Application of a Bi-Geometric Transparent Composite Model to HEVC: Residual Data Modelling and Rate Control

by

Yueming Gao

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Among various transforms, the discrete cosine transform (DCT) is the most widely used one in multimedia compression technologies for different image or video coding standards. During the development of image or video compression, a lot of interest has been attracted to understand the statistical distribution of DCT coefficients, which would be useful to design compression techniques, such as quantization, entropy coding and rate control.

Recently, a bi-geometric transparent composite model (BGTCM) has been developed to provide modelling of distribution of DCT coefficients with both simplicity and accuracy. It has been reported that for DCT coefficients obtained from original images, which is applied in image coding, a transparent composite model (TCM) can provide better modelling than Laplacian.

In video compression, such as H.264/AVC, DCT is performed on residual images obtained after prediction with different transform sizes. What's more, in high efficiency video coding(HEVC) which is the newest video coding standard, besides DCT as the main transform tool, discrete sine transform (DST) and transform skip (TS) techniques are possibly performed on residual data in small blocks. As such, the distribution of transformed residual data differs from that of transformed original image data.

In this thesis, the distribution of coefficients, including those from all DCT, DST and TS blocks, is analysed based on BGTCM. To be specific, firstly, the distribution of all the coefficients from the whole frame is examined. Secondly, in HEVC, the entropy coding is implemented based on the new encoding concept, coefficient group (CG) with size 4×4, where quantized coefficients are encoded with context models based on their scan indices in each CG. To simulate the encoding process, coefficients at the same scan indices among different CGs are grouped together to form a set. Distribution of coefficients in each set is analysed. Based on our result, BGTCM is better than other widely used distributions, such as Laplacian and Cauchy distributions, in both χ^2 and KL-divergence testing.

Furthermore, unlike the way based on Laplacian and Cauchy distribution, the BGTCM can be used to model rate-quantization (R-Q) and distortion-quantization (D-Q) models without approximation expressions. R-Q and D-Q models based on BGTCM can reflect the distribution of coefficients, which are important in rate control. In video coding, rate control involves these two models to generate a suitable quantization parameter without multi-passes encoding in order to maintain the coding efficiency and to generate required rate to satisfy rate requirement. In this thesis, based on BGTCM, rate control in HEVC is revised with much increase in coding efficiency and decrease in rate fluctuation in terms of rate variance among frames for constant bit rate requirement.

Acknowledgements

There are many people I wish to thank for their support and contributions throughout this period of my graduate education and this thesis writing. Firstly and the most importantly, I would like to express my sincerest gratitude to my supervisor, Professor En-hui Yang, for his insightful guidance and great patience. I have learned greatly from the extensive research training he has supplied and from his wisdom and exceptional knowledge of the field.

I would like to thank Professor Zhou Wang and Professor George Freeman for being committee members and their valuable suggestions on my thesis.

I also wish to thank Dr. Xiang Yu, for his insightful suggestion and useful discussion about my research. His extreme proficiency in state of art video coding technologies has always been a great source of inspirations to me. I am extremely grateful to Dr. Jin Meng and Lin Zheng, for all kinds of their help during this period of my life.

I am also grateful to the lab-mates in Multimedia Communication Laboratory at the University of Waterloo: Mr. Chang Sun, Mr. James Ho, Miss Mahshad Eslamifar and Mr. Nan Hu. It has been a great pleasure to be a part of the lab.

Furthermore, I would like to thank my family for their endless understanding and support.

Table of Contents

List of Tables

List of Figures

List of Acronyms

R_T	target rate
AI-Main	all-intra main
BGTCM	bi-geometric transparent composite model
CABAC	context-adaptive binary arithmetic coding
cbf	coded block flag
CG	coefficient group
CU	coding unit
D-Q	distortion-quantization
DCT	discrete cosine transform
DST	discrete sine transform
HEVC	high efficiency video coding
JCT-VC	joint collaborative team on video coding
kbps	kilobits per second
LCU	largest coding unit
LD-Main	low delay main
MAD	mean of absolute difference
MSE	mean square error

pdf	probability density function
PSNR	peak signal-to-noise ratio
PU	prediction unit
QP	quantization parameter
QS	quantization step size
R-Q	rate-quantization
RA-Main	random access main
RDO	rate distortion optimization
SAO	sample adaptive offset
SDQ	soft decision quantization
TCM	transparent composite model
TS	transform skip
TU	transform unit

Chapter 1

Introduction

1.1 Research Motivation and Problem Description

Image and video compression techniques are usually involved in a practical multimedia system. As shown in Figure 1.1, the system consists of a source encoder, a source decoder, and a transmission channel or storage media, which is applied as the connection between the encoder and the decoder. Image and video compression problems have been addressed between the encoder and the decoder so that coding efficiency achieved by the compression technique is as high as possible. Based on this structure, several industry standards have been developed, such as JPEG [40] for image compression, and MPEG [21] [2] [36]and H.26x [1] [34] for video compression. As video applications with high-resolution video sequences are becoming more popular, developing video compression techniques with high coding efficiency in rate-distortion (RD) performance in order to encode high-resolution sequences is becoming more important. Therefore, the joint collaborative team on video coding (JCT-VC) has finalized the newest video coding standard called high efficiency video coding (HEVC) [7], which has reduced the bit rate by more than 20% while keeping the same reconstruction quality compared with H.264/AVC high profile [34].

In the development of these lossy multimedia compression technologies, a transform is normally employed to transfer signals from the space or time domain into the frequency domain where the signals can be represented by a combination of functions with different frequencies, such as a sequences of cosine functions with different frequencies. Compared with the original signals, the correlation among transformed signals can be reduced and the energy is concentrated in only a few transformed signals, which are referred to coefficients. For example, in JPEG, discrete cosine transform (DCT) is performed on image as the first

step. In video coding standard, such as H.264/AVC and HEVC, DCT is performed on residual data obtained by subtracting predictors from original signals. In the encoding process of all these three standards, transform is followed by quantization and entropy coding.

According to that, the transformed signal can be regarded as the source of quantization and entropy coding. Therefore, good understanding of distribution of transformed signals is important to and beneficial for design of quantization, entropy coding to achieve good coding efficiency and to satisfy some output requirement such as rate requirement in rate control.

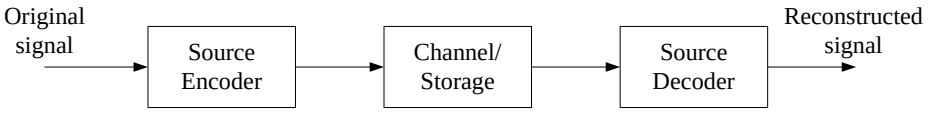


Figure 1.1: Diagram of a Tradition Multimedia System

Unlike image coding, prediction-based video coding technologies, such as H.264/AVC and HEVC, normally perform transform on residual data, which is derived by subtracting predictor from original signals. Furthermore, H.264/AVC and HEVC involve various sizes of blocks to perform prediction, transform, quantization and entropy coding. For example, as in HEVC, each frame is divided into several coding units (CU) . After prediction based on prediction units (PU) , the residual data is obtained, and each frame is further divided into transform units (TU) accordingly. Then, the two-dimensional discrete transform is performed on each TU before quantization and entropy coding. The process to obtain the optimal coding parameters, such as how to partition CUs, PUs and TUs which are determined by minimizing rate distortion (RD) cost is called RD optimization (RDO). This complicated RDO gives rise to challenges to model the distribution of transformed coefficients, especially the coefficients which are encoded into the bit steam. What's more, unlike H.264/AVC where DCT is the only transform method, HEVC employs three different transform methods which are DCT, discrete sine transform (DST) and transform skip (TS). To be specific, DCT is employed in HEVC for all possible sizes of TUs. Discrete sine transform (DST) and transform skip (TS) can be applied to TUs with size of 4×4. These tools also make the modelling difficult.

In previous studies, several models of the distribution of transformed coefficients were reported, including Gaussian [31], generalized Gaussian [29], Laplacian [33] [37],and Cauchy distributions [16]. Among these models, the Laplacian and Cauchy models are the most widely used ones in real applications.

Based on our experimental observation, the actual distribution of transformed coefficients in HEVC differs significantly from both Laplacian and Cauchy distributions in most cases. Laplacian model ignores the fat tails in the distortion of transformed coefficients (large magnitudes of coefficients) because the probability density function (pdf) of Laplacian model drops to zero quickly. However, in the actual entropy coding, these large coefficients contribute quite a few bits to the bit stream, and its contribution to the distortion cannot be ignored either. Furthermore, although the curve of Cauchy distribution drops slower than Laplacian distribution at large coefficients values, it overestimates those probability. In addition, the parameters of Cauchy distribution are normally complicated to be estimated accurately.

Besides compression itself, finding a good model to represent the transformed coefficient probability distributions is also important to the rate control. The rate control is another important application area of video coding technology. Due to limited channel rates, there are rate requirements for some practical video transmission application scenarios. For instance, some applications require a constant transmission rate. In video coding, the rate control scheme is employed to generate required bit rate in the encoded video streams while achieving good coding efficiency.

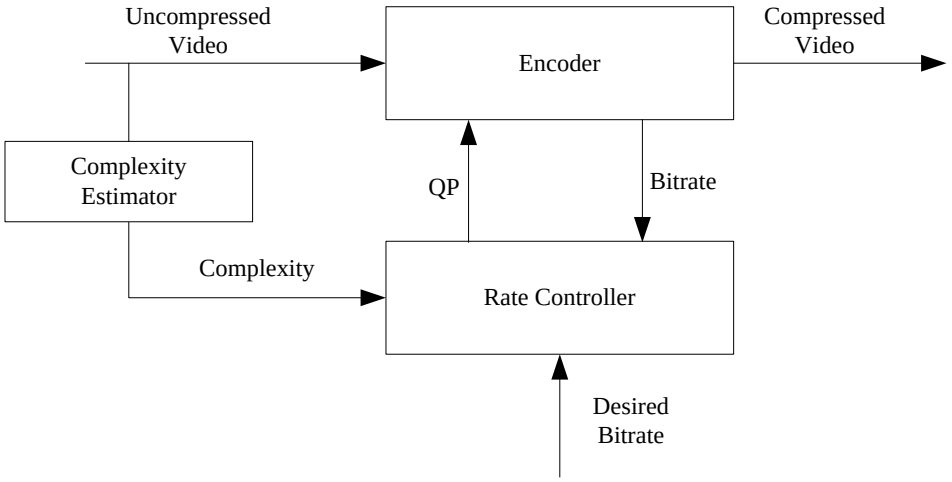


Figure 1.2: Rate Control Diagram in Video Coding [3]

Figure 1.2 shows the diagram of rate control in video coding. In addition to the encoder, a rate controller is employed. During the encoding process, both properties of image and video, and encoding features, such as block size and integer quantization parameter (QP) used for quantization, may affect the final generated rate and the video quality. During

quantization, the transformed coefficients are divided by quantization step size (QS) that is determined by QP. Based on QP, QS can be computed as

$$\mathrm{QS} = q_b\,(\mathrm{mod}(\mathrm{QP}, 6))\,2^{\lfloor \mathrm{QP}/6 \rfloor},\ \mathrm{QP}{=}0{,}1{...}51 \tag{1.1}$$

where, the vector q_b is defined as $q_b = [0.625, 0.6875, 0.8125, 0.8750, 1, 1.125]$ and values in q_b are indexed from 0 to 5.

Since compression can be easily achieved by quantization, the results of choosing a larger QP are that the output bit rate can be reduced, whereas, the video quality is reduced at the same time. The rate controller provides a suitable QP to the encoder according to the information about video features, i.e., the complexity showed in Figure 1.2; the number of bits that have been generated by the encoder; and the target bits that are set by the users or the channel capacity. Therefore, among all coding parameters, QP can be utilized to effectively control the number of generated bits in the encoded stream so that the coding efficiency is maximized without violating the constraints on channel rates.

To derive a suitable QP in rate control, the knowledge of statistic information of transformed coefficients is important. With the carefully selected QP, rate control scheme can achieve good coding efficiency and can generate expected bit rate. However, the output bit rate and distortion cannot be got before encoding finishes. Therefore, the solution to this problem can be obtained by either multi-passes of the encoding process until getting a desired rate and achieving the optimal coding efficiency or using a mathematic model representing the characteristic of the transformed coefficients distribution to estimate the rate and distortion. Since multi-passes solution is not desirable for most applications requiring fast encodings, model-based methods are usually utilized in rate control with the assumption that distribution of coefficients satisfies a certain model.

These mathematical models representing video source statistics can help to describe the relation between the number of generated bits and QP and the relation between the distortion and QP. By modelling the distribution of the transformed residual data, model-based rate control algorithms can estimate the bit rate and the peak signal-to-noise ratio (PSNR) that is most commonly used to measure the quality of reconstruction of lossy compression codecs. With the residual coefficients model, we can get the relation between output bit rate and QP(R-Q model) through the entropy. Also, we can generate the relation between distortion and QP(D-Q model) through PSNR calculation. Based on the R-Q and D-Q models, the suitable QP can be determined to satisfy the rate requirement and to maintain video quality meanwhile without multi-passes encoding. The better the model of the real transformed coefficient distributions fits, the more accurate the estimation of the bit rate and the PSNR will be. Similar to H.264/AVC, HEVC also employs a model-based

rate control method by assuming that the distribution of transformed coefficients satisfies Laplacian density.

The problem involved in the other model-based research in rate control is, that it is hard to get a simple and accurate expression for R-Q and D-Q models with these distributions. Most applications use approximated expressions. Therefore, the parameter used in these models are trained based on different coding standards and testing sequences other than derived from the distribution of actual coefficients from currently encoded sequence or frame directly. Therefore, they cannot reflect the actual coefficient distribution and needs to be adjusted for different compression standards. As the result, the rate and distortion models based on them fail to estimate the actual rate quantization and distortion quantization relationships accurately.

Motivated by these reasons, we want to look for a new distribution model that is a better fit for the video source statistics (the integer transformed coefficients). Furthermore, the parameters are able to be derived easily from the online encoding process which can provide more accurate estimation on statistics information of encoding, such as R-Q and D-Q relations.

1.2 Research Contributions

Recently, a transparent composite model (TCM) [46] has been proposed to provide a model for distribution of DCT coefficients that can handle the fat tail phenomenon in DCT coefficients better with simple ways to derive model parameters. This composite model employs multi-models to describe different parts of the distribution. It is transparent because once the TCM is determined, the specific model for a given coefficient to fall into is clear. It has been reported that TCM can be used to model distribution of coefficients obtained by performing DCT on original images, like the way in JPEG, with both accuracy and simplicity.

Motivated by that, this thesis examines the modelling performance of the newly developed discrete parametric bi-geometric TCM (BGTCM) developed from TCM in HEVC. The BGTCM employs two geometric distributions to composite the model. As described in last section, there are three types of transform implemented in HEVC, i.e., DCT, DST and TS. In this thesis, the data before quantization and after residual derivation, no matter which one out of DCT, DST and TS is performed, is referred to coefficients for simplicity. The coefficients, which are determined after mode decision, are collected through two ways. First, they are collected through the whole frame firstly. In addition, because of the feature

of build the context-model in HEVC, 16 coefficient model are built. In HEVC, no matter what size a TU has, this TU is divided into several non-overlapping coefficient groups (CG) with size of 4×4. CG is the basic unit to run the entropy coding, i.e., context-adaptive binary arithmetic coding (CABAC) , in order to encode the quantized coefficients which contribute the major part to the encoded bit stream. The probability states of different context models in HEVC are initialized at the beginning of encoding each frame. Coefficients at the same scan index (from 0 to 15) among different CGs in each frame share the same set of context models which are used to encode levels and significant bit. Motivated by that, coefficients at the same scan index are grouped into one set.

The BGTCM is employed to model the distribution of coefficients gathered by the two collection methods, i.e., the coefficients from the whole frame and the coefficients grouped based on scan indices. By comparing with the two widely-used models, i.e., Laplacian and Cauchy distributions, in both χ^2 and KL-divergence, BGTCM achieves better modelling accuracy in both data collection methods mentioned above especially for the intra frame.

Because the BGTCM can model the residual coefficients more accurately, it would be helpful for setting up the relation of rate and distortion with the quantization parameter. Therefore, the corresponding R-Q and D-Q models are derived based on BGTCM. To evaluate the R-Q and D-Q models, absolute and relative differences between the actual bits and modelled bits, the actual distortion in terms of MSE and modelled MSE are computed for both intra and inter frames. The better modelling performance of R-Q and D-Q models provided by BGTCM further verifies that the BGTCM is a better model for the residual coefficients than Laplacian and Cauchy distributions.

Furthermore, when implementing BGTCM on HEVC, due to the limited sample range and Laplacian-based discrete probability density function (pdf) of BGTCM, there is no need to approximate the D-Q and R-Q models like Laplacian and Cauchy distributions proposed in other papers. All the model parameters can be obtained online according to actual encoded coefficients rather than being trained offline.The modelling results show that, the estimation of D-Q model is not highly related to coding tools i.e., the distortion can be modelled by BGTCM without approximation. On the other hand, estimation of R-Q model depends on coding techniques. The distortion is modelled directly from BGTCM while R-Q needs scaling.

With the superior of BGTCM when modelling the distribution of coefficients, the rate control scheme in HM-9.0 [5] is revised. Without changing the rate allocation algorithm in codec, the D-Q and R-Q models based on BGTCM can increase coding efficiency while reduce the rate fluctuation in terms of rate variance in the condition of requiring constant rate through all frames when encoding one sequence.

Compared with the HM-9.0 codec rate control, the rate control scheme based on collecting coefficients from the whole frame can reduce the BD-Rate [11] by 13.9% while reduce the rate fluctuation in terms of variance by 64% with turning on the block level rate control. And the rate control scheme based on collecting coefficients are grouped based on scan indices can reduce the BD-Rate by 2.8% while reduce the rate fluctuation by 45% with only turning on the frame level rate control.

1.3 Thesis Organization

This thesis is organized as follows.

Chapter 2 will introduce background knowledge, which will be used in this thesis. Prior work about modeling of distribution of coefficients and rate control in video coding will be presented. In the first section, HEVC and the rate control scheme will be described briefly. Then, the next section will review the previously related work on distribution of transformed coefficients and rate control algorithms. In the last section, because the proposed rate control algorithm is implemented in HEVC codec, the implementation of HEVC rate controller that is based on the Laplacian model will be discussed.

Then, in Chapter 3, the performance of modelling distribution of coefficients in HEVC with BGTCM will be introduced in two parts. The first part is the distributions of transformed coefficients after mode decision from the whole frame. In the second part, distribution of transformed coefficients at different scan indices are modelled. The experiments are performed in both inter and intra frames encoded with different QPs from low ones to high ones. Three models, i.e., BGTCM, Laplacian and Cauchy are employed for modelling. The modelling performance is evaluated according to χ^2 and KL-divergence testings.

After that, in Chapter 4, the proposed rate control algorithm in HEVC based on BGTCM will be presented. Firstly, the rate distortion optimization problem in rate control will be formulated. Then, based on the BGTCM, the corresponding R-Q and D-Q models are derived. Using R-Q and D-Q models with the assumption that distribution of transformed coefficients satisfies BGTCM, the rate control algorithm will be then proposed and the corresponding experimental results will be presented and compared with other related methods. The experiments are performed based on constant rate requirement. The comparisons are performed in both coding efficiency and rate fluctuation.

Finally, Chapter 5 will draw some conclusions and provide some directions for potential future work.

Chapter 2

Background and Prior Work

In this chapter, the background knowledge and prior work related to this thesis is reviewed. At first, the newest video coding standard HEVC is briefly introduced. Then, some prior works about modelling of transformed coefficients and rate control in video coding are reviewed. At last, the framework of rate control in introduced.

2.1 Overview of HEVC

To achieve better video coding RD performance than the H.264/AVC [34] standard, especially for high resolution sequences, a new video coding standard called HEVC [7] has been finalized by the joint collaborative team on video coding (JCT-VC) .

HEVC has proved its superiority in coding efficiency over its predecessors. Compared to H.264/AVC high profile (JM-18.3), HEVC test model (HM-9.0) [5] can save approximately 33% and 37% of bit rate with the same video reconstruction quality for low delay main (LD-Main) and random access main (RA-Main) test configuration, respectively [22], [6].

To encode sequences with high resolution effectively, HEVC incorporates many advanced coding tools, that have not been included in the conventional video standards. During encoding, a video frame is divided into several non-overlapping 64×64 largest coding units (LCUs) . Each LCU is a basic encoding unit relative to the macroblock (MB) with a size of 16×16 defined in H.264. For each LCU, HEVC adopts quad-tree CU structure with a size ranging from 8×8 up to 64×64. Each CU can be symmetrically or asymmetrically partitioned into PUs, where predictions are applied. The residual signal resulting from a prediction is transformed and quantized by TUs, which also have a variable size

block in a quad-tree structure. To choose the optimal coding parameters, such as inter or intra prediction parameters and block partition structure for CUs, PUs and TUs, are called mode selection, which is the most important and most complicated part of HEVC.

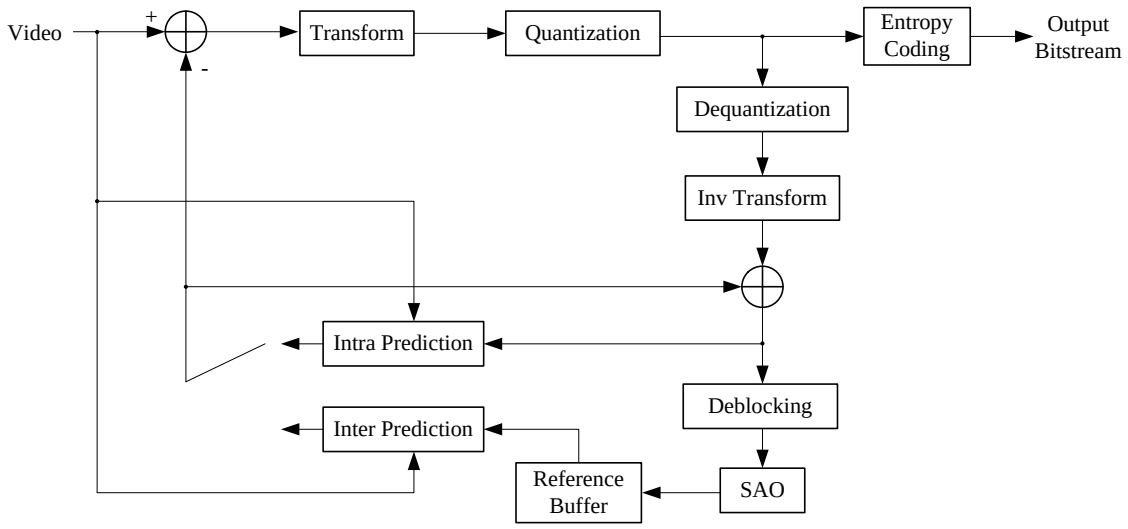


Figure 2.1: Encoding Diagram of HEVC

Figure 2.1 shows the encoding diagram of HEVC, which retains the basic hybrid coding architecture of prior video coding standards, such as H.264/AVC.

Inter/intra prediction is the first step when encoding a frame in HEVC. Inter and intra prediction methods are switched based on configuration and features of current frame and PU. For example, intra frames only allow intra prediction and inter frames allow both inter and intra prediction methods. Inter prediction (motion estimation) reduces temporal redundancy by estimating the current frame from reconstruction of previously encoded frames. This process can achieve accuracy at quarter-pixel level. When an object is encoded in one frame, its appearance in all future frames can be well represented with its shape and displacement. The trajectory of an object (e.g., a moving car) among frames can be estimated, providing an object-based motion field. However, this method of motion estimation is not a practical since it needs a large number of overhead bits to store the arbitrary shape information, and the object shape derivation is very computationally intensive. On the other hand, the block-based motion estimation, due to its fixed shape pattern,utilizes quite a few bits to encode the block partition, and it is computationally tractable. Therefore, the block-based motion estimation is applied in all current video coding standards.

In addition to inter prediction among temporal frames, intra prediction is used to reduce the spatial redundancy. A block is predicted by using intra prediction from reconstruction of previously-coded neighbouring samples in the current frame. When the spatial redundancy is greater than the temporal, intra prediction is used more often than inter prediction. Intra prediction always happens to the frames where scene change occurs.

In video coding, inter/intra prediction determines the residual energy. Therefore, the prediction has a significantly influence on coding efficiency. If the residual energy is high, i.e., the prediction is not good, the bit rate used to encode the residual will be high to maintain the same reconstruction quality. On the contrary, for a limited encoding bit rate, the reconstruction level will degrade.

In prediction, generally speaking, compared with large blocks, blocks with small sizes tend to supply a more accurate prediction, i.e., the prediction for each block is a good match to the actual data in the block. This results in lower residual energy. So fewer bits are required to encode the residual blocks. However, the choice of prediction for every small block must be signalled to the decoder; that is more bits required to code the prediction choices. As such, because of the tradeoff between rate and distortion, how to set the block size is a key problem in the encoding procedure.

After inter/intra prediction, residual information is obtained by subtracting the predictor from the original block in the pixel domain. Then, transform is applied to each TU of this block in order to convert the residual data from the image domain to the frequency domain. The motivation for using transform is to de-correlate signals so that the outputs after transform can be efficiently coded using simple techniques such as scalar quantization (Figure 2.1). It is worthy to point out that, in HEVC, DCT is the major transform method for almost all sizes of TU except when the size of TU is 4×4 where DST and TS can be applied.

Due to the limitations in the human eye's response to spatial details, quantization is applied to video compression. With quantization, a certain amount of distortion may be introduced into reconstruction video sequences where people may not notice it. In addition, even if noticeable distortion occurs, human observers are still able to extract critical information from compressed video signals. Therefore, the quantization is applied in video compression to reduce psycho-visual redundancy.

The simplest scalar quantization is also called hard decision quantization as shown in Figure 2.2, where d is the dead-zone size and s is the QS. Hard decision quantization is implemented to map integer transformed coefficients to some limited indices by simple arithmetical operations based on the input coefficients and QS. By reducing the dynamic range of integer transformed coefficients, quantization can achieve significant bit reduction.

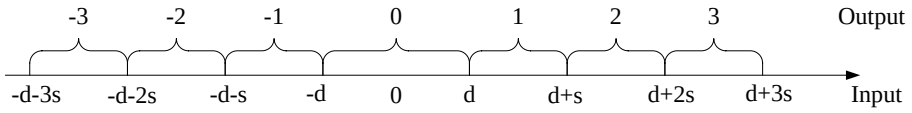


Figure 2.2: An Example of HDQ

In HEVC, rather than the simple HDQ, rate distortion optimized quantization (RDOQ) based on the concept soft decision quantization (SDQ) [44] is applied to increase the RD performance.

In video coding, entropy coding is independent of specific data contents and is developed based on mathematical modelling of the data in order to reduce the statistical redundancy in data. Whereas the quantization makes video coding lossy, entropy coding transfers data in a lossless way.

In HEVC, the quantized coefficients are entropy-coded in groups of 16 coefficients for each TU no matter what size of the TU is. Each group is called a coefficients group (CG). The entropy coding is applied as context adaptive binary arithmetic coding (CABAC) with a similar scheme in H.264/AVC, but it has several improvements to increase the throughput and compression efficiency. In each CG, the scan order may be different, as shown in Figure 2.3. However, when encoding coefficients information, such as levels and significant bits, which context models are used for CABAC will be determined based on the scan index (from 0 to 15) in each CG. In H.264/AVC, a zigzag scan is performed for every 4×4 coefficients. After entropy coding, encoding one frame is finished, and this frame is ready to be transformed or stored.

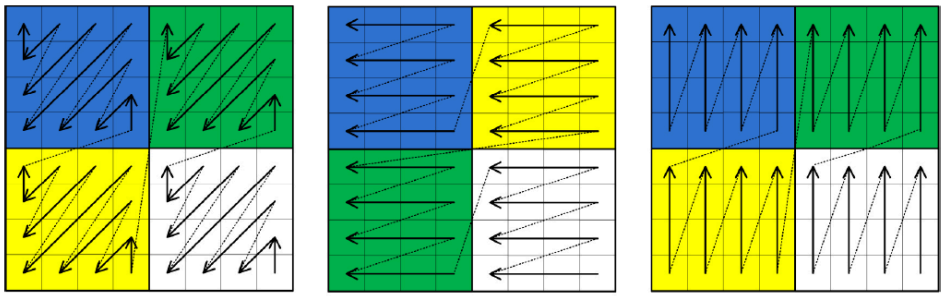


Figure 2.3: Examples of Different Scan Orders in CGs in a 8×8 TU [38]

If an encoded frame will be used as a reference frame for inter prediction in the future encoding procedure, its reconstruction data will be obtained by uniform reconstruction

de-quantization followed by inverse integer DCT, DST or TS. Like other block-based video coding standards, HEVC also suffers from visible discontinuities at block boundaries in reconstructed video sequences. These blocking artifacts are caused by independent coding of blocks, that is the result of applying the block-based inter/intra prediction and transform followed by coarse quantization. To remove such artifacts, HEVC employs a in-loop de-blocking filter and sample adaptive offset (SAO) for better reconstruction. These de-blocking filter in HEVC is implemented to improve the subjective video quality while reducing the complexity of de-blocking method in H.264/AVC, which contributes significantly to the decoder complexity.

The concept of SAO is to reduce mean sample distortion of a region by first classifying the region samples into multiple categories. An offset for each category is derived and added to each sample in that category. For each region, the classifier index and the corresponding offsets are encoded in the bitstream.

2.2 Previous Work on Distribution of Transformed Coefficients

Image and video coding systems with two-dimensional transforms, such as JPEG, H.264/AVC and HEVC, make several different assumptions about the distributions of the transformed coefficients. Based on the central limit theorem, Pratt [31] employs Gaussian distribution to model the pdf of coefficients. Laplacian density is used by Reininger *et al.* [33] and Smooth [37] to estimate that pdf. Based on the quadtree LCU partition structure in HEVC, Seo *et al.* [35] employ separated Laplacian distribution to model DCT from blocks with variable sizes. Kamaci [16] reports that pdf of the DCT coefficients is more conjectured to Cauchy density. These models have simple pdfs that can be used easily for other computations, such as estimation of rate-quantization (R-Q) and distortion-quantization (D-Q) models .

Other studies model the distribution of DCT coefficients using more complicated pdfs. Muller [29] employs a generalized Gaussian function, which combines Gaussian and Laplacian distributions. In [10], Eude *et al.* utilize a linear combination of some of the Laplacian and Gaussian pdfs to model the distribution of actual DCT coefficients. Comparing to pure Gaussian and pure Laplacian density, these complex densities can model the statistics of the DCT coefficients more accurately. However, they cannot be widely applied in practice due to the complication in mathematical analysis.

When pure Gaussian or pure Laplacian density is applied, the probabilities of large

coefficients cannot be estimated properly because the pdf drops to zero quickly and can be ignored for coefficients with large values. However, in actual entropy coding, these large coefficients contribute quite a few bits to the bit stream. If these coefficients cannot be reflected in the pdf, there will be a significant mismatch between the estimated and actually generated rate.

Recently, the TCM and BGTCM [46] have been proposed to modify the estimation of distribution by solving the problem at tails, i.e., large coefficients.

The basic ideas of TCM and BGTCM are the same. Given a sequence of DCT coefficients, they first separate the fat tail of the sequence from the main body of the sequence and, then, use one distribution to model DCT coefficients in the fat tail and a different parametric distribution to model the rest of DCT coefficients. In this way, they can model the fat tail in the DCT coefficient distribution without providing an inaccurate quick drop to zero. The difference between TCM and BGTCM is that how to model the fat tail. Specifically, TCM employs a uniform distribution while BGTCM utilises the same distribution as main body with different parameters, such as geometric model. It has been reported that, by performing DCT directly on original images like the way in JPEG, TCM can provide modelling of distribution of coefficients with both accuracy and simplicity.

In this thesis, target at the transformed coefficients in prediction-based video coding standard, BGTCM is applied to model the distribution of transformed coefficients from the whole frame in HEVC. Furthermore, due to application of CGs in entropy coding, BGTCM is applied to model the distribution of transformed coefficients from the same scan indices from a frame in HEVC.

2.3 Prior Work on Rate Control in Video Coding Based on Modelling of Transformed Coefficients

As introduced in Chapter 1, the rate control scheme is applied to control the generated bit rate by using a specific QP as shown in Figure 1.2. In video coding, the rate controller will determine the QP according to the information of video feature and number of bits have been generated. After one frame is encoded, its rate information will be sent to the controller for future decision. Without several passes or iterative encoding to solve the dilemma, model-based rate control algorithms are usually employed. Based on the assumption that pdf of DCT coefficients can be estimated by some distribution density, the R-Q and D-Q model are established. With the information of rate requirement and

number of generated bits, budget bit for current frame can be derived. Then QP could be computed based on R-Q model.

Several model-based rate control algorithms have been proposed for H.264 and HEVC. The most widely used methods assume Laplacian distribution fit the distribution of DCT coefficients. What's more, due to the easy form of pdf of Laplacian distribution, it is easy to obtain simple expressions of R-Q and D-Q relations. Generally speaking, given a QP, the more complex a sequence is , the less bits are generated. Based on that, some methods involve mean of absolute difference (MAD) as the complexity to combine with Laplacian distribution to estimate R-Q model for rate control.

In [28], a linear R-Q model is proposed, where bits are assumed to be linearly related with the product of QS s and MAD m. In [24], the quadratic rate model proposed in [8] based on Laplacian distribution is employed to determine the QP for a basic encoding unit, which can be either a frame or a MB. This quadratic model is widely used in video coding standards. It involves MAD in the model to illustrate the texture complexity of a unit. MAD for a unit is defined as

$$m = \sum_{i=1}^{n} |O_i - P_i| \tag{2.1}$$

where O and P are the original and predicted signals.

By the Taylor series approximation to the RD function of Laplacian source signals, the quadratic R-Q model proposed in [8] with MAD term is defined as

$$R_T = m\left(\frac{a}{s} + \frac{b}{s^2}\right) \tag{2.2}$$

where R_T , a and b are a target bit rate for the unit and model parameters for the first and second order terms, respectively.

The model parameters a and b are updated every unit by using linear regression [8]. It is worthy to point out that, QS is employed in (2.2) due to the accuracy than QP, which can be mapped to QS in a one-to-one relation. The model in (2.2) can reflect the relation between complexity of contents and QS. Generally speaking, large MAD means residual data is much due to the bad prediction information, if a good reconstruction quality is desired, small QS is preferred. This will cause an increase in bit rate. In rate control problem, small QS should be employed for residual blocks with large MAD to satisfy the requirement of the given target bit rate R_T.

Since the residual signal is not available until RDO finishes, the MAD of a unit is estimated by the MAD of the collocated unit from the previous frame with a linear model

defined as

$$m = c \cdot m_c + d \tag{2.3}$$

where m_c, c and d are MAD of the collocated unit and model parameters. Like a and b in 2.2, c and d are updated based on linear regression as well.

Based on Laplacian distribution, in [13], [14] and [12], He *et al.* presented a rate-distortion model based on the fraction of zeros among the quantized DCT coefficients (denotes as ρ) and a rate-control method based on these models. Based on the observations that the bit rate and ρ have a linear relationship. The proposed rate model is

$$R = \theta(1 - \rho) \tag{2.4}$$

and the distortion model is

$$D = \sigma^2 e^{-\alpha(1-\rho)} \tag{2.5}$$

where θ and α are model parameters and σ^2 is the variance of the coefficients which are assumed to satisfy a Laplacian distribution. The resulting model based on ρ-domain was reported to fit the linear relationship in rate model and to perform better than H.263 and H.264 rate control scheme in terms of both rate fluctuation and coding efficiency.

Based on the model in (2.2),in [15], with the quadratic rate model and the linear MAD estimation model, Jiang *et al.* propose a frame layer bit allocation and improve the complexity measurement by using the MAD ratio and the PSNR drop ratio instead of MAD. In [35], model in (2.2) is separated for blocks of variable sizes due to the introduction of coding tree structure in HEVC. According to that, distributions of transformed coefficients from blocks with different sizes are different. Then the overall R-Q and D-Q models are linear combinations of separated ones.

Besides Laplacian distribution, Cauchy density in (2.6) is widely used in rate control algorithms due to the better fit the fat tails in DCT histograms [16]. In [16], based on zero-mean Cauchy distribution shown as

$$p_c(x) = \frac{\mu}{\pi(\mu^2 + x^2)}, x \in \mathbb{R} \tag{2.6}$$

R-Q and D-Q model are simplified as

$$\begin{aligned} R(Q) &\approx a_c Q^{-\alpha_c} \\ D(Q) &\approx b_c Q^{-\beta_c} \end{aligned} \tag{2.7}$$

where parameters α_c and β_c are related to Cauchy distribution, which only have limited values by setting some intervals for the parameter μ and π, and a_c and b_c are two scaling

factors. Both Laplacian and Cauchy densities are defined in $\mathbb{R}$, however, in video compression, values of DCT coefficients are integers and have limited dynamic range. Furthermore, it is complex to estimate μ accurately. Therefore, Those two densities will not suitable to fit distribution of coefficients. In addition, the R-Q and D-Q models based on Cauchy distribution is approximated. (2.7) cannot reflect information of Cauchy distribution. Furthermore, it may not be proper for every coding standard.

Some other methods employ different model to estimate video complexity, which effects the number of generated bits. For rate control algorithms with focus on intra frames, complexity is highly related to spacial complexity. [20] proposed a complexity measurement to determine QP for an intra frame based on entropy of the block histogram (EBH) which represents signal complexity and uncertainty. Gradient-based complexity models [25] can be used to measure texture distortion. It conducts edge detection in pixel-level and measures bitrate and edge statistical attributes. For inter-frame rate control, Lan *et al.* propose a method based on block histogram difference in [19], which described the difference between two consecutive frames as complexity measurement.

Some methods employs the inter-frame dependency which plays a crucial role in predictive coding to develop rate control algorithms. Ramchandran *et al.* [32] analyze the inter-frame dependency of I, P and B frames in MPEG-2. Lin *et al.* [26] use a piecewise linear approximation model for the rate and distortion dependency among successive frames for rate control. Both [32] and [26] can obtain good coding performance for MPEG-2, but the computational complexity of the proposed algorithm grows exponentially when the number of dependent frames increases.

Since most of the model-based rate control algorithms face the problem how to model statistic information of residual data accurately and effectively, some methods, such as [42] employ two stages in rate control. A subset of RDO is performed in the first stage with a pre-defined QP to derive some rough statistic parameters, such as the motion vectors, the header lengths and information of DCT coefficients. With this rough data, the QP is revised in the rate control algorithm in the second stage by using the information gathered in the first pass. These iterative methods can indeed achieve better rate stabilization and RD performance due to the better estimation in pdf of DCT coefficients. [27] attempts to improve the MAD estimation via a two-pass process with a simplified RDO as the first pass. However, this is at the cost of encoding complexity generated by iterative encoding procedure are mostly used in non-real-time applications.

Furthermore, there are some methods about how to derive a proper QP for the first frame of a video sequence. Rather than default fixed initial QP, [23] determine it based on the bit per pixel (BPP) requirement with some fixed thresholds. [43], [47] and [41] initial

QP is determined by combining BPP and video contents, such as variance and probability of occurrence for grey levels. This is not accurate and based on some ad-hoc methods to fix the relationship between BPP and video contents.

2.4 Rate Controller in HEVC

In the HEVC reference software HM-9.0 [5], the rate control scheme, shown as Figure 2.4, is performed for inter-frames based on the quadratic R-Q model in (2.2) for a basic unit, which can be either a frame or a LCU levels. Each basic unit can be either a frame or a LCU.

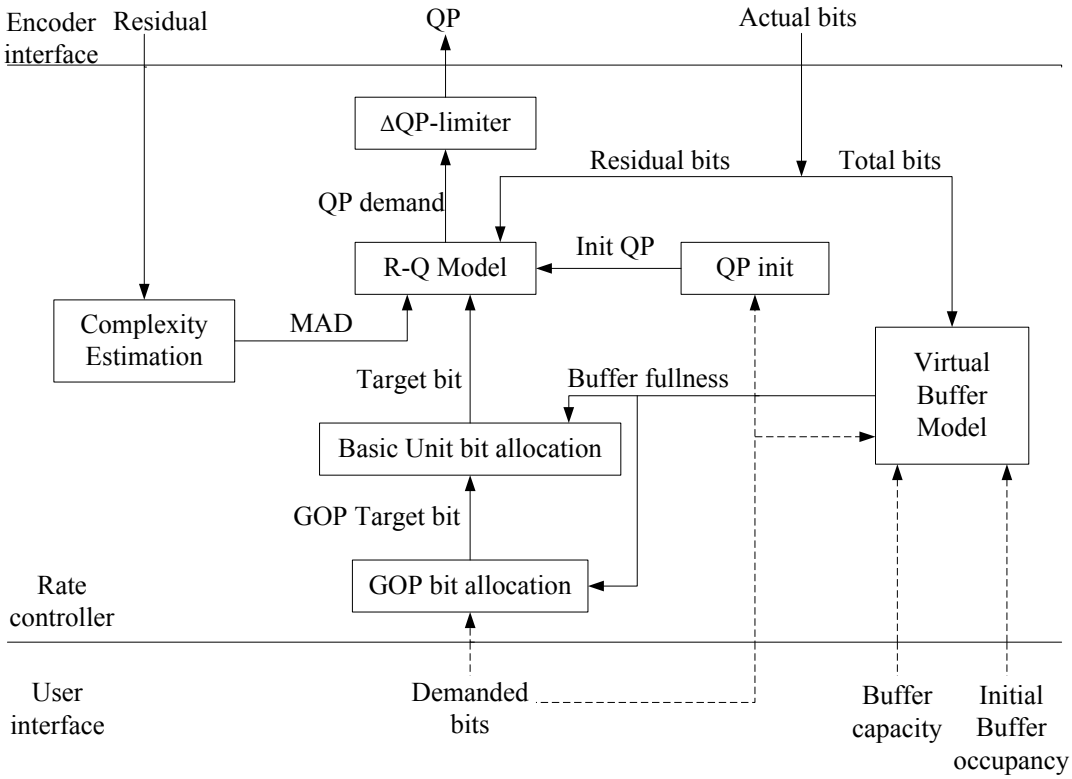


Figure 2.4: Diagram of The Rate Controller in HEVC [3]

The elements of rate controller in Figure 2.4

- R-Q model.

  This is the core of rate controller. QP can only influence the detail of information carried in the transformed residuals. QP has no direct effect on the bits associated with other overhead, such as block partition and motion vectors. As the result, QP in the R-Q model is mainly determined based on residual information. In HM-9.0, R-Q model is based on the quadratic model shown in (2.2).

  Note that a QS value for the current frame is computed from the quadratic R-Q model with the MAD estimated as (2.3). The corresponding QP will be employed in the whole encoding procedure for the unit.

- Complexity estimation.

  MAD is chosen because it is easy to computed and is an inverse measure of prediction quality. Due to the "chicken and egg dilemma", MAD in (2.2) is estimated based on (2.1)

- ΔQP-limiter.

  The rate control system must be implemented to appropriately guarantee stability and to minimize perceptible variations in video quality. For difficult sequences having rapid changes in complexity, QP obtained from R-Q model may oscillate significantly. Therefore, a QP limiter is applied typically to limit changes in QP. To be specific,

  $$|Q_n - Q_{n-1}| \leq 2 \tag{2.8}$$

  where, Q_n and Q_{n-1} are the QPs for current and collocated units respectively.

- QP Initializer.

  Currently, the initial QP is set according to the fixed value in configuration files of testing in HM 9.0. However, it is flexible to develop some methods to set an adaptive QP for different sequences.

- Virtual Buffer Model.

  Any compliant decoder is equipped with a buffer to smooth out variations in the rate and arrival time of incoming data. The corresponding encoder must produce a bitstream that satisfies constraints of the decoder, so a virtual buffer model is used to simulate the fullness of the real decoder buffer.

  This is based on a fluid traffic model and the linear tracking theory introduced in [23]. A mathematical model, also known as a leaky bucket is employed to characterize the

relationship between decoder and its input buffer. Bits stream is filled into the decoder buffer at a constant rate and is removed from this buffer. A bit stream must be decoded in the input buffer without its overflow or underflow. This requirement can be strict satisfied and implemented by the rate control in the encoder.

The change in fullness of the virtual buffer is the difference between the total bits encoded into the stream and a constant removal rate assumed to equal the bandwidth (or demanded bitrate). The buffer fullness is bounded by zero from below and by the buffer capacity from above. The user must specify appropriate values for buffer capacity and initial buffer fullness, consistent with the decoder levels supported.

- GOP Bit Allocation.

  Based upon the required bit rate and the current fullness of the virtual buffer, a target bit rate for the entire group of pictures (GOP) is determined, which are allocated to all frames equally. The QP for the first frame in the GOP is also determined.

- Basic Unit Bit Allocation.

  Based on the buffer status and bits generated by the previous unit, the remaining bits are allocated to all non-coded basic units in the current frame/GOP equally because the MADs of non-coded basic units are not known. If the remaining bits are negative, QP will be simply increased by 2 to reduce the generated bits.

2.5 Summary

In this chapter, some background knowledge is introduced. Firstly, the newest video coding standard HEVC is briefly introduced. After that, some prior work on distribution of transformed coefficients and rate control are reviewed. At last, the general framework of rate control is introduced.

Chapter 3

Modelling Transformed Coefficients in HEVC Based on BGTCM

In this chapter, the application of BGTCM in HEVC is presented. BGTCM is used to model the distribution of both original and quantized coefficients. The modelling provided by BGTCM will be compared with those models provided by Laplacian and Cauchy distributions.

3.1 Modelling of Distribution of Transformed Coefficients

In this section, the modelling of distribution of transformed coefficients in HEVC are discussed with BGTCM, Laplacian and Cauchy distributions. The transformed coefficients are derived after mode decision. Modelling are performed in two different methods. Firstly, the modelling is performed on the coefficients collected from the whole frame. Secondly, due to the CG-based coefficients coding method in HEVC, coefficients are grouped based on the same scan indices among different CGs. Then modelling is performed on each group.

3.1.1 Distribution of Transformed Coefficients from the Whole Frame

Recently, a TCM has been developed to better model DCT coefficient distribution when performing DCT for image and video coding. Compared with some widely used statistics

distribution, such as Laplacian distribution whose probability distribution function (pdf) $p_L(y)$ is defined as

$$p_L(y) = \frac{1}{2\lambda_L} e^{-\frac{|y-\mu_L|}{\lambda_L}} \tag{3.1}$$

where λ_L and μ_L are two model parameters, the results [46] and [45] show that the benefits of BGTCM are due to the fact that the fat tails in DCT coefficient distribution cannot be ignored as they are in other statistic models. If Laplacian distribution is employed to model the DCT coefficient distribution and since the exponential curve drops fast towards zero for the coefficients $\{y \mid |y| > y_c\}$, it does not fit the real distribution very well especially when there are a lot of DCT coefficients s.t. $|y| > y_c$.

For a random variable y s.t. $|y| \leq a$, the pdf $p_T(y)$ of TCM is defined as

$$p_T\left(y|a, y_c, b, \theta, c\right) = \begin{cases} \frac{b}{2F(y_c|\theta)-1} f\left(y|\theta\right), & \text{if } |y| < y_c \\ \frac{1-b}{2(a-y_c)}, & \text{if } y_c < |y| \leq a \\ \frac{bc}{2F(y_c|\theta)-1} f\left(y_c|\theta\right) + \frac{(1-b)c}{2(a-y_c)}, & \text{if } y = y_c \\ 0, & \text{otherwise ,} \end{cases} \tag{3.2}$$

where

$$F\left(y|\theta\right) = \int_{-\infty}^{y} f\left(u|\theta\right) du. \tag{3.3}$$

For computation simplicity, in this proposal, zero-mean truncated Laplacian distribution is employed for $f\left(y|\lambda_T\right)$, i.e.,

$$f\left(y|\lambda_T\right) = \frac{1}{2\lambda_T} e^{-\frac{|y|}{\lambda_T}}. \tag{3.4}$$

Therefore, the pdf $p_T(y)$ of TCM can be rewritten as

$$p_T\left(y|a, y_c, b, \theta, c\right) = \begin{cases} \frac{b}{2\lambda_T\left(1-e^{-y_c/\lambda_T}\right)} e^{-|y|/\lambda_T}, & \text{if } |y| < y_c \\ \frac{1-b}{2(a-y_c)}, & \text{if } y_c < |y| \leq a \\ \frac{bc}{2\lambda_T\left(1-e^{-y_c/\lambda_T}\right)} e^{-y_c/\lambda_T} + \frac{(1-b)c}{2(a-y_c)}, & \text{if } |y| = y_c \\ 0, & \text{otherwise .} \end{cases} \tag{3.5}$$

However, in practice, such as H.264/AVC and HEVC, transform is often implemented to transfer integer values based on the image domain to other integer values based on the frequency domain. Rather than modelling the coefficients in continuous domain in (3.4), a discrete parametric TCM for an integer random variable y based on a truncated geometric distribution is developed as

$$p_T(y) = \begin{cases} bp & \text{if } y = 0 \\ b(1-p)\frac{1}{2}\left[1 - e^{-q_T/\lambda}\right] e^{-q_T \frac{(|y|-1)}{\lambda}} / \left(1 - e^{-q_T y_c/\lambda}\right) & \text{if } |y| \leq y_c \\ \frac{1-b}{2(a-y_c)} & \text{otherwise} , \end{cases} \tag{3.6}$$

where q_T is the corresponding quantization step size used in quantization to concentrate the data range. Compared with (3.5), another benefit from (3.6) lies in dividing the main portion into two parts: zero and non-zero coefficients. Based on the maximum likelihood estimator, $p_T(0)$ in (3.6) is always equals to the actual probability of zero coefficient. As such, this will model the main portion better, which are dominant among all coefficients.

Similar to (3.6), the dynamic range of data is again divided into three sub-ranges in BGTCM. The difference lies in that the tails are modelled by another truncated geometric distribution. This follows the fact that the larger the absolute values of coefficients, the less probabilities are. Specifically, by modifying the tail part in (3.6), the data is modelled by the revised distribution with the following pdf

$$p_T(y) = \begin{cases} bp & \text{if } y = 0 \\ b(1-p)\frac{1}{2}\left[1 - e^{-\frac{q_T}{\lambda_1}}\right] e^{-\frac{q_T}{\lambda_1}(|y|-1)} \Big/ \left(1 - e^{-\frac{q_T}{\lambda_1} y_c}\right) & \text{if } 1 \leq |y| \leq y_c \\ (1-b)\frac{1}{2}\left[1 - e^{-\frac{q_T}{\lambda_2}}\right] e^{-\frac{q_T}{\lambda_2}(|y|-y_c-1)} \Big/ \left(1 - e^{-\frac{q_T}{\lambda_2}(a-y_c)}\right) & \text{otherwise} , \end{cases} \tag{3.7}$$

It is worthwhile to point out that the quantization step size q_T is introduced in (3.6) and (3.7) in order to provide a generic model for both original and quantized transformed coefficients. When considering the R-Q relation based on the BGTCM for a batch of quantized coefficients, q_T is the same as the QS used to quantize coefficients in compression. When the D-Q relation is considered, due to the requirement to know the original coefficients before quantization, q_T is set as 1. Moreover, due to the fact that the larger the values coefficients are the less the probabilities of them are, compared to uniform distribution, a second geometric distribution is more suitable to model the tails. Therefore, in the future analysis and experiments, the BGTCM in (3.7) will be employed.

To justify the proposed BGTCM is a good fit for the distribution of integer transformed coefficients in HEVC, BGTCM, Laplacian and zero-mean Cauchy are employed to model coefficients in blocks whose coded block flags (cbf) equals to 1, i.e., when encoding coefficients there are at least one quantized coefficient whose absolute value are not zero in this block. Nine sequences with various resolutions and motion complexity are encoded with HM-9.0 all-intra main (AI-Main) and low-delay main (LD-Main) for intra and inter experiments. Under AI-Main configuration, ten frames from each sequence are picked up randomly and encoded with QPs from 5 to 50 with 5 as the increase step. In inter case, frame 0 to 4 are encoded with LD-Main configuration. Frame 0 is encoded as an intra frame with a fixed QP, which is 20. Other inter frames are encoded with different QPs from from 5 to 50 with 5 as the increase step as well.

After deriving the transformed coefficients after mode selection from blocks with cbf=1 and encoding each frame, based on maximum likelihood estimation, parameters for Laplacian distributions are estimated as

- μ_L is the median of $\{y_1, y_2, y_3..., y_N\}$
- $\lambda_L = \frac{1}{N}\sum_{i=1}^{N}|y_i - \mu|$

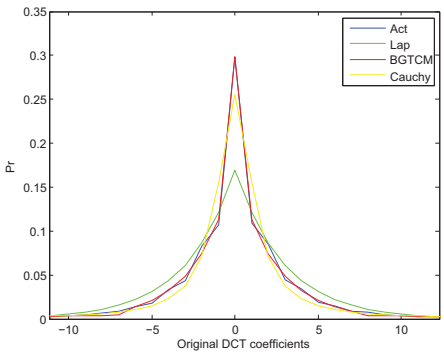
Since 0 is the dominant coefficient, Cauchy distribution is assumed as zero mean. Therefore, the pdf of Cauchy distribution is

$$p_c(x) = \frac{b_c}{\pi(b_c^2 + x^2)} \tag{3.8}$$

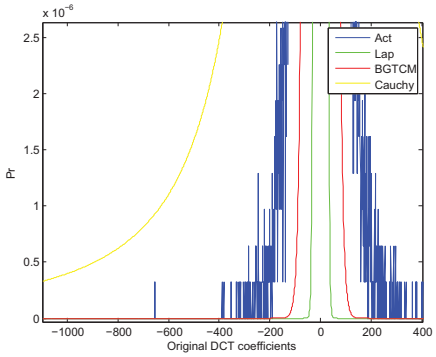
where the parameter b_c is obtained based on maximum likelihood estimation as well.

Figure 3.1 and Figure 3.2 show examples of modelling the distribution of the original and quantized DCT coefficients from the first frame of the sequence Cactus under AI-Main with different QPs by using BGTCM, Laplacian and Cauchy distribution.

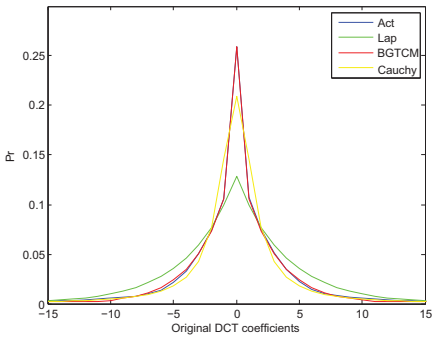
Figure 3.3 and Figure 3.4 show examples of modelling of distribution of original and quantized DCT coefficients from frame 2 of the sequence Cactus with LD-Main configuration respectively.



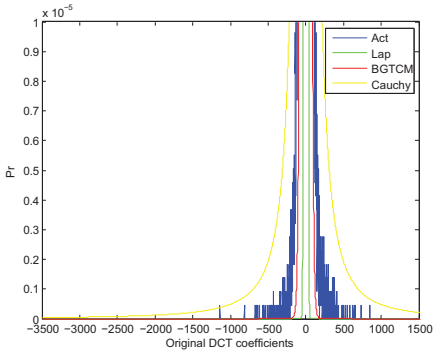
(a) Middle QP=10



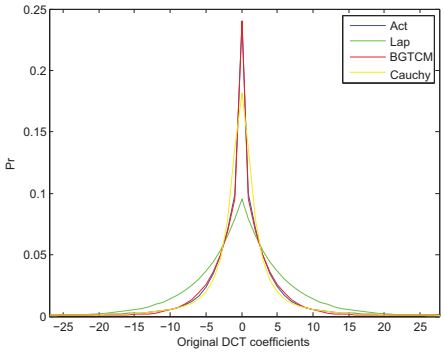
(b) Tail QP=10



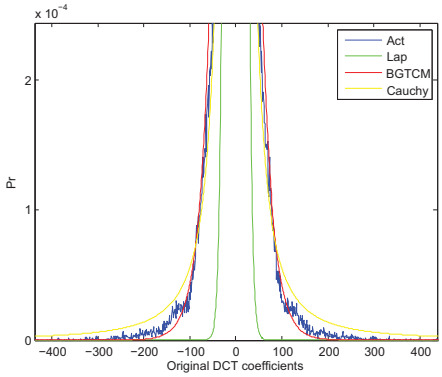
(c) Middle QP=25



(d) Tail QP=25

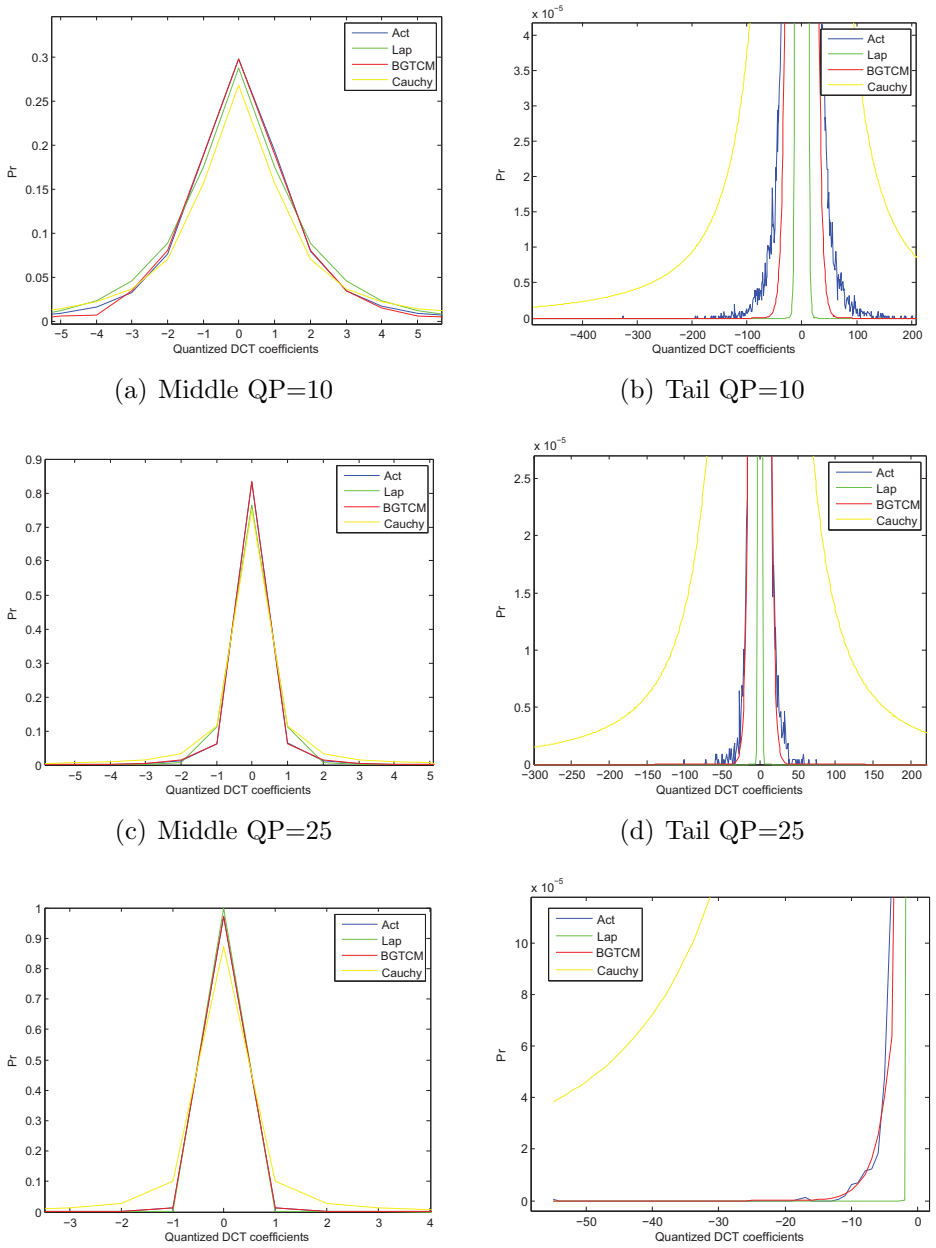


(e) Middle QP=40



(f) Tail QP=40

Figure 3.1: Modelling of Distribution of Original Coefficients in Blocks of cbf=1 in Frame 0 of Cactus Encoded under AI-Main



(a) Middle QP=10

(b) Tail QP=10

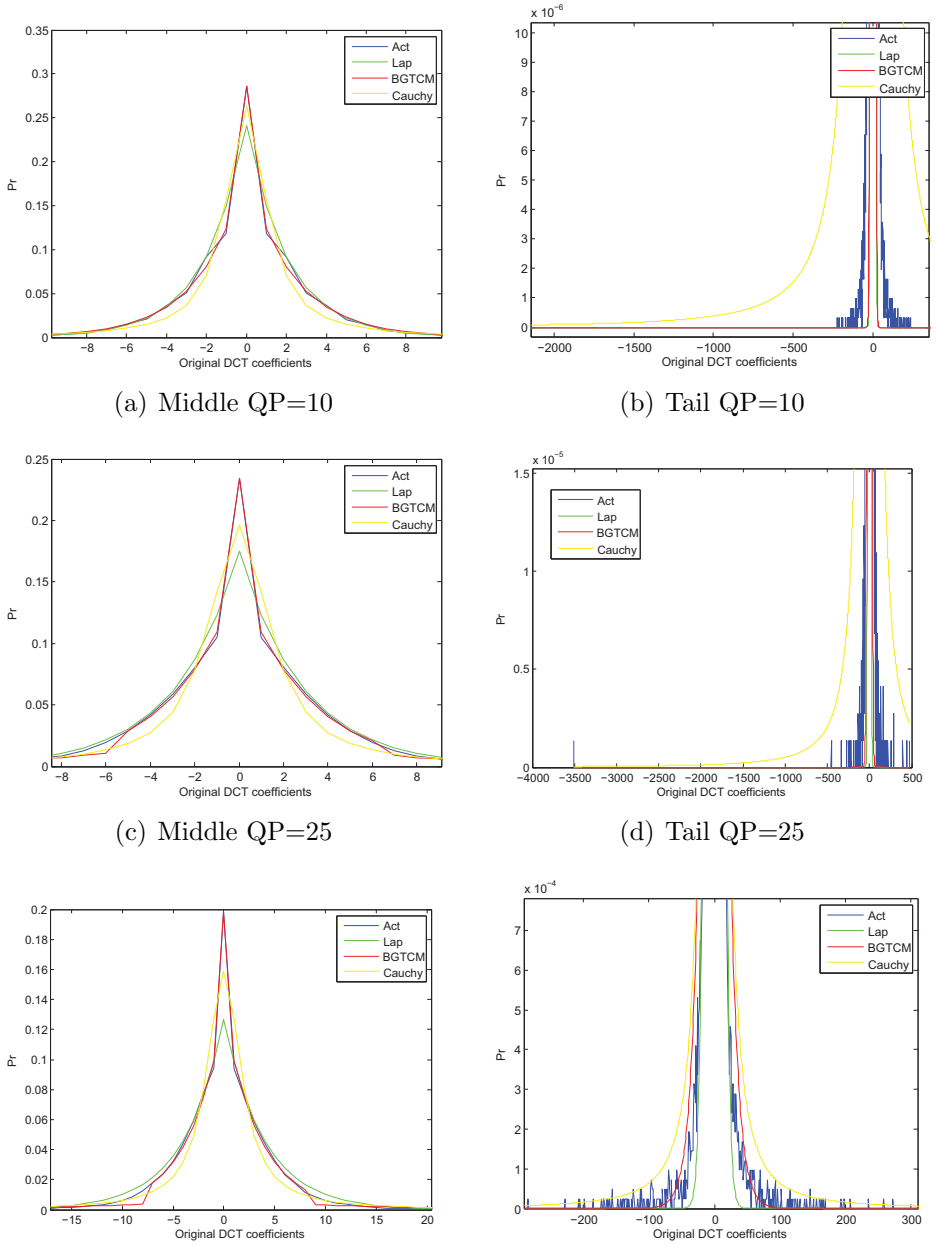
(c) Middle QP=25

(d) Tail QP=25

(e) Middle QP=40

(f) Tail QP=40

Figure 3.2: Modelling of Distribution of Quantized Coefficients in Blocks of cbf=1 in Fame 0 of Cactus Encoded under AI-Main



(a) Middle QP=10

(b) Tail QP=10

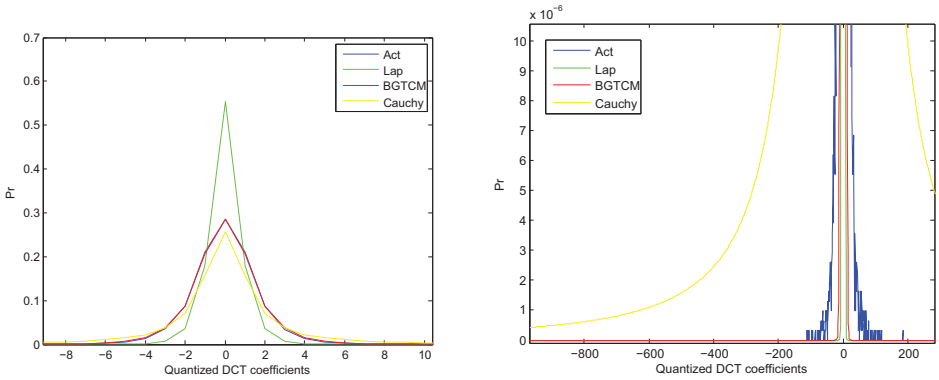
(c) Middle QP=25

(d) Tail QP=25

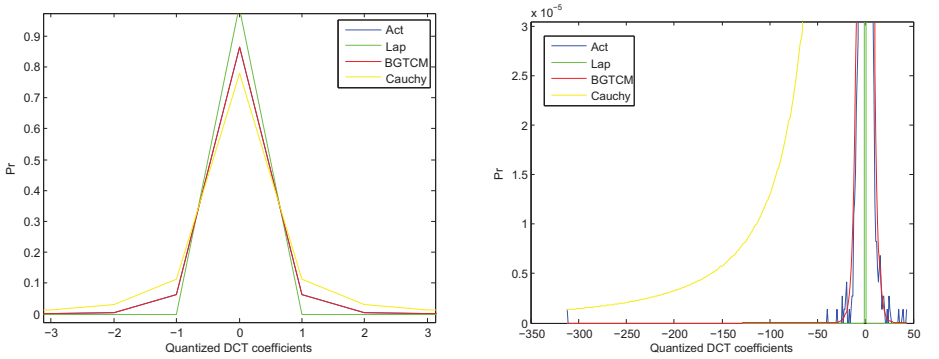
(e) Middle QP=40

(f) Tail QP=40

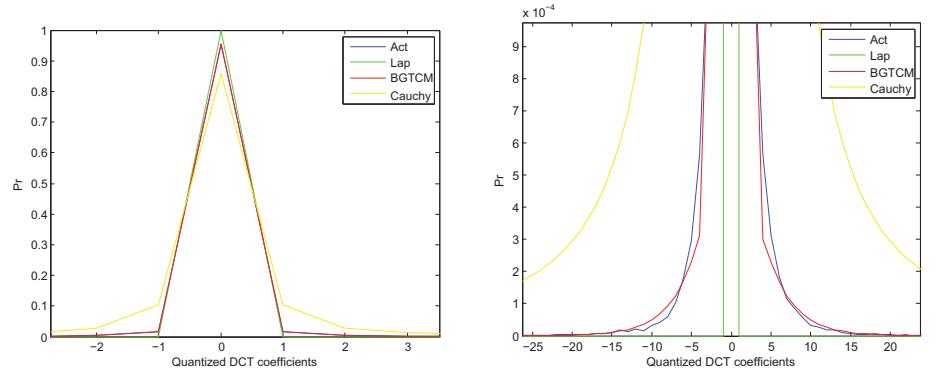
Figure 3.3: Modelling of Distribution of Original Coefficients in Blocks of cbf=1 in Frame 1 of Cactus Encoded under LD-Main



(a) Middle QP=10 (b) Tail QP=10



(c) Middle QP=25 (d) Tail QP=25



(e) Middle QP=40 (f) Tail QP=40

Figure 3.4: Modelling of Distribution of Quantized Coefficients in Blocks of cbf=1 in Frame 1 of Cactus Encoded under LD-Main

In these figures, actual distribution is derived by normalizing the histogram of either original or quantized coefficients. It is worthy to point out that, for save of convenience to compare distribution models, all density curves are illustrated in a continuous way, however, the actual pdf and BGTCM density should be discrete because all coefficients are integers. According to these figures, compared with Laplacian and Cauchy distribution, BGTCM has a better match for middle part of the histograms. For the tails, Cauchy distribution drops less slowly than BGTCM and Laplacian indeed. However, based on the histograms of encoded coefficients, this dropping trend is too slow to follow the real histograms. To be specific, at tails, Cauchy distribution over-estimates the probabilities of large coefficients. This phenomenon is obvious for quantized coefficients especially. Furthermore, to solve the maximum likelihood estimator for Cauchy distribution is complicated and takes much longer time than the other two models.

Besides the histograms of actual pdf and various models shown above, χ^2 and KL-divergence tests are employed to compare modelling accuracy among BGTCM, Laplacian and Cauchy distributions.

Given a sequence of original or quantized coefficients with alphabet set set $\{y_1, y_2, y_3..., y_N\}$, where $y_i \neq y_j$ for $i \neq j$, $p(y_i)$ is the empirical probability of y_i based on the normalized histogram. $p_T(y_i)$, $p_L(y_i)$ can $p_C(y_i)$ are model probabilities obtained by using BGTCM, Laplacian and Cauchy distributions respectively. The KL divergences of the model BGTCM, Laplacian and Cauchy from the observations are computed as

$$\begin{cases} \mathrm{KL}_T &= \sum_{i=1}^{N} p(y_i) \ln \frac{p(y_i)}{p_T(y_i)} \\ \mathrm{KL}_L &= \sum_{i=1}^{N} p(y_i) \ln \frac{p(y_i)}{p_L(y_i)} \\ \mathrm{KL}_C &= \sum_{i=1}^{N} p(y_i) \ln \frac{p(y_i)}{p_C(y_i)} \end{cases} \tag{3.9}$$

where 0ln0 is defined as 0. The χ^2 test is defined as

$$\begin{cases} \chi^2_T &= \sum_{i=1}^{N} \frac{n[p(y_i)-p_T(y_i)]^2}{p_T(y_i)} \\ \chi^2_L &= \sum_{i=1}^{N} \frac{n[p(y_i)-p_L(y_i)]^2}{p_L(y_i)} \\ \chi^2_C &= \sum_{i=1}^{N} \frac{n[p(y_i)-p_C(y_i)]^2}{p_C(y_i)} \end{cases} \tag{3.10}$$

where n is the total number of samples. The smaller the values of KL and χ, the better the modelling is.

Table 3.1: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Original Coefficients from Intra Frames under AI-Main

Sequence	QP	OC_{χ^2}			$OC_{KL} \times 10^{-2}$		
		χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	27392	633070	3529195	3.02	3.53	3.32
	25	4312	51179	67534	4.60	4.85	5.11
	35	1596	27143	25007	5.23	5.4	5.35
Four People	15	2716	113447	235080	0.82	0.94	1.13
	25	106	4291	2443	1.69	1.80	1.79
	35	22	855	94	5.92	8.12	6.11
Race Horses C	15	2029	160668	63045	1.39	1.92	1.79
	25	1666	80409	23570	1.13	1.57	1.24
	35	1455	29246	6134	2.58	2.73	2.68

Table 3.2: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Quantized Coefficients from Intra Frames under AI-Main

Sequence	QP	QC_{χ^2}			$QC_{KL} \times 10^{-4}$		
		χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	10845	184697	289848	36.44	48.39	46.42
	25	1771	162867	249311	14.41	33.52	24.39
	35	78	184697	124953	1.91	2.45	2.22
Four People	15	1808	360724	153757	11.49	16.91	13.44
	25	55	288793	12147	7.86	13.53	8.11
	35	17	715	89	24.33	44.38	30.53
Race Horses C	15	5107	80343	38256	91.97	103.34	111.94
	25	391	10217	41926	10.31	27.32	10.18
	35	24	19246	8699	5.55	11.94	9.62

Table 3.3: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Original Coefficients from Inter Frames under LD-Main

Sequence	QP	OC_{χ^2}			$OC_{KL} \times 10^{-2}$		
		χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	36320	933033	807946	2.67	2.89	2.73
	25	6822	234406	148066	3.14	3.46	3.24
	35	7536	226892	106617	4.51	4.84	4.61
Four People	15	16047	671770	2612675	1.55	1.84	1.75
	25	35161	653683	712375	2.87	3.12	3.27
	35	32180	646594	316113	2.91	3.47	3.41
Race Horses C	15	16908	482353	102928	1.85	2.58	2.25
	25	26943	437712	85945	3.03	4.07	3.73
	35	20150	376349	71900	2.93	3.77	3.14

Table 3.4: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Quantized Coefficients from Inter Frames under LD-Main

Sequence	QP	QC_{χ^2}			$QC_{KL} \times 10^{-4}$		
		χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	16641	603156	262296	58.98	61.09	62.99
	25	2892	2654433	586217	10.15	12.13	10.14
	35	377	294238	610404	2.13	5.03	3.11
Four People	15	11133	638696	118430	100.64	144.96	120.63
	25	916	221829	173737	14.64	19.50	16.61
	35	336	809212	174515	4.60	9.92	6.53
Race Horses C	15	9275	190573	18367	116.84	139.81	126.82
	25	662	794440	60005	23.96	27.76	25.91
	35	264	648899	87994	4.83	8.21	5.71

Table 3.1 to Table 3.4 show χ^2 and KL-divergence testing results when using BGTCM, Laplacian and Cauchy to model distribution of original and quantized coefficients from inter and intra frames encoded with different QPs in HEVC. In most cases, the values of KL_T and χ^2_T are much smaller than those values for Laplacian and Cauchy distributions. For instance, those values for Laplacian and Cauchy distributions are tens and hundreds larger than the values of KL_T and χ^2_T for BGTCM.

Clearly, the BGTCM can model is significantly better than Laplacian and Cauchy model for modelling the transformed coefficients in HEVC. The BGTCM matches the distribution of transformed coefficients very accurately, especially for intra frame.

Above experimental results demonstrate that BGTCM is better than Laplacian density to model both original and quantized DCT coefficients in both inter and intra frames in HEVC.

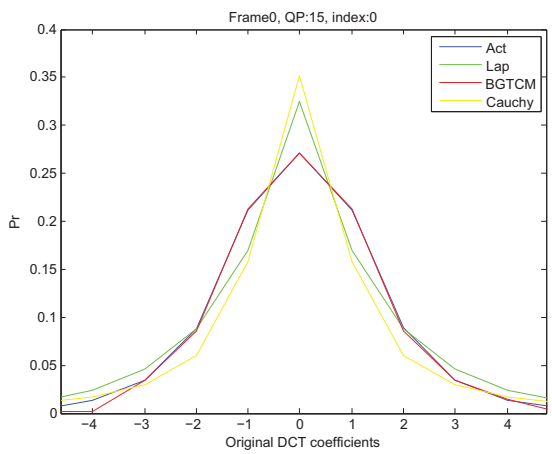
3.1.2 Distribution of Transformed Coefficients from the Same Scan Indices among Different CGs in a Frame

In HEVC, the quantized coefficients are entropy-coded in groups of 16 coefficients for each TU. Each group is a CG. When encoding quantized coefficients information with CABAC, which context models will be used for levels and significant bits of coefficients are determined by the scan indices of coefficients in their CGs. Therefore, it will be useful if the models are established for coefficients with the same scan indices. Since the scan in each CG is from bottom-right to up-left, i.e., bottom-right coefficients are indexed as 0 and up-left coefficients are indexed as 15. According to the energy concentration of transforms, in each CG, generally speaking, coefficients at high indices have higher energy than coefficients at low indices.

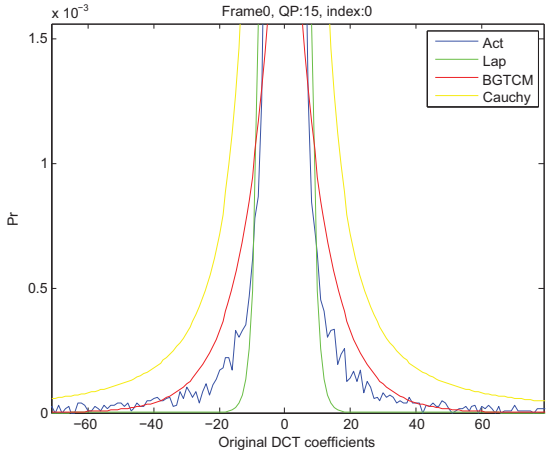
The only difference of testings in this section, compared with testings in last section, is performing models on coefficients at the same indices.

Figure 3.5 and 3.6 show modelling of distributions of original and quantized coefficients at scan index 0 when performing BGTCM, Laplacian and Cauchy. Figure 3.7 and 3.8 show modelling of performing BGTCM, Laplacian and Cauchy on coefficients at scan index 15. The coefficients are from the first frame of sequence Cactus encoded in intra mode with QP set as 15 and 35. Similar to last section, χ^2 and KL-divergence tests are performed. Results are shown in Table 3.5 and Table 3.6 correspondingly.

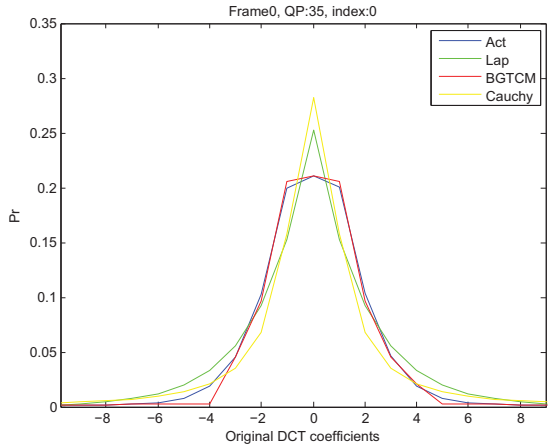
Then D-Q and R-Q relations based BGTCM will be introduced in the following sections.



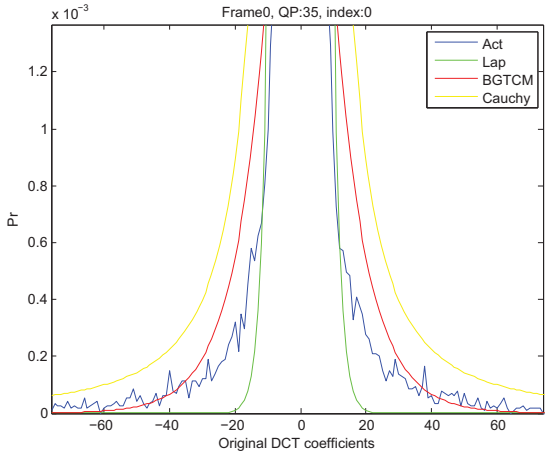
(a) Middle QP=15



(b) Tail QP=15

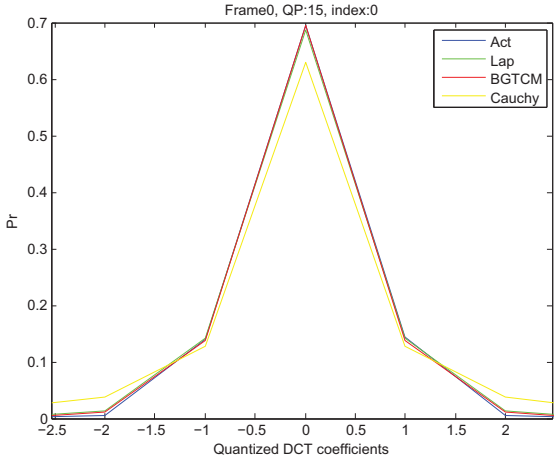


(c) Middle QP=35

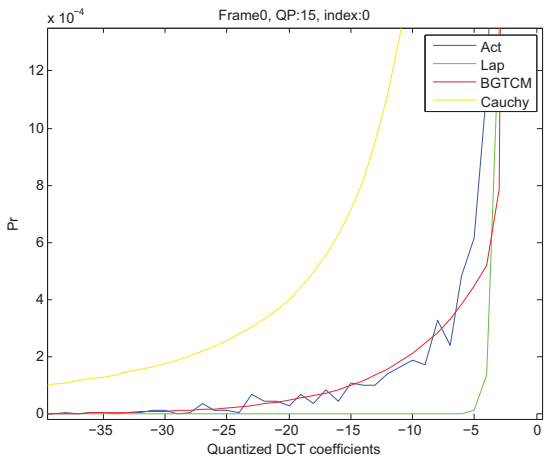


(d) Tail QP=35

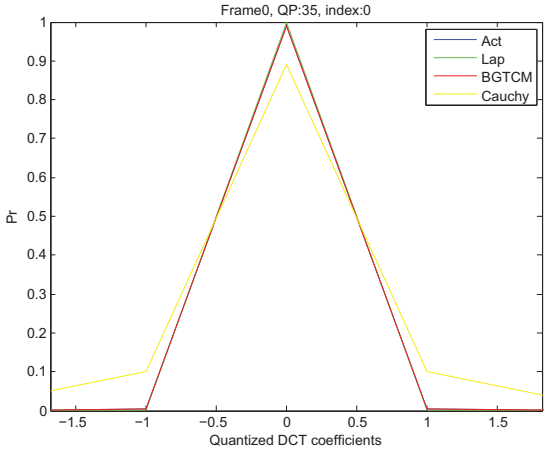
Figure 3.5: Modelling of Distribution of Original Coefficients at Scan Index 0 in Blocks of cbf=1 in Frame 1 of Cactus Encoded under AI-Main



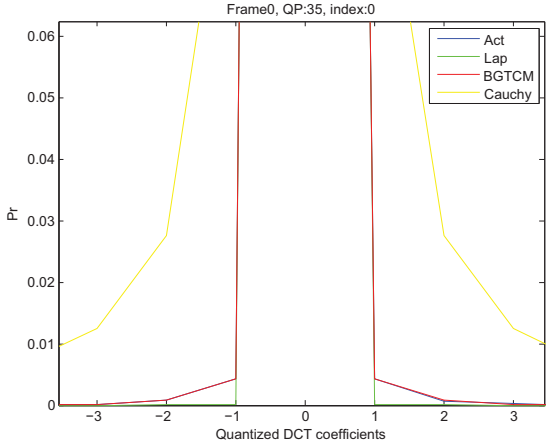
(a) Middle QP=15



(b) Tail QP=15

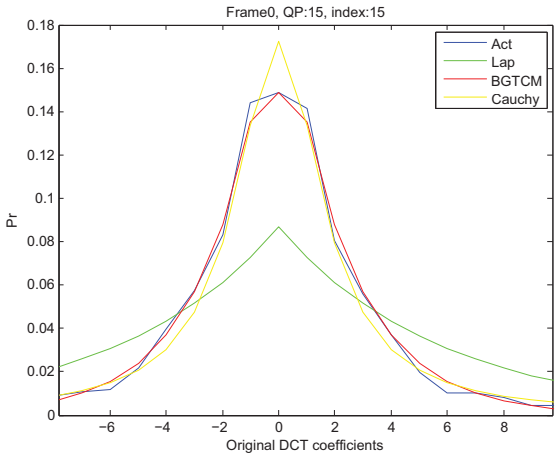


(c) Middle QP=35

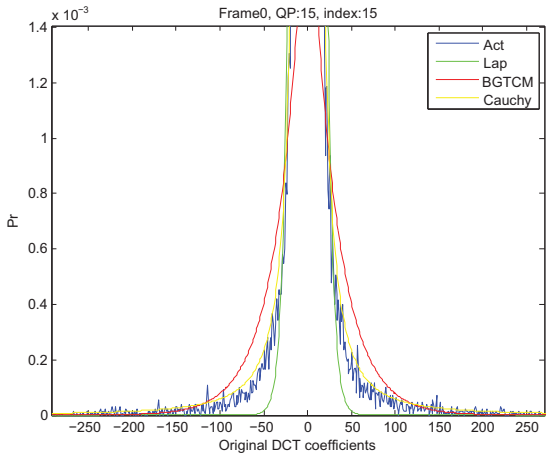


(d) Tail QP=35

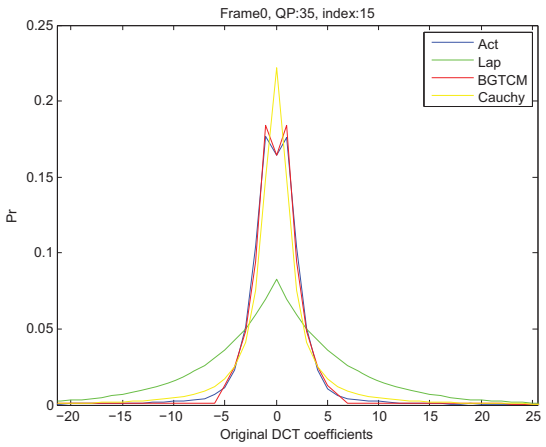
Figure 3.6: Modelling of Distribution of Quantized Coefficients at Scan Index 0 in Blocks of cbf=1 in Frame 1 of Cactus Encoded under AI-Main



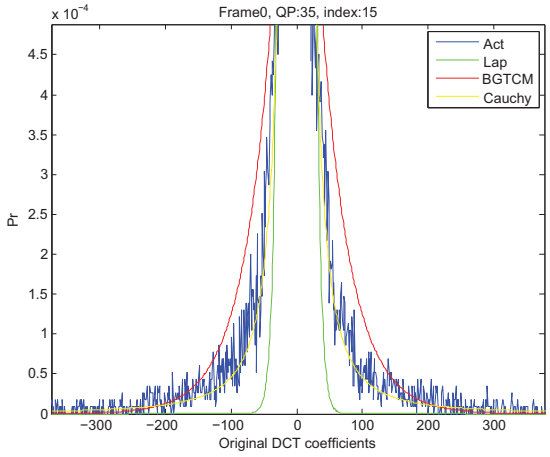
(a) Middle QP=15



(b) Tail QP=15

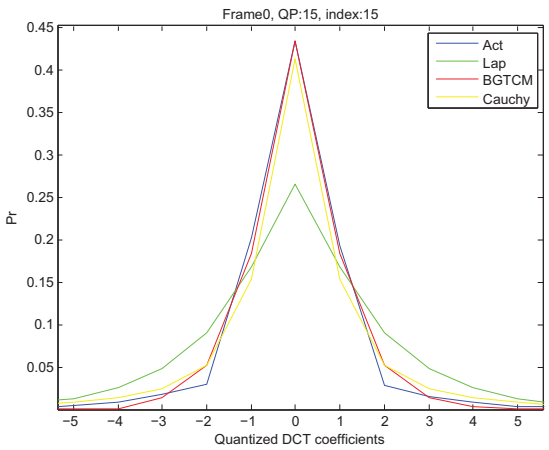


(c) Middle QP=35

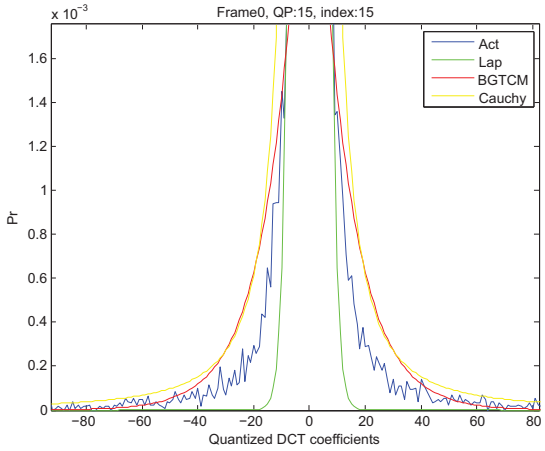


(d) Tail QP=35

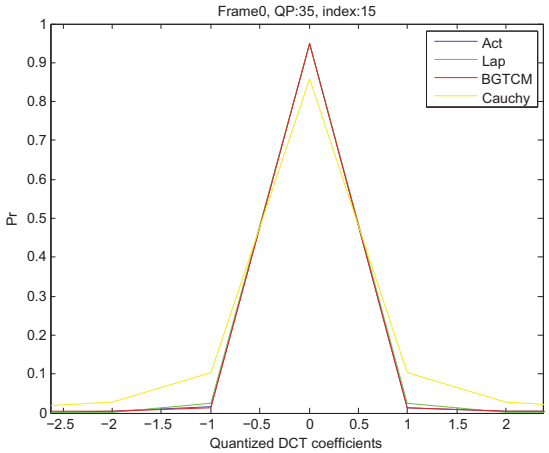
Figure 3.7: Modelling of Distribution of Original Coefficients at Scan Index 15 in Blocks of cbf=1 in Frame 1 of Cactus Encoded under LD-Main



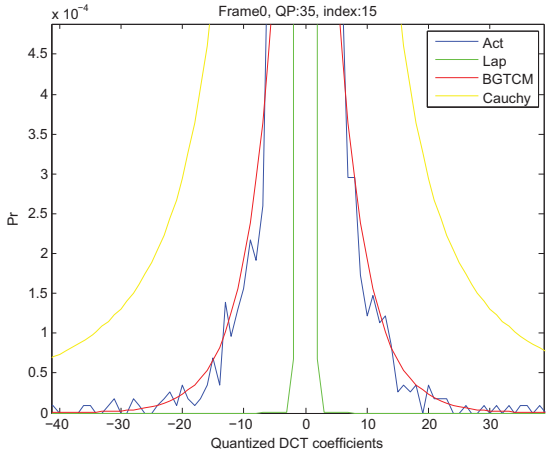
(a) Middle QP=15



(b) Tail QP=15



(c) Middle QP=35



(d) Tail QP=35

Figure 3.8: Modelling of Distribution of Quantized Coefficients at Scan Index 15 in Blocks of cbf=1 in Frame 1 of Cactus Encoded under LD-Main

Table 3.5: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Original Coefficients in Different Scan Indices from Intra Frames under AI-Main

Sequence	QP	Index	OC_{χ^2}			$OC_{KL} \times 10^{-2}$		
			χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	0	10281	31576	45364	3.56	7.73	2.40
		15	4299	20558	8795	3.22	16.92	4.84
	25	0	176	3673	10423	1.27	-0.22	1.17
		15	1315	27783	2943	2.01	30.93	2.72
	35	0	294	4769	10993	1.02	1.26	1.93
		15	552	40624	3936	1.84	26.35	3.38
Four People	15	0	360	3082	39285	1.56	7.31	4.19
		15	1243	25973	13129	2.42	42.3	5.51
	25	0	184	2460	18698	1.97	10.73	5.49
		15	5940	37485	38274	8.24	52.3	15.2
	35	0	173	2637	8612	2.64	14.56	4.08
		15	8045	52264	99754	10.40	75.21	26.54
Race Horses C	15	0	262	3650	1517	0.96	17.33	1.67
		15	2206	9530	8567	7.05	25.67	6.81
	25	0	53	5105	1877	1.28	21.03	1.57
		15	4002	12362	13781	10.85	30.22	11.24
	35	0	84	3340	1655	1.80	16.11	1.86
		15	5369	16501	25972	20.35	42.20	22.18

Table 3.6: Results of χ^2 and KL When Using BGTCM, Laplacian and Cauchy Distributions to Model Distribution of Quantized Coefficients in Different Scan Indices from Intra Frames under AI-Main

Sequence	QP	Index	QC_{χ^2}			$QC_{KL} \times 10^{-2}$		
			χ^2_T	χ^2_L	χ^2_C	KL_T	KL_L	KL_C
Basketball Drive	15	0	282	11315	12447	0.17	2.93	0.19
		15	2610	12547	4337	2.81	14.70	2.86
	25	0	13767	3324419	292876	0.02	0.21	3.02
		15	321	94158	15849	0.42	26.07	-6.05
	35	0	47	$2 * 10^7$	26767	10^{-3}	0.10	10.06
		15	87	$2 * 10^8$	26309	0.17	5.32	-1.30
Four People	15	0	619	20778	6460	3.01	38.21	3.24
		15	450	294441	27456	0.66	11.37	-2.06
	25	0	27	10^7	38236	0.08	9.04	3.68
		15	138	17035	6243	0.95	32.84	-6.11
	35	0	32	10^7	30357	0.03	7.46	7.34
		15	70	303628	15567	0.33	20.7	-5.09
Race Horses C	15	0	343	7119	4142	0.91	14.66	-2.55
		15	1204	7446	3999	3.62	22.37	3.80
	25	0	14	45088	11571	10^{-3}	10.59	-3.13
		15	43	5129	1468	1.05	18.33	-1.85
	35	0	5	8940	1431	10^{-3}	5.21	3.88
		15	59	14910	5834	0.26	10.43	-3.19

3.2 Summary

In this chapter, the performance of modelling transformed coefficients in HEVC are examine by using BGTCM, Laplacian and Cauchy distributions, where the later two are two widely used models. At first, the distribution of all transformed coefficients from a whole frame is modelled with those three models. Since in TUs with different sizes, the transformed coefficients are grouped into 4×4 CGs and coefficients at the same scan indices using the same context model during encoding, models are performed based on scan indices.

Histograms are derived to show the modelling performance intuitively. Also, KL-divergence and χ^2 testing are performed to compare the performance numerically. The results show that, BGTCM outperforms the other two models when modelling both original and quantized transformed coefficients in HEVC.

Chapter 4

The Rate and Distortion Optimized Rate Control Algorithm in HEVC with BGTCM

In this chapter, with the assumption that the distribution of the transformed coefficients in HEVC satisfy BGTCM, the corresponding R-Q and D-Q models are derived. Based on R-Q and D-Q models, a rate and distortion optimized rate control algorithm is developed, which is used to achieved the bit rate requirement and reduce the distortion. Compared with HEVC codec and a newly proposed method, the proposed method in this thesis can achieve less rate fluctuation in terms of rate variance and some gain in RD performance in terms of BD-rate.

4.1 Problem Formulation

As we described in Chapter 1 and 2, a predetermined bit rate is desired in some video coding applications where the bandwidth of the channels and the capacity of storage are limited. However, it is also important to maintain the video quality at the same time. Therefore, the problem can be described as minimizing the distortion, when given a target bit rate for each frame.

$$Q^* = \arg\min\{D(Q)||R(Q) \leqslant R_T\}, \tag{4.1}$$

where Q denotes a quantization step size. R_T and $R(Q)$ are the target bits and the number of bits generated by encoding the current frame with Q, respectively. $D(Q)$ means the distortion resulting from encoding when Q is used. To determine a Q providing R_T, this problem can be transferred to a R-D optimization problem. Therefore, the quantization step size Q for the current frame can be determined as follows:

$$Q^* = \arg\min\{J_{cost} = D(Q) + \lambda \cdot R(Q),\ when\ R(Q) \leqslant R_T\}, \tag{4.2}$$

where J_{cost} denotes the total R-D cost when the current frame is encoded with Q. λ is the Lagrange multiplier. To determine a Q providing R_T, the relationship between $D(Q)$ and Q, $R(Q)$ and Q should be formulated with a D-Q model and a R-Q model, respectively. Not only the predetermined target bit rate for each frame, but also the condition related to the buffer fullness determine the R_T for current frame.

Because HEVC codec has adopted a quadtree-based CU, PU, and TU structure, the transformed coefficients (DCT coefficients) contain less energy compared to H.264/AVC. Some well-developed method to fit distribution of transformed coefficients in H.264 may not work so well in HEVC. Therefore, it is desirable to find a model to better represent the distribution of DCT coefficients in HEVC. In the proposed method, the pdfs of DCT coefficients in HEVC will be modelled according to the skipped and non-skipped CUs separately. In the skipped CU, all the transformed coefficients will be quantized to zero, whereas in the non-skipped CU, at least one coefficient will not be quantized to zero. In the next sections, the R-Q and D-Q models for HEVC based on integer BGTCM are derived.

4.2 D-Q Model Based on BGTCM in HEVC

The distortion D is calculated as the mean square error (MSE) between original and reconstructed sequence. MSE can be computed between original and reconstructed DCT coefficients in frequency domain, as shown as below

$$D = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \, p(y_i) \tag{4.3}$$

where $\hat{y}_i$ is the reconstructed value of y_i and $p(y_i)$ is the empirical probability of y_i.

In our algorithm, actual MSE is estimated based on modelling of distribution of coefficients, D_m.

Therefore, the distortion based on Laplacian model is

$$D_L = 2\lambda_L^2 + \frac{2s}{\lambda_L \left(e^{-\frac{s}{2\lambda_L}} - e^{\frac{s}{2\lambda_L}} \right)} \tag{4.4}$$

Distortion based on Cauchy density can be computed in a similar way.

For BGTCM,

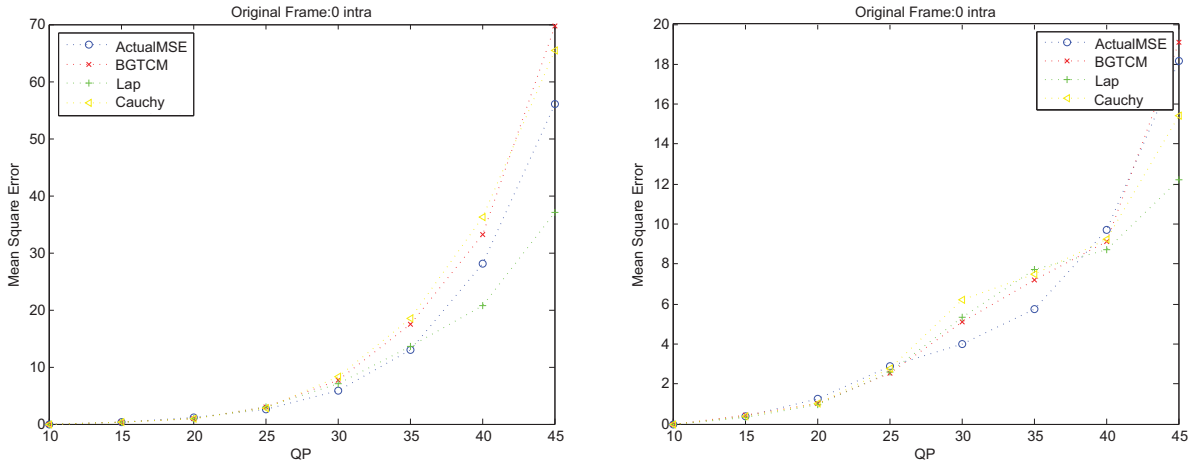
$$D_T = 2 \sum_{y=1}^{a} p_T(y) \left(y - s \lfloor \frac{y}{s} + d \rfloor \right)^2 \tag{4.5}$$

where d is the size of the dead-zone. Note that y in (3.7) is the values of integer DCT coefficients and the quantization step size q_T in (3.7) is set as 1 due to the density of original DCT coefficient is employed when distortion is considered.

Since the ΔQP-limiter in the rate controller is employed to limit the QPs between neighbouring units, in our methods introduced later, the D-Q model in (4.5), which is in the form of summation, will not be approximated to some simple expression. At most 5 QPs are considered for a unit, due to the limited range of y in BGTCM (3.7), D_T for each QP can be computed quickly.

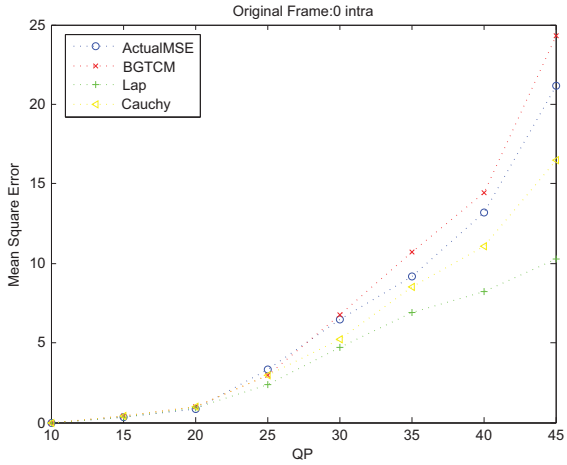
Figure 4.1 shows modelling of D-Q relation in intra frames from four sequences. Each frame is encoded in several QPs from 10 to 45 with 5 as a step. Original DCT coefficients are grouped into two set depends on whether the corresponding blocks are skipped or not. Then parameters of Laplacian and BGTCM densities are derived for each group. Based on the PSNR value, the actual MSE is computed as

$$\text{MSE} = \left[\frac{20 \log_{10}(255) - \text{PSNR}}{10} \right]^{10} \tag{4.6}$$

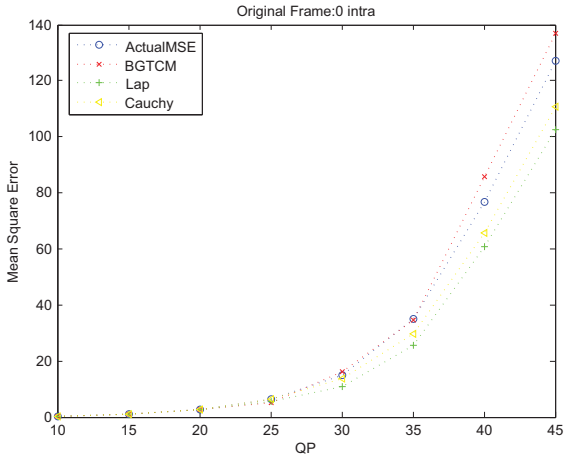


(a) BasketballPass

(b) Kimono

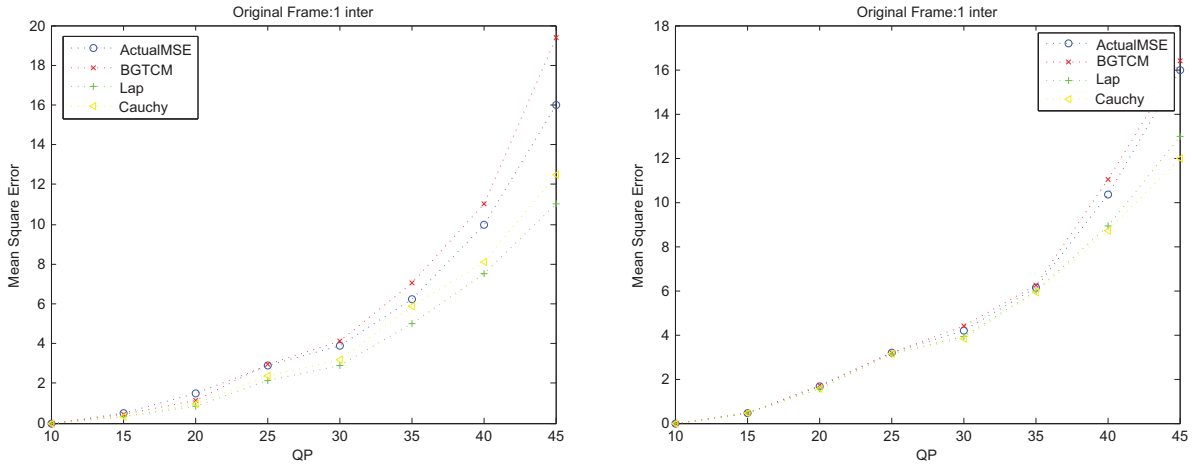


(c) BasketballDrive



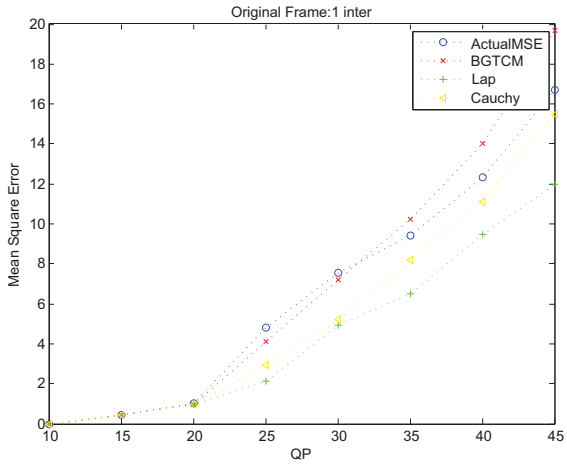
(d) Race Horses C

Figure 4.1: D-Q Model Based on Distribution of Original Coefficients for Frame 0 in Four Sequences Encoded under AI-Main with QP from 10 to 45

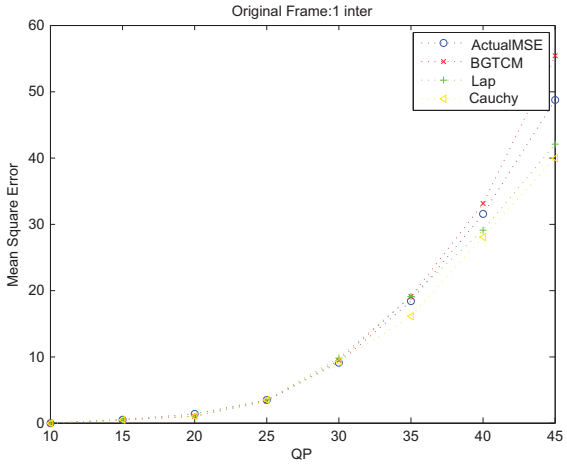


(a) BasketballPass

(b) Kimono



(c) BasketballDrive



(d) Race Horses C

Figure 4.2: D-Q Model Based on Distribution of Original Coefficients for Frame 1 in Four Sequences Encoded under LD-Main with QP from 10 to 45

Absolute (AD) and relative (RD) difference between actual and modelled MSE are also computed as

$$\begin{cases} \text{AD} = \sum_{\text{QP}} |\text{MSE}(\text{QP}) - D_M(\text{QP})|/8 \\ \text{RD} = \sum_{\text{QP}} |\text{MSE}(\text{QP}) - D_M(\text{QP})|/[8 \times \text{MSE}(\text{QP})] \times 100\% \end{cases} \tag{4.7}$$

where D_M, $M \in \{T, L, C\}$, is the modelled MSE. Since, eight QPs (10, 15, 20, 25, 30, 35,

Table 4.1: Results of Absolute and Relative Difference between Modelled MSE by BGTCM, Laplacian and Cauchy and Actual MSE from Intra and Inter Frames under AI-Main and LD-Main Respectively

Sequence	Type	AD_T	AD_L	AD_C	RD_T	RD_L	RD_C
Kimono	Intra	0.73	2.03	1.14	4.82	9.73	6.03
	Inter	0.36	0.78	0.81	2.42	4.12	4.59
Race Horses C	Intra	4.12	6.47	4.75	7.38	11.90	9.01
	Inter	3.29	3.87	4.11	3.97	4.01	5.69
Basketball Drive	Intra	1.21	2.52	1.54	6.95	10.62	9.11
	Inter	0.21	1.05	0.44	3.52	8.20	5.31
Basketball Pass	Intra	5.67	5.89	5.61	5.36	6.21	5.38
	Inter	0.63	0.85	0.69	4.21	7.58	6.96

40, 45) are employed, the sum of difference is divided by eight to obtain the average value. The results are shown in Table 4.1. Based on our observation, the DQ model based on BGTCM is better than the other two.

Similarly, D-Q relation curves in inter frames from those four sequences are shown in Figure 4.2. In this scenario, the second frame in each sequence is encoded with different QPs from 10 to 45 with 5 as a step. For all QPs, the reference frame, i.e., the first frame in each sequence is encoded in intra mode with fixed QP=25.

Figure 4.1 and Figure 4.2 demonstrate that, compared with Laplacian and Cauchy distribution, BGTCM can be employed to model D-Q relation more accurately.

4.3 R-Q Model Based on BGTCM in HEVC

Due to the tools in practical entropy coding, such as context models which considers the conditional distribution, the number of actually generated bit rate may differs a lot from the entropies of sources, in our methods, the bits generated by HEVC encoder can be estimated linearly by using entropy of the encoded signal as follows

$$\text{BPP} = \beta H_T \tag{4.8}$$

where β is the linear scaling factor between H_T and BPP and H_T is entropy of quantized transformed coefficients. Since coefficients from skipped blocks are not encoded into bit

streams, the entropy is computed only for coefficients from non-skipped blocks as

$$H_T = -p_T(0)\log_2(p_T(0)) - 2\sum_{y=1}^{a} p_T(y)\log_2(p_T(y)) \tag{4.9}$$

where y is the value of quantized transformed coefficients. Since the distribution of quantized DCT coefficients are employed when computing rate, the quantization step size q_T in (3.7) is equal to the actually step size used to quantize coefficients. According to (3.7), 4.9 can be written in a closed formula as

$$\begin{aligned}
H_T =& -(bp)\log_2(bp) - 2\sum_{y=1}^{y_c} p_T(y)\log_2(p_T(y)) - 2\sum_{y=y_c+1}^{a} p_i\log_2(p_T(y)) \\
=& -(bp)log_2(bp) \\
& - b(1-p)\{\log_2\left[\frac{b(1-p)(1-e^{-q/\lambda_1})}{2(1-e^{-qy_c/\lambda_1})}e^{q/\lambda_1}\right] - \frac{1}{\lambda_1(1-e^{-q/\lambda_1})} + \frac{y_c e^{-qy_c/\lambda_1}}{\lambda_1(1-e^{-qy_c/\lambda_1})}\} \\
& - (1-b)\{\log_2\left[\frac{(1-b)(1-e^{-q/\lambda_2})}{2\,(1-e^{-q(a-y_c)/\lambda_2})}e^{q(y_c+1)/\lambda_2}\right] - \frac{1}{\lambda_2(1-e^{-q/\lambda_2})} + \frac{a_{NS}e^{-q(a_{NS}-y_c)/\lambda_2}}{\lambda_2(1-e^{-q(a_{NS}-y_c)/\lambda_2})}\}
\end{aligned} \tag{4.10}$$

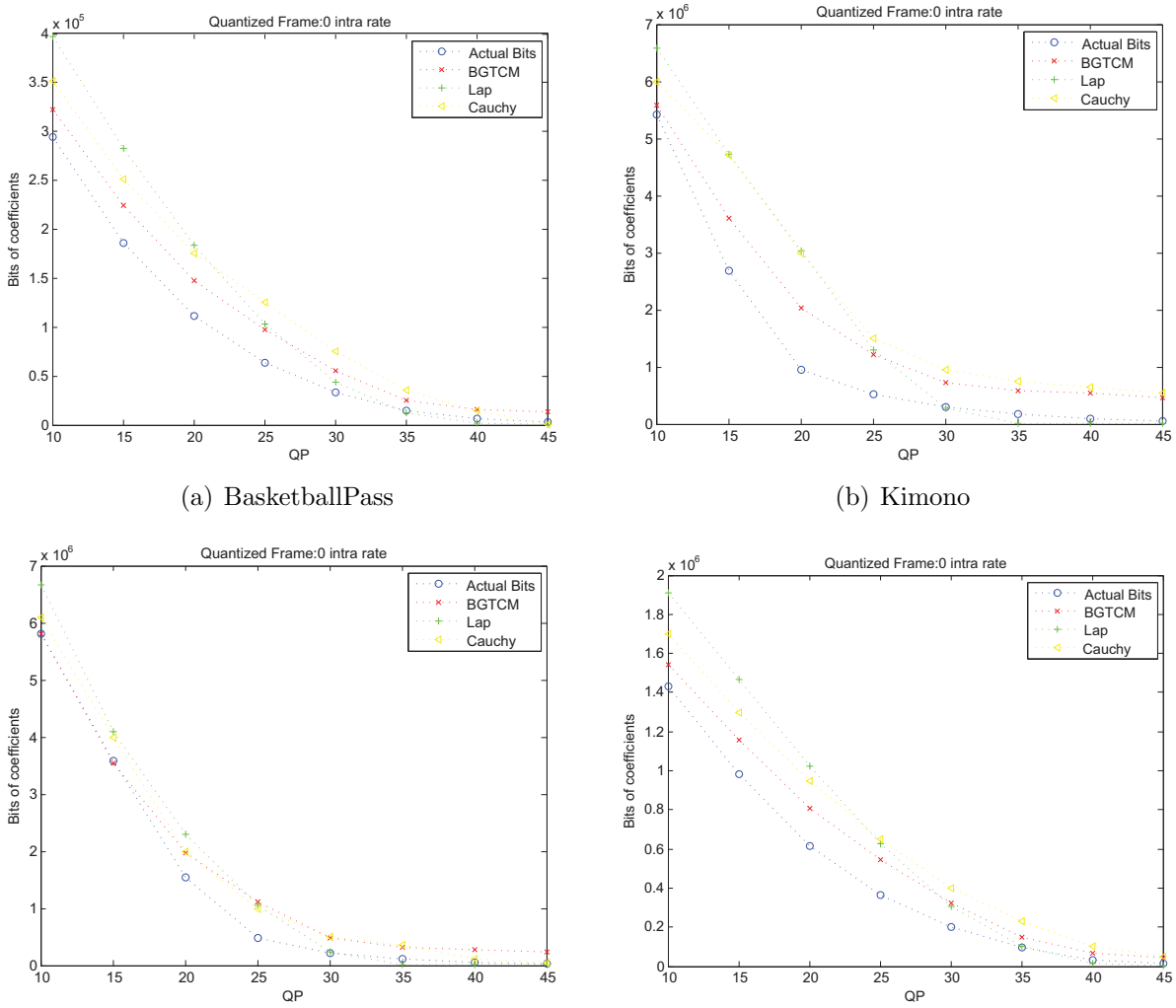
Similarly, the entropy of Laplacian distribution for quantized coefficients from non-skipped blocks is computed as

$$\begin{aligned}
H_L =& \left(1-e^{-\frac{q}{2\lambda_{L,NS}}}\right)\log_2\left(1-e^{-\frac{q}{2\lambda_{L,NS}}}\right) \\
& - 2e^{-\frac{q}{2\lambda_{L,NS}}}\cdot\log_2\left(\frac{e^{\frac{q}{2\lambda_{L,NS}}} - e^{-\frac{q}{2\lambda_{L,NS}}}}{2}\right) \\
& + 2e^{-\frac{q}{2\lambda_{L,NS}}}\cdot\frac{q}{\left(1-e^{-\frac{q}{\lambda_{L,NS}}}\right)\lambda_{L,NS}\ln 2}
\end{aligned} \tag{4.11}$$

and the entropy of Cauchy distribution for quantized coefficients from non-skipped blocks is computed as

$$\begin{aligned}
H_C =& -p_c(0)\log_2(p_c(0)) \\
& - 2\sum_{j=1}^{\infty}\int_{(j-1+d)q}^{(j+d)q} p_c(y)dy\log_2\left(\int_{(j-1+d)q}^{(j+d)q} p_c(y)dy\right)
\end{aligned} \tag{4.12}$$

Under the same testing conditions as those in D-Q model above, Figure 4.3 and Figure 4.4 show the rate curve of R-Q relation in intra and inter encoding mode. Those figures demonstrate that BGTCM can be employed to model R-Q relation more accurately.



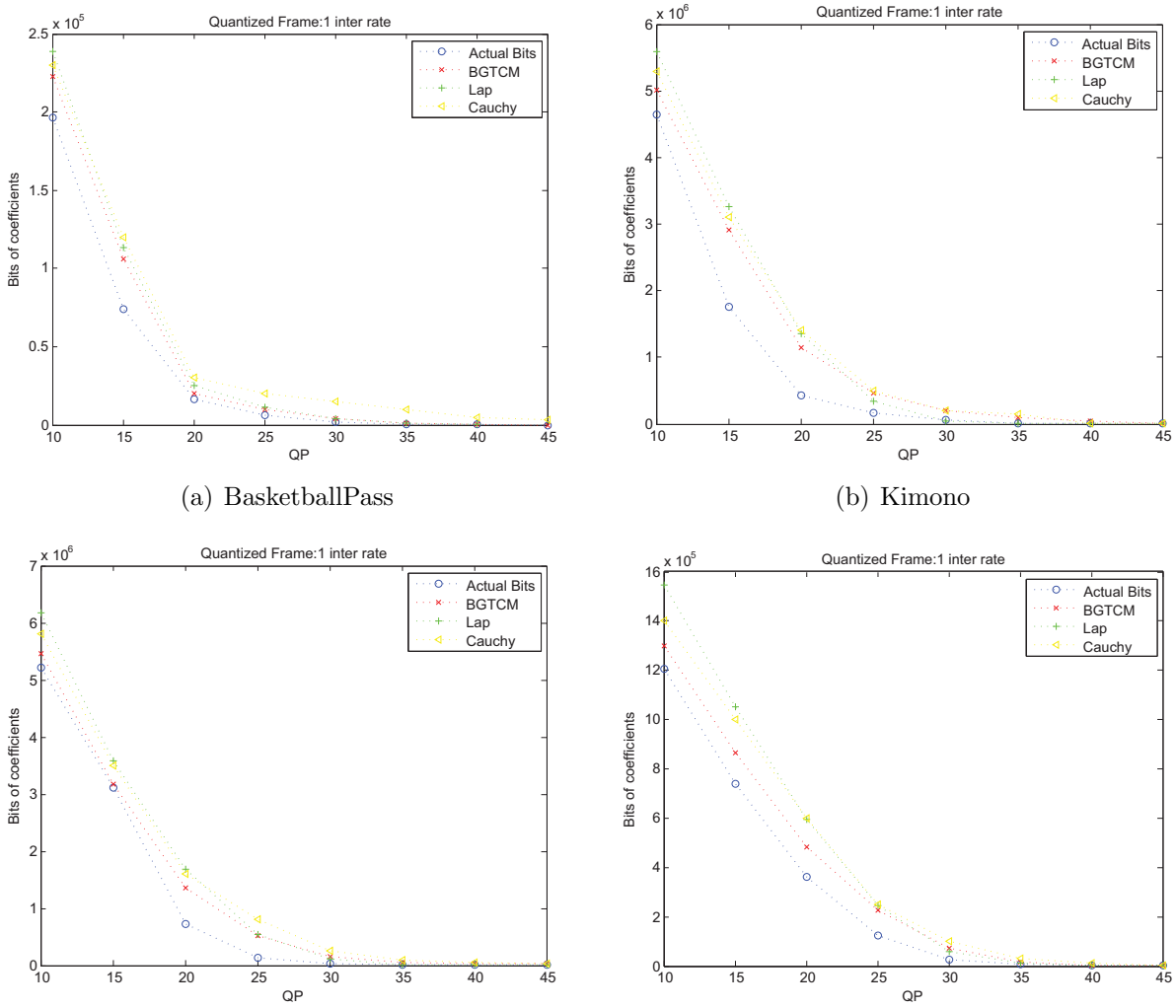
(a) BasketballPass

(b) Kimono

(c) BasketballDrive

(d) Race Horses C

Figure 4.3: R-Q Model Based on Distribution of Quantized Coefficients for Frame 0 in Four Sequences Encoded in Intra Mode with QP from 10 to 45



(a) BasketballPass

(b) Kimono

(c) BasketballDrive

(d) Race Horses C

Figure 4.4: R-Q Model Based on Distribution of Quantized Coefficients for Frame 1 in Four Sequences Encoded in Inter Mode with QP from 10 to 45

4.4 Rate Control Algorithm Based on BGTCM in HEVC

4.4.1 Parameter Estimation

In our method, BGTCM is employed to derive D-Q and R-Q models. As in HEVC, due to the complicated rate distortion optimization (RDO) procedure for motion estimation and mode selection in video coding, the residual data and its related information used for rate control cannot be generated until RDO is performed. On the other hand, RDO employs a pre-defined QP to do quantization in mode selection to choose an optimal compression mode in order to achieve the best coding efficiency. Therefore, the parameters of BGTCM based on actual DCT coefficients cannot be derived before that fixed QP is determined. To overcome this problem, the parameters at frame level are predicted as the same as those for the previous frame and those at block level are predicted as the same as those of the collocated block.

Scaling factors β used in R-Q model (4.8) is updated based on the least mean square algorithm. This happens after encoding one unit when the actual transform coefficients are available.

4.4.2 QP Decision

This section describes the algorithm used for QP decision for current unit (frame or CU), which is shown in Algorithm 1.

The inputs are QP for previous unit $\mathrm{QP}_{\mathrm{pre}}$, target bit rate R_T for current unit and the set of BGTCM parameters S. The target bits allocated for current unit is determined in the same way as HM. Since the actual TCM parameters could be obtained after encoding of current unit, these actual parameters will be used for rate control of future units. Specifically, current frame employs actual S from previous frame in display order and current CU employs actual S from its collocated CU. The output is the QP used to encode current unit.

At the beginning of the algorithm, if the target bit rate $R_T < 0$, i.e., the buffer is full currently, so the generated bit rate should be as little as possible. In this scenario, the QP will be set as $\mathrm{QP}_{\mathrm{pre}} + 2$ directly.

When there is space in the buffer, five candidate QPs are employed, i.e., $\mathrm{QP}_{\mathrm{pre}} - 2, \mathrm{QP}_{\mathrm{pre}} - 1 ... \mathrm{QP}_{\mathrm{pre}} + 2$ to reduce the rate fluctuation among neighbour units. The each

QP, the corresponding estimated bits R_{TCM} is obtained based on (4.9) and (4.10). If $R_{TCM} \leq R_T$, the total cost is computed as

$$J = D_{TCM} + \lambda R_{TCM} \tag{4.13}$$

where the distortion D_{TCM} is computed as (4.5). The QP with minimum cost will be employed for encoding.

However, if there is no QP satisfying the condition $R_{TCM} \leq R_T$, $|(R_{TCM}(\mathrm{QP}_{\mathrm{pre}} + 2) - R_T)|$ and$|R_{TCM}(\mathrm{QP}_{\mathrm{pre}} + 3) - R_T|$ are compared. Then, the QP with smaller rate distance would be taken.

Algorithm 1 QP decision

Input:

The quantization parameter (QP) used for previous frame/block: $\mathrm{QP_{pre}}$;
The target bit rate for current frame/block, R_T;
The set of BGTCM parameters, S;

Output:

The QP used for current frame/block, Q;

if $R_T < 0$ **then**
  $Q = \mathrm{QP_{pre}} + 2$; \∗**when target bit rate is negative,** $Q = \mathbf{QP_{pre}} + 2$ ∗\
else
  $Q_{Temp}[5] = \left\{\mathrm{QP_{pre}} - 2, \mathrm{QP_{pre}} - 1 ... \mathrm{QP_{pre}} + 2\right\}$;
  $HasQ = false$;
  $J_{Best} = MAX$;
  \∗**Choose the** QP **with minimum RD cost which is estimated based on BGTCM parameters**∗\
  for $i = 0; i < 5; i++$ **do**
    if $R_{BGTCM}(Q_{Temp}[i], S) \leq R_T$ **then**
      $HasQ = true$;
      $J = D_{BGTCM}(Q_{Temp}[i], S) + \lambda R_{BGTCM}(Q_{Temp}[i], S)$;
      if $J < J_{Best}$ **then**
        $J = J_{Best}$;
        $Q = Q_{Temp}[i]$;
      end if
    end if
  end for
  if !HasQ **then**
    Compare $|(R_{BGTCM}(\mathrm{QP_{pre}} + 2) - R_T)|$ and$|R_{BGTCM}(\mathrm{QP_{pre}} + 3) - R_T|$ Choose the small one
  end if
end if
return Q;

4.5 Experimental Results

This section presents the experimental results for the proposed rate control algorithm. The testings are implemented in the HEVC reference codec HM-9.0 under default LD-Main configuration. The actually generated bit rate without rate control in HM-9.0 are employed as target bit rates. The initial QPs are set as 22, 27, 32 and 37 for the first intra frame. Twelve sequences [4] with different resolution, motion complexity are utilized. The rate control method in [35] are implemented as well for the purpose of comparison.

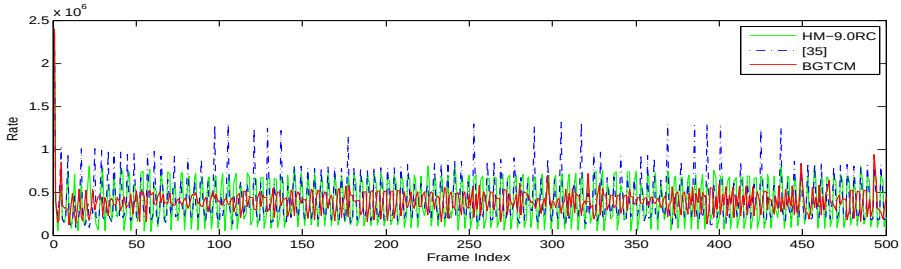
Firstly, the coefficients among the whole frame are grouped to generate a BGTCM with rate control on CU level. The experimental results are shown in Table 4.2. HM-9.0 with rate control on is employed as benchmark. BD-Rate [11] is used to compare those three methods in coding efficiency and variance of number of generated bits among frames is employed to compare them in rate fluctuation.

According to Table 4.2, it is clear that the proposed rate control method based on BGTCM is better than the reference HEVC codec with rate control and [35] in both coding efficiency and rate fluctuation. Compared with HM, on average, the proposed method based on BGTCM can achieve 13.9% reduction in BD-rate with around 67% reduction in variance of bits. On the other hand, when compared with HM, although the method in [35] based on Laplacian achieves 12.9% reduction in BD-rate, the variance of bits is increased by 14.3% as the cost. Figure 4.5 and 4.6 show rate fluctuations of sequence Cactus and BasketballPass with different QPs respectively. This also provides the verification for the performance of proposed methods based on BGTCM. Therefore, the proposed method based on BGTCM outperforms the other two in both rate fluctuation and RD performance.

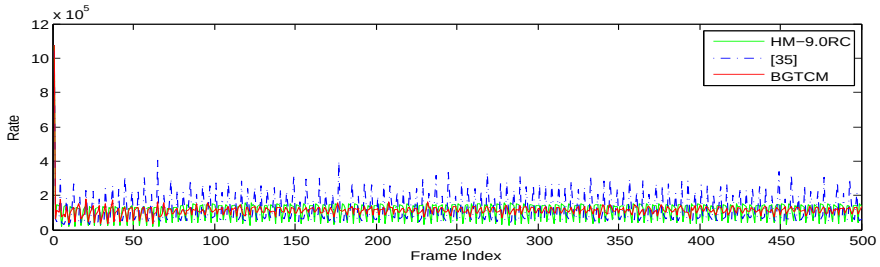
Secondly, due to the CG-based on coefficients encoding method in HEVC, coefficients from the whole frame are separated into 16 groups based on the scan indices. As such, in each group, coefficients come from the same scan indices among different CGs. At this time, only frame level rate control is performed. The experimental results are shown in Table 4.3. Compared with HM, the proposed method can achieve 2.8% and 45.4% reduction in BD-rate and bits variance, respectively.

Table 4.2: Experimental Results with BGTCM on Coefficients from the Whole Frame

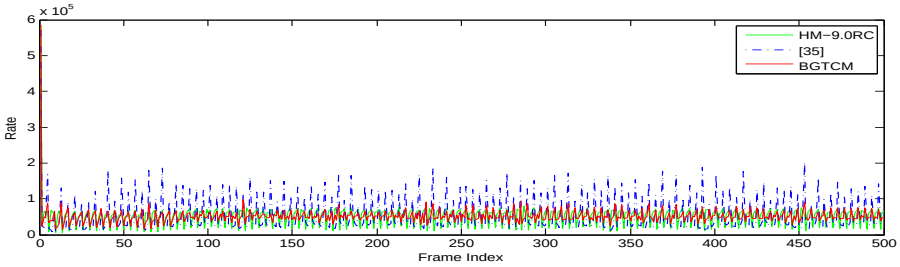
Sequ -ence	Target Rate (kbps)	HM9.0RC		[35]			BGTCM			Variance(Rate)$\times 10^9$		
		Rate (kbps)	PSNR (dB)	Rate (kbps)	PSNR (dB)	BD-Rate%	Rate (kbps)	PSNR (dB)	BD-Rate%	HM9.0 RC	[35]	TCM
Race Horses C	5300	5425.36	39.14	5421.57	39.46	-2.8	5293.74	39.48	-5.8	4.60	4.77	1.63
	2300	2355.97	36.10	2358.06	36.24		2299.55	36.24		1.54	1.79	0.69
	1000	1024.11	32.84	1025.10	32.88		1000.72	32.92		0.73	1.04	0.22
	450	460.59	29.87	460.57	29.90		450.83	29.96		0.65	0.37	0.06
Park Scene	8000	8115.25	39.64	8123.36	39.69	-22.0	7935.41	39.73	-20.1	52.7	46.3	36.3
	3200	3274.35	36.09	3260.99	36.84		3171.16	36.73		7.16	20.2	7.73
	1400	1433.13	33.04	1421.48	34.19		1386.10	33.89		2.02	6.58	2.49
	600	613.69	30.31	606.59	31.57		589.89	31.31		0.49	1.40	0.68
BQ Mall	4200	4302.35	39.86	4302.47	39.87	-8.0	4193.41	39.91	-6.2	0.78	1.44	0.58
	1900	1945.60	36.93	1945.42	37.21		1896.56	37.11		0.26	0.63	0.24
	900	921.46	33.85	920.98	34.30		899.18	34.00		0.12	0.26	0.07
	460	470.80	31.06	470.57	31.54		459.52	31.28		0.06	0.08	0.02
Kimono	5200	5319.38	41.26	5251.53	41.42	-19.9	5191.80	41.56	-20.9	14.7	12.5	2.73
	2400	2435.92	38.84	2433.55	39.414		2394.08	39.39		4.29	5.38	1.21
	1200	1227.58	36.10	1210.44	37.06		1193.52	36.99		2.07	2.44	0.49
	600	611.08	33.02	605.77	34.61		590.50	34.95		0.91	0.94	0.45
Basket ball Pass	1800	1841.68	41.12	1838.81	41.03	+0.8	1801.16	41.26	-3.3	0.045	0.098	0.044
	900	921.70	37.22	922.23	37.25		901.75	37.35		0.017	0.0618	0.016
	450	460.80	33.93	461.15	33.89		451.03	33.95		0.009	0.037	0.007
	200	204.77	30.55	204.86	30.39		200.72	30.45		0.004	0.013	0.002
Race Horses D	1400	1433.72	39.90	1434.75	39.97	-1.7	1398.11	39.94	-3.6	0.09	0.25	0.13
	650	666.14	35.60	666.81	35.74		650.10	35.70		0.06	0.14	0.04
	300	307.34	32.04	307.35	32.11		300.51	32.06		0.015	0.065	0.017
	150	153.57	29.40	153.34	29.27		150.17	29.40		0.005	0.028	0.005
Basket ball Drive	20000	20422.76	39.21	20486.48	39.25	-4.9	19987.57	39.29	-7.4	48.5	33.4	4.05
	6800	6961.22	37.26	6952.58	37.31		6802.19	37.36		1.5	6.06	0.55
	3200	3275.12	35.24	3272.05	35.42		3200.56	35.38		0.42	2.3	0.13
	1600	1637.98	32.81	1635.47	33.25		1600.04	33.20		0.15	0.76	0.12
Cactus	20000	20423.22	38.54	20461.00	38.50	-22.9	20021.412	38.11	-22.8	83.0	84.6	13.5
	5700	5835.46	35.82	5822.64	36.52		5695.33	36.54		10.1	8.59	2.23
	2600	2660.89	33.48	2656.88	34.43		2598.97	34.34		6.21	2.73	0.73
	1300	1330.81	31.03	1328.29	32.14		1299.24	32.13		3.60	0.81	0.25
Johnny	1500	1533.33	41.90	1523.88	42.66	-28.8	1495.60	42.80	-31.4	0.53	0.51	0.27
	500	509.54	40.36	505.33	41.12		498.64	41.19		0.32	0.10	0.069
	200	203.79	38.07	201.52	38.91		199.15	38.97		0.044	0.029	0.023
	100	101.92	35.75	100.83	36.35		99.38	36.41		0.015	0.009	0.008
Four People	2200	2247.34	41.63	2240.78	42.23	-4.8	2194.43	42.43	-7.3	1.13	0.94	0.47
	900	916.88	39.77	914.47	40.03		897.55	40.11		0.39	0.29	0.18
	450	458.28	37.48	456.72	37.50		449.36	37.45		0.13	0.11	0.07
	200	203.37	34.21	202.70	34.13		199.60	34.11		0.04	0.04	0.03
Kristen And Sara	2000	2047.46	42.47	2044.44	43.00	-17.1	1997.39	43.14	-19.9	0.55	0.62	0.26
	700	716.64	40.29	714.47	40.85		699.34	40.90		0.16	0.09	0.16
	350	358.08	38.28	357.27	38.75		349.94	38.77		0.16	0.16	0.09
	200	204.53	35.98	204.09	36.54		200.38	36.50		0.03	0.023	0.018
BQ Terrace	53000	53257.09	38.12	51394.74	38.04	-30.1	51992.56	38.14	-28.4	236	273	90.4
	7400	7574.46	34.75	7607.85	35.10		7395.54	35.11		7.70	43	5.18
	2000	2043.10	31.70	2035.05	33.35		1993.62	32.99		1.86	2.68	1.52
	800	816.39	30.27	813.23	31.30		796.42	30.97		0.72	0.65	0.49
Avarage	/	/	/	/	/	-12.9	/	/	-13.9	9.54	10.9	3.43



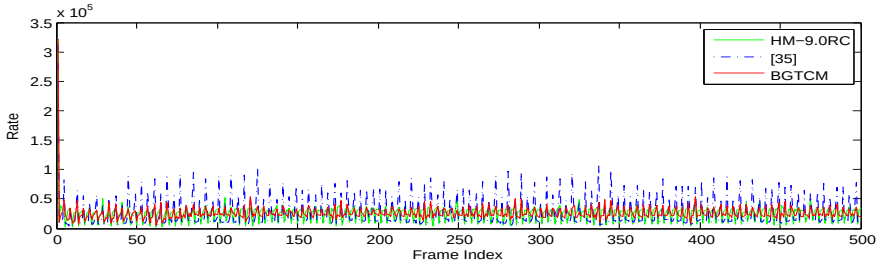
(a) QP=22



(b) QP=27

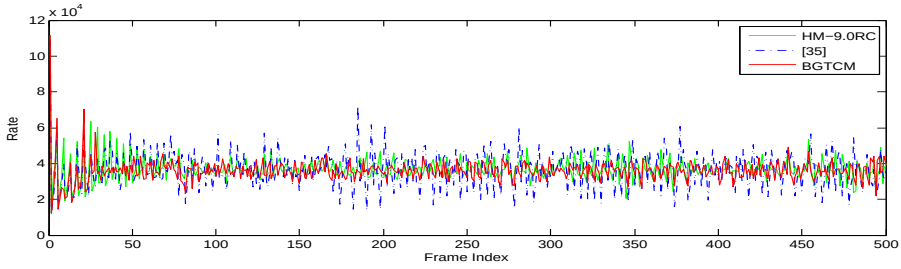


(c) QP=32

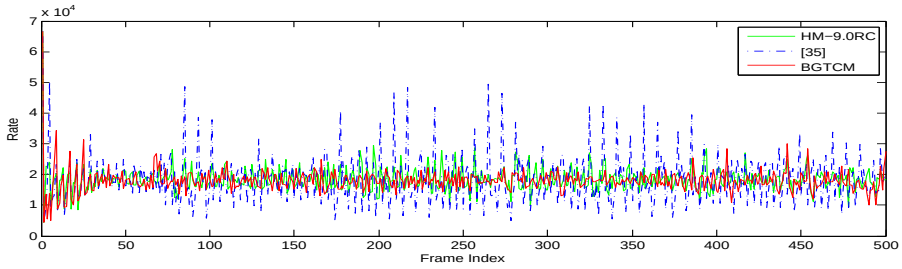


(d) QP=37

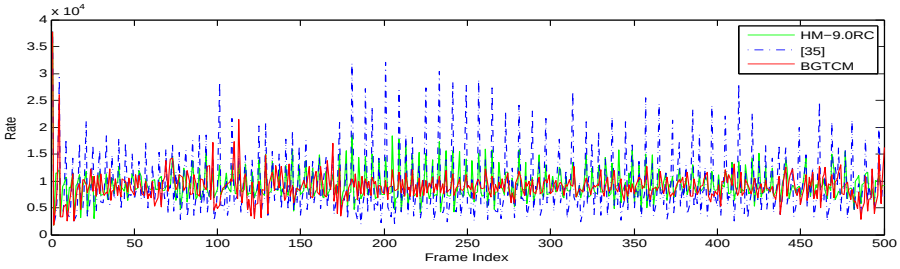
Figure 4.5: Generated Rate Versus Frame for Three Rate Control Algorithms on Sequence Cactus with Different QPs.



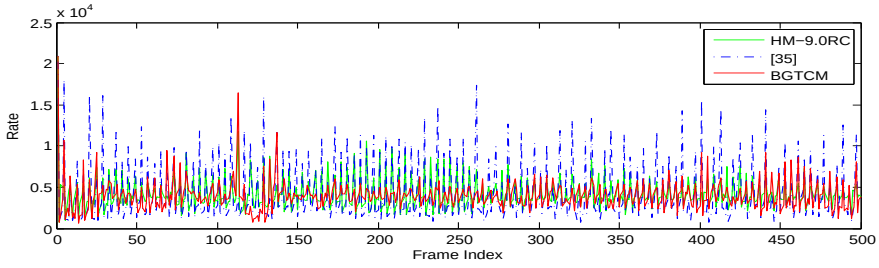
(a) QP=22



(b) QP=27



(c) QP=32



(d) QP=37

Figure 4.6: Generated Rate Versus Frame for Three Rate Control Algorithms on Sequence BasketballPass with Different QPs.

Table 4.3: Experimental Results on Frame Level Rate Control with 16 BGTCMs

Sequence	Target Rate (kbps)	HM9.0RC		BGTCM			Var(Rate)×10^9	
		Rate (kbps)	PSNR (dB)	Rate (kbps)	PSNR (dB)	BD-Rate	HM9.0-RC	BGTCM
Race HorsesC	5300	5421.96	39.71	5295.29	39.67	-2.6%	4.58	3.16
	2300	2357.74	36.48	2301.24	36.45		1.88	1.39
	1000	1025.06	33.11	1002.90	33.19		1.13	0.58
	450	460.65	30.18	450.15	30.19		0.32	0.17
Park Scene	8000	8120.37	39.92	7966.49	39.93	-2.8%	49.1	46.0
	3200	3249.19	37.06	3188.53	37.13		1.94	1.35
	1400	1420.46	34.41	1397.88	34.47		6.66	4.18
	600	607.14	31.79	593.93	31.76		1.40	0.99
BQ Mall	4200	4302.61	40.10	4195.04	40.11	-3.5%	1.43	1.06
	1900	1945.96	37.44	1898.07	37.49		0.65	0.44
	900	921.22	34.55	899.48	34.60		0.26	0.17
	460	470.94	31.79	460.40	31.83		0.08	0.05
Kimono	5200	5250.06	41.64	5185.50	41.62	-0.2%	13.1	13.3
	2400	2428.38	39.63	2392.87	39.59		5.47	4.76
	1200	1214.07	37.31	1194.66	37.27		2.44	2.01
	600	604.54	34.81	596.29	34.70		1.09	0.91
Basketball Pass	1800	1839.34	41.29	1804.81	41.28	-0.8%	0.10	0.09
	900	922.04	37.47	903.95	37.39		0.05	0.04
	450	461.20	34.13	452.90	34.09		0.04	0.03
	200	204.84	30.58	200.83	30.54		0.01	0.009
Race HorsesD	1400	1434.80	40.23	1398.21	40.14	-1.5%	0.22	0.21
	650	667.07	35.98	649.82	35.93		0.15	0.10
	300	307.52	32.34	300.54	32.29		0.07	0.05
	150	153.53	29.48	150.31	29.54		0.03	0.01
Basketball Drive	20000	20482.02	39.46	19987.14	39.42	-1.6%	30.0	20.4
	6800	6959.457	37.53	6803.23	37.49		6.05	3.59
	3200	3271.54	35.65	3202.00	35.69		2.19	1.49
	1600	1635.38	33.49	1601.44	33.42		0.75	0.55
Cactus	20000	20449.74	38.71	19893.54	38.70	-3.3%	82.6	56.2
	5700	5818.23	36.74	5709.21	36.79		8.68	5.56
	2600	2655.42	34.65	2601.94	34.69		2.65	1.85
	1300	1327.35	32.38	1300.26	32.42		0.82	0.66
Johnny	1500	1522.62	42.88	1509.16	42.89	-5.1%	0.51	0.76
	500	504.70	41.34	499.96	41.45		0.10	0.12
	200	201.55	39.13	198.95	39.23		0.03	0.03
	100	100.79	36.57	99.58	36.69		0.01	0.01
Four People	2200	2239.56	42.45	2200.27	42.42	-2.9%	0.97	0.81
	900	912.74	40.26	898.62	40.35		0.29	0.31
	450	456.45	37.72	449.28	37.76		0.11	0.12
	200	202.67	34.36	199.91	34.39		0.04	0.03
Kristen AndSara	2000	2040.70	43.22	2005.75	43.23	-5.0%	0.61	0.72
	700	714.08	41.09	700.44	41.12		0.16	0.15
	350	357.18	38.97	348.82	39.15		0.06	0.06
	200	204.02	36.77	199.99	36.96		0.02	0.02
BQ Terrace	52000	53271.47	38.27	51971.96	38.19	-3.7%	293	138
	7400	7548.27	35.29	7406.11	35.38		43.6	20.4
	2000	2034.53	33.57	2004.38	33.58		2.73	2.83
	800	813.31	31.52	802.14	31.48		0.64	0.61
Avarage	/	/	/	/	/	-2.8%	13.3	7.26

4.6 Summary

In this chapter, based on the assumption that the distribution of both original and quantized transformed coefficients in HEVC satisfy BGTCM, the distortion and rate estimated with BGTCM, Laplacian and Cauchy models are computed. Results show that BGTCM can estimate the bit rate and distortion more accurately compared with Laplacian and Cauchy distributions. a RD optimized rate control algorithm is developed to achieve required bit rate and maintain RD performance. The rate fluctuation is compared in terms of variance and RD performance is compared in terms of BD-rate. Compared with HM and a newly proposed algorithm, the proposed method can reduce rate fluctuation and increase RD performance.

Chapter 5

Conclusion and Future Work

This chapter concludes this thesis by summarizing the contents and contributions and presenting some possible future research topics.

5.1 Conclusion

This thesis begins with brief introduction of video coding and modelling of distribution of transformed coefficients. In Chapter 2, the hybrid encoding diagram of HEVC encoder is described in the first section. Then some prior works on modelling of distribution of transformed coefficients and model-based rate control are presented. Due to the "chicken and egg" dilemma, the rate controller are normally developed based on R-Q and D-Q models. To obtain accurate models, precise estimation of distribution of transformed coefficients is required.

Chapter 3 analyses the modeling of distribution of coefficients in prediction-based HEVC video coding standard. Distribution of all coefficients from a whole frame are firstly modelled. Then, based on the scan indices among different CGs, coefficients at the same scan index are grouped into one set. Distribution of coefficients from each set are modelled. With the newly proposed model BGTCM in discrete version, which has been reported to achieve better modelling in DCT coefficients obtained by performing DCT on original images like the image compression standard JPEG, the model of the distribution of integer transformed coefficients based on residual data in HEVC is better than models provided by Laplacian and Cauchy densities. It benefits from the fat tails in the distribution of coefficients, which traditional models, such as Laplacian and Cauchy densities,

cannot model precisely. The D-Q relation is developed based on distribution of original coefficients and R-Q relation is based on distribution of quantized coefficients. Experimental results have shown that, BGTCM can be employed to fit distribution of both original and quantized coefficients better than Laplacian and Cauchy density especially for large coefficients values which contribute a lot to the number of bits in the bit stream based on χ^2 and KL-divergence tests.

In Chapter 4, a RD optimized rate control algorithm is proposed to determine the QP for current frame or block based on BGTCM. Experimental results have shown that, based on BGTCM, compared with the reference rate control implementation in HM-9.0, the proposed rate control method reduces the rate and quality fluctuation a lot in HEVC. Meanwhile, the RD performance is highly increased.

5.2 Future Work

There are a few possible extensions based the discussion in this thesis.

5.2.1 Estimation of Number of Overhead Bits

The overall bits generated by the encoder can be divided into two parts:

- Texture bits generated by encoding DCT coefficients
- Overhead bits generated by encoding other information

This thesis gives a solution to model the texture bits precisely based on BGTCM. However, the overall bits are simply estimated by multiplying the texture bits with a scaling factor. The overhead bits are generated by encoding other information, such as block partition structure, prediction mode, motion vectors and etc. Unlike bits caused by DCT coefficients, the number of overhead bits has relatively little relationship with QP. It can be estimated based the characteristics of video sequences and coding configuration.

Let's take an intra frame as an example. When the frame is complicated, i.e., it contains a lot of texture information, blocks are potential to be encoded as small ones. In this scenario, the overhead bits can be estimated by how complex a frame is. Since the BGTCM provides the outlier information, which can reflect the edge in image domain, the complexity can be estimated by this.

Consequently, one can develop an adaptive method based on spatial and/or temporal complexity to estimate the number of overhead bits, which mainly consist of prediction information.

5.2.2 Approximated Expression of D-Q Model

Compared to continuous Laplacian density, it is hard to obtain precise D-Q model in simple expressions for the BGTCM which is a discrete function. These complex expressions make it hard to resolve the rate distortion optimization problems. This is why all five candidate QPs determined by the ΔQP-limiter are tested in the proposed algorithm. Thus, the D-Q relation based on BGTCM will studied further to derive some simple expression without estimation mismatching.

References

[1] ITU-T H.263: Video coding for low bit rate communication. 1996.

[2] The MPEG-2 international standard. *ISO/IEC, Reference Number ISO/IEC 13 818-2*, 1996.

[3] Rate control and H.264. *http://www.pixeltools.com/rate_control_paper.html*, 2012.

[4] Test sequences. *ftp://ftp.tnt.uni-hannover.de*, 2012.

[5] HEVC test model. *http://hevc.kw.bbc.co.uk/git/w/jctvc-tmuc.git/commit/1204fc83b1f91723762 4e9ec8c3c4d7bebab58d5*, April 2014.

[6] F. Bossen. Common test conditions and software reference configurations. *JCTVC-I1100, Joint Collaborative Team on Video Coding meeting*, May 2012.

[7] B. Bross, W. J. Han, J. R. Ohm, G. J. Sullivan, Y. K. Wang, and T. Wiegand. High efficiency video coding (HEVC) text specification draft 10. *JCTVC-L1003, Joint Collaborative Team on Video Coding meeting*, 2013.

[8] T. Chiang and Y.-Q. Zhang. A new rate control scheme using quadratic rate distortion model. *IEEE Trans. Circuits Syst. Video Technol.*, 7(1):246–250, February 1997.

[9] K. Choi and E. S. Jang. Coding tree pruning based CU early termination. *JCTVC-F092, Joint Collaborative Team on Video Coding 6th meeting*, July 2011.

[10] T. Eude, R. Grisel, H. Cheri, , and R. Debrie. On the distribution of the DCT coefficients. In *IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pages 365–368, April 1994.

[11] G.Bjontegaard. Improvements of the BD-PSNR model. *ITU-T SG16, VCEG-AI11*, July 2008.

[12] Z. He, Y. K. Kim, and S. K. Mitra. Low delay rate control for DCT video coding via ρ-domain source modeling. *IEEE Trans. Circuits Syst. Video Technol.*, 11(8), August 2001.

[13] Z. He and S. Mitra. Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling. *IEEE Trans. Circuits Syst. Video Technol.*, 12(10):840–849, October 2002.

[14] Z. He and S. K. Mitra. A unified rate-distortion analysis framework for transform coding. *IEEE Trans. Circuits Syst. Video Technol.*, 11(12):1221–1236, December 2001.

[15] M. Jiang and N. Ling. On enhancing H.264/AVC video rate control by PSNR-BASED frame complexity estimation. *IEEE Trans. Consum. Electron.*, 51(2):281–286, February 2005.

[16] N. Kamaci, Y. Altunbasak, and R. M. Mersereau. Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models. *IEEE Trans. Circuits Syst. Video Technol.*, 15(8):994–1005, August 2005.

[17] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T.Ishiguro. Motion compensated interframe coding for video conferencing. In *Proc. of IEEE Nat. Telecommunications Conf.*, pages C9.6.1–C9.6.5, New Orleans, LA, 1981.

[18] E. Y. Lam and J. W. Goodman. A mathematical analysis of the DCT coefficient distributions for images. *IEEE Trans. Image Process.*, 9(10):1661–1666, October 2000.

[19] T. Lan and X. Gu. H.264 frame layer rate control based on block histogram difference. In *ICC*, pages 281–284. IEEE, May 2008.

[20] J. Lee, I. Shin, and H. Park. Adaptive intra-frame assignment and bitrate estimation for variable GOP length in H.264. *IEEE Trans. Circuits Syst. Video Technol.*, 16(10):1271–1279, October 2006.

[21] D. LeGall. MPEG: A video compression standard for multimedia application. *Commun. ACM*, 34:46 – 58, 1991.

[22] B. Li, G. J. Sullivan, and J. Xu. Comparison of compression performance of HEVC working draft 7 with avc high profile. *JCTVC-J0236, Joint Collaborative Team on Video Coding meeting*, 2012.

[23] Z. Li, W. Gao, F. Pan, S. Ma, K. P. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. lu. Adaptive rate control with HRD consideration. *JVT-H014, JVT of ISO/IEC and ITU-T*, May 2003.

[24] Z. G. Li, F. Pan, K. P. Lim, and S. Rahardja. Adaptive rate control for h.264. In *IEEE International Conference on Image Processing*, pages 745–748, October 2004.

[25] L.-J. Lin, A. Ortega, and C.-C. J. Kuo. A gradient-based rate control algorithm with applications to MPEG video. In *Int. Conf. Image Processing*, pages 392–395. IEEE, 1995.

[26] J. Liu and A. Ortega. Bit allocation using piecewise approximated rate-distortion characteristics. *IEEE Trans. Circuits Syst. Video Technol.*, 8(4):446–459, August 1998.

[27] Y. Liu, Z. G. Li, and Y. C. Soh. A novel rate control scheme for low delay video communication of H.264/AVC standard. *IEEE Trans. Circuits Syst. Video Technol.*, 17(1):68–78, January 2007.

[28] S. Ma, W. Gao, and Y. Lu. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans. Circuits Syst. Video Technol.*, 15(12):1533–1544, December 2005.

[29] F. Muller. Distribution shape of two-dimensional DCT coefficients natural images. *Electron. Lett.*, 29(22):1935–1936, October 1993.

[30] L. M. Po and W. C. Ma. A novel four-step search algorithm for fast block motion estimation. *IEEE Trans. Circuits Syst. Video Technol.*, 6:313–317, June 1996.

[31] W. K. Pratt. *Digital Image Processing*. New York: Wiley, 1978.

[32] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multi resolution and MPEG video coders. *IEEE Trans. Image Process.*, 3(5):533–545, September 1994.

[33] R. C. Reininger and J. D. Gibson. Distributions of the two-dimensional DCT coefficients for images. *IEEE Trans. Commun.*, 31(6):835–839, June 1983.

[34] Iain E. G. Richardson. *H.264 and MPEG-4 Video Compression*. John Wiley & Sons Ltd, 2003.

[35] Chan-Won Seo, Joo-Hee Moon, and Jong-Ki Han. Rate control for consistent objective quality in HEVC. *IEEE Trans. Image Proc.*, 22(6):2442–2455, June 2013.

[36] T. Sikora. The MPEG-4 video standard verification model. *IEEE Trans. Circuits Syst. Video Technol.*, 7(1), 1997.

[37] S. R. Smooth and R. A. Lowe. Study of DCT coefficients distributions. In *SPIE*, pages 403–311. IEEE, January 1996.

[38] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Due nas. Transform coefficient coding in HEVC. *IEEE Trans. on Circuits and Syst. for Video Technol.*, 22(12):1765–1777, December 2012.

[39] Y. H. Tan, C. Yeo, and Z Li. Single-pass rate control with texture and non-texture rate-distortion models. *IEEE Trans. Circuits Syst. Video Technol.*, 22(8):1236–1244, August 2012.

[40] G. K. Wallace. The JPEG still picture compression standard. *Commun. ACM*, 34:30 – 44, 1991.

[41] H. Wang and S. Kwong. Rate-distortion optimization of rate control for h.264 with adaptive initial quantization parameter determination. *IEEE Trans. Circuits Syst. Video Technol.*, 18(1):140–144, January 2008.

[42] N. Wang and Y. He. A new bit rate control strategy for H.264. In *IEEE Int. Conf. Inf. Commun. Signal Process.*, pages 1370–1373, December 2003.

[43] Z. Wu, H. Yu, B. Tang, and C. W. Chenan. Adaptive initial quantization parameter determination for H.264/AVC video transcoding. *IEEE Trans. Broadcasting*, 58(2):277–284, June 2012.

[44] E.-H. Yang and X. Yu. Soft decision quantization for H.264 with main profile compatibility. *IEEE Trans. Circuits and Systems for Video Technology*, 19:122–127, 2009.

[45] E.H. Yang, X. Yu, and J. Meng. Multiple segment transparent composite model for DCT coefficients.

[46] E.H. Yang, X. Yu, J. Meng, and C. Sun. Transparent composite model for DCT coefficients: Design and analysis. *IEEE Trans. Image Processing*, 23:1303 – 1316.

[47] S. Zhou, J. Li, J. Fei, and Y. Zhang. Improvement on rate-distortion performance of H.264 rate control in low bit rate. *IEEE Trans. Circuits Syst. Video Technol.*, 17(8):996–1006, August 2007.

[48] Y. Zhou, Y. Sun, Z. Feng, and S. Sun. PID-based bit allocation strategy for H.264/AVC rate control. *IEEE Trans. Circuits Syst. Video Technol.*, 58(3):184–188, March 2011.