

一個針對可調視訊編碼中跨層編碼與位元流擷取之
位元率-失真最佳化模型
A Rate-Distortion Optimization Model for SVC Inter-layer
Encoding and Bitstream Extraction

研究生：黃雪婷

Student : Hsueh-Ting Huang

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

July 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

一個針對可調視訊編碼中跨層編碼與位元流擷取之位元率-失真最佳化模型

研究生：黃雪婷

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

可調視訊編碼標準(SVC)使觀看裝置可以使用位元流擷取機制調整其視訊接收內容。正因可調視訊編碼提供了結合空間、時間與畫質上的可調性，為不同觀看裝置擷取適當的位元流時需要經過特別考慮，不適當的選擇經常會產生粗劣的觀賞品質。在本論文中，我們提出了一個針對可調視訊編碼位元流進行位元率-失真(R-D)最佳化擷取的方法。精確地說，我們針對可調視訊編碼壓縮時的量化參數與跨層編碼相依性之設定，發展一組可適應性規則，遵循此規則所產生的良適性可調視訊編碼位元流(Well-adapted SVC bitstream)可在連續增益步驟中所擷取之可調層產生明顯較好的位元率-失真平衡。我們亦正式定義最佳化與近似最佳化擷取路徑的概念，並設計了在運算量上相當有效率的擷取路徑搜尋策略。實驗的結果展示出我們的位元率-失真最佳化的適應性設定方法與擷取策略可在不同觀賞裝置的播放畫面品質達到重大的改善。特別的是，我們的可適應性規則可保證沿著所找出的最佳擷取路徑之位元率-失真曲線具有凸狀的特性，並使得貪婪探索式擷取策略(Greedy Heuristic)可用以找出最佳化或近似最佳化之路徑，而此最簡單之搜尋策略比起暴力搜尋法(Exhaustive Search)只需要大約一半的運算複雜度。

A Rate-Distortion Optimization Model for SVC Inter-layer Encoding and Bitstream Extraction

Student : Hsueh-Ting Huang

Advisors : Wen-Hsiao Peng

Institute of Multimedia Engineering

National Chiao Tung University

ABSTRACT

The Scalable Video Coding (SVC) standard enables viewing devices to adapt their video reception using bitstream extraction. Since SVC offers spatial, temporal, and quality combined scalability, extracting proper bitstreams for different viewing devices can be a non-trivial task, and naive choices usually produce poor playback quality. In this thesis, we propose an approach for performing rate-distortion (R-D) optimal extraction of SVC bitstreams. Specifically, we developed a set of adaptation rules for setting the quantization parameters and the inter-layer dependencies among the SVC encoding layers. A well-adapted SVC bitstream thus produced manifest good R-D trade-offs when its scalable layers are extracted in successive refinement steps. We also formalized the notion of optimal and near-optimal extraction paths and devised computationally efficient strategies to search for the extraction paths. Experimental results demonstrated that our R-D optimized adaptation schemes and extraction strategies offer significant improvement in playback picture quality on various viewing devices. In particular, our adaptation rules promise R-D convexity along optimal extraction paths and permit the greedy heuristic extraction strategy to be used for discovering the optimal/near-optimal paths. This simplest strategy performs only half of the computation necessary for an exhaustive search.

誌 謝

首先我要感謝彭文孝老師與邵家健老師一直以來給予我研究上的指導，經由一次次與老師的討論，研究內容才得以逐漸趨於完整；若沒有兩位老師所給予的方向指引與協助，我將無法完成這項工作。這兩年來，老師對於研究的熱情與嚴謹態度、解決問題的思考方式，一直是我學習的典範；老師的指導與建言，經常使我獲益良多。對於老師平時給我的鼓勵與支持，這份感激之情更是筆墨難以形容。

其次，我要感謝王澤瑋學弟幫忙跑了很多的實驗與整理數據；感謝林哲民同學參與討論中提出的意見；感謝林岳進同學提供實驗中的時間差補法程式；感謝李志鴻學長、林鴻志學長、陳漪紋學長在我有疑問時為我解答；感謝所有 MAPL 實驗室裡的成員，包含：林岳進、陳敏正、Eric、陳俊吉、陳建穎、林哲永、詹家欣、王澤瑋、吳思賢，無論是研究上的討論或生活中的苦樂分享，你們是我研究生涯不可或缺的伙伴。

最後也最重要的，我要感謝長久以來給我關心與鼓勵、支援我完成碩士班學業的父母：黃慶峯先生與黃吳素真女士，若沒有您們的支持就不會有今日的我；也要感謝我的姐妹黃慧玲與黃綉雯、以及 Puppy，感謝你們一路的陪伴打氣給了我很多力量，謝謝你們！

A Rate-Distortion Optimization Model for SVC Inter-layer Encoding and Bitstream Extraction

Advisors: Prof. Wen-Hsiao Peng

Prof. John Kar-Kin Zao

Student: Hsueh-Ting Huang


Institute of Multimedia Engineering

National Chiao-Tung Univeristy

1001 Ta-Hsueh Rd., 30010 HsinChu, Taiwan

July 2008

Contents

Contents		i
List of Tables		iv
List of Figures		v
1 Research Overview		1
1.1 Introduction		1
1.2 Problem Statement		2
1.3 Contributions and Organization of Thesis		2
2 Background		4
2.1 Scalable Video Coding		4
2.1.1 Concept		4
2.1.2 Transport Interface of SVC		5
2.2 Related Works		7
2.2.1 Basic Extraction		7
2.2.2 Quality Information Table (QIT)		7
2.2.3 Quality Index (QI)		8
2.2.4 Quality Layer Optimized Extraction		9

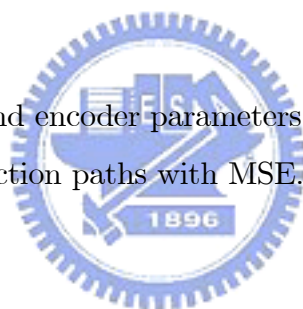
3	Rate-Distortion Optimization of SVC Bitstream Extraction	11
3.1	Extraction Paths through SVC Bitstream	11
3.1.1	Successive Refinement	12
3.1.2	Incremental and Cumulative Rate-Distortion Performance	12
3.2	Rate-Distortion Optimal Extraction Path	14
3.3	Near-optimal Extraction Paths	14
4	Searching for Optimal Extraction Paths	16
4.1	Graphical Tools	16
4.2	Search Strategy	19
4.2.1	Dynamic Programming Algorithm	19
4.2.2	Greedy Heuristic Scheme	21
4.3	Analysis of Greedy Heuristic Scheme	23
4.3.1	Convex Segments and Global Condition	23
4.3.2	Strong Local Conditions	24
4.3.3	Weak Local Conditions	25
4.3.4	Fractional Violation of Local Conditions	26
4.4	Summary	27
5	Production of Well-adapted SVC Bitstreams	28
5.1	Settings of Quantization Parameters	28
5.2	Settings of Inter-layer Dependencies	29
6	Experiments	32
6.1	Implementation of Well-adapted SVC Bitstream	32
6.1.1	Prediction of R-D Convexity	32
6.1.2	Degradation in Coding Efficiency	34
6.2	Analysis of Optimal Extraction Paths	35
6.2.1	Optimal Paths versus Video Contents	36
6.2.2	Optimal Paths versus Distortion Measures	36
6.2.3	Optimal Paths versus Spatiotemporal Interpolation	38
6.3	Performance of Greedy Heuristic Scheme	39
6.3.1	Extraction Paths and R-D Performance	39
6.3.2	Computational Complexity	40

6.4 Comparisons with Other Extraction Schemes	41
7 Conclusions	45
Bibliography	47



List of Tables

6.1	Testing conditions and encoder parameters	36
6.2	Comparison of extraction paths with MSE.	41



List of Figures

2.1	SVC dependency structure	5
2.2	Scalable layers corresponding to Figure 2.1	6
2.3	Preference path of perceptual quality [4]	8
2.4	Quality-Layer-based extraction [2]	9
3.1	Measuring components of the <i>deviation from convexity</i> of a <i>NAL cluster</i> along an SVC R-D curve	15
4.1	R-D mesh and trellis diagram of an SVC test bitstream, Akiyo (CIF30)	18
4.2	Example: using dynamic programming algorithm to find the optimal extraction path.	20
4.3	Example: using greedy heuristic scheme to find the optimal extraction path.	22
4.4	A trellis diagram with convex segments satisfying <i>strong</i> intra-trellis (local) and inter-trellis (global) R-D conditions	25
4.5	A trellis diagram with convex segments satisfying <i>weak</i> intra-trellis (local) and inter-trellis (global) R-D conditions	26
4.6	R-D mesh and trellis diagram of an SVC bitstream with fractional violation of intra-trellis (local) R-D conditions	27

5.1	R-D performance of SVC bitstreams with different inter-layer dependency settings. Labels A, B, C, D, and E denote five coding layers of different SNR levels with E being the target layer for reconstruction. . .	30
6.1	Comparison of SVC dependency settings: (a) Mobile and (b) Foreman. The results were produced with bottom-up encoding process and fixed-quality configurations.	33
6.2	Comparison of SVC dependency settings: (a) Mobile and (b) Foreman. The results were produced with bottom-up encoding process and fixed-rate configurations.	34
6.3	Comparison of total bit rate for different dependence settings. Fixed-quality (FQ) and fixed-rate (FR) configurations were used.	35
6.4	Comparison of optimal extraction paths for different viewing devices: (a) Mobile, (b) Foreman, (c) Akiyo, and (d) ICE. B.Direct and MSE are used for temporal interpolation and distortion measure, respectively. . .	37
6.5	Comparison of optimal extraction paths found by using MSE and MOS as the distortion criterion: (a) Foreman CIF@30 and (b) Mobile CIF@30.	38
6.6	Comparison of optimal extraction paths using frame replication (F.R.) and B_Direct_16x16 (B.Direct) for temporal interpolation: (a) Akiyo CIF@30 and (b) Foreman QCIF@30.	39
6.7	Comparison of extraction paths for the steepest-descent method and exhaustive search: (a) R-D trellis diagram and (b) R-D curves.	41
6.8	R-D performance comparison of the proposed scheme with the Quality Layer and Basic extractions in JSVM 9: (a) QCIF SNR Scalability, (b) QCIF/CIF Combined Scalability.	42
6.9	Bitstream extraction (a) with and (b) without successive refinement. R1-R4 indicate the extracted NAL sets associated with increasing bit rate.	43

CHAPTER 1

Research Overview



1.1 Introduction

Production of scalable bitstreams that can be played back by a garden variety of viewing devices has been a long pursued goal of video compression technology. The new scalable extension of H.264/AVC standard (referred hereafter as SVC) [12][16] promises to achieve that goal by employing *adaptive inter-layer prediction* along with *hierarchical temporal reference*. By encoding a video sequence into an inter-dependent set of *network abstraction layer (NAL) units*, SVC allows different viewing devices to extract and decode different scalable layers according to their display formats, processing power, and/or transport network throughput. However, the parts of a bitstream needed for providing good quality playback at different devices may differ significantly depending on the visual characteristics of video programs, the quantization and dependency settings of SVC encoders as well as the display formats of viewing devices. This problem has prompted an intensified study of bitstream adaptation for viewing quality optimization.

1.2 Problem Statement

While offering the flexibility for discretionary bitstream extraction, the current standard does not specify what to produce and what to use if there are several extraction possibilities. Several approaches have thus been proposed for finding optimal bitstream adaptation/extraction schemes that ensure the best playback quality on a viewing device while making the best use of available transport bandwidth. Although the extraction process can be improved by R-D optimization, the playback quality may still be far from satisfactory. This is because the pre-encoded SVC bitstreams may not be well-adapted, which could easily give rise to poor R-D performance. As a result, in this thesis we propose a novel R-D optimization model to tackle the problem from both encoder settings and extraction process. Experiments were conducted to illustrate

1. How the tuning of quantization parameters coupled with the changing of inter-layer dependencies affects the R-D performance of SVC bitstreams,
2. What criteria on SVC encoder/decoder settings may ensure the existence of optimal or near-optimal extraction paths for different viewing devices,
3. And how the optimal extraction paths of different viewing devices can be found using computationally efficient strategies especially when the SVC bitstream is to be extracted through successive refinements.

Aiming at viewing quality optimization for bitstream adaptation of different devices, this thesis provides an in-depth study on the relationship among video contents, viewing device capability and searching strategies for finding an optimal extraction path. Moreover, SVC inter-layer dependency and quantization parameter settings during SVC encoding were investigated to discover some rules to produce well-adapted SVC bitstream.

1.3 Contributions and Organization of Thesis

Specifically, our main contributions in this work include the following:

- We define the rate-distortion optimal bitstream extraction problem as a constrained optimization problem and create a R-D trellis diagram to model the bitstream extraction process.
- We employ dynamic programming algorithm and propose a fast greedy heuristic

search strategy for searching optimal extraction paths.

- We develop a set of adaptation rules for setting quantization parameters and inter-layer dependencies during SVC encoding.
- We analyze a lot of experimental results to figure out how video contents, device types, distortion measures and interpolation algorithms may affect the optimal extraction paths.

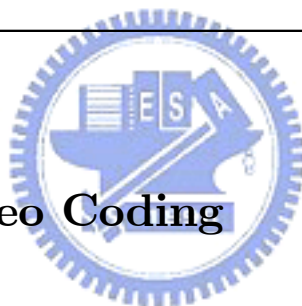
Experimental results indicate that our optimization scheme makes a significant difference in improving viewing quality. Our adaptation rules promise the R-D convexity of optimal extraction paths and enable the greedy heuristic scheme to achieve the same or similar performance as the dynamic programming algorithm while reducing the complexity by 50% or more.

The remaining of this thesis is organized as follows: Chapter 2 contains a review of SVC dependency structure and related works for finding optimal bitstream extraction schemes. Chapter 3 presents our R-D optimization model for bitstream extraction. Chapter 4 introduces and analyses our strategies for finding an optimal/near-optimal extraction path. Chapter 5 further describes the necessary criteria that must be satisfied during SVC encoding in order to guarantee the existence of optimal/near-optimal extraction paths. Chapter 6 addresses the implementation issues of establishing well-adapted inter-layer dependencies and provides a detailed analysis on the optimal extraction paths and evaluates the performance of the greedy heuristic scheme in search for the optimal path. The differences between our extraction scheme and other previous works are also compared. This thesis ends with a summary of our observations and a list of future works in the conclusion.

CHAPTER 2

Background

2.1 Scalable Video Coding



2.1.1 Concept

The scalable video coding (SVC) standard [3][12][16] is an scalable extension of the H.264/AVC standard developed by the Joint Video Team (JVT) that makes a single bitstream to provide multiple frame sizes, frame rates and quality levels while achieving a reasonable coding efficiency. A subset of SVC bitstreams can be extracted and decoded to produce a lower playback quality rather than failed to decode under some constraints of resources such as network throughput or power of devices.

SVC supports three types of scalabilities: spatial, temporal and quality scalabilities. An SVC bitstream is organized into one base layer and one or more enhancement layers in corresponding dimension if it provides certain scalability. The spatial scalability bases on multilayer coding that uses separate encoder loops for different spatial resolution layers and develops adaptive inter-layer prediction techniques to exploit correlations among the layers. For each coding layer, the temporal scalability is provided by hierarchical temporal prediction structures. Quality scalability in SVC is provided

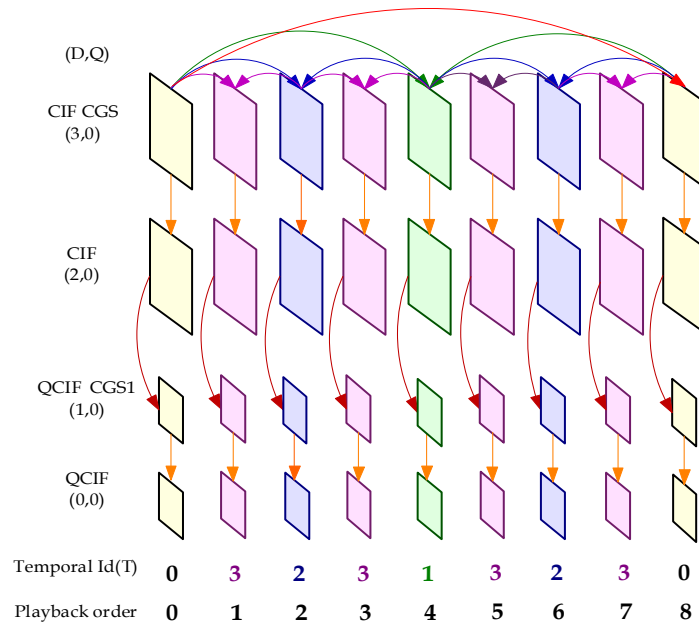


Figure 2.1: SVC dependency structure

by two approaches: *Coarse-grain quality 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 quality scalable coding* (MGS), which provides quality refinement layers inside each spatial layer and allows packet-based quality scalable coding.

Figure 2.1 depicts an example of SVC dependency structure. Each block denotes a coded picture. The horizontal order presents playback order of frames and the vertical stack appears the coding layers, as known as *dependency layers*, in spatial/CGS scalabilities. The arrows present the dependency relations due to coding prediction structures. Every dependency layer may choose one of lower layers as reference layer for inter-layer prediction. To decode correctly, all of lower layers which target layer directly or indirectly depends on for reference should appear while bitstream decoding.

2.1.2 Transport Interface of SVC

The coded video data and other side information in SVC bitstreams are encapsulated as *network abstraction layer* (NAL) units. The NAL unit consists of a header followed by payload data. The SVC NAL header consists of one-byte H.264/AVC header and three-byte extended SVC header. The extended header includes syntax elements *de-*

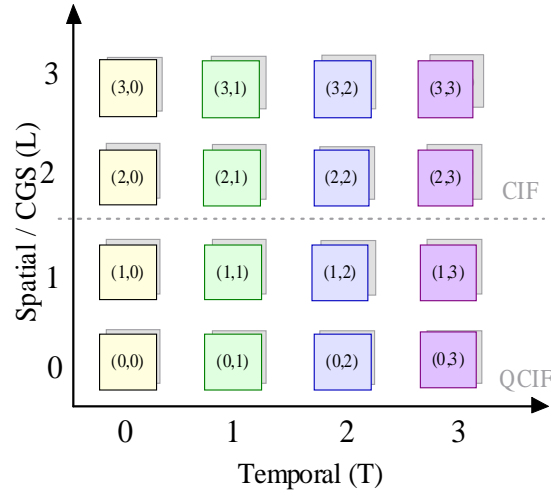


Figure 2.2: Scalable layers corresponding to Figure 2.1

dependency_id (D), *temporal_id* (T) and *quality_id* (Q), which denote the identifier of dependency layers, temporal layers and quality refinement layers respectively, as well as other assisting information to support easy bitstream extraction. Another important syntax element is the priority identifier *priority_id*, which can be used to signal the importance of NAL unit.

The sets of NAL units with identical D , T and Q information are organized into *scalable layers*. Here, the dependency and quality identifiers are combined as coding layer identifier L . As shown in Figure 2.2, the NAL units in the SVC bitstream which is depicted in Figure 2.1 can be grouped into scalable layers using coding layer identifier L and temporal identifier T . A set of scalable layers which are required for decoding certain corresponding *scalable layer* is known as *scalable layer representation* and defined as $S(L, T)$ in this thesis. For instance, $S(3, 2)$ includes all scalable layers with identifiers $L \leq 3$ and $T \leq 2$ in Figure 2.2.

SVC also designs *Scalability information Supplemental Enhancement Information* (SSEI) messages to carry the scalable layers information of bitstream such as spatial resolution, bit rate and priority information of layers for assisting bitstream adaptation processes.

2.2 Related Works

2.2.1 Basic Extraction

Currently, the Joint Scalable Video Model (JSVM) [11][15] provides three different ways to perform bitstream extraction. The first one is to extract a substream according to a bit rate constraint. The scalable layer representation thus extracted will have a bit rate that is closest to but not greater than the target bit rate. The second one is to choose a target scalable layer. The extractor will return the layer representations on which the target layer directly or indirectly depends. The last one is to explicitly specify the desired frame rate, frame size, and bit rate. However, the current standard does not specify what to produce if there are several extraction possibilities.

In following subsection, we reviewed some approaches that have been proposed for finding optimal bitstream extraction schemes.

2.2.2 Quality Information Table (QIT)

Kim *et al.* [4] evaluated the perceptual preference for spatial and temporal quality over a range of bit rates to find preference paths of perceptual quality for bitstream extraction. The spatiotemporal switching points were recorded using Quality Information Tables (QIT), which were further provided to the extractor.

The main idea is to figure out the optimal bit rate allocation strategy for three scalabilities of SVC according to video classes. First of all, video segments are classified and represented using semantic concepts. Then, quality preference paths between multidimensional scalabilities of different semantic concepts are determined by subjective testing while bit rate decreasing. For example, Figure 2.3[4] shows the preference paths of scenery and active concepts in three-dimensional scalability. The quality preference path of each video class is recorded in quality information table, which contains scalable layers information and relative bit rate of every switching point. After all, the QITs are provided to extractor for bitstream adaptation.

This approach can find quality preference paths of perceptual quality for different video classes. However, the display formats of target devices are not considered. Furthermore, subjective testing is time consuming and hardly performed for all video sequences.

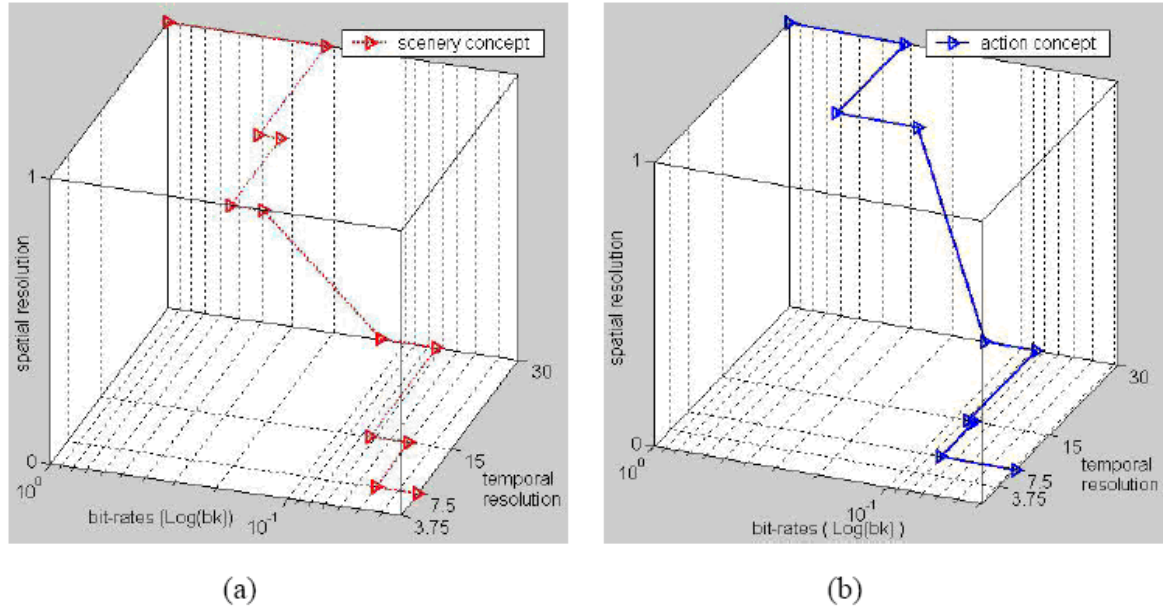


Figure 2.3: Preference path of perceptual quality [4] : (a) scenery concept, (b) action concept

2.2.3 Quality Index (QI)

Unlike QIT used subjective testing as measurement, Lim *et al.*[7] defined an objective Quality Index (QI) to measure the perceptual quality and performed bitstream extraction by maximizing the quality index of the resulting bitstream subject to the bit rate constraint. The total QI is composed of weighted quality indexes of spatial, temporal and quality scalabilities (denote as QI_{SR} , QI_{FR} and QI_{PSNR} , respectively) of extracted bitstream. Among them, quality indexes for spatial scalability QI_{SR} and quality scalability QI_{PSNR} can be measured by PSNR value. While measuring QI_{SR} , video segments are interpolated first to matching the playback format of target devices. Quality index for temporal scalability QI_{FR} , on the other hand, employs an expo-logarithm function [5] as model to estimate subjective perceptual quality MOS.

This scheme measures QI of every scalable layer representations that can be extracted subject to the bit rate constraint and chooses the one that has maximum total QI value. It obtains the sub-stream with best viewing quality measured by QI given any bit rate. But, the arbitrary extracted scalable layers at different bit rates may not support multiple adaptation of single extracted bitstream, which is an important feature in some network applications such as video multicasting.

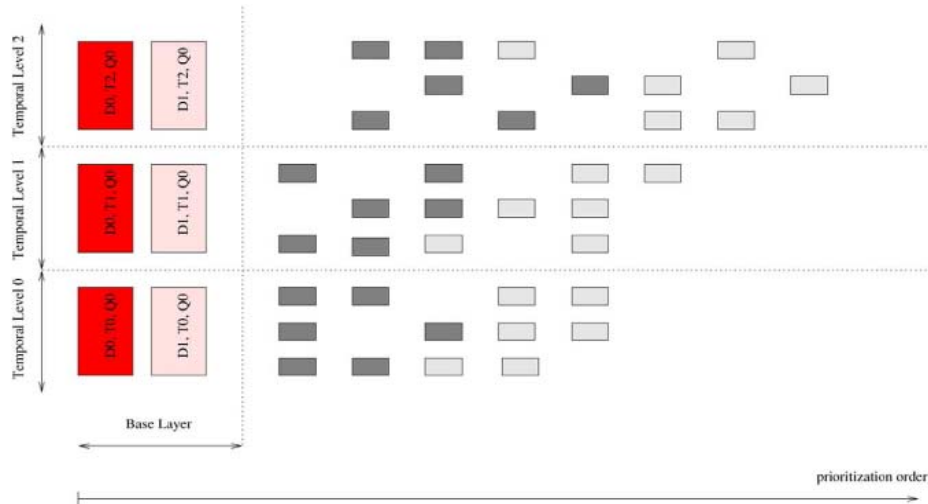


Figure 2.4: Quality-Layer-based extraction [2]

2.2.4 Quality Layer Optimized Extraction

Amonou *et al.* [2] formulated the problem as a rate-distortion (R-D) optimization process and shuffled the quality increments in an R-D sense for MGS/FGS enhancement layers. The idea is similar to Quality Layers in JPEG 2000 [14].

Priorities are assigned to NAL units in SVC bitstream to represent virtual layered organization of stream for further bitstream adaptation. First of all, R-D information is calculated for quality increment of each picture at each quality refinement level using independent or dependent distortion calculation. In dependent distortion calculation, the distortion of a picture and the distortion of pictures which were predicted from it are all considered. Namely, the impact on total rate and on the global reconstruction quality of each quality increment is computed to measure its R-D performance (slope). Based on the R-D information, the quality increments are sorted while the constraints of temporal prediction dependency are respected. Finally, Quality Layers are assigned to the quality increments according to the sorting results and stored in NAL header using *priority_id* field or in SEI messages.

The Quality Layer optimized extraction can even apply to multiresolution bitstream. Figure 2.4 [2] illustrates Quality-Layer-based extraction. Each big block represents a scalable layer referred to as (D_d, T_t, Q_q) where D_d indicates the spatial resolution, T_t for temporal layer and Q_q for the quality level. The small blocks represent the NAL units of quality enhancement layer in different spatial resolution: dark-gray

blocks for D_0 and gray ones for D_1 . The blocks are ordered according to their Quality Layer information rather than quality levels. Therefore, NAL units with lower R-D performance will be dropped first when bitstream extraction happened.

Quality Layer assignment makes quality increments are well prioritized, which insures a simple parsing of the stream that can be performed in network transmission. Nevertheless, the trade-off between spatial and temporal scalabilities is not considered in this approach.

In summary, all of prior studies were designed to determine the bitstream extraction order through different optimization schemes except the Basic Extraction approach. Between them, the Quality-Layers-based extraction is the only one approach that can produce extracted sub-streams which can support multiple adaptations. Moreover, they all can be treated as post-processing of pre-encoded bitstreams. No suggestions for proper parameter settings during SVC encoding have been proposed for benefiting the bitstream extraction.



CHAPTER 3

Rate-Distortion Optimization of SVC Bitstream Extraction



Our investigation began with an attempt to devise strategies for finding an *optimal extraction path* of an SVC bitstream for a viewing device. The extraction path should be amenable to successive refinement of the SVC bitstream for supporting multiple adaptations. In this chapter, we describe the notion of successive refinement of optimal extraction paths and define the R-D optimization of SVC bitstream extraction problem as a constrained optimization problem. We further introduce a *R-D trellis diagram* to model the bitstream extraction process. Based on R-D trellis diagrams, we can employ dynamic programming algorithm to find the solution, and furthermore propose a greedy heuristic scheme to achieve the same or similar performance while reducing the complexity significantly.

3.1 Extraction Paths through SVC Bitstream

While playing back an SVC bitstream, a viewing device may choose to extract and decode various sets of scalable layers (with possible use of error concealment) based

on its display format, decoding capability and network throughput. A sequence of these scalable layer sets arranged from lowest scalable layer (referred to as *base unit*, $S(\bar{L}, \bar{T})$) to the target scalable layer (referred to as *target layer*, $S(\hat{L}, \hat{T})$) according to their dependence relations is known as an *extraction path* Π_{φ} for the viewing device. The subscript φ indicates a denotation of the extraction path.

3.1.1 Successive Refinement

Beside of satisfying the dependence relations, one may want to fulfill some additional criteria while choosing the extraction paths for one or more viewing devices:

1. One may want to feed a viewing device with scalable layer representations of lower bit rates when the network throughput deteriorates. Such an act of bit-rate adaptation enables a viewing device to support graceful degradation of playback quality.
2. One may want to perform *successive extraction* en-route a multicasting tree. Significant reduction of transport bandwidth can be achieved by having an up-stream provider extracts only the scalable layers needed by its down-stream subscribers. Careful selection of extraction paths for different down-stream subscribers may minimize the bandwidth consumption of a multicasting session [9].

The two criteria of *successive refinement* of SVC bitstream imply that every element along the extraction path must have the previous element being its proper subset [Figure 6.9 (a)] for supporting multiple adaptations.

3.1.2 Incremental and Cumulative Rate-Distortion Performance

Several extraction paths are available for traversing an SVC bitstream between the base unit and a target layer. These extraction paths are differentiated by their *rate-distortion (R-D) performance*, which measures the *effectiveness* that an extracted bitstream uses their data bits to enhance the quality of their playback pictures. The R-D performance of an SVC bitstream can be quantified in two ways using: (1) a ratio between the increase in bit rate and the decrease in playback distortion at every refinement step and (2) the area underneath the R-D curve that spans the refinement steps. The two

measurements are defined below and used in Chapter 4.

The first (incremental) measurement of R-D performance evaluates the *R-D improvement*¹ Γ incurred through successive refinement²:

$$\Gamma(L, T; L'', T'') \triangleq -\frac{d(L'', T'') - d(L, T)}{r(L'', T'') - r(L, T)} \quad (3.1)$$

where $d(L, T)$ is the distortion value and $r(L, T)$ is the total bit rate of $S(L, T)$. Note that R-D improvement is path independent because each $S(L, T)$ has unique r, d values.

We further define the *local R-D improvements* γ of a single refinement step in either L or T dimensions as

$$\gamma_L(L, T) \triangleq -\frac{d(L', T) - d(L, T)}{r(L', T) - r(L, T)} \quad (3.2a)$$

$$\gamma_T(L, T) \triangleq -\frac{d(L, T') - d(L, T)}{r(L, T') - r(L, T)} \quad (3.2b)$$

where L' and T' denote the subsequent spatial or quality and temporal layers reached through a single refinement step. Note that these local R-D improvements are uniquely identified by their *reference identifiers* (L, T) . We also define the R-D improvement Γ' of two successive refinement steps (one in each of L and T dimensions):

$$\Gamma'(L, T) \triangleq \Gamma(L, T; L', T') = -\frac{d(L', T') - d(L, T)}{r(L', T') - r(L, T)} \quad (3.3)$$

This is the R-D improvement incurred during the traversal of a four-node trellis in the grid of $S(L, T)$ [Section 4.1]. These traversals play a pivotal role in our proposed strategies to search for an optimal extraction path.

The second (cumulative) measurement of R-D performance is the *underlying area* $\Omega_\varphi(\bar{L}, \bar{T}; \hat{L}, \hat{T})$ of an R-D curve corresponding to an extraction path $\Pi_\varphi(\bar{L}, \bar{T}; \hat{L}, \hat{T})$. Unlike R-D improvement, Ω_φ depend on the chosen extraction path φ . Also, rather than measuring the rate of R-D improvement in a single refinement step, Ω_φ measures the *efficiency* of an SVC bitstream in using its data bits to enhance its play-

¹The *negation* of the slope is used to ensure that a *positive* value reflects an improvement in playback picture quality.

²In the definitions of *R-D improvements* and the equations hereafter, we use indices L', T' to denote the scalable layer representations that it can be reached through a *single refinement step* in L or T dimensions from the reference representation $S(L, T)$ and use L'', T'' to denote that it can be reached through *multiple refinement steps*.

back quality through a series of refinement steps along the path φ . Furthermore, $\Omega_\varphi(\bar{L}, \bar{T}; \hat{L}, \hat{T}) / (r(\hat{L}, \hat{T}) - r(\bar{L}, \bar{T}))$ can be interpreted as the *average playback quality* along the extraction path.

3.2 Rate-Distortion Optimal Extraction Path

When we select an extraction path across an SVC bitstream for a specific viewing device, we intend to choose the *optimal extraction path* that offers the viewing device with best *rate-distortion (R-D) performance* as prescribed by the following criteria.

Criterion 1 *Minimum Underlying Area for Corresponding R-D (MSE) Curve*³. The optimal extraction path Π_φ produced by successive refinement should be the one that has *minimum total underlying area* Ω_φ for the corresponding R-D curve if *mean square errors*⁴ (MSE) are used to measure the playback distortion of the extracted bitstream.

Criterion 2 *Convexity of Corresponding R-D (MSE) Curve*. The optimal extraction path Π_φ produced by successive refinement should have the corresponding R-D curve maintains its *convexity*⁵ at every refinement step. More precisely, the R-D curve should have *monotonically decreasing MSE values and R-D improvement* γ at every step.

Note that among the two criteria, the first one is used as the optimization criterion, which means the optimal path should have best average R-D performance over a bit rate range, while the second one serves as a constraint to ensure that the optimal extraction path has good properties in bitstream adaptation. Namely, it should produce maximal quality improvement within least bit rate increasing or minimal quality degradation within largest bit rate reduction.

3.3 Near-optimal Extraction Paths

In our experiments, we discovered in some rare cases (especially when subjective measures such as *mean opinion scores* are used to quantify playback picture quality), some

³In the cases that the *peak signal-to-noise ratios (PSNR)* are used as the measurement of playback distortion, the *minimum/maximum* conditions of the criteria must be reversed.

⁴The uncompressed videos that match the display format of target devices are used as the references for MSE computation. Also, we interpolate each intermediate representation to the same format before measuring its MSE.

⁵A R-D curve with distortion measured in terms of mean square errors (MSE) is *convex* or *concave upward* if and only if its *epigraph* (the sets of points lying on or above the curve) is a *convex set*.

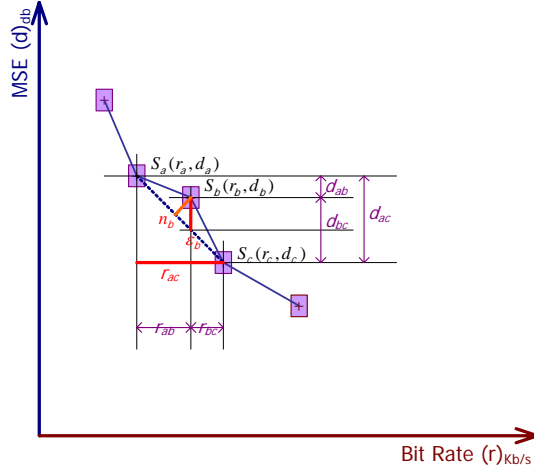


Figure 3.1: Measuring components of the *deviation from convexity* of a *NAL cluster* along an SVC R-D curve

extraction paths with slightly non-convex R-D curves may have better performance than the ones with convex R-D curves. In those cases, we should choose a *near-optimal extraction path* that has the smallest area underneath its R-D curve while the deviation from convexity of the R-D curve falls below a tolerance limit.

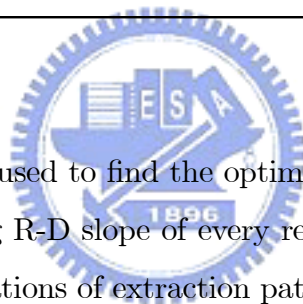
Criterion 3 *Tolerance Limit for Deviation from Convexity.* An SVC extraction path can be considered as *near optimal* if and only if the *deviation from convexity* ζ of its R-D curve at any refinement step (as defined by the following formula) lies within a specified *tolerance limit* and the *total underlying area of its R-D curve* is *minimum* among all the satisfying paths.

$$\zeta(S_b) \triangleq \frac{\epsilon_b}{r_{ac}} = \frac{r_{ac}d_{ab} - r_{ab}d_{ac}}{r_{ac}^2} \quad (3.4)$$

Figure 3.1 illustrates the quantities appeared in Equation 3.4 and offers a physical interpretation of the measurement ζ . As shown in the figure, $\zeta(S_b)$ is a ratio between the increment in MSE distortion ϵ_b and the increment in bit rate r_{ac} within a non-convex segment $[S_a, S_b, S_c]$ of an R-D curve. This ratio must be *small* in order for the deviation from convexity to be deemed *acceptable*. This is particularly true at the early refinement steps, in which the increases in bit rates are moderate while the decreases in distortion measures are step. Only minute deviation of convexity can be tolerated in those early steps.

CHAPTER 4

Searching for Optimal Extraction Paths



An exhaustive search can be used to find the optimal extraction path by decoding all scalable layers and measuring R-D slope of every refinement step. While R-D performance of all possible combinations of extraction paths are computed, it is easy to find out which one not only maintains its convexity but also has minimal underlying area for the corresponding R-D curve by exhaustively comparing. However, the decoding process of all $S(L, T)$ embedded in an SVC bitstream is extremely time consuming. Hence, the number of scalable layers required for decoding through searching processes is treated as complexity measure in this thesis. In the following paragraphs, we introduce graphical tools to model the process of successively refined bitstream extraction and then design more efficient search strategies for searching optimal extraction path.

4.1 Graphical Tools

To aid our search for the optimal extraction path of an successively refined SVC bitstream, we developed two graphical tools and named them, the *R-D mesh* and the *trellis diagram* of the bitstream. Following paragraphs explain the essence and the uses of these tools.

For the sake of examining the *R-D improvement* contributed by different refinement steps, we displayed in a single diagram all the piecewise-linear *R-D curves* of the extraction paths produced by successive refinement of an SVC bitstream. The R-D curves form a mesh, which we call the *R-D mesh* of the SVC bitstream. Every node in the R-D mesh represents a scalable layer representation $S(L, T)$ in the bitstream and is labeled explicitly by its layer L and temporal T identifiers. The coordinates (r, d) of the node represent the bit rate and the distortion of $S(L, T)$, which is decoded and interpolated to fit the display format of target devices to measure viewing quality. Every line segment in the mesh, on the other hand, corresponds to a refinement step π in either L or T dimension:

$$\pi_L(L, T) : S(L, T) \rightarrow S(L', T) \quad (4.1a)$$

$$\pi_T(L, T) : S(L, T) \rightarrow S(L, T') \quad (4.1b)$$

where L' and T' denote the subsequent spatial/CGS and temporal layers. The slope of each segment equals to the negation of the R-D improvement contributed by the corresponding refinement step.

Similarly, for the sake of exhibiting all possible extraction paths of an SVC bitstream, we superimpose them onto a grid of all scalable layer representations $S(L, T)$ embedded in the bitstream, and call the composite diagram, the *trellis diagram* of the SVC bitstream. Again, every node and edge in the trellis diagram represents a scalable layer representation and a refinement step respectively. In the trellis diagram, however, the coordinates of the nodes are their identifier values (L, T) while the edges are explicitly labeled with the R-D improvement $\gamma_L(L, T)$ and $\gamma_T(L, T)$ offered by the corresponding refinement steps. Plausible extraction paths and their segments are also drawn on top of the trellis diagram to illustrate the process of searching for the optimal path.

Figure 3.1 displays the R-D mesh and the trellis diagram of the Aikyo test sequence. Box (a) shows the *R-D mesh*; box (b) shows the *trellis diagram*, and box (c) gives a conceptual rendering of a simple trellis diagram. These tools are used in the rest of this thesis both to expound the search strategies and to interpret the experiment results.

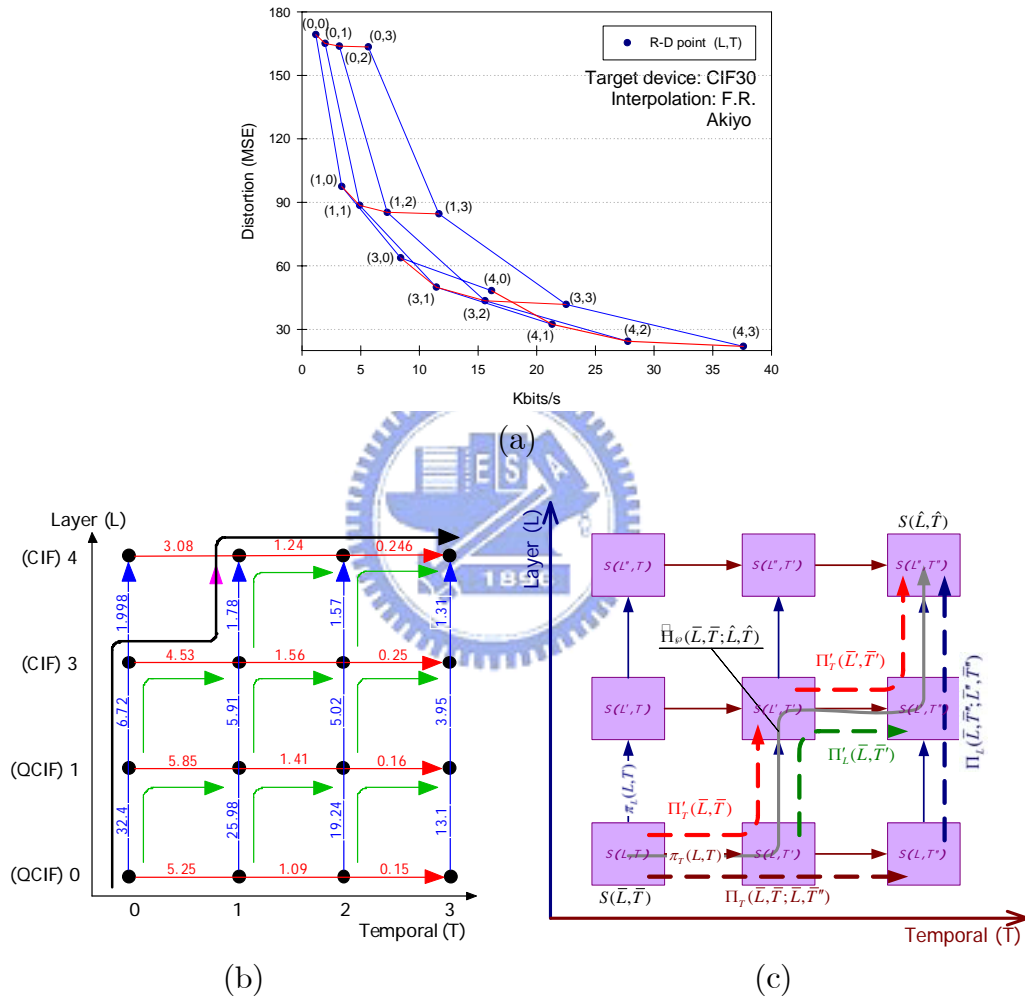


Figure 4.1: R-D mesh and trellis diagram of an SVC test bitstream, Akiyo (CIF30)

4.2 Search Strategy

4.2.1 Dynamic Programming Algorithm

Based on trellis diagrams, we can design search strategies to discover extraction paths who have maintained the convexity for corresponding R-D curve (named as *convex extraction paths*) by examining R-D performance of every refinement step from base unit to target layer. After that, the underlying area of their R-D curve can be computed and compared to find the optimal extraction path. Thinking of this process, dynamic programming algorithm, which is the most classic optimization method, is an available search strategy that can surely find the optimal extraction path if it is existent.

Utilizing dynamic programming to discover convex extraction paths is composed of two iterative phases:

1. The trellis grows in both spatial/CGS and temporal dimensions from each existent path until the paths reach target layer. The word "grow" means to decode subsequent scalable layers with one more spatial/CGS or temporal enhancement layer and evaluate the incremental R-D ratio.
2. Non-convex paths are figured out and pruned at each stage by comparing R-D performance with those of previous refinement step. Due to transitivity of inequality ($A > B \wedge B > C \Rightarrow A > B > C$), the convexity of extraction paths can be maintained even if the incremental R-D performance only compared with preceding one stage while pruning.

After all paths reaching the target scalable layer, the maintained paths are candidates of optimal extraction path. Finally, the one with the smallest total area underneath its R-D curve is the optimal extraction path.

Figure 4.2 shows an example of process using dynamic programming algorithm as search strategy to find the optimal extraction path. Each box in figure illustrates a step. In these trellis diagrams, we denote block nodes as scalable layers that have been decoded and white ones as those not have been decoded. Moreover, the refinement steps depicted as thin edges to represent that their R-D information are evaluated. If R-D curve of the extended path is convex, it would be maintained and depicted as broader edge. Otherwise, the edge would be pruned and eliminated.

The process is described below.

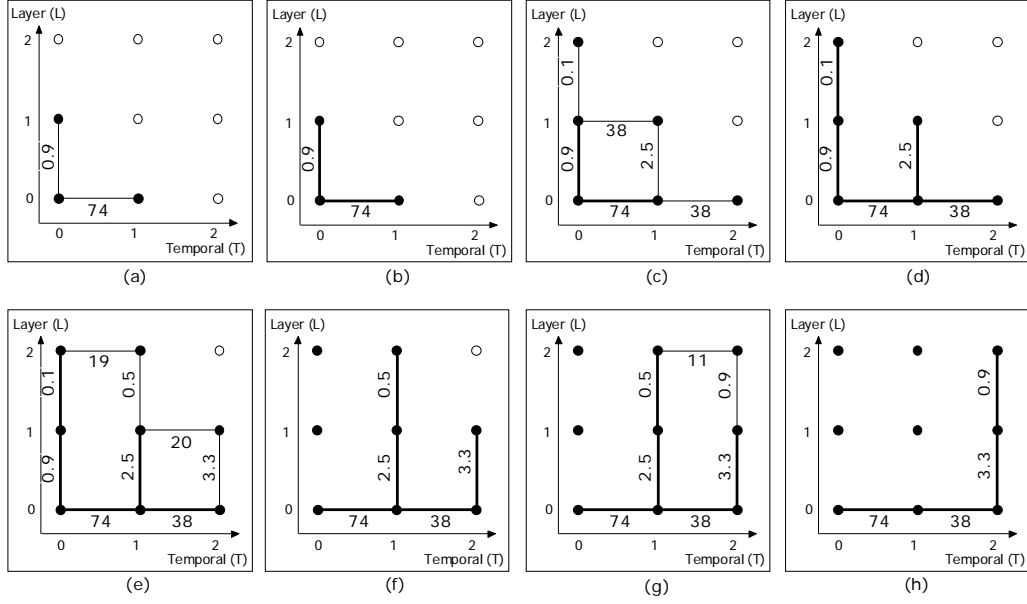


Figure 4.2: Example: using dynamic programming algorithm to find the optimal extraction path.

Step 1) Trellis starts from base unit and grows in both L and T dimensions as shown in Figure 4.2(a).

Step 2) These two paths are both kept as shown in Figure 4.2(b).

Step 3) Next iteration, trellises grow in two dimensions from existent two paths as shown in Figure 4.2(c).

Step 4) Pruned $\pi_T(1,0)$ since it can not construct a convex extraction path ($0.9 < 38$). The other three edges become broader to represent the paths are maintained as shown in Figure 4.2(d).

Step 5) Trellises grow from existent three paths as shown in Figure 4.2(e).

Step 6) Pruned $\pi_T(1,1)$ since $\gamma_L(0,1) < \gamma_T(1,1)$ (as shown in Figure 4.2(e), $2.5 < 20$) and pruned the path with $\pi_T(2,0)$ due to $\gamma_L(1,0) < \gamma_T(2,0)$. The other paths are kept as shown in Figure 4.2(f).

Step 7) Trellises grows in only L or T dimension because each of existent paths had reached their target layer in another dimension as shown in Figure 4.2(g).

Step 8) Pruned the path with $\pi_T(2,1)$ due to $\gamma_L(1,1) < \gamma_T(2,1)$. After four grew and pruned iterations, all paths reach target layer in both L and T and the process stop.

In this example, the process left only one convex extraction path in the end, which

is the optimal extraction path.

Although dynamic programming algorithm can ensure finding all convex extraction paths, its computational complexity is still considerable. It works better than exhaustive search since some paths may be pruned without evaluation and even some nodes (scalable layers) may be skipped without decoding. However, as we can see in Figure 4.2, the gain over exhaustive search is insignificant in case that the size of trellis diagrams are small because almost every nodes are required for decoding. On the contrary, if the number of scalable layers is large, the complexity of dynamic programming algorithm grows exponentially. Therefore, we need more aggressive pruning rules in a search strategy to discover convex extraction paths.

4.2.2 Greedy Heuristic Scheme

Since the efficiency of dynamic programming is not much better than exhaustive search, we propose a greedy heuristic scheme to tackle the problem. The main concept of "greedy" scheme is that every refinement step of extraction path is decided at every stage without looking ahead. This approach also consists of two iterative phases:

1. The same as dynamic programming algorithm, trellises grow in both spatial/CGS and temporal dimensions from existent paths.
2. The refinement step with worse incremental R-D improvement is pruned and *only one* path would be kept at each stage.

In other words, the greedy heuristic scheme is performed as *steepest-descent* method. While the path reaching target layer in any dimension, no more scalable layers are needed to decode and evaluate since there is only one choice for further refinement steps.

For instance, Figure 4.3 presents a process using greedy heuristic scheme as search strategy to find the optimal extraction path. The process is described below.

Step 1) Trellis starts from base unit and grows in both L and T dimensions as shown in Figure 4.3(a).

Step 2) Since $\gamma_L(0,0)$ is worse than $\gamma_T(0,0)$ (as shown in Figure 4.3(a), $0.9 < 74$), $\pi_L(0,0)$ is pruned yet only $\pi_T(0,0)$ is kept as shown in Figure 4.3(b).

Step 3) Trellis grows from the only one existent path as shown in Figure 4.3(c).

Step 4) $\pi_L(0,1)$ is pruned yet $\pi_T(0,1)$ is kept due to $\gamma_L(0,1) < \gamma_T(0,1)$. As shown

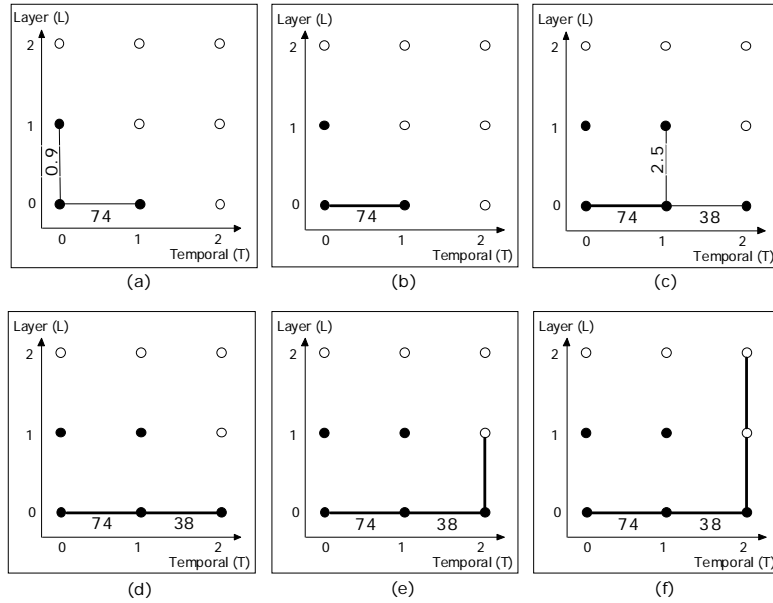


Figure 4.3: Example: using greedy heuristic scheme to find the optimal extraction path.

in Figure 4.3(d), the existent path reach target layer in temporal dimension.

Step 5) No more scalable layers are needed to decode. The refinement step in spatial dimension is included in the path as shown in Figure 4.3(e).

Step 6) Again, no more scalable layers are needed to decode. The path include the left refinement step in spatial dimension thus reach the target layer in both dimensions and end the process as shown in Figure 4.3(f).

The greedy heuristic scheme presents significant complexity reduction about 50% or more. Even number of scalable layers is small, almost half of them not have to be decoded [Figure 4.3 (f)]. Moreover, since always only one path is maintained through whole process, the complexity of greedy heuristic scheme grows linearly while the number of scalable layers increasing.

Intuitively speaking, this approach seems no guarantee of optimality of solution. Its pruning rules are designed neither with verification of R-D convexity nor with comparison of underlying area for corresponding R-D curve. However, empirical finding from our experimental results exposes that greedy heuristic scheme often can obtain the optimal extraction paths or reveal a path with comparable R-D performance. For instance, the previous example in Figure 4.3 obtained the same answer as the optimal extraction path produced by dynamic programming algorithm in Figure 4.2. The

surprisingly good performance of this scheme is theoretically analyzed in next section.

4.3 Analysis of Greedy Heuristic Scheme

We analyze the effectiveness of greedy heuristic scheme based on studying properties of trellis diagrams. Started with convex segments across single trellis and along one dimension, we discovered some satisfying conditions to construct convex extraction paths or even optimal extraction paths.

4.3.1 Convex Segments and Global Condition

All R-D convex extraction paths can be constructed from two *elementary* types of *R-D convex segments* as shown in Figure 3.1 (c):

1. *Intra-trellis (local) convex segments*, which consist of two refinement steps, one of each in L and T dimensions:

$$\Pi'_L(L, T) = \pi_L(L, T) \parallel \pi_T(L', T) \quad (4.2a)$$

$$\Pi'_T(L, T) = \pi_T(L, T) \parallel \pi_L(L, T') \quad (4.2b)$$

Each of these convex segments traverses a single four-node trellis.

2. *Inter-trellis (global) convex segments*, which also consist of two refinement steps, both of them in either L or T dimensions:

$$\Pi_L(L, T; L'', T) : \pi_L(L, T) \parallel \pi_L(L', T) \quad (4.3a)$$

$$\Pi_T(L, T; L, T'') : \pi_T(L, T) \parallel \pi_T(L, T') \quad (4.3b)$$

Each of these inter-trellis convex segments traverses two connected trellises in L or T dimensions.

The existence of intra-trellis segments Π'_L and Π'_T cannot be controlled directly by the setting of SVC encoding process. However, they can be verified by comparing the R-D improvement γ_L or γ_T of their first refinement steps $\{\pi_L, \pi_T\}$ against the R-D

improvement Γ' of the intra-trellis segments $\{\Pi'_L, \Pi'_T\}$:

$$\Pi'_L(L, T) \text{ exists iff } \gamma_L(L, T) \geq \Gamma'(L, T) \quad (4.4a)$$

$$\Pi'_T(L, T) \text{ exists iff } \gamma_T(L, T) \geq \Gamma'(L, T) \quad (4.4b)$$

The existence of inter-trellis segments Π_L and Π_T , nonetheless, can be manipulated indirectly by the setting of *quantization parameter* QP, *inter-layer dependencies* and *temporal dependencies* among the SVC coding layers. In fact, as mentioned in Chapter 5, R-D convex paths in L and T dimensions may exist at every L and T values if parameter setting satisfy certain constraints for *well-adapted SVC encoding*. The discovery of this correlation between SVC encoder setting and decoder (extraction) operation is a major contribution of this thesis. Here, since the existence of convex R-D curves in every spatial/quality and temporal layer was essential for forming convex extraction paths, we referred it as the *global condition*.

4.3.2 Strong Local Conditions

The simplest composition of trellis diagram is single four-node trellis. We looked into four-node trellises to figure out the conditions for existence of intra-trellis (local) convex segments. We defined that it is *strong local condition* satisfied if one and *only one intra-trellis convex segment* exists in every trellis. This situation arises when there is a *clear domination of R-D improvements* in either L or T dimension:

$$\text{Only } \Pi'_L(L, T) \text{ exists iff } \min(\gamma_L(L, T), \gamma_L(L, T')) > \max(\gamma_T(L, T), \gamma_T(L', T)) \quad (4.5a)$$

$$\text{Only } \Pi'_T(L, T) \text{ exists iff } \min(\gamma_T(L, T), \gamma_T(L', T)) > \max(\gamma_L(L, T), \gamma_L(L, T')) \quad (4.5b)$$

With this *strong local condition* and the *global condition*, the search for the optimal extraction path can be perfectly performed using greedy heuristic scheme (*steepest descent* method). This simple search strategy is feasible because there exists a *unique convex extraction path* between the base unit $S(\bar{L}, \bar{T})$ and any target layer $S(\hat{L}, \hat{T})$ if *both strong local and global conditions of R-D performance* are satisfied in an SVC bitstream. Figure 4.4 illustrates a typical example. Notice that the intra-trellis convex

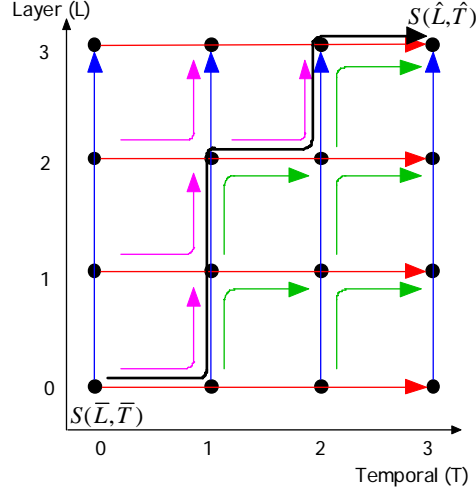


Figure 4.4: A trellis diagram with convex segments satisfying *strong* intra-trellis (local) and inter-trellis (global) R-D conditions

segments Π'_L and Π'_T tend to concentrate in two separate regions of the trellis diagram: Π'_L (drawn as magenta arrows) gathers in the upper-left corner while Π'_T (drawn as green arrows) gathers in the lower-right corner. Both types of convex segments bend their paths towards the boundary that separates the two regions. This is owing to the contradiction between global and strong local conditions. The inequalities in Equations 4.5a and 4.5b eliminate the chance for Π'_L (a magenta arrow) to appear underneath or to the right of Π'_T (a green arrow). The boundary between the two regions defines a convex and optimal extraction path (with maximum convexity and minimum underlying area) of the SVC bitstream because any other extraction path between the same end points would inevitably traverse at least one intra-trellis non-convex segment and thus yield a worse R-D performance. Hence, the traversal from $S(\bar{L}, \bar{T})$ to $S(\hat{L}, \hat{T})$ through any four-node trellis would follow the intra-trellis convex segments, which can be reduced as choosing *steepest descent* refinement steps at any steps.

4.3.3 Weak Local Conditions

Among all the R-D trellises of an SVC bitstream, some of them contain R-D convex segments but lack a clear domination of R-D performance in either L or T dimension. We named it *weak intra-trellis (local) condition* if R-D performance of the four refinement steps in four-node trellis satisfies Equations 4.4a and 4.4b but not Equations 4.5a

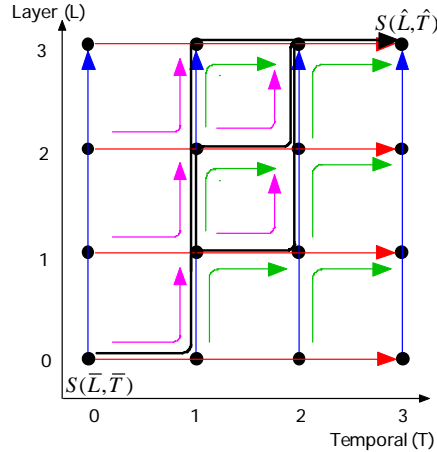


Figure 4.5: A trellis diagram with convex segments satisfying *weak* intra-trellis (local) and inter-trellis (global) R-D conditions

and 4.5b. In these cases, both Π'_L and Π'_T exist in each of these trellises. The existence of multiple convex segments in one or more trellises revokes the unique existence of convex extraction path. Hence, the greedy heuristic scheme could not promise to find the optimal extraction path. However, the difference in underlying area of two convex R-D curves in single four-node trellis is usually insignificant. Furthermore, the trellises that satisfied weak local condition almost appeared along the boundary between two regions of strong local conditional trellises empirically. Figure 4.5 shows an example of this situation. As a result, all the convex extraction paths may have similar underlying area of R-D curves and the greedy heuristic scheme can find one of them. Even though it may not be the optimal extraction path, it would have similar R-D performance.

4.3.4 Fractional Violation of Local Conditions

In some rare cases (when a subjective measures such as the mean opinion scores is used to quantify playback picture quality), the local R-D condition (i.e. the existence of intra-trellis R-D convex segments) may fail to be upheld. As a result, no convex extraction path exists between some $S(\bar{L}, \bar{T})$ and $S(\hat{L}, \hat{T})$ pairs. A *near-optimal extraction path* with a slightly non-convex R-D curve [Section 3.3] may have to be accepted as a substitute instead. In the search for the near-optimal extraction path, extraction path segments with R-D curves that contain slight deviation from convexity [Criterion 3] are included into consideration. Figure 4.6 provides an example that contains a violation

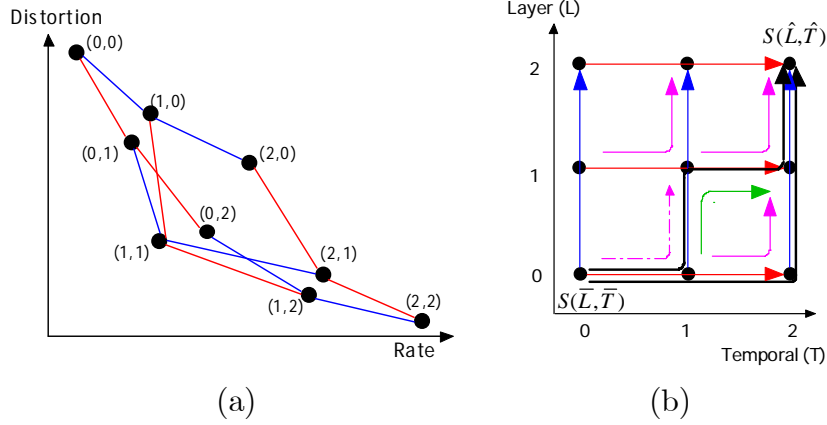


Figure 4.6: R-D mesh and trellis diagram of an SVC bitstream with fractional violation of intra-trellis (local) R-D conditions


of local R-D condition in the lower-left trellis. A slightly non-convex segment $\Pi_T'(0, 0)$ shown as a dashed magenta arrow would not be pruned during searching. In these cases, the greedy heuristic scheme generally has no promise to find the optimal/near-optimal extraction paths. However, the violation of local conditions rarely occurred.

4.4 Summary

In this chapter, we exploited dynamic programming algorithm and proposed greedy heuristic scheme for searching optimal extraction paths. We reveal the effectiveness and limit of the greedy heuristic scheme by analyzing proposed R-D trellis diagrams. The global and local conditions promise the existence of optimal extraction paths. To satisfy the global and strong local condition ensures that the optimal extraction path can be found using efficient greedy heuristic scheme. The local condition depends heavily on video contents and local R-D variations causing by measure schemes or interpolation approaches, while global condition relies on well-adapted temporal and inter-layer dependencies. In next chapter, we introduce proper settings of encoding parameters to produce well-adapted SVC bitstreams for efficient searching optimal extraction paths.

CHAPTER 5

Production of Well-adapted SVC Bitstreams



The second part of our investigation aims at establishing the necessary criteria that must be satisfied during SVC encoding in order to guarantee the existence of optimal extraction paths. Specifically, we examined the combined effects of *quantization parameter* (QP) *setting* and *inter-layer dependence relations* on the *R-D performance* of an SVC bitstream.

5.1 Settings of Quantization Parameters

One important issue in SVC encoding is to determine the QP values for spatial and quality layers so that the resulting bitstream can meet the predefined quality or bit rate constraints. While the application requirements seem to be arbitrary, it should be noted that improper QP settings may produce ill-formed R-D performance and redundant representations. To this end, we proposed two criteria for evaluating the properness of QP assignment when combined scalability is in use.

Criterion 4 *Monotonic Decrease in QP Value for Successive Refinement.* In a given spatial resolution, the QP value should *decrease* monotonically from one quality layer to the next in order to *successively refine texture information*.

Criterion 5 *Elimination of Redundant Representations.* For different spatial resolutions, the high-resolution layers should have higher fidelity than the *spatially interpolated* low-resolution layers in order to eliminate *redundant representations*.

Criterion 4 requires the picture quality to be successively refined as the size of the bitstream increases by extracting more quality layers. Criterion 5 further prohibits redundant layers from being encoded. We say that a high-resolution layer is redundant if there exists another low-resolution layer that can provide the same or even higher fidelity by spatial interpolation. Clearly, such redundancy should be detected and removed during SVC encoding.

In particular, the two criteria specify only the relative QP level among the spatial and quality layers—i.e., the exact values still need to be decided by the intended applications. For instance, by focusing our attention on mobile streaming applications, in our experiments the PSNR of spatial/quality layers is set to fall between 27dB and 35dB. Exhaustive encoding was carried out off-line to obtain the QP values for different test sequences.



5.2 Settings of Inter-layer Dependencies

In our efforts to devise efficient search strategies for optimal/near-optimal extraction paths, we discovered that the global condition can be satisfied by maintaining the *convexity of R-D curves across spatial/quality and temporal layers* during SVC encoding.

With hierarchical and dyadic temporal dependencies, the cascading QP assignment in current JSVM [11] can already make the R-D curves across temporal layers convex in most cases, especially when MSE is used for distortion measure. This is because higher temporal layers are coded with larger QP values, which inherently leads to diminishing R-D improvement with increasing temporal level.

On the other hand, among the spatial and quality layers, the convexity of their R-D curves can be guaranteed by satisfying the following criterion.

Criterion 6 *Convexity of Rate-Distortion Curves across Spatial and Quality Layers.* An SVC encoder should produce an SVC bitstream according to a *well-adapted inter-layer (spatial and quality) dependence relation* that ensures every successive refinement of scalable layer representations exhibits a *monotonic decrease in MSE dis-*

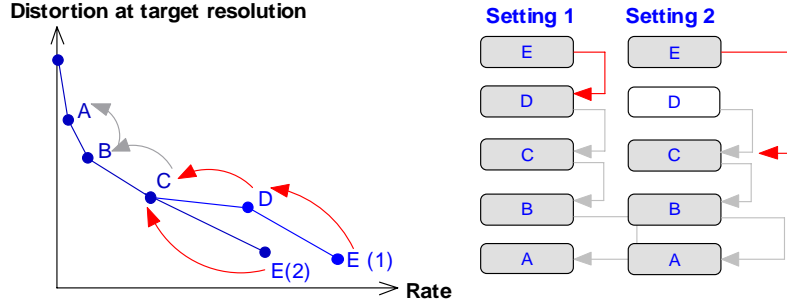


Figure 5.1: R-D performance of SVC bitstreams with different inter-layer dependency settings. Labels A, B, C, D, and E denote five coding layers of different SNR levels with E being the target layer for reconstruction.

tortion $d(L_i, \hat{T}) > d(L_{i+1}, \hat{T})$ as well as a *monotonic decrease of R-D improvement* $\gamma_L(L_i, \hat{T}) > \gamma_L(L_{i+1}, \hat{T}) > 0$.

This criterion forbids the *slope* of the R-D curves to steepen (or equivalently their *R-D improvement* to rise) as a viewing device takes in a sequence of coding layers in successive refinement steps. Its practical implication can be explained using an example shown in Figure 5.1. In the example, each layer (from B to E) in Setting #1 depends on its previous layer; hence, the reconstruction of layer E requires the decoding of all its dependent layers from A to D. However, because the R-D improvement produced by D is not as good as the one produced by E, Setting #1 cannot maintain the R-D convexity. In contrast, Setting #2, which links C directly to E by skipping D, is a well-adapted dependency setting.

We must advise readers to exercise caution when they try to set up a well-adapted inter-layer dependence relation because the adaptation can easily be overdone. In Figure 5.1, although Setting #2 (which ensures R-D convexity along the spatial/quality dimension) produces a better R-D performance for a single viewing device even if it takes layer E in one moment and layer D in another, Setting #1 (which fails to maintain R-D convexity) consumes less bandwidth when it comes to serving two viewing devices existing in the same network. This observation confirms a well-known fact that the SVC coding gain over simulcasting is at the cost of the R-D performance of individual layers. Our advice of caution can be summarized in the following proposition.

Proposition 1 *Minimal Adaptation of Successive Inter-layer Dependencies.* An SVC encoder should choose a *successive inter-layer dependence relation*, which usually pro-

duces the lowest bit rates, to be the *default dependency setting*. The dependence relation should only be modified at the refinement steps that produce *non-convex R-D improvements*. At those refinement steps, the *reference layers* should be chosen to be the *nearest spatial/quality layers* that can produce *convex R-D improvements*.

Again using the example in Figure 5.1, a proper adjustment of inter-layer dependencies is to make layer E depend on layer C rather than layer B. This minimal adjustment of inter-layer dependence relations shall only cause a small increase in the total data rate of the SVC bitstream. We would like to emphasize that such strategy is to ensure the global condition rather than to optimize the R-D performance of individual layers. For the later case, readers are referred to the paper by Yao and Li [17] for more complete discussion.



CHAPTER 6

Experiments

6.1 Implementation of Well-adapted SVC Bitstream

Having described our criteria for well-adapted bitstreams, this section further presents a practical approach for generating well-adapted inter-layer dependencies.

6.1.1 Prediction of R-D Convexity

To predict the R-D performance of SVC along the spatial/quality dimension, one effective approach is to evenly add 10% or more redundancies¹ to the R-D points of H.264/AVC [13]. The results generally hold when *multi-loop encoder control* and *fixed-quality* configurations are used [6][13]. Moreover, the predictability remains valid with *bottom-up encoding process* [11] after taking into consideration that the enhancement layers usually suffer more coding efficiency losses than the base layer. The observations enable us to predict the R-D convexity of SVC without the need of exhaustive encoding.

¹Comparing with the single layer coding, the coding efficiency loss of SVC is generally proportional to the number of coding layers. In some cases, the R-D gap between H.264/AVC and SVC can be much greater than 10%.

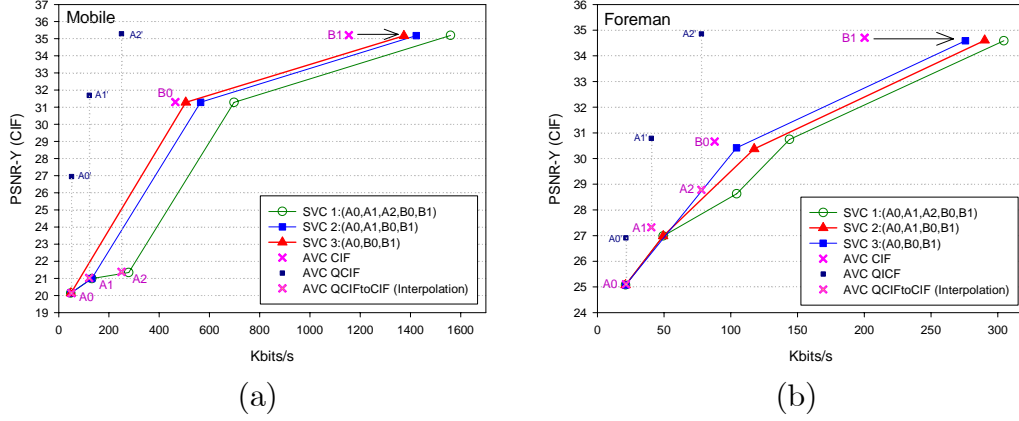


Figure 6.1: Comparison of SVC dependency settings: (a) Mobile and (b) Foreman. The results were produced with bottom-up encoding process and fixed-quality configurations.

For validation, several SVC bitstreams, each corresponds to one of the following dependency settings, were encoded using bottom-up encoder control and fixed-quality configurations. In particular, Setting #1 denotes the default dependency setting (which yields a minimal total bit rate), whereas Settings #2 and #3 adapt the default setting by merely changing the reference layer of layer B0. The R-D performances of these dependency settings are compared with that of H.264/AVC in Figure 6.1.

- Setting #1: (QCIF $A_0 \leftarrow A_1 \leftarrow A_2$), (CIF $\underline{A_2} \leftarrow B_0 \leftarrow B_1$). (Default Setting)
- Setting #2: (QCIF $A_0 \leftarrow A_1 \leftarrow A_2$), (CIF $\underline{A_1} \leftarrow B_0 \leftarrow B_1$).
- Setting #3: (QCIF $A_0 \leftarrow A_1 \leftarrow A_2$), (CIF $\underline{A_0} \leftarrow B_0 \leftarrow B_1$).

Looking at the R-D points of H.264/AVC in Figure 6.1, one can readily predict that Setting #3 would be a well-adapted setting for Mobile sequence, and the prediction was confirmed by the corresponding SVC R-D curve. Likewise, in Foreman sequence, both Settings #2 and #3 are likely to ensure R-D convexity. Although Setting #3 has better R-D performance, we choose Setting #2 because, as will be seen in the next section, the increase in total bit rate is minimized.

In Figure 6.2 we further present the results with *fixed-rate* configurations, in which the quality (and the QP) of each layer is not fixed; rather, the cumulative rate to each layer is kept constant regardless of dependency settings. Comparing with the H.264/AVC, the coding efficiency loss of SVC can be seen from the drop of R-D curves. Similar to the bit rate increase in fixed-quality configurations, the distribution of PSNR drops helps to predict the R-D convexity of SVC. From Figure 6.2, we obtain exactly

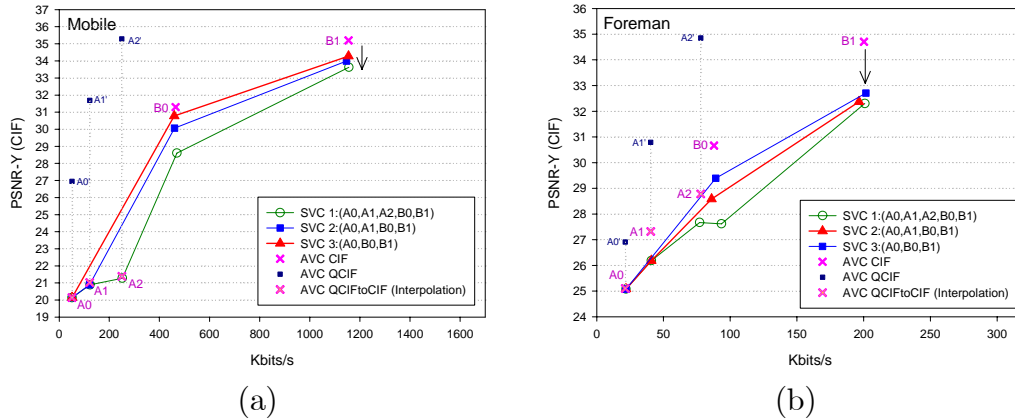


Figure 6.2: Comparison of SVC dependency settings: (a) Mobile and (b) Foreman. The results were produced with bottom-up encoding process and fixed-rate configurations.

the same dependency settings as with fixed-quality configurations. Interestingly, in Foreman sequence there is a “bump” in the R-D curve with Setting #1. This is because the QP value of layer B0 is improperly chosen to meet the bit rate constraint. The result stresses the importance of proper QP settings.

The preceding discussions assume the availability of H.264/AVC R-D points. The assumption does not generally hold unless each layer is pre-encoded with H.264/AVC. Collecting these R-D data is indeed time-consuming, but performing exhaustive SVC encoding is even worse. In addition, in our approach the R-D convexity is guaranteed *only at full frame rate*. Nevertheless, the global condition requires R-D convexity at *all possible frame rates*. We have found empirically that the convexity at full frame rate would also likely to ensure the convexity at lower frame rates. After all, the R-D behavior at full frame rate represents the average performance of all video frames.

6.1.2 Degradation in Coding Efficiency

The previous section has analyzed the SVC R-D convexity under various dependency settings. We now turn our attention to the overall coding efficiency, which is characterized by the total bit rate of an SVC bitstream. As described previously, long-term inter-layer reference may be needed for the sake of R-D convexity. It is natural then to question whether and to what extent the total bit rate will increase. The answers can be found by the comparison shown in Figure 6.3. From there it can be seen that the well-adapted dependency settings (Setting #2 for Foreman; Setting #3 for Mobile)

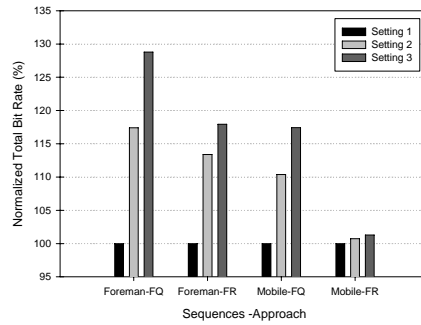


Figure 6.3: Comparison of total bit rate for different dependence settings. Fixed-quality (FQ) and fixed-rate (FR) configurations were used.

incur, on average, 15~20% bit rate increase in comparison with Setting #1 (default setting). The penalty arises mostly because layers A1 and A2 are not utilized for the inter-layer prediction of layer B0 in Settings #2 and #3.

6.2 Analysis of Optimal Extraction Paths

In this section we present a detailed analysis on the optimal extraction paths in regard to the following factors. The analysis is to understand how these factors may affect the choice of optimal extraction paths.

- Video Contents: Static vs. Motion.
- Device Types: QCIF@30/15Hz, CIF@30/15Hz, and 4CIF@30/15Hz.
- Distortion Measures: Mean Squared Error vs. Mean Opinion Score.
- Temporal Interpolations: Frame Replication (F.R.) vs. B_Direct_16x16 (B.Direct).

Table 6.1 lists our testing conditions, in which the QP assignments and the inter-layer dependence settings comply with the guidelines in Chapter 5. To simulate the actual use of SVC, extracted videos were interpolated to the highest spatiotemporal resolutions available on all viewing devices. The interpolation was accomplished by the standard-compliant spatial filtering [11], followed by frame replication (F.R.) or motion field estimation (B.Direct). While sophisticated interpolation techniques could be used, we chose the straightforward implementation because of its simplicity and popularity. In addition, in the experiments comparing subjective and objective distortion measures, we adopted the VQM software [1][10] to predict subjective quality

Table 6.1: Testing conditions and encoder parameters

Software	JSVM 9	
Spatial Scalability	QCIF (176x144), CIF (352x288), 4CIF (704x576)	
Temporal Scalability	GOP Size = 8, Frame Rate = 3.75Hz~30Hz, Hierarchical B Pictures	
SNR Scalability	Coarse Granularity Scalability (CGS)	
Inter-layer Encoding	Adaptive motion, residual, textural predictions	
Sequences	Inter-layer Dependency	QP Settings
Akiyo	QCIF(A0←A1←A2), CIF(A1←B0←B1)	QCIF(50, 43, 37), CIF(44, 40)
Foreman	QCIF(A0←A1←A2), CIF(A1←B0←B1)	QCIF(46, 40, 34), CIF(41, 34)
Football	QCIF(A0←A1←A2), CIF(A0←B0←B1)	QCIF(41, 35, 30), CIF(36, 30)
Mobile	QCIF(A0←A1←A2), CIF(A0←B0←B1)	QCIF(41, 35, 30), CIF(34, 28)
Harbor	CIF(A0←A1←A2), 4CIF(A0←B0←B1)	CIF(41, 36, 31), 4CIF(37, 29)
ICE	CIF(A0←A1←A2), 4CIF(A1←B0←B1)	CIF(45, 40, 35), 4CIF(41, 33)

and computed the Mean Square Error (MSE) between the original and the compressed videos as an objective criterion.

6.2.1 Optimal Paths versus Video Contents

The optimal extraction paths depend heavily on video contents. This is because the spatiotemporal characteristics of video signals crucially affect the efficiency of interpolation algorithm performed on viewing devices. Refining temporal quality normally results in better R-D performance in fast-motion sequences, whereas maintaining spatial or SNR quality is more beneficial in slow-motion sequences. The results can be seen by comparing the optimal paths in Figure 6.4, where MSE and F.R. are used for distortion measure and temporal interpolation, respectively. Interestingly, most of the optimal paths preferentially improve temporal quality except the ones for Akiyo sequence. The reasons are twofold. Firstly, MSE has difficulties in appreciating temporal quality. Secondly, F.R. yields erroneous results in video frames undergoing rapid temporal changes. The two facts together explain the dramatic increase in MSE if video frames are skipped, and thereby justify the tendency of optimal paths to improve temporal quality.

6.2.2 Optimal Paths versus Distortion Measures

In addition to video contents, distortion measures also influence the choice of optimal paths. Figure 6.5 compares the paths found by using MSE and MOS criteria. It can be readily seen that the MSE-based extraction paths are biased towards temporal quality in comparison with the MOS-based solutions. The observation agrees with the general fact that MSE is likely to overestimate the quality degradation caused by temporal

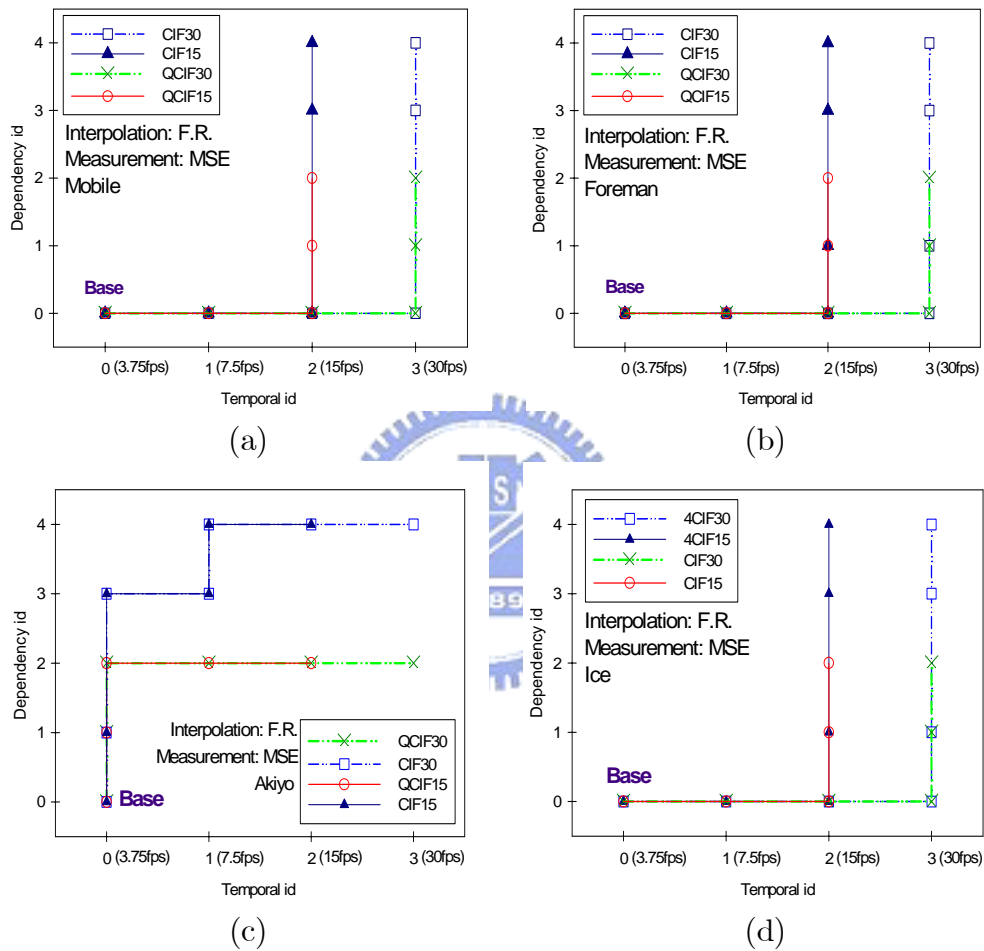


Figure 6.4: Comparison of optimal extraction paths for different viewing devices: (a) Mobile, (b) Foreman, (c) Akiyo, and (d) ICE. B.Direct and MSE are used for temporal interpolation and distortion measure, respectively.

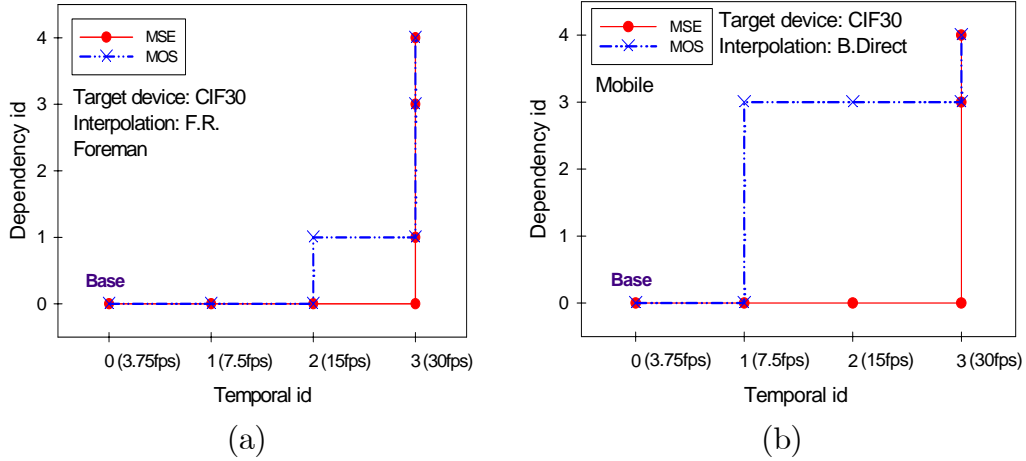


Figure 6.5: Comparison of optimal extraction paths found by using MSE and MOS as the distortion criterion: (a) Foreman CIF@30 and (b) Mobile CIF@30.

jerkiness even if the impairment in perceptual quality is insignificant. On the other hand, the results using MOS, although correlate much well with perceptual quality, are generally less analytical owing to the unpredictable nature of MOS. In view of the pros and cons of each measure, experimental results that follow are provided with both distortion criteria.

6.2.3 Optimal Paths versus Spatiotemporal Interpolation

Besides video contents and distortion measures, interpolation algorithms performed by viewing devices also have a significant effect on the optimal paths. Moreover, the temporal interpolation is more critical than the spatial interpolation because poor efficiency could easily give rise to significant distortion and visible artifacts. To this end, the influences on optimal paths are analyzed in Figure 6.6 by assuming the use of frame replication (F.R.) and B_Direct_16x16 (B.Direct) on viewing devices. In general, the B.Direct method provides better quality than the straightforward F.R. due to better estimation of motion fields. The fact also explains why the B.Direct method allows the extraction to improve more in spatial quality, while the F.R. causes it to extract more temporal layers. The results also confirmed that further optimization would be made possible if the interpolation algorithms performed by viewing devices are provided.

Summarizing, in this section, we have shown that the choice of optimal extraction

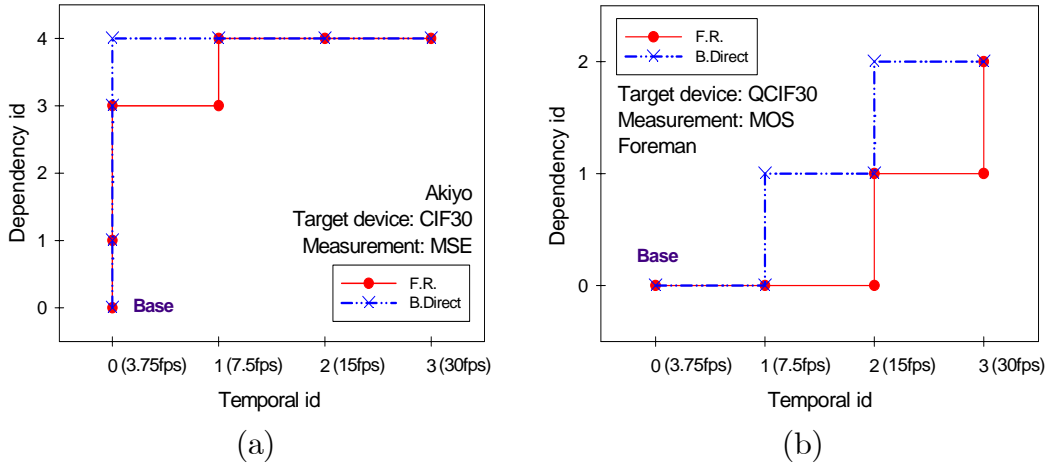


Figure 6.6: Comparison of optimal extraction paths using frame replication (F.R.) and B_Direct_16x16 (B.Direct) for temporal interpolation: (a) Akiyo CIF@30 and (b) Foreman QCIF@30.

paths is determined by several factors: the visual characteristics of video contents, the distortion measures, and the spatiotemporal interpolation algorithms performed by viewing devices. All these factors are related directly or indirectly to the final playback quality and should be considered jointly in the extraction optimization process.

6.3 Performance of Greedy Heuristic Scheme

After the optimal extraction paths have been studied in details, this section evaluates the performance of greedy heuristic scheme in search for optimal paths. Exhaustive search is used as baseline for comparison.

6.3.1 Extraction Paths and R-D Performance

Based on the MSE criterion and well-adapted bitstreams, Table 6.2 compares the optimal extraction paths found by the greedy heuristic scheme and exhaustive search. The differences in path indices are contrasted utilizing exclusive-OR operation.

Clearly, from the table the two methods produce almost identical results. It has been found from the R-D trellis diagrams that both *global* and *strong local* conditions are met in most test sequences, which explains the fairly good performance of the greedy heuristic scheme. The global condition results largely from the well-adapted settings. The local condition, on the other hand, is more intricate in that it represents

local R-D variations and may not be precisely controlled. In fact, no effort was made to adapt the QP or coding to take into account the local condition. The reason that the *strong local condition* holds in this particular set of experiments is mostly due to the MSE effect. Most local trellises are found to have a much higher preference for temporal quality.

The *strong local condition*, however, may be violated. One such example is the extraction of Akiyo sequence for QCIF30 devices, in which *only* the *weak local condition* is satisfied. It has been shown in our theoretical framework that the greedy heuristic scheme may fail to find the optimal solution in such case. This can also be seen practically from Figure 6.7 (a), where a wrong decision was made when encountering the two convex R-D segments at the upper-left corner. However, even if the optimal solution is not reached, we usually end up with a suboptimal path having very similar R-D performance to the optimal one (See the R-D comparison in Figure 6.7 (b)). This is because the greedy nature of the greedy heuristic scheme causes it to always pick the R-D points that are closer to the convex hull.

Before closing this section, it is worth remarking on a few phenomena exhibited by Figure 6.7 (b). First, there are R-D points violating the general expectation that distortion should decrease as the bit rate increases, which is usually true when considering the R-D performance of video codecs. However, Figure 6.7 (b) describes the *true* R-D behavior when decoded videos are presented on viewing devices; the distortion is measured with respect to the interpolated videos rather than the decoded videos. Apparently, the results depend not only on the encoding algorithm, but also on the interpolation schemes implemented on viewing devices. Second, the R-D optimized extraction offers significant improvement in playback picture quality. Without optimization, one may possibly choose an extraction path that has extremely poor R-D performance. An example of such a path is illustrated by the dash curve in Figure 6.7 (b).

6.3.2 Computational Complexity

The computationally most demanding part in search for optimal extraction paths is to collect the R-D data associated with each decodable NAL set. While the exhaustive search needs to actually decode all possible representations, the greedy heuristic scheme

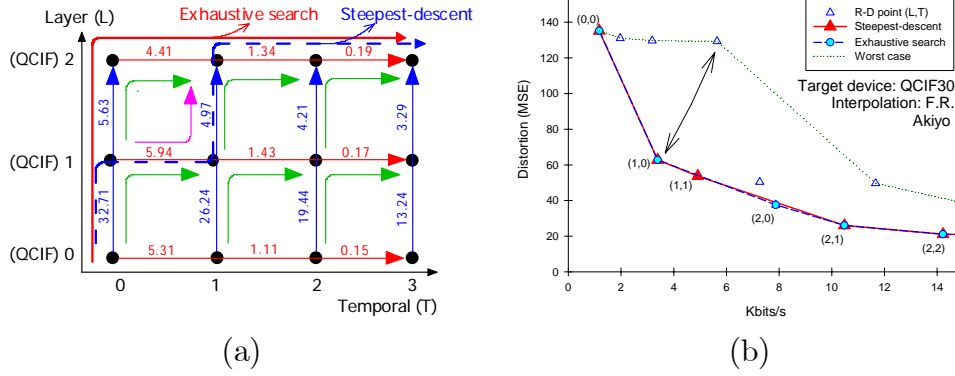


Figure 6.7: Comparison of extraction paths for the steepest-descent method and exhaustive search: (a) R-D trellis diagram and (b) R-D curves.

Table 6.2: Comparison of extraction paths with MSE.

	CIF30			CIF15			QCIF30			QCIF15		
	Exh.	S.D.	XOR	Exh.	S.D.	XOR	Exh.	S.D.	XOR	Exh.	S.D.	XOR
MSE + F.R.												
Akiyo	110100	110100	0	11010	11010	0	11000	10100	01100	1100	1100	0
Foreman	000111	000111	0	00111	00111	0	00011	00011	0	0011	0011	0
Mobile	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
Football	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
4CIF30			4CIF15			CIF30			CIF15			
Harbor	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
ICE	000111	000111	0	001111	001111	0	00011	00011	0	0011	0011	0
MSE + B.Direct												
Akiyo	111000	111000	0	11100	11100	0	11000	11000	0	1100	1100	0
Foreman	000111	000111	0	00111	00111	0	00011	00011	0	0011	0011	0
Mobile	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
Football	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
4CIF30			4CIF15			CIF30			CIF15			
Harbor	00011	00011	0	0011	0011	0	00011	00011	0	0011	0011	0
ICE	000111	000111	0	001111	001111	0	00011	00011	0	0011	0011	0

reduces the computation by lazy evaluation. On average, only *half* (42 ~ 58%) the number of decodable NAL sets are required for evaluation in order to achieve the same or similar performance. The gain is most obvious when an SVC bitstream contains a large number of decodable NAL sets.

6.4 Comparisons with Other Extraction Schemes

We conducted experiments to compare our adaptation scheme with the Quality-Layers-based approach [2] and Basic Extraction implemented in JSVM [11][8]. In our experiments, we examine two types of scalability: (1) QCIF SNR and (2) QCIF/CIF combined scalability. Two quality enhancements from the base quality are encoded for QCIF SNR scalability, while each spatial resolution in QCIF/CIF combined scalability

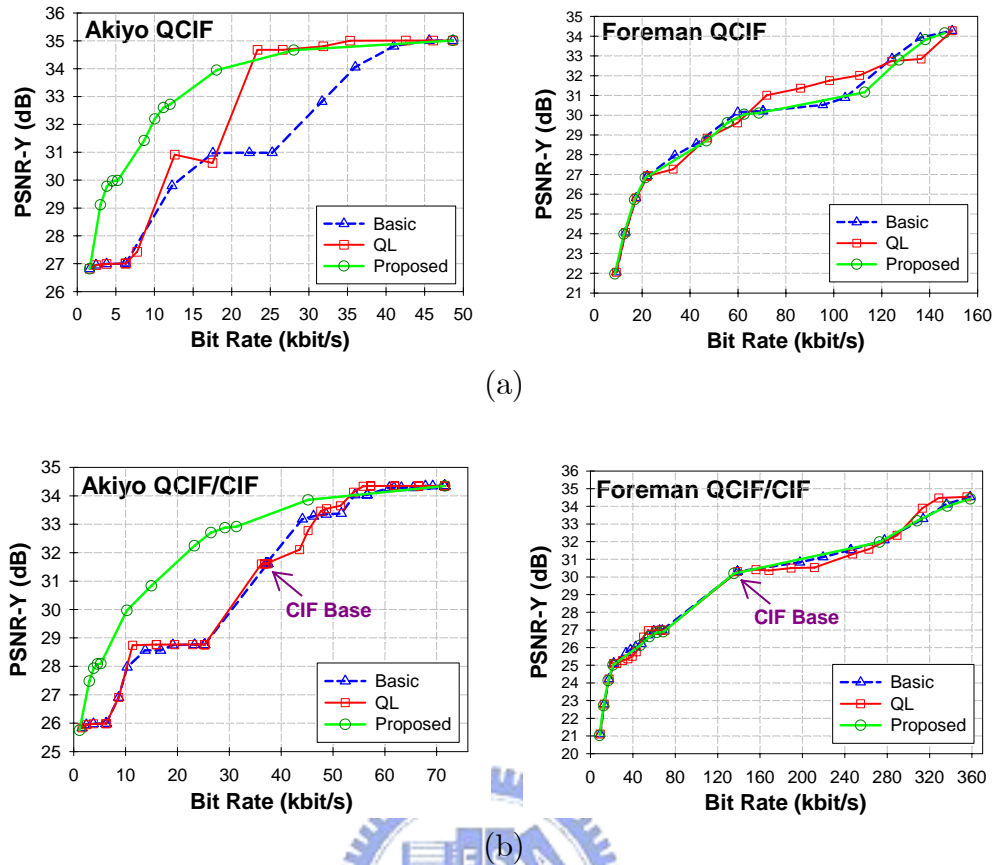


Figure 6.8: R-D performance comparison of the proposed scheme with the Quality Layer and Basic extractions in JSVM 9: (a) QCIF SNR Scalability, (b) QCIF/CIF Combined Scalability.

is encoded with a base quality and one quality enhancement. Both experiments use the MGS vector mode $\{3, 3, 4, 6\}$ without key pictures. In addition, each layer is simply predicted from the previous layer and the Quality Layers are assigned independently across spatial layers, i.e., the QCIF substreams must be entirely extracted prior to the extraction of the CIF layers.

From Figure 6.8, the proposed scheme is far superior to the other two approaches in Akiyo sequence while showing comparable performance in Foreman sequence. The reasons are twofold. Firstly, our scheme allows optimal extraction paths to preferentially improve spatial quality without extracting the entire base layer. However, both the Quality-Layers-based extraction and Basic Extraction must initially extract the base layer at full frame rate. Secondly, our extraction paths are derived based on the *real* R-D costs of scalable layers. Contrarily, the Quality Layers are computed by estimating the R-D information.

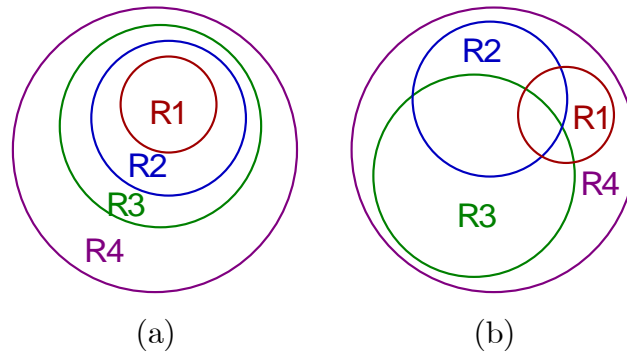


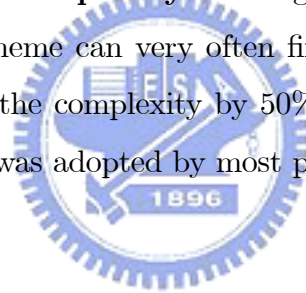
Figure 6.9: Bitstream extraction (a) with and (b) without successive refinement. R1-R4 indicate the extracted NAL sets associated with increasing bit rate.

Finally, we compare and contrast the major differences of our proposed scheme with other previous works, including the Basic Extraction in JSVM [11][8], the Quality Information Table [4], the Quality Index [7], as well as the Quality-Layers-based approach [2].

- Applications:** The Quality-Layers-based extraction [2] aims at *medium-grain quality adaptation*, while the other schemes focus on *multi-dimensional adaptation with combined scalability*. In particular, the Quality-Layers-based approach [2] is conditioned on the full extraction of the base layer, whereas the others allow performing R-D optimal extraction without the presence of the entire base layer, so does ours.
- Extraction Constraints:** Both our scheme and the Quality-Layers-based extraction must incrementally extract NAL units for *successive refinement*, while the others allow discretionary extraction. Through *successive refinement*, coarser representations are always embedded in finer ones, which leads to more efficient use and share of extracted NAL sets among viewing devices. The differences in bitstream extraction with and without *successive refinement* are shown in Figure 6.9 using Venn diagram.
- Extraction Criteria:** All schemes perform bitstream extraction based on the R-D performance of NAL units except the Basic Extraction approach, which carries out extraction in such a way that the resulting bitstream must have a bit rate that is closest to but not greater than the target bit rate. As it has been shown in our R-D analysis, decoding a substream with a higher bit rate does not necessarily produce better playback quality, especially when spatiotemporal

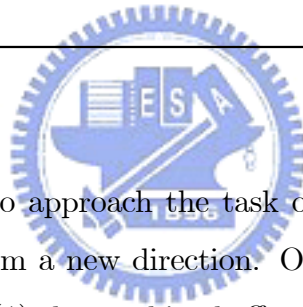
interpolation is involved.

- **Distortion Measurement:** Both our scheme and the Quality-Index-based approach compute the R-D data with respect to *interpolated videos* rather than *decoded videos*. Also, as indicated in our analysis, the *interpolated videos* can more realistically reflect playback quality on viewing devices. An even more direct approach is to acquire the perceptual preference, as used in the Quality Information Table. However, it would be impossible to have subjective evaluation for every video sequence.
- **Rate-Distortion Performance:** While most previous works simply try to construct an R-D optimized extraction path for pre-encoded SVC bitstreams, in this thesis we further recommended a set of criteria for generating well-adapted bitstreams, which together with strong or weak local condition promise the R-D convexity of optimal extraction paths.
- **Search Strategy and Complexity:** Through the use of well-adapted settings, our greedy heuristic scheme can very often find the optimal/near-optimal candidates while reducing the complexity by 50% or more in comparison with the exhaustive search that was adopted by most previous works.



CHAPTER 7

Conclusions



In our work, we attempted to approach the task of rate-distortion (R-D) optimized SVC bitstream extraction from a new direction. Our approach was characterized by three unique considerations: (1) the combined effect of proper encoder setting coupled with matching bitstream extraction and decoding mechanisms, (2) the computation efficiency of search strategies for R-D optimized extraction paths, and (3) the choice of extraction paths amenable to successive refinement of SVC bitstreams.

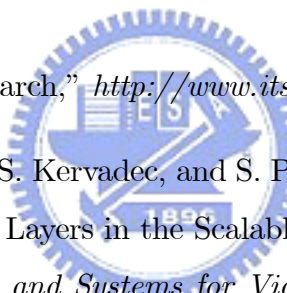
Through theoretical analysis of SVC inter-layer dependence relations and empirical study of the R-D performance of different encoded/extracted bitstreams, we obtain the following discoveries:

1. An optimal extraction path (corresponding to a convex R-D curve with minimal underlying area) can be found for an SVC bitstream if convex R-D performance can be maintained at every spatial/quality layer as well as temporal layers (referred as the global conditions) and in every pair of successive refinement steps (referred as the local conditions). If the convexity of R-D performance is violated only by minor deviations occur in a small fraction of all refinement steps then a near-optimal extraction path can be found.

2. Convex R-D performance can be maintained across spatial/quality layers by adapting the inter-layer dependencies between different layers and the quantization parameter QP of individual layer during SVC encoding. The R-D convexity of SVC layers (especially the spatial layers) can be predicted by referring to the R-D performance of corresponding H.264/AVC bitstreams encoded with fixed-quality or fixed-rate settings. On the other hand, convex R-D performance across temporal layers can be ensured by the proper cascade of QP values over the hierarchy of temporal layers.
3. The greedy heuristic scheme can be employed to search for the unique optimal extraction path if the SVC bitstream can satisfy both global R-D conditions and strong local R-D conditions. The greedy heuristic scheme is most computationally efficient as it decodes only half of the scalable layer representations in comparison with the exhaustive search strategy that was adopted by most previous works. Beside of being efficient, our experiments showed that the greedy heuristic strategy is also relatively robust with respect to its search results. The strategy can always find a sub-optimal extraction path close to the optimal path even under weak local R-D conditions. The strategy can even find the near-optimal extraction path when the global and local R-D conditions are violated in parts as when a subjective quality measure such as mean opinion scores (MOS) is used to quantify R-D performance.

Our work is still in its early stage, we plan to extend our investigation in several directions: (1) to study R-D optimized encoding and bitstream extractions for the SVC bitstreams with medium-grain scalability (MGS) support, (2) to conduct experiments with error concealment techniques and finally, (3) to devise computationally efficient strategies to search for optimal/near-optimal extraction paths under weak or fractional violation of global and local R-D conditions.

Bibliography

- 
- [1] “ITS Video Quality Research,” <http://www.its.bldrdoc.gov/n3/video/index.php>.
- [2] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, “Optimized Rate-Distortion Extraction With Quality Layers in the Scalable Extension of H.264/AVC,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1186 – 1193, September 2007.
- [3] 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.
- [4] Y. S. Kim, Y. J. Jung, T. C. Thang, and Y. M. Ro, “Bit-stream Extraction to Maximize Perceptual Quality Using Quality Information Table in SVC,” *SPIE Conference on Visual Communications and Image Processing (VCIP)*, vol. 6077, January 2006.
- [5] Z. La, W. Lin, B. C. Heng, S. Kato, S. Yao, and X. K. Yang, “Measuring the negative impact of frame dropping on perceptual visual quality,” *Human Vision and Electronic Imaging X, SPIE-IST*, vol. 5666, pp. 554 – 562, January 2005.

- [6] Z. G. Li, S. Rahardja, and H. Sun, "Implicit Bit Allocation for Combined Coarse Granular Scalability and Spatial Scalability," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 12, pp. 1449 – 1459, December 2006.
- [7] J. Lim, M. Kim, S. Hahm, K. Lee, and K. Park, "An Optimization-theoretic Approach to Optimal Extraction of SVC Bitstreams," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-U081*, October 2006.
- [8] H. Liu, H. Li, and Y. K. Wang, "Showcase of Scalability Information SEI Message," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-Q067*, October 2005.
- [9] W. H. Peng, J. K. Zao, T. W. Wang, and H. T. Huang, "Multidimensional SVC Bitstream Adaptation and Extraction for Rate-Distortion Optimized Heterogeneous Multicasting and Playback," *IEEE International Conference on Image Processing (ICIP)*, October 2008.
- [10] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312 – 322, September 2004.
- [11] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-9," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-V202*, January 2007.
- [12] 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.
- [13] H. Schwarz and T. Wiegand, "Further Results for an RD-optimized Multi-loop SVC Encoder," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-W071*, April 2007.
- [14] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE International Conference on Image Processing (ICIP)*, October 1999.
- [15] Y.-K. Wang, M. Hannuksela, S. Pateux, A. Eleftheriadis, and S. Wenger, "System and Transport Interface of SVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1149 – 1163, September 2007.

- [16] 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.
- [17] W. Yao, Z. G. Li, and S. Rahardja, “Balanced Inter-Layer Prediction for Combined Coarse Granular Scalability and Spatial Scalability,” *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2007.

