NTNU
Norwegian University of
Science and Technology

Quality Evaluations and algorithmic Improvement of the next Generation Video Coding - HEVC

Morten Flå

Master of Science in Electronics
Submission date: June 2011
Supervisor: Andrew Perkis, IET
Co-supervisor: Leif Einar Aune, Vizrt

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem description

High Efficiency Video Coding (HEVC) is a new video coding standard under development by MPEG and VCEQ [1]. The vision is to create a new standard that is able to provide the same quality as H.264/AVC at half the bit rate.

The task is to evaluate the emerging video coding standard and contribute to its development. The evaluation shall be done using sport content on mobile clients as the platform. Both subjective and objective quality assessment shall be performed.

Additionally, the deblocking filter adopted in the current HEVC test model should be further explored and improved.

Assignment given: 13. January 2011
Supervisor: Andrew Perkis, IET
Co-Supervisor: Leif Einar Aune, Vizrt

Preface

This thesis presents my work carried out in my final semester at the Master of Science program in Electronics, at the Department of Electronics and Telecommunications, NTNU. The purpose of this thesis was to investigate the performance of and contribute to the High Efficiency Video Coding standard.

I would like to thank my supervisor Professor Andrew Perkis at the Centre for Quantifiable Quality of Service in Communication Systems (Q2S), my co-supervisor Leif Einar Aune at Vizrt and doctoral Student Liyuan Xing for guidance and technical advice throughout this thesis. I would also like to thank all of those who participated in the subjective quality assessment, in addition to Ida Onshus and Øystein Auli for rigorously proofreading.

Trondheim, June 13, 2011.
Morten Flå

Abstract

The increased processing power and screen sizes of mobile devices has made it desirable to watch multimedia presentations on the go. On such devices the data network bandwidth is usually the limiting factor, which imposes a tradeoff between quality and resolution on the presented content. A new video compression system called High Efficiency Video Coding (HEVC) is currently under development. The vision of HEVC is to create a compression system that achieves the same quality at half the bit rate compared to the existing H.264/AVC standard [2].

The goal of this thesis is to investigate how HEVC performs compared to H.264/AVC using mobile platforms and sport content as the scenario. The subjective test was conducted on an Apple iPad. It indicated that HEVC has a clear gain in compression compared to H.264/AVC. On average at a resolution of 640x368, HEVC achieved a good quality rating at approximately 550 kilobit per second while H.264/AVC did almost reach this quality at 1000 kilobit per second. However, it was shown that subjective quality gain varied over content.

The objective measurements showed an overall reduction in bit rate of 32% for the luma component. However, the reduction of bit rate was highly variable over content and resolution. A high correlation between the subjective and objective measurements was found, which indicates that it was almost a linear relationship between the reported subjective and objective results.

In addition, a proposed deblocking filter was implemented. The filter applies a new filter function of the luma samples and performs line based filtering decision. On average the reduction in bit rate was reported to be 0.4%, with a maximum reduction of 0.8% for the luma component. The decoding time relative to the second version of the HEVC test model was reported to be 1.5% higher. This is most likely due to the line based filtering decision.

The general impression of HEVC is that it has the ability to reach the stated vision, and perhaps even surpass, when finalized.

Contents

List of Figures

List of Tables

Abbreviations

3G	Third generation mobile telecommunication
4G	Fourth generation mobile telecommunication
ALF	Adaptive loop filter
AMVP	Advanced Motion Vector Prediction
AVI	Audio Video Interleave (Multimedia container)
BD-rate	Bjøntegaard delta rate
BS	Boundary strength
CABAC	Context Adaptive Binary Arithmetic Coding
CAVLC	Context Adaptive Variable Length Coding
CE	Core experiment
CfE	Call for Evidence
CfP	Call for Proposals
CU	Coding unit
DCT	Discrete cosine transform
DSIS	Double-stimulus impairment scale
FFmpeg	Open source video toolkit
fps	Frames per second
GOP	Group of pictures
H.264/AVC	H.264 / MPEG-4 Advanced Video Coding
HD	High definition
HEVC	High Efficiency Video Coding
HM	High Efficiency Video Coding test Model

HSDPA	High Speed Data Access Network
ISO	International Organization for Standardization
ITU-T	International Telecommunication Union - Telecommunication standardization section
JCT-VC	Joint Collaborative Team on Video Coding
LCEC	Low Complexity Entropy Coding
LTE	Long Term Evolution mobile network technology
MOS	Mean opinion score
MPEG	Moving Picture Expert Group
MPEG-2	MPEG video coding standard
NTIA	National Telecommunications and Information Administration
PPMC	Pearson product-moment correlation coefficient
PSNR	Peak signal-to-noise ratio
PU	Prediction unit
Q2S	The Centre for Quantifiable Quality of Service in Communication Systems
QP	Quantizer parameter
SRCC	Spearman's rank correlation coefficient
TE	Tool experiment
TMuC	Test Model under Consideration
TU	Transform unit
VCEG	Video Coding Expert Group
VOD	Video-on-demand
x264	A practical implementation of H.264/AVC
YUV	Color space with separate luminance and chrominance

Chapter 1

Introduction

Video compression has been the corner stone for delivery of video content to the consumers for many decades. Ray Davis Kell described a novel form of video coding already in 1929 [4]. Since then the technology has evolved and the introduction of color TV and at later stages High Definition (HD) TV, has changed the way people consume visual information.

During the last decade, the consumer base has driven forth a new demand in how video content is delivered, namely video-on-demand (VOD). VOD systems enables users to watch video content such as TV shows, football matches and humoristic movie clips whenever wanted. One of the most successful providers in this business is YouTube, which enables users to share their content to over 100 millions of users [5]. Broadcasting networks have also started to deliver content over the Internet, such as NRK Nett-TV and TV2 Sumo [6][7]. These services are either free of charge or subscription based. VOD has for many years been limited to computers and set-top boxes because it requires a certain processing power and network bandwidth.

The recent development of smart telephones and tablet computers has pushed the limits of processing power and screen sizes. These technology advancements paves the way for new and exciting streaming applications directed to a mobile audience. However, such clients are often bandwidth limited because they use cellular or wireless data networks. This limitation sets a series of constraints on the content delivered, such as resolutions, frame rates and bit rates, to ensure a seamless viewing experience. The introduction of high speed wireless data networks, such as 3G, 4G and LTE, has made it easier for the content providers to deliver higher quality video, but due to the increased screen size and resolution, the content requires more bandwidth now than ever.

The introduction of modern tablet computers that supports HD resolutions up to 1280x720, challenges the limits of what is possible to achieve with the current compression systems and cellular networks. Currently, the state of the art

H.264/Advanced video coding (H.264/AVC) is widely adapted on handheld devices such as the Apple iPad [8][9].

In 2010 the development of a new video compression standard under the name of High Efficiency Video Coding (HEVC) was started. The stated vision by the Moving Picture Experts Group (MPEG) and the Video Coding Expert Group (VCEQ) is to create a new standard that is able to provide the same quality as H.264/AVC at half the bit rate [2]. MPEG anticipates that HD applications will be introduced to the mobile sector, and that HEVC should be ready to conform with such applications.

Streaming of live sports is a highly popular service, but also one of the most demanding type of content to compress due to rapid motion and scene changes. In these days the content is available through wireless connections to smart phones and tablet computers. Accordingly, this thesis presents an evaluation of the performance on the current state of HEVC using a mobile platform with sports as the targeted scenario.

In order to contribute to the development of HEVC, an in depth analysis of the deblocking filter is presented. When coding video for bandwidth limited systems, visual disturbances known as blocking artifacts can manifest themselves at low bit rates. A deblocking filter is proposed in order to reduce these artifacts.

1.1 Development of a new standard

Throughout the years the two major standardization working groups, namely MPEG under International Organization for Standardization (ISO) and VCEQ under International Telecommunication Union - Telecommunication standardization section (ITU-T), have been developing and standardizing video compression technologies. Although at some point they have been competing rivals, they have the last ten years started to see the profit from joint collaborations when developing new standards. The current standard H.264/AVC is now 8 years old and thus the time has come for a new video compression standard. The following elaboration sheds light on how the ongoing HEVC standardization initiative proceeds.

The first step in the standardization initiative is a Call for Evidence (CfE). This includes an investigation of whether there has been significant development of video compression technology since the last released standard. The conclusion of the CfE was that the technology had matured enough to reach a significantly higher efficiency compared to H.264/AVC and thus a joint Call for Proposals (CfP) was issued in January 2010 [10]. The purpose of this call is to gather and evaluate compression technologies from companies and organizations all around the world. This call was carefully designed with a set of coding conditions and test material ranging from low resolution up to ultra HD. Each respondent to the call had to provide the coded test material for all conditions and a full technical description of the compression scheme.

Further in the standardization process a code base was formed that included promising coding tools. Several tool experiment (TE) groups and ad hoc groups were also initiated in order to further investigate the proposed technology. Given the vast amount of proposals, these groups acts as a technical committee to evaluate the proposals. Other important work done by the groups is verification of the results by cross referencing of the proposals. During the following meetings, video coding technologies was continuously proposed and after the third meeting held in Guangzhou, the first draft of the HEVC test model was released.

After the first release of the official test model the TEs was renamed core experiments (CE). The CE is divided into several subgroups that investigate and handle proposals for a specific technology area such as motion compensation, in loop filtering and intra prediction to mention a few. The key technology area of each subgroup is subject to change from meeting to meeting. After each new version of the test model, a working draft is released. This draft aims to describe the decoding process for the current test model. The draft is to be considered as work in progress as it contains several editorial notes and missing descriptions.

The reader should be aware of that due to the continuously changing nature of the unfinished HEVC standard, the presented information may not reflect the final video coding standard.

1.2 Outline

This thesis is structured in the following matter:

- **Chapter 2** presents the theory applied in this thesis. A short overview of general video coding and H.264/AVC is given in addition to an overview of the coding tools in HEVC.
- **Chapter 3** contains a description of the subjective and objective quality measurements performed. Also, the proposed deblocking filter is described in detail.
- **Chapter 4 and 5** report the obtained results from the subjective and objective experiments conducted. Results obtained by the implemented deblocking filter are also given here. The results are further discussed and evaluated, and a general impression of HEVC is given.
- **Chapter 6** draws conclusions from the results and important aspects of this thesis.

Chapter 2

Theory

This chapter includes the theory of video compression systems used in this thesis. Section 2.1 gives an overview of generic video coding and blocking artifacts. A short summary of H.264/AVC is given in Section 2.2, while a complete overview of the High Efficiency Video Coding standard with particular emphasis on the deblocking filter is given in Section 2.3.

2.1 Hybrid video coding

Hybrid video coding is the most commonly adapted scheme of todays video compression systems. The notion *hybrid* comes from the fact that it combines still image coding and temporal exploitation in the same compression system. Figure 2.1 shows the layout of a generic hybrid coding scheme with motion estimation and compensation.

The encoding block generally consists of a transform that performs spatial decorrelation, followed by a quantization of the transform coefficients. Likewise, the decoder performs an inverse transform and quantization.

In general hybrid video coding it is common to denote three types of frames depending on the coding condition, namely I-frame, P-frame and B-frame.

I-frame is an intra coded frame. This frame is a full reference frame, which means that it does not rely on the other frames to be decoded. The purpose of the I-frame is to provide a reference frame for the predicted frames and also to create random access points in the stream. Due to the fact that this is a full reference frame it requires most bits of all the frame types.

P-frame is a predicted frame and thus it requires a previous reference frame to be decoded. A P-frame may contain image data and/or motion vectors displacements. It can also be used as a reference frame for prediction of frames.

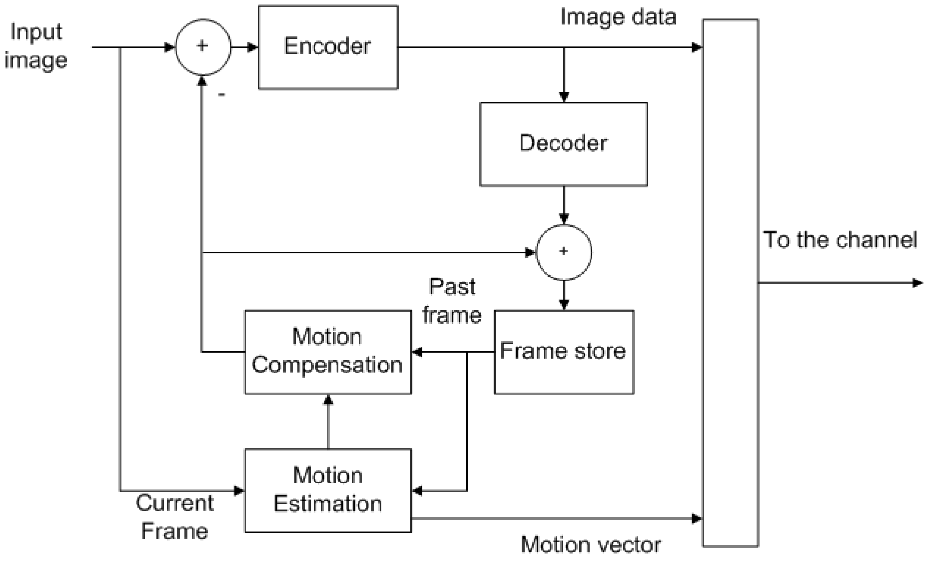


Figure 2.1: A generic hybrid coding scheme.

B-frame is also a predicted frame. However, this frame can be bi-predicted, which means that it can exploit temporal prediction from previous and future reference frames. In newer compression schemes it can also be used as a refrence frame. B-frames are known to use the least amount of bits of the three frame types.

The three frame types denotes the structure of a compressed video stream, which is referred to as a group of pictures (GOP). The GOP always starts with an I-frame and is often subsequently filled with P and B-frames. The GOP is often parametrized with an I-frame period that defines the amount of frames between each I-frame and effectively how far a prediction error can propagate throughout the GOP and the amount of B-frames between each P-frame. As an example GOP structures can be IBBP, IPPP or only I-frames depending on the application.

2.1.1 Blocking artifacts

Many video coding schemes rely on block based coding using transforms such as the discrete cosine transform (DCT). The frames are fragmented into blocks which in turn are transformed and quantized separately. This causes discontinuities at the boundaries between adjacent blocks. In addition, prediction of motion compensated blocks may contribute to further enhancement of these discontinuities because the motion vectors may come from several reference frames. These vectors seldom provides a perfect match and can propagate the discontinuities from the reference frame throughout the GOP. These effects are commonly known as blocking artifacts, which can be quite visible and may lower the experienced quality.

To counter blocking artifacts a so-called deblocking filter is deployed. The deblocking filter aims to smoothen the transition between two adjacent blocks, and thus reduce the visual disturbance created by the artifacts. Deblocking filters are mainly adapted as a in-loop filter or a post-filter. Post deblocking filters are employed at the output of the decoding loop. Such scheme gives maximum freedom and allows third party companies to implement their own deblocking algorithms. It is also possible to adapt the filters to special types of content and quality levels.

When adapted as an in-loop filter, the deblocking takes place both at the encoder and the decoder side. This has several benefits compared to the post-filtering process. Firstly, the presence of a deblocking filter on the encoder side smoothen block boundary transitions on the reference frames before they are used to predict motion. This increases the prediction quality. Secondly, the deblocking filter is standardized, which ensures a thoroughly tested and documented quality level of the filtering process. On the other hand, deblocking filtering increases the decoder complexity by a significant amount. It has been reported that the deblocking filter in H.264/AVC accounts for one-third of the computational complexity at the decoder [11].

2.2 H.264 / MPEG-4 Advanced Video Coding

H.264/AVC is the latest video coder from ITU and MPEG [12]. This compression standard is widely used in streaming applications and has also been adopted as the standard for coding video on Blu-ray. The H.264/AVC standard is a block based hybrid video coding approach with an arsenal of features to obtain better compression gain than its predecessors. A short summary of the feature highlights is given below.

- **Improved motion prediciton:** In contrast to previous standards, H.264/AVC has the ability to use multiple reference frames when predicting motion. In addition, different block sizes varying from 4×4 to 16×16 are used to obtain a higher precision of moving regions. The prediction is further refined with quarter pixel precision for luma and one-eight pixel precision for chroma.

- **Integer transform with small block sizes:** H.264/AVC is based on a 4×4 transform that makes the compression system highly adaptive in small regions within a frame. In addition the transform is an exact match integer transform, which means that the transform does not introduce any rounding errors when an inverse transform is applied. On high profiles the block size is adaptive between 4×4 and 8×8.

- **Adaptive in-loop deblocking filter:** An adaptive in-loop deblocking filter is deployed to reduce blocking artifacts. This has only been present as an optional feature in previous ITU/MPEG coding standards. The deblocking filter is applied in the encoder as well as in the decoder. The presence of a deblocking filter in the encoder loop increases the performance of motion estimation and predicition. The deblocking operation is performed on blocks of size 4×4.

- **Context adaptive entropy coding:** Two different entropy coding schemes are adopted in H.264/AVC, namely context adaptive binary arithmetic coding (CABAC) and context adaptive variable length coding (CAVLC). CAVLC is intended to be applied where low complexity is of importance, while CABAC has been found to achieve better compression with usage of arithmetic coding [13].

The reader is referred to [8] and [12] for more details on the H.264/AVC standard.

2.3 An overview of the emerging High Efficiency Video Coding standard

In January 2010 a joint call for proposals on video coding technology was issued by MPEG and VCEQ. A collaboration between these two groups was formed under the name Joint Collaborative Team on Video Coding (JCT-VC). A total of 27 proposals were submitted and extensive subjective testing was done to compare the proposals against H.264/AVC references. The results indicated that the best performing proposals had significantly improved compression gain compared to H.264/AVC, and that they were quite ahead on the subjective mean opinion score as well [14].

A closer look at the proposals revealed that all the algorithms were based on the traditional hybrid coding approach with variable block sizes, motion compensated prediction and estimation, quantization, in-loop filtering and advanced entropy coding. Key elements from seven of the best performing proposals were chosen to create a model for the new High Efficiency Video Coding development. This model was named Test Model under Consideration (TMuC) [1].

TMuC was further refined over the next JCT-VC meetings and in January 2011 an official test model named HEVC test model (HM) was publicly released. The difference from this model compared to TMuC was that it only contained a minimum set of well tested, good performing tools, and thus reduced the overall computational resources required for encoding and decoding. Although HEVC is built on the same basic notion as H.264/AVC (i.e hybrid video coding), it differs when it comes to adaptivity, flexibility and complex coding tools.

The general outline of HEVC is that there is a set of low complexity tools, and a set of high efficiency tools. The low complexity tools are aimed at low end platforms where processing power, memory bandwidth and power consumption are sparsely resources. The high efficiency tools offers a higher compression gain, but at a higher complexity. Although the two complexity levels share many of the same coding tools, some of them are mutual exclusive. In addition there is three coding cases, namely intra only, random access and low delay. High efficiency and low complexity configurations exists for all three of them [15].

The following sections provides an overview of HEVC based on the HM 2.0.

2.3.1 Coding, transform and prediction units

The first step in partitioning of a frame is so-called slicing. Each frame can consist of one or more slices that are coded independently. This allows for content adaptiveness already at the slice level by coding spatial regions differently. Each slice is further divided into treeblocks which plays a similar role as macroblocks known from older video coding schemes such as MPEG-2 and H.264/AVC [8]. A treeblock is always square and has a size of up to 64×64 luma samples.

The basic processing block in HEVC is called Coding Unit (CU). The CU is always square and can take a size from 8×8 to 64×64 luma samples. It can further be split into a quad-tree segmentation of regions, which allows for a highly content adaptive tree structure as shown in Figure 2.2. The current test model allows for a maximum CU depth of 4. It is important to notice that CU sizes are only defined as a maximum, and thus can take on several different sizes in the same picture. For example, if a depth of 4 is chosen and the maximum size is set to 64x64, then four CU sizes are possible: 64×64, 32×32, 16×16 and 8×8.

A CU consists of one or more prediction units (PU) that always are located in the leaf nodes of a CU. The PU contains all the data related to prediction such as motion vectors, intra prediction and frame reference indexes. A PU can have three different prediction types, namely skip, intra and inter. Intra and inter PUs are flagged to be intra and inter predicted, while skip mode can be seen as a special case of the inter mode where coding of motion vector differences are absent. A PU can be of size from 4×4 up to the size of the CU. The PU can have different rectangular shapes depending on which prediction mode that is used. The skip mode can only obtain a shape of $2N\times 2N$ and the intra mode may be of shapes $2N\times 2N$ or $N\times N$ where N is derived from the size of the CU. Finally, the inter prediction mode can use four different shapes, namely $2N\times 2N$, $N\times N$, $2N\times N$ and $N\times 2N$.

The CU also consists transform units (TU). TU is the basic building block for transformation and quantization. The size of a TU can be larger than PU, but not exceed the size of a CU, ranging from 4×4 up to 32×32. In the case of multiple TUs within a CU, they can be ordered in a similar quad-tree fashion as the CU.

2.3.2 Intra prediction

Intra prediction in HEVC is an extension of the intra prediction existing in H.264/AVC [8]. Earlier, large block sizes (16×16 and above) did not represent spatial homogeneous regions on low resolution content and thus intra prediction was generally

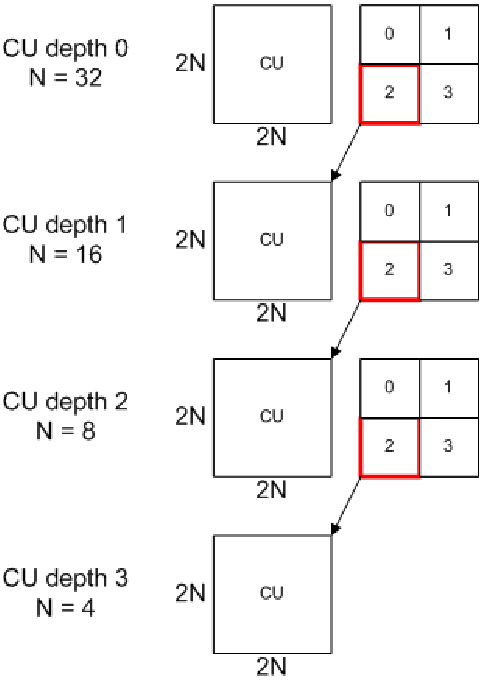


Figure 2.2: Quad-tree segmentation of coding units.

not performed. However, on HD resolutions, large block sizes up to 64×64 can represent homogeneous regions and patterns can be recognized and predicted. The prediction has up to 33 angular and one DC (flat average) prediction mode. The number of allowed prediction modes is regulated by the PU sizes as presented in Table 2.1. The 33 directions are presented in Figure 2.3. Regardless of the number of modes based on the PU size, the accuracy is 1/32th pixel using linear interpolation.

Table 2.1: Relationship between PU size and number of intra modes [3].

Size of PU	Number of intra modes
4×4	17
8×8	34
16×16	34
32×32	34
64×64	3

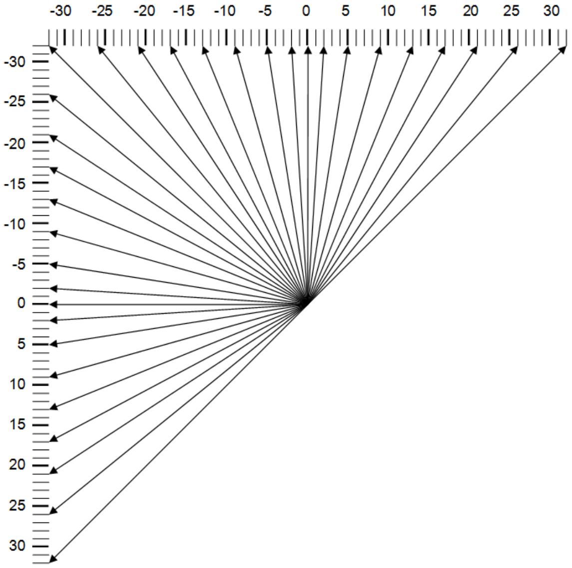


Figure 2.3: Intra prediction angles [3].

2.3.3 Motion estimation and prediction

The motion prediction engine in HEVC is called Advanced Motion Vector Prediction (AMVP). AMVP is based on a motion competition scheme that exploits both temporal and spatial motion vectors. By taking advantage of the fact that each PU contains a set of motion vectors and frame reference indexes, it is possible to further refine the prediction precision. AMVP makes it possible to exploit both temporal and spatial correlation by creating a candidate set of the best predictors from the PU neighbors. AMVP searches through the left and top candidates (A_i, B_i) in addition to temporally located PU to create a best possible candidate set. Figure 2.4 shows a sketch of the candidates. After the candidate set is created the encoder finds the best predictor for the specific PU and transmits it.

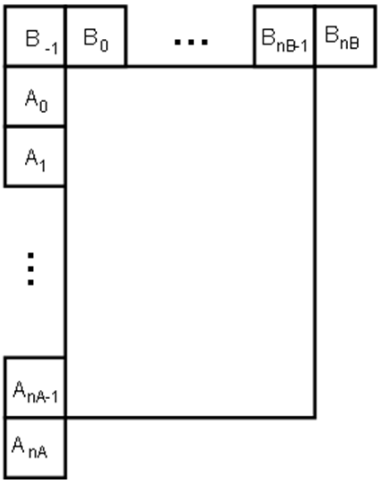


Figure 2.4: AMVP prediction neighbors [3].

In addition to the AMVP algorithm, a motion merge mode is adopted. The merge mode is similar to the AMVP concept, except that the motion parameters for the current PU can be inferred from motion parameters located inside neighboring PUs. In the case of motion merging the motion parameters are not explicit transmitted, but derived at the decoder based on the prediction mode and merge mode parameters. This can reduce the overall data needed to transmit motion parameters in a code stream. Motion merge shall be used when the PU is coded with skip mode type, but it can also be applied to inter mode PUs.

The precision of the motion vectors is up to quarter pixel for luma and one-eight pixel for chroma using a 8 tap and 4 tap DCT interpolation filter, respectively.

2.3.4 Transform and quantization

HEVC is able to perform transformation on sizes from 4×4 to 32×32. In the case of 4×4 and 8×8 the same integer based transformation as H.264/AVC uses is adopted (see Section 2.2). For transform sizes of 16×16 and 32×32 an integer based fast DCT algorithm with a butterfly structure is adopted. This transformation is based on Chen and Smith factorization [16].

The quantization of the transform coefficients is directly inherited from H.264/AVC [12]. This means that a scalar quantizer with a dead zone is deployed. Additional scaling matrices are added in order to support the larger transform size of 16x16 and 32x32.

2.3.5 Entropy coding

HEVC has two different entropy coding methods implemented. For the high efficiency configuration, CABAC is used while Low Complexity Entropy Coding (LCEC) is used for the low complexity configuration. The basic idea behind having two different entropy coders at different complexity levels is taken from H.264/AVC, with LCEC similar to the CAVLC [8]. These engines are further enhanced and optimized in order to support the larger block sizes.

2.3.6 Deblocking filter

The deblocking filter in HEVC is largely based on the deblocking filter present in H.264/AVC. Consequently it inherits many of its features. As mentioned in Section 2.1.1 the deblocking process was reported to be up to one-third of the computational complexity. This is mainly because deblocking in H.264/AVC i performed on a block size of 4×4. HEVC addresses this by defining the smallest block size value as 8×8 and thus the complexity is reduced substantially by decreasing the amount of samples loaded from memory. The choice of 8×8 deblocking goes hand in hand with new applications such as Ultra HD.

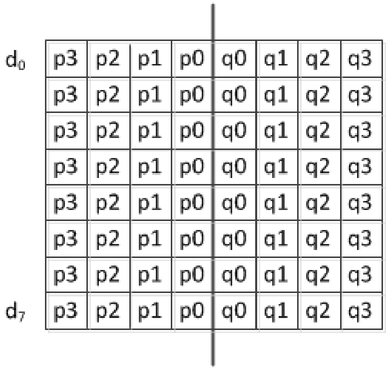
The deblocking is performed on a CU level. In each CU, vertical and horizontal edges of each PU and TU is deblocked. For each edge in a 8×8 block, a Boundary-Strength (BS) is calculated. The BS is parametrized by a value from 0 to 4, which indicates a certain edge condition. Table 2.2 shows the relationship between edge condition and BS.

Table 2.2: Relationship between different edge conditions and the Boundary-Strength.

Edge condition	BS
One of the blocks is intra and the edge is on a CU boundary	4
One of the blocks is intra and the edge is not on a CU boundary	3
One of the blocks has coded residuals	2
Difference in motion vector luma samples ≥ 1	1
Else	0

In order to further elaborate the deblocking filter, let p_3, p_2, p_1, p_0, q_0, q_1, q_2 and q_3 denote the sample values across a specific line d_i in a 8×8 block, where p and q is placed on each side of the edge boundary as presented in Figure 2.5.

$$BS > 0 \tag{2.1}$$



d₀	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
	p3	p2	p1	p0	q0	q1	q2	q3
d₇	p3	p2	p1	p0	q0	q1	q2	q3

Figure 2.5: Denotation of sample values across the lines in a 8x8 block.

$$d_{25} < \beta \tag{2.2}$$

$$d_{25} = \sum_{i=2,5} \Big(\Big| p_{2i} - 2 \cdot p_{1i} + p_{0i} \Big| + \Big| q_{2i} - 2 \cdot q_{1i} + q_{0i} \Big| \Big) \tag{2.3}$$

At the sample level the actual algorithm differs on luma and chroma. Filtering of luma samples within a block takes place if equations (2.1), (2.2) and (2.3) are fulfilled where β is a defined by a look up table and is dependent on the quantization parameter (QP) multiplied with the bit depth. The filtering is invoked on each line d_i where the samples on one line is either strong or weak filtered. If the variation of sample values inside the p and q block is sufficient small and the the difference between p_0 and q_0 is low, then a strong filter is applied on the line d_i. The strong filter alters all the sample values from p_2 to q_2 in order to create a smooth transition across the whole block.

In the case of weak filtering, a delta is computed by weighting the difference in the sample values according to equation (2.4) where tc is derived from a look up table and is based on the QP value multiplied with the bit depth. The function $Clip3$ ensures that the weighted sample values are clipped between $-tc$ and tc. The new sample values are calculated by adding or subtracting factors of Δ according to equations (2.5), (2.6), (2.7) and (2.8). $Clip$ ensures that the sample values p_0', q_0',p_1' and q_1' are clipped between zero and the maximum intensity specified by the bit depth of the frame (i.e 255 for 8 bit).

$$\Delta = Clip3\Bigg(-tc, tc, \Big(13 \cdot (q_0 - p_0) + 4 \cdot (q_1 - p_1) - 5 \cdot (q_2 - p_2) + 16) >> 5 \Big) \Bigg) \tag{2.4}$$

$$p_o' = Clip\Big(p_o + \Delta\Big) \tag{2.5}$$

$$q'_0 = Clip\Big(q_0 - \Delta\Big) \tag{2.6}$$

$$p'_1 = Clip\Big(p_1 + \frac{\Delta}{2}\Big) \tag{2.7}$$

$$q'_1 = Clip\Big(q_1 - \frac{\Delta}{2}\Big) \tag{2.8}$$

Chroma is only deblocked for a BS value greater than 2, which indicates that it is only deblocked in intra mode. The filtering of chroma samples is executed on each color component. Here, only p'_0 and q'_0 are filtered in the same way as the luma formula. However, a new delta is computed described in equation (2.9).

$$\Delta = Clip3\Bigg(-tc, tc, \Big(\big(((q_0 - p_0) << 2) + p_1 - q_1 + 4\big) >> 3\Big)\Bigg) \tag{2.9}$$

2.3.7 Adaptive loop filter

An adaptive loop filter (ALF) is employed after the deblocking filter in the coding sceheme. The purpose of the ALF is to reduce the the distortion between the original frame and the reconstructed frame caused by lossy compression. This is done by applying two-dimensional diamond shaped filters using a Wiener-filter approach, which is well known to reduce the mean square error.

The filtering of luma samples are done at the CU level. For luma samples in each CU the encoder decides whether or not the ALF should be applied based on a calculation of variance between the original and reconstructed frame. Three filters of sizes 3x3, 5x5 and 9x7 are implemented in HM 2.0 as depicted in Figure 2.6. In the case of chroma filtering the decision is performed on a frame basis instead of at each CU. Here, rectangular shaped filters are used and no variance calculation is performed.

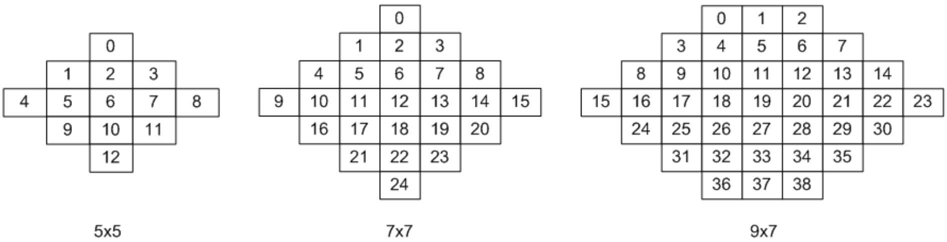


Figure 2.6: Two-dimensional filter shapes for luma samples.

The ALF improves the quality of the decoded frame in addition to improve the quality for further motion estimation and prediction for successive frames. Currently, this is one of the most complex, but at the same time efficient coding tools included in HM 2.0. With up to 4.5% bit rate reductions, it accounts for up to 38% of the decoding time in some specific coding cases [17].

2.3.8 Internal bit depth increase

Internal bit depth increase is a method used to increase the internal accuracy of the calculations involved when compressing a sequence. This technique can improve the overall compression efficiency by for example finding a better prediction match, higher transform precision and deblocking filter precision. A typical scenario would be to input a sequence with 8 bit sample accuracy and add 2 bits of additional precision. Each sample value is then multiplied by 4 before further calculations are done. At the decoder the sample values are then rounded down to 8 bit again and clipped between $[0, 255]$. A generalized view can be seen in Figure 2.7 where M is the original bit depth and N is the added precision.

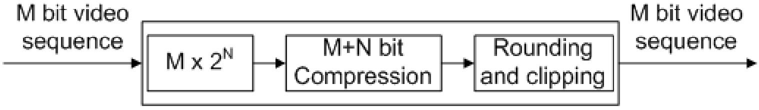


Figure 2.7: A general view on internal bit depth increase in a video compression scheme.

While this might be highly advantageous in terms of higher compression efficiency, it adds complexity to the compression chain. The majority of this complexity comes in the form of a larger memory footprint because the system has to store frames with higher precision. HM 2.0 supports increased precision from 0 to 4 bits. However, it has been shown that 80% - 90% of the potential gain is captured by incrementing the precision by 2 bits (i.e from 8 to 10 bits of accuracy) [18].

Chapter 3

Method

This chapter presents the methods applied in this thesis. Section 3.1 gives a background and scenario for the specific task at hand, while Section 3.2 provides information of the video content used in this thesis. The subjective assessment is described in detail in Section 3.3 and further analyzed in Section 3.4 and 3.5. The objective assessment is described in Section 3.6 and finally the proposed deblocking filter is elaborated in Section 3.7.

3.1 Scenario

The increased usage of smart phones and tablet computers has created a new way for consumers to experience digital content. These devices are no longer only limited to simple applications and mobile services, but can be used as fully featured media hubs. It is now possible to enjoy multimedia experiences at home or on the move.

A limiting factor, in addition to the device itself, is the amount of wireless data throughput it is possible to achieve. A direct effect of this is that it is crucial to find a good tradeoff between content resolution and coding bit rate to obtain the best possible quality. High Speed Data Packet Access (HSDPA) is a wireless technology supported by most of the new smart phones and tablet computers, which has been measured from 1000 to 2500 kilobit per second (kbps) in live networks [19][20]. The introduction of services such as streaming of TV channels and YouTube has set a new standard on how to deliver content to a handheld device using wireless data networks. Streaming of sports content is a highly popular service, but is among one of the most difficult types of media to compress.

The scenario is a set with a person watching sports content on a tablet computer. The video is compressed with bit rates common for wireless connections.

3.2 Dataset

Three video sequences has been chosen to represent the scenario at hand. The video sequences are picked in such a way that a variety of complexity ranging from moderate to difficult is represented. The video sequences are owned by National Telecommunications and Information Administration (NTIA), an American federal agency. They are however free to use for research purposes. The sequences consists of two American football sequences and one boxing training session sequence. They are named *Touchdown day*, *Touchdown pass* and *Heavy bag* respectively. All three sequences has an run time of 19 seconds. A brief description of the video sequences is found in Table 3.1.

Table 3.1: Brief description of the video test material.

Name	Resolution	Frame rate	Content type	Coding complexity
Heavy bag	1920x1080	30	Boxing	Difficult
Touchdown day	1920x1080	30	American football	Moderate/Difficult
Touchdown pass	1920x1080	30	American football	Moderate

Heavy bag is a sequence showing a boxing training session. The focus is on a boxer hitting a boxing bag. It contains very fast motion, close up footage and several scene changes, which makes this a very complex sequence to encode.

Touchdown day shows a touchdown in a football game. The view is narrow with a high motion tackle that involves several players. Due to the lack of scene changes, the coding complexity is rated between moderate and difficult.

Touchdown pass is a continuous sequence showing a football game that ends in a touchdown resulting from a pass. It is captured from above with an wide angle with focus on the ball and the player who possesses it. The video complexity is rated to moderate.

All three sequences were captured with Panasonic P2HD AJ-HPX3000G with a Fujinon HA22x7.8 BERM-M48 lens. This camera records in H.264/AVC intra-frame coding at 100 Megabit per second. The published video files from NTIA is stored in a progressive 4:2:2 UYVY 8 bit format packed into an AVI header with a resolution of 1920 by 1080 and a frame rate of 30 frames per second (fps) [21]. Regarding the file format, Y corresponds to the luminance (luma) component and the U and V corresponds to the chrominance (chroma) components. A frame from each of the sequences can be found in Figure B.1, B.2 and B.3.

3.2.1 Post processing

In order to convert the sequences to the appropriate resolution and make them compatible with the current HM a number of operations had to be done. The original content included an audio track, which had to be removed. The sequences were also packed into an unsupported AVI-format. The AVI header were removed and the video files were converted into a header-less 8 bit 4:2:2 UYVY format. It was also necessary to perform chroma subsampling in order to produce a HEVC compatible input format. The file format was converted from UYVY 4:2:2 to YUV 4:2:0. The reader is referred to [22] for more information about chroma subsampling.

In addition to the above changes a section of 10 seconds was extracted from each of the sequences, and rescaled to the appropriate resolutions. The popular open-source FFmpeg video toolkit was used to remove the audio and to convert the video files into a suitable formats, lengths and resolutions [23].

3.2.2 Resolutions and bit rates

In order to simulate content and screen sizes of smart phones and tablet computers three different resolutions were chosen. In addition to the various resolutions, each video sequence was encoded with six different bit rates. The bit rates are 150, 300, 400, 500, 750 and 1000 kbps and the resolutions are 640x368, 480x272 and 320x176. The resolutions combined with the appropriate bit rates sets a realistic video streaming scenario over cellular data network ranging from low end smart phones to high end tablet computers.

All sequences were coded with a maximum deviation from the specific target rate by $\pm 1\%$. The H.264/AVC encoder has built-in rate control mechanisms to specify the target rate, while HM 2.0 has not. In order to obtain the desired bit rate, quantizer values were carefully selected for a given sequence at a given resolution. The selection of HEVC quantizer values and the resulting bit rates are summarized in Table 3.2.

A closer look at Table 3.2 reveals the proclaimed complexity differences between the sequences. It is clear that *Heavy bag* needs higher quantization values than *Touchdown day* followed by *Touchdown pass* to meet the targeted bit rates.

3.2.3 Encoder settings

Two different encoders were used to compress the content described. To produce a H.264/AVC compliant bit stream, the open source encoder x264 was used [24]. x264 has proven to be among the best available practical software implementations of H.264/AVC and is considerable faster than the H.264/AVC reference model [25]. To properly verify the compliance of the compressed x264 bit stream it was decoded using the H.264/AVC reference software [26]. HM 2.0 was used for HEVC compression [27].

Both encoders use a GOP structure of IBBBP with hierarchical-B structure enabled and an intra period of 3 seconds (i.e 90 frames with a sequence of 30 frames per

Table 3.2: Summary of HEVC quantizer values and bit rates for sequences Heavy bag, Touchdown day and Touchdown pass. The bit rates are given in kbps.

(a) Quantizer values and the corresponding bit rates for Heavy bag and Touchdown day.

	Heavy bag			Touchdown day		
	640x368	480x272	320x176	640x368	480x272	320x176
Bit rate	Quantizer values			Quantizer values		
150	40	37.2	33.1	37.6	34.6	30.5
300	34.8	31.9	28	32.5	29.6	25.3
400	32.53	29.8	25.8	30.5	27.53	31.1
500	30.9	28.1	24.1	28.8	25.9	21.5
750	27.9	25.1	20.9	25.9	23	17.9
1000	25.9	22.9	18.2	23.9	20.8	15.3

(b) Quantizer values and the corresponding bit rates for Touchdown pass.

	Touchdown pass		
	640x368	480x272	320x176
Bit rate	Quantizer values		
150	36.8	34	29.7
300	32.1	29.1	24.7
400	30.15	27.15	22.6
500	28.5	25.6	21
750	26	22.8	17.4
1000	24	20.7	14.8

second). x264 offers adaptive B-frame and I-frame decision. However, this was disabled in order to keep the GOP structure static for both encoders. The number of reference frames used for motion compensation and estimation was 4. Since the content is shown on a high end tablet, x264 was allowed to use the integer 8x8 DCT transform. This corresponds to a high profile setting, which is normally not supported by a typical smart phone or a tablet computer. However, the first generation Apple iPad, which was used as the viewing equipment in the subjective test, had no problem decoding the resulting H.264/AVC stream. Both encoders had their respective in-loop filters enabled and CABAC was used as the entropy coding method. Detailed configurations can be found in Appendix D.

3.3 Subjective assessment

The only way to truly determine the quality of an image or a video sequence is to perform subjective assessment on the content in mind. In subjective assessments one gathers a group of people and asks them to the rate the quality of the presented material. This group should preferably consist of a high number of people, all of whom are assumed non-experts in video coding. The group should also have a vast difference in background, age and sex. While subjective assessment usually provide a good estimation of the quality at hand the test, itself can often be very troublesome and time consuming. Accordingly, rules and guidelines are laid down on how to conduct a proper subjective test [28]. The test is often aimed towards rating of video sequences on a certain platform, such as TV, computers or smart phones at different coding conditions. One wants to to determine the difference in quality based on human observation, by comparing compression algorithms or other factors such as encoder settings.

3.3.1 Test equipment

To create a realistic viewing scenario, the first generation Apple iPad was used to evaluate H.264/AVC and HEVC. The iPad bolsters a screen size of 9.7 inches with a native resolution of 1024x768 running on a 1GHz processor with 256 MB internal memory. It also has a built in H.264/AVC decoder capable of progressively scanned video up to 1280x720 in 30 fps (720p30). The specific model used in the subjective test had 64GB of storage, WiFi and 3G support, running on iOS v4.2.1. Since HEVC is still under development and no implementation of a decoder exist on the iPad, raw YUV streams were used as playback for both encoders.

In order to playback raw YUV streams, a third party application called *Oplayer HD* was installed. This application is available from the official application store [29]. The iPad together with Oplayer HD, is able to read and display YUV 4:2:0 chroma subsampled 640x368 video sequences at 30 fps. At 720p30 the playback was not working properly, most likely due to processing power and hard drive read speed. The raw YUV streams had to be packed into an AVI header that specified resolution, frame rate and color space before they were uploaded to the iPad. This was neccesary so that Oplayer HD could properly recognize the sequences. The brightness level of the iPad display was kept at the default level, and the display was wiped clean before each test session.

3.3.2 Room setup and environment

The test session was held at the Quantifiable Quality of Service (Q2S) media lab, Café Media. Each participant was led into a confined area of the media lab. This area was sealed off with dark blue drapes in order to provide minimal distractions for the participants during the subjective evaluation. The drapes also made it possible to keep the light intensity at the same level for all the participants. The participant were positioned at the far end of the area, keeping the same direction towards the light source. This was particular useful as the Apple iPad has a very glossy display, which easily creates a lot of reflections and poor viewing conditions. Due to the position, reflections were avoided.

3.3.3 Test methodology

The double-stimulus impairment scale (DSIS) method was used to assess the quality of the coded sequences[28]. The choice of test method was based on the wide spread in expected quality of the presented content. This method is based on pairs of stimuli, where an unimpaired sequence A is followed by an impaired sequence B. The participant is asked to evaluate sequence B in comparison to sequence A. The participant is aware about the reference sequence and that it represents the best expected quality. Each sequence last for 10 seconds and after each pair of stimuli the participant has 5 seconds to rate the quality before a new pair of stimuli is shown. The time pattern of the DSIS method adopted is presented in Figure 3.1.

Figure 3.1: Description of the subjective test time pattern.

A five grade quality scale with a corresponding score, as described in Table 3.3, was used to rate the material. In accordance to [28] the sequences and impairments were shown in a random order to avoid possible effects of the sequence order. The unimpaired sequences were also included in the subjective test to be quality rated.

Score	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Table 3.3: Five grade quality scale.

3.3.4 Test session

Each test session consisted of one participant due to only one available iPad. As presented in Figure 3.1, the total time of a single stimuli pair was 27 seconds. The participants were asked to evaluate 39 test conditions for each resolution (i.e 3 reference pairs, 3 different sequences, 6 bit rates, and two encoders). A total of 117 stimuli pairs were presented for 3 resolutions. This results in a total session time of

52 minutes and 39 seconds. However, exposing viewers to test sessions longer than 30 minutes may cause fatigue [28]. Due to the share amount of test conditions, resolutions and compression technologies it was decided to split the session into three parts, one for each resolution. This reduced the evaluation time to 17 minutes and 33 seconds. Between each resolution the participants were asked to take a short break. Each test session was randomized such that no participant had the same sequence order. The scoring sheet for the subjective quality assessment can be found in Figure A.1.

The video sequences were presented on the iPad in 16:9 full screen with the aspect ratio maintained. This was done to get a more realistic user scenario, whereas users tend to view video in full screen regardless of content resolution. This implies that the stimuli pairs were not shown in their native resolution, but rather in an upscaled way. The iPad handled the upscaling of the sequences. Due to the 16:9 aspect ratio, black borders were present at the top and bottom of the screen.

3.3.5 Participants

A total of 25 participants were used in the subjective video test. The group consisted mainly of students that were to be considered non-experts in terms of video coding. The age of the participants ranged from 19 to 29 years and 32% of the viewers were female. 28% of the participants used some forms of visual correction, either glasses or lenses.

Prior to each test session the participants were briefed about the goal, test methodology and the grading scale. A training session was done in order to familiarize the participant with what types of impairments that could be expected, and the assessment procedure. The training sequences were viewed together with a test supervisor which had rated the training sequences in advance. The training session lasted for approximately 5 minutes and the participants were allowed to ask questions during this time. The training instructions can be found in Figure A.2.

3.4 Statistical analysis of the subjective quality assessment

Statistical analysis is an import aspect when interpreting the outcome of a subjective quality assessment. At the first glance, the outcome of a subjective test is merely a mean opinion score with an associated confidence interval for each test condition. However, the test procedure can produce variations in quality ratings between each test subject. Each subject may react differently to the content or test conditions and people have a different subjective perception about what is good and bad quality. In addition to these variations the result can be biased by fatigue and random errors that are caused by external uncontrollable events.

The aim of the analysis is to understand whether or not the results are reliable and distinguishable from random ratings, and if a general conclusion can be drawn despite the low number of subjects.

3.4.1 Distribution analysis

The analysis of the distribution is important in order to use proper statistical methods on the raw data. Statistical methods are often limited by the way the data is distributed, and thus one want to check the normality of the data. There are two ways to analyze the distribution of the data. One can either look at the distribution for each subject across the conditions or all the distributions for each test condition across the subjects. In order to analyze the distribution of data a study of the Kurtosis coefficients, called the β_2 test was used. A full elaboration of method is presented in Section 3.4.3. The result of the test was that the distribution for each subject across the conditions was found not be normally distributed. However, the majority of the distributions for each test condition across the subjects (70%) had a normal or close to normal distribution. The following sections are justifed based on the assumption of normality.

3.4.2 Mean opinion score and confidence interval

A mean opinion score (MOS) is calculated by taking the average of the score across the subjects for each test condition. The MOS value was calculated as:

$$MOS_j = \frac{1}{N}\sum_{i=1}^{N} \mu_{ij} \tag{3.1}$$

where μ_{ij} is the score by subject i for test condition j and N is the number of subjects. To establish the relationship between the estimated mean and the true mean a corresponding confidence interval for each MOS value j was calculated using the two tailed Student's t-distribution. The Student's t-distribution is appropriate when then number of samples are low and independent [30, pp. 257-261].

$$CI_j = t_{(\alpha/2,N-1)}\frac{\sigma_j}{\sqrt{N}} \tag{3.2}$$

Here, α is the confidence level, $N-1$ is the degree of freedom and N is the total number of subjects. In this specific experiment the CI_j was computed with a 95% confidence interval ($\alpha = 0.05$) and 24 degrees of freedom. The associated standard deviation σ_j of each MOS value j is given by:

$$\sigma_j = \sqrt{\sum_{i=1}^{N} \frac{(MOS_j - \mu_{ij})^2}{N-1}} \tag{3.3}$$

3.4.3 Outliers detection and screening of the observers

Detection of outliers was done in accordance to [28, Annex 2, Section 2.3]. This is the recommended procedure for the DSIS method. First the distributions for each test conditions across the subjects were checked for normality by calculating the Kurtosis coefficient as described in equation (3.4). The distribution is normal if the coefficient is between 2 and 4.

$$\beta_j = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mu_{ij} - MOS_j)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(\mu_{ij} - MOS_j)^2\right)^2} \tag{3.4}$$

For each condition, the score of each subject was compared to the respective MOS value plus and minus the standard deviation of the MOS value times 2 or $\sqrt{20}$ for normal and non-normal distributions respectively. Each time a subject i is above the upper threshold, a counter P_i is incremented. Likewise, a counter Q_i is incremented when the subject i is below the lower threshold. An outlier is detected and removed if and only if the ratio $P_i + Q_i$ divided by total number of scores (i.e test conditions for this specific experiment) is greater than 5% and the absolute value of $P_i - Q_i$ divided by $P_i + Q_i$ is below 30%.

3.4.4 Comparison of the estimated mean opinion scores

In order to establish if the MOS values of each encoder pair at the same coding condition (i.e at the same bit rate, resolution and sequence) were significantly different, a paired t-test was used. Two hypotheses were created to see if there was a significant difference between the MOS values of HEVC and H.264/AVC for each test condition as stated below.

$$H_0 : MOS_{hevc} = MOS_{avc}$$
$$H_1 : MOS_{hevc} \neq MOS_{avc}$$

If H_0 is accepted it states that there is no significant difference between the MOS values for HEVC and H.264/AVC for a certain condition and thus we cannot conclude that one of the codecs are better than the other in terms of subjective quality. However, if H_1 is accepted there exists a significant difference between the two codecs and further conclusions can be drawn. The paired t-test was calculated with a significance level of 95%, which indicates a wrongly detected significant difference can occur up to 5%.

3.5 Variables of the subjective quality assessment

Subjective quality assessment always involve some variables, either controlled or uncontrolled. These can range from the human perspective and all the way to the objective perspective such as encoder settings. A brief discussion of the variables in this specific experiment is presented below.

Evaluation area: This criteria is the least controllable variable. As there are noise from multiple computers and noise from the hallway outside, it can influence the result of the subjective quality assessment. However, this was kept at a minimum and the participant was located in a closed off area of the room.

Test methodology: DSIS was adopted with a five level quality scale. Using a continuously rating scale would most likely alter the result, but not necessarily give a more accurate result. Since the test conditions were based on low bit rates with small differences between them, a decimal scale might cause confusion on how to vote.

Viewing equipment: The literature rigorously elaborates factors such as viewing distance, viewing angle and screen brightness [28]. In most cases a subjective quality assessment is performed on a TV or a computer screen where such parameters can be met by finding the best possible equipment for the task at hand. However, when working with tablet computers such luxury cannot be afforded since there only exist a limited amount of devices and not all of them provide precise enough technical specifications. Also, subjective tests tend to be in a very static environment with a mounted screen located in a straight angle in front of the participant. When dealing with hand held equipment, a mounted device does not represent a very natural way of watching video. In addition the angle is often not 90 degrees to avoid reflection of your own mirror image. Given these reasons, the participants were allowed to move the iPad during the test.

Demography: In this experiment 25 people were used. It is important to match the age and multimedia habits to the given scenario. In the case of streaming video on mobile platform, youth seems to be highly represented in terms of usage of the services, and there is a general sense of importance to have the latest and the greatest device. As the majority of the viewers were enrolled at a master program in Electronics or Cybernetics there was a high degree of amusement when they were allowed to test the latest technology.

Content: A proper selection and a correct amount of content is of high importance. Only three sequences were chosen for the experiment, with three different resolutions of each sequence. This was done in order to fit the specific scenario, as content aimed at mobile platforms often are distributed with different bit rates and resolutions.

Compression: In order to provide meaningful results the settings for each encoder should match each other as good as possible in terms of allowed coding tools and complexity. A further elaboration on this matter is discussed in Section 5.2.

Upscaling: The sequences presented in the subjective test were shown in full screen on the iPad. This was done in order to create a more realistic scenario. However, the degree of difficulty for the viewer will increase because the reference frame is also upscaled to full size. The upscaling effect is most apparent at the lowest resolution, and least apparent on the highest resolution.

3.6 Objective assessment

Although subjective assessment is the only way to truly determine video quality, objective assessment can provide a number of advantageous features. Subjective assessment can often be troublesome and time consuming due to the fact that one often need careful planning of the test itself and a group of people to assess the video quality. Objective assessment on the other hand can provide numerical estimations on how good the system is performing. These estimations are very valuable at an algorithmic or design level in the sense that one get a quantifiable metric to describe the quality of the video content. Such metrics can be used to (but not restricted to) optimize algorithms on the fly when dealing with video over error prone channels or improving a compression system at the same relative bit rate. The following section presents the objective metrics used to quantify the video material in this thesis.

3.6.1 Objective metrics

Throughout the years several objective metrics have been proposed. Some aim to strictly look at the signal-to-noise ratio or the mean square error between two images or video frames, while others try to look at the visual similarities (i.e structural similarities) between two difference sources [31]. The most commonly used metric is the peak signal-to-noise ratio (PSNR) given by (3.5).

$$PSNR = 10 \cdot \log_{10}\left(\frac{M_I^2}{MSE}\right) \tag{3.5}$$

Here, M_I^2 is the maximum pixel value and MSE is given by the mean squared error between the reference frame M and the compressed frame C given by (3.6). The total amount of pixels are denoted N.

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^{N} (M_i - C_i)^2 \tag{3.6}$$

PSNR is a powerful tool to use in benchmarks and optimization of algorithms, but it does not consider the visual aspect of the image or video sequence. When using PSNR in describing video quality we often talk about the average over a number of frames. It is quite common to calculate an average over the three color components Y, Cb and Cr or state the PSNR for each individual component.

Bjøntegaard delta rate (BD-rate) is a tool to compute the average PSNR differences between rate-distortion curves [32]. The purpose is to calculate the difference between two coding conditions and present the average difference in bit rate between these two conditions. This method has been highly adopted by JCT-VC, and is the only objective metric presented for each proposal. BD-rate provides a very simple way to analyze the difference between two different systems or two different algorithms over different bit rates and quality levels without having to interpret rate-distortion plots. The values are presented in percent and a positive value would mean that the reference condition is better performing, while a negative value would indicate that the tested condition is better performing. BD-rate has been adopted to present meaningful and simple analysis of the objective assessment.

3.6.2 Correlation of data sets

In order to investigate the correlation and the monoticity between the obtained MOS values and PSNR values, Pearson product-moment correlation coefficient (PPMC) and Spearman's rank correlation coefficient (SRCC) were used.

PPMC is a measure of the linear dependence between two variables. A value of 1 indicates that the relationship between the two data sets can be described as a linear function with perfect correlation. A value of 0 indicates that there is no correlation between the data sets. The equation is given by (3.7), where X_i and Y_i is the data pair of length N and $\bar{X}$ and $\bar{Y}$ are the respective means.

$$Pearson = \frac{\sum_{i=1}^{N} (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}} \tag{3.7}$$

SRCC denotes how well two data sets can be described by using a monotonic function. The SRCC becomes 1 when the tendency is that data set Y increases when X increases. The calculation is similar to the PPMC, except that it orders the values (X_i, Y_i) and assign ranks based on their position. The calculation of

SRCC is given in equation (3.8). Here, the ranks are denoted χ_i and γ_i, and the mid ranks are denoted $\bar{\chi}$ and $\bar{\gamma}$.

$$Spearman = \frac{\sum_{i=1}^{N}(\chi_i - \bar{\chi}) \cdot (\gamma_i - \bar{\gamma})}{\sqrt{\sum_{i=1}^{N}(\chi_i - \bar{\chi})^2} \cdot \sqrt{\sum_{i=1}^{N}(\gamma_i - \bar{\gamma})^2}} \tag{3.8}$$

3.7 Proposed deblocking filter

In order to investigate if there is possible to improve the deblocking filter scheme currently implemented in HM 2.0, a modified deblocking filter was proposed. The proposed filter aims to improve the subjective and objective quality by increasing the compression gain. The main features of the filter is modification of the luma filter decision and filtering of luma and chroma. The proposed design was built on HM 2.0 and does not change the signaling between the encoder and decoding in any way.

3.7.1 Luma filter decision

Recall in Section 2.3.6 that only line number 2 and 5 of an 8x8 block were evaluated and used as a decision of whether or not to deblock the whole 8x8 block. Using only two lines for luma decision has been justified by the fact that high resolution content has larger homogeneous areas. It also reduces some computations and condition queries that may have a impact on the complexity. The usage of only two lines in a block has been accepted as a good enough filter condition due to the fact that more and more content is getting available in higher resolutions. However, on lower resolution content used on mobile platforms, this might not always be the case. Due to this fact the proposed algorithm checks each line, and if it meets the conditions the line is filtered. This behavior is similar to the deblocking in H.264/AVC, where three conditions has to hold to filter a line [11]. In this proposal only one condition has to hold in order to filter the samples. Each line in a 8x8 block is calculated as described in (3.9).

$$d_i = \left|p_{2i} - 2 \cdot p_{1i} + p_{0i}\right| + \left|q_{2i} - 2 \cdot q_{1i} + q_{0i}\right| \text{ for } i = 0 \ldots 7 \tag{3.9}$$

A line is filtered if and only if the condition (3.10) is true for all lines d_i. In a worst case scenario this gives a total of 48 operations (i.e 8 lines, 8 conditions, 8 filter operations for both horizontal and vertical direction). However, since the filter decision has to hold for each line the total number of operations are expected to be lower.

$$2 \cdot d_i < \beta \tag{3.10}$$

3.7.2 Filtering of luma samples

The main goal is to smoothen out high variations inside a block, except when there is a natural edge. In the case of too much attenuation, it is possible to skew the samples too far from the original value. This might be the correct decision for a particular block, but as the rest of the coding chain is dependent on the deblocked sample values other stages of the coding procedure might be affected. One of the main aims are to create the best possible transition between the boundaries. In other words one wants a close to linear transition between the p_i and q_i sample values. Figure 3.2 shows a practical example of deblocking of a line using the proposed luma filter. The luminance sample values are taken from a line in the sequence Heavy bag.

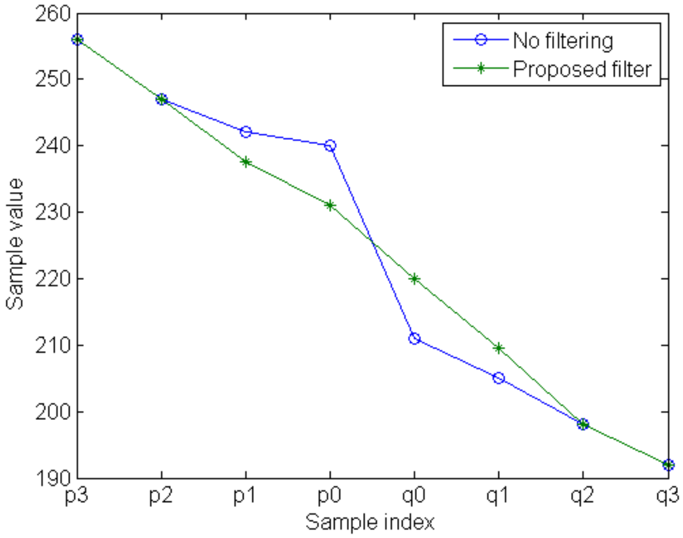


Figure 3.2: Filtered and unfiltered luminance sample values from the sequence Heavy bag.

The strong filtering process is kept intact while the weak filtering process has been manipulated according to (3.11). The sample values are altered in the same way as described in equation (2.5), (2.6), (2.7) and (2.8) discussed in Section 2.3.6.

$$\Delta = Clip3\Bigg(- tc, tc, \Big(14 \cdot (q_0 - p_0) - 3 \cdot (q_1 - p_1) + 10) >> 5\Big)\Bigg) \tag{3.11}$$

In comparison to the current weak luma filter in HM 2.0, the proposed filter only relies on p_1, p_0, q_0 and q_1. A higher weighting of the difference between the innermost samples to the edge (p_0, q_0) is applied. A direct result of this is that the Δ will be higher on average, and thus impose a harder filtering of the block.

3.7.3 Filtering of chroma samples

The filtering of chroma samples are done on a component basis, and only if the frame is an intra frame as specified in Section 2.3.6. The same basic idea of having a transition as smooth as possible applies here as well. The calculation of Δ is equal to the luma filter calculation given by (3.11). Only the innermost pixel values (p_0 and q_0) are filtered as previously stated in (2.5) and (2.6).

Chapter 4

Results

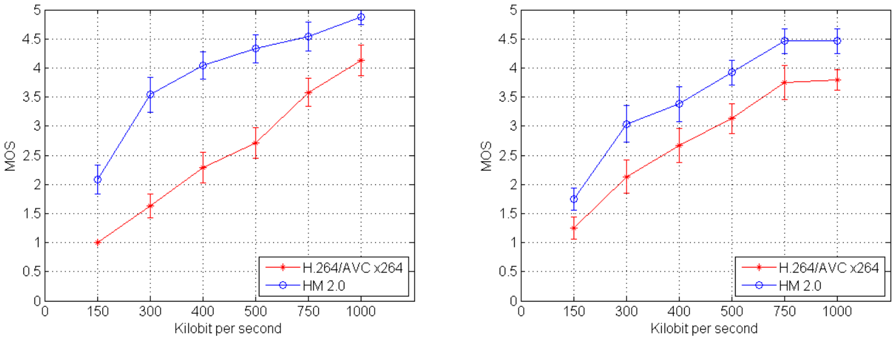
This chapter presents the results of the subjective and objective assessments, as well as the proposed deblocking filter. For the related discussion, see Chapter 5. The results from the subjective quality assessment are given in Section 4.1. This includes subjective rate distortion plots and a comparison of the obtained MOS values. Section 4.2 presents the objective rate distortion plots and the corresponding BD-rate, including observations on the correlation between the presented objective and subjective quality metrics. Lastly, Section 4.3 presents the obtained BD-rate and decoding speed of the proposed deblocking filter and of two deblocking filter from Ericsson and SKKU, already proposed to the JCT-VC committee [33][34].

4.1 Subjective results

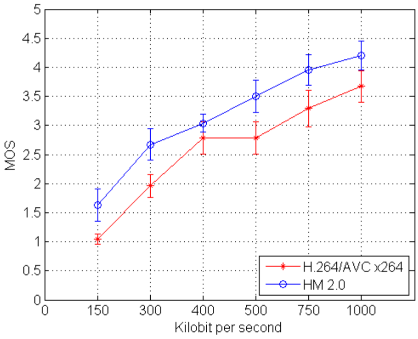
Twenty-five subjects were asked to assess the subjective quality between H.264/AVC and HEVC on an iPad. Three outliers were detected, one for each resolution. The score of these outliers were discarded from the results. The subjective results presented in the next sections are the processed scores. For detailed MOS scores for each participant, see Appendix E.

4.1.1 Rate distortion plots

The MOS values for each resolution are presented in Figure 4.1, 4.2 and 4.3 respectively. Values are shown at different bit rates for the two encoders. The maximum confidence interval over all resolutions was below one unit (±0.4). This indicates that it was good consistency from subject to subject and a general agreement about the quality. The grand average over all resolutions was $\mu = 3.33$. No coding condition reached the absolute maximum on average. This was expected since the sequences are coded with a rather low bit rate. However, the majority of the data distribution was normally distributed, which indicates that the whole scale was used for most of the test conditions.



(a) Heavy bag **(b)** Touchdown day

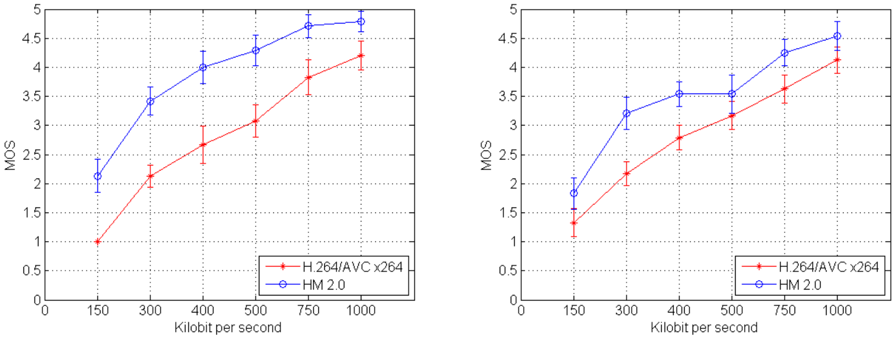


(c) Touchdown pass

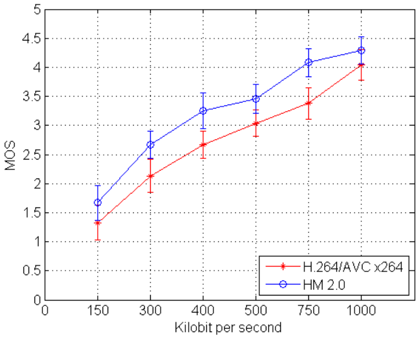
Figure 4.1: Rate distortion plots for Heavy bag, Touchdown day and Touchdown pass sequences with resolution of 640x368. Full size plots can be seen in Figure C.1, C.2 and C.3.

4.1.2 Results of the MOS comparison

The hypothesis test between the means of each test condition was performed as described in Section 3.4.4. A paired T-test was used where each test pair consisted of H.264/AVC and HEVC on the same resolution, bit rate and sequence. The T-test was performed with a confidence level of 95%, which means that H_0 is accepted for p-values above 0.05. H_0 was accepted only once for test pairs at 640x368. The difference of MOS values between the two encoders was not significant at Touchdown pass encoded at 400 kbit/sec with a p-value of 0.056. At 480x272 the H_0 hypothesis was accepted twice. This occurred at Touchdown day encoded at



(a) Heavy bag **(b)** Touchdown day



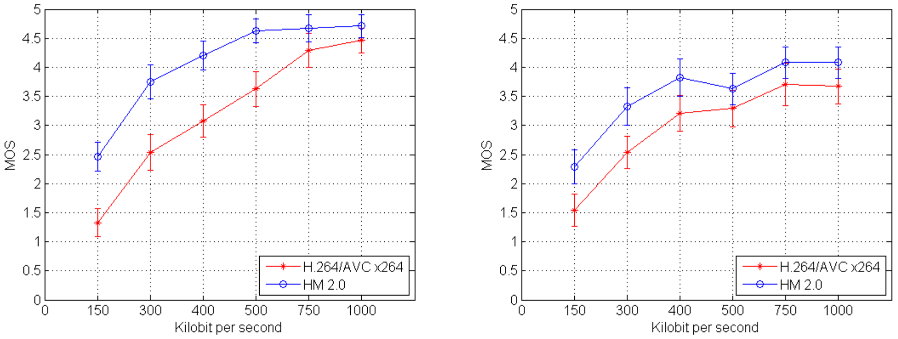
(c) Touchdown pass

Figure 4.2: Rate distortion plots for Heavy bag, Touchdown day and Touchdown pass sequences with resolution of 480x272. Full size plots can be seen in Figure C.4, C.5 and C.6.

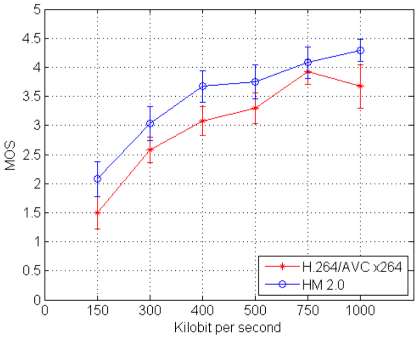
500 kbit/sec and at Touchdown pass encoded at 1000 kbit/sec. The p-values were 0.059 and 0.110 respectively. For sequences encoded at 320x176, H_0 was accepted three times. There was no significant difference between the means for Heavy bag encoded at 1000 kbit/sec, Touchdown day encoded at 500 kbit/sec and Touchdown pass encoded at 750 kbit/sec. The corresponding p-values were 0.056, 0.057 and 0.213. The acceptance of the H_0 hypothesis can be further verified by analysis of Figure 4.1, 4.2 and 4.3.

4.2 Objective results

In addition to the MOS values, average PSNR values for Y, Cb and Cr were logged during the preparation for the subjective test. These were analyzed and



(a) Heavy bag **(b)** Touchdown day



(c) Touchdown pass

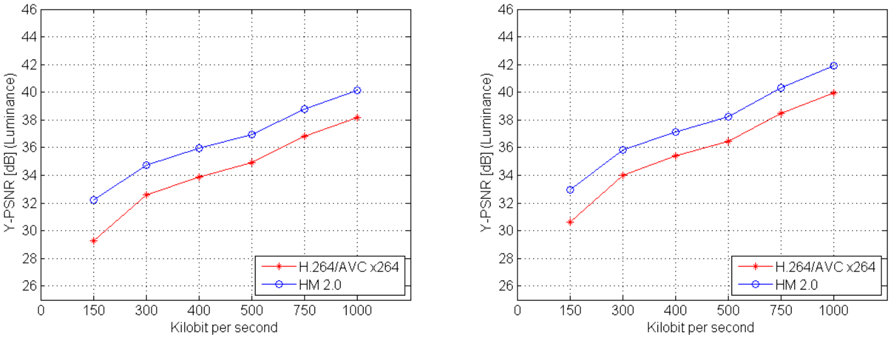
Figure 4.3: Rate distortion plots for Heavy bag, Touchdown day and Touchdown pass sequences with resolution of 320x176. Full size plots can be seen in Figure C.7, C.8 and C.9.

average over sequences is plotted in Figure 4.4. The human eye is more sensitive to brightness than colors and this manifest itself in video compression [35]. Due to this fact and also simplicity, only luma (Y-PSNR) was plotted. For detailed PSNR values and BD-rate computations, see Appendix E.

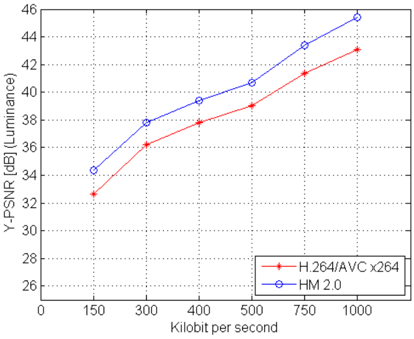
Table 4.1: Average BD-rate computation between H.264/AVC and HEVC HM 2.0 for all sequences and resolutions.

Resolution	Y	Cb	Cr
640x368	-37,6	-9,1	-15,2
480x272	-31,8	-16,8	-19,9
320x176	-26,4	-23,5	-23,1
Average	-32,0	-16,5	-19,4

The average BD-rate for each color component is presented in Table 4.1 together with the rate distortion plots. As explained in Section 3.6.1, a negative number indicates a reduction in bit rate at the same objective quality. The H.264/AVC x264 encoder was used as a reference which means that a negative value is in HEVC's favor. The maximum reported Y-BD-rate was -46.4 for Heavy bag at 640x368 and the minimum Y-BD-rate was -22.5 for Touchdown pass at 320x176.



(a) Resolution of 640x480. **(b)** Resolution of 480x272.



(c) Resolution of 320x176.

Figure 4.4: Rate distortion plots of the average Y-PSNR value over sequences with resolutions of 640x480, 480x272 and 320x176. Full size plots can be seen in Figure C.10, C.11 and C.12.

4.2.1 Correlation between objective and subjective results

The correlation between MOS values and Y-PSNR was computed. The results are reported in Table 4.2. A value of 0 indicates no correlation between the data sets, while a value of 1 indicates perfect correlation between the two data sets. The correlation was computed both with Spearman's and Pearson's method as described in Section 3.6.2. The presented values are close to 1, which indicates that the MOS values increases as the Y-PSNR increases. For additional scatter plots of the data sets, the reader is referred to Figure C.13, C.14 and C.15.

Table 4.2: Correlation between MOS values and Y-PSNR (luminance) at the three different resolutions.

Viteria	640x368	480x272	320x176
Pearson	0.9490	0.9538	0.8645
Spearman	0.9464	0.9422	0.8441

4.3 Deblocking filter results

PSNR values were calculated with the implemented filter as well as two already existing proposed filters. Average BD-rate calculation was performed and the results are reported in Table 4.3 and 4.4. The official unmodified HM 2.0 [27] was set as the reference which indicates that a negative value is in the proposals favor. Detailed results for each sequence can be found in Appendix E.

Complexity assessment of the implemented filter is reported as decoding run times. The decoding run times are reported without writing the output to the hard drive to minimize the inaccuracy. The measurements are reported in Table 4.3. The decoding was performed on a computer running 64-bit Linux 2.6.38-8 with an Intel Core i5 Quad Core 760 at 2.8 GHz and 4GB memory.

Table 4.3: Average BD-rate computation for the proposed deblocking filter over all sequences and resolutions. Decoding time is given as a percentage relative to the official HM 2.0 decoding time.

Resolution	Y	Cb	Cr	Decoding time
640x368	-0.4	-0.1	-0.5	
480x272	-0.5	-0.2	-0.2	
320x176	-0.3	0.1	0,0	
Average	-0.4	-0.1	-0.2	101.5%

Table 4.4: Average BD-rate computation for the proposed deblocking filter by Ericsson and SKKU over all sequences and resolutions.

	SKKU			Ericsson		
Resolution	Y	Cb	Cr	Y	Cb	Cr
640x368	-0.7	-0.3	-0.3	-0.6	-0.2	-0.4
480x272	-0.9	-0.6	-0.4	-0.8	-0.1	-0.4
320x176	-0.7	-0.2	-0.1	-0.7	0.1	0.1
Average	-0.8	-0.4	-0.3	-0.7	-0.1	-0.2

Chapter 5

Discussion

This chapter presents a discussion based on the results presented in Chapter 4. A Discussion on the subjective assessment is presented in Section 5.1, followed by regards on the objective performance in Section 5.2. In Section 5.3, the proposed deblocking filter is discussed. Finally, an overall impression of HEVC and what one should expect in streaming of video is presented in Section 5.4.

5.1 Subjective results

5.1.1 Interpretation of MOS plots

The subjective rate distortion plots presented in Figure 4.1, 4.2 and 4.3, clearly indicates that HEVC has an overall gain in subjective quality compared to H.264/AVC. At the highest resolution (640x368), the sequence Heavy bag was rated more than twice as high at H.264/AVC on the two lowest bit rates (150 kbps and 300 kbps). This trend is perceivable at the higher bit rates as well, though the gap between the encoders are decreasing. This is most likely due to bit rate saturation. It is worth mentioning that at 150 kbps the voting was unanimous for H.264/AVC. On this test condition every subject rated the quality to 1 (Bad), which clearly indicates that at complex scenes such as Heavy bag the H.264/AVC encoder is not capable to produce watchable content that people would likely pay for. At the same test condition HEVC had an MOS value of 2.08. The Heavy bag sequence at 640x368 represent the largest gap between HEVC and H.264/AVC over the various coding conditions. This can be seen when comparing HEVC at 300 kbps with H.264/AVC at 750 kbps, which yields a bit rate reduction factor of 2.5 with the same subjective quality.

The results for Touchdown pass and Touchdown day at 640x368 were not so extreme as the results for Heavy bag. The typical gap between H.264/AVC and HEVC for these sequences was of one unit length in favor of HEVC. As pointed out in Section 4.1.2, there was no significant difference between the MOS values at 500 kbps for the Touchdown pass sequence. Thus, it cannot be concluded that any of the encoders are performing better than the other at this bit rate for Touchdown pass. However, since the difference in MOS values on the higher and lower bit rates are significant, it is very unlikely that there is no significant difference at this specific bit rate. This could be corrected by introducing more subjects to the quality assessment. Another peculiar detail for Touchdown pass is that the subjects were unable to distinguish between H.264/AVC at 400 kbps and 500 kbps. In fact, at these two rates the MOS value and the corresponding confidence interval were perfectly overlapping. The difference between the curves for both Touchdown pass and Touchdown day varied between 1.25 and 1.75.

Subjective assessment at medium resolution was in general more difficult for the participants due to the screening of content in full size on the iPad. For further elaboration of this effect see Section 3.5. Looking at the MOS values for Heavy bag at 480x272, the trend is generally the same as for the large resolution. HEVC has a clear advantage over H.264/AVC for all the bit rates. Also, for this resolution the same unanimous rating happened at 150 kbps for H.264/AVC. A closer look at the rate distortion plots for Heavy bag reveals that HEVC encoded with 400 kbps had a higher MOS value than H.264/AVC at 750 kbps. The same can be seen for HEVC at 500 kbps and H.264/AVC 1000 kbps. On these bit rates one can draw the conclusion that it is possible to obtain the same average subjective quality at half the bit rate using HEVC.

One MOS pair was found to have insignificant difference for the sequence Touchdown day at medium resolution. The insignificance manifested itself at 500 kbps. At this bit rate there was a mutual agreement between the subjects that the quality was equal to the quality at 400 kbps for HEVC. For the Touchdown pass sequence one MOS value pair was found to have insignificant difference at 1000 kbps. This is a normal effect of subjective rate distortion plots because of the saturation of bit rate and quality. The difference between the rate distortion curves was approximately the same as for the largest resolution, between 1.25 and 1.75.

At the lowest resolution it was hard to draw any general conclusion due to the upscaling. However, as before, HEVC was high rated when looking at the the Heavy bag sequence plot. The paired t-test revealed that there was no significant difference between HEVC and H.264/AVC at 1000 kbps in the Heavy bag sequence. The plot further shows that HEVC had approximately half the bit rate at the same subjective quality, except for the highest bit rates.

The plots for Touchdown pass and Touchdown day had some odd dips in performance which indicates that the subjects found it hard to rate the distorted material after looking at an upscaled reference frame. These dips resulted in two insignificant MOS pairs, namely at 500 kbps for Touchdown day and at 750 kbps for Touchdown pass. Despite these two irregularities the MOS values seemed to increase very rapidly at low bit rates indicating that an early bit rate saturation was met. At 300 kbps the MOS value was near good (4), which indicate that the general opinion is that screening of upscaled low resolution content on an iPad is somewhere between a fair (3) and good (4) experienced quality, compared to the reference frame.

5.2 Objective results

The rate distortion plots of the objective performance were presented in Figure 4.4. As explained in Section 4.2 only the Y-PSNR values are plotted as these contain most information. The plots clearly indicate that HEVC has got a foothold over H.264/AVC. This is true for all the resolutions. The objective compression gain is highest at a resolution of 640x368. This was most likely due to the fact that the x264 encoder was struggling to encode the Heavy bag sequence at the lowest bit rates. In fact, the maximum allowed quantizer value had to be increased to 56 in order to meet the bit rate of 150 kbps. By using the standard quantizer values, x264 was not able to reach such a low bit rate at the appropriate compression settings using the built-in rate control mechanisms. This can be explained by the rapid scene changes and movement in the Heavy bag sequence which makes motion estimation and compensation hard to predict.

Further, an interesting parallelism of the two curves can be observed. It seems that they follow each other with high correlation. The only difference between the plots for each resolution is that the gap between the two curves are less at the lower resolutions, in addition to appear at a different region of the Y-PSNR axis, which is to be expected. This parallelism may lead to the conclusion that the settings used for x264 and HEVC has a good match between coding tools and complexity and that the difference in Y-PSNR at each bit rate occurs because HEVC is a better performing video compression system.

The average BD-rate differences between H.264/AVC and HEVC were presented in Table 4.1. When comparing BD-rates they are usually calculated at exactly the same settings and objective quality, i.e with exactly the same quantizer values and quantizer matrix. However, in this thesis the coded sequences were tuned to the bit rate instead of the same quantizer settings. When doing this it is possible to calculate average bit rate reduction at each color component for each content and resolution. The average bit rate reduction was highest for the content at a resolution of 640x368, and lowest for 320x176. This is consistent with Figure 4.4. The extreme bit rate reduction of 46.4% was calculated for Heavy bag 640x368. This is most likely due to the increased maximum allowed quantizer value as described above. However, the rate control mechanisms in x264 are considered to be competitive to JM and thus this extreme value is treated as a proper result [25].

The total average BD-rate reduction was reported to be 32.0%, 16.5% and 19.4% for Y, Cb and Cr respectively. The trend was that the Y BD-rate decreases for each content when the resolution decreases. The Cb and Cr BD-rate follows the opposite trend as the Y BD-rate. The low average Y BD-rate result is most likely due to H.264/AVC's ability to use 4x4 integer transforms which is more efficient on lower resolutions. A weakness of the BD-rate values is that the same average bit rate is used for all the image components, and thus a gain in Y can represent a decrease in Cb and Cr. This issue has been addressed in [36]. An optimal way would be to present the average bit rate for each of the components, however due to the vast difference in coding of luma and chroma this can be a challenging task.

A difference of approximately 10% average Y-BD-rate reduction from the highest to the lowest resolution indicate that the total average is somewhat low. In order to get a better average the results should have been calculated for a higher amount of content, resolutions and encoder settings. In addition, the content should not be limited to sport and it should preferably include different frames per second. However, this was out of the scope for this thesis. A test on broader variety of content and test conditions can be found for some of the submissions to the CfP [37].

5.3 Deblocking filter

A proposed deblocking filter was implemented using the HM 2.0 code base. The BD-rate relative to HEVC HM-2.0 was presented in Table 4.3. The average reduction for Y, Cb and Cr are 0.4%, 0.1% and 0.2% respectively. The source code of two already proposed deblocking filters, was compiled and calculated for the specified coding conditions[33][34]. The results for these proponents are presented in Table 4.4.

The highest peak was found for Touchdown pass at a resolution of 480x272. This peak bit rate reduction was reported to be 0.8%, 0.6% and 0.5% for Y, Cb and Cr respectivly, for the proposed deblocking filter. The same peak was found at same sequences and resolutions for the other proponents, although with slightly higher gains. These BD-rates showed that deblocking is highly dependent on resolution and content. The same weakness of BD-rate discussed in the previous section is apparent here, where a reduction in luma can represent a gain in chroma. Currently, one of the best performing deblocking proposals is a combination of the ones from Mediatek and Ericsson [38]. Unfortunately, the source code was not published and thus no results are presented with this filter. A summary of the various deblocking proposals that are currently being evaluated can be found in [39].

Despite the compression gain of the proposed filter, it has a slight increase in decoding time. The decoding time is reported to be 1.5% higher than HM 2.0 on average. This is most likely due to the implementation of line based filtering. Line based filtering increases the number of calculations compared to the decision perfomed in HM 2.0. However, it is important to specify that this is the case for random access coding with an intra period of 90 frames. The decoding time could have been different if other settings had been used. In addition, due to the low bit rates, deblocking is likely to occur frequently in each sequence. The average decoding time would most probably be lower if a a larger variety of bit rates and content had been used.

During the development of the deblocking filter, experiments were also conducted on filtering of chroma for all boundary strengths. This was found to decrease the overall gain, and in some cases the BD-rate was positive meaning it performed worse than HM 2.0.

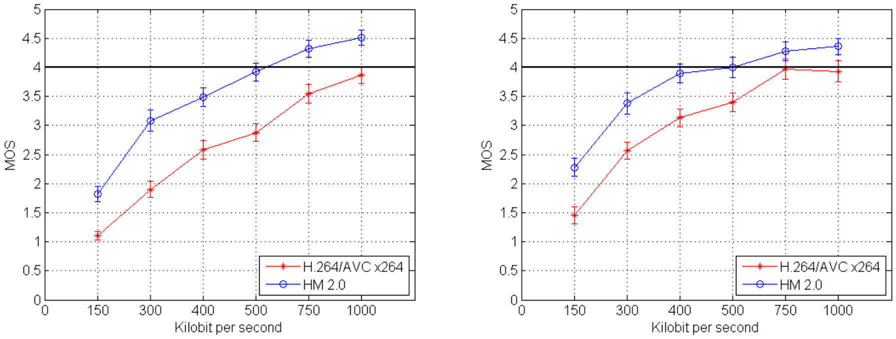
Subjectively it was hard to see a difference between the deblocking in HM 2.0 compared to the proposed filter. It was possible to detect the difference if two identical frames were studied at a high zoom level. The outcome of a subjective test would most probably be the same if the proposed deblocking filter had been included. This is also the general conclusion of a break out group on subjective assessment on deblocking filters within JCT-VC [40].

5.4 Overall impression of HEVC

The overall impression of HEVC is that it has both a subjective and an objective compression gain compared to H.264/AVC. As stated in the introduction the vision is to create a video compression system that has the same quality as H.264/AVC, at half the bit rate. This has been reached to some extent. However, the compression gain was highly variable over different content and resolution. HEVC is only half way through the development and proposals will continue to improve the compression gain. Although many of the proposed techniques only provides minor increases of compressing gain, it is the sum of all the proposals that will make HEVC reach the stated vision.

Because of the highly variable compression gains and MOS values between content it is hard to conclude with definitive bit rate values regarding coding of video content in a streaming environment. As an attempt to anticipate the needed bit rate, a plot of MOS values averaged over content for resolutions of 640x368 and 320x176 can be seen in Figure 5.1. In this calculation 72 values (1 participant discarded) for each test condition is used with a confidence level of 95% and 71 degrees of freedom.

By setting the threshold level to good quality (4) and drawing a line across the plot it can be seen that approximately 550 kbps is enough to provide good quality at 640x368 at this specific content. At the same bit rate x264 is only rated as fair (3). As discussed in this chapter the content includes both high and low coding complexity. Although this is an average and may not be representative for all types of content, it provides a good indication of what service providers should expect when HEVC is finalized. The 320x176 plot shows that the MOS values for HEVC approximate the threshold in a much higher paste. After 400 kbps is reached the steepness of the curve declines which indicates that increasing the bit rate has very little effect on the perceived quality. Since the sequences at this resolution were



(a) 640x368 (b) 320x176

Figure 5.1: Rate distortion plots averaged over MOS values for content at resolutions 640x368 and 320x176. Full size plots can be seen in Figure C.16 and C.17.

highly upscaled it was hard to draw any definite conclusion on which bit rate that may reach good quality.

The improvement of the deblocking filter proves that there is still room for further improvements of the coding tools in HEVC. A reduction in bit rate of 1% should be obtainable based on the already existing proposed filters. This bit rate reduction is possible with little to none increase of coding complexity.

Because of the higher compression gain HEVC offers, it is possible to compress larger resolutions at lower bit rates. While this not necessarily means that the compressed content is of good quality, it provides possibilities to deliver high resolution content to extreme bandwidth limited systems. This is especially apparent by analyzing the MOS values for Heavy bag in Figure 4.1.

Due to the very high correlation between the MOS values and the Y-PSNR values, presented in 4.2, it is possible to draw a general conclusion that when the objective metric increases, so does the perceived subjective quality. This can further prove that there is a close connection between the reported objective BD-rate gain and the MOS values.

5.4.1 Regarding complexity and encoder settings for next generation devices

The ultimate goal of a service provider is to reach as many customers as possible. Some obtain this by using overly simple encoder settings in order to support a large amount of devices. However, these simple settings have an impact on quality because of lower compression efficiency. This in turn create a tradeoff between resolution and bit rate in order to not exceed the capacity of a wireless transmission channel. Such simple profiles typically do not include B-frames and CABAC entropy coding. For many years this have been a reasonable decision due to small screen sizes and low processing power of mobile devices. On the other hand, such simple settings are hardly good enough to present good quality content on newer devices with a much higher screen size and resolution within the constraints of a wireless channel.

In this thesis the x264 encoder was set to use a high profile which includes CABAC, B-frames and adaptive decision of 4×4 and 8×8 transforms while HEVC was configured with similar high efficiency settings. The iPad was able to play back the x264 high profile flawlessly which is a sign of companies adapting to the higher resolution and quality of the available content. In fact, YouTube has started to provide HD content encoded with H.264/AVC high profile [41][1]. This is particularly interesting, because this means that in order to stream 720p or above from YouTube, the device has to support H.264/AVC high profile.

On the topic of HEVC it is difficult to conclude which coding tools will be aimed at smart phones and tablet computers. The complexity of the compression system will most likely change throughout the development. In addition, HM does not serve the purpose of being a practical implementation, but rather a showcase and conformance of the coding tools included. The timeline of the final draft has been set to January 2013 and practical hardware implementations will most likely appear a couple of years after that.

It is clear that simple profiles will still be used at the lower resolutions, but in order to fit the compressed stream of HD content into cellular data networks higher profiles should be adopted.

---

[1] This has been confirmed by the author on several YouTube 720p videos.

Chapter 6

Conclusion

A subjective quality assessment between H.264/AVC and the second version of the official HEVC test model has been presented. The test was performed on an iPad with streaming of sport content at bit rates ranging from 150 to 1000 kbps. The analysis of the assessment indicates that HEVC has a clear gain in subjective quality compared to H.264/AVC. Further, the MOS values indicates that the gain in subjective quality is highly fluctuating over content and resolutions. The bit rate reduction is from 30% to 50% at the same subjective quality compared to H.264/AVC. Results show that in some special cases HEVC was almost rated two quality levels higher than H.264/AVC. This is highly consistent with the objective estimated bit rate reduction.

Based on the objectively measured PSNR, the reduction in bit rate was on average 32.0% for the luma component. The reduction was highest at content with a resolution of 640x368 and lowest at a resolution of 320x176. The maximum achieved reduction of bit rate reduction for the luma component is reported to be 46.4%.

A proposed deblocking filter was implemented based on the second version of the official HEVC test model code base. The proposed filter performs line based decision on filtering of luma and includes a new filtering method for the luma and chroma component. The average reduction in bit rate over content and resolution is reported to be 0.4%, 0.1% and 0.2% for Y, Cb and Cr respectively. The maximum reduction of bit rate is reported to be 0.8% for the luma component. This is considered fine tuning of the existing filter. The decoding time was measured to be 1.5% higher than HM 2.0 which indicates a slightly higher complexity. The increase in decoding time comes from the introduction of line based decision. A 1% reduction of bit rate can be achived with negligible complexity increasement based on already proposed deblocking filters to the HEVC development. The difference in subjective quality between the proposed filter and the official test model was in most cases not apparent.

The overall impression of HEVC after conducting subjective and objective measurements was that it can deliver good quality content at a resolution of 640x368 at a rate of approximately 550 kbps on average over sequences. For the same conditions H.264/AVC did not reach the same quality level for any of the tested bit rates. With this in mind it is clear that HEVC will reach the stated vision of half the bit rate at the same quality as H.264/AVC.

6.1 Future work

In order to provide better results on how HEVC performs in comparison to H.264/AVC, a larger variety of content and test conditions should be tested. This is done to some extent by the JCT-VC team, but there have not been any large subjective quality assessment between the two compression systems since the the call for proposals.

The proposed deblocking filter indicates that there is still room for improvements of the HEVC coding tools. Recent development of the deblocking filter has shown that there is a high effort to further improve the compression gain, but also to reduce the complexity.

During the work with this thesis, a third test model has been released with a higher compression gain over the second version. Further research on how the coding tools evolve and how it relates to subjective quality using the new test model as a reference is encouraged.

Bibliography

[1] Gary J. Sullivan and Jens-Rainer Ohm. Recent developments in standardization of High Efficiency Video Coding (HEVC). volume 7798. SPIE, 2010.

[2] ISO/IEC JTC1/SC29/WG11 Coding of moving pictures and audio. Vision, Applications and Requirements for High Efficiency Video Coding (HEVC). *Video and Requirements Subgroups and JCT-VC, N11872*, 2011.

[3] Ken McCann, Benjamin Bross, Shun ichi Sekiguchi, and Woo-Jin Han. High Efficiency Video Coding (HEVC) Test Model 2 (HM 2) Encoder Description. *Output Document (draft000), JCTVC-D502*, 2011.

[4] Raymond Davis Kell. Improvements relating to electric picture transmission systems. *British Patent 341 811*, 1929.

[5] Google Inc. YouTube Statistics. `http://www.youtube.com/t/press_statistics`, June 13, 2011.

[6] Norwegian Broadcasting Corporation. NRK nett-tv. `http://www.nrk.no/nett-tv/`, June 13, 2011.

[7] TV 2. TV 2 Sumo. `http://webtv.tv2.no`, June 13, 2011.

[8] G.J. Sullivan and T. Wiegand. Video Compression - From Concepts to the H.264/AVC Standard. *Proceedings of the IEEE*, 93(1):18 –31, January 2005.

[9] Apple Inc. iPad. `http://www.apple.com/ipad/specs/`, June 13, 2011.

[10] ITU-T VCEG and ISO/IEC MPEG. Joint call for proposals on video compression technology. *Document VCEG-AM91 of VCEG and N11113 of MPEG*, January 2010.

[11] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz. Adaptive deblocking filter. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):614 –619, july 2003.

[12] ITU-T and ISO/IEC 11496-10 (MPEG-4). Advanced video coding for generic audiovisual services. *Recommendation H.264 (03/10), version 5.0*, March 2010.

[13] D. Marpe, H. Schwarz, and T. Wiegand. Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):620 – 636, july 2003.

[14] Vittorio Baroncini, Gary Sullivan, and Jens-Rainer Ohm. Report of subjective testing of responses to Joint Call for Proposals (CfP) on video coding technology for High Efficiency Video Coding (HEVC). *Document JCTVC-A204*, 2010.

[15] Frank Bossen. Common test conditions and software reference configurations. *Output document, JCTVC-D600*, 2011.

[16] Wen-Hsiung Chen, C. Smith, and S. Fralick. A fast computational algorithm for the discrete cosine transform. *Communications, IEEE Transactions on*, 25(9):1004 – 1009, sep 1977.

[17] Madhukar Budagavi, Vivienne Sze, and Texas Instruments Inc. Minhua Zhou. CE8 Subtest 5: Luma ALF with reduced vertical filter size. *Input Document to JCT-VC, JCTVC-E060*, 2011.

[18] Minhua Zhou Texas instruments Inc. Evaluation results on IBDI. *Input Document to JCT-VC, MPEG m18764 and JCTVC-D025*, January 2011.

[19] M. Jurvansuu, J. Prokkola, M. Hanski, and P. Perala. Hsdpa performance in live networks. In *Communications, 2007. ICC '07. IEEE International Conference on*, pages 467 –471, june 2007.

[20] Roger Pierre Fabris Hoefel and Grace Braz. Throughput and Coverage Analysis of WCDMA and HSDPA Networks. In *Digital Telecommunications (ICDT), 2010 Fifth International Conference on*, pages 70 –76, june 2010.

[21] National Telecommunications and Information Administration Institute for Telecommunication Sciences. The consumer digital video library. `http://www.cdvl.org`, June 13, 2011.

[22] Charles Poynton. *Digital video and HDTV: Algorithms and interfaces.* Morgan Kaufmann Publishers, 2003.

[23] FFmpeg. The FFmpeg project. `http://www.ffmpeg.org`, June 13, 2011.

[24] VideoLan organization x264 development team. x264. `http://www.x264.com`, June 13, 2011.

[25] L. Merritt and R. Vanam. Improved Rate Control and Motion Estimation for H.264 Encoder. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 5, pages V –309 –V –312, 16 2007-oct. 19 2007.

[26] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. H.264/AVC Reference Software. `http://iphome.hhi.de/suehring/tml/`, June 13, 2011.

[27] Joint Collaborative Team on Video Coding (JCT-VC) of ISO/IEC MPEG & ITU-T VCEG. High efficiency video coding test model 2.0. `http://hevc.kw.bbc.co.uk/svn/jctvc-hm/tags/HM-2.0/`, June 13, 2011.

[28] ITU-R. Methodology for the subjective assessment of the quality of television pictures. *Technical Report BT.500-11*, 2002.

[29] Olimsoft. Oplayer HD. `http://www.edavs.com/OPlayer.shtml`, June 13, 2011.

[30] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye. *Probability & Statistics for Engineers & Scientists.* Pearson Prentice hall, eighth edition, 2006.

[31] Zhou Wang, Bovik A.C., Sheikh H.R., and Simoncelli E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, pages 600–612, April 2004.

[32] Gisle Bjøntegaard. Calculation of average PSNR differences between RD-curves. *ITU-T Video Coding Experts Group (VCEG) VCEG-M33*, March 2001.

[33] Andrey Norkin, Kenneth Andersson, and Rickard Sjöberg (Ericsson). Ce12.1: Ericsson deblocking filter. *Input Document to JCT-VC, JCTVC-E276*, 2011.

[34] Jungyoup Yang, Kwanghyun Won, Byeungwoo Jeon (Sungkyunkwan University), and Jeongyeon Lim (SK Telecom). Ce12 subset1: SKT/SKKU deblocking filter. *Input Document to JCT-VC, JCTVC-E417*, 2011.

[35] Peter G. J. Barten. *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality.* SPIE Publications, first edition, 1999.

[36] Dzung Hoang (Zenverge). On the computation of PSNR and BD-Rate. *Input Document to JCT-VC, JCTVC-C315*, 2010.

[37] T. Wiegand, J.-R. Ohm, G. J. Sullivan, W.-J. Han, R. Joshi, T. K. Tan, and K. Ugur. Special section on the joint call for proposals on high efficiency video coding (hevc) standardization. *Circuits and Systems for Video Technology, IEEE Transactions on*, 20(12):1661 –1666, dec. 2010.

[38] Jicheng An, Qian Huang, Xun Guo, Yu-Wen Huang, and Shawmin Lei (MediaTek). Ce12 subtest1: Improved deblocking filter. *Input Document to JCT-VC, JCTVC-E079*, 2011.

[39] Andrey Norkin, Byeungwoo Jeon, and Matthias Narroschke. Ce12: Summary report of core experiment on deblocking filtering. *Input Document to JCT-VC, JCTVC-E032*, 2011.

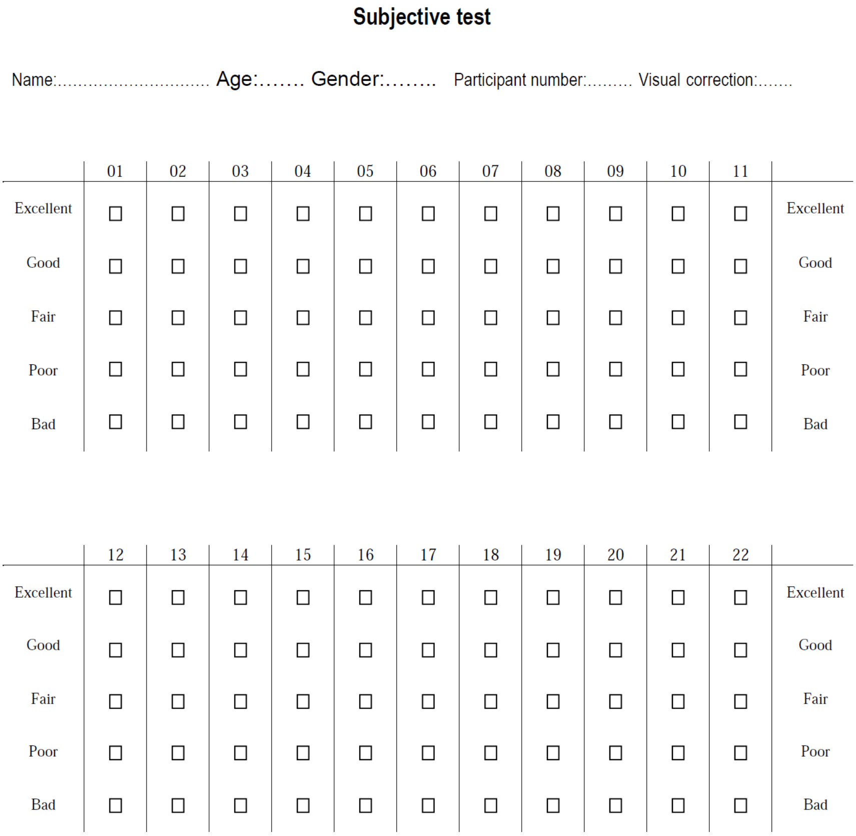
[40] Andrey Norkin, Kenneth Andersson, Keiichi Chono, In Suk Chong, Matthias Narroschke, Byengwoo Jeon, David Flynn, and Teruhiko Suzuki. Bog report on subjective viewing test for deblocking filter proposals. *Input Document to JCT-VC, JCTVC-E501*, 2011.

[41] Streaming learning center. Jan Ozer. YouTube does 720P HD useing H.264. `http://www.streaminglearningcenter.com/articles/youtube-does-720p-hd-using-h264.html`, June 13, 2011.

Appendices

Appendix A

Subjective test documents



Subjective test

Name:............................... Age:....... Gender:........ Participant number:......... Visual correction:.......

	01	02	03	04	05	06	07	08	09	10	11	
Excellent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Excellent
Good	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Good
Fair	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Fair
Poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Poor
Bad	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Bad

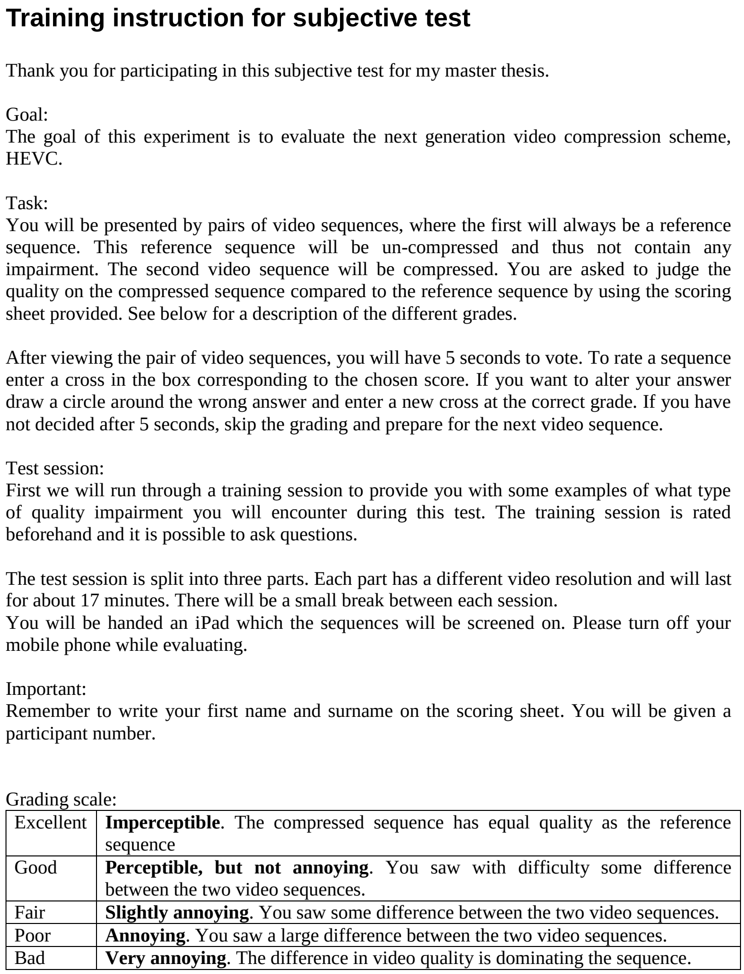
	12	13	14	15	16	17	18	19	20	21	22	
Excellent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Excellent
Good	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Good
Fair	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Fair
Poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Poor
Bad	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Bad

Figure A.1: Scoring sheet for the subjective quality assessment.

	23	24	25	26	27	28	29	30	31	32	33	
Excellent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Excellent
Good	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Good
Fair	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Fair
Poor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Poor
Bad	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Bad

	34	35	36	37	38	39	
Excellent	☐	☐	☐	☐	☐	☐	Excellent
Good	☐	☐	☐	☐	☐	☐	Good
Fair	☐	☐	☐	☐	☐	☐	Fair
Poor	☐	☐	☐	☐	☐	☐	Poor
Bad	☐	☐	☐	☐	☐	☐	Bad

Figure A.1: Scoring sheet for the subjective quality assessment.



Training instruction for subjective test

Thank you for participating in this subjective test for my master thesis.

Goal:
The goal of this experiment is to evaluate the next generation video compression scheme, HEVC.

Task:
You will be presented by pairs of video sequences, where the first will always be a reference sequence. This reference sequence will be un-compressed and thus not contain any impairment. The second video sequence will be compressed. You are asked to judge the quality on the compressed sequence compared to the reference sequence by using the scoring sheet provided. See below for a description of the different grades.

After viewing the pair of video sequences, you will have 5 seconds to vote. To rate a sequence enter a cross in the box corresponding to the chosen score. If you want to alter your answer draw a circle around the wrong answer and enter a new cross at the correct grade. If you have not decided after 5 seconds, skip the grading and prepare for the next video sequence.

Test session:
First we will run through a training session to provide you with some examples of what type of quality impairment you will encounter during this test. The training session is rated beforehand and it is possible to ask questions.

The test session is split into three parts. Each part has a different video resolution and will last for about 17 minutes. There will be a small break between each session.
You will be handed an iPad which the sequences will be screened on. Please turn off your mobile phone while evaluating.

Important:
Remember to write your first name and surname on the scoring sheet. You will be given a participant number.

Grading scale:

Excellent	**Imperceptible**. The compressed sequence has equal quality as the reference sequence
Good	**Perceptible, but not annoying**. You saw with difficulty some difference between the two video sequences.
Fair	**Slightly annoying**. You saw some difference between the two video sequences.
Poor	**Annoying**. You saw a large difference between the two video sequences.
Bad	**Very annoying**. The difference in video quality is dominating the sequence.

Figure A.2: Training instructions for the subjective quality assessment.

Appendix B

Video test material

Figure B.1: Frame 58 of the Heavy bag sequence.

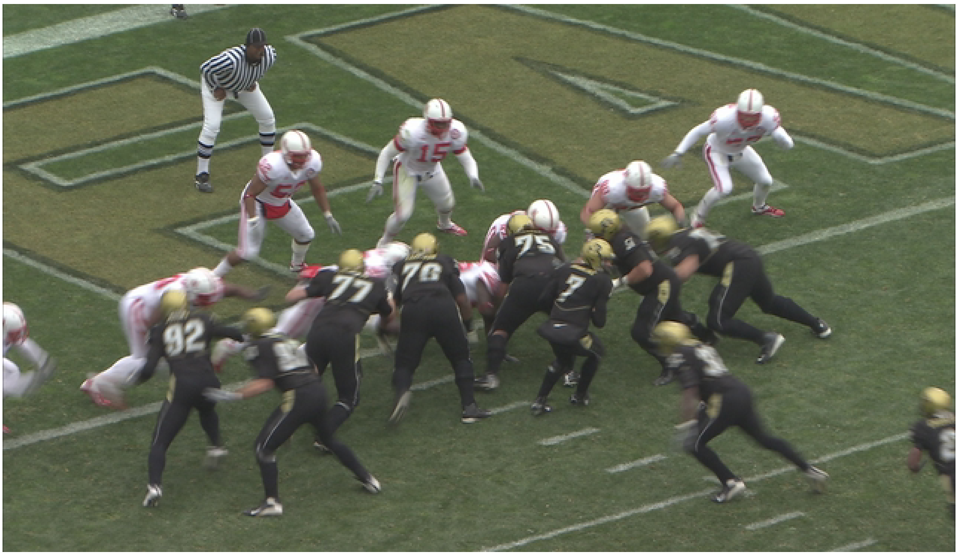


Figure B.2: Frame 121 of the Touchdown day sequence.

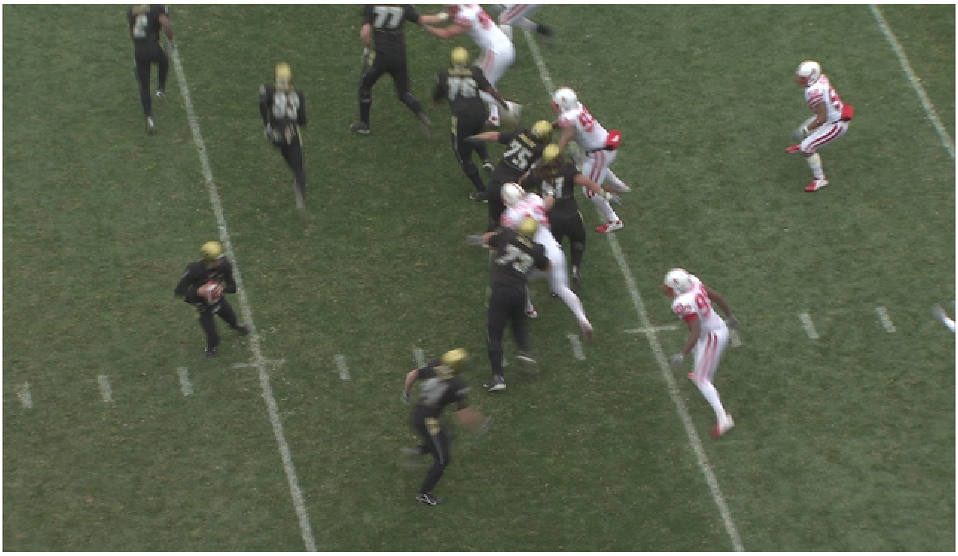


Figure B.3: Frame 57 of the Touchdown pass sequence.

Appendix C

Results

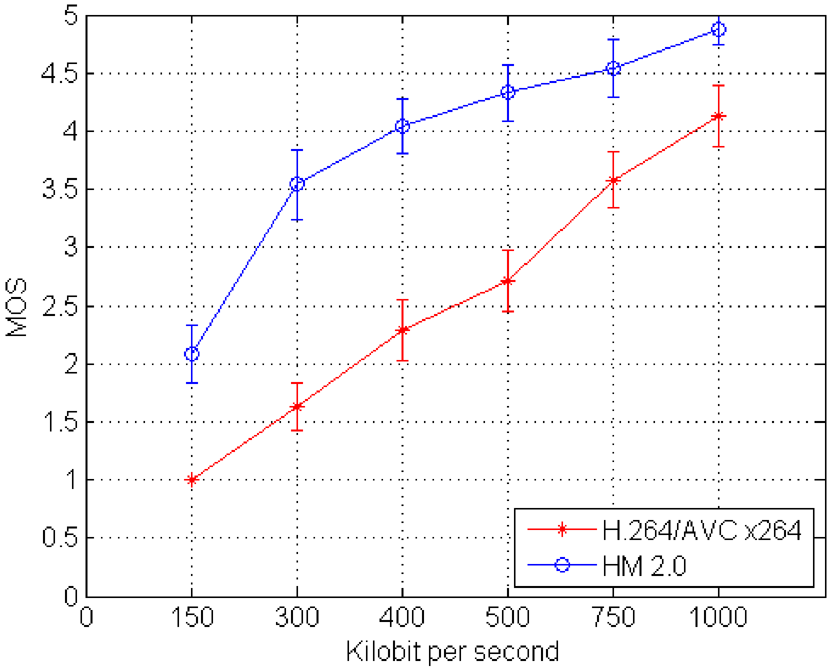


Figure C.1: Rate distortion plot for the Heavy bag sequence with a resolution of 640x368.

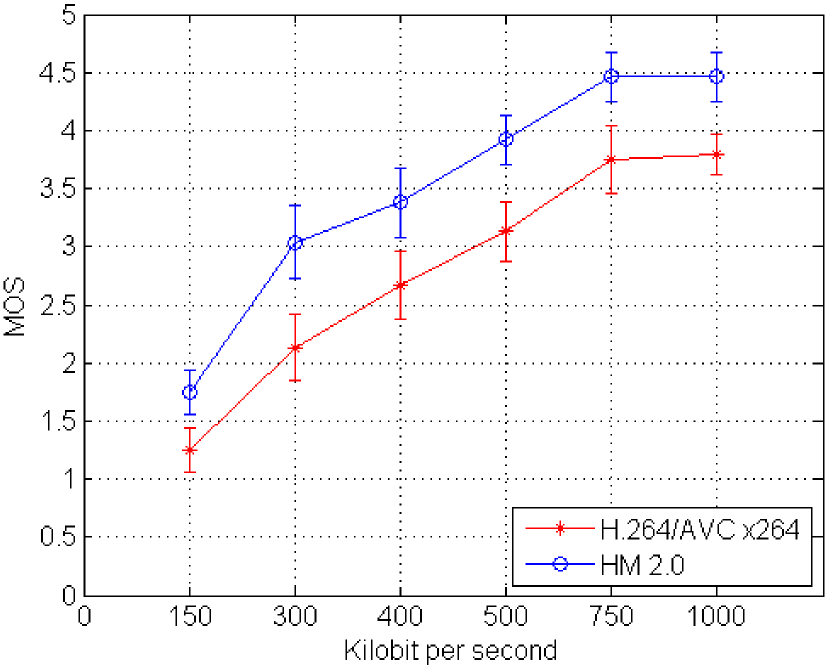


Figure C.2: Rate distortion plot for the Touchdown day sequence with a resolution of 640x368.

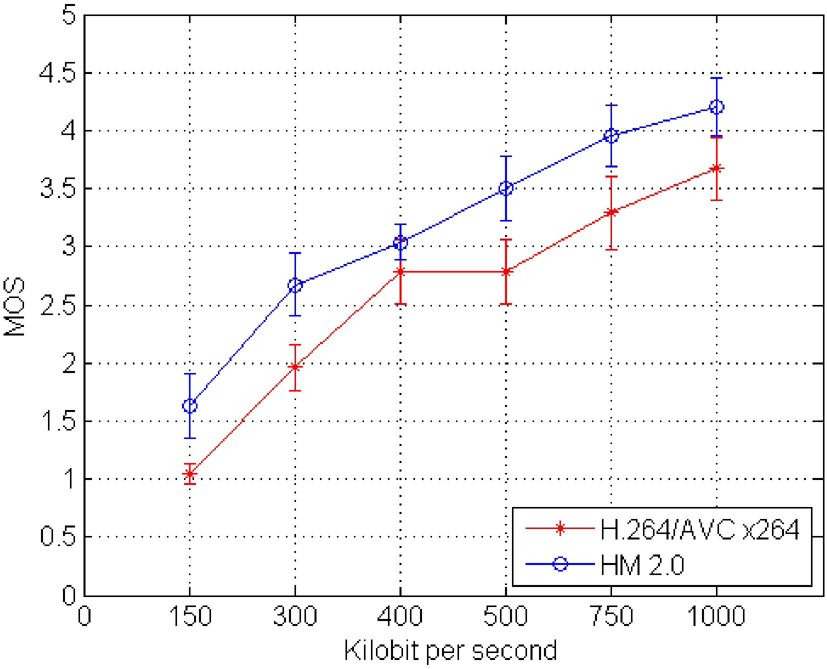


Figure C.3: Rate distortion plot for the Touchdown pass sequence with a resolution of 640x368.

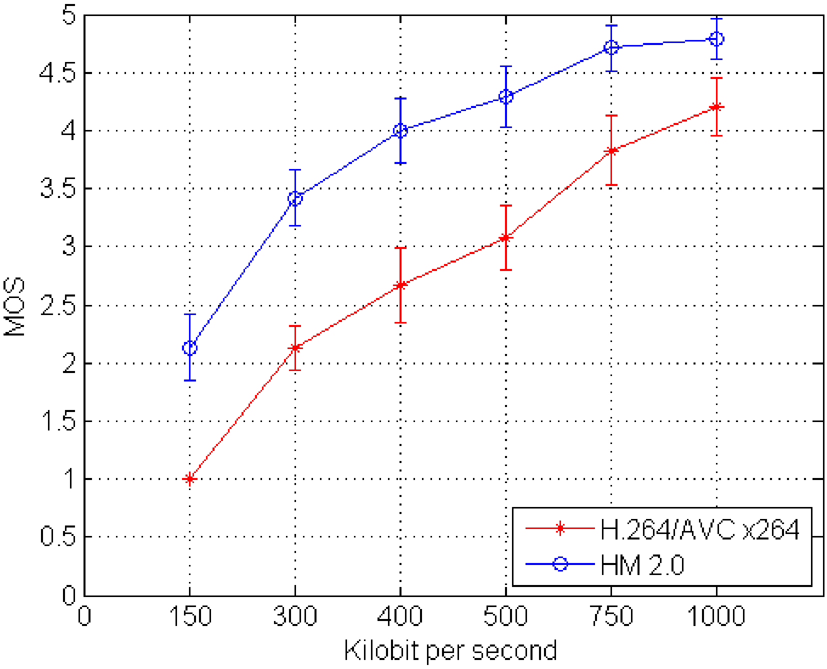


Figure C.4: Rate distortion plot for the Heavy bag sequence with a resolution of 480x272.

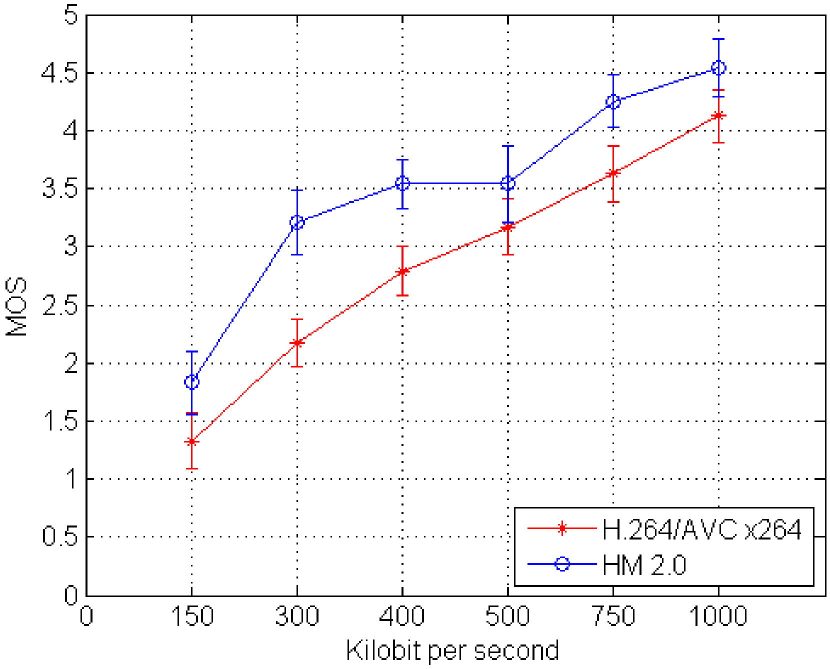


Figure C.5: Rate distortion plot for the Touchdown day sequence with a resolution of 480x272.

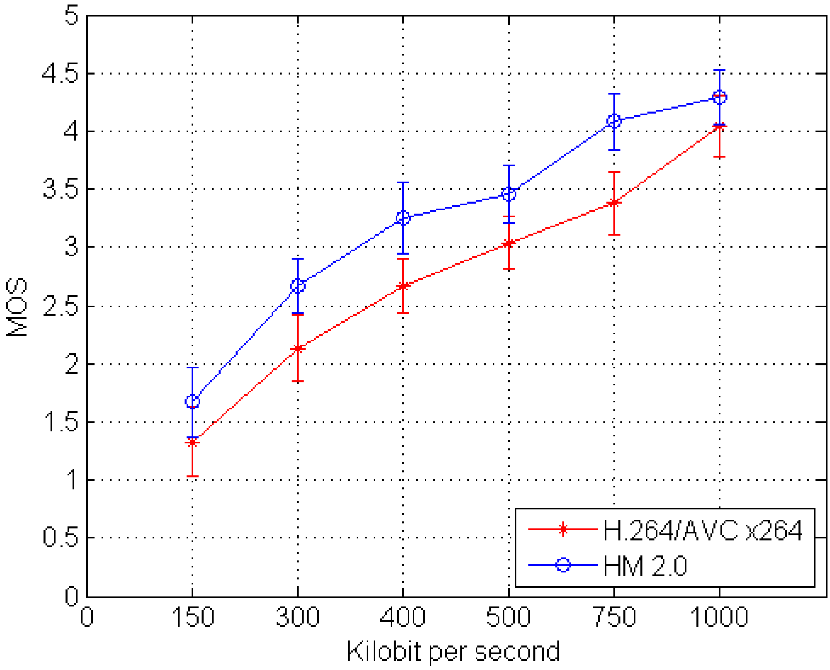


Figure C.6: Rate distortion plot for the Touchdown pass sequence with a resolution of 480x272.

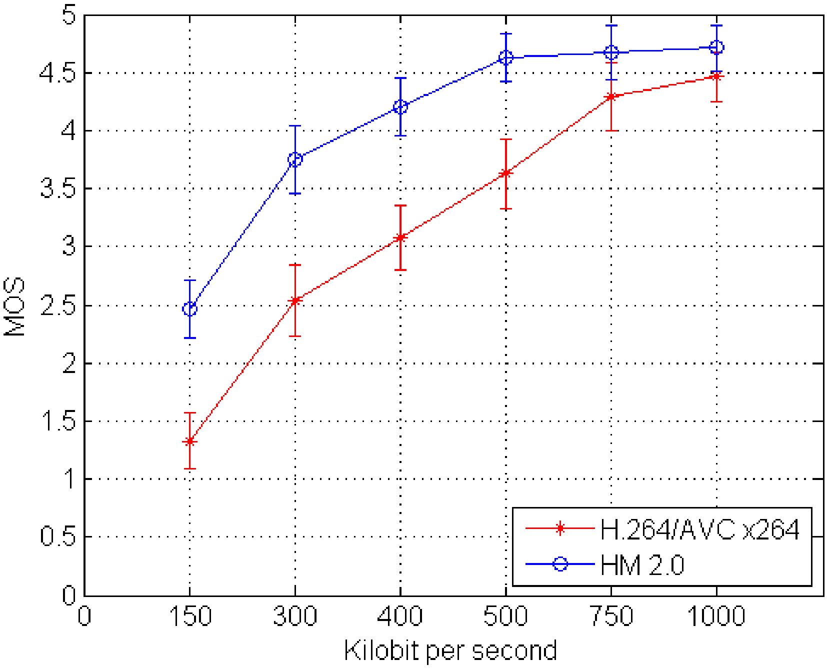


Figure C.7: Rate distortion plot for the Heavy bag sequence with a resolution of 320x176.

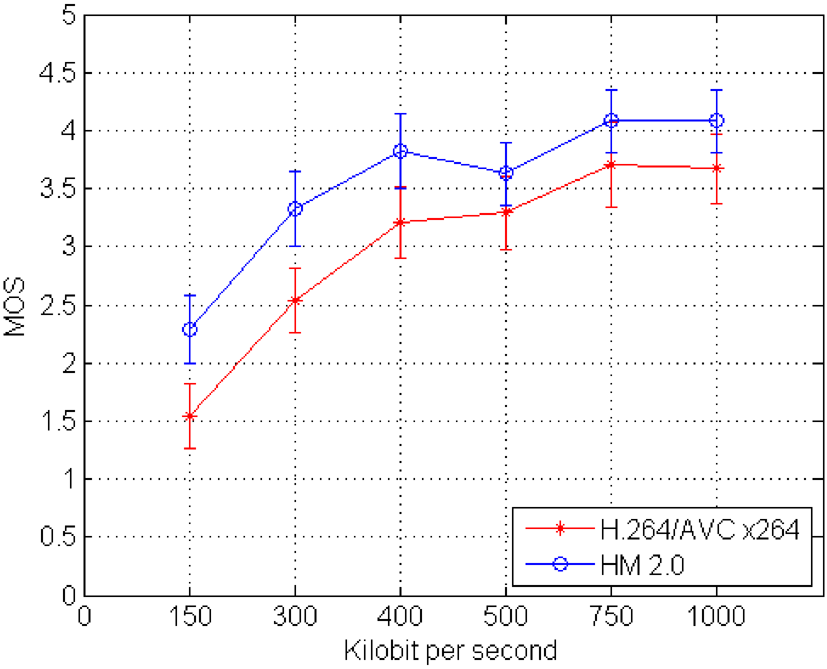


Figure C.8: Rate distortion plot for the Touchdown day sequence with a resolution of 320x176.

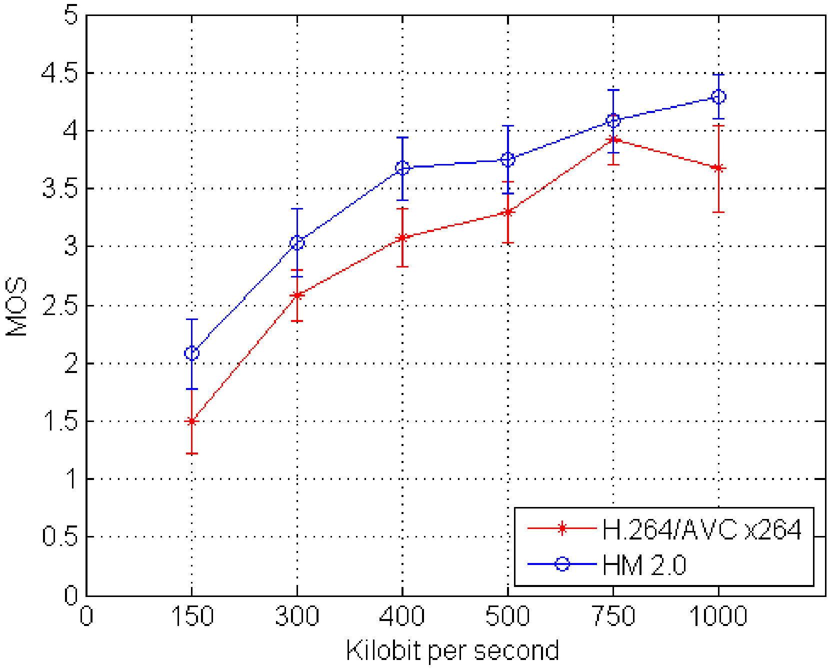


Figure C.9: Rate distortion plot for the Touchdown pass sequence with a resolution of 320x176.

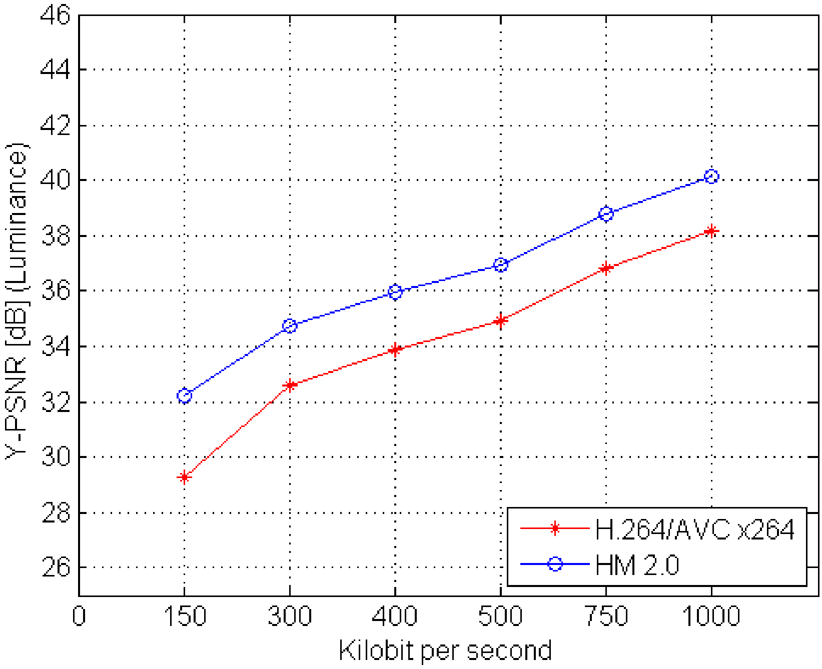


Figure C.10: Rate distortion plot of the average Y-PSNR over sequences with a resolution of 640x480.

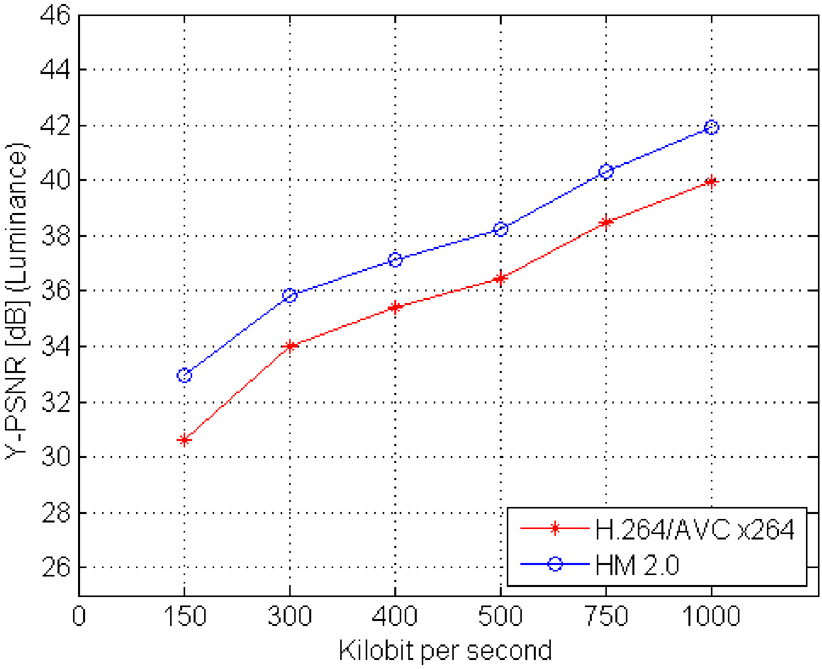


Figure C.11: Rate distortion plot of the average Y-PSNR over sequences with a resolution of 480x272.

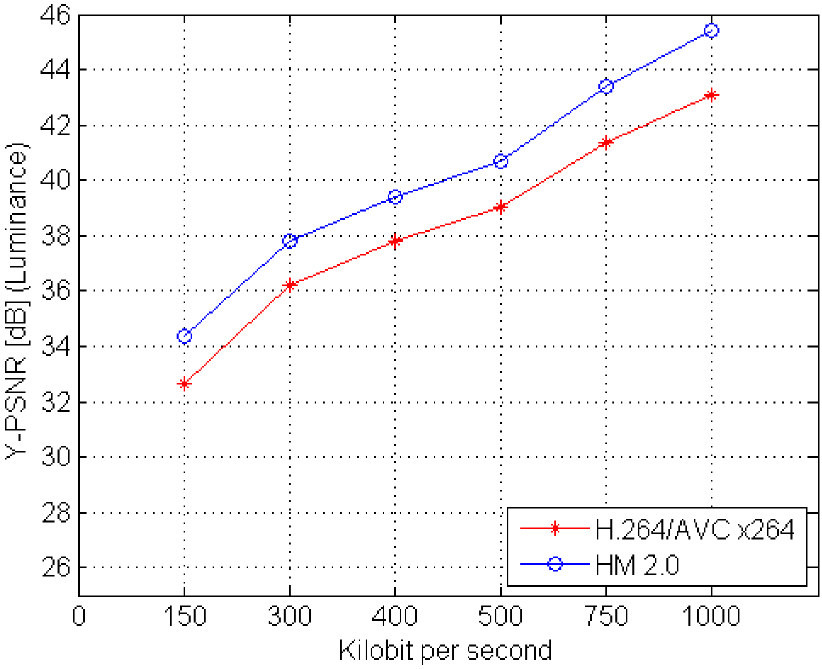


Figure C.12: Rate distortion plot of the average Y-PSNR over sequences with a resolution of 320x176.

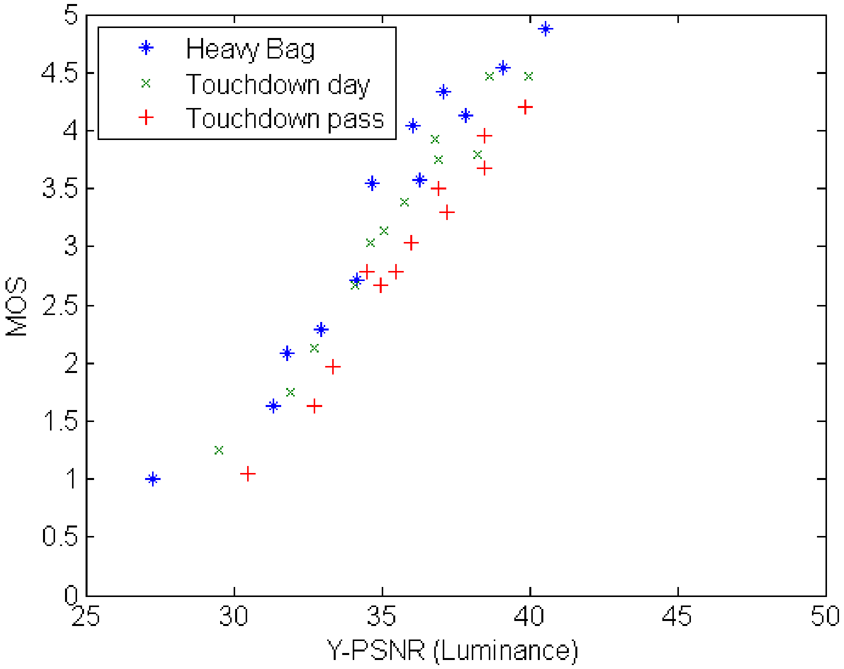


Figure C.13: Scatter plot between MOS values and Y-PSNR of the sequences with a resolution of 640x480.

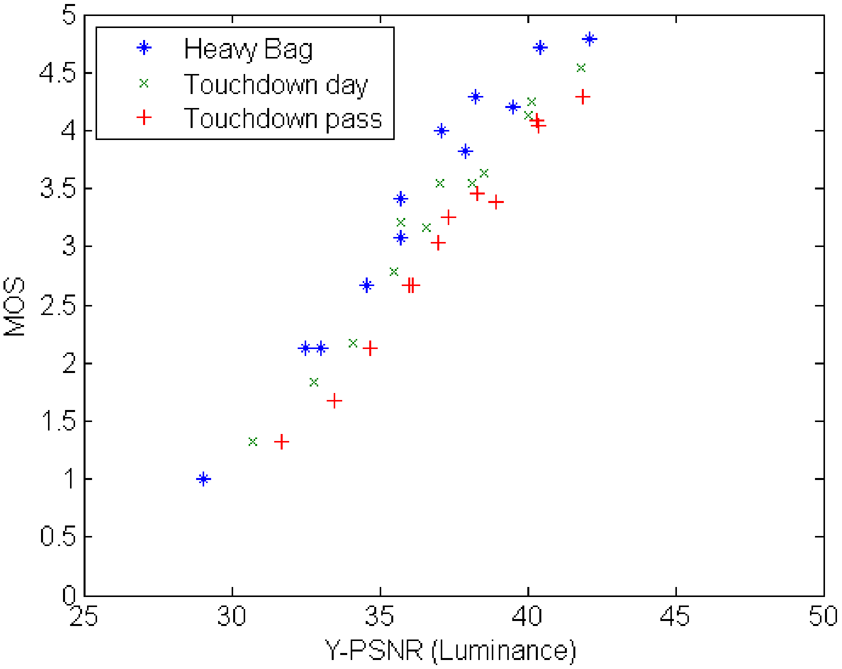


Figure C.14: Scatter plot between MOS values and Y-PSNR of the sequences with a resolution of 480x272.

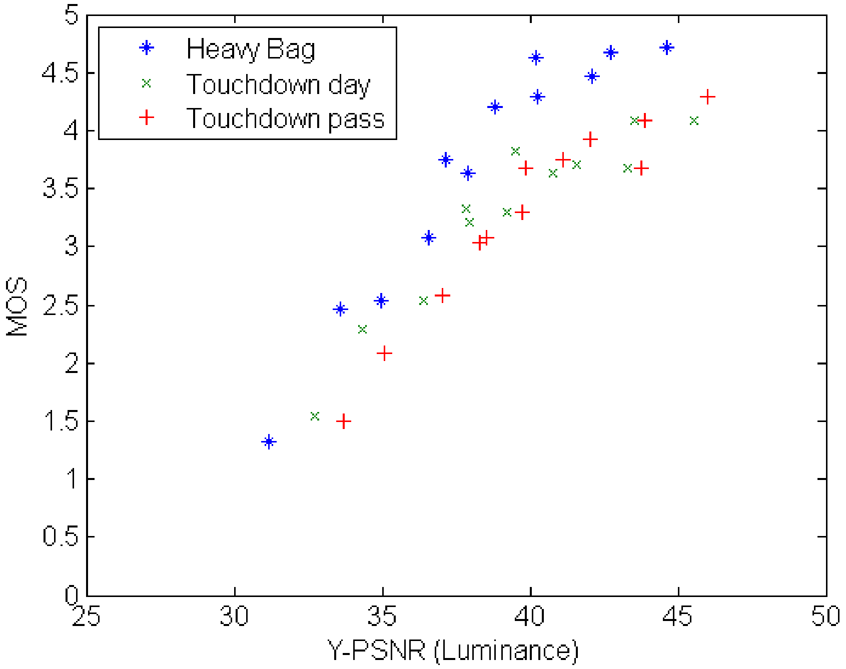


Figure C.15: Scatter plot between MOS values and Y-PSNR of the sequences with a resolution of 320x176.

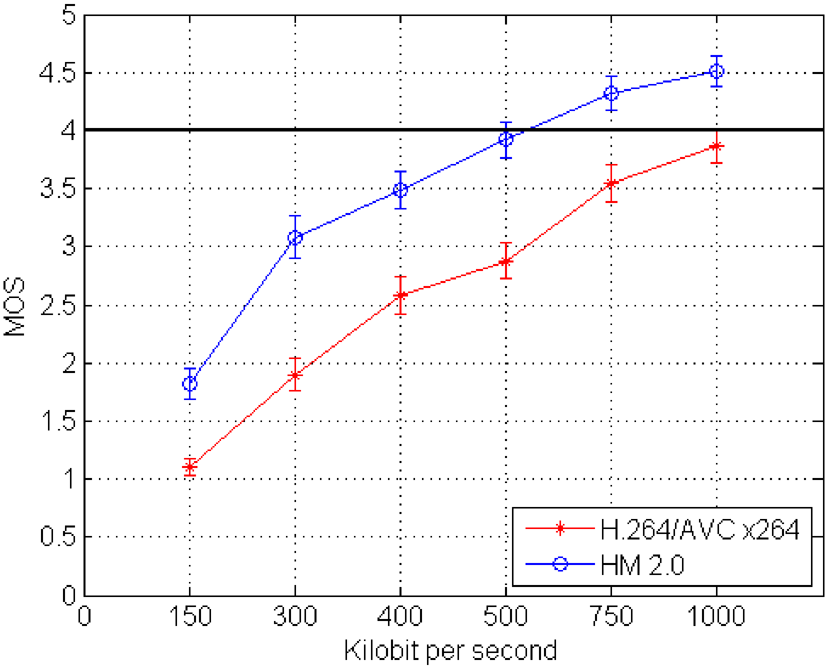


Figure C.16: Rate distortion plot of MOS values averaged over content with a resolution of 640x368.

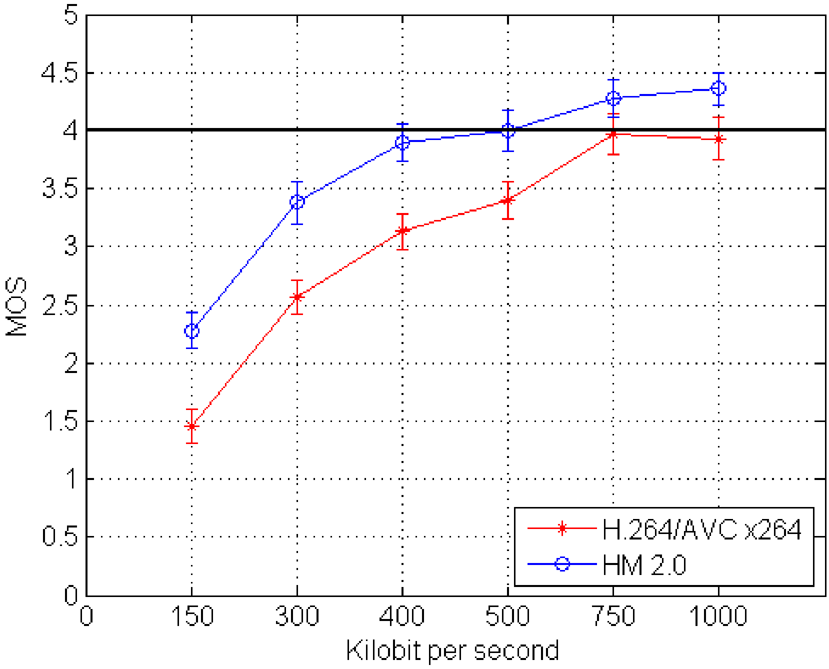


Figure C.17: Rate distortion plot of MOS values averaged over content with a resolution of 320x176.

Appendix D

Encoder settings

D.1 H.264/AVC x264 settings

```
x264.exe -o test.264 --fps 30 --input-res 640x368
..\testseq\yuv\420\ntia_heavybag_640x368_30.yuv --psnr --keyint 90
--scenecut 0 --deblock 0:0 --b-adapt 0 --bframes 3 --qpmin 10 --qpmax 56
--qcomp 0.6 --me umh --merange 64 --partitions partb8x8,parti4x4,parti8x8,partp8x8
--trellis 2 --ref 4 --bitrate 150 --threads 4 --subme 7 --verbose --no-fast-pskip
--mixed-refs --dump-yuv ntia_heavybag_640x368_30_150k.yuv
```

Frames per second(--fps): 30
Input Resolution(--input-res): 640x368
Verbose PSNR (--psnr)
Intra frame period (--keyint): 90
Scene cut detection (--scenecut): 0
Deblock offset (--deblock): 0:0
B-frame adaption (--b-adapt): 0
Number of B-frames (--bframes): 3
Minimum quantizer value (--qpmin): 10
Maximum quantizer value (--qpmax): 56
Quantizer compression curve factor (-qcomp): 0.6
Motion estimation search (--me): umh (uneven multi-hex)
Motion search range (--merange): 64
Block partitions (--partitions): partb8x8,parti4x4,parti8x8,partp8x8
Trellis quantization (--trellis): 2
Number of reference frames (--ref): 4
Average bit rate (--bitrate): 150
Number of CPU threads (--threads): 4
Subpixel estimation complexity (--subme): 7
Display statistics (--verbose)
Disable early skip detection of P-frames (--no-fast-pskip)
Mixed motion references (--mixed-refs)
Write YUV output (--dump-yuv)

All other settings were kept as default.

D.2 HEVC settings

```
#======== File I/O =====================
InputFile                    : ntia_heavybag_640x368_30.yuv
InputBitDepth                : 8
OutputBitDepth               : 8
FrameRate                    : 30
FrameSkip                    : 0
SourceWidth                  : 640
SourceHeight                 : 368
FrameToBeEncoded             : 300
BitstreamFile                : ntia_heavybag_640x368_30.bin
ReconFile                    : ntia_heavybag_640x368_30.yuv
#======== Unit definition ========
MaxCUWidth                   : 64
MaxCUHeight                  : 64
MaxPartitionDepth            : 4
QuadtreeTULog2MaxSize        : 5
QuadtreeTULog2MinSize        : 2
QuadtreeTUMaxDepthInter      : 3
QuadtreeTUMaxDepthIntra      : 3
#======== Coding Structure =======
IntraPeriod                  : 90
DecodingRefreshType          : 1
GOPSize                      : 4
RateGOPSize                  : -1
NumOfReference               : 4
NumOfReferenceB_L0           : 2
NumOfReferenceB_L1           : 2
HierarchicalCoding           : 1
LowDelayCoding               : 0
GPB                          : 1
NRF                          : 1
BQP                          : 0
ListCombination              : 1
#=========== Motion Search =======
FastSearch                   : 1
SearchRange                  : 64
BipredSearchRange            : 4
HadamardME                   : 1
FEN                          : 1
#======== Quantization ===========
QP                           : 30
MaxDeltaQP                   : 0
DeltaQpRD                    : 0
RDOQ                         : 1
#=========== Entropy Coding ======
SymbolMode                   : 1
#=========== Deblock Filter ======
LoopFilterDisable            : 0
LoopFilterAlphaC0Offset      : 0
LoopFilterBetaOffset         : 0
#=========== Misc. ===============
InternalBitDepth             : 10
#=========== Coding Tools ========
MRG                          : 1
ALF                          : 1
```

Appendix E

Zip-file attachment

Content of attached zip-file:

```
Figures/Figure_2_1.png
Figures/Figure_2_2.png
Figures/Figure_2_3.png
Figures/Figure_2_4.png
Figures/Figure_2_5.png
Figures/Figure_2_6.png
Figures/Figure_2_7.png
Figures/Figure_3_1.png
Figures/Figure_3_2.png
Figures/Figure_4_1_heavybag.png
Figures/Figure_4_1_touchdownday.png
Figures/Figure_4_1_touchdownpass.png
Figures/Figure_4_2_heavybag.png
Figures/Figure_4_2_touchdownday.png
Figures/Figure_4_2_touchdownpass.png
Figures/Figure_4_3_320x176.png
Figures/Figure_4_3_heavybag.png
Figures/Figure_4_3_touchdownday.png
Figures/Figure_4_3_touchdownpass.png
Figures/Figure_4_4_480x272.png
Figures/Figure_4_4_640x368.png
Figures/Figure_5_1_320x176.png
Figures/Figure_5_1_640x368.png
Figures/Figure_B_1.png
Figures/Figure_B_2.png
Figures/Figure_B_3.png
Figures/Figure_C_13.png
Figures/Figure_C_14.png
Figures/Figure_C_15.png
Results/BD_rate_computation.xls
Results/HEVC_and_x264_rates.xls
Results/Subjective_test.xls
Results/Subjective_test_outliers.xls
HM-2.0-Dev-avclike/
```