

國立交通大學

多媒體工程研究所

碩士論文

針對可調視訊編碼粗略可調性之模式相依
的位元與失真解析模型
Analytical Mode-Dependent Rate and Distortion Models
for H.264/SVC Coarse Grain Scalability

研究生：曾于真

指導教授：彭文孝 教授

中華民國 一 百 年 十 月

針對可調視訊編碼粗略可調性之模式相依的位元與失真解析模型
Analytical Mode-Dependent Rate and Distortion Models for H.264/SVC
Coarse Grain Scalability

研究生：曾于真

Student : Yu-Chen Tseng

指導教授：彭文孝

Advisor : Wen-Hsiao Peng

國立交通大學
多媒體工程研究所
碩士論文



Submitted to Institute of Multimedia Engineering
College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

Oct. 2011

Hsinchu, Taiwan, Republic of China

中華民國一十年十月

針對可調視訊編碼粗略可調性之模式相依的位元與失真解析模型

研究生：曾于真

指導教授：彭文孝

國立交通大學多媒體工程研究所 碩士班

摘 要

可調視訊編碼的層間預測和動作補償預測中執行動作預測的區塊其不同的分割模式會導致位元與失真上差異。然而現今只有少數模型可以解釋可調式視訊編碼位元與失真行為，更遑論有任何方法可以讓我們針對可調視訊編碼中不同的區塊分割模式分析其位元與失真關係。針對可調視訊編碼粗略可調性，本論文推導出了一個解析性及模式相依的位元與失真關係模型。考慮到加強層為可採用層間殘留預測的壓縮方式，我們對可調視訊編碼中基礎層與加強層各提供了一取決於區塊分割模式和影像特性的位元與失真模型。在我們所提出的位元與失真模型推演過程中採納了一個向前信道模型以及一個時間上穩態的過程假設，我們藉由一個動作預測軌跡詮釋重建區塊，並且將殘留變異數設計成一個統計量模型。實驗結果顯示，我們提出的模型可以很準確的估量出不同區塊分割模式，其真實壓縮出的基礎層及加強層的位元與失真曲線。並且最後針對不同區塊分割層間殘留預測的效能分析後，所提出的模型也呈現與真時壓縮相似的位元與失真趨勢。

Analytical Mode-Dependent Rate and Distortion Models for H.264/SVC Coarse Grain Scalability

Student : Yu-Chen Tseng

Advisor : Wen-Hsiao Peng

Institute of Multimedia Engineering

National Chiao Tung University

ABSTRACT


In Scalable Video Coding (SVC), the inter-layer prediction and the variable motion estimation block partition modes for motion-compensated prediction (MCP) cause differences in rate and distortion behavior; however, there are just few models could explain the rate and distortion behavior of SVC, not to mention methods which focus on analyzing the rate and distortion of different partition mode pairs in SVC. In this thesis, we derive analytical mode-dependent rate and distortion models for Coarse-grain scalable video coding techniques. The rate and distortion models for base and enhancement layer both depend on the partition mode and sequence characteristics with consideration of the inter-layer residual prediction capability in enhancement layer. Adopting a forward channel model and an assumption of temporal-stationary process in the derivation of proposed models, we interpret the reconstructed block by a motion prediction trajectory and model the transformed residual variance into a mode-dependent statistic. Our experimental results show that the proposed model can estimate the actual-coded R-D curves of different partition modes in base layer and enhancement layer with high accuracy. In addition, similar tendencies between model and actual-coded curve are observed over the performances of different mode pair encoded with inter-layer residual prediction.

誌 謝

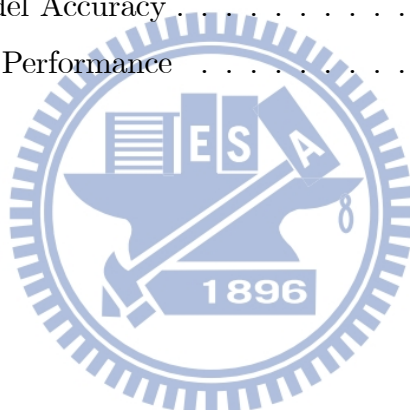
在兩年前我有榮幸進入多媒體架構與處理實驗室，首先，我要感謝我的指導教授—彭文孝 博士，從我大三時起給予我於學問研究上的指導。彭老師的認真勤勉與他的絕頂思路高度相當，做事情追根究柢並且仔細嚴謹的態度，已經成為我在學習與研究路上的典範與楷模。其次，我要感謝吳崇豪同學，在龐大的課業壓力下也不辭辛勞地回答我的問題，給予許多珍貴的意見，並且總是從旁督促、修正我的研究觀念，給予我許多修課、學問上的幫助，使我在這兩年的碩士生涯，不再舉步維艱。在多媒體架構與處理實驗室這個優良的環境下不斷學習，還得感謝曾經一起做研究的學長們—陳漪紋學長、王澤偉學長，以及在我座位旁邊的李宗霖同學；他們總是讓我感受到實驗室的溫暖，不論是在學業上或是生活上的支持，對我來說都是研究生生涯中難以忘懷的。

最後，我要感謝我的家人—曾正義 先生、劉培良 女士的栽培，在爭取碩士學位的路上，讓我有自由的空間，也不責怪我能力不足畢業時期拖得太晚。感謝我的室友們—朱庭玉和單師涵，總在我遇到挫折的時候第一時間給我關懷，生活中互相扶持了兩年，情感如同家人。更要感謝單師涵的媽媽單懷靈 女士，百忙之中也願意幫我的論文做英文文法上的糾正。感謝我的老師、家人、與朋友們，是你們的支持，陪伴我取得學位的一路艱辛，謝謝你們。

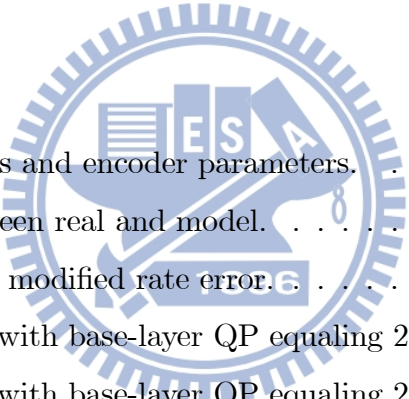
Contents

Contents		i
List of Tables		iii
List of Figures		iv
1 Research Overview		2
1.1 Introduction		2
1.2 Problem Statement		3
1.3 Contributions and Organization of Thesis		4
2 Background		6
2.1 Overview of Hybrid Video Coding		6
2.2 Overview of Scalable Video Coding		7
2.2.1 Concept		7
2.2.2 Coarse-grain Scalable Coding (CGS)		7
2.3 Rate and Distortion Model Based on Laplace Distribution		8
3 Rate Distortion Model: Single Layer		10
3.1 Derivation Outline		10

3.2	Distortion Model for Single Layer	11
3.3	Rate Model for Single Layer	19
3.4	Rate and Distortion Summary for Single Layer	19
4	Rate Distortion Model: Multiple Layers	21
4.1	Distortion Model for Multiple Layers	21
4.2	Approximation Distortion Solution for Multiple Layers	32
4.3	Rate Model for Multiple Layers	34
4.4	Rate and Distortion Summary for Multiple Layers	34
5	Experiments and Analyses	36
5.1	Comparison of Estimation Accuracy	36
5.1.1	Distortion Model Accuracy	37
5.1.2	Rate Model Accuracy	40
5.2	SVC R-D Curve Performance	43
6	Conclusions	46
	Bibliography	47



List of Tables



5.1	Testing conditions and encoder parameters.	37
5.2	PSNR error between real and model.	40
5.3	Entropy rate and modified rate error.	43
5.4	Base-layer usage with base-layer QP equaling 28., Bus(CIF).	44
5.5	Base-layer usage with base-layer QP equaling 28., Football(CIF).	45
5.6	Base-layer usage with base-layer QP equaling 28., Foreman(CIF).	45

List of Figures

2.1	hybrid coding diagram.	6
2.2	H.264 quantization scheme.	8
3.1	Forward channel model and models in matrix notation.	12
3.2	3-D Model of Hybrid Coder.	13
3.3	Motion trajectory for a 16x16 predicted block along the time axis.	13
3.4	Single-layer residual signal generating for a 16x8 predicted block	17
4.1	3-D model of SVC hybrid coder with inter-layer residual prediciton.	22
4.2	EL and BL motion trajectory starting from \mathbf{f}_{k-m}^E for a 16x16 predicted block.	25
4.3	Four relativities of \mathbf{f}_{k-n}^E and $\mathbf{f}_{k-m,k-m-p}^B$ for a 16x16 predicted block.	30
4.4	Multi-layer residual signal generating for BL 16x16 predicted block and EL 16x8 predicted block	32
4.5	Laplace distortion and its approximation.	33
5.1	PSNR v.s. QP curves of BL and EL applying different configuration regression result. (Foreman)	38
5.2	Real v.s. Model D-Q curves applying one regression result. (mobile and foreman sequence)	39

5.3 Entropy curves compared with actual curves (Foreman) 40

5.4 Linearity relationship between $\ln(R)$ and H^* (a) Base mode16x16 (b) Base mode 16x8 (c) Base mode 8x8 form Forman(CIF) 41

5.5 Modified rate(R^*) compared with actual rate, entropy(H) as a contrast. Blue lines with solid squares are BL R-Q curve. Red lines with hollow squares are the curves for entropy v.s. QP. Green line with hollow triangles are the curves for modeified rate v.s. QP 42

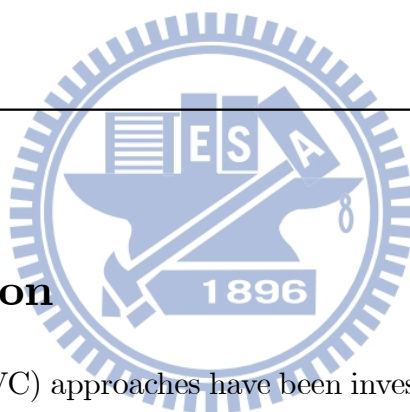
5.6 SVC performance (Football), the dotted lines indecate the distance between simulcast and base layer, as well as the distance between simulcast and inter-layer residual prediction 44



CHAPTER 1

Research Overview

1.1 Introduction



Scalable Video Coding(SVC) approaches have been investigated for more than 20 years to answer demand from various video transmission channels and heterogeneous viewing devices.

In the aspect of traditional non-scalable video coding, hybrid coding has drawn most of the attention in the past several decades, which derives well-known H.264/AVC standard. In hybrid coding, motion-compensated prediction (MCP) is used to exploit temporal similarities between successive video frames (inter-frame coding). Transform coding is then implemented in two steps, first, converting spatial values into transform coefficient values and second, quantizing the coefficients to achieve a lossy compression. In the block-based MCP, each macroblock is split into one or more partition, referred to partition mode, for motion compensation. Different partition mode causes different coding efficiency; when it comes to mode decision, the rate and distortion behavior of each partition mode is then desirable for decision criterion.

Scalable Video Coding as proposed in [1][2] is an extension of H.264/AVC standard.

Instead of independently encoding consecutive spatial layers using MCP based coders, SVC adopts additional inter-layer prediction to exploit statistical dependences between different layers. In comparison to simulcasting different qualities or resolutions, in inter-layer prediction method, pictures with higher quality or resolution levels utilize the information from the lower levels in order to improve coding efficiency. The issue of how to analyse the performance of inter-layer prediction then catches the attention and becomes critical.

1.2 Problem Statement

In transform coding of images and videos, two important factors are coding bit rate R and picture distortion denoted by D . Analysis and estimation of the R-D performance are significant in image and video coding. For example, based on the rate and distortion models, optimum bit allocation as well as other R-D optimization procedures can be adopted to improve the coding efficiency and, consequently, to improve the image quality or video presentation quality. In typical hybrid video coding, the rate and distortion behaviors are relevant to motion estimation partition mode of MCP and quantization method.

Many efforts have been made on deriving rate and distortion model for non-scalable hybrid coding [3][4][5]. Basically, these methods provide analytical or empirical approach to the rate and distortion of overall video sequences. Among the non-scalable rate and distortion models, [4] proposes a quantization-distortion model for H.264/AVC with particular consideration of the motion-compensated prediction effect, however, the non-linear numerical computation required by this model is impractical, and it cannot be used to model the R-D variation between different partition mode in block level. In addition to the non-scalable coding, the rate distortion analysis in [6] gives a framework for evaluating the rate-distortion theoretic lower bound for spatially scalable video coding in general. The approach in [6] is simply an extension of that in an earlier work by B. Girod [7] and the authors in [6] propose ideal assumptions for theoretical analysis which are far from adequate to describe real SVC codec's. For capacity of practical application, an operational and analytical rate distortion model is still needed. In this thesis we derive rate and distortion models, which depend on block partition mode, to

approach the behavior of H.264/AVC coding and its extension SVC with inter-layer residual prediction for Coarse-grain scalable coding. When considering the inter-layer residual prediction, the problem of rate and distortion modeling in H.264/SVC is very challenging. This study aims to provide answers to the following questions:

1. How do we determine the single-layer mode-dependent rate and distortion models based on prediction and quantization schemes of H.264/AVC.
2. How do we extend the single-layer mode-dependent rate and distortion models for SVC inter-layer residual prediction, given quantization parameters q_B for base layer and q_E for enhancement layer?
3. How do the rate and distortion behave with different partition mode and different characteristic factors extracted from an input sequence?
4. How does inter-layer residual prediction perform when applied to different partition mode pairs for MCP?

Since R-D behavior is affected by features of input sequence and quantization parameter, this thesis provides an in-depth study on the relationship between rate distortion and video contents, as well as relationship between rate distortion and quantization parameter in SVC for characterizing the rate-quantization (R-Q) and distortion-quantization (D-Q) models.

1.3 Contributions and Organization of Thesis

Specifically, our main contributions in this work are:

- Two mode-dependent distortion-quantization (D-Q) models are proposed for non-scalable and scalable video coder.
- Two mode-dependent rate-quantization (R-Q) models are conducted for non-scalable and scalable video coder.
- Our analysis is capable of evaluating the R-D performance of different mode pair with SVC inter-layer prediction.

The remaining of this thesis is organized as follows: Chapter 2 contains a review of hybrid coding, Coarse-grain scalability in SVC and a rate distortion model based on Laplace distortion. Chapter 3 presents a derivation of single layer rate and distortion models. And the rate distortion model for multi-layer encoder is introduced in Chapter

4. Chapter 5 provides simulation results of examining the accuracy of proposed rate and distortion models and analyzing the performance of inter-layer prediction. Finally, the thesis is concluded with a summary.



CHAPTER 2

Background

2.1 Overview of Hybrid Video Coding

In hybrid video coding depicted in Fig.2.1, a video sequence is temporally segmented into several groups of pictures (GOP). Each picture is divided into numbers of macroblocks (MBs); each MB is split into one or more MB partitions and an intra or inter prediction is applied on each partition. The error generated as difference between predictor and current block is called motion-compensated prediction (MCP) residual.

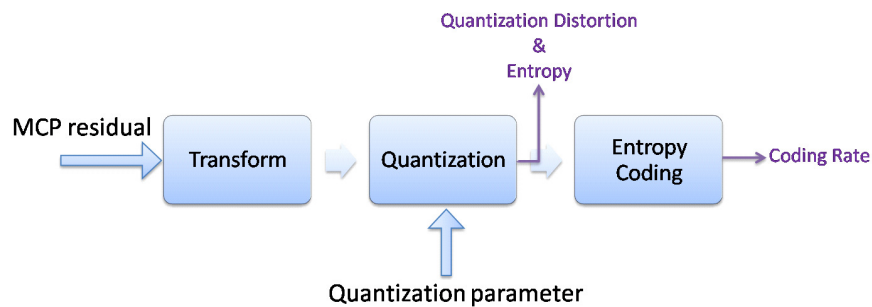


Figure 2.1: hybrid coding diagram.

Then a fixed $M \times M$ block transform, which is commonly a DCT transform, is applied to the prediction residual of inter and intra-prediction modes; the prediction residual blocks are therefore transformed into DCT coefficients. Note that the size of an $M \times M$ transform block is always less than or equal to the partition sizes in a MB. After that, scalar quantization followed by entropy coding is applied to the DCT coefficients. Finally, the quantizer causes the main quality loss of compression, which is quantization distortion D .

2.2 Overview of Scalable Video Coding

2.2.1 Concept

Scalable Video Coding (SVC) standard [2][8][1] is a scalable extension of the H.264/AVC standard developed by the Joint Video Team (JVT), which allows a single bitstream to provide multiple frame sizes, frame rates and quality levels while achieving a reasonable coding efficiency. An SVC bitstream is organized into one base layer (BL) and one or more enhancement layers (EL) in corresponding dimension if it provides certain scalability. A subset of SVC bitstreams can be extracted to form another valid bitstream for a given decoder and be decoded to produce a playback with a reduced reconstruction quality compared to the original bitstream.

SVC supports three types of scalabilities: spatial, temporal and quality scalabilities. Subsets in the spatial scalability bit-stream represent the source content with a reduced picture size (spatial resolution). The temporal scalability is provided by hierarchical temporal prediction structures for each coding layer while quality scalability is achieved by two approaches: *Coarse-grain scalable coding* (CGS), which can be considered as a special case of spatial scalability with identical frame sizes for base and enhancement layer, and *medium-grain scalable coding* (MGS), which provides quality refinement layers inside each spatial layer and enables packet-based quality scalable coding.

2.2.2 Coarse-grain Scalable Coding (CGS)

SVC performs CGS through encoding series of quality layers, which have the same spatial and temporal resolutions. At first, the texture information is encoded into an AVC

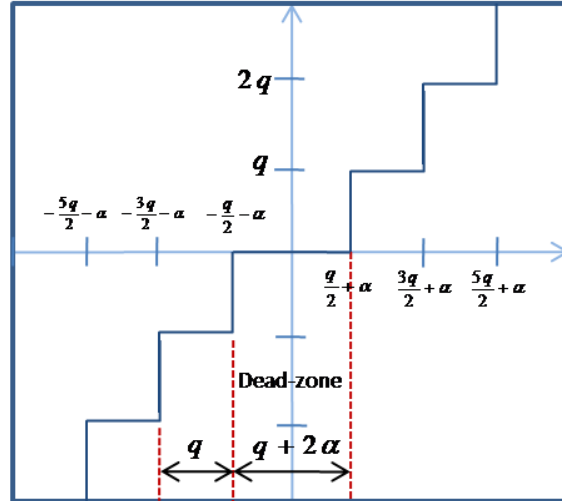


Figure 2.2: H.264 quantization scheme.

compatible bitstream to provide a base layer (BL) with the minimum quality among layers at a given quantization level. At enhancement layers (EL), CGS decreases the quantization step sizes and encodes successive refinements of the transform coefficients. For residual information, inter-layer prediction is employed. The base layer signal of the co-located block is used as prediction for the residual signal of the current enhancement layer macroblock, so that only the corresponding difference signal is coded.

2.3 Rate and Distortion Model Based on Laplace Distribution

Laplace distribution [3][4][9][10] is a well-known distribution which bears resemblance to the distribution for DCT coefficients of images. Due to its low computational complexity and high accuracy, in this thesis, we choose Laplace distribution as the base distribution of transform coefficient in the proposed derivation. A zero-mean Laplace-distributed random variable with probability density function (pdf) is:

$$f(x) = \frac{1}{2\Lambda} e^{-\frac{|x|}{\Lambda}},$$

$$\Lambda = \frac{\sigma}{\sqrt{2}},$$

where x represents the transformed residual, and Laplace parameter Λ is a function of σ , which is their standard deviation indicating the property of the input sequence. Recent coding standards usually adopt the uniform quantizer depicted in Fig. 2.2. The probability that transform coefficient x fall inside each quantization bin i are calculated by

$$P(i) = \begin{cases} \int_{iq-\frac{q}{2}+\alpha}^{iq+\frac{q}{2}+\alpha} (x-iq)^2 f(x)dx & \text{if } i > 0, \\ \int_{-\frac{q}{2}-\alpha}^{\frac{q}{2}+\alpha} x^2 f(x)dx & \text{if } i = 0. \end{cases} \quad (2.1)$$

where q is quantization step size and α is the quantizer *dead-zone* parameter. For H.264/AVC inter frame coding, $\alpha = \frac{q}{3}$. The above probability (2.1) can be computed and then represented by a close form. the close form then introduces the distortion function:

$$D = \sigma^2 - (2\alpha + \sqrt{2}\sigma) \frac{\exp(\frac{-q-2\alpha}{\sqrt{2}\sigma})}{1 - \exp(\frac{-\sqrt{2}q}{\sigma})} q, \quad (2.2)$$

In addition to the distortion model, the entropy of the quantized transformed residuals can also be computed according to the entropy definition with the probability function.

$$\begin{aligned} H &= -\sum_i P(i) \log P(i) \\ &= -P(0) \log P(0) - 2 \sum_{i=1}^{\infty} P(i) \log P(i). \end{aligned}$$

And the closed form of entropy is obtained by:

$$H = -p_0 \log p_0 - \left(\log \frac{c}{p} \right) (1 - p_0) - \frac{2c \log p}{(1 - p)^2}, \quad (2.3)$$

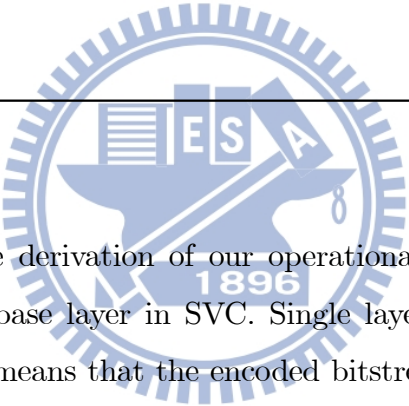
where

$$\begin{aligned} p &= \exp\left(-\frac{\sqrt{2}q}{\sigma}\right), \\ p_0 &= 1 - \exp\left(-\frac{\sqrt{2}\alpha}{\sigma}\right)\sqrt{p}, \\ c &= \frac{1}{2} \exp\left(-\frac{\sqrt{2}\alpha}{\sigma}\right)\sqrt{p}(1 - p). \end{aligned}$$

In the next chapter, the proposed rate distortion model will be discussed in detail based on Laplace distribution.

CHAPTER 3

Rate Distortion Model: Single Layer



This chapter presents the derivation of our operational rate distortion model for a single layer, or so-called base layer in SVC. Single layer, which is equivalent to the well-known AVC coding, means that the encoded bitstream does not contain scalable resolutions or scalable bit-rates.

3.1 Derivation Outline

Based on Laplace distribution and H.264 quantization scheme, a coefficient distortion function of residual variance σ^2 and quantization step size q can be developed as in (2.2). As well as the entropy function is obtain by (2.3).

Given a quantization step size q in the distortion model, the influence of MCP method can only be revealed in the variance of residual transform coefficients σ^2 . To model the impact of MCP schemes on distortion and rate behavior, we introduce a *forward channel model* to help us conveniently construct a hybrid coding flow. Then we formulate a closed-form residual variance function $\sigma^2(q)$ and use it to attain the rate and distortion model.

The MCP prediction error in block base is generated as difference between prediction block in coded reference frame and the current block; current residual is then encoded and the current distortion is formed. Considering that current block distortion is also relevant to distortion of reference block, in this thesis we assume that video sequence is a locally temporal-stationary process.

The statistical models proposed by Tao *et al.* [11] are used to characterize the motion and intensity fields of video signals. These models provide parameters of motion, intensity, and block-partition mode to analyze the block-level motion-compensation predictor; therefore, the closed-form residual variance function $\sigma^2(q)$ we have can also be controlled by those motion, intensity, and block-partition mode parameters. Eventually, a rate distortion model that react to the MCP method is achieved.

3.2 Distortion Model for Single Layer

H.264/AVC is based on the block-based hybrid coding approach. The motion estimation is performed to find the prediction of each macroblock(MB) partition, and DCT transform followed by quantization is applied on each $M \times M$ segment block inside a macroblock individually. It follows that the rate distortion model of a MB can be reduced to modeling an $M \times M$ transform block coverage. Therefore, the derivation of the distortion model of an inter mode is depicted on the basis of an $M \times M$ transformed block. To model the distortion for a whole MB, we only need to model each $M \times M$ transform block separately.

To evaluate the transform domain residual variance σ^2 for distortion function of Laplace-distributed source, we first formulate the prediction error by subtracting the reconstructed reference frame from the current k th original frame. Let \mathbf{f}_k be the vectorization of an $M \times M$ intensity block of an MB to be coded in (current) frame k , and $\tilde{\mathbf{f}}_{k-1}$ be the vectorized motion-compensated prediction of \mathbf{f}_k in the reference frame $k - 1$. The corresponding residual vector \mathbf{e}_k is

$$\mathbf{e}_k = \mathbf{f}_k - \tilde{\mathbf{f}}_{k-1}. \quad (3.1)$$

Let \mathbf{e}_k^T , \mathbf{f}_k^T , and $\tilde{\mathbf{f}}_{k-1}^T$ represent the transformed vectors of \mathbf{e}_k , \mathbf{f}_k , and $\tilde{\mathbf{f}}_{k-1}$, respectively; DCT, \mathbf{H} , is used to transform vertically and horizontally each $M \times M$ block correspond-

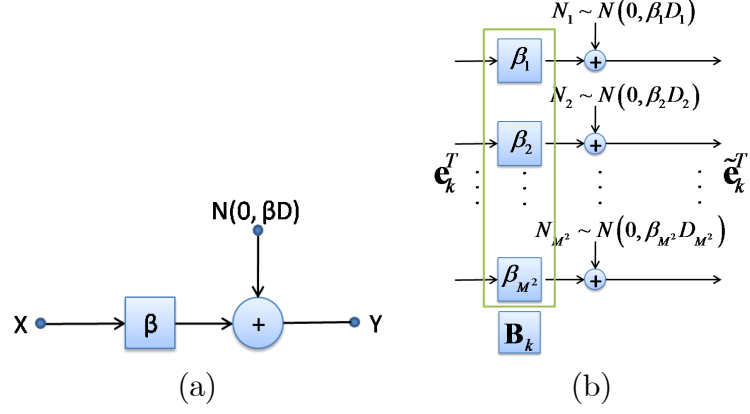


Figure 3.1: Forward channel model and models in matrix notation.

ing to \mathbf{e}_k , \mathbf{f}_k , or $\tilde{\mathbf{f}}_{k-1}$, and the equivalent transform after vectorization is $\mathbf{H} \otimes \mathbf{H}$, where \otimes is the Kronecker product. Since the equivalent transform is linear, (3.1) implies

$$\mathbf{e}_k^T = \mathbf{f}_k^T - \tilde{\mathbf{f}}_{k-1}^T, \quad (3.2)$$

where \mathbf{e}_k^T contains M^2 transform coefficients in column-major order.

Although being followed by specific quantization and entropy coding in hybrid video coding, the transform is instead accompanied with the forward channel model, as shown in Fig.3.1 [12], in the following derivations. It is well-known that if a Gaussian source with mean zero and finite variance σ^2 and an additive Gaussian noise are given, then with proper scaling $\beta = 1 - \frac{D}{\sigma^2}$ of the channel input, direct connection of the source to the channel results in a system that provides an ideal rate distortion function of the source with respect to the squared-error criterion, where D is the squared-error distortion between input and output. Though the forward channel needs a Gaussian source as the input to achieve its ideality and is not quite suitable for the transformed residual signal, it is still adopted in our framework for mathematical tractability.

Based on the optimum forward channel as shown in Fig.3.1(a), we can give a model of hybrid coder as Fig.3.2 with dimension M^2 extension obtained by Fig.3.1(b) for applying intensity input \mathbf{e}_k^T to it. Thus, the reconstruction $\tilde{\mathbf{e}}_k^T$ of prediction error \mathbf{e}_k^T is obtained by:

$$\tilde{\mathbf{e}}_k^T = \mathbf{B}_k \mathbf{e}_k^T + \mathbf{n}_k^T, \quad (3.3)$$

where \mathbf{n}_k^T is a memoryless additive Gaussian noise vector, which is statistically inde-

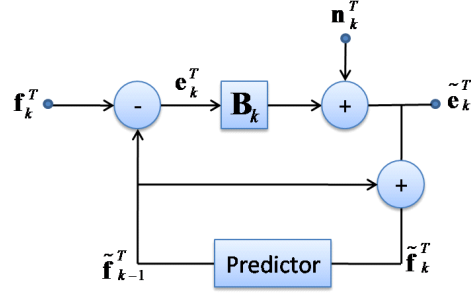


Figure 3.2: 3-D Model of Hybrid Coder.

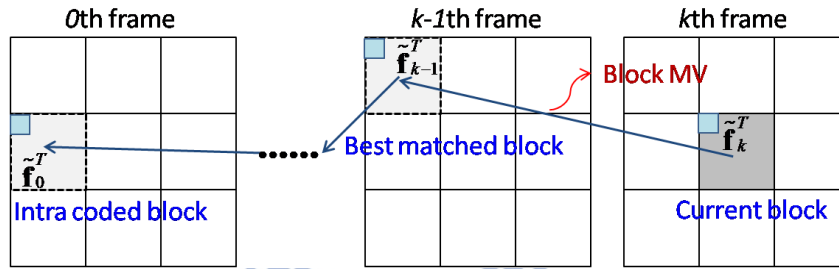


Figure 3.3: Motion trajectory for a 16x16 predicted block along the time axis.

pendent of the input vector. \mathbf{B}_k is an $M^2 \times M^2$ diagonal scaling matrix whose diagonal entries represent the scaling β 's for each coefficient.

Let $\tilde{\mathbf{f}}_k^T$ be the reconstruction of \mathbf{f}_k^T , $\tilde{\mathbf{e}}_k^T$ can be rewritten as

$$\tilde{\mathbf{e}}_k^T = \tilde{\mathbf{f}}_k^T - \tilde{\mathbf{f}}_{k-1}^T. \quad (3.4)$$

Substituting (3.2) and (3.4) into (3.3) gives

$$\tilde{\mathbf{f}}_k^T = \mathbf{B}_k \mathbf{f}_k^T + (\mathbf{I} - \beta_k) \tilde{\mathbf{f}}_{k-1}^T + \mathbf{n}_k^T, \quad (3.5)$$

which shows an affine relation between $\tilde{\mathbf{f}}_k^T$ and $\tilde{\mathbf{f}}_{k-1}^T$. In general, the $M \times M$ predicted block corresponding to $\tilde{\mathbf{f}}_{k-1}^T$ is not aligned with an $M \times M$ transformed block in the reference frame, so (3.5) is only valid for $\tilde{\mathbf{f}}_k^T$. However, we adopt the alignment assumption that (3.5) is valid for every nonnegative integer k , and thus the reconstructed intensity sequence $\left\{ \tilde{\mathbf{f}}_k^T \right\}_{k \geq 0}$ satisfies a recurrence relation and shows a motion trajectory, as shown in Fig.3.3.

Substituting (3.5) into (3.2) recursively gives the closed-form

$$\begin{aligned} \mathbf{e}_k^T &= \mathbf{f}_k^T - \sum_{n=1}^{k-1} \left(\prod_{i=0}^{n-1} (\mathbf{I} - \mathbf{B}_{k-1-i}) \right) \mathbf{B}_{k-1-n} \mathbf{f}_{k-1-n}^T - \sum_{n=1}^{k-1} \left(\prod_{i=0}^{n-1} (\mathbf{I} - \mathbf{B}_{k-1-i}) \right) \mathbf{n}_{k-1-n}^T \\ &\quad - \mathbf{B}_{k-1} \mathbf{f}_{k-1}^T - \mathbf{n}_{k-1}^T - \left(\prod_{i=0}^{k-1} (\mathbf{I} - \mathbf{B}_{k-1-i}) \right) \tilde{\mathbf{f}}_{-1}^T. \end{aligned} \quad (3.6)$$

As different characteristics appear between frames, the scaling matrices \mathbf{B}_k 's may be unequal. Since characteristics between frames do not vary greatly except for some special cases, e.g., scene changes, a temporal-stationary assumption, which assumes the scaling matrices are all equal, i.e., $\mathbf{B}_k = \mathbf{B}$ for every k , is introduced to simplify the very complicated (3.6):

$$\begin{aligned} \mathbf{e}_k^T &= \mathbf{f}_k^T - \sum_{n=0}^{k-1} (\mathbf{I} - \mathbf{B})^n \mathbf{B} \mathbf{f}_{k-1-n}^T - \sum_{n=0}^{k-1} (\mathbf{I} - \mathbf{B})^n \mathbf{n}_{k-1-n}^T - (\mathbf{I} - \mathbf{B})^k \tilde{\mathbf{f}}_{-1}^T \\ &= \mathbf{f}_k^T - \sum_{n=0}^{k-1} \mathbf{A}_n \mathbf{f}_{k-1-n}^T - \sum_{n=0}^{k-1} \mathbf{C}_n \mathbf{n}_{k-1-n}^T, \end{aligned} \quad (3.7)$$

where $\mathbf{A}_n = (\mathbf{I} - \mathbf{B})^n \mathbf{B}$, $\mathbf{C}_n = (\mathbf{I} - \mathbf{B})^n$, $(\mathbf{I} - \mathbf{B})^0 = \mathbf{I}$, and $\tilde{\mathbf{f}}_{-1}^T = \mathbf{0}$. That $\tilde{\mathbf{f}}_{-1}^T$ equals $\mathbf{0}$ indicates the $M \times M$ block corresponding to \mathbf{f}_0^T along the motion trajectory in the first frame ($k = 0$) is intra-coded:

$$\begin{aligned} \tilde{\mathbf{f}}_0^T &= \mathbf{B}_0 \mathbf{f}_0^T + (\mathbf{I} - \mathbf{B}_0) \tilde{\mathbf{f}}_{-1}^T + \mathbf{n}_0^T \\ &= \mathbf{B} \mathbf{f}_0^T + \mathbf{n}_0^T. \end{aligned} \quad (3.8)$$

In general, \mathbf{e}_k is a zero-mean random vector and so is \mathbf{e}_k^T . Thus, the covariance matrix of \mathbf{e}_k^T can be computed as follows:

$$\begin{aligned} E \left\{ \mathbf{e}_k^T (\mathbf{e}_k^T)^t \right\} &= \mathbf{R}_f^T(0) - \sum_{n=0}^{k-1} \mathbf{A}_n \mathbf{R}_f^T(n+1) - \sum_{n=0}^{k-1} (\mathbf{A}_n \mathbf{R}_f^T(n+1))^t \\ &\quad + \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \mathbf{A}_n \mathbf{R}_f^T(n-m) (\mathbf{A}_m)^t + \sum_{n=0}^{k-1} \mathbf{C}_n \mathbf{R}_N^T(0) (\mathbf{C}_n)^t, \end{aligned} \quad (3.9)$$

where

$$\mathbf{R}_f^T(n-m) = E \left\{ \mathbf{f}_{k-n}^T (\mathbf{f}_{k-m}^T)^t \right\} \text{ for } k \geq n, m \geq 0,$$

and

$$\begin{aligned} \mathbf{R}_N^T(n-m) &= E \left\{ \mathbf{n}_{k-n}^T (\mathbf{n}_{k-m}^T)^t \right\} \text{ for } k \geq n, m \geq 0 \\ &= \begin{cases} \mathbf{R}_N^T(0) & \text{for } n = m \\ \mathbf{0} & \text{otherwise} \end{cases}, \end{aligned}$$

where $\mathbf{0}$ is a zero matrix.

While $M \times M$ blocks along the motion trajectory, regarded as a vector-valued random process \mathbf{f}_k^T , is assumed to be a vector-valued wide-sense stationary process, it implies that the autocovariance $\mathbf{R}_f^T(n-m)$ depends only on frame interval $n-m$. Therefore, the two autocovariance functions are independent of the specific frame number k . Noting that any intensity vector and noise vector are statistically independent, we have

$$E \left\{ \mathbf{f}_{k-n}^T (\mathbf{n}_{k-m}^T)^t \right\} = \mathbf{0} \text{ for } k \geq n, m \geq 0.$$

The covariance matrix of \mathbf{e}_k^T can be seen as a generation of the scalar-valued variance $\sigma_k^2(i)$ of i th coefficient of \mathbf{e}_k^T , which is extracted from the i th diagonal element of $E \left\{ \mathbf{e}_k^T (\mathbf{e}_k^T)^t \right\}$:

$$\begin{aligned} \sigma_k^2(i) &= r_f^T(0; i) - 2\beta_i \sum_{n=0}^{k-1} (1-\beta_i)^n r_f^T(n+1; i) + \beta_i^2 \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} (1-\beta_i)^{n+m} r_f^T(n-m; i) \\ &\quad + \sum_{n=0}^{k-1} (1-\beta_i)^{2n} r_N^T(0; i), \end{aligned} \quad (3.10)$$

where $r_f^T(n; i)$ and $r_N^T(0; i)$ are the i th diagonal elements of $\mathbf{R}_f^T(n)$ and $\mathbf{R}_N^T(0)$, respectively.

Instead of deriving variance in a specific frame k , i.e., $\sigma_k^2(i)$, it is more useful to consider the convergent behavior of hybrid coding. Let k goes to infinity and adopt the following Markov-like assumption:

$$r_f^T(n; i) = \begin{cases} r_f^T(0; i) & \text{for } n = 0 \\ (\alpha_i)^{|n|-1} r_f^T(1; i) & \text{otherwise} \end{cases} \text{ where } 1 \geq \alpha_i \geq 0.$$

We then obtain

$$\begin{aligned}\sigma_i^2 &= \lim_{k \rightarrow \infty} \sigma_k^2(i) \\ &= \left(\frac{2}{2-\beta_i}\right) r_f^T(0; i) - \left(\frac{2}{2-\beta_i}\right) \frac{\beta_i}{1-\alpha_i(1-\beta_i)} r_f^T(1; i) + \frac{r_N^T(0; i)}{1-(1-\beta_i)^2}.\end{aligned}\quad (3.11)$$

Equation (3.11) shows a convergent form for the variance of the i th coefficient. Before substituting (3.11) to Laplace distortion function (??), we need to fill up the parameters shown up in (3.11), i.e., α_i , β_i , $r_f^T(0; i)$, $r_f^T(1; i)$, and $r_N^T(0; i)$.

The parameters α_i are the impact factors in temporal correlation between frames which motion-compensated prediction (MCP) tries to exploit for a better prediction efficiency. MCP usual shows a better prediction efficiency when α_i 's are all closed to 1 than that when closed to 0. To model a single-layer video coding with a good MCP scheme, each α_i is approximated as 1 for all coefficients, i.e. $\alpha_i \approx 1$ for all i . As the definition in the theorem T. Berger[12] $\beta_i \approx 1 - \frac{D_i}{\sigma_i^2}$ and $r_N^T(0; i) \approx \beta_i D_i$ (3.11) is deduced:

$$\sigma_i^2 = 2r_f^T(0; i) - 2r_f^T(1; i).\quad (3.12)$$

From the equation (3.12), we conclude that the transformed residual variance σ_i^2 is a function of $r_f^T(0; i)$ and $r_f^T(1; i)$. According to the definition in (3.9), $r_f^T(0; i)$ and $r_f^T(1; i)$ are the i th diagonal elements of $\mathbf{R}_f^T(0)$ and $\mathbf{R}_f^T(1)$, respectively. Regardless of the independent factor, frame number k , $\mathbf{R}_f^T(0)$ is autocovariance of the $M \times M$ transform block \mathbf{f}_k^T and $\mathbf{R}_f^T(1)$ is covariance between \mathbf{f}_k^T and motion compensated reference block \mathbf{f}_{k-1}^T .

In order to analyze the distribution of motion-compensated residuals, Tao *et al.* [11] assumes that the autocorrelation function of the intensity and motion fields can be approximated with a quadratic function and an exponential function, respectively:

$$\begin{aligned}E\{I_k(\mathbf{s}_i) I_k(\mathbf{s}_j)\} &= \sigma_I^2 \left(1 - \frac{\|\mathbf{s}_i - \mathbf{s}_j\|_2^2}{K}\right) \\ E\{v_x(\mathbf{s}_i) v_x(\mathbf{s}_j)\} &= E\{v_y(\mathbf{s}_i) v_y(\mathbf{s}_j)\} = \sigma_m^2 \rho_m^{\|\mathbf{s}_i - \mathbf{s}_j\|_1} \text{ and } E\{\mathbf{v}(\mathbf{s}_i)\} = E\{\mathbf{v}(\mathbf{s}_j)\},\end{aligned}$$

where $I_r(\mathbf{s})$ represents the intensity value of pixel $\mathbf{s} = (x(\mathbf{s}), y(\mathbf{s}))$ in reference frame; $\mathbf{v}(\mathbf{s}) = (v_x(\mathbf{s}), v_y(\mathbf{s}))$ denotes the motion of \mathbf{s} , and $\{\sigma_I^2, K\}$ and $\{\sigma_m^2, \rho_m\}$ are their

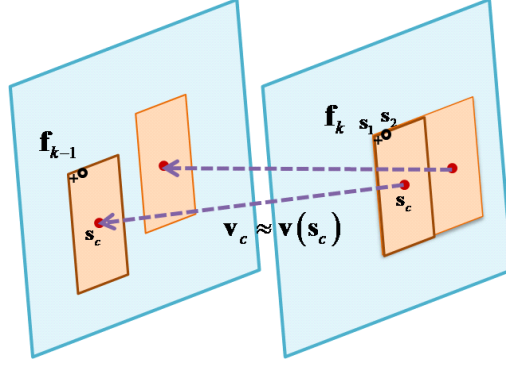


Figure 3.4: Single-layer residual signal generating for a 16x8 predicted block

(respective) variances and correlation coefficients. In [13], these model are further extended to address the motion sampling efficiency for MCP. In the temporal dimension, it is further assumed that $I_k(\mathbf{s}) = I_{k-1}(\mathbf{s} + \mathbf{v}(\mathbf{s}))$, where $I_k(\mathbf{s})$ represents the intensity value of pixel \mathbf{s} in current frame; moreover, a block motion vector \mathbf{v}_c is approximated as the motion at the block center, i.e., $\mathbf{v}_c \approx \mathbf{v}(\mathbf{s}_c)$, and in that regard, block-based motion estimation is seen as a motion sampler.

The \mathbf{f}_k^T and \mathbf{f}_{k-1}^T are the transformd intensity vectors of current block and reference block; however, Tao's model[11] approximates the intensity fields in the spatial domain. To derive the transform domain autocovariance matrices $\mathbf{R}_f^T(0)$ and $\mathbf{R}_f^T(1)$, we first illustrate a current intensity block in matrix form \mathbf{F}_k , and each element of \mathbf{F}_k is in correspondence with each pixel intensity of block. For processing with covariance matrix, we need current intensity vector $\mathbf{f}_k = \text{vec}(\mathbf{F}_k)$ in column-major order:

$$\mathbf{F}_k = \begin{bmatrix} I_k(\mathbf{s}_1) & I_k(\mathbf{s}_{M+1}) & \cdots & I_k(\mathbf{s}_{M^2-M}) \\ I_k(\mathbf{s}_2) & I_k(\mathbf{s}_{M+2}) & \cdots & I_k(\mathbf{s}_{M^2-M+1}) \\ \vdots & \vdots & \ddots & \vdots \\ I_k(\mathbf{s}_M) & I_k(\mathbf{s}_{2M}) & \cdots & I_k(\mathbf{s}_{M^2}) \end{bmatrix}$$

$$\longrightarrow \mathbf{f}_k = \left[I_k(\mathbf{s}_1) \quad I_k(\mathbf{s}_2) \quad \cdots \quad I_k(\mathbf{s}_{M^2}) \right]^t$$

Considering for the transformed block \mathbf{F}_k^T in matrix form, which is a 2-D transformation by $M \times M$ DCT matrix \mathbf{H} , we can achieve the transformed vector $\mathbf{f}_k^T = \text{vec}(\mathbf{F}_k^T)$ in column-major order:

$$\mathbf{F}_k^T = \mathbf{H}\mathbf{F}_k\mathbf{H}^t \longrightarrow \mathbf{f}_k^T = (\mathbf{H} \otimes \mathbf{H}) \mathbf{f}_k$$

$$\begin{aligned}
 \mathbf{R}_f^T(0) &= E \left\{ \mathbf{f}_k^T (\mathbf{f}_k^T)^t \right\} \\
 &= E \left\{ (\mathbf{H} \otimes \mathbf{H}) \mathbf{f}_k ((\mathbf{H} \otimes \mathbf{H}) \mathbf{f}_k)^t \right\} \\
 &= (\mathbf{H} \otimes \mathbf{H}) E \left\{ \mathbf{f}_k \mathbf{f}_k^t \right\} (\mathbf{H} \otimes \mathbf{H})^t
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{R}_f^T(1) &= E \left\{ \mathbf{f}_{k-1}^T (\mathbf{f}_k^T)^t \right\} \\
 &= E \left\{ (\mathbf{H} \otimes \mathbf{H}) \mathbf{f}_{k-1} ((\mathbf{H} \otimes \mathbf{H}) \mathbf{f}_k)^t \right\} \\
 &= (\mathbf{H} \otimes \mathbf{H}) E \left\{ \mathbf{f}_{k-1} \mathbf{f}_k^t \right\} (\mathbf{H} \otimes \mathbf{H})^t
 \end{aligned}$$

We define the spatial domain covariance matrix $E \{ \mathbf{f}_k \mathbf{f}_k^t \}$ and $E \{ \mathbf{f}_{k-1} \mathbf{f}_k^t \}$ to be $\mathbf{R}_f(0)$ and $\mathbf{R}_f(1)$. The (i, j) th element of $\mathbf{R}_f(0)$, or $[E \{ \mathbf{f}_k \mathbf{f}_k^t \}]_{ij}$, is then computed with assistance of Tao's model [11]:

$$\begin{aligned}
 E \{ I_k(\mathbf{s}_i) I_k(\mathbf{s}_j) \} &= E \{ I_{k-1}(\mathbf{s}_i + \mathbf{v}(\mathbf{s}_i)) I_{k-1}(\mathbf{s}_j + \mathbf{v}(\mathbf{s}_j)) \} \\
 &= E \left\{ \sigma_I^2 \left(1 - \frac{\|(\mathbf{s}_i - \mathbf{s}_j) + (\mathbf{v}(\mathbf{s}_i) - \mathbf{v}(\mathbf{s}_j))\|_2^2}{K} \right) \right\} \\
 &= \sigma_I^2 \left(1 - \frac{\|\mathbf{s}_i - \mathbf{s}_j\|_2^2}{K} - \frac{4\sigma_m^2 (1 - \rho_m^{\|\mathbf{s}_i - \mathbf{s}_j\|_1})}{K} \right). \tag{3.13}
 \end{aligned}$$

The (i, j) th element of $\mathbf{R}_f(1)$, or $[E \{ \mathbf{f}_{k-1} \mathbf{f}_k^t \}]_{ij}$:

$$\begin{aligned}
 E \{ I_{k-1}(\mathbf{s}_i + \mathbf{v}_c) I_k(\mathbf{s}_j) \} &\approx E \{ I_{k-1}(\mathbf{s}_i + \mathbf{v}(\mathbf{s}_c)) I_{k-1}(\mathbf{s}_j + \mathbf{v}(\mathbf{s}_j)) \} \\
 &= \sigma_I^2 \left(1 - \frac{\|\mathbf{s}_i - \mathbf{s}_j\|_2^2}{K} - \frac{4\sigma_m^2 (1 - \rho_m^{\|\mathbf{s}_c - \mathbf{s}_j\|_1})}{K} \right) \tag{3.14}
 \end{aligned}$$

Finally, we estimate the parameters σ_I^2 , K , σ_m^2 , and ρ_m , which account for the sequence characteristics, and the distortion model is obtained by substituting residual variance (3.12) to the Laplace distortion function (2.2). The algorithm is detail in the following section and the estimation of parameter is described in experiment ch.5.

3.3 Rate Model for Single Layer

The entropy (2.3), denoted as H , of the quantized transformd residuals is actually far from the true coding rate such it is the measure of independent coding. In hybrid coding, quantized transform residuals are always dependently coded, like run-length coding, at a block level. And it is extremely difficult to redeem this inaccuracy caused by dependent coding. The authors in [9] have noticed a stable relationship between the real coded rate R and entropy H and modified the rate model by involving some correcciton factor. However, we bring out a more steady and more significant linearity relationship between natural logarithm of real rate $\ln(R)$ and $\ln\left(\left(\frac{1}{q}\right)^{\left(1+\frac{\sqrt{2}}{\sigma}\right)} H^{\frac{\sqrt{2}}{\sigma}} \exp\left(-\frac{\sqrt{2}}{\sigma}\right)\right)$. Then we proposed a new and more accurate rate model:

$$\begin{aligned}\ln R &\approx a \ln \left(\left(\frac{1}{q} \right)^{\left(1 + \frac{\sqrt{2}}{\sigma} \right)} H^{\frac{\sqrt{2}}{\sigma}} \exp\left(-\frac{\sqrt{2}}{\sigma}\right) \right) + b \\ R &\approx q^{-\left(1+\frac{\sqrt{2}}{\sigma}\right)a} H^{\frac{\sqrt{2}a}{\sigma}} \exp\left(b - \frac{\sqrt{2}a}{\sigma}\right)\end{aligned}\quad (3.15)$$

where the a and b are both constans at prediction mode level, and σ^2 is the estimated residual variance by computing (3.12).

3.4 Rate and Distortion Summery for Single Layer

In this section, we give a summerized algorithm for a single layer modeling through previously proposed rate and distortion model. Some external parameters need to be provided as inputs.

Input: Variance of motion field, σ_m^2 ,
 Correlation coefficient of motion field, ρ_m ,
 Variance of Intensity field, σ_I^2 ,
 Positive number, K ,
 Quantization step of q ,
 Block partition mode, or prediction block center, s_c .

Output: Coding bit rate, $R(q)$,
 Quantization distortion $D(q)$.

1. Residual variance for a $M \times M$ transform block coefficients:

1.1 Compute spatial domain parameters:

$\mathbf{R}_f(0)$ by (3.13), $\mathbf{R}_f(1)$ by (3.14).

1.2 Transform residual covariance matrix:

$$\mathbf{R}_f^T(0) = (\mathbf{H} \otimes \mathbf{H}) \mathbf{R}_f(0) (\mathbf{H} \otimes \mathbf{H})^t.$$

$$\mathbf{R}_f^T(1) = (\mathbf{H} \otimes \mathbf{H}) \mathbf{R}_f(1) (\mathbf{H} \otimes \mathbf{H})^t.$$

1.3 Compute single layer i th variance of coefficients:

σ_i^2 by (3.12),

2. R-D for a $M \times M$ transform block:

2.1 Compute quantization distortion of i th coefficient:

$D_i(q)$ by substituting σ_i^2 into (2.2).

2.2 Compute entropy of i th coefficient:

$H_i(q)$ by (2.3),

2.3 Output:

Average variance per pixel, $\sigma^2 = \frac{1}{M^2} \sum_{i=1}^{M^2} \sigma_i^2$.

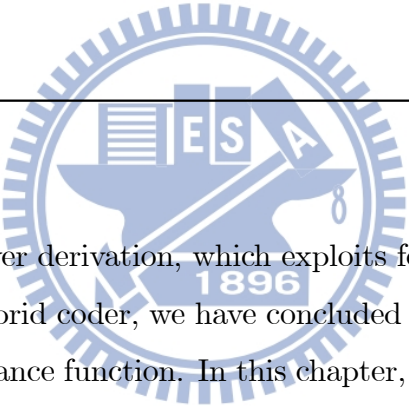
Average entropy per pixel, $H(q) = \frac{1}{M^2} \sum_{i=1}^{M^2} H_i(q)$.

Quantization distortion, $D(q) = \frac{1}{M^2} \sum_{i=1}^{M^2} D_i(q)$.

Average bit per pixel, $R(q)$ by substituting $H(q)$ and σ^2 to (3.15).

CHAPTER 4

Rate Distortion Model: Multiple Layers



In the preceding single layer derivation, which exploits forward channel model to form a model of single-layer hybrid coder, we have concluded the D-Q and R-Q function by an estimated residual variance function. In this chapter, we present the rate distortion model of the enhancement layer in Coarse-Grain scalable video coding, which can utilize inter-layer residual prediction to achieve a better prediction efficiency. Without loss of generality, a two-layer scenario is studied in the thesis for simplicity.

4.1 Distortion Model for Multiple Layers

The framework of a two-layer CGS coder base on the forward channel model is depicted in Fig. 4.1

We denote the transform-domain base-layer signal with a superscript B and transform-domain enhancement-layer signal with E . For example, \mathbf{e}_k^B and \mathbf{e}_k^E represent the transform-domain prediction residual vector for the base layer and enhancement layer, respectively, of an $M \times M$ transform block of the current k th original frame. Note that in CGS, the base layer input vector \mathbf{f}_k^B and the input of enhancement layer \mathbf{f}_k^E

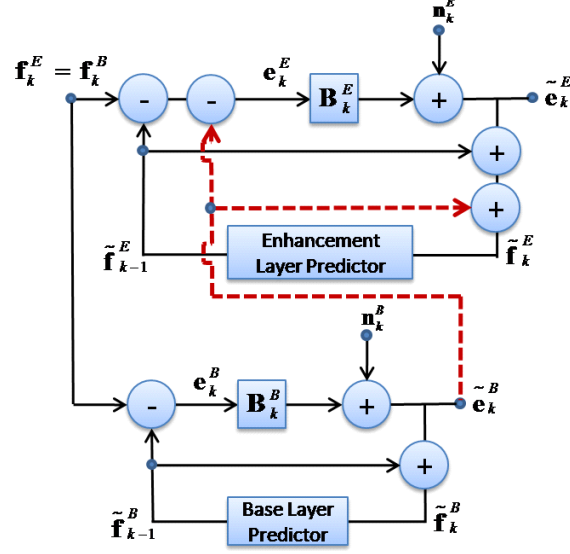


Figure 4.1: 3-D model of SVC hybrid coder with inter-layer residual prediction.

are vectorization of co-located $M \times M$ transform blocks of the current k th original frame, which are identical because both of base and enhancement layers are in the same resolution.

Fig. 4.1 illustrates the framework of a two-layer scalable coder with inter-layer residual prediction. The dashed lines in Fig. 4.1 indicate the inter-layer prediction propagation path of reconstructed base-layer residual $\tilde{\mathbf{e}}_k^B$. In that way, the enhancement layer prediction residual \mathbf{e}_k^E to be coded exploits the spatial redundancies by subtracting the base layer coded residual $\tilde{\mathbf{e}}_k^B$. The enhancement layer transformed residual vector is

$$\mathbf{e}_k^E = \mathbf{f}_k^E - \tilde{\mathbf{f}}_{k-1}^E - \tilde{\mathbf{e}}_k^B. \quad (4.1)$$

The reconstruction $\tilde{\mathbf{e}}_k^E$ of inter-layer prediction error \mathbf{e}_k^E is then made through the forward channel model:

$$\tilde{\mathbf{e}}_k^E = \beta_k^E \mathbf{e}_k^E + \mathbf{n}_k^E, \quad (4.2)$$

or it can also be made as:

$$\tilde{\mathbf{e}}_k^E = \tilde{\mathbf{f}}_k^E - \tilde{\mathbf{f}}_{k-1}^E - \tilde{\mathbf{e}}_k^B. \quad (4.3)$$

Substituting (4.1) and (4.3) into (4.2) gives transform domain representation of enhancement layer coded frame with Inter-layer residual prediction based on one base-

layer

$$\tilde{\mathbf{f}}_k^E = \mathbf{B}_k^E \mathbf{f}_k^E + (\mathbf{I} - \mathbf{B}_k^E) \tilde{\mathbf{f}}_{k-1}^E + (\mathbf{I} - \mathbf{B}_k^E) \tilde{\mathbf{e}}_k^B + \mathbf{n}_k^E. \quad (4.4)$$

Adopting the same alignment assumption in single-layer derivation and temporal-stationary assumption, which assumes $\mathbf{B}_k^E = \mathbf{B}^E$ for every k , we substitute (4.4) into (4.1) recursively and then give the closed-form

$$\begin{aligned} \mathbf{e}_k^E &= \mathbf{f}_k^E - \sum_{n=1}^{k-1} \left(\prod_{i=0}^{n-1} (\mathbf{I} - \mathbf{B}_{k-1-i}^E) \right) \mathbf{B}_{k-1-n}^E \mathbf{f}_{k-1-n}^E - \sum_{n=1}^{k-1} \left(\prod_{i=0}^{n-1} (\mathbf{I} - \mathbf{B}_{k-1-i}^E) \right) \mathbf{n}_{k-1-n}^E \\ &\quad - \mathbf{B}_{k-1}^E \mathbf{f}_{k-1}^E - \mathbf{n}_{k-1}^E - \left(\prod_{i=0}^{k-1} (\mathbf{I} - \mathbf{B}_{k-1-i}^E) \right) \tilde{\mathbf{f}}_{-1}^E - \sum_{n=0}^{k-1} \left(\prod_{i=0}^n (\mathbf{I} - \mathbf{B}_{k-1-i}^E) \right) \tilde{\mathbf{e}}_{k-1-n}^B - \tilde{\mathbf{e}}_k^B \\ &= \underbrace{\mathbf{f}_k^E - \sum_{n=0}^{k-1} \mathbf{A}_n^E \mathbf{f}_{k-1-n}^E - \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{n}_{k-1-n}^E}_{\triangleq \omega_k^E} - \underbrace{\sum_{m=0}^{k-1} \mathbf{C}_m^E \tilde{\mathbf{e}}_{k-m}^B}_{\triangleq \omega_k^B}, \end{aligned} \quad (4.5)$$

where $\mathbf{A}_n^E = (\mathbf{I} - \mathbf{B}^E)^n \mathbf{B}^E$, $\mathbf{C}_n^E = (\mathbf{I} - \mathbf{B}^E)^n$, and $\mathbf{f}_{k-n}^E, 1 \leq n \leq k$ are reference blocks in the EL motion trajectory starting from \mathbf{f}_k^E , with $\tilde{\mathbf{e}}_{k-n}^B, 1 \leq n \leq k$ denoting their co-located BL coded residual blocks defined as:

$$\tilde{\mathbf{e}}_{k-m}^B = \mathbf{B}_k^B \mathbf{e}_{k-m}^B + \mathbf{n}_{k-m}^B.$$

To explain the reconstructed block in the first intra-coded frame ($k = 0$), we define $\tilde{\mathbf{f}}_{-1}^E = \mathbf{0}$ and $\tilde{\mathbf{e}}_0^B = \mathbf{0}$. And the definition $\tilde{\mathbf{e}}_0^B = \mathbf{0}$ indicates that there is no inter-layer residual prediction in intra coding. Then $\tilde{\mathbf{f}}_0^E$ is obtained by:

$$\begin{aligned} \tilde{\mathbf{f}}_0^E &= \mathbf{B}_0^E \mathbf{f}_0^E + (\mathbf{I} - \mathbf{B}_0^E) \tilde{\mathbf{f}}_{-1}^E + (\mathbf{I} - \mathbf{B}_0^E) \tilde{\mathbf{e}}_0^B + \mathbf{n}_0^E \\ &= \mathbf{B}_0^E \mathbf{f}_0^E + \mathbf{n}_0^E. \end{aligned}$$

The original residual signal of k th frame in single layer \mathbf{e}_k^T is modeled as (3.7); each $M \times M$ transform block $\mathbf{f}_{k-n}^T, 1 \leq n \leq k$ is along an independent single-layer motion trajectory predicted starting from \mathbf{f}_k^T . Nevertheless, in the inter-layer prediction, \mathbf{e}_{k-m}^B is the BL original residual vector of the block which corresponds with the EL block \mathbf{f}_{k-m}^E in $(k-m)$ th frame. And \mathbf{f}_{k-m}^E is the reference block in the EL motion trajectory

starting from \mathbf{f}_k^E . Therefore, we define \mathbf{e}_{k-m}^B as:

$$\mathbf{e}_{k-m}^B = \mathbf{f}_{k-m}^E - \sum_{p=0}^{k-1} \mathbf{A}_p^B \mathbf{f}_{k-m,k-m-1-p}^B - \sum_{p=0}^{k-1} \mathbf{C}_p^B \mathbf{n}_{k-m,k-m-1-p}^B,$$

where $\mathbf{A}_p^B = (\mathbf{I} - \mathbf{B}^B)^p \mathbf{B}^B$, $\mathbf{C}_p^B = (\mathbf{I} - \mathbf{B}^B)^p$, and $\mathbf{f}_{k-m,k-m-1-p}^B$, $0 \leq p \leq k-1$ are reference blocks in the BL motion trajectory starting from \mathbf{f}_{k-m}^E illustrated in Fig. 4.2, in other words, the starting block in the BL prediction trajectory is the co-located block \mathbf{f}_{k-m}^E in EL. And $\mathbf{n}_{k-m,k-m-1-p}^B$, $0 \leq p \leq k-1$ are corresponding Gaussian noise vectors. Hence, $\tilde{\mathbf{e}}_{k-m}^B$ can be rewritten as:

$$\tilde{\mathbf{e}}_{k-m}^B = \mathbf{B}^B \left(\mathbf{f}_{k-m}^E - \sum_{p=0}^{k-m-1} \mathbf{A}_p^B \mathbf{f}_{k-m,k-m-1-p}^B - \sum_{p=0}^{k-m-1} \mathbf{C}_p^B \mathbf{n}_{k-m,k-m-1-p}^B \right) + \mathbf{n}_{k-m}^B. \quad (4.6)$$

For tractability, we separate the right-hand side of (4.5) into two terms represented as ω_k^E and ω_k^B . According to (4.6), the last term in (4.5) is obtained by:

$$\begin{aligned} \omega_k^B &= \sum_{m=0}^{k-1} \mathbf{C}_m^E \tilde{\mathbf{e}}_{k-m}^B \\ &= \sum_{m=0}^{k-1} \mathbf{C}_m^E \left(\mathbf{B}^B \left(\mathbf{f}_{k-m}^E - \sum_{p=0}^{k-m-1} \mathbf{A}_p^B \mathbf{f}_{k-m,k-m-1-p}^B - \sum_{p=0}^{k-m-1} \mathbf{C}_p^B \mathbf{n}_{k-m,k-m-1-p}^B \right) + \mathbf{n}_{k-m}^B \right) \end{aligned} \quad (4.7)$$

Since \mathbf{e}_k^E is a zero-mean random vector, the transform-domain EL residual covariance is shown as:

$$E \{ \mathbf{e}_k^E (\mathbf{e}_k^E)^t \} = E \left\{ \omega_k^E (\omega_k^E)^t \right\} - E \left\{ \omega_k^E (\omega_k^B)^t \right\} - \left(E \left\{ \omega_k^E (\omega_k^B)^t \right\} \right)^t + E \left\{ \omega_k^B (\omega_k^B)^t \right\}, \quad (4.8)$$

and we can observe that the first squared expectation term is arithmetically identical to the derivation of single layer covariance matrix of \mathbf{e}_k^E (3.9).

$$\begin{aligned} E \left\{ \omega_k^E (\omega_k^E)^t \right\} &= \mathbf{R}_f^E(0) - \sum_{n=0}^{k-1} \mathbf{A}_n^E \mathbf{R}_f^E(n+1) - \sum_{n=0}^{k-1} (\mathbf{A}_n^E \mathbf{R}_f^E(n+1))^t \\ &\quad + \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \mathbf{A}_n^E \mathbf{R}_f^E(n-m) (\mathbf{A}_m^E)^t + \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{R}_N^E(0) (\mathbf{C}_n^E)^t, \end{aligned} \quad (4.9)$$

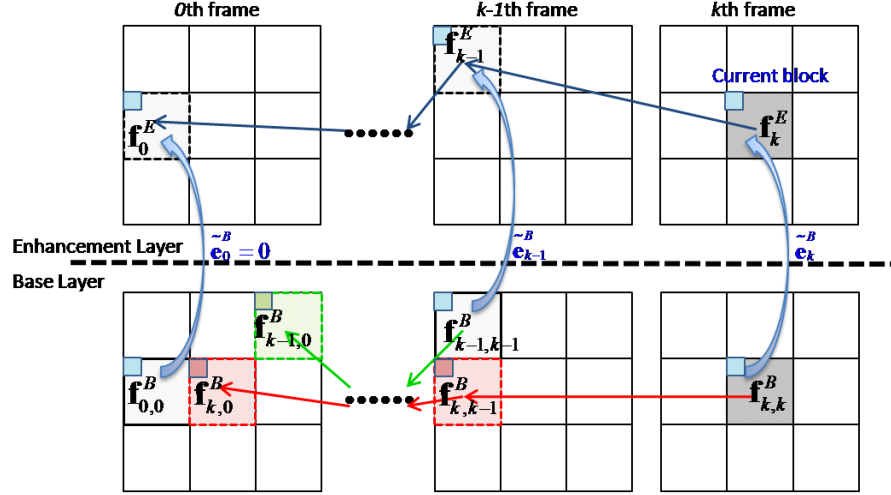


Figure 4.2: EL and BL motion trajectory starting from \mathbf{f}_{k-m}^E for a 16x16 predicted block.

where

$$\mathbf{R}_f^E(n-m) = E \left\{ \mathbf{f}_{k-n}^E (\mathbf{f}_{k-m}^E)^t \right\} \text{ for } k \geq n, m \geq 0,$$

and

$$\begin{aligned} \mathbf{R}_N^E(n-m) &= E \left\{ \mathbf{n}_{k-n}^E (\mathbf{n}_{k-m}^E)^t \right\} \text{ for } k \geq n, m \geq 0 \\ &= \begin{cases} \mathbf{R}_N^E(0) & \text{for } n = m \\ \mathbf{O} & \text{otherwise} \end{cases} \end{aligned}$$

where the autocovariances $\mathbf{R}_f^E(n-m)$ and $\mathbf{R}_N^E(n-m)$ depend only on frame interval $n-m$ according to the assumption of wide-sense stationary process as in single layer. Therefore, we maintain our focus on the cross expectation term

$$E \left\{ \omega_k^E (\omega_k^B)^t \right\} = E \left\{ \left(\mathbf{f}_k^E - \sum_{n=0}^{k-1} \mathbf{A}_n^E \mathbf{f}_{k-1-n}^E - \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{n}_{k-1-n}^E \right) \left(\sum_{n=0}^{k-1} \mathbf{C}_n^E \tilde{\mathbf{e}}_{k-n}^B \right)^t \right\}. \quad (4.10)$$

Since intensity vectors and noise vectors are statistically independent, by substitut-

ing (4.7), we expand (4.10) into:

$$\begin{aligned}
 & E\{\omega_k^E (\omega_k^B)^t\} \\
 &= E \left\{ \underbrace{\left(\mathbf{f}_k^E - \sum_{n=0}^{k-1} \mathbf{A}_n^E \mathbf{f}_{k-1-n}^E \right) \left(\sum_{n=0}^{k-1} \mathbf{C}_n^E \beta^B \mathbf{f}_{k-n}^E - \sum_{n=0}^{k-1} \sum_{p=0}^{k-n-1} \mathbf{C}_n^E \mathbf{B}^B \mathbf{A}_p^B \mathbf{f}_{k-n,k-n-1-p}^B \right)}_{\Phi_k^f} \right\}^t \\
 &+ E \left\{ \underbrace{\left(\sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{n}_{k-1-n}^E \right) \left(\sum_{n=0}^{k-1} \sum_{p=0}^{k-n-1} \mathbf{C}_n^E \mathbf{B}^B \mathbf{C}_p^B \mathbf{n}_{k-n,k-n-1-p}^B - \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{n}_{k-n}^B \right)}_{\Phi_k^N} \right\}^t, \tag{4.11}
 \end{aligned}$$

Noting that \mathbf{n}_k^B and \mathbf{n}_k^E are memoryless additive Gaussian noise vectors, which can be deemed as being uncorrelated with the input sequence and independent of any other signals. We conclude

$$E \left\{ \mathbf{n}_{k-n}^E (\mathbf{n}_{k-m}^B)^t \right\} \approx \mathbf{O} \text{ for } k \geq n, m \geq 0.$$

The expected value of Φ_k^N can be easily worked out:

$$\begin{aligned}
 E \{ \Phi_k^N \} &= \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} \mathbf{C}_n^E E \left\{ \mathbf{n}_{k-1-n}^E (\mathbf{n}_{k-m,k-m-1-p}^B)^t \right\} (\mathbf{C}_m^E \mathbf{B}^B \mathbf{C}_p^B)^t \\
 &- \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \mathbf{C}_n^E E \left\{ \mathbf{n}_{k-1-n}^E (\mathbf{n}_{k-m}^B)^t \right\} (\mathbf{C}_m^E)^t \\
 &\approx \mathbf{O}
 \end{aligned}$$

The cross expectation term (4.11) is then computed as follows:

$$\begin{aligned}
 E\{\omega_k^E (\omega_k^B)^t\} &= E\{\Phi_k^f\} + \mathbf{O} \\
 &= \sum_{m=0}^{k-1} \mathbf{R}_f^E(m) (\mathbf{C}_m^E \mathbf{B}^B)^t - \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} \mathbf{R}_f^{BE}(m+1, p) (\mathbf{C}_m^E \mathbf{B}^B \mathbf{A}_p^B)^t \\
 &\quad - \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \mathbf{A}_n^E \mathbf{R}_f^E(m-n-1) (\mathbf{C}_m^E \mathbf{B}^B)^t \\
 &\quad + \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} \mathbf{A}_n^E \mathbf{R}_f^{BE}(m-n, p) (\mathbf{C}_m^E \mathbf{B}^B \mathbf{A}_p^B)^t
 \end{aligned} \tag{4.12}$$

where

$$\mathbf{R}_f^{BE}(m-n, p) = E\left\{\mathbf{f}_{k-n}^E (\mathbf{f}_{k-m, k-m-p}^B)^t\right\} \text{ for } k \geq m, n, p \geq 0.$$

The matrix $\mathbf{R}_f^{BE}(m-n, p)$ is the covariance between EL block \mathbf{f}_{k-n}^E and BL reference block $\mathbf{f}_{k-m, k-m-p}^B$ in the BL motion trajectory starting from \mathbf{f}_{k-m}^E , where the argument $m-n$ signifies the EL motion trajectory distance between \mathbf{f}_{k-m}^E and \mathbf{f}_{k-n}^E .

The last expected value of (4.8) is:

$$\begin{aligned}
 E\left\{\omega_k^B (\omega_k^B)^t\right\} &= E\left\{\left(\sum_{n=0}^{k-1} \mathbf{C}_n^E \tilde{\mathbf{e}}_{k-n}^B\right) \left(\sum_{n=0}^{k-1} \mathbf{C}_n^E \tilde{\mathbf{e}}_{k-n}^B\right)^t\right\} \\
 &= \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \mathbf{C}_n^E E\left\{\tilde{\mathbf{e}}_{k-n}^B (\tilde{\mathbf{e}}_{k-m}^B)^t\right\} (\mathbf{C}_m^E)^t.
 \end{aligned} \tag{4.13}$$

Here we assume that coded BL residuals of different frames are uncorrelated, that is

$$E\left\{\tilde{\mathbf{e}}_{k-n}^B (\tilde{\mathbf{e}}_{k-m}^B)^t\right\} = \mathbf{O}, \text{ for } m \neq n.$$

Therefore, (4.13) is derived into:

$$\begin{aligned}
 E\left\{\omega_k^B (\omega_k^B)^t\right\} &\approx \sum_{n=0}^{k-1} \mathbf{C}_n^E E\left\{\tilde{\mathbf{e}}_{k-n}^B (\tilde{\mathbf{e}}_{k-n}^B)^t\right\} (\mathbf{C}_n^E)^t \\
 &= \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{B}^B E\left\{\mathbf{e}_{k-n}^B (\mathbf{e}_{k-n}^B)^t\right\} (\mathbf{C}_n^E \mathbf{B}^B)^t + \sum_{n=0}^{k-1} \mathbf{C}_n^E \mathbf{R}_N^B(0) (\mathbf{C}_n^E)^t
 \end{aligned} \tag{4.14}$$

where

$$E \left\{ \mathbf{e}_{k-n}^B (\mathbf{n}_{k-m}^B)^t \right\} = \mathbf{O} \text{ for } k \geq n, m \geq 0$$

and

$$\begin{aligned} \mathbf{R}_N^B(n-m) &= E \left\{ \mathbf{n}_{k-n}^B (\mathbf{n}_{k-m}^B)^t \right\} \text{ for } k \geq n, m \geq 0 \\ &= \begin{cases} \mathbf{R}_N^B(0) & \text{for } n = m \\ \mathbf{O} & \text{otherwise} \end{cases} \end{aligned}$$

We substitute (4.9), (4.12), and (4.14) to (4.8) and consequently obtain the covariance matrix $E \left\{ \mathbf{e}_k^E (\mathbf{e}_k^E)^t \right\}$ of transformed EL residual. The EL variance $\sigma_{E,k}^2(i)$ of i th coefficient is extracted from the i th diagonal element of $E \left\{ \mathbf{e}_k^E (\mathbf{e}_k^E)^t \right\}$. And the diagonal elements of $E \left\{ \omega_k^E (\omega_k^B)^t \right\}$ equal the diagonal elements of $\left(E \left\{ \omega_k^E (\omega_k^B)^t \right\} \right)^t$, that is, $\left[E \left\{ \omega_k^E (\omega_k^B)^t \right\} \right]_{ii} = \left[\left(E \left\{ \omega_k^E (\omega_k^B)^t \right\} \right)^t \right]_{ii}$. The residual variance $\sigma_E^2(i)$ is then obtain by letting k goes to infinity:

$$\begin{aligned} \sigma_E^2(i) &= \lim_{k \rightarrow \infty} \sigma_{E,k}^2(i) \\ &= \lim_{k \rightarrow \infty} \left[E \left\{ \mathbf{e}_k^E (\mathbf{e}_k^E)^t \right\} \right]_{ii} \\ &= \lim_{k \rightarrow \infty} \left[E \left\{ \omega_k^E (\omega_k^E)^t \right\} \right]_{ii} - \lim_{k \rightarrow \infty} 2 \left[E \left\{ \omega_k^E (\omega_k^B)^t \right\} \right]_{ii} + \lim_{k \rightarrow \infty} \left[E \left\{ \omega_k^B (\omega_k^B)^t \right\} \right]_{ii} \end{aligned}$$

and adopt the following assumption:

$$r_f^B(n; i) = \begin{cases} r_f^B(0; i) & \text{for } n = 0 \\ (\alpha_i^B)^{|n|-1} r_f^B(1; i) & \text{otherwise} \end{cases} \text{ where } 1 \geq \alpha_i^B \geq 0, \quad (4.15)$$

$$r_f^E(n; i) = \begin{cases} r_f^E(0; i) & \text{for } n = 0 \\ (\alpha_i^E)^{|n|-1} r_f^E(1; i) & \text{otherwise} \end{cases} \text{ where } 1 \geq \alpha_i^E \geq 0. \quad (4.16)$$

$$r_f^{BE}(m-n, p; i) = \begin{cases} r_f^{BE}(-1, 1; i) & \text{for } m-n = -1, p = 1, \\ (\alpha_i^B)^{|p|} r_f^B(1; i) & \text{for } m-n = 0, \\ (\alpha_i^B)^{|p|} (\alpha_i^E)^{|m-n|-1} r_f^E(1; i) & \text{for } m-n > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (4.17)$$

where $1 \geq \alpha_i^B, \alpha_i^E \geq 0$.

where $r_f^B(n; i)$ and $r_f^E(n; i)$ are the i th diagonal elements of $\mathbf{R}_f^B(n)$ and $\mathbf{R}_f^E(n)$. In (4.17), $r_f^{BE}(m-n, p; i)$ is i th diagonal element of $\mathbf{R}_f^{BE}(m-n, p)$, which is a cross-covariance matrix between the EL block \mathbf{f}_{k-n}^E and the BL reference block $\mathbf{f}_{k-m, k-m-p}^B$ in the BL motion trajectory starting from \mathbf{f}_{k-m}^E . Based on parameters m , n , and p , The relativity of \mathbf{f}_{k-n}^E and $\mathbf{f}_{k-m, k-m-p}^B$ is explained by four classified cases. A Markov-like assumption is utilized for the calculation as well.

1. when $m-n = -1$, $p = 1$, or $n = m+1$, $p = 1$, the matrix $\mathbf{R}_f^{BE}(-1, 1)$ can be evaluated with assistance of Tao's model [11], and $r_f^{BE}(-1, 1; i)$ is the i th diagonal element of $\mathbf{R}_f^{BE}(-1, 1)$. This trajectory situation is depicted in Fig. 4.3(a).
2. Since \mathbf{f}_{k-m}^E and $\mathbf{f}_{k-m, k-m}^B$ are co-located blocks, we can conclude that $\mathbf{f}_{k-m}^E = \mathbf{f}_{k-m, k-m}^B$. In the case of $m-n = 0$, $\mathbf{R}_f^{BE}(0, p) = E \left\{ \mathbf{f}_{k-m}^E (\mathbf{f}_{k-m, k-m-p}^B)^t \right\}$ can be rewritten as $\mathbf{R}_f^{BE}(0, p) = E \left\{ \mathbf{f}_{k-m, k-m}^B (\mathbf{f}_{k-m, k-m-p}^B)^t \right\}$. According to the assumption of wide-sense stationary process, $\mathbf{R}_f^{BE}(0, p)$ is equivalent to $\mathbf{R}_f^B(p) = E \left\{ \mathbf{f}_k^B (\mathbf{f}_{k-p}^B)^t \right\}$, where p is the frame interval between $\mathbf{f}_{k-m, k-m-p}^B$ and \mathbf{f}_{k-m}^E . The above trajectory situation is illustrated in Fig. 4.3(b).
3. Fig. 4.3(c) is the trajectory situation of $m-n > 0$. We suppose that the target reference block depends only on the last block and not on the entire past trajectory.
4. In addition to above three cases, the remain case shown in Fig. 4.3(d) is approximated as zero.

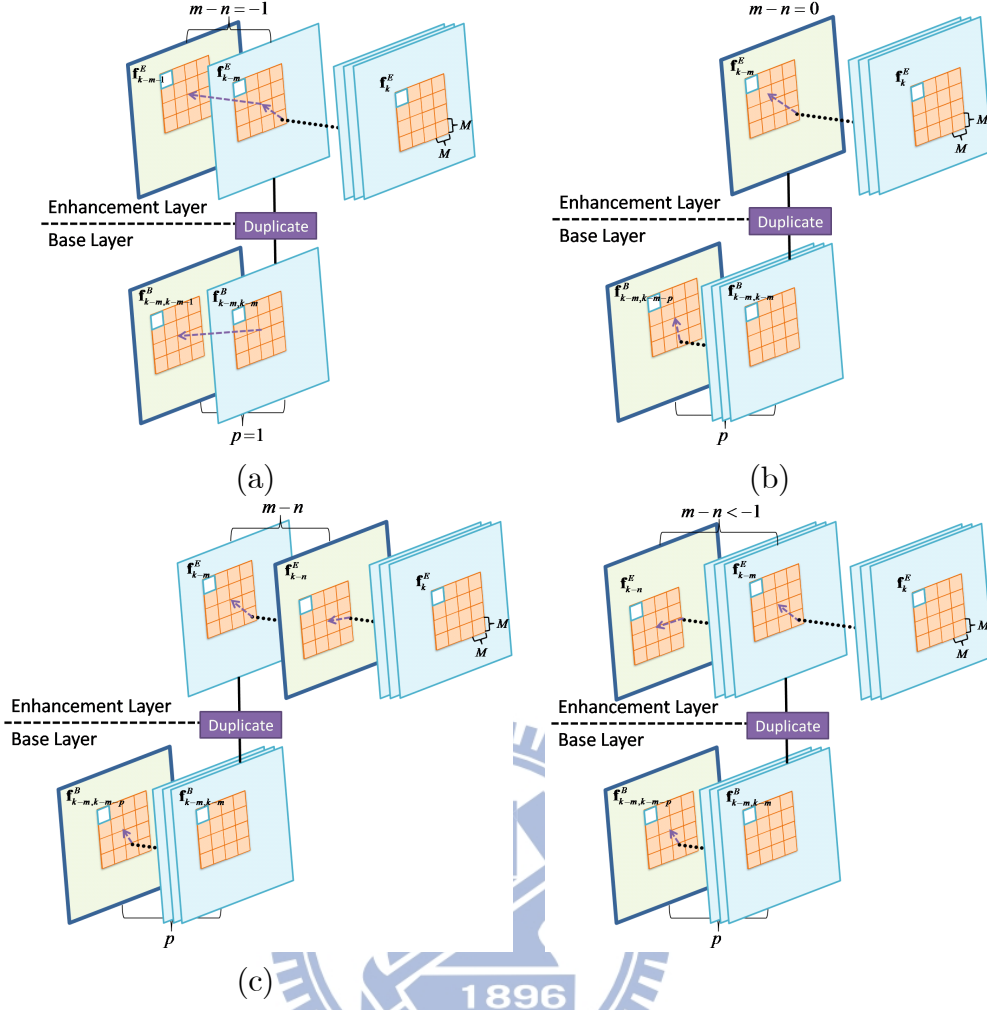


Figure 4.3: Four relativities of \mathbf{f}_{k-n}^E and $\mathbf{f}_{k-m,k-m-p}^B$ for a 16x16 predicted block.

After computation, we obtain

$$\begin{aligned}
 \sigma_E^2(i) &= \left(\frac{2}{2 - \beta_i^E} \right) \left(r_f^E(0; i) - r_f^E(1; i) \frac{\beta_i^E}{1 - \alpha_i^E(1 - \beta_i^E)} \right) \\
 &+ \frac{r_N^E(0; i)}{1 - (1 - \beta_i^E)^2} - \frac{\beta_i^B}{\beta_i^E (\beta_i^E - 2)} \sigma_B^2(i) \\
 &- \frac{2\beta_i^B}{2 - \beta_i^E} \left(r_f^E(0; i) - \frac{1}{1 - \alpha_i^E(1 - \beta_i^E)} \left(\beta_i^E + \frac{\beta_i^B \alpha_i^B (1 - \beta_i^E)}{1 - \alpha_i^B(1 - \beta_i^B)} \right) r_f^E(1; i) \right) \\
 &+ \frac{2(\beta_i^B)^2}{2 - \beta_i^E} \left(\frac{1}{1 - \alpha_i^B(1 - \beta_i^B)} r_f^B(1; i) - r_f^{BE}(-1, 1; i) \right)
 \end{aligned}$$

where $r_f^B(n; i)$, $r_f^E(n; i)$, $r_f^{BE}(0; i)$ and $r_N^E(0; i)$ are the i th diagonal elements of $\mathbf{R}_f^B(n)$, $\mathbf{R}_f^E(n)$, $\mathbf{R}_f^{BE}(0)$ and $\mathbf{R}_N^E(0)$, respectively. The parameters α_i^B and α_i^E are the impact factors in temporal correlation between frames of base and enhancement layer. There are the same reasons in single layer derivation, α_i^B and α_i^E are approximated as 1 for all

coefficients, i.e. $\alpha_i^B = \alpha_i^E \approx 1$ for all i . As the definition in the theorem T. Berger[12] $\beta_i^E \approx 1 - \frac{D_E(i)}{\sigma_E^2(i)}$ and $r_N^E(0; i) \approx \beta_i^E D_E(i)$, the EL residual variance of i th coefficient is

$$\begin{aligned} \sigma_E^2(i) &= \frac{\sigma_E^2(i)}{\sigma_E^2(i) + D_E(i)} \frac{\sigma_E^2(i)}{\sigma_E^2(i) - D_E(i)} (\beta_i^B \sigma_B^2(i) + 2\beta_i^B r_f^E(1; i)) \\ &+ \frac{\sigma_E^2(i)}{\sigma_E^2(i) + D_E(i)} (2(1 - \beta_i^B) r_f^E(0; i) - 2r_f^E(1; i)) \\ &+ \frac{\sigma_E^2(i)}{\sigma_E^2(i) + D_E(i)} (2\beta_i^B r_f^B(1; i) - 2(\beta_i^B)^2 r_f^{BE}(-1, 1; i) + D_E(i)) \end{aligned}$$

which is a function of enhancement layer distortion $D_E(i)$ and base layer information $\beta_i^B = 1 - \frac{D_E(i)}{\sigma_E^2(i)}$; we rewrite it into a quadratic equation:

$$(\mathcal{A} - \mathcal{B}) D_E(i) + (\mathcal{C} - \mathcal{A} + \mathcal{B} + D_E(i)) \sigma_E^2(i) - (\sigma_E^2(i))^2 = 0 \quad (4.18)$$

where

$$\begin{aligned} \mathcal{A} &= (2r_f^E(0; i) + 2r_f^{BE}(-1, 1; i) \beta_i^B - 2r_f^B(1; i)) \beta_i^B, \\ \mathcal{B} &= 2r_f^E(0; i) - 2r_f^E(1; i), \\ \mathcal{C} &= (\sigma_B^2(i) + 2r_f^E(1; i)) \beta_i^B. \end{aligned}$$

The parameters $r_f^E(0; i)$, $r_f^E(1; i)$, $r_f^B(0; i)$, and $r_f^B(1; i)$ have been explicated in previous chapter as (3.13) and (3.14). Now $r_f^{BE}(-1, 1; i)$, which is the (i, i) th element of $\mathbf{R}_f^{BE}(-1, 1)$, is obtained by the same method in previous chapter. First, we need to compute the spatial domain covariance matrix. Then we transform the covariance matrix into DCT domain to achieve $\mathbf{R}_f^{BE}(-1, 1)$. Note that \mathbf{f}^B and \mathbf{f}^E are co-located $M \times M$ transform blocks of the current original frame, i.e., $\mathbf{f}^B = \mathbf{f}^E$, however, for different prediction mode in BL and EL, they have different corresponding motion $\mathbf{v}_B(\mathbf{s}_B)$ and $\mathbf{v}_E(\mathbf{s}_E)$ since a block motion vector is approximated as the motion at the block center. Fig. 4.4 illustrates the prediction situation. Thus, the (i, j) th element of the

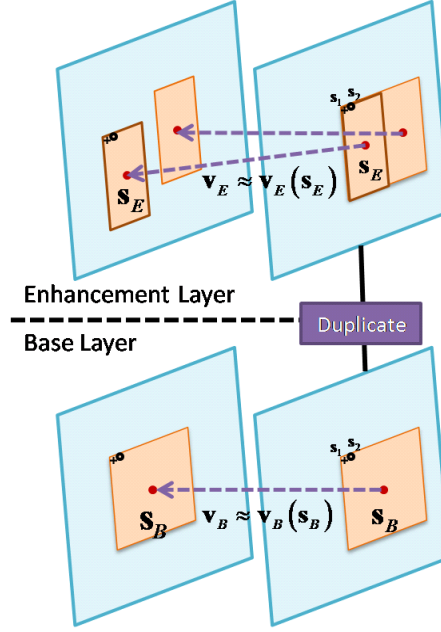


Figure 4.4: Multi-layer residual signal generating for BL 16x16 predicted block and EL 16x8 predicted block

spatial domain covariance matrix is

$$\begin{aligned}
 & E \{ I_{k-1}(\mathbf{s}_i + \mathbf{v}_E(\mathbf{s}_E)) I_{k-1}(\mathbf{s}_j + \mathbf{v}_B(\mathbf{s}_B)) \} \\
 &= E \left\{ \sigma_I^2 \left(1 - \frac{\|(\mathbf{s}_i - \mathbf{s}_j) + (\mathbf{v}_E(\mathbf{s}_E) - \mathbf{v}_B(\mathbf{s}_B))\|_2^2}{K} \right) \right\} \\
 &= \sigma_I^2 \left(1 - \frac{\|\mathbf{s}_i - \mathbf{s}_j\|_2^2}{K} - \frac{4\sigma_m^2 (1 - \rho_m^{\|\mathbf{s}_E - \mathbf{s}_B\|_1})}{K} \right). \tag{4.19}
 \end{aligned}$$

4.2 Approximation Distortion Solution for Multiple Layers

A problem that arises in solving EL variance equation is that the equation (4.18) has unknown parameter, distortion of EL $D_E(i)$. In this thesis, we approximate that the residual coefficients of inter-layer residual prediction obey Laplace distribution. Therefore, the distortion and entropy function is the same as single layer's, and we can obtain the EL variance by solving the simultaneous equation which contains EL variance function (4.18) and Laplace distortion function (2.2). The equation is very

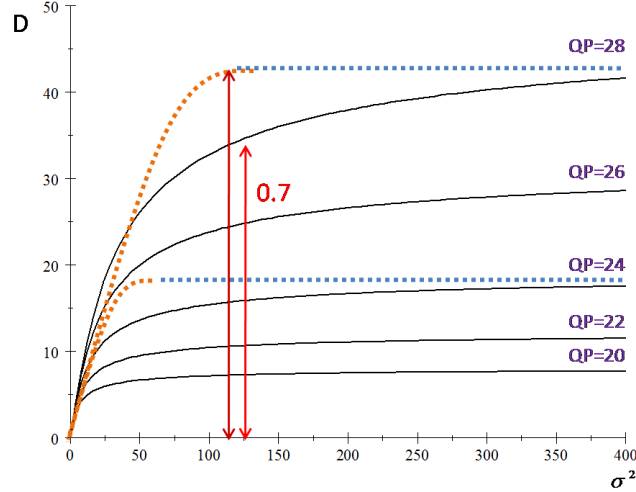


Figure 4.5: Laplace distortion and its approximation.

hard to solve because Laplace distortion is a nonlinear function (2.2). The Laplace distortion of five different QP values are schematized in Fig.4.5, and it has the same curve tendency of the other QP values. We can observe that given different QP, there is a distortion upper bound quickly reached by growing σ^2 . We can easily bring out the bounding value of distortion for different QP

$$D_{upper} = \frac{q^2}{12} + \alpha^2.$$

where q is the corresponding quantization step size. In the case of very small error variance σ^2 , however, the distortion would make a great deal of difference from its upper bound D_{upper} . We use parabolas through the origin to approach the distortion curves in the region of small σ^2 :

$$\begin{aligned} D_{para} &= m\sigma^2 + n\sigma \\ &= m \left(\sigma + \frac{n}{2m} \right)^2 - \frac{n^2}{4m}, \end{aligned}$$

where m and n depend on the vertex (a, b) of the parabola. The vertex (a, b) is designed as the point when Laplace distortion is 0.7 times the distortion upper bound D_{upper} sketched in Fig. 4.5. Thus, m and n are given as:

$$n = \frac{2b}{a}, m = -\frac{b}{a^2}.$$

And the approximation curves are illustrated in Fig. 4.5.

Instead of nonlinearly solving the simultaneous equation of (4.18) and (2.2), we substitute D_{upper} or D_{para} into (4.18) to solve the EL residual variance. A table of the change points $\delta(q)$ for each QP is builded up beforehand. As a result, if the solution of residual variance $\sigma_E^2(i)$ by substituting D_{upper} into (4.18), which is

$$(\mathcal{A} - \mathcal{B}) \left(\frac{q^2}{12} + \alpha^2 \right) + \left(\mathcal{C} - \mathcal{A} + \mathcal{B} + \frac{q^2}{12} + \alpha^2 \right) \sigma_E^2(i) - (\sigma_E^2(i))^2 = 0, \quad (4.20)$$

is smaller than $\delta(q)$ looked up from the change point table, i.e., $\sigma_E^2(i) < \delta(q)$, we re-solve it by substituting D_{para} :

$$\mathcal{C} - (\mathcal{A} - \mathcal{B}) + n(\mathcal{A} - \mathcal{B}) + (n + m(\mathcal{A} - \mathcal{B}) - 1) \sigma_E^2(i) + m(\sigma_E^2(i))^2 = 0. \quad (4.21)$$

where

$$\begin{aligned} \mathcal{A} &= (2r_f^E(0; i) + 2r_f^{BE}(-1, 1; i) \beta_i^B - 2r_f^B(1; i)) \beta_i^B, \\ \mathcal{B} &= 2r_f^E(0; i) - 2r_f^E(1; i), \\ \mathcal{C} &= (\sigma_B^2(i) + 2r_f^E(1; i)) \beta_i^B. \end{aligned}$$

The final answer of residual variance $\sigma_E^2(i)$ is substituted into Laplace distribution distortion function (2.2) ultimately for achieving enhancement layer distortion.

4.3 Rate Model for Multiple Layers

The rate model of multi-layer is identical to the rate function of single layer in Sec. 5.1.2. We substitute the EL residual variance with inter-layer residual prediction into the rate function.

4.4 Rate and Distortion Summary for Multiple Layers

To summarize, we present an algorithm for modeling a two-layer coding with inter-layer residual prediction. Single-layer model is used as the base-layer model in SVC.

Input: Variance of motion field, σ_m^2 ,
 Correlation coefficient of motion field, ρ_m ,
 Variance of intensity field, σ_I^2 ,
 Positive number, K ,
 Quantization step of BL, q_B , and EL, q_E ,
 Turning point $\delta(q_E)$,
 Block partition mode pair, s_B and s_E .

Output: Coding bit rate of BL and EL, $R_B(q_B)$, $R_E(q_E)$,
 Quantization distortion of BL and EL, $D_B(q_B)$, $D_E(q_E)$.

1. Compute base layer rate and distortion:

$R_B(q_B)$, $D_B(q_B)$ and $\sigma_B^2(i)$ for each coefficient by Sec.3.4.

2. Compute residual variance for a EL $M \times M$ transform block:

2.1 Spatial domain EL parameters:

by (3.13), (3.14), and (4.19)

2.2 Transform residual covariance matrix:

$$\mathbf{R}_f^T = (\mathbf{H} \otimes \mathbf{H}) \mathbf{R}_f (\mathbf{H} \otimes \mathbf{H})^t.$$

2.3 Obtain EL i th coefficient of residual variance:

$\sigma_E^2(i)$ = solution of (4.20),

if $\sigma_E^2(i) > \delta(q_E)$ then $\sigma_E^2(i)$ = solution of (4.21).

3. Rate distortion for a EL $M \times M$ transform block:

3.1 Compute quantization distortion of i th coefficient:

$D_E(q_E, i)$ by (2.2)

3.2 Compute entropy of EL i th coefficient:

$H_E(q_E, i)$ by (2.3),

Average entropy per pixel, $H_E(q_E) = \frac{1}{M^2} \sum_{i=1}^{M^2} H_E(q_E, i)$.

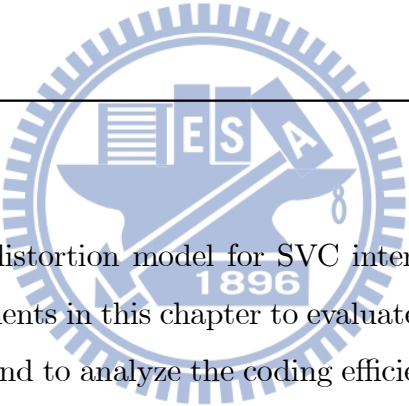
3.3 Output:

Quantization distortion, $D_E(q_E) = \frac{1}{M^2} \sum_{i=1}^{M^2} D_E(q_E, i)$.

Average bit per pixel, $R_E(q_E)$ by substituting $H_E(q_E, i)$ to (3.15) .

CHAPTER 5

Experiments and Analyses



Having derived our rate distortion model for SVC inter-layer residual prediction, we conduct extensive experiments in this chapter to evaluate the accuracy of the proposed R-D estimation method, and to analyze the coding efficiency of the inter-layer residual prediction by using our model. We compare our model with eight common test video sequences in CIF and 4CIF format encoded by SVC reference software JSVM 9.19.8[14] into two quality layers. The proposed model is based on the analysis of MCP coding for different partition mode. In our experiments, we test the proposed method in IPPP coding structure. Table 5.1 details above encoder setting. And the mode 16×8 and the mode 8×16 are viewed as the same due to their symmetry.

5.1 Comparison of Estimation Accuracy

As shown in the R-D function summary of single layer in Sec. 3.4 and multi-layer in Sec. 4.4, we need to estimate the sequence characteristics σ_I^2 , K , σ_m^2 , and ρ_m . Those parameters are addressed from the statistical models, and their estimations have been described in the Tao *et al.* [11]. Instead of the estimation methods in [11], we measure

Sequence	CIF@30Hz, 4CIF@30Hz (120 frames)
DCT transform size	4×4
Prediction structure	1 Reference Frame + IPPP...
Intra period	-1
ME partition mode pair (BL/EL)	$16 \times 16/16 \times 16$, $16 \times 16/16 \times 8$, $16 \times 16/8 \times 8$, $16 \times 8/16 \times 16$, $16 \times 8/16 \times 8$, $16 \times 8/8 \times 16$, $16 \times 8/8 \times 8$, $8 \times 8/16 \times 16$, $8 \times 8/16 \times 8$, $8 \times 8/8 \times 8$
Base layer QP	24, 26, 28, 30, 32, 34, 36, 38, 40
Enhancement layer QP	BL_QP-4, BL_QP-6, BL_QP-8
Inter-layer residual prediction	On

Table 5.1: Testing conditions and encoder parameters.

each parameter through a PSNR curves regression, which is described in a following, to find the accuracy of the proposed rate distortion model in the better condition.

The designed experiment can be described as following steps:

1. Encode sequence into two-layer SVC bitstream with inter-layer residual prediction based on the setting in Table 5.1.
2. PSNR curves regression for the variables σ_I^2 , K , σ_m^2 , and ρ_m : Apply conditions with BL mode 16×16 , EL mode 16×16 , and QP difference 4 to the algorithm in Sec. 3.4 and Sec. 4.4. Then sum up the PSNR values of every BL QP and find the parameter set that gives the least sum-up PSNR difference from the real.
3. Take the solutions of variables σ_I^2 , K , σ_m^2 , and ρ_m from step 2 as inputs for the algorithm in Sec. 3.4 and Sec. 4.4 to generate rate distortion model of every configuration in Table 5.1.

As our rate distortion model is a combination of distortion function of QP(D-Q) and rate function of QP(R-Q), we examine the accuracy of distortion and rate separately, and the following experiments are presented with considering only Luma component. Besides, considering the simulcast for enhancement layer, we stretch the QP range of BL PSNR curves regression up to 16, and demonstrate the experiment results in this QP range.

5.1.1 Distortion Model Accuracy

The proposed rate distortion model, can be used to estimate the D-Q curves of the encoded video sequences. The encoding distortion is measured in terms of PSNR between the encoded video and the original one.

As stated before, the proposed experiment requires estimations for σ_I^2 , K , σ_m^2 , and

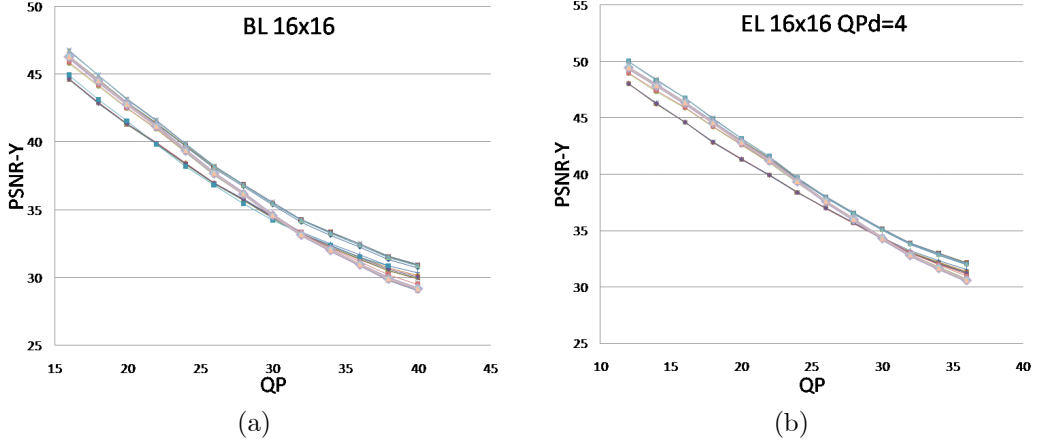


Figure 5.1: PSNR v.s. QP curves of BL and EL applying different configuration regression result. (Foreman)

ρ_m . Despite that different mode pair has slightly different regression parameters, we only apply the regression result from the situation of BL mode 16×16 , EL mode 16×16 , and QP difference 4. It is reasonable to adopt this simple and convenient method since the parameters associate with sequence characteristics not the ME mode pair. And the D-Q curves of the mode pair of BL mode 16×16 and EL mode 16×16 under different configurations (mode pair and QP difference) regression result is demonstrated in Fig.5.1.

The example is the result of *Foreman*(CIF). We can observe some divergences of curves under different regression results in both BL and EL.

The D-Q curves estimated with the only one regression result of mode pair of BL 16×16 and EL 16×16 are compared with the actual D-Q curves in Fig.5.2. It can be seen that for the two testing sequences with distinct characteristics, the D-Q curves estimated by the proposed model fit the actual curves very well; the preciseness is maintained at other QP differences and sequences. To substantiate our claim, Table 5.2 provides the results of all testing sequences in different QP differences; the numbers represent the average PSNR error of every mode. It is seen that for both base and enhancemet layer, there is no more than 0.5dB PSNR error across many sequences and QP differences except, the worst case, BL of *Mobile*(CIF).

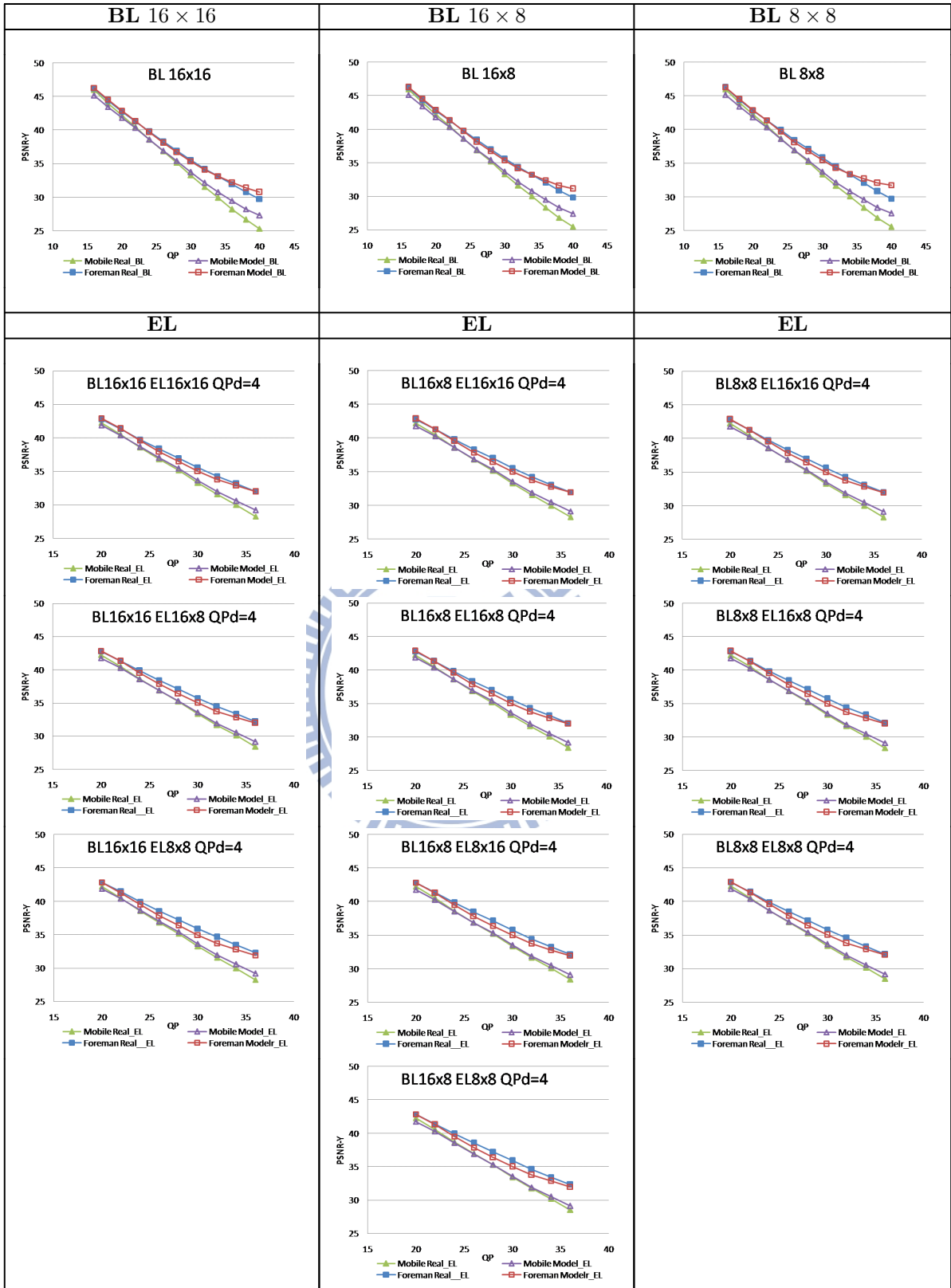


Figure 5.2: Real v.s. Model D-Q curves applying one regression result. (mobile and foreman sequence)

Resolution	Sequence	PSNR error			
		BL(Δ dB)	EL(Δ dB)		
			Qpd=4	Qpd=6	Qpd=8
CIF	Bus	0.36	0.32	0.32	0.34
	Football	0.40	0.24	0.18	0.16
	Foreman	0.33	0.31	0.25	0.20
	Mobile	0.67	0.49	0.50	0.50
4CIF	City	0.22	0.29	0.24	0.22
	Crew	0.41	0.24	0.19	0.16
	Harbour	0.25	0.30	0.28	0.27
	Soccer	0.33	0.26	0.20	0.17
Average		0.37	0.31	0.27	0.25

Table 5.2: PSNR error between real and model.

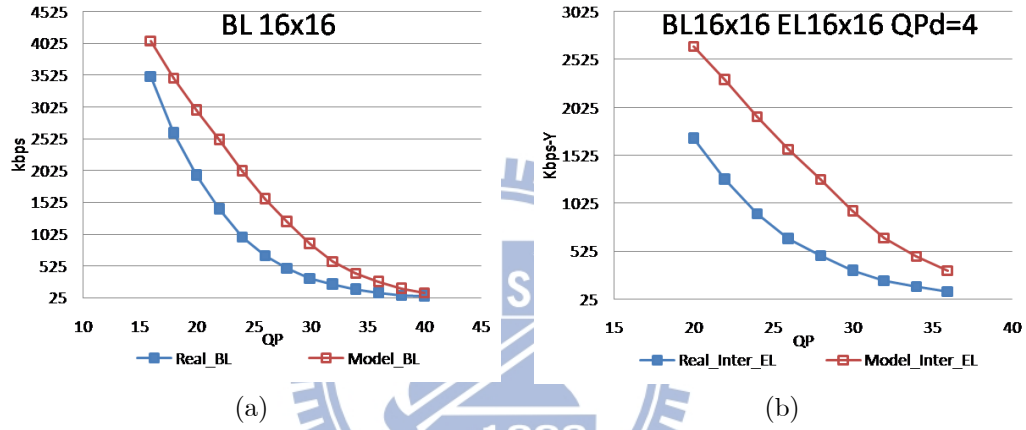


Figure 5.3: Entropy curves compared with actual curves (Foreman)

5.1.2 Rate Model Accuracy

In JSVM, the entropy coding design includes *Context-adaptive binary arithmetic coding* (CABAC) and *Context-adaptive variable-length coding* (CAVLC). As the bitrate in video compression is highly related to the entropy coding method and the dependency of quantized coefficients at block level, it is a very difficult problem to conclude a rate model.

Based on the assumption of Laplace distribution, the entropy can be obtain as (2.3). Since entropy is a measurement for the case of independent coding, we can observe wide discrepancies between the actual R-Q curves and entropy versus QP curves of *Foreman*(CIF) in Fig.5.3 as example. Note that we only present the residual rates in the actual R-Q curves.

The actual rates are encoded by CABAC and the entropy curves is computed with

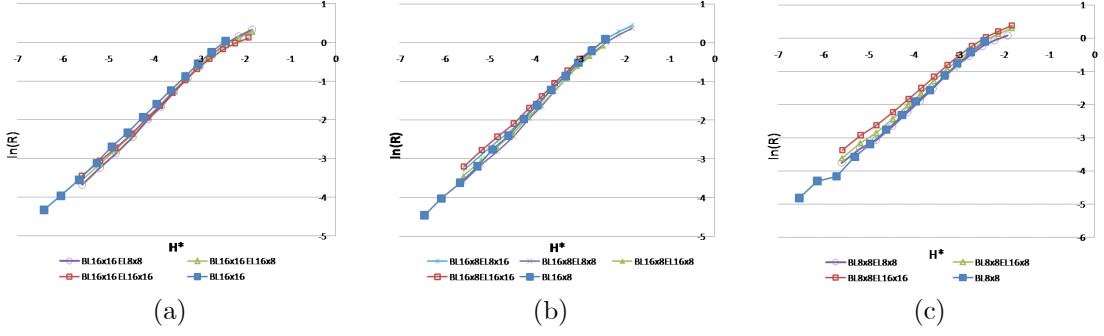


Figure 5.4: Linearity relationship between $\ln(R)$ and H^* (a) Base mode 16x16 (b) Base mode 16x8 (c) Base mode 8x8 form Foreman(CIF)

the same parameter set of distortion model, which is the regression result by the mode pair of BL 16×16 and EL 16×16 .

To compensate this inaccuracy of entropy, we exploit a relationship between the real coded rate R and entropy H . A linearity relationship between natural logarithm of real rate $\ln(R)$ and $\ln\left(\left(\frac{1}{q}\right)^{\left(1+\frac{\sqrt{2}}{\sigma}\right)} H^{\frac{\sqrt{2}}{\sigma}} \exp\left(-\frac{\sqrt{2}}{\sigma}\right)\right)$, which is represented as H^* , is observed as in Fig.5.4, *Foreman*(CIF), for example. The blue lines with solid squares represent the linearity relationship in single-layer coding, and the lines with hollow tokens indicate the linearity relationship for inter-layer prediction coding of different modes.

The other testing sequences have the similar relationship between $\ln(R)$ and H^* as Fig.5.4; we can achieve a modified rate model, which is represented as R^* and is described before in Sec.. The constants a and b in (3.15) are the coefficients of the approximate line function of $\ln(R)$ and H^* , and the constants vary according to mode pair and coded sequence. We provide some example curves of modified rate model compared with real rate curves in Fig.???. Contrast the modified rate with entropy H , it can be observed that the rate modification has a great improvement in fitting the real rate curves across different sequences. For comprehensive analysis, the results of all testing sequences in different QP differences appear in Table 5.5; the enormous errors of entropy H is corrected by the proposed rate model R^* throughout the difference sequences.

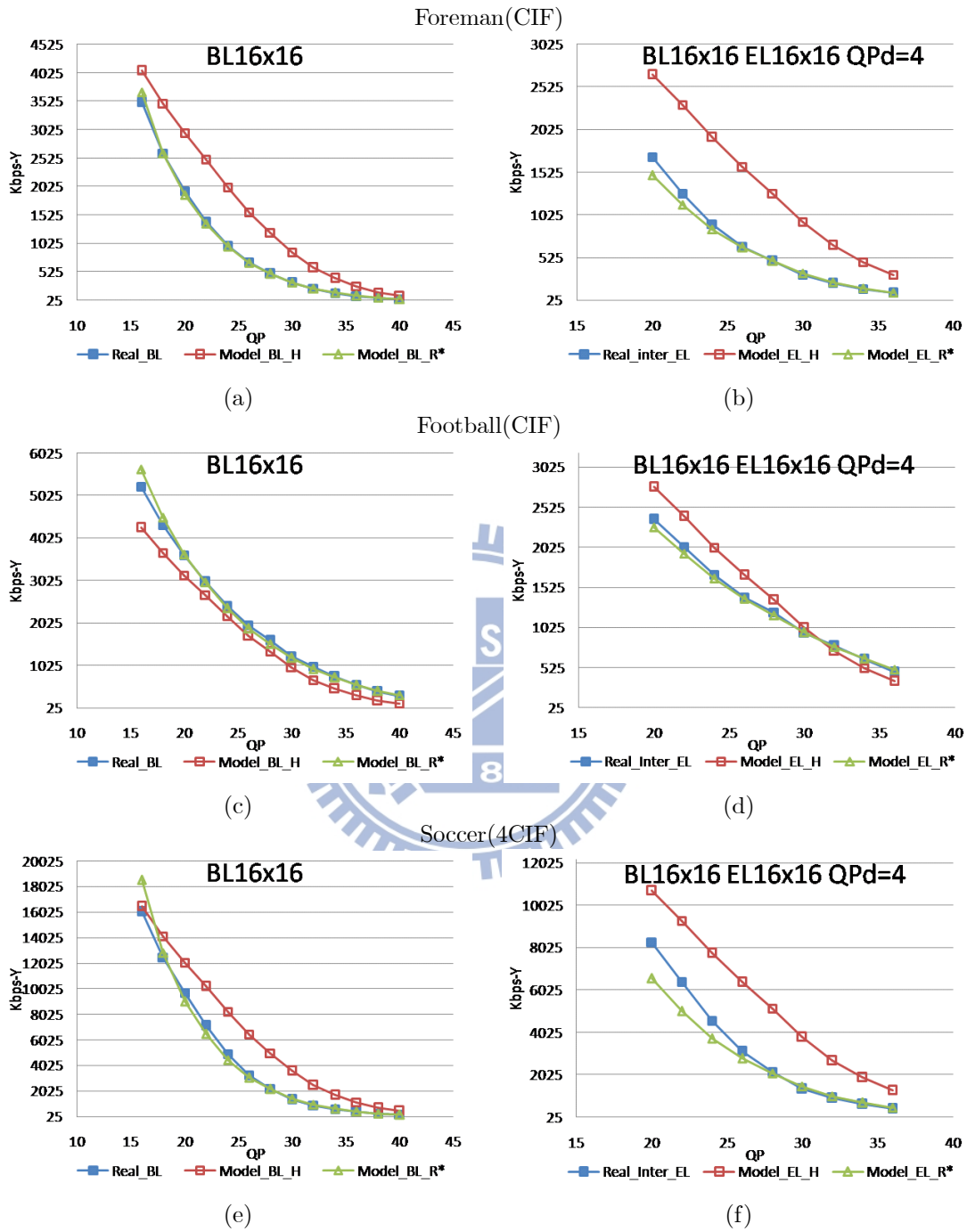


Figure 5.5: Modified rate(R^*) compared with actual rate, entropy(H) as a contrast. Blue lines with solid squares are BL R-Q curve. Red lines with hollow squares are the curves for entropy v.s. QP. Green line with hollow triangles are the curves for modified rate v.s. QP

Resolution	Sequence	Rate error							
		BL(%)				EL(%)			
				Qpd=4		Qpd=6		Qpd=8	
		H	R^*	H	R^*	H	R^*	H	R^*
CIF	Bus	29.4	7.1	40.1	11.0	28.9	11.8	20.2	12.3
	Football	26.2	3.8	11.5	3.9	10.1	5.0	10.2	6.0
	Foreman	122.4	4.3	124.4	7.4	100.3	10.6	80.4	13.4
	Mobile	28.1	20.9	28.2	17.3	16.6	15.0	9.6	13.2
4CIF	City	433.4	42.2	149.7	29.8	102.2	32.1	67.4	31.3
	Crew	24.5	11.5	42.2	9.9	32.4	13.2	26.4	16.2
	Harbour	64.3	12.9	57.4	18.0	38.6	19.2	26.3	19.0
	Soccer	108.7	7.1	108.7	14.1	82.3	17.6	61.0	19.7
Average		104.6	13.7	70.3	13.9	51.4	15.6	37.7	16.4

Table 5.3: Entropy rate and modified rate error.

5.2 SVC R-D Curve Performance

The performance of SVC with inter-layer residual prediction is demonstrated in Fig.5.6, *Bus*(CIF) for example. The simulcast curve expresses the enhancement layer coding without residual prediction, in other words, it is a single layer coding applied lower QP. The following experiments use "base-layer usage" to measure the benefits of inter-layer residual prediction; the base-layer usage (BLU), which is an alternative measure of the utilization of base-layer information, is calculated by

$$BLU = \frac{R_{simu} - R_{inter}}{R_{simu} - R_{base}}$$

where R_{base} , R_{inter} , and R_{simu} denote the kbit-rates of base-layer, inter-layer prediction, and simulcast, respectively. It should be noted that a base layer usage of 0% corresponds to the coding efficiency of simulcast, while BLU less than 0% represents the inter-layer prediction is inefficient than simulcast. Besides, R_{inter} is the rate required for obtaining the same PSNR of simulcast; this rate is calculated by interpolating the inter-layer rate-distortion curve for the simulcast resolution using linear interpolation.

Table 5.5-5.6 display the comparison of base-layer usage between real and model, *Bus*(CIF) *Football*(CIF), and *Foreman*(CIF) as testing sequence. Keeping realistic, we compute BLU with BL QP equaling 28 for each testing condition. Despite numerically inaccurate fitting between real and model, they still show some similar tendencies. First, it can be observed that when BL has the same mode as EL, inter-layer prediction attains the best performance. Second, viewing the change of QP difference, we can see

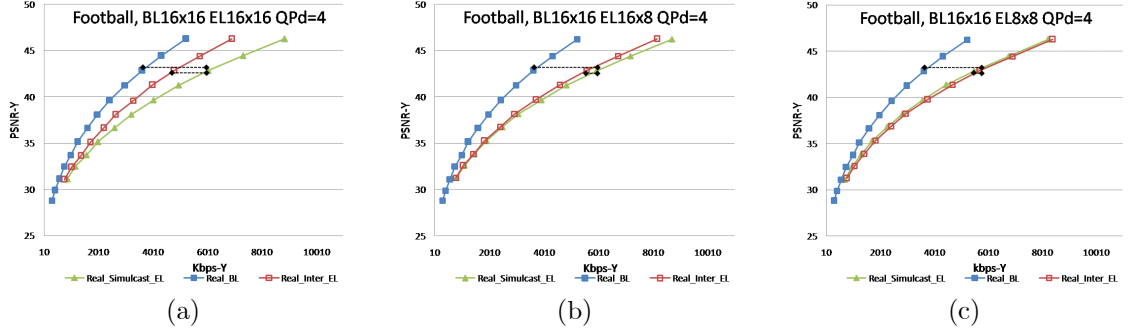


Figure 5.6: SVC performance (Football), the dotted lines indicate the distance between simulcast and base layer, as well as the distance between simulcast and inter-layer residual prediction

Mode pair		<i>Bus</i> (CIF) BLU(%)					
		Qpd=4		Qpd=6		Qpd=8	
BL	EL	Real	Model	Real	Model	Real	Model
16 × 16	16 × 16	24.9	31.0	47.9	29.7	15.7	30.0
	16 × 8	9.6	16.5	8.4	17.7	5.7	19.8
	8 × 8	-7.3	-1.0	-5.7	5.3	-5.3	11.6
16 × 8	16 × 16	-2.7	3.1	-1.9	7.1	-2.3	11.5
	16 × 8	22.0	29.2	18.1	27.1	13.3	26.9
	8 × 16	-7.4	1.6	-5.5	5.2	-5.3	9.3
8 × 8	8 × 8	-1.5	3.9	-1.7	8.0	-2.1	12.8
	16 × 16	-2.1	4.4	-1.5	8.7	-1.8	13.2
	16 × 8	7.3	14.9	6.6	16.4	4.3	18.8
	8 × 8	19.3	25.9	14.7	26.0	11.3	27.7

Table 5.4: Base-layer usage with base-layer QP equaling 28., *Bus*(CIF).

the inter-layer prediction performance of most mode pair is decreasing as QP difference grows.

Mode pair		<i>Football(CIF)</i> BLU(%)					
		Qpd=4		Qpd=6		Qpd=8	
BL	EL	Real	Model	Real	Model	Real	Model
16×16	16×16	30.1	30.6	49.8	28.3	20.7	26.7
	16×8	9.3	7.5	7.6	10.1	7.3	12.2
	8×8	-8.8	-10.5	-8.0	-4.3	-7.0	0.4
16×8	16×16	-0.5	1.1	-0.1	4.2	-0.4	6.8
	16×8	27.9	27.6	22.5	36.4	19.2	24.8
	8×16	-7.0	-3.8	-5.7	0.9	-5.5	4.7
	8×8	-1.9	-1.6	-3.0	2.1	-3.0	5.1
8×8	16×16	-3.1	-0.9	-2.7	2.7	-2.37	5.7
	16×8	5.8	6.1	4.5	9.0	4.7	11.4
	8×8	25.1	24.8	18.8	23.2	14.67	22.1

Table 5.5: Base-layer usage with base-layer QP equaling 28., Football(CIF).

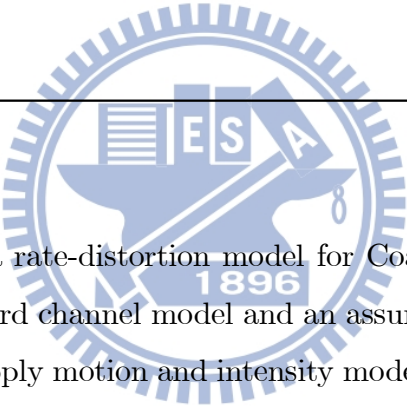


Mode pair		<i>Foreman(CIF)</i> BLU(%)					
		Qpd=4		Qpd=6		Qpd=8	
BL	EL	Real	Model	Real	Model	Real	Model
16×16	16×16	6.8	9.9	43.3	12.9	2.3	15.8
	16×8	-5.4	-6.0	-6.3	1.8	-5.3	7.4
	8×8	-30.9	-26.7	-19.9	-14.3	-16.6	-5.4
16×8	16×16	-9.0	-7.4	-6.3	0.7	-5.6	6.6
	16×8	5.3	4.3	1.1	8.7	0.1	12.8
	8×16	-20.4	-15.6	-13.5	-4.8	-11.3	3.0
	8×8	-21.7	-18.2	-13.7	-8.6	-12.1	-1.3
8×8	16×16	-7.4	-7.3	-5.1	0.9	-4.6	6.8
	16×8	-1.6	-3.0	-3.0	4.0	-2.7	9.2
	8×8	-4.9	-2.2	-2.4	2.8	-3.7	7.4

Table 5.6: Base-layer usage with base-layer QP equaling 28., Foreman(CIF).

CHAPTER 6

Conclusions

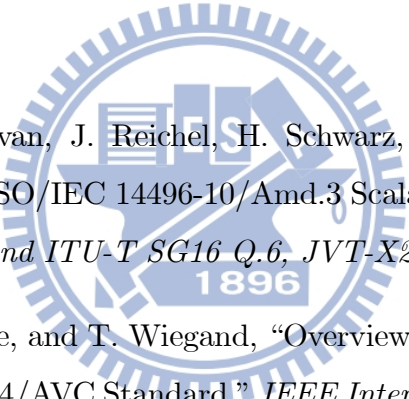


In this thesis, we derive a rate-distortion model for Coarse-Grain Quality Scalability by construction of a forward channel model and an assumption of temporal-stationary process. Afterwards we apply motion and intensity models to interpret the autocovariance factors in proposed model. These models provide parameters of motion, intensity, and block-partition mode to analyze the block-level motion-compensation predictor. Through the proposed R-D model for SVC, we estimate the R-D curves of inter-layer residual with different mode pairs, different sequences, and different QP differences. Experiments shows a high accuracy of distortion error less than 0.5dB; a great correctness of rate model by an entropy modification function is also demonstrated.

The proposed R-D model for SVC will be extended to several applications:

1. To propose a rate-distortion function for optimal bit allocation of SVC.
2. To find the Lagrange multiplier for R-D optimization in SVC with Multi-layer encoder control.
3. To give a theoretical analysis for fast mode decision in SVC.

Bibliography

- 
- [1] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, “Joint Draft ITU-T Rec. H.264 | ISO/IEC 14496-10/Amd.3 Scalable Video Coding,” *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-X201*, July 2007.
- [2] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” *IEEE International Conference on Image Processing (ICIP)*, October 2006.
- [3] D. S. Turaga, Y. Chen, and J. Caviedes, “No reference PSNR estimation for compressed pictures,” *Signal Process. Image Commun.*, vol. 19, pp. 173–184, September 2004.
- [4] L. Guo, O. C. Au, M. Ma, Z. Liang, and P. H. W. Wong, “A Novel Analytic Quantization-Distortion Model for Hybrid Video Coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 627–641, May 2009.
- [5] J. Liu, Y. Cho, Z. Guo, and C.-C. J. Kuo, “Bit Allocation for Spatial Scalability Coding of H.264/SVC With Dependent Rate-Distortion Analysis,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 967–641, July 2010.

- [6] R. Zhang and M. L. Comer, "Rate Distortion Analysis for Spatially Scalable Video Coding," *IEEE Trans. on Image Processing*, vol. 19, pp. 2947–2957, November 2010.
- [7] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, August 1987.
- [8] H. C. Huang, W. H. Peng, T. Chiang, and H. M. Hang, "Advances in the Scalable Amendment of H.264/SVC," *IEEE Communications Magazine*, vol. 45, pp. 68 – 76, 2007.
- [9] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 193–205, February 2009.
- [10] A. Eden, "No-Reference Estimation of the Coding PSNR for H.264-Coded Sequences," *IEEE Trans. on Consumer Electronics*, vol. 53, pp. 667–674, May 2007.
- [11] B. Tao and M. Orchard, "A Parametric Solution for Optimal Overlapped Block Motion Compensation," *IEEE Trans. on Image Processing*, vol. 10, pp. 341–350, Mar. 2001.
- [12] T. Berger, *Rate Distortion Theory*. Prentice-Hall, 1971.
- [13] Y. C. Tseng, C. H. Wu, Y. W. Chen, T. W. Wang, and W. H. Peng, "On The Analysis And Design of Motion Sampling Structure for Advanced Motion-Compensated Prediction," *IEEE Int'l Conf. on Image Processing*, September 2010.
- [14] "Joint Scalable Video Model 10 (JSVM 10)," *Joint Video Team, doc. JVT-W202*, April 2007.