Investigations of the Error Resilient Rate-Distortion Optimization for H.264 Video Transport over Packet Lossy Channels

Von der Fakultät Informatik, Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Hao Zeng

aus Zizhong, China

Hauptberichter:	Prof. Dr.-Ing. J. Speidel
Mitberichter:	Prof. Dr. rer. nat. habil. P. Levi
Tag der mündlichen Prüfung:	11. Mai 2010

Institut für Nachrichtenübertragung der Universität Stuttgart

2010

Abstract

This dissertation first presents a rigorous analysis of the rate-distortion optimization (RDO) framework in H.264 and other similar video coding standards. The analysis is based on the rate-distortion theory and provides a deep theoretical insight into the RDO practice. In the analysis, the crucial assumptions and models on which the RDO is based, are formulated and discussed. Then the validity of deploying the RDO in error resilient video encoding is analyzed. Finally, the dissertation proposes and investigates two modification schemes to the error resilient RDO (ER-RDO) framework in the H.264 reference codec. The schemes are inspired by the aforementioned analyses and seek to find better complexity-performance balance points. The simulation results suggest that the performance of the existing H.264 ER-RDO framework has already been highly optimized. On the other side, algorithm complexity reduction is still possible using one of the proposed schemes.

Kurzfassung

Diese Dissertation stellt zunächst eine eingehende Analyse der Rate-Distortion-Optimierung (RDO) für H.264 und ähnliche Standards der Videocodierung. Die Analyse basiert auf der Rate-Distortion-Theorie und liefert einen tiefen theoretische Einblick in ihre praktische Anwendung. Bei der analytische Betrachtung werden die wichtigsten Annahmen und Modelle, auf denen die RDO beruht, wiedergegeben und diskutiert. Danach wird die Gültigkeit des RDO-Einsatzes für fehlertolerante Videocodierung analysiert. Schließlich werden zwei Änderungsvorschläge für die Grundstruktur der fehlertoleranten RDO (ER-RDO) diskutiert und mittels des H.264-Referenz-Codecs untersucht. Die Änderungsvorschläge wurden auf der Basis der ausgeführten Analysen entwickelt und sollen ein besseres Gleichgewicht zwischen Komplexität und Leistungsfähigkeit finden. Die Simulationsergebnisse zeigen, dass die Leistungsfähigkeit der bestehenden H.264-ER-RDO bereits stark optimiert ist. Jedoch kann mit einem der vorgeschlagenen Verfahren die Komplexität des Algorithmus reduziert werden.

Contents

List of Abbreviations

CABAC	context adaptive binary arithmetic coding
CAVLC	context adaptive variable length coding
DCT	discrete cosine transform
DMS	discrete memoryless source
DR	distortion-rate
ER-RDO	error resilient rate-distortion optimization
FEC	forward error correction
FMO	flexible macroblock ordering
HVS	human visual system
IDCT	inverse discrete cosine transform
JM	joint model (the H.264 reference codec)
JSCC	joint source channel coding
LA-RDO	loss aware rate-distortion optimization
MAD	mean absolute difference
MB	macroblock
MSE	mean squared error
NAL	network abstraction layer
NALU	NAL unit
PSNR	peak signal-to-noise ratio
QCIF	quarter common intermediate format
QP	quantization parameter
QS	quantization step size
RDO	rate-distortion optimization
ROI	region of interest
RVLC	reversible variable length coding
SAD	sum of absolute difference
SDU	service data unit
SSE	sum of squared error
UEP	unequal error protection
VCL	video coding layer

List of Symbols

$P\{\cdot\}$	probability
$E(\cdot)$	mathematical expectation
$d(\cdot,\cdot)$	distortion
$I(\cdot;\cdot)$	mutual information
$H(\cdot)$	information entropy
n	discrete time index
S_n	an output symbol of the information source S
$\check{S}_n$	decoded or reconstructed S_n
$\boldsymbol{S_n}$	vector of information symbols $(S_1, S_2, \ldots, S_n)$
$\check{\boldsymbol{S}}_{\boldsymbol{n}}$	decoded or reconstructed $\boldsymbol{S_n}$
Φ, Υ, Θ, Ω	set sizes
α	complexity tuning parameter
p	slice loss ratio
$\psi_{\max}$	peak signal value for a given bit depth
σ_e^2	MSE value in video quality evaluation
K	number of MBs in a video frame
k	MB index in a video frame
MB_k	k-th MB
R_0	frame budget rate
M	set of available modes
m_k	mode of MB_k
$\hat{m}_k$	optimal value of m_k
$m_{k,\mathrm{INIT}}(q_k)$	mode of a point from $O_k(q_k)$ with the least distortion
Q	set of available QSs
q	global QS
q_k	QS of MB_k
$\hat{q}_k$	optimal value of q_k
$\tilde{q}_k$	continuous QS of MB_k
$\hat{\tilde{q}}_k$	optimal value of $\tilde{q}_k$
q'	global QP
q'_k	QP of MB_k

$r_k(m_k, q_k)$	achievable rate of MB_k as a function of m_k and q_k
$\tilde{r}_k(\tilde{q}_k)$	operating rate-quantization function of MB_k
R_k	continuous rate of MB_k
$\hat{R}_k$	optimal value of R_k
$\acute{r}_k$	quantization constrained continuous rate of MB_k
$\hat{\acute{r}}_k$	optimal value of $\acute{r}_k$
$D(R)$	DR function
$D_k(R_k)$	distortion-rate function of MB_k
$d_k(m_k, q_k)$	achievable distortion of MB_k as a function of m_k and q_k
$\tilde{d}_k(\tilde{q}_k)$	operating distortion-quantization function of MB_k
$\tilde{D}_k(R_k)$	operating DR function of MB_k
$\acute{d}_k(\acute{r}_k\|\tilde{q}_k$	quantization constrained DR of MB_k
$\bar{D}_{k,\mathrm{INTRA}}(R_k)$	expected intra-only operating DR function of MB_k with slice losses
$\bar{D}_{k,\mathrm{INTER}}(R_k)$	expected inter-only operating DR function of MB_k with slice losses
$D_{k,\mathrm{CON}}$	distortion of MB_k after concealment
$D_{k,\mathrm{PROP}}$	distortion caused by the propagated errors in the reference regions of MB_k
$d_{k,\mathrm{SUM}}(m_k, q_k)$	distortion of MB_k for an inter mode m_k with slice losses when MB_k is received
$\bar{D}_k(R_k)$	expected operating DR function of MB_k with slice losses
J_{QUANT}	global cost function in QS selection
J_{MODE}	global cost function in mode selection
J_k	local cost function in the mode selection for MB_k
$J_{k,\mathrm{LOSS}}$	local cost function in the mode selection for MB_k with slice losses
$J_{k,\mathrm{LOSS,\ JM}}$	JM version of $J_{k,\mathrm{LOSS}}$
$J_{k,\mathrm{LOSS,FAST}}$	local cost function in the fast mode selection for MB_k with slice losses
λ	Lagrange multiplier in both QS and mode selection
$\Lambda(q)$	Lagrange multiplier-quantization step size function
λ_{QUANT}	Lagrange multiplier in QS selection
$\hat{\lambda}_{\mathrm{QUANT}}$	optimal value of λ_{QUANT}
λ_{MODE}	Lagrange multiplier in mode selection
$\hat{\lambda}_{\mathrm{MODE}}$	optimal value of λ_{MODE}
λ_{LOSS}	Lagrange multiplier in ER-RDO

λ_{MB}	Lagrange multiplier used in the modeling of $\Lambda(q)$
q_{MB}	QS of a MB in the modeling of $\Lambda(q)$
D_{MB}	distortion of a MB in the modeling of $\Lambda(q)$
R_{MB}	rate of a MB in the modeling of $\Lambda(q)$
J_{MB}	local cost function of a MB in the modeling of $\Lambda(q)$
$\boldsymbol{u}$	vector $(m_1, q_1, m_2, q_2, \ldots, m_K, q_K)$
$\boldsymbol{q}$	vector $(q_1, q_2, \ldots, q_K)$
$\boldsymbol{r}$	vector $(R_1, R_2, \ldots, R_K)$
$\boldsymbol{m}$	vector $(m_1, m_2, \ldots, m_K)$
$\acute{\boldsymbol{r}}$	vector $(\acute{r}_1, \acute{r}_2, \ldots, \acute{r}_K)$
$\hat{\boldsymbol{u}}$	vector $(\hat{m}_1, \hat{q}_1, \hat{m}_2, \hat{q}_2, \ldots, \hat{m}_K, \hat{q}_K)$
$T_k(\tilde{q}_k)$	tangent point of the unconstrained and quantization constrained DR functions of MB_k
$O_k(q_k)$	set of the achievable DR points of MB_k as a function of q_k

List of Figures

List of Tables

Chapter 1

Introduction

H.264 is the latest member of the H.26x video coding standard series and gradually gaining popularity in many applications. Since H.264 is a typical block based hybrid video coding method and intensively exploits inter and intra frame redundancy, videos encoded with it are very sensitive to transport errors. Consequently, error control measures should be taken to reduce the influence of such errors in unreliable communication networks.

Although most communication networks provide some error control tools such as channel coding and retransmission, the extra delay caused by the error control tools is sometimes intolerable for real time video applications. In those applications the video encoder and the video decoder have to cope with transport errors. H.264 provides some error control tools which make encoded videos more robust to transport errors. Among the tools, encoder-side intra refreshing and decoder-side error concealment are widely adopted because of their simplicity and effectiveness. A key issue of intra refreshing is to find the optimal positions of forced intra macroblocks (MBs). The H.264 ER-RDO framework has proved to excel in locating forced intra MBs.

Error resilient rate-distortion optimization (ER-RDO) is a extension of rate-distortion optimization (RDO). RDO uses convex optimization techniques to select optimal video encoding parameters. Although RDO has been thoroughly studied, there are still links between the theory and the practice that open up chances for improvements. This dissertation provides a rigorous analysis of the H.264 RDO, with emphasis on its usage in MB quantization parameter and prediction mode selection. The extension from

RDO to ER-RDO is also discussed.

In a complexity constrained video application, the computational load of the video codec, especially the encoder has to be considered. H.264 has the most complex codecs among all the H.26x standards. On the basis of the analyses of the H.264 RDO and ER-RDO frameworks in this dissertation, two opposite modification schemes are put forward for better complexity-performance balances. One modification is to broaden the search range for the optimal distortion-rate (DR) point for each MB, in order to improve the compression and error resilience performance. The other is to conditionally narrow the search range, in order to lower the computational complexity. The suggested schemes are checked by simulations. The simulation results show that moderately broadening the search range might not be effective while narrowing the range is viable.

The rest of this dissertation is organized as follows. Chap. 2 is a survey of the technical background and related works. Chap. 3 analyses the H.264 RDO and ER-RDO. Chap. 4 introduces the suggested modification schemes and presents the simulation results. Chap. 5 concludes this dissertation. Appendix A shows the test video sequences. Appendix B presents a new formulation for optimal intra-refreshing.

Chapter 2

Error Control in H.264 Video Transport

This chapter introduces the background of this dissertation. Beginning with a brief explanation of the H.264 video coding standard, this chapter reviews various H.264 error control tools and identifies those suitable for complexity and delay constrained applications. Finally, some important previous works are listed. For a comprehensive survey of error control in video communications, see [1].

2.1 Raw Video Formats

An H.264 video encoder takes a raw video as the input signal. The raw video can be directly fetched from either a video acquisition device or a stored binary file which has been acquired before. The contents of the raw video form a hierarchic data structure which is described below.

The raw video is a sequence of photographic images. Usually the raw video is called a video sequence and the images are called frames. The frames are sequentially acquired with a certain time interval between two adjacent frames. The number of frames acquired per second is called the frame rate of the video. The frames are expected to be displayed in the same order and with the same frame rate, as they are acquired.

Although gray level videos (sometimes called black and white videos) are still used in some applications, e.g. surveillance, in this dissertation only color videos are considered. Therefore, a frame is conceptually a matrix of colored pixels. According to the properties of the human visual system (HVS), the color and the brightness of a pixel can be specified by three values: R for red, G for green and B for blue. The three components form the so called RGB color space. The RGB components of a visual scene are physically detected by a sensor in the video acquisition device. However, a color space transform is normally performed before the video compressing operations in order to separate the brightness information from the color information. The pixelwise transform generates another triple from the RGB values of the pixel : Y, Cb and Cr. Y is called the luma component and represents the brightness of the pixel. Cb and Cr are called the chroma components and jointly represent the color of the pixel. The luma and chroma components form the so called YCbCr color space.

Because the HVS is more sensitive to brightness than color, the Cb and Cr components of the pixels in a frame are normally subsampled to reduce the bandwidth requirement. As a consequence, a frame in the raw video is composed of three matrices. The luma matrix has a larger size than the chroma matrices.

Each element in the matrices occupies a certain number of binary bits. The number is called the bit depth of the raw video and determines the precision of the displayed colors and brightness levels. H.264 supports a variety of chroma subsampling schemes and bit depths. More details about color spaces and video acquisition can also be found in [1].

2.2 The H.264 Standard

In H.264, the functionality of a compliant encoder is conceptually divided into two layers: a video coding layer (VCL) and a network abstraction layer (NAL). The VCL compresses a raw video and the NAL formats the encoded data for transport.

2.2.1 The Video Coding Layer

The VCL of H.264 inherits the traditional hybrid video coding framework which originates from H.261 and is shown in Fig. 2.1.

An input video frame is divided into macroblocks (MBs) each containing one 16×16 luma pixel block and two corresponding chroma pixel blocks. The MBs are usually encoded one by one in the raster scan order. A MB is predicted either from some encoded pixels within the same frame or from some in a previously encoded frame. These two cases are called intra and inter prediction, respectively. In both cases the standard allows the encoder to select the sizes of prediction blocks for the MB. In this dissertation, the intra or inter selection together with the prediction block sizes are called the prediction mode of the MB. The prediction residue undergoes the discrete cosine transform (DCT) and quantization to bear less information entropy. Finally, the quantized DCT coefficients, the prediction mode information and some other control information are entropy encoded and multiplexed to form the output bit stream of the encoder.

At the same time, the virtual decoder within the encoder, demarcated by the rectangle in Fig. 2.1, decodes the current MB via inverse quantization, the inverse discrete cosine transform (IDCT) and residue compensation. The decoded or reconstructed MB which approximates the raw MB is stored into the encoded frame buffer for future prediction.

Though H.264 shares the same coding framework as H.263 [2], MPEG-2 [3] and MPEG-4 [4], it roughly doubles the compression efficiency of the previous standards for comparable quality levels according to the experiments in [5]. This improvement is a result of more complex syntactic structures and codec algorithms which unavoidably incur heavier computational burden on H.264 codecs. It was reported that the computational complexity of an H.264 encoder is about four to five times higher than that of an MPEG-2 encoder while the complexity of an H.264 decoder is about two to three times higher that of an MPEG-2 decoder [6]. Such high complexity would be challenging for some real time or mobile video applications.

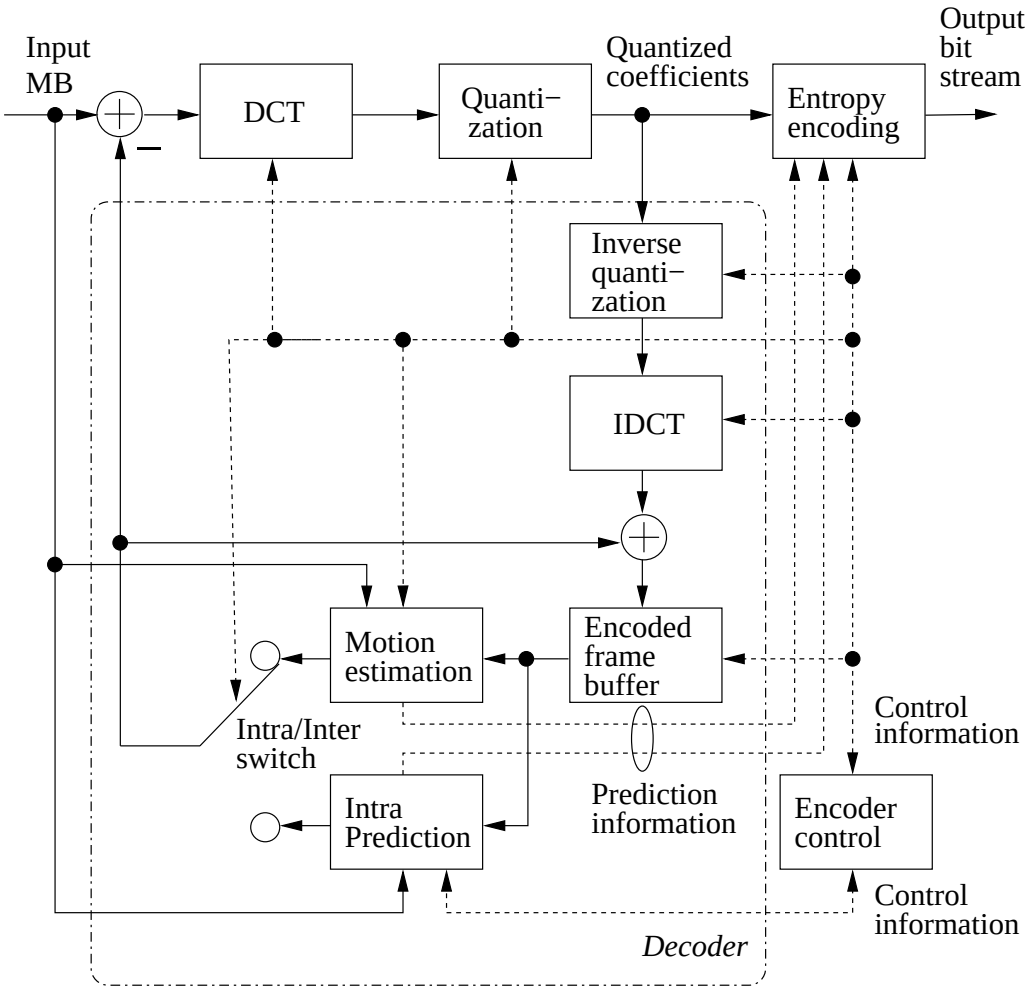


Figure 2.1: An H.264 encoder.

2.2.2 The Network Abstraction Layer

The network abstraction layer (NAL) of H.264 defines a minimum set of rules for transporting H.264 encoded videos in various network and protocol environments.

Network Abstraction Layer Units

The service data unit (SDU) of H.264 transport is named a NAL unit (NALU). The syntax elements, e.g., the motion vectors and the quantized DCT coefficients, of an encoded video are grouped into different NALUs. A NALU has a size of an integer number of bytes and contains one of the following items, possibly also some padding bits.

- an encoded slice or slice group (see Sec. 2.4),
- one of the three partitions of an encoded slice or slice group (see Sec. 2.4),
- a set of parameters applied to a whole video sequence,
- a set of parameters applied to a whole frame,
- supplemental enhancement information such as timing information for decoder control.

The above scheme helps to increase the robustness of H.264 encoded videos against transport related errors.

Transport of NALUs

NALUs are expected to be transported over packet oriented or bit oriented networks. In a packet oriented transport, the NALUs are individually passed down as SDUs to the transport layer. In a bit oriented transport, the NAL inserts delineating byte patterns between the NALUs so that the decoder extract them from the bit stream.

The NAL is designed to be simple and generic to accommodate a broad range of network infrastructures. Rather than imposing complicated efficiency and robustness requirements on the lower protocol layers, it leaves much space for application specific optimization of lower layer operations, such as NALU aggregation, NALU segmentation, packet scheduling and forward error correction (FEC).

2.3 Error Types in Video Communication

Fig. 2.2 illustrates a simplified model of video commutation systems. The network transports the encoded video to the decoder and may transport feedback information to the encoder.

There are three types of transport errors arising in the network. The first type is bit errors including bit inversions and bit erasures due to physical signal corruption. The second is losses of complete NALUs due to queue or buffer overflow within the network. The last type is late arriving of NALUs in real time applications which makes the NALUs in question to be considered lost.

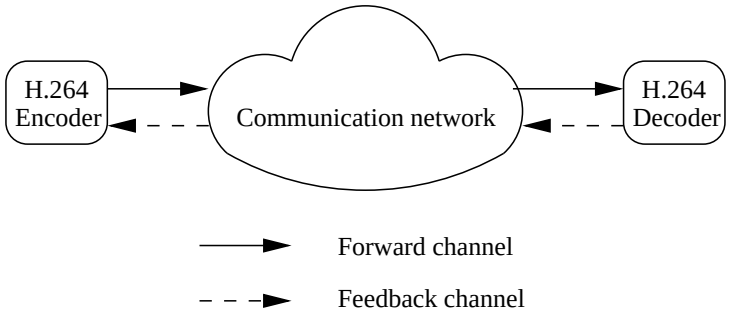


Figure 2.2: A video communication system.

2.3.1 Error Propagation

As the case with previous video coding standards, H.264 encoded videos are prone to error propagation in unreliable networks. This phenomenon is depicted in Fig. 2.3. It can be seen that transport errors could be temporally

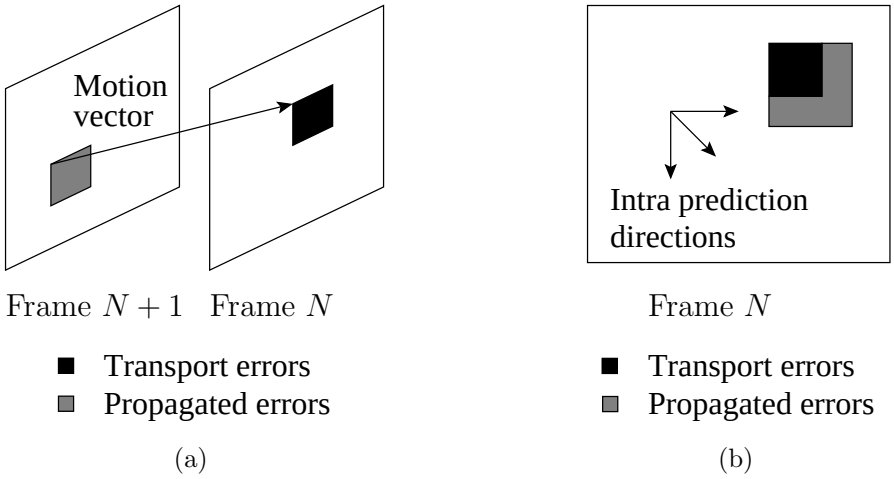


Figure 2.3: Error propagation due to (a) inter prediction, and (b) intra prediction.

and spatially propagated to even correctly received regions. Transport errors and propagated errors are collectively called transport-introduced errors in this dissertation.

2.3.2 Error Control Paradigms

In order to combat transport-introduced errors in video communications, error control measures are usually taken in different communication layers. Following are some common error control paradigms differentiated by the involved communication layers:

Application Layer Error Control This paradigm includes error resilient encoding in the video encoder and error concealment in the decoder. In error resilient encoding, the encoder sets the encoding parameters so that the encoded video is more resilient to transport-introduced errors while still maintaining an acceptable compression efficiency. In error concealment, the decoder estimates the erroneous video regions based on correctly received regions. In this paradigm, the operations of the lower layers are hidden from the application layer, but the application layer may obtain error statistics or per-NALU error status from the transport layer.

Lower Layer Error Control This paradigm uses only network-side content unaware error control measures such as retransmission and FEC. The same measures are usually applied indiscriminately to any transported contents.

All-layer Error Control This paradigm is a combination of the above two paradigms but the measures are taken independently in the application layer and the lower layers.

Crosslayer Error Control In this paradigm, the application layer and the lower layers use their error control tools in a coordinated way to achieve higher robustness of the video contents. Some unequal error protection (UEP) methods such as joint source channel coding (JSCC) and selective retransmission are examples of this paradigm.

The basis of this dissertation is application layer error control. This paradigm provides a flexible error control solution to video transport over any communication infrastructures, especially heterogeneous networks. If computing and communication resources are sufficient, the discussions in this dissertation can be extended to cross-layer designs for specific communication protocol stacks and network environments as in [7], [8].

2.3.3 Constraints on Application Layer Error Control

Two characteristics of video applications form major constraints on the applicability of various error control measures:

End-to-end Delay This is the maximum allowed delay between the acquisition and the display of a video frame. A small end-to-end delay would disallow certain encoding options such as the bi-prediction MB modes in which a MB is partly predicted from regions of a future frame.

Codec Complexity As stated in Sec. 2.2.1, the computational complexity of H.264 codecs is challenging for some real time or mobile video applications. Since error control measures tend to add extra burden to the codecs, the complexity of such measures has to be taken into consideration in these cases.

2.4 H.264 Error Control Tools

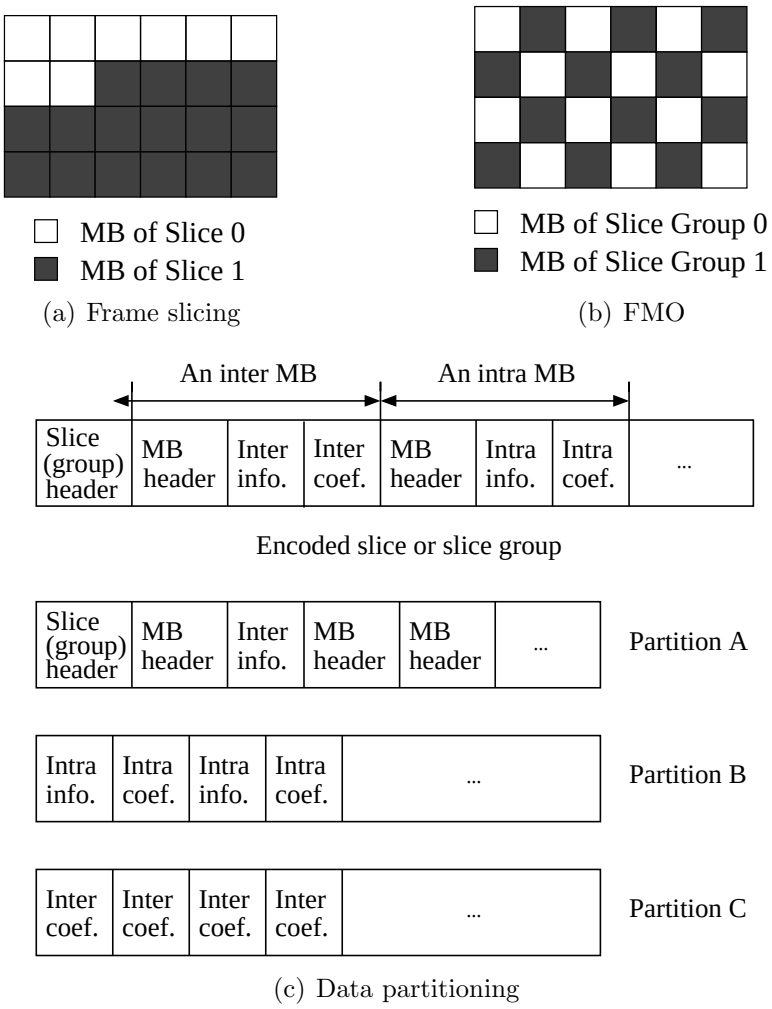
H.264 inherits the hybrid video coding framework, thus naturally most error control tools from the previous standards. At the same time, an important error control tool in H.263 and MPEG-4, named reversible variable length coding (RVLC), is discarded by H.264. RVLC [9] is a special entropy coding scheme in which a sequence of binary codewords can be decoded in either direction, which helps to recover data from bit streams or packets corrupted by bit errors. Instead of RVLC, H.264 adopts two context adaptive entropy coding schemes, context adaptive binary arithmetic coding (CABAC) and context adaptive variable length coding (CAVLC), in favor of higher compression efficiency. This choice reflects the trend of packet oriented communication and makes it difficult for the decoder to handle bit errors. Actually, the computational burden and the gain in compression efficiency do not fully justify bit error detecting and processing by the video decoder [7]. Generally, discarding NALUs with bit errors is a simple and efficient solution. Therefore, this dissertation considers only error control tools for NALU losses, regardless of whether the losses are caused by physical signal corruption, transmission queue overflow or delayed NALU transport.

2.4.1 Encoder-side Tools

The major error resilience tools at the encoder side are listed below. Note that the compression efficiency is normally compromised by using these tools due to either more residual redundancy or extra header bits, or both.

Information Segmentation In a packet lossy environment, the loss of a NALU that contains a whole video frame normally leads to an abrupt degradation of video quality in the lost frame and the coming frames even after error concealment. Information segmentation distributes the contents of a video frame to multiple NALUs to reduce the impairment caused by a single NALU loss.

H.264 provides three information segmentation schemes: frame slicing, flexible macroblock ordering (FMO) and data partitioning. Frame slicing and FMO apply to the raw video while data partitioning applies to the encoded syntax elements. Fig. 2.4 illustrates these segmentation schemes.



(a) Frame slicing

(b) FMO

(c) Data partitioning

Figure 2.4: The H.264 information segmentation schemes.

In frame slicing, a frame is divided into slices each containing some MBs that are successive in the raster scan order. Without data partitioning which is explained below, a slice is encapsulated in a NALU. The number of MBs in a slice is freely determined by the encoder and may vary from slice to slice. There is no syntax prediction, e.g., intra prediction or motion vector prediction, crossing slice borders so that a correctly received slice can be independently decoded.

FMO is similar to frame slicing in that a frame is divided into groups of MBs. In FMO, such a group is named a slice group and the MBs of it need not to be, and seldom are, successive in the raster scan order. A slice group is also encapsulated in a NALU unless data partitioning is used. FMO is essentially a MB interleaving scheme in which the correlation among spatially adjacent MBs could be better utilized by error concealment than in slicing.

Data partitioning encapsulates the syntax elements of an encoded slice or slice group in three NALUs of different importance degrees. Each of the NALUs is called a data partition. Partition A carries the header information and motion information of the slice or slice group. Partition B carries the prediction information and the quantized DCT coefficients of the intra MBs. Partition C carries the quantized DCT coefficients of the inter MBs. For a partly received slice, any of the three partitions, if received, could improve the error concealment result to some extent. In terms of potential contribution to error concealment, the three partitions can be ranked as A>B>C. The difference in the importance of the partitions forms an interface of UEP support to lower layers.

Intra Refreshing A MB can be encoded in either an intra mode or an inter mode as explained in Sec. 2.2.1. The MBs of the first frame of a video sequence have to be encoded in the intra modes. If the transport channel is reliable, most MBs in the following frames will be encoded in the inter modes. The reason is that statistically inter-frame similarity is stronger than intra-frame similarity in natural video sequences, thus inter MBs occupy less bits than intra MBs. In the presence of NALU losses, however, a common practice is to force some MBs to be encoded in the intra modes to alleviate inter-frame error propagation. Accompanying intra refreshing, intra MBs are commonly prevented from referencing to inter encoded regions so as to completely stop intra-frame propagation.

Reference Frame Selection H.264 allows an inter MB to select its reference frame from a number of previously encoded frames. In the deci-

sion, a MB can take into consideration the reliability of the reference regions in the reference frames for error resilience.

Redundant Slices Redundant slices are encoded slices that are not integral parts of a standard H.264 video. They serve as backups of some original slices in case of slice losses. A received redundant slice is decoded only if the corresponding original slice is lost. A redundant slice is normally encoded with larger quantization step sizes than the original slice so that its bandwidth consumption can be tolerated.

2.4.2 Decoder-side Concealment

At the decoder side, although video quality is inevitably degraded by NALU losses, concealment can be performed to make the reconstructed video sequence visually acceptable. Error concealment is possible because video compression can not remove all redundant information from raw video sequences.

2.5 Error Control in Delay and Complexity Constrained Applications

H.264 finds usage in video telephony, DVB [10] and many other applications. These applications are usually exposed to transport errors in real world environments. This dissertation is targeted at error control for delay and complexity constrained H.264 applications such as video telephony, video conferencing and live Internet video streaming [1].

[1] Although some Internet streaming services may allow a larger end-to-end delay than interactive services such as video telephony, it is conservative, albeit reasonable, to use only low delay encoding options to cope with the unpredictable transport delay of the packets.

2.5.1 Selected Error Control Tools

Not all the available error control tools introduced in Sec. 2.4 are suitable for low delay and low complexity applications. A rough analysis of the applicability of those tools is provided below.

The three information segmentation schemes generally do not significantly increase the encoder complexity and are not delay sensitive. However, effective implementation of FMO or data partitioning requires complicated error concealment algorithms at the decoder side. Furthermore, data partitioning is especially designed for UEP featured transport. Therefore, frame slicing is selected as a suitable information segmentation scheme for the target applications.

Intra refreshing has been widely used in applications of previous standards for error resilience. This tool is not delay sensitive. Additionally, it is simple to implement. Furthermore, sometimes it could even reduce the computational burden of the encoder since encoding a MB in an intra mode is normally faster than in an inter mode. For all its benefits, intra refreshing is adopted in this dissertation.

Reference frame selection is both delay and complexity sensitive. For low delay communications, referencing to future frames (in the display order) should be avoided. Moreover, the complexity of motion estimation, which is the most time consuming encoding operation, is proportional to the number of searched reference frames. Consequently in this dissertation a MB is only allowed to reference to the closest encoded frame.

Redundant slices are especially useful in region of interest (ROI) error resilience. ROIs are regions of a video that are more important or attractive to viewers than the other parts. Transporting redundant slices of ROIs makes them more robust against transport-introduced errors while keeping a small need for extra bandwidth. However, identifying ROIs in a video sequence is not a pure technical issue and may depend on the interest of individual viewers. For this reason, redundant slices are not considered hereafter.

Among all the existing error concealment methods, the simplest conceal-by-copy method is selected for the targeted applications. In this method, a lost MB is concealed by copying from the co-located MB in the previously decoded frame. Of course, the previously decoded frame may also contain

concealed regions or correctly received regions with propagated errors.

With the above analysis, the scope of this dissertation is limited to the joint usage of frame slicing, intra refreshing and conceal-by-copy concealment method to combat NALU losses.

The conceal-by-copy operation is straightforward and leaves little space for improvement. In addition, optimal frame slicing is usually a cross-layer design topic and requires precise information about some transport layer features such as packetization, comprehensive statistics of packet losses. For the purpose of simplicity, a fixed slicing scheme is used in this dissertation in which each slice is encapsulated in a NALU. It is further assumed that slices are subject to independent losses and the slice loss probability or loss ratio is invariant during the transport of a video sequence. Therefore, the actual topic of this dissertation is narrowed down to the efficient usage of the intra refreshing method for error resilience.

2.5.2 Previous Works

Although periodically encoding a whole frame in the intra modes can somehow stop error propagation, its usage is limited by the delay constraint. An intra encoded frame normally has a size several times larger than that of an inter frame. Consequently, in a low delay application, timely transporting the bits of the intra frames would require the instantaneous throughput of the transport channel to be quite dynamic or over provisioned. To achieve a more stable bit rate, it is better to force some MBs of each frame to be intra encoded. Following are various ways to decide the positioning of forced intra MBs for a good balance between compression efficiency and error resilience.

A simple method is to force the same number of intra MBs in all the frames of a video sequence. The percentage of intra MBs and equivalently the intra refreshing frequency of each MB [2], are heuristically determined in this case. For a given intra refreshing frequency, there are variant refreshing patterns specifying the positions of intra MBs in an individual frame. The H.263 test model [11] and H.264 reference codec [12] adopt random patterns. [13] tested some special patterns whose refreshing units are of different

[2]Depending on the context, a MB may refer to either a single MB or the position of a set of co-located MBs. Here the latter meaning is used.

shapes, e.g, 1×9 or 3×3 MBs. The results in [13] indicated that the intra refreshing frequency, rather than the pattern of refreshing, dominates the performance.

For more efficient usage of intra MBs, it was advised in [14], [15] to adapt the intra refreshing frequency of every MB to its activity. [16] further took into account the refreshing history of each pixel in intra refreshing.

Some works [17]–[20] combined the intra refreshing tool with the rate-distortion optimization (RDO) framework. RDO [20]–[22] is an optimization framework based on the method of Lagrange multipliers. It determines some important encoding options and parameters for each MB so that an optimal global distortion-rate (DR) point of the encoded video can be closely achieved. The combination of intra refreshing and RDO is realized in the encoder by estimating the expected distortion of the reconstructed MB after random slice losses and error concealment. The combined framework is named error resilient rate-distortion optimization (ER-RDO), or loss aware rate-distortion optimization (LA-RDO). ER-RDO performs better than other intra refreshing methods at the cost of higher computational complexity [1], [23]. A detailed description about the RDO and the ER-RDO in H.264 is provided in Chap. 3.

Chapter 3

The H.264 Error Resilient Rate-Distortion Optimization

This chapter describes the existing ER-RDO framework in the H.264 reference codec. The H.264 reference codec [24] which is named the joint model (JM) has been developed and upgraded from the beginning of the standardization work. The ER-RDO part of the JM has become relatively stable and serves as a widely accepted basis for adaptation and improvement.

A short introduction to the rate-distortion theory leads this chapter, followed by analyses of RDO and ER-RDO. Finally some implementation details of the H.264 ER-RDO are explained.

3.1 The Rate-Distortion Theory

The rate-distortion theory studies the performance bounds of lossy compression of various kinds of information sources. Below are the concepts and conclusions from the theory that are fundamental to video compression. More details can be found in [25]–[27].

In video coding, a commonly used information source model is the discrete memoryless source (DMS) model. The application of DMSs in video compression is illustrated in Sec. 3.3. An information source S is a DMS if

1. it emits a symbol S_n at the n-th time instant,
2. the value of S_n is taken from a finite source alphabet $A = \{A_1, A_2, \ldots, A_\Theta\}$ where Θ is the size and $A_i \in \mathbb{R} \, \forall \, i$,
3. $S_1, S_2, \ldots$ are independent, identically distributed random variables sharing the same symbol probabilities $P_i \equiv P\{S_n = A_i\}$.

Consider a source coding scheme in which the original sequence $S_1, S_2, \ldots$ is first encoded and then decoded as $\check{S}_1, \check{S}_2, \ldots$ where $\check{S}_n \in A$. Let $\boldsymbol{S_n}$ and $\check{\boldsymbol{S}}_{\boldsymbol{n}}$ denote the vectors $(S_1, S_2, \ldots, S_n)$ and $(\check{S}_1, \check{S}_2, \ldots, \check{S}_n)$, respectively. Suppose for each sample pair of $(S_n, \check{S}_n)$, a nonnegative value $d(S_n, \check{S}_n)$, named a distortion measure, is defined to measure the difference between S_n and$\check{S}_n$ and the definition of the function d is extended to measure the difference between $\boldsymbol{S_n}$ and $\check{\boldsymbol{S}}_{\boldsymbol{n}}$ by

$$d(\boldsymbol{S_n}, \check{\boldsymbol{S}}_{\boldsymbol{n}}) = \sum_{i=1}^{n} d(S_i, \check{S}_i). \tag{3.1}$$

The distortion-rate (DR) function of the source is then defined by

$$D(R) = \lim_{n \to \infty} \frac{1}{n} \min E(d(\boldsymbol{S_n}, \check{\boldsymbol{S}}_{\boldsymbol{n}})) : I(\boldsymbol{S_n}; \check{\boldsymbol{S}}_{\boldsymbol{n}}) \leq nR \text{ and } 0 \leq R, \tag{3.2}$$

where $E(\cdot)$ being the mathematical expectation operator and $I(\cdot;\cdot)$ being the mutual information operator. The minimization takes place over all possible encoding schemes, each of which is characterized by a set of conditional probabilities $P(\check{\boldsymbol{S}}_{\boldsymbol{n}}|\boldsymbol{S_n})$ for all the possible values of $\check{\boldsymbol{S}}_{\boldsymbol{n}}$ and$\boldsymbol{S_n}$.

The rate-distortion theory yields

1. the support of $D(R)$ is $[0, H(S_n)]$ where $H(S_n)$ being the entropy of S_n;
2. The function $D(R)$ is continuous, decreasing and convex on its support interval;
3. The range of $D(R)$ is $[0, D_{max}]$ where $D_{max} = \min_j \sum_{i=1}^{\Theta} P_i \cdot d(A_i, A_j)$;
4. and $D(R)$ is the smallest achievable average distortion (or distortion for short) for a given average number of bits per symbol (or rate for short) R through an ideal source coding scheme.

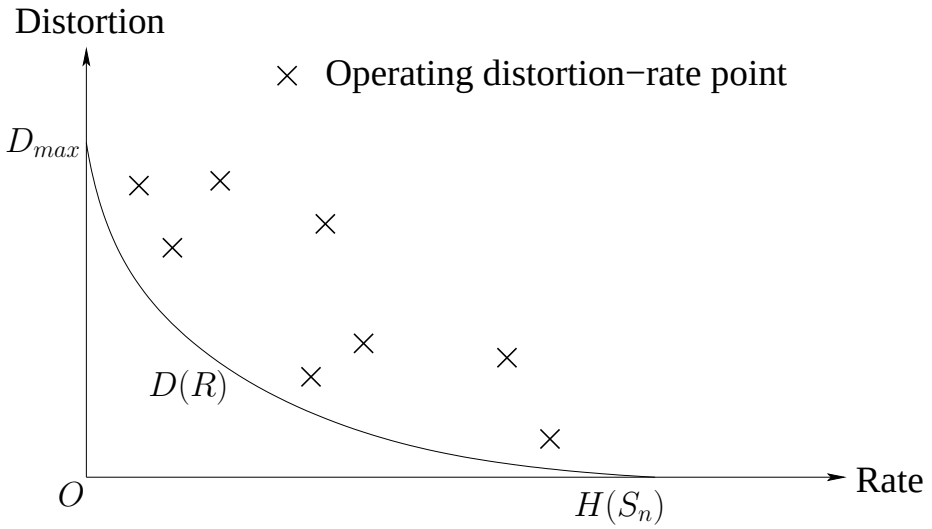


Figure 3.1: The DR function and the operating DR points of a DMS.

The major characteristics of $D(R)$ are illustrated in Fig. 3.1. Note that $D(R)$ only serves as a lower bound of the finite set of all the achievable DR points of a real world source coding system. The achievable points of a real world system are called the operating DR points which are depicted in Fig. 3.1 by $\times$ for comparison to $D(R)$.

3.2 Distortion Measures in Video Coding

Let X and $\check{X}$ denote two corresponding video regions with the same dimensions. They each contains totally Φ pixels of Y, Cb and Cr components. The pixels are indexed by X_i and $\check{X}_i$. The following distortion measures are commonly used to quantify the difference between X and $\check{X}$.

- sum of absolute difference (SAD) $= \sum_{i=1}^{\Phi} |X_i - \check{X}_i|$;
- mean absolute difference (MAD)$= \frac{1}{\Phi} \sum_{i=1}^{\Phi} |X_i - \check{X}_i|$;

- sum of squared error (SSE) $= \sum_{i=1}^{\Phi}(X_i - \check{X}_i)^2$;
- mean squared error (MSE)$= \frac{1}{\Phi}\sum_{i=1}^{\Phi}(X_i - \check{X}_i)^2$.

It can be seen that SAD and MAD, SSE and MSE are pairwise equivalent.

3.3 Rate-Distortion Optimization in Video Encoding

RDO is a powerful tool in video encoder control. Although [20]–[22] have explained the theory and applications of RDO in video compression, a more rigorous analysis of RDO is provided below to better reveal its rationality. The analysis is valid not only for H.264, but also for the other similar standards.

3.3.1 Problem Formulation

In hybrid video coding the number of adjustable encoding parameters is very large. Besides the video sequence level and frame level configurations such as the slicing scheme and the entropy coding method, there are many encoding options for an individual MB. For example, the encoder has to determine for each inter MB the sizes of motion prediction blocks and the motion vectors for each of the blocks. In addition, there exists inter-dependence among the parameters of the same MB and among the parameters of different MBs. Such a complex system is extremely difficult to model and optimize. In practice, the optimization task is divided into smaller subtasks and in each subtask, only a small number of parameters is adjusted. In this dissertation, RDO specifically refers to the parameter optimization for MB level bit allocation and mode selection. Consequently, the following assumptions are adopted for the discussion in this section.

1. Each MB in a frame is assumed to be an output symbol of an individual

DMS. Such a symbol is a vector composed of the luma and the chroma pixels of a MB.

2. The DMSs are independent and have different statistical characteristics.

3. Only two parameters are used to control the encoding result of a MB: the prediction mode (or mode for short) and the quantization step size (QS). The two parameters can be adjusted independently. The intra modes are not considered in this section since they seldom appear in encoded videos unless in error resilience scenarios.

With the above assumptions, the following notation is defined.

1. There are totally K MBs in the current frame. The k-th MB is denoted by MB_k.

2. Due to a limited bandwidth, the current frame is allocated with a total budget of R_0 bits.

3. MB_k is encoded in a mode m_k and a QS q_k. The values of m_k and q_k are taken from the set of all the usable modes $M = \{M_1, M_2, \ldots, M_\Upsilon\}$ and the set of all the usable QSs $Q = \{Q_1, Q_2, \ldots, Q_\Omega : Q_1 < Q_2 < \ldots < Q_\Omega\}$, respectively. All the parameters of the current frame form a vector $\boldsymbol{u} = (m_1, q_1, m_2, q_2, \ldots, m_K, q_K)$.

4. The achievable number of encoded bits of MB_k, as a function of m_k and q_k, is denoted by $r_k(m_k, q_k)$.

5. The achievable distortion between the original and the reconstructed MB_k, as a function of m_k and q_k, is denoted by $d_k(m_k, q_k)$ for a given distortion measure.

6. The DR function of MB_k is denoted by $D_k(R_k)$. The $D_k(R_k)$'s can not be found due to the lack of perfect mathematical models for natural video signals.

The RDO problem is now formulated as

$$\arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} d_k(m_k, q_k), \text{ subject to } \sum_{k=1}^{K} r_k(m_k, q_k) \leq R_0. \tag{3.3}$$

Solving Eq. 3.3 is very complex even if fast algorithms such as dynamic programming are used. In practice, with reasonable assumptions, sub-optimal solutions can usually be found with affordable computational complexity. Following are the steps of the most popular RDO method in hybrid video coding.

3.3.2 Step 1: Search for the Optimal Quantization Step Sizes

It can be seen that for a MB, say MB_k, and a given value of q_k, the possible values of m_k determine a set of operating points in the DR plane of MB_k. Let $O_k(q_k)$ denote this set of operating points. Experiments show that q_k has a more significant effect than m_k on the rate and the distortion. Therefore it can be reasonably assumed that $O_k(q_k)$ is a relatively compact set of points in the DR plane for a given q_k value, as illustrated in Fig. 3.2. Since the target of Eq. (3.3) is minimizing the global distortion, a point with the smallest distortion from $O_k(q_k)$ can be initially taken as the representative of $O_k(q_k)$. The mode of such a representative point of MB_k is now denoted by $m_{k,\mathrm{INIT}}(q_k)$. The problem is then simplified to

$$\arg\min_{\boldsymbol{q}} \sum_{k=1}^{K} d_k(m_{k,\mathrm{INIT}}(q_k), q_k), \text{ subject to} \sum_{k=1}^{K} r_k(m_{k,\mathrm{INIT}}(q_k), q_k) \leq R_0, \tag{3.4}$$

where $\boldsymbol{q} = (q_1, q_2, \ldots, q_K)$.

Experiments show that in most cases, the relation between r_k and q_k as defined by $r_k(m_{k,\mathrm{INIT}}(q_k), q_k)$ can be well fitted by a decreasing and convex function $\tilde{r}_k(\tilde{q}_k)$ for $\tilde{q}_k \in [Q_1, Q_\Omega]$. Similarly, $d_k(m_{k,\mathrm{INIT}}(q_k), q_k)$ can also be fitted by an increasing and convex function $\tilde{d}_k(\tilde{q}_k)$. The two functions are separately called the operating rate-quantization function and the operating distortion-quantization function of MB_k. They are depicted in Fig. 3.3.

It is then assumed that the QSs can be adjusted continuously in the encoder and the functions $\tilde{r}_k(\tilde{q}_k)$ and $\tilde{d}_k(\tilde{q}_k)$ can precisely describe the achievable performance bounds of the encoder for MB_k. It can be derived from the characteristics of $\tilde{r}_k(\tilde{q}_k)$ and $\tilde{d}_k(\tilde{q}_k)$ that the achievable distortion is a decreasing and convex function of the achievable rate. Let $\tilde{D}_k(R_k)$ denote the function. It is named the operating DR function of MB_k. If the encoder

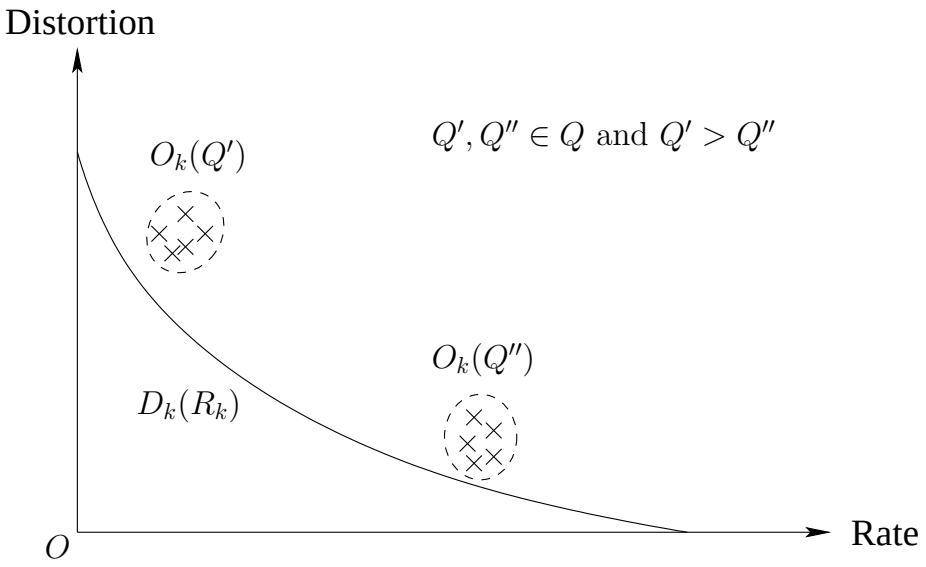


Figure 3.2: The operating DR point sets dependent on the QS.

performs well enough, $\tilde{D}_k(R_k)$ should approximate $D_k(R_k)$, as illustrated in Fig. 3.4.

With the above simplification, the optimization problem is now

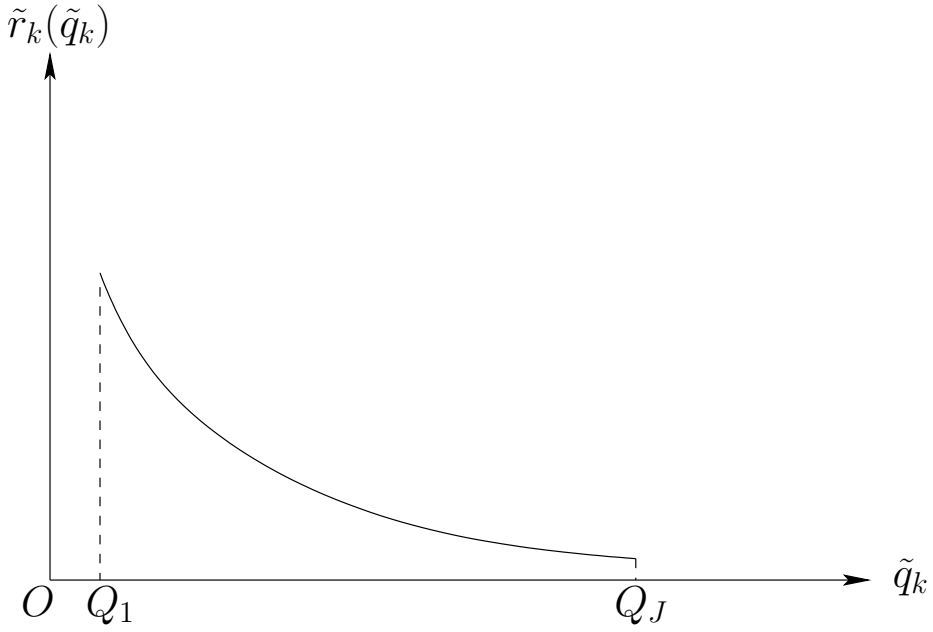
$$\arg\min_{\boldsymbol{r}} \sum_{k=1}^{K} \tilde{D}_k(R_k), \text{ subject to } \sum_{k=1}^{K} R_k = R_0, \tag{3.5}$$

where $\boldsymbol{r} = (R_1, R_2, \ldots, R_K)$. Considering the convexity of the $\tilde{D}_k(R_k)$'s, the Lagrange multiplier method is used. A global cost function is first constructed as

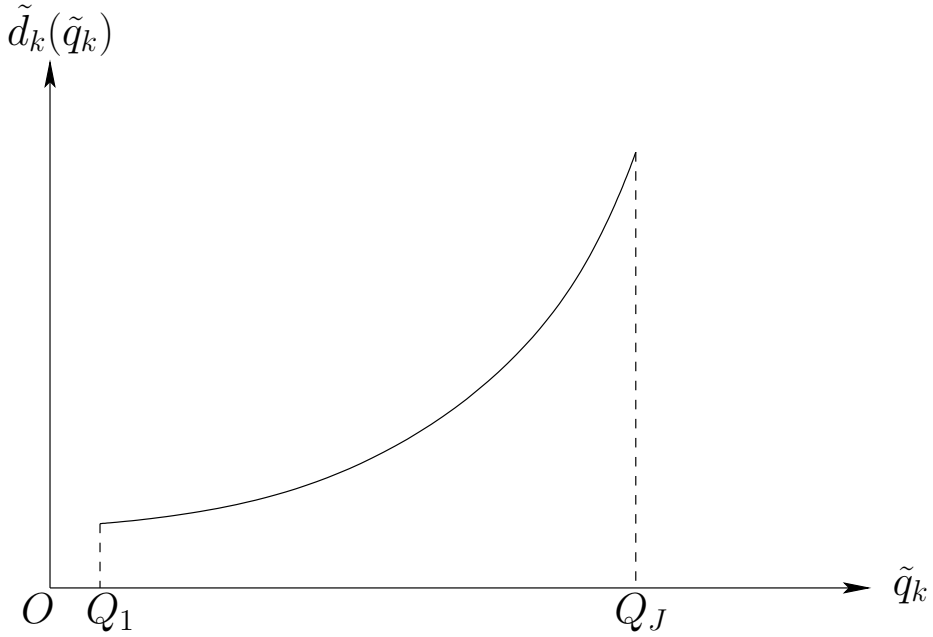
$$J_{\text{QUANT}} = \sum_{k=1}^{K} \tilde{D}_k(R_k) + \lambda_{\text{QUANT}} \sum_{k=1}^{K} R_k, \tag{3.6}$$

where λ_{QUANT} is the Lagrange multiplier. Solving Eq. (3.5) is then equivalent to minimizing J_{QUANT}. The Lagrange multiplier method yields

$$\begin{cases} \dfrac{d\tilde{D}_k(R_k)}{dR_k} = -\lambda_{\text{QUANT}} \,\forall\, k, \\ \displaystyle\sum_{k=1}^{K} R_k = R_0. \end{cases} \tag{3.7}$$



(a)



(b)

Figure 3.3: An operating (a) rate-quantization function, and (b) distortion-quantization function.

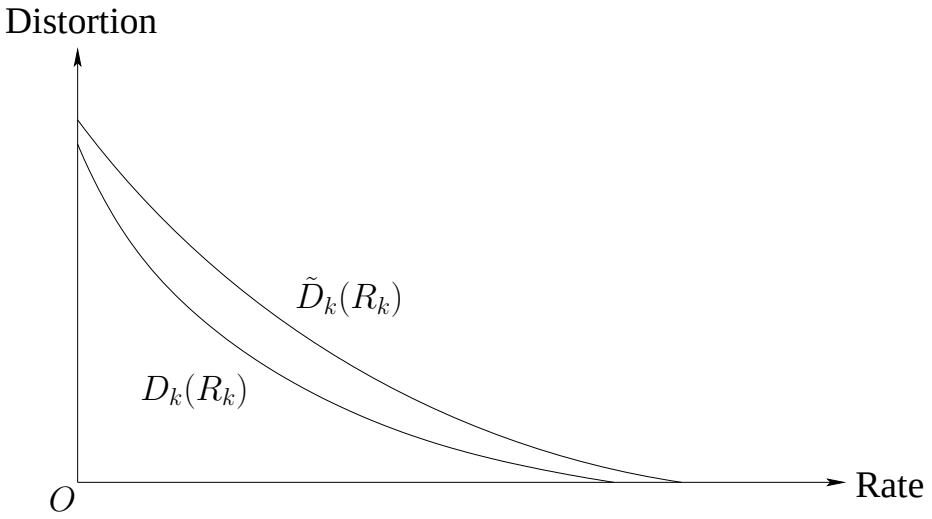


Figure 3.4: An operating DR function.

For the known $\tilde{D}_k(R_k)$'s, the solution to Eq. (3.7) can be analytically calculated. Let $(\hat{\lambda}_{\text{QUANT}}, \hat{R}_1, \hat{R}_2, \ldots, \hat{R}_K)$ denote the solution. For MB_k, the value of the corresponding $\tilde{q}_k$ can be found from the relation $\tilde{r}_k(\tilde{q}_k) = R_k$. Let $\hat{\tilde{q}}_k$ denote the corresponding value. Next, $\hat{\tilde{q}}_k$ is mapped to the nearest element in Q which is denoted by $\hat{q}_k$. The $\hat{q}_k$'s are now taken as components of the solution to Eq. (3.3).

3.3.3 Step 2: Search for the Optimal Modes

Since $m_{k,\text{INIT}}(\hat{q}_k)$ is only a tentative solution in the mode selection, improvement is possible through solving

$$\arg\min_{\boldsymbol{m}} \sum_{k=1}^{K} d_k(m_k, \hat{q}_k), \text{ subject to} \sum_{k=1}^{K} r_k(m_k, \hat{q}_k) \leq R_0, \tag{3.8}$$

where $\boldsymbol{m} = (m_1, m_2, \ldots, m_K)$.

Assuming the DR points within $O_k(q_k)$ can be tightly enclosed by a smooth and convex hull. It is also assumed that all the points on the hull are now achievable. For MB_k, the corresponding DR point of the solution to Eq. (3.8) is obviously on a segment of its hull. Let $\acute{d}_k(\acute{r}_k)|\tilde{q}_k$ denote the

segment. By using $\tilde{q}_k$ instead of q_k as the condition, it is again assumed that the encoder has a continuous domain for the QS, and for each QS value, a compact point set of achievable DR points like $O_k(q_k)$ exists. The function $\acute{d}_k(\acute{r}_k)|\tilde{q}_k$ is called the quantization constrained DR function of MB_k, and shown in Fig. 3.5.

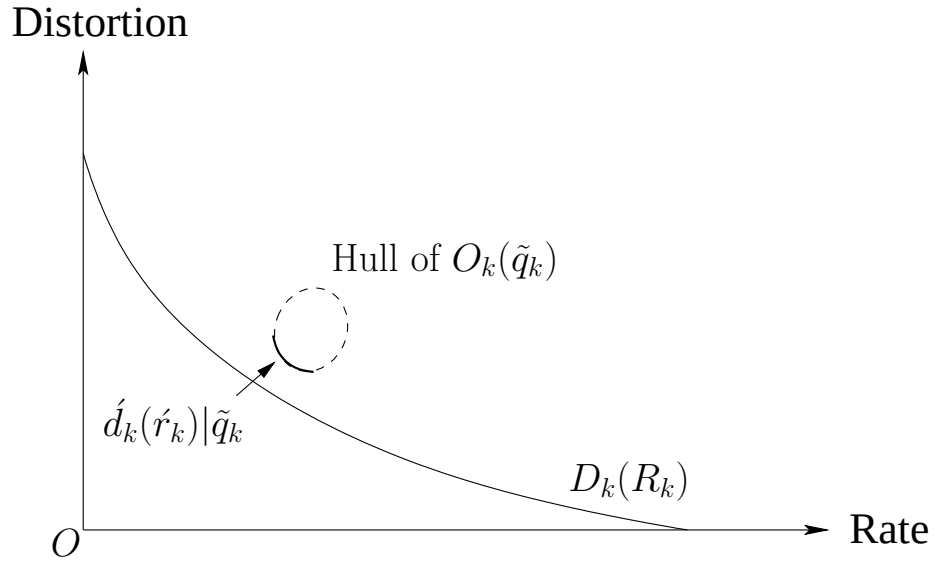


Figure 3.5: A quantization constrained DR function.

The problem in Eq. (3.8) is then changed to

$$\arg\min_{\acute{\boldsymbol{r}}} \sum_{k=1}^{K} \acute{d}_k(\acute{r}_k)|\hat{q}_k, \text{ subject to} \sum_{k=1}^{K} \acute{r}_k = R_0, \tag{3.9}$$

where $\acute{\boldsymbol{r}} = (\acute{r}_1, \acute{r}_2, \ldots, \acute{r}_K)$ and the domains of the $\acute{r}_k$'s are dependent on the $\hat{q}_k$'s. The problem in Eq. (3.9) can be solved in the same way as the first step. A global cost function

$$J_{\mathrm{MODE}} = \sum_{k=1}^{K} \acute{d}_k(\acute{r}_k)|\hat{q}_k + \lambda_{\mathrm{MODE}} \sum_{k=1}^{K} \acute{r}_k, \tag{3.10}$$

has to be minimized, where λ_{MODE} is the Lagrange multiplier. The solution

should then satisfy

$$\begin{cases} \dfrac{d\acute{d}_k(\acute{r}_k)|\hat{q}_k}{d\acute{r}_k} = -\lambda_{\text{MODE}} \,\forall\, k, \\ \displaystyle\sum_{k=1}^{K} \acute{r}_k = R_0. \end{cases} \tag{3.11}$$

Similar to Eq. (3.7), the solution of Eq. (3.11) can be calculated analytically. Let $(\hat{\lambda}_{\text{MODE}}, \hat{\acute{r}}_1, \hat{\acute{r}}_2, \ldots, \hat{\acute{r}}_K)$ denote the solution. For MB_k, among all the operating points in $O_k(\hat{q_k})$, one that has the smallest distance to $(\hat{\acute{r}}_k, \acute{d}_k(\hat{\acute{r}}_k)|\hat{q}_k)$ is selected as the optimal operating point.. Let $\hat{m}_k$ denote the corresponding mode.

Finally, the solution to the original problem in Eq. (3.3) is provided as $\hat{\boldsymbol{u}} = (\hat{m}_1, \hat{q}_1, \hat{m}_2, \hat{q}_2, \ldots, \hat{m}_K, \hat{q}_K)$.

3.3.4 Further Simplification

In the above two steps, the original discrete problem is largely replaced by two analytic problems. Although such a replacement reduces the average search depth, the computational complexity is still very high. There are two reasons for the high complexity. First, all the operating points have to be known in advance, i.e.,the MBs have to be tested with all the possible combinations of modes and QSs to obtain the corresponding syntax elements and the reconstructed MB so that the rate and the distortion data can be collected . Second, the curve fitting computation needed for determining the operating functions is also complex. To lower the computational burden, the following treatment is normally practiced in stead.

Simplification of Step 1

In Step 1, there are two simpler methods to find the optimal QSs.

In the first method, all the MBs of the current frame are assumed to have the identical statistics. Therefore they should be assigned with the same QS according to Eq. (3.7). Let q denote the global QS. With a frame level rate-quantization model such as those in [28], [29], the value of q can be calculated

from R_0. Testing other QS values is not needed in this method. The JM uses this method.

In the second method, curve fitting is still required to obtain the operating functions. However, for lower complexity, only a small number of QSs are tested as in [30]. With the collected operating data, the QSs are calculated from Eq. (3.7). It is not used in this dissertation since the computational burden is still unfordable.

Simplification of Step 2

In the second step, a MB is anyway tested in all the available modes since the number of the modes is much smaller than that of the available QSs. E.g., H.264 has 52 QSs, but only 5 inter modes.

Since all the modes have to be tested, the only possibility of simplification is to avoid the curve fitting for $\acute{d}_k(\acute{r}_k)|\hat{q}_k$ and the analytic solution to Eq. (3.11).

If the value of $\hat{\lambda}_{\text{MODE}}$ can be obtained without solving Eq. (3.11), the rate constraint in Eq. (3.11) can be removed and the problem can be simplified to some extent. In the following, an indirect way to find $\hat{\lambda}_{\text{MODE}}$ is described.

It is now assumed that the encoder is ideal in a sense that each point on $D_k(R_k)$ is achievable through and only through a specific $\tilde{q}_k$ value. It can then be inferred that the hull of $O_k(\tilde{q}_k)$ or $\acute{d}_k(\acute{r}_k)|\tilde{q}_k$ is tangent to $D_k(R_k)$, as in Fig. 3.6. Let $T_k(\tilde{q}_k)$ denote the tangent point. It is shown below that the rate of $T_k(\hat{q}_k)$ is just a component of the solution to Eq. (3.11).

In Fig. 3.7, the curve of $\acute{d}_k(\acute{r}_k)|\hat{q}_k$ contains the candidate operating points for Eq. (3.11). If a point other than $T_k(\hat{q}_k)$, e.g., the point C, is selected, it has to lie above $D_k(R_k)$ because of the convexity of $\acute{d}_k(\acute{r}_k)|\hat{q}_k$. It can be seen that on the curve of $D_k(R_k)$, the points between E and F have both smaller rates and smaller distortions than C. Since the whole curve of $D_k(R_k)$ is achievable, C can not be an optimal operating point. Therefore, $T_k(\hat{q}_k)$ is the optimal operating point of MB_k.

Because

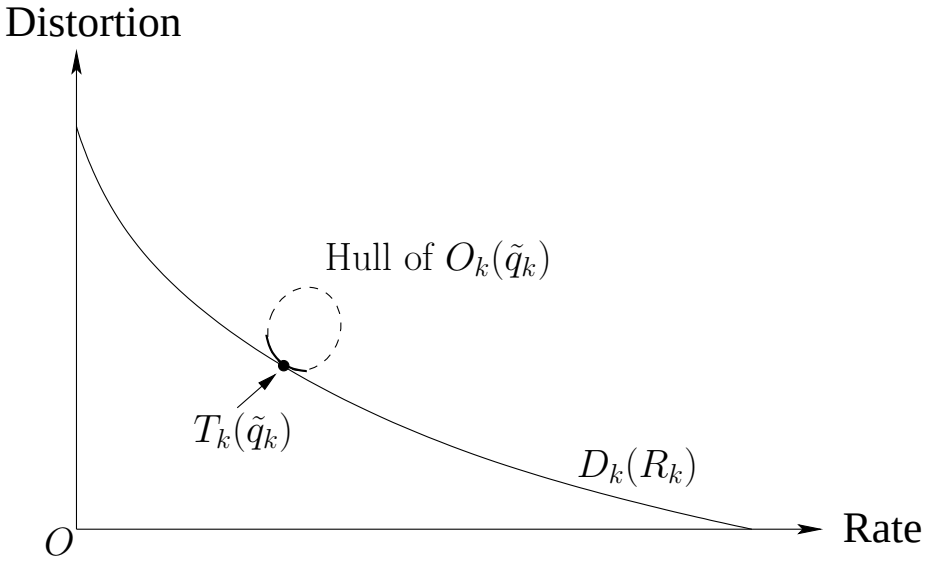


Figure 3.6: The tangency between an unconstrained DR function and one of the corresponding quantization constrained DR functions.

1. $T_k(\hat{q}_k)$ is close to $(\hat{R}_k, \tilde{D}_k(\hat{R}_k))$ in the scale of the domain and range of $\tilde{D}_k(\hat{R}_k)$,
2. $\acute{d}_k(\acute{r}_k)|\hat{q}_k$ is tangent to $D_k(R_k)$ at $T_k(\hat{q}_k)$,
3. $\tilde{D}_k(R_k)$ closely approximates $D_k(R_k)$,

the slope of $\acute{d}_k(\acute{r}_k)|\hat{q}_k$ at $T_k(\hat{q}_k)$ can be approximated by the slope of $\tilde{D}_k(R_k)$ at $(\hat{R}_k, \tilde{D}_k(\hat{R}_k))$. From Eq. (3.7) and Eq. (3.11), it can be seen that the slopes of these curves at the optimal points simply equal the negatives of the corresponding Lagrange multipliers. Therefore,

$$\hat{\lambda}_{\text{MODE}} = \hat{\lambda}_{\text{QUANT}}. \tag{3.12}$$

Although $\hat{\lambda}_{\text{MODE}}$ can be approximated by $\hat{\lambda}_{\text{QUANT}}$, the value of $\hat{\lambda}_{\text{QUANT}}$ can not be precisely obtained without thoroughly testing all the MBs as stated before. The so far prevailing method was introduced in [31]. It establishes a statistical relation between $\hat{\lambda}_{\text{QUANT}}$ and q. The relation is fixed by testing many video sequences with various characteristics. The testing process is described blow.

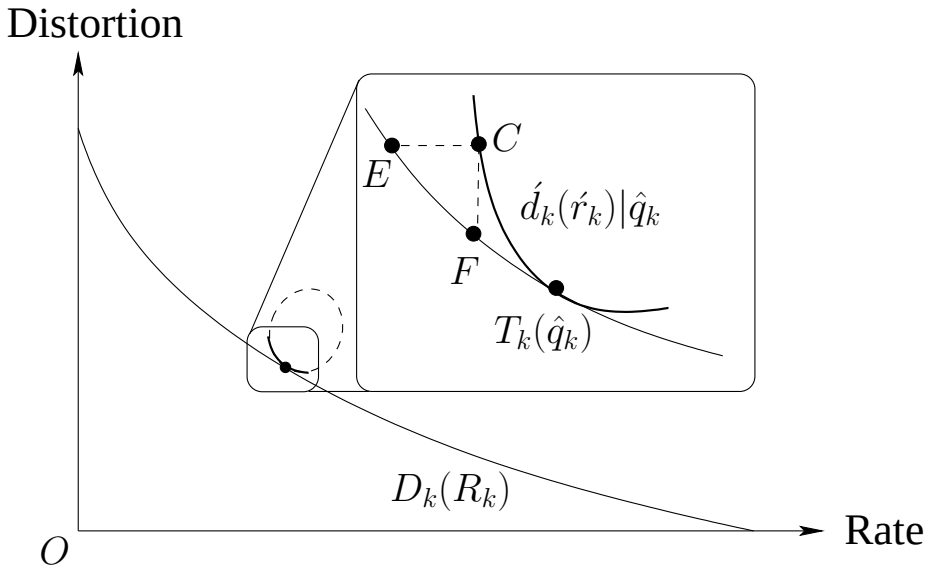


Figure 3.7: The tangent point as the optimal operating point.

For each MB, a variable λ_{MB} is assigned with a certain value. The MB is then tested with all the available QSs and in all the available modes so that all the operating points can be obtained. Let D_{MB} and R_{MB} denote the distortion and rate of the MB. Obviously, D_{MB} and R_{MB} are functions of the QS q_{MB}. If λ_{MB} is regarded as the Lagrange multiplier of a problem similar to that in Eq. (3.7), a q_{MB} value that minimizes the local cost function $J_{\text{MB}} = D_{\text{MB}} + \lambda_{\text{MB}} R_{\text{MB}}$ can be viewed as a component of the solution to the same problem. By testing the MBs of a large number of test video sequences with a variety of λ_{MB} values, enough samples can be collected to establish a statistical relation between λ_{MB} and q_{MB}. The relation is denoted by

$$\lambda = \Lambda(q), \tag{3.13}$$

where λ is the Lagrange multiplier in both QS and mode selection and q is the global QS. The usage of λ is the same as the above testing process except that the variable is the mode. In detail, among all the available modes of MB_k, a mode minimizing the MB level cost function

$$J_k = d_k(m_k, q) + \lambda r_k(m_k, q), \tag{3.14}$$

is selected as the optimal mode. Therefore, the curve fitting for the operating functions and the analytical solution to Eq. (3.11) are avoided.

3.4 From Rate-Distortion Optimization to Error Resilient Rate-Distortion Optimization

With slice losses, the video decoder has to use reference frames that contain transport-introduced errors, as explained in Sec. 2.3.1. Because of the randomness of slice losses, the encoder has no precise information about the contents of the reconstructed MB in the decoder. Therefore the applicability of RDO in such a situation has to be checked.

3.4.1 Problem Formulation

The aim of RDO is to minimize the overall distortion of a video frame for a given rate. The rate reflects the network resource consumption and the distortion reflects the reconstructed video quality. In the presence of slice losses, the rate generated by the encoder may be different from the rate received by the decoder, but the expected network resource consumption can still be assumed to be positively related to the rate. Similarly, the expected distortion can be used to measure the quality of the reconstructed video. Therefore, the problem in Eq. (3.3) can be modified to

$$\arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} E(d_k(m_k, q_k)), \text{ subject to } \sum_{k=1}^{K} r_k(m_k, q_k) \leq R_0, \tag{3.15}$$

for slice lossy transport. In this new problem, both the inter and the intra modes should be considered.

3.4.2 Applicability of Rate-Distortion Optimization in Slice Lossy Environments

Let p denote the slice loss ratio and $D_{k,\mathrm{CON}}$ denote the distortion caused by concealment of MB_k. Note that $D_{k,\mathrm{CON}}$ is independent of the encoding parameters.

For an intra setting (m_k, q_k), the expected distortion in a slice lossy environment is

$$E(d_k(m_k, q_k)) = (1-p)d_k(m_k, q_k) + pD_{k,\mathrm{CON}}. \tag{3.16}$$

For an inter setting (m_k, q_k), the expected distortion in a slice lossy environment is

$$E(d_k(m_k, q_k)) = (1-p)d_{k,\mathrm{SUM}}(m_k, q_k) + pD_{k,\mathrm{CON}}, \tag{3.17}$$

where $d_{k,\mathrm{SUM}}(m_k, q_k)$ is the joint distortion of the compression errors and the transport-introduced errors in the referenced regions. Experiments show that if SSE or MSE is used as the distortion measure, $d_{k,\mathrm{SUM}}(m_k, q_k)$ can be decomposed by

$$d_{k,\mathrm{SUM}}(m_k, q_k) = d_k(m_k, q_k) + D_{k,\mathrm{PROP}}, \tag{3.18}$$

where $D_{k,\mathrm{PROP}}$ is an item which does not change much with m_k and q_k. The term $D_{k,\mathrm{PROP}}$ can be regarded as the distortion caused by the transport-introduced errors in the referenced regions. It is treated below as a constant. With Eq. (3.18), Eq. (3.17) can be rewritten as

$$E(d_k(m_k, q_k)) = (1-p)d_k(m_k, q_k) + (1-p)D_{k,\mathrm{PROP}} + pD_{k,\mathrm{CON}}. \tag{3.19}$$

Without transport-introduced errors, the distortions of an intra mode and an inter mode are solely controlled by the QS, thus roughly equal each other. With slice losses, however, $D_{k,\mathrm{PROP}}$ normally causes a significant difference between the expected distortions of them. At the same time, the intra mode normally generate far more bits than the inter mode does. As a result, the set of the DR points for a give q_k value is no longer compact and the points for different q_k values are more heavily mixed together. Therefore, the RDO framework is not equally suitable for the problem in Eq. (3.15).

3.4.3 Error Resilient Rate-Distortion Optimization Principles

Although RDO can not be directly used in slice lossy environments, some slight modification can make it more suitable for error resilience. The modified RDO is called ER-RDO in this dissertation.

With the observation that even in error resilient encoding, the majority of the MBs of a video sequence are inter MBs, ER-RDO almost duplicates RDO and takes the intra modes into account only in its very last operation.

In ER-RDO, the solution to Eq. (3.15) is also divided into two consecutive steps. The first step is to find the optimal values for the q_k's and the second one for the m_k's. When the intra modes are excluded, the problem can be rewritten as

$$\begin{aligned} \arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} [(1-p)d_k(m_k, q_k) + (1-p)D_{k,\mathrm{PROP}} + pD_{k,\mathrm{CON}}], \\ \text{subject to} \sum_{k=1}^{K} r_k(m_k, q_k) \leq R_0. \end{aligned} \tag{3.20}$$

Since $D_{k,\mathrm{PROP}}$ and $D_{k,\mathrm{CON}}$ are constants, it is simplified to

$$\arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} d_k(m_k, q_k), \text{ subject to} \sum_{k=1}^{K} r_k(m_k, q_k) \leq R_0. \tag{3.21}$$

The problem is identical to that in Eq. (3.3). Therefore, the inter-only solution $\hat{\boldsymbol{u}}$ is also the same as in error free transport.

If an intra mode has a sufficiently low rate or expected distortion, it can be a better choice than the inter modes. Empirically, ER-RDO uses a cost function

$$J_{k,\mathrm{LOSS}} = E(d_k(m_k, q)) + \lambda_{\mathrm{LOSS}}\, r_k(m_k, q), \tag{3.22}$$

to select the optimal mode among all the intra and inter modes, where λ_{LOSS} is the weight of the rate and to be fixed.

For inter-only mode selection,

$$J_{k,\mathrm{LOSS}} = (1-p)d_k(m_k, q) + (1-p)D_{k,\mathrm{PROP}} + pD_{k,\mathrm{CON}} + \lambda_{\mathrm{LOSS}}\, r_k(m_k, q). \tag{3.23}$$

By comparing Eq. (3.23) with Eq. (3.14), it can be found that

$$\lambda_{\mathrm{LOSS}} = (1-p)\lambda, \tag{3.24}$$

guarantees the results of the two mode selection methods are compatible in inter-only mode selection.

As a conclusion, ER-RDO can

1. switch a MB to an intra mode if the mode generates a sufficiently low rate or expected distortion;

2. for the set of inter encoded MBs, maintain an optimality comparable to that of RDO .

It can be seen that the handling of the intra modes in ER-RDO is rather empirical although the performance of ER-RDO has been justified by previous works. Appendix B suggests that the problem of optimal intra-refreshing can be formulated as a piecewise convex programming problem which is more difficult than the case of RDO.

3.5 Details of the H.264 Error Resilient Rate-Distortion Optimization

The process of H.264 ER-RDO is basically the same as that in Sec. 3.3 and Sec. 3.4. However, some details relevant to the investigations in this dissertation are explained below. A complete coverage of the encoding process can be found in [32].

3.5.1 Calculation of the Lagrange Multiplier

The Lagrange multiplier is a function of the QS of a frame, which is given by Eq. (3.13). In H.264, the QS is specified by an index named the quantization parameter (QP). There exists a one-to-one relation between the QS and the QP, as given in Tab. 3.1. The relation is also depicted in Fig. 3.8. The exponential relation between the QP and the QS is designed to reduce the memory requirement for the quantizing and inverse quantizing operations [33]. Since the QP is more conveniently used in the encoding process than the QS, it is better to convert Eq. (3.13) to a relation between the Lagrange multiplier and the QP. Experiments yield

$$\lambda = 0.85 \cdot 2^{(q'-12)/3}, \tag{3.25}$$

where λ denotes the Lagrange multiplier and q' denotes the QP. The Lagrange multiplier for QP selection and the one for mode selection are not differently denoted since they are equal.

Table 3.1: The QP to QS mapping.

QP	0	1	2	3	4	5	...	51
QS	0.625	0.6875	0.8125	0.875	1	1.125	...	224

For ER-RDO, the Lagrange multiplier λ, instead of λ_{LOSS} is used in the JM since the difference between the resulted performances is nearly ignorable.

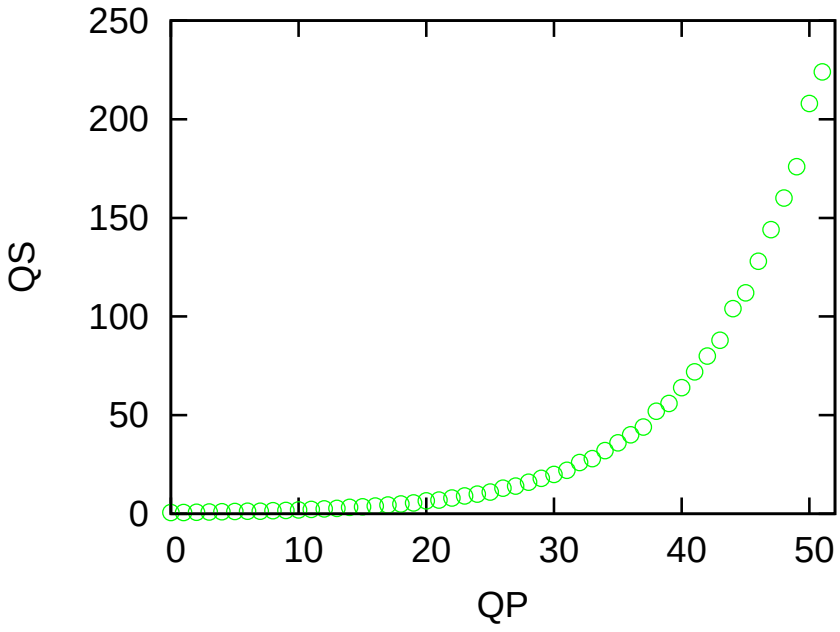


Figure 3.8: The QP to QS mapping.

3.5.2 The Available Macroblock Modes

There are seven available modes for a MB in H.264. The standard specifies only the sizes of the luma prediction blocks of a mode. The sizes of the corresponding chroma prediction blocks can be derived from the chroma subsampling rate. The modes are described below where the length unit is a luma pixel.

SKIP This is a special inter mode in which no bits are encoded. When the decoder finds that a MB has been skipped, it infers the motion vector of the whole MB from some neighboring MBs.

P16×16 In this inter mode, a MB itself is a prediction block.

P8×16 In this inter mode, a MB is divided into two 8×16 prediction blocks.

P16×8 In this inter mode, a MB is divided into two 16×8 prediction blocks.

P8×8 In this inter mode, a MB is divided into four 8×8 sub-blocks. Each sub-block can be treated as a prediction block or further divided into four 4×4 prediction blocks.

I4×4 In this intra mode, a MB is divided into 16 4×4 prediction blocks.

I16×16 In this intra mode, a MB itself is a prediction block.

The MB division schemes of the inter modes are illustrated in Fig. 3.9. Those of the intra modes are not depicted here for conciseness.

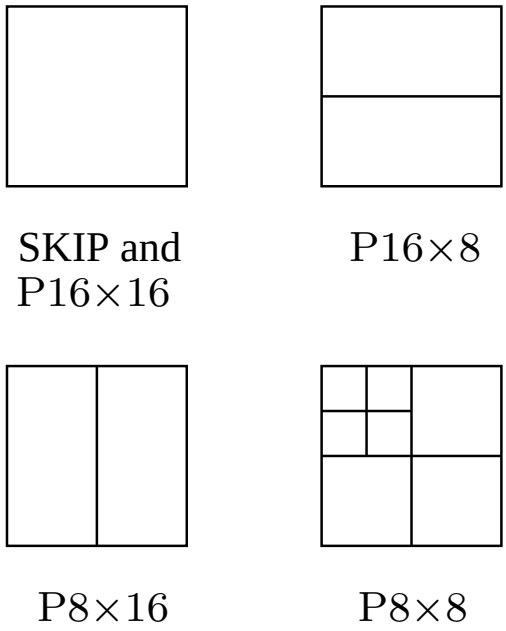


Figure 3.9: The MB division of the inter modes.

Chapter 4

Modifications of the H.264 Error Resilient Rate-Distortion Optimization Framework

As stated in Sec. 2.2.1, the computational complexity of H.264 codecs is very high. When ER-RDO is used for error resilience, the complexity is even higher due to the calculation of the expected distortions of different MB modes. This chapter investigates two opposite modification schemes of the existing ER-RDO framework in H.264. The purpose of the schemes is to improve the simplicity-performance compromise.

The principles and the simulation results of the schemes are presented below.

4.1 Multiple Intra Quantization Parameter Value Testing

In the H.264 ER-RDO, after the QP value is designated, all the intra and inter modes of a MB are tested and a mode minimizing a cost function

$$J_{k,\text{LOSS, JM}} = E(d_k(m_k, q)) + \lambda\, r_k(m_k, q), \tag{4.1}$$

is selected. It is emphasized here that

1. the same λ value should be applied to all the MBs in a frame to satisfy Eq. (3.11),

2. the relation in Eq. (3.25) is only statistical and does not provide the optimal QP values for the given λ value for all the MBs,

3. the computation complexity for inter modes (except the SKIP mode) testing is much higher than intra modes testing.

It is then suggested here that in addition to the designated QP value, some extra QP values combined with only the intra modes be tested to improve the performance. Note that although the testing of the SKIP mode is very simple, its rate and distortion is independent of the QP. Therefore it needs only to be tested once.

The above modification is called multiple intra QP value testing and referred to as Scheme 1 hereafter. Scheme 1 actually seeks to improve the performance of the ER-RDO by broadening the search range for the optimal encoding QP and mode for a MB.

There are two side effects of Scheme 1. First, the complexity is certainly increased. Second, more bits have to be used to represent the difference between the QPs of two adjacent MBs. Because the performance gain and the side effects can not be definitely predicted, the viability of this scheme has to be checked by simulations.

4.2 Fast Macroblock Mode Selection

Contrary to Scheme 1, the modification introduced in this section tries to reduce the complexity while keeping the same performance. The modification is a fast MB mode selection scheme and called Scheme 2 hereafter. It is described below.

The P8×8, P8×16, P16×8 and P16×16 modes have the highest computational complexity among all the modes. Experiments show that for a given QP value, the operating DR points of the four modes are close to one another in the DR plane, as assumed in Sec. 3.3. It is then reasonable to choose one of them as the representative in mode selection. Without a priori knowledge

about the motion features of a MB, a reasonable choice is the P8×8 mode with 8×8 prediction blocks. The first step of Scheme 2 is then selecting from the SKIP, I4×4, I16×16 and P8×8 modes, one minimizing a cost function

$$J_{k,\text{LOSS, FAST}} = \begin{cases} J_{k,\text{LOSS, JM}}, \text{ if } m_k = \text{SKIP, I4×4 or I16×16,} \\ \alpha J_{k,\text{LOSS, JM}}, \text{ if } m_k = \text{P8×8,} \end{cases} \tag{4.2}$$

where m_k is the mode of the MB; q is the global QS; $r_k(m_k, q)$ is the achievable rate; and $d_k(m_k, q)$ is the achievable distortion. The parameter α is real valued and lies between 0 and 1.

If P8×8 is not selected in the first step, the mode selection is early terminated and the encoding process simply continues with the selected mode. Otherwise, a second step is required.

In the second step, the untested modes, i.e., the P8×16, P16×8 and P16×16 modes are eventually tested. The cost $E(d_k(m_k, q)) + \lambda r_k(m_k, q)$ of all the inter and intra modes are compared and the optimal mode is selected in the same way as the JM.

It can be seen that α controls the complexity and the performance. When α is smaller, the P8×16, P16×8 and P16×16 modes are more likely to be tested; the complexity tends to be higher; a better mode might be found; and vice versa. Therefore it is called the complexity tuning parameter.

Note that an early termination of the first step is also possible. The principle is explained as follows. The cost functions of the SKIP, I4×4 and I16×16 modes are calculated firstly. Then the four 8×8 sub-blocks of the P8×8 mode are tested one by one. After testing a sub-block, the rate and the distortion of the already tested sub-blocks are used to calculate a partial cost according to Eq. (4.2). Once the partial cost is greater than the costs of the SKIP, I4×4 and I16×16 modes, the testing of the P8×8 mode can be terminated. Of course, some information such as the MB header information of the P8×8 mode can not be obtained before the whole MB is tested in the mode. However, the information occupies only a small portion of the total rate, thus should have little influence on the effectiveness of the early-termination method.

Scheme 2 is also checked by simulation in this chapter.

Table 4.1: Test video sequences.

Low activity	Middle activity	High activity
Akiyo, Mother and Daughter	Carphone, Foreman	Mobile, Coastguard

4.3 Simulation Settings

In this section, the key settings of the simulations are explained.

4.3.1 Test Video Sequences

In the simulations to verify the two modification schemes, six of the most popular test video sequences are used. The used video sequences are grouped by motion activity as in Tab. 4.1. A sample frame of each sequence is provided in Appendix A.

The test sequences are of the standard quarter common intermediate format (QCIF) resolution of 176×144. Their frame rate is 10Hz. The length of each sequence is 10sec or equivalently 100 frames. The sequences use the YCbCr color space and their chroma subsampling scheme is 4:2:0 which means the subsampling ratio is 1/2 along both the vertical and the horizontal axes. The bit depth of the sequences is eight.

4.3.2 Video Quality Evaluation

The most popular video quality criterion peak signal-to-noise ratio (PSNR) is used in this dissertation for video quality evaluation. Consider a raw video sequence X. After video encoding, transporting and decoding (possibly also error concealment), the video is reconstructed as $\breve{X}$. The PSNR of $\breve{X}$ is then defined by

$$\mathrm{PSNR} = 10 \log_{10} \frac{\psi_{\max}^2}{\sigma_e^2}, \tag{4.3}$$

where $\psi_{\max}$ is the peak value for a given bit depth and σ_e^2 is the MSE between X and $\breve{X}$. Since the bit depth of the test sequences is eight, $\psi_{\max}$ equals 255

in this dissertation.

Although PSNR does not fully reflect subjective video quality, the evaluation results based on the two measures are rather consistent for the simulations in this dissertation. Therefore PSNR is adopted as the video quality measure.

4.3.3 Other Settings

Scheme 1 and 2 are compared with the scheme of the JM version 10.2 in the simulations. In all the simulations, a frame is divided into nine equally sized slices for error resilience. A slice is encapsulated into a NALU. The tested slice loss ratios are 5%, 10%, 15% and 20%. The sequence parameter set and the first frame in a simulation are always correctly transported. A single QP value is applied to all the frames in a simulation. The tested QP values are selected so that the average bit rates are in the range of 32 to 256kbps. For each pair of encoder configuration and slice loss ratio, the random channel is simulated for 50 times and the average PSNR of the 50 reconstructed sequences is used to evaluate the encoder configuration.

In the Scheme 1 simulations, in addition to the original QP value q', $q'-3$ and $q'+3$ are also tested for the intra modes. The value of α is set to 0.9 for Scheme 2 simulation.

4.4 Simulation Results

Because the human visual system (HVS) is more sensitive to the luma component of images and videos, only the PSNR of the luma or Y component is used in video quality evaluation. The performance of an encoder is demonstrated by the corresponding PSNR-rate curves.

The complexity of the schemes is measured by the number of CPU clock ticks consumed on the mode testing operations in a common simulation environment. Although the measured values are dependent on the hardware and software infrastructure, they can somehow reveal the relation between the complexity of the schemes.

The results for the slice loss ratios $p = 5\%$ and 20% are illustrated in Fig. 3.2 to Fig. 3.2. The figures for the other loss ratios have similar characteristics and are omitted here. The results are analyzed below.

4.4.1 Performance Comparison

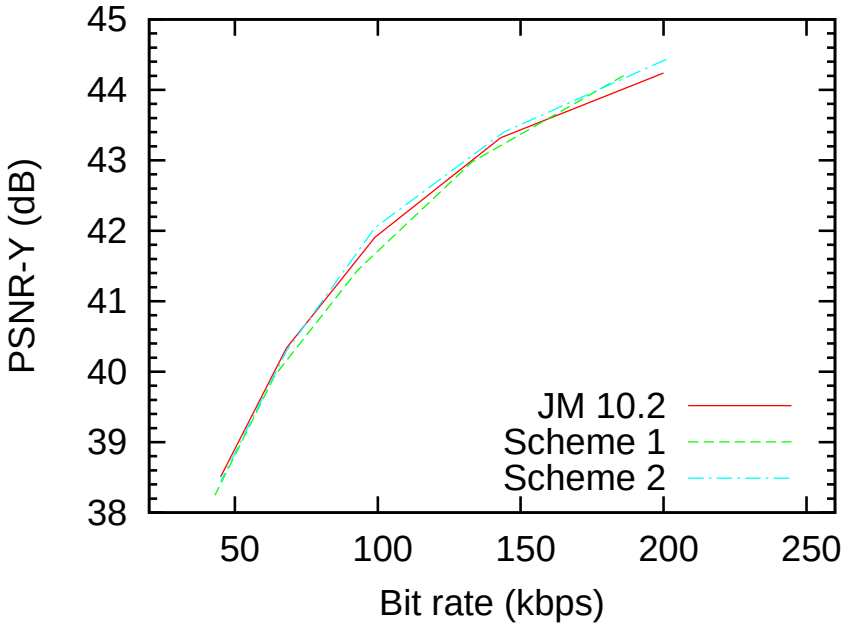
It can be seen that Schemes 1 and 2 have only marginal and indecisive influence on the performance. Theoretically, broadening the search range would improve the performance and vice versa. However, the PSNR-rate curves of the schemes in the diagrams are generally close to and sometimes entangled with one another. This phenomenon suggests that the search range of the JM is large enough to cover some nearly-optimal solutions while the efficiency of the search algorithm can still be improved.

4.4.2 Complexity Comparison

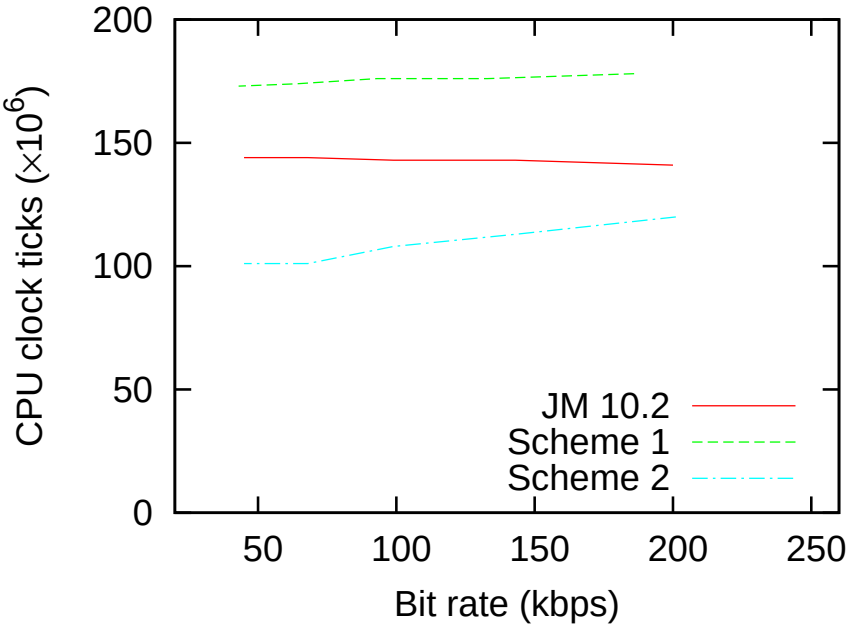
The influence of Schemes 1 and 2 on the complexity is obvious. Since Scheme 1 proves to be ineffective in performance improvement, only the complexity of Scheme 2 is compared to that of the JM. In all the complexity-rate diagrams, the complexities of Scheme 2 and the JM change slowly with the rate. Therefore, the relative complexity reduction calculated from the average complexities over all the tested rates is a good measure for the effectiveness of Scheme 2. The reduction values are given in Tab. 4.1.

It can be found in Tab. 4.1 that when the slice loss ratio is high, Scheme 2 is more effective. The reason is that a higher loss ratio generally causes greater distortion for the inter modes. In this case, the probability of early terminations in mode selection is higher.

The relative complexity reduction for the video sequence Mobile is remarkably lower than for the other video sequences. The reason is that Mobile has a very complex spatial texture which results in great rates for the intra modes. Therefore early terminations in mode selection are less probable than for the other video sequences.

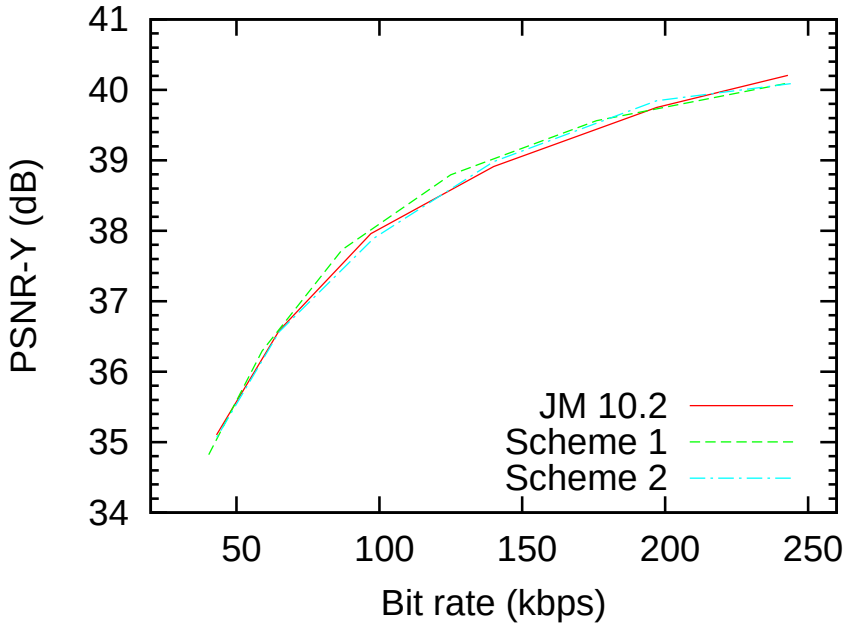


(a)

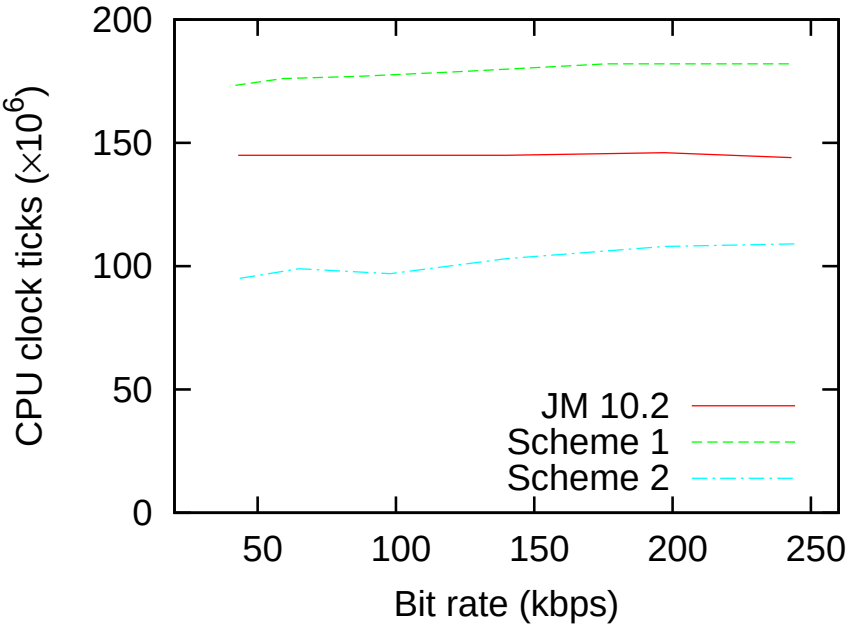


(b)

Figure 4.1: The (a) performance and (b) complexity of the schemes for Akiyo and slice loss ratio $p = 5\%$.

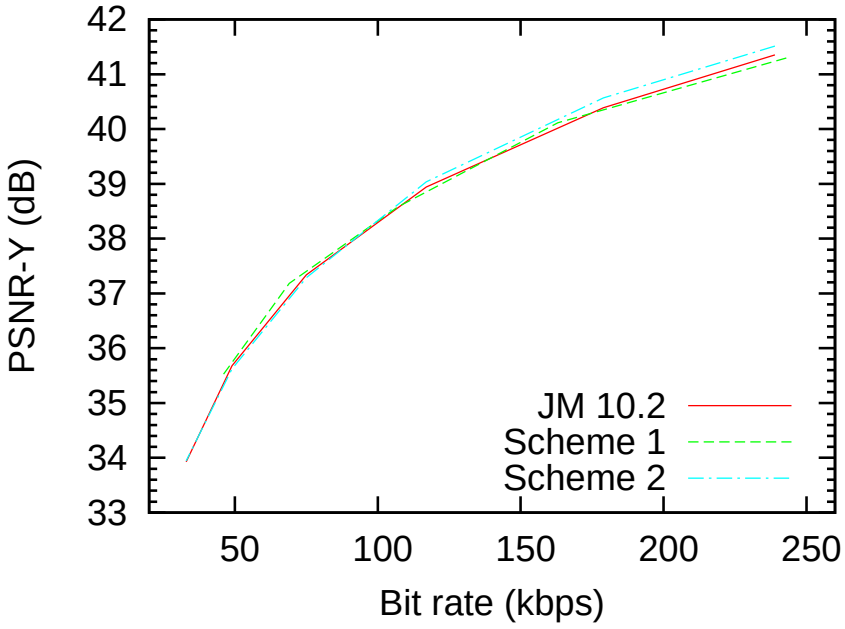


(a)

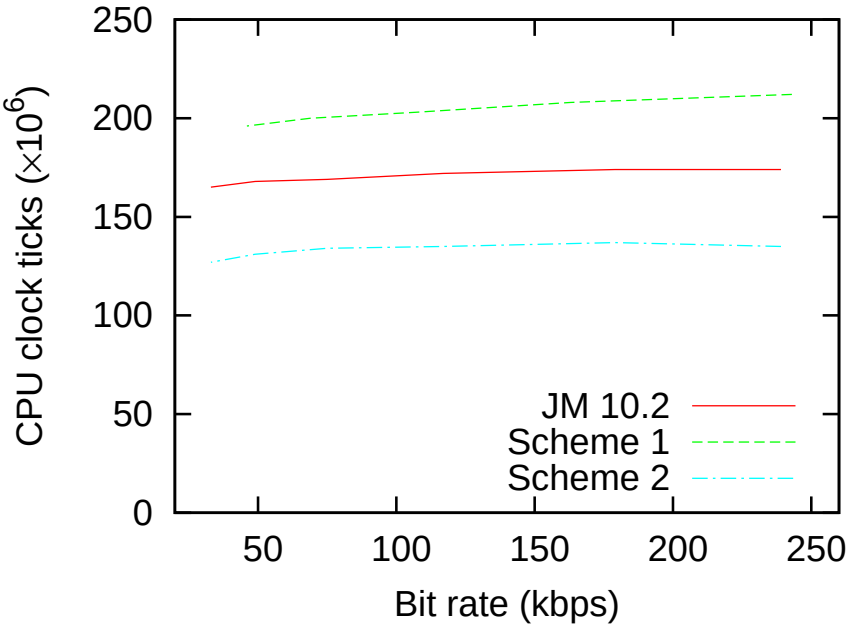


(b)

Figure 4.2: The (a) performance and (b) complexity of the schemes for Akiyo and slice loss ratio $p = 20\%$.

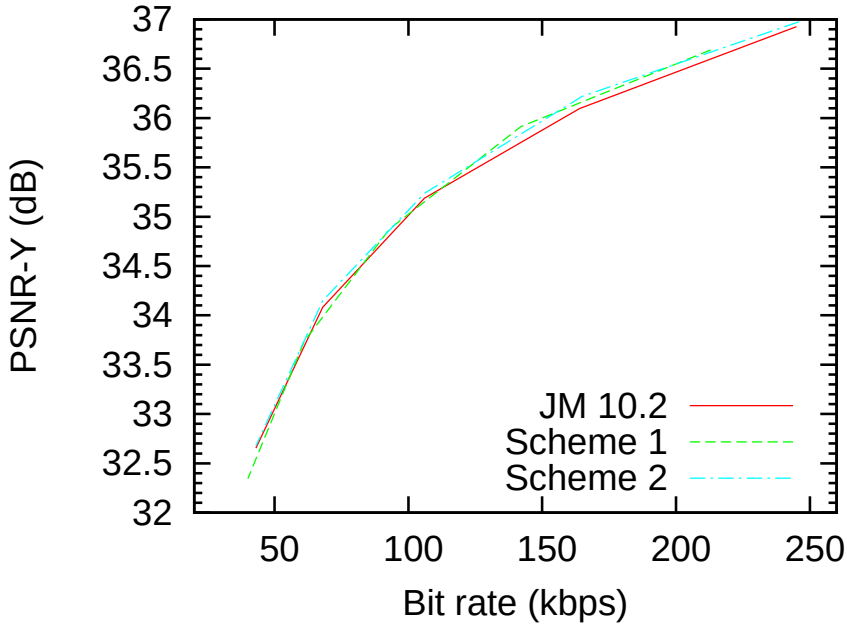


(a)

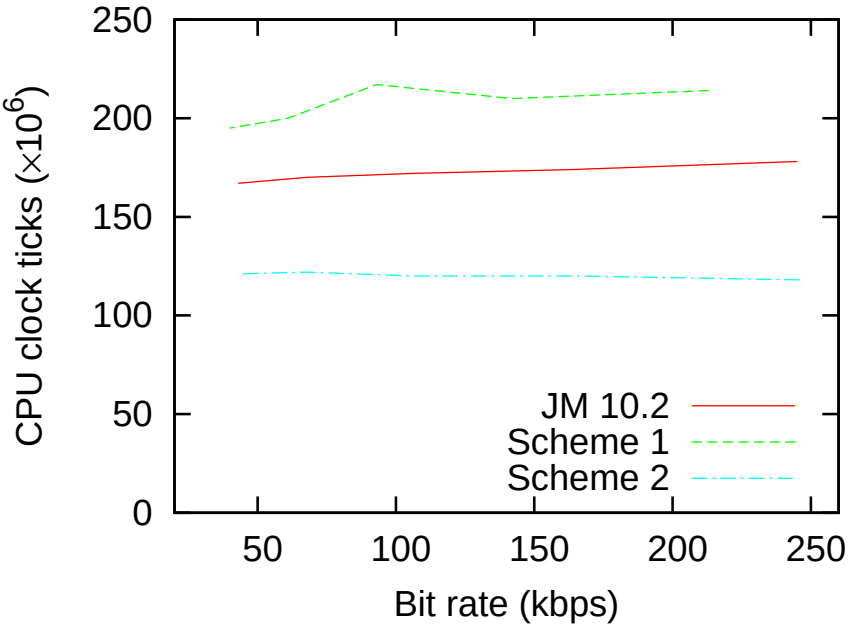


(b)

Figure 4.3: The (a) performance and (b) complexity of the schemes for Mother and Daughter, and slice loss ratio $p = 5\%$.

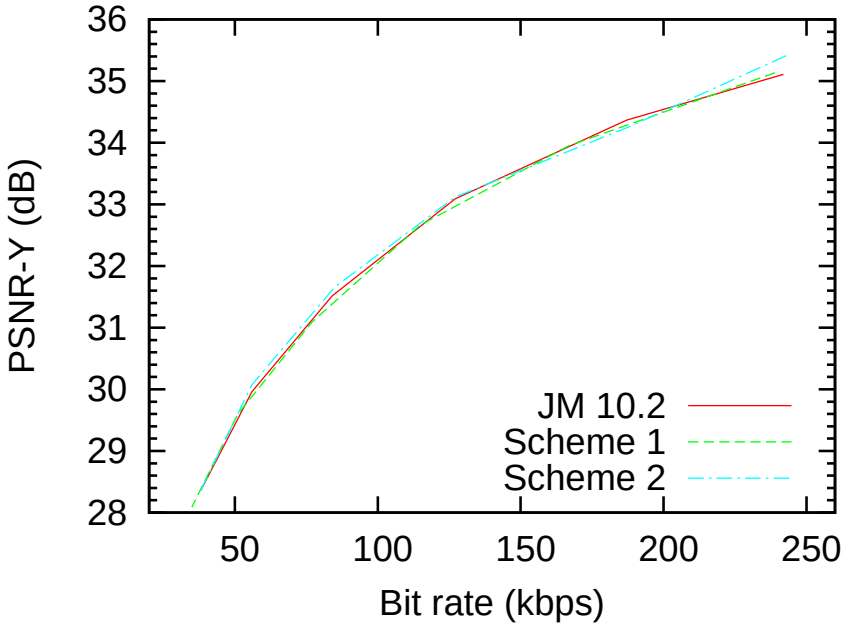


(a)

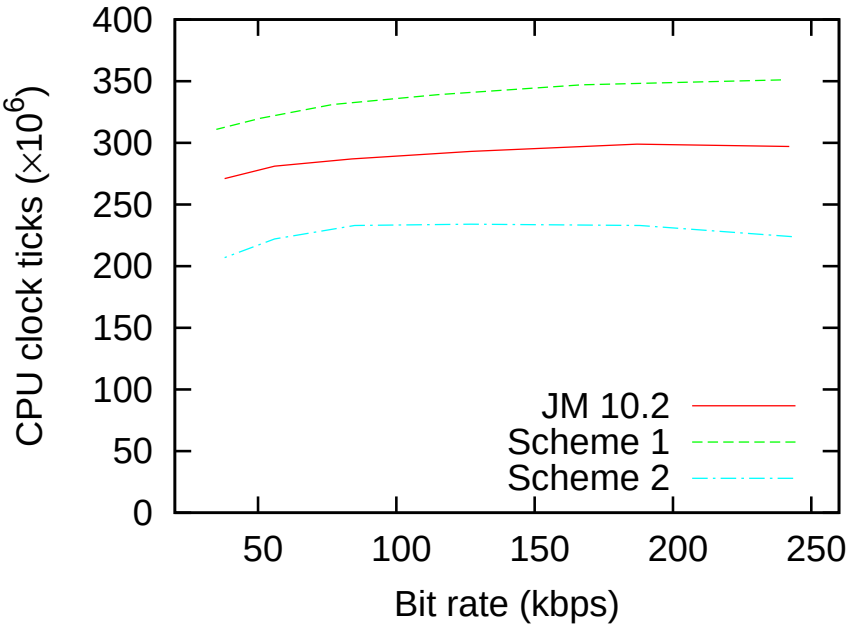


(b)

Figure 4.4: The (a) performance and (b) complexity of the schemes for Mother and Daughter, and slice loss ratio $p = 20\%$.

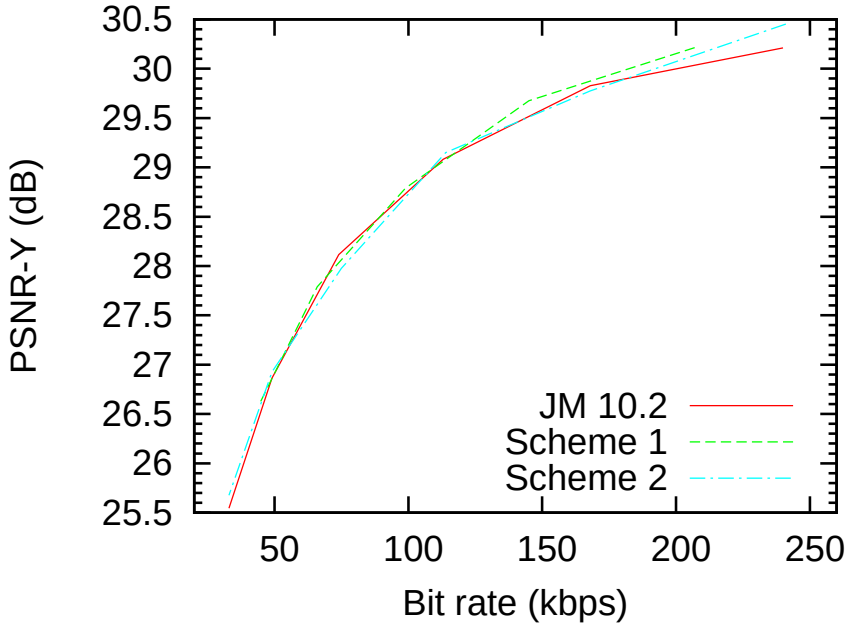


(a)

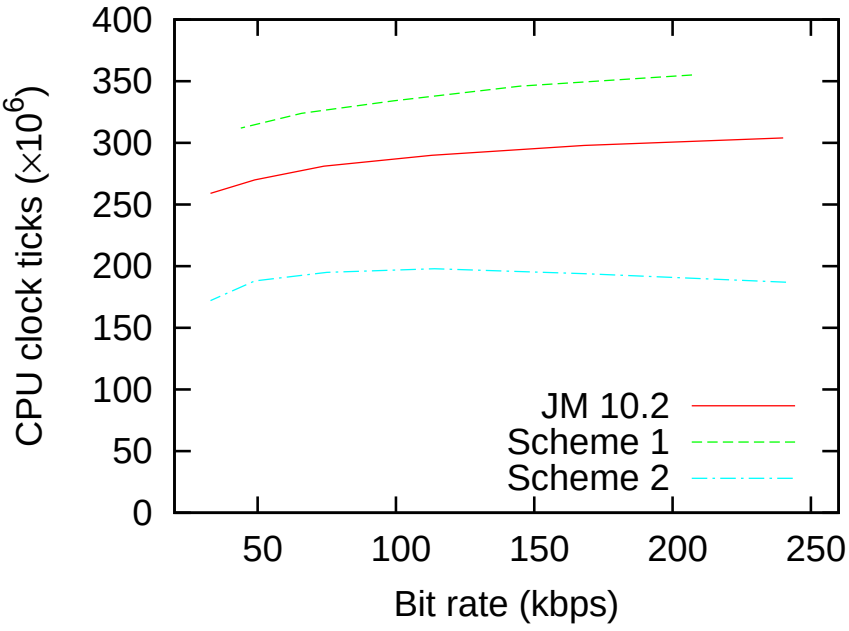


(b)

Figure 4.5: The (a) performance and (b) complexity of the schemes for Carphone and slice loss ratio $p = 5\%$.

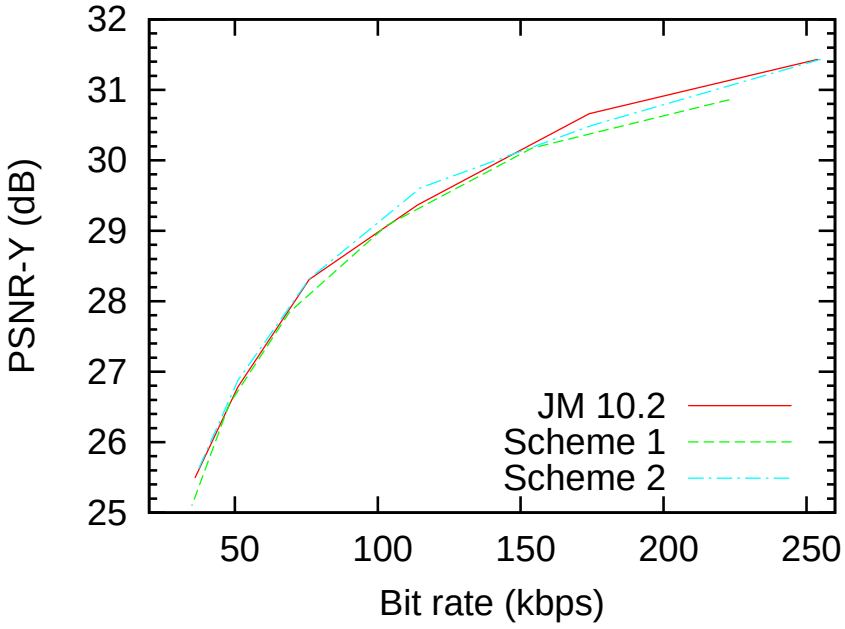


(a)

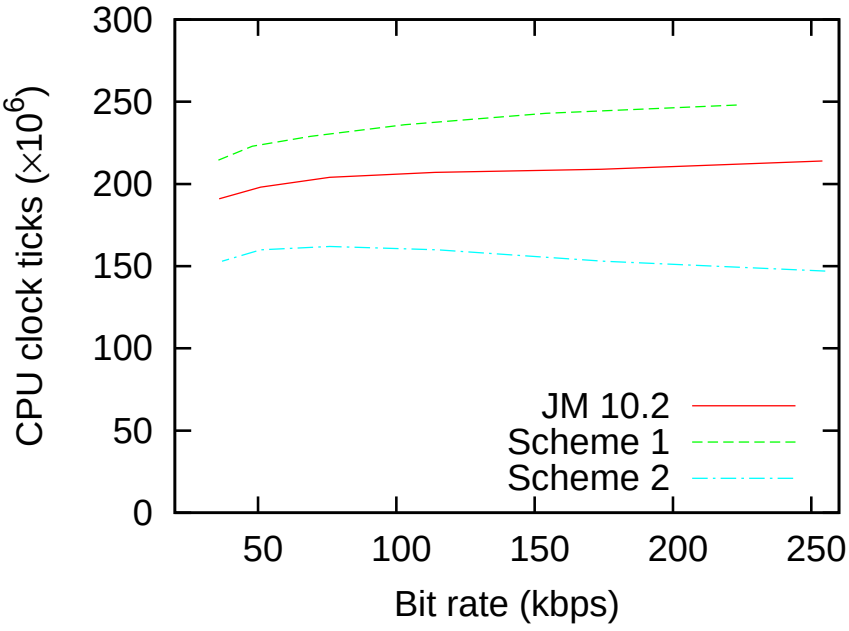


(b)

Figure 4.6: The (a) performance and (b) complexity of the schemes for Carphone and slice loss ratio $p = 20\%$.

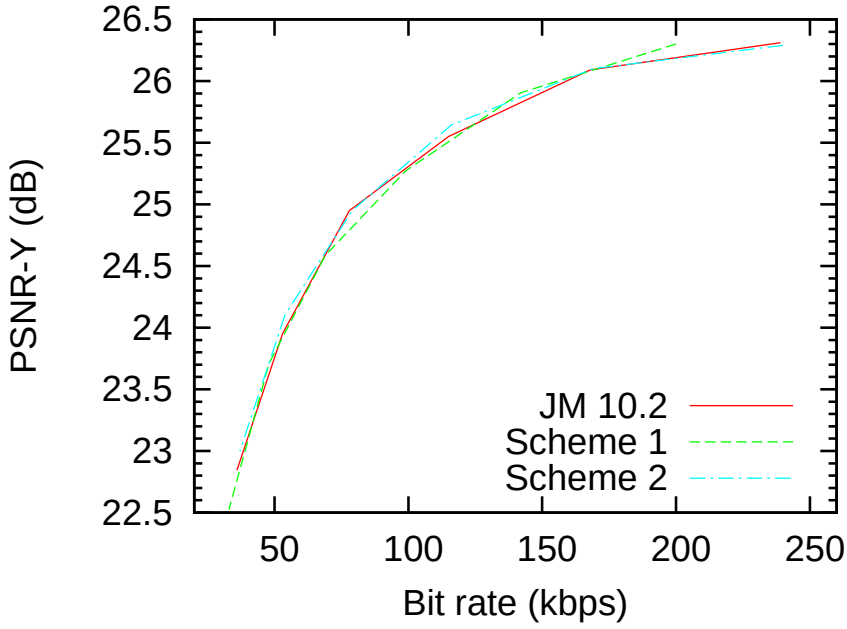


(a)

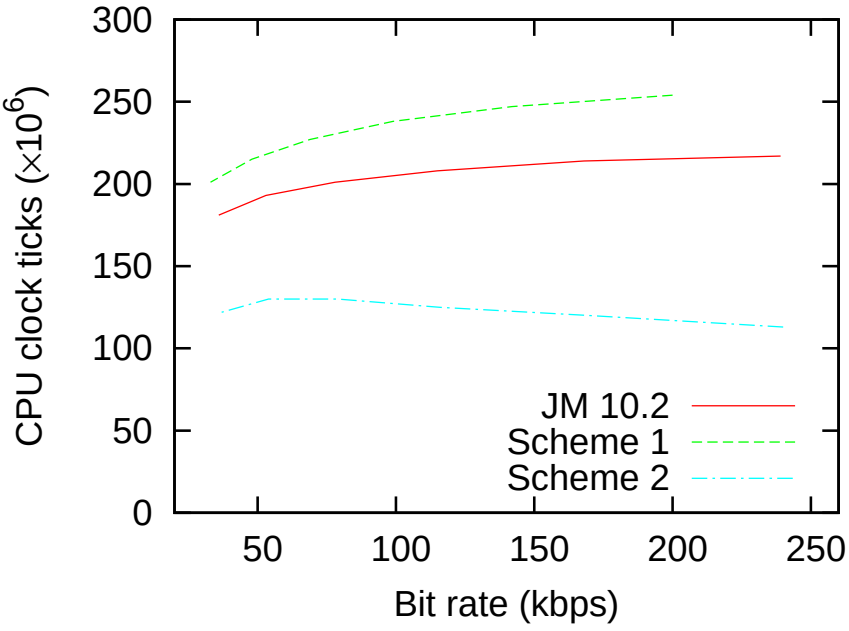


(b)

Figure 4.7: The (a) performance and (b) complexity of the schemes for Foreman and slice loss ratio $p = 5\%$.

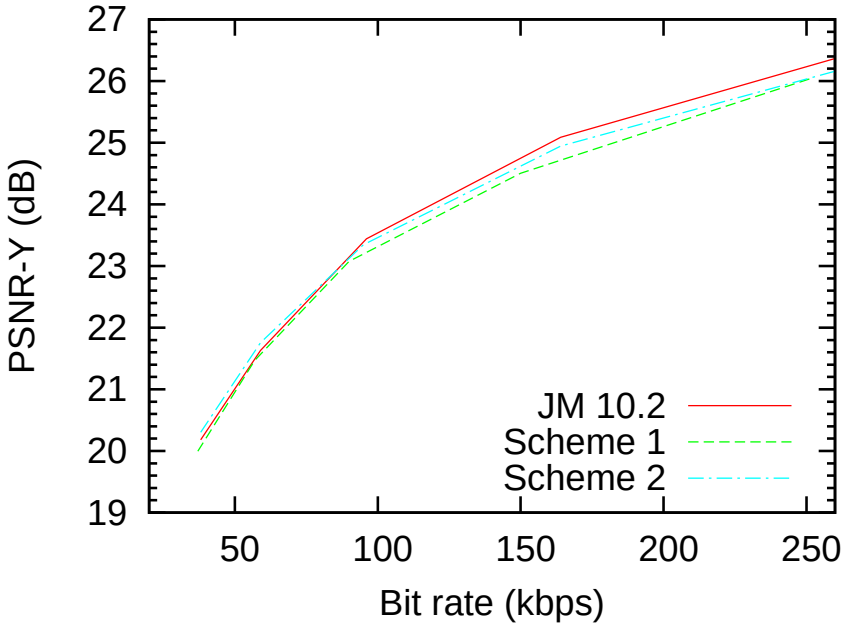


(a)

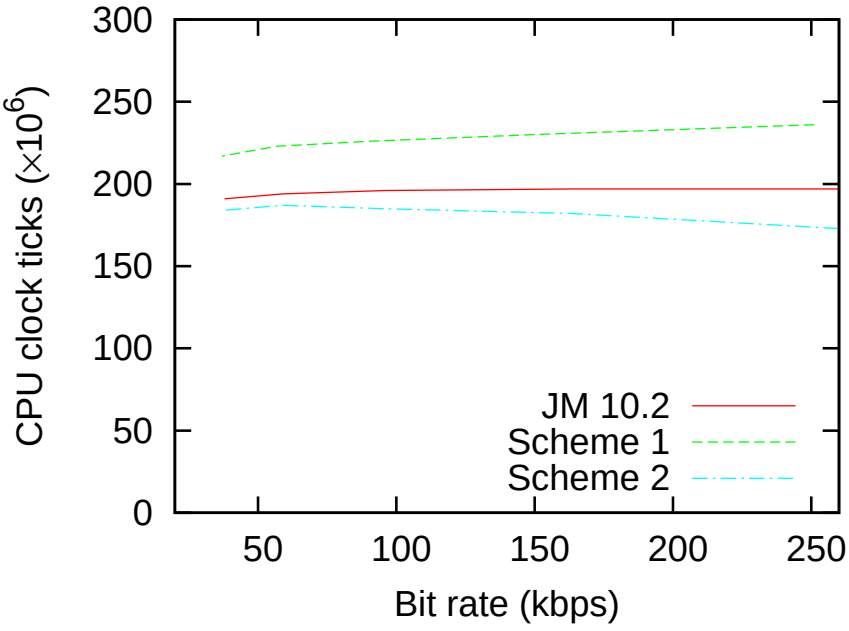


(b)

Figure 4.8: The (a) performance and (b) complexity of the schemes for Foreman and slice loss ratio $p = 20\%$.

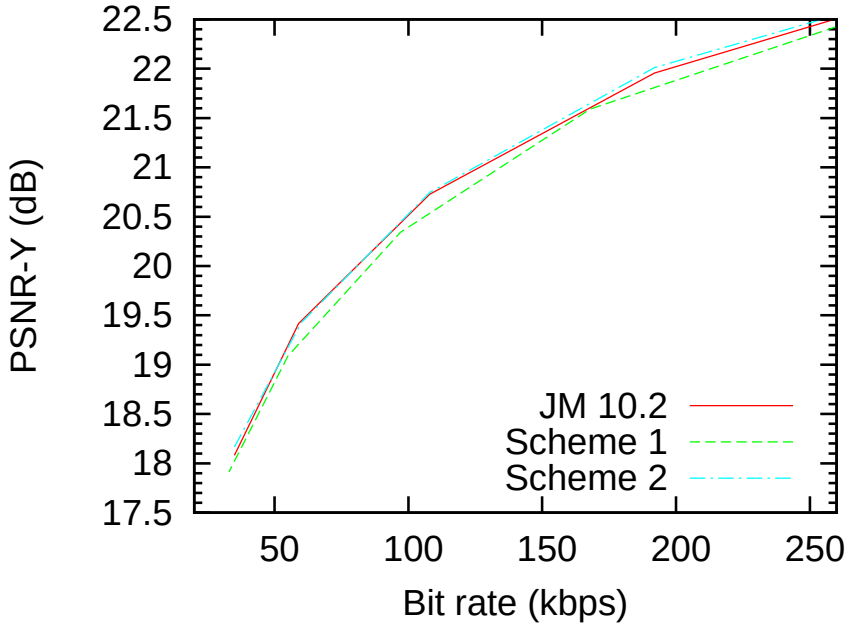


(a)

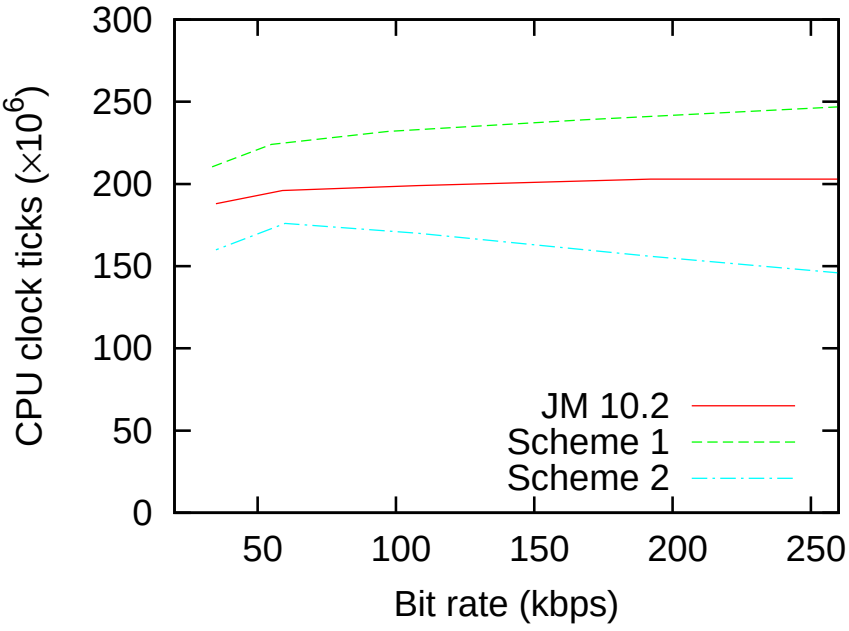


(b)

Figure 4.9: The (a) performance and (b) complexity of the schemes for Mobile and slice loss ratio $p = 5\%$.

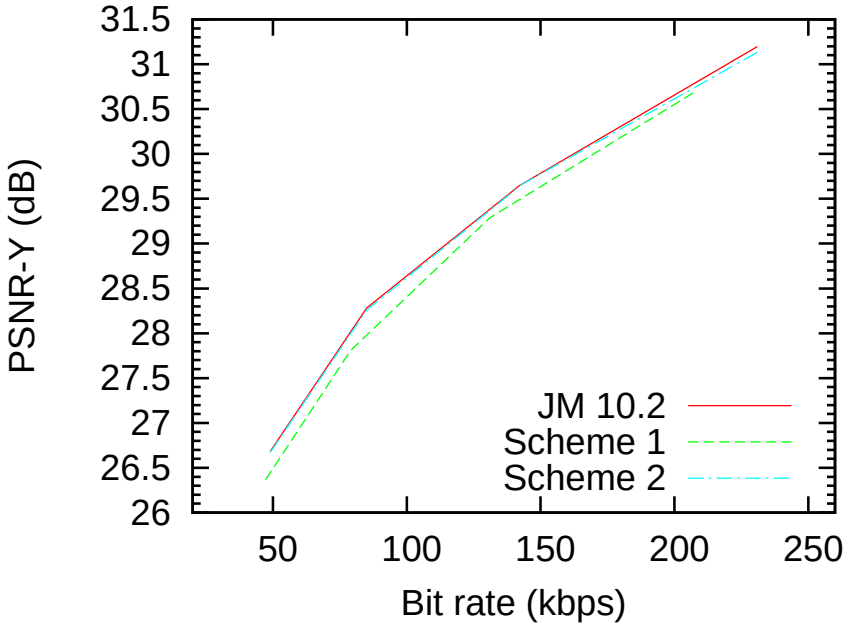


(a)

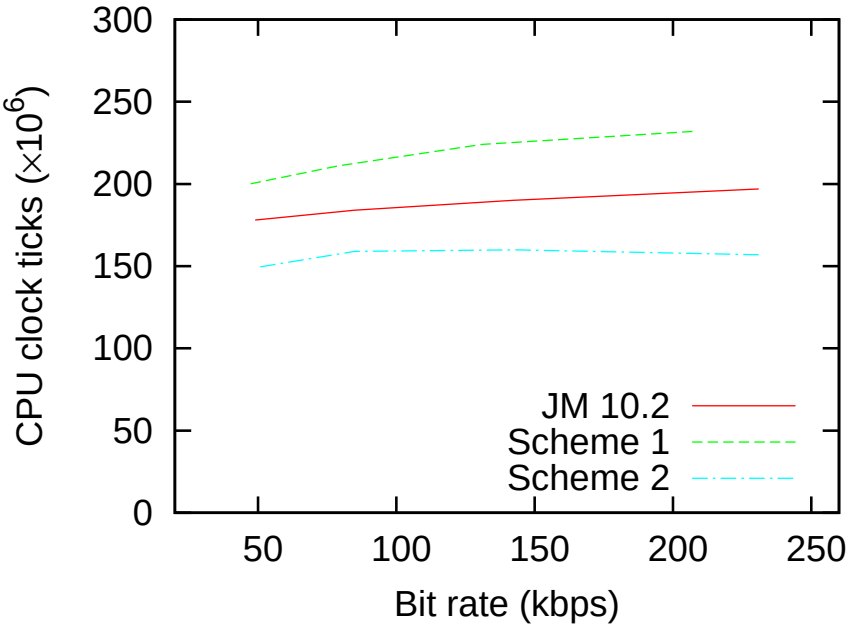


(b)

Figure 4.10: The (a) performance and (b) complexity of the schemes for Mobile and slice loss ratio $p = 20\%$.

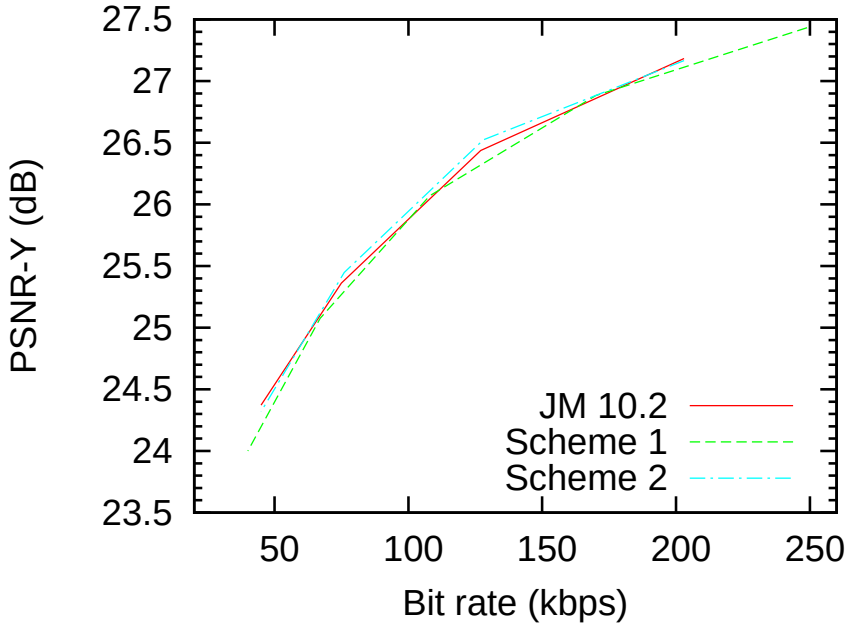


(a)

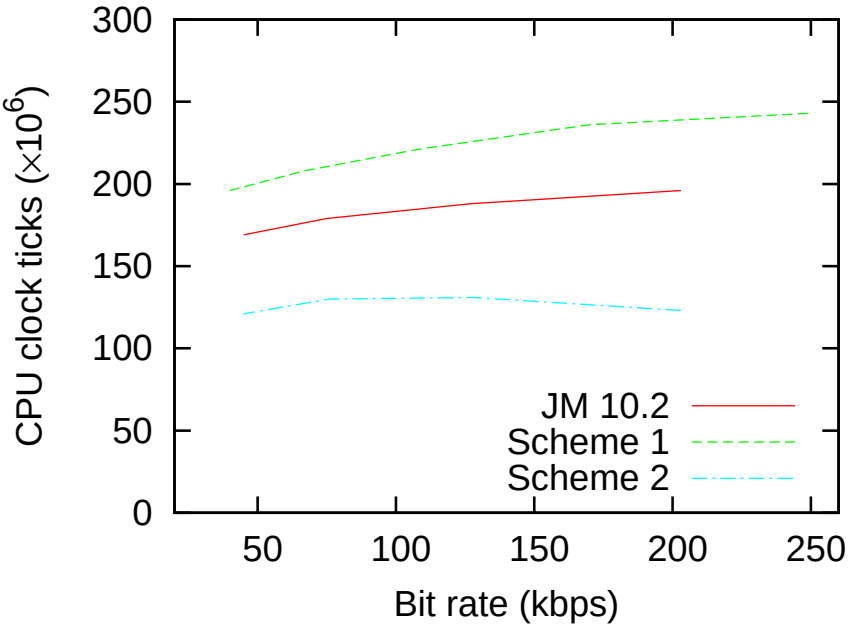


(b)

Figure 4.11: The (a) performance and (b) complexity of the schemes for Coastguard and slice loss ratio $p = 5\%$.



(a)



(b)

Figure 4.12: The (a) performance and (b) complexity of the schemes for Coastguard and slice loss ratio $p = 20\%$.

Table 4.2: Relative complexity reduction through Scheme 2 compared to JM.

Video Sequence	$p = 5\%$	$p = 10\%$	$p = 15\%$	$p = 20\%$
Akiyo	24%	27%	28%	29%
Mother and Daughter	21%	24%	27%	30%
Carphone	21%	25%	30%	33%
Foreman	23%	30%	34%	39%
Mobile	6%	9%	14%	18%
Coastguard	16%	23%	28%	31%

Chapter 5

Conclusion

In this dissertation, the existing H.264 rate-distortion optimization (RDO) and error resilient rate-distortion optimization (ER-RDO) frameworks are investigated and two modified schemes to the ER-RDO framework are proposed and tested.

The RDO framework has proven to be efficient in encoder parameter selection in hybrid video coding. Generally it achieves a satisfying balance between the computational simplicity and the compression ratio. The analysis in this dissertation helps to better understand the source of the efficiency of RDO.

ER-RDO is an extension of RDO which considers error propagation in packet lossy environments, such as communication networks or storage devices. It is shown that ER-RDO can be viewed as a special intra refreshing method. Like RDO, it has been successful in macroblock (MB) mode selection in error resilient encoding. However, the extension from RDO to ER-RDO is not straightforward. In RDO, the convexity of the operating distortion-rate functions makes the problem suitable to be solved with the Lagrange multipliers method. In packet lossy environments, the operating distortion-functions are no longer convex due to the different rate and distortion characteristics of intra and inter modes. For the purpose of simplicity, ER-RDO ignores the non-convex condition and adopts the same optimization procedures as RDO. This technical handling is a practical solution but the performance may be farther from the optimum.

The first proposed modification of the ER-RDO framework called Scheme 1 is testing two more QPs with only intra modes. Ideally, this modification would not increase the complexity too much. At the same time, the search range is broadened and an improvement in performance is expected. However, the simulation results show that the video quality can not be definitely improved by the modification. The reason might be that the existing QP and mode search range is large enough to cover some nearly-optimal solutions.

The second modification referred to as Scheme 2 is a fast mode selection scheme. Since the clusters of the intra operating points and the inter points are generally far from each other, using an inter mode to represent all the inter modes would not significantly affect intra refreshing decision. This modification uses the P8×8 mode as the representative inter mode in the decision. The simulation results show that the video quality is almost unchanged while the computational burden is substantially lowered.

As a conclusion, the performance of the existing H.264 ER-RDO framework can hardly be improved by moderately increasing the complexity, but it can be maintained even with substantially lowered complexity.

Appendix A

Test Video Sequences

Akiyo

Mother and Daughter

Carphone

Foreman

Mobile

Coastguard

Appendix B

A New Formulation of the Intra-refreshing Problem

In Sec. 3.4, it is pointed out that the existing ER-RDO framework is efficient in locating forced intra MBs although its theoretical basis is incomplete. Here a new formulation of the intra-refreshing problem is developed for possible improvement or alternatives to ER-RDO.

B.1 The Original Problem

In a packet lossy environment, the optimal QS and mode Selection including intra-refreshing problem is given by Eq. (3.15), i.e.,

$$\arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} E(d_k(m_k, q_k)), \text{ subject to} \sum_{k=1}^{K} r_k(m_k, q_k) \leq R_0.$$

B.2 Expected Operating DR Functions

Since for a give q_k value, the expected DR points of the inter modes and of the intra modes form two distinctive sets in the DR plane, it is feasible

to study the expected DR bounds of the inter modes and the intra modes separately.

For the inter modes,

$$E(d_k(m_k, q_k)) = (1-p)d_k(m_k, q_k) + (1-p)D_{k,\text{PROP}} + pD_{k,\text{CON}}.$$

Therefore the lower bound of the expected inter DR points can be fitted by

$$\bar{D}_{k,\text{INTER}}(R_k) = (1-p)\tilde{D}_k(R_k) + (1-p)D_{k,\text{PROP}} + pD_{k,\text{CON}}, \tag{B.1}$$

where $\bar{D}_{k,\text{INTER}}(R_k)$ is called the expected inter-only operating DR function of MB_k. It is also a decreasing and convex function like $\tilde{D}_k(R_k)$.

Similarly the lower bound of the expected intra DR points of MB_k can be fitted by a creasing and convex function $\bar{D}_{k,\text{INTRA}}(R_k)$ which is called the expected intra-only operating DR function.

When both the inter modes and the intra modes are allowed, the expected lower performance bound is given by

$$\bar{D}_k(R_k) = \min\{\bar{D}_{k,\text{INTER}}(R_k), \bar{D}_{k,\text{INTRA}}(R_k)\}, \tag{B.2}$$

where $\bar{D}_k(R_k)$ is called the expected operating DR function.

Because the intra modes normally generate more bits than the inter modes and the inter modes are subject to transport-introduced errors, $\bar{D}_k(R_k)$ is typically a piecewise convex function, as shown in Fig. B.1. In the figure, $\bar{D}_k(R_k)$ contains two pieces of curves: one from $\bar{D}_{k,\text{INTER}}(R_k)$, the other from $\bar{D}_{k,\text{INTRA}}(R_k)$.

B.3 A New Formulation

With the defined expected DR functions, the problem in Eq. (3.15) can be approximated by

$$\arg\min_{\boldsymbol{u}} \sum_{k=1}^{K} \bar{D}_k(R_k), \text{ subject to} \sum_{k=1}^{K} R_k = R_0. \tag{B.3}$$

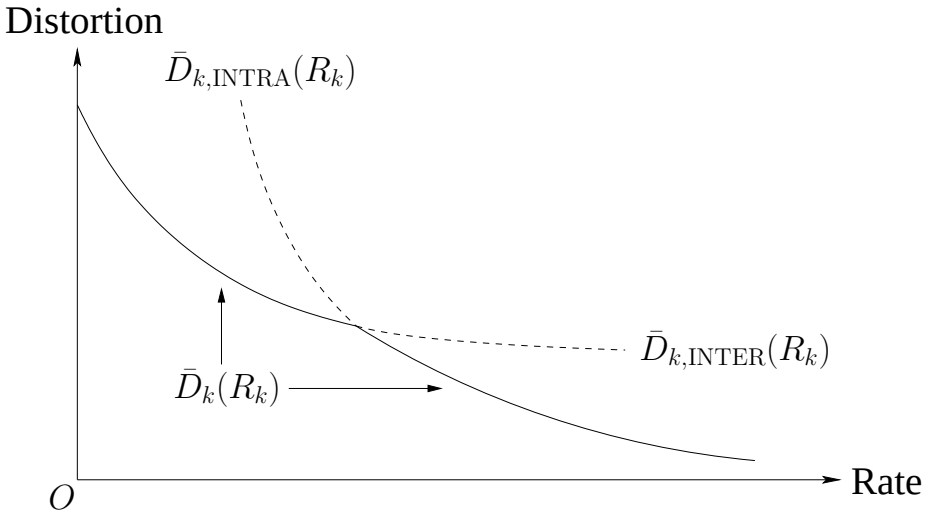


Figure B.1: The expected operating DR functions.

It can be seen that Eq. (B.3) is an optimization problem which is only piecewise convex. The problem is intrinsically much more difficult than the convex optimization problem in Eq. (3.3), thus unsuitable for complexity constrained applications. If computational complexity is not constrained, solving Eq. (B.3) should result in better rate-PSNR performance than the existing ER-RDO method. However, application of the new formulation is out of the scope of this dissertation and can be a topic of future work.

Bibliography

[1] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications*. Prentice Hall, 2002.

[2] *Video Coding for Low Bit Rate Communication*, ITU-T Std. H.263, 2005.

[3] *Generic Coding of Moving Pictures and Associated Audio Information*, ISO/IEC Std. ISO/IEC 13 818, 2007.

[4] *Coding of Audio-Visual Objects Part 2: Visual*, ISO/IEC Std. ISO/IEC 14 496, 2004.

[5] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688–703, Jul. 2003.

[6] R. Schäfer, T. Wiegand, and H. Schwarz, "The emerging H.264/AVC standard," *EBU Technical Review*, vol. 293, Jan. 2003.

[7] T. Stockhammer, M. Hannuksela, and T. Wiegand, "H.264 in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 657–673, Jul. 2003.

[8] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Commun. Mag.*, vol. 12, pp. 59–65, Aug. 2005.

[9] Y. Takishima, M. Wada, and H. Murakami, "Reversible variable length codes," *IEEE Trans. Commun.*, vol. 43, pp. 158–162, Feb. 1995.

[10] "DVB project," `http://www.dvb.org/index.xml`.

[11] S. Wenger, "Test Model 12 / Appendix II of H.263 Version 3," VCEG, Tech. Rep., 2000.

[12] A. M. Tourapis, A. Leontaris, K. Sühring, and G. Sullivan, "H.264/MPEG-4 AVC reference software manual," JVT, Tech. Rep., 2009.

[13] Q. Zhu and L. Kerofsky, "Joint source coding, transport processing, and error concealment for H.323-based packet video," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, San Jose, USA, 1999, pp. 52–62.

[14] P. Haskell and D. Messerschmitt, "Resynchronization of motion-compensated video affected by ATM cell losses," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1992, pp. 545–548.

[15] J. Liao and J. Villasenor, "Adaptive intra update for video coding over noisy channels," in *Proceedings of the IEEE International Conference on Image Processing*, 1996, pp. 763–766.

[16] C. Zhao, K. Ngan, J. Zhang, R. Mathew, and X. Zhang, "Using inter-frame dependence history to select intra-refresh blocks," *Electronics Letters*, vol. 38, pp. 1337–1338, Oct. 2002.

[17] R. Zhang, S.L.Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 966–976, Jun. 2000.

[18] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 952–965, Jun. 2000.

[19] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion-compensated prediction," *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 1050–1062, Jun. 2000.

[20] T. Wiegand, M. Lightstone, D. Mukherjee, T. Campbell, and S. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 182–190, Apr. 1996.

[21] G. Schuster and A. Katsaggelos, *Rate-distortion Based Video Compression*. Boston, USA: Kluwer Academic Publishers, 1997.

[22] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, pp. 23–50, Nov. 1998.

[23] T. Stockhammer, D. Kontopodis, and T. Wiegand, "Rate-distortion optimization for JVT/H.26L video coding in packet loss environment," in *Proceedings of International Packet Video Workshop*, 2002.

[24] "H.264 reference software," http://iphome.hhi.de/suehring/tml/.

[25] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, USA: Prentice-Hall, 1971.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[27] R. McEliece, *The Theory of Information and Coding*. Beijing, China: Publishing House of Electronics Industry, 2003.

[28] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp. 878–894, Sep. 2000.

[29] N. Kamaci, Y. Altunbasak, and R. 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.*, vol. 15, pp. 994–1006, Aug. 2005.

[30] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 446–459, Aug. 1998.

[31] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," in *Proceedings of the IEEE International Conference on Image Processing*, 2001.

[32] K.-P. Lim, G. Sullivan, and T. Wiegand, "Text description of joint model reference encoding methods and decoding concealment methods," JVT, Tech. Rep., 2005.

[33] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 598–602, Jul. 2003.