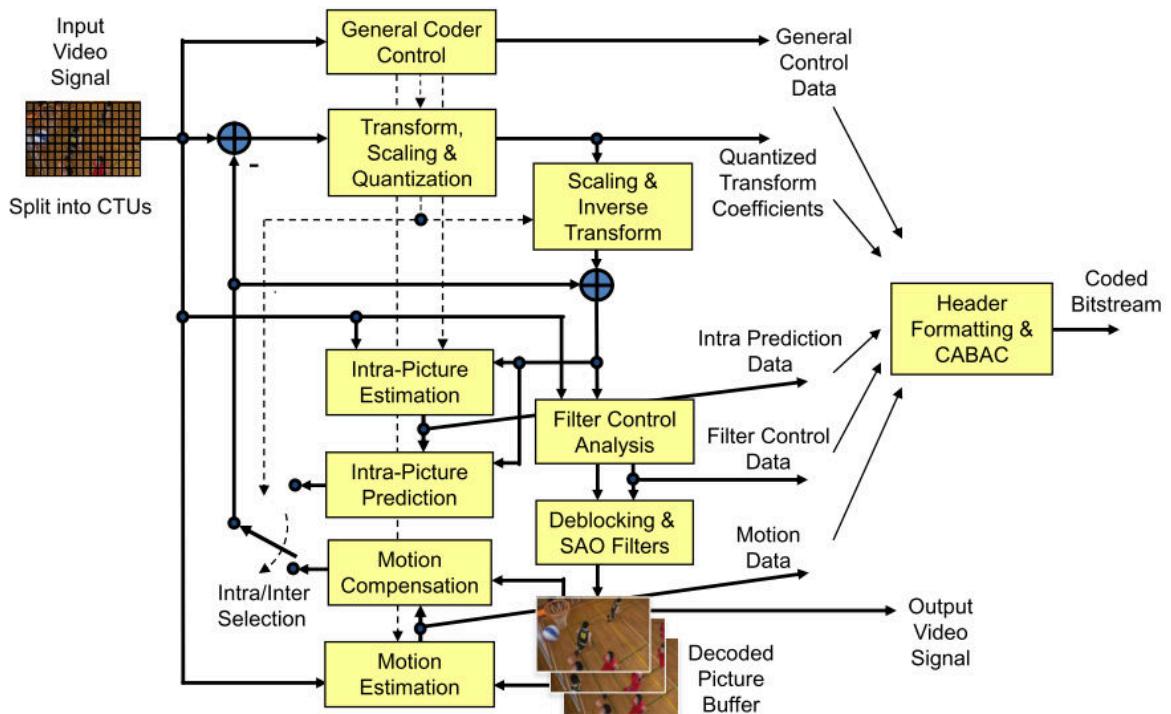


Figure 2.9: Block diagram of the H.265/HEVC standard [63]

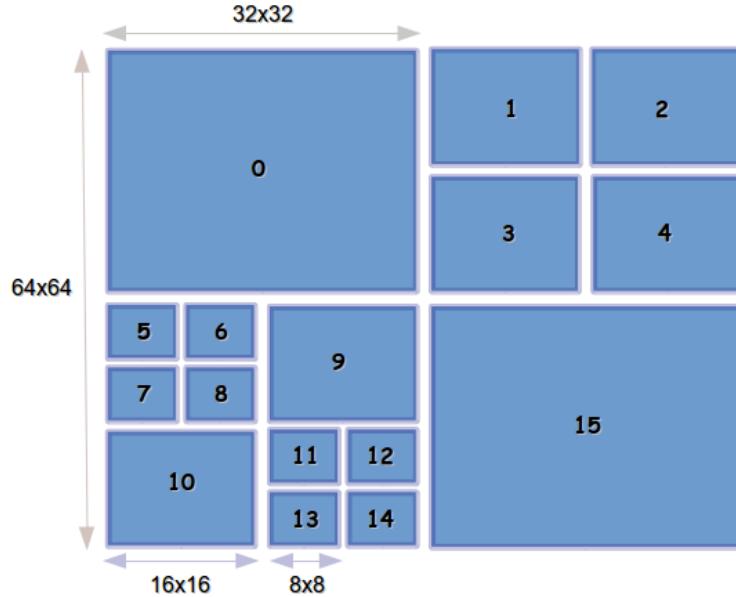


Figure 2.10: Example of CTU partitioning into smaller CUs

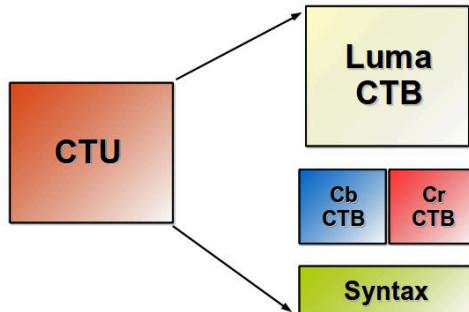


Figure 2.11: Illustration of the type of data contained within a Coding Tree Unit (CTU)

2.4.1 Coding Units

Many features in HEVC have been inherited from AVC, including the Network Abstraction Layer (NAL) units and parameter sets. The HEVC standard also introduces a new coding structure. The macroblock, as used in previous standards, has been replaced by the Coding Tree Unit (CTU). The size of the CTUs can vary for each sequence and is determined by the encoder in regard to video content. The largest size the CTUs can take is 64×64 pixels and is known as a Largest Coding Unit (LCU). Alternatively, the CTUs can be 32×32 or 16×16 pixels. In any case the CTU size is signalled in a Sequence Parameter Set (SPS) – this means that the dimensions of each CTU is fixed for the entire video sequence. Figure (2.9) shows a HEVC block diagram and also illustrates how the input picture can be split into a number of CTUs.

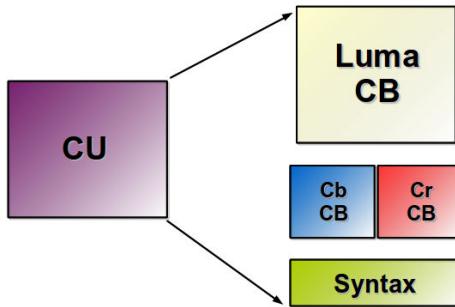


Figure 2.12: Illustration of the data contained within a Coding Unit (CU)

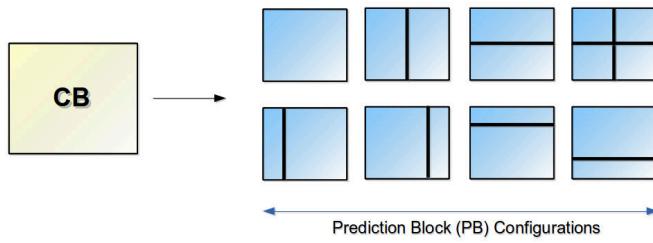


Figure 2.13: Prediction Block (PB) Configurations

A CTU is often too large to be able to accurately encode fine details within the frame. Each CTU can therefore be independently split into multiple partitions of equal size, each known as a Coding Unit (CU). An initial partitioning will produce four CUs of size 32×32 , given that the CTU is of size 64×64 . Each CU can then be split a further two more times until the minimum CU size of 8×8 has been reached. As suggested, a CU may be of any size from 64×64 , 32×32 , 16×16 or 8×8 . Figure (2.10) shows an example of how a CTU can be segmented into multiple CUs. CUs are therefore also highly beneficial for efficiently segmenting the frame into areas of different prediction types – inter and intra.

From Figure (2.11) it can be seen that a CTU contains three types of Coding Tree Blocks (CTB): luma, chroma blue (Cb) and chroma red (Cr). It also contains syntax to specify coding parameters. Similarly, as shown in Figure (2.12), a CU contains luma and chroma Coding Blocks (CB). It also contains syntax containing parameters such as the prediction type, which is decided at the CB level. In HEVC a ‘xxxUnit’ is a logical unit that is encoded into the bitstream, whereas a ‘xxxBlock’ is a section of the video frame buffer. For fine detail, an 8×8 CB may still be too large to accurately encode the video content. Each CB can be split into multiple Prediction Blocks (PB), where each PB is created based on spatial-temporal prediction requirements. Figure (2.13) illustrates the different ways in which a PB can be partitioned. As with previous standards, after prediction the residual is encoded using a transform function. If the CB contains both high and low frequency content it would be beneficial to split it into multiple Transform Blocks (TB) to improve coding efficiency. Figure (2.14) shows the different configurations that a TB can take.

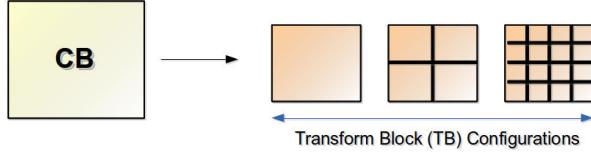


Figure 2.14: Transform Block (TB) Configurations

2.4.2 Prediction

2.4.2.1 Inter-Prediction

HEVC supports Advanced Motion Vector Prediction (AMVP) which utilises PUs to make more accurate predictions compared to previous standards. Similar to AVC, quarter pixel motion vector accuracy can be performed but with a 8-tap filter, rather than a 6-tap filter, for interpolation of fractional sample positions. Additionally, HEVC also uses multiple reference frames for which one or two motion vectors can be transmitted for each PB, thus allowing prediction from either past frames or from past and future frames. Weighted prediction can also be applied to the prediction signals as in AVC.

2.4.2.2 Intra-Prediction

Prediction from spatial samples within the PB partitions is performed from previously decoded adjacent blocks. As an enhancement to AVC, HEVC supports 35 modes (including Planar and DC) as opposed to just 9 (including DC). Figure (2.15) shows which angles of prediction are available for the directional modes. The increased number of directional prediction modes enables higher accuracy of predictions for objects with angular structures. Planar and DC prediction are not directional modes and are used as a predictor for smooth image content. Angular prediction can produce visual contouring in areas containing low frequency content. The DC prediction mode, as applied in AVC, averages the the values of referenced samples immediately above and to the left. DC can still produce block-like artefacts given highly compressed reference samples. The planar mode was introduced to counteract this problem by averaging reference samples only in the horizontal and vertical direction. This reduces the effect of artefacts on the block boundaries.

These prediction modes can be applied to blocks of all sizes. Reference samples are filtered as a pre-processing step prior to prediction according to the selected prediction mode, direction and block size. The number of reference samples available is restricted by a number of factors. Firstly, only samples coded in the same picture can be considered, even if the picture is a P or B picture – this is also the case for tiles and slices, samples outside of the tile or slice can not be referenced. To reduce the probability of error propagation, intra-coded blocks are also not permitted to reference a PU coded in inter mode.

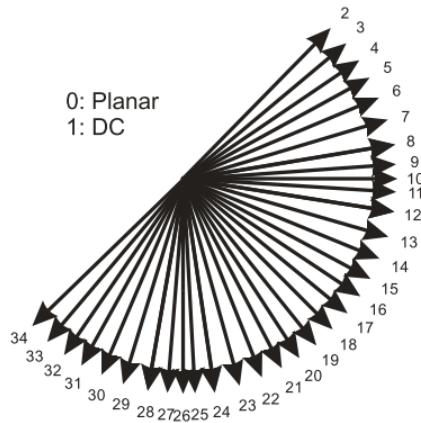


Figure 2.15: Directional modes for intra prediction in H.265/HEVC [63]

2.4.3 Performance Comparisons

Many of the features that make HEVC superior to any prior video coding standard are only slight improvements upon the AVC standard. However, the combination of these enhancements result in a significant increase in the overall compression and coding efficiency. The standard also includes a few completely new additions which enhance parallel processing capability and improve the structure of frame slicing. For a more in-depth review of the standard it is suggested that the reader refers to [5, 39, 49, 50, 62, 63, 65] for more details.

Table 2.3: Comparison of video coding standards. Average bitrate savings for video streaming applications [49]

Coder	Bit Rate Savings Relative To:			
	H.264/AVC HP	H.263 CHC	MPEG-4 ASP	MPEG-2/H.262 MP
HEVC MP	40.3%	67.9%	72.3%	80.1%
H.264/AVC MP	-	46.8%	54.1%	67.0%
H.263 CHC	-	-	13.2%	37.4%
MPEG-4 ASP	-	-	-	27.8%

AVC provides approximately a 50% increase in compression performance relative to previous standards. From [49] it is stated that HEVC encoders can achieve an equivalent level of subjective quality as AVC when using approximately 50% less bitrate on average. Tables (2.3) and (2.4) provide comparative results of various standards and profiles for different applications [49], namely High Profile (HP), Main Profile (MP), Advanced Simple Profile (ASP), High Latency Profile (HLP), and Conversational High Compression (CHC). Results also demonstrate that HEVC outperforms AVC especially at high definition resolutions [69].

Table 2.4: Comparison of video coding standards. Average bitrate savings for video entertainment applications [49]

Coder	Bit Rate Savings Relative To:			
	H.264/AVC HP	MPEG-4 ASP	H.264 HLP	MPEG-2/H.262 MP
HEVC MP	35.4%	63.7%	65.1%	70.8%
H.264/AVC HP	-	44.5%	46.6%	55.4%
MPEG-4 ASP	-	-	3.9%	19.7%
H.263 HLP	-	-	-	16.2%

2.5 Scalable Video Extensions

Scalable Video Coding (SVC) [57, 58] and Scalable HEVC (SHVC) [6, 21, 60] are extensions of H.264/AVC and H.265/HEVC, respectively. Scalable coding enables the decoder to extract and decode sub-streams from the full bitstream thus providing increased adaptability and versatility for receiving devices. In SHVC, standard HEVC coding is applied at the Base Layer (BL) with Enhancement Layers (EL) providing increased spatial and/or quality gains. Temporal scalability is no longer incorporated into the scalable extensions as it is already part of the HEVC standard. Like SVC, SHVC is developed by the JCT-VC.

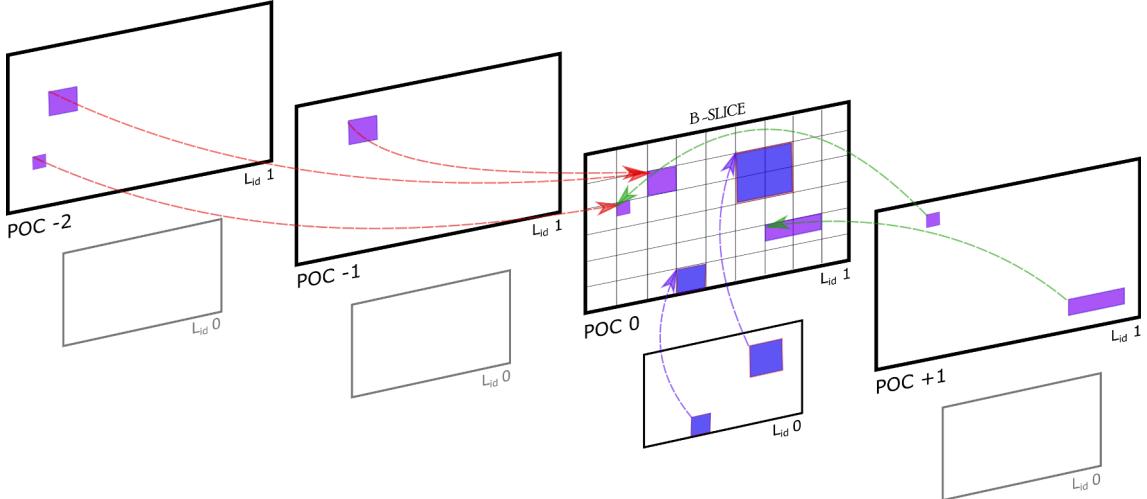


Figure 2.16: Inter-layer prediction in SHVC with two spatial layers. The current frame is able to take reference from coded pictures from both future and past pictures in display order. It also can take reference from the upsampled picture from the lower spatial layer with the same time index

2.5.1 Spatial Scalability

Spatial scalability provides multiple layers coded at different resolutions. Depending on the receiving device's specifications, decoding at HD or UHD may be inefficient provided that a naive downscaling operation would need to be performed in real-time as a post-processing stage in order to display on the device's low resolution screen. SHVC performs spatial scalability by coding a BL at a low resolution using standard HEVC and then applying one or more ELs to enhance the resolution iteratively. This enhancement process is performed using inter-layer prediction, which essentially supplies the upper spatial layer with an additional reference picture after the corresponding BL picture is upsampled. This is illustrated in Figure (2.16) – $L_{id}0$ is the BL and $L_{id}1$ is an EL, which in this example is coded at the full resolution. The current frame is a B-picture taking reference from both future and past frames. It is also referencing a BL picture with the same time index but coded at a lower resolution. Inter-layer prediction applies inter-coding to predict the current frame from the upsampled picture at the lower spatial layer. Given that the upsampled BL picture represents the same instance of time, it should provide a good predictor unless significant information is lost during the downsampling operation performed prior to encoding.

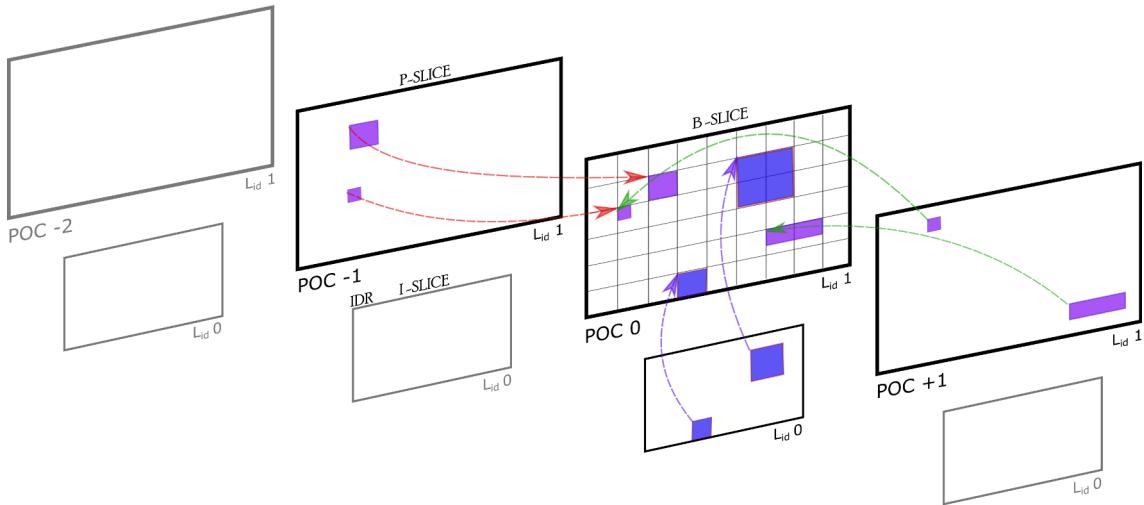


Figure 2.17: Inter-layer prediction in SHVC with an IDR picture at the BL. The current frame is able to take reference from coded pictures from both future and past pictures in display order but limited due to the IDR picture. It also can take reference from the upsampled picture from the lower spatial layer with the same time index

The work presented in this thesis utilises a coding structure with frequently coded IDR pictures. As IDR pictures prevent referencing of previously coded pictures, in SHVC this coding structure can be represented by Figure (2.17). Note that any coded IDR pictures are provided at the BL; frames coded at higher spatial layers but with the same time index are coded in inter

mode but are only able to reference the corresponding lower resolution pictures – this means that they are coded as P-pictures. Despite that the IDR picture is only contained within the BL, at the EL future frames are still prevented from using reference pictures coded prior to the IDR picture.

2.6 Test Sequences

To evaluate the performance of the proposed techniques, within this thesis a series of widely used and easily acquirable test sequences are used. A full detailed list of the selected sequences are given in Appendix-A and can be found from the following links:

- <ftp://ftp.tnt.uni-hannover.de/>
- <https://media.xiph.org/video/derf/>
- <https://data.bris.ac.uk/data/dataset/1if54ya4xpph81fb01gkpk5kk4> [52]
- <http://ultravideo.cs.tut.fi/#testsequences>

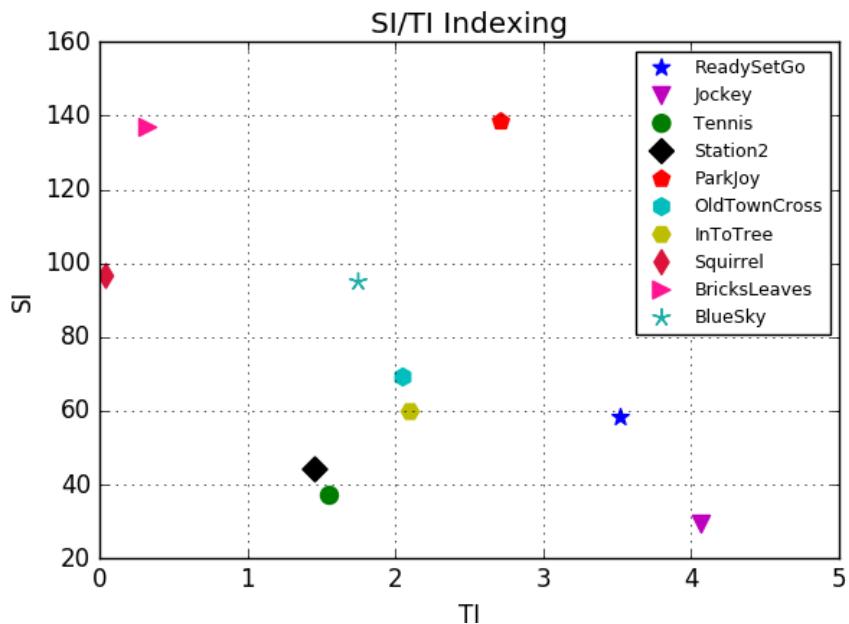


Figure 2.18: Spatial and temporal information indexing for test sequences

To categorise these sequences, the Spatial Information (SI) and Temporal Information (TI) indexes are calculated, as used in [52, 73], to give indication of the spatial and temporal information within each sequence. The SI index uses calculated gradient images to produce an average result over all frames:

$$(2.18) \quad SI = \sqrt{\frac{L}{1080} \sum_k^N \frac{I_{G_k}}{P}},$$

where P is the total number of pixels, N is the total number of frames, L is the number of pixels in the vertical direction and I_G is the gradient image of frame k given by:

$$(2.19) \quad I_G = \sqrt{(\delta_x I)^2 + (\delta_y I)^2},$$

$$(2.20) \quad \delta_x I = \delta_x * I,$$

$$(2.21) \quad \delta_y I = \delta_y * I,$$

where I is the luminance channel of the picture and the two gradient operators are given by:

$$\delta_x = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix} \quad \delta_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{pmatrix}$$

TI takes into account the magnitude of motion vectors relative to the frame rate, it is given as:

$$(2.22) \quad TI = \frac{f}{L} \sqrt{\sum_n^M \frac{|\mathbf{MV}_n|^2}{M}},$$

where M is the total number of motion vectors, $|\mathbf{MV}|$ is the magnitude of motion vector n and f is the frame rate. Motion vectors are calculated using 8×8 blocks within a search window of 16×16 and determined based on minimising the MSE.

Figure (2.18) shows how each test sequence can be categorised according to their SI and TI index. It is expected that sequences with a low SI index will be coded efficiently in intra mode at the original HR. However, these pictures may be oversampled and therefore could be coded more efficiently at a lower spatial resolution. A high TI index may reduce the over efficiency of the proposed technique as temporal correlation between each IDR picture and the succeeding inter-coded pictures will likely be low. Throughout this thesis, these results are referenced to provide insight into why performance can vary for each sequence.

2.7 Prior Art

As highlighted in Chapter 1, previous work has shown that coding at lower resolutions given a tight bit-budget can provide greater efficiency gains compared to simply coding at the original HR. In this section, further details are given regarding the methodology and performance of prior art within this area.

Some of the earliest research into this area was proposed in [7] by Bruckstein et al. This work explores the potential performance gain of image compression using JPEG on a downsampled image prior to upsampling back to the original resolution, as opposed to coding without scaling. JPEG utilises many of the tools discussed earlier in this chapter, such as the DCT applied to blocks of 8×8 pixels, quantisation of the high frequency coefficients, zigzag reordering and Entropy coding. The problem of coding given a tight bit-budget is described and examples are provided which show that high levels of quantisation produces images with highly noticeable artefacts around block boundaries. This is due to the fact that, as the level of quantisation increases, more and more frequency coefficients are removed until only the DC coefficient remains. By reducing the spatial resolution of an image before coding, fewer bits are required to represent a given area; for a scale factor of 0.5 each 8×8 DCT block will in fact represent 16×16 pixels of the original image. It is also mentioned that applying interpolation after coding will have a positive effect on the coding artefacts produced around block boundaries as the low pass filter will create a smoother transition between the pixels. The proposed method aims to predict the optimal downsampling factor given the bit-budget. Resampling is achieved using '*Matlab's standard image resizing function*'. Further details on resampling are not given but in recent versions of Matlab the default method of image resizing uses Bicubic interpolation. A model for compression error is devised using information of image statistics and it is this model which determines the optimal scale factor. Resampling distortion is somewhat ignored and it is also stated that the method for predicting the parameters based on image statistics is inefficient. Detailed results comparing the proposed method with standard JPEG are not provided but the evaluation of the proposed model suggests it should perform well. The proposed method is very much designed for JPEG compression as it assumes fixed block sizes throughout the entire image. More recent coding standards, such as AVC and HEVC, use variable block sizes according to image content. Because of these issues, such a technique can not be accurately applied to newer coding standards effectively. However, this work illustrates how coding distortion can be reduced by reducing the spatial resolution of the image and has sparked further research into this area, including coding entire video sequences at lower spatial resolutions for greater coding efficiency at low bitrates.

In [42], Lin et al. follow the work of Bruckstein but propose an alternative approach to downsampling prior to coding with JPEG. In this work, an adaptive downsampling method is described where the decision regarding the best scale factor and quantisation parameter are

made for each 8×8 macroblock. Each block is coded multiple times; after coding at the original resolution the block is downsampled by 0.5 in the horizontal direction, by 0.5 in the vertical direction, and by 0.5 in both directions. For each scaled version, up to four QP values are tested which are derived from the original with the aim of conforming to the number of bits produced at the original resolution. Unlike the method proposed in [7], a compression model is not used but an exhaustive coding approach is applied to find the best resolution and QP combination. The decision is made by computing the MSE of the decoded block compared to the original. Results show significant improvement over standard JPEG and JPEG2000, and also Bruckstein's method given in [7]. The weaknesses of this method lie within its reliance on coding each block numerous times. A combination of making decisions based on information from a predefined model and exhaustive coding might provide a better balance between performance and computational complexity. However, recently hardware technology would enable this method to be applied in parallel, thus providing a more efficient implementation.

In a later paper, Lin co-authored a paper with Nguyen et al. [48] which proposes further enhancement to the previous work and also expands the application to video coding using MPEG-2. One of the major improvements is the use of a Rate-Quantisation (R-Q) model to predict the required number of bits for any given QP. This significantly reduces the computational load compared to their previous work by only requiring three additional coding processes. As before, these processes could be run in parallel but now with a reduced number of processing cores. In this paper, results are compared to standard MPEG-2 and show significant improvements at low bit-rates.

Wu et al. also explore the potential gains of downsampling prior to coding images using JPEG and JPEG2000 [74]. However, they very much focus on the resampling process to achieve greater level of coding efficiency. In the previously described work [7, 42, 48], the methods used for downsampling and upsampling are not considered to be very important. Wu et al. devise a *collaborative adaptive downsampling and upconversion* method which filters the image prior to downsampling according to frequency content such that more edge information is maintained. After decoding, the interpolation process takes into account that filtering was applied in a spatially variant manner in order to reconstruct the edges at the original HR. Results are compared to resampling using Bicubic (more details on Bicubic are given in Chapter 5), in combination with JPEG and JPEG2000, which show better objective and subjective levels of quality. However, the proposed method does not provide a complete and practical solution as there is no explanation of how the best resolution and QP are chosen.

In [59], Shen et al. expand upon the idea of downsampling video data to enable greater levels of compression efficiency at low bitrates. This more recent work modifies and compares against H.264/AVC. Additionally, a Multi-Image Super Resolution (MISR) technique is proposed to enhance

the reconstruction of the HR frames (see Chapter 5 for more discussion on Super Resolution techniques). Interestingly the authors choose to code intra pictures at the original HR and only code inter pictures at a lower spatial resolution. After decoding each intra-coded picture, the picture is downsampled to provide a reference for the following inter-coded frames. This means that all motion estimation and compensation is applied at the LR. This is the opposite to the proposed solution given in this thesis, which was briefly highlighted in Chapter 1 and described in more detail later in Chapter 3. Results show improvement at low bitrates compared to standard AVC. However, theoretically, enabling coding of intra pictures at lower spatial resolutions will provide greater performance gains per frame as there is more information to be coded and therefore greater levels of quantisation would be applied when coding at the original HR given a tight bit-budget. In most cases, inter-coded pictures contain relatively less information within the residual after motion compensation. These pictures are therefore coded more efficiently at the original resolution. It is also expected that motion estimation and compensation would be better applied at the HR as there is more information available to enable more accurate prediction.

Dong et al. [12, 13] propose some of the more recent work related to the solution given in this thesis. The work also demonstrates a more comprehensive and practical solution compared to previous methods, but it is not necessarily more efficient. Unlike the solutions given in [48, 59], which also describe methods of resampling video data, the work provided in [12, 13] treats the AVC video codec as a black box by downsampling and upsampling before and after encoding and decoding, respectively, and maintaining a fixed spatial resolution in order to comply with the standard and enable the use of existing rate control algorithms. A key contribution is the notion that the overall distortion can be estimated from the combination of both the coding and resampling distortions. To this end they devise methods for predicting each of these distortions in order to define the optimal scale factor. They then design filters based on a truncated sinc function with knowledge of the scale factor applied. Due to the fact that the resolution is not adaptive, the only way to truly determine the optimal scale factor is by analysing the whole sequence. This would limit its use to offline applications, not real-time. However, they do note that video content typically does not largely change over time and therefore analysing the information in the first frame can provide a good prediction overall. Much better objective and subjective results are presented compared to standard AVC. However, unless the optimal scale factor for a sequence truly does not change over time this method would not be suitable for real-time applications.

2.8 Performance Evaluation

To evaluate the performance of the proposed techniques within this thesis, various tools are used to indicate levels of picture quality and the average performance gain when coding within a given range of bitrates.

2.8.1 Peak Signal-to-Noise Ratio

The most widely used objective quality metric is the Peak Signal-to-Noise-Ratio (PSNR) [27]. The PSNR is a MSE based similarity metric, which indicates how similar two images are on a pixel by pixel basis. It is given as:

$$(2.23) \quad PSNR = 20 \log_{10} \left(\frac{2^B - 1}{MSE} \right),$$

where B is the number of bits used to represent each pixel (typically $B = 8$) and MSE is the Mean Squared Error as given in 2.15.

2.8.2 Structural Similarity Index Metric

PSNR is widely considered an unreliable metric for evaluating subjective quality. For this reason the Structural Similarlity Index Metric (SSIM) [70] was proposed which aims to analyse differences between the features within two images that a human observer would notice. These features are contained within the luminance channel of an image and should therefore not be applied to the chrominance channels. SSIM first evaluates the level of luminance and contrast masking as studies in Cognitive Psychology have identified the phenomenons which indicate that errors are more noticeable in areas of high luminance (brightness) and also less noticeable in areas containing high frequencies that form textures. The formula for SSIM is given as:

$$(2.24) \quad SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

where x and y are the two images (or a section of two images) that are compared, μ_x and μ_y are the average values and σ_x^2 and σ_y^2 are the variance values of x and y , respectively. σ_{xy} is the covariance of x and y , and c_1 and c_2 are typically defined by $c_1 = 0.01(2^B - 1)$ and $c_2 = 0.03(2^B - 1)$.

2.8.3 Best Rate-Distortion Curves

Throughout this thesis sequences are coded using different parameters, including spatial resolution. In some cases an algorithm is not applied to predict the optimal output prior to coding as it

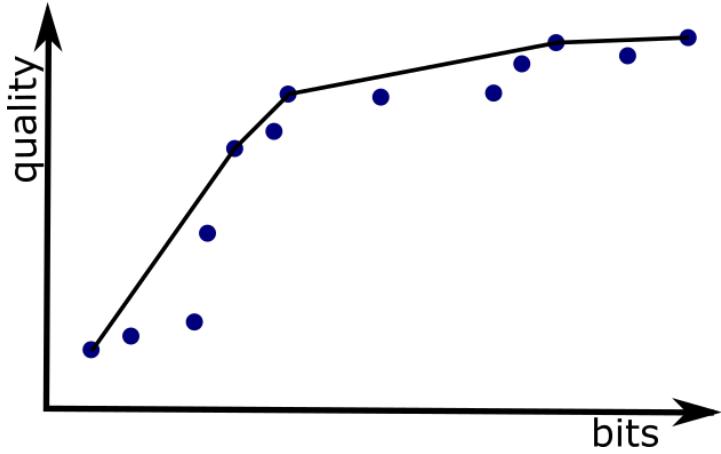


Figure 2.19: Illustration of how the best rate-distortion curves are generated from all available data points.

is often desirable to first code using a large selection of these parameters to identify potential performance gains. In such cases the best curve is plotted, given all available data points, by maximising the gradient between any two points. This process is illustrated in Figure (2.19).

2.8.4 BD-PSNR

In 2001, Gisle Bjontegaard proposed a metric for calculating the average difference between two rate-distortion curves – MATLAB code is given in [3]. Known in the literature as BD-PSNR, this metric is widely used to evaluate the compression efficiency over a range of bitrates. Throughout this thesis, BD-PSNR is used to analyse the average increase in performance that the proposed techniques provide over standard coding.

2.9 Chapter Summary

Video coding and compression is a large area of research containing numerous algorithms and techniques, all of which contribute to the overall level of compression efficiency and provide a robust video bitstream. Over time video coding standards have increased in complexity but more advanced algorithms provide a much higher level of compression efficiency. HEVC is the latest video coding standard produced by the JCT-VC. The research in this thesis is carried out by applying the HEVC standard as a baseline and applies modifications and additions to the standard in order to achieve better performance at low bitrates.

Previous work has demonstrated the potential for improving rate-distortion performance given a tight bit-budget. However, the majority of these methods lack details to provide a fully operational and practical solution. Also, these methods either compare against (and are imple-

mented within) older image or video codecs which lack the compression performance achievable with HEVC at high resolutions. As compression efficiency increases, the performance gained by applying an adaptive resolution approach would decrease. In order to significantly improve rate-distortion performance compared to standard HEVC we also need to focus on reducing resampling distortion by implementing more intelligent resampling methods.

The scalable video coding extensions are of interest as they apply functions for coding pictures at multiple resolutions. These functions are analysed in the later Chapters of this thesis and compared to alternative methods. Scalable coding is not designed to provide greater levels of compression efficiency, however given that previous work has shown that spatial resampling of video sequences can increase rate-distortion performance at low bitrates, it is possible that some elements of these extensions could be applied for this purpose.

Chapter 3 describes the proposed solution and presents initial findings which provide insight into the level of performance achievable. Unlike the previous work described in this chapter, the solution proposed in this thesis not only aims to provide greater performance gains but also aims to take further steps into designing a more practical real-time solution.

SPATIAL RESAMPLING OF IDR PICTURES

As the demand for higher quality and higher resolution video increases, many applications fail to meet this demand due to low bandwidth restrictions. One factor contributing to this problem is the high bitrate requirement of the intra-coded IDR picture featuring in all video coding standards. Frequent coding of IDR pictures is essential for error resilience in order to prevent the occurrence of error propagation. However, as each one consumes a huge portion of the available bitrate, the quality of future coded frames is hindered by high levels of compression. This chapter presents a novel coding technique and shows how it can increase rate-distortion performance by providing a higher and more consistent level of video quality at low bitrates.

HEVC provides significantly higher compression efficiencies at HD and UHD resolutions compared to previous standards [49]. Despite increased coding performance, bandwidth restricted video applications often fail to achieve an acceptable and useful level of video quality. This is especially true of HR video, warranting the need for additional coding and processing.

Frequent coding of IDR pictures can impact the performance of the rate control algorithm, especially at low bitrates – rate control will assign a high QP to inter-coded pictures to ensure the video bitrate matches the target bitrate, which could potentially reduce the overall level of video quality. This is illustrated in Figure (1.1) and Figure (3.12a). For these reasons, the proposed method focuses on resampling IDR pictures only and encodes inter-coded pictures at the original HR. This means that, unlike previous video resampling methods [7, 12, 13, 42, 48, 59, 74] which were discussed in Chapter 2, the majority of the video will be decoded at the original HR (either

HD or UHD) and not upsampled afterwards. As will be shown in the results, reducing the required data size of IDR pictures can enable a more effective use of rate control and provide a better and more consistent level of video quality at low bitrates.

This chapter presents the initial design and develop stages of the proposed coding technique. Bicubic resampling (see Chapter 5) is used to generate the LR input pictures and also to reconstruct the HR picture after coding to be displayed and placed into the Reference Picture Buffer (RPB). In Chapter 5 alternative filters and techniques are discussed in order to enhance the resampling process. Relative to its complexity, Bicubic resampling is capable of generating high quality resampled images but may fail to maintain the fidelity of high frequency information. The results presented in this chapter are by no means optimal but along with the methodology they provide a *proof of concept* and what can be considered as the minimum level of performance achievable with this initial version of the proposed coding technique.

3.1 Performance of Resampled Intra Coded Pictures

Coding IDR pictures at a lower spatial resolution will only provide a benefit if the overall distortion is lower than the amount of coding distortion introduced at the original HR given the same bitrate. To ensure that spatial resampling of intra-coded pictures can increase rate-distortion performance an initial test can be performed by comparing sequences coded in intra mode at the original resolution and at multiple lower resolutions.

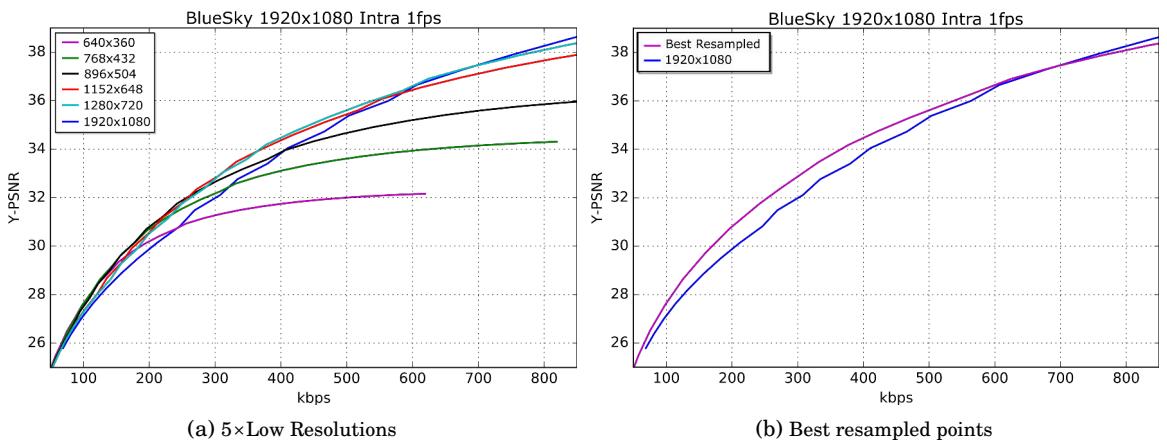


Figure 3.1: BlueSky (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Results show a clear benefit over a wide range of bitrates by coding intra pictures at lower resolutions

3.1. PERFORMANCE OF RESAMPLED INTRA CODED PICTURES

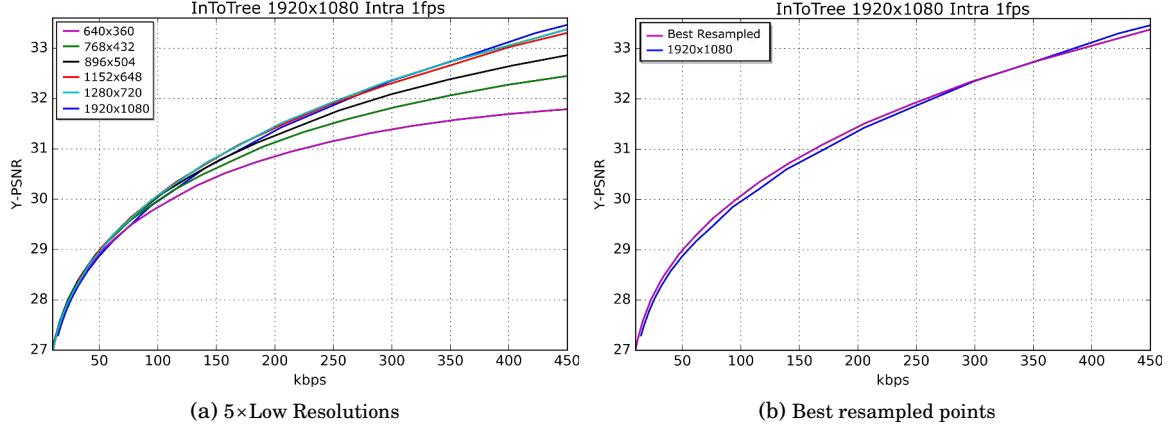


Figure 3.2: InToTree (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Some gain in performance is possible by coding intra pictures at lower resolutions

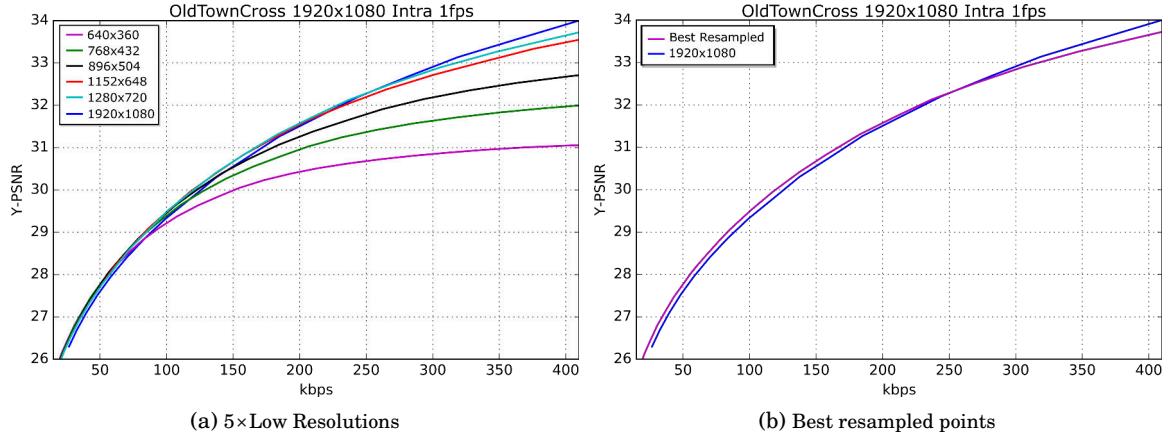


Figure 3.3: OldTownCross (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Some performance gains can be achieved by coding intra pictures at lower resolutions

3.1.1 Intra Coding at Low Spatial Resolutions

Using Bicubic resampling, each of the tested raw input sequences are first downsampled to several lower spatial resolutions. The resolutions that can be tested are limited as HEVC requires that both the height and width are multiples of the smallest coding unit dimension, which is 8×8 pixels. Additionally, the resolutions selected maintain the original aspect ratio. Table 3.1 lists the widths and corresponding heights of the spatial resolutions tested along with the upsampling scale factor required to reconstruct the picture at the original resolution.

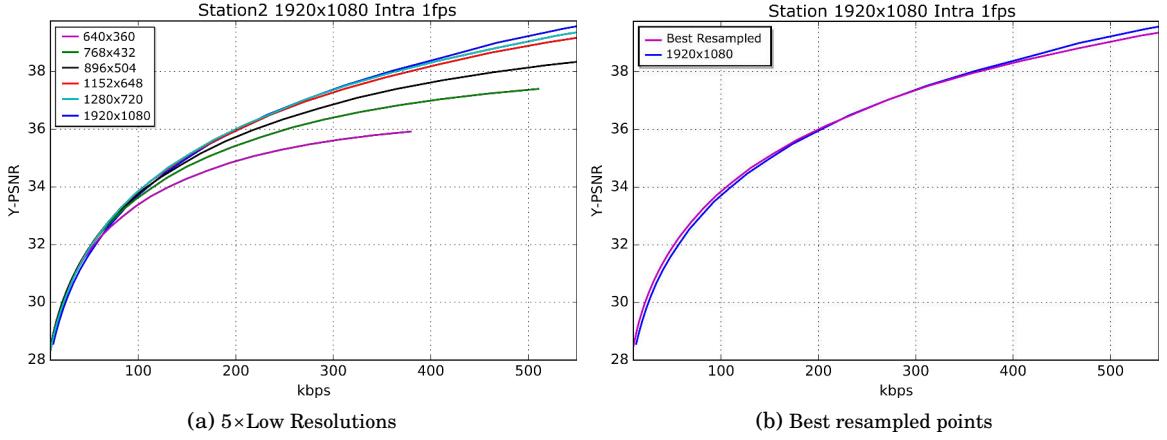


Figure 3.4: Station2 (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. For this sequence, resampling intra-coded pictures provides only a small benefit

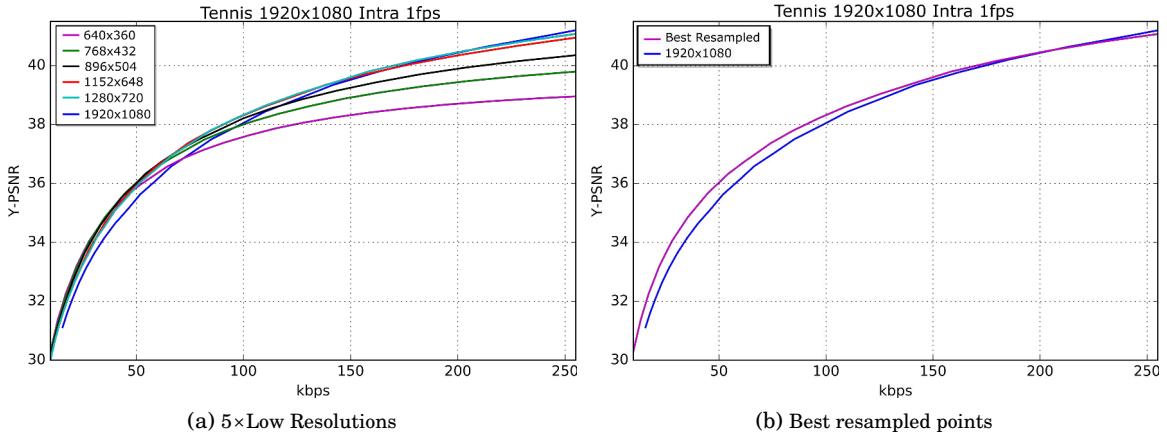


Figure 3.5: Tennis (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Results show a clear benefit by coding intra pictures at lower spatial resolutions given low bitrates

Each scaled version of the sequence is then coded in intra mode. Reducing the spatial resolution may remove some high frequency information originally contained within the HR picture and therefore increase spatial correlation. As a result of this and the fact that fewer coding blocks are required, the LR picture can be coded more efficiently. In other words it can be said that given a number of encoded bits, less quantisation needs to be applied to a lower resolution picture than a corresponding higher resolution picture. To some degree this is content dependent but as the number of coding blocks decreases there will also be less signalling overhead.

3.1. PERFORMANCE OF RESAMPLED INTRA CODED PICTURES

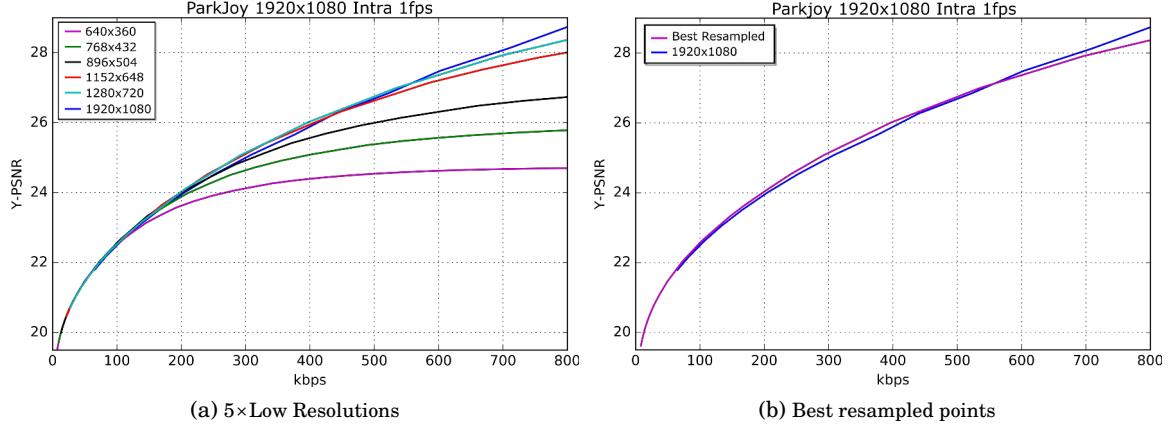


Figure 3.6: ParkJoy (1920×1080) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Results show that some benefit can be achieved by coding intra pictures at spatial lower resolutions

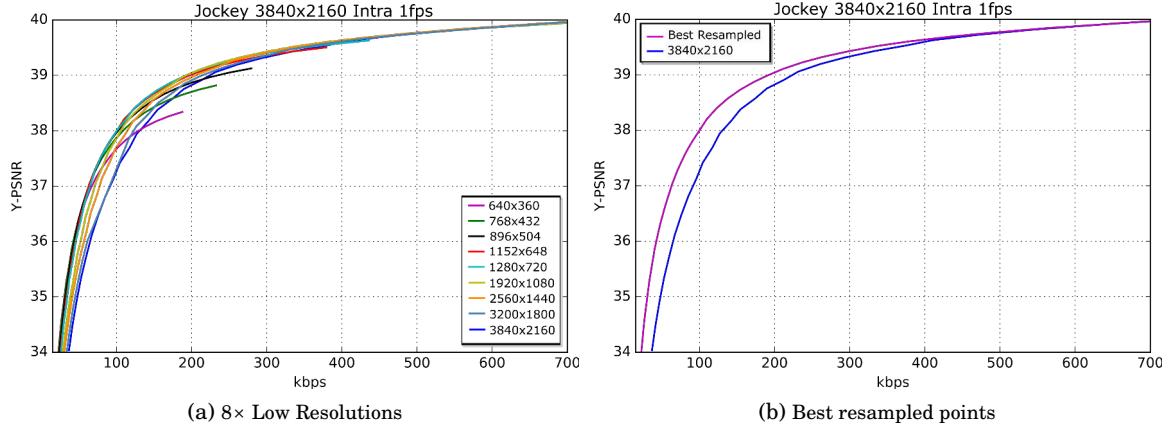


Figure 3.7: Jockey (3840×2160) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Greater performance gains can be achieved by coding intra pictures at lower resolutions

After encoding, each sequence is then decoded and upsampled back to the original HR using Bicubic interpolation. Note that for a practical real-time system each frame will need to be decoded and upsampled in succession, however this is not necessary for these tests. Although coding at lower spatial resolutions can be more efficient, fewer samples and greater scale factors may result in a lower quality reconstructed image. Quality metrics can be used to analyse the similarity of the reconstructed image compared to the original, or predict perceptually noticeable differences in accordance to the HVS – these are discussed in Section 2.8.

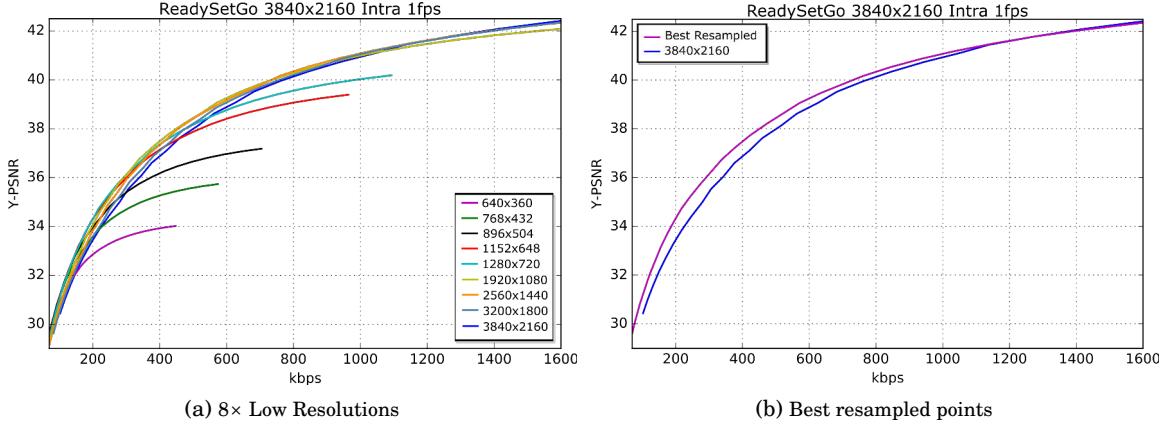


Figure 3.8: ReadySetGo (3840×2160) intra-coded at the original and lower spatial resolutions using Bicubic for resampling. Results show a clear benefit, over a very wide range of bitrates, by coding intra pictures at lower resolutions

Table 3.1: List of spatial resolutions and corresponding scale factors used for testing performance of resampled intra-coded pictures

Width	Height	HD Scale	UHD Scale
640	360	3	6
768	432	2.5	5
896	504	2.14	4.28
1152	648	1.66	3.33
1280	720	1.5	3
1920	1080	1	2
2560	1440	—	1.5
3200	1800	—	1.2
3840	2160	—	1

3.1.2 Analysis of Intra Coded Results

Figure (3.1-3.8) provide rate-distortion results of intra-coded sequences. Each point on each *Best Resampled* curve represents the best resolution given the coded bitrate and quality metric result, and therefore maximises the gradient of the curve. Table 3.2 contains corresponding BD-PSNR [3] results calculated up to the *critical bitrate* [48] – the highest bitrate to provide a benefit. See Section 2.8 for details on generating the *best rate-distortion curves* and information about the BD-PSNR metric. As these results represent intra-coded data and therefore each frame is coded independently, for clarity the frame rate has been normalised to 1Hz. Generally at lower bitrates, coding at lower spatial resolutions provides the best overall performance and as the bitrate increases the optimal resolution also increases. For all tested sequences an increase in rate-distortion performance is possible by resampling to lower spatial resolutions when coding distortion is high. However, the amount gained is ultimately dependent on spatial content. Using the SI index data given in Section 2.6, it can be seen that a lower average amount of spatial

3.2. VISUAL COMPARISONS OF INTRA CODED PICTURES

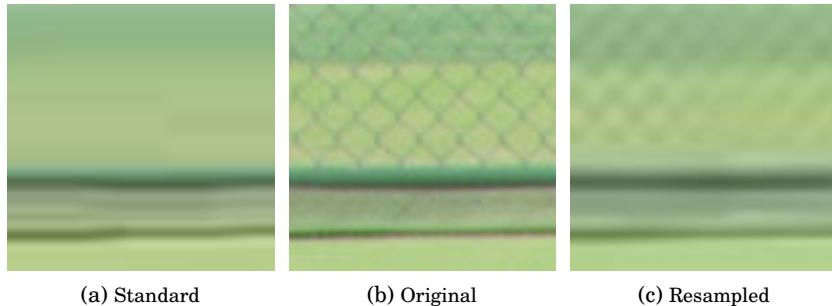


Figure 3.9: Cropped section of an intra-coded picture from the sequence *Tennis*. a) Standard coding at 1920×1080 with a QP of 43 which produces 60.8kbits of data for the entire frame b) original without resampling or coding c) Downsampled to 768×432 prior to coding with a QP of 34, which produces 59.5kbits of data for the entire frame, and then upsampled back to 1920×1080 . The resampled result clearly preserves more high frequency information contained within the original picture and requires less encoded bits

information within a sequence will likely result in a high average PSNR at the critical bitrate. For example, the sequence *ParkJoy* has the highest SI index and also has the lowest average PSNR of 27dB at the bitrate where the curve of the resampled and the standard result intersects. The sequences *Tennis* and *Station2* have the lowest SI indexes out of all the HD sequences and also benefit by improving coding efficiency up to an average PSNR of around 40dB. This suggests that as the amount of spatial information within a sequence increases, the performance of coding at the original HR relative to a LR will increase. Applying more intelligent and optimal sampling techniques will reduce overall distortion and provide greater performance gains, and also increase the critical bitrate – this is explored more in Chapter 5.

Table 3.2: BD-PSNR Metric [3] results – average bitrate savings for comparison of intra-coding performance when coding at lower spatial resolutions prior to upsampling back to the original high resolution. Calculations applied to the best data points from all tested spatial resolutions

Video	BD-PSNR	Critical Bitrate
BlueSky	13.197%	693kbps
InToTree	7.320%	332kbps
OldTownCross	6.097%	253kbps
Tennis	15.156%	206kbps
ParkJoy	4.242%	557kbps
Station	7.959%	279kbps
ReadySetGo	14.217%	1311kbps
Jockey	42.374%	715kbps

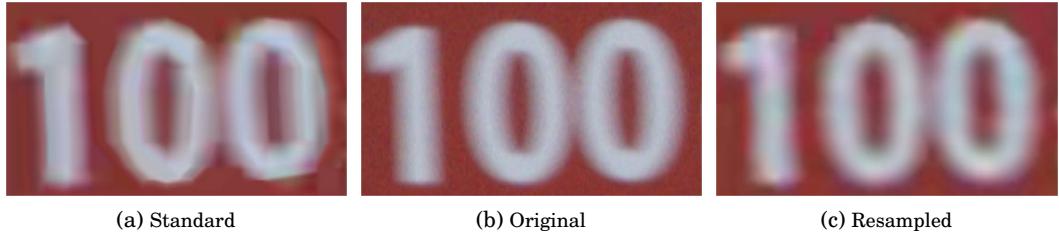


Figure 3.10: Cropped section of an intra-coded picture from the sequence Jockey. a) Standard coding at 3840×2160 with a QP of 45 which produces 106.12kbits of data for the entire image b) original without resampling or coding c) Downsampled to 1280×720 prior to coding with a QP of 34, which produces 104.11kbits of data for entire frame, and then upsampled back to 3840×2160 . The resampled result lacks the high frequencies that form the sharp edges. The standard result has also lost the original fidelity of the edges but has also produced unnatural sharp edges due to ‘blocky’ artefacts

3.2 Visual Comparisons of Intra Coded Pictures

Objective quality metrics, such as PSNR, enable quick and easy analysis of visual quality – or at least provide a measure of how accurately the picture has been reconstructed. However, a more accurate reconstruction may not be perceived to provide a higher level of quality. In this Section, visual comparisons are provided of intra-coded pictures coded using a similar number of bits but at different spatial resolutions.

Figure (3.9) contains cropped images of intra-coded pictures from the sequence Tennis. Standard coding at 1920×1080 with a QP of 43 is compared with the best resampled picture with a bit-budget set by the standard result. Spatial resampling is able to preserve a larger portion of the original high frequency information. Standard coding completely removes the grid structure that forms the fence within the original picture. However, it could be argued that the preserved information within the resampled result appears unnatural and therefore perceptually hinders visual quality – for a more clear evaluation of how resampling can affect perceptual quality subjective tests would need to be performed, but this is beyond the scope of this thesis.

Figure (3.10) contains cropped images of intra-coded pictures from the sequence Jockey, centered around a sign containing the number ‘100’. The standard coded picture was coded with a QP of 45 and produced 106.12kbits of data. The resampled picture was coded at 1280×720 using a QP of 34 and produced 104.11kbits of data. Both methods cause loss of structural fidelity that forms the edges of the numbers but the standard approach also produces ‘blocky’ artefacts thus generates high frequencies not contained within the original picture, which could be perceived as detrimental to visual quality.

3.3 Methodology and Proof-of-Concept

In HEVC, the term *intra period* is used to indicate the frequency at which an intra-coded picture occurs (including IDR pictures) and it is required that the intra period is a multiple of the GOP size, although in the literature these terms are often used interchangeably.

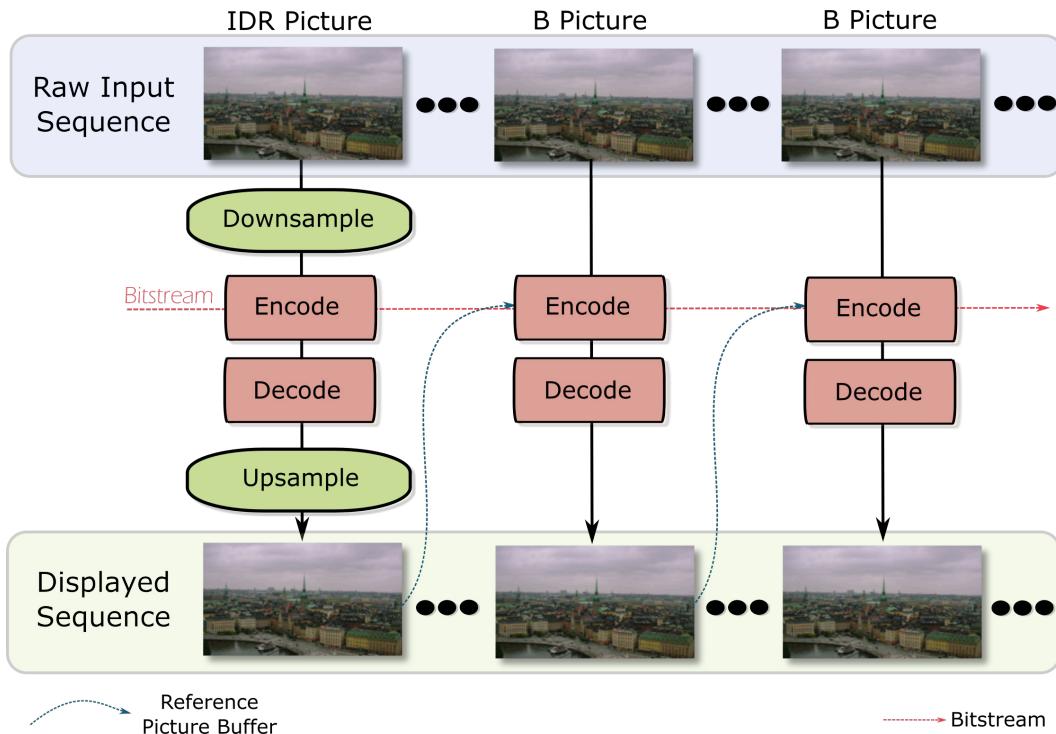


Figure 3.11: Block diagram of the processes involved for spatial resampling of IDR pictures. B-pictures are coded as standard at the original HR. IDR pictures are first downsampled to a lower spatial resolution before coding. After each IDR picture is decoded, it is upsampled back to the original HR to be displayed and placed into the RPB

3.3.1 Proposed Coding Technique

Algorithm 1 Spatial Resampling of IDR Pictures

- 1: Downsample each frame flagged as an IDR picture
- 2: Encode the LR frame as an IDR picture
- 3: Decode and upsample the LR IDR picture back to the original HR.
- 4: Insert the IDR picture into the RPB.
- 5: Encode inter-coded pictures, with the option of referencing the upsampled IDR picture stored in the RPB . Rate control should take into account the reduced bits required to encode the LR IDR picture
- 6: At the decoder, repeat steps 3 and 4 before decoding the remaining inter-coded pictures.

Figure (3.11) and Algorithm 1 highlights the key steps involved in the initial build of the proposed coding technique. Rate control, as available within the HM Reference Software, is applied with the purpose of targeting a specified bitrate. This means that a bit-budget is set prior to coding each frame in succession and the best QP is predicted that will generate a number of coded bits within that budget. Without modifying the existing rate control algorithm it will not take into account the reduced spatial resolution and therefore will not be able to determine the best QP for IDR pictures. As an initial solution to this problem, for the data collected in this chapter, the whole of each sequence is coded at each LR using the rate control algorithm and then each LR IDR picture is used in substitution of the IDR pictures at the original HR. This means that, when coding at the original HR, the rate control algorithm will be affected as the number of bits used when coding the LR IDR pictures is likely to be reduced. In Chapter 4 a more practical solution is developed and then is applied later in Chapter 6.

A major difference between the proposed coding technique and other existing video resampling techniques is that only IDR pictures are resampled and not the entire sequence. It is therefore required that both downsampling and upsampling are performed at the encoder in order to encode inter-predicted pictures from the upsampled IDR picture. Although the encoded data size of each IDR picture will likely be reduced, a portion of the bits saved could potentially be lost as resampling distortion will increase the amount of residual information between the IDR picture and the first inter-predicted picture.

This work bears some similarities to scalable coding but ultimately aims to tackle a different issue – increasing rate-distortion performance. As described in Chapter 2, scalable coding provides increased adaptability for receiving devices. The spatial base layer in SHVC with a frame rate equal to the intra period is comparable to the coding structure as described in this chapter. However, the aim of the proposed coding technique is to always display video at the original HR and also no inter-layer prediction is performed.

3.3.2 Experimental Results

3.3.2.1 Encoder Configurations

The proposed technique is applied to 5 test sequences and compared to standard HEVC. Details of each test sequence are given in Appendix-A. As discussed in Chapter 1, the Low Delay coding configuration is used to reduce latency, thus all frames are coded in display order. Alternative coding structures, such as *Random Access*, encode inter-predicted pictures out of display order – this can enhance compression efficiency but also increases latency. As well as using the Low Delay configuration, a frequent intra period is used to provide a robust video bitstream that is highly resilient against errors. All sequences are encoded with a frame rate of 25Hz and results are given for a bitrate range between 100kbps - 3Mbps. As mentioned, the QP for each LR IDR

3.3. METHODOLOGY AND PROOF-OF-CONCEPT

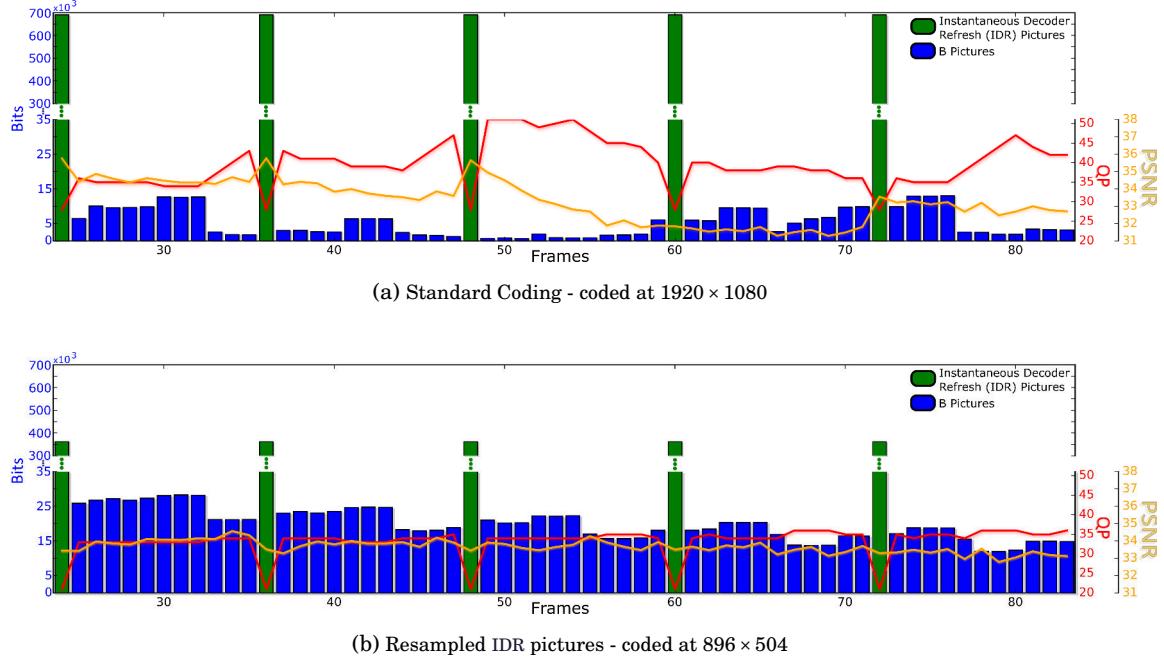


Figure 3.12: Rate control performance according to bit consumption, assigned QP, and corresponding PSNR within five intra periods – Sequence: InToTree, 1920×1080 , 25Hz, Target Bitrate 1000kbps. Without any modification as given in a), rate control allocates a relatively large bit-budget for IDR pictures. The assigned QP is low resulting in very few bits allocated to the remaining B pictures within the GOP. Inter-coded frames are therefore highly compressed resulting in large variations of PSNR. The proposed result, as given in b), reduces the bit-budget of IDR pictures and encoded them at a spatial resolution of 896×504 . The PSNR for each IDR picture is reduced as a result of lowering the bit-budget, however it is likely that more information is retained by coding these pictures at a lower spatial resolution. It is clear from these graphs that the proposed method produces a more consistent level of quality and bitrate

frame is currently determined by first encoding the entire sequence at the LR using rate control and the same coding parameters, however, for practical use this would be determined by the rate control algorithm in real-time. To evaluate the performance of the technique the objective quality metrics PSNR and SSIM [71] are used (see Section 2.8).

3.3.2.2 Results

Figure (3.12) shows the bit consumption and rate controlled QP values within five intra periods for standard HEVC coding and also coded using the proposed technique. The plots represent data from the sequence *InToTree* coded with a target bitrate of 1000kbps. For the proposed coding technique, IDR pictures are downsampled to a spatial resolution of 896×504 , as this resolution provides the best results at this bitrate. The reduced bit consumption of IDR pictures allows rate control to allocate a much lower and more consistent level of compression. This can lead to

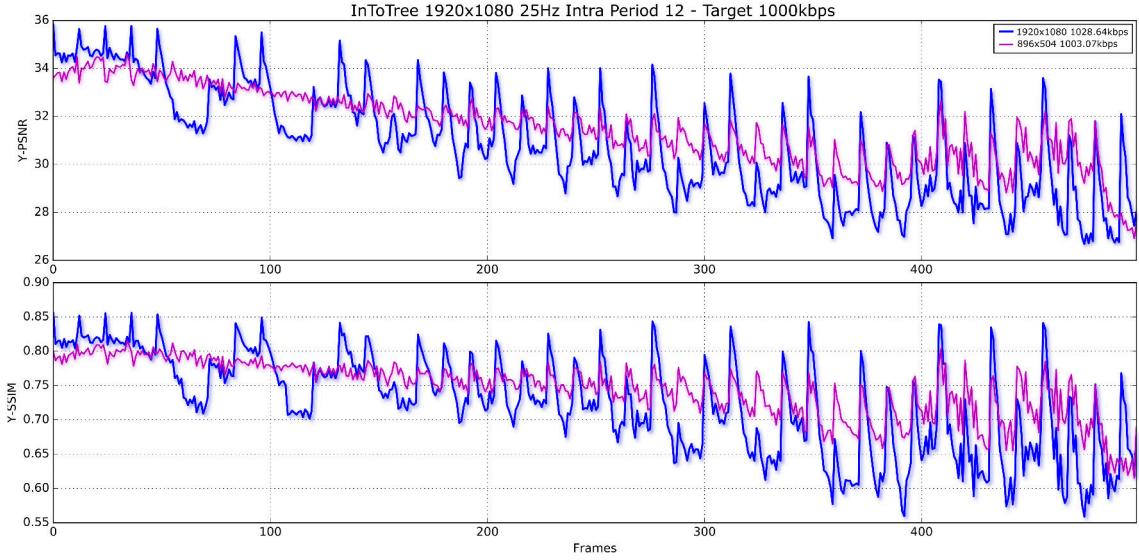


Figure 3.13: Frame by frame quality comparison between standard coding and the proposed coding technique – Sequence: InToTree, 1920x1080, 25Hz, Target Bitrate 1000kbps. Results clearly show that the proposed technique produces a coded sequence that has a more consistent level of video quality. The quality of standard coding with the unaltered rate control algorithm is much more varied with higher peaks and lower troughs

Table 3.3: Statistical results of temporal variation of quality for sequence *InToTree* with a Target bitrate of 1000kbps. The proposed method produces a higher mean of both PSNR and SSIM but also less variation and a higher minimum

	Standard HEVC				Resampled IDR 896 × 504			
	Mean	σ	Min	Max	Mean	σ	Min	Max
PSNR	30.87dB	2.25	26.68dB	35.83dB	31.52dB	1.57	26.92dB	34.58dB
SSIM	0.724	0.071	0.559	0.856	0.746	0.041	0.615	0.815

less variation and a more consistently high level of quality, as shown in Figure (3.13). Figure (3.13) and Table 3.3 provide graphical and statistical results, of the same coded sequence, in order to compare temporal variation of quality. With a target bitrate of 1000kbps, an IDR picture resolution of 896×504 produces a higher average of both PSNR and SSIM but also a smaller standard deviation and a higher minimum. Furthermore, as shown in Figure (3.14), standard coding produces a highly inconsistent video bitrate that will likely cause high levels of latency and possibly skipping of frames within a real-time communications link. These periods of high video bitrate tend to correspond to the high quality peaks shown in Figure (3.13). It can therefore be said that the rate-distortion results, which are given in Figures (3.15 - 3.19), do not represent true performance as the quality levels of the actual frames displayed, when applying standard coding, are likely to be lower than the average.

3.3. METHODOLOGY AND PROOF-OF-CONCEPT

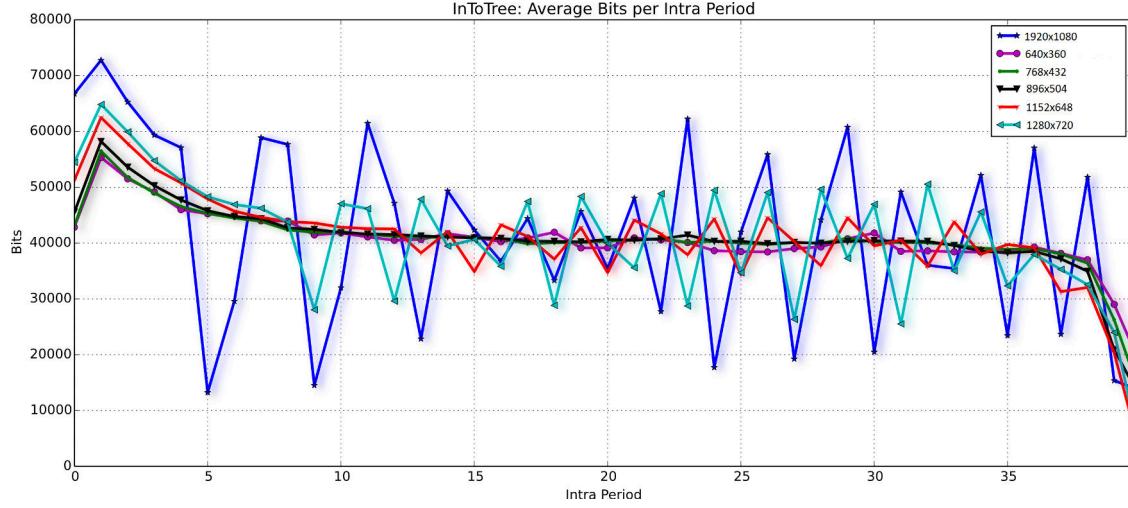


Figure 3.14: Average number of bits per intra period for standard coding and the proposed coding method at each resolution – Sequence: *InToTree*, 1920×1080 , 25Hz, Target Bitrate 1000kbps. At higher spatial resolutions for IDR pictures, the corresponding bit-budget determined by the rate control algorithm is greater. This leads to greater variation of the video bitrate as the defined bit-budget of each GOP changes abruptly in order to account for previous inaccuracies when attempting to match the target bitrate

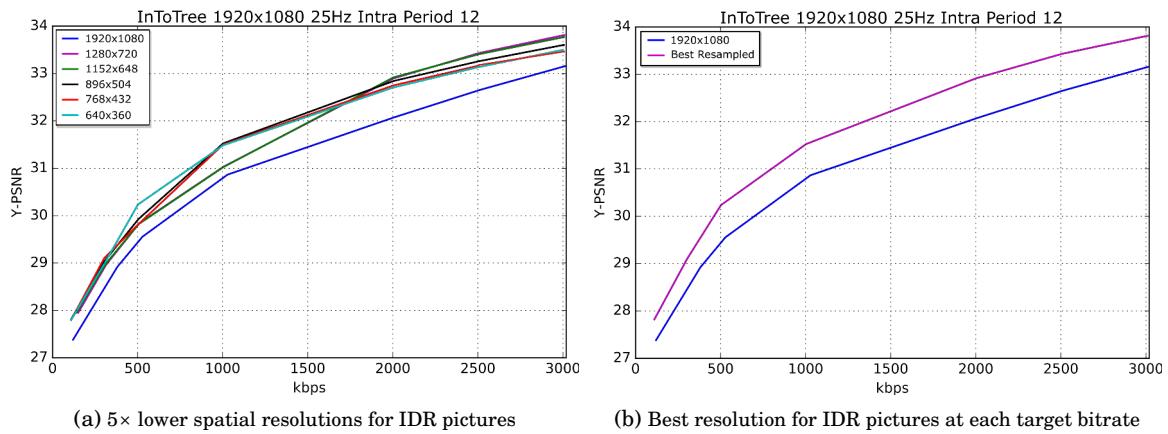


Figure 3.15: Rate distortion performance comparisons – Sequence: *InToTree* 1920×1080 25Hz Intra Period 12. Results show a clear benefit by applying the proposed technique. However, given that intra-coded results only indicated a small increase in performance for resampling intra picture, it is suspected that the majority of this performance gain is due to better allocation of the IDR bit-budget

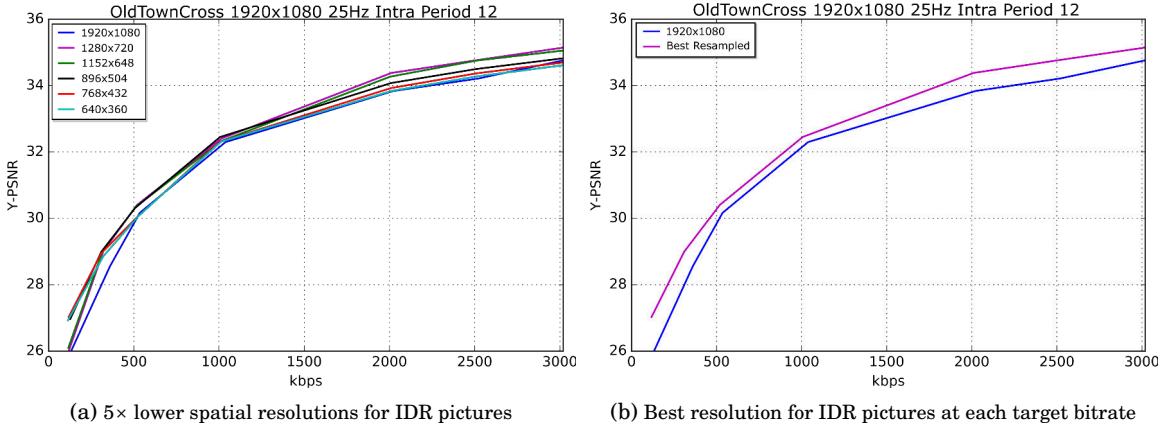


Figure 3.16: Rate distortion performance comparisons – Sequence: *OldTownCross* 1920×1080 25Hz Intra Period 12. Results show considerable benefit when applying the proposed technique. Like the case for sequence InToTree, it is suspected that much of this performance gain is due to better allocation of the IDR bit-budget as only a small gain in resampling intra-coded pictures was shown

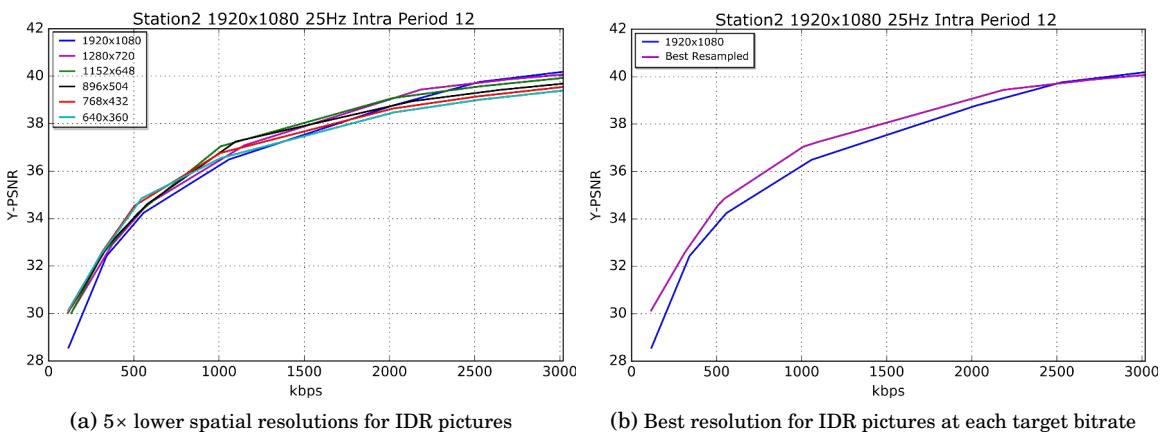


Figure 3.17: Rate distortion performance comparisons – Sequence: *Station2* 1920×1080 25Hz Intra Period 12. Significant performance gains can be achieved with the proposed coding method. At this stage it is unclear how much is due to the increase in efficiency as a result of resampling or due to the fact that the bit-budget of IDR pictures is reduced

3.3. METHODOLOGY AND PROOF-OF-CONCEPT

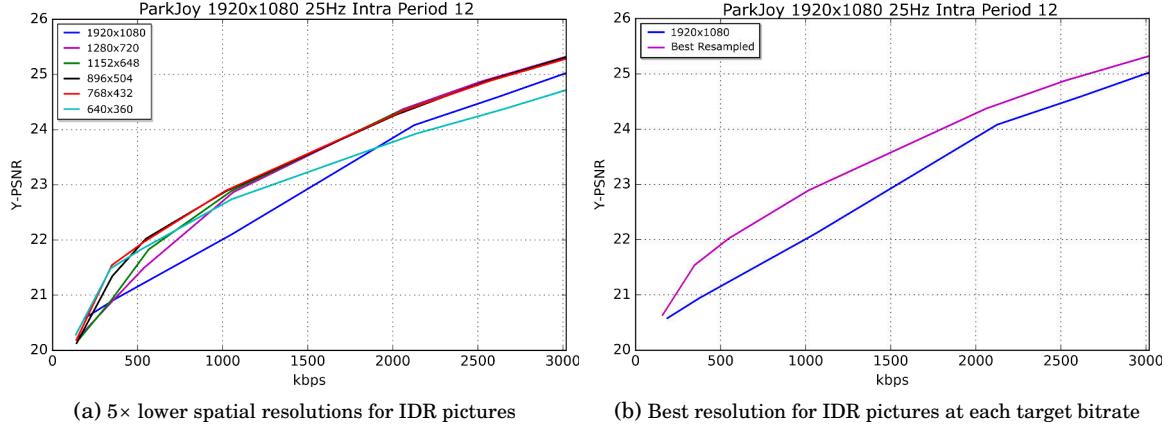


Figure 3.18: Rate distortion performance comparisons – Sequence: *ParkJoy* 1920×1080 25Hz Intra Period 12. As only a small benefit was observed by resampling intra-coded pictures within this sequence, it is suspected that the majority of this performance increase is due to the reduction of the IDR bit-budget

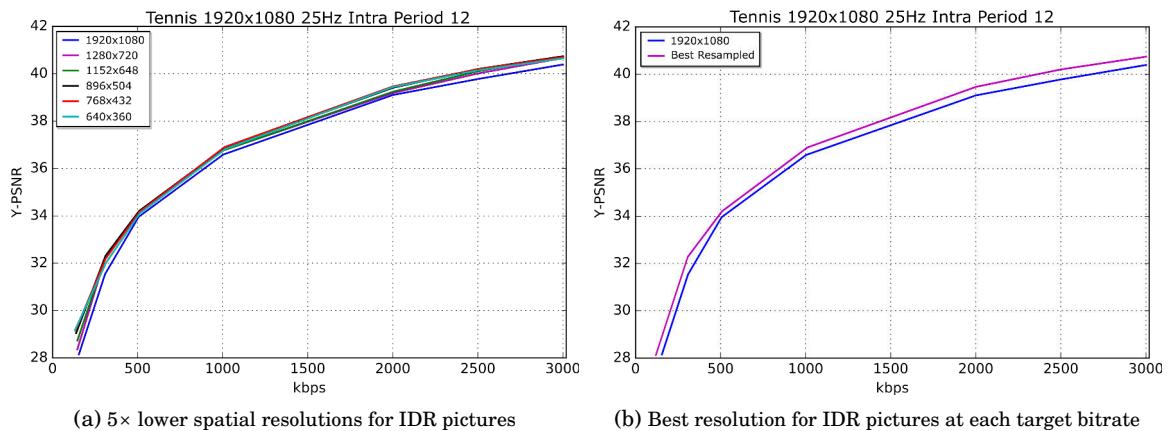


Figure 3.19: Rate distortion performance comparisons – Sequence: *Tennis* 1920×1080 25Hz Intra Period 12. Results show performance gains across the entire bitrate range. Although intra results only showed benefit for a small range of bits, they indicate that the potential performance gained by resampling IDR pictures is high given a low bit-budget. It is possible that much of this level of rate-distortion performance gained using the proposed technique is due to the resampling of IDR pictures. However, at higher bitrates it becomes increasingly likely that rate control will select a higher IDR bit-budget and therefore any performance gained is likely due to the reduction of this budget

Each tested sequence is encoded using 5 different resampling scale factors, which are the same as those given in Table 3.1. From the rate-distortion curves plotted in Figures (3.15 - 3.19) it can be seen that the critical bitrate is generally higher for smaller scale factors – this trend can also been seen in the intra-coding results. Therefore, for varying channel conditions the optimal scale factor will likely change. The critical bitrate is also dependent on the complexity of video content; results suggest that as the TI index (as given in Section 2.6) increases the average PSNR decreases at the critical bitrate, however, spatial information also has an affect on the results as the intra-coded pictures are coded at a lower resolution using the proposed method. In each case, an increase in video quality for a range of bitrates is achieved. It should also be noted that the rate-distortion plots show the average level of quality and do not show the variance of quality for each coded sequence. Therefore a sequence of equal average quality, or possibly marginally lower, is likely to be of a higher subjective level of quality when applying the proposed coding technique due to greater consistency.

3.4 Chapter Summary

This chapter has demonstrated that resampling and coding IDR pictures at lower spatial resolutions can improve rate-distortion performance at low bitrates. Coding in intra mode gives some indication of the level of quality that can be gained for IDR pictures. The results presented in this chapter use fixed scale factors for each frame and the optimal is selected according to the average result which maximises the gradient of the curve. As content within a sequence can vary over time, coding each intra picture at the same resolution may not produce the best results. However for each sequence it is shown that a considerable gain in the efficiency of intra-coded pictures is possible using Bicubic resampling.

When coding in inter mode using the proposed technique, each tested sequence shows benefit for a wide range of bitrates. Any gain in performance is dependent on two factors: the reduced bit-budget which is selected by the rate-control algorithm when coding at the lower spatial resolution, and also any benefit gained by resampling the IDR pictures. Reducing the IDR bit-budget can also improve rate-distortion performance, even without resampling. These two factors should therefore be independent – the bit-budget should not be dependent on the spatial resolution. With the current design of the proposed technique used in this chapter it is unclear which of these factors is contributing most to the overall increase in performance. As the IDR bit-budget is determined by the rate control algorithm this is not the focus of this thesis, however this is discussed more in Chapter 6. Chapter 4 presents a solution to the problem of selecting the best spatial resolution for each IDR picture given a bit-budget. In Chapter 6 this technique is then applied within an inter-coding framework with independent modification of the bit-budget.

The work in this chapter also demonstrates that a more consistent bitrate and level of quality can be achieved by reducing the IDR bit-budget at low target bitrates. Due to the fluctuations in the bitrate, when applying standard HEVC, many of the frames are likely to be skipped when transmitted over a communications link with a finite channel capacity. It can therefore be argued that the rate-distortion results do not represent a fair comparison as the average level of quality of frames displayed would be much lower without modification of the IDR bit-budget, as is the case with the standard HEVC result. Greater performance gains should be possible by implementing a better resampling technique that is able to more accurately reconstruct high frequency content, thus enhancing compression efficiency. This is discussed more in Chapter 5.

To summarise, the issues that have arisen as result of the work achieved in this chapter, which need to be addressed in the later chapters, are as follows:

1. Decisions are made offline, not in real-time, after coding exhaustively at multiple resolutions.
2. The selected resolution is fixed for all IDR pictures and not adapted independently given the set bit-budget and the combine resampling and coding distortion produced.
3. A basic sampling technique is used which is not always capable of creating high quality reconstructed pictures after coding at a lower spatial resolution.
4. The bit-budget of IDR pictures is selected by the rate control algorithm when coding at each spatial resolution and therefore the specified budget is determined with the assumption that the remaining inter-coded frames will also be coded at the same resolution and not at the original HR.
5. The bit-budget selected is dependent on spatial resolution, however, these parameters should be independent.

Chapter 4 addresses points 1 and 2 as it presents a technique that aims to select the best resampled intra-coded picture for each frame independently and can be run in real-time. Chapter 5 presents a high performance resampling technique well suited for reconstruction of HR IDR pictures and therefore addresses point 3. Chapter 6 addresses points 4 and 5 by first proposing a slight modification to the rate control algorithm that reduces the bit-budget of IDR pictures and allows a greater number of bits to be allocated to B pictures. Then, to increase the rate distortion performance and to provide a better reference for future inter-coded pictures, the Adaptive Resolution R-Q model, as described in Chapter 4, is applied to select the best QP for each resolution and then select the best coded picture.

ADAPTIVE RESOLUTION R-Q MODEL FOR INTRA-CODED PICTURES

Rate control schemes aim to regulate the video bitrate in order to prevent overabundant or excessively compressed coded bitstreams. While in most cases the former will result in frame skipping, the latter can cause unnecessary degradation of video quality. This chapter presents a novel rate control method for intra-coded pictures that adapts the spatial resolution according to the combine resampling and coding distortion.

Rate Distortion Optimisation (RDO) techniques aim to produce the best overall level of video quality given the available channel capacity. Lower QP values reduce the amount of distortion by limiting high frequency degradation. On a frame level, rate control first assigns a bit-budget to the current frame given the buffer status and the available bandwidth. For best results, we wish to select the lowest QP that minimises distortion while preventing overflow of the channel buffer. Compression efficiency is determined from a combination of video content and the effectiveness of the encoder, so without coding there is no way of determining the resulting number of coded bits that each QP will generate with absolute certainty. However, for most applications it is only desirable to apply a single-pass rate control method where the QP is determined prior to coding and not subsequently after exhaustively coding with a range of QP values, as is the case with multi-pass methods. Multi-pass methods provide a means of deducing the optimal QP with a higher degree of certainty but at the cost of increased computational complexity. Selecting the optimal QP that meets these requirements without any prior coding is a challenging problem and one that has encouraged a great deal of research, as is described in [11]. It is also often

referred to as a *chicken and egg* problem; *how can we select the optimal QP without knowledge of the resulting number of coded bits that each QP will generate?*

Both resampling and coding introduces distortion. The work presented in Chapter 3 demonstrated the benefits of coding IDR pictures at a lower spatial resolution for low bitrate video coding. However, the results are not necessarily optimal as each IDR picture is coded with a fixed spatial resolution and is selected based on the one that produces the best overall performance for the sequence. The selected spatial resolution is also determined offline and no real-time rate-distortion optimisation techniques are used. For a practical and real-time system, the best scale factor and QP value would be determined for each IDR picture independently.

4.1 Complexity Measures

In [36], Kim et al. proposed a fast bit allocation method for still image coding using a *complexity* measure, also called a *picture activity* measure. Using an effective complexity measure, it is possible to estimate the number of bits required to code each frame for a selected QP value. This removes the need of an exhaustive search which requires the picture to be coded with multiple QP values in order to find the optimal QP given the channel or storage constraints. For real-time applications this is necessary as coding requires a lot of computation and time.

Image complexity refers to frequency content; the higher the frequencies an image contains, the more bits required to code the image given a fixed QP. This is illustrated in Figures (4.1a, 4.1c, 4.1e, 4.1g). Note that the variation of bits required per frame is a result of the changes in content complexity. It therefore stands to reason that an image with a higher complexity will suffer from more distortion given a fixed bit-budget.

Kim et al. [36] also provide comparisons between various complexity measures, including: the average Gradient per Pixel (GPP) as given in equation 4.1, the Edge Point Ratio (EPR) as given in equation 4.6 , the variance in luminance on a block basis and a DCT-coefficient-based method. It was found that the gradient measure provided the most statistically accurate result which has led to later work on rate control algorithms for intra-coded pictures applying the same approach [31, 32, 66, 78].

4.1. COMPLEXITY MEASURES

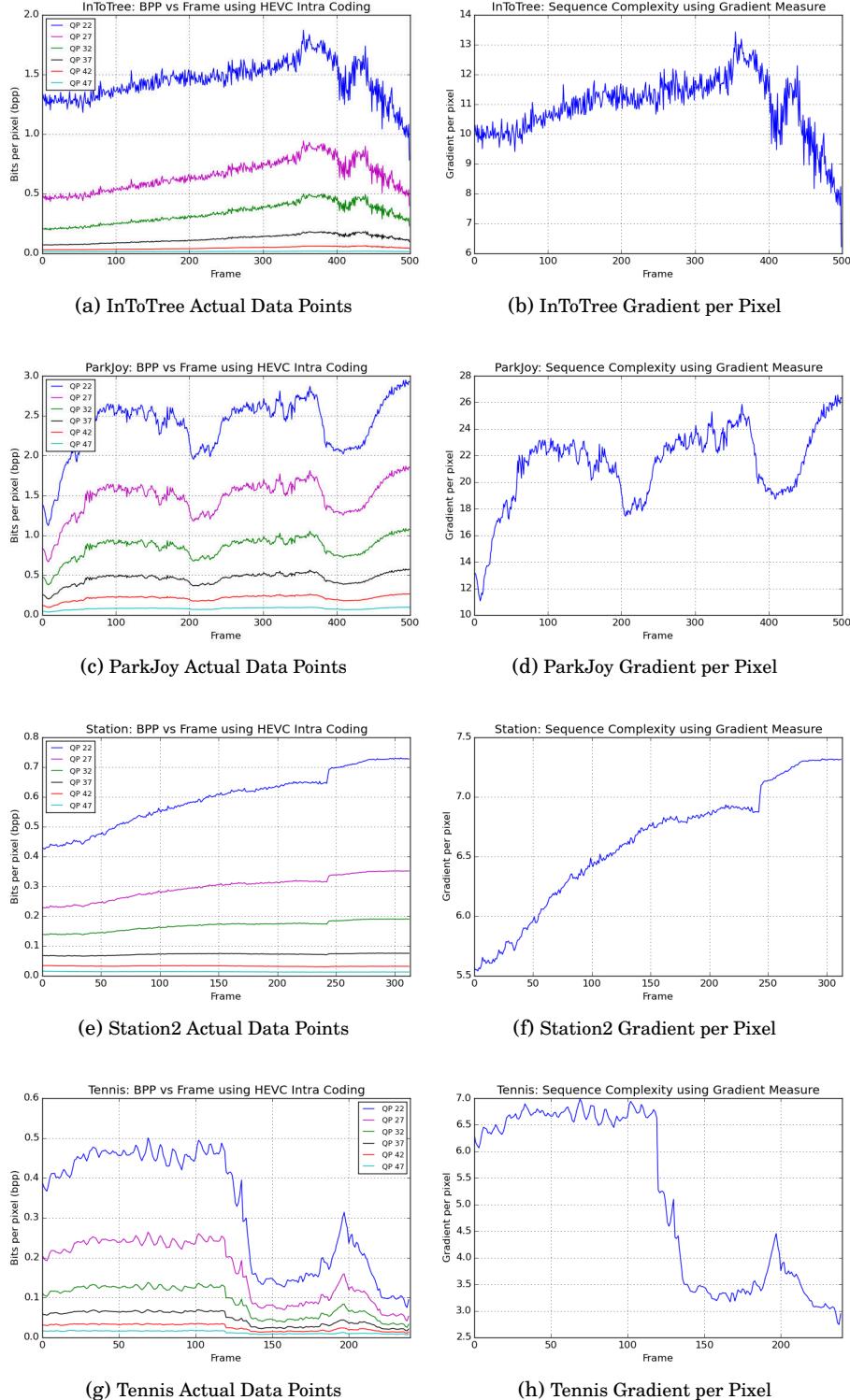


Figure 4.1: Actual coded complexity represented by the average BPP in (a), (c), (e) and (g), and the predicted complexity determined by the average GPP in (b), (d), (f) and (h). It is clear that, even by visual comparison of the two graphs corresponding to each sequence, there is a high level of correlation between BPP and GPP

4.1.1 Average Gradient Per Pixel

The standard formula for the GPP is given as:

$$(4.1) \quad G = \frac{1}{N_w \times N_h} \sum_{i=1}^{N_w-1} \sum_{j=1}^{N_h-1} \sqrt{(I_{i,j} - I_{i+1,j})^2 + (I_{i,j} - I_{i,j+1})^2},$$

where N_w and N_h are the number of pixels in the horizontal and vertical dimensions and I is the luminance image.

In practice, to reduce computation while preserving statistical accuracy, it is often more desirable to use the simplified formula as demonstrated in [31] and given as follows:

$$(4.2) \quad G = \frac{1}{N_w \times N_h} \sum_{i=1}^{N_w-1} \sum_{j=1}^{N_h-1} (|I_{i,j} - I_{i+1,j}| + |I_{i,j} - I_{i,j+1}|).$$

Figure (4.1) shows the variance of complexity on a frame by frame basis for various sequences according to the actual coded data and the GPP. It is clear from these results that this measure is proportional to the average number of Bits per Pixel (BPP). Lower QP values will reduce the amount of quantisation and therefore result in a greater amount of BPP variation, whereas high QP values will remove high frequency components and therefore reduce the BPP as well as the BPP variation. This means that the correlation between the GPP and the BPP will be affected by the QP value. This can be seen from Figure (4.2) – high correlation will produce a graph with points plotted linearly.

Correlation can be calculated using the Coefficient of Determination, denoted as R^2 . The R^2 value determines how well the data fits a statistical model, which in this case is linear. Given a line of best fit for the data presented in Figure (4.2), the corresponding R^2 value can be calculated as follows:

$$(4.3) \quad R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

$$(4.4) \quad SS_{res} = \sum_i (X_i - \hat{X}_i)^2,$$

$$(4.5) \quad SS_{tot} = \sum_i (X_i - \bar{X})^2,$$

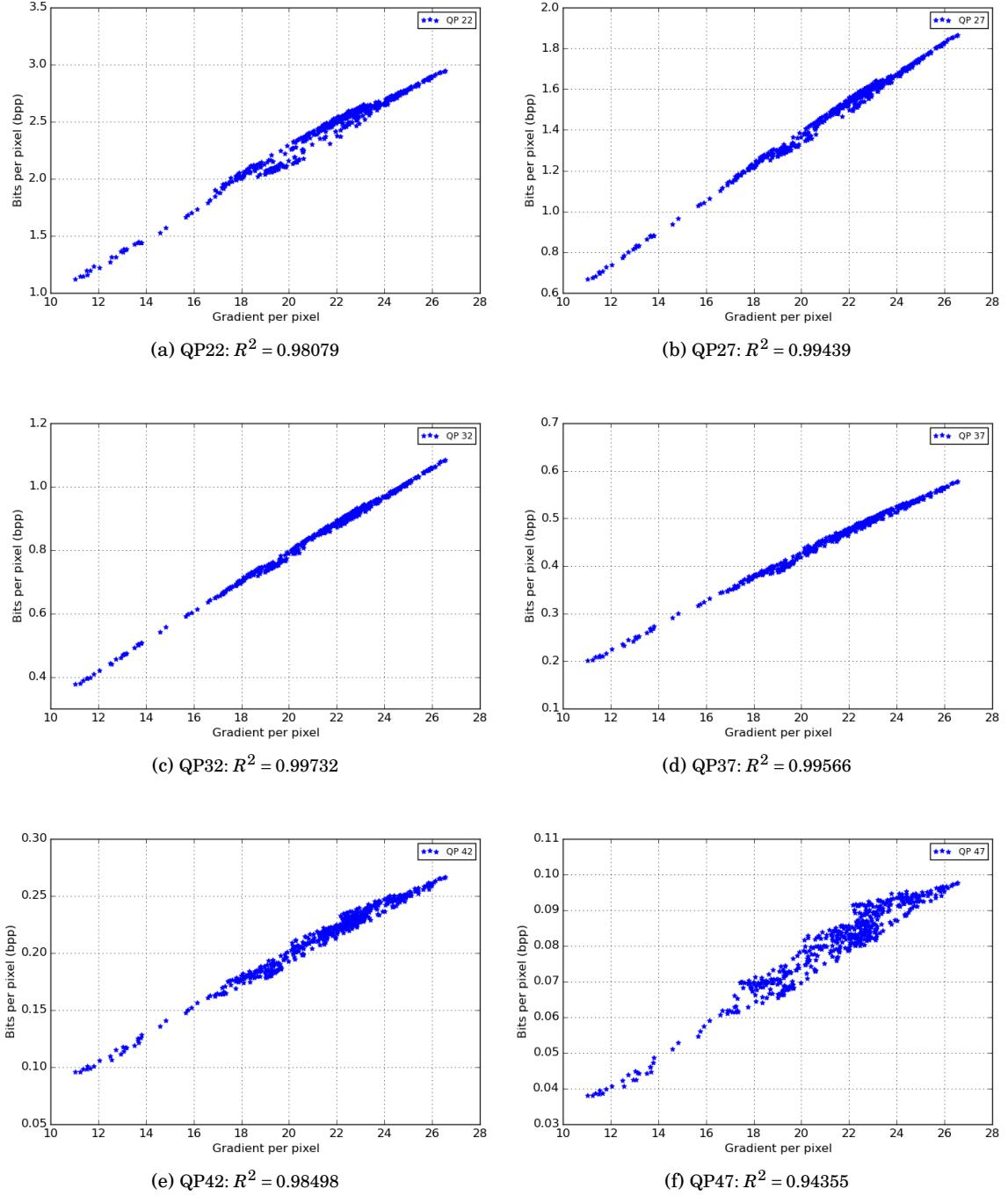


Figure 4.2: Scatter plots of BPP vs GPP for all frames within the test sequence *ParkJoy*. Higher levels of correlation are indicated by data points plotted linearly

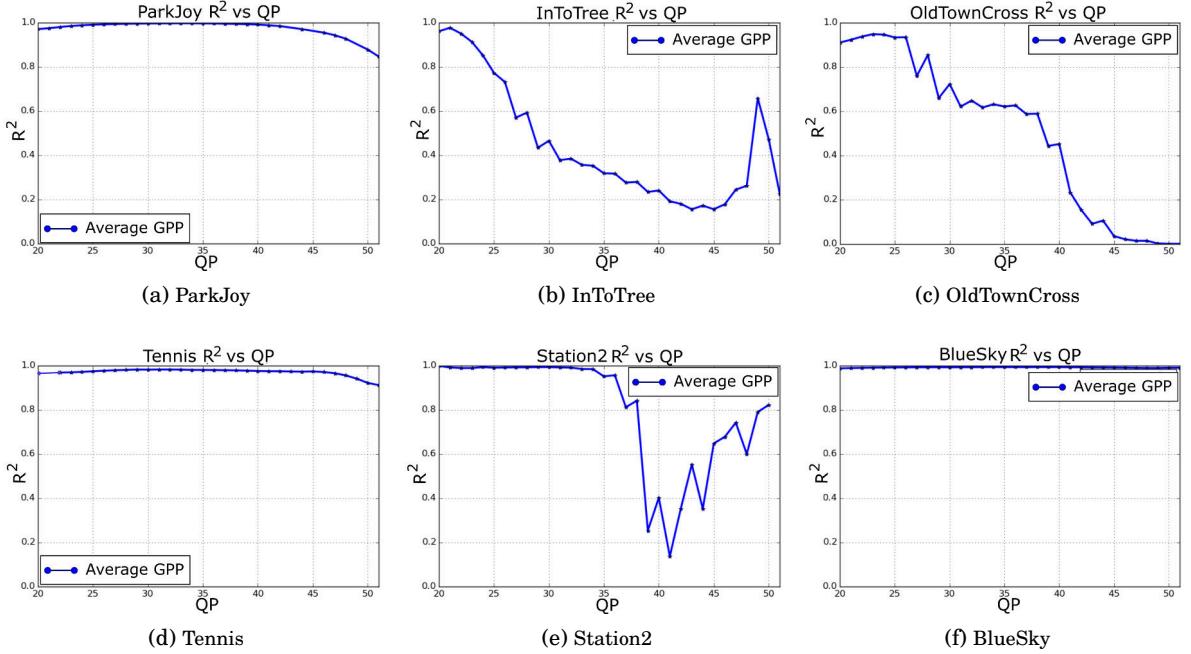


Figure 4.3: The variation of R^2 over QP, using the GPP complexity measure, for six different sequences. These graphs show that the GPP can be an accurate measure of picture complexity for some sequences over a wide range of QP values. However, for other sequences it is not always reliable

where SS_{res} is the Residual Sum of Squares (RSS), SS_{tot} is the Total Sum of Squares (RSS), X is the set of all real data points, \hat{X} is the set of all the estimated data points along the line of best fit and \bar{X} is the mean of all data points in X .

R^2 is given in the range of $[0,1]$ with 1 indicating high levels of correlation. Figure (4.3) shows how R^2 varies over QP for six different sequences. Results are somewhat inconclusive but generally it can be said that the GPP is a more accurate measure when coding using a low QP. It can therefore be said that the GPP is not always reliable. This phenomenon is also highlighted in [66] and it is suggested that an adaptive complexity measure is used.

4.1.2 Edge Point Ratio

In [36] the EPR was found to be a *fair* measure in terms of statistical accuracy. Sun et al. [66] proposed a combined complexity measure using a weighting of the EPR in addition to the GPP. EPR is given as:

4.1. COMPLEXITY MEASURES

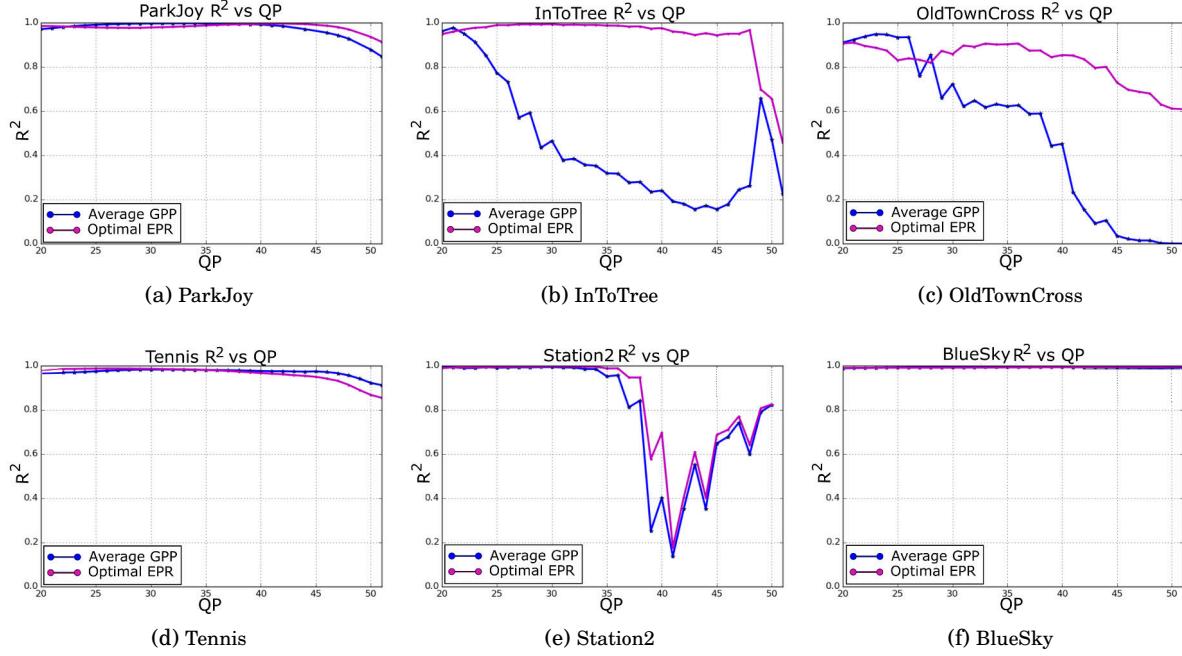


Figure 4.4: The variation of R^2 over QP, using the EPR and GPP complexity measures, for six difference sequences. Given the optimal threshold T , EPR outperforms GPP in most cases

$$(4.6) \quad EPR = \frac{1}{N_w \times N_h} \sum_{i=1}^{N_w-1} \sum_{j=1}^{N_h-1} B_{i,j},$$

$$B_{i,j} = \begin{cases} 1 & |I_G(i,j)| > T, \\ 0 & |I_G(i,j)| < T, \end{cases}$$

where N_w and N_h are the width and height of the image, B is a binary image representing the locations of edges, I_G is the gradient image as given in 2.19 and T is a threshold value limiting the gradient intensities that represent an edge pixel.

In [66], the threshold is chosen according to the value that gives the best correlation result. However, it is not clear if this value is selected independently for each QP or a single threshold is selected that produces the best results across all QPs, for the given sequence. Neither can be applied in a real-time scenario as the whole sequence needs to be coded prior to determining the R^2 correlation value. In Figure (4.4), the optimal EPR threshold is plotted for each QP value and, for comparison, the GPP is also given. With knowledge of the optimal EPR threshold, it is possible to significantly improve the correlation between the complexity measure and the BPP. For most of the tested sequences, the optimal EPR correlates much better overall compared to the

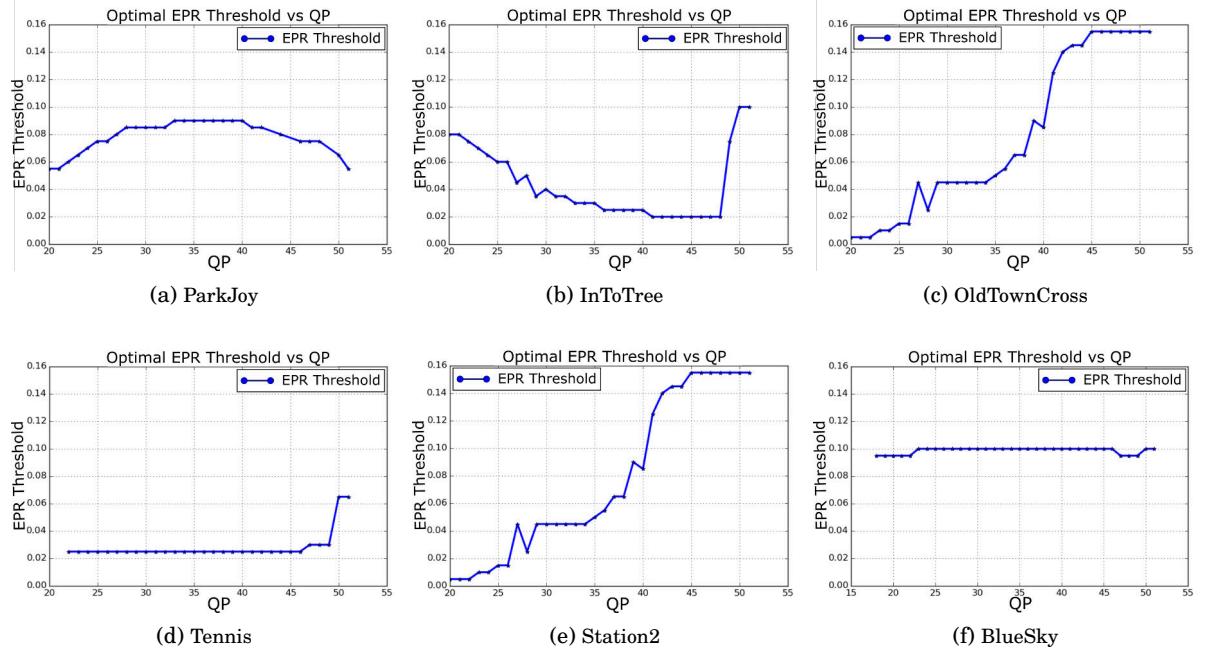


Figure 4.5: Optimal EPR threshold over QP. The changes in the optimal threshold value appear unpredictable and therefore it is unlikely that the EPR will provide an accurate and practical measure for real-time applications

GPP complexity measure. However without any means of predicting the optimal threshold, which is shown in Figure (4.5), the GPP complexity measure is more practical and therefore will be used within the R-Q model described in the next section.

4.2 Single Resolution R-Q Model

For standard HEVC intra-coding, coded at a single spatial resolution, the complexity measures described in Section 4.1 can be used as a means of predicting the resulting number of coded bits given a selected QP. More practically speaking, given a bit-budget we wish to select the QP that will minimise distortion and maximise channel throughput, or select the lowest QP without causing channel overflow.

It was found in [32] that the required number of bits for an intra-coded picture can be predicted by:

$$(4.7) \quad R_{pred}(Q_{step}) = G \cdot \alpha \cdot Q_{step}^{\beta},$$

where R_{pred} is the predicted bitrate normalised to the average number of BPP, G is the average GPP as given in 4.2, $\alpha > 0$ and $\beta < 0$ are parameters that depend on content and Q_{step} is the

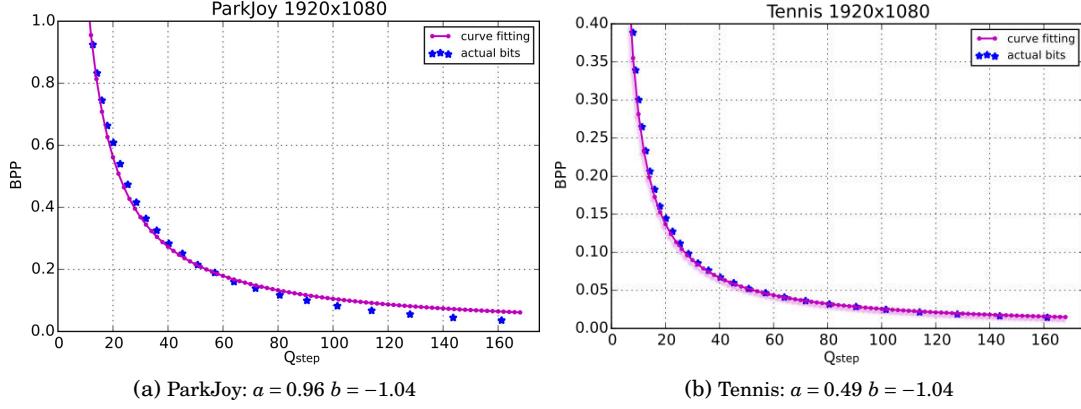


Figure 4.6: Accurate curve fitting using a fixed value of β . Average BPP over Q_{step} size for sequences (a) *ParkJoy* and (b) *Tennis*

quantisation step size which has the following relationship:

$$(4.8) \quad Q_{step} = 2^{\frac{QP-4}{6}}.$$

For any given frame, the optimal values for parameters α and β can be found by solving:

$$(4.9) \quad [\alpha_{opt}, \beta_{opt}] = \operatorname{argmin}_{\alpha, \beta} \sum_{i=0}^{D-1} (R_{actual,i} - R_{pred,i})^2,$$

where D is the total number of data points and R_{actual} contains all actual coded data points.

After solving equation 4.9 for a variety of sequences it can be found that in each case the value of β is fairly consistent and only α is subject to change. The parameter β can therefore be fixed and only α need be adjusted to produce a minimum close to the result produced in equation 4.9, as shown in Figure (4.6). After coding each frame, α can be recalculated by:

$$(4.10) \quad \alpha = \frac{R_{actual}}{G \cdot Q_{step}^\beta}.$$

In [31, 32, 66] an update procedure is performed to account for the changes in video content – QP selection is dependent on α , as indicated by equation 4.7, which is determined from a weighting of values calculated from previous frames. The update procedure is given as:

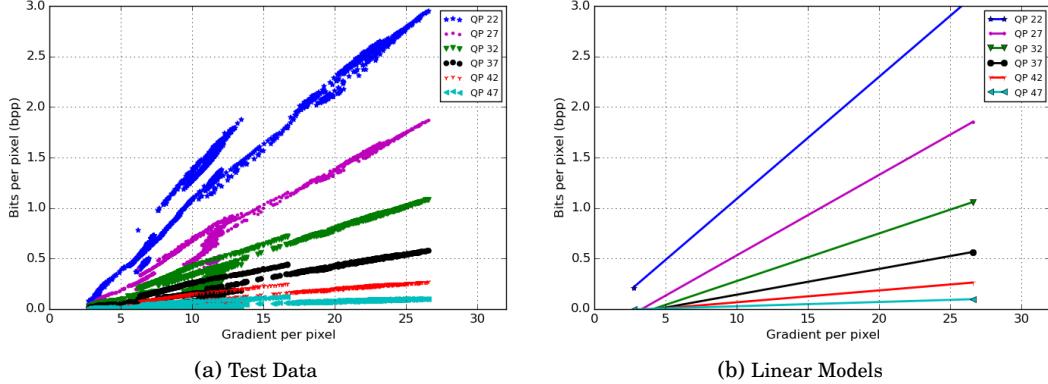


Figure 4.7: By coding a large number of sequences in intra mode, using all possible QP values, a set of linear models can be generated for the purpose of predicting the best QP for the first frame of any sequence. This exploits the high correlation between the GPP and the BPP. Although the best QP will not always be selected, this approach at least provides a higher probability of selecting the optimal QP compared to selecting at random

$$(4.11) \quad \alpha_{k+1} = \lambda \cdot \alpha_k + (1 - \lambda) \cdot \frac{R_{actual,k}}{G_k \cdot Q_{step_k}^{\beta}},$$

where λ is a forgetting factor which is commonly stated to have a typical value of 0.5 [32, 66] and k is the frame index.

When calculating α for the next frame, the forgetting factor determines how much weight is given to the value generated from the current frame over the value generated from the previous frame. A factor of 1 results in infinite memory and therefore the parameter is never updated – the initial value of α is applied to all future frames. Alternatively, a factor of 0 applies the value calculated from the current frame only when estimating the optimal value for the preceding Q_{step} size.

4.2.1 Initial QP Selection

To determine the QP for the first frame, prior calculations can be used to generate a set of linear models. We know from [36] and Section 4.1 that the correlation between the average GPP and the average number of BPP is high. Using data from a range of different video sequences, a set of linear models for each QP can be produced to provide an initial prediction of the optimal QP given the target BPP and the measured complexity of the frame calculated in equation 4.2. Note that these models are only generated once and then applied to all future coded sequences. All proceeding QP selections will be determined based on the results of previously coded frames. For illustration,

Figure (4.7a) shows a scatter graph of coded data points collected from 7 HD sequences for a selection of QP values in the range [22, 47] and Figure (4.7b) shows the corresponding set of linear models generated from these data. In actual fact, data were collected and models generated for all QP values in the range [18, 51] and for all high and low resolution variations of each sequence.

4.2.2 Practical Single Resolution Algorithm

The full algorithm is given in Algorithm 2 which assumes Q_{step} for the first frame is determined using the predetermined linear models, as described in Section 4.2.1, and the parameters for proceeding frames are determined using the update procedure as given in equation 4.11.

Algorithm 2 Intra Single Resolution Rate Control

```

1: procedure RATECONTROL( $f_{in,k}, R_{target}, \alpha_k$ )  $\triangleright$  input frame, target BPP and parameter  $\alpha$  (if available)
2:    $G_k \leftarrow$  average GPP for frame  $f_{in,k}$  using equation 4.2
3:   if  $k == 0$  then
4:      $Q_{step} \leftarrow LinearModels(G_0, R_{target})$ 
5:      $QP \leftarrow$  corresponding QP using equation 4.18
6:      $R_{actual}, f_{out,k} \leftarrow Encoder(f_{in,k}, QP)$ 
7:      $\alpha_{k+1} \leftarrow$  calculate  $\alpha$  using equation 4.10
8:   else
9:      $Q_{step} \leftarrow$  using equation 4.8
10:     $QP \leftarrow$  corresponding QP using equation 4.18
11:     $R_{actual}, f_{out,k} \leftarrow Encoder(f_{in,k}, QP)$ 
12:     $\alpha_{k+1} \leftarrow$  update  $\alpha$  using equation 4.11

```

4.2.3 Proposed Modified Updating Procedure

The performance of the update procedure given in equation 4.11 relies heavily on the correlation between successive frames. Better results can be achieved by applying a weighting of parameters calculated from frames with similar complexities. The rate control algorithm can learn from previously coded frames and make more accurate predictions. The proposed modified updating procedure is given as:

$$(4.12) \quad \alpha_{k+1} = (\lambda \cdot \alpha_k + (1 - \lambda) \cdot \frac{R_{actual,k}}{G_k \cdot Q_{step_k^\beta}}) \cdot (1 - \tau) + \tau \cdot \alpha_G,$$

where α_G is a weighted value determined from the complexities of all previous frames (see equation 4.13) that satisfy certain conditions and τ is selected based on the availability of frames with similar complexities stored.

A normal distribution can be used to determine the weights of previously calculated parameters:

$$(4.13) \quad \alpha_G = \frac{1}{\omega} \sum_{i=0}^{N_G-1} \mathcal{N}(|G_{k+1} - G_i|) \alpha_i,$$

where ω is the normalising factor equal to the sum of the weights and N_G is the number of previously stored parameters with similar complexities as given in 4.16.

GPP complexity and α parameter pairs are only stored if the result from the corresponding coded frame satisfies the conditions: $R_{actual} \leq R_{target}$ and $M \leq \gamma$ where R_{target} is the target number of BPP, M is the mismatch ratio given in equation 4.14 and γ is a threshold.

$$(4.14) \quad M = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\%.$$

Initially τ is set to 0 meaning that the calculated value of α is determined by the original portion of the formula, as given in 4.11. As more frames are coded satisfying the required conditions, there is a greater likelihood that a portion of the stored parameters will have similar picture complexities as the next picture to be coded. The value of τ is increased as the number of parameters within two standard deviations also increases, up to a predetermined value T .

$$(4.15) \quad \tau = \min\left(\frac{N_G}{T}, 1\right),$$

$$(4.16) \quad N_G = \begin{cases} +1 & (|G_{k+1} - G_i|) < 2\sigma \text{ for } i = 0, 1, \dots, K-1, \\ +0 & \text{Otherwise,} \end{cases}$$

where T is used as a limiter, such that as $N_G \rightarrow T$ then $\tau \rightarrow 1$, σ is the standard deviation of the normal distribution and K is the total number of previously coded frames.

It should be noted that the implementation of the modified update procedure described in this section is designed for a constant target bitrate. The same approach can be applied to variable targets by storing parameters in different sets according to R_{target} . This will increase memory requirements and also potentially reduce the effectiveness of the technique if the target bitrate varies greatly over time (a lower and upper bound would provide a balance between memory requirements and the level of performance achieved).

4.3 Adaptive Resolution R-Q Model

Contrary to previous works on rate control for intra-coding, which provide a single resolution solution, adaptation of the spatial resolution requires a rate control algorithm that also has the functionality to select the best resolution by analysing the affect on rate-distortion that each coded picture would have on the overall sequence. It is therefore required to adapt both the QP and the spatial resolution while also generating a bitrate that complies to the set target. For the multi-resolution rate control problem, each frame has multiple solutions; a frame may be encoded at the original HR or downsampled to a lower spatial resolution prior to encoding – frames coded at a lower spatial resolution are then upsampled after coding. This approach has two stages: the lowest QP possible at each resolution is first predicted for a given bit-budget and then the combination that minimises distortion is determined.

The Adaptive Resolution R-Q model uses the same approach as the Single Resolution R-Q model for predicting the best QP. This means that, in order to calculate the GPP at each resolution, the original HR picture is downsampled multiple times to produce each low resolution input $I_{k,s}$. In this chapter, Bicubic is used for the downsampling and upsampling operations. The GPP $G_{k,s}$ can then be used to predict the best QP for each scaled picture, where s signifies the scale factor:

$$(4.17) \quad Q_{step_{k,s}} = e^{\frac{\ln(\frac{R_{target}}{G_{k,s} \cdot \alpha_{k,s}})}{\beta}},$$

and corresponding QP, determined by:

$$(4.18) \quad QP_{k,s} = round(6 \cdot \log_2(Q_{step_{k,s}}) + 4),$$

except for the first frame, which uses the approach described in Section 4.2.1. For detailed steps of how equation 4.17 is derived, see Appendix-B. This means that each intra picture needs to be coded at each resolution as the optimal value of $\alpha_{k,s}$ will be different for each scale s and the value is updated from $\alpha_{k-1,s}$ using equation 4.11.

When selecting the best resolution and QP combination the combine resampling and coding distortions need to be considered. Similar work on predicting the optimal scale factor has been carried out by Dong et al. [12, 13] who show that the combine MSE distortion can be estimated from the addition of both the resampling and coding distortions. The key difference between their approach and the Adaptive Resolution coding method is that they perform resampling on the entire sequence, including inter-coded frames. It is stated that all parameters, including the scale factor, are determined once per sequence meaning that the whole sequence is coded with a fixed resolution to comply with the H.264 standard. This also means that determining

the optimal scale factor can only be achieved by analysing the entire sequence and therefore limiting its use to offline applications, not real-time. As a fixed scale factor is applied to the whole sequence, a standard rate control algorithm can be used to regulate the bitrate and select the best QP given the set target. The technique essentially provides an accurate prediction of the optimal scale factor for the entire sequence without the need to code. Such a technique would be useful for storage of video given a set limitation of the data file size. Alternatively, if it were desirable to transmit a video post-capture, an optimal scale factor could be determined given knowledge of the Constant Bitrate (CBR) that the channel provides. For the scenarios outlined in this thesis, this technique is not suitable but some of the processes involved may be applied in the future for real-time solutions given that the best QP at any resolution can be predicted accurately even if the previously coded frames were each coded at a different resolution. The reason why this is challenging is because equation 4.17 contains two parameters that can not be predicted accurately without previously coding using the same scale factor s , namely $G_{k,s}$ and $\alpha_{k,s}$, and therefore multiple scaled versions of each IDR picture are coded and the decision regarding the optimal one is made after.

Fluctuations in video quality can be seen as a temporal artefact, thus reducing perceptual quality. A coded sequence that has a high average PSNR but is also highly varied over time will not necessarily be preferred over one that produces a lower average and a lower variance of PSNR. Adaptive Resolution coding therefore adds some much needed redundancy; although only one scaled picture $I_{k,s}$ will be displayed and uploaded into the RPB, there is a much higher probability that the selected picture will enable greater consistency of quality, ensure a low mismatch ratio, and also increase rate distortion performance at low bitrates.

The process can be summarised as follows:

1. Generate all LR pictures $I_{k,s}$ for the current frame k
2. Calculate the complexity $G_{k,s}$ of each LR picture using equation 4.2
3. Predict the QP that best matches the set target R_{target} for each picture $I_{k,s}$ using equation 4.17
4. Code $I_{k,s}$ and upsample (for $\uparrow s > 1$) to produce the reconstructed HR pictures $\hat{I}_{k,s}$
5. Select the best coded picture $\hat{I}_{k,s}$

The challenges and processes behind step 5 are described in Sections 4.3.1 and 4.3.2.

4.3.1 Best Coded Picture Selection Problem

When selecting the best IDR picture we need to maximise the level of quality obtained given the number of bits required to code the picture while also aiming to conform to the bit-budget. Another way to think about it is to maximise the gradient of the rate-distortion curve generated by coding at a range of bits – this is illustrated in Figure (4.8). For any point on the curve or any target bitrate, we wish to increase the gradient of the tangent by selecting the best IDR picture. In turn this will increase the overall compression efficiency and therefore enable greater savings in bitrate for a given level of quality.

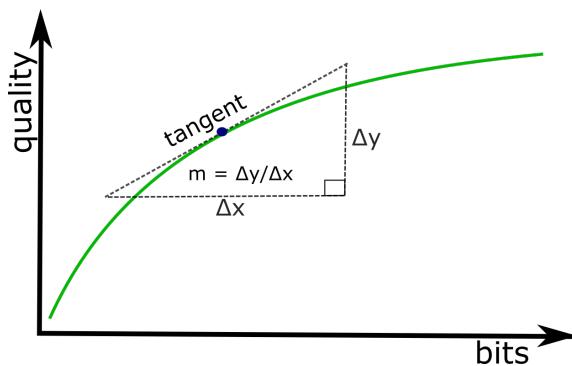


Figure 4.8: To gain the best performance we need to maximise the level of quality obtained for any given number of bits. In other words, maximise the gradient of the tangent at any point on the rate-distortion curve

This is actually quite problematic and there is no simple solution that will always provide good results. The ideal scenario would be to have a coded IDR picture that provides the highest level of quality and requires the smallest number of bits. This is illustrated in Figure (4.9a), in this case there is no question that the optimal point is p_1 . Depending on the number of lower resolution variations of the picture that are coded, it is possible that some of these can instantly be discarded given that another point is available that is unquestionably more efficient. However, in some cases a scenario may present itself as shown in Figure (4.9b). In this case there is no way to determine which point represents the most efficiently coded picture, at least not without any additional information. A relationship between the PSNR, or any other quality metric used, and the number of bits required for each picture needs to be determined. Given this knowledge, the optimal point in Figure (4.9b) can be predicted. However, this relationship would change depending on the target bitrate. We know this because we know that the gradient of the tangent changes along the rate-distortion curve. This problem can be represented using the Lagrangian Multiplier [15]:

$$(4.19) \quad J = D + \lambda R$$

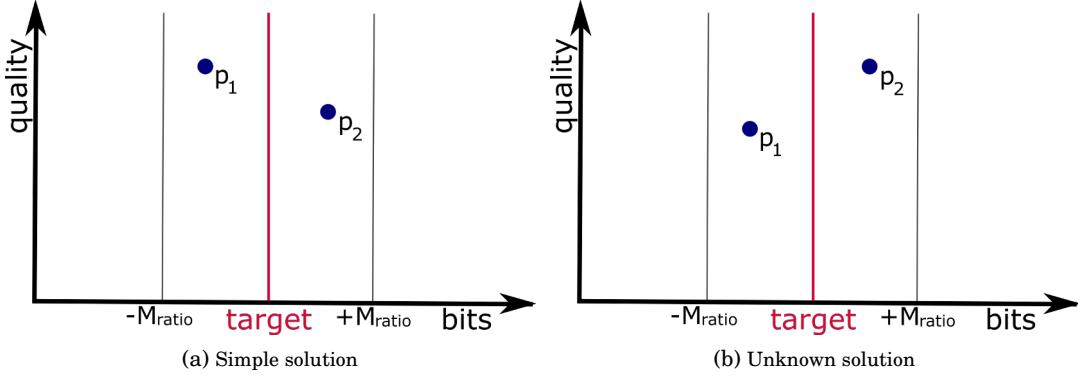


Figure 4.9: Two possible scenarios of data points available for selection. Scenario a) is simple, data point p_1 has a higher level of quality and has been coded with fewer bits compared to data point p_2 . Scenario b) is not so straight forward; without knowing or defining the relationship between quality and bits, we can not determine which point will provide the best results

where J is a cost function that we wish to minimise, D is the amount of distortion introduced, R is the number of bits required and λ describes the relationship between D and R and should be adapted.

However, without an effective means of defining λ , this approach will not necessarily provide optimal results. One solution might be to use a multi-pass rate control approach rather than a single-pass. This means that each frame would be coded using multiple levels of quantisation in order to accurately model the rate-distortion curve for each spatial resolution. Doing so would mean that modelling the relationship between quantisation and bits, as discussed in Section 4.2, could also be improved. However this would require considerably greater computational complexity and increased latency. For the purpose of analysing the problem in greater detail, here it is assumed that additional information is available despite only using the single-pass approach. Given a known optimal data point that represents a lower but similar target number of bits would at least provide some means for predicting the best point at the intended target. Figure (4.10) illustrates the scenario as previously shown in Figure (4.9b) but with a known data point. Given the known data point, the gradient between itself and each available point for the actual target can be calculated in order to predict the best picture. Using this method would mean that in Figure (4.10a) the best picture is represented by point p_1 and in Figure (4.10b) the best picture is represented by point p_2 . This is effectively how the optimal rate-distortion curves are produced throughout this thesis; starting from the data point at the lowest bitrate, the curve is plotted by connecting points that maximise the gradient.

If we assume a more comprehensive model of each rate-distortion curve for each resolution, it is possible to predict the best resolution for a given target. Figures (4.11a) and (4.11b) provide

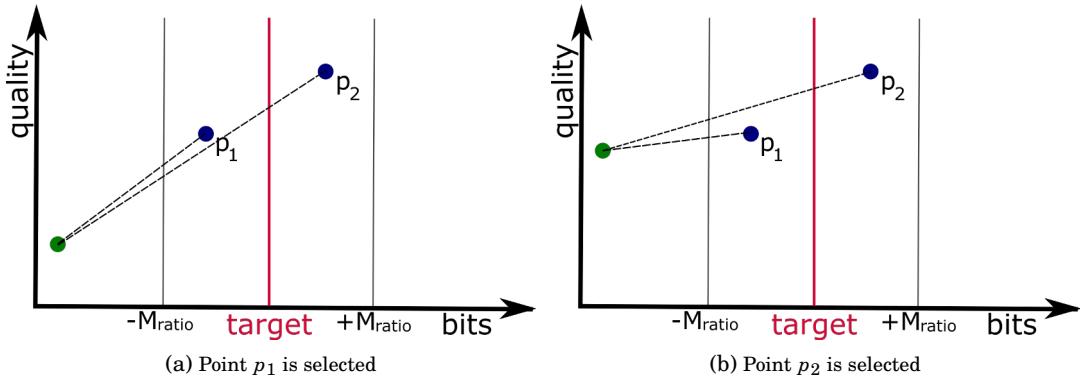


Figure 4.10: If we had prior knowledge of the position of an optimal data point on the rate-distortion curve that represents a lower target rate, the optimal point for the current target could be predicted with a higher degree of certainty. This is effectively how the optimal rate-distortion curve can be plotted given all data points for a varying number of bits

two examples of scenarios where p_3 is rejected either due to a lower level of efficiency compared to another existing point or a lower level of quality and a high mismatch ratio – the specified rules that determined whether or not a data point is eliminated will be described in more detail in Section 4.3.2. In both cases the rate-distortion models are the same but p_3 would have been coded using a different QP. This leaves two other possible points; p_1 is positioned under the curve of the resolution corresponding to p_3 , and p_2 is positioned on its curve at a point where no other rate-distortion curve provides a higher level of quality for the the same number of bits. If the optimal path is taken to define the optimal rate-distortion curve, as illustrated in Figure (4.11c), then it can be seen that only p_2 is located on this curve. However, this does not necessarily mean that p_2 is more efficient than p_1 ; the only thing we can deduce from these models is that the resolution corresponding to p_2 is more efficient than the resolution corresponding to p_1 when coded with a number of bits equal to p_2 , and vice versa. It can also be said that, for the actual target the optimal resolution is actually the resolution corresponding to p_1 as quality is higher at this point than any other resolution rate-distortion model.

In addition to coding efficiency the mismatch ratio is also highly important – this is especially true when used in an inter-coding framework. Assuming the rate control algorithm has selected the best target for the IDR picture that enables effective coding of all frames by taking into account the level of temporal correlation within the GOP, the IDR picture selected should closely match this target unless it produces a higher quality picture for fewer coded bits. In this thesis this target is often referred to as the IDR bit-budget as a picture requiring fewer bits but produces a higher level of quality would provide better results compared to one that provides a lower level of quality at the specified target.

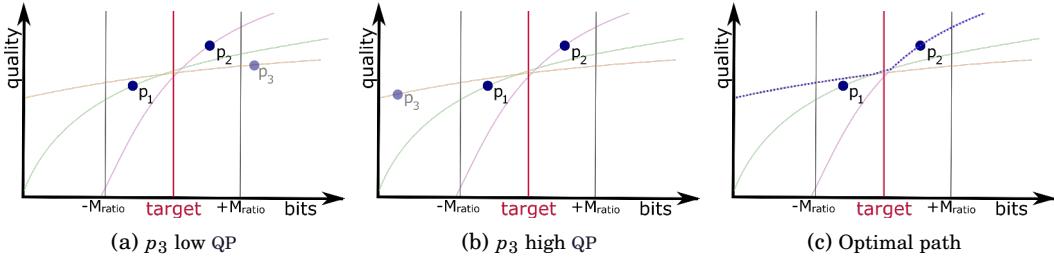


Figure 4.11: In the case of applying a multi-pass rate control approach, rate-distortion models could provide a means of deducing the point that represents the best resolution for any number of bits. a) and b) represent scenarios where p_3 is coded with a low QP and a high QP, respectively. Note that in both cases point p_3 is rejected. These models only give indication of the optimal resolution for any number of bits but do not really give any indication of the optimal data point. As shown in c), p_2 is situated on the optimal path and would therefore be the data point selected when plotting the optimal rate-distortion curve. However, this still does not imply that p_2 is more efficient than p_1 as given the case where the resolution corresponding to p_3 is not available, both p_1 and p_2 would be situated on the optimal path

4.3.2 Best Coded Picture Selection Method

As discussed in Section 4.3.1, to determine the optimal resolution at the desired target, a multi-pass rate control approach would be required in order to generate rate-distortion models. This is not desirable as it would incur a significant increase in computational complexity and latency. It is also not really beneficial; even with these models, it is not possible to determine the best coded picture in terms of its coding efficiency – just the best resolution at any number of coded bits. However, with just the data points available using the single-pass rate control approach, a number of these points may be rejected due to a clear lower level of efficiency compared to other available points. In cases where an obvious choice can not be made, given that two or more points have not been ruled out, this proposed method applies a series of conditions in order to make a decision.

As it is also important not to exceed the set target, any points beyond a number of bits equal to a predetermined mismatch ratio limit $+M_{ratio}$ are ignored. A lower bound $-M_{ratio}$ is also applied to any point that has a lower level of quality than any other point available requiring a number of bits less than $+M_{ratio}$. This is illustrated in Figure (4.12a); p_4 is ignored as it is coded beyond the upper bound $+M_{ratio}$, p_1 is ignored as p_3 is more efficient, and finally p_3 is rejected as it is below the lower bound $-M_{ratio}$ and p_2 is available with a higher level of quality and requires fewer bits than the upper bound $+M_{ratio}$. As the ideal case would be to have a point representing a picture with the highest level of quality and the lowest number of required bits, if p_3 provided a greater level of quality compared to p_2 then it would be selected instead as shown in Figure (4.12b).

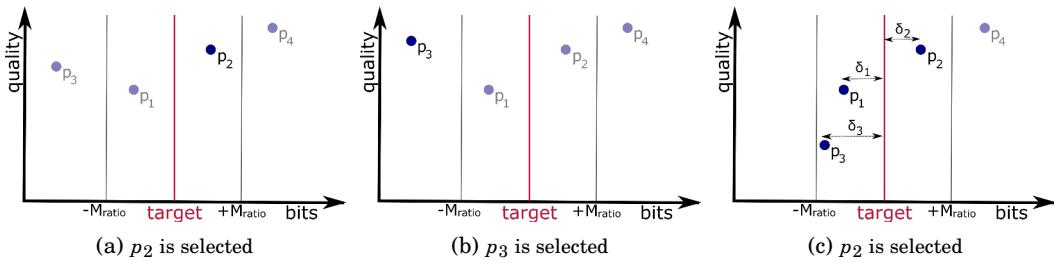


Figure 4.12: Scenarios for selecting the best picture. a) p_2 is selected – p_1 is rejected as p_3 is more efficient, p_3 is rejected as it has a lower quality level than p_2 and also requires fewer bits than the minimum mismatch ratio $-M_{ratio}$, and p_4 is rejected as it requires more bits than the maximum mismatch ratio $+M_{ratio}$. b) p_3 is selected as it is the most efficient point below $+M_{ratio}$. c) p_2 is selected – as only p_4 can be eliminated, the next criteria is to minimise the mismatch ratio to ensure a more consistent and controllable bitrate. In this scenario p_2 is coded with a number of bits closest to the target

In some cases even after the initial elimination stage, two or more data points may still remain. This problem is illustrated in Figure (4.12c) – p_4 is eliminated for the reasons mentioned above and therefore the decision falls down to either p_1 , p_2 or p_3 . Given this scenario, selecting the best picture in terms of efficiency becomes a gamble and in this case there is a $\frac{1}{3}$ probability of choosing the best one. Depending on the number and distribution of data points, and therefore the effectiveness of the initial elimination stage, this probability may decrease. Within an inter-coding configuration, we have to assume that the defined IDR bit-budget (or target) is optimal. For the scenario specified in Figure (4.12c) the most efficient picture can not be determined and therefore in such a case the proposed method selects the data point closest to the target. Using this approach will at least minimise the mismatch ratio and provide a more consistent bitrate. It also enables rate control to allocate closer to the number of intended bits for the remaining inter-coded pictures within the GOP. If a picture is not selected given the aforementioned criteria, then the mismatch ratio limits are increased and the process is repeated.

4.3.3 Computational Complexity

The Adaptive Resolution R-Q model provides a number of benefits including higher rate-distortion performance and greater quality and bitrate consistency. However, as it is required that every scaled version of the original HR input picture is coded and resampled, there is an increase in computational complexity. The additional complexity depends on the number of LR pictures coded, thus providing a common trade-off between performance and computation.

Bossen et al. [4] analyse the complexity of the HEVC standard for different encoding configurations, including *All Intra* (AI), *Random Access* (RA), and *Low Delay* (LD). The RA configuration

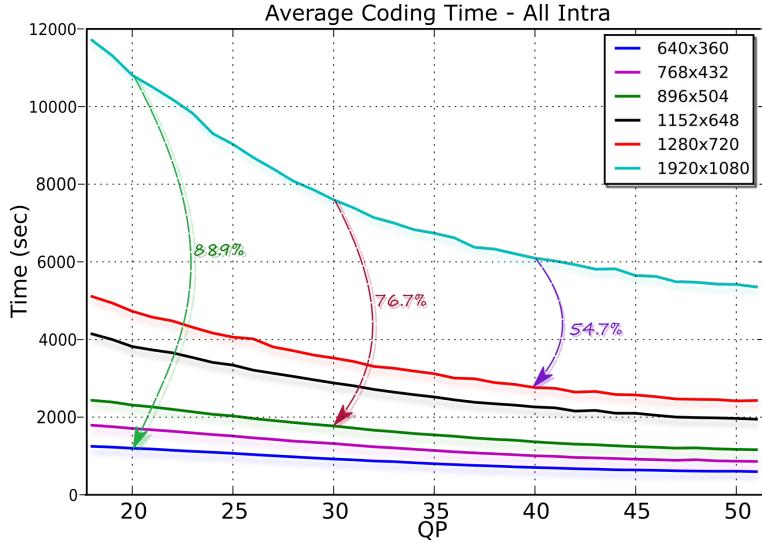


Figure 4.13: Intra-coding time – Average coding time of 6 HD sequences coded at 6 different spatial resolutions. Results show that lowering the spatial resolution of an intra-coded picture can significantly reduce computational complexity (All sequences coded using an Intel Core i7-3770S CPU @ 3.10GHz)

uses a pyramidal coding structure with an IRAP coded approximately once every 1 second. The LD configuration codes pictures in display order and only the first frame is coded as an intra-picture. Table 4.1 contains the average coding time results for 9 different HD sequences coded using fixed QP values 27 and 32 – the original data was taken from [4].

Table 4.1: Average coding time (seconds) for three different coding configurations and also the reduced computation time (%) when coding in intra mode compared to the inter configurations (using data from [4])

AI27	AI32	RA27	RA32	LD27	LD32
5217.7	4496.6	14946.6	12833.3	21260	18060
All Intra Savings:		65%	64.9%	75.5 %	75.1%

These results show that the computational complexity of an intra-coded picture is approximately 75% or 4× less than an inter-coded B picture. As we wish to maintain error resilience, in this thesis the aim is to code IDR pictures two times a second. Using standard HEVC, this means a lower level of computational complexity is required compared to the LD configuration used in [4], as there is a higher frequency of IRAP pictures coded.

Figure (4.13) shows the average coding time of 6 HD sequences for a QP range [18,51]. For comparison, each sequence was also coded at 5 lower spatial resolutions. Table 4.2 shows the reductions in computational complexity for each resolution at various points within the QP range.

Table 4.2: Average reduction of computation time when coding at lower spatial resolutions using the AI configuration

	QP20	QP30	QP40	QP50
640 × 360	88.9%	87.8%	88.5%	88.8%
768 × 432	84.2%	82.6%	83.5%	84.0%
868 × 504	78.6%	76.7%	77.6%	78.4%
1152 × 648	64.7%	62.1%	62.9%	63.8%
1280 × 720	56.3%	53.7%	54.7%	55.4%

It should also be noted that the resampling and coding process for each IDR picture $I_{k,s}$ can be performed independently and therefore can be run in parallel. From these data it can be said that coding additional IDR pictures at lower spatial resolutions will not incur a significant increase in processing power per frame, although it would still be economical to limit the scale factor and also not code pictures with similar spatial dimensions. In addition to coding complexity there is also the added computational requirements of resampling. However, the SHVC extension provides proof that this is achievable with a number of spatial layers up to 3 [28]. From the intra-coded data given in Chapter 3 it can be seen that on average a scale factor of 3, producing a low resolution of 640×360 for HD sequences, rarely provides the best results. This is due to the restrictive cut-off frequency of the low pass filters applied prior to downsampling (see Section 2.1.1). The level of filtering required to prevent aliasing when reducing the sampling rate by 3 in each direction tends to significantly remove large amounts of high frequency information and therefore greatly increases resampling distortion. Based on the analysis given in this section, and knowledge that SHVC is designed to operate with up to 3 spatial layers which includes inter-layer prediction, to ensure a more practical system in terms of computational complexity, the number of additional LR intra-coded pictures will be limited to 3 for the remaining chapters within this thesis – although results are still given for 5 additional spatial resolutions in the results section of this chapter. This is also to reduce redundancy in the form of coding at resolutions that provide similar levels of performance when the difference between each scale factor is small. The additional resolutions will therefore be limited to: 736×432 , 1280×720 and 1536×864 for HD sequences and 1536×864 , 1920×1080 and 2560×1440 for UHD sequences.

4.4 Experimental Results

As in [32, 66], evaluation of performance is partly determined by analysing the average mismatch ratio, given as:

$$(4.20) \quad \overline{M} = \frac{1}{N} \sum_i^N \frac{|R_{target} - R_{actual,i}|}{R_{target}} \times 100\%,$$

where N is the total number of frames within the sequence

Table 4.3: Comparison of intra-coded sequences using the proposed Adaptive Resolution R-Q model and the Single Resolution R-Q model, with frame rates normalised to 1Hz. The proposed method outperforms the single resolution approach in every case in terms of minimising the mismatch ratio and also produces a higher PSNR in most cases

Video	Target kb/s	Single Resolution R-Q			Adaptive Resolution R-Q		
		\overline{M} -Ratio%	Rate	Y-PSNR	\overline{M} -Ratio%	Rate	Y-PSNR
BlueSky	829.44	9.212	756.86	38.22	5.074	792.12	38.52
	414.72	10.792	371.17	33.82	5.191	396.06	34.65
	207.36	11.292	184.55	29.94	5.673	196.99	30.98
InToTree	1036.80	10.305	995.33	35.89	4.822	1032.65	35.73
	622.08	8.289	603.42	34.28	3.899	624.15	34.12
	207.36	11.788	207.36	31.45	4.722	207.36	31.47
Station	207.36	8.225	201.14	36.00	6.122	196.99	35.95
	103.68	8.588	101.61	33.74	6.191	99.53	33.76
	20.74	13.56	20.74	29.53	7.437	20.74	29.72
ParkJoy	622.08	8.113	584.76	27.41	4.890	599.27	28.88
	207.36	16.825	205.29	24.03	4.391	205.29	24.09
	103.68	24.713	107.83	22.62	4.483	105.75	22.67

A total of 6 spatial resolutions were selected based on the conditions that the original aspect ratio remains the same and that the width and height are multiples of the smallest coding unit size in HEVC – which is 8×8 . Given these criteria, the tested spatial resolutions were: 640×360 , 768×432 , 896×504 , 1152×648 , 1280×720 , 1920×1080 . All resampling was performed using Bicubic, although better methods can be applied to improve rate-distortion performance – these will be explored in Chapter 5. Details of the four test sequences are given in Appendix-A. For the updating procedure given in equation 4.11 and equation 4.13, λ was set to 0.1 so to give more weight to the value calculated from the current frame as this was shown to produce better results. For the modified procedure, γ was set to 10 so that any coded picture with a mismatch ratio less than 10% and $R_{actual} \leq R_{target}$ are considered to have a near optimal value of α . Note that this is a different use of the mismatch ratio as used when selecting the best coded picture as described in Section 4.3.2. For this process the mismatch ratio is set to 5% initially and then is incremented by 5% if necessary until a picture is selected. T in 4.15 is set as 10 and σ in 4.16 is set as 8.

Figure (4.14) shows a comparison of the proposed Adaptive Resolution R-Q model alongside the more standard Single Resolution R-Q model approach using the update procedure given in equation 4.11. Table 4.3 includes additional comparisons between these methods and it is shown

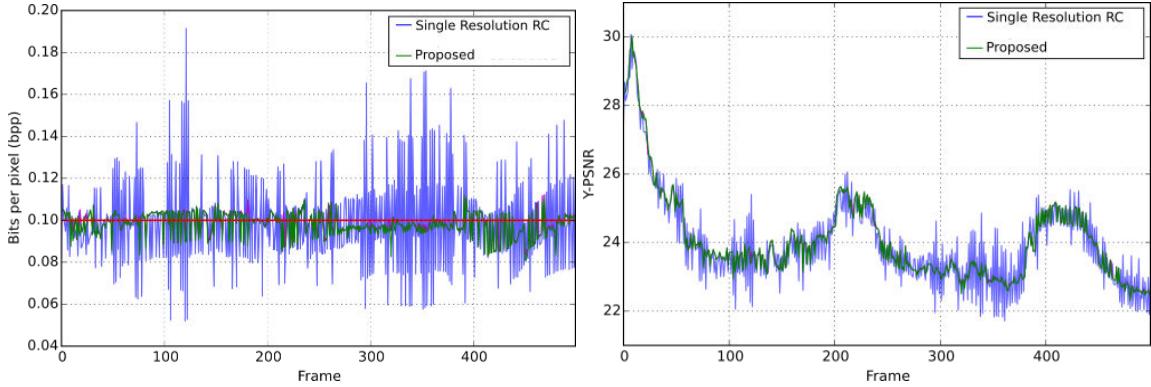


Figure 4.14: Reduced variation of rate and quality using the proposed method – *ParkJoy* coded using the Adaptive R-Q model and the Single Resolution R-Q model in intra mode. Target: 0.1 BPP / 207.36 kb/s (1Hz)

that the Adaptive Resolution R-Q model significantly outperforms the standard single resolution approach by providing a much lower average mismatch ratio and also a higher level of quality for the majority of the tested target bitrates for each sequence. It should also be noted that, due to the fluctuations in the video bitrate in the standard single resolution case, a high number of frames are likely to be skipped resulting in a much lower average PSNR than those stated in Table 4.3. This also causes variance of video quality over time, as indicated in Figure (4.14).

4.5 Chapter Summary

This chapter demonstrates that the proposed Adaptive Resolution R-Q method outperforms the single resolution approach by generating a coded bitstream with a far better regulated bitrate. Further evidence is also provided to support the previous claim, as first stated in Chapter 3, that spatial resampling of intra-coded pictures can provide a higher and more consistent level of video quality at low bitrates.

The proposed Adaptive Resolution R-Q model also benefits from a modified update procedure which makes predictions of the optimal QP for the next frame based on the complexities of previously coded pictures. The method requires that parameters are stored into memory provided that the corresponding pictures were coded efficiently by conforming to the set target. It should be noted that the modified updating procedure, as given in equation 4.12, will be less effective in an inter-coding configuration. This is because its performance is dependent on the number of *successfully* coded intra pictures. Therefore, as the intra period increases, this technique will become less effective. It also means that its memory requirements will decrease and so it can still be implemented without any adverse effects. An increase in the intra period would also reduce the performance of the standard approach as it relies heavily on correlation between consecutive

intra-coded pictures. In this case the proposed modified update procedure could potentially enable far greater prediction of the optimal QP for the next frame.

In Chapter 5 resampling techniques are analysed and applied to intra-coded pictures to reduce the level of distortion introduced for each scaled picture. Based on the intra-coded results given in Chapter 3, and evaluation of the computational complexity that the Adaptive Resolution R-Q model requires, for the remaining chapters within the thesis the number of additional LR intra-coded pictures is limited to 3 and the difference between each scale is also increased. Coding 3 additional IDR pictures should be well within acceptable limits. This is especially true considering that coding at lower spatial resolutions will also reduce the complexity of an intra-coded picture. Fewer resolutions should therefore provide a good balance between the amount of additional processing power required and the level of performance gained.

TECHNIQUES FOR SPATIAL RESAMPLING OF IDR PICTURES

Image and video rescaling is performed as part of many applications. A common example is when a DVD is viewed on a HD screen; the spatial resolution of the video is equal to 720×480 and therefore it will need to be upscaled, either by the DVD player or television, in order to display it at 1920×1080 on the HD screen. Computation is therefore a big factor when selecting the resampling technique implemented within real-time applications. However, as processing power increases, more intelligent methods can be applied enabling better performance.

In Chapter 3 it was shown that spatial resampling of IDR pictures can improve rate-distortion performance by providing a higher and more consistent level of video quality at low bitrates. The level of performance gained is partly dependent on, and also limited by, the resampling techniques used – the more accurately an IDR picture is resampled, the more efficiently the GOP can be encoded. Traditionally, resampling an image involved applying spatially invariant filters, such as Bilinear or Bicubic [35], to reconstruct an image at a desired resolution. Such techniques often produce images with noticeably reduced structural fidelity due to the removal of high frequency information. More recently, adaptive image interpolation techniques have been introduced to counteract this problem by varying the filter parameters according to frequency content [2, 20, 37]. Super Resolution SR techniques also aim to generate enhanced HR images from one or more LR sources. These can generally be categorised into either Single-Image SR (SISR) [14, 22, 40, 45, 56, 61, 68, 76] or Multi-Image SR (MISR) [1, 9, 10, 16, 29, 30]. The latter, however, would be counter-productive if used to enhance IDR pictures as it requires the use of pixels from adjacent inter-coded frames which may have been subjected to errors.



Figure 5.1: Test image *Lena* - a) *full* image scaled to fit on the page b) cropped section printed at native resolution

Low sampling rates will reduce the number of bits spent on coding IDR pictures. However, aliasing and loss of detail will affect prediction performance and the quality of subsequent inter-coded pictures. The best choice of filter and sampling rate will produce a LR picture that maintains enough detail, lacks any strong artefacts due to aliasing and leads to high compression efficiency when applying intra-coding. An alternative solution to the downsampling problem would be to capture IDR pictures at multiple resolutions. In [44] Lu et al. demonstrates the advantages of using a multi-resolution camera over the use of resampling techniques.

5.1 Spatially Invariant Filtering

Spatially invariant filtering techniques provide a low computational means for downsampling and upsampling images. In this section, some of the most common techniques are described and examples are given to demonstrate performance of interpolation. An example of interpolation as used in the SHVC extension is also given for comparison.

For these tests the *Lena* image is used and a cropped section is provided to clearly show the effect of each technique, see Figure (5.1). All sampling processes in this thesis are applied in the YCrCb colour space. Therefore the RGB *Lena* image is first converted using the formulas in Section 2.2.3 as provided in Chapter 2. In this section, only interpolation performance is compared and so the same downsampled image is used as the input for each technique. Figure (5.2) shows the luminance component of the *Lena* image and the corresponding downsampled version used as the input.

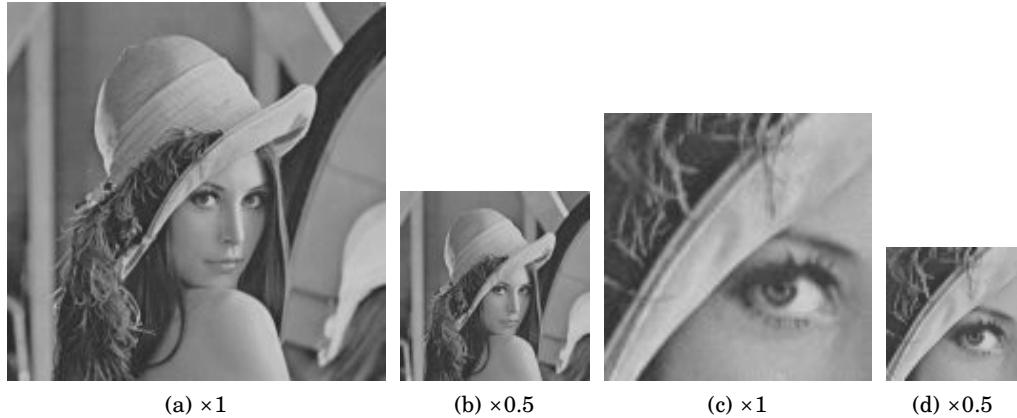


Figure 5.2: Test image *Lena* converted to the YCrCb colour space in order to extract the luminance channel. a) Full luminance image scaled to fit on the page b) half size luminance image downsampled using Bicubic c) cropped section of luminance image printed at native resolution d) half size cropped section downsampled using Bicubic

5.1.1 Nearest Neighbour

The Nearest-Neighbour (NN) algorithm is the simplest of all sampling techniques; the sample to be interpolated takes the value of the nearest input sample. In the spatial domain this forms a box filter and in the frequency domain it forms a sinc filter which has infinite support – this is the opposite of an ideal filter [67]. Figure (5.3) shows that, as a result of this method, edges appear *jagged* and unnatural. The NN technique can be given as:

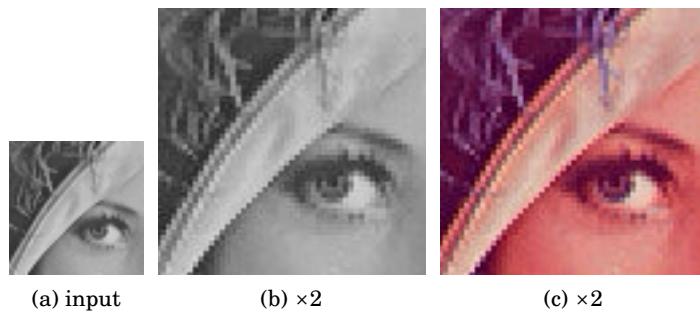


Figure 5.3: Nearest Neighbour interpolation - the simplest of all interpolation methods; samples are interpolated by copying the nearest corresponding input sample in the LR sample space. Edges appear *jagged* as there is no smooth transition between pixel intensities

$$(5.1) \quad I_h(x, y) = I_l(\text{round}(\frac{x}{s}), \text{round}(\frac{y}{s})),$$

where I_h is the HR interpolated image, I_l is the LR input image, x and y are the coordinates within the 2-dimensional HR sample space, and s is the scale factor. Note that *round* is a simple function that takes the value of the nearest integer.

5.1.2 Bilinear

Bilinear interpolation applies the same principles as linear interpolation but instead is performed in two dimensions, one at a time. The pixel to be interpolated is a calculated average of the four neighbouring pixels. Figure (5.4) shows the result of this technique. It can be seen that the edges appear much smoother than the NN result.

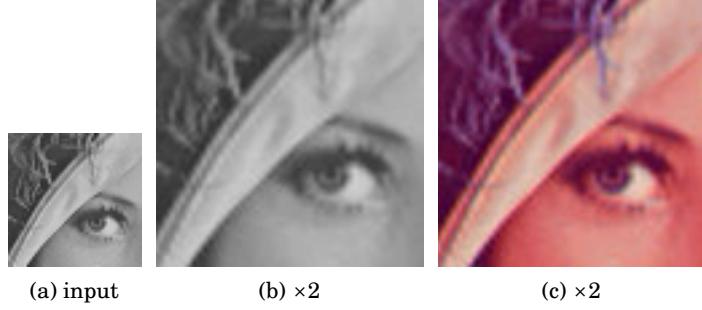


Figure 5.4: Bilinear interpolation - samples are interpolated as a weighted average of the four nearest input samples (two in each dimension). Low frequency areas appear more natural compared to the NN result but edges and textures can appear overly smooth

Let $p_{(1,1)}$, $p_{(2,1)}$, $p_{(1,2)}$ and $p_{(2,2)}$ be the four nearest LR samples (two in each dimension) relative to the position of the interpolated sample in the LR sample space, with corresponding coordinates (x_1, y_1) , (x_2, y_1) , (x_1, y_2) and (x_2, y_2) . Then Bilinear interpolation can be formulated as:

$$(5.2) \quad I_h(x, y) = p_{(1,1)} \frac{(x_2 - \frac{x}{s})(y_2 - \frac{y}{s})}{(x_2 - x_1)(y_2 - y_1)} + \\ p_{(2,1)} \frac{(\frac{x}{s} - x_1)(y_2 - \frac{y}{s})}{(x_2 - x_1)(y_2 - y_1)} + \\ p_{(1,2)} \frac{(x_2 - \frac{x}{s})(\frac{y}{s} - y_1)}{(x_2 - x_1)(y_2 - y_1)} + \\ p_{(2,2)} \frac{(\frac{x}{s} - x_1)(\frac{y}{s} - y_1)}{(x_2 - x_1)(y_2 - y_1)}.$$

5.1.3 Bicubic

Bicubic interpolation, like bilinear interpolation, calculates a weighted average from neighbouring pixels but is based on multidimensional cubic interpolation [35]. Bicubic interpolation requires additional processing power, compared to the previously mentioned techniques, due to the fact it takes into account the nearest sixteen pixels rather than just one or four. The end result is a smooth image that preserves high frequencies far better than Bilinear interpolation. The result can be seen in Figure (5.5).

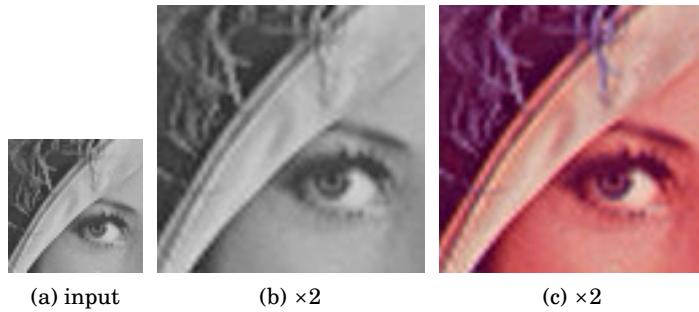


Figure 5.5: Bicubic interpolation - samples are predicted using multidimensional cubic interpolation. Bicubic enables better prediction compared to NN and Bilinear as it takes in account 16 of the nearest input samples in order to better reconstruct the original 2-dimensional signal. Edges appear sharper than the Bilinear result without highly noticeable artefacts

Consider first interpolation in one dimension using Cubic interpolation. Let $p_{(1)}, p_{(2)}, p_{(3)}$ and $p_{(4)}$ be values of pixels at positions x_1, x_2, x_3 and x_4 , respectively. A value at position x within the interval $[x_2, x_3]$ can be interpolated by:

$$(5.3) \quad I_h(x) = \left(-\frac{1}{2}p_{(1)} + \frac{3}{2}p_{(2)} - \frac{3}{2}p_{(3)} + \frac{1}{2}p_{(4)} \right)x^3 + \left(p_{(1)} - \frac{5}{2}p_{(2)} + 2p_{(3)} - \frac{1}{2}p_{(4)} \right)x_2 + \left(-\frac{1}{2}p_{(1)} + \frac{1}{2}p_{(3)} \right)x + p_{(2)}.$$

Note that x is calculated as a sub-pixel position such that $0 \leq x < 1$, if the phase is equal to zero then $I_h(x)$ will take the value of $p_{(2)}$. Cubic interpolation is applied in one direction (horizontal or vertical) and then applied again on the result in the other direction to produce a Bicubic interpolated result.

5.1.4 SHVC Sampling

Chapter 2 contains details of the SHVC scalable extension for the HEVC standard, including details of the filters used for the resampling operations. Table 5.1 and Table 5.2 contain the filter

Table 5.1: SHVC downsampling coefficients - low range

Phase i	Downsampling filter coefficients for ratios in the range $[\frac{5}{4}, \frac{5}{3}]$											
	F[i,0]	F[i,1]	F[i,2]	F[i,3]	F[i,4]	F[i,5]	F[i,6]	F[i,7]	F[i,8]	F[i,9]	F[i,10]	F[i,11]
0	0	5	-6	-10	37	76	37	-10	-6	5	0	0
1	0	5	-4	-11	33	76	40	-9	-7	5	0	0
2	-1	5	-3	-12	29	75	45	-7	-8	5	0	0
3	-1	4	-2	-13	25	75	48	-5	-9	5	1	0
4	-1	4	-1	-13	22	73	52	-3	-10	4	1	0
5	-1	4	0	-13	18	72	55	-1	-11	4	2	-1
6	-1	4	1	-13	14	70	59	2	-12	3	2	-1
7	-1	3	1	-13	11	68	62	5	-12	3	2	-1
8	-1	3	2	-13	8	65	65	8	-13	2	3	-1
9	-1	2	3	-12	5	62	68	11	-13	1	3	-1
10	-1	2	3	-12	2	59	70	14	-13	1	4	-1
11	-1	2	4	-11	-1	55	72	18	-13	0	4	-1
12	0	1	4	-10	-3	52	73	22	-13	-1	4	-1
13	0	1	5	-9	-5	48	75	25	-13	-2	4	-1
14	0	0	5	-8	-7	45	75	29	-12	-3	5	-1
15	0	0	5	-7	-9	40	76	33	-11	-4	5	0

Table 5.2: SHVC downsampling coefficients - high range

Phase i	Downsampling filter coefficients for ratios in the range $[\frac{5}{2}, 2]$											
	F[i,0]	F[i,1]	F[i,2]	F[i,3]	F[i,4]	F[i,5]	F[i,6]	F[i,7]	F[i,8]	F[i,9]	F[i,10]	F[i,11]
0	2	-3	-9	6	39	58	39	6	-9	-3	2	0
1	2	-3	-9	4	38	58	43	7	-9	-4	1	0
2	2	-2	-9	2	35	58	44	9	-8	-4	1	0
3	1	-2	-9	1	34	58	46	11	-8	-5	1	0
4	1	-1	-8	-1	31	57	47	13	-7	-5	1	0
5	1	-1	-8	-2	29	56	49	15	-7	-6	1	1
6	1	0	-8	-3	26	55	51	17	-7	-6	1	1
7	1	0	-7	-4	24	54	52	19	-6	-7	1	1
8	1	0	-7	-5	22	53	53	22	-5	-7	0	1
9	1	1	-7	-6	19	52	54	24	-4	-7	0	1
10	1	1	-6	-7	17	51	55	26	-3	-8	0	1
11	1	1	-6	-7	15	49	56	29	-2	-8	-1	1
12	0	1	-5	-7	13	47	57	31	-1	-8	-1	1
13	0	1	-5	-8	11	46	58	34	1	-9	-2	1
14	0	1	-4	-8	9	44	58	35	2	-9	-2	2
15	0	1	-4	-9	7	43	58	38	4	-9	-3	2

Table 5.3: SHVC Luma coefficients for interpolation

Phase i	Luma coefficients							
	F[i,0]	F[i,1]	F[i,2]	F[i,3]	F[i,4]	F[i,5]	F[i,6]	F[i,7]
0	0	0	0	64	0	0	0	0
1	0	1	-3	63	4	-2	1	0
2	-1	2	-5	62	8	-3	1	0
3	-1	3	-8	60	13	-4	1	0
4	-1	4	-10	58	17	-5	1	0
5	-1	4	-11	52	26	-8	3	-1
6	-1	3	-9	47	31	-10	4	-1
7	-1	4	-11	45	34	-10	4	-1
8	-1	4	-11	40	40	-11	4	-1
9	-1	4	-10	34	45	-11	4	-1
10	-1	4	-10	31	47	-9	3	-1
11	-1	3	-8	26	52	-11	4	-1
12	0	1	-5	17	58	-10	4	-1
13	0	1	-4	13	60	-8	3	-1
14	0	1	-3	8	62	-5	2	-1
15	0	1	-2	4	63	-3	1	0

Table 5.4: SHVC Chroma coefficients for interpolation

Phase i	Chroma coefficients			
	F[i,0]	F[i,1]	F[i,2]	F[i,3]
0	0	64	0	0
1	-2	62	4	0
2	-2	58	10	-2
3	-4	56	14	-2
4	-4	54	16	-2
5	-6	52	20	-2
6	-6	46	28	-4
7	-4	42	30	-4
8	-4	36	36	-4
9	-4	30	42	-4
10	-4	28	46	-6
11	-2	20	52	-6
12	-2	16	54	-4
13	-2	14	56	-4
14	-2	10	58	-2
15	0	4	62	-2

coefficients used for downsampling, and Table 5.3 and Table 5.4 contain the filter coefficients used for upsampling. As mentioned before, SHVC uses a 12-tap polyphase filter for downsampling and an 8-tap polyphase filter for upsampling. Details of how a polyphase filter is constructed is given in [51]. It should be noted that a polyphase filter does not actually produce a different result but enables much more efficient implementations.

To some degree SHVC does not adjust filter coefficients according to the scale factor. As described in [8], when performing downsampling there are two sets of coefficients for which each support a range of scale factors. Each set contains 16 subsets for 16 degrees of phase, therefore the interpolated sample position within the LR sample space is rounded to the nearest $\frac{1}{16}^{th}$ sample position. Within the source code of the SHVC reference software <https://hevc.hhi.fraunhofer.de/shvc> additional sets of coefficients can be found that support a wider range of scale factors. However, two similar scale factors may apply the same set of filter coefficients.

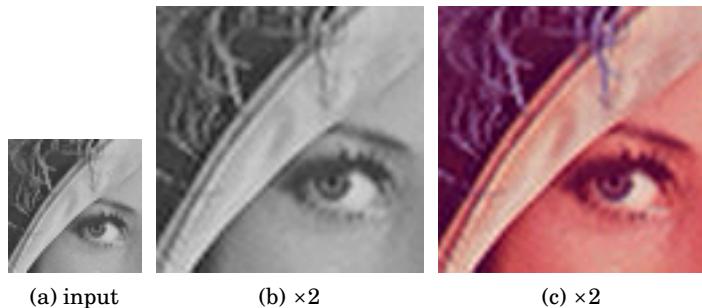


Figure 5.6: Interpolation using SHVC sampling functions - samples are interpolated using the filters coefficients as used in the SHVC scalable coding extension. The predicted sample is calculated from a weighting of up to 64 of the nearest pixels. The interpolation function is capable of producing sharper edges compared to Bicubic but may also introduce *ringing* artefacts in some areas

The interpolation function within SHVC only uses one set of filters for the luminance channel and another set for the chrominance channels, regardless of the scale factor. Similar to downsampling, filters for 16 degrees of phase are specified. The result of the interpolation function can be seen in Figure (5.6). SHVC interpolation produces sharper edges but may also produce more noticeable artefacts such as *ringing* compared to Bicubic.

5.2 Comparison of Sampling Techniques

Both Bicubic and the SHVC sampling functions produce promising results. In this section, objective quality measures are used to evaluate which of these techniques performs best for a wide range of image content. Still images are taken from the HD (1920×1080) BVI Texture Database [52] and resampled using each technique by applying three different scale factors. Thumbnails of each image are provided in Figure (5.7).

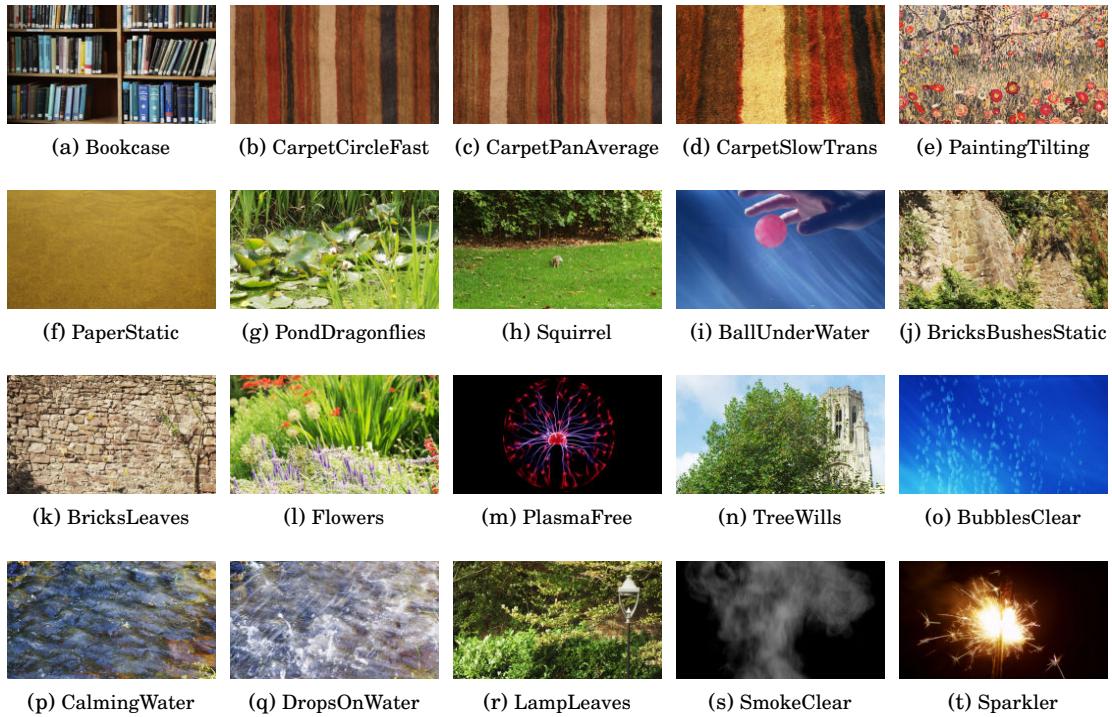


Figure 5.7: Thumbnails of still images taken from the BVI Texture Database [52] and used as test images for the comparison of sampling techniques

5.2.1 Objective Quality Analysis

Each image is first downsampled to three different lower resolutions: 1536×864 , 1280×720 and 768×432 . With corresponding (up) scale factors: 1.25, 1.5 and 2.5. Note that these are the scale factors defined in Section 4.3.3 to be provide a good balance between performance and computational complexity. For these tests, the combination of downsampling and upsampling distortion is measured and therefore three LR variations of each image are produced by each sampling technique. The LR images are then upsampled back to the HR and compared with the original using PSNR and SSIM objective quality metrics (see Section 2.8). As there are 120 resampled images in total, results are split into three groups according to those in the associated

Table 5.5: Objective quality comparison between Bicubic and SHVC sampling for test images within *Group A*. The majority of images are better reconstructed using the SHVC sampling functions

Image	Scale	Bicubic		SHVC	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
Bookcase	1.25	46.582	0.9954	47.046	0.9954
	1.5	43.256	0.9921	44.491	0.9929
	2.5	36.131	0.9671	35.906	0.9646
CarpetCircleFast	1.25	52.630	0.9954	52.439	0.9952
	1.5	51.265	0.9937	51.620	0.9942
	2.5	46.766	0.9835	46.794	0.9836
CarpetPanAverage	1.25	52.036	0.9966	51.993	0.9966
	1.5	50.122	0.9949	51.078	0.9958
	2.5	42.137	0.9721	42.588	0.9749
CarpetSlowTrans	1.25	49.623	0.9945	49.866	0.9947
	1.5	47.051	0.9905	48.016	0.9922
	2.5	40.578	0.9607	40.377	0.9593
PaintingTilting	1.25	46.363	0.9984	46.458	0.9984
	1.5	42.692	0.9967	44.815	0.9977
	2.5	31.359	0.9577	31.834	0.9608
PaperStatic	1.25	49.553	0.9921	49.783	0.9926
	1.5	46.945	0.9856	47.933	0.9886
	2.5	40.875	0.940	40.721	0.9389
PondDragonflies	1.25	45.856	0.9954	46.498	0.9957
	1.5	42.440	0.9912	43.978	0.9931
	2.5	34.887	0.9527	34.779	0.9507
Squirrel	1.25	42.648	0.9905	43.501	0.9920
	1.5	38.971	0.9783	40.632	0.9845
	2.5	31.385	0.8753	31.257	0.8709

paper [52]. However, it should be noted that the grouping procedure used in this paper takes into account both spatial and temporal information within the whole sequence.

Tables 5.5-5.7 contain measured quality results for each sampling technique and for each scale. From these results it can be seen that the majority of resampled images are reconstructed best using the SHVC sampling functions. In many cases Bicubic still performs better but usually only by a small margin. This is most commonly seen for results of images upsampled by $\times 2.5$.

Results given in Table 5.8 confirm that on average Bicubic resampling reconstructs the test images better than SHVC resampling for the larger scale factor of $\times 2.5$. However, according to both PSNR and SSIM, performance is only marginally better. When scaling by $\times 1.5$ or $\times 1.25$ greater performance gains can be obtained by using the SHVC resampling functions across a wider range of images, especially for a scale factor of $\times 1.5$. A possible reason for this could be due to the fact that Bicubic calculates filter coefficients based on scale and the phase of the sampled pixel

5.2. COMPARISON OF SAMPLING TECHNIQUES

Table 5.6: Objective quality comparison between Bicubic and SHVC sampling for test images within *Group B*. Most of the scaled images are better reconstructed using the SHVC sampling functions, however when scaled by $\times 2.5$ Bicubic produces better results

Image	Scale	Bicubic		SHVC	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
BallUnderWater	1.25	52.285	0.9944	52.094	0.9941
	1.5	50.708	0.9924	51.034	0.9926
	2.5	46.828	0.9846	46.654	0.9839
BricksBushesStatic	1.25	39.949	0.9924	40.844	0.9935
	1.5	36.187	0.9821	37.952	0.9873
	2.5	28.401	0.8866	28.295	0.8827
BricksLeaves	1.25	39.813	0.9920	40.589	0.9931
	1.5	36.143	0.9816	37.723	0.9865
	2.5	28.646	0.8900	28.508	0.8854
Flowers	1.25	49.836	0.9970	49.718	0.9970
	1.5	46.469	0.9957	47.769	0.9962
	2.5	39.070	0.9833	38.988	0.9831
PlasmaFree	1.25	55.605	0.9988	55.448	0.9988
	1.5	53.514	0.9982	54.068	0.9984
	2.5	46.919	0.9952	46.874	0.9950
TreeWills	1.25	39.602	0.9920	40.469	0.9929
	1.5	35.901	0.9822	37.523	0.9868
	2.5	28.512	0.9001	28.374	0.8961

Table 5.7: Objective quality comparison between Bicubic and SHVC sampling for test images within *Group C*. Bicubic consistently performs better for a scale factor of $\times 2.5$ by a small margin and the SHVC sampling functions provide much greater performance gains when scaling by $\times 1.5$

Image	Scale	Bicubic		SHVC	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
BubblesClear	1.25	49.594	0.9944	49.381	0.9939
	1.5	47.589	0.9921	47.841	0.9921
	2.5	43.526	0.9828	43.349	0.9819
CalmingWater	1.25	49.731	0.9955	50.113	0.9958
	1.5	46.958	0.9920	48.254	0.9938
	2.5	39.683	0.9614	39.636	0.9611
DropsOnWater	1.25	50.311	0.9962	50.660	0.9964
	1.5	47.673	0.9935	48.955	0.9949
	2.5	40.390	0.9683	40.319	0.9682
LampLeaves	1.25	43.726	0.9958	44.561	0.9962
	1.5	40.094	0.9911	41.891	0.9935
	2.5	32.044	0.9412	31.977	0.9398
SmokeClear	1.25	51.399	0.9931	51.284	0.9930
	1.5	49.393	0.9890	49.782	0.9900
	2.5	45.574	0.9735	45.396	0.9724
Sparkler	1.25	48.593	0.9952	48.582	0.9948
	1.5	46.108	0.9930	46.780	0.9934
	2.5	40.774	0.9822	40.663	0.9816

Table 5.8: Average objective quality comparison between Bicubic and SHVC sampling. Overall the SHVC sampling functions provide better performance, although Bicubic produces slightly better results when scaling by $\times 2.5$

Group	Scale	Bicubic		SHVC	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM
A	1.25	48.161	0.9948	48.448	0.9951
	1.5	45.343	0.9904	46.570	0.9924
	2.5	38.015	0.9512	38.032	0.9505
B	1.25	46.099	0.9944	46.527	0.9949
	1.5	43.153	0.9887	44.345	0.9913
	2.5	36.396	0.9400	36.282	0.9377
C	1.25	48.892	0.9950	49.097	0.9950
	1.5	46.303	0.9918	47.250	0.9930
	2.5	40.332	0.9682	40.223	0.9675
All	1.25	47.717	0.9948	48.024	0.9950
	1.5	44.933	0.9903	46.055	0.9922
	2.5	38.248	0.9531	38.179	0.9519

relative to the source samples. As discussed in Section 5.1, the SHVC downsampling function uses predetermined sets of filter coefficients for different ranges of scale factors and quantises the phase to the nearest $\frac{1}{16}^{th}$ sample position. For interpolation, only one set of filter coefficients is applied to all scales. It may be the case that the interpolation operation is better optimised for a scale factor of $\times 1.5$ and not larger factors such as $\times 2.5$.

For many images, resampling using the SHVC functions with an upscaling factor of $\times 1.5$ produces far better reconstructions at the original HR compared to Bicubic. PSNR improvements of greater than 1dB are common and on average SHVC resampling provides 1.122dB improvement over Bicubic. On average, scaling by $\times 1.25$ is also better achieved using the SHVC resampling functions but by a much smaller margin compared to scaling by $\times 1.5$.

Although the results are not entirely consistent, it can be seen that SHVC resampling provides a better level of objective quality on average. It is therefore expected that a reduced level of resampling distortion can be achieved within the proposed Adaptive Resolution coding technique by applying the SHVC resampling functions. These functions will be used as the underlying sampling technique within the following chapters.

5.3 Spatially Variant Filtering

Spatial invariant filtering techniques filter the input image without any adaptation of the filter coefficients. When interpolating samples in areas of high frequencies, a low cut-off frequency may produce overly smooth transitions across edges and textures. Spatially variant filtering

techniques aim to rectify this problem by adapting filters given the input image and possibly additional information regarding image content.

5.3.1 Joint Bilateral Upsampling

The Bilateral filter adapts filter coefficients according to pixel intensities as well as the Euclidean distance from the source input samples. It is composed as a combination of a range filter and a spatial filter. The filter is often used for noise reduction within areas of low frequency. When filtering across edges boundaries, less weight is given to source pixels with a greater difference of intensity, thus preserving high frequency content. Figure (5.8) illustrates how the combination of filters can affect the weighting of input samples.

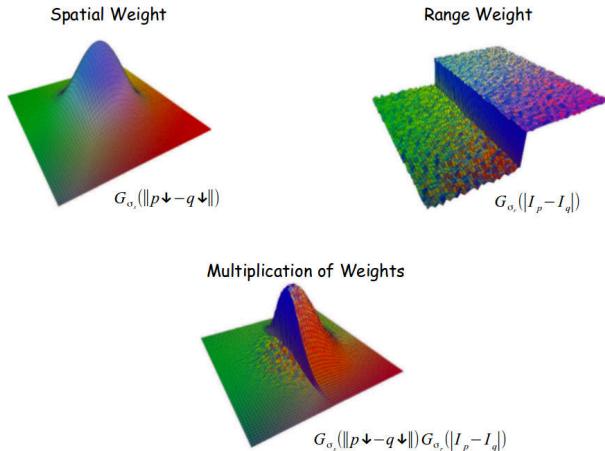


Figure 5.8: JBU multiplication of the spatial and range filters to produce a spatial variant filter that adapts according to content

On its own, Bilateral filtering can not be used for interpolation as the intensities of the samples being interpolated are unknown. However, with additional information it is possible to apply the Bilateral filter for this purpose. In [37] the Joint Bilateral Upsampling (JBU) method was proposed for the purpose of preserving the fidelity of edges when applying interpolation. The technique involves using a guide image, at the desired spatial resolution, in order to define the location of edges and provide sample intensities for the range filter. Therefore the spatial filter is applied to the LR image while a range filter is applied to the full resolution guide image. The upsampled image can be obtained by:

$$(5.4) \quad JBU[\tilde{S}]_p = \frac{1}{W_p} \sum_{q \in \Omega} S_{q \downarrow} G_{\sigma_s}(\|p \downarrow - q \downarrow\|) G_{\sigma_r}(|\tilde{I}_p - \tilde{I}_q|),$$

where \tilde{I} is the guide image, p and q denote coordinates of pixels in \tilde{I} , G_{σ_s} is the spatial filter kernel centred over $p \downarrow$, G_{σ_r} is the range filter centred over p , Ω is the spatial support of the

kernel G_{σ_s} and W_p is a normalising factor calculated as the sum of the $(G_{\sigma_s} \cdot G_{\sigma_r})$ filter weights. Note that p and q coordinates with a \downarrow correspond to the LR sample space.

A major drawback for this technique is its reliance on a guide image at the desired resolution. For most applications a HR image is not available. This technique was proposed with the intention that it can be used to reduce computational complexity of image enhancement techniques by applying them to a downsampled version of the image and then upsampled using the JBU technique with the addition of the original HR image as the guide for the range filter. Such a technique is not suitable for resampling of IDR pictures as the original HR image is not available at the decoder.

In [38] a method for creating a HR guide image from a LR solution was proposed using vector graphics. In vector form, an image can be stretched without any loss of image quality. Representing an image in vector form can be achieved via bitmap tracing, where the image contours are located and their vectors calculated. Vector information will not change regardless of the image dimensions.



Figure 5.9: Vector graphic generated using *Potrace* <http://potrace.sourceforge.net/> to perform bitmap tracing, and *Cairo* <https://cairographics.org/> to combine each intensity layer. Bitmap tracing fails to capture textured areas accurately but performs well for edges and low frequency content

After scaling the vector graphic to the desired spatial resolution it should be converted back into a bitmap image. Note that only a greyscale guide image is necessary as the formula 5.4 only refers to the guide image for pixel intensities. Figure (5.9) shows the result of a bitmap→vector→bitmap conversion. Detail maintained after the initial conversion to vector graphics will not be lost after scaling. The result in Figure (5.9) will provide a reasonable guide image for the JBU technique. However, the conversion to vector graphics does not always produce such desirable results. This can be seen within the area containing the fur hat; bitmap tracing

fails to reproduce high frequency textured areas. It should also be noted that high quality conversion to vector graphics is highly computational and therefore not suitable within a video codec given the current limitations on processing power.

5.4 Super Resolution & Enhanced Sampling Techniques

SR techniques aim to increase the spatial resolution of an image to a higher level of fidelity compared to standard interpolation approaches. The proposed Adaptive Resolution coding technique for resampling IDR pictures only uses information from the decoded LR picture in order to reconstruct the HR picture. Although MISR techniques may produce better results under ideal channel conditions, applying SISR will ensure a high level of error resilience is maintained by only considering the independently coded IDR picture.

5.4.1 Iterative Back Projection

Although a relatively old technique, Iterative Back Projection (IBP) is both effective and computationally simple. The technique was first proposed in 1991 by Irani et al. [29] and has been applied extensively ever since – many SR techniques use the IBP technique in some form [14, 56, 68]. It is essentially an optimisation process that maximises the information within the resampled HR image given the information contained within the LR image and the performance limitations of the filters used.

Given a HR image I_h , a LR image I_l can be formulated by:

$$(5.5) \quad I_l = (h_d * I_h) \downarrow s,$$

where h_d is the downsampling kernel, $*$ is the convolution operator and $\downarrow s$ indicates the down-sampling operation by a scale factor s .

The challenge of any upsampling technique is to reconstruct the HR image I_h as accurately as possible, or to a high perceptual level of quality, given knowledge of the LR image I_l . It can be seen that, if the reconstructed image \hat{I}_h is identical to the original HR image I_h , then downsampling \hat{I}_h in the same manner as equation 5.5 will produce I_l . In practise the filters used for resampling are not perfect and so some resampling error will exist.

$$(5.6) \quad (h_d * I_h) \downarrow s = (h_d * \hat{I}_h) \downarrow s + e_l,$$

where e_l indicates the difference of the two images at the low resolution.

This formula can be simplified and rearranged to make e_l the subject by:

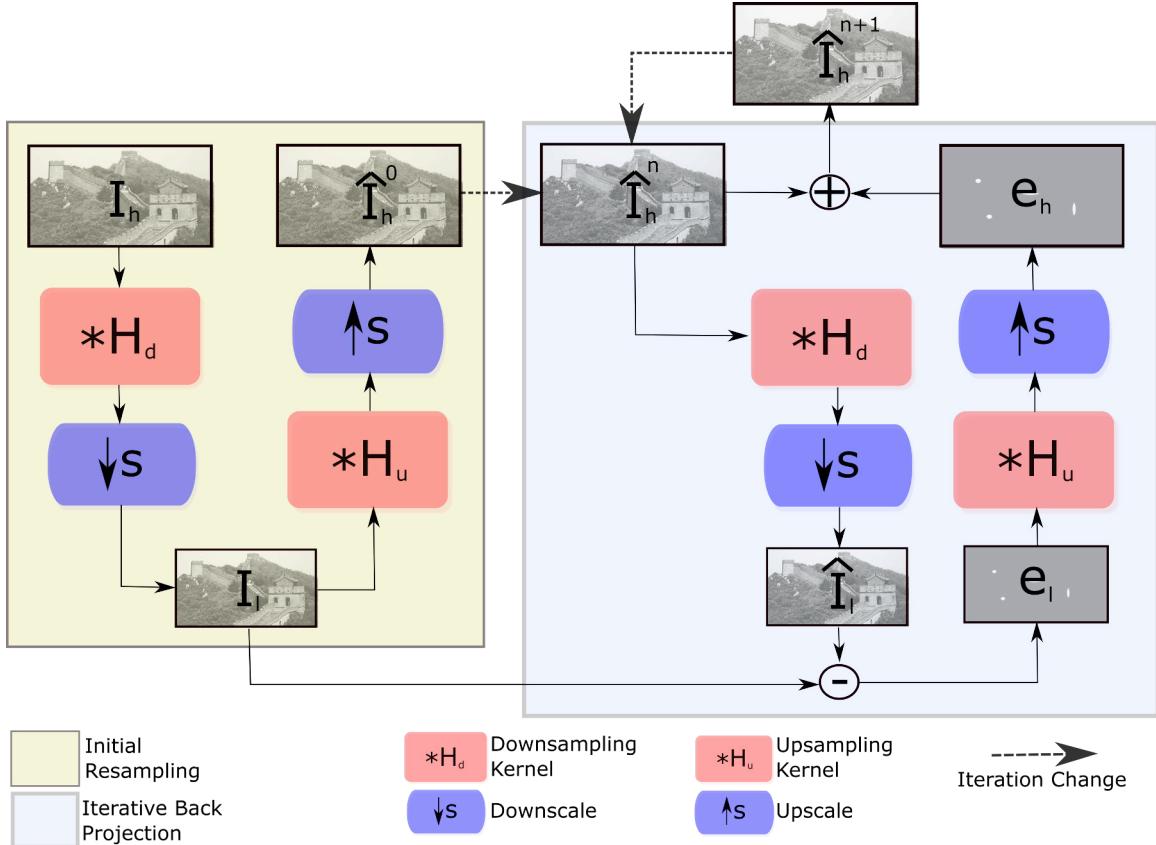


Figure 5.10: Block diagram of the IBP upsampling technique. By applying the same downsampling operation at each point, the error signal will reduce as the reconstructed result \hat{I}_h^n approaches the original I_h

$$(5.7) \quad e_l = I_l - (h_d * \hat{I}_h) \downarrow s.$$

IBP minimises the error in equation 5.7 by iteratively adding more of the information contained within the LR image I_l to the reconstructed HR image \hat{I}_h . As a result, the difference between I_h and \hat{I}_h is reduced. The full IBP formula can be given as:

$$(5.8) \quad \hat{I}_h^{n+1} = \hat{I}_h^n + (h_u * (I_l - (h_d * \hat{I}_h^n) \downarrow s) \uparrow s),$$

where n denotes the current iteration, h_u is the upsampling kernel and $\uparrow s$ is the upsampling operation by a scale factor s . It should be noted that the s in $\downarrow s$ and $\uparrow s$ do not represent the same numerical value but are in fact reciprocals of one another.

By combining equations 5.7 and 5.8, it can be seen that, in order enhance the reconstructed HR image \hat{I}_h , the error e_l is upsampled and applied to the current iteration \hat{I}_h^n .

5.4. SUPER RESOLUTION & ENHANCED SAMPLING TECHNIQUES

$$(5.9) \quad \hat{I}_h^{n+1} = \hat{I}_h^n + (h_u * e_l^n) \uparrow s.$$

Figure (5.10) shows a block diagram illustrating how the IBP technique works. The initial resampling procedure is contained within the yellow box on the left and the actual technique is contained within the blue box on the right. The better the initial resampling process performs, the fewer iterations it should take for the technique to converge. It should be noted that IBP is an upsampling technique and can therefore be applied without initially downsampling.

To test the performance of the IBP technique, it is applied to the BVI test images as used before in Section 5.2. Figure (5.11) shows the PSNR gains after each iteration of the IBP technique for the 20 HD (1920×1080) images which are resampled using the SHVC filter coefficients as specified in [8] and also provided in Section 5.1.4. Two low resolutions are tested: 1280×720 and 768×432 with corresponding upscaling factors 1.5 and 2.5. The IBP algorithm is simple but effective – in many cases PSNR gains between 1 – 2dB are possible compared to the initial downsampled and upsampled result. However, it should be noted that the PSNR will never fully converge; IBP minimises the error e_l but the addition of the upsampled error $(h_u * e_l) \uparrow s$ will not necessarily produce a more accurate reconstruction of the image I_h compared to the previous iteration due to the fact that the upsampling procedure h_u is not ideal.

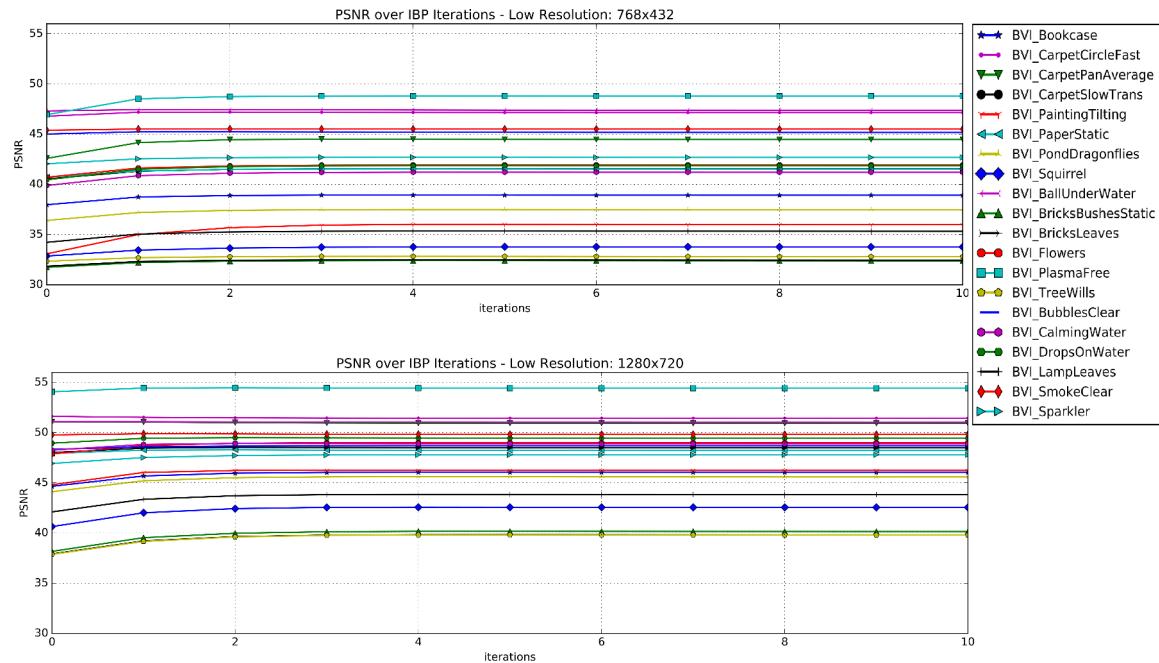


Figure 5.11: PSNR plotted for each iteration using the IBP technique, upsampled from two different spatial resolutions: 768×432 and 1280×720 . Generally, best results can be achieved within 3 or 4 iterations

5.4.2 Interpolation Dependent Image Downsampling

Within a video codec the original HR input frame is available at the encoder. The downsampling and upsampling operations therefore need not be entirely independent processes. It is possible to optimise the downsampling process for a given upsampling operation.

Given knowledge of the original HR input image I_h , it is possible to modify the IBP formula 5.8 such that the optimal downsampled image I_l is generated for a given upsampling kernel h_u .

$$(5.10) \quad I_l^{n+1} = I_l^n + (h_d * (I_h - (h_u * I_l^n) \uparrow s) \downarrow s),$$

where I_l^n is the current iteration of the optimised LR image.

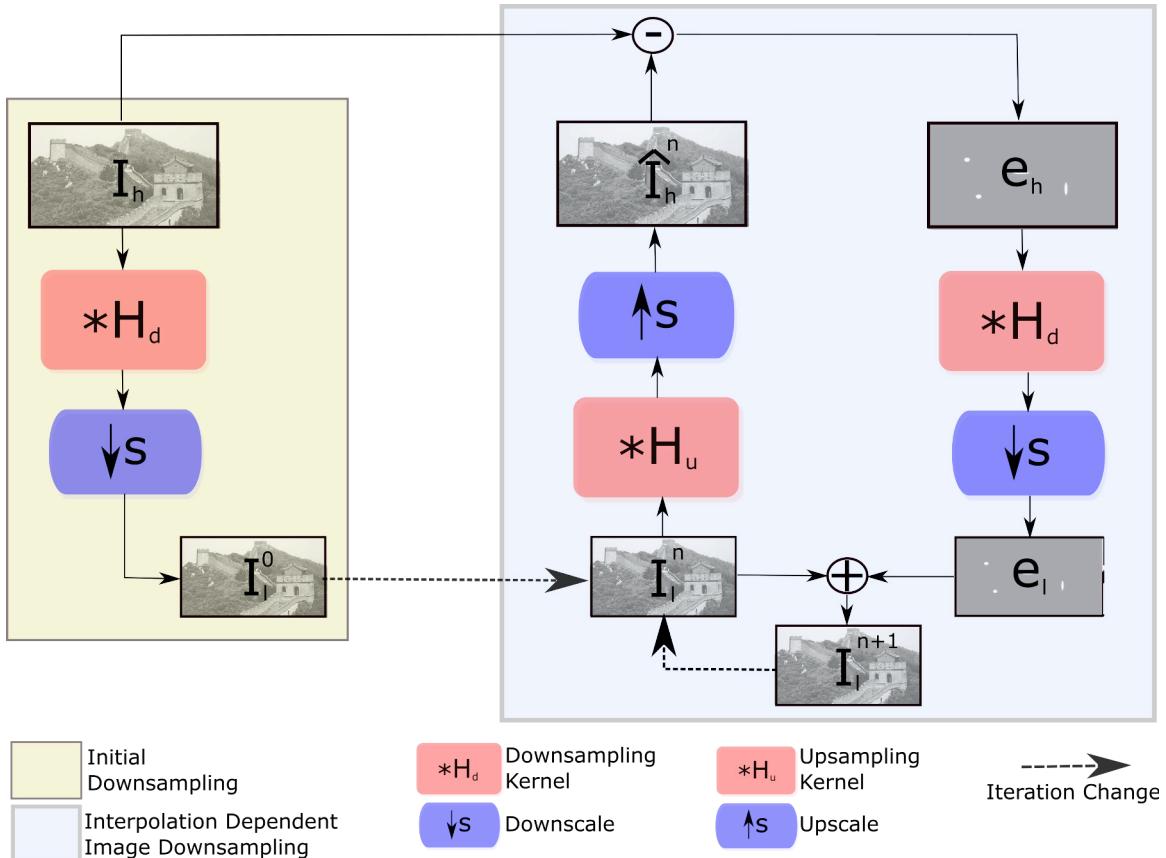


Figure 5.12: Block diagram of the IDID downsampling technique. IDID minimises the error signal between the reconstructed picture \hat{I}_h^n and the original I_h by iteratively adding the downsampled error signal to the LR result to compensate for the information lost during interpolation

Similar ideas have been proposed in the literature [77] [17]. The term Interpolation Dependent Image Downsampling (IDID) was coined by Zhang et al. in [77]. Zhang's IDID approach is to construct the downsampled image using a matrix formulation of the normal equation, given the interpolation kernel and original HR image, such that:

5.4. SUPER RESOLUTION & ENHANCED SAMPLING TECHNIQUES

$$(5.11) \quad \mathbf{I}_l = (\mathbf{h}_u^T \mathbf{h}_u)^{-1} \mathbf{h}_u^T \mathbf{I}_h.$$

However, inversion of large matrices is highly computational. An alternative approach is also proposed involving spitting the image into smaller blocks to reduce the size of the matrix $\mathbf{h}_u^T \mathbf{h}_u$. This approach still requires far more computational power compared to the proposed IDID approach given in 5.10 and is therefore not suitable for real-time video coding.

The additional information contained within the IDID image compensates for the low pass filtering properties of the upsampling kernel h_u resulting in a more accurate reconstructed image. However, IDID tends to create a sharper LR image, thus reducing spatial correlation and potentially reducing compression efficiency when coding in intra mode. Applying the IDID technique in a video coding scenario has the added benefit that it is only required at the encoder and therefore has an overall lower computational cost compared the IBP technique.

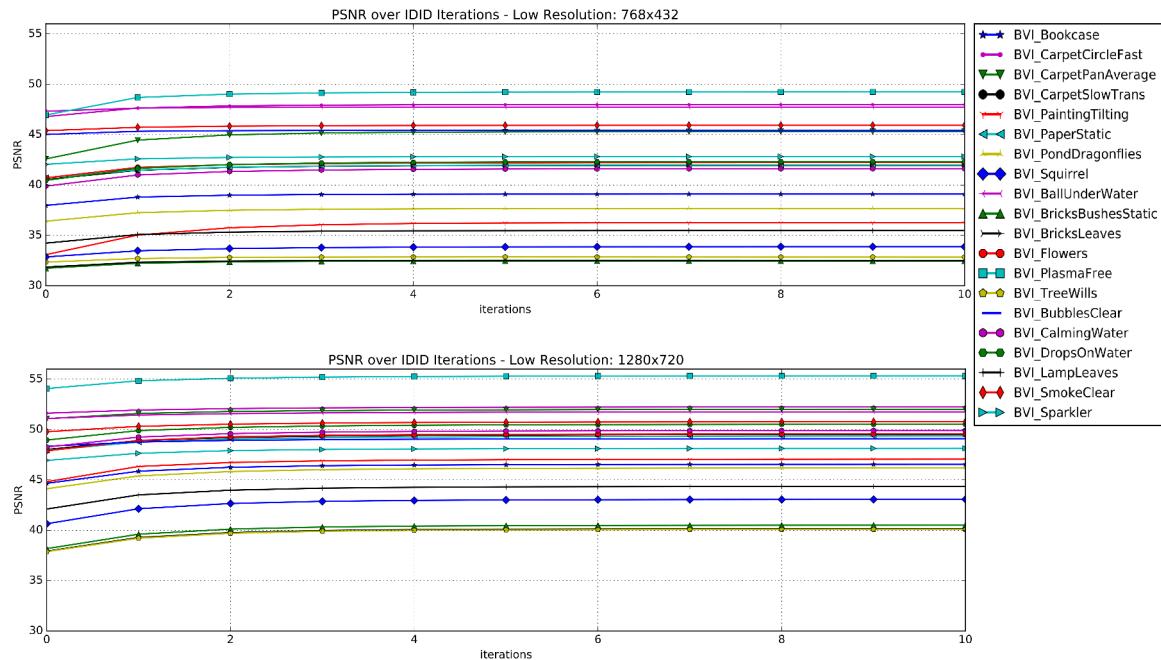


Figure 5.13: PSNR of the reconstructed HR picture plotted for each iteration of the IDID technique. Tested using two different scale factors producing spatial resolutions: 768×432 and 1280×720 . Generally, the largest gains can be achieved after the first iteration but small gains are possible even after 10 iterations.

Figure (5.12) contains a block diagram of the proposed IDID technique. Note that unlike the IBP technique, IDID minimises the difference between the original HR image and the reconstructed HR image. The information contained within the LR image will not necessarily produce a perceptually good representation of the original HR image, which is usually the requirement of a

downsampling operation. However, visual quality of the LR image is not important as only the reconstructed HR image will be displayed and used for reference.

The BVI test images are again used to test resampling performance. Figure (5.13) shows the PSNR gain of the reconstructed HR images after each iteration of the IDID technique. Unlike the IBP technique, the PSNR of the IDID technique will converge but this may require a large number of iterations. IDID has a similar computational cost as IBP and also enables PSNR gains between 1–2dB. A more detailed comparison between these two techniques is given in Section 5.5.

5.4.3 Cross-Scale Self-Similarity

SR techniques can vary greatly in both methodology and performance. Another popular SR subgroup consists of those that exploit the *cross-scale self-similarity* property. This states that similar patches can be found across different spatial resolutions of the same image. Standard interpolation methods are limited in their performance due to their low-pass filtering behaviour; much of the high frequency content contained within the LR image is lost after resampling. These SR techniques therefore attempt to enhance the HR samples by applying a weighted average of LR pixels that are found to have some relation to the local HR patch.

5.4.3.1 Patch Matching

Algorithm 3 Patch Matching

```
1: procedure MATCHING( $I_l, I_h$ )            $\triangleright$  low and high resolution images, respectively
2:    $wrange \leftarrow$  half of the search window length
3:    $height \leftarrow$  height of HR image
4:    $width \leftarrow$  width of HR image
5:   for  $j \leftarrow 1, height$  do
6:     for  $i \leftarrow 1, width$  do
7:        $P_{I_h}(i, j) \leftarrow$  local patch located at  $I_h(i, j)$ 
8:        $(ci, cj) \leftarrow$  closest corresponding LR coordinates
9:       for  $j_l \leftarrow cj - wrange, cj + wrange$  do
10:        for  $i_l \leftarrow ci - wrange, ci + wrange$  do
11:           $P_{I_l}(i_l, j_l) \leftarrow$  local patch located at  $I_l(i_l, j_l)$ 
12:          if  $Cost(P_{I_h}(i, j), P_{I_l}(i_l, j_l)) < Criteria$  then
13:            Apply to SR method
```

Many techniques that exploit the *cross-scale self-similarity* property use patch matching to locate groups of pixels contained in one or more LR images that exhibit some form of relationship to the local pixel data in the HR image to be enhanced. The initial HR image is constructed using a standard interpolation technique – such as Bicubic or by using the SHVC resampling functions. Given that the LR image(s) will usually contain a greater range of frequencies than

the interpolated HR image, it is expected that applying these data will enhance the HR image's spectral content and thus produce better quality results at edge boundaries and textured regions. Patch matching can be achieved using Algorithm 3. A patch within the LR data that fulfils a specified criteria, and therefore deemed '*similar*', can either be applied independently or alongside other similar patches according to the SR technique.

5.4.3.2 Salvador's SISR

In [56], Salvador et al. proposed a method of applying Patch Matching to estimate the high frequency content that is inherently missing from a HR image constructed using standard interpolation techniques. As highlighted in the paper, interpolation-based upsampling methods produce a HR image with shrunk frequency support. Standard interpolation techniques can never fully predict the missing high frequency band up to the wider Nyquist limit. The initial resampled HR image is given by:

$$(5.12) \quad \hat{I}_{h,low} = (h_u * I_l) \uparrow s.$$

The LR image I_l can be separated into its low and high frequency components by first applying a low pass filter:

$$(5.13) \quad I_{l,low} = h_s * I_l,$$

$$(5.14) \quad I_{l,high} = I_l - I_{l,low},$$

where h_s is a smoothing kernel and $I_{l,low}$ and $I_{l,high}$ denote the low and high frequency bands of the LR input image I_l , respectively.

This work utilises the *cross-scale self-similarity* property by finding the patch within $I_{l,low}$ that most closely matches the current patch within $\hat{I}_{h,low}$, using Algorithm 3. $I_{l,low}$ is obtained by applying a filter with the same properties as h_u so that better performance can be gained by matching two images with the same normalised bandwidth. Once the best LR patch is located, the corresponding patch within $I_{l,high}$ is then mapped to a new bitmap to form $\hat{I}_{h,high}$, the high frequency HR image. Assuming accurate HR synthesis of the LR high frequency data $I_{l,high}$, $\hat{I}_{h,high}$ will contain all the high frequency information present in I_l that is missing from $I_{h,low}$. The best match is found within a search window according to the patch with minimal cost:

$$(5.15) \quad P_{I_{l,low}(i_l, j_l)} = \arg \min ||P_{I_{l,low}(i_l, j_l)} - P_{\hat{I}_{h,low}(i, j)}||,$$

where $P_{I_{l,low}(i_l, j_l)}$ is the patch within $I_{l,low}$ that best matches the current patch $P_{\hat{I}_{h,low}(i, j)}$ within $\hat{I}_{h,low}$.

The cost function used in [56] is the P-norm:

$$(5.16) \quad ||x||_p = \left(\sum_{i=1}^n |x_i|^P \right)^{1/p},$$

where $p = 1$ is known as the *taxicab* norm and $p = 2$ is the *Euclidean* norm.

For every pixel in the HR image, the local patch is matched to a LR patch. This results in a sliding window and so the patches in $\hat{I}_{h,high}$ overlap one another. It is therefore necessary to normalise the pixel values according to the number of high frequency estimates used for each pixel:

$$(5.17) \quad p_i = \sum_{j=1}^n \frac{e_{i,j}}{n},$$

where p_i is the normalised pixel, n is the number of high frequency estimates used for pixel i and e_i is a vector containing all the overlapping high frequency pixel values.

The final step is to combine the estimated high frequency HR image $\hat{I}_{h,high}$ with the initial interpolated low frequency HR image $\hat{I}_{h,low}$. According to the paper, better results can be achieved by first removing any low frequency content within $\hat{I}_{h,high}$:

$$(5.18) \quad \hat{I}_h = \hat{I}_{h,low} + \hat{I}_{h,high} - h_s * \hat{I}_{h,high}.$$

A full evaluation of this technique's performance is not included in this thesis. However, initial tests showed that, when applied to Bicubic resampling, only small PSNR and SSIM gains are possible. The authors of the paper also apply IBP as a post-processing technique which is what actually appears to provide the greatest increase in performance. When applied to SHVC resampled images, no gains in performance were observed. The reason for this is that when filtering the LR image I_l using the SHVC interpolation filter coefficients, the result produces exactly the same image and therefore $I_{l,high}$ will not contain any information. This technique is more effective when used with sampling techniques such as Bilinear which removes greater amounts of high frequency content and causes smoothing of sharp edges.

5.4.3.3 Dong's SISR

Although not mentioned in their paper, Dong et al. [14] also exploit the *cross-scale self-similarity* property by applying a patch matching process. This technique aims to reduce the noise contained within the initial interpolated HR image, which can be constructed using the same operation given in equation 5.12. Rather than synthesising the missing high frequency content, as described in Section 5.4.3.2, the HR pixel values are recalculated as a weighting of the interpolated pixel and a number of matched pixels within the LR image determined using Algorithm 3. Similar to equation 5.12, the initial HR image can be constructed by:

$$(5.19) \quad \hat{I}_h = (h_u * I_l) \uparrow s.$$

Patch matching is applied to the LR input image I_l and the initial reconstruction of the HR image \hat{I}_h . Any patch with a *cost* less than a given threshold is deemed similar and the central pixel is grouped within a set $S_{i,j}$ of other similar pixels and the corresponding interpolated HR pixel. The *cost* is determined by:

$$(5.20) \quad \text{cost} = \|P_{I_l(i_l, j_l)} - P_{\hat{I}_h(i, j)}\|.$$

The cost function applied in [14] is the L_2 norm, which equates to the P-norm as given in (5.16) when $p = 2$. A weighting for each pixel within the set of similar pixels is calculated using the exponential function of the $\text{cost}_{(i_l, j_l)}$:

$$(5.21) \quad \omega(i_l, j_l) = e^{-\frac{\text{cost}(i_l, j_l)}{t}}.$$

The weight of the initial HR pixel $\hat{I}_h(i, j)$ is given as $\omega(i, j) = 1$. Before calculating the new HR pixel value, the weights need to be normalised:

$$(5.22) \quad C(i, j) = \omega(i, j) + \sum_{I_l(i_l, j_l) \in S_{(i, j)}} \omega(i_l, j_l),$$

$$(5.23) \quad \omega(i_l, j_l)_{\text{norm}} = \frac{\omega(i_l, j_l)}{C(i, j)},$$

$$(5.24) \quad \omega(i, j)_{\text{norm}} = \frac{1}{C(i, j)},$$

where $C(i, j)$ is the normalisation factor.

This is effectively an image denoising technique but in this case it filters samples according to weightings calculated from the LR image. The result removes noise within low frequency areas and maintains the high frequencies within the areas containing edges and textures. Perceptually such a technique may appear to produce desirable results but it does not necessarily produce an image that is more faithful to the original – which is the requirement for the proposed Adaptive Resolution coding technique.

5.5 Comparison of Enhanced Sampling Techniques

In this section the two most promising enhanced resampling techniques out of those discussed in Section 5.4, namely IBP and IDID, are compared. Firstly by comparing the average PSNR result for each iteration from the results given in Section 5.4. Visual examples are then provided to demonstrate how the reconstructed image may affect perceptual quality. In the final part of this section, regional objective quality analysis is provided to illustrate how well each technique performs given different areas of content and by applying different scaling factors.

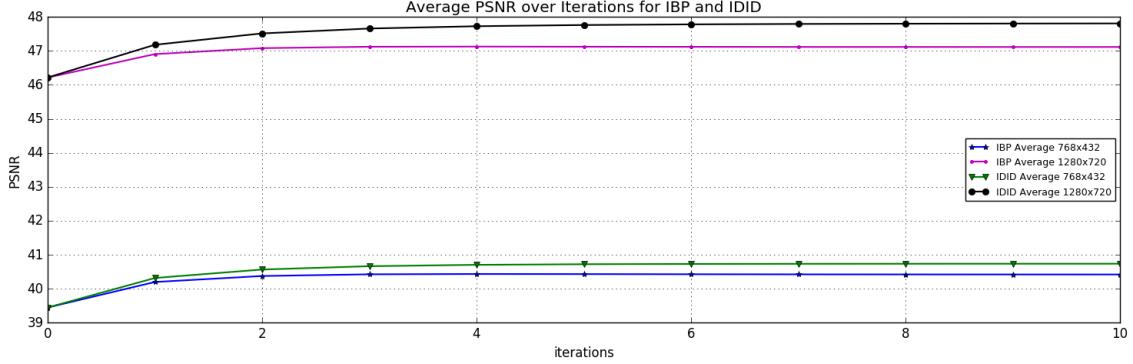


Figure 5.14: Average PSNR plotted for each iteration of the IBP and IDID techniques, upsampled from two different low spatial resolutions: 768×432 and 1280×720 . IDID outperforms IBP for both scale factors but on average greater gains can be achieved when scaling by a factor of $\times 1.5$

5.5.1 Objective Comparison of IBP and IDID

The average PSNR results and the two tested scale factors ($\times 1.5$ and $\times 2.5$) are shown in Figure (5.14). For both scale factors the IDID technique has a greater average PSNR for all iterations. The reason for this is that IBP does not take into account the information lost during the final upsampling process and only attempts to reduce the error between the initial LR image and subsequent iterations of the downsampled reconstructed image. Alternatively, IDID minimises the error between the reconstructed and original high resolution images.

5.5.2 Visual Comparison of IBP and IDID

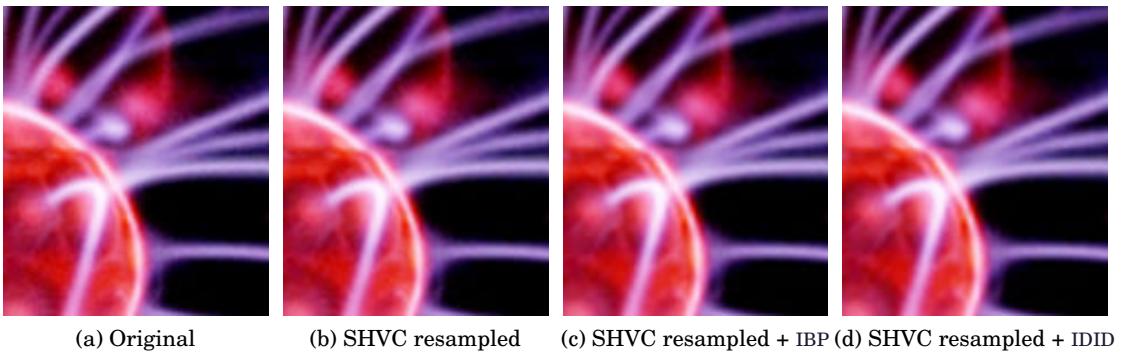


Figure 5.15: Visual comparison of sampling techniques, with scale factor $\times 2.5$, using a cropped section from the test image taken from the sequence *PlasmaFree*. Very little difference can be observed between these techniques as the initial resampled image is already a good reconstruction of the original

High PSNR values, or even high SSIM values, do not necessarily mean that the image will be perceived to be of high quality or at least any better than an image with a lower calculated

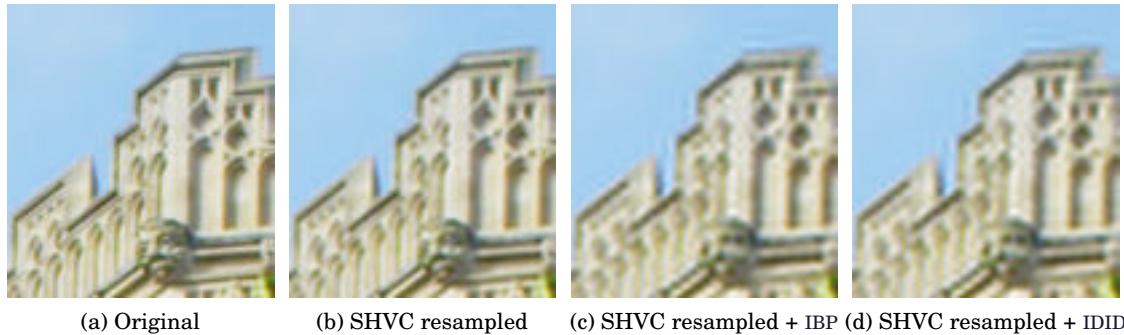


Figure 5.16: Visual comparison of sampling techniques, with scale factor $\times 2.5$, using a cropped section from the test image taken from the sequence *TreeWills*. The initial resampled result, as given in a), fails to reproduce the high frequencies that form the sharp edges of the building structure. Therefore the error signal, when applying IBP or IDID, contains large amounts of information which can not be reconstructed effectively given the large scale factor, thus produces unwanted ringing artefacts

value. These metrics produce a normalised value representing a measured average quality level across the entire image. Figure (5.15) shows visual comparisons of a cropped section of the test image taken from the sequence *PlasmaFree* scaled down to a spatial resolution of 768×432 before upsampling back to the original HR ($\times 2.5$). Figures (5.11) and (5.13) indicate that the initial resampled image, produced using the SHVC sampling functions, has a high PSNR and applying multiple iterations of either the IBP or IDID technique can further increase PSNR. However, there is very little visual difference between the results of these techniques and the initial resampled image is clearly a good reconstruction of the original.

Figure (5.16) shows a cropped section from the test image taken from the sequence *TreeWills* scaled down to a spatial resolution of 768×432 before being reconstructed at the original HR ($\times 2.5$). This cropped section highlights an area of the image subjected to high levels of visually perceived distortion. The initial resampled image using the SHVC functions fails to reproduce the structural fidelity of the captured building – edges appear smooth and much of the original detail is lost. Although both IBP and IDID produce better PSNR results, it could be argued that in this portion of the image applying either of these enhancement techniques is not beneficial. Due to the loss of high frequency information within the SHVC resampled image, greater amounts of information will be contained within the error signals when applying either of the enhanced sampling techniques. Both techniques produce ringing artefacts at edge boundaries as a result of resampling this error and applying the result to the previous iteration. This is because the error signal can not be reconstructed without significant distortion when scaling by large scale factor.



Figure 5.17: Visual comparison of sampling techniques, with scale factor $\times 1.5$, using a cropped section from the test image taken from the sequence *TreeWills*. The smaller scale factor can potentially reduce the amount of information lost in the initial resampled picture, shown in a). However, visually it is apparent that much of the high frequency information is lost resulting in smoothing of sharp edges. Due to the smaller scale factor and greater sampling rate at the LR, compared to the result in Figure (5.16), IBP and IDID are able to reproduce more of this information with minimal artefacts

Figure (5.17) shows the same cropped section of the test image taken from the sequence *TreeWills* but with the scale factor reduced to $\times 1.5$ and therefore the LR input image has a spatial resolution of 1280×720 . The SHVC resampled result is clearly still missing high frequencies that form the sharp edges within the original image. However, the IBP and IDID techniques are capable of reconstructing much more of this information given the smaller scale factor applied to the error signal. Both enhanced techniques produce a visibly better reconstruction of the original HR image without producing noticeable artefacts.

5.5.3 Region Objective Quality Performance of IDID and IBP

The enhanced sampling techniques IBP and IDID minimise the overall MSE, thus enhancing PSNR. However, as mentioned before, this is a normalised value representing objective quality across the entire image. It is possible for some areas of the image to be subjected to more distortion as a result of applying these techniques. In this section, PSNR is applied on a block by block basis to evaluate local objective quality.

Figures (5.18) and (5.19) show the result of a block-PSNR function calculated using a block size of 9×9 for which the PSNR result is stored within the sample positioned in the centre of this block. This function is applied to the whole image from the sequence *TreeWills* after resampling using the three methods: resampling using the SHVC sampling functions, and IBP and IDID also using these same sampling functions. Figure (5.18) contains the results after resampling by a scale factor of $\times 2.5$. These results show that in areas of low frequency, the enhanced sampling

5.5. COMPARISON OF ENHANCED SAMPLING TECHNIQUES

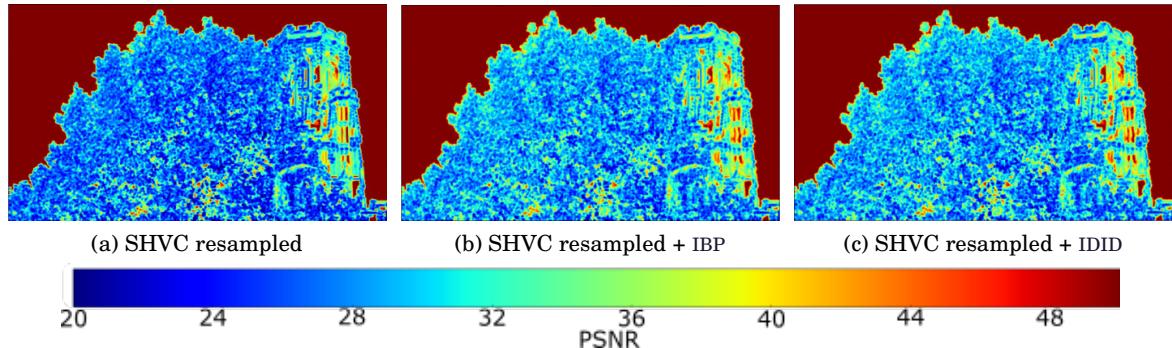


Figure 5.18: Local PSNR colourmap for the test image from the sequence *TreeWills* scaled by $\times 2.5$. Low frequency content is well reconstructed, however, high frequency textures and edges are subjected to high levels of distortion. Both IBP and IDID are able to reduce this distortion to provide an overall higher PSNR

techniques do not provide any benefit – areas producing PSNR values around $50dB$ are likely to be perceived to be identical to the original. The higher frequency content that forms the tree and the building are subjected to much greater levels of distortion but it can be seen that overall the enhanced sampling techniques increase local PSNR within these regions. In Figure (5.19), the results after resampling by a lower scale factor of $\times 1.5$ are given. No difference in quality can be observed within the low frequency regions but it is clear that the information within the regions comprising of the tree and building are much better preserved by only down-sampling to a spatial resolution of 1280×720 . Overall the enhanced sampling techniques are again able to reconstruct these regions more accurately and therefore produce a higher local PSNR.

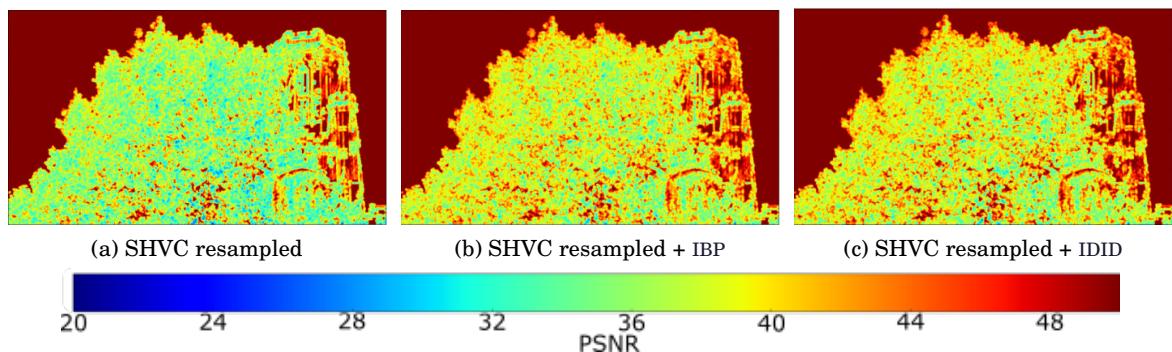


Figure 5.19: Local PSNR colourmap for the test image from the sequence *TreeWills* scaled by $\times 1.5$. By reducing the scale factor little difference can be observed in the low frequency areas but a big improvement can be seen in the areas consisting of high frequency content. Further reductions in distortion can be observed by applying the IBP or IDID techniques

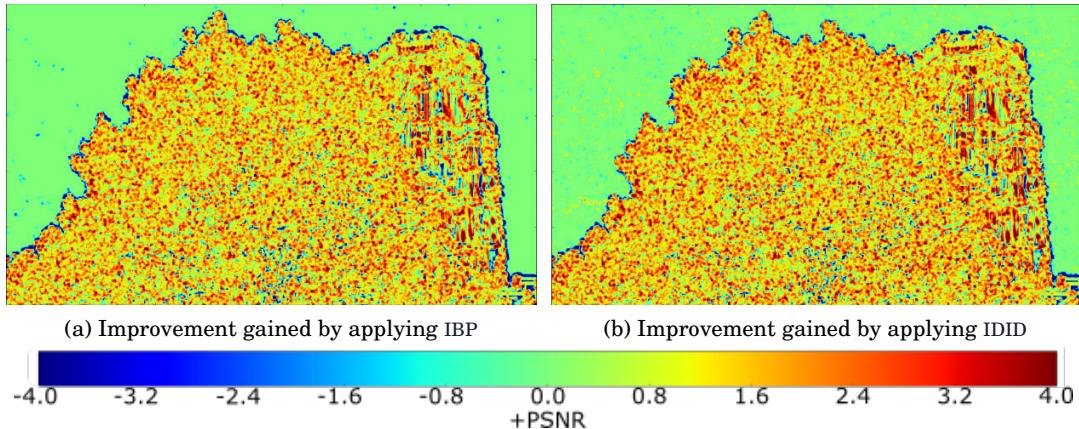


Figure 5.20: Local PSNR improvement provided by applying IBP and IDID when scaling by $\times 2.5$. Positive values indicate a measured local PSNR gain compared to just resampling using the SHVC functions. As expected, the majority of the high frequency regions are enhanced by applying either of these techniques. However, some areas are subjected to greater levels of distortion – especially the edges that form the tree and building structures

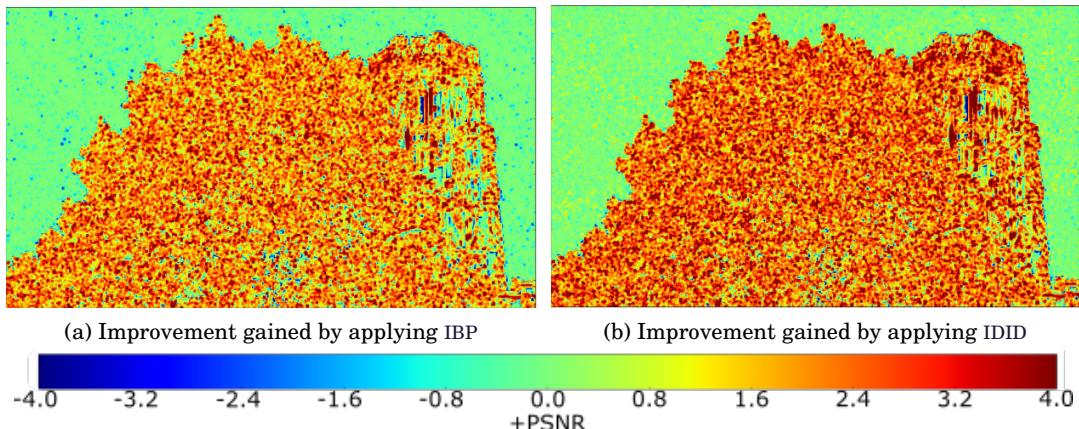


Figure 5.21: Local PSNR improvement provided by applying IBP and IDID when scaling by $\times 1.5$. Positive values indicate a measured local PSNR gain compared to just resampling using the SHVC functions. Even greater benefits can be observed by applying these techniques using the smaller scale factor, compared to the result in Figure (5.20). It can be observed that fewer areas are subjected to distortion, especially by applying IDID

Figures (5.20) and (5.21) show the difference in local PSNR for the IBP and IDID techniques compared to just resampling using the SHVC functions. Positive PSNR means that the enhanced sampling technique provides a benefit, whereas negative values show a decrease in local PSNR. Figure (5.20) represents resampled images by a scale factor of $\times 2.5$. Very little difference can be seen within the low frequency regions forming the blue sky. There are however some regions that clearly do not benefit from these techniques. These correlate with the findings given in Section

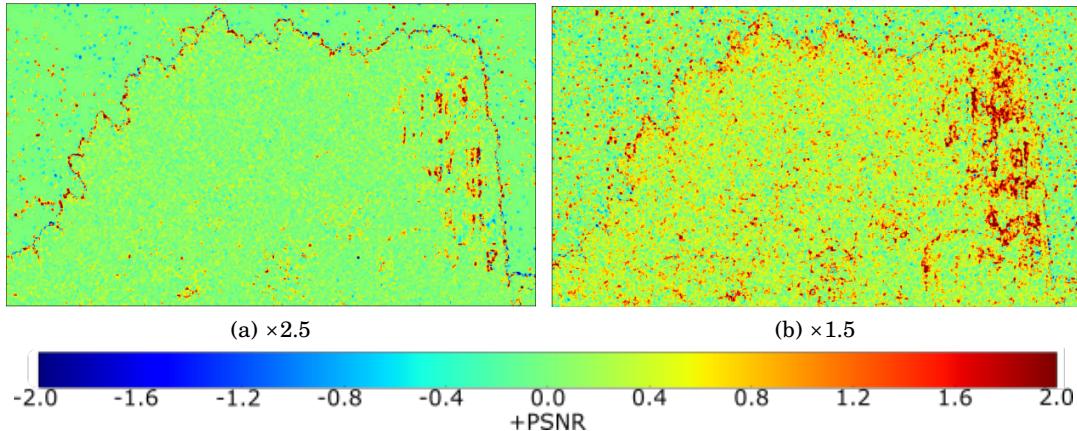


Figure 5.22: Direct comparison of local PSNR for IBP and IDID, resampled from both of the tested low spatial resolutions. Positive values indicate an increase in measured local PSNR for the technique IDID, whereas negative values indicate better performance using the IBP technique. For each scale factor it is clear that IDID is the better resampling technique, especially when scaling by $\times 1.5$

5.5.2 regarding the results in Figure (5.16). The information forming the sharp edges around the structures is lost during downsampling and the SHVC reconstructed HR image produces a smoother transition across these edges. Both IBP and IDID introduce significant distortion in these areas and therefore the content would be better represented without applying these techniques. Figure (5.21) contains resampled images by a scale factor of $\times 1.5$. Some difference can be seen in the low frequency regions; IBP causes local PSNR to drop in some areas within the blue sky whereas IDID produces slightly better results. It can also be seen that only small amounts of distortion is introduced as a result of applying these techniques around the tree and building structures.

As a final evaluation of spatial variability of PSNR, Figure (5.22) shows the local PSNR difference between IBP and IDID for both of the tested scale factors. Positive values represent an increase in the performance for IDID over IBP while, conversely, negative values represent an increase in IBP performance over IDID. These results show little benefit of applying the IBP technique over the IDID technique. With the larger scale factor of $\times 2.5$, the two techniques produce more similar results compared to scaling by $\times 1.5$. However, it can be noted that for both scale factors, the sharp edges that form the main structures are reconstructed better using IDID. When scaling by $\times 1.5$ much larger areas can be seen to benefit from applying IDID over IBP.

5.5.4 IBP vs IDID for Enhancement of Intra Coded Sequences

Coding can affect the performance of a resampling technique. For the Adaptive Resolution coding method, upsampling is applied after coding the LR picture and therefore any coding artefacts introduced will also be upsampled. Figure (5.23) shows a block diagram of the IBP technique with

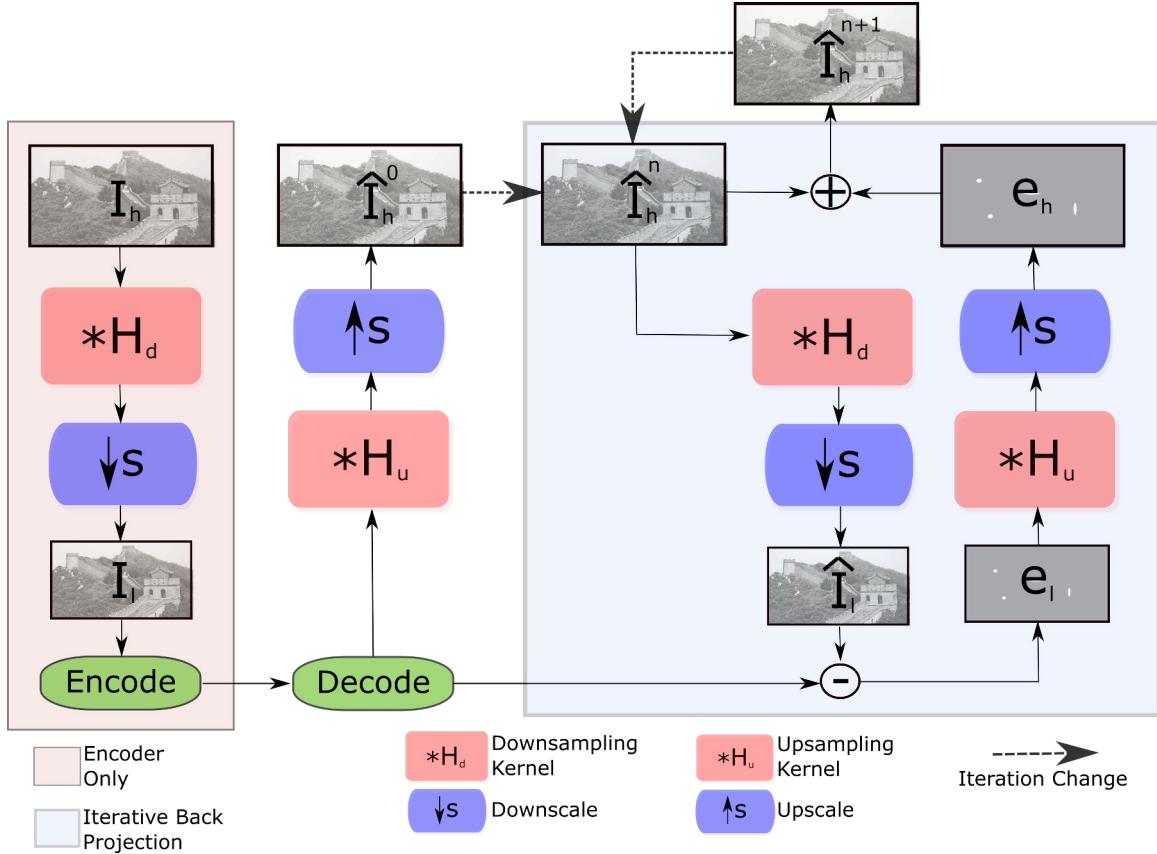


Figure 5.23: Block diagram of the IBP upsampling technique within a video codec. IBP effectively enhances any coding artefacts introduced at the LR

the addition of the encoding and decoding operations. It can be seen that the compressed LR image will affect the performance of the technique as it is contained within the resampling loop. In effect the technique will now minimise the error between the downsampled reconstructed image \hat{I}_l , which will contain some degree of compression and resampling distortion, and the initial downsampled and coded picture I_l . In Figure (5.24), a block diagram of the IDID technique with the addition of the encoding and decoding operations is shown. Note that these coding operations are not contained within the resampling loop as they are in IBP. IDID can therefore be performed without any affect to the actual optimisation process. However, coding efficiency may be reduced due to the additional information contained within the LR picture. Also, quantisation may remove some of this additional information thus reducing the benefit of applying this technique.

Figure (5.25) and Table 5.9 shows the rate-distortion performance of four intra-coded sequences using standard HEVC and also resampled variations using SHVC resampling functions and the IBP and IDID techniques. Results are calculated up to the critical bitrate [48] – the maximum bitrate to provide a benefit. Possibly the most surprising result is that IBP performs worse than just SHVC resampling without any enhancement, indicated as *Resampled*. This is

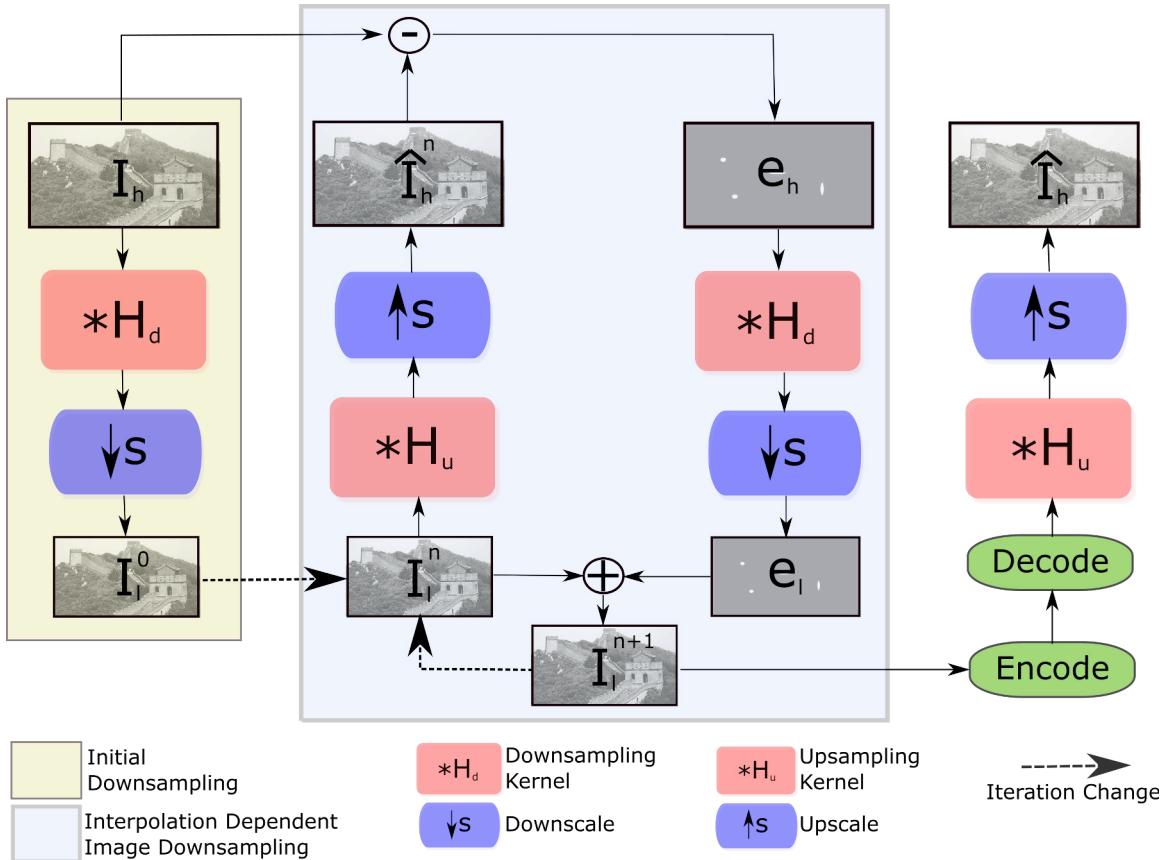


Figure 5.24: Block diagram of the IDID downsampling technique within a video codec. Coding is performed outside of the resampling loop and therefore does not effect the optimisation process. However, quantisation may remove additional information added to the LR picture I_l^n , thus reducing performance

because at lower bitrates higher QP values are applied and therefore more coding distortion is introduced. IBP effectively enhances this distortion, thus produces a lower quality reconstruction. When applying lower QP values at higher bitrates, IBP performs better but the overall distortion is generally greater than the standard coding approach at the original HR. High QP values also affect the performance of the IDID technique as the high frequency information introduced to compensate for low-pass filtering will be removed. Therefore as QP increases, the performance of IDID will approximate the SHVC *Resampled* result. Overall, the IDID technique provides a significant gain in rate-distortion performance and will therefore be applied as the underlying resampling technique within the proposed Adaptive Resolution coding method in Chapter 6.

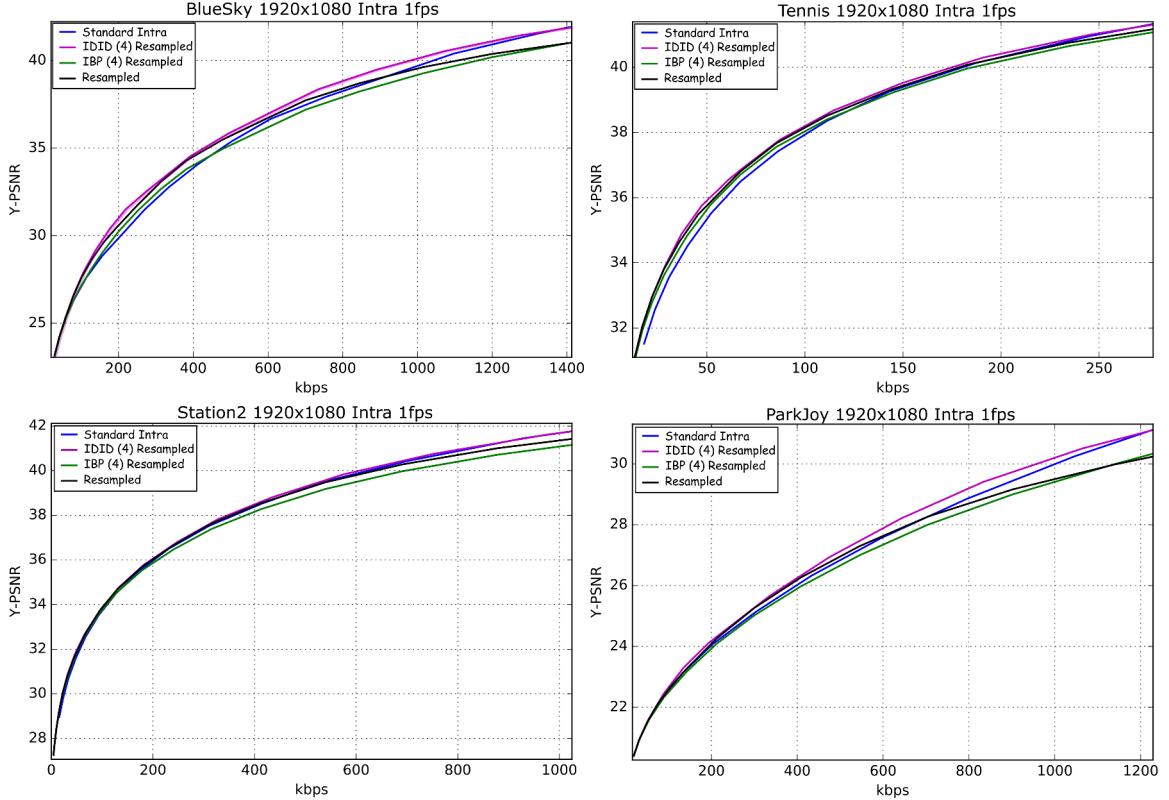


Figure 5.25: Rate-distortion performance comparisons when coding at lower spatial resolutions prior to upsampling back to the original HR using the 3 resampling methods. Standard HEVC results are also provided and all coding is applied in intra mode with a normalised frame rate of 1Hz. (#) indicates the number of iterations applied for each technique. In each case, IDID produces the greatest level of performance

Table 5.9: BD-PSNR [3] average bitrate savings, compared to standard HEVC, when coding at lower spatial resolutions prior to upsampling back to the original HR using the 3 resampling methods

	BlueSky	Tennis	Station2	ParkJoy
Resampled	8.476%	10.483%	4.302%	0.380%
IBP 4*	-0.012%	6.112%	-0.746%	-5.525%
IDID 4*	13.389%	12.378%	5.991%	6.439%
Critical Bitrate**	1375kbps	276kbps	1009kbps	1210kbps

*Indicates the number of iterations applied in each method

**The highest bitrate to provide a benefit as described in [48], also used when calculating the BD-PSNR metric result

5.6 Intra Coding Performance of IDID Resampled Sequences

It was found in Section 5.5 that the IDID technique reduces overall resampling distortion and can increase rate-distortion performance of intra-coded pictures. In this section results are given for each of the test sequences, detailed in Appendix-A. As it was discussed in Chapter 4, the number of LR IDR pictures will be limited to 3: 1536×864 , 1280×720 and 768×432 for HD sequences and 2560×1440 , 1920×1080 and 1536×864 for UHD sequences. This is in order to reduce computational complexity whilst maintaining a wide range of scaling factors.

Table 5.10: BD-PSNR [3] average savings for IDID resampled and coded sequences compared to standard HEVC. Maximum bit-budget is determined by the highest number of bits to provide a benefit

Video	Bit Range	BD-PSNR	Video	Bit Range	BD-PSNR
Squirrel	0-3600kb	6.279%	BricksLeaves	0-3353kb	5.424%
	0-1500kb	8.126%		0-1250kb	5.931%
	1000-2500kb	5.563%		1000-2250kb	6.017%
	2000-3600kb	3.207%		2000-3353kb	3.378%
BlueSky	0-1748kb	13.249%	Station2	0-1164kb	5.907%
	0-700kb	16.911%		0-400kb	7.681%
	500-1200kb	8.505%		300-700kb	2.392%
	900-1748kb	3.915%		600-1164kb	1.455%
Tennis	0-344kb	11.921%	ParkJoy	0-2312kb	7.379%
	0-150kb	14.738%		0-1000kb	9.014%
	75-225kb	5.628%		500-1500kb	7.971%
	150-344kb	3.356%		1000-2312kb	4.964%
InToTree	0-2975kb	5.131%	OldTownCross	0-1305kb	5.680%
	0-1500kb	5.793%		0-500kb	6.977%
	500-2000kb	3.388%		400-900kb	2.716%
	1000-2975kb	3.715%		800-1305kb	0.851%
ReadySetGo	0-1349kb	13.985%	Jockey	0-747kb	26.500%
	0-500kb	17.630%		0-275kb	31.213%
	400-900kb	9.429%		250-525kb	10.382%
	800-1349kb	3.746%		500-747kb	0.838%

Table 5.10 contains BD-PSNR [3] average bitrate savings. To provide a more meaningful result, the frame rate for each sequence is normalised to 1Hz in order to give an indication of the expected performance increase for a given bit-budget – it is in fact the average number of intra-coded bits per sequence. As in Section 5.5.4, the maximum *Bit Range* is defined according to the highest number of bits to provide a benefit over standard HEVC intra-coding. This maximum is dependent on content and it should be noted that a low maximum may still provide a high level of picture quality. Additional bit ranges are given, within the larger range, to give indication of the instantaneous bit savings for varying bit-budgets.

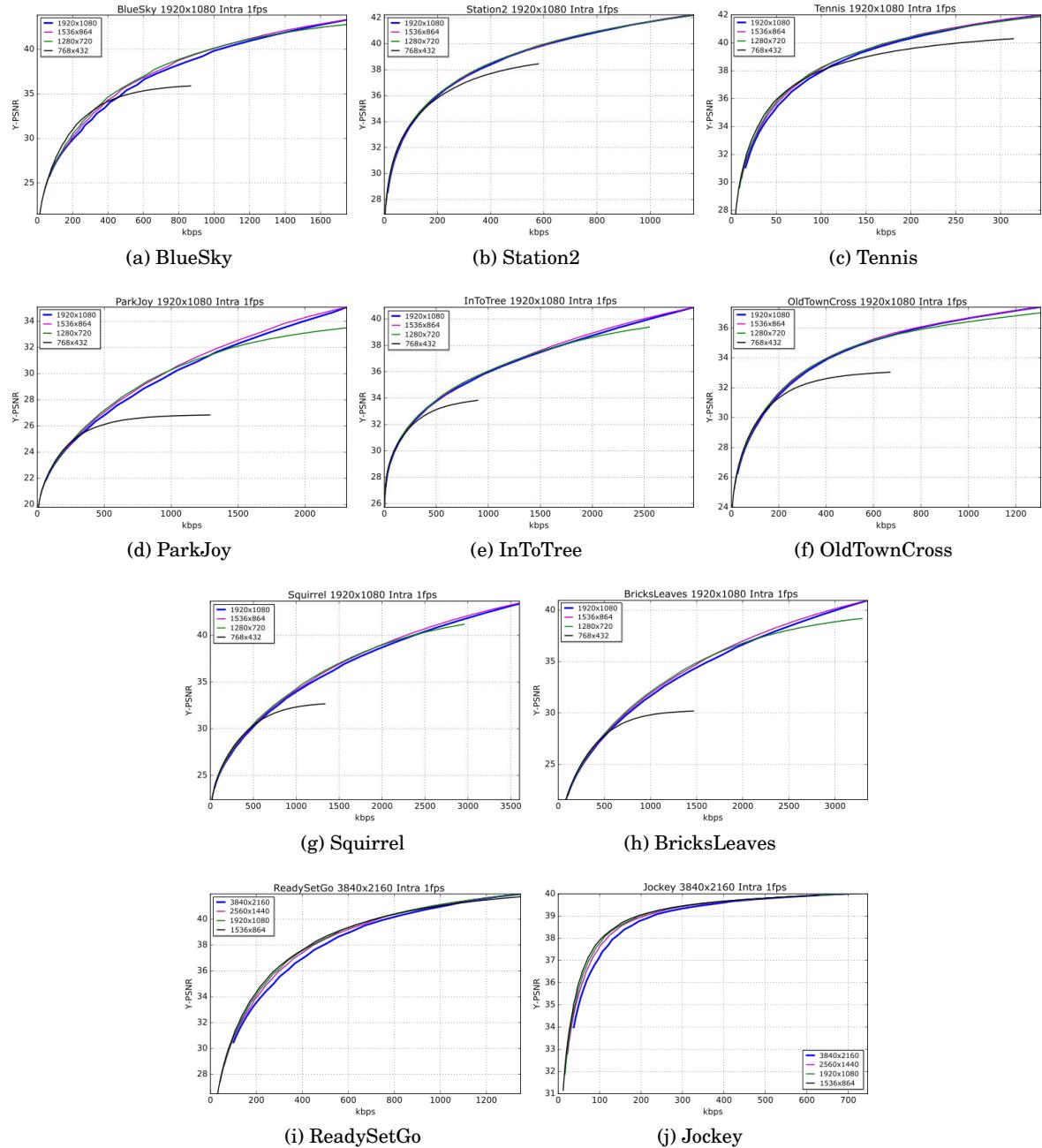


Figure 5.26: IDID rate-distortion curves for each sequence and scale factor. Generally, lower rates benefit more from greater scale factors and as the number of coded bits increases, the best scale factor reduces

5.6. INTRA CODING PERFORMANCE OF IDID RESAMPLED SEQUENCES

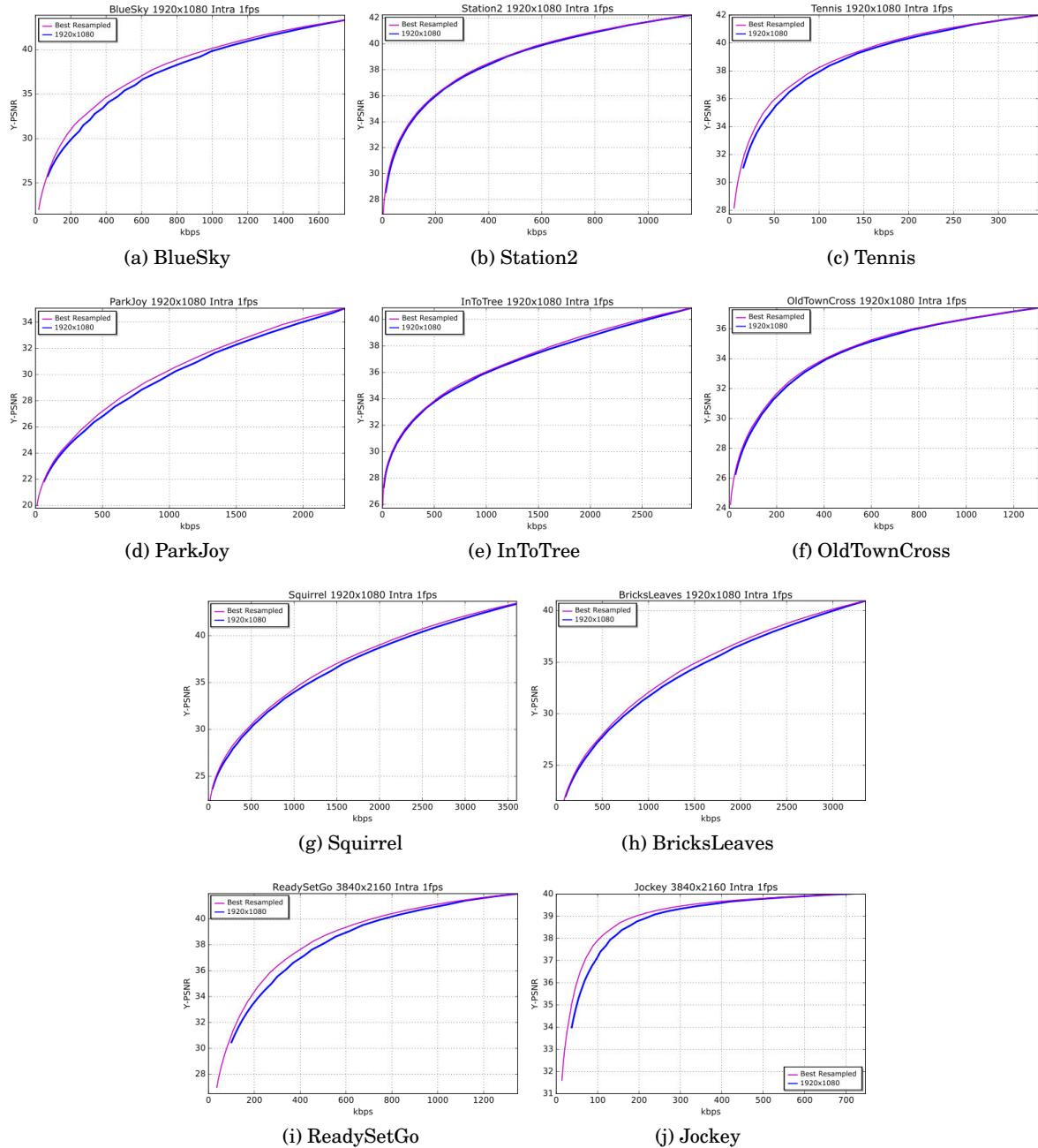


Figure 5.27: Best IDID rate-distortion curve for each sequence given each scale factor. In each case for a wide range of PSNR, resampling and coding using IDID reduces the requirement of coded bits

These results show that for each sequence a significant increase in intra-coding performance can be expected when resampling IDR pictures within the proposed Adaptive Resolution coding technique. The level of performance tends to be higher for lower bit-budgets and drops as the maximum beneficial rate is reached. In some cases the average bit savings within the smaller bit ranges is much higher than the overall average. This suggests that the lower the IDR bit-budget within an inter-coding framework, the greater the chance of increasing rate-distortion performance compared to standard HEVC. As the IDR bit-budget increases, typically as the overall target bitrate increases, the level of performance gained will decrease. However, the IDR bit-budget is determined by the rate control algorithm which is not part of the standard and so different algorithms may select different budget ratios between IDR pictures and the remaining inter-coded pictures within the GOP. Figure (5.26) contains rate-distortion curves for each of the test sequences. Each LR is plotted independently in order to demonstrate how performance varies for each scale. Typically larger scale factors provide the greatest gains at lower rates and as the rate increases smaller scale factors become more optimal. Figure (5.27) contains corresponding rate-distortion curves where the best resampled data points are plotted within one curve. These data points are used to calculate the BD-PSNR metric results in Table 5.10. A key additional piece of information that we can observe from these graphs is the level of PSNR at the maximum beneficial bitrate. As mentioned before, this maximum will vary for different picture content – in some cases only a small range of bitrates will provide a benefit. However, provided that a high level of image quality can be achieved at these low bitrates, the rate control algorithm can allocate a low bit-budget for each IDR picture and still produce high quality video. These PSNR values range between roughly 35 – 43dB, which is considered high for compressed picture data and will likely correspond to a perceptually high level of quality due to the close similarity to the original picture. These results therefore also show that resampling of intra-coded pictures can reduce the number of bits required for producing images within a wide range of quality levels. In almost all cases, the average level of PSNR at the maximum beneficial rate is above, or equal to, 40dB. Compared to the SI index data given in Section 2.6, the level of performance achieved for each sequence very loosely correlates with the order of measured spatial information. The reason for this is because the proposed IDID technique performs well within textured regions but can actually reduce quality at edge boundaries, as shown in Section 5.5. Therefore sequences such as *BlueSky*, which has a relatively high SI index due to large textured regions, can benefit from greater performance gains compared to sequences with a lower SI index which are formed from a higher ratio of edges to textures. To provide a more accurate measure of spatial information, a formula that differentiates textures from edges to apply different weightings, may give better indication of the expected level of coding efficiency.

5.7 Chapter Summary

In this chapter a variety of resampling techniques were reviewed. Many applications apply spatially invariant techniques due to their relatively low computational complexity. Perhaps the most common of these techniques is Bicubic sampling. However in this chapter it was shown that the sampling functions implemented into the SHVC scalable extension of HEVC reduce the average resampling distortion for the tested dataset.

A method of spatially invariant filtering, SR techniques and enhanced sampling approaches are also discussed and reviewed. In particular the upsampling technique IBP, which lead to the proposal of a IDID technique for optimal downsampling. It was found that IDID performs better as it minimises the error between the HR reconstructed image and the original. Local PSNR performance was also calculated in order to evaluate each technique's performance given different areas of image content. Although IDID can produce an overall better quality image, for some types of image content and for larger scale factors, this technique can introduce unwanted artefacts. Although not analysed further in this thesis, an adaptive IDID approach might provide better results by first identifying regions where the technique may perform badly and then by only applying the technique to the rest of the image.

As the purpose of this chapter was to find a better alternative to Bicubic for resampling IDR pictures, these techniques were also tested by coding the LR pictures in intra mode before upsampling. IDID again proved to be the best technique and it was also found that IBP can actually introduce greater levels of distortion when applying high QP values. In the final section of this chapter, a full set of results are provided for the purpose of demonstrating the rate-distortion performance for each test sequence coded in intra mode. These results can give some indication of the level of performance expected when applying IDID to the proposed Adaptive Resolution coding technique. For a given bit-budget determined by rate control, if the picture can be coded more efficiently at a lower spatial resolution then the IDR picture will provide a better reference for inter-coded pictures within the GOP. Results also show that, for each sequence, the maximum average bitrate can produce very high quality compressed pictures. This suggests that even for high target bitrates, rate control can reduce the bit-budget of IDR pictures without loss of quality and allocate more bits to inter-coded pictures.

CHAPTER



ADAPTIVE RESOLUTION CODING

Delivering HR video in restricted bandwidth scenarios can be challenging. High levels of coding distortion can render a video transmission useless by failing to provide enough detail from the original captured scene. Some degree of this distortion can be alleviated by coding at lower spatial resolutions. Both coding and resampling introduces distortion and therefore best results can be obtained by minimising the distortion produced by these two processes. True video quality is assessed over the duration of the sequence; high variability of picture distortion can alter the perceived level of coding efficiency. A well coded sequence is therefore one that maintains a high but also consistent level of quality, and also a well regulated bitrate that prevents skipping of frames due to the over fullness of the channel buffer

As first discussed in Chapters 1 and 3, Figure (1.1) highlights a major problem with the rate control algorithm currently used within the HM reference software (HM16) when targeting low bitrates. Given that each IDR picture prevents itself and future frames referencing prior coded pictures, it is important that they are coded to a high level of quality to provide a good reference that will enable effective coding of the remaining inter-coded B pictures. For this reason, a relatively low QP value is assigned to reduce coding distortion of each IDR picture. At low bitrates this could potentially consume almost the entirety of the bit-budget allocated to the GOP, thus resulting in very high quantisation of B pictures and greatly reducing the quality of these frames. This chapter combines the work carried out in previous chapters and demonstrates a more practical solution compared to the initial version of the proposed coding technique presented in Chapter 3.

6.1 Rate Control Modifications

As can be seen from Figure (6.1a), at low bitrates the quality of IDR pictures far exceeds the quality of the inter-coded B pictures, resulting in fluctuations in PSNR and potentially reducing perceptual quality. Reducing the bit-budget of IDR pictures will enable more bits to be coded for the remaining inter-coded B pictures. However, at the original HR, allocating fewer bits to the IDR pictures requires greater quantisation and may therefore reduce the overall level of video quality, as IDR pictures are initially the only reference within the GOP.

6.1.1 The Rate Control Problem

For the proposed Adaptive Resolution coding technique, the optimal bit-budget for each IDR picture depends on a number of factors. To simplify the problem, consider first standard HEVC with a fixed resolution. In this case, rate control is required to set a bit-budget for the current IDR picture given the number of bits allocated to the entire GOP and knowledge of the picture to be encoded. The bit-budget will determine the QP that is predicted to generate a number of coded bits equal to or less than the bit-budget itself. In turn the QP will affect the amount of information that is lost from the original picture and therefore affect image quality. The quality of the IDR picture and the number of bits remaining ultimately determines the quality of the inter-coded B pictures within the GOP. The overall coding efficiency is therefore dependent on how well the IDR bit-budget and the resulting level of picture quality is balanced.

To determine the optimal bit-budget, the rate control algorithm also requires knowledge of the temporal correlation between the IDR picture and the inter-coded pictures that reference it. This presents a number of problems which can not be solved without knowledge only available after coding the entire GOP, therefore presenting another *chicken and egg* dilemma – the only way to optimise the coding process is by having knowledge of how the coded IDR picture will affect future inter-coded pictures. The effectiveness of any rate control algorithm therefore depends on spatial and temporal video content. Assuming high temporal correlation it would be beneficial to code a high quality IDR picture with a high bit-budget as inter-coded pictures can be coded efficiently while only comprising of a small number of bits. Reduced temporal correlation would mean that inter-coded pictures would require additional bits for coding motion vectors and possibly intra-coded data, therefore a lower bit-budget for the initial IDR picture would provide better results.

For the proposed Adaptive Resolution coding technique the rate control problem becomes more complicated. This is because the IDR bit-budget at the original HR is not necessarily going to be optimal if the picture were coded at a lower spatial resolution and produced a higher quality reconstruction. Better results might be achieved by reducing the bit-budget further to

6.1. RATE CONTROL MODIFICATIONS

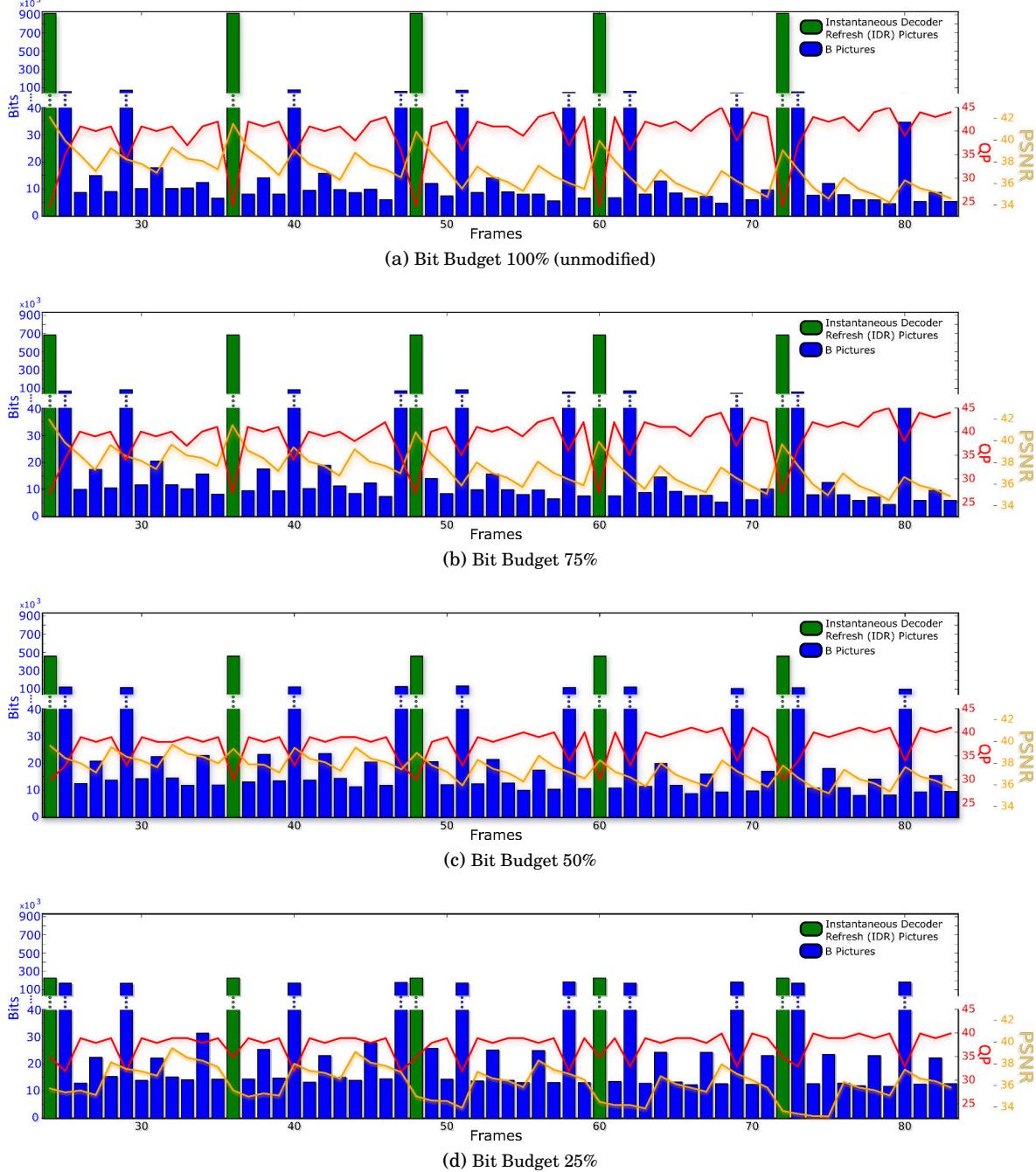


Figure 6.1: The effect of reducing the IDR bit-budget to a percentage of the original budget determined by rate control over 5 GOPs within the sequence *BlueSky* – coded with a target bitrate of 1500kbps and using the Low Delay coding configuration. Reducing the bit-budget causes the rate control algorithm to shift the amount of quantisation from the subsequent B pictures to the initial IDR picture, thus potentially reducing variation of the bitrate and increasing the overall average level of quality. Over quantisation of IDR pictures will have the opposite effect as larger portions of B pictures will be coded in intra mode to account for the lack of good reference pictures and therefore reduce compression efficiency

the extent that a larger number of bits can be used for coding the inter-coded pictures while the quality of the referenced IDR picture remains roughly the same. Therefore, in addition to having knowledge of future frames and the level of distortion that each QP would produce, to select the optimal bit-budget you would also require knowledge of the amount of distortion that each resampled version of the picture would produce. Although selecting the optimal bit-budget may be impossible under normal circumstances, for many of the test sequences used in this thesis a reduced bit-budget will likely improve performance at low target bitrates.

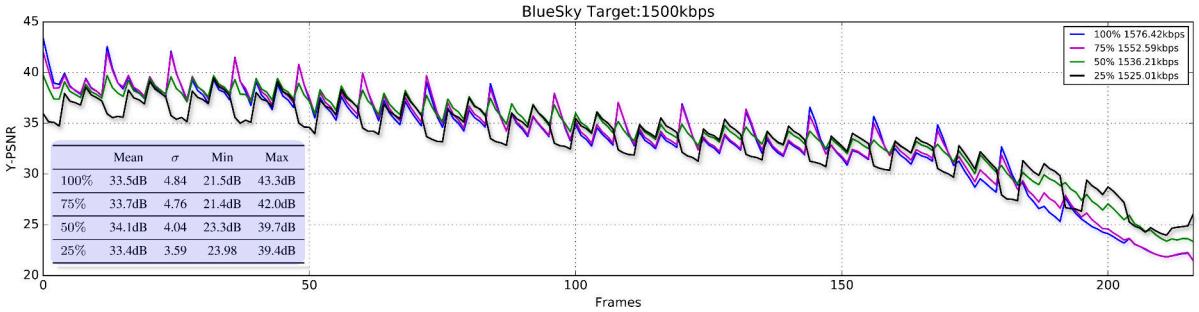


Figure 6.2: The affect on PSNR as a result of reducing the bit-budget of IDR pictures to a percentage of the original budget determined by rate control for sequence *Bluesky* – coded with a target bitrate of 1500kbps and the using Low Delay coding configuration. In this example, the most perceptually desirable result is likely to be with a reduction of 50% as this provides the highest average PSNR and also reduced variation of PSNR, thus providing a better viewing experience

Table 6.1: Comparison of PSNR statistics for different percentages of the original IDR bit-budget determined by rate control, based on data illustrated in Figure (6.2)

	Mean	σ	Min	Max
100%	33.5dB	4.84	21.5dB	43.3dB
75%	33.7dB	4.76	21.4dB	42.0dB
50%	34.1dB	4.04	23.3dB	39.7dB
25%	33.4dB	3.59	24.0dB	39.4dB

6.1.2 Adapting the Bit-Budget

The aim of this chapter is not to develop a new rate control algorithm but rather to demonstrate that greater performance can be achieved at low bitrates by adapting the IDR bit-budget to enable more effective use of the Adaptive Resolution coding technique. Figure (6.1) shows the effect of reducing the bit-budget of IDR pictures for the HD sequence *BlueSky* with a low target bitrate of 1500kbps using standard HEVC coding with a fixed spatial resolution. Figure (6.2) shows corresponding PSNR variation over time and Table 6.1 provides PSNR statistics for the entire sequence. Figure (6.1a) illustrates hows rate control performs without any modifications

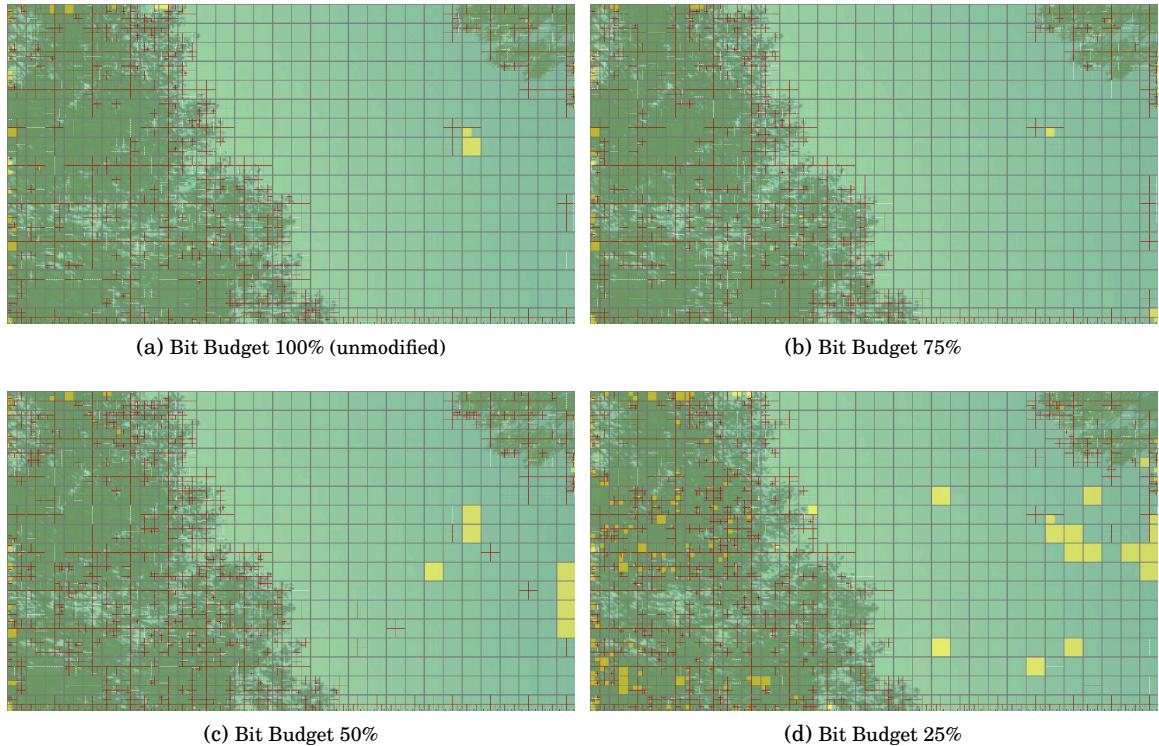


Figure 6.3: Comparison of block partitioning and coding mode selection for the inter-coded B picture at frame 28 when applying a reduced IDR bit-budget compared to the unmodified rate control result. Yellow represents intra-coded blocks, while all other blocks are coded in inter mode

– due to low quantisation of the IDR pictures and high quantisation of the remaining B pictures, both the video bitrate and PSNR varies greatly over time. By reducing the IDR bit-budget, as seen in Figures (6.1b - 6.1d), this variation is reduced. It can also be seen in Figure (6.2) and Table 6.1 that a decreased level of quantisation applied to the inter-coded B pictures can increase overall rate-distortion performance. In this example, reducing the IDR bit-budget to 25% produces the smallest standard deviation of the PSNR data. However, the average PSNR is lower than the unmodified rate control result due to the fact that the IDR pictures no longer provide a high quality reference leading to greater amounts of intra-coded data within the B pictures.

This is shown in Figure (6.3) which illustrates how each variation partitions and codes the inter-coded B picture at frame 28. Note that the coding structure is shown as an overlay on top of the original picture and therefore some blocks may appear to be represented by a different shade of the same colour. Although this particular frame is coded with more bits and a lower QP after restricting the IDR bit-budget to 25%, the PSNR is still lower than the unmodified coded sequence. Figure (6.3d) clearly shows a greater number of intra-coded blocks which indicates that the preceding pictures provide poor quality reference data and therefore lowers coding efficiency.

Table 6.2: BD-PSNR [3] average savings for different reductions of the IDR bit-budget compared to the original. Calculations are made using data within the range 100kbps - 5Mbps. Results show that better performance can be achieved by applying better allocation of bits within each GOP for the majority of the test sequences

IDR Budget:	25%	50%	75%
BlueSky	10.3%	25.7%	5.5%
Tennis	11.5%	2.8%	-0.9%
Station2	-3.1%	7.5%	5.7%
ParkJoy	52.1%	24.0%	3.1%
InToTree	24.0%	18.2%	7.9%
OldTownCross	6.1%	13.3%	5.2%
Squirrel	-26.9%	-13.5%	-3.75%
BricksLeaves	-11.4%	2.7%	2.1%
ReadySetGo	3.2%	3.9%	1.94%
Jockey	11.6%	5.6%	2.9%

6.1.3 Performance After Bit-Budget Modifications

Figure (6.4) contains rate-distortion curves for each of the test sequences, detailed in Appendix-A, that show the overall average result of lowering the IDR bit-budget by a percentage of the original budget. These results are not optimal but demonstrate that at low bitrates the rate control algorithm does not perform well. Although, this is not always the case; for the sequences *Squirrel* and *BricksLeaves*, results show that for a wide range of bitrates the unmodified rate control result produces the greatest rate-distortion performance. This is because these sequences have very high temporal correlation and therefore the assigned high IDR bit-budget provides a more optimal result. Using the unmodified rate control algorithm at these target bitrates may result in low performance when applying the Adaptive Resolution coding technique due to the fact that greater intra-coding efficiency can be achieved given a low bit-budget, as shown in Chapter 5. For most of these sequences and for a wide range of target bitrates, better results can be achieved by lowering the IDR bit-budget. Table 6.2 provides details of the BD-PSNR average bitrate savings for IDR bit-budget reductions calculated as a percentage of the original set by the rate control algorithm. These results indicate in some cases that huge savings are possible given a better regulated bitrate and better allocation of bits within each GOP. For the purpose of demonstrating the performance of the Adaptive Resolution coding technique, in the following sections the best bit-budget for each target bitrate will be used as defined in these results. It should be noted that, apart from assigning the IDR bit-budget which is a rate control problem and beyond the scope of this thesis, all other processes are performed in real-time to the extent that decisions are made without prior coding or knowledge of future frames.

6.1. RATE CONTROL MODIFICATIONS

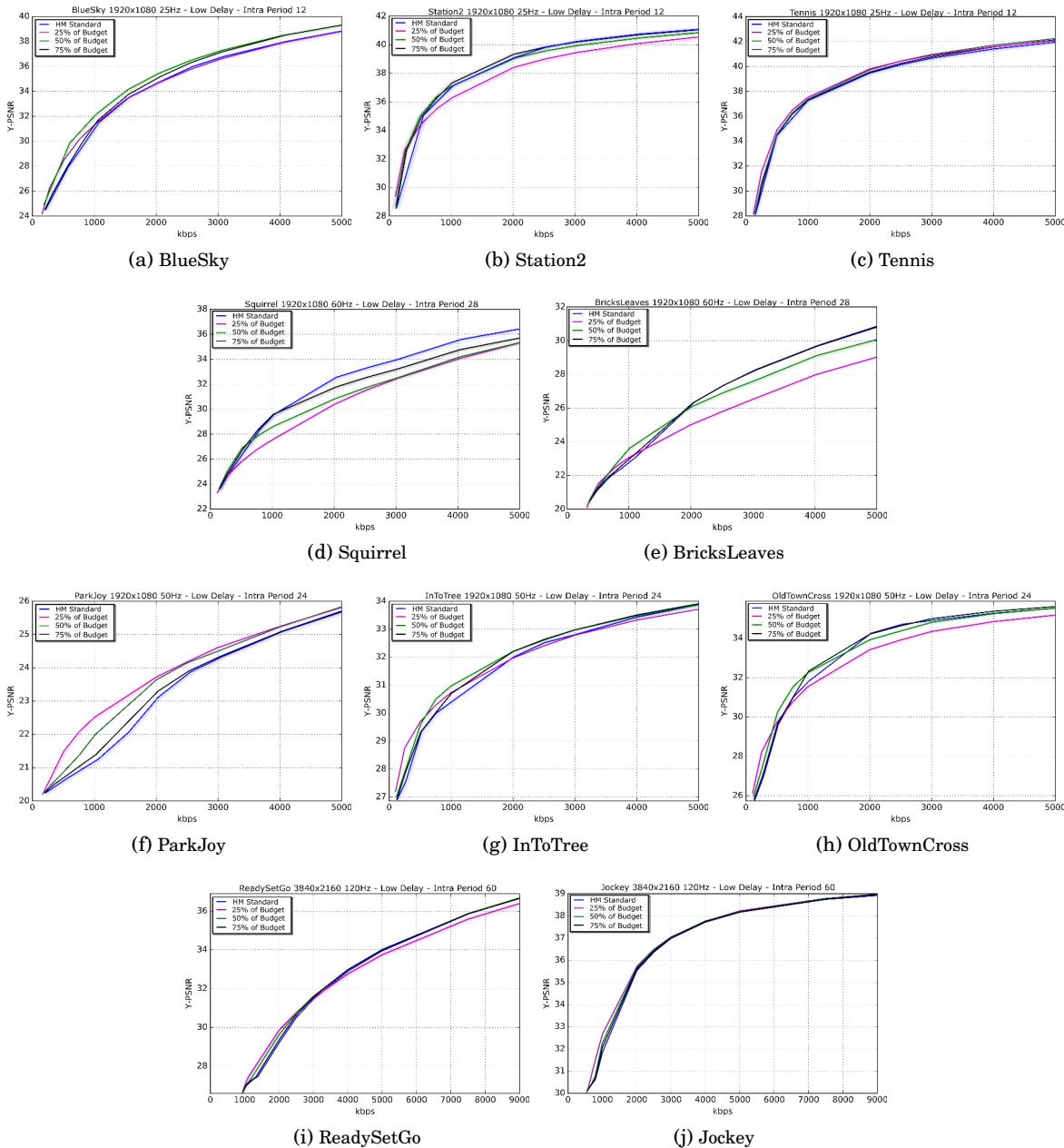


Figure 6.4: The affect on the average rate-distortion performance after reducing the IDR bit-budget to a percentage of the budget originally set by the rate control algorithm. At low bitrates the unmodified algorithm generally performs badly by assigning large IDR bit-budgets – better performance can be achieved by reducing this budget. As the bitrate increases, the rate control algorithm operates more optimally

6.2 Adaptive Resolution of IDR Pictures

Using the Adaptive Resolution Intra R-Q model described in Chapter 4 will not only enhance the rate-distortion performance of IDR pictures themselves but also enable better prediction and provide better reference pictures for inter-coded frames.

6.2.1 Performance Factors

The level of performance achieved by the Adaptive Resolution coding technique depends on a number of factors:

- **Intra Period** - The shorter the intra period, and therefore the more frequent IDR pictures are coded, the greater the benefit Adaptive Resolution coding will provide compared to standard HEVC. This is because as the intra period increases, Adaptive Resolution is applied to fewer frames and although B pictures can benefit from higher quality IDR pictures, content may change over time thus reducing the IDR pictures usefulness as a reference.
- **IDR Bit-Budget** - As rate control sets a high bit-budget for IDR pictures, coding distortion at the original HR is relatively low compared to the remaining pictures and therefore reducing the spatial resolution may not provide any benefit. This is not always optimal, as discussed in Section 6.1. As the optimal bit-budget for IDR pictures is often lower than the budget determined by rate control at low bitrates, the Adaptive Resolution coding technique is more likely to select a lower spatial resolution in order to alleviate some of the coding distortion introduced at the original HR.
- **Resampling Technique** - The distortion introduced as a result of downsampling, coding and then upsampling, will not only affect the rate-distortion performance of IDR pictures but also the remaining inter-coded B pictures. Downsampling will also have an affect on coding. The IDID technique, as described in Chapter 5, introduces additional high frequency information compared to the initial downsampled picture which may reduce coding efficiency. A resampling technique that minimises overall distortion may not contain the best results throughout the entire picture. It is possible that some areas of a picture would be better preserved using an alternative resampling method and if these areas are used for prediction then the remaining B pictures could be coded more efficiently.
- **Video Content** - the information contained within the original video sequence has a huge impact on the level of compression performance achieved by a codec. The proposed Adaptive Resolution coding technique targets IDR pictures only and therefore if the content contained within the IDR picture largely differs from the content contained within the B pictures to be coded, the only frames that will benefit from the technique will be the IDR pictures

Table 6.3: BD-PSNR average bitrate savings against standard HEVC for the best fixed bit-budget reduction for each target bitrate, and the proposed Adaptive Resolution coding technique using the same bit-budget reduction – results given within the range 100kbps-5Mbps. Results show that the majority of the increased performance gained is a result of greater efficiency of bit allocation for coding frames within each GOP. However, developing a rate control algorithm that is capable of producing these same performance gains in real-time is a challenging problem. Adaptive Resolution coding is capable of providing performance gains given that the overall distortion of an intra-coded picture can be reduced by coding at a lower spatial resolution. As shown in Chapter 5, this occurs more often when coding at the original HR generates high levels of coding distortion due to a low IDR bit-budget

	Reduced Budget	Adaptive Resolution	Gain
BlueSky	26.0%	28.2%	+2.2%
Tennis	12.1%	14.3%	+2.2%
Station2	16.8%	18.0%	+1.2%
ParkJoy	54.6%	60.5%	+5.9%
InToTree	32.3%	39.2%	+6.9%
OldTownCross	26.7%	33.2%	+6.5%
Squirrel	1.2%	13.7%	+11.5%
BricksLeaves	6.8%	16.7%	+9.9%
ReadySetGo	5.6%	9.4%	+3.8%
Jockey	11.7%	14.3%	+2.6%

themselves. This may be the case for sequences coded with a low frame rate and a relatively fast moving camera. Alternatively, if a sequence is captured using a static camera and a high frame rate then the information contained within each IDR picture will greatly benefit the compression efficiency of the inter-coded B pictures by providing a large portion of the information required.

6.3 Experimental Results

The results presented in this section are produced using an intra period equal to roughly half of the frame rate, the low delay coding configuration, and fixed reductions of the IDR bit-budget as described in Section 6.1 – at each target bitrate the best reduction is selected (25%, 50%, 75% or 100% of the original budget determined by the rate control algorithm). For the proposed Adaptive Resolution coding technique each IDR picture is coded at 3 additional low resolutions: 736×432 , 1280×720 and 1536×864 for HD sequences and 1536×864 , 1920×1080 and 2560×1440 for UHD sequences.

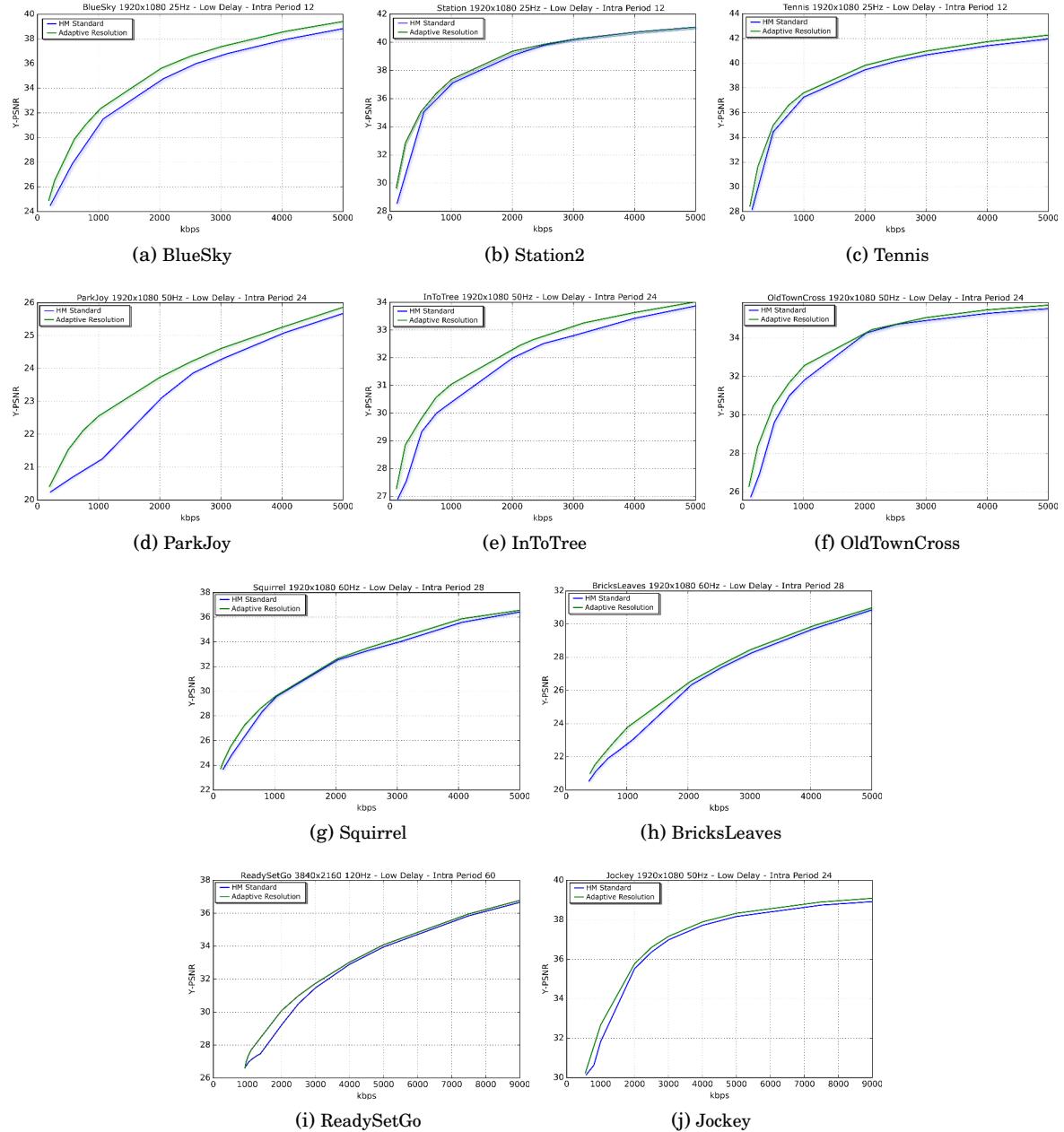


Figure 6.5: Rate-distortion performance using the best fixed bit-budget reduction for each target bitrate and the Adaptive Resolution coding technique to enhance IDR pictures. For each sequence, a considerable increase in performance can be achieved compared to standard HEVC coding

6.3.1 Average Rate-Disotrtion Performance

Table 6.3 contains BD-PSNR average bitrate savings for standard HEVC coding and the proposed Adaptive Resolution coding technique, both with modified IDR bit-budgets, and Figure (6.5) shows corresponding rate distortion performance plotted for each sequence. Results show a vast improvement compared to standard HEVC with the unmodified rate control algorithm. Reducing the bit-budget of IDR pictures not only enables more effective use of coding resources but also provides the Adaptive Resolution coding technique with a greater probability of improving the efficiency of IDR pictures by coding at lower spatial resolutions in order to reduce overall distortion. For all tested sequences it has been shown that rate-distortion performance can be improved by reducing the bit-budget of IDR pictures, although some more than others. Determining a more optimal budget for each IDR picture independently should provide even greater efficiency gains. However, developing a rate control algorithm that is capable of producing higher performance gains by more effective bit allocation in real-time is a challenging problem. In each case the proposed Adaptive Resolution coding technique adds additional coding efficiency but the amount gained is highly content dependent – results show that an increase of up to 11.5% is achievable. Greater performance gains are possible given that the overall distortion of an intra-coded picture can be reduced by coding at a lower spatial resolution. As shown in Chapter 5, this occurs more often for low bit-budgets as coding at the original HR generates high levels of coding distortion.

Sequences *BlueSky*, *Tennis* and *Station2* were captured using a temporal sampling frequency of 25Hz (frames per second). Temporal content within *Station2* changes very quickly due to camera zoom and *BlueSky* and *Tennis* were captured while subject to camera panning. As a result, the usefulness of the information contained within each IDR picture when coding the remaining B pictures within the GOP is reduced. From the intra-coding results given in Chapter 5 we can see that spatial content will also impact the overall performance gained. While the IDR pictures within the sequences *BlueSky* and *Tennis* greatly benefit from spatial resampling using the IDID technique, for the sequence *Station2* this benefit is relatively small. When applied within an inter-coding configuration, the performance of the Adaptive Resolution coding technique ultimately depends on the rate-distortion benefits of applying intra-coding at lower spatial resolutions. If there is little or no gain in performance, inter-coding will provide similar results as standard HEVC.

The two sequences that benefit most from the proposed technique are *Squirrel* and *BrickLeaves*. These sequences contain very high temporal correlation due to being captured with a fixed camera and a high temporal sampling rate (60Hz). For a wide range of target bitrates the original bit-budget provides the best results, as shown in Section 6.1. This means that a high bit-budget is assigned which can reduce the efficiency of the coded IDR pictures. However, as shown in Chapter 5, coding efficiency of these intra-coded pictures can be increased for a wide

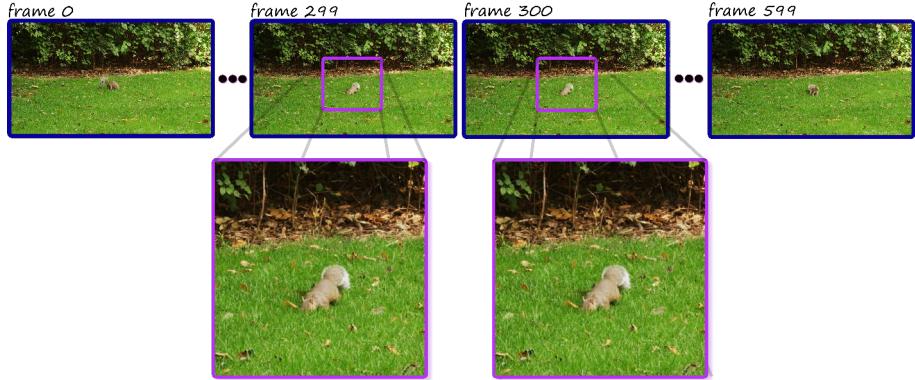


Figure 6.6: Illustration of high temporal correlation within the sequence *Squirrel*. This sequence produces very little residual energy between frames due to being captured with a fixed camera and a high frame rate. Within the scene itself there is very little movement; the squirrel is really the only part of the sequence that exhibits change from one frame to the next but its movements are slow relative to the frame rate

range of bit-budgets by applying the IDID sampling technique. Compared to other sequences, resampling the IDR pictures may only provide a small increase in coding efficiency when only considering the IDR pictures themselves, but a greater amount of this higher quality data is applied to the remaining inter-coded pictures without the need for coding large number of motion vectors and residual information.

6.3.2 Temporal Variation of Quality & Bitrate

To gain a better understanding of the principles behind why the Adaptive Resolution coding technique performs well for some sequences but not others, we have to consider more than just the average rate-distortion performance. According to the associating paper [52], the BVI sequence *Squirrel* has one of the lowest TI indexes of the whole database. This result can also be seen from the TI index given in Section 2.6. This means that very little motion energy is present within the sequence, resulting in very high temporal correlation between frames. This is illustrated in Figure (6.6) which contains thumbnails of 4 frames – the first and last frame, and two consecutive frames positioned in the middle of the sequence. It can be seen that there is very little change from one frame to the next. Given that the quality of an IDR picture can be enhanced via spatial resampling and that it will provide a large portion of the information required for coding the following inter-predicted pictures, all frames within the GOP will therefore benefit.

Figures (6.7 - 6.10) compares temporal variation of PSNR and SSIM image quality along with the instantaneous bitrate between standard HEVC coding and the proposed Adaptive Resolution coding technique. The instantaneous bitrate gives indication of bitrate at any point in time. To produce a smooth representation of the instantaneous bitrate, equation 6.1 calculates an average

6.3. EXPERIMENTAL RESULTS

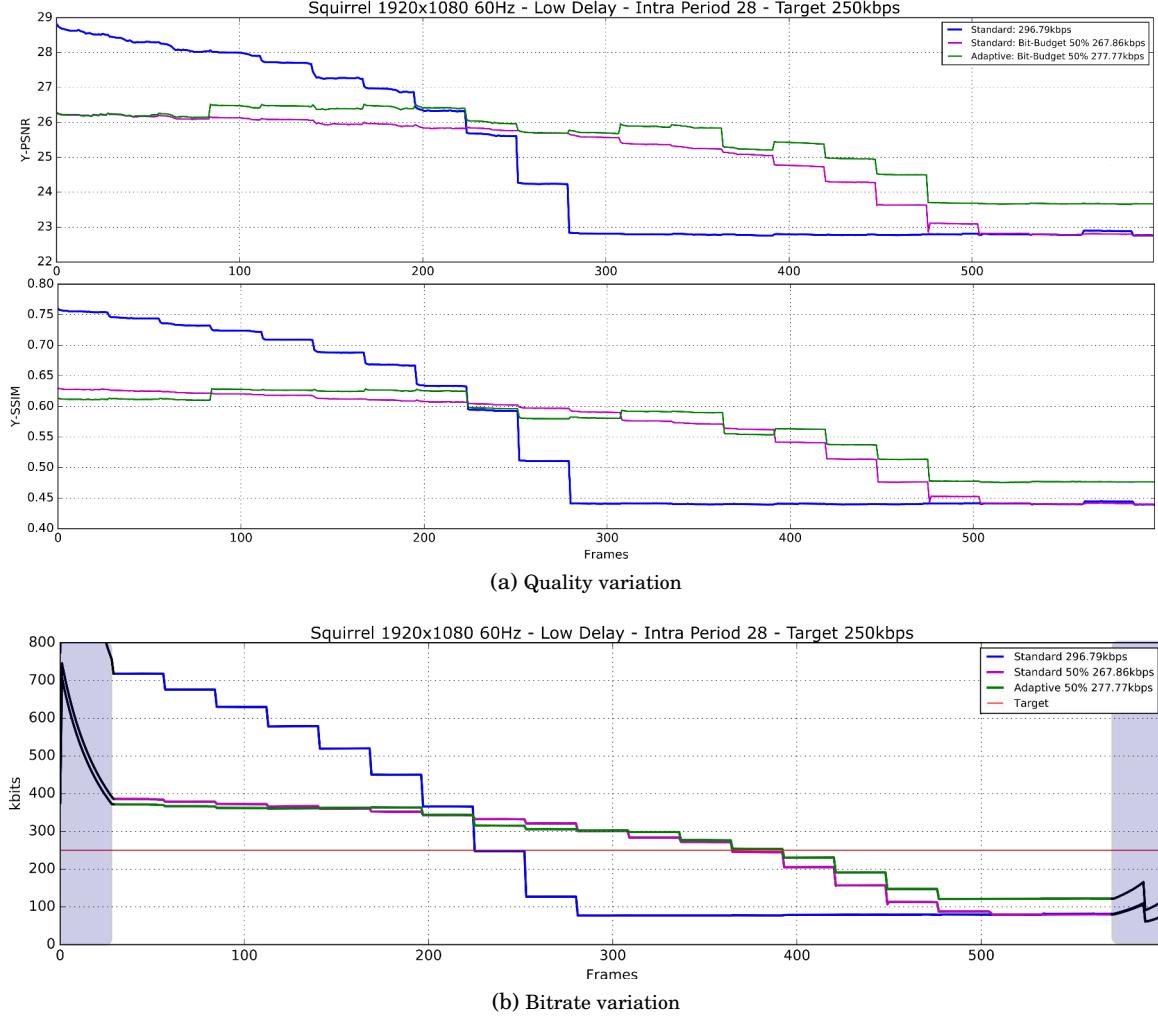


Figure 6.7: Sequence: *Squirrel* 1920 × 1080 60Hz Low Delay Intra Period 28 Target 250kbps. Temporal variation of quality and instantaneous bitrate. For the first section of the sequence, the unmodified rate control algorithm sets a bit-budget for each GOP such that the instantaneous bitrate is far greater than the target. This is rectified by significantly reducing this budget over time in order to produce an average bitrate closer to the target. Reducing the IDR bit-budget increases performance by producing a more consistent bitstream in terms of quality and bitrate variation – although for a large portion of the sequence, the instantaneous bitrate is still greater than the target. The Adaptive Resolution coding technique generally provides an increased level of quality for the same or similar bitrate. Although an IDR bit-budget of 50% produces a high average level of rate-distortion performance, these results suggest that a more optimal budget would be somewhere between 25 – 50%. Note that, given equation (6.1), there is insufficient data within the shaded areas of graph (b) and therefore the results at these points are not accurate

over a time window centred over each instance of time for which a frame is displayed. The window is set according to the intra period such that 2 IDR pictures and $(2 \times \text{intra period}) - 2$ inter-coded pictures are taken into account. This is so that large spikes during IDR pictures are averaged out.

$$(6.1) \quad R_k = \frac{f}{2P} \sum_{i=-P}^{P-1} (x_{k+i}),$$

where R is the Instantaneous Bitrate at the time of frame k , f is the temporal sampling frequency measured in frames per second, and P is the intra period.

Figure (6.4d) indicates that on average the rate control algorithm performs well. Above a target bitrate of 1000 kbps, best results can be achieved with no reduction of the IDR bit-budget – this is at least true compared to fixed reductions of 25%, 50% or 75% of the original budget. It is possible that better overall performance can be achieved with a slight reduction, or possibly a slight increase, of the IDR bit-budget. However, further modification to the rate control algorithm is beyond the scope of this thesis. The reason why rate control performs well was initially discussed in Section 6.2.1; due to high temporal correlation between frames within the sequence *Squirrel*, a high bit-budget will enable coding of a high quality IDR picture to be used as a reference for inter-coded pictures. These inter-coded pictures can be coded very efficiently as they will require very few data bits due to the fact that the majority of each frame can simply be coded in skip mode or with very little information in the form of a coded residual and motion vectors. To obtain high coding efficiency using the proposed Adaptive Resolution coding technique it is only necessary that the temporal correlation within each GOP is high; a scene change that occurs at the time of coding an IDR picture will not affect performance as pictures coded prior can not be referenced. It is therefore more important that the sequence is captured at a high frame rate and that the intra period is short.

At very low bitrates, some reduction of the IDR bit-budget can produce more desirable results. Figure (6.7a) shows variation of PSNR and SSIM when coding with a target bitrate of 250 kbps, and Figure (6.7b) shows corresponding variation of the instantaneous bitrate. Initial quality results show no real benefit of the Adaptive Resolution coding technique and according to the SSIM metric, quality is in fact worse than standard HEVC coding using the same bit-budget reduction. As selection of the coded IDR picture resolution is based on the PSNR metric, it is understandable that the SSIM result does not always favour the proposed technique. However, observing the instantaneous bitrate at these points in time reveals that greater efficiency is in fact achieved as a lower bitrate produces a similar level of quality. At frame 84 the selected IDR picture provides a higher quality image and as a result the successive inter-coded pictures also benefit, thus providing a fairly consistent level of quality for the entire GOP. The bitrate at this point is still lower than standard coding using the same reduced bit-budget, indicating much greater rate-

6.3. EXPERIMENTAL RESULTS

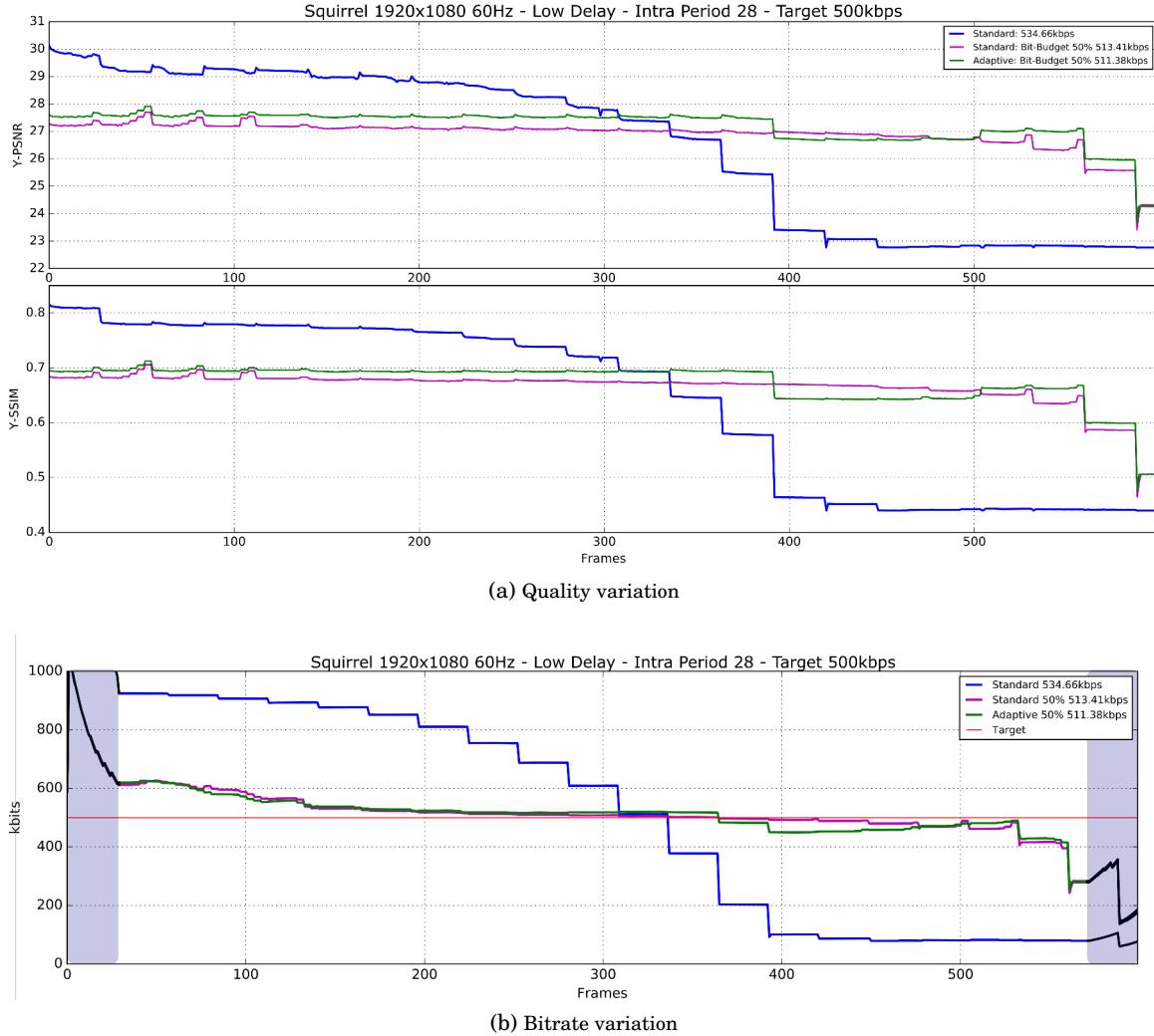


Figure 6.8: Sequence: *Squirrel* 1920 × 1080 60Hz Low Delay Intra Period 28 Target 500kbps. Temporal variation of quality and instantaneous bitrate. Similar to the result shown in Figure 6.7b, the unmodified rate control algorithm initially sets a bit-budget for each GOP such that the instantaneous bitrate is far greater than the target and as a result it is required to drop this budget significantly over time to produce an average rate closer to the target. A bit-budget reduction of 50% has a more desirable effect as the coded bitstream conforms much closer to the target for a large portion of the sequence. The Adaptive Resolution coding technique improves quality performance for almost the entirety of the sequence except for a short period around frame 400 when the bitrate decreases – the IDR selection does not take into account the bitrate during the previous GOP. Note that, given equation (6.1), there is insufficient data within the shaded areas of graph (b) and therefore the results at these points are not accurate

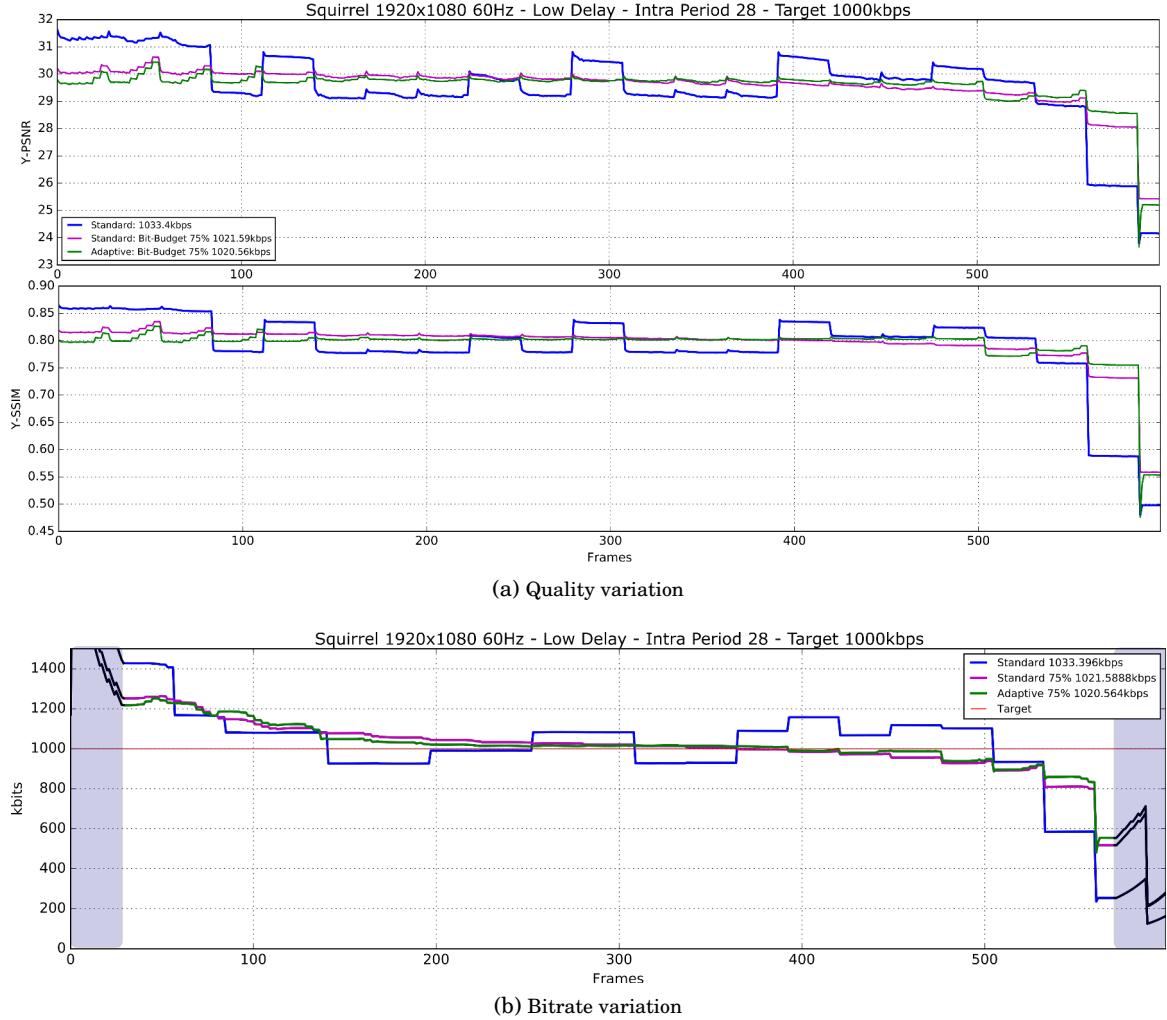


Figure 6.9: Sequence: *Squirrel* 1920×1080 60Hz Low Delay Intra Period 28 Target 1000kbps. Temporal variation of quality and instantaneous bitrate. As the target bitrate increases, the unmodified rate control algorithm does a much better job at conforming to this target. This is partly due to the fact that the IDR bit-budget selected is closer to optimal. Therefore, it is understandable that 75% of the original bit-budget was found to produce better rate-distortion performance when compared to coding with lower target bitrates. Quality results show that initially the Adaptive Resolution coding technique performs worse than standard coding using the same bit-budget. This is because a resolution is selected that requires fewer bits to code but it may also be the case that the original resolution is the more optimal choice given the budget allocated. The Adaptive Resolution coding technique still manages to increase quality performance during some sections of the sequence but overall there is little improvement when targeting this bitrate. Note that, given equation (6.1), there is insufficient data within the shaded areas of graph (b) and therefore the results at these points are not accurate

distortion performance. It can be seen that, the unmodified rate control result causes quality to drop significantly throughout the sequence. This is due to the fact that over time the rate control algorithm compensates for prior inaccurate predictions regarding the bit-budget for previously coded GOPs. The initial bitrate far exceeds the target and so to produce a closer average bitrate, the remainder of the sequence is coded at a very low bitrate. For video transmission applications, the average video bitrate is of little importance; if the instantaneous bitrate is greater than the channel capacity, a number of frames at this point will be skipped and reducing the video bitrate later in the sequence will only result in further inefficient use of the channel buffer. The rate control algorithm is clearly designed with data storage in mind and not transmission. For applications transmitting video in real-time, only the instantaneous bitrate is important and rate control should aim to match the bitrate capacity of the channel at all times. Reducing the IDR bit-budget generates a more consistent bitrate but a large portion of the sequence is still coded at a rate greater than the target, resulting in over compression of the remaining sequence.

Figure (6.8), which represents the target bitrate 500kbps, shows similar patterns as seen in Figure (6.7). For the standard and unmodified rate control result, bitrate and quality begins high and ends low, and much more consistency can be achieved with a reduced IDR bit-budget. The only time the quality of the Adaptive Resolution coding result drops below the standard coding result with the same budget is when there is a drop in bitrate. This is caused by the availability of an alternative IDR picture coded with fewer bits but greater efficiency. As an improvement to the proposed technique it would be beneficial to take into account the bitrate prior to coding the current IDR picture to try and ensure a greater level of consistency whilst not largely affecting rate-distortion performance.

In Figure (6.9) better performance can be seen when targeting 1000kbps using the unmodified rate control algorithm compared to lower target bitrates. It is clear that as the target bitrate increases, the video bitrate is much more consistent. However, there are still large variations in the number of bits allocated to each GOP and therefore fluctuations in quality are somewhat inevitable given the temporal consistency of content within this sequence. In terms of average rate-distortion, a bit-budget reduction of 75% of the original budget provides the best overall performance. This budget also provides a more consistent level of quality and of the video bitrate. At this target bitrate the Adaptive Resolution coding technique does not show any obvious benefit and on average it shows similar performance as standard coding with the same budget.

At higher target bitrates, such as 4000kbps as shown in Figure (6.10), the allocation of bits within each GOP appears to be much more optimal. None of the tested reductions in the IDR bit-budget show any increase in performance. For this reason, the original bit-budget set by the rate control algorithm is used and applied to the Adaptive Resolution coding technique. However,

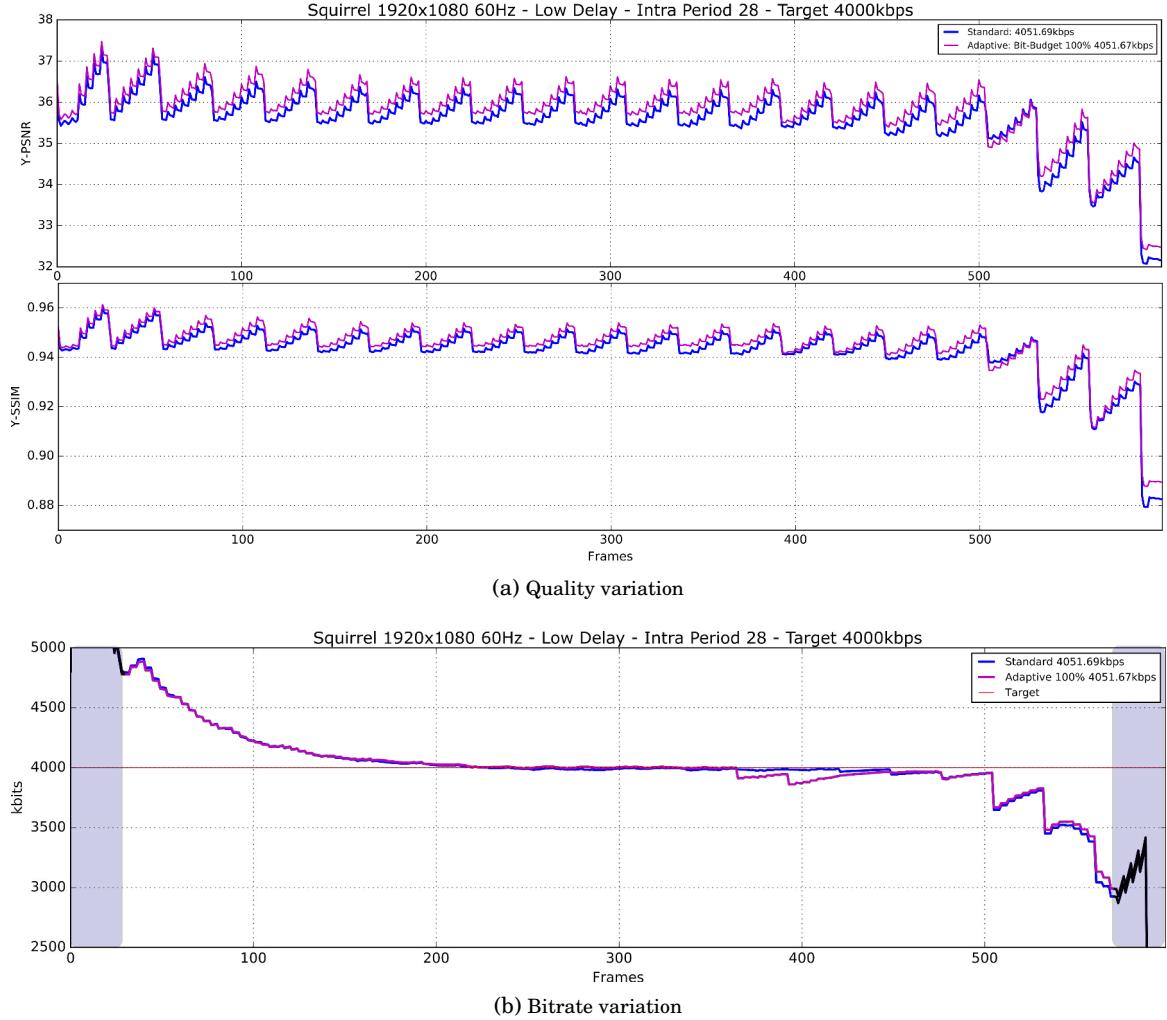


Figure 6.10: Sequence: *Squirrel* 1920 × 1080 60Hz Low Delay Intra Period 28 Target 4000kbps. Temporal variation of quality and instantaneous bitrate. At this target bitrate the IDR bit-budget is found to produce the best results, at least when compared to the fixed reductions of 25%, 50% or 75% of the original budget. Initially the instantaneous bitrate is far too high but over time the rate control algorithm adapts and produces a more consistent bitrate that conforms to the target. Almost throughout the entire sequence the Adaptive Resolution coding technique provides greater rate-distortion performance. However, despite a fairly consistent bitrate, the quality of both standard and the proposed coding methods fluctuates and produces what looks like a ‘staircase’ pattern. The reason for these fluctuations is because while coding the remaining frames within each GOP, some inter-coded frames are coded with a large amount of intra-coded data that improves the quality at those spatial regions within the sequence. This means that over time, better reference pictures are available and there are still enough bits remaining that were allocated to the GOP to improve other regions of the frame. This suggests that the IDR bit-budget set by the rate control algorithm is now too low and increasing this budget may produce a more consistent level of quality. Note that, given equation (6.1), there is insufficient data within the shaded areas of graph (b) and therefore the results at these points are not accurate

although the instantaneous bitrate result shows only gradual changes, in actual fact there is a great deal of fluctuation within each GOP. The result of equation 6.1 fails to illustrate this but is still useful for showing the general trend of how the bitrate varies over time. At lower bitrates, the quality of the GOP is somewhat restricted by the quality of the initial reference picture as it is coded as an IDR picture and the majority of the information in the succeeding frames are coded in inter mode. In this case, even though the IDR bit-budget has not been restricted, a large number of bits are still available to the GOP after coding the IDR picture that areas of scene are further enhanced with the addition of intra coded blocks. These blocks are coded using a lower level of quantisation and therefore improve upon the IDR picture as a reference. These results suggest that reduced variation in quality, and possibly an increase in the overall average level of quality, could be achieved by increasing the IDR bit-budget rather than reducing it at this target. It should be emphasised that this is only true for this sequence given the coding configuration and channel constraints – the optimal IDR bit-budget depends on a number of factors, as discussed in Section 6.2.1.

6.3.3 Percentage of Selected Resolutions

The IDR picture selected for each GOP will have an affect on all future coded frames. Although at each resolution the target bit-budget remains the same, each available version of the IDR picture will be coded with a different number of bits and provide a different level of quality. For this reason the decision made will determine the number of bits remaining for inter-coded pictures and in turn will likely alter the number of bits allocated to the following GOP. The Adaptive Resolution R-Q model, as described in Chapter 4, makes decisions purely based on the available bit-budget and the level of rate-distortion performance that each variation of the picture provides. It also aims to limit the selection to those that conform to a set mismatch ratio and therefore helps to maintain a well regulated bitrate.

Figure (6.11) gives the percentage of resolutions selected for each sequence and each target bitrate. Finding a pattern amongst these figures would be difficult as there are many factors that influence this decision. Firstly, the IDR bit-budget may have been altered for each target bitrate depending on whether or not better average rate-distortion performance could be achieved by reducing this budget by a fixed percentage. It was found in Chapter 5 that generally for low bit-budgets of intra-coded pictures a low spatial resolution can provide greater performance and for higher bit-budgets a higher spatial resolution would be more optimal. When coding in inter mode, a lower target bitrate does not necessarily mean that a lower IDR bit-budget would be applied, this depends on the rate control algorithm and how well it performs. Another factor affecting this decision is the ability to predict accurately the best QP for each spatial resolution variation of the frame. Even if a particular scale factor and QP were to provide the optimal reconstructed picture for the given frame and bit-budget, unless that QP is predicted accurately,

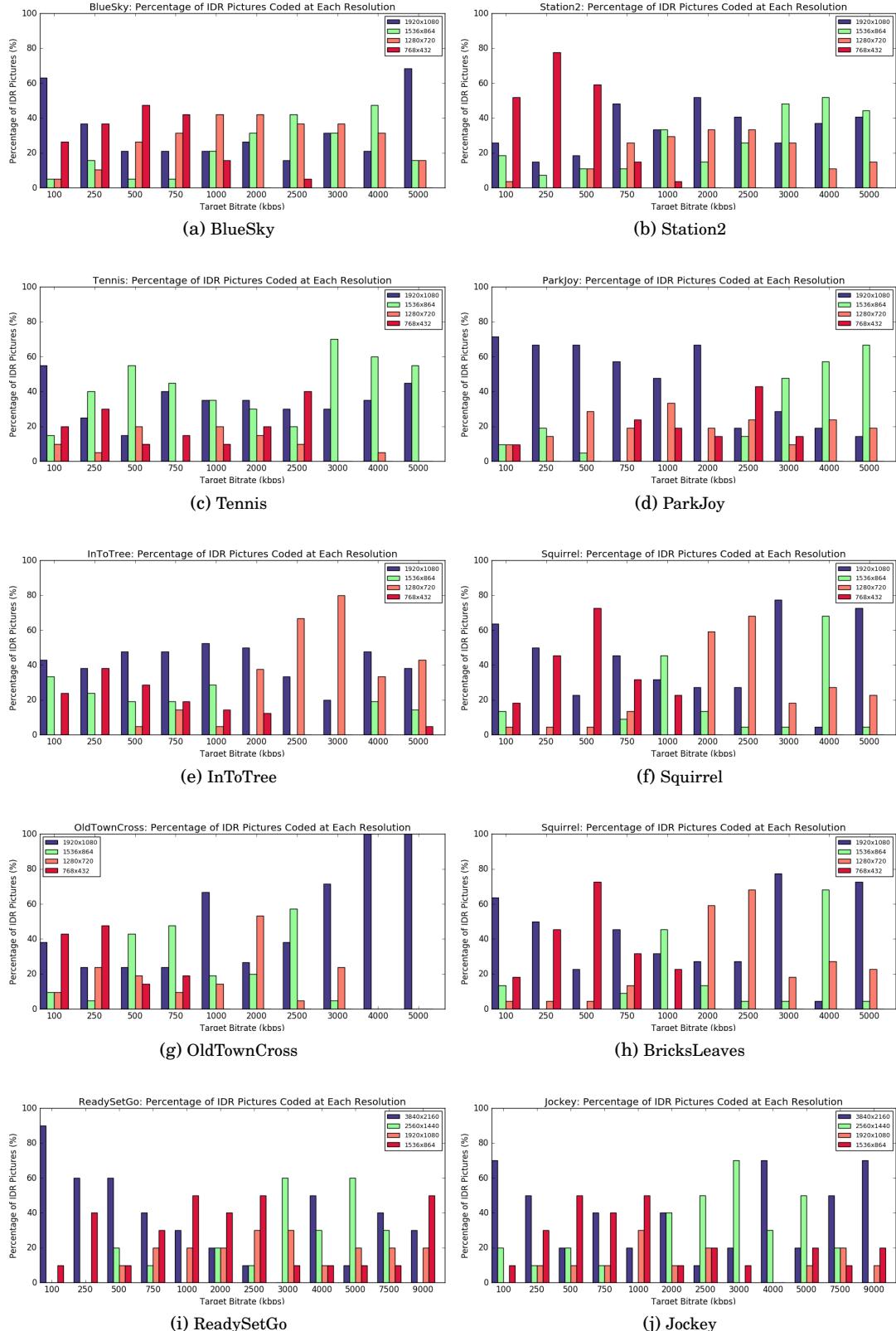


Figure 6.11: Percentage of each spatial resolution selected per sequence using the Adaptive Resolution coding technique for IDR pictures

an alternative scaled picture may be selected instead. A third factor influencing this decision is the challenge of determining the best IDR picture given the number of coded bits and the corresponding level of quality. As explained in Chapter 4, in some cases determining the best reconstructed picture is simple but in other cases it is not always possible to determine with absolute certainty.

What can be said about the data presented in Figure (6.11) is that the Adaptive Resolution coding technique selects lower spatial resolutions for the majority of the tested target bitrates and sequences. Even if a small number of IDR pictures selected were coded at lower spatial resolutions, a large benefit could potentially be gained if these pictures also enabled greater efficiency of inter-coded pictures. As it is expected that on average the selected IDR resolution would increase for increasing target bitrates, these results can also indicate how effectively the RDO algorithms are performing – namely the Adaptive Resolution R-Q model and the IDR bit-budget allocation determined by rate control. When targeting bitrates between 100kbps-2000kbps for the sequence *ParkJoy*, the IDR bit-budget is reduced to 25% of the original budget. It is therefore expected that the Adaptive Resolution R-Q model would select low resolutions. However, it can be seen from Figure (6.11d) that, within this bitrate range, the most selected resolution is the original HR. One possible reason for this could be that the R-Q model frequently failed to find more efficiently coded pictures and therefore chose the ones that minimised the mismatch ratio, which happened to be the pictures coded at the original HR. This seems unlikely because as the resolution decreases, each quantisation step would produce a smaller difference in the number of coded bits generated and therefore the mismatch ratio should generally be lower for lower resolutions. It is also possible that even though the average video bitrate matches the target, the bit-budget for each IDR picture could vary greatly from the first frame to the last such that initially each IDR picture is coded given a large bit-budget, and therefore coded at a high resolution, and the later pictures are coded given a very low bit-budget. This behaviour can be seen in Figure (6.7) and Figure (6.8) from the unmodified rate control result. Chapter 4 provides results of using the Adaptive Resolution R-Q model in intra mode and it is also highlighted that the model would be less effective when used in inter mode with a long intra period. This is because the correlation between consecutive intra-coded pictures would be decreased, given content that varies over time, and therefore the accuracy of predicting the optimal QP would likely reduce. This could explain why the selected resolution does not always conform to what is expected but more research is required to analyse the performance of the Adaptive Resolution R-Q when applied to inter-coding. It would also be beneficial to remove the influence of rate control, either by using a fixed QP or by fixing the bit allocation for each method. This is beyond the scope of this thesis but discussed in more detail in the further work section of Chapter 8.

6.4 Chapter Summary

This chapter combines the work and findings from all previous chapters to demonstrate the result of the proposed Adaptive Resolution coding technique using independent IDR bit-budgets, the Adaptive Resolution R-Q model as defined in Chapter 4, and by applying the IDID resampling technique described in Chapter 5 to produce LR pictures optimised for interpolation using the SHVC sampling functions. As the rate control problem is beyond the scope of this thesis, the IDR bit-budget is set offline after exhaustedly coding a set of defined reductions of the bit-budget set by the rate control algorithm within the HM reference software. Although these reductions in the IDR bit-budget do not represent optimal results, they demonstrate that better performance can be achieved and also increase the probability of selecting a lower spatial resolution for each IDR picture.

By applying the Adaptive Resolution coding technique, further gains in rate-distortion performance can be achieved. It was found that sequences of high temporal correlation benefit more from the proposed technique. This is because any gain in coding efficiency obtained after coding each IDR will be carried over to the remaining inter-coded pictures. Very few bits are required to code the remainder of the GOP whilst maintaining the level of quality set by the IDR picture. Sequences of low temporal correlation, due to a low sampling rate relative to motion within the scene, mostly only benefit as a result of higher coding efficiency for each IDR picture. As only a small fraction of frames within each sequence are coded as an IDR picture, on average only a small benefit can be observed. Results show that performance gains of up to 11.5% can be achieved for HD sequences within the range 100kbps-5Mbps.

Analysis of temporal variation of quality and of the bitrate for the sequence *Squirrel* at low target bitrates shows that a lower IDR bit-budget can improve rate-distortion performance for standard HEVC and also provide a more consistent level of quality and a much better regulated bitrate. Relatively, the proposed Adaptive Resolution coding technique provides a small increase in performance but to develop a rate control algorithm that can achieve these performance gains is essentially an impossible task without knowledge of how video content changes for future frames. Another issue observed is the fact that the rate control algorithm is not designed for real-time video transmission. Each sequence is coded with the aim of providing an average bitrate that closely conforms to the set target. When transmitting over a communications channel with finite capacity only the instantaneous bitrate is of concern. Regardless of how well the rate control algorithm was able to regulate the bitrate for the preceding GOPs, the target bitrate should always be matched.

The results presented in this chapter highlight some key strengths and weaknesses of the proposed technique. In some cases the additional complexity may not be economical given the low level of coding efficiency gained. However, given knowledge that the sequence to be coded will

contain a high level of temporal correlation, the proposed technique will likely provide a much greater level of coding efficiency. As a further improvement and enhancement to this technique, it would be beneficial to also resample intra-coded data (if any) within the inter-coded pictures. This would further differentiate the technique from the standard but could potentially increase coding efficiency. Additionally, given that for the proposed application video is captured from a UAV, it may be possible to make calculated predictions regarding the temporal correlation of future frames using the onboard motion sensors and also guidance information from the UAV's controller. This would enable the rate control algorithm to define the IDR bit-budget more efficiently for each GOP. These ideas are discussed more in Chapter 8.

MIXED RESOLUTION ENHANCEMENT CODING

Allocating low bit-budgets to intra-coded pictures can often lead to high quantisation and loss of important high frequency information. Coding at lower resolutions can alleviate some of this distortion when utilising effective resampling techniques and filters that minimise the additional distortion introduced as a result of spatial resampling. However, it is not always possible to reconstruct each portion of a picture to the same or similar level of quality; as low pass filtering is applied prior to downsampling, lowering the sampling rate by a large factor can significantly increase distortion due to the loss of high frequencies. Within natural scenes frequency content tends to vary spatially. High frequency content, such as edges and textures, may only occupy small portions of the picture and therefore the selected scale factor will either cause loss of high frequencies or oversampling of low frequency content. It therefore makes sense not only to adapt resolution for each frame but also according to the spectral properties within each region

Adaptive Resolution Coding, as described in Chapter 6, aims to select the best resolution based on an average of the combine coding and resampling distortion across the entire frame. As content can vary spatially, some of these regions could potentially have been coded more efficiently at an alternative resolution. For this reason a mixed resolution coding approach would be more optimal – this is illustrated in Figure (7.1). The idea is to code low frequency content at a selected LR and enhance the areas that have been subjected to high levels of distortion, due to downsampling, at higher spatial resolutions. A similar coding approach was first decribed in Chapter 2 using SVC's or SHVC's spatial scalability functions. However, neither of these extensions were built with the intention of enhancing compression efficiency but rather to provide greater versatility for

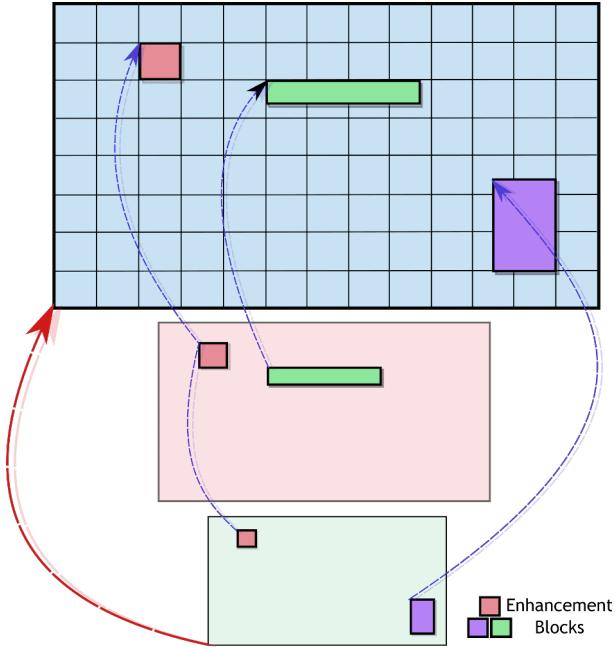


Figure 7.1: Illustration of Mixed Resolution Enhancement Coding. Low frequency content is only coded at the lowest spatial resolution, whereas high frequency content is enhanced by coding these areas at higher spatial resolutions. This can be achieved using functionality already available within the SHVC scalable extension but modifications would be required to enable adaptation of the spatial resolution for each layer according to content

the decoder. Unlike the resampling techniques described in Chapter 5, spatial scalability utilises inter-prediction to generate ELs so that receiving devices have some flexibility regarding the decoded spatial resolution. Previous work has shown that scalable coding within an inter-coding framework provides no additional performance gains [41, 72] but very little work exists demonstrating the benefits of applying ELs to intra-coded pictures. The contribution within this chapter therefore lies within the comparison of intra-coding methods and the potential performance gains that generating HR ELs may provide over simply upsampling the LR IDR picture to the original HR.

Although only intra-coded results are provided in this chapter, this technique could potentially replace the existing approach of downsampling, encoding, decoding and then upsampling. The Mixed Resolution Enhancement coding method is illustrated in Figure (7.2). With some modifications to the encoder, this coding method could also be used to enhance specified areas. In some respects this is similar to Region of Interest (ROI) coding where the QP is adapted according to a prioritisation of regions ranked in order of visual importance [47, 75]. A Mixed Resolution ROI approach for intra-coded pictures is also described in this chapter but it should not be considered as a comprehensive analysis of the technique but as an introduction and proof-of-concept.

7.1. ENHANCEMENT OF RESAMPLED INTRA-CODED PICTURES

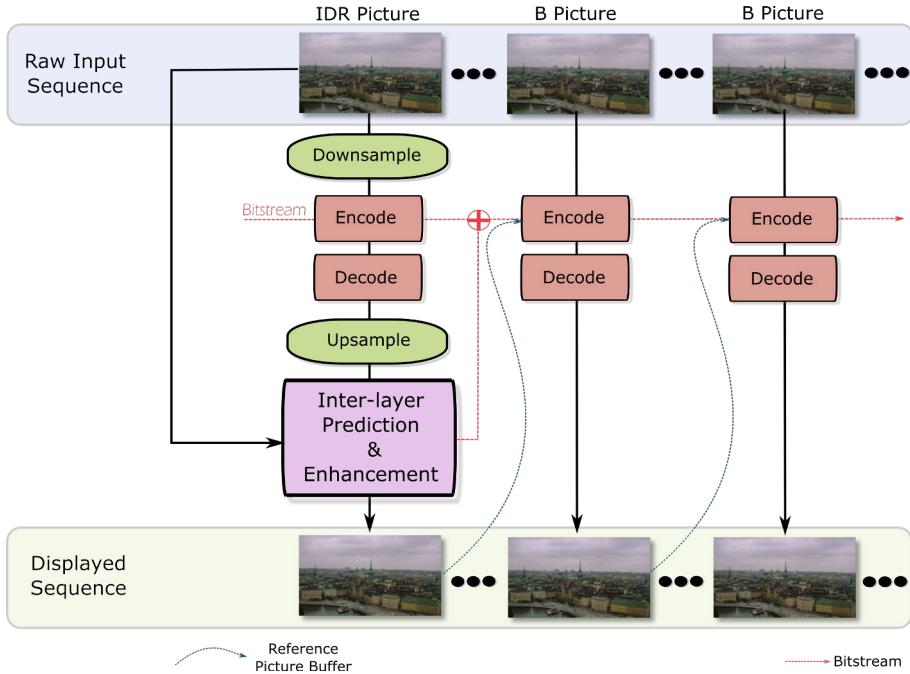


Figure 7.2: Block diagram of Mixed Resolution Enhancement Coding using two spatial layers for each IDR picture. After upsampling the decoded intra picture, inter prediction is applied to generate motion vectors and a coded residual from the original HR picture.

7.1 Enhancement of Resampled Intra-Coded Pictures

Enhancement Coding utilises inter-layer prediction, as used in SHVC, to provide an inter-coded EL that adds additional information to the decoded and upsampled picture. Figure (7.3) illustrates this process for a LR picture coded in intra mode that is iteratively upsampled and enhanced two times. Prior to encoding, the input picture is downsampled to one or more lower spatial resolutions. In SHVC the picture coded at the lowest resolution forms the BL and contains the data for which all receiving devices must at the very least decode. After upsampling to the next spatial layer, the picture is enhanced by applying the information contained within the EL. Each EL is generated by inter-prediction of the non-coded picture given the corresponding upsampled picture of the same spatial resolution.

As we are only concerned with the quality of the reconstructed picture at the original HR, any intermediate resolution data would be discarded after being applied at the decoder. For the work produced in this chapter, the number of spatial layers is limited to 2 but results are provided for a range of resolutions at the BL. Performance is affected by a number of factors:

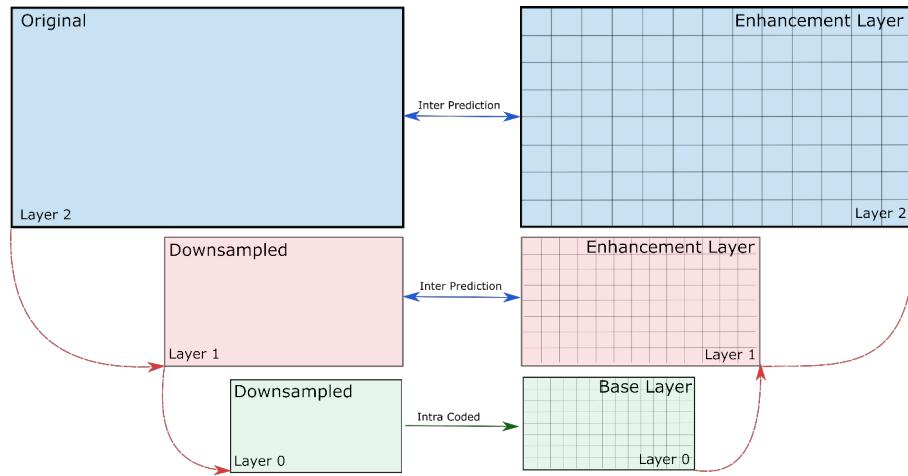


Figure 7.3: Enhancement of an intra-coded picture using inter-layer prediction. Inter-layer prediction is used to generate an EL at all possible resolutions beyond the BL. In this example the BL picture is coded in intra mode and is iteratively upsampled and enhanced until the final HR picture is reconstructed. In SHVC inter-layer prediction can also be used to enhance an inter-coded BL

- **Picture Content and BL Resolution** - High frequency information contained within the original picture will be lost as a result of downsampling. Larger scale factors will reduce the number of bits spent on coding the BL but aliasing and loss of detail could potentially increase the amount of information coded in the EL.
- **The Allocated Bit-Budget** - As coding an EL requires additional signalling overhead, given a low bit-budget, high levels of quantisation may reduce the overall efficiency compared to coding without any enhancement.

Within an EL, signalling information needs to be coded for each block. At the very least a block will be flagged as SKIP provided that it contains no residual information. Assuming a perfectly reconstructed picture after interpolation, the entire EL would be coded in SKIP mode and therefore would be completely redundant. In this case resampling alone would be more efficient. SHVC always applies an EL to the upsampled picture as its purpose is to provide greater adaptability to devices of varying specifications. However in some circumstances resampling alone, or an alternative selection of EL resolutions, could provide a greater level of rate-distortion performance. Additionally the resolution of the BL should also be adapted for each frame as described in Chapter 4 – but this is not possible with the SHVC standard.

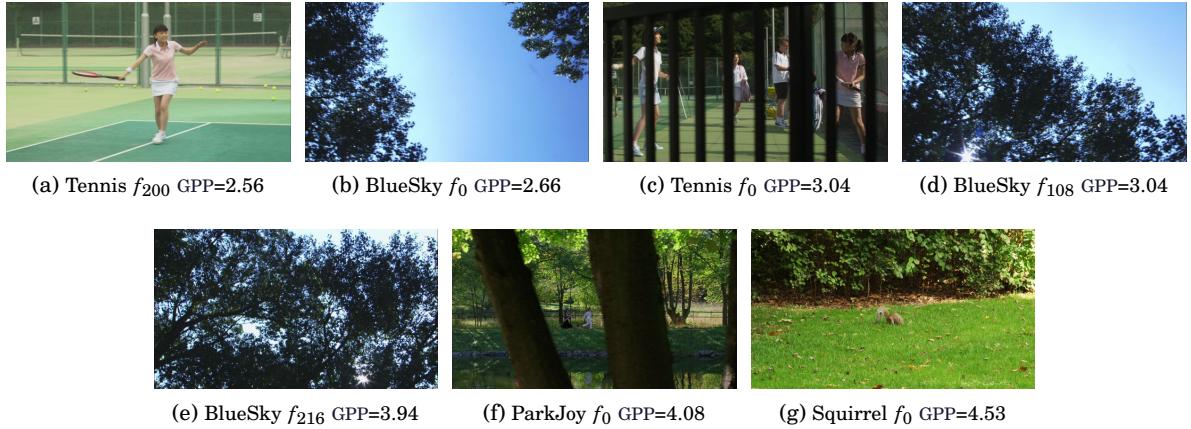


Figure 7.4: Pictures taken from 4 different sequences, used for testing the performance of intra-coding methods, ordered according to their complexity calculated by the average GPP

For the Enhancement Coding results produced in this chapter, the process can be summarised as follows:

1. Downsample each picture to each LR using the IDID technique and the SHVC filter coefficients given in Chapter 5.
2. Encode each scaled picture in intra mode.
3. Decode each picture.
4. Upsample each picture using the SHVC filter coefficients.
5. Perform inter-layer prediction between the upsampled result and the original HR non-coded picture and apply the EL.

7.2 Experimental Results

7.2.1 Test Data and Coding Parameters

To evaluate the performance of applying an EL at the original HR, 7 HD (1920×1080) input pictures are used which are taken from 4 different sequences: *Tennis*, *BlueSky*, *ParkJoy* and *Squirrel*. Each sequence is detailed in Appendix-A. In Figure (7.4) these pictures are ordered according to their spatial complexity [36], calculated by the GPP given in Chapter 4. The tested BL resolutions are 736×432 , 1280×720 and 1536×864 , and each picture is coded for a range of QPs [18, 51]. Comparisons with standard HEVC intra-coding are also provided for which each picture is coded at the original resolution 1920×1080 . For enhancement coding, each EL is coded with a QP according to an offset of the BL QP value. The tested offset values are within the range $[0, 20]$ but restricted by the following condition:

$$(7.1) \quad QP_{EL} = QP_{BL} + QP_{offset} \leq 51$$

7.2.2 Average Bit Savings

In this section coding performance is analysed with and without an EL and also compared to standard HEVC. Given all data points, the optimal rate-distortion curves are generated for each coding technique by maximising the gradient between any two points. The optimal combination of QP_{BL} and QP_{EL} depends on picture content, the target bit-budget, and the resolution of the BL. In [6, 57, 64] a fixed QP offset is applied to each EL given the QP applied at the lower spatial layer. In [33, 43] adaptive rate-quantisation methods are presented that aim to select the best QP for each layer in SVC.

Results are presented in Table 7.1 by calculating the average PSNR difference between the rate-distortion curves as determined by the BD-PSNR metric [3]. A maximum bit-budget is set according to the largest number of coded bits to provide a benefit over standard HEVC, thus both the bit range and associated bit savings are indicated. Smaller bit-budget windows are also analysed to give an idea of the instantaneous difference and demonstrate how performance varies.

For each picture, it can be seen that, simply upsampling the BL is more efficient for low bit-budgets and in most cases this approach also provides the greatest average result overall. As discussed in Section 7.1, this is because as the bit-budget decreases the ratio between enhancement data and signalling data within each EL will also decrease as a result of higher quantisation. Distribution of residual information will also affect the efficiency of an EL. The most effective use of each EL would be if all residual information were confined to the fewest number of coding blocks possible. This would maximise the number of blocks coded in SKIP mode and therefore minimise signalling data given the amount of enhancement data. It should also be noted that the greater level of sampling performance provided by the IDID technique, defined in Chapter 5, may reduce the efficiency that enhancement coding provides as it can potentially reduce the amount of residual information at the higher spatial layer.

For higher bit-budgets, the additional EL increases rate-distortion performance. In each case, the maximum bit-budget to provide a benefit is also increased as a result of applying enhancement coding. Given the data available, results suggest that this maximum budget is related to picture complexity. A higher average GPP generally results in a wider range for which coding at a lower spatial resolution is more efficient. However this metric is not an exact measure of complexity as it is an average and therefore does not take into account spatial variation of content. It also does not take into account the complexity of coding the chrominance channels. Given a bit-budget it is not an easy task to determine which coding method will provide the best rate-distortion

Table 7.1: Average PSNR difference using the BD-PSNR metric. Two resampling based techniques compared to standard HEVC intra-coding. BL represents the base layer upsampled and +EL represents the upsampled BL with the addition of an enhancement layer. Results show large compression efficiency gains can be achieved by either technique. Resampling without enhancement produces the best results overall but enhancement coding can increase rate-distortion performance at higher bitrates

Video	Frame	Bit Range	BL	+EL
BlueSky	f_0	0-1380kb	14.14%	13.09%
		0-500kb	17.31%	16.31%
		250-1000kb	8.20%	8.86%
		500-1380kb	6.96%	7.56%
	f_{108}	0-1820kb	13.05%	12.55%
		0-500kb	19.31%	18.09%
		250-1000kb	9.69%	11.32%
		750-1820kb	4.69%	5.51%
	f_{216}	0-3200kb	12.56%	12.84%
		0-500kb	18.80%	18.41%
		250-1000kb	16.91%	16.96%
		750-2000kb	10.41%	11.57%
		1000-3200kb	6.94%	7.64%
Tennis	f_0	0-612kb	7.66%	6.74%
		0-150kb	11.07%	9.18%
		100-300kb	4.31%	3.94%
		250-450kb	1.63%	1.80%
		300-612kb	0.76%	1.31%
	f_{200}	0-520kb	10.90%	12.12%
		0-150kb	15.23%	14.76%
		100-300kb	2.61%	8.10%
		250-450kb	-1.96%	2.93%
		300-520kb	-2.47%	1.96%
ParkJoy	f_0	0-1454kb	7.76%	7.42%
		0-750kb	9.55%	9.01%
		500-1000kb	5.55%	5.52%
		750-1250kb	5.23%	5.17%
		1000-1454kb	3.74%	3.87%
Squirrel	f_0	0-3800kb	6.26%	6.19%
		0-750kb	9.30%	8.67%
		500-1500kb	6.02%	5.97%
		1000-2250kb	5.77	5.82%
		1500-3000kb	4.24	4.27%
		2250-3800kb	2.82%	2.89%

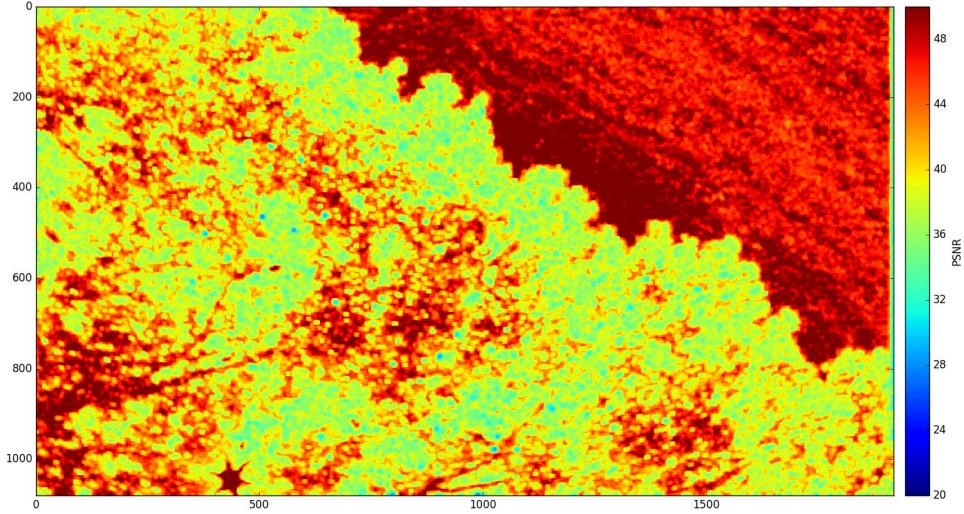


Figure 7.5: Local block-PSNR illustrating spatial variation of quality over the frame f_{108} of the sequence *BlueSky* coded at 1280×720 with a QP value of 25 and upsampled to 1920×1080 . The result shows low distortion in areas of low frequency and much higher levels of distortion in areas of high frequency

performance. However, these results demonstrate that an adaptive approach incorporating all three methods, including standard HEVC intra-coding, could potentially produce the best overall result.

7.2.3 Regional Quality Evaluation

Areas containing oversampling of the 2-dimensional signal can be coded at a lower spatial resolution without introducing greater levels of distortion. If these areas expand over one or more coding blocks within the EL then these blocks will be coded in SKIP mode. Areas of high frequency that contain a reduced amount of information as a result of downsampling may be enhanced within the EL. This effectively means that the reconstructed picture is composed of content coded as a mixture of resolutions.

Table 7.2: Total number of coded bits and corresponding PSNR for frame f_{108} of the sequence *BlueSky* coded at 1280×720 with a QP of 25 and with an EL coded at 1920×1080

QP_{EL}	Bits Coded	Average PSNR
29	977.31kb	40.373dB
31	949.90kb	40.249dB
33	930.56kb	40.190dB
35	926.72kb	40.133dB
None	924.61kb	40.106dB

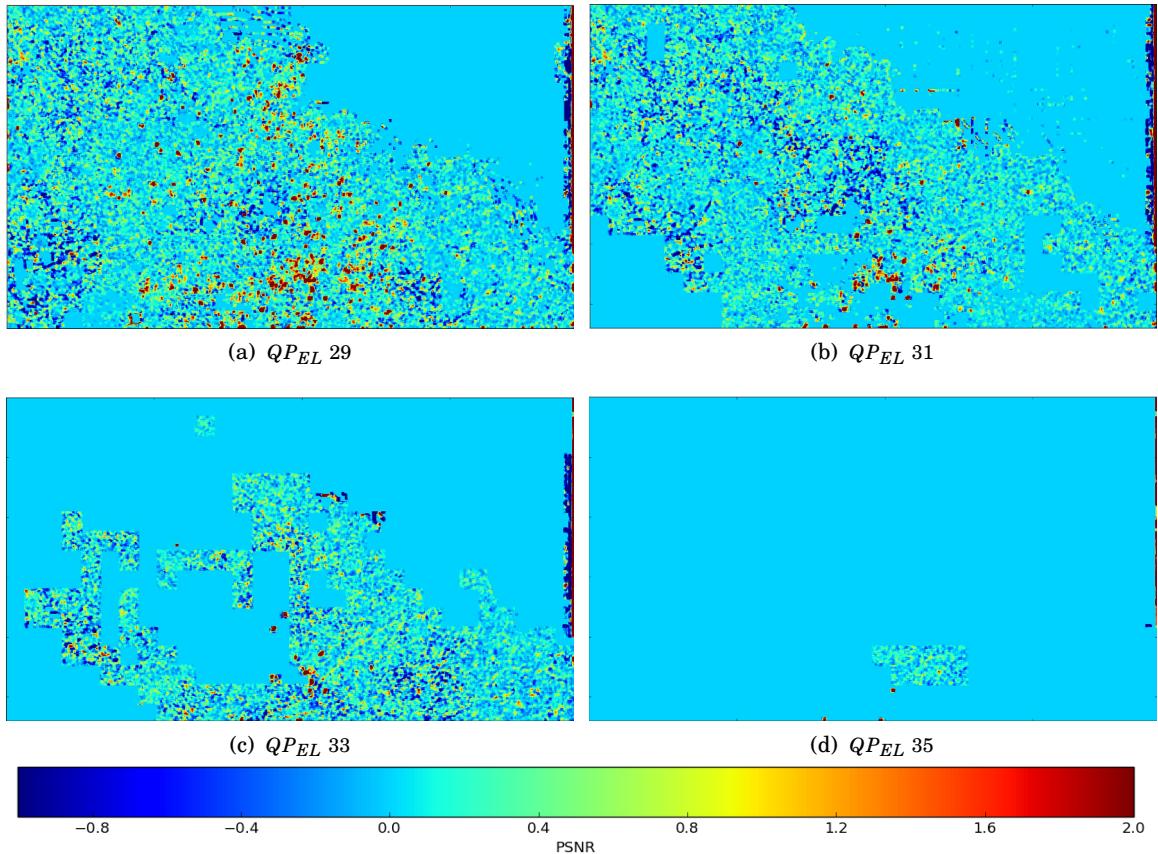


Figure 7.6: Quality gained or lost with the addition of an EL. Difference in local block-PSNR for frame f_{108} of sequence *BlueSky* coded at 1280×720 with $QP_{BL}\ 25$ compared to with the addition of an EL at the original HR (1920×1080). Positive values indicate a gain in local PSNR as a result of applying the EL

Figure (7.5) shows the result of a block-based PSNR function, in which local PSNR values are calculated for each pixel based on a 9×9 window, to demonstrate spatial variation of quality. For illustration the data is represented using a colour map. This function is applied to frame f_{108} of the sequence *BlueSky* after downsampling to a spatial resolution of 1280×720 and coding in intra mode using a QP value of 25. The result clearly shows lower levels of quality within the areas of high frequency content. Figure (7.6) illustrates how an EL can be used to enhance these areas by coding at the original resolution. Each figure shows the level of local PSNR gained or lost after applying an EL, using 4 different $QPEL$ values, compared to coding without any enhancement – dark blue indicates loss of quality whereas dark red indicates an increase in quality. The large areas of light blue colourisation represent no increase or decrease in quality because quantisation removed the data within the residual. It can be seen that it is possible that the EL applied will actually increase the amount of distortion in some areas and therefore it may be more efficient to code these corresponding blocks in SKIP mode. Table 7.2 provides the corresponding rate and

average PSNR for each coding combination. According to the results as presented in Section 7.2.2, the combination QP_{BL} 25 and QP_{EL} 33 is the most efficient.

7.3 Mixed Resolution Region-of-Interest Coding

Given a ROI binary map, we can force the inter-layer prediction function within SHVC to apply a SKIP flag to specified low priority regions and therefore ensure that enhancement is only applied to areas of importance. Prior to coding, the coding-tree structure is unknown as it depends on both content and the selected QP. However, the positioning of the LCUs are known as they have a fixed dimension of 64×64 pixels. By limiting the ROI area to the boundaries of the LCUs, we can ensure that, if an LCU is to be coded then all information within that block has the potential to be enhanced and therefore provide more effective use of signalling overhead data.

Figure (7.7) gives an example of Mixed Resolution ROI coding using frame f_{360} of the sequence *ParkJoy*. Figure (7.7a) highlights the ROI area to be enhanced along with the positions of LCUs, and Figure (7.7b) shows the spatial variation of local PSNR difference between a Mixed Resolution coded picture and a standard HEVC coded picture requiring a similar number of bits. Positive PSNR values corresponds to an increase in the quality of the ROI result and negative values correspond to better performance using standard HEVC. It can be seen from this Figure that within the non-ROI area, local PSNR varies between positive and negative values. The ROI area clearly shows a much higher level of PSNR quality using the Mixed Resolution ROI coding approach.

Figure (7.8) provides visual comparisons between the same two coded variations of the frame. A much greater level of detail is maintained within the ROI area using Mixed Resolution ROI coding compared to standard HEVC. The standard HEVC result is very *burry* in parts and also introduces high frequencies not present in the original picture. Within the example of the non-ROI area, there is very little noticeable difference. This may partly be due to textural masking within the areas comprising of the tree leaves. However, it can be argued that the image of the lamp is better maintained within the standard HEVC result.

Similar ideas have been proposed in the literature [18, 53, 54]. However, these methods aim to apply ROI within each spatial layer of SVC to maintain full scalability and adaptivity for receiving devices. Applying this approach for inter-coded pictures may typically reduce compression efficiency as it has been shown that single layer coding produces better results at the HR. As a continuation of the work presented in this chapter, it is suggested that inter-coded pictures are coded using the more common single layer QP adaptation methods and only apply Mixed Resolution ROI coding for intra-coded pictures.

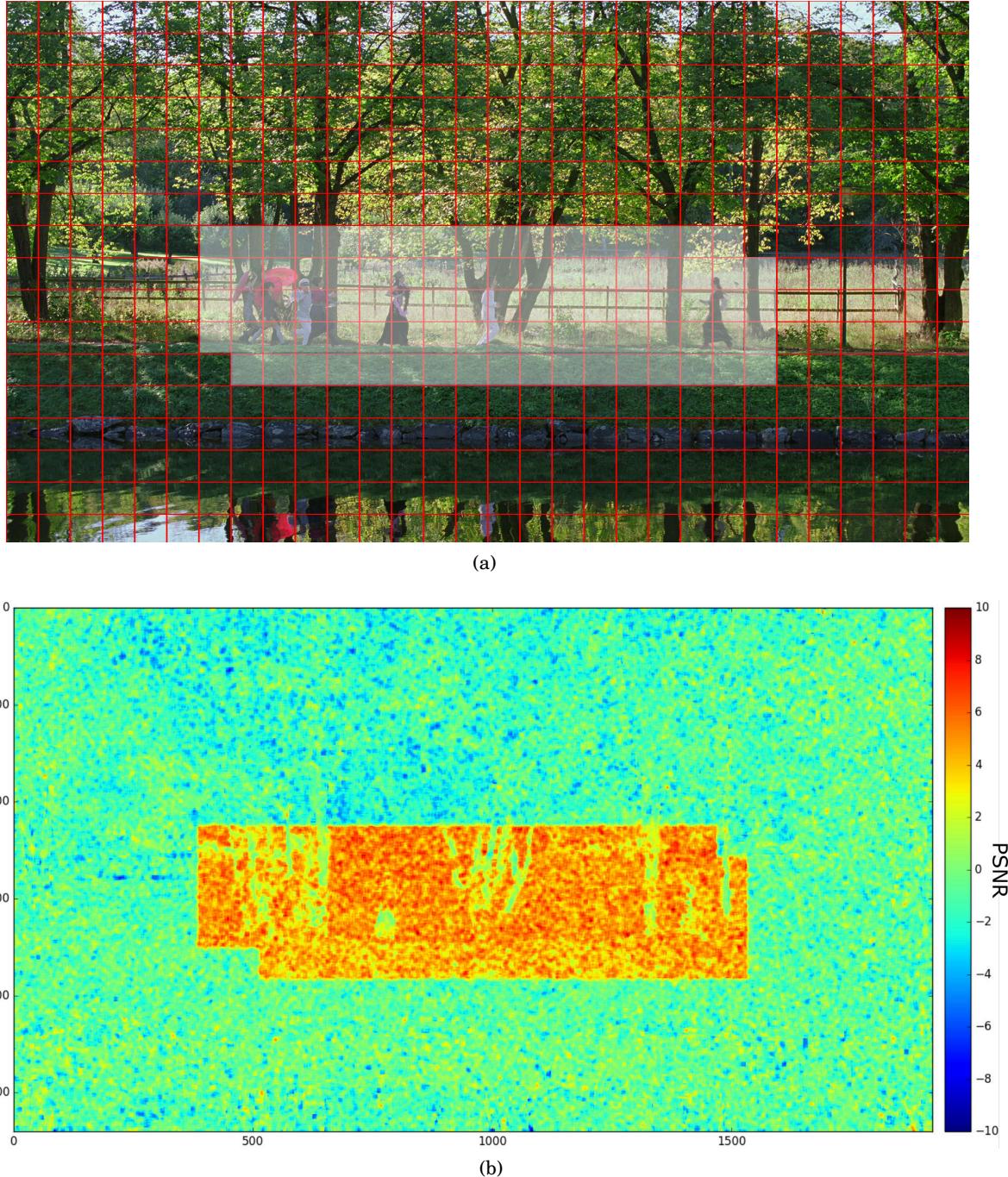


Figure 7.7: a) ROI mapping according to the position of LCUs. b) Local block-PSNR difference between standard HEVC coding: 1920×1080 QP 34 1613.25kb, and Mixed Resolution ROI Coding: 1280×720 QP_{BL} 32 $\rightarrow 1920 \times 1080$ QP_{PEL} 28 1610.98kb. Due to greater efficiency of coding intra-coded pictures at lower spatial resolutions, the non-ROI area can be coded without large loss in PSNR. The reduced QP at the HR within the ROI produces a much higher level of quality



Figure 7.8: Visual comparisons between standard HEVC and Mixed Resolution ROI coding. (a), (b) and (c) are contained within the defined ROI area, and (d), (e) and (f) are contained within the non-ROI area

7.4 Chapter Summary

This chapter demonstrates how enhancement coding using SHVC's inter-layer prediction functionality can further improve rate-distortion performance for a range of bit-budgets. Coding without any enhancement is still beneficial given restricted bandwidths but as more coding resources are made available the addition of an enhancement layer can be more effective. Enhancement coding also increases the range for which coding at lower spatial resolutions is more beneficial than standard HEVC. Using SHVC in this manner can not be performed within the standard as it requires adaptation of the BL spatial resolution according to picture content and the target bit-budget. However, most of the functionality is already available and therefore only small modifications would be required to achieve this greater level of coding efficiency for intra-coded pictures.

A Mixed Resolution ROI coding method is also introduced which shows very promising initial results. As spatial resampling of intra-coded pictures can provide greater rate-distortion performance, it is possible to significantly enhance quality within ROI areas whilst maintaining a similar level of quality within the non-ROI areas compared to standard HEVC. Further work is required to provide a more detailed evaluation of performance compared to alternative ROI methods.

CONCLUSIONS & FURTHER WORK

The demand for higher spatial resolutions, faster frame rates and greater dynamic range for video transmissions has led to the need for more intelligent and efficient coding algorithms. HEVC is the latest state-of-the-art video coding standard but it is yet to overtake AVC in terms of deployment; although more efficient algorithms are being developed all the time, the encoder needs to comply with a standard decoder which is typically implemented in hardware within consumer devices. HEVC has been shown to provide efficiency gains of around 50% for HD sequences compared to AVC. However at low bitrates, high compression rates can render the video useless by removing important information contained within the original.

8.1 Conclusions & Discussions

The key contribution provided in this thesis lies within the defined coding method. All coding standards at the time of writing maintain a fixed spatial resolution throughout the coded bit-stream. Previous work has highlighted that, depending on video content, a sequence may be coded more efficiently at a lower spatial resolution and upsampled afterwards. As resampling is applied before and after encoding and decoding, respectively, this approach can be implemented without changes to the coding standard. However, as inter-coded data is coded far more efficiently than intra-coded data, much of this performance gain will be a result of resampling the latter. For this reason the proposed coding technique applies spatial resampling to intra-coded pictures only and therefore needs to be applied within the coding loop and not as pre- and post-processing operations. This means that it does not comply with the standard but enables inter-coded pictures

to be coded efficiency at the original HR.

Chapter 3 presented the initial version of the proposed technique. The Low Delay configuration is applied to produce all inter-coding results throughout this thesis and IDR pictures are coded with an intra period equal to roughly 0.5 seconds. This version of the coding technique is neither optimal nor practical for real-time video transmission. However, it does demonstrate that spatial resampling of intra-coded pictures can increase rate-distortion performance at low bitrates. It also indicates that the rate-control algorithm currently implemented within the HM reference software does not always perform well. Results show vast improvements in rate-distortion performance at low bitrates but it is unclear how much of this gain is due to a greater level of coding efficiency provided by the resampling of IDR pictures and how much is due reducing the target bit-budget of these pictures. As the IDR bit-budget is determined by the rate-control algorithm when coding at the lower spatial resolution, it is dependent on the scale factor and therefore is not necessarily going to provide the best result when coding at the original HR – the selected scale factor and the bit-budget should be independent.

By performing resampling within the coding standard we have the added benefit that spatial resolution can be adapted in real-time, according to picture content and the available bandwidth. In Chapter 4 an Adaptive Resolution R-Q model is presented for the resampling of intra-coded pictures. Each intra-coded picture is coded at multiple resolutions by first predicting the best QP given a model based on a picture complexity measure. After coding, each picture variation is upsampled and the overall distortion is calculated. The best coded picture is selected according to distortion, the number of coded bits, and the mismatch ratio which indicates how well the actual rate conforms to the target rate. This R-Q model is later applied within an inter-coding configuration, in Chapter 6, to adapt the resolution of IDR pictures.

The performance of the proposed Adaptive Resolution coding technique is partly dependent on the performance of the resampling operations; reducing the distortion introduced by downsampling and upsampling will improve overall compression efficiency. Bicubic is used for resampling pictures within the work presented in both Chapter 3 and 4. However, it is shown in Chapter 5 that the SHVC sampling functions produce better results. Various image enhancement techniques are also analysed and it is found that the IDID technique can further reduce MSE by optimising the LR picture for the given upsampling operation. This technique is also effective for coded LR pictures and is therefore highly suitable for the proposed Adaptive Resolution coding technique.

Chapter 6 brings together the best parts of the work achieved in the previous chapters to present the main results for this thesis. As development of a more efficient rate control algorithm is beyond the scope of this thesis, the IDR bit-budget is modified offline, not in real-time, to enable

more effective use of the Adaptive Resolution coding technique. Results vary for each of the tested sequences, however best results can be seen within sequences with high temporal correlation. Temporal correlation is determined by the rate of change/movement within the scene relative to the temporal sampling rate. The sequence that benefits most by the proposed Adaptive Resolution coding technique is the sequence *Squirrel*. This is because it was captured using a fixed camera and a very high sampling rate relative to the amount of motion within the scene. Due to coding of IDR pictures, only high temporal correlation within each GOP is important as pictures coded prior can not be referenced. For a GOP with high temporal correlation, a high IDR bit-budget is more optimal as large portions of the inter-coded pictures can be efficiently predicted from the IDR picture. Lower temporal correlation would result in coding of more residual information, motion vectors, and possible intra-coded data. Therefore a lower IDR bit-budget would enable rate control to allocate more bits to the remaining inter-coded pictures and provide a more consistent and high level of quality. Results show that the proposed method increases rate-distortion performance up to 11.5% within the range 100kbps-5Mbps for HD sequences. Given that the scenario for the proposed work involves using a UAV to capture video from a high vantage point, there will likely be high correlation between frames and therefore enable high performance using the Adaptive Resolution coding technique. However, the proposed method may fail to increase performance by a large factor when tracking a fast moving object while flying at a low altitude or applying camera zoom.

There are similarities between the work presented in this thesis and the scalable extensions of the coding standards, for example SVC and SHVC. Scalable video coding provides greater versatility for receiving devices, including the ability to restrict the decoded spatial resolution. This is achieved by spatial resampling and use of inter-layer prediction to enhance upsampled pictures after coding. Previous work has shown that scalable coding does not provide greater rate-distortion performance at the original HR within an inter-coding configuration. However, in Chapter 7 results show that at low bitrates applying an EL to a LR intra-coded picture can provide greater performance gains than simply coding at the original HR. This approach is also compared to the previous approach of resampling without enhancement, as originally presented in Chapter 3. Results show that resampling alone is more efficient overall, especially at very low bitrates. However, as the bitrate increases the addition of an EL becomes more optimal and can also increase the range for which resampling is more beneficial than coding at the original HR.

The work presented in this thesis demonstrates a more effective method of coding intra-coded pictures. Although the specifications of HEVC will not change, such a technique could be applied to future standards. Adaptive Resolution coding does incur some additional coding complexity due to resampling and coding at multiple lower resolutions, including coding at the original resolution. However, these additional complexity gains are relatively small in comparison to other

coding processes such as inter-coding. There are still modifications that could be made in order to improve the overall efficiency of the technique. Coding intra pictures at multiple resolutions is required as accurate prediction of the overall distortion after reconstruction at the original HR is not an easy task and an effective method was not discovered. Solving this problem could actually reduce coding complexity, compared to the standard, when coding at lower spatial resolutions. However, currently prediction of the best QP for each resolution also relies on previously coded pictures, coded at the same resolution. Other improvements include better rate control to better define the IDR bit-budget, better resampling so not to reduce quality in some areas of the picture, and increased efficiency of intra-coded data within inter-coded pictures. These ideas are discussed further in Section 8.2. Although not a major focus of this thesis, computational complexity is actually a big issue for the proposed UAV scenario as battery power is very limited. Increasing the computational complexity of the video encoder would decrease the UAV's flight time. Solving the issue of needing to code at multiple resolutions for each IDR picture would therefore provide a more desirable solution than standard HEVC in terms of the amount of processing power required, as well as the added rate-distortion benefits that the technique provides.

8.2 Further Work

8.2.1 Adaptive Rate Control Algorithm for Varying Video Content

The major part of this thesis that prevents the proposed technique being used for real-time applications, such as UAV video transmission, is the lack of a suitable and high performance rate control algorithm. There are two major flaws with the current algorithm currently used in the HM Rerference software when used for real-time applications.

1. The bitrate is regulated such that the overall average bitrate matches the target and not the instantaneous bitrate at any point in time.
2. The optimal ratio of IDR bits to inter-coded bits, within each GOP, depends on video content – primarily temporal correlation between frames. However, without prior knowledge, the optimal ratio can not be determined.

To elaborate, problem 1. means that if the coded video bitrate is greater than the set target at any point in the sequence then the rate control algorithm will compensate by reducing the video bitrate to a lower level in order to produce an overall average that matches the set target. This is acceptable for storage of compressed video bitstreams as, provided that the hardware is capable of processing the substantial changes in the instantaneous bitrate, the only thing that matters is the average bitrate or the total number of coded bits required for storage. For video transmission, the instantaneous bitrate should closely match the target at all time to prevent frame skipping and maximise bandwidth efficiency.

Problem 2. is much more complicated and under normal conditions it would be impossible to accurately predict the optimal bit-budget with absolute certainty. However, for the defined scenario of transmitting video from a UAV, additional information is available that can give some indication of the expected level of temporal correlation within the immediate future. UAVs utilise sensors such as accelerometers and gyroscopes to provide information of acceleration and angle of tilt. A UAV that is hovering in a stationary position with a camera that is fixed is likely to capture video content with higher temporal correlation compared to a UAV that is in flight and has a camera that is panning and tilting. The information provided by the on-board sensors could therefore be used to determine a more optimal IDR bit-budget for the next GOP. Some latency is expected but this approach should provide a means of predicting characteristics of near future video content.

8.2.2 Spatially Invariant IDID

The IDID technique is shown to reduce the overall MSE of a resampled image. It is also effective for resampling of downsampled compressed video given that it will be upsampled back to the original HR after decoding. The work presented in this thesis demonstrates that IDID can actually increase MSE distortion in some areas of the reconstructed image. By identifying these areas, it may be possible to restrict application of the IDID technique only to areas that will benefit from its use.

8.2.3 Adaptive Resolution of Intra-Coded Data within Inter-Coded Pictures

For sequences with low temporal correlation the benefits of applying the proposed Adaptive Resolution coding technique are minimal. By defining a low IDR bit-budget, there is a high chance that each IDR picture will be coded more efficiently. However, inter-coded pictures will not directly benefit given that the information in these frames can not be predicted efficiently from the IDR picture available for reference. In such cases, areas of these pictures may be coded in intra mode. It would therefore also make sense to have the option of coding this information at lower spatial resolutions and have the potential to gain greater coding efficiency as achieved within the IDR pictures.

8.2.4 Evaluation of the Adaptive Resolution R-Q model for Inter-Coded Video

Due to the fact that the Adaptive Resolution R-Q model relies on similarities in previously coded intra pictures, as the intra period increases the accuracy of QP prediction could potentially reduce. Although a performance gain of up to a 11.5% is demonstrated using the proposed Adaptive Resolution coding technique, a detailed evaluation of how well the model operates in an inter-coding configuration is not provided in this thesis. As a result it is unclear if performance could

be increased further by exploring methods of increasing the accuracy of QP prediction or if performance was affected by inefficient allocation of the bit-budget due to rate control.

8.2.5 More Effective Models for Picture Complexity

Predicting the optimal QP for each intra-coded picture is achieved using a picture complexity measure and parameters learnt from previously coded intra frames. A more effective picture complexity model could enable better prediction without the need of learning from previous frames. This would be beneficial when content varies greatly over time and the intra period is long, meaning that the parameters calculated for previous frames may actually reduce the accuracy of prediction. Current measures do not take into account the information lost as a result of quantisation and can therefore further reduce the accuracy of prediction for large QP values.

BIBLIOGRAPHY

- [1] K. AL ISMAEIL, D. AOUADA, B. MIRBACH, AND B. OTTERSTEN, *Dynamic super resolution of depth sequences with non-rigid motions*, in IEEE International Conference on Image Processing (ICIP), 2013, pp. 660–664.
- [2] J. BAE, O. LEE, AND J. KIM, *Image interpolation using gabor filter*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 986–990.
- [3] G. BJONTEGAARD, *Calculation of average PSNR differences between RD-curves (VCEG-M33)*, Video Coding Experts Group (VCEG) Meeting: Austin, Texas, USA, 2-4 April, (2001).
- [4] F. BOSSEN, B. BROSS, S. KARSTEN, AND D. FLYNN, *HEVC Complexity and Implementation Analysis*, IEEE Trans. on Circuits and Systems for Video Technology, 22 (2013), pp. 1685–1696.
- [5] F. BOSSEN, B. BROSS, S. MEMBER, S. KARSTEN, AND D. FLYNN, *HEVC Complexity and Implementation Analysis*, IEEE Transactions on Circuits and Systems for Video Technology, 22 (2013), pp. 1685–1696.
- [6] J. M. BOYCE, Y. YE, J. CHEN, AND A. K. RAMASUBRAMONIAN, *Overview of SHVC: scalable extensions of the High Efficiency Video Coding (HEVC) Standard*, IEEE Transactions on Circuits and Systems for Video Technology, (2015), pp. 1–15.
- [7] A. M. BRUCKSTEIN, M. ELAD, AND R. KIMMEL, *Down-scaling for better transform compression*, IEEE Transactions on Image Processing, 12 (2003), pp. 1132–1144.
- [8] J. CHEN, J. BOYCE, Y. YE, AND M. M. HANNUKSELA, *International organization for standardization organisation internationale de normalisation iso / iec jtc1 / sc29 / wg11 coding of moving pictures and audio*, JCTVC-R1007_v1 (tutorial information document), (2014).
- [9] J. CHEN, J. NUNEZ-YANEZ, AND A. ACHIM, *Video super-resolution using generalized Gaussian Markov random fields*, IEEE Signal Processing Letters, 19 (2012), pp. 63–66.

BIBLIOGRAPHY

- [10] J. CHEN, J. L. NUNEZ-YANEZ, AND A. ACHIM, *Bayesian Video Super-Resolution With Heavy-tailed Prior Models*, IEEE Transactions on Circuits and Systems for Video Technology, 24 (2014), pp. 905–914.
- [11] Z. CHEN AND K. N. NGAN, *Recent advances in rate control for video coding*, Signal Processing: Image Communication, 22 (2007), pp. 19–38.
- [12] J. DONG AND Y. YAN, *Adaptive downsampling for high-definition video coding*, IEEE International Conference on Image Processing (ICIP), (2012), pp. 2925–2928.
- [13] J. DONG AND Y. YE, *Adaptive downsampling for high-definition video coding*, IEEE Transactions on Circuits and Systems for Video Technology, 24 (2014), pp. 480–488.
- [14] W. DONG, L. ZHANG, G. SHI, AND X. WU, *Nonlocal back-projection for adaptive image enlargement*, IEEE International Conference on Image Processing (ICIP), (2009), pp. 349–352.
- [15] H. EVERETT, *Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources*, Operations Research, (1963), pp. 399-417.
- [16] S. FARSIU, M. D. ROBINSON, M. ELAD, AND P. MILANFAR, *Fast and robust multiframe super resolution.*, IEEE Transactions on Image Processing, 13 (2004), pp. 1327–44.
- [17] T. FRAJKA AND K. ZEGER, *Downsampling dependent upsampling of images*, Signal Processing: Image Communication, 19 (2004), pp. 257–265.
- [18] D. GROIS, E. KAMINSKY, AND O. HADAR, *ROI adaptive scalable video coding for limited bandwidth wireless networks*, in Wireless Days IFIP, 2010, pp. 0–4.
- [19] W.-J. HAN, J. MIN, I.-K. KIM, E. ALSHINA, A. ALSHIN, T. LEE, J. CHEN, V. SEREGIN, S. LEE, Y. M. HONG, M.-S. CHEON, N. SHLYAKHOV, K. MCCANN, T. DAVIES, AND J.-H. PARK, *Flexible Unit Representation and Corresponding Extension of Coding Tools*, 20 (2010), pp. 1709–1720.
- [20] K. HE, J. SUN, AND X. TANG, *Guided image filtering.*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2013), pp. 1397–409.
- [21] P. HELLE, H. LAKSHMAN, M. SIEKMANN, J. STEGEMANN, T. HINZ, H. SCHWARZ, D. MARPE, AND T. WIEGAND, *A Scalable Video Coding Extension of HEVC*, 2013 Data Compression Conference, (2013), pp. 201–210.
- [22] T. HO-PHUOC, A. DUPRET, AND L. ALACOQUE, *Super Resolution method adapted to spatial contrast*, International Conference on Image Processing (ICIP), (2013), pp. 976–980.

- [23] B. HOSKING, D. AGRAFIOTIS, D. BULL, AND N. EASTON, *Spatial resampling of IDR frames for low bitrate video coding with HEVC*, IS&T/SPIE Electronic Imaging, 9410 (2015), pp. 1–6.
- [24] B. HOSKING, D. AGRAFIOTIS, D. BULL, AND N. EASTON, *Adaptive Resolution of Intra-Pictures for Bandwidth Constrained Video Coding*, IEEE Transactions on Circuits and Systems for Video Technology - Submitted, (2016).
- [25] B. HOSKING, D. AGRAFIOTIS, D. BULL, AND N. EASTON, *An adaptive resolution rate control method for Intra Coding in HEVC*, IEEE International Conference on Acoustics, Speech and Signal Processing, (2016), pp. 1486–1490.
- [26] B. HOSKING, D. AGRAFIOTIS, D. BULL, AND N. EASTON, *Enhancement of Intra-Coded Pictures for Greater Coding Efficiency*, Picture Coding Symposium (PCS) - Accepted, (2016).
- [27] Q. HUYNH-THU AND M. GHANBARI, *Scope of validity of PSNR in image/video quality assessment*, Electronic Letters, 44 (2008), pp. 1–2.
- [28] IMTC, *Specification of Interoperable Scalability Modes for the HEVC Video Coding Standard for Unified Communication Applications*, Technical Document (IMTC1014), (2016).
- [29] M. IRANI AND S. PELEG, *Improving resolution by image registration*, CVGIP: Graphical models and image processing, 53 (1991), pp. 231–239.
- [30] M. M. ISLAM, V. K. ASARI, M. N. ISLAM, AND M. A. KARIM, *Super-resolution enhancement technique for low resolution video*, IEEE Transactions on Consumer Electronics, (2010), pp. 919–924.
- [31] X. JING AND L. CHAU, *A novel intra-rate estimation method for H.264 rate control*, IEEE International Symposium on Circuits and Systems, (2006).
- [32] X. JING, L. CHAU, AND W. SIU, *Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control*, IEEE Signal Processing Letters, 15 (2008), pp. 373–376.
- [33] X. JING, J. Y. THAM, Y. WANG, K. H. GOH, AND W. S. LEE, *Efficient rate-quantization model for frame level rate control in spatially scalable video coding*, in IEEE International Conference on Networks, ICON, 2012, pp. 339–343.
- [34] S. KAPPAGANTULA AND K. R. RAO, *Motion Compensated Interframe Image Prediction*, IEEE Transactions on Communications, 33 (1985), pp. 1011–1015.
- [35] R. KEYS, *Cubic convolution interpolation for digital image processing*, IEEE Transactions on Acoustics, Speech, and Signal Processing, (1981), pp. 1153–1160.

BIBLIOGRAPHY

- [36] W. J. KIM, J. W. YI, AND S. D. KIM, *A bit allocation method based on picture activity for still image coding.*, IEEE Transactions on Image Processing, 8 (1999), pp. 974–977.
- [37] J. KOPF AND M. COHEN, *Joint bilateral upsampling*, ACM Transactions on Graphics, 26 (2007).
- [38] S. KRISHNAN AND J. KLOSOWSKI, *Guided image upsampling using bitmap tracing*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 650–654.
- [39] 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, 22 (2012), pp. 1792–1801.
- [40] H. LAKSHMAN, W.-Q. LIM, H. SCHWARZ, D. MARPE, G. KUTYNIOK, T. WIEGAND, H. LAKSHMAN, W.-Q. LIM, AND H. SCHWARZ, *Image interpolation using Shearlet based iterative refinement*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 1–11.
- [41] H. LEE, J. KANG, J. LEE, J. CHOI, J. KIM, AND D. SIM, *Scalable Extension of HEVC for Flexible High-Quality Digital Video Content Services*, ETRI Journal, 35 (2013), pp. 990–1000.
- [42] W. LIN AND L. DONG, *Adaptive downsampling to improve image compression at low bit rates.*, IEEE Transactions on Image Processing, 15 (2006), pp. 2513–21.
- [43] J. LIU, Y. CHO, Z. GUO, AND C. KUO, *Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis*, IEEE Transactions on Circuits and Systems for Video Technology, 20 (2010), pp. 967–981.
- [44] X. LU AND X. LI, *Multiresolution imaging*, IEEE Transactions on Cybernetics, 44 (2014), pp. 149–60.
- [45] S. MANDAL AND A. SAO, *Edge Preserving Single Image Super Resolution in Sparse Environment*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 967–971.
- [46] G. MARCONI, *Radio telegraphy*, Journal of the American Institute of Electrical Engineers, 41 (1922), pp. 561–570.
- [47] M. MEDDEB AND M. CAGNAZZO, *ROI-based rate control using tiles for an HEVC encoded video stream over a lossy network*, IEEE International Conference on Image Processing (ICIP), (2015), pp. 1389–1393.
- [48] V. NGUYEN, Y. TAN, AND W. LIN, *Adaptive downsampling/upsampling for better video compression at low bit rate*, IEEE International Symposium on Circuits and Systems, (2008), pp. 1624–1627.

- [49] J. OHM, G. J. 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, 22 (2012), pp. 1669–1684.
- [50] J.-R. OHM AND G. J. SULLIVAN, *High Efficiency Video Coding: The Next Frontier in Video Compression*, IEEE Signal Processing Magazine, (2013), pp. 152–158.
- [51] P. P. VAIDYANATHAN, *Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications*, Proceedings of the IEEE, 78 (1990), pp. 56–93.
- [52] M. A. PAPADOPOULOS, F. ZHANG, D. AGRAFIOTIS, AND D. BULL, *A video texture database for perceptual compression and quality assessment*, IEEE International Conference on Image Processing (ICIP), (2015), pp. 2781–2785.
- [53] Y. PATNAIK, R. MAHARAJU, AND D. PATRA, *Region of interest based scalable image and video coding a superlative study*, Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN), (2014), pp. 1–6.
- [54] W.-H. PENG, T. CHIANG, AND H.-M. HANG, *Adding Selective Enhancement in Scalable Video Coding for Region-of-Interest Functionality*, IEEE International Symposium on Circuits and Systems, 2006. ISCAS 2006. Proceedings., (2006), pp. 3089–3092.
- [55] I. RICHARDSON, *The H.264 Advanced Video Coding Standard*, John Wiley & Sons Inc, 2010.
- [56] J. SALVADOR, E. PÉREZ-PELLITERO, AND A. KOCHALE, *Fast single-image super-resolution with filter selection*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 640–644.
- [57] H. SCHWARZ, D. MARPE, AND T. WIEGAND, *Overview of the scalable video coding extension of the H.264/AVC standard*, IEEE Transactions on Circuits and Systems for Video Technology, 17 (2007), pp. 1103–1120.
- [58] C. SEGALL AND G. SULLIVAN, *Spatial scalability within the H. 264/AVC scalable video coding extension*, IEEE Trans. on Circuits and Systems for Video Technology, 17 (2007), pp. 1121–1135.
- [59] M. SHEN, P. XUE, AND C. WANG, *Down-sampling based video coding using super-resolution technique*, IEEE Transactions on Circuits and Systems for Video Technology, 21 (2011), pp. 755–765.
- [60] Z. SHI, X. SUN, AND F. WU, *Spatially Scalable Video Coding for HEVC*, 2012 IEEE International Conference on Multimedia and Expo, (2012), pp. 1091–1096.

BIBLIOGRAPHY

- [61] A. SINGH AND N. AHUJA, *Single image super resolution using adaptive domain transformation*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 947–951.
- [62] R. SJÖBERG, Y. CHEN, S. MEMBER, A. FUJIBAYASHI, M. M. HANNUKSELA, J. SAMUELSSON, T. K. TAN, Y.-K. WANG, AND S. WENGER, *Overview of HEVC High-Level Syntax and Reference Picture Management*, 22 (2012), pp. 1858–1870.
- [63] G. SULLIVAN, J. OHM, W. HAN, AND T. WIEGAND, *Overview of the High Efficiency Video Coding (HEVC) Standard*, IEEE Transactions on Circuits and Systems for Video Technology, (2012), pp. 1–19.
- [64] G. J. SULLIVAN, J. M. BOYCE, Y. CHEN, J. OHM, C. A. SEGALL, AND A. VETRO, *Standardized extensions of High Efficiency Video Coding (HEVC)*, IEEE Journal on Selected Topics in Signal Processing, 7 (2013), pp. 1001–1016.
- [65] G. J. SULLIVAN AND T. WIEGAND, *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 10th Meeting: Stockholm, SE, 11-20 July*, (2012), pp. 11–20.
- [66] L. SUN, O. C. AU, W. DAI, Y. GUO, AND R. ZOU, *An adaptive frame complexity based rate quantization model for intra-frame rate control of High Efficiency Video Coding (HEVC)*, Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), (2012), pp. 1–6.
- [67] B.-D. TSENG, *Specifications of Ideal Prototype Filters for Designing Linear Phase FIR Filters*, Signals, Systems and Computers, (1997), pp. 29–33.
- [68] M. TÜRKAN, D. THOREAU, AND P. GUILLOTEL, *Optimized neighbor embeddings for single-image super-resolution*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 645–649.
- [69] K. UGUR, K. ANDERSSON, A. FULDSETH, G. BJØNTEGAARD, L. P. ENDRESEN, J. LAINEEMA, A. HALLAPURO, J. RIDGE, D. RUSANOVSKYY, C. ZHANG, A. NORIKIN, C. PRIDDLE, T. RUSERT, J. SAMUELSSON, AND R. SJ, *High Performance , Low Complexity Video Coding and the Emerging HEVC Standard*, 20 (2010), pp. 1688–1697.
- [70] Z. WANG, A. C. BOVIK, H. R. SHEIKH, S. MEMBER, E. P. SIMONCELLI, AND S. MEMBER, *Image Quality Assessment : From Error Visibility to Structural Similarity*, 13 (2004), pp. 600–612.
- [71] Z. WANG, A. C. BOVIK, H. R. SHEIKH, AND E. P. SIMONCELLI, *Image quality assessment: from error visibility to structural similarity.*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–12.

- [72] M. WIEN, H. SCHWARZ, AND T. OELBAUM, *Performance analysis of SVC*, IEEE Transactions on Circuits and Systems for Video Technology, 17 (2007), pp. 1194–1203.
- [73] S. WINKLER, *Analysis of Public Image and Video Databases for Quality Assessment*, IEEE Journal of Selected Topics in Signal Processing, 6 (2012), pp. 616–625.
- [74] X. WU, X. ZHANG, AND X. WANG, *Low bit-rate image compression via adaptive down-sampling and constrained least squares upconversion*, IEEE Transactions on Image Processing, 18 (2009), pp. 552–561.
- [75] M. XU, X. DENG, S. LI, AND Z. WANG, *Region-of-Interest Based Conversational HEVC Coding with Hierarchical Perception Model of Face*, IEEE Journal of Selected Topics in Signal Processing, 8 (2014), pp. 475–489.
- [76] K. ZHANG, X. GAO, D. TAO, X. LI, AND I. SCIENCE, *Image super resolution via non-local steering kernel regression*, IEEE International Conference on Image Processing (ICIP), (2013), pp. 943–946.
- [77] Y. ZHANG, D. ZHAO, J. ZHANG, R. XIONG, AND W. GAO, *Interpolation-Dependent Image Downsampling*, IEEE Transactions on Image Processing, 20 (2011), pp. 3291–3296.
- [78] Y. ZHOU, L. TIAN, AND X. NING, *Intra frame constant rate control scheme for high efficiency video coding*, International Conference on Computing, Networking and Communications (ICNC), (2013), pp. 648–652.

APPENDICES

A - Test Sequences



Sequence: BlueSky
Resolution: 1920 × 1080
Frame Rate: 25Hz
Frames: 217
Link: <https://media.xiph.org/video/derf/>
Database: Xiph



Sequence: Station2
Resolution: 1920 × 1080
Frame Rate: 25Hz
Frames: 313
Link: <https://media.xiph.org/video/derf/>
Database: Xiph



Sequence: ParkJoy
Resolution: 1920 × 1080
Frame Rate: 50Hz
Frames: 500
Link: <https://media.xiph.org/video/derf/>
Database: Xiph



Sequence: Squirrel
Resolution: 1920 × 1080
Frame Rate: 60Hz
Frames: 600
Link: <https://data.bris.ac.uk/data/dataset/>
Database: Visual Information Laboratory BVI Textures



Sequence: BricksLeaves
Resolution: 1920×1080
Frame Rate: 60Hz
Frames: 600
Link: <https://data.bris.ac.uk/data/dataset/>
Database: Visual Information Laboratory BVI Textures



Sequence: InToTree
Resolution: 1920×1080
Frame Rate: 50Hz
Frames: 500
Link: <https://media.xiph.org/video/derf/>
Database: Xiph



Sequence: OldTownCross
Resolution: 1920×1080
Frame Rate: 50Hz
Frames: 500
Link: <https://media.xiph.org/video/derf/>
Database: Xiph



Sequence: Tennis
Resolution: 1920×1080
Frame Rate: 25Hz
Frames: 240
Link: <ftp://ftp.tnt.uni-hannover.de/>
Database: TNT Hannover



Sequence: ReadySetGo
Resolution: 3840×2160
Frame Rate: 120Hz
Frames: 600
Link: <http://ultravideo.cs.tut.fi/#testsequences>
Database: Ultravideo



Sequence: Jockey
Resolution: 3840×2160
Frame Rate: 120Hz
Frames: 600
Link: <http://ultravideo.cs.tut.fi/#testsequences>
Database: Ultravideo

D - Quantisation formula for Rate Control

$$(A-1) \quad R_{pred}(Q_{step}) = C.a.Q_{step}^b$$

$$(A-2) \quad Q_{step}^b = \frac{R}{C.a}$$

$$(A-3) \quad b = \log_{Q_{step}} \frac{R}{C.a}$$

$$(A-4) \quad \ln Q_{step} = \frac{\ln \frac{R}{C.a}}{b}$$

$$(A-5) \quad Q_{step} = e^{\frac{\ln(\frac{R}{C.a})}{b}}$$