

COMPRESSION EFFICIENCY IMPROVEMENT FOR 2D AND 3D VIDEO

by

Sima Valizadeh

B.Sc., Sharif University of Technology, 2005

M.Sc., The University of Tehran, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2017

© Sima Valizadeh, 2017

Abstract

Advances in video compression technologies have resulted in high visual quality at constrained amounts of bitrate. This is crucial in video transmission and storage, considering the limited bandwidth of communication channels and storage media with limited capacities. In this thesis, we propose new methods for improving the compression efficiency of HEVC and its 3D extension for stereo and multiview video content.

To achieve high video quality while keeping the bitrate within certain constraints, the characteristics of the human visual system (HVS) play an important role. The utilization of video quality metrics that are based on the human visual system and their integration within the video encoder can improve compression efficiency. We, therefore, propose to measure the distortion using a perceptual video quality metric (instead of sum of squared errors) inside the coding unit structure and for mode selection in the rate distortion optimization process of HEVC. Experiments show that our method improves HEVC compression efficiency by 10.21%.

Next, we adjust the trade-off between the perceptual distortion and the bitrate based on the characteristics of the video content. The value of the Lagrange multiplier is estimated from the first frame for every scene in the video. Experimental results show that the proposed approach further improves the compression efficiency of HEVC (up to 2.62% with an average of 0.60%).

Furthermore, we extend our work to address the HEVC extension for 3D video. First, we integrate the perceptual video quality in the rate distortion optimization process of stereo video coding where the dependencies between the two views are exploited to improve coding

efficiency. Next, we extend our approach to multiview video coding for auto-stereoscopic displays (where 3D content can be viewed without using 3D glasses). In this case, two or three views and their corresponding depth maps need to be coded. Our proposed perceptual 3D video coding increases the compression efficiency of multi-view video coding by 2.78%.

Finally, we show that compression efficiency of stereoscopic videos improves if we take advantage of asymmetric video coding. The proposed approach reduces the amount of bitrate required for transmitting stereoscopic video while maintaining the stereoscopic quality.

Lay Summary

Advances in video compression technologies have resulted in high visual quality at constrained amounts of bitrate. This is crucial in video transmission and storage, considering the limited bandwidth of communication channels and storage media with limited capacities. In this thesis, we propose new methods for improving the compression efficiency of HEVC and its 3D extension for stereo and multiview video content. First, we integrate a perceptual video quality metric in the high efficiency video coding standard. Next, we adjust the trade-off between the perceptual distortion and the bitrate based on the characteristics of the video content. Furthermore, we extend our work to address the 3D extension of HEVC for stereo and multiview videos. Finally, we show that compression efficiency of stereoscopic videos improves if we take advantage of asymmetric video coding. The proposed approach reduces the amount of bitrate required for transmitting video while maintaining the video quality.

Preface

This thesis presents research conducted by Sima Valizadeh, under the guidance of Dr. Panos Nasiopoulos and Dr. Rabab Ward. A list of publications resulting from the work presented in this thesis is provided on the following page.

The content of Chapter 2 is published in two conference papers [P4]-[P5] and one submitted journal paper [P1]. The content of Chapter 4 appears in one conference [P3] and one submitted journal publication [P2]. The main body of Chapter 5 is taken from our previous publications in [P6]-[P7]. The work presented in all of these manuscripts was performed by Sima Valizadeh, including literature review, designing and implementing the proposed algorithms, performing all experiments, analyzing the results and writing the manuscript. Maryam Azimi collaborated on the subjective tests and discussion of results for work conducted in [P7]. The entire work was conducted under the supervision and with editorial input from Dr Panos Nasiopoulos and Dr Rabab Ward. The thesis was written by Sima Valizadeh, with editing assistant from Dr. Panos Nasiopoulos and Dr. Rabab Ward.

The work in Chapter 3 is the continuation of the work in Chapter 2. Also, the work in Chapter 4 is the extension of the work in Chapter 2 for stereo and multiview videos. Chapter 5 presents stand-alone work for stereo videos.

- [P1] S. Valizadeh, P. Nasiopoulos and R. Ward, "Improving Compression Efficiency of HEVC Using Perceptual Coding," *submitted*.
- [P2] S. Valizadeh, P. Nasiopoulos and R. Ward, "Perceptual Rate Distortion Optimization of 3D-HEVC using PSNR-HVS," *submitted*.
- [P3] S. Valizadeh, P. Nasiopoulos and R. Ward, "Perceptual distortion measurement in the coding unit mode selection for 3D-HEVC," in *2016 IEEE International Conference on Consumer Electronics (ICCE), Vegas, USA, 2016*, pp. 347-350.
- [P4] S. Valizadeh, P. Nasiopoulos and R. Ward, "Optimizing the lagrange multiplier in perceptually-friendly high efficiency video coding for mobile applications," in *2016 International Conference on Computing, Networking and Communications (ICNC), Kauai, USA, 2016*, pp.1-4.
- [P5] S. Valizadeh, P. Nasiopoulos and R. Ward, "Perceptually-friendly rate distortion optimization in high efficiency video coding," in *2015 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 2015*, pp. 115-119.
- [P6] S. Valizadeh, M. Azimi and P. Nasiopoulos, "Bitrate reduction in asymmetric stereoscopic video with low-pass filtered slices," in *2012 IEEE International Conference on Consumer Electronics (ICCE), Vegas, USA, 2012*, pp. 170-171.
- [P7] M. Azimi, S. Valizadeh and , P. Nasiopoulos, "Subjective study on asymmetric stereoscopic video with low-pass filtered slices," in *2012 International Conference on Computing, Networking and Communications (ICNC), Maui, USA, 2012*, pp. 719-723.

Table of Contents

Abstract.....	ii
Lay Summary	iv
Preface.....	v
Table of Contents	vii
List of Tables	x
List of Figures.....	xiii
List of Abbreviations	xvii
Acknowledgements	xx
Dedication	xxi
Chapter 1: Introduction and Overview	1
1.1 High Efficiency Video Coding	3
1.1.1 Coding Units	4
1.1.2 Prediction Structures	6
1.1.3 Rate Distortion Optimization.....	7
1.2 Video Quality Metrics.....	13
1.2.1 2D Video Quality Metrics.....	13
1.2.2 3D Video Quality Metrics.....	15
1.3 Multiview Video Coding	18
1.3.1 Coding of Dependent Views.....	20
1.3.2 Coding of Depth Maps.....	23
1.4 Asymmetric Stereoscopic Video Coding.....	27
	vii

1.5	Thesis Contributions	30
Chapter 2: Perceptual Coding of HEVC		33
2.1	Proposed Method	34
2.1.1	Perceptual Mode Decision	34
2.1.2	Scaled Lagrange Multiplier.....	37
2.1.3	Optimal Lagrange Multiplier	41
2.2	Experimental Results	44
2.2.1	Test Video Sequences	44
2.2.2	Scaled Lagrange Multiplier.....	47
2.2.3	Optimal Lagrangian Multiplier	51
2.2.4	Higher Resolution Tests.....	53
2.2.5	Prediction Structures	54
2.2.6	Visual Comparison.....	56
2.2.7	Subjective Tests	56
2.2.8	Encoding Time.....	59
2.3	Conclusions.....	60
Chapter 3: Content Adaptive Perceptual Video Coding.....		62
3.1	Observations	63
3.2	Lagrangian Multiplier Estimation.....	65
3.3	Experimental Results	66
3.4	Conclusions.....	70
Chapter 4: Perceptual 3D Video Coding		71
4.1	Overview of Inter-component Dependencies in 3D-HEVC	72

4.1.1	Stereo Video Coding.....	72
4.1.2	Multi-view Video Coding.....	74
4.2	Proposed Method	76
4.3	Experimental Results	82
4.3.1	Test Setup.....	82
4.3.2	Compression Performance for Stereo Videos.....	83
4.3.3	Asymmetric Coefficients	85
4.3.4	Compression Performance of Multi-view Videos	87
4.3.5	Subjective Evaluation	90
4.3.6	Complexity Overhead.....	92
4.4	Conclusions.....	93
Chapter 5: Asymmetric 3D Video Coding.....		94
5.1	A New Method for Asymmetric Video Coding.....	94
5.2	Experimental Results	97
5.3	Conclusions.....	105
Chapter 6: Conclusions and Future Work		107
6.1	Significance and Potential Application of the Research.....	107
6.2	Summary of Contributions.....	108
6.3	Directions of Future Work.....	110
References.....		112

List of Tables

Table 1.1 QP values for corresponding quantization step sizes.....	9
Table 1.2 weighting factor for the Lagrangian multiplier in HEVC.....	13
Table 2.1 Bitrate saving of the proposed approach compared to the reference HEVC software for different scaling factors for the <i>BQSquare</i> video sequence.....	40
Table 2.2 Test video sequences	45
Table 2.3 Spatial and Temporal Information Content of the Test Video Sequences	47
Table 2.4 Bitrate and quality of the proposed method along with the HEVC reference software for the test video sequence BQSquare. The quality is measured by PSNR-HVS.....	48
Table 2.5 Bitrate and quality of the proposed method along with the HEVC reference software for various test video sequences	49
Table 2.6 Rate reduction of the proposed method with scaled Lagrangian multiplier compared to HEVC reference software (over QPs of 22, 27, 32 and 37)	51
Table 2.7 Rate reduction of the proposed method with optimal Lagrangian multiplier compared to HEVC reference software (over qps of 22, 27, 32 and 37)	52
Table 2.8 Bitrate saving of the proposed approach compared to the reference HEVC for the Higher Resolution test video sequences	53
Table 2.9 Bitrate saving of the proposed approach compared to the reference HEVC for the test video sequences with AI, RA, LDP and LDB prediction structure	55
Table 2.10 Comparison between encoding time of the proposed algorithm and HEVC reference software.....	59
Table 2.11 Encoding time Geomean ratio for the proposed approach compared with the reference HEVC software for different prediction structures.....	60

Table 3.1 Rate distortion curve fitting parameters.....	64
Table 3.2 Bitrate and quality of the proposed content-adaptive method along with the HEVC reference software for various test video sequences.....	68
Table 3.3 Rate reduction of the proposed method compared to the HEVC reference software...	69
Table 4.1 BD Rate saving based on scaling factors for the Lagrangian multiplier in coding depth maps	82
Table 4.2 Test video sequences	83
Table 4.3 Bitrate and quality of the reference 3D-HEVC and the proposed approach for stereo <i>PoznanStreet</i> test video sequence	84
Table 4.4 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for natural scene stereoscopic videos.....	84
Table 4.5 Bitrate saving of the base view (video0) for stereo test sequence <i>PoznanStreet</i>	85
Table 4.6 Bitrate saving of the dependent view (video1) for stereo test sequence <i>PoznanStreet</i>	86
Table 4.7 Overall bitrate saving of the texture views and the synthesized views for multi-view test sequence <i>PoznanStreet</i>	87
Table 4.8 Bitrate and quality of the reference 3D-HEVC and the proposed approach for multiview <i>PoznanStreet</i> test video sequence for texture views.....	87
Table 4.9 Bitrate and quality of the reference 3D-HEVC and the proposed approach for multiview <i>PoznanStreet</i> test video sequence for texture views and synthesized views.....	88
Table 4.10 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for multi-view videos. Video0 is the base view, video1 and video2 refer to the dependent views.....	89

Table 4.11 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for multi-view videos. Video(2v) and video(3v) refer to two and three texture views. Synth(2v) and synth(3v) correspond to the synthesized views. All(2v) and All(3v) refer to all the texture views and synthesized views.	89
Table 4.12 Comparison between encoding time of the proposed approach and 3D-HEVC reference software for the multi-view video coding	92
Table 5.1 Parameters employed for our experiments	99
Table 5.2 Bitrate in kbps for the two video sequences	103

List of Figures

Figure 1.1 block diagram of an HEVC encoder	3
Figure 1.2 Example for the partitioning of a 64*64 coding tree unit (CTU) into coding units (CUs) and its associated quad-tree.....	5
Figure 1.3 Graphical representation of an All Intra (AI) configuration	6
Figure 1.4 Graphical representation of Random Access (RA) configuration.....	6
Figure 1.5 Graphical representation of Low Delay P (LDP) configuration	7
Figure 1.6 Rate distortion curve with cost function $D + \lambda R$	8
Figure 1.7 (a) Flow chart for coding unit mode decision in HEVC. (b) HEVC quad-tree decomposition: coding tree unit (CTU), coding unit (CU), and prediction unit (PU).....	10
Figure 1.8 Stereoscopic quality (Q_s) and monoscopic quality (Q_m) for a stereo pair	16
Figure 1.9 Quality estimation of stereo pairs using original left and right views (R_{ref} and L_{ref}) compared with the degraded versions (R_{test} and L_{test}) along with the original disparity map compared to the degraded disparity map	17
Figure 1.10 $Q_{cyclopean}$ and Q_{depth} for multiview plus depth format	18
Figure 1.11 3D-HEVC structure and its inter-component dependencies for one frame.....	19
Figure 1.12 3D-HEVC prediction structure for various frames for a 3-view case	20
Figure 1.13 Disparity compensated prediction	21
Figure 1.14 Deriving motion vector for a block in the current picture (blue rectangle) based on the motion vector of the disparity compensated block in the base view	22
Figure 1.15 Sample depth map (left picture) associated with a texture view (right picture).....	23
Figure 1.16 Partitioning of a block in depth map (a) wedgelet (b) contour.....	24
Figure 1.17 Overview of 3D-HEVC coding tools for texture and depth components	25

Figure 1.18 Synthesized view distortion change calculated for choosing depth map coding modes	26
Figure 2.1 Rate-Distortion Curves for <i>BQSquare</i> video sequence for three different scaling factors. Quality is measured with Mean PSNR-HVS. (a) Scaling factor=1.6. (b) Scaling factor=5.6. (c) Scaling factor=8 and (d) Scaling factor=16.....	39
Figure 2.2 Bitrate savings versus the scaling factor for different video sequences	41
Figure 2.3 Equal quality contours (blue lines) for the BQSquare video sequence. On each projected contour, the point associated with the least amount of bitrate (yellow dot) is selected. These points show the optimal operating points and lead to minimum amount of bitrate at the same level of distortion.....	43
Figure 2.4 Optimal coefficients based on the quantization parameter. Based on this plot the (linear) relation between $\log(\text{coef})$ and QP is obtained.	44
Figure 2.5 First frame of test video sequences	46
Figure 2.6 Quality versus bitrate plots comparing the proposed approach with scaled Lagrange multiplier with HEVC reference.....	50
Figure 2.7 Rate-distortion curves for our proposed approach with optimal Lagrange multiplier compared to the HEVC reference	52
Figure 2.8 Rate-distortion curves for 4k video sequence, <i>Traffic</i> . Our proposed approach with scaled lambda is shown on the left and our method with optimal lambda is shown on the right side.	54
Figure 2.9 Visual comparison of the proposed method with the reference HEVC software for part of the BQSquare video with QP=37. The picture on the left is the original picture, the picture in	

the middle is compressed with HEVC and the picture on the right is compressed with our proposed method.....	56
Figure 2.10 Subjective tests structure	57
Figure 2.11 Subjective tests results. Quality (measured by mean opinion score) versus bitrate plots comparing the proposed approach with HEVC reference for different video sequences....	58
Figure 3.1 Quality versus bitrate for perceptual coding and HEVC.....	63
Figure 3.2 Rate distortion curve fitting for various test sequences.....	64
Figure 3.3 Tangent line approximation.....	66
Figure 3.4 Lagrange multiplier versus quantization parameter	67
Figure 3.5 Quality versus bitrate plots comparing the proposed content-adaptive approach with the HEVC reference	69
Figure 4.1 Basic codec structure for compression of the base view and the dependent view	73
Figure 4.2 3D-HEVC prediction structure for various frames for a 2-view case.....	73
Figure 4.3 3D video data format for auto-stereoscopic displays	74
Figure 4.4 coding 3D content with depth map.....	77
Figure 4.5 Bitrate savings for different natural-scene 3D video sequences	80
Figure 4.6 Bitrate savings for different computer-generated 3D video sequences.....	80
Figure 4.7 BD Rate saving based on scaling factors for the Lagrangian multiplier in coding depth maps	82
Figure 4.8 First frame of the test video sequences. (a) PoznanStreet (b) Newspaper (c) Kendo (d) PoznanHall (e) GhostTownFly (f) UndoDancer.....	83
Figure 4.9 Bitrate saving of the dependent view (video1) for stereo test sequence <i>PoznanStreet</i>	86
Figure 4.10 Subjective tests results for video <i>PoznanHall</i>	91

Figure 4.11 Subjective tests results for video <i>PoznanStreet</i>	91
Figure 5.1 Illustration of an example of our filtering pattern. Grey areas represent filtered slices in each view of the stereoscopic video.	95
Figure 5.2 (a) left and right frames with unsmoothed slice borders. (b) left and right frames with smoothed borders	97
Figure 5.3 The first frame of our two video sequences. (a)First video sequence: Mother and Kid (b) Second video sequence: Two Dolls. These frames are the right view of our stereoscopic videos	98
Figure 5.4 Sharpness, depth and quality of (a) video 1 and (b) video 2 averaged over the viewers.	101
Figure 5.5 Sharpness, depth and quality of (a) video 1 and (b) video 2 averaged over the viewers. Smoothing the borders of the slices provide better quality and sharpness.	103
Figure 5.6 Mean Opinion Score of the first stereoscopic video sequence versus the required bitrate with H.264 compression	104
Figure 5.7 Mean Opinion Score of the second stereoscopic video sequence versus the required bitrate with H.264 compression	105

List of Abbreviations

AI	All Intra
AVC	Advanced Video Coding
BD	Bjontegaard's Delta
CB	Coding Block
CfP	Call for Proposals
CSF	Contrast Sensitivity Function
CTB	Coding Tree Block
CU	Coding Unit
CTU	Coding Unit Tree
DBBP	Depth-Based Block Partitioning
DCP	Disparity Compensation Prediction
DCT	Discrete Cosine Transform
DIBR	Depth-Image-Based Rendering
DM	Distance Measure
DSCQ	Double Stimulus Continuous Quality
DSIS	Double Stimulus Impairment Scale
DVQ	Digital Video Quality
HD	High Definition
HEVC	High Efficiency Video Coding
HVS	Human Visual System
JCT-3V	Joint Collaborative Team on 3D Video coding
JNDs	Just Noticeable Differences
LDB	Low Delay B Pictures
LDP	Low Delay P Pictures
MB	Macroblock
MCP	Motion Compensation Prediction
MOS	Mean Opinion Score
MOVIE	Motion-based Video Integrity Evaluation

MPEG	Moving Picture Experts Group
MPI	Motion Parameter Inheritance
MPQM	Motion Picture Quality Metric
MS-SSIM	Multi-scale Structural SIMilarity index
MVC	Multi-view Video Coding
MVD	Multi-view plus Depth
PB	Prediction Block
PEVQ	Perceptual Evaluation of Video Quality
PQSM	Perceptual Quality Significance Map
PSNR	Peak Signal to Noise Ratio
PVQM	Perceptual Video Quality Measure
PU	Prediction Unit
QP	Quantization Parameter
RA	Random Access
RDO	Rate Distortion Optimization
RDOQ	Rate–distortion optimized quantization
SAD	Sum of Absolute Differences
SCS	spatial contrast sensitivity
SSD	Sum of Squared Differences
SSE	Sum of Squared Error
SI	Spatial Information
SSIM	Structural SIMilarity index
TB	Transform Block
TI	Temporal Information
UHD	Ultra-High Definition
VCEG	Visual Coding Experts Group
VIF	visual information fidelity
VSNR	Visual signal to Noise Ratio
VSSIM	Video Structural SIMilarity index
WVGA	Wide Video Graphics Array

WQVGA

Wide Quarter Video Graphics Array

3D-QoE

3D Quality of Experience

3D-HEVC

3 Dimensional High Efficiency Video Coding

Acknowledgements

I want to express my sincere gratitude to my supervisors, Dr. Panos Nasiopoulos and Dr. Rabab Ward, for their guidance and support throughout my PhD program. I thank Dr. Nasiopoulos and Dr. Ward for their encouragement, motivation and insightful suggestions. They have been great mentors and role models to me.

I also thank my lab-mates at the Digital Multimedia Lab, Dr. Lino Coria, Dr. Di Xu, Dr. Bambang Sarif, Dr. Hamidreza Tohidipour, Dr. Amin Banitalebi Dehkordi, Dr. Ronan Boitard, Basak Oztas, Anahita Shojaei, Pedram Mohammadi, Cara Dong, Maryam Azimi, Mohsen Amiri, Sergio Infante, Ilya Koreshev, Abrar Wafa, Ilya Ganelin, Stelios Ploumis, Fujun Xie, Ahmad Khaldieh, Joseph Khoury and all other colleagues.

I would like to thank my friends at the University of British Columbia, Dr. Parisa Behnamfar, Dr. Neda Eskandari, Dr. Mani Malek Esmaeili, Dr. Yashar Komijani, Dr. Mohammad Ghasemi, Dr. Mehdi Piltan, Dr. Mohammad Najafi, Saeid Allahdadian, Hamid Palangi, and Hossein Bashashati.

Special thanks are owed to my parents, my sister and my dear friend, Farzin Ferdis, for their love and support.

Dedication

To my parents, *for your incredible strength and love*

To my sister, *for your constant faith and kindness*

And,

To Farzin, *for your inspiration and support*

Chapter 1: Introduction and Overview

Advances in video compression techniques have enabled the development of videos of high visual quality with constrained bitrates. These advances are important for video transmission and storage purposes, considering the limited bandwidth of communication channels and the limited capacity in storage media. They are also crucial for many new services that depend on very high video compression efficiency at very high resolution video, such as in sophisticated multimedia applications and ultra-high definition television. For very high video compression, a new video coding standard was recently (2013) developed by collaboration between the ITU-T Visual Coding Experts Group (VCEG) and the ISO-IEC Moving Picture Experts Group (MPEG). The new standard is called High Efficiency Video Coding (HEVC).

In order to achieve high video quality while keeping the bitrate value within constraints, exploiting the characteristics of the human visual system (HVS) can play an important role. As for the vast majority of applications, humans are the ultimate consumers and judges of the video quality, it is how people perceive the quality of the video that matters the most. Therefore, it is important to use video quality metrics that are based on the human visual system. It is also crucial to integrate such metrics within the video encoder to improve the quality of the compressed video.

One of the exciting extensions of the HEVC standard is its application to 3D videos (enabled by stereoscopic and multiview representations). The 3D-HEVC standard considers new capabilities such as the use of depth maps for view-synthesis techniques. In stereoscopic videos, two views (one for the left eye and the other for the right eye) need to be transmitted or stored. This doubles the amount of data required for transmission or storage compared to monoscopic

video. For multiview videos, two or three different views of the same scene as well as the corresponding depth maps of these views need to be transmitted so that the receiver can create multiview videos. The huge amount of data in stereoscopic and multiview videos necessitate the employment of efficient compression techniques (for stereo and multiview videos) in practical applications.

In this thesis, we propose new methods for improving the compression efficiency of HEVC and 3D-HEVC which is the extension of HEVC to 3D videos. In Chapter 2, we investigate the integration of a perceptual video quality metric (instead of sum of squared errors) inside the rate distortion optimization process of HEVC and find the optimal Lagrange multiplier that leads to the highest compression efficiency in HEVC. In Chapter 3, we adjust the value of the Lagrange multiplier based on the characteristics of the video content. This content-adaptive approach can further improve the compression efficiency of HEVC. In Chapter 4, we extend our work to 3D-HEVC, for stereo and for multiview videos. Finally, in Chapter 5, we focus on stereoscopic videos and evaluate a novel asymmetric method for increasing their compression efficiency.

The following sections in this introductory chapter provide background information and a literature review on the topics addressed in each of the research chapters. Section 1.1 provides basic background information on HEVC, the new coding structure it uses and the process of rate distortion optimization. Section 1.2 surveys video quality metrics including the most popular perceptual video quality metrics. Section 1.3 gives an overview of 3D-HEVC tools. Asymmetric approaches for 3D video coding are presented in section 1.4. Section 1.5 gives an overview of the research contributions presented in this thesis.

1.1 High Efficiency Video Coding

HEVC is a block based video coding standard as illustrated in Figure 1.1. An HEVC encoder partitions each picture into blocks. Each block is coded in either an intra or an inter mode. Intra prediction uses data from spatial neighbor blocks while inter prediction uses data from other frames by exploring temporal dependencies between frames using motion estimation. The first picture of a video sequence is encoded as an intra mode. For all the other remaining frames of the video sequence, the encoder has to decide whether to use intra or inter mode for each partition within the frames. The difference between the original block and its predicted block forms the residual signal. Each residual block is DCT transformed, scaled, quantized and entropy coded to form the output bitstream.

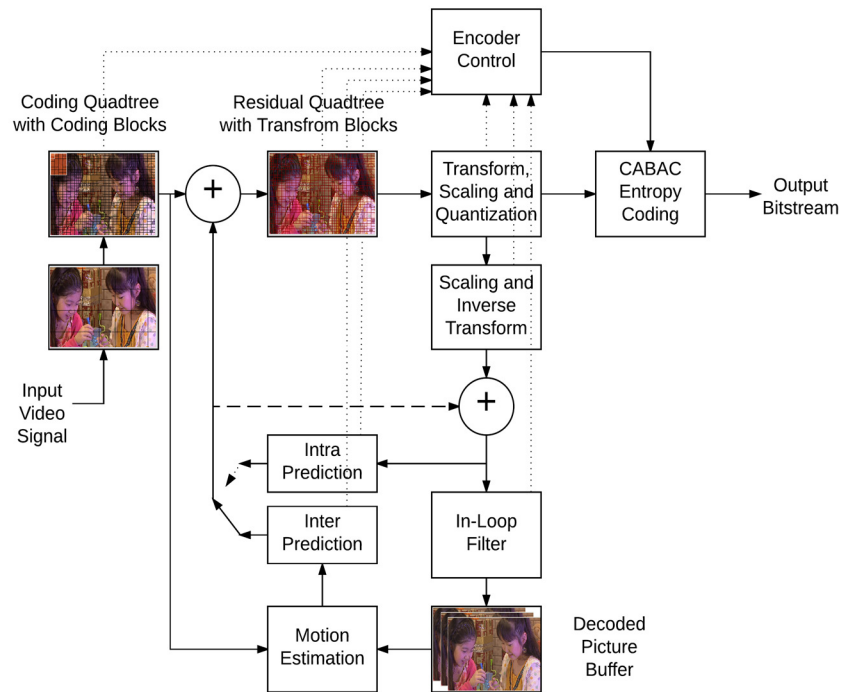


Figure 1.1 block diagram of an HEVC encoder

The HEVC project had great participation from industry and research groups. The first draft of HEVC was finalized in 2013 [1]. The ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) issued a joint call for proposal in January 2010 [2]-[3]. Prior to this call, investigations were carried out to study the feasibility of obtaining significant improvements [4] in the compression efficiency of the video coding standards compared to the previous major standard called H.264/AVC[5]-[8].

HEVC introduces new tools such as variable size for coding units and prediction units, extended intra prediction modes, advanced motion vector prediction and a block merge method [9]. HEVC was shown to provide 35.4% bitrate savings (at the same image fidelity) in comparison to the previous video coding standard, H.264/AVC[10]. It is reported that HEVC provides comparable video quality at half or less than half the bit rate of the H.264/AVC in 92.5% of the test cases [11]. Experiments with the new video coding standard, HEVC, show that more than half of its average bit-rate savings relative to its predecessor, H.264/MPEG-4 AVC, comes from its increased flexibility in block partitioning for prediction and transform [12].

1.1.1 Coding Units

HEVC provides more flexibility in block partitioning for prediction and transform coding compared to the previous video coding standard, H.264 [13]. This flexibility is offered by the introduction of the coding tree blocks (CTBs), coding blocks (CBs), prediction blocks (PBs), transform blocks (TBs) and quad-tree based block partitioning in HEVC. Based on the experiments in [12], more than half of the average bitrate savings of HEVC compared to H.264 comes from its increased flexibility in block partitioning for prediction and transform.

Each coding tree block (CTB) is the root of a coding quad-tree structure. The coding quad-tree structure partitions CTB into coding blocks (CBs) for which the encoder has to decide between using intra or inter modes in each unit. Figure 1.2 illustrates how CTBs can be split into multiple CBs. The quad-tree structure provides a hierarchical block partitioning to support larger sizes in dealing with high definition (HD) and ultra-high definition (UHD) videos, and at the same time, adapt to the local properties of the frames. The size of the coding block can range from 8*8 to the size of CTB (typically 64*64 pixels).

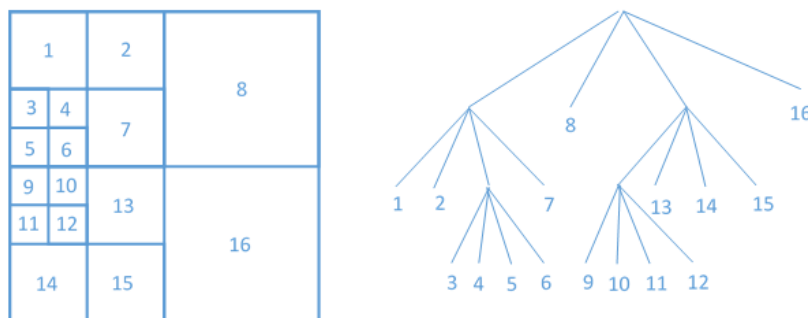


Figure 1.2 Example for the partitioning of a 64*64 coding tree unit (CTU) into coding units (CUs) and its associated quad-tree

In H.264, the previous video coding standards, macroblocks (MBs) with a fixed size of 16*16 pixels were used as the basic processing units. For each MB, the coding mode (inter or intra) was selected by the encoder. The MB size of 16*16 luma samples were the largest for motion prediction. As the interest in HD and UHD videos grew, restricting the largest block size (used for signaling prediction parameters) to 16*16 macroblock would consume a lot of bitrate. Also, transform sizes that are larger than 16*16 can better explore the spatial correlation in neighboring blocks in high resolution videos. In HEVC, the size of the coding tree block can be 16*16, 32*32 or 64*64.

1.1.2 Prediction Structures

HEVC supports four different prediction structures [14]: All Intra (AI), Random Access (RA), Low Delay P pictures (LDP) and Low Delay B pictures (LDB). In AI, every frame is coded as I (intra) frame. There is no inter prediction in AI configuration setting. Thus, it is suitable for low delay and higher bitrate applications. Figure 1.3 shows an All Intra configuration. The quantization parameter (QP) will stay constant for all the frames (QP offset=0).

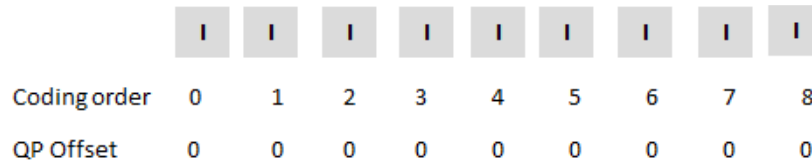


Figure 1.3 Graphical representation of an All Intra (AI) configuration

In the RA prediction configuration, a hierarchical B structure is used as the frame prediction order. Figure 1.4 shows an example of this prediction structure. This structure achieves high coding efficiency but has a larger delay due to reordering of the pictures. Quantization Parameter (QP) is modified by adding a QP offset value. QP offset values for each picture is shown in Figure 1.4. These QP offset values are set by the HEVC standard.

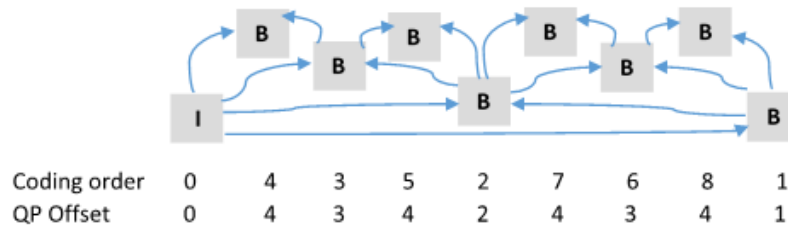


Figure 1.4 Graphical representation of Random Access (RA) configuration

The Low Delay P pictures (LDPs) and the Low Delay B pictures (LDBs) prediction structures do not allow the re-ordering of pictures. The first picture is coded as I picture and the subsequence pictures are coded as P pictures in LDP and B pictures in LDB. LDB achieves higher compression efficiency than LDP because bi-prediction is allowed in LDB. Figure 1.5 shows an example of a Low Delay P configuration.

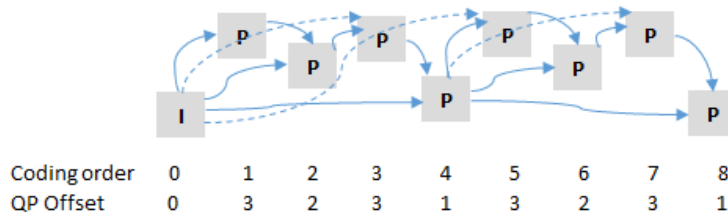


Figure 1.5 Graphical representation of Low Delay P (LDP) configuration

1.1.3 Rate Distortion Optimization

The job of the encoder is to select coding parameters in a way that results in the best coding efficiency. In video coding, the bitrate is minimized for a certain fixed amount of distortion, or the distortion is minimized for a certain fixed bitrate:

$$\min(D) \text{ subject to } R \leq R_T \tag{1.1}$$

where the minimization is done over the coding parameters, D is distortion, R is the number of bits required for the compressed video and R_T is the upper bitrate allowed. This minimization process is formulated via a non-negative Lagrange multiplier λ in the Rate Distortion Optimization (RDO) [15] process:

$$\min(D + \lambda \times R) \tag{1.2}$$

Any solution of equation (1.2) is a true solution for the minimization problem in equation (1.1). The proof for the general problem is given in [16].

A typical rate distortion curve is shown in Figure 1.6. Distortion is a non-increasing convex function of rate. In this figure, the straight line denotes Lagrangian cost function. On the rate-distortion curve, the minimum cost occurs at the tangent point to the line with slope of $-\lambda$.

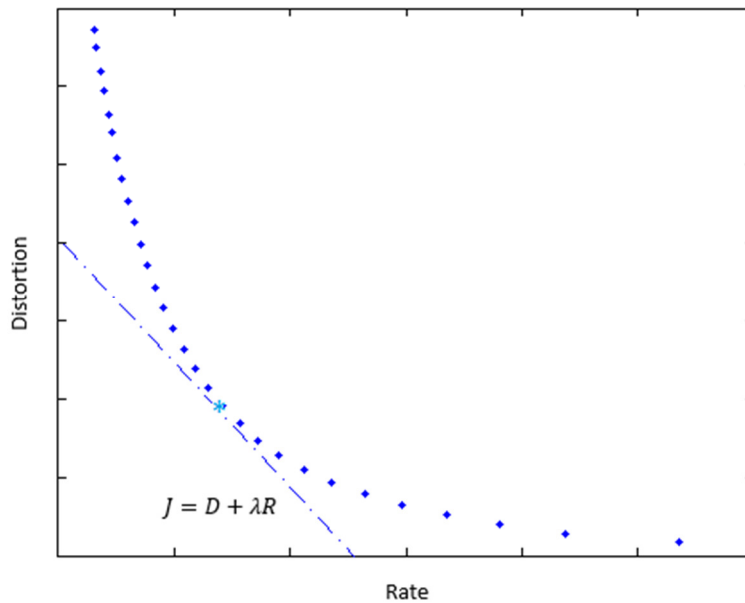


Figure 1.6 Rate distortion curve with cost function $D + \lambda R$

λ acts as a knob in the rate distortion tradeoff. Smaller values for λ , leads to minimizing mainly the distortion term while larger values for λ lead to minimizing the rate term. The quantization step size is defined as twice of Quant ($Quant = \frac{1}{2}q_{step}$). In [15], a relationship between λ and Quant is established:

$$\lambda = 0.85 \times Quant^2 \quad (1.3)$$

$$Quant = \frac{1}{2} q_{step} \quad (1.4)$$

In HEVC and AVC, the quantization step size is controlled by a Quantization Parameter (QP). Quantization step size is doubled by every increase of 6 in QP. The quantization step size associated with each QP value is shown in Table 1.1.

Quantization Parameter (QP)	6	12	18	24	30	36	42	48	51
Quantization Step Size	1.25	2.5	5	10	20	40	80	160	224

Table 1.1 QP values for corresponding quantization step sizes.

The choice of the quantization step size is sent to the decoder. However, parameter λ is only used at the encoder.

There are spatial/temporal dependencies between blocks. However, for practical applicability, these dependencies are ignored when the minimization is divided into stages. Otherwise, minimization of each frame distortion or minimization of an average frame distortion taken (over many video frames) can be considered. In such cases, the increased computational complexity and the excessive delay that such approaches incur prevent their practical implementation [17]. Instead, the overall minimization process is split into a series of smaller minimization problems. Partitioning the frame into coding units (CUs) with various sizes is one of the improvements offered by HEVC compared to H.264.

In HEVC, the coding parameters of each block have to be determined for the four stages:
 1. Coding unit (CU) quad-tree structure 2. Intra prediction mode candidates 3. Inter prediction

motion parameters and reference list for motion estimation 4. Rate-distortion optimized quantization (RDOQ), for quantization process [12].

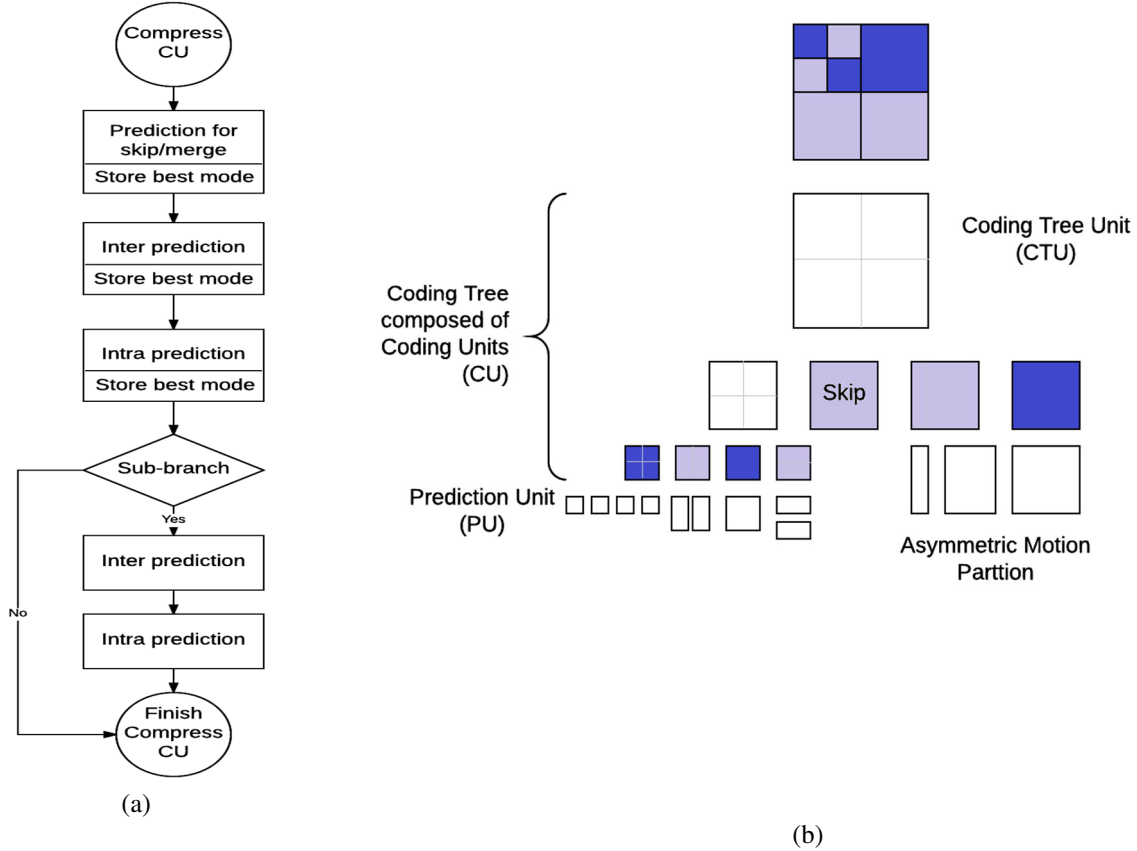


Figure 1.7 (a) Flow chart for coding unit mode decision in HEVC. (b) HEVC quad-tree decomposition: coding tree unit (CTU), coding unit (CU), and prediction unit (PU)

In the rate distortion optimization of HEVC, coding unit level (CU) mode decision (intra vs. inter) is based on finding the coding parameters that minimizes:

$$J_{mode} = SSE + \lambda_{mode} \times R_{mode} \tag{1.5}$$

where SSE is the sum of squared errors between the original and reconstructed CU blocks, R_{mode} is the total number of bits used for coding the CU. For finding the best inter CU coding

cost, J_{mode} is evaluated for all possible partition modes and a partition that gives the minimum coding cost is chosen. The best intra prediction mode is the one that gives the minimum J_{mode} among candidate intra prediction modes in the candidate list. By applying this CU level mode decision at each level of CU recursion tree, coding unit structure is obtained. Figure 1.7.a shows the flow chart for this process. HEVC quad-tree decomposition is illustrated in Figure 1.7.b.

In the rate distortion optimization process, motion estimation is carried out by:

$$mp^* = \arg \min D_{mp} + \lambda_{pred} \times R_{mp} \quad (1.6)$$

where D_{mp} is calculated by the sum of absolute differences (SAD) between the original PU block and its motion-compensated block. R_{mp} is the estimation of the number of coded bits required to transmit motion prediction. Motion estimation for each inter PU partition is broken into two parts: integer-sample precision and sub-sample precision. For integer-sample precision, sum of absolute differences (SAD) is used. For sub-sample precision, the distortion is measured by applying Hadamard transform to the difference between the original block and the motion-compensated reference block.

Intra prediction in the rate distortion optimization process is carried out by:

$$p^* = \arg \min D_p + \lambda_{pred} \times R_p \quad (1.7)$$

where D_p is measured by applying Hadamard transform to the difference between the original block and the prediction block. R_p is the number of bits required for the prediction mode. For intra prediction, the minimization process is performed in two parts. In the first part, the prediction cost function is minimized over 33 angular prediction directions, the planar and the

DC mode giving a fixed number of mode candidates for intra prediction. In this part, distortion is measured by applying Hadamard transform to the block difference between the original block and its prediction block. Rate is measured by the number of coded bits required to indicate the prediction mode of the block. In the second part, the list of candidate modes is augmented by a list of three most probable modes. The best block candidate is determined in the rate distortion optimization process.

During the last stage (i.e. the quantization process) of the rate distortion optimization, the levels of the transform coefficients are adjusted. The distortion term is determined in the transform domain.

For coding unit mode decision, the distortion is measured by the sum of squared errors (SSE) between the original and the reconstructed blocks while λ_{mode} is defined as [18]:

$$\lambda_{mode} = \alpha \times w_k \times 2^{((QP-12)/3.0)} \quad (1.8)$$

where QP is the quantization parameter, α is defined as [19]:

$$\alpha = \begin{cases} 1.0 - Clip3(0,0.5,0.05 \times number_of_B_frames) & \text{for referenced pictures} \\ 1.0 & \text{for non_referenced pictures} \end{cases} \quad (1.9)$$

$$Clip3(x, y, z) = \begin{cases} x; z < x \\ y; z > y \\ z; otherwise \end{cases} \quad (1.10)$$

w_k is a weighting factor based on encoding configuration and QP offset hierarchy level of current video frame [19]. Table 1.2 reports the values for w_k .

k	QP offset hierarchy level	Slice type	Referenced	W_k
0	0	I	-	0.57
1	0	GPB	1	RA: 0.442 LD: 0.578
2	1, 2	B or GPB	1	RA: $0.3536 * \text{Clip3}(2.0, 4.0, (QP-12)/6.0)$ LD: $0.4624 * \text{Clip3}(2.0, 4.0, (QP-12)/6.0)$
4	3	B	0	RA: $0.68 * \text{Clip3}(2.0, 4.0, (QP-12)/6.0)$

Table 1.2 weighting factor for the Lagrangian multiplier in HEVC

Measuring the distortion by SSE and minimizing this SSE in the rate distortion optimization process, leads to better PSNR in the coded video. However, the PSNR has been shown to have limited correlation with subjective tests. Various other video quality metrics have been developed to better represent how subjects perceive video quality. The study in [20] demonstrates the scope of validity of PSNR as a video quality metric. Based on this study, PSNR is reliable only when the video content and the codec are fixed in the test conditions. For a given video codec and a specified video content, the performance of different optimization settings can be measured by PSNR. However, PSNR is inaccurate in measuring the quality of video content that is encoded at different frame bitrates. Furthermore, PSNR is unreliable when several video streams are jointly assessed [21].

1.2 Video Quality Metrics

1.2.1 2D Video Quality Metrics

Perceptual video quality metrics are classified into methods that work in the pixel domain and the frequency domain [23]. Metrics that rely on the discrete cosine transform (DCT), wavelet transform and Gabor filter banks are used in the frequency domain. Pixel domain methods use local gradient changes around a pixel or extract visual features based on computational models of the low level vision. One of the earliest perceptual video quality metrics in the frequency domain consists of filtering and masking processes [24]. For filtering, it uses a nonlinear spatio-temporal

model and for masking, it uses point-by-point weighting. The Motion Picture Quality Metric (MPQM) [25] is based on a multi-channel model of human spatio-temporal vision. Digital Video Quality (DVQ) [26] estimates the local contrast in the image based on the ratio of its DCT amplitude to its DC component. From the local contrast, Just Noticeable Differences (JNDs) are estimated. An extension of DVQ, uses the fact that the human eyes' sensitivity to spatio-temporal patterns decrease with high spatial and temporal frequencies. This metric uses a spatial contrast sensitivity (SCS) matrix for static frames and SCS raised to a power for dynamic frames [27]. Another perceptual video quality metric in the frequency domain is based on the Wavelet Transform [28]. The Motion-based Video Integrity Evaluation (MOVIE) metric models the response characteristics of the middle temporal visual area with separable Gabor filter banks [29]. PSNR-HVS [30] is a full-reference frequency-domain perceptual video quality metric based on the characteristics of the human visual system. PSNR-HVS will be explained in detail in section 2.1.1.

In addition to the perceptual video quality metrics in the frequency domain, various pixel-domain perceptual quality metrics have been introduced in the literature. The Perceptual Video Quality Measure (PVQM) [31] uses a linear combination of three indicators including the edginess of the luminance computed using local gradient filter. Based on visual attention, the Perceptual Quality Significance Map (PQSM) measure has been introduced. This metric performs feature extraction, stimuli integration and post processing [32]. Another video quality metric is based on a combined measure of distortion-invisibility, block fidelity and content-richness fidelity [33]-[34]. Perceptually salient feature points are used in a full reference metric which uses Sobel filter to approximate the gradient of local luminance [35]. The Visual signal to

Noise Ratio (VSNR) [36] is based on the near-threshold and the supra-threshold of human vision. Wavelet-based models of visual masking and visual summation are used to compute contrast thresholds for detection of distortions in the presence of natural images. Perceptual Evaluation of Video Quality (PEVQ) [37] extracts regions of interest by pre-processing. After aligning them spatially and temporally, four distortion measures are calculated and using a sigmoid approach the distortions are mapped to the Mean Opinion Score (MOS) video measure.

Some of the video quality metrics use natural visual statistics or features. Examples of quality metrics based on natural visual statistics are SSIM [38], VSSIM [39], MS-SSIM [40], VIF [41], a method based on DCT and region classification [42] and a method based on Singular Value Decomposition (SVD) [43]. In order to integrate a video quality metric inside a block-based video encoder, the metric has to be easily adaptable to blocks of data and should not be too computationally complex. In Chapter 2, we explain our method of integrating a perceptual video quality metric inside the rate distortion optimization of HEVC, with the goal of increasing the compression efficiency of the video encoder. We extend our approach to 3D video coding in Chapter 4. Therefore, in the following section, we give a brief overview of 3D video quality metrics.

1.2.2 3D Video Quality Metrics

For integrating a perceptual quality metric inside the rate distortion optimization in 3D-HEVC, we need to understand the factors that affect the quality of 3D video. 3D quality of experience (3D-QoE) is affected by various factors such as naturalness, depth perception, image quality, viewing experience and visual comfort. Subjective tests have evaluated the effect of each visual cue on 3D-QoE. In [44], camera base distances, screen disparity, Gaussian noise and

Gaussian blur with various levels have been subjectively tested on stereoscopic images to build a 3D quality model as a weighted sum of image quality and perceived depth. In another work, monoscopic and stereoscopic quality components from a cyclopean image and a disparity map are combined into an overall score for the 3D quality [45]. Figure 1.8 shows the stereoscopic quality (Q_s) and monoscopic quality (Q_m) for a stereo pair. For a block in the left view, the best horizontal match in the right view is found. The peak position represents the perceived disparity and the peak value gives the stereo-similarity of the matched blocks. A cyclopean image is the average of a block in the left view with its disparity-matched block in the right view. Corresponding weights for combining Q_s and Q_m are estimated in [45] based on subjective experiments.

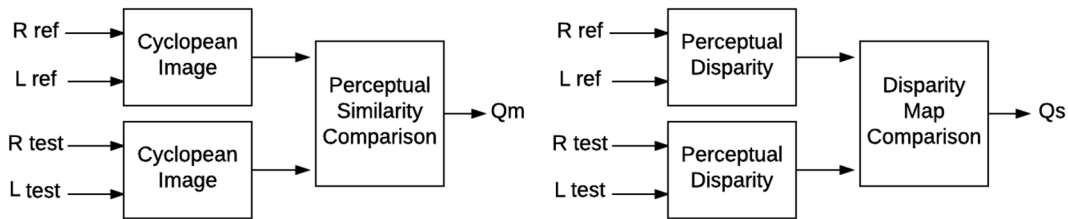


Figure 1.8 Stereoscopic quality (Q_s) and monoscopic quality (Q_m) for a stereo pair

In [46], quality scores of both stereo-pair and disparity map are computed by 2D quality metrics and are combined into a final score. This paper presents two approaches to obtain the final score. The first approach is presented in Figure 1.9 where the quality of the left and right views is calculated with a 2D quality metric and the average of the two values is obtained. Afterwards, the disparity distortion score is combined with the average image distortion score.

For disparity estimation, two disparity computation algorithms which are based on Markov random fields are selected. The first algorithm uses belief propagation for inference [47] and the second one uses graph cuts [48]. The correlation coefficients are calculated and used for comparing the original and the degraded disparity maps. Their work is then further extended by combining the local disparity distortion with the left and right quality estimations before taking the mean value to obtain the final quality score locally.

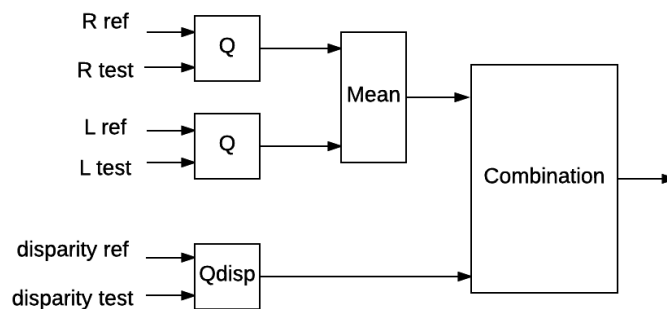


Figure 1.9 Quality estimation of stereo pairs using original left and right views (Rref and Lref) compared with the degrades versions (Rtest and Ltest) along with the original disparity map compared to the degraded disparity map

In [49], the quality of the cyclopean view as well as the quality of the depth maps are combined, and temporal pooling is used to account for temporal variations in the 3D video quality. Figure 1.10 shows the block diagram for $Q_{Cyclopean}$ and Q_{Depth} . The cyclopean view is generated from 3D-DCT of matching blocks. The lower frequency coefficients are multiplied by a matrix resembling the contrast sensitivity function (CSF). For depth maps, the quality as well as the local variance in each block is taken into account.

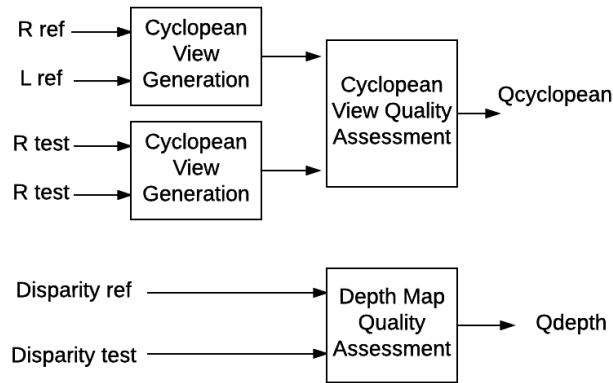


Figure 1.10 Qcyclopean and Qdepth for multiview plus depth format

The effectiveness of 2D quality metrics in stereoscopic image quality assessment is investigated in [50]. Experimental results show that better results are obtained using an appropriate combination of disparity information and original images. Moreover, this study in [50] shows that the performance of 2D quality metrics is promising for assessing the quality of stereoscopic images with the same type of distortion. The work in [51]-[53], tested the effectiveness of three 2D quality metrics on measuring the quality of compressed 3D video. Although subjective tests remain as the best judgement of 3D video quality, these works conclude that the average of the objective 2D quality measure of the left and the right view shows similar trend as the 3D perceived quality for their test videos.

1.3 Multiview Video Coding

Multiview and Stereo video coding standards aim at exploiting all the redundancies in the 3D content. Disparity compensation had been introduced in the Multi-view Video Coding (MVC) extension of the H.264/AVC [54]-[55]. However, there are more redundancies in 3D

videos, such as the redundancies in the motion parameters and in the residuals of the dependent views. Also, the depth data in the multi-view plus depth (MVD) format have unique characteristics that need to be exploited. In 2011, MPEG issued a call for proposals (CfP) for 3D video coding techniques with a specified set of requirements and the defined evaluation procedure [56]. ISO and ITU formed a joint collaborative team on 3D video coding extension development (JCT-3V). This team developed two 3D video coding standards, one based on the AVC (3D-AVC) framework and one based on HEVC (3D-HEVC). 3D-HEVC removes the correlation between texture and the depth components of the 3D video, by efficient predictions or inheritances of coding information from other components as well as by exploiting the unique characteristics of the depth maps [57]-[59]. Figure 1.11 shows the inter-component dependencies in time in 3D-HEVC for one frame. Figure 1.12 shows the dependencies between various frames and components.

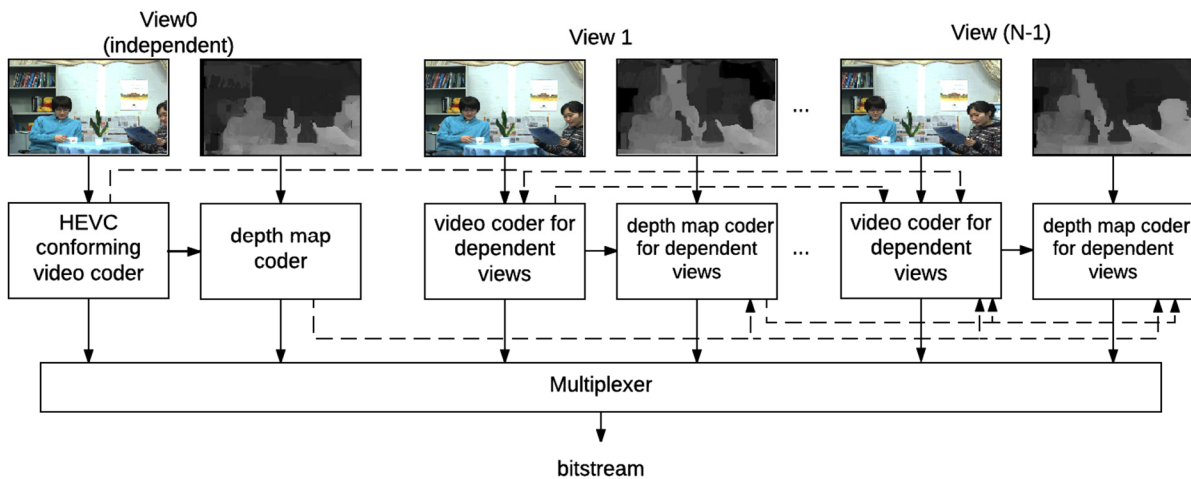


Figure 1.11 3D-HEVC structure and its inter-component dependencies for one frame

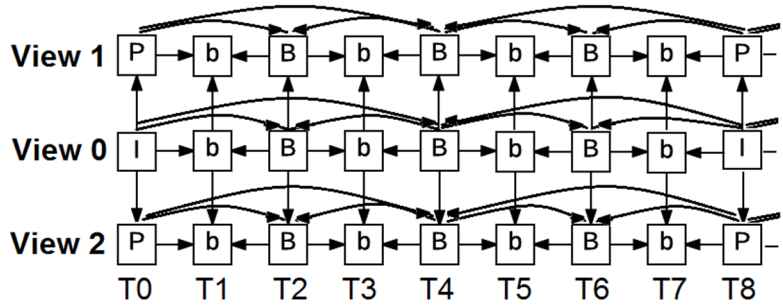


Figure 1.12 3D-HEVC prediction structure for various frames for a 3-view case

In comparison to HEVC simulcast, 3D-HEVC provides 50% bitrate savings. The inter-component tools in 3D-HEVC provide 20% bitrate savings, in comparison to disparity compensated coding [60].

The set of tools explained below were chosen as the basis for the test model under consideration [61] for HEVC-based 3D video coding, called MV-HEVC [62]. In MV-HEVC, two or more captured views plus their associated depth maps are transmitted. Additional views that are suitable for autostereoscopic displays will be generated in the display side. MV-HEVC can also be used for conventional multiview video coding without coding the depth data [63].

1.3.1 Coding of Dependent Views

In MV-HEVC, dependent views that take advantage of disparity compensation prediction (DCP) are used in addition to motion compensation prediction (MCP) to improve the coding efficiency. Figure 1.13 illustrates disparity compensation for the dependent view. In disparity compensation, pictures of the same access unit that have already been coded are added to the reference picture list. This modification does not need any change in the macroblock syntax or

decoding process. Disparity compensation prediction usually wins over motion compensation prediction in image areas that differ in two sequential frames due to motion. Other than that, for most cases MCP is used. Thus, there is a need to take advantage of other views more effectively. Inter-view motion parameter prediction and inter-view residual prediction have been implemented to increase the coding efficiency.

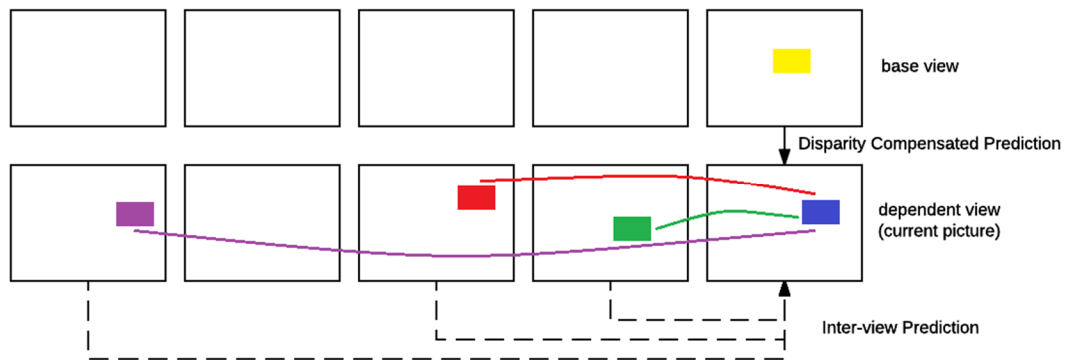


Figure 1.13 Disparity compensated prediction

Different views of a multiview video capture the same scene but from different angles. Thus, in most cases, there are similarities in the motion in the different views. Based on this fact, the motion parameters of a dependent view can be predicted from other views. For each block in the current view, a corresponding block in the reference view is found and the motion parameters of that block are added to the list of candidate motion parameters of the current block [64]. For determining the corresponding block, the disparity vector is used. A disparity vector is based on an estimate of the depth map. Even in the multiview plus depth coding, the depth map needs to be estimated since the pictures for each view are coded before the depth map is coded. Depth

map estimation is based on the already coded depth data of the previously-coded view by warping. In another configuration, the depth map is estimated from the already transmitted disparity vectors and motion parameters. Deriving the motion vector of a block, based on the motion vector of the disparity compensated block in the base view, is illustrated in Figure 1.14.

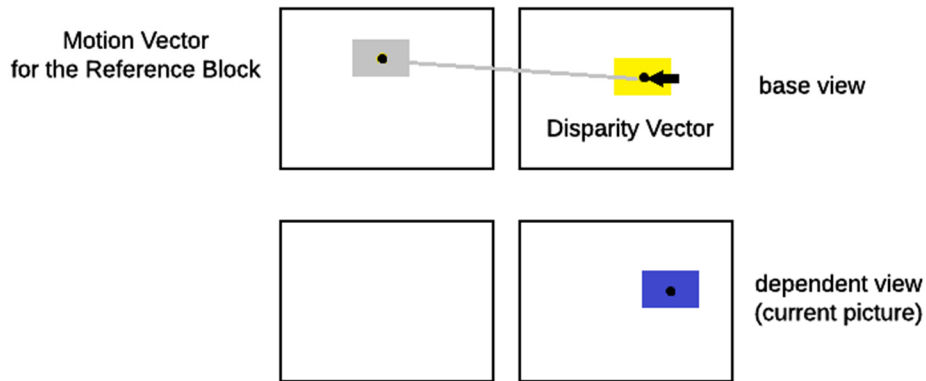


Figure 1.14 Deriving motion vector for a block in the current picture (blue rectangle) based on the motion vector of the disparity compensated block in the base view

In addition to inter-view motion parameter prediction, taking advantage of the similarities between inter-view residuals improves the coding efficiency. The use of inter-view residual prediction is signaled by adding a flag to the syntax of inter coded blocks. When inter-view residual prediction is used, the estimated depth map is employed to determine the disparity vector. The disparity vector points the corresponding block in the reference view. The residuals of that block are subtracted from the current residual and the difference is coded. A bilinear filter is used in cases where the disparity vector points to a sub-sample location.

In order to improve inter-view prediction, illumination mismatches between views should be removed. This process is called illumination compensation. Depth-based block partitioning (DBBP) assigns different motion compensation modes to both sides of a boundary in the texture view based on the available information in an already coded depth map.

1.3.2 Coding of Depth Maps

In general, the depth maps can be coded the same way as the video pictures. However, depth maps have different characteristics such as sharp edges and large areas of constant values as shown in Figure 1.15. HEVC is optimized for natural video. It can code areas of constant values well enough but for better representation of sharp edges, new intra coding modes have been introduced to the coding algorithm. Also, the partitioning and motion data of a picture view can be used for its depth map. In addition to these changes, to avoid ringing artifacts at the depth map edges, there is no interpolation in the motion compensation part. For depth map coding, the motion vectors are coded with sample (instead of quarter-sample) accuracy. Also, no in-loop filtering is used for depth map coding.

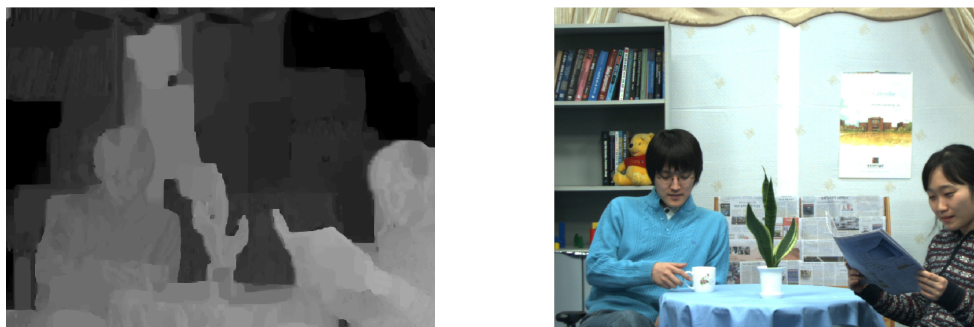


Figure 1.15 Sample depth map (left picture) associated with a texture view (right picture)

For better coding of depth maps, a new skip mode and three new prediction modes called Intra-single, Intra-Wedge and Intra-Contour have been added. Intra-Wedge and Intra-Contour modes partition a depth block into two areas. Each area has constant value and is not necessarily rectangular. Wedgelets and Contours are two types of partitioning. The first one is used for straight line partitioning and the second one is used for arbitrary division of the two areas within a block as shown in Figure 1.16. After partitioning, constant values in the two areas are predicted from neighboring blocks and their difference is transmitted [65].

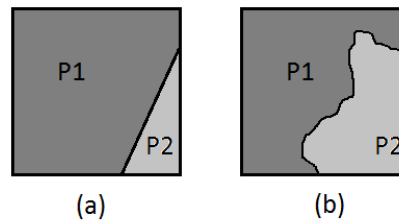


Figure 1.16 Partitioning of a block in depth map (a) wedgelet (b) contour

The motion characteristics of a depth map are similar to those of its associated video picture. Thus, motion parameters and the sub-block partitioning of a block can be inherited from its corresponding block in the video picture [66]-[67]. For each block, either the inherited information or a new partitioning and the motion data are transmitted. In the Motion Parameter Inheritance (MPI) mode, the merge mode is modified such that merging the current block with its corresponding block in the video picture is the first candidate. The skip mode is modified as well. Motion vectors with quarter or half sample accuracy are rounded to full-sample precision.

Furthermore, in coding the depth maps, the DC values are more important for view synthesis than the high frequency components. Thus, in DC-Only Coding, the DC value of the

prediction residual is explicitly signaled. In 3D-HEVC, DC-Only Coding is extended by a Depth Lookup Table.

Figure 1.17 summarizes 3D-HEVC coding tools for texture and depth components.

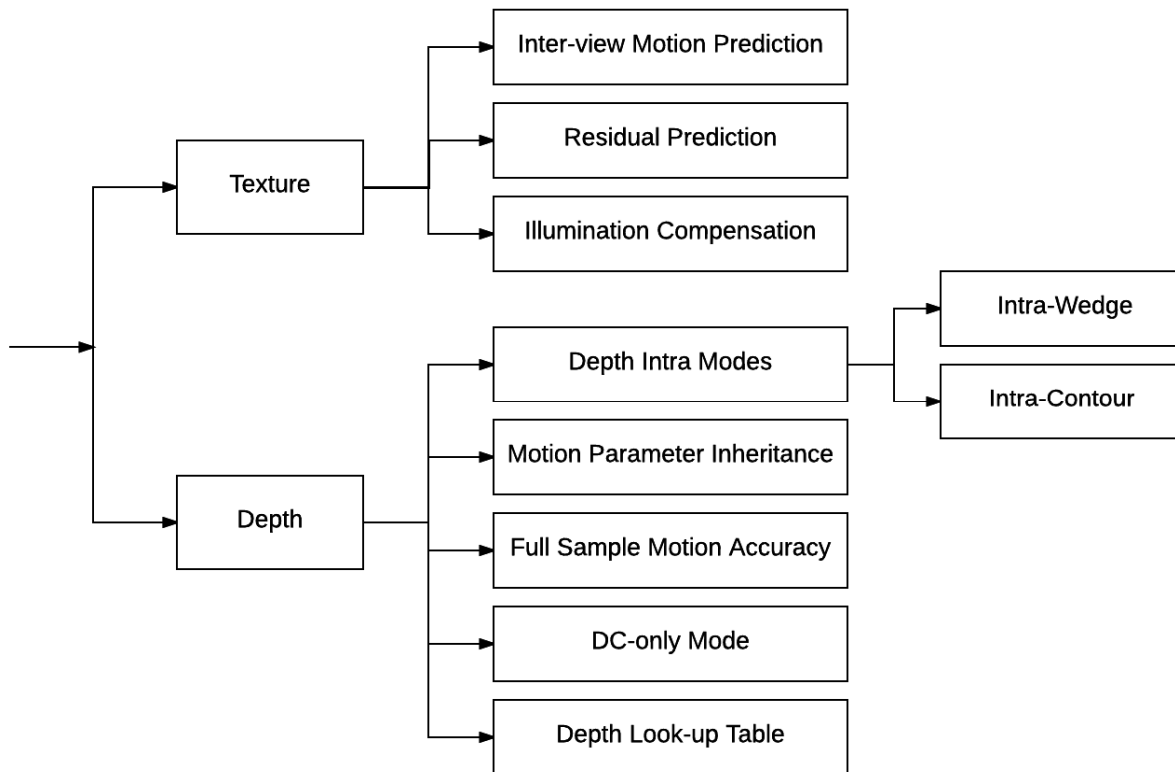


Figure 1.17 Overview of 3D-HEVC coding tools for texture and depth components

For coding the depth maps, the dis-similarities between the reconstructed and the original view should be taken into account rather than the dis-similarities between the original depth map and the coded depth map. Since the depth maps are not what is viewed, the distortion measure for the depth maps needs to be modified. The distortion in the synthesized view is used in the Rate-Distortion Optimization in the encoding process of the depth maps [68]. To select the

current coding mode in the depth map, two versions of the reference view are compared. The first version is synthesized based on the reconstructed depth values of the already coded blocks while the original depth values are used for the rest of the blocks. For the second one, the current mode is used to reconstruct the depth data of the current block. Then, the dis-similarities between the two are measured with SSD. The change in the distortion of the synthesized view is illustrated in Figure 1.18. In this figure, S'_{Ref} denotes a reference texture rendered from original video and depth data and SSD stands for Sum of Squared Differences. ΔD gives a measure of the distortion caused by the current mode. In 3D-HEVC, a fast rendering mechanism is used to re-render only those parts that are affected by the depth block [69].

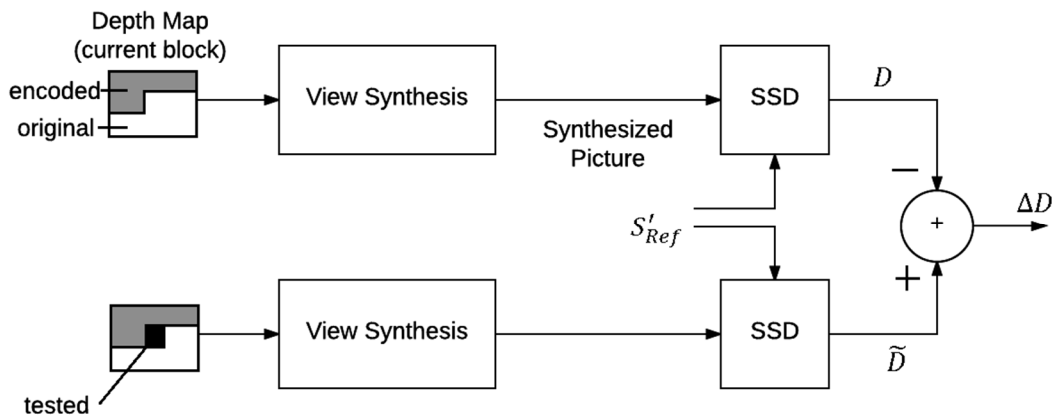


Figure 1.18 Synthesized view distortion change calculated for choosing depth map coding modes

In this section, we gave an overview of the additional coding tools introduced in 3D-HEVC. These tools provide 50% bitrate savings in comparison to HEVC simulcast and need to be taken into account in modifying and improving 3D-HEVC. Another approach for increasing compression efficiency of 3D video coding is asymmetric stereoscopic video coding.

1.4 Asymmetric Stereoscopic Video Coding

In video compression, spatial and temporal redundancies within a sequence of frames are removed as much as possible. In stereoscopic video, the similarity between the two views should also be taken into account to improve the compression efficiency. An auspicious technique for reducing the amount of data required for transmitting stereoscopic video is to reduce the quality of one of the views and keep the other view at the original quality. This is based on the suppression theory of the binocular vision [70]. Sharp edges in the high quality image masks the blur in the low quality view and the overall depth impression is close to the sharper view. Asymmetric coding for stereoscopic video is based on this characteristic of the HVS.

Perkins [71] introduced the concept of asymmetric coding, where a low-resolution picture is presented to one eye and a high-resolution picture to the other eye. For the low-resolution, the picture is subsampled by a factor of 4. Thus, the bitrate required for the asymmetric coding will only be 6% more than the bitrate that a single sequence of high resolution needs. Bilinear interpolation is proposed for reconstructing the low-resolution image, because it is easy to implement. In asymmetric stereoscopic video, MSE is not the best distortion measure. The viewer's perception of depth and quality of the 3D video is what matters and should be assessed subjectively. Subjective tests show that the resulting stereo pair is pleasing to watch. The eye or brain fuses the asymmetric stereo pairs and the final quality and sharpness perception is similar to the high-resolution picture.

In [72], the quality of asymmetric stereo video is examined. Instead of downsampling, low-pass filter is used to reduce the quality of the right view. It is a worthwhile test since the coding efficiency is improved by low-pass filtering and the asymmetric sequence can be

compressed more efficiently. The quarter and half resolution sequences of one view (while keeping the other view unfiltered) are compared to the original stereo video for two test sequences. Also, temporal averaging and drop-and-repeat frame modes are tested for stereo and non-stereo videos. Two groups of 21 subjects each have rated the quality and sharpness of the 10-second video sequences. They used double stimulus continuous quality (DSCQ) scale method described in ITU-R recommendation BT.500 [100]. These subjective tests show that for the spatial filtering, the quality and sharpness of the filtered stereo video is rated better than the monoscopic video and close to the original stereoscopic video. For temporal filtering, subjective tests show that there is a noticeable quality drop: 2-field averaged sequences get blurred and drop-and-repeat sequences have a jerky appearance. They concluded that spatial filtering is more promising than temporal filtering.

In [73], asymmetric quality with respect to asymmetric luminance and chrominance qualities are experimentally studied. They proposed an algorithm to reconstruct the chrominance quality of one pair at the decoder end due to the less visual acuity for chrominance relative to luminance. They ran subjective tests to examine whether the chrominance degradation in the right view affects the stereoscopic effect and whether a qualitative visibility threshold between right and left views exists. The results showed a significant amount of bitrate saving while the perception quality degradation was not noticeable for viewers.

Asymmetric stereo video is shown to be a promising approach for reducing the amount of bandwidth or memory required for transmission or storage of the stereoscopic video [74][75] but it is not a fair approach for people with a right or left dominant eye [76][77]. If the high quality sequence is shown to their weak eye, the overall impression of the 3D video is not close to the

high quality sequence. In addition, sustained imbalance in the two views has negative effects on children's premature visual system and causes fatigue in the viewers. One idea to reduce this imbalance is to interleave low and high quality views in time i.e. the quality of one view is reduced for a certain number of frames and the quality of the other view is reduced for the following time interval and this procedure continues. In [78], degraded picture is distributed to both of the views in a balance way over time where a GOP is taken as an interval to cross-switch the resolution of the two views. In [79], they suggest to interleave high-quality images with reduced-quality images within each stream. They ran subjective tests to see when the subjects notice the cross-switch. Their results show that the cross-switch is noticeable and annoying unless it occurs at scene cuts.

In [80], they examined the subjective overall of the traditional method of mixed resolution coding in which one of the two stereo views of a video is spatially down-sampled, as well as alternating the blurry view at every frame. They found that the perceived quality can be strongly affected by temporal frequency, whether due to motion or frame rate. While at 60 frame per second, the alternation of blur between the eyes is relatively imperceptible, at 30 frame per second, alternating method led to lower quality than the traditional asymmetric method for the low motion video. They also compared two methods of mixed resolution coding, single-eye and alternating-eye blur, in terms of overall quality for short exposures and visual fatigue level for long exposures [81]. Their subjective results showed higher quality for the single-eye blur. However, for long exposures, viewers experienced less fatigue from the alternating-eye blur for animated scenes.

1.5 Thesis Contributions

In this thesis, we investigate novel methods for improving the compression efficiency yielded by the 2D and 3D high efficiency video coding (HEVC) standard.

In order to achieve high video quality while keeping the bitrate value within constraints, we propose to take advantage of the characteristics of the human visual system. Perceptual video quality metrics are modeled based on some characteristics of the human visual system. In Chapter 2, we integrate a perceptual video quality metric in the rate distortion optimization process of HEVC. HEVC offers greater flexibility compared to the previous video coding standard, H.264. Instead of macroblocks with fixed size of 16×16 pixels in H.264, a coding tree unit structure with coding unit sizes varying between 64×64 to 8×8 pixels is used in HEVC. The structure of coding units is selected in the rate distortion optimization process. Traditionally, distortion is measured by the sum of squared differences (SSD) in the rate distortion optimization process of HEVC which correlated poorly with the human visual system. Our novel approach suggests the use of a perceptual video quality metric in the selection of the coding unit structure. We adapt the perceptual video quality metric so that it is applicable to blocks of data. This adaption allows the integration of the perceptual video quality metric in the rate distortion optimization process of HEVC. Also, we have selected to use the perceptual video quality metric in specific stages of the rate distortion optimization process in HEVC. We do not integrate the perceptual quality metric where the rate distortion optimization process selects the best motion vector in the inter mode or the candidate list for intra modes. Due to the recursive nature of rate distortion optimization process, computational complexity of the proposed approach should be taken into account. We have selected to use the perceptual video quality metric in HEVC stages

where the resulting complexity does not increase significantly. Experiments show that our method improves HEVC compression efficiency by 10.21%.

Our developed approach in Chapter 2 can benefit from adjusting the Lagrangian multiplier (which represents the trade-off between the perceptual distortion and the bitrate) based on the characteristics of the video content. In Chapter 3, we first report our observations that show this adjustment based on the video content can further improve our proposed approach. Next, we develop a model for the bitrate and the perceptual distortion of the first frame of the video. From this model, the best Lagrangian multiplier can be obtained. Finally, we propose to estimate the value of the best Lagrange multiplier from the first frame. This value is used for the frames in that scene. Experimental results show that the proposed approach further improves the compression efficiency of HEVC (up to 2.62% with an average of 0.60%).

In Chapter 4, we extend our work in chapter 2 to 3D extension of HEVC mainly to stereoscopic and multi-view videos. First, we integrate the perceptual video quality in the rate distortion optimization process of stereo video coding. Stereoscopic video coding compresses the base view using the unmodified HEVC encoder and use additional tools for coding the dependent view. We investigate the effects of asymmetric coding on the base view and the dependent view in the stereoscopic video coding. Additionally, we extend our approach to multi-view video coding for auto-stereoscopic displays. Auto-stereoscopic displays require two or three views plus their corresponding depth maps to generate intermediate views. In 3D-HEVC, new tools are introduced in coding depth maps. Also, rate distortion optimization is replaced by view synthesis optimization. Our perceptual 3D video coding increases the compression efficiency of stereo and multi-view video coding.

Chapter 5 studies the compression performance of asymmetric coding of stereoscopic videos. This approach lowers the quality of all parts of one of the views but does not alter the other view and keeps it at its original quality. In our implementation, we modify the asymmetric coding so that lower quality parts are distributed in each frame of both views. This avoids the fatigue experienced by viewers due to sustained imbalance in one of the views. In addition, we applied low-pass filtering to slices on each of the two views while the corresponding slice in the other view is kept unaltered and of high quality. For the same visual quality, performance evaluations have shown that the asymmetric approach yields better compression performance compared to the original symmetric 3D video.

Chapter 2: Perceptual Coding of HEVC

In previous work on perceptual video coding, properties of the human visual system have been integrated into some aspects of H.264 such as quantization, motion estimation or coding of intra frames. The study in [82] introduces the properties of HVS into the quantization process of H.264. In this study, the human eye sensitivity to contrast and spatial/temporal masking effects are taken into account for macroblock quantization adjustment. As the sensitivity of HVS differs for different frequencies, a frequency-weighting scheme at the macroblock level has been used in the quantization process of H.264 [83]. An improved motion estimation method for H.264 inter coding is proposed in [84]-[86] based on the structural similarity index (SSIM). Moreover, SSIM has been integrated into H.264 for coding intra frames [87]-[89]. Based on the DCT domain SSIM, the study in [90] normalized the transform coefficients in H.264 to improve its quantization process. In another study, reduced reference SSIM estimation is used to select best coding in the rate distortion optimization of H.264 [91]. The study in [92] proposed a local scaling of the Lagrangian multiplier based on SSIM estimation. The study in [93] have integrated PSNR-HVS in three different settings including mode selection with or without motion estimation as well as in the I-frames. These studies focus on integration of a perceptual video quality metric in the rate distortion optimization of H.264/AVC or multiview video coding (MVC). In more recent studies, SSIM has been used as a perceptual video quality metric in quantization and mode selection of the rate distortion optimization in HEVC [94]-[95].

Our proposed method suggests to integrate a perceptual quality metric inside the rate distortion optimization process of HEVC. For our modified rate distortion optimization process, we drive the proper Lagrangian multiplier. In our proposed approach, first we modify the

Lagrange multiplier by a scalar coefficient. In the second step, we optimize the modified Lagrange multiplier based on the quantization parameter.

Our proposed approach is presented in detail in section 2.1. The experimental results are presented in section 2.2. Finally, conclusions are made in section 2.3.

2.1 Proposed Method

2.1.1 Perceptual Mode Decision

Experiments with HEVC show that more than half of its average bit-rate savings over H.264/AVC can be attributed to its increased flexibility of block partitioning for prediction and transform [12]. As discussed in section 1.1.1, introduction of the tree structure in HEVC for the coding units (CU), for which an encoder has to decide between intra and inter prediction modes, provides more flexibility in HEVC compared to H.264/AVC. This quad-tree structure supports various block sizes for coding units. Larger block sizes are advantageous for the high and ultra-high resolution videos, while smaller block sizes are critical to adapt to the local characteristics of the picture. Hierarchical block partitioning in HEVC addresses both objectives efficiently while in H.264/AVC, macroblocks with fixed size were used. Due to its novelty, this stage needs to be further investigated.

We improve the coding unit structure and mode selection in HEVC by integrating the PSNR-HVS perceptual video quality metric in the rate distortion optimization process. We modify the distortion measurement based on a perceptual video quality metric to incorporate the characteristics of the human visual system in the coding unit selection process. PSNR-HVS is based on the characteristics of the human visual system [18] and shows higher correlation with subjective video quality evaluations compared to PSNR. PSNR-HVS has the advantage of being

easily adoptable to blocks of data and is not too computationally complex to be integrated inside the video encoder.

In HEVC, the rate distortion optimization process is split into a series of smaller minimization problems as discussed in section 1.1.3. In coding unit (CU) quad-tree structure and mode selection, distortion is measured by the sum of squared errors (SSE) between the original and reconstructed CU blocks. In the inter prediction process of the rate distortion optimization, distortion is measured between the original block and its motion-compensated block. Similarly, in the intra prediction process of the rate distortion optimization, distortion is measured between the original block and its prediction block. By changing the distortion measurement in coding unit structure and mode selection process, we control how close the reconstructed CU block is to the original block. The difference between the original block and the prediction block (i.e. the residuals) are transformed, quantized, entropy coded and transmitted to the decoder. At the decoder, the prediction block is not displayed to the viewer but rather the reconstructed block. By integration of a perceptual video quality metric in the coding unit structure and mode selection process, the reconstructed blocks are selected perceptually and provide better visual quality to the viewers.

For coding unit structure and mode selection, sum of squared errors (SSE) is used for measuring distortion in HEVC. Measuring the distortion by SSE and minimizing it in the rate distortion optimization process, leads to better PSNR in the coded video. However, the PSNR has been shown to have limited correlation with subjective tests. Various other video quality metrics have been developed to better represent how subjects perceive video quality. The study in [20] demonstrates the scope of validity of PSNR as a video quality metric. Based on this

study, PSNR is reliable only when the video content and the codec are fixed in the test conditions [21]. For a given video codec and a specified video content, the performance of different optimization settings can be measured by PSNR. In case of modifying the video encoder, PSNR will not be able to provide reliable objective assessment of the video quality.

PSNR-HVS [30] is a full reference video quality metric, which takes into account the characteristics of the human visual system. One of the characteristics of the HVS is that its sensitivity decreases at high spatial frequencies. PSNR-HVS is defined as:

$$PSNR - HVS = 10 \log \left(\frac{255^2}{MSE_{HVS}} \right) \quad (2.1)$$

$$MSE_{HVS} = K \sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^8 \sum_{n=1}^8 ((X[m, n]_{ij} - X[m, n]_{ij}^e) T_c[m, n])^2 \quad (2.2)$$

where $K=1/[(I-7)(J-7)64]$, I and J are the image width and height, X_{ij} is the DCT coefficient of an 8×8 image block with its upper left corner at (i,j), X_{ij}^e is the DCT coefficient of the corresponding block in the original image and T_c is a matrix adopted from the JPEG quantization table proposed in the JPEG [96].

PSNR-HVS considers a window size of 8×8 . A matrix of correcting factors adopted from the JPEG quantization table gives more weights to lower frequency coefficients. It has been shown that PSNR-HVS has higher correlation with the subjective results than PSNR [30]. PSNR-HVS has desirable properties that allow its application to coding units in the video compression based on HEVC.

PSNR-HVS is a Distance Measure (DM). It is important to note that not all metrics satisfy the requirements of distance measures. PSNR is a distance measure in the pixel domain and it satisfies the following properties:

$$D[A, A'] = f(|a_{11} - a'_{11}|, |a_{12} - a'_{12}|, \dots, |a_{nn} - a'_{nn}|) \quad (2.3)$$

$$\forall i, j, x_{11}, \dots, x_{nn}: |a_{ij}| < |a'_{ij}| \Rightarrow f(|x_{11}|, \dots, |a_{ij}|, \dots, |x_{nn}|) \leq f(|x_{11}|, \dots, |a'_{ij}|, \dots, |x_{nn}|) \quad (2.4)$$

where A and A' are two matrices, a_{ij} and a'_{ij} are the elements at (i, j) index, and D measures the distance between matrices A and A' . The distance measure, D , is defined by the function f . x_{11}, \dots, x_{nn} are the elements of the arbitrary matrix X which is of the same size as A and A' . Triangular inequality does not imply that the second condition is met [97].

PSNR-HVS satisfies these two conditions in the frequency domain. However, not all metrics belong to this class. Only video quality metrics that satisfy requirements of being a Distance Measure can be used to measure distortion in the rate distortion optimization process.

2.1.2 Scaled Lagrange Multiplier

In the rate distortion optimization process, the Lagrange multiplier λ acts as a knob that controls the trade-off between rate decreases versus distortion increases. Distortion measurements with a video quality metric changes this trade-off based on the range of output values. The new optimal value of λ needs to be determined to have the best quality of video while minimizing the required bitrate. We consider a scaling factor to denote the relationship between the proposed λ and the λ used in HEVC. In the next subsection, we optimize λ based on the quantization parameter.

Various scaling factors are tested to find the best trade-off between the rate and distortion. In our tests, the Lagrangian multiplier is modified by a scaling factor, c , as:

$$\lambda_{proposed} = c \times \lambda_{mode} \quad (2.5)$$

where $\lambda_{proposed}$ is the proposed Lagrangian multiplier in our approach, λ_{mode} is the Lagrangian multiplier in HEVC as defined in equation (1.8) and c is the scaling factor. In the rate distortion optimization process, λ acts as a knob that controls the trade-off between the video bitrate and the distortion. High values of the scaling factor forces low bitrates for the compressed video at the expense of having a low quality video. On the other hand, low values of the scaling factor results in low distortion or high quality video provided that the high bitrate can be handled in the transmission or storage media. Figure 2.1 shows the Rate Distortion performance for the video sequence BQSquare for three different values of λ . Figure 2.1 shows that our proposed method achieves higher quality at the same bitrate compared to HEVC reference software over the Quantization Parameter (QP) values. Figure 2.1 shows that as λ increases, the compressed video needs less bitrate. Comparing the three plots in Figure 2.1 shows that there is a scaling factor which balances the bitrate and distortion leading to the most bitrate savings over the range of QPs. Quantization Parameters (QP) of values 22, 27, 32 and 37 have been tested. These QP values have been selected in the MPEG tests during standardization process of HEVC.

The average bitrate difference between the proposed and reference rate-distortion curves is referred to Bjontegaard's Delta (BD) Rate [99]. A cubic polynomial approximation is derived using four data points (quality and bitrate points). The difference between the two curves is

integrated in the horizontal direction for the BD Rate. The BD Rate measures the average bitrate savings by the proposed approach compared to the HEVC reference software.

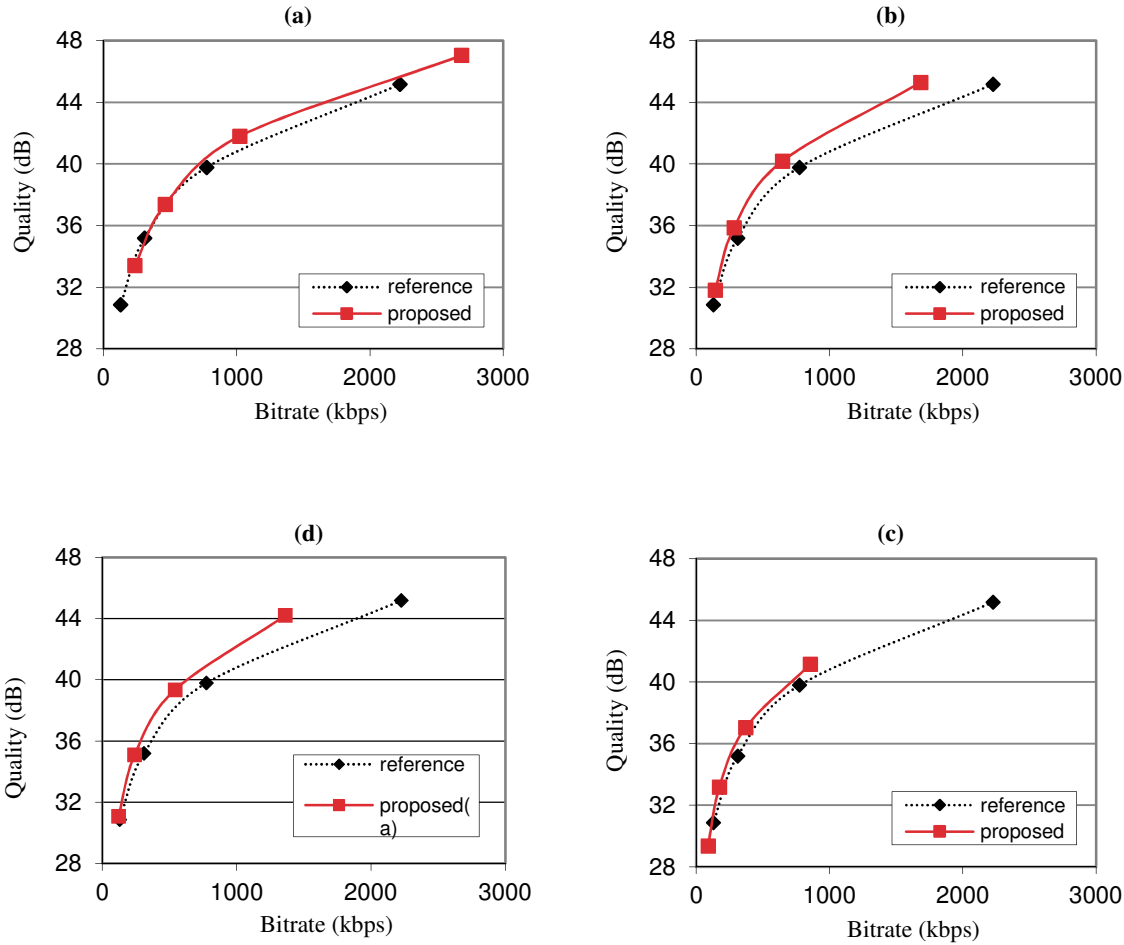


Figure 2.1 Rate-Distortion Curves for *BQSquare* video sequence for three different scaling factors. Quality is measured with Mean PSNR-HVS. (a) Scaling factor=1.6. (b) Scaling factor=5.6. (c) Scaling factor=8 and (d) Scaling factor=16.

The bitrate savings depends on the scaling factor. There is a trade-off between minimizing distortion and the required bitrate. This trade-off is controlled by the scaling factor applied to λ . As the scaling factor increases, more emphasis is given to the bitrate in the minimization process, while lower scaling factors emphasizes more on lower distortion values.

The bitrate savings achieved by the proposed approach for the BQSquare video sequence based on different scaling factors are summarized in Table 2.1.

Scaling factor	1.6	4.0	5.6	8.0	12	16
BD Rate	-5.78%	-18.13%	-20.69%	-21.26%	-19.31%	-15.86%

Table 2.1 Bitrate saving of the proposed approach compared to the reference HEVC software for different scaling factors for the *BQSquare* video sequence

To find the optimal scaling factor, the bitrate savings versus scaling factors are plotted in Figure 2.2 for different video sequences. For the video sequence RaceHorses, scaling factor 5.6 yields the largest bitrate savings. However, scaling factor 8 shows the highest amount of bitrate savings for the BQSquare, BQTerrace and PartyScene video sequences. Thus, scaling factor 8 is selected for our proposed approach. Based on Figure 2.2, for BQSquare, the amount of bitrate saving changes very minimally as the scaling factor sweeps the range of 4 to 12. The sensitivity of the bitrate savings to the scaling factor can be analyzed based on Figure 2.2. The slope of the line tangent to the curves increases monotonically by moving towards lower or higher scaling factors. Figure 2.2 shows that regardless of the content of the video sequence and its complexity, the minimum of BD rate savings happen around the same value for the Lagrangian multiplier.

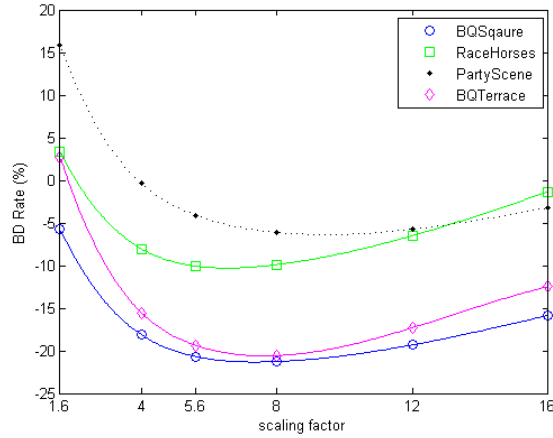


Figure 2.2 Bitrate savings versus the scaling factor for different video sequences

2.1.3 Optimal Lagrange Multiplier

An optimal λ should lead to the best quality of video while minimizing the required bitrate. In the previous section, a constant scaling factor is used to denote the relationship between the new λ and the λ used in HEVC. However, the Lagrange multiplier can be optimized based on the quantization parameter (QP). In this paper, we find the optimal Lagrange multiplier in the rate distortion optimization process based on:

$$\lambda_{optimal}(QP) = c(QP) \times \lambda_{mode}(QP) \quad (2.6)$$

The quantization parameter is monotonically dependent on the quantization step size. An increase of one in the quantization parameter means an increase of the quantization step by approximately 12%. A larger quantization step size leads to a higher amount of distortion, but requires lower bitrate. Thus, the choice of step size is closely related to the choice of the relative emphasis on rate and distortion. That is the choice of the Lagrangian multiplier, λ .

Originally, the relationship between λ and step size was derived by looking at the relative occurrence of λ based on Quant (step size = 2Quant). For sequences with widely varying content, λ versus the obtained average Quant was observed to follow a similar pattern:

$$\lambda_{mode} = 0.85. (Quant)^2 \quad (2.7)$$

This observation formed the basis for the relationship between λ and the quantization parameter [15]. It was further justified by considering high rate approximation for a typical quantization curve:

$$R(D) = a. \ln\left(\frac{\sigma^2}{D}\right) \quad (2.8)$$

where a is a constant that depends on the source distribution function with variance σ^2 . To obtain λ , we set the derivative of $D + \lambda \times R$ with respect to D equal to zero:

$$\frac{dR(D)}{dD} = -\frac{a}{D} = -\frac{1}{\lambda_{mode}} \quad (2.9)$$

At sufficiently high rates, a reasonably well-behaved source probability distribution can be approximated as a constant within each quantization interval:

$$D = \frac{(2. Quant)^2}{12} \quad (2.10)$$

Thus, λ is obtained as:

$$\lambda_{mode} = \frac{D}{a} = \frac{4}{12.a} (Quant)^2 = c. (Quant)^2 \quad (2.11)$$

This derivation reveals qualitatively that λ is proportional to the square of the quantization step size [15]. Considering the logarithmic relationship between the Quantization parameter and step size, we derive our proposed Lagrangian multiplier based on the logarithmic relationship between λ and QP. By considering eq. 2.8, we model $c(QP)$ as:

$$\log_2(c(QP)) = \alpha \cdot QP + \beta \quad (2.12)$$

where α and β are constants to be determined. In order to find the optimal Lagrange multiplier based on the quantization parameter, contours of equal quality are plotted in the plain of the two axes of the quantization parameter (QP) and coefficient c in Figure 2.3. On each contour, the point that leads to the least amount of bitrate is determined. We derive the relationship between the optimal Lagrange multiplier and the quantization parameter by connecting these points as shown in and fitting a curve to them.

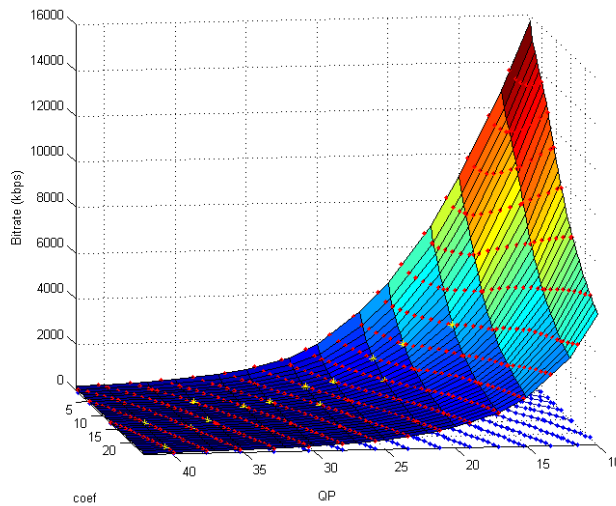


Figure 2.3 Equal quality contours (blue lines) for the BQSquare video sequence. On each projected contour, the point associated with the least amount of bitrate (yellow dot) is selected. These points show the optimal operating points and lead to minimum amount of bitrate at the same level of distortion

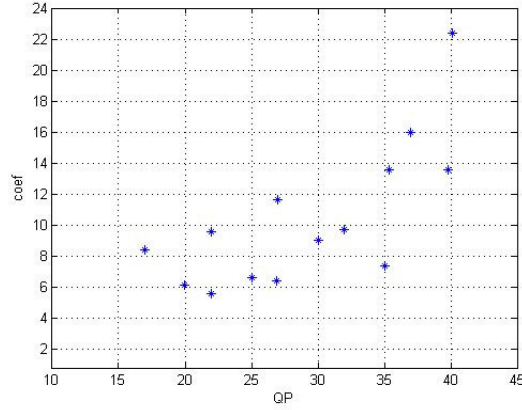


Figure 2.4 Optimal coefficients based on the quantization parameter. Based on this plot the (linear) relation between $\log(\text{coef})$ and QP is obtained.

By curve fitting, we can find the relationship between the optimal coefficient and the quantization parameter. In eq. 2.12, α and β will be obtained as:

$$\alpha = \frac{0.22}{3}, \beta = \frac{3.49}{3} \quad (2.13)$$

By replacing $c(QP)$ from eq. 2.12 in eq. 2.6, $\lambda_{optimal}$ can be found as:

$$\lambda_{optimal} = \alpha \times w_k \times 2^{((1.22QP-8.51)/3.0)} \quad (2.14)$$

2.2 Experimental Results

2.2.1 Test Video Sequences

To validate the efficiency of the proposed approach, PSNR-HVS is integrated into the HEVC reference software HM9.2. For the test video sequences, we used the standard test video sequences of MPEG as summarized in Table 2.2. All test video sequences are in the YCbCr 4:2:0 format. The length of each sequence is 10 seconds.

Class B has the picture size of 1920×1080 pixels and is used for evaluation of 1080p HDTV. Classes C and D have videos of 832×480 pixels and 416×240 pixels, respectively. Test video sequences in these two classes are for measuring performance in mobile applications.

	Video sequence	Total Frames	Frame Rate	Resolution
Class D WQVGA	BQSquare	600	60	416x240
	BlowingBubbles	500	50	
	RaceHorses	300	30	
Class C WVGA	PartyScene	500	50	832x480
	BQMall	600	60	
	BasketballDrill	500	50	
Class B 1080p	Cactus	500	50	1920x1080
	BasketballDrive	500	50	
	BQTerrace	600	60	
Class A 4K	PeopleOnStreet	150	30	2560 x1920
	Traffic	150	30	

Table 2.2 Test video sequences

First frame of the test sequences are shown in Figure 2.5. All tests are run on a cluster containing quad-core processors running at 3GHz. The 8 cores in a single node in the cluster share 16GB of RAM.



BQSquare



BlowingBubbles



RaceHorses



PartyScene



BQMall



BasketballDrill



Cactus



BasketballDrive



BQTerrace



PeopoleOnStreet



Traffic

Figure 2.5 First frame of test video sequences

Our proposed approach has been tested on video sequences with various amounts of detail, texture and motion. Table 2.3 reports the amount of perceptual Spatial Information (SI) and Temporal Information (TI) in each of the test sequences. Spatial details of a picture is measured by SI. Higher values of SI are associated with higher spatial complexity in scenes. SI is based on the Sobel filter. Then, the standard deviation over the pixels in each Sobel-filtered

frame is computed. The maximum value of this time series is selected as SI. TI measures the amount of temporal changes of a video sequence. High motion sequences generally have higher TI values. TI is based on the motion difference between frames. The perceptual spatial information and temporal information are critical parameters in determining the amount of video compression that is possible. Thus, the set of test sequences should span should the range of spatial and temporal information.

	Video Sequence	Spatial Information	Temporal Information
Class D WQVGA	BQSquare	160.76	18.04
	BlowingBubbles	100.72	26.06
	RaceHorses	103.52	34.95
Class C WVGA	PartyScene	106.57	17.95
	BQMall	109.35	34.20
	BasketballDrill	76.86	21.29
Class B 1080p	Cactus	67.27	17.25
	BasketballDrive	79.78	23.74
	BQTerrace	107.45	27.04
Class A 4K	PeopleOnStreet	82.73	23.19
	Traffic	62.20	13.29

Table 2.3 Spatial and Temporal Information Content of the Test Video Sequences

2.2.2 Scaled Lagrange Multiplier

Table 2.4 shows the bitrate and the corresponding video quality for the test video sequence BQSquare, using HEVC and the proposed approach. Quantization parameters of 22, 27, 32 and 37 are used. These results show that the proposed approach can deliver the same video quality at a lower bitrate. The average bitrate saving is 21% for the BQSquare video sequence. The results show that the proposed approach improves the encoding efficiency.

QP	Reference HEVC		Proposed		BD-rate (piecewise cubic)	BD-rate (cubic)
	Bitrate (kbps)	Quality (dB)	Bitrate (kbps)	Quality (dB)		
22	2225.06	45.18	1362.82	44.20		
27	772.30	39.79	541.38	39.34	-21.26%	-21.19%
32	308.49	35.19	240.54	35.10		
37	126.62	30.87	120.57	31.09		

Table 2.4 Bitrate and quality of the proposed method along with the HEVC reference software for the test video sequence BQSquare. The quality is measured by PSNR-HVS

Table 2.5 shows the bitrate and the corresponding video quality for the proposed approach and the reference HEVC software for two sequences of each class of the test video sequences. The test video sequences in Table 2.5 have different resolutions (416×240 , 832×480 or 1920×1080). The amount of motion details and texture of the video sequences differ between the test video sequences. Bjontegaard’s Delta (BD) Rate [99] measures the average bitrate savings by the proposed approach compared to the reference HEVC software. The quality is measured by PSNR-HVS.

Video Test Sequences	QP	Reference video		Proposed approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Bitrate (kbps)	Quality (dB)	Bitrate (kbps)	Quality (dB)		
Blowing Bubbles	22	1942.13	45.28	1630.37	44.63	-9.89%	-9.82%
	27	825.84	38.74	743.31	38.84		
	32	353.12	32.80	347.45	33.43		
	37	148.83	27.73	170.63	28.65		
Race Horses	22	1338.03	43.46	1379.89	44.38	-5.24%	-5.18%
	27	635.79	37.07	672.03	38.09		
	32	298.27	31.43	335.54	32.53		
	37	142.95	27.08	175.08	28.19		
Party Scene	22	8054.40	45.83	6492.80	44.88	-9.31%	-9.23%
	27	3447.49	39.25	3038.62	39.12		
	32	1504.46	33.37	1431.29	33.74		
	37	643.55	28.31	705.21	29.01		
BQ Mall	22	4200.41	44.45	4083.42	45.06	-6.11%	-6.06%
	27	1870.84	39.22	1949.41	40.15		
	32	903.03	34.17	1018.44	35.29		
	37	457.74	29.68	571.39	30.84		
Basketball Drive	22	19835.95	42.40	18066.34	42.85	-6.63%	-6.45%
	27	6750.85	38.40	7349.24	39.21		
	32	3111.42	34.46	3636.96	35.41		
	37	1584.96	30.98	1946.03	31.87		
BQ Terrace	22	52793.42	42.27	28543.60	42.10	-20.57%	-20.33%
	27	7558.00	38.66	6094.38	39.08		
	32	1989.84	35.05	2211.78	35.75		
	37	760.90	31.11	1008.34	32.05		

Table 2.5 Bitrate and quality of the proposed method along with the HEVC reference software for various test video sequences

Figure 2.6 shows the quality versus bitrate for *BQSquare*, *PartyScene*, *BasketballDrive* and *BQTerrace* sequences. We see improvement in quality across the tested range of rates.

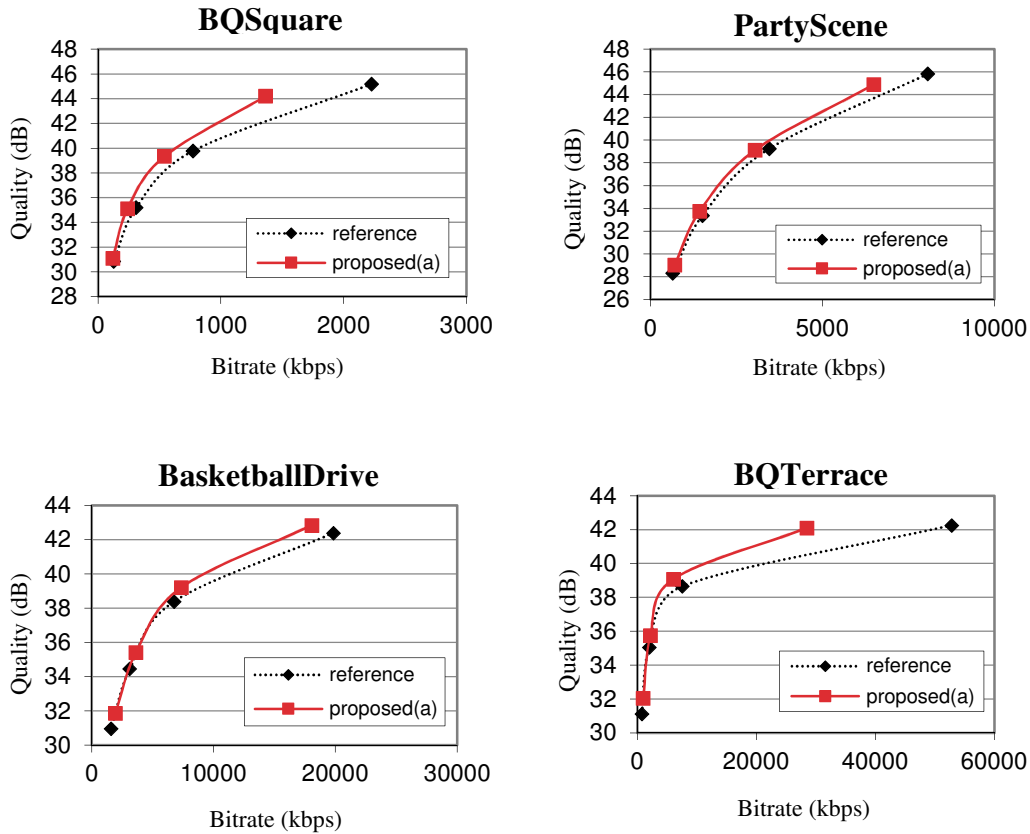


Figure 2.6 Quality versus bitrate plots comparing the proposed approach with scaled Lagrange multiplier with HEVC reference

Table 2.6 summarizes the total bitrate savings for the test video sequences. Test video sequences are grouped into three classes: class B with HD resolution, and classes C and D, resolutions of Wide Video Graphics Array (WVGA) and Wide Quarter Video Graphics Array (WQVGA), respectively. The average bitrate savings over all the video sequences is 9.79%.

	Video sequence	$\Delta rate$
Class D WQVGA 416x240	BQSquare	-21.26%
	BlowingBubbles	-9.89%
	RaceHorses	-5.24%
Class C WVGA 832x480	PartyScene	-9.31%
	BQMall	-6.11%
	BasketballDrill	-3.05%
Class B 1080p 1920x1080	Cactus	-6.04%
	BasketballDrive	-6.63%
	BQTerrace	-20.57%
	Average	-9.79%

Table 2.6 Rate reduction of the proposed method with scaled Lagrangian multiplier compared to HEVC reference software (over QPs of 22, 27, 32 and 37)

2.2.3 Optimal Lagrangian Multiplier

In this section, we take advantage of the optimal lambda found in section 2.1.3. Table 2.7 summarizes total bitrate savings of the proposed approach with the optimal λ for the test video sequence. The average bitrate savings over all the video sequences is 10.21%. Optimal λ increases the amount of bitrate savings compared to the scaled λ . By comparing the results in Table 2.7 and Table 2.6, we see that, for almost all video sequences, the bitrate savings have increased with the optimal λ . However, the result of the scaled λ closely follows that of the optimal λ . Figure 2.7 shows rate distortion curves for our proposed approach with optimal Lagrange multiplier compared to the HEVC reference for different video sequences.

	Video sequence	$\Delta rate$
Class D WQVGA 416x240	BQSquare	-21.61%
	BlowingBubbles	-10.13%
	RaceHorses	-5.72%
Class C WVGA 832x480	PartyScene	-8.59%
	BQMall	-6.55%
	BasketballDrill	-3.82%
Class B 1080p 1920x1080	Cactus	-7.03%
	BasketballDrive	-7.01%
	BQTerrace	-21.44%
	Average	-10.21%

Table 2.7 Rate reduction of the proposed method with optimal Lagrangian multiplier compared to HEVC reference software (over qps of 22, 27, 32 and 37)

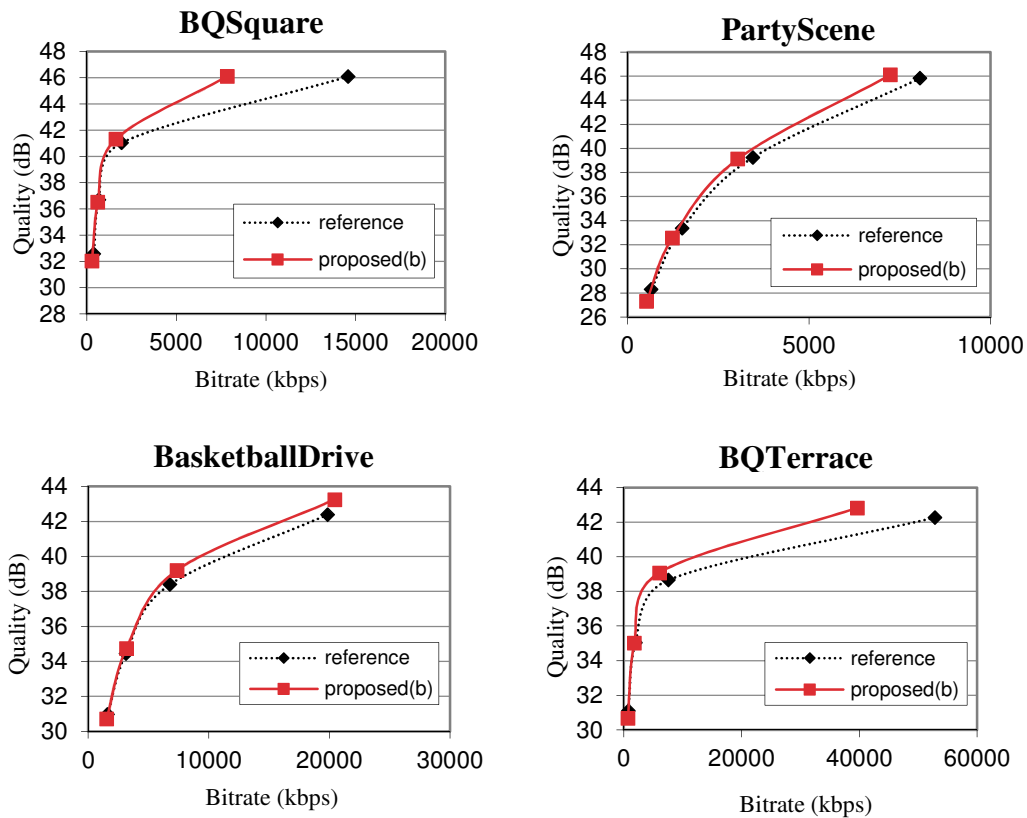


Figure 2.7 Rate-distortion curves for our proposed approach with optimal Lagrange multiplier compared to the HEVC reference

2.2.4 Higher Resolution Tests

The performance of the proposed algorithm is evaluated for higher resolution than 1080p. The resolution of the test video sequences of *PeopleOnStreet* and *Traffic* is 1920×2560 . These video sequences are used to measure the coding efficiency on 4K/8K video. Their picture sizes have been cropped by MPEG to reduce computation time. Total of 150 frames are coded.

Table 2.8 shows the simulation results. We observe that coding improvements are achieved by the proposed approaches compared to the HEVC reference software. The test video sequences have a lot of details. In the proposed approach with scaled λ , bitrate savings of 2.60% and 2.67% are reported for the *PeopleOnStreet* and *Traffic* video sequences, respectively. With the optimal λ in the proposed approach, the bitrate savings are further improved to 2.98% and 4.56% for the test video sequences, as summarized in Table 2.8. Figure 2.8 shows rate distortion curves for our proposed approach with scaled lambda as well as with optimal lambda compared to the HEVC reference software for a 4k video sequences.

4K Video Test Sequences	Proposed approach with scaled Lambda	Proposed approach with optimal Lambda
	BD-rate (piecewise cubic)	BD-rate (piecewise cubic)
People on Street	-2.60%	-2.98%
Traffic	-2.67%	-4.56%

Table 2.8 Bitrate saving of the proposed approach compared to the reference HEVC for the Higher Resolution test video sequences

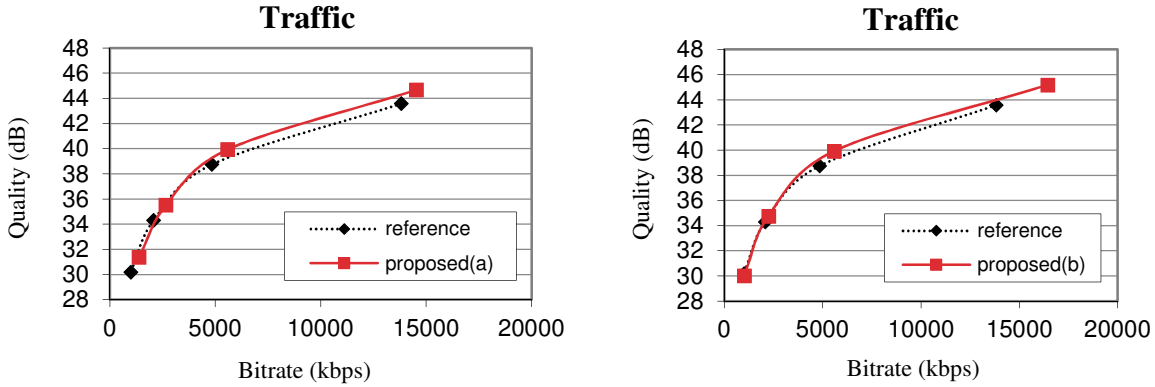


Figure 2.8 Rate-distortion curves for 4k video sequence, *Traffic*. Our proposed approach with scaled lambda is shown on the left and our method with optimal lambda is shown on the right side.

2.2.5 Prediction Structures

HEVC supports three types of prediction structures: all intra, random access and low delay. These structures are defined in the common test conditions [14] and are used for performance evaluations. Based on the temporal dependencies of the video frames, these prediction structures are categorized as All Intra (AI), Random Access (RA), Low Delay P picture (LDP) and Low Delay B pictures (LDB).

In the AI configuration, there is no inter-picture prediction and each picture is encoded as an I-frame. The average bitrate saving of the proposed approach for this configuration is 2.64% over all test video sequences.

A hierarchical B structure is used in the RA configuration. The coding efficiency achieved by the proposed approach for the RA configuration is 8.03% on average. In this configuration, the bitrate savings are higher than the AI configuration since inter-prediction is allowed. The QP in the common test conditions is set for the first picture. The QPs of the other

pictures are derived by adding an offset value (depending on the picture type) to the pre-set QP value.

Video sequence		<i>Δrate</i>			
		Low Delay B	Intra	Random Access	Low Delay P
Class D WQVGA 416x240	BQSquare	-21.61%	-1.97%	-15.12%	-24.40%
	BlowingBubbles	-10.13%	-4.57%	-9.83%	-11.26%
	RaceHorses	-5.72%	-2.94%	-5.29%	-5.80%
Class C WVGA 832x480	PartyScene	-8.59%	-3.75%	-11.03%	11.49%-
	BQMall	-6.55%	-2.97%	-6.37%	-7.49%
	BasketballDrill	-3.82%	-3.02%	-4.07%	-4.99%
Class B 1080p 1920x1080	Cactus	-7.03%	-3.19%	-6.65%	-7.99%
	BasketballDrive	-7.01%	-4.31%	-6.99%	-7.89%
	BQTerrace	-21.44%	-4.79%	-17.75%	-26.25%
Class A 2560 x1920	People on Street	-2.98%	1.76%	-2.24%	-3.56%
	Traffic	-4.56%	0.69%	-2.97%	-5.32%
Average		-9.04%	-2.64%	-8.03%	-10.50%

Table 2.9 Bitrate saving of the proposed approach compared to the reference HEVC for the test video sequences with AI, RA, LDP and LDB prediction structure

In the LDP and LDB configurations, the first picture is coded as an I frame and the subsequent pictures are coded as P and B frames. Low coding delay stems from the fact that re-ordering of pictures is not allowed in these configurations. LDP and LDB improve the average coding efficiency by 10.50% and 9.04%, respectively. Low delay configuration achieves the highest bitrate saving for the proposed approach compared to other configurations. Table 2.9 lists the bitrate savings for the test video sequences in each class, compressed with AI, RA, LDP or LDB configuration.

2.2.6 Visual Comparison

Figure 2.9 shows a zoomed portion of the BQSquare coded sequence in the low-delay encoding configuration. The picture on the left is the original picture. The picture in the middle is compressed with HEVC and the picture on the right is compressed with our proposed method with optimal λ . Both pictures are compressed with QP of 37. The bitrate required for the HEVC video is 126.62 kbps while the proposed approach needs 120.57 kbps. We observe that our method preserves more details compared to the original HEVC.

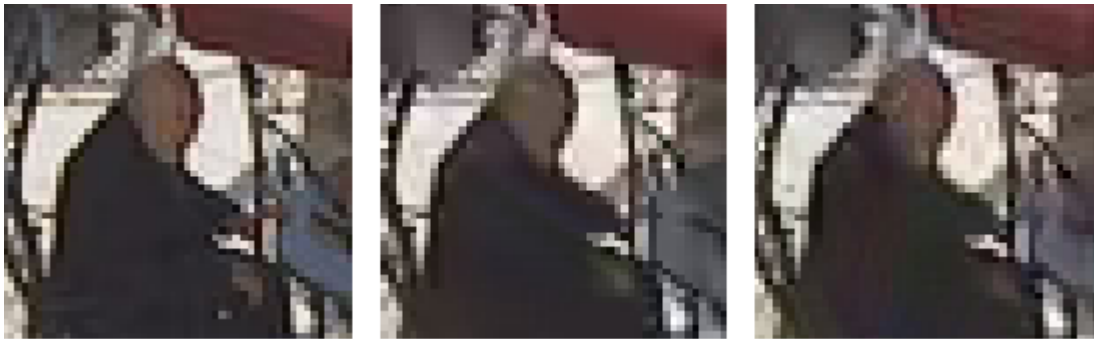


Figure 2.9 Visual comparison of the proposed method with the reference HEVC software for part of the BQSquare video with QP=37. The picture on the left is the original picture, the picture in the middle is compressed with HEVC and the picture on the right is compressed with our proposed method.

2.2.7 Subjective Tests

We run subjective tests to examine the performance of our proposed approach. Standard viewing conditions outlines in ITU-R BT.500 [100] were used in our tests. Our display is a 55-inch curved display. Distance of the viewers from display was 5 times the height of the display. 15 subjects took part in our tests with average age of 29. Minimum age of our subjects was 24 and their maximum age was 33. All subjects in our test have been assessed and passed the visual acuity test (using Snellen charts), color blindness test (using Ishihara chart) and stereovision

acuity test (via Randot test–graded circle test 100s of arc). For each sequence, four quantization parameters were used. Quantization parameters were selected as is suggested by MPEG standard tests. Each video sequence was coded both with the proposed approach and with the reference HEVC resulting in eight different videos. We used the double stimulus impairment scale (DSIS) method. In DSIS, the subjects are presented with a series of pictures with various levels of impairment in random order. An un-impaired picture is included to serve as the reference for assessments. The grading scale has five labels: excellent, good, fair, poor and bad. The structure of test material is shown in Figure 2.10.

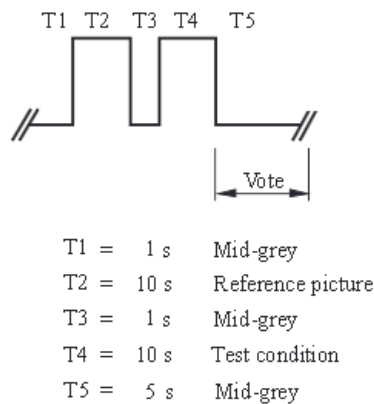


Figure 2.10 Subjective tests structure

Prior to the test, there was a training session for subjects to familiarize them with the test procedure. Subjective assessments are converted to scores in the range of 0 to 5. The mean opinion score (MOS) over all subjects is plotted for the proposed approach and is compared with HEVC. The rate distortion curves obtained from the subjective tests are shown in Figure 2.11 with their 95% confidence intervals. From these figures, it can be seen that the proposed approach achieves visual quality improvement over HEVC. The intersection of any horizontal

line with the proposed and reference plots will give us two points with equal subjective quality scores. The amount of bitrates that are associated with these two points will give the amount of bitrate required for the proposed approach and the reference software. The difference between these two amounts will give us the bitrate reduction offered by the proposed approach at that quality score. For example, for the *PartyScene* video sequence, at the subjective quality score equal to 4, bitrate reduction of almost 600 kbps is observed.

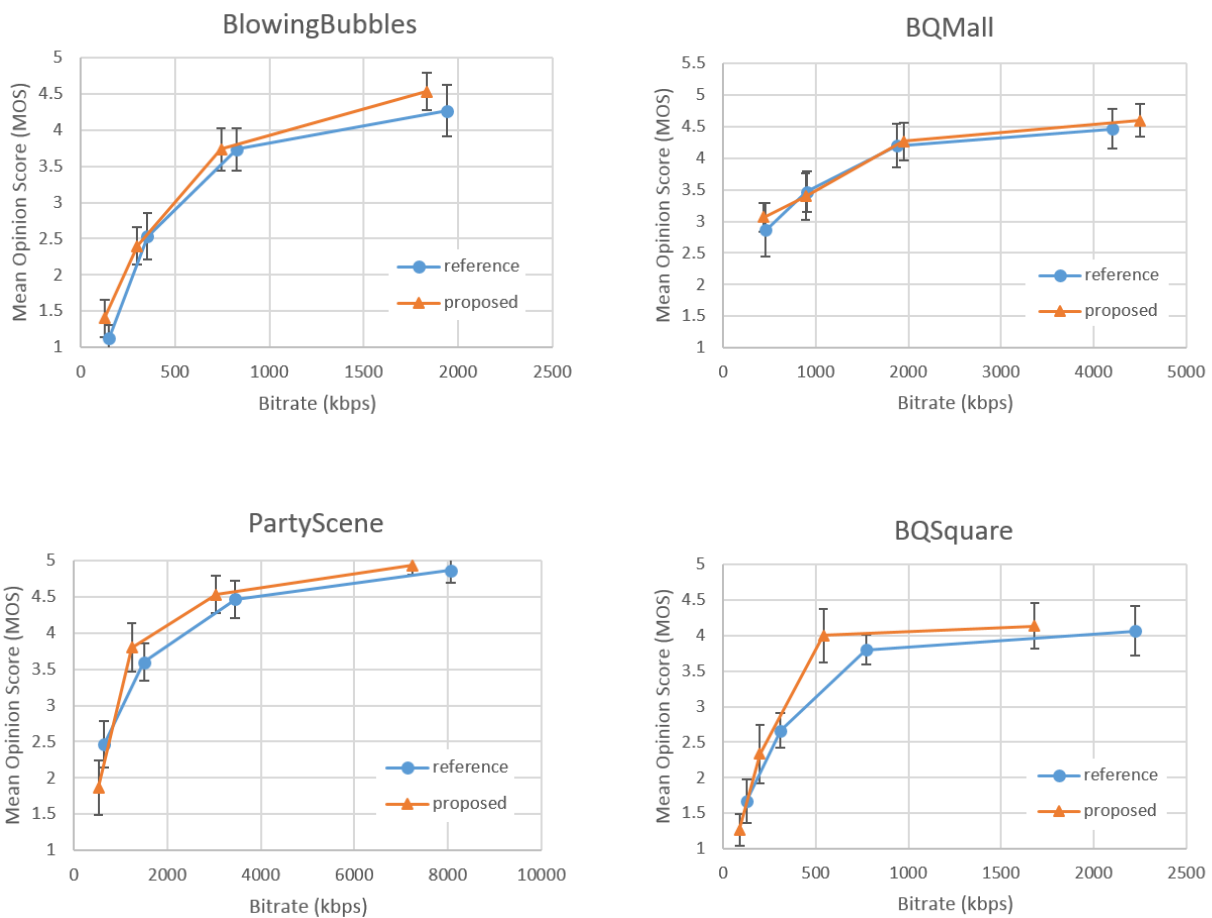


Figure 2.11 Subjective tests results. Quality (measured by mean opinion score) versus bitrate plots comparing the proposed approach with HEVC reference for different video sequences.

2.2.8 Encoding Time

Integration of the quality metric inside the encoder achieves higher compression efficiency at the expense of more complexity. The encoding time of the proposed approach along with the reference HEVC software are reported in Table 2.10. It is worth noting that the same QP will result in different bitrates with the reference software and the proposed approach as reported in Table 2.10. For instance, encoding the BQSquare video with the reference software results in 2,225 kbps for QP=22. The same QP value of 22, results in the bitrate of 1,362 kbps when encoded with the proposed approach. Thus, different bitrates also contribute to different encoding times although the QP is the same. If we take the Geomean of the encoding times of HEVC and the proposed approach, the ratio for the two Geomean values is 1.08 for BQSquare video sequence. This approach is used by MPEG, in their comparisons as they introduce new tools or propose complexity reduction approaches.

Encoding Time (s)		
QP	Reference HEVC	Proposed
22	3083.78	3218.59
27	2352.79	2533.15
32	1853.21	2020.79
37	1583.61	1764.24
Geomean	2148.12	2321.94
ratio	1.08	

Table 2.10 Comparison between encoding time of the proposed algorithm and HEVC reference software

The average of the Geomean ratio for all video sequences encoded in Low Delay B configuration is 1.12. Geomean ratio for RaceHorses video sequence is 1.16 while Geomean

ration for BQTerrace video sequence is 1.07. For other configuration settings, similar trend is observed for the Geomean ratio. The average Geomean ratio for Intra, Random Access and Low Delay P encoding configuration is 1.20, 1.16 and 1.16, respectively, as shown in Table 2.11.

Video sequence		Encoding Time			
		Low Delay B	Intra	Random Access	Low Delay P
Class D WQVGA 416x240	BQSquare	1.08	1.16	1.01	1.12
	BlowingBubbles	1.11	1.16	1.12	1.15
	RaceHorses	1.16	1.16	1.16	1.16
Class C WVGA 832x480	PartyScene	1.09	1.16	1.15	1.12
	BQMall	1.12	1.18	1.28	1.18
	BasketballDrill	1.15	1.19	1.18	1.17
Class B 1080p 1920x1080	Cactus	1.19	1.19	1.21	1.18
	BasketballDrive	1.11	1.23	1.13	1.14
	BQTerrace	1.07	1.19	1.14	1.14
Class A 2560 x1920	People on Street	1.10	1.18	1.18	1.18
	Traffic	1.16	1.37	1.23	1.18
Average		1.12	1.20	1.16	1.16

Table 2.11 Encoding time Geomean ratio for the proposed approach compared with the reference HEVC software for different prediction structures

2.3 Conclusions

In this chapter, we integrate the PSNR-HVS video quality metric inside the HEVC video coding standard to control the rate distortion optimization process. PSNR-HVS is used to measure distortion in the coding unit structure and mode selection of HEVC. In the first step, the Lagrange multiplier was selected based on linear scaling. Furthermore, the optimal Lagrange multiplier was found based on the quantization parameter. The proposed approach was tested for different prediction structures of HEVC and on different standard video sequences with various

resolutions. The results show that, for the same perceived video quality, our proposed scheme requires on average 10.21% lower bitrate than the unmodified HEVC.

Chapter 3: Content Adaptive Perceptual Video Coding

HEVC achieves better performance compared to the previous video coding standards. Through continuous efforts, each video coding standard aims to maximize compression capability with the available computational resources that were practical at the time of standardization. As discussed in the previous chapter, the Sum of Squared Errors (SSE) is not a good metric for distortion measurement as it correlated poorly with perceptual quality. As humans are the ultimate viewers of the video, the perceived quality by them is what matters the most. In the previous chapter, we developed a perceptual-based RDO. In this chapter, we further develop our proposed approach by adjusting the trade-off between the perceptual distortion and the bitrate based on the characteristics of the video content. Experimental results show that the proposed approach further improves the compression efficiency of HEVC

In the rate distortion optimization, the Lagrangian technique is used to convert a constrained optimization problem (minimum distortion with constraint amount of bitrate) to an unconstrained optimization problem as follows:

$$\min\{J\} = \min(D + \lambda R) \quad (3.1)$$

where J is the cost function, D and R are the distortion and bit-rate of the current coding unit. λ is the Lagrangian multiplier that controls the trade-off between distortion and the bitrate. If we take the derivative of J and set it to zero, minimum J can be found:

$$\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0 \quad (3.2)$$

which leads to:

$$\lambda = -\frac{dD}{dR} \quad (3.3)$$

The Lagrangian multiplier, λ , is the negative slope of the tangent to the R-D curve. For our perceptual-based rate distortion optimization method developed in the previous chapter, we used the same λ for all sequences. In this chapter, first we demonstrate some observations that justify having a content-adaptive λ in section 3.1. Afterwards, we investigate possible avenues for deriving the content-adaptive λ and explain our proposed approach in section 3.2. The experimental results are presented in section 3.3. Finally, conclusions are made in section 3.4.

3.1 Observations

Figure 3.1 compares the quality of the perceptual RDO with the MSE-based RDO traditional method. Each of four shorter RD curves represent manual modifications of the Lagrangian multiplier for each fixed quantization parameter. Optimal Lagrangian multiplier can be found by the tangent slope of the R-D curve at each point. As figure 3.1 shows, higher amounts of bitrate savings can be achieved by appropriate choice of λ for a specific video sequence at each quantization parameter.

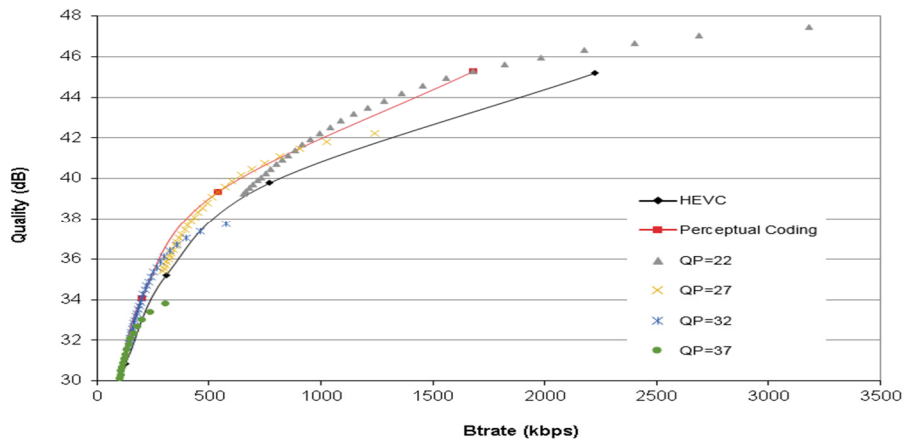


Figure 3.1 Quality versus bitrate for perceptual coding and HEVC

Furthermore, Figure 3.2 shows R-D curves for three different video sequences: *BQSquare*, *BlowingBubbles* and *PartyScene*. We fit a power function ($D = \alpha R^\beta$) to the RD points for each video sequence. Table 3.1 shows the values of α and β for each video sequence. The closeness of the curve fitting is measured by R^2 metric [98]. Maximum value for R^2 is 1 and higher values of R^2 means better fitting. Based on Table 3.1, it can be seen that all R^2 values are close to 1. This result allows us to use power approximation for our R-D curves. Another very important observation from Figure 3.2 is that the Lagrangian multiplier varies with QP and is not necessarily the same for different video sequences. This is the motivation for examining content-adaptive Lagrange multiplier estimation.

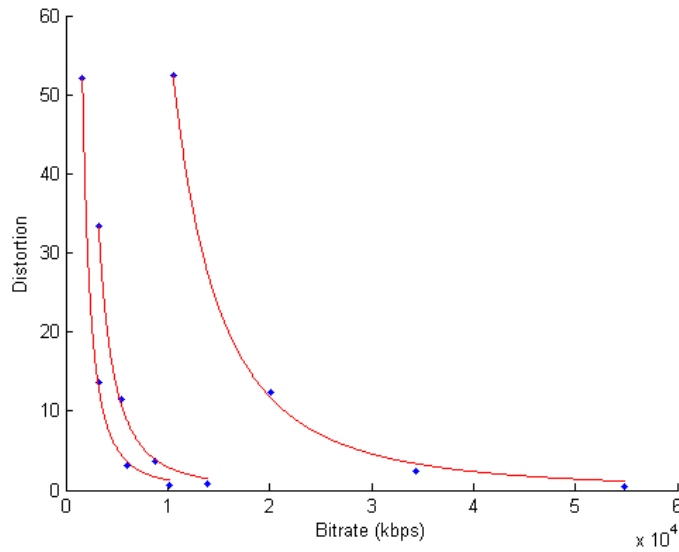


Figure 3.2 Rate distortion curve fitting for various test sequences

	α	β	R^2
BQSquare	9.759e+08	-2.129	0.9991
BlowingBubbles	1.806e+08	-2.036	0.9994
PartyScene	1.184e+11	-2.325	0.9990

Table 3.1 Rate distortion curve fitting parameters

3.2 Lagrangian Multiplier Estimation

In our proposed approach, we take the first frame of the video as a key frame. Parameters of our model are obtained based on the key frame. Those parameters remain constant for the following frames since consecutive frames generally have high correlation. We will encode the key frame with two different QP values to get two distinct points, (R_1, D_1) and (R_2, D_2) . These two points are used to estimated model parameters α and β :

$$\alpha = D_1 \exp \left\{ -\frac{\ln(R_1) \times \ln(D_2/D_1)}{\ln(R_2/R_1)} \right\} \quad (3.4)$$

$$\beta = \frac{\ln(D_2/D_1)}{\ln(R_2/R_1)} \quad (3.5)$$

By having the model parameters, the Lagrangian multiplier will be found based on the tangent line on the R-D curve.

We find the Lagrangian multiplier at each point by its horizontal and vertical projection on the R-D curve. This is based on the Lagrange multiplier estimation method by slope approximation in [101]. As illustrated in figure 3.3, (R_v, D_v) and (R_h, D_h) denote the horizontal and vertical projections of point P , respectively. Based on the power approximation for the RD curve, (R_v, D_v) and (R_h, D_h) can be found as:

$$R_v = R_p \quad (3.6)$$

$$D_v = \alpha R_p^\beta \quad (3.7)$$

$$R_h = \left(\frac{D_p}{\alpha} \right)^{\frac{1}{\beta}} \quad (3.8)$$

$$D_h = D_p \quad (3.9)$$

From (R_v, D_v) and (R_h, D_h) , λ is calculated as:

$$\lambda = -\frac{D_v - D_h}{R_v - R_h} \quad (3.10)$$

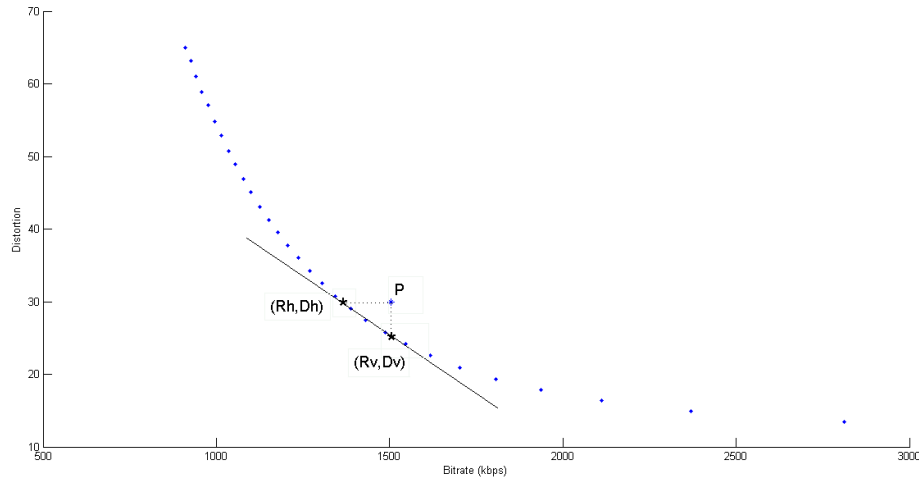


Figure 3.3 Tangent line approximation

3.3 Experimental Results

Figure 3.4 shows the estimated values for the Lagrangian multiplier coefficient for different video sequences. At a certain quantization parameter, higher value for the Lagrangian multiplier means that we can save more percentage of bits at the cost of a relatively small increase of distortion for that video sequence. On the other hand, smaller values for the Lagrangian multiplier shows that small percentage of bitrate increase leads to significant distortion reduction in that video sequence. Thus, the rate distortion performance will be improved. In case of scene changes, characteristics of the subsequent frames differ significantly from the previous key frame. In those cases, new key frames will be inserted.

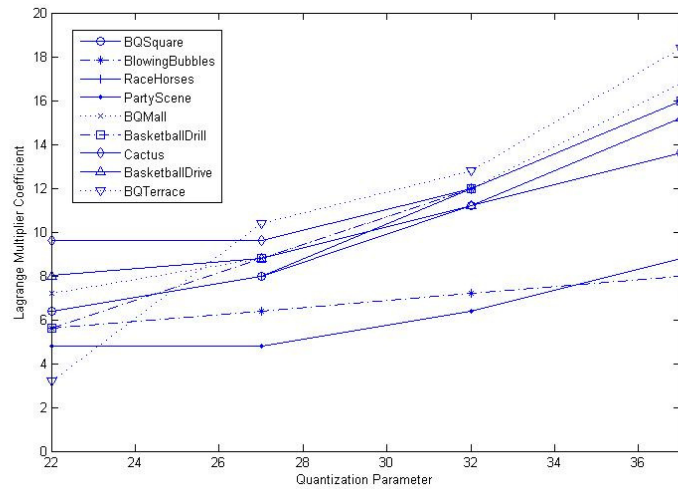


Figure 3.4 Lagrange multiplier versus quantization parameter

Table 3.2 shows the bitrate and the corresponding video quality for the proposed content adaptive approach and the reference HEVC software for test video sequences. The test video sequences in Table 3.2 have different resolutions (416×240, 832×480 or 1920×1080). The amount of motion details and texture of the video sequences differ between the test video sequences. Bjontegaard’s Delta (BD) Rate [99] measures the average bitrate savings by the proposed approach compared to the reference HEVC software. The quality is measured by PSNR-HVS.

Video Test Sequences	QP	Reference video		Proposed approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Bitrate (kbps)	Quality (dB)	Bitrate (kbps)	Quality (dB)		
BQSquare	22	2225.06	45.18	1559.88	44.93	-21.80%	-21.69%
	27	772.30	39.79	541.38	39.34		
	32	308.49	35.19	205.4144	34.28		
	37	126.62	30.87	95.1944	29.824		
Party Scene	22	8054.40	45.83	7573.59	46.52	-11.21%	-11.13%
	27	3447.49	39.25	3528.88	40.49		
	32	1504.46	33.38	1547.59	34.304		
	37	643.55	28.31	677.63	28.819		
Basketball Drive	22	19835.95	42.40	18066.34	42.85	-7.16%	-6.97%
	27	6750.85	38.40	7142.30	39.08		
	32	3111.42	34.46	3245.83	34.87		
	37	1584.96	30.98	1541.98	30.83		
BQTerrace	22	52793.42	42.27	53739.42	43.58	-22.65%	-22.69%
	27	7558.00	38.66	5212.41	38.69		
	32	1989.84	35.05	1756.07	34.90		
	37	760.90	31.12	660.77	30.33		

Table 3.2 Bitrate and quality of the proposed content-adaptive method along with the HEVC reference software for various test video sequences

Figure 3.5 shows the quality versus bitrate for *BQSquare*, *PartyScene*, *BasketballDrive* and *BQTerrace* sequences.

Table 3.3 presents the bitrate savings of the proposed approach in comparison to the reference HEVC software. Test video sequences are grouped into three classes: class B with HD resolution, and classes C and D, resolutions of Wide Video Graphics Array (WVGA) and Wide Quarter Video Graphics Array (WQVGA), respectively. Based on results in table 3.3, on average 10.81% bitrate savings can be achieved by the proposed approach.

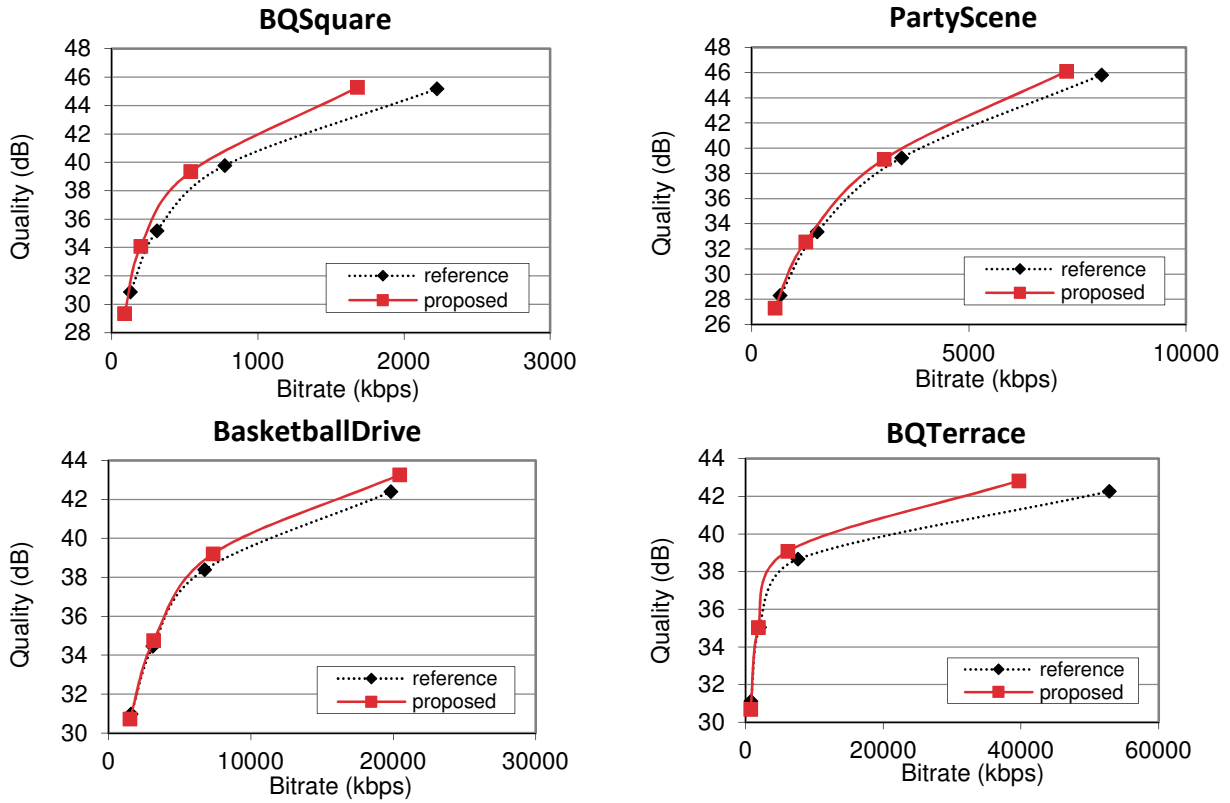


Figure 3.5 Quality versus bitrate plots comparing the proposed content-adaptive approach with the HEVC

reference

	Video sequence	$\Delta rate$
Class D	BQSquare	-21.80%
	BlowingBubbles	-11.01%
	RaceHorses	-5.76%
Class C	PartyScene	-11.21%
	BQMall	-6.62%
Class B	BasketballDrill	-3.85%
	Cactus	-7.26%
1080p	BasketballDrive	-7.16%
	BQTerrace	-22.65%
1920x1080	Average	-10.81%

Table 3.3 Rate reduction of the proposed method compared to the HEVC reference software

3.4 Conclusions

In this chapter, we developed a content adaptive approach for finding the Lagrangian multiplier for our proposed perceptual video coding. Model parameters of the video sequence are found for the key frames and are used to estimate the Lagrangian multiplier. Our experimental results show 10.81% bitrate savings in comparison to the HEVC reference software over wide range of video sequence.

Chapter 4: Perceptual 3D Video Coding

In recent years, 3D video coding format has seen a rapid development from the industry leaders and research communities. On the consumer side, cinema screens capable of showing 3D movies and various movies captured in 3D are already available. 3D-capable TV sets, 3D broadcast channels and the release of the 3D Blu-ray discs have brought 3D video to the consumers' homes. Additionally, auto-stereoscopic displays allow more views to be displayed to the viewer, giving the depth perception present in 3D TVs and at the same time, enabling the viewer to see the scene from different angles. Technology present in the auto-stereoscopic displays removes the need for wearing 3D glasses making it more convenient for the consumer to view 3D.

In both HEVC and 3D-HEVC, the rate distortion optimization (RDO) process selects the best coding mode and partitioning size for a coding unit. Distortion is minimized between the reconstructed and original video sequences with constrained bitrate. This minimization problem is split into a series of smaller minimization problems to keep the encoding delay within an acceptable range [12]. The distortion measurement is done by sum of squared differences (SSD) for mode selection and sum of absolute differences (SAD) for motion compensation in both HEVC and 3D-HEVC. Mathematical distances such as SSD or SAD are straightforward to implement, but poorly correlate with subjective evaluations. Integration of perceptual metrics that have higher correlation with subjective quality of experience can increase coding efficiency of the encoder.

In chapter 2, we proposed to integrate PSNR-HVS in the rate distortion optimization of HEVC. PSNR-HVS [30] is a full reference perceptual video quality metric based on the

characteristics of the human visual system. Experimental results showed that compression efficiency of HEVC will be improved by integration of PSNR-HVS in the coding unit structure and mode selection process of HEVC. In 3D-HEVC, each component signal is coded using an HEVC-based codec. However, direct extension of the proposed approach in our work will not improve the compression efficiency of 3D-HEVC. Compared to HEVC, 3D-HEVC introduces additional coding tools and inter-component prediction techniques. These dependencies are the main root for the 2D approach not being applicable to 3D video coding. In this chapter, we investigate inter-component dependencies in 3D-HEVC. These dependencies are the root differences between HEVC and 3D-HEVC. After a brief overview of the inter-component dependencies in 3D-HEVC in section 4.1, we will state our proposed approach along with necessary modification for stereo and multiview video coding in section 4.2. Experimental results are presented in section 4.3. Finally, conclusions are made in section 4.4.

4.1 Overview of Inter-component Dependencies in 3D-HEVC

4.1.1 Stereo Video Coding

In 3D-HEVC, the base view is coded using an unmodified HEVC. The dependent view uses all the tools of an HEVC encoder as well as additional tools such as inter-view motion prediction and the inter-view residual prediction. In 3D-HEVC, layers of HEVC-coded video sequences are multiplexed into one bitstream. These layers can depend on each other. Inter-layer dependencies increase compression efficiency of 3D-HEVC by removing redundancies in different layers [57]. Figure 4.1 shows the basic structure of the 3D video codec for compressing two views for the stereoscopic video. Figure 4.2 illustrates the dependencies between various frames of the base view and the dependent view in stereoscopic video coding.

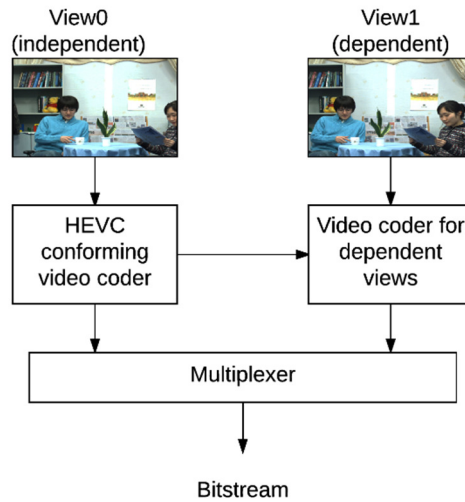


Figure 4.1 Basic codec structure for compression of the base view and the dependent view

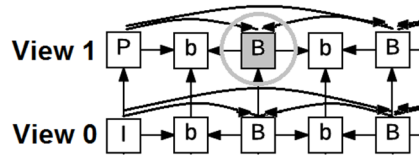


Figure 4.2 3D-HEVC prediction structure for various frames for a 2-view case

Different views of a multiview video capture the same scene but from different angles. Thus, in most cases, there are similarities in the motion in the different views. Based on this fact, the motion parameters of a dependent view can be predicted from previous views. Therefore, inter-view motion parameter prediction has been implemented in 3D-HEVC to increase the coding efficiency.

In addition to inter-view motion parameter prediction, taking advantage of the similarities between inter-view residuals improves the coding efficiency. The disparity vector points the

corresponding block in the reference view. The residuals of that block are subtracted from the current residual and the difference is coded.

For determining the corresponding block, the disparity vector is used. A disparity vector is based on an estimate of the depth map. Depth map estimation is based on the already coded depth data of the previous view by warping.

Disparity compensation prediction, inter-view motion parameter prediction and inter-view residual prediction are some of coding tools in 3D-HEVC that are not present in HEVC. In addition to these tool, multi-view video coding takes advantage of new depth coding modes, motion parameter inheritance and view synthesis optimization. In the next section, we briefly summarize the new coding tools in the multiview video coding.

4.1.2 Multi-view Video Coding

Auto-stereoscopic displays allow viewing the scene from various angles without wearing 3D glasses. Intermediate views are generated at the receiver from two or three views and their corresponding depth maps as shown in Figure 4.3. 3D-HEVC introduces new coding modes for depth maps and rate distortion optimization is replaced by view synthesis optimization.

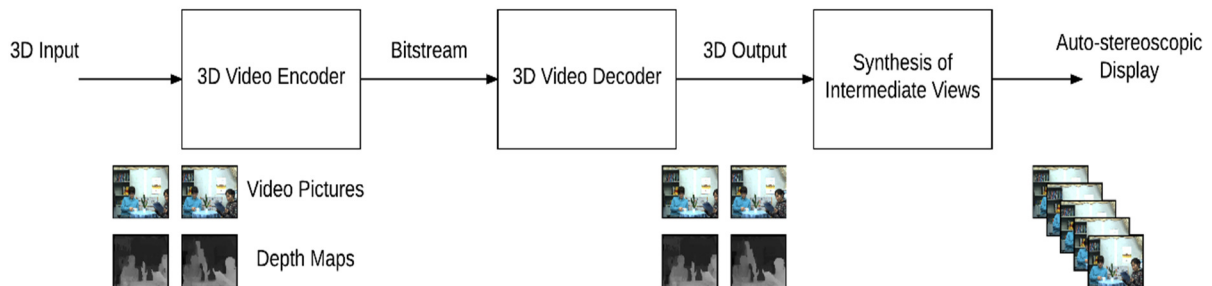


Figure 4.3 3D video data format for auto-stereoscopic displays

In general, the depth maps can be coded the same way as the video pictures. However, depth maps have different characteristics such as sharp edges and large areas of constant values. HEVC is optimized for natural video. It can code areas of constant values well enough but for better representation of sharp edges, new intra coding modes (Intra-Wedge and Intra-Contour) have been introduced to the coding algorithm. These modes partition a depth block into two areas. Each area has constant value and is not necessarily rectangular. Wedgelet partitioning is done with a straight line while contour partitioning is with arbitrary division of the two areas within a block. Also, the partitioning and motion data of a picture view can be used for coding its depth map.

In addition to new coding modes for depth maps, rate distortion optimization is replaced by view synthesis optimization in coding depth maps. For coding the depth maps, the dissimilarities between the reconstructed and the original view should be taken into account rather than the dissimilarities between the original and coded depth map. Since the depth maps are not viewed directly, the distortion measure for the depth maps needs to be modified. The distortion in the synthesized view is used in the rate distortion optimization in the encoding process of the depth maps [69]. The current coding mode in the depth map is selected by comparing two versions of the reference view. The first one is synthesized based on the reconstructed depth values of the already coded blocks while the original depth values are used for the rest of the blocks. For the second one, the current mode is used to reconstruct the depth data of the current block. Then, the dissimilarities between the two are measured with SSD. View synthesis optimization and depth coding modes are explained in section 1.3.2 in details. Next, we explain

our proposed approach for improving compression efficiency of 3D-HEVC for stereo as well as multi-view video coding.

4.2 Proposed Method

For coding unit mode decision, the smallest cost is selected among all mode candidates [12]. In the Lagrangian cost function, $D + \lambda_{mode} \times R$, D is the distortion caused by coding the considered block in a given mode and R is the number of the required bits. λ_{mode} is the Lagrangian multiplier that balances the trade-off between bitrate and distortion. λ_{mode} is selected as discussed in section 1.1.3 (equation 1.8-1.10).

The dependency of λ_{mode} on the quantization parameter has been derived empirically [15],[102]. As a measure for distortion, the sum of squared errors (SSE) is used in HEVC and 3D-HEVC. SSE is easy to implement and has low computational complexity in a video encoder but poorly correlates with the subjective video quality evaluation. Measuring the distortion by SSE in the rate distortion optimization process, leads to better PSNR of the coded video. However, the PSNR has been shown to have limited correlation with subjective tests [20],[22]. Various other video quality metrics have been developed to better represent how subjects perceive video quality.

For 3D video content, various 3D quality metrics have been developed in the literature as discussed in section 1.2.2. However, in the 3D-HEVC encoder, different views and their associated depth maps are put one after another and are coded orderly as is shown in Figure 4.4. An access unit includes all video pictures and depth maps that correspond to the same time

instant. For efficient coding of the current access unit, the data from previous access units can be used. These dependencies are explained in the previous section (section 4.1).

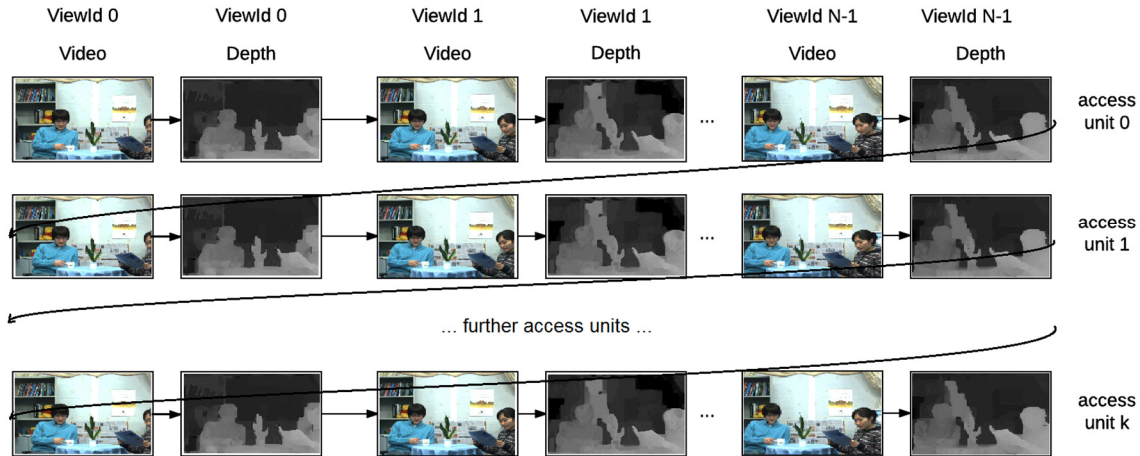


Figure 4.4 coding 3D content with depth map

To improve the compression efficiency of 3D-HEVC, we looked into integrating a 3D video quality metric inside the encoder. However, most of the 3D quality metrics take into account the quality of the depth maps or the quality of the generated disparity maps. In the 3D-HEVC encoding process, the qualities of the synthesized views with various coding modes are measured rather than the quality of depth maps. The reason behind this approach is that synthesized views are shown to the viewer and their quality matters the most. Depth maps or disparity maps are not viewed by the viewers directly. In line with this approach, we integrated a perceptual quality metric in the view synthesis optimization of 3D-HEVC.

In our proposed approach, PSNR-HVS is used for measuring the distortion in the coding unit mode selection. The concept of coding unit (CU) has been introduced in HEVC. CUs are coded as intra or inter modes and have different sizes. The best coding mode and partition size of a Coding Unit (CU) is determined in a quad-tree structure within each Coding Tree Block (CTB). In all ITU-T and ISO/IEC JTC 1 video coding standards prior to HEVC, macroblocks with fixed size were used [55]. High resolution video content benefits from the larger size of CTBs. The size of the CTB is chosen based on the needs of encoders in terms of memory and computational requirements. 3D-HEVC utilizes the same quad-tree structure as HEVC.

PSNR-HVS is a full-reference quality metric that is based on characteristics of the human visual system and shows higher correlation with the subjective video quality evaluations than PSNR [30]. The formulation for PSNR-HVS is explained in section 2.1.1 (equations 2.1 and 2.2). PSNR-HVS is adaptable to block sizes with different sizes and is not too computationally complex. These characteristics allows its application inside the rate distortion optimization process of 3D-HEVC.

In the rate distortion optimization process, λ_{mode} balances the trade-off in decreasing the bitrate or distortion. In selecting the coding modes, lower bitrate comes at the expense of more distortion. Thus, effective choice of λ_{mode} plays an important role in reaching higher compression efficiency. In our approach, the modified Lagrangian multiplier is the product of a scalar coefficient and the traditional λ_{mode} used in 3D-HEVC reference software:

$$\lambda_{proposed} = c_{view} \cdot \lambda_{mode} \quad (4.1)$$

The stereoscopic videos present different views to the left and right eye. These two views are fused in the brain to provide corresponding depth perception of the scene. In order to mimic this fusion, some 3D video quality metrics suggest to form a cyclopean view by averaging each block with its corresponding block in the dependent view [45]-[46],[49]. Similarly, in the stereoscopic video encoder, the corresponding block is found in the base view. However, rather than averaging it with the current block, the corresponding block is used as one of the candidates for inter-view motion prediction. In the encoding process, inter-view motion prediction is one of the additional coding tools in coding the dependent views as explained in the previous section 1.3.1.

In order to find c_{view} for our approach, we find the most amount of bitrate savings for test video sequences based on different values for the quantization parameter. We run our tests for four different QPs: 25, 30, 35 and 40. These values are used in the standard MPEG 3D-HEVC tests [103]. By running the reference 3D-HEVC software for these QP values, four points in the rate distortion plot are obtained. We used Bjontegaard's Delta (BD) Rate [99] in order to compare the proposed approach with the reference 3D-HEVC software. BD rate measures the average difference along the bitrate axis, between the proposed and reference curves.

Figure 4.5 and Figure 4.6 show the BD rate savings of the proposed approach for the natural-scene and computer-generated videos, respectively. BD savings show a similar trend in the three natural-scene videos of *PoznanStreet*, *Newspaper* and *PoznanHall*. We are interested in the most negative value of BD rate saving. Based on these figures, c_{view} equal to 15 leads to the most amount of BD rate savings in the proposed approach compared with unmodified 3D-HEVC. This value is selected for c_{view} in our approach.

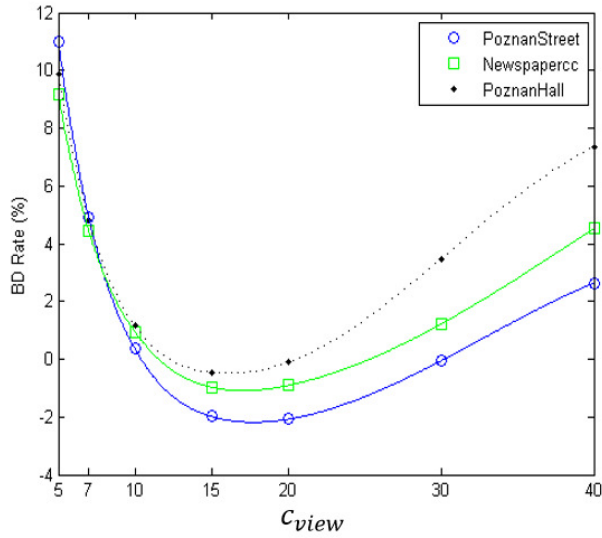


Figure 4.5 Bitrate savings for different natural-scene 3D video sequences

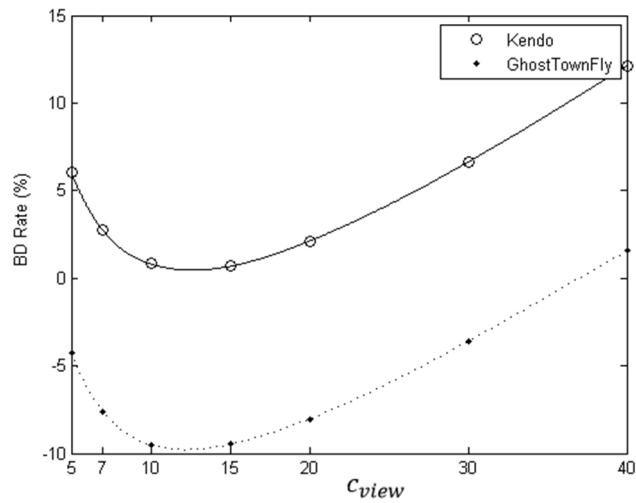


Figure 4.6 Bitrate savings for different computer-generated 3D video sequences

A stereo display needs two views for the left and right eyes. Auto-stereoscopic displays which allow viewing multiview content without glasses, require video and depth data to synthesize intermediate views. Intermediate views are synthesized by methods known as depth-image-based rendering (DIBR). DIBR techniques take the video and depth data as input and

produce intermediate views to show on an auto-stereoscopic display which allows viewing from various angles and do not need 3D glasses [104]-[105].

We integrate PSNR-HVS in coding texture views as well as their corresponding depth maps. As explained in the previous section, 3D-HEVC has new intra modes for coding depth maps [103] and rate distortion optimization is replaced by view synthesis optimization [54] in coding depth maps.

To find the optimal coefficient for the Lagrangian multiplier in coding depth maps, the quality of the synthesized views are considered. We measure the quality of the synthesized views with PSNR-HVS. Bjontegaard's Delta (BD) Rate [99] is used to measure bitrate savings of the proposed approach based on different values of c_{depth} .

Figure 4.7 shows that BD Rate savings for three video sequences: *PoznanStreet*, *NewsPaper* and *PoznanHall*. Numerical values of the BD Rate savings are reported in Table 4.1. The optimal scaling factor is found as 70 based on Table 4.1. This value leads to the highest amount of BD Rate saving for all three sequences. In the proposed approach the optimal scaling factor of the depth is higher than the optimal scaling factor for coding the texture view. This observation is in accordance with coding the views with higher quality (smaller quantization parameter) than depth maps [106].

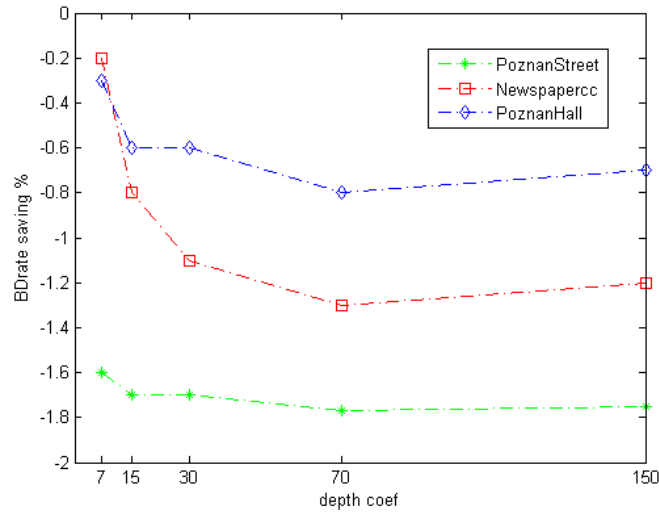


Figure 4.7 BD Rate saving based on scaling factors for the Lagrangian multiplier in coding depth maps

Scaling factor	7	15	30	70	150
PoznanStreet	-1.6%	-1.7%	-1.7%	-1.77%	-1.75%
Newspapercc	-0.2%	-0.8%	-1.1%	-1.3%	-1.2%
PoznanHall2	-0.3%	-0.6%	-0.6%	-0.8%	-0.7%

Table 4.1 BD Rate saving based on scaling factors for the Lagrangian multiplier in coding depth maps

Next, we evaluate the compression performance of our proposed approach for stereoscopic and multiview videos on standard MPEG video test sequences.

4.3 Experimental Results

4.3.1 Test Setup

To test the performance of the proposed approach, PSNR-HVS is integrated into the 3D-HEVC reference software HTM10 [107]. Test video sequences are selected from the standard MPEG test sequences [108] as summarized in Table 4.2. The test video sequences are in the YCbCr format. The length of each sequence is 10 seconds. The first frame of each video sequence is shown in Figure 4.8.

	Resolution	Frame rate	Camera order	Number of frames
PoznanStreet	1920×1088	25	4, 5, 3	250
Newspapercc	1024×768	30	4, 2, 6	300
Kendo	1024×768	30	3, 1, 5	300
PoznanHall	1920×1088	25	6, 7, 5	200
GhostTownFly	1920×1088	25	5, 9, 1	250
UndoDancer	1920×1088	25	5, 1, 9	250

Table 4.2 Test video sequences

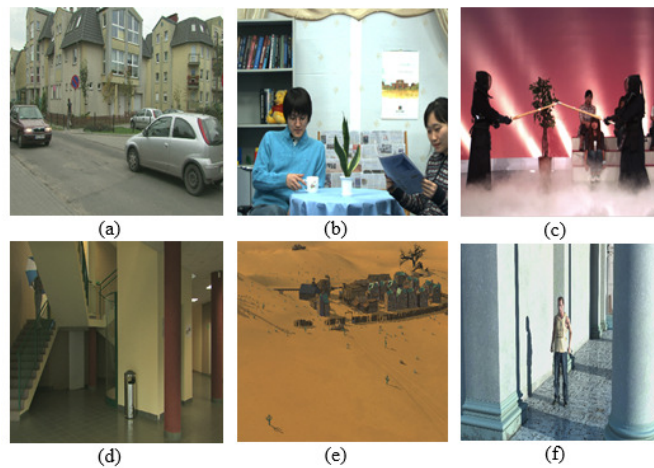


Figure 4.8 First frame of the test video sequences. (a) PoznanStreet (b) Newspaper (c) Kendo (d) PoznanHall (e) GhostTownFly (f) UndoDancer

4.3.2 Compression Performance for Stereo Videos

In Table 4.3, we summarized the results (bitrate and quality) of the proposed approach and the reference software for the stereoscopic videos. The quality is measured by PSNR-HVS. Video0 is the base view while video1 refers to the dependent view. Video (2v) contains both base view (video0) and dependent view (video1). The quality of the video(2v) is the average of the quality of video0 and video1. The bitrate of the video(2v) is equal to summation of the bitrate of video0 and video1. Video(2v) is reported in all MPEG 3D-HEVC tests.

PoznanStreet	QP	Reference 3D-HEVC		Proposed Approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Kbps (kbps)	Quality (dB)	Kbps (kbps)	Quality (dB)		
Video0	25	2438.8984	39.5482	2264.8912	39.4768	-1.8%	-1.8%
	30	956.9616	36.2907	942.2680	36.3013		
	35	461.1040	33.2217	458.1376	33.2359		
	40	241.1088	30.2582	240.9576	30.2788		
Video1	25	702.7456	38.2428	612.4696	38.1710	-4.3%	-4.3%
	30	196.1960	35.2568	183.1904	35.2264		
	35	75.4288	32.3799	73.6064	32.3647		
	40	33.5168	29.5985	33.1120	29.5919		
Video(2v)	25	3141.6440	38.8955	2877.3608	38.8239	-2.0%	-2.0%
	30	1153.1576	35.7738	1125.4584	35.7638		
	35	536.5328	32.8008	531.7440	32.8003		
	40	274.6256	29.9283	274.0696	29.9353		

Table 4.3 Bitrate and quality of the reference 3D-HEVC and the proposed approach for stereo *PoznanStreet* test video sequence

Table 4.4 summarizes the BD-rate savings for the stereo video sequences with natural scene (Figure 4.8 (a), (b) and (d)). The BD rates are reported for the base view (video0), the dependent view (video1) and also for both views together, video (2v). The average bitrate savings of the proposed approach compared with the reference 3D-HEVC software for the stereoscopic video (two views together) is 1.17%. Our proposed approach needs on average 1.17% less bitrate to transmit/store the stereoscopic video compared to the reference 3D-HEVC.

VIDEO SEQUENCES	VIDEO 0	VIDEO 1	VIDEO (2V)
	BD-RATE	BD-RATE	BD-RATE
PoznanStreet	-1.8%	-4.3%	-2.0%
NewsPapercc	-0.4%	-3.1%	-1.0%
PoznanHall2	-0.5%	-0.8%	-0.5%
Average	-0.9%	-2.7%	-1.17%

Table 4.4 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for natural scene stereoscopic videos

4.3.3 Asymmetric Coefficients

We investigate the effect of asymmetric coefficients for the Lagrangian multiplier for the base view and the dependent view. In 3D-HEVC, base view is coded with the un-modified HEVC codec but the dependent view takes advantage of additional coding tools. This difference in their coding structure motivates us to test the effect of asymmetric choice of coefficients for the base view and dependent view. Table 4.5 reports the total bitrate saving in the base and dependent view.

Base View Coef	Dependent view coefficient				
	5	7	10	15	30
5	11.0%	9.1%	8.4%	8.8%	12.8%
7	7.1%	4.9%	3.7%	3.7%	6.8%
10	4.8%	2.1%	0.4%	-0.1%	2.1%
15	4.7%	1.3%	-1.0%	-2.0%	-0.7%
30	10.4%	5.8%	2.4%	0.2%	-0.1%

Table 4.5 Bitrate saving of the base view (video0) for stereo test sequence *PoznanStreet*

Maximum amount of bitrate saving happens at the base and dependent view coefficient of 15. We observe that the symmetric choice of coefficients for the base view and the dependent view leads to the most amount of overall bitrate saving in stereo video coding.

Asymmetric coefficients for the Lagrangian multiplier has interesting results on the dependent view bitrate saving. Maximum amount of bitrate saving for the dependent view happens at very low base view coefficient and very high dependent view coefficient. With high values for the dependent view coefficients, very low amount of bitrate is assigned to the dependent views. However, the quality of the dependent view will not decrease as much due to dependency on the base view. Thus, higher bitrate savings can be seen in the dependent view

(video1). Table 4.6 reports the results of bitrate saving in the dependent view for asymmetric coefficients for the base and dependent view. Figure 4.9 shows the results in Table 4.6, visually.

Base View Coeff	Dependent view coefficient				
	5	7	10	15	30
5	31.4%	10.9%	-5.6%	-20.0%	-37.8%
7	38.7%	16.9%	-1.0%	-16.2%	-35.5%
10	48.7%	24.9%	5.1%	-11.5%	-32.6%
15	62.9%	36.5%	14.6%	-4.3%	-28.0%
30	97.3%	66.5%	37.9%	13.0%	-17.6%

Table 4.6 Bitrate saving of the dependent view (video1) for stereo test sequence *PoznanStreet*

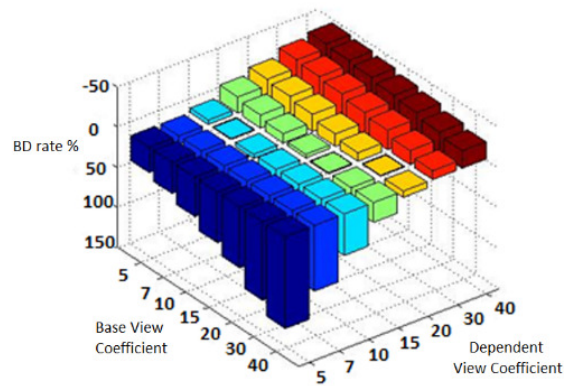


Figure 4.9 Bitrate saving of the dependent view (video1) for stereo test sequence *PoznanStreet*

Next, we test the asymmetric coefficients for the Lagrangian multiplier for depth maps of the base view and the dependent views. Figure 4.7 shows the BD rate savings for all the texture views and synthesized views. Based on Figure 4.7, higher coefficients for the Lagrangian multiplier in coding depth maps yield better results. The coefficient of Lagrangian multiplier is selected symmetrically for the corresponding depth maps of the base view and dependent views.

Depth0 Coef	Depth1 and Depth2 coefficient				
	5	7	10	15	30
5	-1.45%	-1.51%	-1.55%	-1.58%	-1.60%
7	-1.51%	-1.58%	-1.61%	-1.63%	-1.66%
10	-1.49%	-1.54%	-1.59%	-1.62%	-1.64%
15	-1.51%	-1.57%	-1.63%	-1.66%	-1.68%
30	-1.47%	-1.55%	-1.61%	-1.66%	-1.69%

Table 4.7 Overall bitrate saving of the texture views and the synthesized views for multi-view test sequence

PoznanStreet

4.3.4 Compression Performance of Multi-view Videos

Table 4.9 and Table 4.10 report bitrate and quality of the proposed approach and the reference software 3D-HEVC for the multiview video, *PoznanStreet*.

PoznanStreet	QP	Reference 3D-HEVC		Proposed Approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Kbps (kbps)	Quality (dB)	Kbps (kbps)	Quality (dB)		
Video0	25	2438.8984	39.5482	2264.8912	39.4768	-1.77%	-1.76%
	30	956.9616	36.2907	942.2680	36.3013		
	35	461.1040	33.2217	458.1376	33.2359		
	40	241.1088	30.2582	240.9576	30.2788		
Video1	25	702.7456	38.2428	611.9488	38.1700	-4.14%	-4.13%
	30	196.1960	35.2568	184.0704	35.2236		
	35	75.4288	32.3799	73.4160	32.3643		
	40	33.5168	29.5985	33.1968	29.5923		
Video2	25	691.8848	38.2764	598.0424	38.1940	-4.20%	-4.20%
	30	193.3552	35.1205	180.6056	35.0857		
	35	73.6088	32.1370	72.4144	32.1344		
	40	31.4928	29.3075	31.1672	29.3048		

Table 4.8 Bitrate and quality of the reference 3D-HEVC and the proposed approach for multiview

PoznanStreet test video sequence for texture views

PoznanStreet	QP	Reference 3D-HEVC		Proposed Approach		BD-rate (piecewise cubic)	BD-rate (cubic)
		Kbps (kbps)	Quality (dB)	Kbps (kbps)	Quality (dB)		
Video(2v)	25	3141.6440	38.8955	2876.8400	38.8234	-1.95%	-1.93%
	30	1153.1576	35.7738	1126.3384	35.7624		
	35	536.5328	32.8008	531.5536	32.8001		
	40	274.6256	29.9283	274.1544	29.9355		
Video(3v)	25	3833.5288	38.6891	3474.8824	38.6136	-2.23%	-2.21%
	30	1346.5128	35.5560	1306.9440	35.5369		
	35	610.1416	32.5795	603.9680	32.5782		
	40	306.1184	29.7214	305.3216	29.7253		
Synth(2v)	25	3405.3760	39.9827	3121.5128	39.8349	-1.14%	-1.12%
	30	1227.4672	36.3912	1195.0152	36.3353		
	35	566.9720	33.2272	560.1752	33.1879		
	40	290.5952	30.2542	289.1784	30.2206		
Synth(3v)	34	4213.8056	39.9734	3830.1112	39.8216	-1.46%	-1.45%
	39	1449.2304	36.3440	1402.0256	36.2854		
	42	650.9504	33.1638	642.4496	33.1247		
	45	326.9560	30.1811	325.1632	30.1490		
All(2v)	34	3405.3760	39.5478	3121.5128	39.4303	-1.54%	-1.52%
	39	1227.4672	36.1442	1195.0152	36.1061		
	42	566.9720	33.0566	560.1752	33.0328		
	45	290.5952	30.1238	289.1784	30.1066		
All(3v)	34	4213.8056	39.5453	3830.1112	39.4190	-1.77%	-1.76%
	39	1449.2304	36.0814	1402.0256	36.0359		
	42	650.9504	32.9690	642.4496	32.9425		
	45	326.9560	30.0279	325.1632	30.0078		

Table 4.9 Bitrate and quality of the reference 3D-HEVC and the proposed approach for multiview

PoznanStreet test video sequence for texture views and synthesized views

The performance evaluation of our proposed approach compared to the reference 3D-HEVC software is presented in Table 4.10 and Table 4.11. These two tables summarize the BD-rate savings for the multiview video sequences. The BD rates are reported for the base view (video0), the dependent views (video1-video2) and their corresponding depth maps (depth0-depth2). Video(2v) and video(3v) accounts for BD rate savings of two and three views, respectively. Synth(2v) and synth(3v) report the BD rate savings for all the synthesized views

based on two and three views, respectively. Finally, all(2v) and all(3v) accounts for BD rate saving of texture views and synthesized views all together.

	Video0	Video1	Video2	Depth0	Depth1	Depth2
PoznanStreet	-1.77%	-4.14%	-4.20%	52.18%	83.68%	79.82%
Newspapercc	-0.44%	-3.18%	-3.20%	65.83%	86.85%	82.33%
Kendo	0.76%	-0.36%	-0.24%	68.64%	74.39%	77.27%
PoznanHall	-0.48%	-0.86%	-1.78%	67.73%	86.23%	75.60%
GhostTownFly	-2.79%	-12.37%	-12.22%	106.22%	150.90%	155.51%
UndoDancer	-9.02%	-13.32%	-13.99%	41.84%	58.78%	52.44%

Table 4.10 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for multi-view videos. Video0 is the base view, video1 and video2 refer to the dependent views.

	Video(2v)	Video(3v)	Synth(2v)	Synth(3v)	All(2v)	All(3v)
PoznanStreet	-1.95%	-2.23%	-1.14%	-1.46%	-1.54%	-1.77%
Newspapercc	-0.98%	-1.39%	-0.06%	-0.62%	-0.95%	-1.29%
Kendo	0.71%	0.54%	-0.16%	-1.03%	-0.87%	-1.59%
PoznanHall	-0.49%	-0.71%	-0.19%	-0.50%	-0.54%	-0.78%
GhostTownFly	-3.42%	-4.18%	-2.98%	-3.49%	-3.11%	-3.61%
UndoDancer	-9.47%	-10.01%	-6.01%	-6.89%	-7.16%	-7.66%

Table 4.11 Bitrate saving of the proposed approach compared to the 3D-HEVC reference software for multi-view videos. Video(2v) and video(3v) refer to two and three texture views. Synth(2v) and synth(3v) correspond to the synthesized views. All(2v) and All(3v) refer to all the texture views and synthesized views.

The average bitrate savings of the proposed approach compared with the reference 3D-HEVC software for the multi-view video (three views together) is 2.78%. Our proposed approach needs on average 2.78% less bitrate to transmit/store the multi-view video compared to the reference 3D-HEVC Computer generated contents such as *GhostTownFly* and *UndoDancer* show higher bitrate savings in our proposed approach. By comparing Table 4.10 with Table 4.3, small difference in the Video1 BD rate saving can be noticed for test sequence *PoznanStreet*. The reason for that is the dependency of video1 on depth0 in multi-view coding in Table 4.10. In

the tests reported in Table 4.3, no depth map was coded for the stereo video. In multi-view video coding in Table 4.10 and Table 4.11, we searched for optimizing all texture views as well as all synthesized views.

4.3.5 Subjective Evaluation

We examine the performance of our proposed approach by subjective tests. Subjective assessment methods are used to more directly measure the reaction of the ultimate viewers of the video content. We run our tests in the standard viewing conditions outlines in ITU-R BT.500 [100]. Based on this standard, methodology for 3D video quality assessment is developed in [109]. Our display is a 55-inch curved polarized 3D display. Distance of the viewers from display was 5 times the height of the display. 15 subjects took part in our tests with average age of 29. Minimum age of our subjects was 24 and their maximum age was 33. All subjects in our test have been assessed and passed the visual acuity test (using Snellen charts), color blindness test (using Ishihara chart) and stereovision acuity test (via Randot test–graded circle test 100s of arc). For each sequence, four quantization parameters were used. Quantization parameters were selected as is suggested by MPEG standard tests. Each video sequence was coded both with the proposed approach and with the reference 3D-HEVC resulting in eight different videos. We used the double stimulus impairment scale (DSIS) method. In DSIS, the subjects are presented with a series of pictures with various levels of impairment in random order. An un-impaired picture is included to serve as the reference for assessments. The grading scale has five labels: excellent, good, fair, poor and bad.

Prior to the test, there was a training session for subjects to familiarize them with the test procedure. Subjective assessments are converted to scores in the range of 0 to 5. The mean

opinion score (MOS) over all subjects is plotted for the proposed approach and is compared with 3D-HEVC. The rate distortion curves obtained from the subjective tests are shown in Figure 4.10 and Figure 4.11 with their 95% confidence intervals.

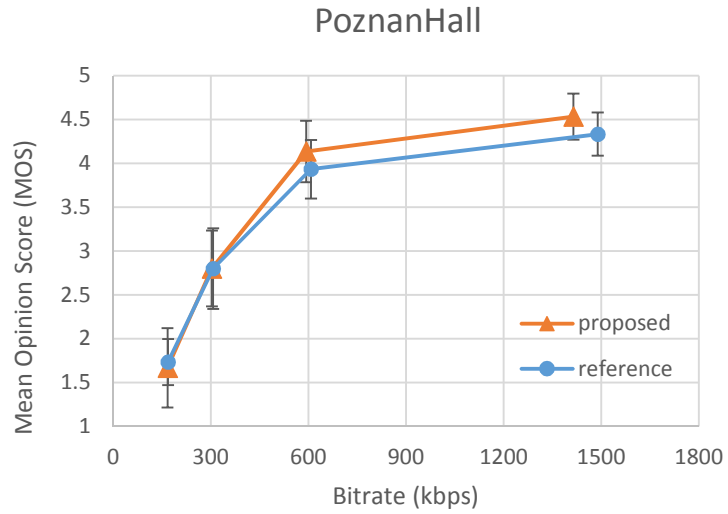


Figure 4.10 Subjective tests results for video *PoznanHall*

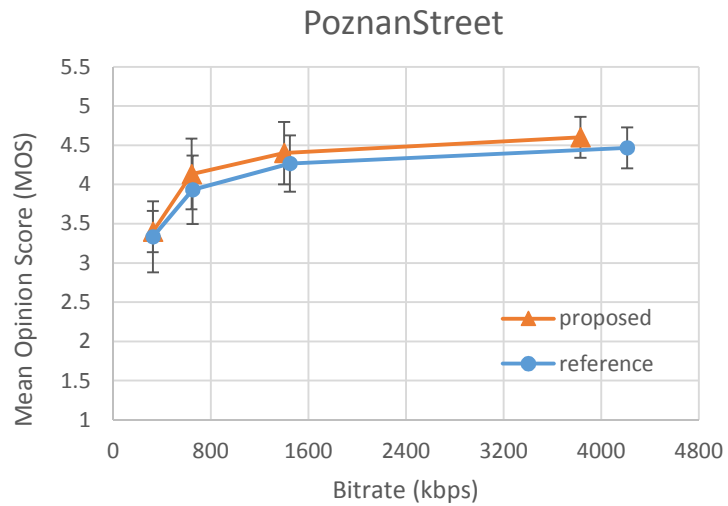


Figure 4.11 Subjective tests results for video *PoznanStreet*

From these figures, it can be seen that the proposed approach achieves visual quality improvement over 3D-HEVC.

4.3.6 Complexity Overhead

Integration of the quality metric inside the encoder achieves higher compression efficiency at the expense of more complexity. The encoding time of the proposed approach along with the reference 3D-HEVC software are reported in Table 4.12. By taking the geomean of the encoding times of 3D-HEVC and the proposed approach, the ratio for the two geomeans is 1.25 on average for the two test video sequences. This approach is used by MPEG, in their comparisons as they introduce new tools or propose complexity reduction approaches.

	QP	Encoding Time (s)		Geomean Ratio
		Reference 3D-HEVC	Proposed Approach	
PoznanStreet	25	53388.89	71489.96	1.23
	30	43990.69	54890.36	
	35	36790.86	44360.87	
	40	34106.29	38429.87	
Newspaper	25	31352.09	37572.18	1.19
	30	24916.3	31376.64	
	35	21766.78	25722.04	
	40	19449.43	21920.65	
Kendo	25	34950.32	44985.98	1.30
	30	25852.92	33992.25	
	35	21705.1	28400.3	
	40	19698.43	25261.31	
GhostTownFly	25	64040.92	89856.95	1.27
	30	48412.77	62694.36	
	35	41495.84	51396.46	
	40	36879.44	42706.3	
average				1.25

Table 4.12 Comparison between encoding time of the proposed approach and 3D-HEVC reference software for the multi-view video coding

4.4 Conclusions

In this section, we improved the compression efficiency of 3D-HEVC for the stereoscopic video and multi-view video by integrating the PSNR-HVS perceptual video quality metric inside the encoder. We used PSNR-HVS as a measure of distortion in the rate distortion optimization process for the coding unit structure and mode selection. Compared to the 2D case, 3D-HEVC deals with more complex structure and inter-component dependencies between different texture views and depth maps. Dependent views are coded with additional tools and depth maps have new coding modes. Also, rate distortion optimization is replaced by view synthesis optimization in coding the depth maps. Our proposed approach was tested on a variety of standard 3D video sequences. The results show that our proposed scheme provides on average 2.78% bitrate savings compared to the unmodified 3D-HEVC.

Chapter 5: Asymmetric 3D Video Coding

An auspicious technique for reducing the amount of data required for the storage and transmission of the stereoscopic video is to low-pass filter one of the views while keeping the other view at the original state. This technique is based on the fact that the human visual system (HVS) perceives high quality 3D even if only one of the views is of high quality. In the literature review in section 1.4, we have discussed the related research that takes advantage of this characteristic of HVS. However, sustained imbalance in the two views causes fatigue in the viewers and has negative effects on children's premature visual system. While splitting the low-pass filtered frames between the right and left views over time is a valuable test, time-interleaving did not achieve acceptable results. Here, we evaluate the compression performance of a new method for asymmetric stereoscopic video. We have applied low-pass filtering to slices of both views while the corresponding slice in the other view is of high quality. We tested the perceived sharpness, depth and quality of the filtered stereoscopic video subjectively and compared it to the original video. It is a propitious test because low pass filtering slices of both views reduces the amount of bitrate required for transmission and storage of stereoscopic video while maintaining the stereoscopic quality at high levels.

5.1 A New Method for Asymmetric Video Coding

An auspicious technique for reducing the amount of data required for transmitting stereoscopic video is to reduce the quality of one of the views and keep the other view at the original quality. Based on the suppression theory of the binocular vision [70], sharp edges in the high quality image masks the blur in the low quality view and the overall depth impression is close to the sharper view. Asymmetric stereo video is a promising approach for reducing the

amount of bandwidth or memory required for transmission or storage of the stereoscopic video but it is not a fair approach for people with a right or left dominant eye [74]. If the high quality sequence is shown to their weak eye, the overall impression of the 3D video is not close to the high quality sequence. In addition, sustained imbalance in the two views has negative effects on children’s premature visual system and causes fatigue in the viewers.

In our implementation, we modify the asymmetric coding so that lower quality parts are distributed in each frame of both views. We have applied low-pass filtering to slices of both views while the corresponding slice in the other view is of high quality. We examined the perceived sharpness, depth and quality of stereoscopic videos after low-pass filtering alternate horizontal slices in the right and left views. A large variety of filter levels and sizes of horizontal slices were considered. Figure 5.1 shows one example where the odd slices of the left view and the even slices of the right view are filtered. We considered only horizontal slices since the horizontal disparity between the two views has the potential of causing the same object to be filtered in both views in the case of other directions (e.g., vertical).

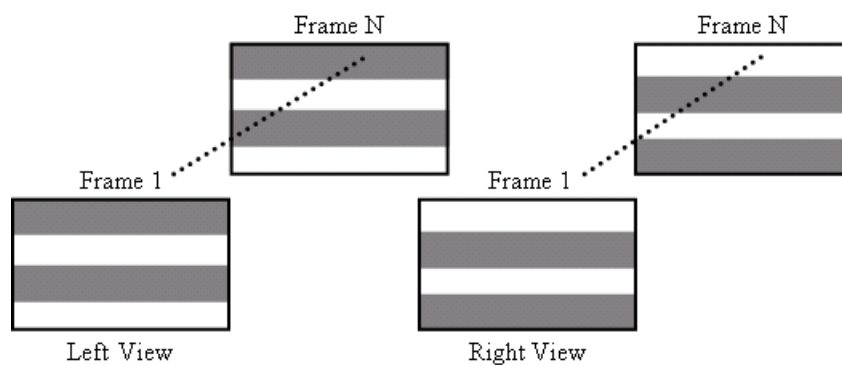


Figure 5.1 Illustration of an example of our filtering pattern. Grey areas represent filtered slices in each view of the stereoscopic video.

To reduce the discontinuity at the slice borders, we reduced the strength of filtering along these borders. In order to do that, we selected a filtering strength that follows a bell shaped pattern so that the center of the slice is strongly filtered compared to its borders.

In our study, we considered several numbers of horizontal slices in each frame: 2, 4, 10, 40 and 72. Performance evaluations have shown that 10 slices per frame provide the best visual quality for all different levels of low-pass filtering, with a fair distribution of reduced quality in both views. An excessive number of slices – such as 40 and 72 - seemed to annoy the viewers.

Our choice of a low-pass filter was a 15×15 Gaussian filter which has the following form:

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5.1)$$

We used a Gaussian filter as it is a typical and well-known low-pass filter. The size of the slicing was 15×15. In order to generate filters for the different slice sizes with less and more strength the sigma in eq. 5.1 was varied. In the first set of our tests, the strength of the filter followed a pulse-train pattern, which allowed us to apply it to every other slice while keeping the in-between slices at their original quality. In the second set of our tests, we ensured a smooth transition between the filtered and unfiltered slices by applying weaker filtering at the borders compared to the center of the slice. For the pixels near the slice borders, we applied weak filters with sigma close to zero. As the pixels get further from the borders, the sigma of the filter increases such that the strongest filtering is applied to the pixels in the center of every slice. Figure 5.2.a and Figure 5.2.b show the filtering pattern of the right and left views in the first and the second set of tests, respectively.

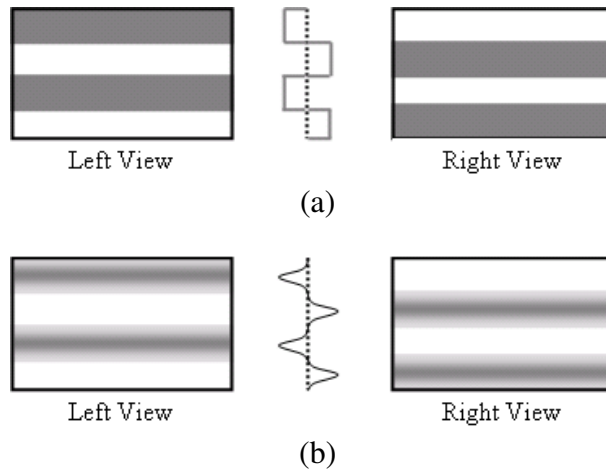


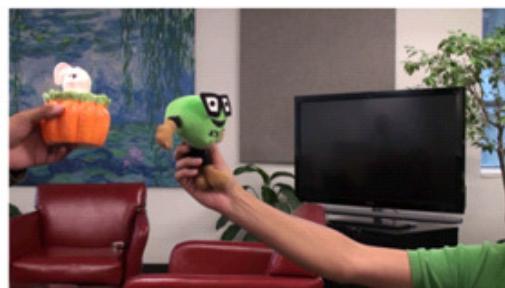
Figure 5.2 (a) left and right frames with unsmoothed slice borders. (b) left and right frames with smoothed borders

5.2 Experimental Results

We considered two representative stereo video sequences for our tests. The first video, “Mother and Kid” was taken outdoors. It contains different levels of details such as human faces and textures such as bushes and grass. The second video, “Two Dolls”, has fewer details and includes two dolls being moved in front of the camera. It was shot indoors. Each sequence is 10 second long with 30 fps. The resolution of each view is 1920×1080 pixels. Our test videos were captured with two identical HD camcorders (1080i, 60Hz, NTSC) set up in parallel. Figure 5.3.a and Figure 5.3.b show the first frame of the right views of each.



(a)



(b)

Figure 5.3 The first frame of our two video sequences. (a)First video sequence: Mother and Kid (b) Second video sequence: Two Dolls. These frames are the right view of our stereoscopic videos

We applied four Gaussian filters with four different sigma values of 1, 3, 10 and 30, to each of test sequences. The original video along with its four low-pass filtered videos form five different versions for each sequence, resulting in ten unique stereo sequences.

To compare how people perceive the quality degradation obtained by the same filters in 2D and 3D videos, we used the left view of each sequence as their 2D version and applied the above-mentioned filters to them. This formed 5 unique 2D test videos for each sequence.

As a next step, in order to have smoothed slice borders, we considered the three sigma values (3, 10 and 30) for the maximum filter strength. That strength is applied at the center row of each slice and gradually decreases to the minimum sigma value of 1 for pixel rows closer to the slice borders. This implementation follows the pattern mentioned before.

For each sequence, there were eight stereoscopic test sequences as well as five monoscopic ones. We asked the viewers to rate the overall quality, depth and sharpness of sequence after viewing the original one. All test sequences were shown in a random order and

the subjects were not aware of the test objectives. Table 5.1 summarizes the parameters used in our experiments.

Filter type	Gaussian	
Number of horizontal slices	10	
Display method	Stereo and non-stereo	
Video sequences	Mother and Kid (outdoor) Two Dolls (indoor)	
Filter strength	Pulse Pattern Sigma of 1, 3, 10 and 30	Bell Shape Max sigma of 3, 10 and 30

Table 5.1 Parameters employed for our experiments

We showed our tests to 14 viewers. These viewers were between 23 and 38 years old with mean age of 28. Gender distribution was not controlled. All viewers were screened for visual acuity, color vision and contrast sensitivity. All viewers passed the screening tests.

We used a 65” 3D HD TV with 16:9 aspect ratio to show the videos to the viewers. We inserted a 10-second grey field between test sequences to allow the viewers’ eyes to rest and also to give them enough time to rate perceived sharpness, depth and quality of the videos. The viewers’ distance from the display was 4 times the height of the display. The room, in which we conducted the tests, was consistent with the ITU-R recommendation 500. The duration of the test was approximately 12 minutes for each participant.

We asked the viewers to rate the sharpness, overall quality and depth perception for the stereoscopic video sequences, whereas for non-stereo videos, they were asked to rate only the quality and sharpness on a vertical rating scale.

The actual scale used in the tests had five equal-length labels: Excellent, Good, Fair, Poor and Bad. We used a linear transformation from the scale to numbers between 0 and 100. These numbers were used to calculate the average ratings over the viewers for each video sequence. The ratings were obtained using the double-stimulus continuous-quality method described in ITU-R recommendation 500. The original video was shown to viewers prior to each modified video.

Figure 5.4(a) shows the result of our subjective test for the “Mother and Kid” video sequence (video 1). The vertical axis shows the averaged ratings of both the stereo and non-stereo video sequences. The horizontal axis shows the sigma value of the Gaussian filter applied to the sequences. Sharpness and quality are shown in the top and bottom plots, respectively, while depth perception for the stereo video sequence is shown in the middle figures. Viewers’ evaluations of the non-stereoscopic (2D) videos provide an indication of how strong the filtering is and how it affects the quality and sharpness of the non-stereo tested content.

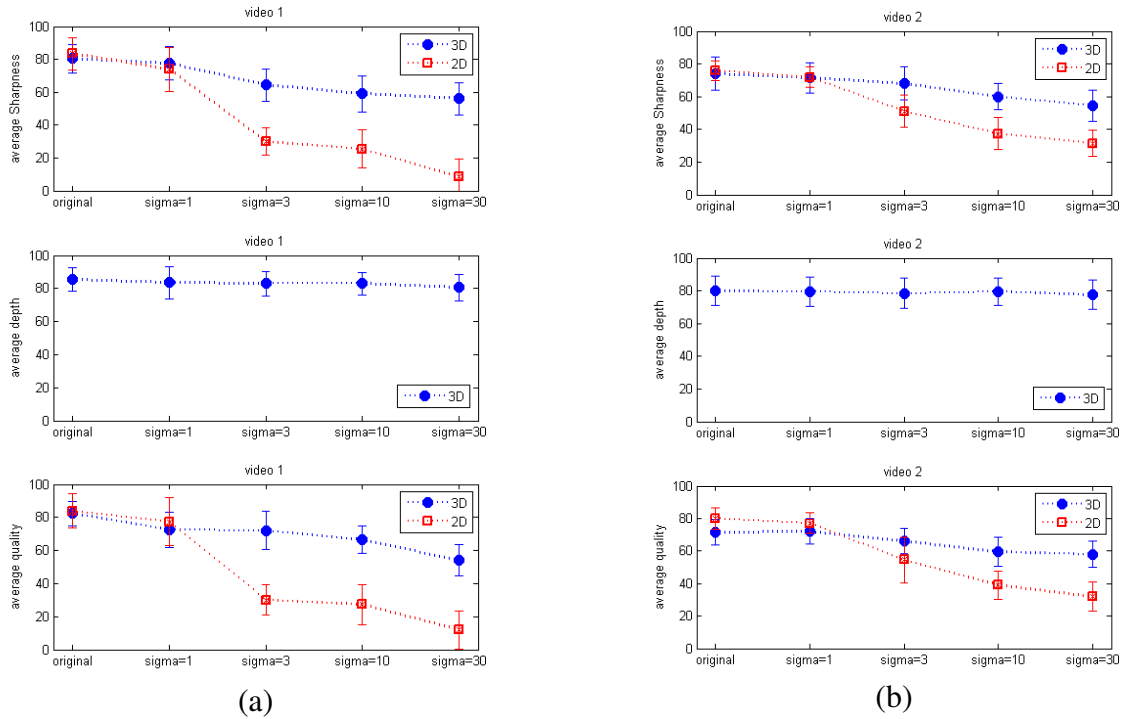


Figure 5.4 Sharpness, depth and quality of (a) video 1 and (b) video 2 averaged over the viewers.

An overall observation for video 1 is that the quality and sharpness of the filtered stereo are much better than those of the 2D corresponding videos. This is because the high quality slices in one view mask the blur in the low-pass filtered slices in other view in the case of stereoscopic videos. This does not apply to monoscopic videos since there is only one view.

Figure 5.4(a) also implies that the quality and sharpness of the filtered stereo video are rated close to those of the original video, up to a certain threshold in filtering strength. These results indicate that we can low-pass filter the alternate slices of both views without significantly reducing the overall perceived quality of stereo pair. In this case, we may conclude that a Gaussian filter of 15×15 size, and $\sigma = 3$ is a safe bound up to which most people cannot perceive the quality degradation. Beyond this point, degradation in the perceived quality and sharpness is observed.

Figure 5.4(b) shows the results for “Two Dolls”. Similar to the case of the “Mother and Kid”, the sharpness and quality of the low-pass filtered stereo videos are rated higher than those of the filtered non-stereo counterparts. Additionally, sharpness and quality of filtered stereo pairs are rated close to the original stereo, up to filter strength of $\sigma = 3$. The sharpness and quality seem to be just a bit higher than the one for the “Mother and Kid.” This may be due to the less relevant details (faces and textures) that this video has compared to “Mother and Kid.”

We also notice from the depth plots (Figure 5.4(a) and Figure 5.4(b)) that low-pass filtering does not affect the perceived depth in stereo pairs and it has remained unchanged for all applied filtering levels.

Our next test was designed to determine whether better ratings can be achieved for quality and sharpness of videos by gradually smoothing the slice borders. To achieve this, we applied strong filtering in the middle of each slice and reduced the amount of filtering as we got closer to the borders. Figure 5.5(a) shows the results of our subjective tests on sharpness, depth and video quality for Mother and Kid with smooth borders and compares them to those with the original filter (unsmoothed borders). We observe that smoothing the slice borders results in slightly better stereo video quality and sharpness. Figure 5.5(b) shows the results for Two Dolls. For this sequence as well, our subjects rated the sharpness and quality of videos with smoothed borders slightly better than those without smoothing.

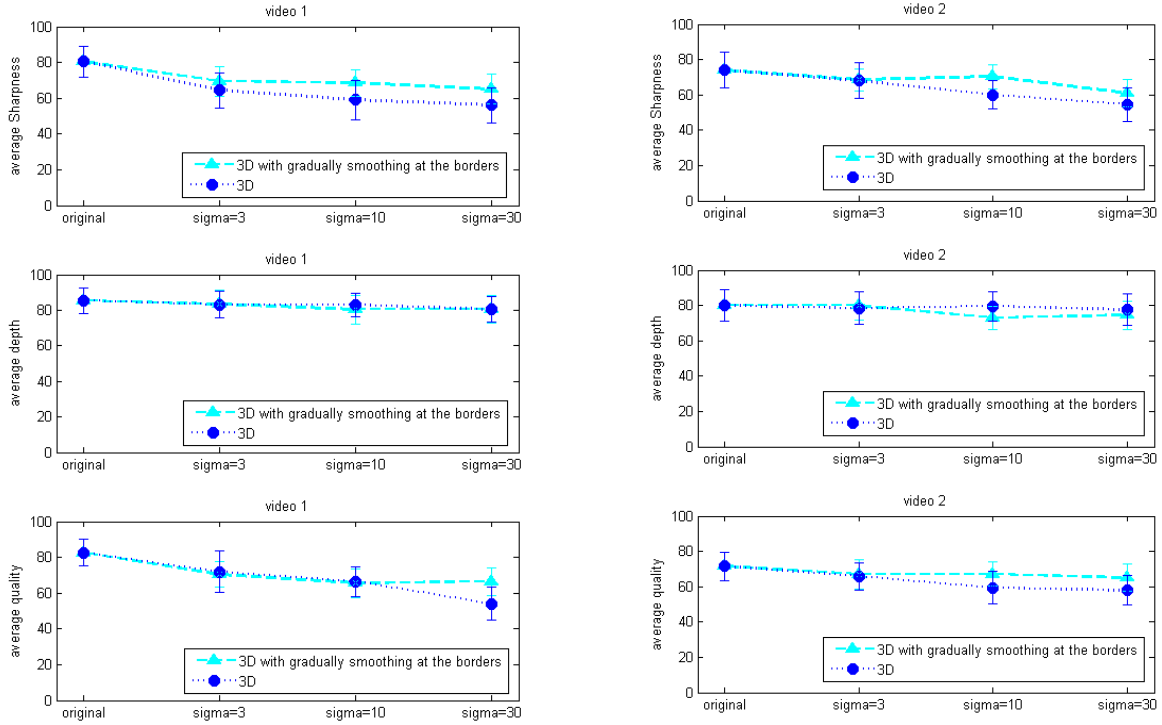


Figure 5.5 Sharpness, depth and quality of (a) video 1 and (b) video 2 averaged over the viewers. Smoothing the borders of the slices provide better quality and sharpness.

Next, we examine the bitrate reduction in our asymmetric stereoscopic videos. The resulting video sequences are compressed with H.264 encoder and the total bitrate of the low-pass filtered sequences is compared with the bitrate of the original video. Table 5.2 shows the compression results for the two video sequences. As we can see from this table we can achieve reduced bitrate by applying low-pass filtering to horizontal slices in the stereoscopic videos.

	original	Sigma=1	Sigma=3	Sigma=10	Sigma=30
Video1	2393	1985	1598	1523	1525
Video2	1238	1137	1004	972	970

Table 5.2 Bitrate in kbps for the two video sequences

Figure 5.6 shows the results of our subjective test for video1. The vertical axis shows the averaged rating for the quality of the asymmetric stereoscopic video. The horizontal axis shows the bitrate required for its transmission after compressing with H.264 encoder.

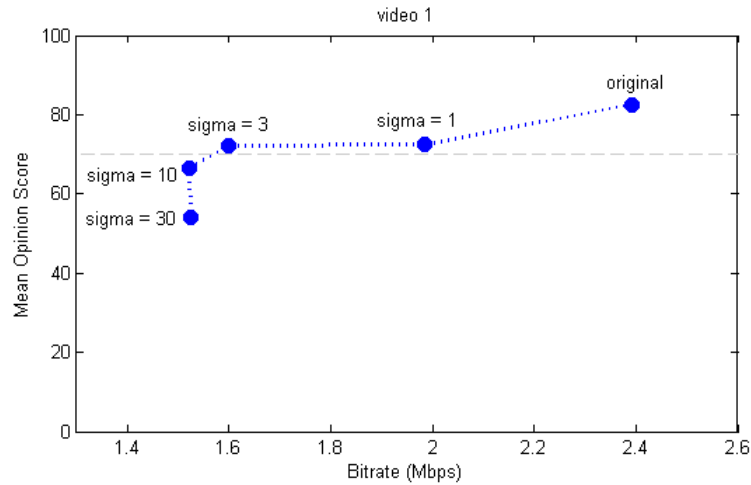


Figure 5.6 Mean Opinion Score of the first stereoscopic video sequence versus the required bitrate with H.264 compression

For video1, the original video requires almost 2.4 Mbps. As we increase the filtering strength, the required bitrate decreases at the cost of reduced quality. This figure shows that we can achieve bitrate saving and still have the video quality close to the original video. Furthermore, we can see that there is a threshold beyond which increasing the filtering strength does not result in reduced bitrate. As we can see from Figure 5.6, if we increase the filtering strength from sigma 10 to sigma 30, it will not reduce the bitrate. Results for the second video are shown in Figure 5.7.

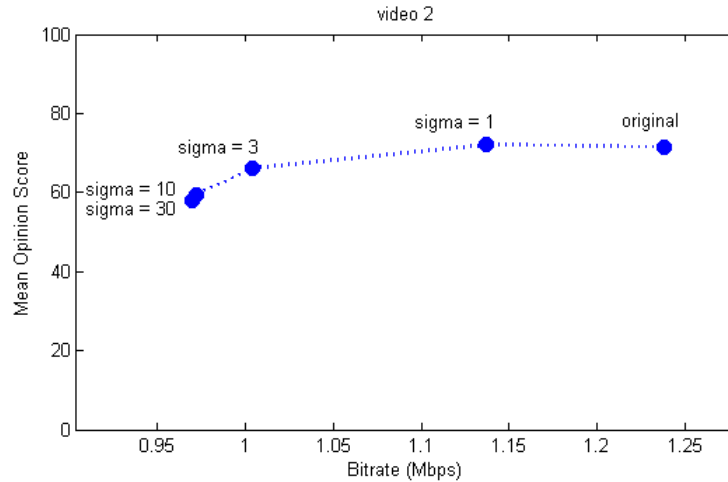


Figure 5.7 Mean Opinion Score of the second stereoscopic video sequence versus the required bitrate with H.264 compression

Low-pass filtering reduces the required bitrate while the overall quality remains close to the original video. This is based on the fact that human visual system perceives high quality 3D even if only one of the views is of high quality. Sharp edges in the high quality image mask the blur in the low quality view and overall impression is close to the sharper view.

5.3 Conclusions

We evaluated the compression performance of a modified scheme for asymmetric coding of the stereoscopic video. In our implementation, the videos are divided into horizontal slices in both the left and right views. Half of these slices are low-pass filtered while the corresponding slices in the other view have the original quality. We tested the perceived sharpness, quality and depth of the video sequences subjectively. Subjective evaluation demonstrated that sharpness and quality of our modified asymmetric videos is close to those of the original stereoscopic videos up to a filtering strength threshold (Gaussian 15×15 with $\sigma=3$) while the same amount of filtering was quite apparent in the monoscopic videos. Our implementation of asymmetric video

coding has the advantage over the conventional asymmetric methods of being distributed over the two views. We measured the bitrate reduction in this asymmetric stereoscopic video coding. Performance evaluations show that a significant bitrate reduction can be achieved.

Chapter 6: Conclusions and Future Work

6.1 Significance and Potential Application of the Research

High efficiency video coding (HEVC) is the most recent video coding standard developed by collaboration between Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG). Our proposed approach for improving the compression efficiency of HEVC can see mass market adoption since consumer playback devices with the standard HEVC video decoder are capable of receiving and displaying video data encoded with our proposed approach. Our approach modifies only the encoder side. Standard TV sets and setup boxes are capable of receiving and displaying video compressed with our proposed approach.

Additionally, achieving higher compression efficiency is crucial for limited capacity communication channels and storage media in consumer electronics. Also, offering new services such as higher resolution video, more sophisticated multimedia applications and ultra-high definition television depend on higher video compression efficiency. The importance of achieving higher compression efficiency is highlighted by considering that video streaming has grown significantly in recent years and has rapidly become the largest consumers of network capacity in both fixed and mobile networks.

Our novel approach looks at the coding unit structure and mode selection process in HEVC. The concept of coding units is a new concept introduced in HEVC and was not present in the previous video coding standards. Introduction of the coding units in HEVC leads to higher flexibility of the encoder for various video resolution. HEVC achieves higher compression efficiency due to the increased flexibility and introduction of new tools. Our research investigates this new concept in the latest video coding standard.

In recent years, 3D video coding format has seen a rapid development from the industry leaders and research communities. 3D video can provide a more immersive and realistic visual experience. For stereoscopic video, two views for the left and right eye need to be sent to the receiver. Without efficient 3D compression techniques, the amount of information in stereoscopic video is double that of a monoscopic video. 3D-HEVC introduced new tools in coding two views of the stereoscopic video to make it more efficient. Our perceptual approach makes 3D-HEVC more efficient for coding stereoscopic videos.

Auto-stereoscopic displays allow more views to be displayed to the viewer, giving the depth perception present in 3D TVs and at the same time, enabling the viewer to see the scene from different angles. Technology present in the auto-stereoscopic displays removes the need for wearing 3D glasses making it more convenient for the consumer to view 3D. For auto-stereoscopic displays, two or three views plus their corresponding depth maps should be compressed. Coding depth maps requires different tools than texture views due to their different characteristics. Our proposed approach improves the compression efficiency of 3D-HEVC for auto-stereoscopic displays.

6.2 Summary of Contributions

This thesis includes extensive investigations for increasing compression efficiency of 2D and 3D video coding in the following avenues: 1) Perceptual coding of the new video coding standard, HEVC, in order to increase its compression efficiency. 2) Extension of the proposed perceptual approach and making its rate and distortion trade-off adaptive to the input video sequences. 3) Extension of the perceptual coding to 3D-HEVC. 4) Asymmetric coding of stereoscopic video coding with the goal of reaching higher compression efficiency.

We integrate the PSNR-HVS perceptual video quality metric in the rate distortion optimization process of HEVC. We investigate the newly added concept of coding units in HEVC and modify the coding unit structure and mode selection in our proposed approach. Our proposed approach improves the compression efficiency of HEVC by 10.21% while maintaining compatibility with the standard HEVC decoders for various video sizes from mobile size videos up to HD and 4K videos.

We find the proper Lagrangian multiplier for our proposed approach. Lagrangian multiplier balances the trade-off between rate and distortion in the video coding process. Its selection is critical for any video encoder. We first find the best coefficient for the Lagrangian multiplier for our proposed approach. Then, we find the optimal Lagrangian multiplier based on different quantization parameters for our proposed approach to reach the highest amount of compression efficiency.

We present a modified version of our perceptual video coding for adapting the Lagrangian multiplier to the input video sequence. We fit a rate distortion model to the data from first frame of the input video sequence and estimate the content-adaptive Lagrangian multiplier based on this model. Our content-adaptive Lagrangian multiplier adjustment further improves the compression efficiency of our perceptual proposed approach (up to 2.62% with an average of 0.60%).

We propose the extension of our perceptual video coding approach for coding stereoscopic and multi-view videos. Unlike previous methods that looked at integration of a perceptual video quality metric in the mode selection for fixed size macroblocks or for motion estimation, our method integrates the perceptual video quality metric in the mode selection for

variable size coding units. Our proposed perceptual 3D video coding increases the compression efficiency of multi-view video coding by 2.78%.

We evaluate the compression performance of a novel method for asymmetric coding of stereoscopic videos. This method tries to distribute the lower quality in both views spatially by applying low-pass filtering to slices in both views. Our evaluations showed that this asymmetric approach reduces the amount of bitrate required for transmission and storage of stereoscopic video while maintaining the stereoscopic quality at high levels.

6.3 Directions of Future Work

One direction for future research would be to improve the compression efficiency of video coding for high dynamic range (HDR) videos. The goal of the HDR technology is to capture, distribute and display a wider range of luminance and color values compared to the standard dynamic range (SDR) videos. In fact, HDR captures and displays 6 orders of luminance, the same magnitude as the one perceived by the human visual system. Existing video compression standards are capable of compressing HDR video content but their performance is optimized for SDR content [110]. Integration of a perceptual video quality metric (such as PSNR-HVS) in the rate distortion optimization process of HEVC can potentially improve its compression efficiency for HDR content as well. For such integration, a new optimal Lagrangian multiplier should be found based on the HDR video content.

Development of highly efficient video compression techniques is a continuous effort from the international standardization body. MPEG has already received expression of interest for improving compression efficiency of the video coding techniques beyond HEVC in various existing and emerging application areas [111]. Investigations towards the next generation of

video compression standards have begun. The Joint Video Exploration Team (JVET) is a joint collaboration effort to evaluate compression technology designs proposed by the experts in this area. In these explorations, the coding structure of HEVC is kept unchanged. However, various HEVC design elements are modified [112]. For instance, larger coding tree blocks and larger transform units are considered in the future video coding; Quad-tree plus binary tree (QTBT) block structure is suggested. Also, intra and inter prediction improvements have been investigated. All these modifications change the rate distortion optimization process of the new video coding. Therefore, a new Lagrangian multiplier should be investigated for the integration of any perceptual quality metric (such as PSNR-HVS) inside the rate distortion process of the future video coding standard.

One of the emerging technologies for digital video is in augmented and virtual reality platforms [113]. In virtual reality applications, a scene is captured from different angles. The content from each camera can be stitched together before coding to make one single scene prior to encoding. Alternatively, different views can be coded separately but the redundancies between them should be removed. In some use cases, the capturing device is not provided with a feedback channel and does not know which perspective the end user will take. Therefore, efficient compression techniques are extremely important due to high quantity of captured video data from different perspectives. Integration of a perceptual video quality metric (such as PSNR-HVS) can improve the compression efficiency for virtual reality use cases and needs to be explored.

References

- [1] Rec, "H. 265 and ISO/IEC 23008-2: High efficiency video coding," ITU-T and ISO/IEC JTC, vol. 1, 2013.
- [2] VCEG, "Joint Call for Proposals on Video Compression Technology," VCEGAM91 and MPEG N, vol. 11113, 2010.
- [3] T. Wiegand, J. Ohm, G. J. Sullivan, W. Han, R. Joshi, T. K. Tan and K. Ugur, "Special section on the joint call for proposals on High Efficiency Video Coding (HEVC) standardization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 1661-1666, 2010.
- [4] Call for Evidence on High-Performance Video Coding (HVC), MPEG document N10553, ISO/IEC JTC 1/SC 29/WG 11, Apr. 2009.
- [5] Draft, "recommendation and final draft international standard of joint video specification (ITU-T Rec. H. 264| ISO/IEC 14496-10 AVC)," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, vol. 33, 2003.
- [6] G. J. Sullivan and T. Wiegand, "Video compression—from concepts to the H. 264/AVC standard," *Proc IEEE*, vol. 93, pp. 18-31, 2005.
- [7] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H. 264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, 2003.
- [8] I. E. Richardson, *The H. 264 Advanced Video Compression Standard*. John Wiley & Sons, 2011.
- [9] G. J. Sullivan, J. Ohm, Woo-Jin Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649-1668, 2012.
- [10] J. Ohm, G. J. Sullivan, H. Schwarz, Thiow Keng Tan and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1669-1684, 2012.
- [11] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J. Ohm and G. J. Sullivan, "Video quality evaluation methodology and verification testing of HEVC compression performance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 76-90, 2016.
- [12] V. Sze, M. Budagavi, G. Sullivan, *High Efficiency Video Coding (HEVC)*, Springer, 2014.

- [13] I. Kim, J. Min, T. Lee, W. Han and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1697-1706, 2012.
- [14] F. Bossen, "HM 9 common test conditions and software reference configurations," Joint Collaborative Team on Video Coding (JCT-VC), Document JCTVC-J1100, Stockholm, July 2012.
- [15] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74-90, 1998.
- [16] H. Everett III, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Oper. Res.*, vol. 11, pp. 399-417, 1963.
- [17] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Magazine.*, vol. 15, pp. 23-50, 1998.
- [18] K. McCann, B. Bross, W. Han, I. Kim, K. Sugimoto, G. Sullivan, High Efficiency Video Coding (HEVC) Test Model 13 (HM 13) Encoder Description, Joint Collaborative Team on Video Coding (JCT-VC), Document JCTVC-O1002, Geneva, Oct. 2013.
- [19] McCann K, Bross B, HanWJ, Kim IK, Sugimoto K, Sullivan GJ (2013), High Efficiency Video Coding (HEVC) Test Model 13 (HM 13) Encoder Description, Joint Collaborative Team on Video Coding (JCT-VC), Document JCTVC-O1002, Geneva, Oct. 2013.
- [20] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, pp. 800-801, 2008.
- [21] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication Systems*, vol. 49, pp. 35-48, 2012.
- [22] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, 26(1), pp.98-117, 2009.
- [23] S. Chikkerur, V. Sundaram, M. Reisslein and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, pp. 165-182, 2011.
- [24] F. Lukas and Z. L. Budrikis, "Picture Quality Prediction Based on a Visual Model," *IEEE Transactions on Communications*, vol. 30, pp. 1679-1692, 1982.
- [25] Van den Branden Lambrecht, Christian J and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," in *Electronic Imaging: Science & Technology*, 1996, pp. 450-461.

- [26] A. B. Watson, J. Hu and J. F. McGowan, "Digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, pp. 20-29, 2001.
- [27] F. Xiao, "DCT-based video quality evaluation," Final Project for EE392J, vol. 769, 2000.
- [28] C. Lee and O. Kwon, "Objective measurements of video quality using the wavelet transform," *Optical Engineering*, vol. 42, pp. 265-272, 2003.
- [29] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Transactions on Image Processing*, vol. 19, pp. 335-350, 2010.
- [30] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. New full-reference quality metrics based on HVS. In *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, Arizona, USA, 2006.
- [31] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. De Caluwe, S. Kohler, R. Koenen, S. Rihs, M. Ehrsam and D. Schlauss, "PVQM—A perceptual video quality measure," *Signal Process Image Commun*, vol. 17, pp. 781-798, 2002.
- [32] Z. Lu, W. Lin, E. Ong, X. Yang and S. Yao, "PSQM-based RR and NR video quality metrics," in *Visual Communications and Image Processing 2003*, 2003, pp. 633-640.
- [33] E. Ong, X. Yang, W. Lin, Z. Lu and S. Yao, "Video quality metric for low bitrate compressed videos," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*, 2004, pp. 3531-3534.
- [34] E. Ong, W. Lin, Z. Lu and S. Yao, "Colour perceptual video quality metric," in *IEEE International Conference on Image Processing 2005*, 2005, pp. III-1172-5.
- [35] P. Ndjiki-Nya, M. Barrado and T. Wiegand, "Efficient full-reference assessment of image and video quality," in *2007 IEEE International Conference on Image Processing*, 2007, pp. II-125-II-128.
- [36] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, pp. 2284-2298, 2007.
- [37] I. T. S. Sector, "Objective Perceptual Multimedia Video Quality Measurement in the Presence of a Full Reference," *ITU-T Recommendation J*, vol. 247, 2008.
- [38] 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*, vol. 13, pp. 600-612, 2004.
- [39] Z. Wang, L. Lu and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process: Image Communication*, vol. 19, pp. 121-132, 2004.

- [40] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2003, pp. 1398-1402.
- [41] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing.*, vol. 15, pp. 430-444, 2006.
- [42] L. Lu, Z. Wang, A. C. Bovik and J. Kouloheris, "Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video," in *IEEE International Conference on Multimedia and Expo*, 2002. ICME'02, pp. 61-64.
- [43] P. Tao and A. M. Eskicioglu, "Video quality assesment using M-SVD," in *Electronic Imaging 2007*, pp. 649408-649408-10.
- [44] M. Lambooi, W. IJsselsteijn, D. G. Bouwhuis, and I. Heynderickx, "Evaluation of stereoscopic images: Beyond 2D quality," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 432-444, Jun. 2011.
- [45] A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, and G. B. Akar, "Towards compound stereo-video quality metric: A specific encoder-based framework," *Proc. IEEE Southwest Symp. Image Anal. Interpretation Denver, CO, USA*, 2006, pp. 218-222.
- [46] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP J. Image Video Process.*, vol. 2008, Jan. 2009, Art. ID 659024.
- [47] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 41-54, 2006.
- [48] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," in *Proceedings of the 7th European Conference on Computer Vision*, pp. 82-96, Copenhagen, Denmark, May 2002.
- [49] A. Banitalebi-Dehkordi, M. T. Pourazad and P. Nasiopoulos, "An efficient human visual system based quality metric for 3D video," *Multimedia Tools and Applications*, vol. 75, pp. 4187-4215, 2016..
- [50] J. You, L. Xing, A. Perkis, and X. Wang, "Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis," *Proc. Int. Workshop Video Process. Qual. Metrics Consum.Electr.*, Scottsdale, AZ, USA, 2010, pp. 61-66.
- [51] C. Hewage, S. T. Worrall, S. Dogan and A. Kondoz, "Prediction of stereoscopic video quality using objective quality models of 2-D video," *Electron. Lett.*, vol. 44, pp. 963-965, 2008.
- [52] S. Yasakethu, C. T. Hewage, W. A. C. Fernando and A. M. Kondoz, "Quality analysis for 3D video using 2D video quality models," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 1969-1976, 2008.

- [53] C. T. Hewage, S. T. Worrall, S. Dogan, S. Villette and A. M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, pp. 304-318, 2009.
- [54] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging MVC standard for 3D video services," *EURASIP Journal on Applied Signal Processing*, vol. 2009, no. 1, p. 786015, Jan. 2009.
- [55] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626-642, Apr. 2011.
- [56] ISO/IEC MPEG Video and Requirements Group, "Call for Proposals on 3D Video Coding Technology," document N12036, Geneva, Switzerland, Mar. 2011.
- [57] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro and Y. K. Wang, "Overview of the Multiview and 3D Extensions of High Efficiency Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 35-49, 2016.
- [58] 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 of Selected Topics in Signal Processing*, vol. 7, pp. 1001-1016, 2013.
- [59] H. Schwarz, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, D. Marpe, P. Merkle, K. Müller, H. Rhee, G. Tech, M. Winken and T. Wiegand, "3D video coding using advanced prediction, depth modeling, and encoder control methods," in *Picture Coding Symposium (PCS)*, pp. 1-4, 2012.
- [60] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle and F. H. Rhee, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing*, vol. 22, pp. 3366-3378, 2013.
- [61] "Test Model under Consideration for HEVC based 3D video coding," MPEG N12350, 2011.
- [62] H. Schwarz, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, D. Marpe, P. Merkle, K. Muller, H. Rhee, G. Tech, M. Winken, T. Wiegand , "3D video coding using advanced prediction, depth modeling, and encoder control methods," *Picture Coding Symposium (PCS)*, pp.1-4, 7-9 May 2012.
- [63] C. Bartnik et al., "HEVC Extension for Multiview Video Coding and Multiview Video plus Depth Coding", Document of ITU-T SG16/Q6 VCEG, VCEG-AR13, Feb. 2012
- [64] H. Schwarz and T. Wiegand, "Inter-View Prediction of Motion Data in Multiview Video Coding," *Picture Coding Symposium (PCS)*, pp. 101 - 104, 7-9 May 2012.

- [65] P. Merkle, C. Bartnik, K. Muller, D. Marpe, and T. Wiegand, "3D Video: Depth Coding Based on Inter-component Prediction of Block Partitions," Picture Coding Symposium (PCS), pp. 149 - 152, 7-9 May 2012.
- [66] M. Winken, H. Schwarz, and T. Wiegand, "Motion Vector Inheritance for High Efficiency 3D Video plus Depth Coding," Picture Coding Symposium (PCS), pp. 53 - 56, 7-9 May 2012.
- [67] E. G. Mora, J. Jung, M. Cagnazzo and B. Pesquet-Popescu, "Initialization, limitation, and predictive coding of the depth and texture quadtree in 3D-HEVC," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, pp. 1554-1565, 2014.
- [68] G. Tech, H. Schwarz, K. Muller, and T. Wiegand, "Synthesized View Distortion Based 3D Video Coding for Extrapolation and Intrappolation of Views," In the proceeding of IEEE International Conference on Multimedia and Expo (ICME), pp. 634 - 639 , 2012.
- [69] G. Tech, H. Schwarz, K. Muller, and T. Wiegand, "3D Video Coding using the Synthesized View Distortion Change," Picture Coding Symposium (PCS), pp. 25 - 28, 7-9 May 2012.
- [70] B. Julesz, "Foundations of Cyclopean Perception," Chicago, IL: Univ. Chicago Press, 1971.
- [71] M. G. Perkins, "Data compression of stereopairs," IEEE Transactions on Communication, vol. 40, pp. 684-696, Apr. 1992.
- [72] L. Stelmach, W. J. Tam, D. Meegan, A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution, " IEEE Transactions on Circuits and Systems for Video Technology, vol.10, no.2, pp.188-193, Mar 2000.
- [73] S. Feng, J. Gangyi, W. Xu, Y. Mei, C. Ken, "Stereoscopic video coding with asymmetric luminance and chrominance qualities," IEEE Transactions on Consumer Electronics, vol.56, no.4, pp.2460-2468, 2010.
- [74] A. K. Jain, "Perceived Blur in Stereoscopic Video: Experiments and Applications," 2014.
- [75] J. Wang, S. Wang and Z. Wang, "Asymmetrically Compressed Stereoscopic 3D Videos: Quality Assessment and Rate-Distortion Performance Evaluation," in IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1330-1343, March 2017.
- [76] M. Reiss, and G. Reiss, "Ocular Dominance: Some Family Data," Laterality Journal, Psychology Press, volume 2, issue 1, pages 7-16, 1997.
- [77] A. Kondo and D. Tasos, "3D future internet media," Springer Science & Business Media, 2013. (Ch 2.4.2.2)

- [78] S. Liu, F. Liu, J. Fan, and H. Xia, "Asymmetric stereoscopic video encoding algorithm based on subjective visual characteristic," in Proc. Int. Conf. Wireless Commun. Sig. Process., 2009, pp. 1–5.
- [79] W. J. Tam, L. B. Stelmach, and S. Subramaniam, "Stereoscopic video: Asymmetrical coding with temporal interleaving," *Stereoscopic Displays and Virtual Reality Systems VIII*, Vol. 4297, pp. 299-306, 2001.
- [80] A. K. Jain, C. Bal, A. Robinson, D. MacLeod, and T. Q. Nguyen, "Temporal aspects of binocular suppression in 3-D video," in Proc. 6th Int. Workshop Video Process. Quality Metrics Consum. Electron., 2012, pp. 93–98.
- [81] A. K. Jain, A. E. Robinson and T. Q. Nguyen, "Comparing Perceived Quality and Fatigue for Two Methods of Mixed Resolution Stereoscopic Coding," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 3, pp. 418-429, March 2014.
- [82] Z. Chen and C. Guillemot, "Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 806-819, 2010.
- [83] J. Chen, J. Zheng and Y. He, "Macroblock-Level Adaptive Frequency Weighting for Perceptual Video Coding," *IEEE Transactions on Consumer Electronics*, vol. 53, pp. 775-781, 2007.
- [84] Z. Mai, C. Yang, K. Kuang and L. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, vol. 2, Feb. 2006, pp. 913-916, 2006.
- [85] C. Yang, R. Leung, L. Po and Z. Mai, "An SSIM-optimal H.264/AVC inter frame encoder," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, 2009*, pp. 291-295.
- [86] C. Yang, H. Wang and L. Po, "Improved inter prediction based on structural similarity in H.264," in *Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on, 2007*, pp. 340-343.
- [87] Z. Mai, C. Yang, L. Po, and S. Xie, "A new rate-distortion optimization using structural information in H.264 I-frame encoder," in Proc. ACIVS 2005, pp. 435–441.
- [88] Z. Mai, C. Yang and S. Xie, "Improved best prediction mode(s) selection methods based on structural similarity in H.264 I-frame encoder," in Proc. IEEE Int. Conf. Sys. Man Cybern., May 2005, pp. 2673-2678 Vol. 3.
- [89] B. H. K. Aswathappa and K. R. Rao, "Rate-distortion optimization using structural information in H.264 strictly intra-frame encoder," in *System Theory (SSST), 2010 42nd Southeastern Symposium on, 2010*, pp. 367-370.

- [90] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, "Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization," *IEEE Transactions on Image Processing*, vol. 22, pp. 1418-1429, 2013.
- [91] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, "SSIM-Motivated Rate-Distortion Optimization for Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 516-529, 2012.
- [92] C. Yeo, H. Li Tan and Y. H. Tan, "On Rate Distortion Optimization Using SSIM," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, pp. 1170-1181, 2013.
- [93] G. Tech et al, "Final report on coding algorithms for mobile 3DTV," *Mobile 3DTV*, Technical Report D2.6, 2011.
- [94] A. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2012, pp. 497-502.
- [95] T. Zhao, K. Zeng, A. Rehman and Z. Wang, "On the use of SSIM in HEVC," in *Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 1107-1111.
- [96] G. Wallace, "The JPEG still picture compression standard," *Comm. of the ACM*, vol. 34, No.4, 1991.
- [97] D. W. Paglieroni, "Distance transforms: Properties and machine vision applications," *CVGIP: Graphical Models and Image Processing*, vol. 54, pp. 56-74, 1992.
- [98] J. L. Devore and N. R. Farnum, *Applied Statistics for Engineers and Scientists*. New York: Duxbury, 1999.
- [99] G. Bjontegaard, "Calculation of average PSNR difference between RD curves," in *Proc. 13th Meeting ITU-T Q.6/SG16 VCEG*, Austin, TX, Apr. 2001.
- [100] I. Recommendation, "500-11," "Methodology for the Subjective Assessment of the Quality of Television Pictures," Recommendation ITU-R BT. 500-11," *ITU Telecom. Standardization Sector of ITU*, 2002.
- [101] Y. Huang et al, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 1614-1624, 2010.
- [102] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, pp. 542-545, 2001.
- [103] K. Müller, P. Merkle, G. Tech and T. Wiegand, "3D video coding with depth modeling modes and view synthesis optimization," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012 Asia-Pacific, 2012, pp. 1-4.

- [104] Y. Chen, G. Tech, K. Wegner, and S. Yea, Test Model of 3D-HEVC and MV-HEVC, document JCT3V-J1003, Geneva, Switzerland, Feb. 2015.
- [105] P. Benzie et al., "A Survey of 3DTV Displays: Techniques and Technologies," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, pp. 1647-1658, Nov. 2007.
- [106] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, and A. M. Knodoz. "Prediction of stereoscopic video quality using objective quality models of 2-D video," Electr. Lett., vol. 44, no. 16, pp. 963-965, Jul. 2008.
- [107] 3D-HEVC Reference Software, HTM-10. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSsoftware/tags/HTM-10.0. (Retrieved at July 2014)
- [108] K. Müller and A. Vetro, Common Test Conditions of 3DV Core Experiments, document JCT3V-G1100, San Jose, CA, USA, Jan. 2014.
- [109] F. Lewandowski et al, "Methodology for 3D Video Subjective Quality Evaluation," International Journal of Electronics and Telecommunications, vol. 59, pp. 25-32, 2013.
- [110] R. Boitard et al, "Demystifying High-Dynamic-Range Technology: A new evolution in digital media." IEEE Consumer Electronics Magazine, vol. 4, pp. 72-86, 2015.
- [111] "Request for contributions on future video compression technology," ISO/IEC JTC1/SC29/WG11 CODING OF MOVING PICTURES AND AUDIO, ISO/IEC JTC1/SC29/WG11 N15273, Geneva, February 2015.
- [112] "Algorithm Description of Joint Exploration Test Model 1 (JEM 1)," ISO/IEC JTC1/SC29/WG11 CODING OF MOVING PICTURES AND AUDIO, N15790, Geneva, October 2015.
- [113] "Requirements for a Future Video Coding Standard v4," ISO/IEC JTC 1/SC 29/WG 11 CODING OF MOVING PICTURES AND AUDIO, ISO/IEC JTC1/SC29/WG11, N16359, Geneva, June 2016.