ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

NANYANG TECHNOLOGICAL UNIVERSITY

ERROR CONTROL AND JOINT RATE ALLOCATION TECHNIQUES FOR SCALABLE VIDEO TRANSMISSION

ERROR CONTROL AND JOINT RATE ALLOCATION TECHNIQUES FOR SCALABLE VIDEO TRANSMISSION

WANG YU

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

2010

WANG YU

2010

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ERROR CONTROL AND JOINT RATE ALLOCATION TECHNIQUES FOR SCALABLE VIDEO TRANSMISSION

WANG YU

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

To all in my family,
for their love.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Acknowledgements

First and foremost, I would like to express my deepest appreciation and sincere gratitude to my supervisor, Prof. Lap-Pui Chau, for his professional guidance and invaluable assistance. His unique insights and perspectives have enriched my research experience. I highly appreciate his kind patience and instant response to my queries even in his busy schedule. His enthusiasm for the research is a constant source of encouragement throughout my entire candidature.

Appreciation is also extended to all my friends in the Nanyang Technological University, for their valuable support and help during those long hours spent in the laboratory, and for their kind friendship in the daily life. They have made my life and work at NTU an enjoyable experience. I would also like to acknowledge School of EEE, NTU, Singapore, for awarding me the research scholarship and providing me with the excellent research facilities.

At last, this thesis is dedicated to my family, for the tremendous love and constant encouragement they have given me in the course of my studies. Without their persistent support, the achievement of my thesis is impossible.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Summary

Recent advances in video coding technology have led to a dramatic growth in network-based video applications. In these applications, users may access and interact with video content on different types of terminals and networks. One critical need in such a ubiquitous environment is the ability to handle the huge variation of resource constraints such as bandwidth, display capability, CPU speed, power, etc. Scalable video coding intends to encode the signal once at highest resolution, but enables decoding from partial streams depending on the specific rate and resolution required by a certain application. This allows for simple and flexible solutions for transmission over heterogeneous networks, additionally providing adaptability for bandwidth variations and error conditions. It further allows simple adaptation for a variety of storage devices and terminals. In this thesis, we study various fundamental techniques in error control and joint rate allocation for scalable video transmission. The goal of our research is to produce the best visual quality of the reconstructed video with a given resource constraint and achieve better tradeoff between computing complexity and quality.

In this work, we first give a brief introduction of the fundamentals of video coding as well as an overview of the scalable extension of H.264/AVC. After that, the proposed error control and joint rate allocation schemes are presented, which we believe to be the state-of-the-art technologies for scalable video transmission. The contributions are summarized as follows.

The first aspect focuses on channel coding scheme for transmission of scalable video over packet-erasure channel. Although today's broadband wireless networks can transmit

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

video data at high bit rates, there are still major challenges existing, such as fluctuations in channel quality and high error rate, which greatly affect the perceived video quality. This could be solved with an efficient unequal error protection (UEP) scheme. We propose a novel two-dimensional (2-D) UEP scheme for the video with combined scalability. Unequal amounts of protection are dynamically allocated to different parts of the compressed video stream considering the importance of each sub-stream and the channel conditions. In this way, the video quality has been much improved to achieve a graceful degradation in the presence of packet loss.

The second aspect of this thesis is the adaptive resynchronization approach for scalable video transmission. After scalable video coding, the compressed bit-stream contains base layer and enhancement layers, where the base layer is usually small and of high importance. Error-free transmission could be realized for the base layer through high-priority protection. Therefore, the overall video quality greatly depends on the enhancement layers. An adaptive resynchronization method is developed in this thesis. The enhancement layer bit-stream is separated into a set of units according to the temporal levels and quality levels. By measuring the utility-cost ratio, we arrange all the units hierarchically from the most important unit to the least important one. Moreover, an algorithm is designed to optimally insert different amounts of resynchronization markers to different units considering the time-varying channel conditions and the significance of each unit. Experimental results show that, by using the proposed resynchronization approach, a robust transmission of the enhancement layer is achieved and the overall quality is also improved.

The third aspect is bit-rate allocation for broadcasting of scalable video over wireless networks. Wireless broadcasting enables various mobile users with different platforms to access to the multimedia information simultaneously. A single transmission rate is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

unlikely to satisfy the heterogeneous requirements from all the receivers. Therefore, we develop a new system for wireless broadcasting using scalable video in this thesis. To realize a reliable transmission of the scalable video over the error-prone channel, we design different channel error protection schemes for different quality layers. Given the clients' bandwidth distribution, a novel algorithm is proposed to determine both the source coding bit-rate and the channel coding bit-rate for each layer to maximize a system-defined utility function. We implement the algorithm to verify its advantage and show how various allocation structures affect the overall utility.

The fourth aspect is joint rate allocation for multi-program video coding using scalable video codec. The main objective of joint rate allocation is to distribute the channel capacity among video sequences according to their respective complexities, thus a more uniform picture quality and a more efficient utilization of channel capacity are achieved. Most of the existing approaches are based on non-scalable video coding platforms, where computationally expensive encoding or transcoding is demanded to adjust the bit-rate of each video program. This thesis presents a new statistical multiplexing system, where the scalable extension of H.264/AVC is applied to compress the video sequences. We design an algorithm to dynamically allocate the channel bandwidth to different video sequences. In addition, the coding structure of each video is properly determined. With our algorithm, the joint rate allocation for all the video sequences is achieved without consuming much computation. We show that the variance of quality of different video sequences is dramatically reduced comparing with the existing method.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Contents

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Figures

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Tables

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 1

Introduction

1.1 Background and Motivation

Multimedia is one of the most important aspects of the information era. The demands for multimedia services are rapidly increasing and the expectation of the quality for these services is becoming higher. The conventional analogue video, due to its high bandwidth requirement for transmission and the difficulty in storage and manipulation, is no longer suitable for today's video communication applications. Compared with analogue video format, digital video can be more easily manipulated. Since the digital representation of raw video signals requires a high capacity, low complexity video coding algorithms must be defined to efficiently compress video sequences for storage and transmission purposes.

The advances in video compression technologies have led to an explosive growth in digital video applications [1], such as video conferencing, video telephony, digital video broadcasting (DVB) [3] [4], high definition television (HDTV) [2], interactive TV, telemedicine, etc. Video signals naturally contain a number of redundancies that could be

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

exploited in the digital compression process. These redundancies are either statistical due to the likelihood of occurrence of intensity levels within the video sequence, spatial due to similarities of luminance and chrominance values within the same frame or even temporal due to similarities encountered amongst consecutive video frames. Video compression is the process of removing these redundancies from the video content for the purpose of reducing the size of its digital representation. The statistic redundancy could be removed by entropy coding methods, such as Huffman coding [5], arithmetic coding [6] and run-level coding (RLC) [7]. The spatial redundancy could be reduced by transform coding, such as discrete cosine transform (DCT) [8] and discrete wavelet transform (DWT) [9]. The temporal redundancy could be explored by either the traditional motion estimation and motion compensation technique of MPEG standards [10] or the recently proposed motion compensated temporal filtering (MCTF) [11] [12].

Over the last two decades, video coding technology has witnessed an evolution. Both the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) and the International Standards Organization (ISO) have released standards for video coding algorithms that employ waveform-based compression techniques to trade-off the compression efficiency and the quality of the reconstructed signal. These standards include H.261 [13], H.262/MPEG-2 [14], H.263 [15], MPEG-1 [16], MPEG-4 [17] and the latest H.264/AVC [18]. With the introduction of the H.264/AVC video coding standard, significant improvements have been demonstrated in rate-distortion efficiency. The H.264/AVC could provide half bit-rate savings when compared with MPEG-2, which is the most common standard used for video storage and transmission, and the compression gain of H.264/AVC over H.263 is in the range of 25% to 50% due to different type of applications. However, such improvement is achieved at the expense of extremely high computing complexity caused by the newly adopted technologies in H.264/AVC.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Besides the coding efficiency, to satisfy the increasing demands of video streaming over computing networks, the transmission of video over heterogeneous networks should be more reliable, which requires efficient coding, as well as scalability to different client capabilities, system resources and network conditions [19] [20]. For instance, clients may have different display resolutions, systems may have different caching or intermediate storage resources, and networks may have varying bandwidths, loss rates, and best-effort or quality of service (QoS) capabilities. Scalable video coding is one solution to the problems posed by the characteristics of modern video transmission systems [21]. There are many applications that can benefit from scalable video coding: video streaming, video conferencing, video surveillance, video broadcasting, storage applications, and so on. The Joint Video Team (JVT) of the ITU-T VCEG and the ISO/IEC MPEG has now standardized a Scalable Video Coding (SVC) extension of the H.264/AVC standard. A scalable compressed bit-stream is one that contains multiple embedded sub-streams, each of which represents the original video content at a particular temporal rate, spatial resolution or quality represent. Due to its flexibility in coping with bandwidth variations, we are mainly concerned with transmission of scalable video in this work.

Compressed video streams are intended for transmission over communication networks. There has been a growing need for the support of multimedia services, which require the real time transmission of video data over fixed and mobile networks of varying bandwidth and error rate characteristics. Since the compressed video data is highly sensitive to information loss and channel bit errors, the decoded video quality is bound to suffer dramatically at high error rate. This quality degradation will be exacerbated when no error control mechanism is employed to protect coded video data against the hostility of error-prone environments. As shown in Fig. 1.1, due to the temporal and spatial predictions used in the video coding standards, a single error that hits a coded

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

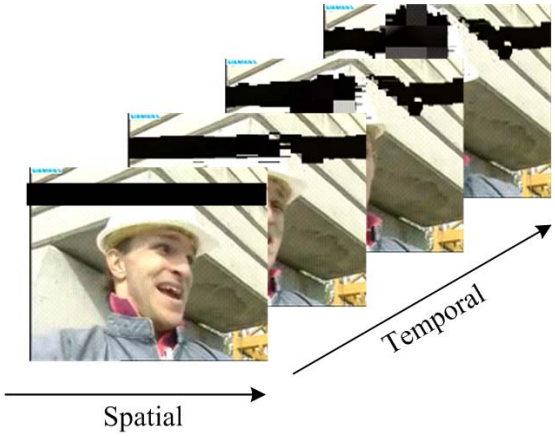


Figure 1.1. Error propagation.

video stream could lead to disastrous quality degradation for extended periods of time. To mitigate the effects of channel errors on the decoded video quality, error control schemes must be efficiently applied.

The advances in digital video compression and digital transmission technology have also enabled delivery of several digitally compressed video programs in a channel, which was used to transmit a single analog program. Fig. 1.2 shows a block diagram of independent coding of multiple video programs, where the programs are coded independently and each of them has a separate rate control. However, independent coding suffers from two major drawbacks: potentially large variations in picture quality among programs as well as within a program, and inefficient use of channel capacity. On contrast, it has been shown that joint coding of multiple video programs (see Fig. 1.3) is able to achieve more uniform picture quality and more efficient use of channel capacity by dynamically allocating the channel capacity among video programs. Therefore, joint rate control schemes should be investigated to reasonably allocate bit-rate.

Motivated by the above mentioned problems, this research work mainly focuses on designing efficient algorithms for error control and joint rate allocation for transmission of scalable video.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

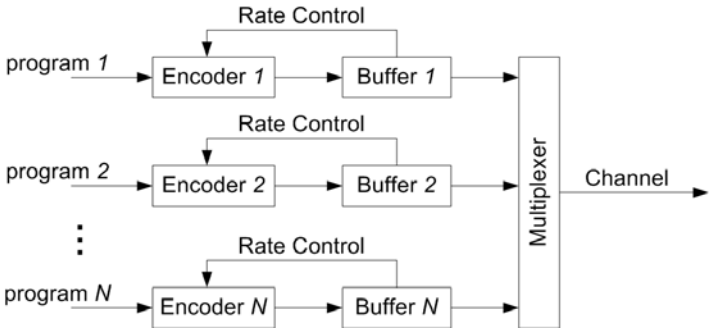


Figure 1.2. Independent coding of multiple video programs.

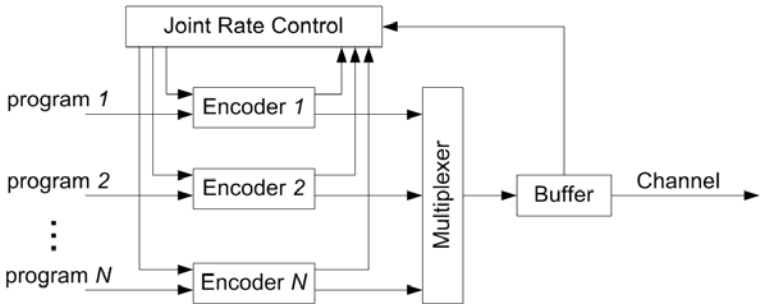


Figure 1.3. Joint coding of multiple video programs.

1.2 Objective and Main Contributions

The primary objective of this thesis can be mainly classified into two aspects.

The first objective is to provide sufficient robustness for video applications to ensure that the quality of the decoded video is not overly affected by the channel unreliability. Hence, error control techniques must be developed to alleviate the effect of transmission errors. Fig. 1.4 shows the block diagram of a generic video transmission system. The raw video data is firstly compressed by the video encoder into bit-stream, which is converted by transport coder into data units suitable for transmission over communication channels. Typical operations carried out in the transport coder include channel coding, framing of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

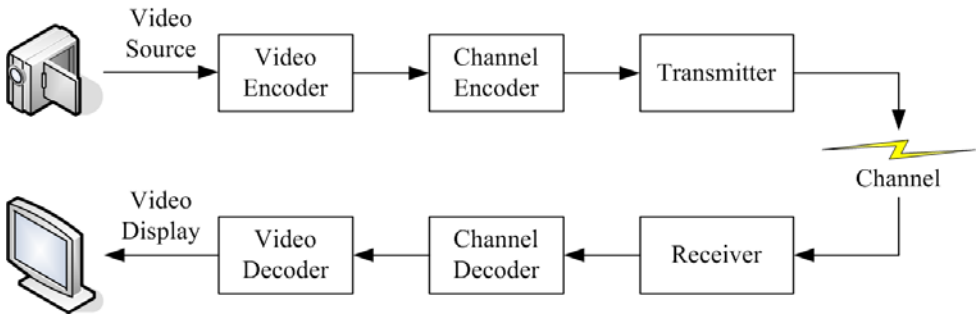


Figure 1.4. A generic video transmission system.

data, modulation and control operations required for accessing the communication channel. At the receiver side, the inverse operations are performed to reconstruct the video for display. Error control schemes can take place at different stages and normally they can be categorized into three classes:

(i) *Error resilience*. In such techniques, the video encoder plays an important role in improving the error robustness, typically by introducing redundant information in the transmitted data.

(ii) *Error protection*. In such techniques, the transport coder plays the primary part, where forward error correction (FEC), packetization and transport protocols should be designed to minimize the distortion due to transmission errors. Essentially, they all add a controlled amount of redundancy in the transport coder.

(iii) *Error concealment*. Error concealment techniques are purely decoder-based, whereby the video decoder attempts to benefit from previous received error-free video information for the approximate recovery or lost or erroneous data without relying on additional information from the encoder.

In this thesis, the first two classes of the error control techniques are explored.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The second objective is to develop efficient joint rate allocation algorithms for multi-program video compression system. For these systems, encoding of video programs at an equal constant bit-rate results in uneven picture quality due to different picture and motion content of the programs. How to reasonably allocate bit-rate among programs within a constrained fixed-rate network bandwidth becomes a great deal to achieve consistent optimized video quality in the distributive application. An efficient joint rate allocation algorithm can dynamically distribute the channel capacity among video programs according to their respective complexities, thus a more uniform picture quality and a more efficient utilization of channel capacity are achieved. In this work, the joint rate allocation techniques are extensively studied and the new algorithms are developed using the scalable video coding.

The main contributions of this thesis are as follows:

1. **Proposal of 2-D channel coding scheme for MCTF based scalable video coding:**

   In digital communications, channel coding is broadly used to deal with the transmission errors. In this work, a 2-D channel coding scheme is proposed to provide robustness for the scalable video with combined scalability over packet-erasure channel. Through allocation of different amount of channel protections to different part of the scalable bit-stream, the video quality has been much improved to achieve a graceful degradation in the presence of packet loss.

2. **Development of adaptive resynchronization approach for scalable video over wireless channel:**

   For scalable video, error-free transmission could be realized for the base layer through high-priority protection. Therefore, the overall quality greatly depends

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

on the enhancement layers. Resynchronization is proven to be a very effective tool among the state-of-art error-resilience techniques. In this work, an adaptive resynchronization approach is developed to achieve a robust transmission of the enhancement layer information. With the resynchronization method, the video exhibits robustness to the transmission errors and performs a graceful degradation over error-prone channel.

3. **Presentation of bit-rate allocation scheme for broadcasting of scalable video over wireless networks:**

   Wireless broadcasting enables various mobile users with different platforms to access to the multimedia information simultaneously. A single transmission rate is unlikely to satisfy the heterogeneous requirements from all the receivers. Therefore, we develop a new system for wireless broadcasting using scalable video in this thesis. To realize a reliable transmission of the scalable video over the error-prone channel, different channel error protection schemes are designed for different quality layers. Given the clients' distribution, the algorithm can determine both the source coding bit-rate and the channel coding bit-rate for each layer to maximize a system-defined utility function.

4. **Design of joint rate allocation algorithm for multi-program video coding using scalable video codec:**

   Joint rate allocation is an efficient tool to distribute the channel capacity among video sequences. Most of the existing approaches are based on non-scalable video coding platforms, where computationally expensive encoding or transcoding is demanded to adjust the bit-rate of each video program. In this work, we present a new statistical multiplexing system, where the scalable extension of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

H.264/AVC is applied to compress the video sequences. A joint rate allocation algorithm is designed to dynamically distribute the channel bandwidth to different video sequences. In addition, the coding structure of each video is properly determined. With our algorithm, a more uniform picture quality and a more efficient utilization of channel capacity are achieved without consuming much computation.

1.3 Outline of the Thesis

The thesis is organized as follows:

Chapter 2 first overviews the basic ideas behind the hybrid video coding structure. Related techniques and developments are introduced. It then gives a review of the scalable video coding. The H.264/AVC scalable extension is briefly introduced and different types of scalability are summarized, i.e. temporal scalability, spatial scalability, quality or SNR scalability and combined scalability.

Chapter 3 proposes a novel 2-D UEP scheme for MCTF based scalable video coding. The background and related works are first described. Then, the generation of layers using the scalable video with combined scalability is presented. A novel method is proposed to assign the channel protection codes considering the dependency of different layers. Unequal amounts of protection are allocated to different temporal layers based on the MCTF structure and in each temporal layer, unequal amounts of protection are allocated to different SNR layers to provide a graceful degradation of video quality as packet loss rate varies. To solve the allocation problem, an efficient algorithm is further designed to quickly get the allocation pattern, which is hard to get with other conven-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

tional methods. Finally, several experimental results and relevant discussions are provided.

Chapter 4 develops an adaptive resynchronization approach for scalable video transmission. It firstly gives a brief introduction of the related background. And then the separation of the enhancement layer bit-stream is described, where a set of units are generated according to the temporal levels and quality levels. By measuring the utility-cost ratio, all the units are arranged hierarchically from the most important unit to the least important one. Meanwhile, an efficient algorithm is designed to optimally insert different amounts of resynchronization markers to different units considering the time-varying channel conditions and the significance of each unit. At the end of this chapter, experimental results and discussions are presented.

Chapter 5 presents a bit-rate allocation scheme for broadcasting of scalable video. An overview of the related works is given firstly. A new system is developed for wireless broadcasting using scalable video. To deal with the transmission errors, two different channel coding schemes are designed for different quality layers. Moreover, a novel algorithm is proposed to determine both the source coding bit-rate and the channel coding bit-rate for each layer to maximize a system-defined utility function. The algorithm is implemented to verify its advantage and show how various allocation structures affect the overall utility.

Chapter 6 proposes a joint rate allocation algorithm for multi-program video coding using scalable video codec. At the beginning, the existing joint rate control schemes are investigated, most of which are based on the non-scalable video coding platforms. Then, a new statistical multiplexing system is presented, where the scalable extension of H.264/AVC is applied to compress the video programs. An novel algorithm is designed to dynamically allocate the channel bandwidth to different video programs. In addition, the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

coding structure of each video is also properly decided. The efficiency of our proposed joint rate allocation algorithm can be demonstrated in the experimental results.

Chapter 7 concludes this thesis with concluding summaries and discussions for future research directions.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2

Overview of Scalable Video Coding

2.1 Fundamentals of Video Coding

Digital video compression technologies have been growing rapidly in recent years due to the increasing demand of visual communication applications. Unlike audio signals, digital representation of raw video signals requires a large number of bits. Since video data is either to be saved on storage devices such as Compact Disk (CD) and Digital Versatile Disk (DVD) or transmitted over a communication network, the size of digital video data is an important issue in multimedia technology. Network bandwidth continues to increase, high bit-rate connections are commonplace and the storage capacity of hard disks, flash memories and optical media is greater than before. With continual falling of price per transmitted or stored bit, perhaps it is not immediately obvious why video compression is highly desirable for storage and transmission purposes. However, video compression has two important benefits. First, it makes it possible to use digital video in transmission and storage environments that would not support raw video. For instance, current Internet

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

throughput rates are insufficient to handle uncompressed video in real time. A DVD can only store a few seconds of raw video at television-quality and so video storage using DVD would be unpractical without video compression. Second, video compression enables more efficient use of transmission and storage resources. When the channel's available bandwidth is high, it is preferable to send a high-resolution compressed video or multiple compressed video programs instead of sending a single, low-resolution, uncompressed stream.

2.1.1 Principles of Video Compression

In most of the video coding schemes, a video is considered as a sequence of rectangular frames (or pictures) [22]. A frame typically consists of three rectangular arrays of integer-valued samples, one array for each of the three components of a tristimulus color representation for the spatial area represented in the image. The statistical analysis of video signals indicates that there is a strong correlation both between successive frames and within the frame elements themselves. Video compression could be realized by removing the redundant information from the video signal. If lossless compression schemes are employed, the statistical redundancy will be removed so that the original signal can be reconstructed perfectly at the receiver end. Unfortunately, the compression ratio achieved using lossless methods cannot satisfy the application requirement. As the Human Visual System (HVS) is not sensitive to loss of certain spatio-temporal visual information, lossy compression techniques can be used to reduce the video bit-rate while maintaining an acceptable image quality.

There are several types of redundancies existed in the video signals. These redundancies are either spatial due to similarities of luminance and chrominance values within the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

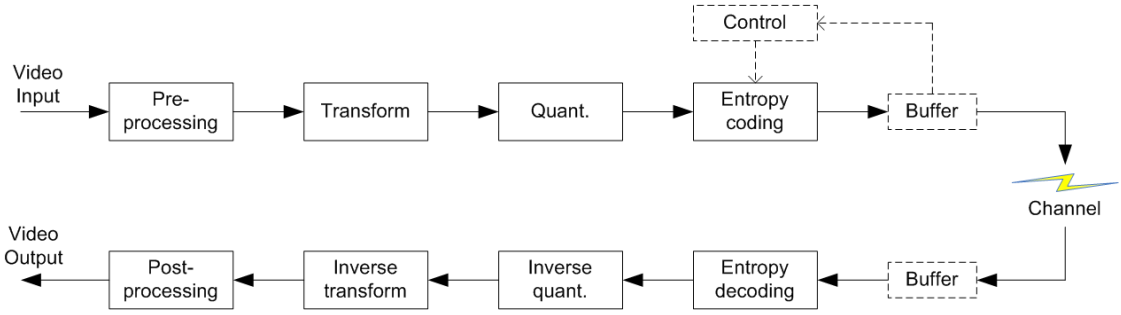


Figure 2.1. Block diagram of a typical video encoder and decoder.

same frame, statistical due to the likelihood of occurrence of intensity levels within the video sequence or temporal due to similarities encountered amongst successive video frames. Fig. 2.1 depicts a simplified block diagram of a typical video encoder and decoder. A number of techniques will be applied on each input video frame before the compression process is completed. The typical compression techniques can be classified as follows.

1. Prediction:

Prediction is a process by which a set of prediction values is created to predict the values of the input samples. Normally the encoder will indicate how to form the prediction based on analysis of the input samples and the types of prediction that can be selected in the system. Through this way, the values that need to be represented become only the differences from the predicted values, thus the encoding bit-rate could be decreased. These differences are called residual values.

2. Transformation:

Transformation is a process that maps the pixels into a transform domain prior to data reduction. A transformation can prevent the need to repeatedly represent similar values and can capture the essence of the input signal by using frequency analysis. The strength of transform coding in achieving video compression is that the image energy of most

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

natural scenes is mainly concentrated in the low frequency region, and hence into a few transform coefficients. These coefficients can then be quantized with the aim of discarding insignificant coefficients without significantly affecting the reconstructed image quality. The Discrete Cosine Transform (DCT) is broadly used because it has smoothly varying basis vectors that resemble the intensity variations of most natural images, such that image energy is matched to a few coefficients.

3. Quantization:

After transformation, a significant part of the image energy is concentrated at the lower frequency components, while less for the high frequency ones. The human eye is less sensitive to picture distortions at higher frequencies. This fact allows one to discard a great amount of information in the high frequency components. It is realized by simply dividing each component in the frequency domain by a constant for that component, and then rounding to a nearest integer. This is the so-called quantization. As a result, many high frequency components are rounded to zero and the rest become small positive or negative values.

4. Entropy coding:

Entropy coding is a process by which discrete-valued source symbols are represented in a manner that takes advantage of the relative probabilities of the various possible values of each source symbol. Variable Length Code (VLC) is a well-known entropy code. In VLC, short code words are assigned to the highly probable values while long code words to the less probable ones. There are two types of VLC in the standard video codecs. They are *Huffman* coding and *Arithmatic* coding. Huffman coding is a practical VLC code, but its compression cannot reach as low as the entropy. However, the arithmetic coding can approach the entropy since the symbols are not coded individually [23].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

To further improve the compression performance, the large amount of temporal redundancy should be taken into account. Usually, most of the natural scene is essentially just repeated in picture after picture without any significant changes. Thus, video can be represented more efficiently by sending only the changes in the video scene rather than coding all regions repeatedly. This is referred to as *Inter* coding. The most important techniques in inter coding are *motion estimation* [24] and *motion compensation* [25]. The frame difference is mainly caused by illumination variation and motion of the objects. Frame difference due to motion can be significantly reduced if the motion of the object can be estimated, and the difference is taken on the motion compensated frame.

2.1.2 Video Structure

An example of the data structure in a video sequence is shown in Fig. 2.2. In a video sequence, successive frames are similarly coded. A picture is the top level of the coding hierarchy. Due to the existence of several picture types, a Group of Pictures, called GOP, is the highest level of the hierarchy. A GOP is a series of one or more pictures to assist random access to the video sequence. Basically, there are three types of video frames

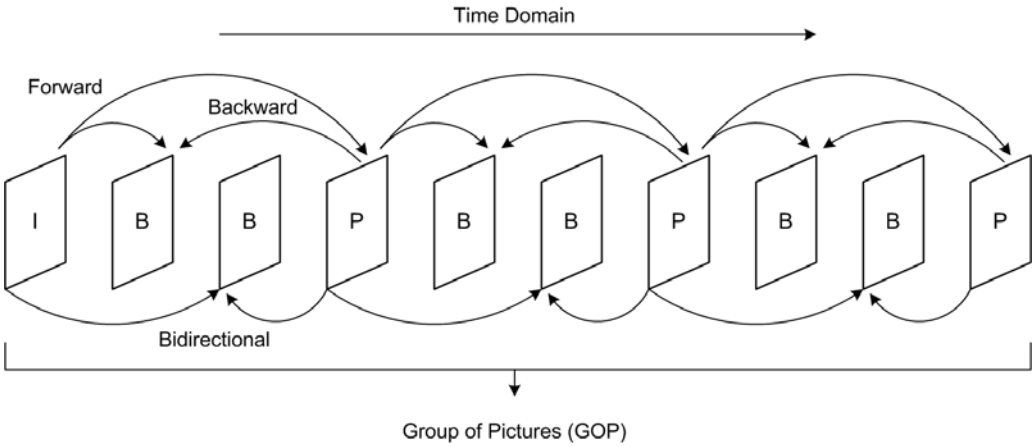


Figure 2.2. An example of frame configuration in a group of pictures (GOP).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

depending on which prediction methods are used.

(1) I-frame: The first coded frame in the GOP is an intra-coded (i.e., I-) frame, where all the macro-blocks (MBs) are intra-coded without reference to other frames. The I-frame is followed by an arrangement of P and B frames. Generally, I-frames result in the lowest compression ratio as they only exploit spatial redundancy.

(2) P-frame: P-frames are predictive-coded frames. Each P-frame is coded with reference to the nearest preceding I-frame or P-frame. Each MB in a P-frame can either be intra-coded or inter-coded using motion estimation and compensation. The intra-/inter-decision is made based on the rule that the mode which can generate the smallest coding cost will be chosen. Normally, P-frames lead to moderate compression ratio since they exploit both the spatial and the temporal redundancies.

(3) B-frame: B-frames are bi-directionally predictive-coded frames. Each B-frame makes use of both the nearest preceding and following I- or P- frames as reference. Each MB can be coded as an intra-block, an inter-block with forward prediction only, an inter-block with backward prediction only, or an inter-block which is bi-directionally predicted. Usually, B-frames achieve the highest compression ratio by further reducing temporal redundancy with bi-directional motion compensation.

The GOP length is defined as the distance between successive I-frames. A GOP may have any length, but it should be at least one I frame in each GOP.

2.1.3 Video Quality Assessment

The lossy video compression techniques introduce distortion to the reconstructed video quality. It is of great importance to know whether the distortion is acceptable to the viewers. In the past years, many subjective assessment methodologies are developed to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

evaluate the video quality, where a group of people could be asked to judge the quality of the decoded video against the original one and the levels of distortion are rated according to some rating system [26]. Although subjective assessments can give reliable indications of the perceived video quality, these methods are time-consuming and expensive.

Objective measurements are much faster and cheaper than subjective assessments. The two most widely used measurements are *mean square error* (MSE) and *peak signal-to-noise ratio* (PSNR) [26], both of which are numerical difference between the reconstructed frame and the original one. The definition of MSE is as follows,

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left[x(i,j) - x'(i,j) \right]^2 \tag{2.1}$$

with M and N being the dimensions (in pixels) of the picture and $x(i,j)$ and $x'(i,j)$ are the original and reconstructed pixel values at position (i,j), respectively. PSNR can be derived from MSE and is measured in *decibel* (*dB*):

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \tag{2.2}$$

where 255 is the peak signal with an 8-bit resolution. Although it has been claimed that in some cases PSNR's accuracy is doubtful, its relative simplicity makes it a very popular choice. If accuracy is a main concern, some more sophisticated models than simple pixel differences could be used [27].

2.1.4 Contemporary Video Coding Schemes

Over past years, video coding technology has witnessed a significant evolution. Although there exist several other video coding schemes, such as the wavelet based sub-band video coding [29] [30] [31], the hybrid video coding is the most common algorithm with successful applications [28]. Both ITU and ISO have released a number of video coding

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

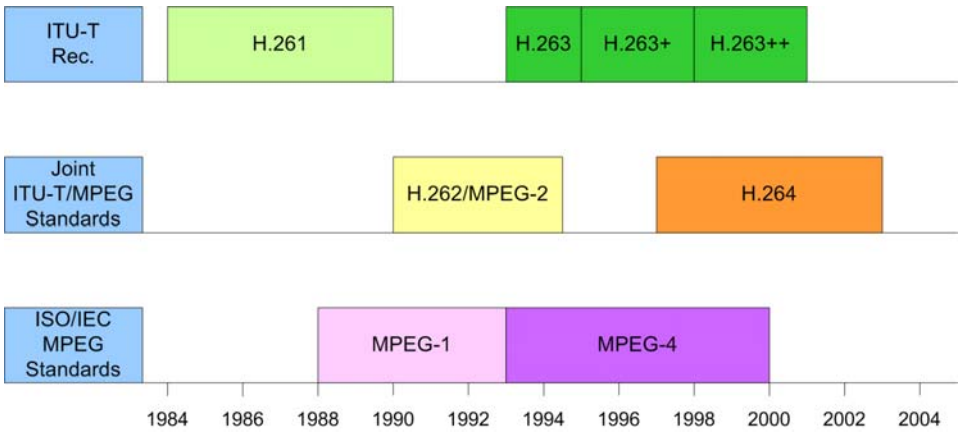


Figure 2.3. Development of various video coding standards.

standards, which have shaped the development of the visual communication industry. Fig. 2.3 illustrates the development of various video coding standards.

H.261 [13] is the first video coding standard, which is recommended by ITU-T for low bit-rate communications over ISDN at $p \times 64\,(1 \le p \le 30)$ kbit/s. This standard aims at meeting application requirements for videophone, video conferencing and other audio-visual services. It is an inter-frame DCT-based coding algorithm. Inter-frame prediction is first carried out in the pixel domain. The prediction error is then transformed into frequency domain, followed by quantization to reduce bit-rate.

ITU-T Recommendation H.263 [15] [33] was proposed in 1995 and designed for low bit-rate (less than 64 kbit/s) communication. The primary goal of the H.263 codec is to achieve a low or very low coding bit-rate for video applications such as mobile networks, public switched telephone network (PSTN) and the narrowband Integrated Services Digital Network (ISDN). The basic configuration of H.263 is based on H.261. Opposed to the full pixel precision used in H.261, half pixel precision is used for motion compensation in H.263. The performance of H.263 is improved in four aspects: unrestricted Motion

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Vectors (MVs), syntax based arithmetic coding, advanced prediction, and forward and backward frame prediction, which is similar to that in MPEG.

MPEG-1 [16] is the first generation of video coding standard proposed by MPEG. It is largely an extension of H.261, and many of the features are common. MPEG-1 is developed in response to industry needs for an efficient way of storing visual information on storage media such as CD-ROM and VCD. In most applications, the MPEG-1 video bit-rate is in the range of 1-1.5 Mbit/s. As the coding delay is not a major concern in storage applications, it can be traded for a higher coding efficiency.

Unlike MPEG-1, which is basically a standard for storing and playing video on a single computer at relatively low bit-rate, MPEG-2 [14] is a standard proposed for digital TV and meets the requirements of HDTV and DVD. The target bit-rate of MPEG-2 is 4-15 Mbit/s. To serve a wide range of applications, MPEG-2 introduces the concepts of *profile* and *level* as a way of encouraging interoperability without restricting the flexibility of the standard [32].

MPEG-4 Visual part 2 [34] [35] supports all functions that are already provided by MPEG-1 and MPEG-2. It is the first standard that considers content-based coding and introduces the concept of video object, where each video frame could be segmented into several objects and coded separately. MPEG-4 video aims at providing tools and algorithms for efficient storage, transmission and manipulation of video data in multimedia environments.

H.264 [18] [36], known variously as Advanced Video Coding (AVC), MPEG-4 part 10, is the newest video compression standard that is becoming the worldwide digital video coding standard for consumer electronics and personal computers. Besides providing higher coding efficiency than previous standards (MPEG-1, MPEG-2, MPEG-4

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Visual part 2, H.261 and H.263), the H.264/AVC standard is a straightforward video coding strategy designed to achieve a network-friendly video representation, simple syntax specifications, seamless integration of video coding into all current protocols and error robustness. Although most of the basic functional elements (prediction, transform, quantization, entropy coding) of H.264/AVC are also presented in the previous video coding standards, some new features are included, which enable the enhanced performance in coding efficiency and network adaptation.

2.2 Scalable Video Coding

With rapid advances in computing and communications, various user devices are accessible to visual content via different networks. Fig. 2.4 illustrates a heterogeneous video communication environment. In the environment, clients may have different display resolutions, systems may have different caching or intermediate storage resources, and networks may have varying bandwidths, loss rates, and best-effort or quality of service (QoS) capabilities. For example, the processing power of a personal digital

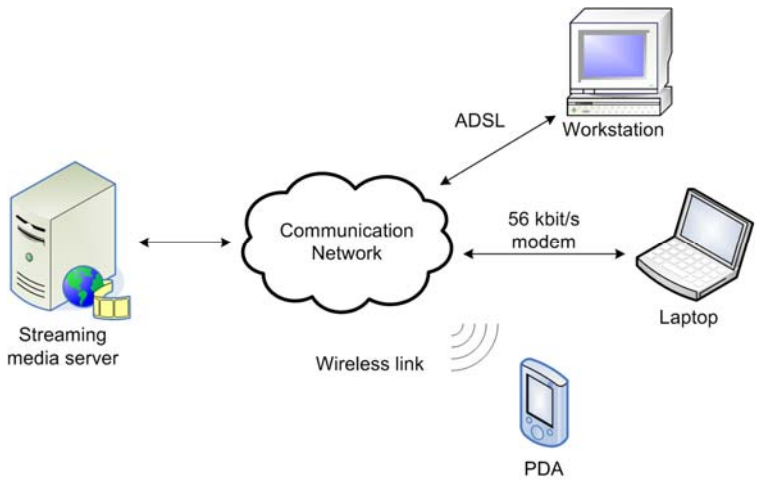


Figure 2.4. Heterogeneous video communication environment. Different end-users may have different network access speeds and different processing capabilities.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

assistance (PDA) cannot match that of a workstation. Reliable transmission of video over heterogeneous networks requires efficient coding, as well as scalability to different client capabilities, system resources, and network conditions [19]. Scalable video coding has been proposed to increase its adaptability to network and client conditions [21] [48]. A scalable compressed bit-stream is one that contains multiple embedded sub-bit-streams, each of which represents the original video content at a particular spatial resolution, temporal resolution, or quality represents. To be more specific, in scalable video coding, the input video is encoded into a *base layer* and one (or multiple) *enhancement layer(s)*. The base layer can provide a coarse playback video quality, while the enhancement layers can be progressively decoded to further improve the video quality. The standalone availability of an enhancement layer is useless unless all the corresponding lower layers (including the base layer) can be correctly received and decoded.

Early video compression standards such as ITU-T H.261 and ISO/IEC MPEG-1 produce monolithic bit-streams without scalability attributes [37]. In this case, if the decoder doesn't receive all the compressed bit-stream generated by the encoder, it can do nothing.

MPEG-2 is the first general-purpose video compression standard which includes a number of tools providing scalability. It intends to be forward compatible with MPEG-1, where eventually base information could be encoded and decoded by the old standard, while higher quality enhancement information is processed by the new standard [38]. All dimensions of scalability including spatial, temporal and SNR scalability are supported by MPEG-2. The scalability is realized by introducing multiple motion compensation loops, which will result in a decrease of the compression efficiency. Hence the number of scalable layers is generally restricted to a maximum of three in any of the existing MPEG-2 profiles. Fig. 2.5 and Fig. 2.6 give examples of two-layer SNR scalable encoder and two-layer spatial scalable encoder in MPEG-2 respectively.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

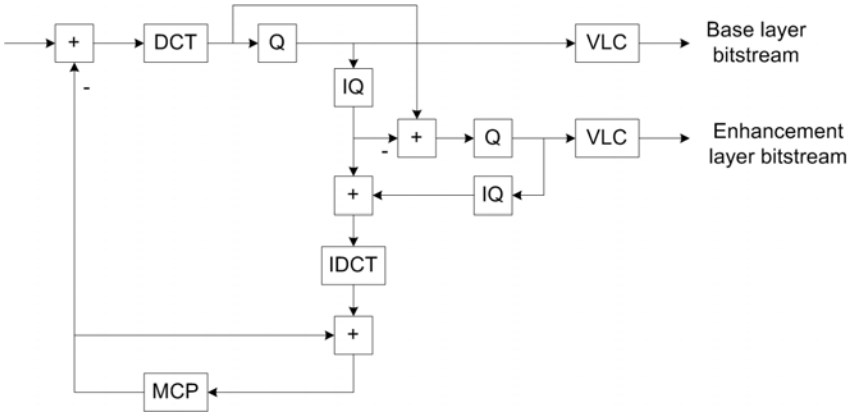


Figure 2.5. A two-layer SNR scalable encoder.

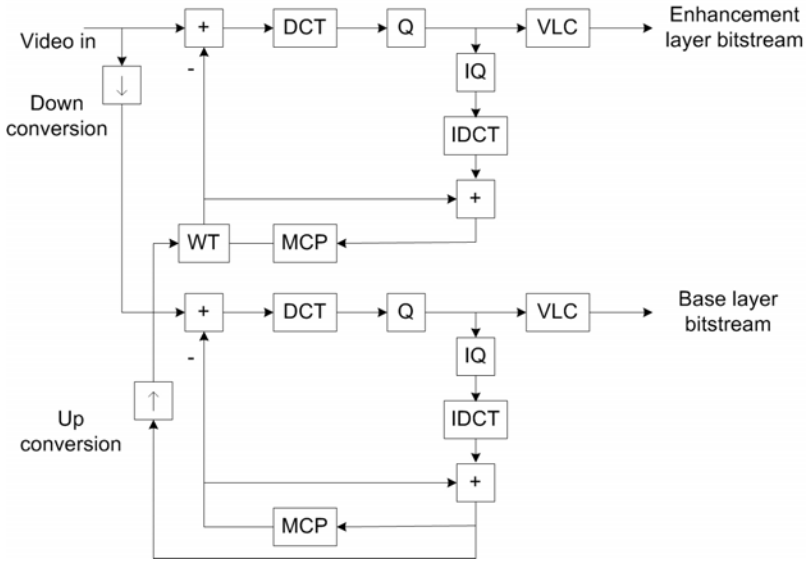


Figure 2.6. A two-layer spatial scalable encoder.

The video codec of the ISO/IEC MPEG-4 standard provides even more flexible scalability tools. MPEG-4 “Simple Profile” mode is equivalent to the ITU-T H.263 baseline codec, which provides no scalability. Extensions of H.263 define spatial, temporal, and SNR scalabilities as well. H.264/AVC, as recently defined as part 10 of the MPEG-4 standard, can in principle be run in different temporal scalability modes, due to its flexibility in the definition of prediction frame references. The most influential scalability

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

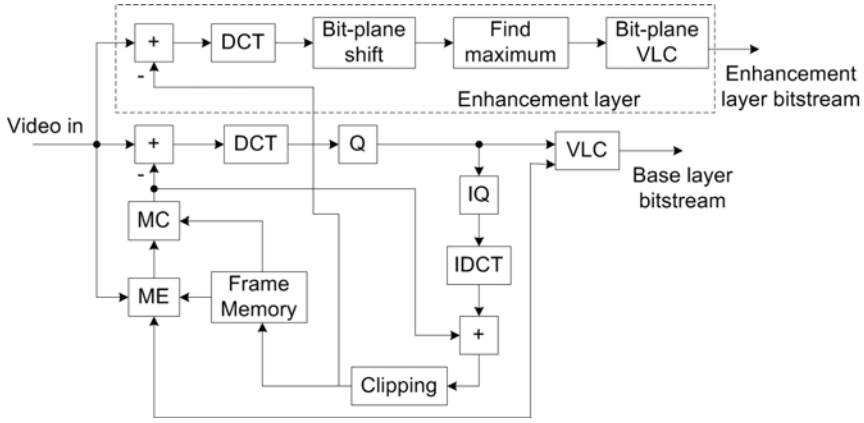


Figure 2.7. Fine granularity scalability in MPEG-4.

tool in MPEG-4 is fine granularity scalability (FGS) [39]. The basic idea of FGS is to encode a video sequence into a base layer and an enhancement layer. The base layer uses non-scalable coding to reach the lower bound of the bit-rate range. The enhancement layer is to encode the difference between the original picture and the reconstructed picture using bit-plane coding of the DCT coefficients. The block diagram of fine granularity scalability is illustrated in Fig. 2.7.

Although scalability has already been provided in MPEG-2 and MPEG-4, the decrease in coding efficiency greatly restricts its application. Hence, research for more efficient scalable coding techniques is a demanding topic in video compression. H.264/AVC is the most recent international video coding standard that provides significantly improved coding efficiency in comparison to all prior standards [42]. H.264/AVC has attracted a lot of attention from industry and has been adopted by various applications. Given the high degree of acceptance of the new standard and taking into account the large investments that have been taken place for the H.264/AVC-based products, a scalable video coding (SVC) scheme [40] [41] is developed as the scalable extension of H.264/AVC. In this thesis, our contributions for scalable video transmission are based on the scalable extension of H.264/AVC proposed by Fraunhofer Institute for Telecommuni-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

cations – Herinrich Hertz Institute (HHI). In comparison to the H.264/AVC standard, the scalable extension provides the temporal, spatial and quality scalability, as well as any combination of these scalabilities. Both the MCTF (Motion Compensated Temporal Filtering) and the "hierarchical B frame" can be employed to remove the temporal redundancy.

2.2.1 MCTF

Because a high correlation exists in the successive frames, efficient video compression technique requires an effective removal of temporal redundancy. In the traditional hybrid video coding algorithms [28], closed-loop motion compensation is employed in the temporal domain and 2-D DCT algorithm is applied in the spatial domain as described in section 2.1. In this scheme, the previous frame is reconstructed and used as reference to predict the current frame before motion compensation. The resulting displaced frame difference, which has much lower energy as compared with the original frame, is then encoded and transmitted to the decoder side. There are some limitations in this method. First, the reconstructed frame has lower quality than the original frame, especially when the transmission bit-rate is low. Thus, using this distorted reconstructed frame to perform ME/MC will result in decrease in the temporal prediction efficiency. Second, the closed-loop based video coding scheme employs the layered coding and the bit-plane coding schemes to achieve quality (SNR) scalability, such as FGS method [39] and its improvement versions [43]-[45]. Because only the reconstructed frames in the base layer are used as reference for motion estimation, the prediction gain is relatively low at high bit-rate. Third, when there is error happening during video transmission, the error will propagate to the following frames and lead to dramatic decrease of reconstructed video quality until

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the next I-frame is received. Therefore, this kind of closed-loop ME/MC scheme is unsuitable for the heterogeneous network environments, such as Internet.

To overcome these problems, an open-loop video coding scheme is proposed and becomes one important issue in scalable video coding. For a long time in the development of SVC, motion compensated temporal filtering (MCTF) was considered a useful coding structure for temporally scalable coders. Although investigations have revealed that benefits of MCTF are limited compared to hierarchical B pictures [47], MCTF is still an option to remove the temporal redundancy in SVC. MCTF was first introduced by Ohm [12]. After that, it is further improved by Choi and Woods [11]. MCTF is based on the lifting schemes [30] [31]. The lifting schemes may ensure a perfect reconstruction of the input signal in the absence of quantization of the decomposed signal even if non-linear operations are employed during the lifting process.

There are three steps involved in the lifting scheme: decomposition, prediction and update. At the decomposition stage, the input video signal is firstly separated into even signals and odd ones. Let L^n denote the input signal, where n is the temporal decomposition level ($n = 0$ for the original frames), $L^n[2k+1]$ and $L^n[2k]$ represent the odd signals and even signals respectively. The odd signals are predicted by a linear combination of the even signals using a predictor operation $\mathbf{P}(\bullet)$ and the high-pass sub-band is output as prediction residuals. The corresponding low-pass sub-band is achieved by adding a linear combination of the high-pass sub-band to the even signals using the update operation $\mathbf{U}(\bullet)$. The high-pass and low-pass sub-bands at temporal level n+1 are derived as follows (note that for simple expression, the spatial coordinates and the corresponding motion vectors are not specified):

$$H^{n+1}[k] = L^n[2k+1] - \mathbf{P}\left(L^n[2k]\right) \tag{2.1}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$L^{n+1}[k] = L^{n}[2k] + \mathbf{U}\left(H^{n+1}[k]\right) \tag{2.2}$$

where k is the temporal coordinate for each temporal level, $H^{n+1}[k]$ and $L^{n+1}[k]$ are the high-pass sub-band and the low-pass sub-band at temporal level n+1 respectively. The prediction and update operators for the temporal decomposition using the lifting representation of the Haar wavelet are given by

$$\mathbf{P}_{Haar}\left(L^{n}[2k]\right) = L^{n}[2k] \tag{2.3}$$

$$\mathbf{U}_{Haar}\left(H^{n+1}[k]\right) = \frac{1}{2}H^{n+1}[k] \tag{2.4}$$

For the 5/3 transform, the prediction and update operators are derived as

$$\mathbf{P}_{5/3}\left(L^{n}[2k]\right) = \frac{1}{2}\left(L^{n}[2k] + L^{n}[2k+2]\right) \tag{2.5}$$

$$\mathbf{U}_{5/3}\left(H^{n+1}[k]\right) = \frac{1}{4}\left(H^{n+1}[k] + H^{n+1}[k-1]\right) \tag{2.6}$$

As can be seen from the above equations, the Haar transform based temporal decomposition refers to temporal information of one directional neighboring frame, while 5/3 transform based temporal decomposition covers both directions. As the unidirectional

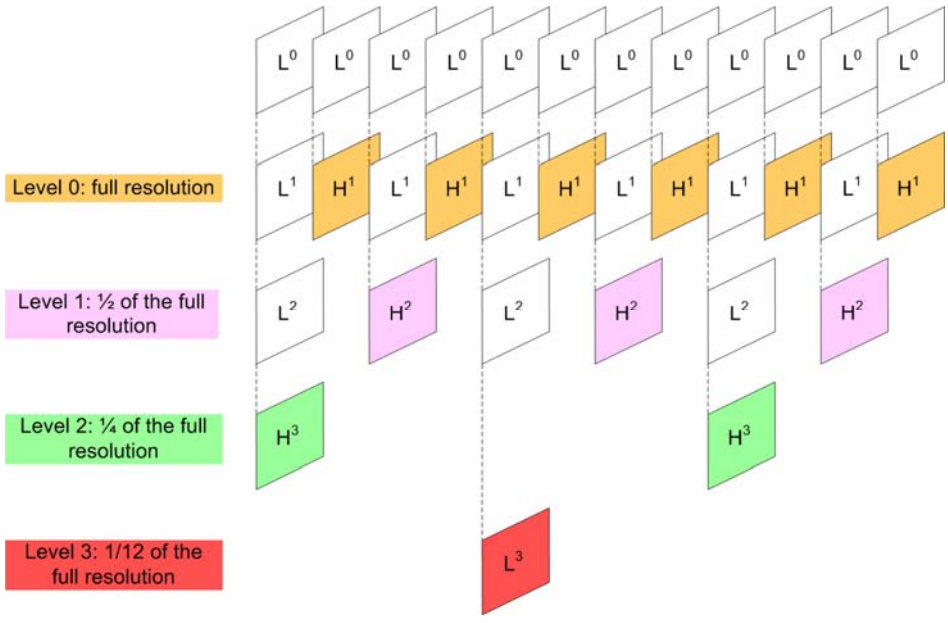


Figure 2.8. An example of MCTF structure.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

prediction produces less motion data compared with bidirectional prediction, it could save more bits for the spatial information at a given bit-rate. However, unidirectional prediction explores less temporal correlation and increase the energy of the residual data in the high-pass sub-band. To attain a tradeoff between these two effects, it is desirable to switch dynamically between unidirectional prediction and bidirectional prediction to achieve a high coding efficiency.

In Fig. 2.8, an example for the temporal decomposition of a group of 12 frames with 3 decomposition stages is illustrated, where H^n and L^n denote the high-pass and low-pass sub-bands at temporal level n.

Since both the prediction and the update are fully invertible, the reconstruction (synthesis) simply consists of the prediction and the update operations in reverse order with the inverted signs in the summation process. The equations can be derived as follows

$$L^n[2k] = L^{n+1}[k] - \mathbf{U}\left(H^{n+1}[k]\right) \tag{2.7}$$

$$L^n[2k+1] = H^{n+1}[k] + \mathbf{P}\left(L^n[2k]\right) \tag{2.8}$$

2.2.2 Hierarchical B-Pictures

When the update step is removed from the 5/3 transform based MCTF, a predictive video coding structure comes into being with hierarchical B-pictures [46] [47]. A typical hierarchical prediction structure with 4 dyadic decomposition stages is shown in Fig. 2.9. In each GOP, the key frame (e.g. 0^{th}, 8^{th}, 16^{th} display order in Fig. 2.9) is first coded as an intra frame or inter frame with reference to the key frame of the previous GOP. The remaining frames in a GOP are hierarchically predicted as illustrated in Fig. 2.9. It should be noted that the key frames are only predicted from other key frames, and each of the non-key frames is predicted using two reference frames, which are the nearest frames of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

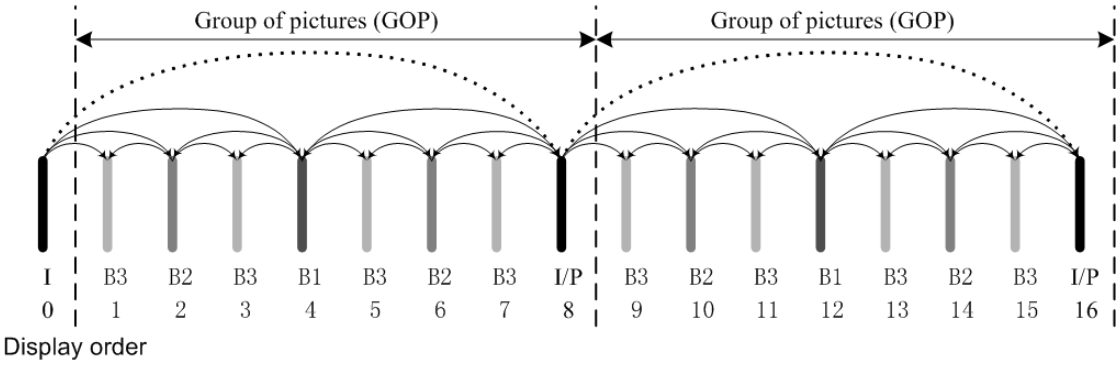


Figure 2.9. A typical hierarchical prediction structure with 4 temporal levels and a GOP size of 8.

the lower temporal level from the past and the future. As in Fig. 2.9, the coding order within each GOP is "I/P, B1, B2, B2, B3, B3, B3, B3". In this way, the hierarchical B pictures could also provide the temporal scalability as MCTF.

2.2.3 Temporal Scalability

Temporal scalability is defined to represent the same video content in different temporal resolutions or frame rates. A bit-stream provides temporal scalability when it can be partitioned into a temporal base layer and one or more temporal enhancement layers. Both the MCTF and the hierarchical B pictures inherently provide temporal scalability. With n decomposition stages, up to n levels of temporal scalability can be provided.

In Fig. 2.8, a group of 12 pictures is decomposed into 4 temporal levels with 3 decomposition stages. If only the low-pass sub-band L^3 is obtained after transmission, the video can be reconstructed at the decoder at 1/12 of the temporal resolution of the input sequence. By additionally transmitting the high-pass sub-bands H^3, the decoder can reconstruct the video that has 1/4 of the original temporal resolution. By further adding the high-pass sub-bands H^2, the video with half the temporal resolution can be recon-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

structed. Finally, if the remaining high-pass sub-bands H^1 are transmitted, a reconstructed video with a full temporal resolution is obtained. Therefore, the low-pass sub-band L^3 is the temporal base layer, while H^3, H^2 and H^1 refer to the progressive temporal enhancement layers.

Temporal scalability with a number of temporal enhancement layers can also be efficiently provided with the hierarchical B pictures. As illustrated in Fig. 2.9, four temporal layers can be generated with a group of 8 pictures using the hierarchical prediction structure. The key frames (I/P) represent the coarsest supported temporal resolution, which forms the temporal base layer. The enhancement layer pictures are encoded as B frames. The temporal resolution could be refined by further including the B frames of the next temporal levels.

2.2.4 Spatial Scalability

Spatial scalability is defined to represent the same video content in different spatial resolutions or frame sizes. To support spatial scalability [50], SVC employs the multi-layer coding method, which is also adopted in conventional video coding standards, such as MPEG-2, H.263 and MPEG-4 Visual. Each spatial layer corresponds to a supported spatial resolution. The lowest resolution video data is referred to as the base layer, and the higher resolution video data is referred to as the enhancement layer. In each spatial layer, both the motion-compensated prediction and the intra-prediction are implemented as for single-layer coding. To further improve the coding efficiency in comparison to simulcasting different spatial resolutions, inter-layer prediction mechanism is incorporated to explore the correlation between spatial layers as illustrated in Fig. 2.10.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

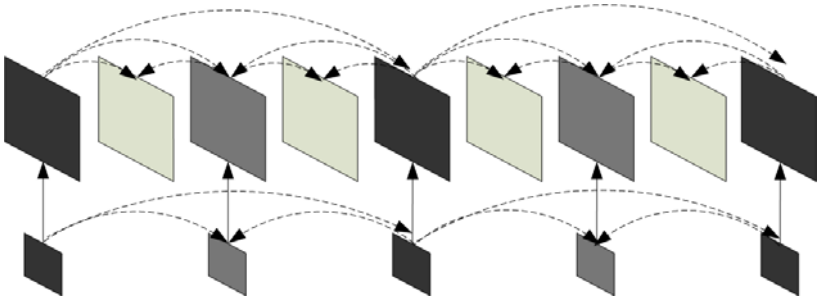


Figure 2.10. Illustration of multi-layer structure with inter-layer prediction to enable the spatial scalable coding.

As seen in Fig. 2.10, the inter-layer prediction enables reusing of the encoded lower resolution video sequence for coding of the corresponding higher resolution video sequence. The purpose is to use as much lower layer information as possible for improving rate-distortion efficiency of the enhancement layers. In traditional video coding standards, the inter-layer prediction is realized by simply up-sampling the reconstructed lower layer signal or by averaging such an up-sampled signal with a temporal prediction signal. In order to improve the coding efficiency for spatial scalable coding, two additional inter-layer prediction concepts [51] have been included in SVC: prediction of motion and prediction of residual data. The inter-layer prediction approaches in SVC can be mainly categorized into 3 aspects:

(1) *Inter-Layer Motion Prediction*: In comparison to H.264/AVC, two more macro-block modes are included for prediction of the motion vectors from the available motion information of the lower resolution spatial layer: a *base layer mode* and a *quarter pel refinement mode* [49]. For both of these two modes, the macro-block partitioning is obtained by up-sampling the corresponding co-located 8×8 block in the reference layer (see Fig. 2.11). The reference index for the up-sampled macro-block partitions are the same as for the co-located reference layer blocks. The associated motion vectors are derived by scaling the corresponding reference layer motion vector by a factor of 2. For

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

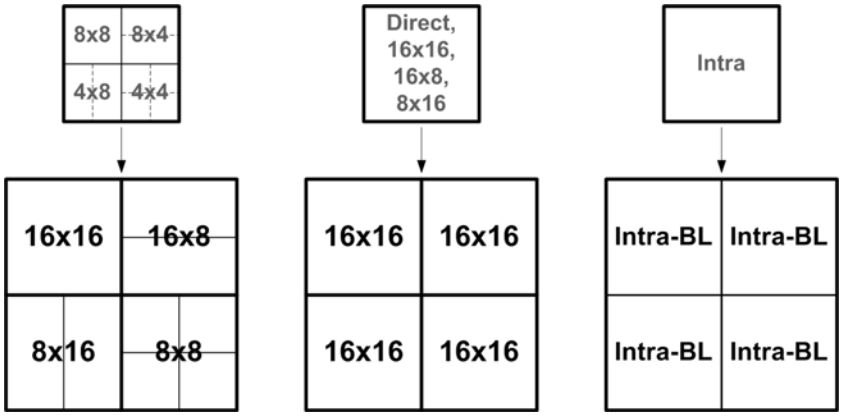


Figure 2.11. The up-sampling of MB partitions.

MB with *base layer mode*, no additional motion information is coded or transmitted, while for MB with *quarter pel refinement mode*, a quarter-sample motion vector refinement is further transmitted for each motion vector. Meanwhile, the up-sampled motion vector from the reference layer could also be used as motion vector predictor for the MB in the enhancement layer. If none of the above techniques are used, the macro-block modes as well as the corresponding reference indices and motion vector differences are encoded according to the H.264/AVC syntax.

(2) *Inter-Layer Residual Prediction*: Inter-layer residual prediction [49] can be employed for all inter-coded macro-blocks. If the motion vector for a block of the current layer is identical or nearly identical to the motion vector of the corresponding block in the previous layer, it is with a high probability that the coding efficiency can be increased when the coded previous layer residual is used as prediction for the current residual and thus only the difference between the current residual signal and the previous layer reconstruction is coded. However, when the motion vectors are not similar it is unlikely that a prediction of the residual signal could improve the coding efficiency. Consequently, the inter-layer residual prediction approach should be adaptively used.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(3) *Inter-Layer Intra-Prediction*: The inter-layer intra-prediction [49] can be applied to macro-blocks for which the corresponding blocks of the base layer are located inside intra-coded macro-blocks. With this mode, the intra texture information is predicted by up-sampling the reconstructed intra texture information from the lower spatial layer using the 6-tap filter that is defined in H.264/AVC for the purpose of half-sample interpolation. The prediction residual is transmitted using the H.264/AVC residual coding. Through this way, the intra prediction signal is directly obtained by de-blocking and up-sampling the corresponding 8×8 luminance block in the lower spatial layer. Therefore, the decoding complexity is significantly reduced.

2.2.5 Quality Scalability

Quality or SNR scalability is defined to represent the same video content in different perceptual quality. With quality or SNR scalable coding, the input video can be compressed into multiple layers including a quality base layer and one or more quality enhancement layers. For the quality base layer, H.264/AVC based transform coding is carried out. The predicted frames contain intra or residual macro-blocks as in the hybrid video coding described in Section 2.1. The intra frames are coded independently of each other as H.264/AVC intra frames. The intra macro-blocks are coded using intra coding modes, and the residual macro-blocks are encoded with DCT and quantization, the same as H.264/AVC. Fig. 2.12 illustrates a general structure of a scalable bit-stream with a base layer and L quality enhancement layers. The quality base layer provides the minimal visual quality with an initial quantization step size. To achieve the quality enhancement layers on top of the quality base layer, different scalable coding approaches are employed to provide different granularities of scalability as follows [49].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

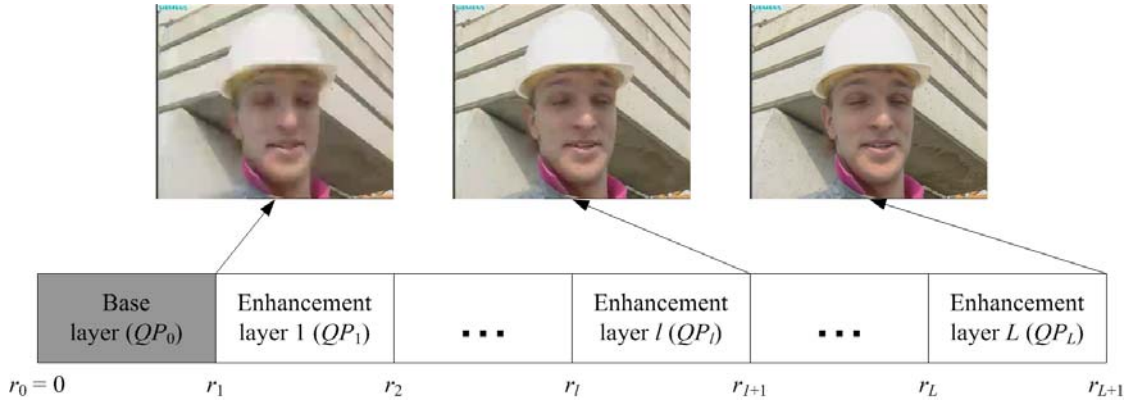


Figure 2.12. General structuring of a scalable bit-stream with a base layer and L quality enhancement layers.

Coarse-Grain Quality Scalability (CGS)

Quality or SNR scalability can be regarded as a special case of spatial scalability with identical frame size for all the layers. It is supported by the general concept for spatial scalable coding and referred to as *coarse-grain quality scalable coding*. To achieve CGS, the same inter-layer prediction mechanism as for spatial scalable coding is employed without using the corresponding up-sampling operations and the inter-layer de-blocking for intra-coded reference layer macro-blocks. Furthermore, the inter-layer intra- and residual- prediction are directly performed in the transform domain. When utilizing the inter-layer prediction for CGS in SVC, a refinement of texture information is typically achieved by re-quantizing the residual texture signal in the enhancement layer with a smaller quantization step size relative to that used for the preceding CGS layer.

In this way, a number of quality or SNR enhancement layers can be generated. However, the coarse-grain scalable coding only allows for a few truncation points in a scalable bit-stream, where the number of supported bit-rates is identical to the number of layers. Another limitation of coarse-grain scalable coding is that the enhancement layer has to be completely transmitted or decoded to refine the picture quality. If the available bandwidth

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

is not sufficient for transmission of the enhancement layer, the server will drop the entire enhancement layer for rate adaptation, which may lead to under-utilization of the channel bandwidth.

Fine-Grain Quality Scalability (FGS)

In fine-grain quality scalable coding, the base layer is similar to its counterpart in coarse-grain scalable coding in the sense that it has to be completely received and decoded. The difference lies in the enhancement layers. Within each spatial resolution, fine-grain quality scalability is achieved by encoding successive refinements of the transform coefficients, starting with the minimum quality provided by the H.264/AVC compatible base layer coding. This is done by repeatedly decreasing the quantization step size and applying a modified entropy coding process akin to sub-bit-plane coding. This coding mode is referred to as *progressive refinement*.

The FGS layers provide progressive refinement to the quality base layer. The progressive refinement can be truncated at any rate point, so that the decoded video can be improved in a fine-granular way, which is similar to the fine granularity scalability proposed in MPEG-4 visual [39]. Such fine-grained scalability enables a more precise rate adaptation. It should be noted that FGS coding mode has been removed from the SVC amendment that has been finalized in July 2007. A phase 2 SVC project is under study, which may contain FGS coding mode.

2.2.6 Combined Scalability

The concepts of temporal, spatial and quality scalability as described in Sections 2.2.3-2.2.5 can be easily combined to a general scalable video coding scheme, which provides a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

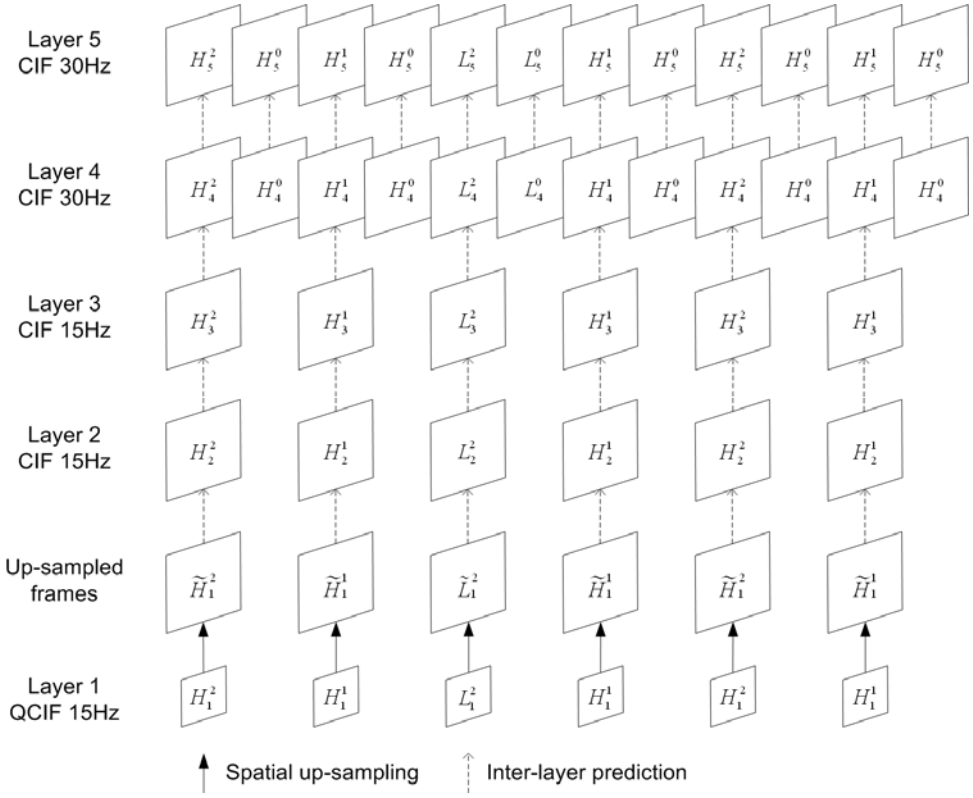


Figure 2.13. An example of combined scalability.

wide range of scalability. The coding scheme depends on the scalability that is required by a certain application.

Fig. 2.13 gives an example of combined scalability, where L_m^n and H_m^n represent the temporal low-pass and high-pass frames, respectively. *m* indicates the layer number, and *n* is the index of the temporal level. The video sequence with QCIF resolution is firstly encoded at the frame rate of 15Hz as layer 1 (base layer). The video data in layer 1 is up-sampled to be used as reference for layer 2. Through inter-layer prediction, layer 2 is encoded with CIF resolution. With further encoding of the quality enhancement layer, layer 3 is achieved based on layer 2. By additionally encoding the high-pass frames H_4^0, the temporal resolution is increased to 30Hz in layer 4. Layer 5 has the same frame size

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

and temporal resolution as layer 4, but the picture quality is improved with quality refinement layers.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 3

Two Dimensional Channel Coding Scheme for MCTF based Scalable Video Coding

3.1 Introduction

In pervasive media environments, users may access and interact with multimedia content from different terminals and via different networks. One critical need in such ubiquitous environments is the ability to handle the huge variation of resource constraints [52]. Recently, MPEG Committee has developed related tools and protocols to support development and deployment of video adaptation applications. In Joint Scalable Video Model (JSVM), the three main scalability aspects, i.e. temporal, spatial and SNR

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

scalability, are implemented to make the video streaming fully scalable. With the development of broadband wireless networks, such as IEEE 802.11b wireless LAN, delivering video over wireless networks has gained increasing attention. Although today's broadband wireless networks can transmit video data at high bit-rates, there are still major challenges existing, such as fluctuations in channel quality and high bit-error rates compared with wired links [53]. Transmission errors, together with lossy source coding techniques, lead to the distortion of the video sequences at the decoder. Hence, proper allocation of the system resources to minimize the distortion is required to transmit the video efficiently [54]. Unequal Error Protection (UEP), which is based on the Priority Encoding Transmission (PET) [55], has been proven to be very promising to resolve this problem.

In this chapter, for the scalable video with combined scalability, we propose a new two-dimensional (2-D) UEP scheme, which considers the dependency in both the temporal direction and the quality/SNR direction. It can properly allocate protection bits to the temporal layers, and in each temporal layer, unequal amounts of protection are allocated to different SNR layers. However, the optimal allocation for different layers is a complicated issue, especially when (1-dimensional) 1-D UEP is extended to 2-D UEP scheme. Many fast search algorithms have been proposed to achieve proper assignment, such as the hill-climbing method [58]. It has a non-increasing or non-decreasing pattern and is suitable for the 1-D UEP schemes for different layers with clear priority. In the 2-D UEP approach, the importance level of different layers cannot be estimated with a non-increasing or non-decreasing pattern. Hence, the hill-climbing method and other such kind of algorithms are not applicable in solving the 2-D UEP problem, which will be demonstrated later. Genetic Algorithm (GA) [65] is a randomized search technique. GA mimics the processes of natural evolution of species that led to higher organisms. It is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

often employed to solve difficult optimization problems. Hence, in this work GA is implemented in order to achieve a fast channel bits allocation.

The rest of the chapter is organized as follows. An overview of the related background is given in Section 3.2. In Section 3.3, we will present in detail the development of our proposed 2-D UEP scheme. The generation of the 2-D layers is firstly illustrated. After that, the bit-rate allocation problem is formulated and GA is briefly introduced. The experimental results are given in Section 3.4, where our novel algorithm is discussed and compared with traditional methods. Finally, we draw a conclusion in Section 3.5.

3.2 Related Background

3.2.1 Forward Error Correction (FEC)

In telecommunication and information theory, FEC is well known for both error detection and error correction. By adding redundancy to the transmitted information, it allows the receiver to detect and correct errors without the need to ask the sender for additional data. Since FEC has the effect of increasing transmission overhead and hence reducing usable bandwidth for the payload data, it must be used judiciously in video services that are very demanding in bandwidth but cannot tolerate a certain degree of losses. FEC has been studied for error recovery in video communications [66]-[71]. For example, in H.261, an 18-bit error correction code is computed and appended to 493 video bits for detection and correction of random bit errors in integrated services digital network (ISDN). For packet video, because several hundred bits have to be recovered when a packet loss occurs, it is more difficult to apply error correction [72].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

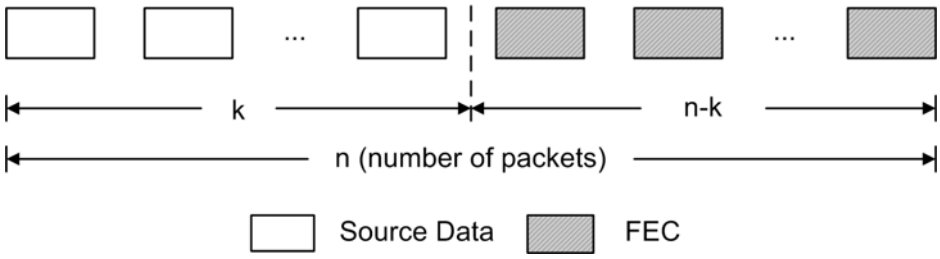


Figure 3.1. An example of Reed-Solomon coding.

Among various FEC codes, Reed-Solomon (R-S) codes are a well-known class of block codes with good erasure correction properties. R-S codes are particularly effective at recovering from erased symbols when the locations of the erased symbols are known. Because we are concerned only with networks in which packets either arrive perfectly intact or are completely discarded, we consider R-S codes that are optimized for erasures and do not correct bit errors [78]. Normally, these maximum distance separable block codes are denoted by a pair (N,k), where N is the block length and k is the number of source symbols. It has the property that a (N,k) code can exactly recover the k data symbols from any size k subset of the N total symbols. However, it will be unable to recover the original k data symbols if more than $N-k$ symbols are lost. Fig. 3.1 gives an example of the Reed-Solomon coding.

Because error recovery is carried out entirely at the receiver, FEC scheme can scale to arbitrary number of receivers in a large multicast group. In addition, due to its ability to recover any packets regardless of which packets are lost, it allows the network and receivers to discard some of the packets that cannot be handled due to limited bandwidth or processing power. Hence, FEC is also applicable to heterogeneous networks and receivers with different capabilities [73]. However, there are also some disadvantages associated with FEC, which are listed as follows,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(1) It increases the transmission bit-rate: The FEC schemes inflict redundancy bits on the transmitted video data. The higher the loss rate is, the higher the transmission rate is demanded to recover from the loss. The higher the transmission rate is, the more congested the network gets, which leads to an even higher loss rate. This makes FEC vulnerable for short-term congestion. However, efficiency may be improved by making use of UEP, which will be introduced below.

(2) It increases delay: One of the reasons is that an FEC must wait for all packets in a segment (i.e., k source packets in Fig. 3.1) before it can generate the redundant packets ($n-k$ FEC packets in Fig. 3.1). The other reason is because the receiver must wait for at least k packets of a block before it can playback the video segment. In addition, recovery from bursty loss, which is very common for wireless channels, requires the use of either longer blocks or techniques like interleaving. In either case, delay will be further increased. However, the video streaming applications can tolerate relatively large delay. Thus the increase in delay may not be an issue for this kind of applications.

(3) It is not adaptive to time-varying loss rate: As stated above, if more than $n-k$ packets of a block are lost, FEC cannot recover any portion of the original segment. This makes FEC useless when the short-term loss-rate exceeds the recovery capability of the protection code. On the other hand, if the loss rate is well below the recovery capability of FEC, the redundant information is more than necessary and a small ratio would be more appropriate. To improve the adaptive capability of FEC, feedback information should be used. That is, if the receiver sends the loss rate to the source, the channel encoder can adaptively add redundancy [73].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.2.2 Unequal Error Protection (UEP)

FEC techniques normally apply equal error protection (EEP) onto various video parameters. In other words, all the bits of the compressed video stream are treated equally, and given an equal number of redundancies, regardless of their sensitivity to errors and their contribution to overall video quality. However, different parts of the compressed video stream are not equally important. EEP makes the protection of highly sensitive data less efficient, while leading to unnecessary waste of bandwidth by overprotecting less important data. To solve this problem, the current research is heavily weighted toward unequal error protection (UEP) schemes, in which the video data can be protected with unequal rates depending on their sensitivity to errors.

UEP schemes have been addressed by many researchers [74]-[77] [56]-[64]. An effective UEP scheme takes advantage of the differential sensitivity of the output bit-stream of video encoder. The existing schemes can be mainly classified into three categories according to the consideration of different aspects of video stream sensitivities:

(1) Bits for headers, motion data and transform coefficients: The headers are more important than the motion vectors, in turn, are more important than the transform coefficients. Hence, the UEP scheme intends to add more redundancy for the motion data than for the DCT coefficients [74].

(2) Different frames in a GOP: UEP scheme is also applied on different types of frames, such as I-, P- or B-frames. Due to the predictive coding structure and the error sensitivities of the three types of frames, an UEP approach is proposed to exploit the dependency among I-, P- and B-frames in [75]. By further considering the temporal dependency of P-frames in a GOP, both [76] and [77] explore the sensitivity of succes-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

sive frames (including I- and P-frames) in order to minimize the overall distortion of the transmitted video sequence.

(3) Different layers of scalable coding: Different layers are not equally important. To provide a graceful degradation for the transmitted video stream, an obvious way is to add more protection to the layers that impact the quality more [56]-[64]. For example, Cheng, et al. [63] study the impact of using UEP to protect Fine Granular Scalability (FGS) compressed video. Sachs, et al. [64] apply UEP on Set Partitioning in Hierarchical Trees (SPIHT) coder and find an algorithm for optimizing the amount of protection bits used to protect progressive data.

3.2.3 Model of Wireless Packet-Erasure Channel

To evaluate the performance of the R-S codes, we need to know the probability that more than k symbols are lost since then the missing symbols cannot be reconstructed. We can compute this probability if we know the probability $P(m, N)$ with m being the number of lost packets within the block of N packets. In this chapter, we model the wireless network channel as a packet-erasure channel at the IP level. The process, which leads to packet losses over wireless packet-erasure channel, is quite complex. We assume the existence of a channel estimator, which indicates the probability that a particular number of packets are lost, given the total number of packets to be transmitted. This estimator could be formulated as any distribution of expected packet loss rate, such as uniform, binomial, exponential, Zipf, Poissen, etc. However, among such distributions, a Markov model [79] approximates the wireless channel's packet loss behavior fairly well.

In Markov model, hidden channel states $S = \{0,1,2,...\} \equiv N$ form a Markov chain with the initial state X_0 and the transitional probabilities $P(X_i \mid X_{i-1})$. The probability of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

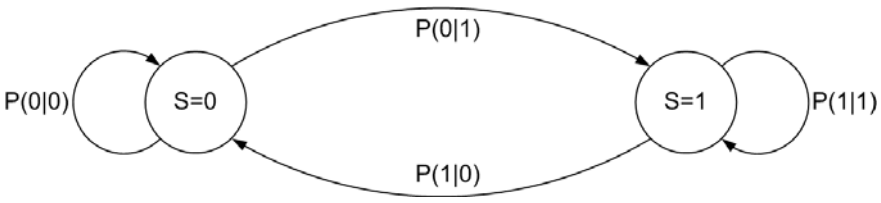


Figure 3.2. Two-state Markov channel model.

a packet erasure is represented by a conditional density $P(erasure \mid X_i)$. The number of states N and the values of the erasure probabilities $P(erasure \mid X_i)$ can be chosen in a number of ways. Here we apply a simple and analytical tractable Markov model with two states, which are denoted as 0 when packets are received correctly and 1 when packets are lost. Transitions between these two states occur with given probabilities and are used to capture the variability of the channel. Fig. 3.2 shows an example of a two-state Markov model, where we assume that the channel can be in only one of the two states (*good* and *bad* channel behavior). $S = 0$ denotes the good channel state, in which no error happens, and $S = 1$ represents the bad channel state when error occurs. The model is fully described by transitional probabilities $P(0 \mid 1)$ from state 1 to state 0 and $P(1 \mid 0)$ from state 0 to state 1. Since these parameters are not very intuitive, we use the average packet loss probability P_1 and the average burst length L_1 instead, both of which are calculated as follows,

$$P_1 = \frac{P(1 \mid 0)}{P(1 \mid 0) + P(0 \mid 1)} \tag{3.1}$$

$$L_1 = \frac{1}{P(0 \mid 1)} \tag{3.2}$$

where L_1 is the average number of consecutively lost packets [80]. With P_1 and L_1, we can calculate the block error density function $P(m, N)$. In the Markov model, the event of a loss frees the memory of the loss process and restarts it. The distribution of error-free gaps can describe such a model. Assuming that a gap of length *v* is the event that a packet

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

is missing after $v-1$ packets are received correctly, the gap density function $g(v)=P(0^{v-1}1|1)$ represents the probability of a gap length v. The gap distribution function $G(v)=P(0^{v-1}|1)$ gives the probability of a gap length, which is not smaller than v. Both $g(v)$ and $G(v)$ can be calculated using the transitional probabilities $P(0|1)$ and $P(1|0)$.

$$g(v)=\begin{cases}1-P(0|1) & \text{for} \quad v=1,\\ P(0|1)(1-P(1|0))^{v-2}P(1|0) & \text{for} \quad v>1,\end{cases} \tag{3.3}$$

$$G(v)=\begin{cases}1 & \text{for} \quad v=1,\\ P(0|1)(1-P(1|0))^{v-2} & \text{for} \quad v>1.\end{cases} \tag{3.4}$$

To facilitate the calculation of $P(m,N)$, we define $R(m,n)$ as the probability that $m-1$ packets will be lost within the next $n-1$ packets, which follows a lost packet. $R(m,n)$ is computed as

$$R(m,n)=\begin{cases}G(n) & \text{for} \quad m=1,\\ \sum_{v=1}^{n-m+1} g(v)R(m-1,n-v) & \text{for} \quad 2\le m\le N.\end{cases} \tag{3.5}$$

Finally, the probability $P(\mathrm{m},\mathrm{N})$ that m packets will be lost within N packets can be calculated as

$$P(m,N)=\begin{cases}\sum_{v=1}^{n-m+1} P_1G(v)R(m,n-v+1) & \text{for} \quad 1\le m\le N,\\ 1-\sum_{m=1}^{n}P(m,N) & \text{for} \quad m=0,\end{cases} \tag{3.6}$$

3.3 Proposed 2-D UEP Scheme

3.3.1 Proposed Framework

In scalable video coding, the input video sequence is divided into groups of pictures (GOPs) with fixed-size m (m is a power of 2). In this work, temporal scalability is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

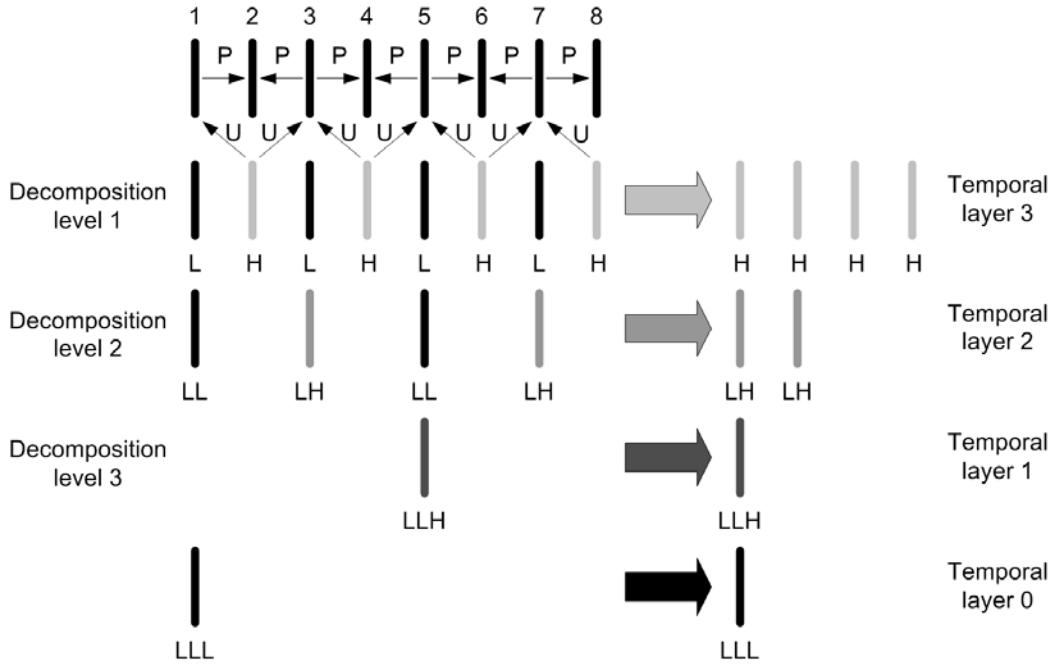


Figure 3.3. MCTF of a group of pictures.

realized by employing MCTF on each GOP although the proposed scheme is also applicable for the hierarchical B structures. MCTF [11] is implemented in each GOP, where even number frames use neighboring frames as reference for prediction and output high-pass subbands. The high-pass subbands are reversely used to update the odd number frames so that the low-pass subbands are generated. This is one level of temporal decomposition. Low pass subbands will be further decomposed in the same way. After n levels of decomposition there will be one low-pass subband and $m-1$ high-pass subbands. Meanwhile, $n+1$ temporal layers are generated, where n is decided by m ($m=2^n$). This process is illustrated in Fig. 3.3. At the decoder side, if only the low-pass subband is received, the video sequence can be reconstructed with a temporal resolution f_{Hz}/m (f_{Hz} is the frame rate of the original video sequence). With receiving of higher temporal layers based on all the lower temporal layers, the temporal resolution increases until a full resolution is achieved. This is denoted as temporal scalability. The enhance-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

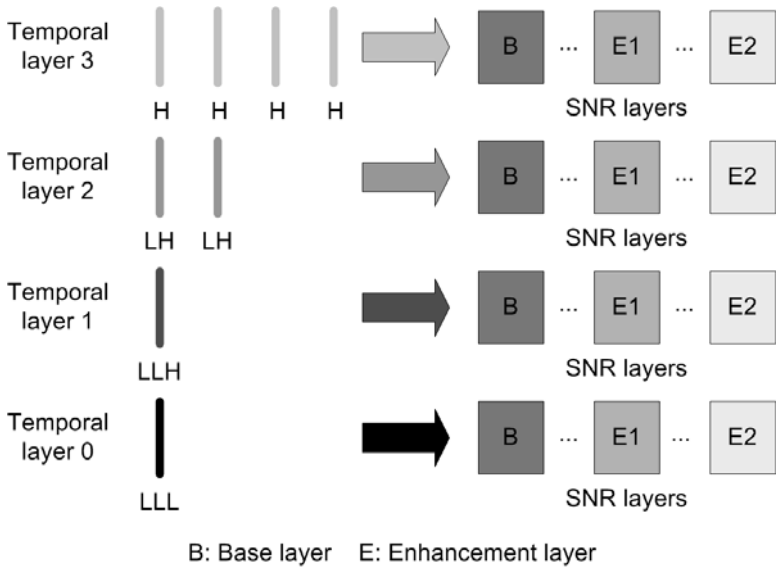


Figure 3.4. SNR scalability in each temporal layer.

ment of temporal resolution improves the visual quality, which is evaluated by PSNR of the reconstructed video sequence in this work.

After MCTF, for each subband, texture is encoded to produce a base layer, which provides a minimum quality at a given quantization level. To achieve quality/SNR scalability, the quantization parameter is repeatedly decreased and the refinements are encoded starting from the base layer. Therefore a number of SNR enhancement layers are formed. With more SNR layers received by decoder, the quality of the reconstructed frame will be further improved. In Fig. 3.4, all high-pass subbands in the same temporal layer are assumed as a whole object. Each temporal layer contains several SNR layers. Each SNR layer in each temporal layer is defined as a scalable unit.

Due to the decomposition structure of MCTF and the dependency between SNR layers, the effect of channel errors on the video could be extremely severe when compressed video data is transmitted over error-prone channels. Thus the video applications should provide sufficient robustness to ensure that the quality of the decoded video is not overly

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

affected by the channel unreliability. In order to provide the video data with some measure of reliability while communicating over error-prone channels, it is necessary to exert some forms of error control. The FEC codes are designed to protect data against channel erasures by introducing parity packets. If the number of erased packets is less than the decoding threshold for the FEC code, the original data can be recovered perfectly. In this work, the well-known Reed-Solomon codes are used to generate the FEC codes.

From the relationship existing among the 2-D scalable units, we can find that: (1) the lower temporal layer an error occurs in, the more the reconstructed frames will be affected, and the lower the quality of the reconstructed sequence is, (2) the lower SNR layer an error occurs in, the lower the quality of the reconstructed sequence is. To minimize the impact of the transmission error, an appropriate choice of protection bits for all the 2-D scalable units is necessary. That is, we should add more protection bits to the 2-D scalable units with lower temporal and SNR levels while adding fewer to the higher ones.

Our proposed 2-D UEP scheme for all the units in a GOP is shown in Fig. 3.5. There are T temporal layers, each of which is further divided into F SNR layers. We add FEC codes for each unit in each temporal layer. Both the source data and the FEC codes are vertically packetized into N packets. The length and the height of the FEC codes for the unit (i, j) are denoted by $k(i, j)$ and $h(i, j)$, respectively. Here i represents the temporal level and j represents the SNR level, where $i = 0,1,2,...,T-1$ and $j = 0,1,2,...,F-1$. Therefore, the length of unit (i, j) is $N - k(i, j)$, with N being the total number of packets in a GOP time interval. The packet size is denoted by M.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

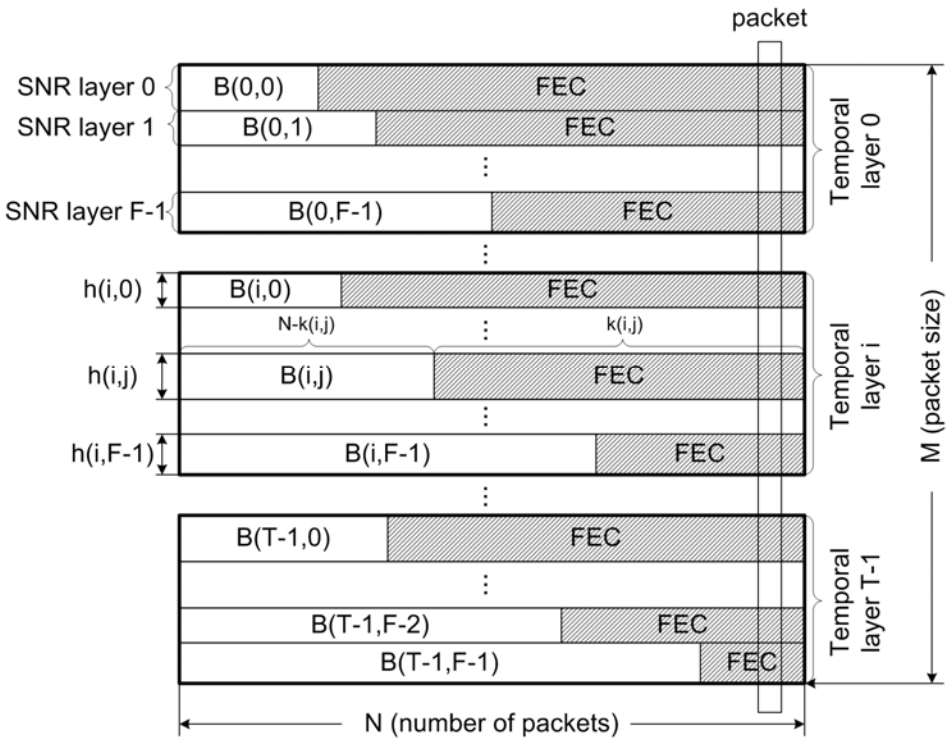


Figure 3.5. Channel coding scheme for the 2-D scalable units.

The problem is: what is the appropriate channel bits allocation for all the units to provide the best video quality under the constraint of a constant transmission rate? This will be resolved in the following section.

3.3.2 Problem Formulation

Next we will discuss the problem of allocating channel bits subject to an overall target bit rate as illustrated in Fig. 3.5. In this chapter, we mainly consider the problem of allocating channel rates to the compressed bit-stream. Therefore, we don't consider source distortion here because it is independent of the channel coding [81]. With all constraints we want to find the best allocation mode such that the overall quality of the decoded video is maximized. It should be noted that using different objective functions to evaluate the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

quality may result in different channel rate allocation patterns. As stated in section 2.1.3, MSE and PSNR are two commonly used metrics. Besides, there are a number of more complicated and precise metrics existed [130]. As PSNR is broadly used as a measure of quality of reconstruction of compressed image due to its relative simplicity, in this work we apply PSNR as the objective function and try to maximize it although other metrics can also be used. The overall PSNR is calculated as follows,

$$PSNR_{\text{overall}} = \sum_{i=0}^{T-1} \sum_{j=0}^{F-1} S(i, j) \times P(i, j) \tag{3.7}$$

where $S(i, j)$ is the PSNR increment provided by unit (i, j) and $P(i, j)$ is the probability of correctly receiving the corresponding unit.

$S(i, j)$ is calculated experimentally. Firstly we consider a simple case where there are no SNR enhancement layers. With l temporal layers received, 2^{l-1} frames are decoded and used to reconstruct the sequence and calculate the PSNR increment $S(i,0)$. Next, we take the SNR enhancement layers into account. In each temporal layer, with more SNR layers received, the PSNR of the reconstructed sequence is increased. In order to calculate $S(i, j)$, $PSNR(i, j)$ is defined as the average value of PSNR of the reconstructed sequence when $i+1$ temporal layers with $j+1$ SNR layers in each temporal layer are received. For example, the sequence size is 8 and after MCTF there are four temporal layers, where three SNR layers are further generated in each temporal layer. We'll show how $PSNR(2,1)$ is calculated as an example. With 3 temporal layers received, four frames can be decoded, including the 1st, 3rd, 5th and 7th frames of the sequence. Each of the other four frames will use the previous frame as replacement to calculate PSNR. Finally with the reconstructed frames and the original ones, $PSNR(2,1)$ is calculated. With $PSNR(i, j)$, $S(i, j)$ can be computed as follows.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For the SNR base layer of the temporal base layer,

$$S(0,0) = PSNR(0,0) \tag{3.8}$$

For any unit in the temporal base layer,

$$S(0,j) = PSNR(0,j) - PSNR(0,j-1) \tag{3.9}$$

For any unit in the SNR base layer,

$$S(i,0) = PSNR(i,0) - PSNR(i-1,0) \tag{3.10}$$

For all the other units,

$$S(i,j) = \left[PSNR(i,j) - PSNR(i,j-1)\right] - \left[PSNR(i-1,j) - PSNR(i-1,j-1)\right] \tag{3.11}$$

In (3.9)-(3.11), $i = 1,2,...T-1$, $j = 1,2,...,F-1$.

The probability $P(i,j)$ is related to the packet loss rate over wireless packet-erasure channel. The process leading to packet loss is very complex. In this work, we use a two-state Markov model to approximate the wireless channel's packet loss behavior [80]. As described in Section 3.2.3, the Markov model can be calculated by $p(m,N)$, which illustrates the probability of losing *m* packets within *N* packets. As long as the number of lost packets doesn't exceed the number of protection packets, the original data can be reconstructed. Thus $P(i,j)$ can be formulated as,

$$P(i,j) = \sum_{m=0}^{k(i,j)} p(m,N) \tag{3.12}$$

Till now the objective of the optimization problem is to find the proper 2-D channel rate allocation matrix **K**,

$$\mathbf{K} = \begin{bmatrix} k(0,0) & k(0,1) & \cdots & k(0,F-1) \\ k(1,0) & k(1,1) & \cdots & k(1,F-1) \\ \cdots & \cdots & \cdots & \cdots \\ k(T-1,0) & k(T-1,1) & \cdots & k(T-1,F-1) \end{bmatrix} \tag{3.13}$$

Searching of appropriate distribution for **K** should follow:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Constraint 1:

$$\sum_{i=0}^{T-1}\sum_{j=0}^{F-1} h(i,j) \le M \tag{3.14}$$

which restricts the total amount of source bit and channel bit not to exceed the target bit rate. $h(i,j)$ can be calculated as

$$h(i,j) = \left\lceil \frac{B(i,j)}{N - k(i,j)} \right\rceil \tag{3.15}$$

with $B(i,j)$ being the number of source data bytes in the unit (i,j).

Constraint 2:

$$k(i,0) \ge k(i+1,0),\ i = 0,1,\ldots,T-2 \tag{3.16}$$

$$k(i,j) \ge k(i,j+1),\ i = 0,1,\ldots,T-1,\ j = 0,1,\ldots,F-2 \tag{3.17}$$

which confines the relationship between the elements in the matrix **K**. On one hand, in the SNR base layer, units with lower temporal levels are more important and should get more protection. On the other hand, in each temporal layer, units with lower SNR levels should get more protection. This constraint has been justified to be true by many works done before, such as [77] [63].

Now the optimization problem with constraints can be expressed as

$$\max\ PSNR_{\text{overall}}(\mathbf{K}),\ \text{subject to Constraints 1,2} \tag{3.18}$$

3.3.3 Application of Generic Algorithm for Fast Channel Rate Allocation

Exhaustive searching can be applied to solve the maximization problem. However, it is unfeasible in reality because this method will consume a large amount of computation. Hence, we need more efficient algorithms than explicit enumeration. There exist some traditional methods, such as hill-climbing method and local search algorithms, which are

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

often employed to solve the optimization problems. Using these methods, the available channel codes can be appropriately allocated to different layers with clear priority and the protection has a non-increasing or non-decreasing pattern. However, for the proposed scheme, the formulated problem is complicated and the channel rate allocation is actually two-dimensional, including the temporal direction and the SNR direction. Therefore, it is quite difficult to estimate the priority among the 2-D units. It is truly straightforward to claim that $k(i,0) \geq k(i+1,0)$, $i = 0,1,\ldots,T-2$ and $k(i,j) \geq k(i,j+1)$, $i = 0,1,\ldots,T-1$, $j = 0,1,\ldots,F-2$. But for the other cases, it may not be so easy to determine the priority. For instance, given unit $(2,2)$ and unit $(3,1)$, it is hard to tell which one is more important and should be allocated with more protection. Due to the above reasons, these conventional methods cannot be applied here to solve the 2-D channel rate allocation problem.

Generic algorithm (GA) is a solution that can be applied here. It can overcome the limits of the traditional algorithms, which require a continuous search space. GA attempts to mimic the processes in nature, which lead to evolution of higher organisms. In GA, every individual string in the population represents an instance of a solution to the problem being solved. Fitness of an individual population member (i.e. a solution instance) reflects how good the solution is to the problem being solved. In the experiments, standard binary encoding [65] is used for GA. Following is the basic scheme.

Step 1. Initialization.

When GA is started for the first time, every individual string will be given a random value so that an initial population is generated. Each individual string denotes a channel rate allocation matrix **K**. The population size is also decided and will be kept constant for all generations. We use *l* to represent it.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Step 2. Evaluation.

In this step, the fitness of each individual string is calculated. In this work, the fitness represents the quality of the reconstructed video sequence, $PSNR_{\text{overall}}$.

Step 3. Reproduction.

Reproduction is to copy solution strings into a mating pool based on the fitness value ($PSNR_{\text{overall}}$) of them. The strings with higher fitness values will most likely be represented in higher numbers in the mating pool so that their genes may have higher probability to be passed on to the next generation. Several alternative ways are available to implement the reproduction. In this work, the roulette wheel method is applied. Detailed description about the method can be found in [65].

Step 4. Crossover.

The crossover is applied to the mating pool, which is generated in the previous step. It is the process of combining the genes of one string with those of another to create 2 offspring, which will inherit the characteristic of the parent strings and replace them. Before the crossover is performed, we have to decide whether a crossover operation will take place. Crossover probability p_c is used to determine if there will be crossover between 2 parent strings. To create l offspring, we will perform crossover $l/2$ times. We can generate a random vector with $l/2$ elements. Each of the element ranges from 0 to 1. After 2 parents strings are randomly selected, the corresponding random number is compared with p_c. If it is smaller than p_c, the crossover is applied. Otherwise, the parents are passed on as offspring.

Step 5. Mutation.

As in nature, random mutations may occur in the genes of some strings. The mutation will introduce some degree of diversity into the population to prevent a premature

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

convergence. Mutation operation is to change the bit values in the generated offspring strings from “1” to “0” or from “0” to “1”. Let r be the number of bits in the generated offspring. The position of bit for mutation is selected according to the mutation probability, p_m. There are r random numbers generated from 0 and 1, which correspond to the r bits in the string. If a generated value is smaller than or equal to p_m, the bit value is changed at that position.

We continue steps 2-5 until the specified number of generations is complete.

Both the size of population and the number of generations affect the computation complexity of GA. These values are carefully selected through the experiments. In the experiments, the cost time of GA can be neglected compared with the coding time of the video sequence.

3.4 Simulation Results

In this section, we will show the simulation results of channel bits allocation in Subsection 3.4.1 and the performance of the proposed scheme in Subsection 3.4.2. In the experiments, we tested six QCIF sequences: *Football* with 128 frames, *Crew* with 144 frames, *Bus* with 72 frames, *Ice* with 120 frames, *Mobile* with 144 frames and *Harbour* with 144 frames, which have the same frame rate of 15Hz and were coded by MCTF based scalable video codec [82].

3.4.1 Channel Rate Allocation with Implementation of GA

The first work is to select the proper population size and the number of generations.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 3.1
CHANNEL RATE ALLOCATION AT DIFFERENT PACKET LOSS RATES FOR DIFFERENT SEQUENCES: (a) FOOTBALL (b) BUS (c) CREW and (d) HARBOUR

Packet Loss Rate	u(0.0)	u(1.0)	u(2.0)	u(3.0)	u(1.1)	u(2.1)	u(3.1)	u(1.2)	u(2.2)	u(3.2)
5%	70%	75%	76%	76%	87%	89%	86%	89%	92%	89%
10%	68%	71%	76%	76%	88%	83%	81%	94%	96%	92%
20%	60%	71%	71%	71%	97%	97%	81%	97%	100%	98%
30%	53%	75%	76%	76%	97%	97%	94%	97%	100%	98%

(a) Sequence: Football Number of Packets: 160 Packet Size: 180

Packet Loss Rate	u(0.0)	u(1.0)	u(2.0)	u(3.0)	u(1.1)	u(2.1)	u(3.1)	u(1.2)	u(2.2)	u(3.2)
5%	58%	59%	63%	66%	78%	63%	69%	78%	75%	76%
10%	57%	57%	57%	60%	78%	63%	69%	90%	78%	76%
20%	53%	55%	55%	57%	78%	75%	76%	96%	83%	76%
30%	51%	51%	51%	51%	88%	63%	55%	96%	98%	95%

(b) Sequence: Bus Number of Packets: 120 Packet Size: 160

Packet Loss Rate	u(0.0)	u(1.0)	u(2.0)	u(3.0)	u(1.1)	u(2.1)	u(3.1)	u(1.2)	u(2.2)	u(3.2)
5%	64%	77%	77%	83%	88%	77%	92%	92%	84%	96%
10%	61%	77%	79%	93%	88%	83%	93%	92%	92%	96%
20%	56%	77%	88%	93%	88%	88%	93%	92%	95%	96%
30%	55%	74%	88%	93%	88%	88%	95%	92%	96%	96%

(c) Sequence: Crew Number of Packets: 75 Packet Size: 90

Packet Loss Rate	u(0.0)	u(1.0)	u(2.0)	u(3.0)	u(1.1)	u(2.1)	u(3.1)	u(1.2)	u(2.2)	u(3.2)
5%	51%	52%	52%	52%	54%	52%	55%	63%	59%	60%
10%	48%	51%	51%	51%	54%	58%	55%	66%	63%	60%
20%	44%	51%	51%	51%	63%	58%	55%	69%	69%	60%
30%	38%	44%	51%	51%	54%	58%	55%	98%	76%	68%

(d) Sequence: Harbour Number of Packets: 120 Packet Size: 160

In the experiment, six sequences are tested. For the two-state Markov channel, the average packet loss rate is set to 10% and the average burst length is given as 9.57. We specify the number of generations as 500 and discover that the value of the fitness will not change much after 300 generations for all the test sequences. Therefore in the following experiments, the number of generations is given as 300. We also attempt to change the population size l and discover that after $l = 100$ the result will converge. Thus the size of population is specified as 100. The probabilities of crossover and mutation, p_c

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

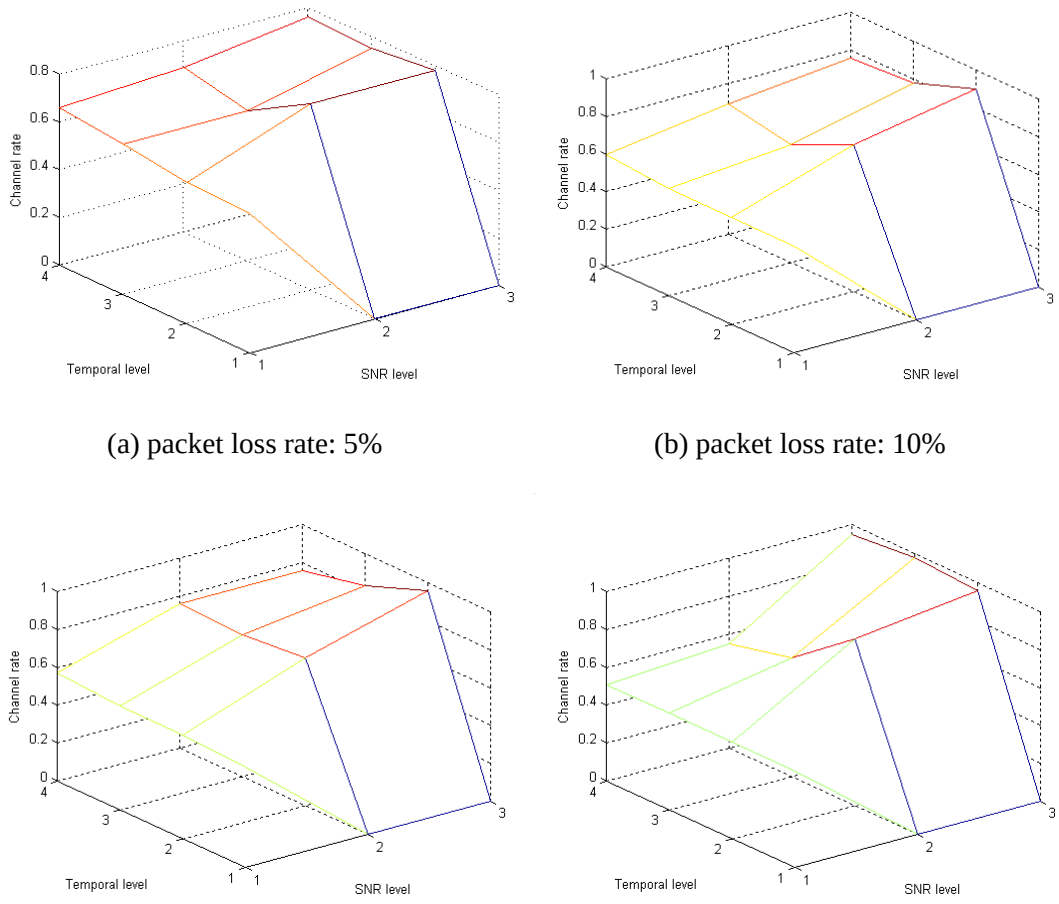
and p_m, are 0.65 and 0.02, respectively. These values are cautiously determined by experiments. All programs were run on an Intel Pentium 4 CPU 3.0G. C language is used for implementation and typically the consumed time for the processing of one group of pictures is about 0.5 s.

Each sequence is divided into several groups with fixed size 8. For each group, there are totally 4 layers in temporal direction and 3 layers in SNR direction. Therefore twelve sub-streams are generated for one group of pictures. $u(i, j)$ is used to denote these twelve units, where i represents the temporal level and j indicates the SNR level, $i = 0,1,2,3$ and $j = 0,1,2$. Units $u(0,1)$ and $u(0,2)$ make use of large numbers of source bits whereas provide small quantity of PSNR increment. It leads to the result that few protection bits are assigned to these units. Thus in the experiment we skip such units and do not take them into account for the allocation problem. Table 3.1 shows the results of channel rate allocation for four sequences in terms of different packet loss rates, 5%, 10%, 20% and 30%, respectively. The channel coding rate defined as the ratio of the source bit rate over the target bit rate [81]. Here the channel coding rate for each unit (i, j) is represented by $R(i, j)$, which is calculated as

$$R(i, j) = \frac{N - k(i, j)}{N} \times 100\% \tag{3.19}$$

The lower the channel coding rate, the more the protection allocated to the unit. From the statistics in Table 3.1 we observe that generally more parity symbols are added to the more important units. We note that in SNR base layer, the channel rate for the temporal base layer is lower than the following temporal layers because the PSNR improvement brought by it is much higher than the others. Meanwhile, the system allocates more protection bits to the former temporal layers. This is due to the predictive relationship between temporal layers. Therefore we can conclude that the channel rate is related to both the PSNR contribution and the temporal level of a unit. On the other hand, in each

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) packet loss rate: 5%

(b) packet loss rate: 10%

(c) packet loss rate: 20%

(d) packet loss rate: 30%

Figure 3.6. 3-D view of the channel rate allocation at different packet-loss rates for sequence "Bus": (a) packet loss rate: 5% (b) packet loss rate: 10% (c) packet loss rate: 20% and (d) packet loss rate: 30%.

temporal layer, more protection bits are assigned to the units with lower SNR levels. One important reason is, the forming of higher SNR layers depends on the lower ones. In addition, compared with the amount of source bits they occupy, the higher SNR layers provide little PSNR increment. It reveals that the channel rate is also related to the SNR level and the source bit occupation of a unit. Of course, it does not mean that a unit coded with more bits will definitely be allocated with a high channel rate since the PSNR increment and its position make much more sense.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Comparing the data under different packet loss rates, we also observe that with the rising of packet loss rate, the protection bits gradually concentrate on the SNR base layer especially on its temporal base layer $u(0,0)$. Fig. 3.6 reveals this tendency in an intuitional way. It can further demonstrate the level of importance for different units.

3.4.2 Performance of the 2-D UEP Scheme

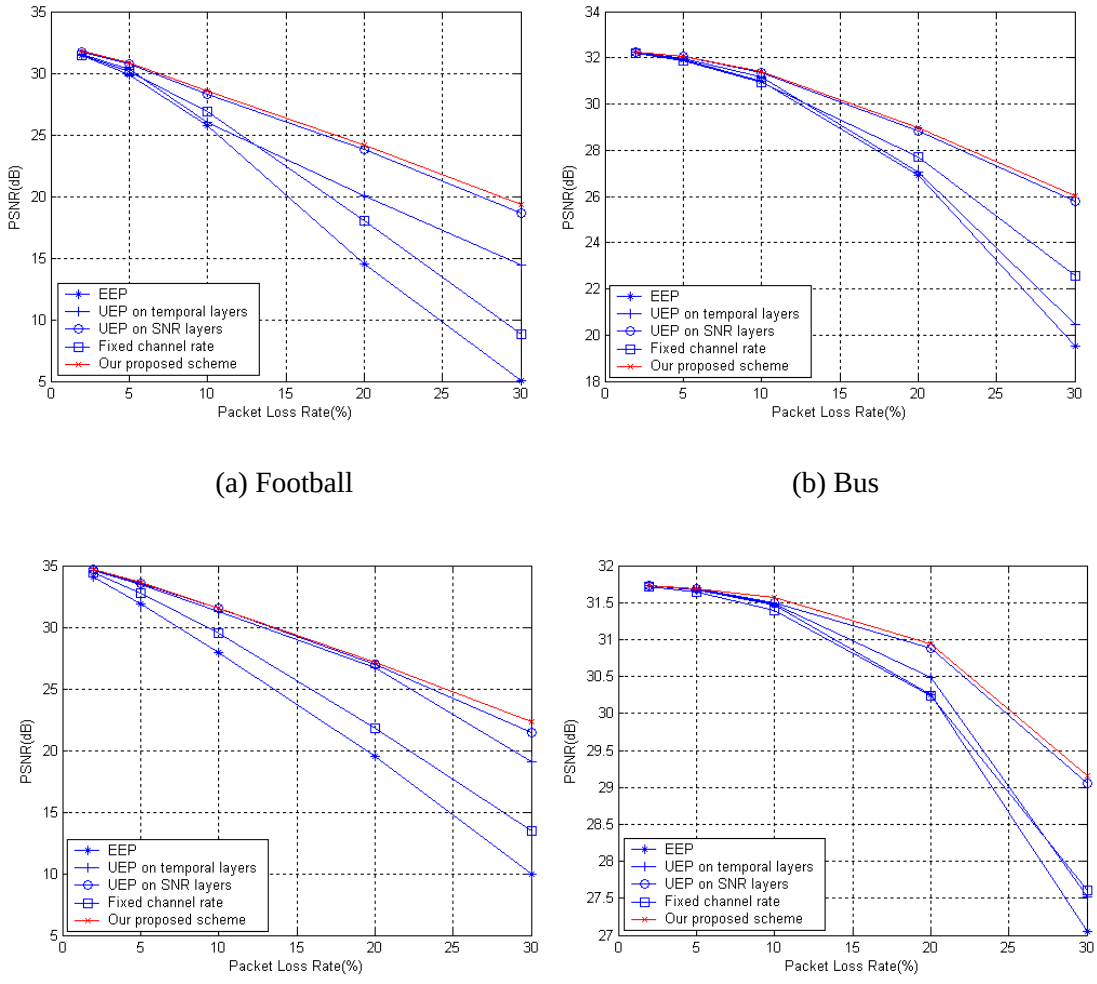
Simulations were performed to transmit video sequences over a Two-state Markov channel. Due to the random nature of such a channel, 100 different runs of the experiments were conducted using different packet loss rate from 2% to 30%. Here we use four sequences with different properties to show our experimental result: *Football*, *Bus*, *Crew* and *Harbour*. Different conditions are prescribed for these sequences. For sequences *Harbour* and *Bus*, $N=120$ and $M=160$. For *Football*, $N=160, M=180$, and for *Crew*, $N=75, M=90$. During decoding of the sequence, a simple temporal replacement is employed as the error concealment method although more sophisticated approaches can be applied.

The performance of our proposed scheme is compared with three other schemes, equal error protection (EEP), UEP on temporal layers, UEP on SNR layers and UEP on 2-D units with fixed channel rate, respectively, over a variety of average packet loss rates.

EEP: equal channel rates are allocated to different units in a GOP without consideration of the error sensitivities of different segments of the bit-stream.

UEP on temporal layers: unequal channel rates are allocated to the units in different temporal layers without consideration of the importance of different SNR layers in each temporal layer. The main difference of this method from the proposed one is that channel rate allocation is found only in the temporal direction.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) Football (b) Bus

(c) Crew (d) Harbour

Figure 3.7. Comparison of the proposed UEP scheme against other three schemes on different video sequences: (a) Football (b) Bus (c) Crew and (d) Harbour.

UEP on SNR layers: unequal channel rates are allocated to the units with different SNR levels without consideration of the importance of different units in the temporal direction. The main difference of this method from the proposed one is that channel rate allocation is found only in the quality/SNR direction.

UEP on 2-D units with fixed channel rate: unequal channel rates are allocated to the 2-D units considering the importance of each unit in both the temporal direction and the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 3.2
COMPARISON OF DIFFERENT UEP SCHEMES ON DIFFERENT VIDEO SEQUENCES:
(a) FOOTBALL and (b) CREW

Packet Loss Rate	PSNR comparison of different methods (dB)				
	EEP	UEP on temporal layers	UEP on SNR layers	Fixed channel rate	Our proposed scheme
2%	31.46	31.56	31.71	31.48	31.79
5%	29.88	30.31	30.75	30.14	30.84
10%	25.77	26.01	28.33	26.92	28.53
20%	14.51	20.07	23.83	18.02	24.20
30%	5.02	14.42	18.61	8.79	19.30

(a) Football

Packet Loss Rate	PSNR comparison of different methods (dB)				
	EEP	UEP on temporal layers	UEP on SNR layers	Fixed channel rate	Our proposed scheme
2%	34.05	34.63	34.68	34.40	34.71
5%	31.86	33.47	33.55	32.75	33.63
10%	27.94	31.25	31.52	29.53	31.57
20%	19.56	26.74	26.99	21.77	27.17
30%	9.97	19.09	21.48	13.45	22.29

(b) Crew

SNR direction. The main difference of this method from the proposed one is that the channel rate allocation is not dynamically changed according to the channel conditions.

The comparison result is illustrated in Fig. 3.7. In contrast, our proposed scheme shows more advantage than the other four schemes for different types of sequences. We can observe that the proposed method exhibits obvious superiority over the EEP, the UEP on temporal layers and the UEP on 2-D units with fixed channel rate, and the improvement is more than 3 dB. Comparing the proposed method with the UEP on SNR layers, our method still gives up to 0.81 dB improvements for some video sequences. We use Table 3.2 to give a detailed comparison for sequences *Football* and *Crew*. Refer to the channel rate allocation in Table 3.1, we can observe that the channel coding rates for different units in the same SNR enhancement layer are very close because of the similar importance of them. In addition we also find that, the SNR base layer is more important than the enhancement layers and most of the channel bits are allocated to it. Therefore,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the channel rate allocation for the units in the SNR base layer will more affect the result. It is noted that for some sequences (e.g., *Bus* and *Harbour*), the channel coding rates for the units in the SNR base layer have small differences. As a result, there are almost equal amount of protection bits allocated to each of them and the protection pattern is close to the UEP on SNR layers. However, for some other sequences (e.g. *Football* and *Crew*), the channel coding rates for these units differ a lot. Therefore, in the experiment this type of sequences has a better performance.

It should be acknowledged that the proposed scheme achieves the improvement of PSNR performance with scarifying of some computational complexity. As elaborated in Section 3.4.1, the complexity of GA greatly depends on the number of generations. For the UEP on temporal layers and UEP on SNR layers, we also apply GA to solve it. The result converges after 100 generations. Therefore, the computation complexity is reduced compared with the proposed scheme. For the EEP scheme and the UEP scheme with fixed channel rate, the complexity is further decreased while the PSNR performance drops a lot.

3.5 Conclusion

Currently a new scalable video coding technique is developed by Heinrich Hertz Institute (HHI). It supports a wavelet based related tool called motion compensated temporal filtering (MCTF). The MCTF based SVC can provide flexibly combined temporal, spatial, SNR and complexity scalability. To the best of our knowledge, the channel rate allocation for the video with combined scalability in the MCTF based SVC has never been considered. In this chapter, a novel 2-D UEP scheme is proposed for this new technology, which can properly allocate the channel protection codes to the combined

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

temporal and SNR scalable units according to the MCTF structure in the temporal direction and the inter-layer dependency between the SNR layers in the quality direction. Given different types of video sequences, the priority of the 2-D scalable units cannot remain the same. Hence, we apply GA to solve the optimization problem efficiently. The scheme is compared with other four methods under different channel conditions for a variety of video sequences. The simulation results demonstrate the advantage of our proposed scheme.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 4

Adaptive Resynchronization Approach for Scalable Video over Wireless Channel

4.1 Introduction

With increasing of the bandwidth in the mobile network, visual communication over wireless channels has become popular and received much attention. However, wireless channels are typically noisy and suffer from various channel degradations such as bit errors, which is caused by small-scale (multipath) and large-scale (shadowing) fades [83]. When compressed video bit-stream is sent over these channels, the effect of channel errors can be very severe. Therefore, delivering quality video over wireless channels is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

really a challenging work and it is highly demanded to develop robust video coding techniques to ensure the quality of the decoded video.

Recently, SVC is developed as scalable extension of H.264/AVC to provide a full scalability including temporal, spatial and quality/SNR scalability with fine granularity. Therefore, it is easy to divide the compressed bit-stream into a number of layers in each scalable dimension. As we discussed in the previous chapter, forward error correction codes can be used to provide protection to the source data. Since the base layer is usually small and of high importance, error-free transmission could be realized for it through high-priority protection. However, the enhancement layer bit-stream may be faced with severe errors when transmitting over error-prone channels. Therefore, the overall quality greatly depends on the enhancement layers.

In this chapter, we aim at improving the error-resilience of the enhancement layers. Various error-resilience techniques can be employed on the compressed bit-stream to improve the robustness of the transmitted video [84]. Among the state-of-art error-resilient techniques, resynchronization is proven to be a very effective tool. In this chapter, we propose an adaptive resynchronization approach to achieve a robust transmission of the enhancement layer information. In each GOP, the enhancement layer bit-stream is separated into a group of units with different temporal levels and quality levels. We measure the importance of each unit and organize them into hierarchical units from the most important one to the least important one. A joint GOP level and picture level resynchronization algorithm is developed to optimally insert resynchronization markers in different units considering both the time-varying channel conditions and the significance of each unit. It is shown from experimental results that the proposed method can perform a graceful degradation under a variety of error conditions and shows advantages over conventional method. We also conduct the experiments to demonstrate that the resyn-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

chronization method can also be employed together with other error-resilient techniques to further improve the quality of the decoded video.

The rest of this chapter is organized as follows. In Section 4.2, we give a brief introduction of the existing resynchronization approaches. In Section 4.3, an overview of the proposed scheme is presented. The optimization problem is formulated mathematically and properly solved. Section 4.4 demonstrates the simulation results for performance comparison. Finally, the conclusion is drawn in Section 4.5.

4.2 Background and Related Works

In coding theory, a variable length code (VLC) is a code which maps source symbols to a variable number of bits. It allows the source data to be compressed and decompressed with zero error and still be read back symbol by symbol. Although VLC can improve the coding efficiency, it may cause the loss of synchronization and a series of erroneous code words due to a single bit error. When the decoder detects an error in a variable length code word, it skips all the forthcoming bits, regardless of their correctness, in the search for the first error-free synch word to recover the state of synchronization. Therefore, the corruption of a single bit is transformed into a burst of channel errors. Residual redundancy in non-compact VLC's can be used to design self-synchronization codes in order to obtain valid symbols again after some slippage [85]. However, even if resynchronization is regained quickly, the appropriate location of the decoded information within the video frame is impossible to be known because the number of missing symbols may not be decided. What's more, the subsequent code words are useless if the information is encoded differentially [84].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

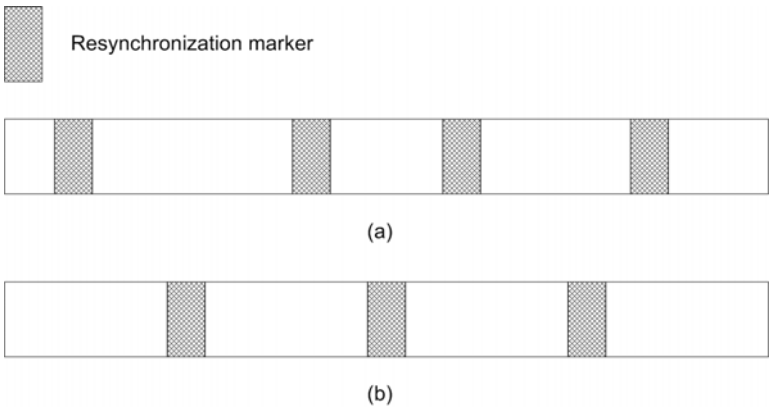


Figure 4.1. Insertion of resynchronization markers into video packets: (a) H.263 and (b) MPEG-4.

To limit the error propagation, resynchronization markers (RMs) are widely used in VLC bit-stream (see Fig. 4.1). Bit patterns of RMs are specially designed and placed at approximately regular intervals in video bit-stream. The aim of insertion of these markers is to divide the compressed bit-stream into independent segments. The decoder can reliably locate each segment without actually decoding the packet by searching for the markers. Therefore, error can be localized in one segment to prevent error propagation across separated segments. However, the insertion of RMs unavoidably introduces some decrease in coding efficiency. To achieve a tradeoff between the overhead to encode the markers and the reliability to detect errors, length of each segment should be appropriately decided [86].

In H.263, resynchronization markers are inserted at certain position in the bit-stream such as the starting point of the group of blocks (GOB) [15]. Once a specific macro-block (MB) location is reached in the encoding process, a resynchronization marker will be inserted into the bit-stream. The disadvantage of this approach is that the RMs are likely to be unevenly spaced because of the variable length coding. As a result, some areas of the picture will be more susceptible to errors. In contrast to H.263, MPEG-4 uses a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

periodic resynchronization approach throughout the bit-stream [17]. To be more specific, the length of a video segment is not determined by the number of MBs as in H.263, but by the number of bits contained in that packet. If the number of bits contained in the current video segment is more than a predefined threshold, a new segment will be created at the start of the next MB [86].

Besides the conventional approaches mentioned above, there are still many literatures considering the insertion of resynchronization markers for the non-scalable video [87]-[92]. Picture level resynchronization is deeply studied to insert resynchronization markers in the bit-stream of a picture. In [87], the resynchronization marker positioning problem is worked over by formulating a cost function mathematically. Later, Lee and Kim [88] propose a method for placement of resynchronization markers based on rate-distortion optimization and the Viterbi algorithm, where the error resilience performance can be substantially improved. Some other resynchronization approaches are based on the data partitioning strategy, which will divide the bit-stream into regions according to their sensitivity to errors. For an instance, in [92], Fang and Chau propose a content-based resynchronization framework to effectively position the resynchronization markers such that the image quality of foreground can be improved at the expense of sacrificing background information. Besides picture level resynchronization methods, GOP level approaches have also been explored. In [90], different sizes of slice are assigned for different pictures in a GOP considering the type and the order of each picture.

The insertion of resynchronization markers has also been applied on scalable video. Yan, et al., design a hierarchical enhancement layer bit-stream structure with resynchronization markers and Header Extension Code (HEC) to achieve a stronger error detection and resynchronization capability [93]. In this scheme, the same number of resynchronization markers is inserted in each bitplane without considering about the significance of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

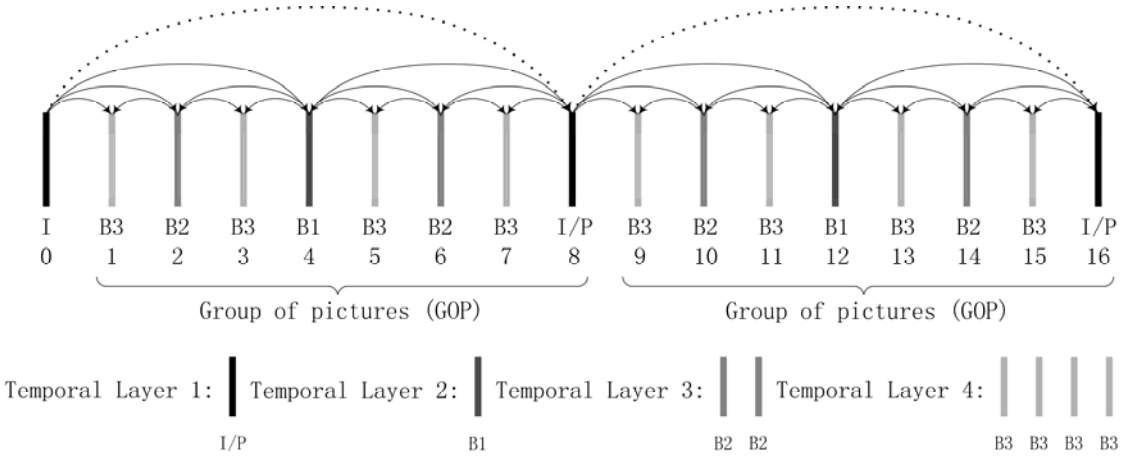


Figure 4.2. Hierarchical B pictures and generation of temporal layers.

different bitplane and different channel condition. In this chapter, we focus on designing an efficient resynchronization approach for the scalable video. Detailed presentation of the algorithm will be given in the following section.

4.3 Proposed Resynchronization Approach

In this section, the structure of the proposed scheme is firstly presented, where the enhancement layer bit-stream in a GOP is divided into a number of units with different temporal and quality levels. After that, a utility based method is introduced to estimate the significance of each unit with the purpose to arrange the enhancement layer units into hierarchical units. Finally, the optimization problem is formulated and properly solved.

4.3.1 System Overview

As introduced in Chapter 2, a key element in SVC is the hierarchical de-composition structure that is applied to realize the temporal scalability [46]. The video can be recon-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

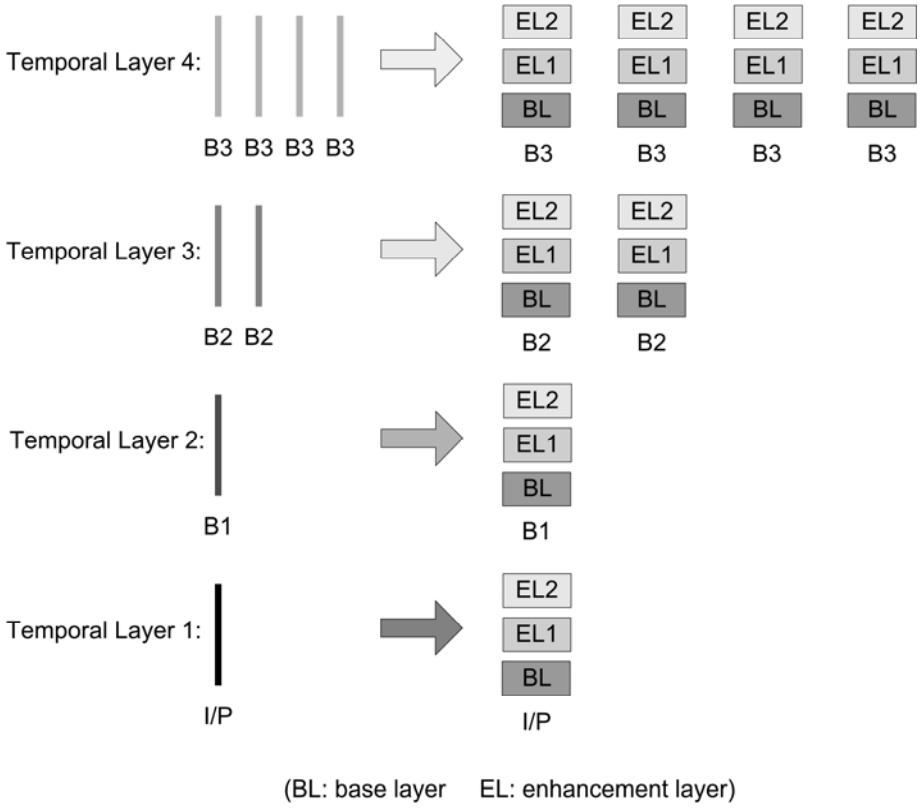


Figure 4.3. Generation of quality/SNR layers of each picture in a GOP.

structed from the temporal base layer, where a minimum temporal resolution or frame rate is achieved. Receiving of temporal enhancement layers will increase the temporal resolution of the decoded video. As illustrated in Fig. 4.2, each GOP consists of eight pictures. Due to the dyadic structure, four temporal layers are generated in each GOP. The pictures of the same type are grouped into a temporal layer.

In this chapter, the compressed bit-stream supports combined temporal and quality/SNR scalability. Both CGS and FGS can be used to achieve quality scalability. The texture of a picture is encoded using a larger quantization parameter to produce a quality base layer, which provides a minimum quality for the decoded video. A refinement of texture information is achieved by re-quantizing the residual signal with a smaller

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

quantization step size relative to that used for the previous quality layer. In this work, we make use of the FGS for encoding of the quality enhancement layers to realize the truncation of the enhancement layers at arbitrary position although the proposed method is also able to work with other coding modes. Fig. 4.3 shows the generation of 3 SNR layers (1 base layer and 2 enhancement layers) for each picture in a GOP.

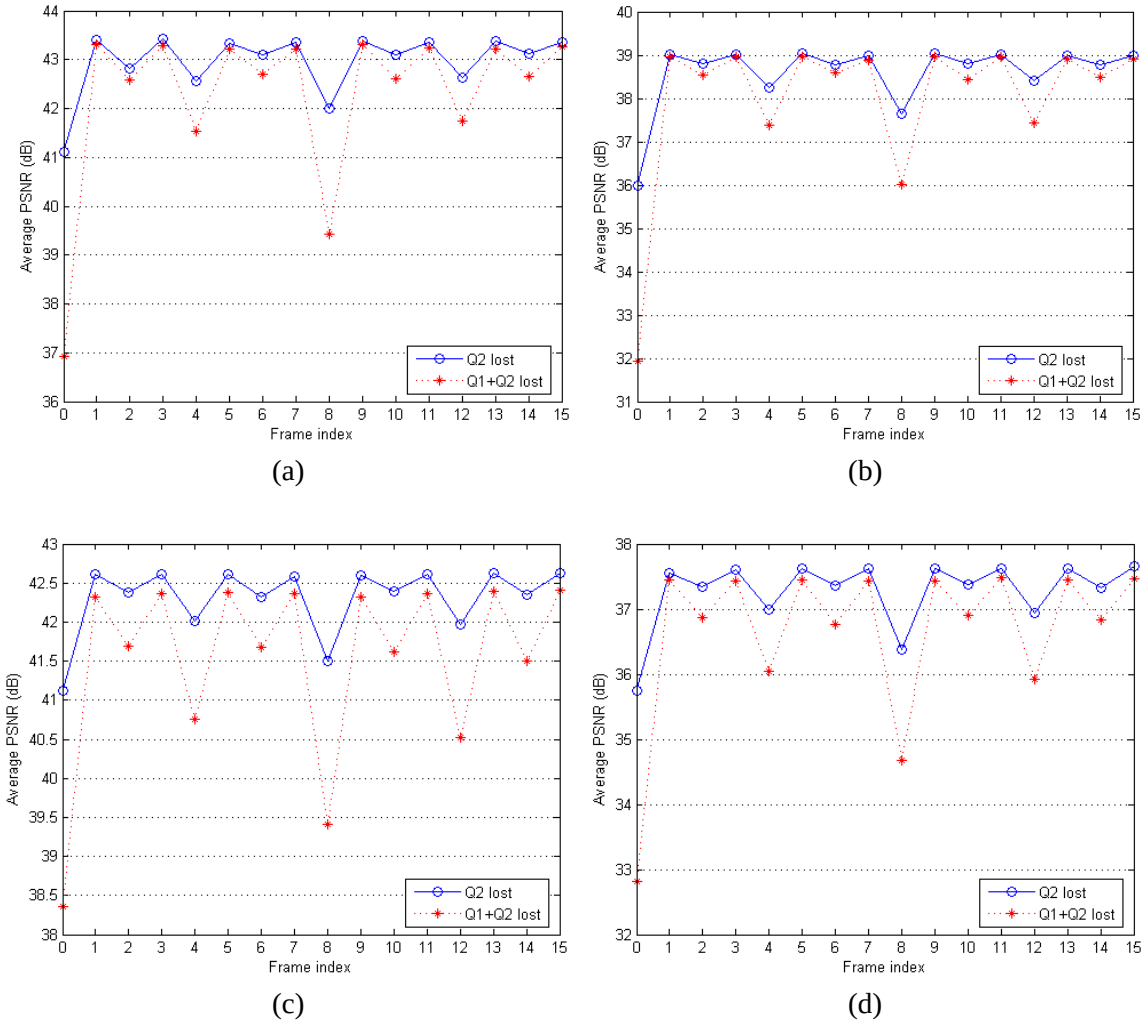


Figure 4.4. Average PSNR of the reconstructed video when different number of enhancement layers of different picture in a GOP is lost: (a) Foreman (347.9 kbps), (b) Foreman (193.5 kbps), (c) Football (786.8 kbps) and (d) Football (486.6 kbps).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

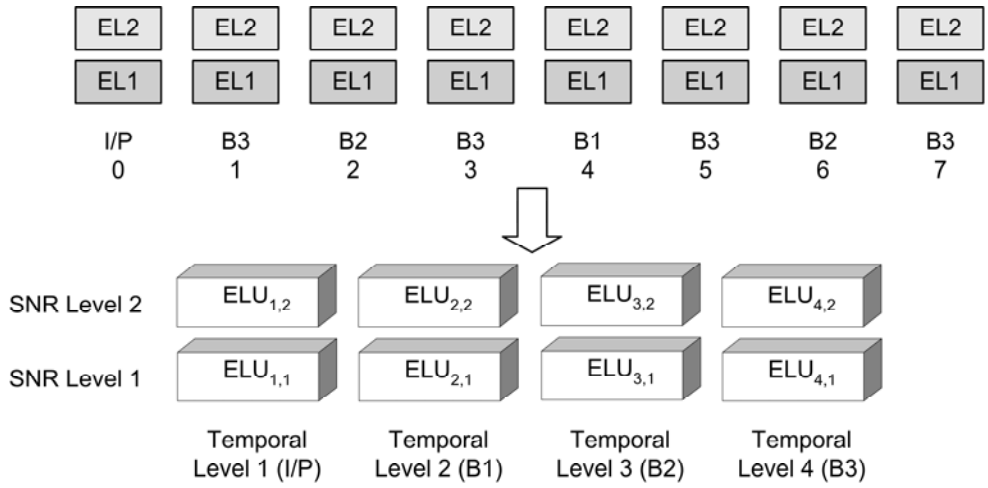


Figure 4.5. Generation of enhancement layer units (ELU) in a GOP.

In the proposed scheme, we assume that an error-free transmission of the base layer information could be realized through high-priority protection. For the GOP-level resynchronization, it is clear that more resynchronization markers should be inserted to the pictures with lower temporal levels. For the picture-level resynchronization, we further allocate different amount of resynchronization markers to different quality layers of a picture, where more markers should be assigned to the lower layers. To optimally insert resynchronization markers in each enhancement layer of each picture of a GOP, the work will be highly complicated. Experiments are conducted to estimate the impairments caused by loss of different slices in different picture in a GOP. Fig. 4.4 shows the average PSNR of the reconstructed video when different number of enhancement layers of different picture in a GOP is lost. “Foreman” and “Football” are used as test videos. There are 96 frames in each video sequence and the GOP size is 16. Each frame is encoded into a quality base layer and two FGS layers, where the enhancement layers are denoted by Q1 and Q2. Each sequence is tested under two different bitrates, which are achieved using QP32 and QP38 respectively. Because of the hierarchical coding structure the impairments caused by loss of information in the enhancement layers of various

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

pictures are different. As shown in Fig. 4.4, the impairments due to loss of information in the same type of pictures are quite similar. Therefore, to save the computational complexity, we group the enhancement layers with the same temporal level and the same SNR level into an enhancement layer unit (ELU). The generation of ELUs in a GOP is depicted in Fig. 4.5. In the figure, for a group of 8 pictures, there are 4 temporal layers created and each picture is encoded into a quality base layer and two quality enhancement layers. All the enhancement layers are combined into ELUs with 4 temporal levels and 2 SNR levels. We use $ELU_{i,j}$ to denote the ELU with temporal level *i* and SNR level *j*.

Now the joint GOP-level and picture-level resynchronization problem is converted to insertion of resynchronization markers to different ELUs within a GOP. There are several factors needed to be considered. First is the channel condition. The limitation of the channel bandwidth makes it an important issue to allocate the channel resources for source coding and resynchronization markers respectively. It is straightforward that with increasing of the channel bit error rate, more resynchronization markers should be inserted to quickly achieve resynchronization and vise versa. It is believed that under different channel conditions, there must be a tradeoff between the bits used for source coding and those used for resynchronization markers. We expect to adaptively and optimally allocate the available bandwidth between the two parts under time-varying channel conditions. Second is the importance of different ELUs. Intuitively, more resynchronization markers should be inserted in the more important ELUs to make the video stream more robust to transmission errors. In Section 4.3.2, the problem of measurement of importance of different ELUs will be addressed.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

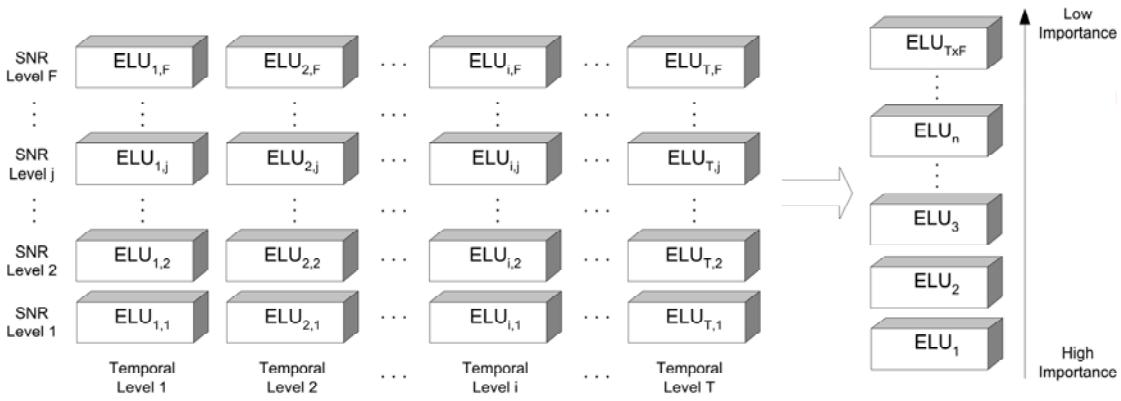


Figure 4.6. Generation of hierarchical units after measurement of importance.

4.3.2 Measurement of Importance of Different ELUs

For the ELUs with the same temporal level, it is obviously that the ELU with a lower SNR level is more important than the higher ones. However, it is not easy to judge the priority of ELUs with different temporal levels. To estimate the importance of a unit, two factors are taken into account in our scheme.

The first factor is utility, which is the impairment caused when the slices belonging to the unit are lost. In this work, the impairment is measured by decreased PSNR due to loss of the information. We use $U_{i,j}$ to denote the utility of the $ELU_{i,j}$, where $i = 1, 2, 3, \ldots, T$ and $j = 1, 2, 3, \ldots, F$.

The second factor is cost, which is the amount of source bits in the unit. We use $B_{i,j}$ to represent the cost of the ELUi,j.

The utility-cost ratio $RT_{i,j}$ is defined to measure the importance of each unit. It is calculated as

$$RT_{i,j} = U_{i,j} / B_{i,j} \tag{4.1}$$

The utility-cost ratio is computed for each unit. For the units with different temporal levels, a larger value of the utility-cost ratio means that the unit contributes more utility

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

while consumes fewer bits. Therefore, a unit with a larger utility-cost ratio is assumed to be more important. For example, if $RT_{2,1} > RT_{1,2}$, $ELU_{2,1}$ is supposed to have a higher priority than $ELU_{1,2}$. Through this utility based method, the significance of each ELU can be properly determined and hierarchical units come into being from the most important one to the least important one as shown in Fig. 4.6. In the figure, the left hand side depicts the ELUs with totally T temporal levels and F quality levels. After measurement of importance of all the ELUs, they are arranged into $T \times F$ hierarchical units from the first important unit ELU_1 to the least important unit ELU_{T*F}, where ELU_n denotes the unit with the nth importance level ($n = 1, 2, 3, \ldots, T \times F$). It is believed that to insert the resynchronization markers optimally to improve the resilience of the bit-stream, we should make the slice size smaller for the more important units while larger for the less important ones.

4.3.3 Formulation of the Problem

In this subsection, we will discuss the problem of inserting resynchronization markers to different ELU of a GOP adaptively and optimally subject to an overall target bit-rate R_{Budget}.

We assume that there are 2^{T-1} pictures in a GOP and each picture is encoded into a base layer and F enhancement layers. With dyadic decomposition structure, ELUs are formed with T temporal levels and F SNR levels. Thus there are totally $T \times F$ ELUs to be transmitted in one GOP. Given the overall coding rate, our objective is to optimally insert resynchronization markers such that the quality of the decoded video is maximized. It also means that the total distortion is minimized. The problem is formulated as

$$\text{Min } D_{overall}, \text{ subject to } R_{overall} \le R_{Budget} \tag{4.2}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The overall bit rate $R_{overall}$ is defined as

$$R_{overall} = R_S + R_{RM} \tag{4.3}$$

with R_S being the source rate and R_{RM} being the rate consumed by resynchronization markers.

The overall expected distortion is defined as

$$D_{overall} = D_S + D_{RM} \tag{4.4}$$

The total distortion $D_{overall}$ consists of two parts. The first one is D_S, which is the distortion due to the loss of source bits caused by channel errors during transmission. It is calculated as follows,

$$D_S = \sum_{i=1}^{T}\sum_{j=1}^{F}\left(P'_{i,j-1}\sum_{l=1}^{k_{i,j}}\delta_{i,j,l}P_{i,j,l} + \left(1-P'_{i,j-1}\right)\delta_{i,j,k_{i,j}} \right) \tag{4.5}$$

In equation (5), i denotes the temporal level and j represents the SNR level. $k_{i,j}$ is the total amount of resynchronization markers in the $ELU_{i,j}$. With insertion of $k_{i,j}$ resynchronization markers, the bit-stream of the unit is separated into $k_{i,j}$ slices. l is the number of slices lost in the $ELU_{i,j}$ while $\delta_{i,j,l}$ represents the average decreased PSNR due to loss of l slices in the $ELU_{i,j}$. $\delta_{i,j,l}$ can be obtained during encoding by calculating the difference between the decoded GOP and the original one. $P'_{i,j}$ is the probability for the first j ELUs with temporal level i to be correctly received from $ELU_{i,1}$ to $ELU_{i,j}$. $P_{i,j,l}$ is the probability that l slices are lost in the $ELU_{i,j}$. $P'_{i,j}$ and $P_{i,j,l}$ are calculated as

$$P'_{i,j} = \begin{cases} 1, & j = 0 \\ \prod_{j'=1}^{j}\left(1-\sum_{l=1}^{k_{i,j'}}P_{i,j',l}\right), & j = others \end{cases} \tag{4.6}$$

$$P_{i,j,l} = \binom{k_{i,j}}{l} SER_{i,j}^{l}(1-SER_{i,j})^{k_{i,j}-l} \tag{4.7}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

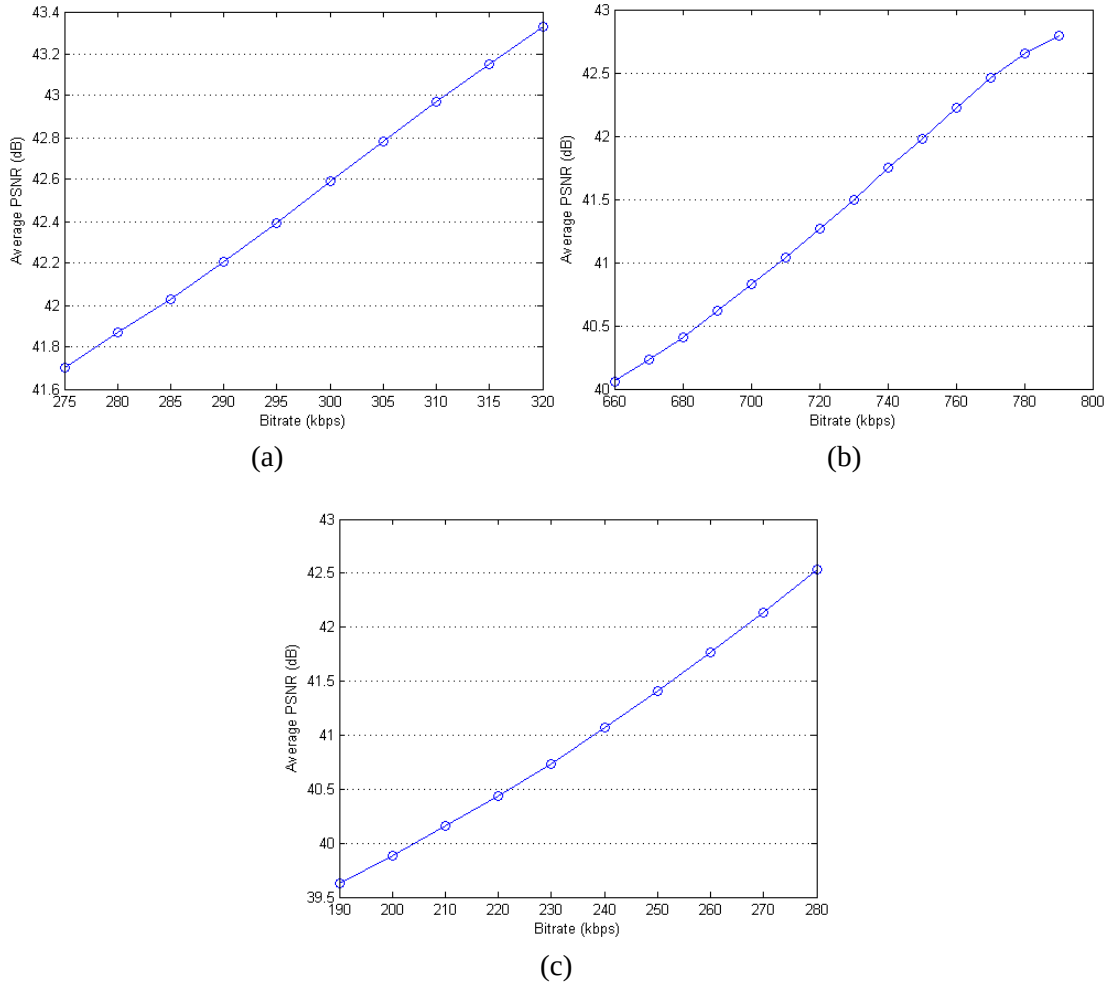


Figure 4.7. Rate-distortion curves of ELU_{T*F} for different video sequences: (a) Foreman (144 frames), (b) Football (128 frames) and (c) City (144 frames).

where $SER_{i,j}$ is the slice error rate, which is the probability that one slice in $ELU_{i,j}$ will be lost. We assume that if an error happens at any location within a slice, the whole slice will be discarded. For a channel modeled with i.i.d. bit errors, $SER_{i,j}$ can be computed as

$$SER_{i,j} = 1-(1-BER)^{b_{i,j}} \tag{4.8}$$

with *BER* being the bit error rate and $b_{i,j}$ being the average number of bits in each slice of $ELU_{i,j}$. Note that we assume the existence of a channel estimator that indicates the bit error rate.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The second part of the distortion, D_{RM}, is resulted from the bits used up by resynchronization markers. D_{RM} is equal to increased distortion due to loss of $\sum_{i=1}^{T}\sum_{j=1}^{F} k_{i,j}$ slices in ELU_{T*F}. Actually this part is not because of loss, but is used on resynchronization markers. It should be noticed that the overall bit rate is limited. During inserting of resynchronization markers, the same amount of source coding bits is required to be removed to satisfy the bandwidth requirement. The distortion due to reduce of the source coding bits is represented by D_{RM}. The more bits consumed by resynchronization markers, the larger the value of D_{RM}. However, the error-resilience performance of the bit-stream will be improved with more resynchronization markers. To minimize the distortion caused by insertion of resynchronization markers, this portion of bits should be taken from the least important unit ELU_{T*F}. Thus, D_{RM} is calculated as the distortion caused by loss of information in ELU_{T*F}, that is

$$D_{RM} = U_{T\times F} \times \left(m \sum_{i=1}^{T}\sum_{j=1}^{F} k_{i,j} \Big/ B_{T\times F} \right) \tag{4.9}$$

where $U_{T\times F}$ is the increased distortion when the whole unit ELU_{T*F} is discarded. $B_{T\times F}$ is the total amount of bits in ELU_{T*F}, respectively. m is the number of bits consumed by one resynchronization marker. In [113], the R-D function of SVC FGS is analyzed and inferred to be linear under MSE criterion within an FGS level. Thus, we assume that D_{RM} can be linearly interpolated from $U_{T\times F}$ as in (4.9). To demonstrate this feature, experiments are carried out on different video sequences. The base layer is encoded using QP32 and two FGS layers are generated as enhancement layers. The rate-distortion curves of ELU_{T*F} for different videos are shown in Fig. 4.7.

With the above deduction, the optimization problem is expressed as

$$\text{Min } D_{overall}(\mathbf{K}), \text{ subject to } R_{overall} \le R_{Budget} \tag{4.10}$$

with

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$\mathbf{K} = \begin{bmatrix} k_{1,1} & k_{1,2} & \dots & k_{1,F} \\ k_{2,1} & k_{2,2} & \dots & k_{2,F} \\ \dots & \dots & \dots & \dots \\ k_{T,1} & k_{T,2} & \dots & k_{T,F} \end{bmatrix} \tag{4.11}$$

As described in Section 4.3.2, all the ELUs can be measured and arranged into $T \times F$ units from the most important unit ELU_1 to the least important one ELU_{T*F}. Therefore, we re-write $\mathbf{K}$ as

$$\mathbf{K} = \begin{bmatrix} k_1 & k_2 & k_3 & \cdots & k_n & \cdots & k_{T\times F} \end{bmatrix} \tag{4.12}$$

where k_n is the number of resynchronization markers in the nth important unit ELU_n.

The problem is deduced to find $\mathbf{K}_{\mathbf{Opt}}$,

$$\mathbf{K}_{\mathbf{Opt}} = \text{argmin } D_{overall}(\mathbf{K}) \tag{4.13}$$

There are two key points in the optimization problem. One is the trade-off between the bits for source coding and those for resynchronization markers. The other is to allocate the available resynchronization markers in different units. By solving these problems properly, we can optimally insert resynchronization markers in each unit and achieve a graceful degradation under time-varying channel condition. During implementation of the algorithm, a constraint needs to be satisfied, which is

$$b_1 \le b_2 \le b_3 \le \cdots \le b_n \le \cdots \le b_{T\times F-1} \le b_{T\times F} \tag{4.14}$$

with b_n being the average number of bits in each slice of ELU_n. The above constraint restricts that the average slice size of the more important unit should not exceed that of the less important one.

Searching for the optimal insertion of resynchronization markers seems to be very time consuming especially when the number of ELUs is very large. To overcome this problem, we develop a simple and effective local hill-climbing algorithm [58], which is briefly described in the following.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1. $\mathbf{K}'^{*} = [1 \quad 1 \quad \cdots \quad 1]$;
2. while $(\mathbf{K}'^{*} \neq \mathbf{K}'_{\mathbf{last}})$
3. $\mathbf{K}'_{\mathbf{last}} = \mathbf{K}'^{*}$;
4. for $n = 1$ to $T * F$
5. for $q = -Q$ to Q
6. $\mathbf{K}'_{\mathbf{temp}} = \mathbf{K}'_{\mathbf{last}}$;
7. $k'_{temp,n} = k'_{temp,n} + q$;
8. if $(k'_{temp,n} < 0$ or $k'_{temp,n} > N)$
9. goto 5;
10. if $(k'_{temp,n} > 0)$
11. for $i = n+1$ to $T * F$
12. $k'_{temp,i} = B'_i / \max(B'_n / k'_{temp,n}, B'_i / k'_{temp,i})$;
13. else
14. for $i = 1$ to $n-1$
15. $k'_{temp,i} = B'_i / \min(B'_n / k'_{temp,n}, B'_i / k'_{tmep,i})$;
16. compute $D_{overall}(\mathbf{K}'_{\mathbf{temp}})$;
17. if $\left(D_{overall}(\mathbf{K}'_{\mathbf{temp}}) < D_{overall}(\mathbf{K}'^{*})\right)$
18. $\mathbf{K}'^{*} = \mathbf{K}'_{\mathbf{temp}}$;
19. $D_{overall}(\mathbf{K}'^{*}) = D_{overall}(\mathbf{K}'_{\mathbf{temp}})$;
20. $\mathbf{K}' = \mathbf{K}'^{*}$;

Figure 4.8. Pseudo code of the algorithm for assignment of resynchronization markers. (N is the maximal number of resynchronization markers in each layer of an ELU. B'_n is the average number of bits in each layer of ELU_n. $\mathbf{K}'^{*}$, $\mathbf{K}'_{\mathbf{last}}$ and $\mathbf{K}'_{\mathbf{temp}}$ are vectors that store the assignments.)

4.3.4 Hill-Climbing Method

Each ELU contains one or several layers depending on its temporal level. Let $\mathbf{K}' = [k'_1 \quad k'_2 \quad \cdots \quad k'_n \quad \cdots \quad k'_{T \times F}]$, where k'_n denotes the average number of resynchronization markers in each layer of ELUn. First, we initialize $\mathbf{K}'$ as $[1 \quad 1 \quad 1 \quad \cdots \quad 1]$. For each iteration, we examine $2QTF$ possible assignments to find certain $\mathbf{K}'$ that can reduce

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the overall distortion, where Q is the maximal number of resynchronization markers that can be added to or subtracted from each layer. The overall distortion is calculated after adding or subtracting 1 to Q resynchronization markers for each layer while satisfying the constraint in (4.14). This process is repeated until $\mathbf{K}'^{*}$ is found, which minimizes the overall distortion.

The pseudo code of the algorithm is given in Fig. 4.8. This algorithm can find a local minimum, which is close to the global minimum with tolerable computation.

4.4 Experimental Results

To test the performance of the proposed scheme, we conduct the experiments on standard video sequences, which are “Foreman” (144 frames), “Football” (128 frames) and “City” (144 frames). The frame rate is 15 Hz and the spatial resolution is QCIF. All sequences are encoded by the scalable video codec [94]. The GOP size is set to 16 for each sequence. Therefore, five temporal layers are generated in each GOP. Furthermore, every picture in a GOP is encoded into a quality base layer and two enhancement layers, where the base layer is compressed using QP32 and the enhancement layers are coded as FGS layers. The ns-2 network simulator [95] was used to study the performance of the proposed algorithm for transmission of the enhancement layers of the scalable video over wireless network. For “Foreman” “Football” and “City”, the available bandwidths are set to 255 kbps, 550 kbps and 235 kbps respectively. Each packet is comprised of 512 bytes. The Gilbert-Elliot bit error model is used to approximate the wireless channel’s bit error behavior. Due to the random nature of such a channel, 50 different runs of the experiments were conducted under each error rate.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 4.1
AVERAGE SLICE SIZE IN EACH ELU UNDER DIFFERENT BER

BER	Average slice size (in bytes)									
	ELU_1	ELU_2	ELU_3	ELU_4	ELU_5	ELU_6	ELU_7	ELU_8	ELU_9	ELU_{10}
10^{-6}	38.85	43.53	44.47	54.97	57.39	96.15	100.51	238.63	300.08	404.94
10^{-5}	37.41	41.45	42.35	54.97	55.92	80.13	85.21	102.27	189.53	404.94
10^{-4}	26.58	33.48	39.82	43.98	54.51	68.68	68.77	75.36	76.62	269.96
10^{-3}	19.8	30.02	36.05	36.65	37.6	40.06	43.56	53.03	62.09	134.98

Currently, there are no error-resilience tools adopted in the joint scalable video model (JSVM). However, we would still like to make a comparison on PSNR performance of our proposed scheme with other schemes. The same with the proposed method, scheme 1 encodes the input video into a number of layers with different temporal and SNR levels. The enhancement layers with the same temporal level and SNR level are grouped into an ELU. The ELUs in each GOP are arranged from the most important unit to the least important one. Different amount of resynchronization markers are inserted to different ELUs. The only difference between scheme 1 and our proposed scheme is that scheme 1 attempts to maximize the total amount of correctly decoded source bits instead of the PSNR of the reconstructed video. For scheme 2, the same amount of resynchronization markers is inserted in each enhancement layer of each frame. One slice consists of three rows of macro-blocks (MBs). In the proposed scheme, to reduce the computational complexity, we confined the minimum slice size to be 3 MBs. To show the performance of the resynchronization method, we also conduct the experiments using SVC when there are no resynchronization markers inserted in the bit-stream. In the experiments, the enhancement layer bit-stream in each GOP is separated into different units. $ELU_{i,j}$ represents the unit with temporal level *i* and SNR level *j*, where i = 1,2,3,4,5 and j = 1,2.

We use “Foreman” to show the average slice size in bytes for different ELUs under various bit error rates in Table 4.1. We can observe that the average slice size of the more

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

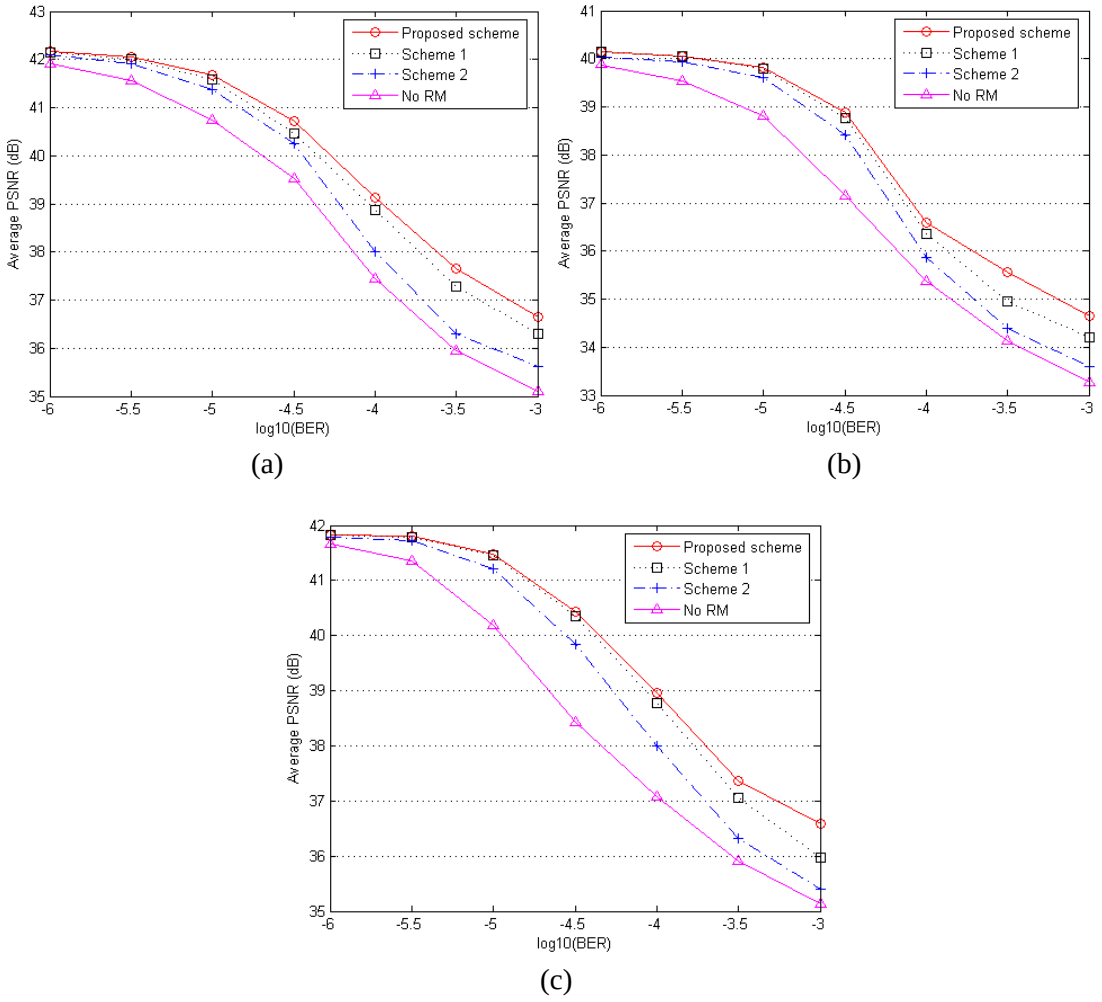


Figure 4.9. Average PSNR of the reconstructed video under different BERs: (a) Foreman, (b) Football and (c) City.

important ELU is smaller than that of the less important one. With rising of the bit error rate, the average slice size for each ELU decreases. This means that more markers are inserted to make the video stream more resilient to the transmission errors.

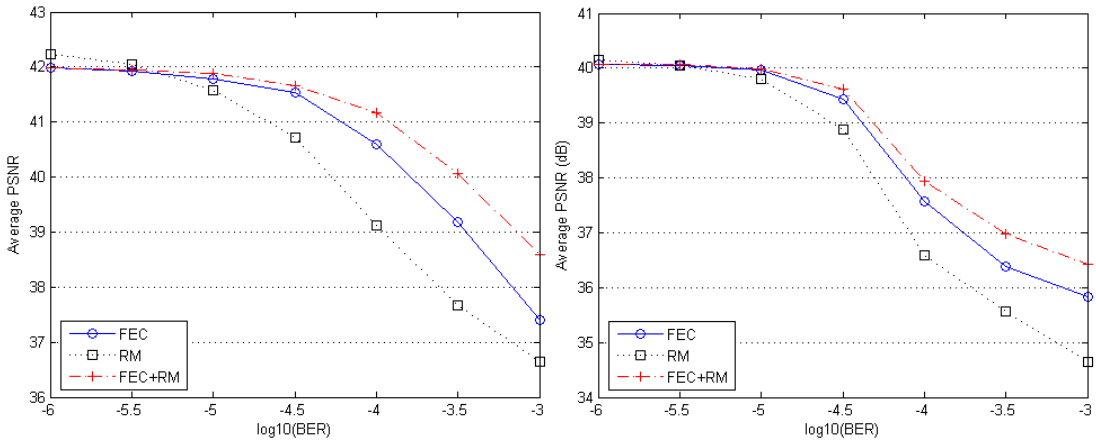
The end-to-end performance of sequences “Foreman” “Football” and “City” is illustrated in Fig. 4.9. The average PSNR of the reconstructed video under the proposed method and the other three schemes are compared. It is obviously that all the schemes that adopt resynchronization markers perform better than the scheme without using resynchronization markers. In contrast, our proposed scheme exhibits superiority over the other

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

three methods under a wide range of bit error rate. When the BER is low, the performance of different schemes is quite similar. However, with increasing of the BER, our proposed scheme can show more advantage than the other three schemes. The average PSNR of the decoded video can be improved more than 1 dB. It should be noted that the performance gain is attained by sacrificing some computational complexity. The complexity of the proposed algorithm has been explained in Section 4.3.4. As scheme 1 is almost the same with the proposed scheme except that it tries to maximize the amount of decodable bits, its complexity is same as the proposed one. For scheme 2, its complexity greatly decreases to $O(1)$. However, its performance also drops a lot.

The proposed resynchronization method can also be adopted together with other error control methods such as forward error correction (FEC) to further improve the end-to-end performance. In the experiments, we employ unequal error protection on different SNR layers as proposed in [96]. The channel coding rates for the two enhancement layers are 75% and 85% respectively. Reed-Solomon (RS) codes are used to generate the FEC codes. To make a fair comparison, the bits used for FEC codes are also taken from the least important unit. We compare the performance of the FEC scheme with the proposed resynchronization method in Fig. 4.10, where “FEC” denotes the FEC scheme and “RM” represents the resynchronization algorithm. As illustrated in Fig. 4.10, when the BER is low, RM is slightly better than FEC. It is because that FEC consumes a lot of bits as redundancy and results in decrease of the source bitrate, which is not necessary when the error rate is not high. However, with increasing of the BER, FEC shows more advantage than RM because it can provide a better protection for the source data. Although the FEC scheme can show good performance, it has unavoidable limitations. Firstly, the generation of RS codes will increase the computational complexity. In addition, for RS (n,k) codes, an FEC encoder must wait for all the k packets before it can generate the redun-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a) (b)

Figure 4.10. Comparison of the performance of different error resilient schemes: (a) Foreman and (b) Football.

dant packets and the receiver must wait for k packets of a block before it can playback the video segment, which will increase delay in the system. Finally, if more than n-k packets of a block are lost, FEC cannot recover any portion of the original segment. This makes FEC useless when the short-term loss rate exceeds the recovery capability of the code. In the experiments, we also employ the resynchronization method together with the FEC scheme. The FEC scheme remains the same as the above one, where the channel coding rates are 75% and 85% respectively for the two enhancement layers. And the bits used for FEC codes are also taken from the least important unit. In addition, the proposed adaptive resynchronization approach is carried out to insert resynchronization markers. It is shown in Fig. 4.10 that the combination of the two schemes is more efficient and can greatly improve the quality of the reconstructed video.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.5 Conclusion

Delivering video over wireless channel is not a trivial task since wireless network typically have harsh channel conditions and time varying throughput. The bit error rate of a wireless channel can range from an ideal case 10^{-5} or a typical case 10^{-4} to a critical case 10^{-3} or instantaneously even higher. In this chapter, a joint GOP level and picture level resynchronization method is proposed for transmission of the scalable video bit-stream with combined scalability over wireless channel. The main aim of the scheme is to adopt error-resilience tool to SVC to enhance the robustness of the compressed bit-stream. After scalable video coding, the input video sequence is encoded into the bit-stream with combined temporal and quality scalability. Different parts of the enhancement layer bit-stream in a GOP have different sensitivity to errors. The scheme groups the enhancement layer bit-stream into a set of units with different temporal levels and quality levels. An efficient method is applied to measure the importance of different units and organize them into hierarchical units from the most important unit to the least important one. The overall distortion is formulated, where a trade-off exists between the distortion caused by channel errors and the distortion due to removal of the source bits used for resynchronization markers. Considering the time-varying channel condition and the significance of different units, a local hill-climbing algorithm is designed to quickly solve this trade-off and optimally allocate the resynchronization markers to different units. With the resynchronization method, the video exhibits robustness against the transmission errors and performs a graceful degradation over error-prone channels. The simulation results demonstrate the efficiency of our proposed method by comparing it with the conventional methods under various error conditions. It is clear that our scheme is superior to the conventional methods and the improvement is up to 1 dB. The proposed resynchronization algorithm can also be adopted together with other error resilient methods to further

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

improve the quality of the reconstructed video, which is proved by the experimental results. It should be noted that the proposed technique is useful when the video decoder sees bit errors. For today's packet switching networks, it will have some limitations. Nevertheless, as network techniques are developed, more and more proposals suggest not to abandon the whole packet (with partially useful information) at the lower layers, but to pass it to the high layers or even application layer, which can save more bits with error correction or other related techniques. More information can be found from Y. Wang, etc [72].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 5

Bit-rate Allocation for Broadcasting of Scalable Video over Wireless Networks

5.1 Introduction

With increasing of the bandwidth in the mobile network, visual communication over wireless channels has become popular and received much attention [115]. Wireless broadcasting enables various mobile users with different platforms to access to the multimedia information simultaneously. A distinctive feature of wireless broadcast system is that the receivers are highly heterogeneous in terms of their bandwidths and processing capabilities. A single transmission rate is unlikely to satisfy the heterogeneous

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

requirements from all the receivers. It is therefore desirable to use multi-rate transmission, in which the receivers can receive video streams at different rates depending on their corresponding bandwidths. Scalable video coding has been shown to be a very attractive solution to solve this problem. As introduced in Chapter 2, it encodes raw video data into a number of layers of different priority. The layer with the highest priority, called the base layer, contains the data with the highest importance, which can provide a minimum video quality. The enhancement layers with lower priorities may be encoded progressively to further refine the quality of the base layer stream. The base layer should be guaranteed to be received by each end-user with very low loss rate. A client can subscribe to all or some of the enhancement layers that best match its bandwidth to improve the quality of the reconstructed video.

Layered transmission has been studied by many researchers [117]-[122]. McCanne et al. firstly proposed the receiver-driven approach for layered video multicast [117]. This approach uses a layered video encoder to generate multiple layers from a single video sequence and transmits each layer over a separate multicast group. The number of layers as well as their bandwidths is predetermined. The adaptation is performed at the receiver's end, where a receiver periodically tries to subscribe to more groups. In the receiver-driven approaches, the sender's coding strategy is predefined, where the source generates a fixed number of layers, each at a fixed rate. The destinations choose from the layers that the source provides. However, the selections may not be adequate enough to optimize the network utilization and the video quality. Therefore, the sender-driven mechanisms were introduced, where the sender can use the feedback information to adjust the coding parameters dynamically to improve both the network utilization and the quality of the video obtained by the end users [120]-[122]. Hsu and Hefeeda applied multilayer scalable coding technique to customize the quality for individual clients in

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[121]. They proposed an algorithm to determine the optimal rate and encoding granularity of each layer in a scalable video stream to maximize a system-defined utility function for a given client distribution. Combining both receiver-driven and sender-driven approaches, Zhang et al. proposed a system for video multicast over Internet [122]. The sender adaptively splits the video data coded by a scalable codec and a channel codec into multiple data streams based on the feedback information on receivers' network parameters. In the meantime, a receiver can estimate the available bandwidth based on a modified packet-pair technique and choose to subscribe to a given part or all of the data streams according to its network conditions.

Another major challenge in video communication is the channel errors. Transmission errors, together with lossy source coding techniques, lead to distortion of reconstructed video at the decoder. Automatic Repeat on reQuest (ARQ) has been widely used to deal with packet loss [116]. Unfortunately, the ARQ-based error control is not realistic for real-time applications due to introduction of delay in the system. As mentioned in Chapter 3, FEC techniques could be employed to reduce the effects of errors on the decoded video quality by adding redundancy codes to the source data [123]-[126]. Lee et al. studied a layered video multicast system with receiver feedback [125]. The layer bandwidth and FEC are properly allocated given the client's heterogeneous bandwidth and error characteristics, subject to a certain overall loss rate requirement. Nevertheless, it simply aims at transmitting each enhancement layer with a target end-to-end loss rate, thus could not optimize the overall system utility. Besides, this method neglects the fact that when burst packet loss occurs in a certain layer and results in loss of the layer, the received packets in the higher layers will become useless. In [126], Schierl et al. presented an approach for wireless video broadcasting using the scalable extension of H.264/AVC with an unequal erasure protection scheme. Different amount of FEC codes are allocated to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

different layers according to their priority to achieve graceful degradation. However, it predetermines the source and channel coding rates without consideration of receivers' statistics.

In this chapter, we develop a new system for wireless broadcasting of scalable video. To cope with the heterogeneity of different receivers, the scalable extension of H.264/AVC, known as SVC, is applied to encode the raw video data into multiple layers including a quality base layer and several enhancement layers. FEC scheme is adopted to generate error protection codes. The proposed joint source and channel coding scheme is designed to maximize the overall system utility and it has the following features. First, the statistics such as the minimum client bandwidth and the maximum packet loss rate are collected to determine the base layer source coding bit-rate and channel coding bit-rate. Through allocation of proper amount of FEC codes, it is guaranteed that the base layer can be correctly received by all the clients with a very low error rate. Second, to deal with the burst packet loss, we propose to apply UEP scheme on the enhancement layers using the interleaved packetization to ensure that a lower layer achieves a higher protection priority. Since the coding structure needs to be continuously adapted given the clients' feedback statistics, fast algorithms are designed to dynamically allocate the source coding bit-rate and the channel coding bit-rate for the enhancement layers to maximize the system-defined utility function. To simplify the problem, a K-means clustering method is carried out to categorize the clients into groups. We also apply an efficient dynamic search algorithm to quickly solve the optimization problem. We implement the algorithm to verify its advantage, and we show how various allocation structures affect the overall utility. A limitation of the proposed scheme compared with [125] is that it introduces more delay to the system due to the interleaved FEC. Besides, because of interleaving of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

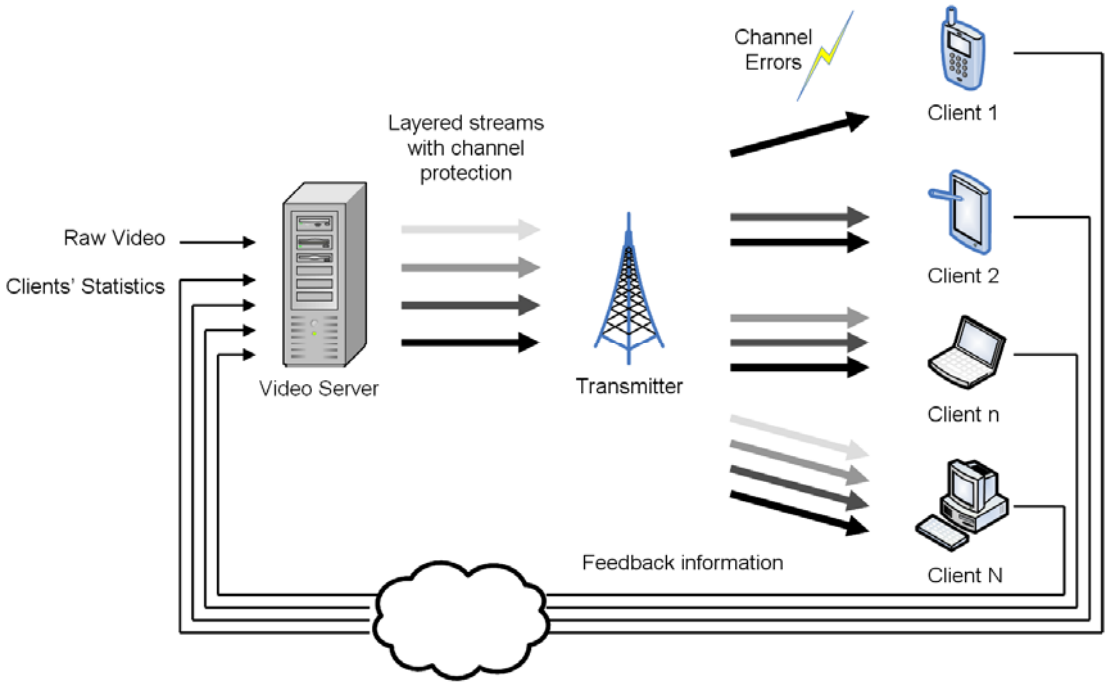


Figure 5.1. Basic framework of the system.

the source data into different packets, receiving of partial packets is useless unless all the source packets can be decoded.

The rest of this chapter is organized as follows. In Section 5.2, we present the architecture of the scheme for video broadcasting. In Section 5.3, the system-level optimization problem is formulated and the proposed algorithm is described. The experimental results are presented and analyzed in Section 5.4. Finally, the conclusion is drawn in Section 5.5.

5.2 System Overview

The basic framework of our proposed system is depicted in Fig. 5.1. Given the input video and the clients' statistics, the server's task is to dynamically encode the raw video

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

into multiple layers using the scalable video coding followed with allocation of channel protection codes to different layers. The layered streams, together with the protection codes, are packetized for transmission over error-prone network to the clients. Each receiver subscribes to all or partial of the packets depending on its bandwidth limitation. The receivers also send sparse feedback information on statistics about their network conditions back to the video server. The most important component in the system is the video server. It needs to analyze the clients' statistics on network parameters such as bandwidth distribution and error conditions and in turn adjust the layered coding structure as well as the channel protection scheme. The joint operations of source coding and channel coding aim at maximizing the system-wide utility for all the clients under their specified network conditions.

In the system, SVC is used for source coding to produce the scalable bit-stream. As introduced in Chapter 2, SVC is the extension of the hybrid video coding approach of H.264/AVC to achieve a wide range of spatio-temporal and quality scalability. Scalability is a functionality that allows the removal of partial stream from the original bit-stream while decoding the video at reduced temporal, SNR or spatial resolution to satisfy the specific rate and resolution required by a certain application. In this work, the quality/SNR scalability is utilized for source coding to make it convenient to adjust the bit-rate in each layer.

To provide robustness for the transmitted video data, the compressed bit-stream will go through the channel encoder. The scalable video coding approach makes it easy to split the bit-stream into multiple layers. It allows transmission of these sub-streams with different protection classes of an UEP transmission profile as proposed in [126]. FEC schemes based on R-S codes are used to protect data transmitted over broadcast channels.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

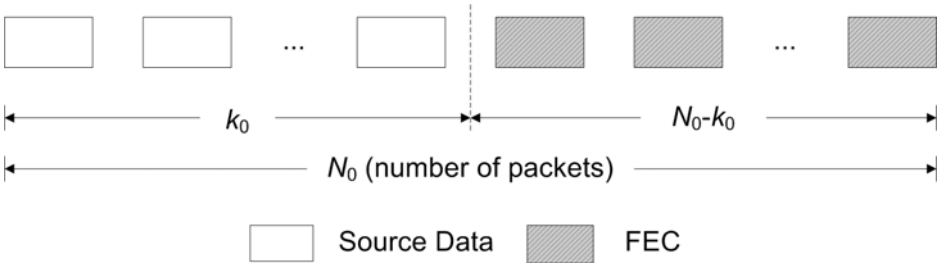


Figure 5.2. Proposed channel protection scheme for the base layer.

Before reading on, it is suggested to review Section 3.2 for a detailed introduction of FEC and UEP.

5.3 Proposed Scheme

In this section, the proposed scheme is described in detail. We design different channel protection architectures for the base layer and the enhancement layers, respectively. The problems are formulated mathematically and we propose algorithms to solve them efficiently.

5.3.1 Bit-rate Allocation for the Base Layer

The base layer provides a minimum quality of video, which should be guaranteed to be received by every client with very low loss rate. Thus, the bit-rate allocated to the base layer should be equal to the minimum end-to-end bit-rate of all the clients. It consists of both the source bits and the channel protection bits. Let C be the total number of clients in the system, $b(c)$ denote the available bandwidth of the cth client, where c ranges from 1 to C. The bit-rate allocated to the base layer, r_0, is determined as

$$r_0 = \min_{1 \le c \le C} b(c) \tag{5.1}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The channel coding scheme for the base layer is illustrated in Fig. 5.2. In the figure, each rectangle denotes a packet. A block of packets is comprised of N_0 packets, in which k_0 of them are source packets. A packet-level FEC is applied on the source packets to generate $N_0 - k_0$ parity packets. Due to the property of R-S code, up to $N_0 - k_0$ packet losses in a block can be corrected. With the R-S (N_0, k_0) code, the video source bit-rate $r_{S,0}$ and the channel coding bit-rate $r_{C,0}$ are computed as

$$r_{S,0} = r_0 \left(\frac{k_0}{N_0} \right) \tag{5.2}$$

$$r_{C,0} = r_0 \left(\frac{N_0 - k_0}{N_0} \right) \tag{5.3}$$

Given each client's end-to-end bit-rate and packet loss rate, the server has to decide the source coding bit-rate and the channel coding bit-rate to make the base layer received by all the clients in the system with a very low loss rate, which is no more than ε. ε is defined as the residual loss rate after error correction.

We use $P_b(c)$ to define the probability that the base layer is lost for client c. The probability is related to the packet loss rate over wireless packet-erasure channel. The process leading to packet loss is very complex. In this work, we assume the existence of a channel estimator that indicating the probability that a particular number of packets are lost, given the total number of packets to be transmitted. As presented in Section 3.2.3, this estimator could be formulated as any distribution of expected packet-loss rate, such as uniform, binomial, exponential, Zipf, Poisson, etc. Among such distributions, a two-state Markov model approximates the wireless channel's packet loss behavior fairly well [80]. The Markov model can be calculated by $p(m, N)$, which illustrates the probability of losing m packets within N packets. As long as the number of lost packets does not exceed the number of protection packets, the original data can be reconstructed. Therefore, $P_b(c)$ can be formulated as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

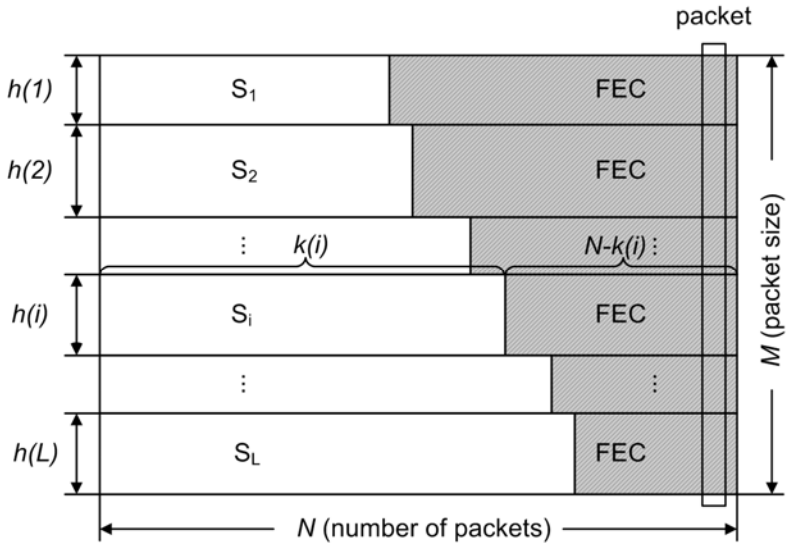


Figure 5.3. Proposed channel protection scheme for the enhancement layers.

$$P_b(c) = \sum_{m=N_0-k_0+1}^{N_0} p_c\left(m, N_0\right) \tag{5.4}$$

where $p_c(m, N_0)$ is the probability for client c to lose m packets within N_0 packets. N_0 is calculated as

$$N_0 = \left\lfloor r_0 \cdot \left(\frac{T}{F}\right) \Big/ 8M \right\rfloor \tag{5.5}$$

Let $P_{b\_\max}$ be the maximum probability for the base layer to be lost among all the clients, which is

$$P_{b\_\max} = \max_{1 \le c \le C} P_b(c) \tag{5.6}$$

The optimization problem is formulated as

$$k_0^* = \max \quad k_0 \tag{5.7a}$$

$$\text{subject to} \quad k_0 \le N_0; \tag{5.7b}$$

$$P_{b\_\max} \le \varepsilon. \tag{5.7c}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The searching of the maximum k_0 is for the purpose of maximizing the source data to achieve a better video quality. Once k_0^* is computed, the source bit-rate and the protection bit-rate of the base layer can be calculated using (5.2) and (5.3).

5.3.2 Bit-rate Allocation for the Enhancement Layers

All the clients can access to the base layer information and those clients with higher bandwidths may subscribe to more enhancement layers to improve the perceived video quality. There exists dependency between enhancement layers. Once a layer is lost, all the higher layers are useless. Therefore, we apply UEP on different enhancement layers of a GOP as illustrated in Fig. 5.2. There are totally L enhancement layers, where S_i denotes the source data for the ith layer ($i = 1, 2, \ldots, L$). In each block of the source data, the source bytes are inserted line by line from the upper left to the lower right. R-S based FEC codes are added as redundancy for protection of the source data. More protections are allocated to the lower layer while less for the higher ones. Both the source data and the FEC codes are vertically split into N packets with M being the size of each packet in byte. The length and the height of the source data in S_i are denoted by $k(i)$ and $h(i)$, respectively. Thus, the length of the FEC codes for S_i is $N - k(i)$. Let $\mathbf{R_S}$ and $\mathbf{R_C}$ denote the source coding and channel coding vectors for the enhancement layers, $\mathbf{R_S} = \left[r_{S,1}\ r_{S,2} \cdots r_{S,L} \right]$ and $\mathbf{R_C} = \left[r_{C,1}\ r_{C,2} \cdots r_{C,L} \right]$, where $r_{S,i}$ and $r_{C,i}$ are the source coding bit-rate and the channel coding bit-rate of the ith enhancement layer respectively. The calculation of $r_{S,i}$ and $r_{C,i}$ are as follows,

$$r_{S,i} = 8k(i) \cdot h(i) \Big/ \left(\frac{T}{F} \right) \tag{5.8}$$

$$r_{C,i} = 8\left(N - k(i) \right) \cdot h(i) \Big/ \left(\frac{T}{F} \right) \tag{5.9}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

with T being the number of frames in a GOP and F being the frame rate of the video. Given the clients' statistics, such as the available bandwidth and the packet loss rate, the server has to determine $\mathbf{R}_S$ and $\mathbf{R}_C$ in order to maximize the system-wide utility function, which is formulated as

$$U_{avg} = \frac{1}{C}\sum_{c=1}^{C} u(c) \tag{5.10}$$

where U_{avg} is the average utility over all the clients and $u(c)$ represents the utility received by client c, $c = 1, 2, \ldots, C$. The proposed scheme can work with any form of user-defined utility function, for instance, client perceived quality in terms of PSNR and mismatch between received stream rate and client bandwidth. In this work, the PSNR of the perceived video is used to measure the utility of each client.

Given the server's total sending bit-rate R and the bit-rate used for base layer coding, r_0, the available bit-rate for the enhancement layers, r_e, is computed as

$$r_e = R - r_0 \tag{5.11}$$

The optimal allocation problem for the whole system is formally stated as follows,

$$U_{avg}^{*} = \max_{\mathbf{R}_S, \mathbf{R}_C} \quad U_{avg} = \frac{1}{C}\sum_{c=1}^{C} u(c) \tag{5.12a}$$

$$\text{s.t.} \quad \sum_{i=1}^{L} \left(r_{S,i} + r_{C,i} \right) \le r_e; \tag{5.12b}$$

$$k(1) \le k(2) \le \cdots \le k(L). \tag{5.12c}$$

The first constraint restricts the total amount of bit-rate in the enhancement layers, including source coding bit-rate and channel coding bit-rate, not to exceed the budget bit-rate. The second constraint confines the protection priority to be non-increasing for the layers from low to high. Although the function is not difficult to design, in practice it is not easy to adjust the sender's parameters in order to obtain the optimal allocation.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

To simplify the problem, we apply a classification method to partition all the clients into several groups based on their available bandwidths. Given the number of layers in the scalable video stream, the total of C clients are clustered into $L+1$ groups with L being the number of enhancement layers according to their accessing bandwidth. Here, we assume that C is far more than L. We use $b_{\min}(i)$ to represent the lowest bit-rate of all the clients in the ith group, where $i = 0,1,2,\ldots,L$. $b_{\min}(i)$ serves as the data rate of group i. We assume that $\forall\, i = 0,1,2,\ldots,L-1$, $b_{\min}(i) < b_{\min}(i+1)$, which implies that a client in the group with a higher index is able to access to more packets or more enhancement layers. As illustrated in Fig. 5.3, if a client can receive up to $k(i)$ packets, it can decode the ith enhancement layer and all the lower layers, where $i = 1,2,\ldots,L$. To simplify the problem, we prescribe the relationship between $b_{\min}(i)$ and $k(i)$ as,

$$k(i) = \left\lfloor b_{\min}(i) \cdot \left(\frac{T}{F}\right) \Big/ 8M \right\rfloor \tag{5.13}$$

It guarantees that even the client with minimum bandwidth in the ith group can access to $k(i)$ packets and the client with higher bandwidth can access to more packets. It also ensures that even the client with maximum bandwidth in the ith group cannot access to $k(i+1)$ packets, therefore cannot decode the (i+1)th enhancement layer. For the first group of clients, they can only access to the base layer. Thus, it will not be considered in the bit-rate allocation for the enhancement layers. We use U'_{avg} to denote the average utility of all except for the clients in the first group. Client (i, j) is used to represent the jth client in the ith group. Let $g(i)$ be the number of clients in the ith group, U'_{avg} is formulated as

$$U'_{avg} = \frac{1}{\sum_{i=1}^{L} g(i)} \sum_{i=1}^{L} U_G(i) \tag{5.14}$$

where $U_G(i)$ is the overall utility in the ith group, which is computed as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$U_G(i) = \sum_{j=1}^{g(i)} u(i,j) \tag{5.15}$$

with $u(i,j)$ being the utility of client (i,j). The calculation of $u(i,j)$ is shown as follows,

$$u(i,j) = \delta(0) \cdot P(i,j,0) + \left(\sum_{l=1}^{i} \delta(l) \cdot P(i,j,l) \right) \cdot P(i,j,0) \tag{5.16}$$

where $\delta(0)$ is the utility contribution of the base layer and $\delta(l)$ is the utility contribution of the lth enhancement layer. Both $\delta(0)$ and $\delta(l)$ are dependent on the source bit-rate in the corresponding layer. We assume that the rate-distortion function of the scalable video codec in each layer is known. $P(i,j,0)$ denotes the probability for the base layer to be received by client (i,j) and $P(i,j,l)$ represents the probability for client (i,j) to correctly receive the lth enhancement layer. The calculation of $P(i,j,0)$ and $P(i,j,l)$ are as follows,

$$P(i,j,0) = \sum_{m=0}^{N_0 - k_0} p_{i,j}(m, N_0) \tag{5.17}$$

$$P(i,j,l) = \sum_{m=0}^{N_{i,j} - k_l} p_{i,j}(m, N_{i,j}) \tag{5.18}$$

where $l = 1, 2, \ldots, L$. $p_{i,j}(m, N_0)$ is the probability for client (i,j) to lose m packets within N_0 packets and $p_{i,j}(m, N_{i,j})$ is the probability for client (i,j) to lose m packets within $N_{i,j}$ packets. As described in the previous sub-section, $p_{i,j}(m, N_0)$ and $p_{i,j}(m, N_{i,j})$ can be achieved using the two-state Markov model. $N_{i,j}$ is the maximum number of packets that can be received by client (i,j), which should be computed as

$$N_{i,j} = \left\lfloor b(i,j) \cdot \left(\frac{T}{F} \right) \Big/ 8M \right\rfloor \tag{5.19}$$

with $b(i,j)$ being the available bandwidth of client (i,j). From the above deduction, it is shown that the changing of $h(i)$ will result in the variation of source coding bit-rate of each layer and consequently affect the overall utility of the system. Therefore, to maxi-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

mize the overall utility, we should look for the proper value of each $h(i)$. The problem is re-formulated as follows,

$$U'^{*}_{avg} = \max \quad U'_{avg} \tag{5.20a}$$

$$\text{s.t.} \quad \sum_{i=1}^{L} h(i) \le M; \tag{5.20b}$$

$$0 \le h(i) \le M, i = 1, 2, \ldots, L. \tag{5.20c}$$

Exhaustive searching can be applied to solve the maximization problem. However, it is unfeasible in reality because of the large amount of computation consumption. Hence, we need more efficient algorithms than explicit enumeration. In this work, we employ a dynamic programming algorithm to solve the problem. Let $U_{overall}(l)$ denote the cumulative utility for the clients from group *1* to group *l*, which is calculated as

$$U_{overall}(l) = \sum_{i=1}^{l} U_G(i) \tag{5.21}$$

Obviously, our purpose is to attain $U^*_{overall}(L)$. The overall utility for the l groups, $U_{overall}(l)$, is the sum of the overall utility in the lth group and the first $(l-1)$ groups. Thus, $U^*_{overall}(L)$ can be computed recursively with the following dynamic program by searching for $h(i)$.

$$U^*_{overall}(L) = \max_{\substack{0 \le h(L) \le M \\ \sum_{j=1}^{L} h(j) \le M}} \left((U_G(L) + U^*_{overall}(L-1) \right) \tag{5.22a}$$

$$\vdots$$

$$U^*_{overall}(L-i) = \max_{\substack{0 \le h(L-i) \le M \\ \sum_{j=1}^{L-i} h(j) \le M}} \left((U_G(L-i) + U^*_{overall}(L-i-1) \right) \tag{5.22b}$$

$$\vdots$$

$$U^*_{overall}(1) = \max_{0 \le h(1) \le M} U_G(1) \tag{5.22c}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Dynamic programming is a method of solving complex problems by breaking them down into simpler sub-problems in a recursive manner. It is applicable to problems that exhibit the properties of overlapping sub-problems, which are slightly smaller and optimal substructure. In the above approach, an overview of the system is first formulated. Each sub-system is then refined in greater detail until the entire specification is reduced to base elements. And then we try to solve the sub-problems first and use their solutions to build on and arrive at solutions to bigger sub-problems. Detailed descriptions on dynamic programming algorithm can be found in [129]. Dynamic programming algorithm can greatly save the computation complexity. In each recursive step in the above dynamic program, there are $O(M)$ possibilities of $h(i)$. Therefore, the search space of the maximization problem is $O(ML)$, which is much lower than the exhaustive search algorithm.

5.4 Experimental Results and Discussions

In this section, we first describe our experimental setup. Then we evaluate the performance of the proposed method by comparing it against other methods.

In our system, there are a large number of clients that are heterogeneous in bandwidth and error rate. The server uses the proposed algorithm to determine the coding structure of the bit-stream based on the clients' statistics. The output of the algorithm is the bit-rate allocation for different layers including the source coding bit-rate and the channel coding bit-rate. The information will be fed to the scalable video encoder, which divides the input video into GOPs and encodes each frame into multiple quality layers with the predetermined coding rates. Then the compressed source bit-stream goes through

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

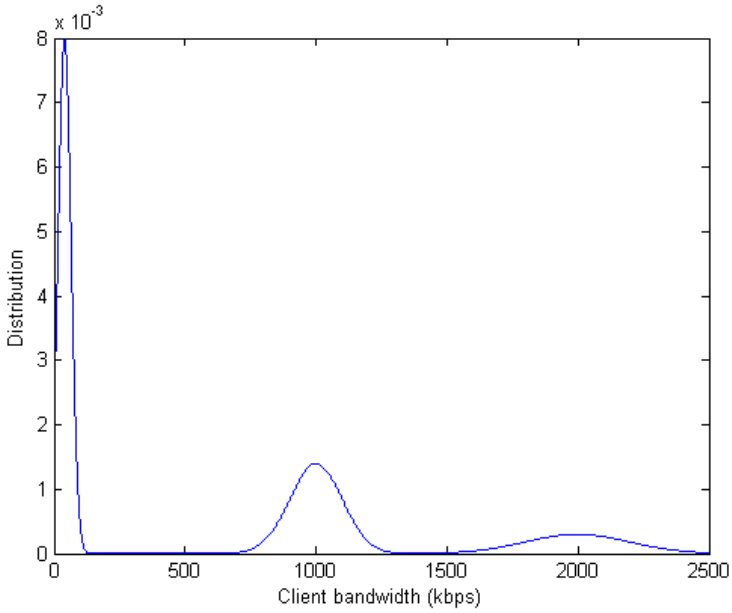


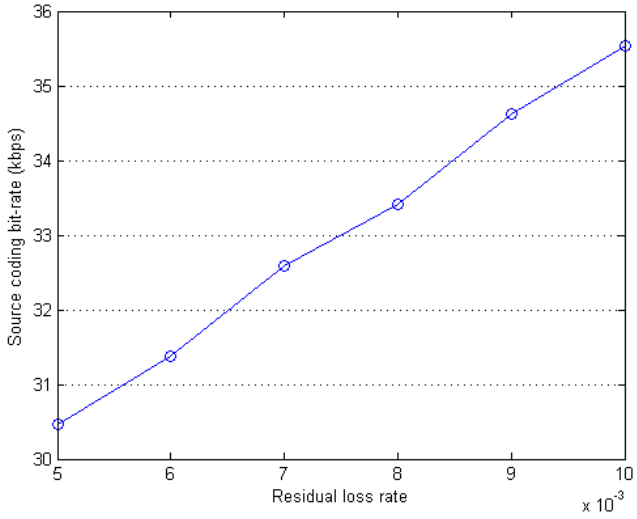
Figure 5.4. Client bandwidth distribution considered in the experiment.

a channel encoder to produce the FEC codes. Both the source codes and the channel codes are packetized for transmission. The packet size used in this work is 128 bytes.

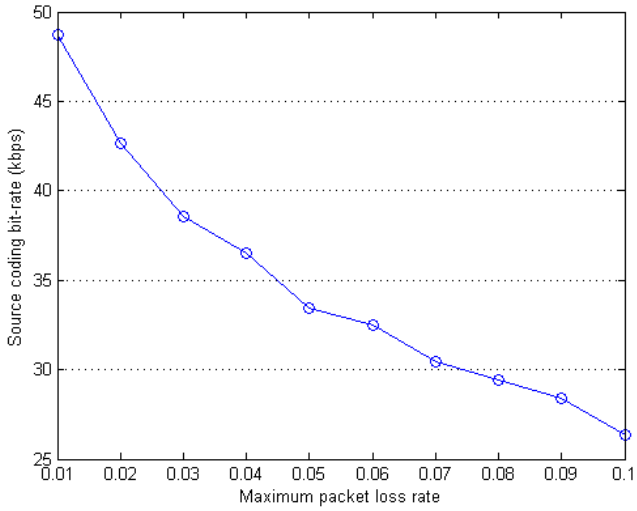
We assume that there are totally 10000 clients in the system, where their bandwidths range from 64 to 2048 kbps. The bandwidth distribution is depicted in Fig. 5.4. It follows a multimodal distribution with three normal distributions [121]: 50% of clients have a normal distribution with mean 40 kbps and a standard deviation of 25 kbps; 35% of clients have a normal distribution with mean 1000 kbps and a standard deviation of 100 kbps; and 15% of clients have a average bandwidth 2000 kbps and a standard deviation of 200 kbps. The packet loss rate of each client is randomly generated, where the mean value is set to 0.03 and the variance is given as 0.015. All the clients are categorized into a given number of groups by using the K-means classification method. K-means clustering is a method of cluster analysis, which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. In this work, we

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

apply the Kaufman approach [131] to partition the clients into groups given their accessing bandwidth, which means the clients with similar bandwidths are clustered into the same group. More details about the algorithm can be found in [131]. The server's sending data rate is 3000 kbps, which should be properly allocated to the source coding and the channel coding of each layer.



(a)



(b)

Figure 5.5. Two factors that affect the source coding bit-rate of the base layer (r_0 = 64 kbps): (a) Source coding bit-rate versus residual loss rate ε (the random packet loss rate has an average of 0.03 and a variance of 0.015) and (b) Source coding bit-rate versus maximum packet loss rate (ε = 0.008).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The video sequence Foreman is used for testing, which is encoded at 15 Hz in CIF format using the scalable video codec [94]. The GOP size is fixed at 16. Experiments are performed to transmit the video sequence over a two-state Markov channel. Due to the random nature of such a channel, 50 different runs of the experiments are conducted.

The available bandwidth for the base layer is the minimum receiving bit-rate among all the clients, which is 64 kbps in the experiment. It consists of both the source coding bit-rate and the channel coding bit-rate. Fig. 5.5 illustrates the results for the source coding bit-rate of the base layer. There are two factors that affect the result. One is the desired residual loss rate after channel protection, the other is the maximum packet loss rate among all the clients. To test the influence of the first factor, we let the random packet loss rate has the average of 0.03 and the variance of 0.015. As shown in the Fig. 5.5 (a), with increasing of the desired residual loss rate ε, the source coding bit-rate arises. It is because that to meet the requirement of a smaller ε, more bits should be allocated to the channel coding and this results in decrease of the source coding bit-rate. As for the second factor, we fix the value of ε to be 0.008. We find in Fig. 5.5 (b) that with increasing of the maximum packet loss rate, fewer bits are assigned to the source coding. It is due to the reason that more bits are needed for the channel coding to attain the desired residual loss rate when the maximum packet loss rate becomes higher.

To test the performance of the proposed method for the enhancement layers, we vary the total number of enhancement layers L from 1 to 4. Under each L, the proposed algorithm is carried out to determine the coding structure that maximizes the average utility in the system. For coding of the base layer, the residual loss rate ε is set to 0.008. The source coding bit-rate of each layer and the average PSNR of all the clients are displayed in Table 5.1. Obviously, the overall utility increases with increasing of L. This

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 5.1
CODING STRUCTURE OF THE SOURCE DATA AND THE AVERAGE PSNR OF THE RECONSTRUCTED VIDEO

Total number of layers	Source coding bit-rate (kbps)					Average PSNR (dB)
	r_{s_0}	r_{s_1}	r_{s_2}	r_{s_3}	r_{s_4}	
2	33.4	1506.5	X	X	X	32.49
3	33.4	208.8	1568.8	X	X	34.38
4	33.4	305.8	614.4	1178.8	X	35.57
5	33.4	498.4	692.8	935.6	1130.4	35.98

is because that a larger number of layers can better adapt to the heterogeneous bandwidth distribution to improve the system-wide utility.

Fig. 5.6 gives the comparison of the proposed method against the other two schemes. For all these three schemes, the base layer is encoded using the same structure to guarantee that it can be received by all the clients with a very low error rate. In scheme 1, UEP is also applied on different enhancement layers. However, the height of the source data is equally allocated to different layers. In scheme 2, traditional method is employed, where equal error protection is added to different layers with fixed coding rate. During increasing of the total number of layers, the resulted system utility of scheme 2 slightly decreases due to the increase of overhead information for all the layers. From Fig. 5.6, we can observe that the variation of the coding structure can greatly affect the overall utility and the proposed method exhibits obvious superiority over the other two schemes. When there is only one enhancement layer, all three schemes have the same performance because they use the same coding structure. With increasing of the number of layers, the proposed method shows more advantage than the other two methods because it can better utilize the available bandwidth. The improvement is up to 2 dB. We would like to point out that there is a tradeoff between the performance gain and the computational complex-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

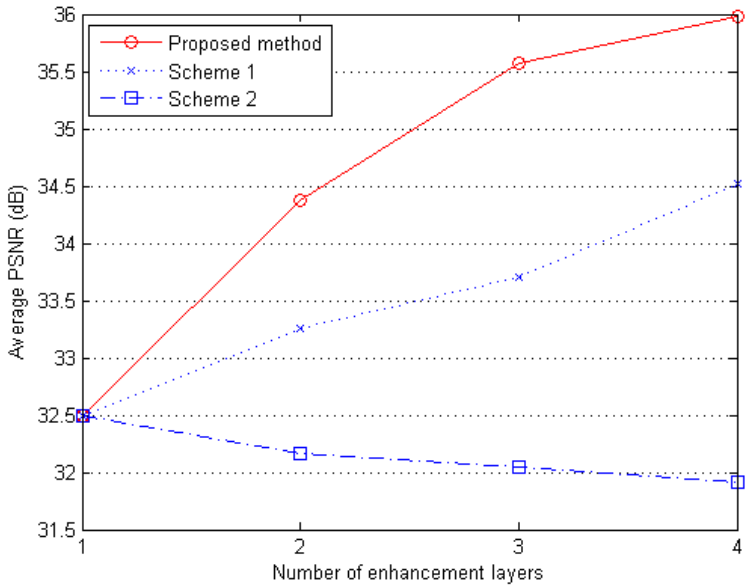


Figure 5.6. Comparison of the proposed method against the other two schemes.

ity. As analyzed in Section 5.3.2, the complexity of the proposed algorithm is $O(ML)$, while for scheme 1 and scheme 2, the complexity will reduce to $O(1)$.

5.5 Conclusion

In this chapter, we deal with the problem of wireless broadcasting of scalable video. To cope with the heterogeneity of different receivers, the scalable extension of H.264/AVC, known as SVC, is applied to encode the raw video data into multiple layers including a quality base layer and several enhancement layers. To the best of our knowledge, the joint source and channel bit-rate allocation for video broadcasting to multiple receivers using the H.264/SVC has never been considered. In this work, we design two different channel error protection schemes based on FEC for the base layer and the enhancement layers, respectively. To guarantee that all the clients can receive the video with a minimum

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

quality, the base layer is highly protected to achieve a very low loss rate. To improve the overall quality of the reconstructed video in the system, UEP scheme is designed for different enhancement layers to ensure that a lower layer can be correctly received with a higher probability. Given the receivers' statistics including the bandwidth and the error rate, we develop a novel algorithm to determine the source coding bit-rate and the channel coding bit-rate of each layer to maximize the system-wide utility. We implement the algorithm to verify its performance and it is shown to be very efficient. The proposed method is compared with the other two schemes to show how various coding structures affect the overall utility. It is demonstrated from the experimental results that the proposed method is more efficient than other schemes and the improvement is up to 2 dB.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 6

Joint Rate Allocation for Multi-Program Video Coding using Fine Granularity Scalability

6.1 Introduction

As stated in Chapter 1, the advances in multimedia technology and digital communications have enabled the broadcasting of multiple programs over a single constant bit-rate (CBR) channel, which was used to transmit a single analog program. It means that different videos are encoded in parallel and the compressed bit-streams will be multiplexed to share the available bandwidth. A simple solution is independent coding of multiple video programs, where the programs are coded independently and each of them

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

has a separate rate control (see Fig. 1.2). However, independent coding suffers from two major drawbacks: potentially large variations in picture quality among programs as well as within a program, and inefficient use of channel capacity. On contrast, it has been shown that joint coding of multiple video programs (see Fig. 1.3) is able to achieve more uniform picture quality and more efficient use of channel capacity by dynamically allocating the channel capacity among video programs. Therefore, rate allocation among the sequences with a constrained bandwidth becomes an important issue to achieve consistent optimized video quality in the distributive application.

In this chapter, we address the problem of joint rate allocation for multi-program video coding using the scalable video encoders. Most of the existing approaches are based on non-scalable video coding platforms, where computationally expensive encoding or transcoding is demanded to adjust the bit-rate of each video program. Different from all these works, we develop a new statistical multiplexing system, where the scalable video coding technique is applied to compress the video programs. In our scheme, each video sequence is separated into GOPs with a fixed length and the rate allocation is updated at the GOP boundaries. First, we propose an efficient look-ahead approach to distribute the base layer coding bit-rate to each video encoder. Then each video is encoded into a base layer and several quality enhancement layers with fine granularity. Second, a piecewise linear model is applied to accurately estimate the rate-distortion (R-D) relationship in the FGS layers. Based on this model, a novel algorithm is designed to dynamically allocate the available channel bandwidth to different video programs for bit-stream adaptation in order to minimize the variation of quality of different video programs in the statistical multiplexing system. Experiments are carried out to verify the performance of the proposed scheme by comparing it with existing methods. The results demonstrate the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

superiority of the proposed scheme and the quality difference between different programs is greatly reduced.

The rest of this chapter is organized as follows. In Section 6.2, we will present the background of joint rate allocation and introduce some related works. In Section 6.3, we give an overview of the multi-program video coding system. The proposed joint rate allocation algorithm is described in Section 6.4. Section 6.5 demonstrates the experimental results for comparison. Finally, we draw a conclusion in Section 6.6.

6.2 Background and Related Work

Statistical multiplexing techniques are broadly used in many video encoding applications, such as digital TV broadcast, video surveillance, and video conferencing. In a statistical multiplexing system, multiple video programs are encoded individually, then multiplexed and transmitted over a bandwidth-limited network to receivers for video decoding and presentation. In this case, not only the bit-rate of each encoder needs to be accurately controlled, but also the total transmission bandwidth needs to be efficiently allocated between video programs to achieve consistent optimized video quality.

The simplest approach to deal with the problem is to encode each of the video programs at an equal constant bit-rate. However, in a multiplexing system, different video programs may have different scene activities, and within each program, the scene activity may change dramatically over time. According to the rate-distortion theory [97], equal rate allocation may lead to uneven distortions between different video programs due to variation of scene complexity. To achieve equal video quality for all programs, the channel bandwidth should be dynamically allocated to different programs in proportion to the complexity of each of the video sources. Therefore, the joint rate allocation algorithm

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

is desired to manage the operation of all the encoders to maintain a uniform picture quality among all video programs.

Joint rate allocation for statistical multiplexing has been studied by many researchers in the past years [98]-[110]. These approaches can be mainly classified into two categories depending on how the R-D statistics of the video frames are obtained: the feed-back approaches and the look-ahead approaches. In the feed-back approaches [98]-[101], [107], [109], statistics are collected during encoding of the previous frames to derive the complexity of the current frame and its subsequent frames. For instance, Böröczky et al. proposed a joint rate control algorithm, where statistics generated by the encoder as a by-product of the compression process are used to control the future bit-rate allocation [98], [110]. However, the feedback approaches assume that the neighboring frames have similar characteristics. Hence, they may suffer from performance degradation when scene change happens. Different from the feed-back approach, the look-ahead approach can greatly extend the range for selection of the statistics, but at the expense of extra computations [102], [103]. In [102], a preprocessing procedure is applied on video frames within the look-ahead window to collect their statistics prior to encoding of the frames. These statistics are then used for joint rate allocation and rate control. These two types of approaches can also be jointly employed to predict more accurately the bit-rate allocation for each of the video programs to achieve a better overall performance [104], [107], [109].

Conventionally, the statistical multiplexing works are based on the single layer coding standards, such as non-scalable mode of MPEG-2 [14] and Annex A profiles of H.264/AVC [18]. In these applications, multiple pre-encoded video streams have to be re-encoded or transcoded in order to fit into the CBR channel, where complicated computations are inevitable. Recently, Jacobs et al. proposed a new system for real-time statistical

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

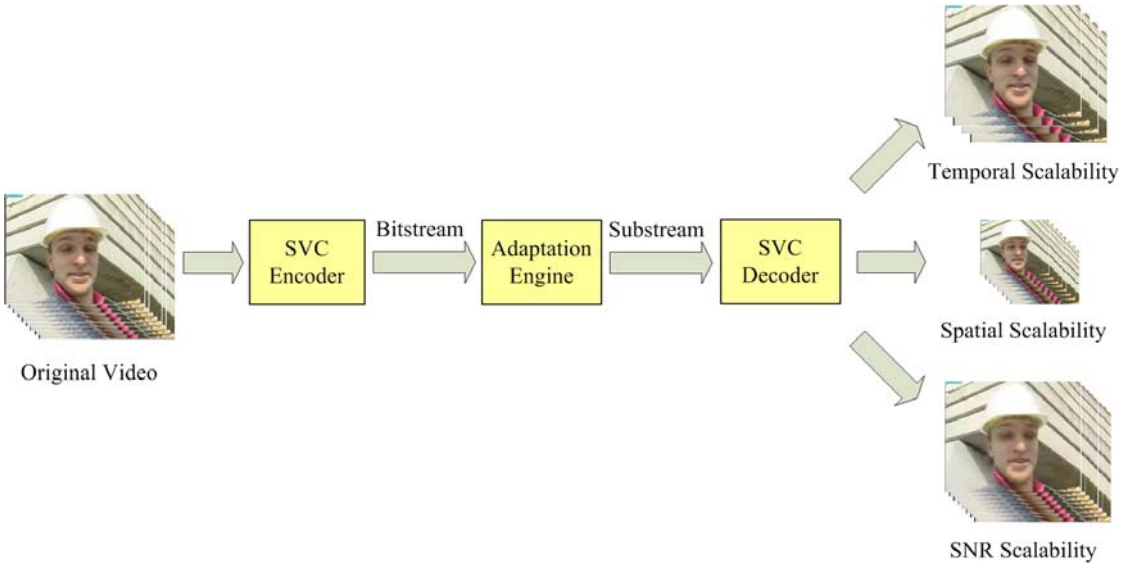


Figure 6.1. Different scalabilities supported by scalable video coding.

multiplexing of video streams in digital video broadcasting for handhelds (DVB-H) using the scalable extension of H.264/AVC [106]. By encoding the video source into scalable bit-stream, a partial stream can be easily extracted to adapt to the bandwidth variation without introducing computationally complex re-encoding or transcoding. Nevertheless, in [106], the bandwidth distribution for the base layer of each video stream is simply decided using equal allocation. In addition, the proposed joint rate allocation algorithm is not efficient enough and will result in a great difference in quality between different video programs, which affects the overall performance of the system.

In this work, we design a new statistical multiplexing system based on the platform in [106]. In the following sections, we will focus on the development of our proposed joint rate allocation algorithm.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

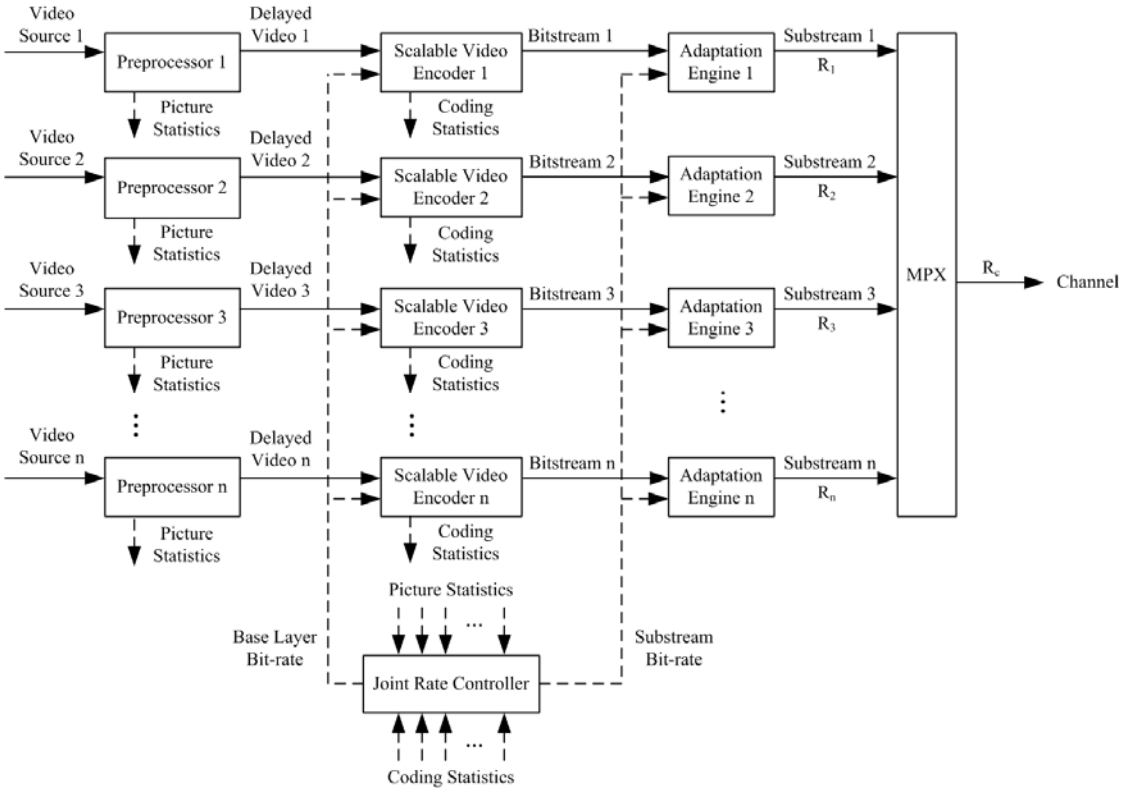


Figure 6.2. The basic framework of the multi-program video coding system.

6.3 System Overview

As introduced in Chapter 2, SVC is developed as the scalable extension of H.264/AVC to achieve a wide range of spatio-temporal and quality scalability. As shown in Fig. 6.1, it allows removal of partial stream from the original bit-stream while decoding the video at reduced temporal, SNR or spatial resolution to satisfy the specific rate and resolution required by a certain application. In this work, we use SVC to encode the video program into compressed bit-stream with quality scalability. We make use of the FGS for encoding of the SNR enhancement layers to realize the truncation of the quality enhancement layers at arbitrary position.

The basic framework of our proposed multi-program video coding system is depicted in Fig. 6.2. The main components in the system comprise a number of preprocessors,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

scalable video encoders, adaptation engines, a joint rate controller and a multiplexer. Firstly, each video source goes through a preprocessor for analysis of the video content on a GOP basis. The picture statistics are derived to characterize the video complexity for each GOP and fed into the joint rate controller. Given the input statistics, the task of the joint rate controller is to dynamically calculate the base layer bit-rate for each program according to the relative complexities of the programs. After receiving the delayed video and the allocated base layer bit-rate, each encoder compresses the video into a base layer and several FGS layers. Meanwhile, the coding statistics are produced as by-product of the encoder. Thereafter, the joint rate controller gathers these coding statistics and calculates the optimal bit-rate allocation to different video programs within a GOP. The compressed bit-stream, together with the computed target bit-rate, is sent to the adaptation engine. Given the target bit-rate, partial stream will be extracted from the input bit-stream in each adaptation engine. Finally, the sub-streams of different video programs are multiplexed to transmit over a single CBR channel.

The most important component in the system is the joint rate controller. The joint rate allocation algorithm aims at minimizing the variation of quality of different video programs. The details of the proposed algorithm will be discussed in the following section.

6.4 Joint Rate Allocation Algorithm

In this section, the proposed joint rate allocation algorithm is described in detail, which consists of two parts. In Section 6.4.1, a look-ahead approach is presented to distribute the base layer coding rate to each video program before the encoding process. In Section 6.4.2 we propose a novel algorithm to allocate the available bandwidth to realize an

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

optimal adaptation of the video streams to minimize the variation of video quality after encoding of the multiple video programs.

6.4.1 Bit-rate Allocation for Video Encoding

In our scheme, within each GOP, the input video is encoded into a base layer and several FGS layers. The base layer provides a minimum quality of video, which is guaranteed to be transmitted when the channel bandwidth is very low. Given the total transmission rate for the base layers, we need to distribute it to different video programs prior to encoding. Instead of equal allocation as employed in [106], we develop a look-ahead approach to determine the base layer bit-rate for each video program.

In our scheme, both the frame activity and the motion activity are used to characterize the video complexity. There are many activity measures for still image coding. In [114], Kim et al. have classified these measures into four categories and suggested that the gradient-based method is more reliable for complexity measure of still image. Therefore, we calculate the frame activity of the I-frames in a GOP as

$$activ_{\text{frame}} = \frac{1}{L \cdot W} \sum_{x=1}^{L-1} \sum_{y=1}^{W-1} \left\{ \left| lum(x,y) - lum(x+1,y) \right| + \left| lum(x,y) - lum(x,y+1) \right| \right\} \tag{6.1}$$

where L and W represent the frame height and the frame width in pixels, respectively. $lum(x,y)$ is the luminance value of pixel (x,y) in the I-frame of the GOP.

The motion activity is calculated as the average frame difference of a GOP, which is computed as follows,

$$activ_{\text{motion}} = \frac{1}{N_{\text{frame}} - 1} \sum_{k=1}^{N_{\text{frame}}-1} Diff_{\text{frame}}(k) \tag{6.2}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

with N_{frame} being the total number of frames in a GOP. $Diff_{\text{frame}}(k)$ denotes the frame difference between the *k*th frame and the (*k*+*1*)th frame in a GOP, which is calculated as

$$Diff_{\text{frame}}(k) = \frac{1}{L \cdot W} \sum_{i=1}^{L} \sum_{j=1}^{W} \left| lum(k, x, y) - lum(k+1, x, y) \right| \tag{6.3}$$

where $lum(k, x, y)$ is the luminance value of pixel (x, y) in the *k*th frame.

These statistics are continuously fed into the joint rate controller on a GOP basis and utilized to dynamically determine the bit-rate allocation for coding of the base layer at each encoder. For the *j*th GOP in the *i*th video program $GOP_{i,j}$, the base layer bit-rate $r_{i,j}$ is determined as

$$r_{i,j} = R_b \left[\alpha \cdot \frac{activ_{\text{frame}}(i,j)}{\sum_{i=1}^{n} activ_{\text{frame}}(i,j)} + (1-\alpha) \cdot \frac{activ_{\text{motion}}(i,j)}{\sum_{i=1}^{n} activ_{\text{motion}}(i,j)} \right] \tag{6.4}$$

with *n* being the total number of the video programs in the system and R_b being the total bit-rate for the base layers. $activ_{\text{frame}}(i,j)$ and $activ_{\text{motion}}(i,j)$ are the frame activity and the motion activity of $GOP_{i,j}$, respectively. The weighting factor α is chosen as 0 at the start of encoding. After encoding a GOP, the bits in different frames are known. α will be dynamically updated using the data from the previous GOP. The value of α is calculated as the proportion of bits used for base layer coding of the I-frames respective to the whole GOP. Once the computed target bit-rate for the base layer is received, each encoder will compress the input video into a high quality bit-stream with a base layer and several FGS layers.

6.4.2 Bit-rate Allocation for Video Adaptation

Since all the *n* video programs are multiplexed into a single channel with a bandwidth R_c, we should have

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$\sum_{i=1}^{n} R_{i,j} \le R_c \tag{6.5}$$

where $R_{i,j}$ is the transmission bit-rate allocated to $GOP_{i,j}$. To satisfy this constraint, all the output bit-streams from the encoders need to be extracted in the adaptation engines. Because of the FGS coding, it is possible to truncate the bit-stream at any positions in the enhancement layers. As described in [111], the adaptation engine employs a simple method to truncate the refinement network abstraction layer (NAL) units. To determine the NAL units to be truncated as well as the truncation points, this method makes use of the calculated target bit-rate and the average bit-rate of different layers of the GOP to be adapted. The average bit-rate can be calculated in the encoder to save the runtime overhead and the results will be inserted in the SVC compliant supplemental enhancement information (SEI) messages at the beginning of each GOP.

Truncation of a video stream at different bit-rates will result in variation of quality of the decoded video. Given the total transmission bandwidth, our objective is to dynamically allocate a target bit-rate to each video program for bit-stream adaptation such that the variation of distortion between different video programs will be minimized. This problem can be formulated as

$$\begin{aligned} \min_{R_{i,j}} \quad & Var = \frac{1}{n}\sum_{i=1}^{n}\left(MSE_{i,j} - \overline{MSE_j}\right)^2 \\ s.t. \quad & \sum_{i=1}^{n} R_{i,j} \le R_c \end{aligned} \tag{6.6}$$

with Var being the variance of distortion between different video programs. $MSE_{i,j}$ denotes the average mean square error (MSE) of the reconstructed video in $GOP_{i,j}$ and $\overline{MSE_j}$ is computed as

$$\overline{MSE_j} = \frac{1}{n}\sum_{i=1}^{n} MSE_{i,j} \tag{6.7}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For the video data in FGS layers, we can build a one-to-one mapping between the video bit-rate and the video quality. Let Φ represent the rate-distortion relationship, we have

$$D = \Phi(R) \tag{6.8}$$

where D is the distortion of the reconstructed video and R is the bit-rate of the video bit-stream. That is to say, we can map each value of $R_{i,j}$ to a value of video distortion $MSE_{i,j}$. In turn, we can map each value of $MSE_{i,j}$ to a value of $R_{i,j}$. Therefore, we assume that searching of the optimal rate allocation equals to searching of the corresponding video distortion under the premise that the mapping function is known.

In [113], the R-D function of SVC FGS is analyzed and inferred to be linear under MSE criterion within an FGS level. We also conduct a number of experiments to verify the piecewise linear R-D relationship in the FGS layers. Fig. 6.3 shows the results using *Foreman* and *Football*. We randomly pick out a GOP from each video sequence and use *QP*38 to encode it into a base layer and three FGS layers. Each figure gives the R-D curve in an individual FGS layer. Fig.6.3 (a)-(c) show the results for *Foreman* and (d)-(f) show the results for *Football*. It is observed that the R-D curve of each FGS layer appears to be nearly linear. Therefore, given both of the end values in an R-D curve, any in-between values can be estimated using linear interpolation as follows,

$$D^* = (1-\eta) D_L + \eta D_U \tag{6.9}$$

where D_L and D_U are the lower bound and the upper bound of the video distortion of a FGS layer. D^* denotes the video distortion of an in-between point in the curve and η is a number between 0 and 1 that represents how far the point is placed between the end points. η is computed as

$$\eta = \frac{R^* - R_L}{R_U - R_L} \tag{6.10}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

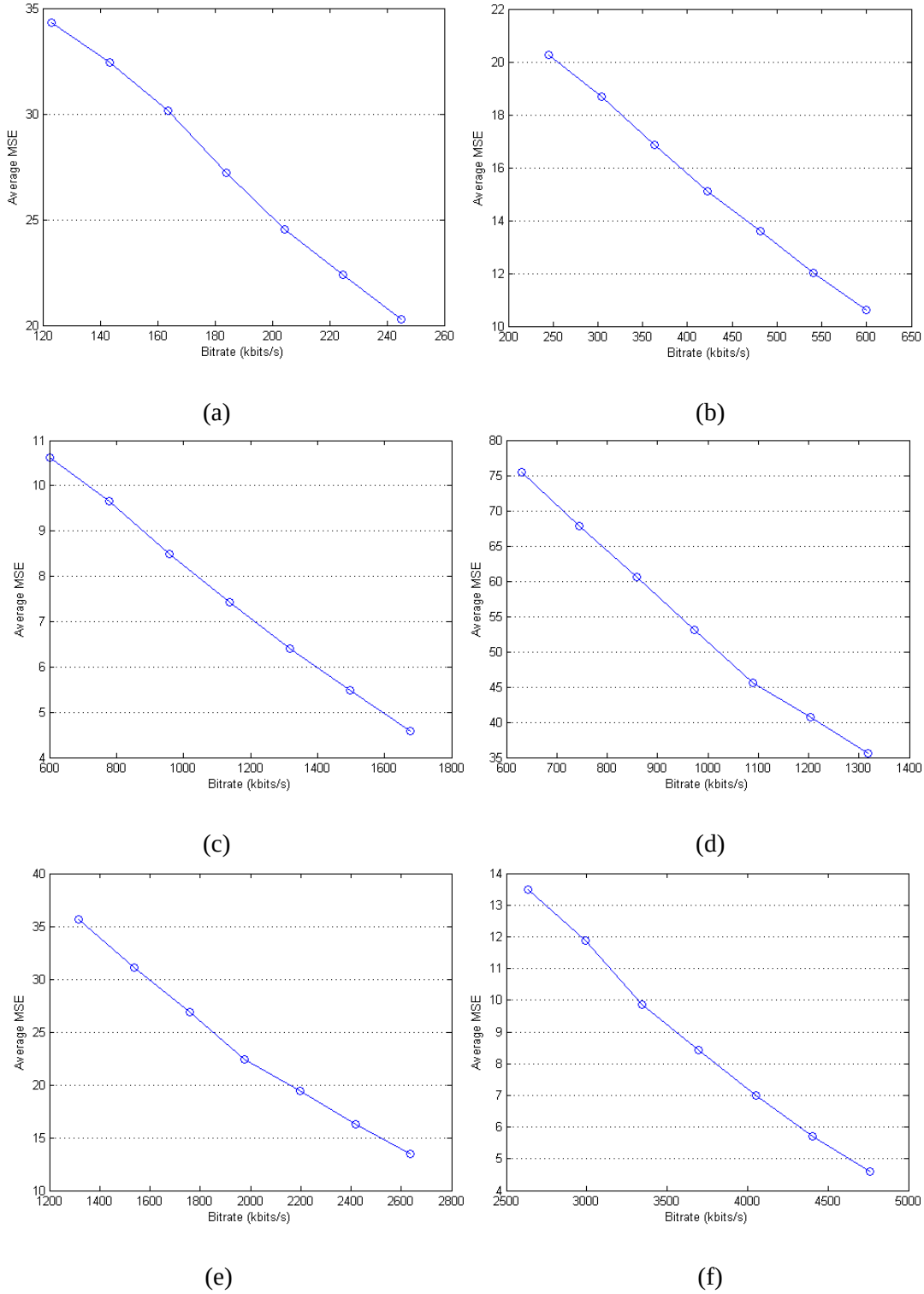


Figure 6.3. R-D curve of different FGS layers: (a) Foreman (the first FGS layer), (b) Foreman (the second FGS layer), (c) Foreman (the third FGS layer), (d) Football (the first FGS layer), (e) Football (the second FGS layer) and (f) Football (the third FGS layer).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

with R_L and R_U being the lower bound and the upper bound of the bit-rate of a FGS layer. R^* represents the bit-rate of the in-between point in the curve. Given the value of D^*, (9) is deduced to calculate R^* as

$$R^* = (1-\eta')R_L + \eta' \cdot R_U \tag{6.11}$$

$$\eta' = \frac{D^* - D_L}{D_U - D_L} \tag{6.12}$$

During the encoding process, the bit-rate and the distortion of the reconstructed video of the base layer or any of the FGS layers can be calculated as by-products. In our scheme, these coding statistics are collected by the joint rate controller for bit-rate allocation. Because the end values of the R-D curve of each FGS layer are available, the value of any truncation point in a FGS layer can be estimated using linear interpolation.

Based on the above observations and derivations, we propose a novel joint rate allocation algorithm. Instead of searching of the bit-rate allocation, we aim at looking for the target video distortion. To minimize the variation of video quality between different video programs, the target video distortion should be the same for the adapted videos. Therefore, the problem in (6) is re-formulated as

$$\begin{aligned} \min_{MSE^j_{\text{target}}} \quad & Diff_R = \left| R_c - \sum_{i=1}^{n} R_{i,j} \right| \\ s.t. \quad & \sum_{i=1}^{n} R_{i,j} \le R_c \end{aligned} \tag{6.13}$$

with MSE^j_{target} being the target MSE of the jth GOP of each video program and $Diff_R$ being the bit-rate difference between the available channel bandwidth and the summation of the bit-rate of the adapted video bit-streams. Given the target MSE, $R_{i,j}$ for the jth GOP in each of the video programs is estimated using (6.11) and (6.12). To optimize the overall performance of the system, our objective is to make full use of the available

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

channel bandwidth under the constraint that the total allocated bit-rate should not exceed the budget bit-rate.

Exhaustive searching can be applied to solve the problem. However it is unfeasible in reality because of the large amount of computation. Hence, we design a golden-section search algorithm to find the sub-optimal solution to this problem. The golden section search is among the most efficient region elimination methods to optimize functions with single variable, provided upper and lower bounds exist [112]. The details of the algorithm are introduced in the following.

Step 1. Initialization.

In this algorithm, the search is based on the number 0.618, known as the golden section. Let [a, b] be the initial interval, the first searching point is located at 0.618(b-a) from b. Let $R_l^{i,j}$ and $D_l^{i,j}$ be the bit-rate and the distortion of the reconstructed video with receiving of the lth layer and all the lower layers of GOP$_{i,j}$, where $l = 0,1,2,\ldots,F$. Here the 0th layer is the base layer and F is the total number of FGS layers. The values of a and b are determined as

$$a = \min\left(D_F^{i,j}\right), i = 1,2,3,\ldots,n \tag{6.14}$$

$$b = \max\left(D_0^{i,j}\right), i = 1,2,3,\ldots,n \tag{6.15}$$

Given a and b, the first searching point is achieved.

Step 2. Calculate the bit-rate allocation.

The new searching point serves as MSE_{target}^{j} and is used to calculate the bit-rate allocation. For each video program, we firstly locate the target MSE in one of the FGS layers.

1) If $MSE_{\text{target}}^{j} \geq D_0^{i,j}$, $R_{i,j} = R_0^{i,j}$.

2) If $MSE_{\text{target}}^{j} < D_F^{i,j}$, $R_{i,j} = R_F^{i,j}$.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3) If $D_l^{i,j} \le MSE_{\text{target}}^{j} < D_{l-1}^{i,j}$, the target MSE is located in the lth FGS layer, where $l = 1, 2, 3, \ldots, F$. The R-D curve of this FGS layer is used to map the target MSE to the allocated bit-rate $R_{i,j}$ as in (6.11).

Let $R_{sum} = \sum_{i=1}^{n} R_{i,j}$, R_{sum} is computed for the current searching point.

Step 3. Update the interval of interest.

1) If $R_{sum} > R_c$, the lower endpoint of the interval is replaced by the current searching point. Go to step 4.

2) If $R_{sum} < R_c - \varepsilon$, the higher endpoint of the interval is replaced by the current searching point. Here, ε is defined as a threshold value, which equals to $0.002R_c$. Go to step 4.

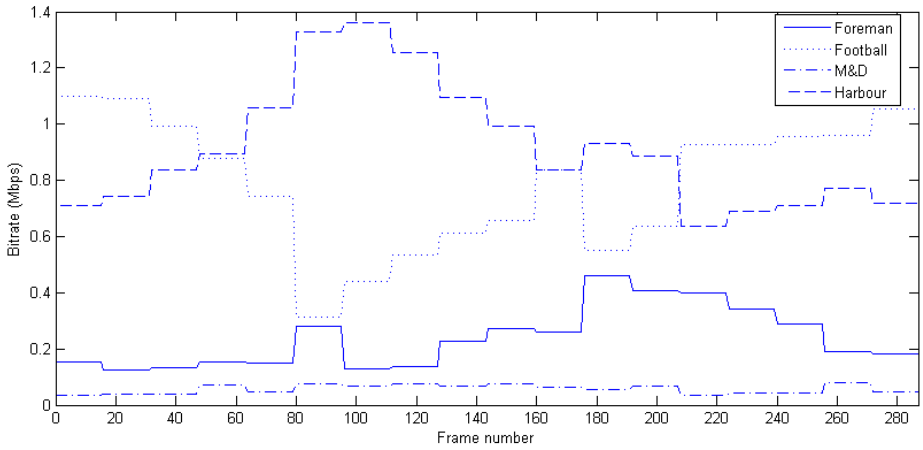
3) If $R_c - \varepsilon \le R_{sum} \le R_c$, the sub-optimal bit-rate allocation is attained. Go back to step 1 to process the next GOP.

Step 4. Calculate the new searching point.

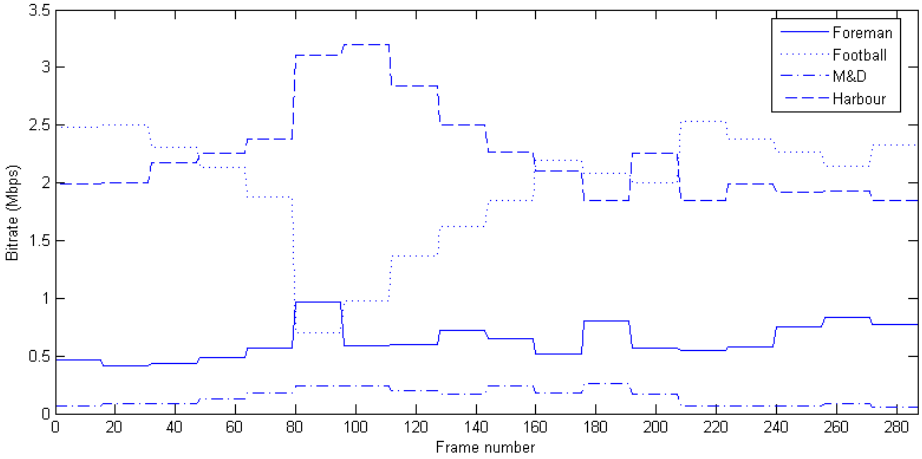
Let L be the length of the new interval, the new searching point is positioned at $0.618L$ from the new endpoint. Go to step 2.

The proposed golden-section search algorithm converges very fast. Normally, it takes only several iterations to attain the sub-optimal solution even if we set a very small value to ε. As described in this section, the proposed joint rate allocation algorithm only involves a number of addition and multiplication operations. Therefore, the overall complexity is very low. The performance of the algorithm will be evaluated in Section 6.5.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



(a)



(b)

Figure 6.4. Allocated bit-rates for different video programs under different channel bandwidths: (a) 2 Mbps and (b) 5 Mbps.

6.5 Experimental Results

To assess the performance of the proposed system, we carry out a number of experiments using the H.264/SVC encoders [94].

In our system, CIF resolution sequences *Foreman*, *Football*, *Mother & Daughter* (*M&D*) and *Harbour* are used as video sources, which have different scene complexities. The input frame rate is 30fps for each encoder. They are multiplexed into a single CBR

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

channel for transmission. Each video program is separated into GOPs and the bit-rate allocation between different video programs is updated at the GOP boundaries.

The proposed joint rate allocation algorithm can be easily employed under different channel bit-rates. In the experiments, we simulate the proposed algorithm under 2Mbps and 5Mbps, respectively. The target bit-rate for coding of the base layers is assumed to be the minimal possible bandwidth and the estimation of the minimum bandwidth is out of scope of this work. In the experiments, we assume that the target bit-rate for coding of the base layers of all the video programs is 1Mbps. Once the base layer coding bit-rate for each video is determined by joint rate controller using the statistics from preprocessors, each encoder compresses the input video program into a base layer and 3 FGS layers. Thereafter, the joint rate controller dynamically allocates the truncation bit-rate for each adaptation engine based on the collected statistics from the encoders. Fig. 6.4 shows the bit-rate allocated to each video program using the proposed joint rate allocation algorithm under different channel bit-rates. It can be seen that the bit-rate allocation is dynamically updated on a GOP basis. In general, the tendency for distribution of the bit-rate under different channel bandwidths remains. Sequences *Foreman* and *M&D* have lower bit-rates because they have lower scene complexities with respect to the other two video programs.

In the experiments, we also implement another two existing methods for comparison. One is equal allocation of bit-rate to different video programs. The other is the algorithm proposed in [106]. In this algorithm, the bit-rate for the base layer coding is equally allocated to different video programs. Several complexity measures are proposed to distribute the channel bandwidth. The most promising one among them is

$$x_{i,j} = 10^5 \frac{size_{BL}^{i,j}}{PSNR_{BL}^{i,j5}} \tag{6.16}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 6.1
AVERAGE MSE AND VARIANCE OF DISTORTION OF EACH GOP UNDER DIFFERENT SCHEMES (CHANNEL BITRATE=2MBPS, GOP SIZE: 16)

GOP index	Average MSE of each video program												Variance		
	Foreman			Football			M&D			Harbour					
	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme
1	13.31	20.61	37.88	129.44	76.55	55.69	2.62	3.11	25.77	73.64	66.10	53.79	2592.8	935.8	150.03
2	11.39	17.05	42.18	130.57	84.42	51.79	3.24	4.71	30.16	74.05	65.65	50.92	2663.5	1092.0	75.78
3	11.74	17.88	41.03	125.24	79.46	51.65	3.10	4.08	28.43	78.09	66.71	54.76	2507.9	1008.5	106.38
4	13.75	20.04	39.49	104.13	75.17	52.51	4.03	6.36	24.85	80.09	63.34	54.39	1815.5	826.3	140.40
5	13.12	20.27	41.44	73.65	56.02	41.86	3.83	6.12	37.06	76.13	58.24	43.04	1114.3	508.3	5.14
6	16.45	21.43	24.98	13.73	19.45	17.70	3.96	6.06	21.78	76.32	43.20	33.83	812.2	177.3	35.26
7	9.12	13.52	30.12	19.87	24.93	20.70	3.92	6.19	23.70	83.22	43.84	35.09	1011.9	201.8	31.30
8	10.20	15.14	30.82	32.97	30.01	27.24	4.04	6.20	21.23	78.14	48.92	33.98	846.2	262.0	22.42
9	16.28	27.03	32.85	54.11	42.99	40.81	3.95	6.26	25.55	77.62	59.07	39.79	865.2	381.8	37.58
10	17.13	27.88	27.21	61.59	49.20	42.80	4.93	8.33	30.87	72.95	56.20	42.63	825.6	353.0	48.43
11	14.42	20.93	23.83	89.47	60.98	48.35	3.91	6.29	28.34	71.76	60.10	47.27	1329.1	577.5	120.71
12	19.95	29.46	15.47	53.09	45.55	46.37	3.67	5.90	25.74	74.45	55.05	48.55	765.2	346.7	194.11
13	17.15	27.58	17.08	62.39	48.99	48.12	3.90	6.09	25.40	75.47	58.25	46.53	896.1	406.7	179.12
14	21.41	36.97	22.70	131.28	75.67	63.84	2.82	3.79	26.59	72.11	67.05	64.63	2486.9	796.7	393.76
15	21.86	35.82	28.62	136.57	77.58	64.78	3.01	3.92	25.36	75.44	71.26	67.56	2700.5	876.0	385.96
16	24.90	39.30	36.10	128.51	74.32	57.86	2.79	3.69	23.02	72.06	67.91	60.00	2327.4	779.1	237.59
17	23.21	37.35	50.71	110.09	65.16	51.83	3.39	4.94	18.32	66.90	61.08	52.55	1695.9	573.8	209.34
18	20.86	35.43	48.77	122.24	66.71	48.94	2.61	3.16	18.44	63.12	62.49	50.88	2116.4	645.6	181.89
Average MSE	16.46	25.76	32.85	87.72	58.51	46.27	3.54	5.29	25.59	74.53	59.69	48.90			

where $size_{BL}^{i,j}$ is the size of the base layer for GOPi,j and $PSNR_{BL}^{i,j}$ is the average of the PSNR between the base layer and the original sequence for $GOP_{i,j}$. Given the estimated complexity, the bit-rate allocation is determined as

$$R_{i,j} = R_c \frac{x_{i,j}}{\sum_{i=1}^{n} x_{i,j}} \tag{6.17}$$

Using this method, the joint rate controller regularly assigns bit-rate that is even lower than the base layer bit-rate to the video program with lower complexity. When this happens, we let the base layer bit-rate be the allocated bit-rate to this program and the remaining channel bandwidth is re-allocated to other programs using the above method.

Table 6.1 shows the results for different schemes when the total available channel bandwidth is 2Mbps. For each GOP, the average MSE of each video program and the variance of distortion of all the programs are displayed. The average MSE of a video program across all the GOPs is also calculated. It can be seen that the variance of video

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

distortion is greatly reduced using our method, which demonstrates the superiority of the proposed algorithm. In contrast, the variance of distortion achieved using the method in [106] is much higher. In addition, both the proposed algorithm and the method in [106] perform much better than equal allocation. We also notice that the video quality of the method in [106] is better than our algorithm for sequences M&D and Foreman. It is because that the method in [106] always allocates more bit-rate to the low complexity video programs. The increase of bit-rate for the low complexity video can greatly improve the video quality. We can observe that the video distortion of M&D is obviously lower than other video programs. Comparing these two algorithms, the quality of the videos with higher complexity, such as Football and Harbour, is always better under our proposed method.

TABLE 6.2
AVERAGE MSE AND VARIANCE OF DISTORTION OF EACH GOP UNDER DIFFERENT SCHEMES (CHANNEL BITRATE =5MBPS, GOP SIZE: 16)

GOP index	Average MSE of each video program												Variance		
	Foreman			Football			M&D			Harbour					
	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme
1	6.66	13.49	18.30	51.49	17.15	19.08	1.96	2.79	18.91	32.19	25.25	17.55	401.41	65.20	0.36
2	5.56	12.12	18.36	58.31	21.65	18.88	2.22	4.09	18.25	33.49	24.11	17.77	519.58	63.40	0.15
3	5.73	12.45	18.04	53.77	18.74	17.58	2.13	3.43	17.67	34.76	25.43	19.47	453.51	65.77	0.58
4	6.44	12.69	17.85	47.69	22.16	17.73	2.37	6.35	17.60	35.42	24.07	19.93	365.90	51.69	0.92
5	6.34	11.97	15.24	31.46	17.06	14.92	2.27	5.85	15.47	33.55	20.88	17.19	201.39	31.79	0.77
6	7.97	9.71	10.14	6.31	10.16	9.74	2.39	4.72	10.15	33.77	13.06	11.45	153.41	8.99	0.41
7	4.64	8.28	9.95	7.97	9.51	9.98	2.36	4.90	10.01	36.48	12.51	11.62	189.90	7.43	0.51
8	5.12	9.24	11.09	13.55	11.88	11.21	2.40	5.08	11.20	35.23	16.16	11.84	166.04	16.21	0.09
9	7.77	12.48	13.94	20.63	14.44	13.78	2.35	5.70	13.77	35.17	20.31	14.77	160.47	27.19	0.17
10	7.67	12.67	14.01	26.37	17.41	13.81	2.49	6.65	14.34	32.39	20.47	13.72	155.52	27.23	0.06
11	6.68	12.50	13.86	37.14	16.69	13.82	2.34	6.09	14.64	31.38	22.71	14.27	227.76	36.71	0.11
12	8.38	11.41	9.36	22.69	16.46	10.24	2.37	5.49	9.42	31.81	18.97	17.02	134.53	26.63	10.23
13	7.55	11.66	13.33	27.86	18.05	13.64	2.33	5.67	13.65	33.23	20.88	13.40	170.92	34.64	0.02
14	9.31	17.04	18.74	52.95	18.92	19.62	2.03	3.28	18.87	31.88	25.45	18.53	399.61	65.15	0.17
15	10.94	19.13	19.96	56.29	19.87	20.99	2.08	3.33	21.17	34.04	27.46	21.79	445.19	77.08	0.43
16	13.09	20.37	18.98	50.87	17.89	19.03	2.05	3.21	19.78	31.57	25.71	20.62	344.95	69.50	0.45
17	11.19	19.12	17.60	43.97	15.42	17.44	2.21	4.37	17.81	30.21	23.36	19.81	264.63	49.70	0.92
18	10.27	17.91	16.80	49.88	18.17	16.39	2.00	2.76	17.13	27.71	24.11	18.92	336.58	62.28	0.94
Average MSE	7.85	13.57	15.31	36.62	16.76	15.44	2.24	4.65	15.55	33.02	21.72	16.65			

The experiments are also conducted under the channel bit-rate 5Mbps to evaluate the performance of these three algorithms. As shown in Table 6.2, the difference of distortion

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between various video programs is dramatically decreased under each of the algorithms. It is because that the base layer quality of different videos may have significant difference. When the channel bandwidth is low, even if most of the available bit-rate is allocated to the video with higher complexity, the reconstructed quality is still much lower than the base layer quality of the low complexity video. With increasing of the channel bandwidth, the quality difference can be made up by allocating more bits to the high complexity video. This further demonstrates the superiority of our proposed algorithm. Instead of equal allocation, we distribute the base layer coding rate to different videos according to their scene complexities, which makes the base layer quality of different videos to be uniform. In Table 6.2, the computed variance of video distortion is very small and in general it is no more than 1. It proves both the efficiency of the proposed joint rate allocation algorithm and the accuracy of the piecewise linear model used to estimate the R-D relationship in the FGS layers.

We further conduct experiments to test the performance of different schemes when GOP size is 30. The results are given in Table 6.3-6.4. Similar results can be seen in the tables compared to the results obtained when GOP size is 16. Under different channel bandwidths, the proposed method always exhibits more advantage than the other two schemes. With increasing of the overall transmission bit-rate, the performance of the proposed method greatly improves.

To compare the computational complexity of our proposed algorithm with the method in [106], we present the running time of these two algorithms in Table 6.5. Both of the two algorithms terminate in negligible time. For instance, our algorithm takes up to 12.8 milliseconds to solve the joint rate allocation problem for one GOP. Because of its low computational complexity, our algorithm is feasible for real-time statistical multiplexing.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

TABLE 6.3
AVERAGE MSE AND VARIANCE OF DISTORTION OF EACH GOP UNDER DIFFERENT SCHEMES (CHANNEL BITRATE =2MBPS, GOP SIZE: 30)

GOP index	Average MSE of each video program												Variance		
	Foreman			Football			M&D			Harbour					
	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme
1	10.80	15.60	30.89	128.57	76.02	50.95	2.68	3.11	28.81	65.21	63.60	55.02	2541.7	952.40	136.41
2	10.99	15.42	30.50	112.95	71.40	45.09	3.25	4.49	18.13	69.39	60.59	52.69	2010.6	814.65	177.31
3	12.28	19.91	30.96	43.25	37.46	30.80	3.62	5.71	23.01	69.64	49.38	35.23	684.2	277.29	19.46
4	9.37	15.09	20.62	21.76	23.39	21.50	3.64	5.49	17.15	73.19	42.61	32.04	754.5	186.63	30.95
5	12.86	21.10	23.23	47.46	38.77	34.21	3.65	5.58	17.88	67.32	48.76	37.43	663.5	274.02	63.17
6	15.20	26.22	20.87	75.72	54.47	39.61	3.73	6.11	24.10	64.20	52.26	43.56	947.9	397.17	94.41
7	17.67	26.48	14.18	60.08	49.73	40.29	3.73	6.10	22.25	67.08	51.20	45.67	729.4	344.25	165.15
8	20.68	29.68	20.08	129.44	70.14	58.48	2.81	3.40	17.56	63.94	64.86	64.43	2380.2	739.11	459.58
9	20.38	26.16	26.16	110.81	61.86	47.17	2.87	3.58	13.65	61.01	57.70	51.94	1727.8	570.21	242.24
10	17.96	22.86	30.47	124.65	64.60	44.17	2.53	2.92	10.60	55.59	59.65	49.94	2220.6	658.82	229.33
Average MSE	14.82	21.85	24.80	85.47	54.78	41.23	3.25	4.65	19.31	65.66	55.06	46.80			

TABLE 6.4
AVERAGE MSE AND VARIANCE OF DISTORTION OF EACH GOP UNDER DIFFERENT SCHEMES (CHANNEL BITRATE =5MBPS, GOP SIZE: 30)

GOP index	Average MSE of each video program												Variance		
	Foreman			Football			M&D			Harbour					
	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme	Equal rate	Scheme in [106]	Proposed scheme
1	5.47	11.75	17.09	54.77	18.10	17.40	2.01	3.11	13.11	29.89	23.78	18.88	451.09	58.95	4.56
2	5.48	10.99	16.49	50.32	18.04	15.37	2.25	4.50	15.88	31.96	23.16	18.11	390.74	49.91	1.07
3	6.11	9.85	12.60	18.65	14.16	12.21	2.38	5.02	12.66	31.99	17.68	13.72	135.01	22.46	0.31
4	5.02	7.80	10.05	9.19	9.35	9.63	2.40	4.52	9.84	26.49	13.56	11.08	88.24	10.56	0.31
5	6.37	10.50	12.36	19.55	13.25	12.79	2.37	5.02	11.54	30.18	17.53	12.26	121.14	20.62	0.20
6	7.17	12.02	13.79	33.04	16.40	13.39	2.33	5.68	13.38	28.02	19.83	11.86	172.24	27.94	0.54
7	7.82	11.02	12.34	25.54	17.81	12.97	2.31	5.48	12.89	28.90	18.45	11.72	127.87	28.28	0.25
8	9.84	16.83	19.74	52.81	19.24	18.54	2.05	3.40	17.27	28.75	25.72	18.72	383.25	66.06	0.77
9	11.93	16.33	16.74	44.91	16.00	16.23	2.04	3.58	14.32	27.71	22.29	16.36	264.23	46.40	0.88
10	10.15	14.95	17.05	51.85	19.00	15.48	1.91	2.92	11.17	24.54	23.23	17.01	360.33	57.42	5.75
Average MSE	7.54	12.20	14.82	36.06	16.14	14.40	2.20	4.32	13.21	28.84	20.52	14.97			

TABLE 6.5
COMPARISON OF RUNNING TIME (IN MILLISECONDS) USING THE PROPOSED SCHEME AND THE SCHEME IN [106] (CHANNEL BITRATE=2MBPS, GOP SIZE: 16)

Scheme	GOP index									
	1	2	3	4	5	6	7	8	9	10
Proposed	8.1	10.9	11.9	12.8	8.6	10.0	11.3	10.4	10.9	9.5
[106]	3.2	3.2	3.1	3.3	3.2	4.7	3.1	3.2	4.4	4.7

We also carry out experiments to compare the performance of the proposed algorithm with the scheme in [104], where non-SVC based joint rate control algorithm is implemented to allocate the channel bandwidth among video programs with various

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

complexities. The results are shown in Table 6.6. The overall distortion of the scheme in [104] is lower than our proposed method. It is mainly because that the scalable coding algorithm will result in decrease of the coding efficiency. With the same amount of bit-rate, the non-scalable coding algorithm can achieve a better video quality. However, it is shown in the table that the variation of quality under the proposed scheme is much lower than the scheme in [104]. In addition, re-encoding or transcoding is unavoidable using the non-scalable coding platform, which consumes a lot of computations compared to the proposed method. Therefore, our proposed algorithm is more practical for real-time applications.

TABLE 6.6
COMPARISON OF THE PROPOSED SCHEME AND THE SCHEME IN [104]
(CHANNEL BITRATE=2MBPS, GOP SIZE: 16)

GOP index	Average MSE of each video program								Average MSE		Variance	
	Foreman		Football		M&D		Harbour					
	Scheme in [104]	Proposed scheme	Scheme in [104]	Proposed scheme	Scheme in [104]	Proposed scheme	Scheme in [104]	Proposed scheme	Scheme in [104]	Proposed scheme	Scheme in [104]	Proposed scheme
1	11.27	37.88	83.84	55.69	2.97	25.77	41.40	53.79	34.87	43.28	1003.8	150.03
2	10.36	42.18	78.78	51.79	4.53	30.16	41.36	50.92	33.76	43.76	871.70	75.78
3	11.25	41.03	68.78	51.65	4.27	28.43	42.86	54.76	31.79	43.97	667.52	106.38
4	11.94	39.49	60.59	52.51	6.79	24.85	46.39	54.39	31.43	42.81	515.20	140.40
5	11.09	41.44	43.51	41.86	5.97	37.06	41.87	43.04	25.61	40.85	295.32	5.14
6	13.25	24.98	10.41	17.70	5.69	21.78	40.67	33.83	17.50	24.57	186.19	35.26
7	7.47	30.12	20.91	20.70	5.14	23.70	41.51	35.09	18.76	27.40	208.77	31.30
8	8.44	30.82	31.66	27.24	5.10	21.23	37.05	33.98	20.56	28.32	195.26	22.42
9	13.85	32.85	44.36	40.81	5.26	25.55	37.12	39.79	25.15	34.75	258.86	37.58
10	13.71	27.21	43.82	42.80	7.12	30.87	36.30	42.63	25.24	35.88	232.25	48.43
11	12.09	23.83	63.77	48.35	5.91	28.34	36.62	47.27	29.60	36.95	521.15	120.71
12	15.66	15.47	39.84	46.37	4.96	25.74	36.81	48.55	24.32	34.03	211.67	194.11
13	16.67	17.08	47.29	48.12	5.29	25.40	34.83	46.53	26.02	34.28	261.81	179.12
14	22.53	22.70	92.05	63.84	3.60	26.59	36.82	64.63	38.75	44.44	1085.6	393.76
15	21.16	28.62	82.63	64.78	4.13	25.36	40.67	67.56	37.15	46.58	856.625	385.96
16	23.33	36.10	76.82	57.86	4.02	23.02	40.19	60.00	36.09	44.25	716.72	237.59
17	22.54	50.71	63.63	51.83	5.52	18.32	39.14	52.55	32.71	43.35	460.14	209.34
18	20.51	48.77	67.97	48.94	3.33	18.44	35.62	50.88	31.86	41.76	565.28	181.89
Average MSE	14.84	32.85	56.70	46.27	4.98	25.59	39.29	48.90				

6.6 Conclusion

In this chapter, we deal with the problem of joint rate allocation for multi-program video coding on a GOP basis. To the best of our knowledge, most of the existing approaches are

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

based on non-scalable video coding platforms, where computationally expensive encoding or transcoding is demanded to adjust the bit-rate of each video program. To save the computational complexity of the system, we develop a new statistical multiplexing system, where the scalable extension of H.264/AVC, known as SVC, is applied to compress the video programs. In video broadcasting, the minimum video quality fluctuation should be achieved while switching from one video program to another. Aiming at this purpose, a novel joint rate allocation algorithm is proposed to dynamically allocate the available channel bandwidth to different programs in order to minimize the variation of quality of the decoded videos. Instead of equal allocation of base layer coding bit-rate, we first design a simple and effective method to distribute the bit-rate to different video encoders based on the video properties. Second, the coding statistics generated from the encoders as by-products are utilized for bit-rate allocation to different video programs. Based on this statistical information, a piecewise linear model is proposed to accurately estimate the R-D relationship in the FGS layers for the statistical multiplexing system. And then we develop a novel golden-section search algorithm to quickly find the sub-optimal solution for the bit-rate allocation problem. Experiments are conducted to evaluate the performance of the proposed joint rate allocation algorithm under different channel bit-rates. The proposed method achieves significant improvement comparing with the existing methods. The experimental results show that the variance of quality of different video programs is dramatically reduced under our algorithm, which demonstrates the efficiency and the superiority of the proposed joint rate allocation scheme.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 7

Conclusions and Future Works

In this chapter, we summarize the main developments and results in Section 7.1, and present some directions for future research in Section 7.2.

7.1 Conclusions

With fast advances in computing and networking technologies, the network-based video applications have attracted more and more interest from both industry and research societies. Due to the variation of user devices and the heterogeneity of transmission networks, it is desirable to introduce an efficient tool to realize video adaptation. Scalable video coding technique allows for simple and flexible solutions for transmission over heterogeneous networks, additionally providing adaptability for bandwidth variations and error conditions. Simple adaptation can be achieved for a variety of storage devices and terminals. In this thesis, a general review of the scalable video coding is given and various fundamental techniques in error control and joint rate allocation are studied for

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

scalable video transmission. The goal of our research is to produce the best visual quality of the reconstructed video with a given resource constraint and achieve better tradeoff between computing complexity and quality. In this regards, this thesis provided the following contributions:

The first aspect was on the error control techniques including error resilience and error protection methods.

1) A 2-D channel rate allocation scheme is proposed for MCTF-based SVC.

In Chapter 3, we address the problem of UEP for scalable video transmission over packet-erasure channel. The proposed method jointly considers the dependency of different temporal layers in a GOP as well as the dependency of different quality layers in each temporal layer. In this way, the available channel protection rate is properly allocated to different layers in both the temporal direction and the SNR direction based on their importance and the channel conditions to provide a graceful degradation in the presence of packet loss. As packet loss rate varies, the channel rate allocation pattern is adaptively changed to achieve the optimal result.

2) An adaptive resynchronization approach is developed for scalable video.

In Chapter 4, a joint GOP level and picture level resynchronization method is proposed to deal with the bit errors that occur in the wireless networks. For transmission of the enhancement layers of the scalable video, we first group different parts of the compressed bit-stream into a number of units according to their sensitivity to errors. An efficient method is applied to measure the importance of each unit and organize them into hierarchical units from the most important unit to the least important one. Considering the time-varying channel condition and the significance of different units, resynchronization markers are properly inserted in different layers. The effectiveness of the proposed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

resynchronization method lies in reducing the loss of information based on significance of each unit of FGS video streaming.

3) A bit-rate allocation scheme is presented for broadcasting of scalable video.

In Chapter 5, we address the problem of wireless broadcasting of scalable video. The joint source and channel bit-rate allocation for video broadcasting to multiple receivers are studied. Two different channel error protection schemes based on FEC are designed for the base layer and the enhancement layers, respectively. To guarantee that all the clients can receive the video with a minimum quality, the base layer is highly protected to achieve a very low loss rate. On the other hand, UEP scheme is designed for different enhancement layers to ensure that a lower layer can be correctly received with a higher probability. Given the receivers' statistics including the bandwidth and the error rate, a novel algorithm is developed to determine the source coding bit-rate and the channel coding bit-rate of each layer to maximize the system-wide utility.

The second aspect was on the joint rate allocation for multi-program video coding. In video broadcasting, the minimum video quality fluctuation should be achieved while switching from one video program to another. Aiming at this purpose, we develop a new statistical multiplexing system in Chapter 6, where SVC is applied to compress the video programs. We first design a simple and effective method to distribute the bit-rate to different video encoders based on the video properties. Second, the coding statistics generated from the encoders as by-products are utilized for bit-rate allocation to different video programs. Based on this statistical information, a piecewise linear model is proposed to accurately estimate the R-D relationship in the FGS layers for the statistical multiplexing system. And then we develop a novel golden-section search algorithm to quickly find the sub-optimal solution for the bit-rate allocation problem. With the proposed scheme, the available channel bandwidth is dynamically allocated and the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

variation of quality of the decoded videos is minimized. In addition, the computational complexity of the system is dramatically reduced comparing with the existing approaches, where computationally re-encoding or transcoding is demanded to adjust the bit-rate of each video program.

7.2 Recommendations for Future Works

In addition to the error control and joint rate allocation approaches discussed in this thesis, there are some interesting extensions for this work which are presented in the following.

- **Hybrid error control for scalable video transmission**

  As introduced in Chapter 1, a number of error control techniques can be used to alleviate the effect of transmission error. In this thesis, we focus on the error resilience and error protection schemes in Chapter 3 and Chapter 4. Different error control technique is employed to provide robustness for transmission of scalable video separately. However, hybrid error control techniques have not been considered. Thus, it would be interesting to investigate the joint implementation of all kinds of error control techniques in order to further enhance the error robustness of the transmitted video. For instance, one promising approach is to combine the UEP scheme and the resynchronization method into a unified framework. Moreover, error concealment approaches could also be jointly considered to further improve the overall performance. It is obviously that, to achieve the optimal performance, the rate allocation problem would be more complicated under such a framework.

- **Cross-layer design for resource allocation in heterogeneous wireless networks**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Traffic carried by future wireless networks is expected to be a mix of real-time traffic such as voice, multimedia conferences and games, and data traffic such as web browsing, messaging and file transfer. All of these applications will require widely varying and very diverse Quality of Service (QoS) guarantees for different types of offered traffic. Traditionally, different protocol layers in the networks are treated as separate entities for specific operations. Different from the traditional works, cross-layer design approaches are critical for efficient utilization of scarce resources with QoS provisioning in the heterogeneous wireless networks [128]. To fully optimize the wireless broadband networks, both the challenges from the physical medium and the QoS demands from the applications have to be taken into account. Rate, power and coding at the physical layer can be adapted to meet the requirements of the applications given the current channel and network conditions. Information has to be exchanged across protocol layers to obtain the highest possible adaptability. Therefore, considerable works on cross-layer design are desired in order to provide efficient end-to-end QoS support for scalable video transmission.

- **Bit-stream extraction for network aware video adaptation**

  H.264/SVC can provide a full scalability comprising temporal, spatial, and quality scalability or any combination of these scalabilities. It allows transmission and decoding of partial streams resulting in lower temporal or spatial resolution or reduced fidelity in order to adapt to the various preferences of end users as well as varying network conditions. An extractor can be used to adapt the bit-stream to the target bit-rate and resolution before the decoder is invoked to decode the extracted partial stream. However, the embedded bit-stream consists of a number of operation points, at which the bit-stream adaptation can be realized. Truncation of the bit-stream at different points will result in different spatio-temporal resolutions and different visual

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

qualities. Given the budget bit-rate, the main objective is to find the optimal operation point that leads to the best visual quality. Although optimal adaptation of the scalable video have been explored using the R-D framework [127], there still lacks in-depth discussion about the selection of the optimal operation point in the cases that the frame size of the extracted stream is not specified. Thus, bit-stream extraction for network aware video adaptation needs further examination.

- **Joint rate allocation among multiple video programs**

  Joint rate allocation algorithm is very important for broadcast systems. In Chapter 6, a novel joint rate allocation algorithm is elaborated for multi-program video coding using FGS. In this work, the available channel bandwidth is dynamically allocated to different video programs to minimize the variation of distortion among multiple videos in the same GOP. However, the quality variation in successive GOPs for the same video has not been taken into account. Since temporal variation may be perceptually annoying, some future works on this issue would be worthy of discussion. Besides, other types of scalabilities, such as spatial scalability and temporal scalability, could also be considered for rate adaptation to further improve the coding efficiency.

Besides the above mentioned extensions, it is also a promising direction to consider the implementation of the proposed schemes based on different video quality evaluation methods. In the current works, objective assessment is used to measure the quality of the perceived video, such as PSNR and MSE. In the future works, subjective quality assessment methods should also be explored. We believe that subjective assessment can constitute a benchmark for evaluation of the performance of objective assessment metrics. By jointly considering both evaluation methods, the overall performance could be further improved.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Publications

Journal Publications

1. Y. Wang, T. Fang, L.-P. Chau, and K.-H. Yap, "Two-dimensional Channel Coding Scheme for MCTF Based Scalable Video Coding," *IEEE Trans. Multimedia*, vol. 9, no. 1, pp. 37-45, Jan. 2007.

2. Y. Wang, L.-P. Chau and K.-H. Yap, "Adaptive resynchronization approach for scalable video over wireless channel," *Journal of Visual Communication and Image Representation*, vol. 21, no. 3, pp. 210-218, Apr. 2010.

3. Y. Wang, L.-P. Chau and K.-H. Yap, "Joint rate allocation for multi-program video coding using FGS", *IEEE Trans. Circuits Syst. Video Technol*., vol. 20, no. 6, pp. 829-837, Jun. 2010.

4. Y. Wang, L.-P. Chau and K.-H. Yap, "Bit-rate allocation for broadcasting of scalable video over wireless networks," *IEEE Trans. Broadcasting*, vol. 56, no. 3, pp. 288-295, Sep. 2010.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Conference Publications

1. Y. Wang, T. Fang, L.-P. Chau and K.-H. Yap, "Two-dimensional Channel Rate Allocation for SVC over Error-prone Channel," in *Proceedings of IEEE International Symposium on Circuits and Systems*, ISCAS'06, Kos, Greece, May 2006.

2. Y. Wang, L.-P. Chau and K.-H. Yap, "A novel resynchronization method for scalable video over wireless channel," in *Proceedings of IEEE International Conference on Multimedia and Expo*, ICME'06, Toronto, Canada, July 2006.

3. Y. Wang, L.-P. Chau and K.-H. Yap, "A motion-based selective error protection method for scalable video over error-prone channel", in *Proceedings of IEEE International Conference on Multimedia and Expo*, ICME'07, Beijing, China, July 2007.

4. Y. Wang, L.-P. Chau and K.-H. Yap, "GOP-based unequal error protection for scalable video over packet erasure channel," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Las Vegas, NV, April 2008.

5. Y. Wang, L.-P. Chau and K.-H. Yap, "Spatial resolution decision in scalable bitstream extraction for network and receiver aware adaptation," in *Proceedings of IEEE International Conference on Multimedia and Expo*, ICME'08, Hannover, Germany, June 2008.

6. Y. Wang, L.-P. Chau and K.-H. Yap, "Broadcast of scalable video over wireless networks," in *Proceedings of IEEE International Symposium on Circuits and Systems*, ISCAS'09, Taipei, Taiwan, May 2009.

7. Y. Wang, L.-P. Chau and K.-H. Yap, "Bit allocation for scalable video coding of multiple video programs," in *Proceedings of IEEE International Conference on Image Processing*, ICIP'10, Hong Kong, Sep. 2010.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Bibliography

[1] J. Chen, U.-V. Koc, and K. J. R. Liu, *Design of Digital Video Coding Systems*. Marcel Dekker, 2001.

[2] Advanced Television Systems Committee, "ATSC digital television standard," Sep. 1995.

[3] U. Reimers, "Digital video broadcasting," *IEEE Communications Magazine.,* vol. 36, no. 6, pp. 104-110, Jun. 1998.

[4] A. J. Stienstra, "Technologies for DVB services on the Internet," *Proc. IEEE*, vol. 94, no. 1, pp. 228-236, Jan. 2006.

[5] M. Nelson and J.-L. Gailly, *The Data Compression Book*. Ed. M&T Books, 1996.

[6] I. Witten, R. Neal, and J. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, pp. 520–540, 1997.

[7] A. G. and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publisher, 1992.

[8] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. New York: Academic, 1990.

[9] L. Debnath, *Wavelet Transform and Their Applications*. Springer, Hardcover, 2000.

[10] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. L. Gall, MPEG *Video Compression Standard*. Chapman and Hall, 1996.

[11] S. J. Choi and J. W. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[12] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 9, pp. 559 – 571, September 1994.

[13] ITU-T SG15, "Video codec for audiovisual services at $p \times 64$ kbit/s," *ITU-T recommendation H.261 Version 3*, Mar. 1993.

[14] JTC 1/SC 29, "Information technology – generic coding of moving pictures and associated audio information: Video," *ISO/IEC 13818-2/ITU-T H.262*, 1996.

[15] ITU-T SG15, "Video coding for low bitrate communication," *ITU-T recommendation H.263*, May 1996.

[16] JTC 1/SC 29, "Information technology - coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s – part 2: Video," *ISO/IEC 11172-2(MPEG-1 Video)*, 1993.

[17] ISO/IEC 14496-2, Coding of Audio-Visual Objects – Part 2: Visual, 2001.

[18] T. Wiegand and G.J. Sullivan, "Draft ITU-T recommendation H.264 and final draft international standard of joint video specification(ITU-T recommendation h.264|ISO/IEC 14496-10 AVC)," Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G050 Pattaya, Thailand, Mar. 2003.

[19] D. Taubman, "Successive refinement of video: fundamental issues, past efforts and new directions", *VCIP 2003*, vol. 5120, July 2003.

[20] "Applications and requirements for scalable video coding," ISO/IEC JTC1/SC29/WG11/N5880, Trondheim, Norway, July 2003.

[21] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Scalable video coding and transport over broadband wireless networks," *Proc. IEEE*, vol. 89, pp. 6-20, Jan. 2001.

[22] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Upper Saddle River, New Jersey: Prentice Hall, 2002.

[23] G. G. Langdon, "An introduction to arithmetic coding," *IBM Journal, Res. Develop.*, vol. 28, no. 2, pp. 135-149, 1984.

[24] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and a new contribution", *Proc. IEEE*, vol. 83, no. 6, pp. 858-876, Jun. 1995.

[25] B.G. Haskell and J.O. Limb, "Predictive video encoding using measured subjective velocity," US Patent no. 3,632,865, Jan. 1972.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[26] K. Ngan, C. Yap, and K. Tan, *Video Coding for Wireless Communication Systems*. New York, NY: Marcel Decker 2001.

[27] K. T. Tan, M. Ghanbari, and D.E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Processing*, vol. 70, no. 3, pp. 279-294, 1998.

[28] B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An introduction to* MPEG-2. Chapam Hall, 1997.

[29] J. Xu, R. Xiong, B. Feng, G. Sullivan, M. Chieh, F. Wu, and S. Li, "3D sub-band video coding using barbell lifting," ISO/IEC JTC1/SC29 WG 11 MPEG2004/M10569/S05, Munich, Germany, Mar. 2004.

[30] B. P. Popescu and V. Bottreau, "Three-dimensional lifting scheme for motion compensated video compression," in *Proc. IEEE Int. Conf. Image Process*., ICIP2001, pp. 1793-1796, Thessaloniki, Greece, Oct. 2001.

[31] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in *Proc. IEEE Int. Conf. Image Process*., ICIP2001, pp. 1029-1032, Thessaloniki, Greece, Oct. 2001.

[32] S. Okuba, K. McCann, and A. Lippman, "MPEG-2 requirements, profile and performance verification," *Signal Processing of* HDTV, Elsevier Science, Amsterdam, pp. 65-73, 1994.

[33] "H.26L test model long term number 9 (TML-9) draft 0," ITU-T, Video Coding Expert Group (VCEG), Dec. 2001.

[34] N. Brady, "MPEG-4 standardized methods for the compression of arbitrarily shaped video objects," *IEEE Trans. Circuits Syst. Video Technol*., vol. 9, no. 8, pp. 1170-1189, 1999.

[35] S. Battista, F. Casalino, and C. Lande, "A multimedia standard for the third millennium," *IEEE Multimedia*, vol. 6, no. 4, pp. 74-83, 1997.

[36] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and L. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol*., vol 13, no. 7, pp. 560-576, Jul. 2003.

[37] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the Internet: approaches and directions," *IEEE Trans. Circuits Syst. Video Technol*., vol. 11, no. 3, pp. 282-300, Mar. 2001.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[38] S. Okubo, "Requirements for high quality video coding standards", *Signal Process. Image Commun.*, vol. 4, pp. 141-151, 1992.

[39] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301-317, Mar. 2001.

[40] H. Schwarz, D. Marpe, and T. Wiegand, "Scalable extension of H.264/AVC," ISO/IEC JTC1/SC29 WG11MPEG2004/M10569/S03, Munich, Germany, Mar. 2004.

[41] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007.

[42] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688-703, Jul. 2003.

[43] M. van der Schaar and H. Radha, "A hybrid temporal-SNR fine-granular scalability for Internet video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 318-331, Mar. 2001.

[44] H. Radha, M. van der Schaar, and Y. Chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 53-68, Mar. 2001

[45] F. Wu, S. Li, and Y.-Q. Zhang, "DCT-prediction based progressive fine granularity scalability," in *Proc. IEEE Int. Conf. Image Process.*, ICIP2000, vol. 3, pp. 556-559, Sep. 2000.

[46] H. Schwarz, D. Marpe, and T. Wiegand, *Hierarchical B pictures*, Joint Video Team, Doc. JVT-P014, Jul. 2005.

[47] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *Proc. ICME*, Toronto, ON, Canada, Jul. 2006, pp. 1929-1932.

[48] J.-R. Ohm, "Advances in scalable video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 42-56, Jan. 2005.

[49] J. Reichel, H. Schwarz, and M. Wien, *Joint Scalable Video Model 11 (JSVM 11)*, Joint Video Team, Doc. JVT-X202, Jul. 2007.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[50] C. A. Segall and G. J. Sullivan, "Spatial scalability within the H.264/AVC scalable video coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1121-1135, Sep. 2007.

[51] H. Schwarz, D. Marpe, and T. Wiegand, *SVC Core Experiment 2.1: Inter-layer Prediction of Motion and Residual Data*, ISO/IEC JTC1/SC29/WG11, Doc. M11043, Jul. 2004.

[52] S. F. Chang and A. Vetro, "Video adaptation: concepts, technologies, and open issues," *Proc. IEEE*, vol. 93, no. 1, pp. 148-158, Jan. 2005.

[53] A. Majumda, D. G. Sachs, I. V. Kozintsev, K. Ramchandran, and M. M. Yeung, " Multicast and unicast real-time video streaming over wireless LANs," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 524-534, Jun. 2002.

[54] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. N. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *Proc. IEEE*, vol. 93, no. 1, pp. 135-147, Jan. 2005.

[55] A. Albanese, J. Blömer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737-1744, Nov. 1996.

[56] M. Srinivasan and R. Chellappa, "Adaptive source-channel subband video coding for wireless channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 9, pp. 1830-1839, Dec. 1998.

[57] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint souce-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1043-1052, Sep. 2002.

[58] M. Mohr, E. A. Riskin, and R. E. Ladner, "Unequal loss protection: graceful degradation of image quality over packet erasure channels through forward error correction," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 819-828, Jun. 2000.

[59] M. van der Schaar and H. Radha, "Unequal packet loss resilience for fine-granular-scalability video," *IEEE Trans. Multimedia*, vol. 3, no. 4, pp. 381-394, Dec. 2001.

[60] U. Horn, K. W. Stuhlmüller, M. Link, and B. Girod, "Robust internet video transmission based on scalable coding and unequal error protection," *Signal Process. Image Commun.*, vol. 15, pp. 77-94, Sep. 1999.

[61] G. Cheung and A. Zakhor, "Bit allocation for joint source/channel coding of scalable video", *IEEE Trans. Image Processing*, vol. 9, no. 3, pp. 340-356, Mar. 2000.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[62] J. Kim, R. M. Mersereau, and Y. Altunbasak, "A multiple-substream unequal error-protection and error-concealment algorithm for SPIHT-coded video bitstreams," *IEEE Trans. Image Processing*, vol. 13, no. 12, 1547-1553, Dec. 2004.

[63] L.-J. Cheng, W.-J. Zhang and L. Chen, "Rate-distortion optimized unequal loss protection for FGS compressed video," *IEEE Trans. Broadcasting*, vol. 50, pp. 126-131, Jun. 2004.

[64] D. G. Sachs, R. Anand and K. Ramchandran, "Wireless image transmission using multiple-description based concatenated codes," in *Proc. SPIE'00*, vol.3974, pp.300-311, Apr. 2000.

[65] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison Wesley, Dec. 1988.

[66] S. H. Lee, P. J. Lee, and R. Ansari, "Cell loss detection and recovery in variable rate video," in *Proc. 3rd Int. Workshop Packet Video*, Morriston, Mar. 1990.

[67] K. Challapali, X. Lebegue, J. S. Lim, W. H. Paik, R. Saint Girons, E. Petajan. V. Sathe, P. A. Snopko, and J. Zdepski, "The grand alliance system for US HDTV," *Proc. IEEE*, vol. 83, no. 2, pp. 158-174, Feb. 1995.

[68] T. Kinoshita, T. Nakahashi, and M. Maruyama, "Variable bit rate HDTV CODEC with ATM cell loss compensation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 3, pp. 230-237, Jun. 1993.

[69] V. Parthasarathy, J. W. Modestino, and K. S. Vastola, "Design of a transport coding scheme for high quality video over ATM networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 358-376, Apr. 1997.

[70] J.-Y. Cochennec, "Method for the correction of cell losses for low bit-rate signals transport with the AAL type 1," ITU-T SG15 Doc. AVC-538, Jul. 1993.

[71] E. Ayanoglu, R. Pancha, and A. R. Reibman, and S. Talwar, "Forward error control for MPEG-2 video transport in a wireless ATM LAN," *ACM/Baltzer Mobile Networks Applicat.*, vol. 1, no. 3, pp. 245-258, Dec. 1996.

[72] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proc. IEEE*, vol. 86, no. 5, pp. 974-997, May 1998.

[73] D. Wu, Y. T. Hou, and Y.-Q. Zhang, "Transporting real-time video over the Internet: challenges and approaches," *Proc. IEEE*, vol. 88, no. 12, pp. 1855-1877, Dec. 2000.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[74] J. T. H. Chung-How and D. R. Bull, "Robust H.263+ video for real-time Internet applications," in *Proc. IEEE Int. Conf. Image Process.*, ICIP2000, vol. 3, pp. 544-547, Sep. 2000.

[75] C. Huang and S. Liang, "Unequal error protection for MPEG-2 video transmission over wireless channels," *Signal Process. Image Commun.*, vol. 19, pp. 67-79, Jan. 2004.

[76] F. Marx and J. Farah, "A novel approach to achieve unequal error protection for video transmission over 3G wireless networks," *Signal Process. Image Commun.*, vol. 19, pp. 313-323, Apr. 2004.

[77] X.-K. Yang, C. Zhu, Z. G. Li, X. Lin, G. N. Feng, S. Wu, and N. Ling, "Unequal loss protection for robust transmission of motion compensated video over the Internet', *Signal Processing, Image Commun.*, vol. 18, pp. 157-167, 2003.

[78] L. Rizzo, "Effective erasure codes for relialble computer communication protocols," *ACM Computer Communication Review*, vol. 27, pp. 24-36, Apr. 1997.

[79] M. Zorzi, R. R. Rao, and L. Milstein, "On the accuracy of a first-order Markov model for data transmission on fading channels," in *Proc. IEEE Int. Conf. Universal Personal Commun.*, ICUPC95, pp. 211-215, Nov. 1995.

[80] E. O. Elliott, "A model of the switched telephone network for data communications," *Bell. Syst. Techn. J.*, pp. 89-109, Jan. 1965.

[81] Z. He, J. Cai, and C. W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 511-523, Jun. 2002.

[82] MPEG Committee, SVM 3.0 Software, ISO/IEC JTC 1/SC 29/WG 11 N6717, Palma de Mallorca, Spain, Oct. 2004.

[83] B. Sklar, "Rayleigh fading channels in mobile digital communication systems, part I: characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90-100, Jul. 1997.

[84] Y. Wang, S. Wenger, J. Wen, and A. K. Katsagellos, "Error resilient video coding techniques," *IEEE Signal Processing Magazine*, vol. 17, no. 4, pp. 61-82, Jul. 2000.

[85] T. J. Ferguson and J. H. Rabinowitz, "Self-synchronizing Huffman codes," *IEEE Trans. Infom. Theory*, vol. IT-30, no. 4, pp. 687-693, Jul. 1984.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[86] J. D. Villasenor, Y.-Q. Zhang, and J. Wen, "Robust video coding algorithms and systems," *Proc. IEEE*, vol. 87, no. 10, pp. 1724-1733, Oct. 1999.

[87] K.-Y. Yoo, "Adaptive resynchronization marker positioning method for error resilient video transmission," *Electronics Letters*, vol. 34, no. 22, pp. 2084-2085, 1998.

[88] S.-H. Lee and J.-K. Kim, "Optimisation-based placement of resynchronization marker for error robust video transmission," *Electronics Letters*, vol. 37, no. 6, pp. 348-350, 2001.

[89] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE Journal on Selected Areas in Commun.*, vol. 18, no. 6, pp. 952-965, Jun. 2000.

[90] L. O.-Barbosa and T. Han, "On the use of frame-based slice size for the robust transmission of MPEG video over ATM networks," *IEEE Trans. Broadcasting*, vol. 46, no. 2, pp. 134-143, Jun. 2000.

[91] O. Harmanci and A. M. Tekalp, "A stochastic framework for rate-distortion optimized video coding over error-prone networks," *IEEE Trans. Image Processing*, vol. 16, no. 3, pp. 684-697, Mar. 2007.

[92] T. Fang and L.-P. Chau, "Efficient content-based resynchronization approach for wireless video," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1021-1027, Dec. 2005.

[93] R. Yan, F. Wu, S.-P. Li, and R. Tao, "Error resilience method for FGS video enhancement bitstream," in *IEEE Pacific-Rim Conference on Multimedia* (PCM), 2000.

[94] MPEG Committee, JSVM 9.0 Software, ISO/IEC JTC 1/SC 29/WG 11 W8752.

[95] Network simulator 2 (Ns-2), available: http://www.isi.edu/nsnam/ns.

[96] T.-W. A. Lee, S.-H. G. Chan, Q. Zhang, W.-W. Zhu and Y.-Q. Zhang, "Allocation of layer bandwidths and FECs for video multicast over wired and wireless networks," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 12, no. 12, pp. 1059-1070, Dec. 2002.

[97] T. Berger, Rate Distortion Theory, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[98] L. Böröczky, A. Y. Ngai, and E. F. Westermann, "Statistical multiplexing using MPEG-2 video encoders," *IBM Journal of Research and Development*, vol. 43, no. 4, pp. 511-520, Jul. 1999.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[99] L. Wang and A. Vincent, "Joint rate control for multi-program video coding," *IEEE Trans. Consumer Electron.*, vol. 42, no. 3, pp. 300-305, Aug. 1996.

[100] G. Keesman, "Multi-program video compression using joint bit-rate control," *Philips J. Res.*, vol. 50, no. 1/2, pp. 21-45, 1996.

[101] J. Yang, X. Fang and H. Xiong, "A joint rate control scheme for H.264 encoding of multiple video sequences," *IEEE Trans. Consumer Electron.*, vol.51, no. 2, pp. 617-623, May 2005.

[102] M. Perkins and D. Arnstein, "Statistical multiplexing of multiple MPEG-2 video programs in a single channel," *SMPTE J.*, vol. 104, no. 9, pp. 596-599, 1995.

[103] A. Guha and D. J. Reininger, "Multichannel joint rate control of VBR MPEG encoded video for DBS applications," *IEEE Trans. Consumer Electron.*, vol. 40, no. 3, pp. 616-623, Aug. 1994.

[104] L. Böröczky, A. Y. Ngai, and E. F. Westermann, "Joint rate control with look-ahead for multi-program video coding," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 10, no. 7, pp. 1159-1163, Oct. 2000.

[105] M. Balakrishan and R. Cohen, "Global optimization of multiplexed video encoders," in *Proc. ICIP'97*, vol. 1, pp. 377-380, Santa Barbara, CA, Oct. 1997.

[106] M. Jacobs, S. Tondeur, T. Paridaens, J. Barbarien, R. Van de Walle, and P. Schelkens, "Statistical multiplexing using SVC", *IEEE Broadband Multimedia*, 2008.

[107] E. N. Linzer and A. Wells, "Statistical multiplexed video encoding using pre-encoding a priori statistics and a priori and a posteriori statistics," U.S. Patent 6094457, Jul. 2000.

[108] L. Wang and A. Luthra, "Dynamic bit allocation for statistical multiplexing of compressed and uncompressed digital video signals," U.S. Patent 6167084, Dec. 2000.

[109] L. Böröczky, A. Y. Ngai, and E. F. Westermann, "Adaptively encoding multiple streams of video data in parallel for multiplexing onto a constant bit rate channel," U.S. Patent 6859496, Feb. 2005.

[110] L. Böröczky, A. Y. Ngai, and E. F. Westermann, "Control strategy for dynamically encoding multiple streams of video data in parallel for multiplexing onto a constant bit rate channel," U.S. Patent 6956901, Oct. 2005.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[111] T. Paridaens, D. De Schrijver, W. De Neve, and R. Van de Walle, "XML-driven bitrate adaptation of SVC bitstreams," *WIAMIS*, 2007.

[112] T. F. Edgar and D. M. Himmelblau, Optimization of chemical processes, McGraw-Hill, New York, 1988.

[113] J. Sun, W. Gao, D. Zhao, and W. Li, "On rate-distortion modeling and extraction of H.264/SVC fine-granular scalable video," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 19, no. 3, pp. 323-336, Mar. 2009.

[114] W. J. Kim, J. W. Yi, and S. D. Kim, "A bit allocation method based on picture activity for still image coding," *IEEE Trans. Image Processing.*, vol. 8, no. 7, pp 974-977, Jul. 1999.

[115] L. Hanzo, P. Cherriman and J. Streit, Wireless video communications: second to third generation systems and beyond, *IEEE Press*, 2001.

[116] D. G. Sachs, I. Kozintsev, M. Yeung, and D. L. Jones, "Hybrid ARQ for robust video streaming over wireless LANs," in *Proc. IEEE ITCC'01*, Las Vegas, NV, Apr. 2001, pp. 317-321.

[117] S. Mccanne, V. Jacobson and M. Vetterli, "Receiver-driven layered multicast," in *Proc. ACM SIGCOMM*, 1996, pp. 117-130.

[118] L. Vicisano, L. Rizzo and J. Crowcroft, "TCP-like congestion control for layered multicast data transfer," in *Proc. IEEE INFOCOM*, 1998, pp. 996-1003.

[119] J. Liu, B. Li and Y.-Q. Zhang, "Optimal stream replication for video simulcasting," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 162-169, Feb. 2006.

[120] B. J. Vickers, C. Albuquerque and T. Suda, "Adaptive multicast of multi-layered video: rate-based and credit-based approaches," in *Proc. IEEE INFOCOM*, 1998, pp. 1073-1083.

[121] C.-H. Hsu and M. Hefeeda, "Optimal coding of multilayer and multiversion video streams," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 121-131, Jan. 2008.

[122] Q. Zhang, Q. Guo, Q, Ni, W. Zhu and Y.-Q. Zhang, "Sender-adaptive and receiver-driven layered multicast for scalable video over the Internet," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 15, no. 4, pp. 482-495, Apr. 2005.

[123] W.-T. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 11, no. 3, pp. 373-386, Mar. 2001.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[124] T. Fang and L.-P. Chau, "GOP-based Channel Rate Allocation Using Genetic Algorithm for Scalable Video Streaming over Error-prone Networks", *IEEE Trans. Image Processing*, vol. 15, no. 6, pp. 1323-1330, Jun. 2006.

[125] T.-W. A. Lee, S.-H. G. Chan, Q. Zhang, W.-W. Zhu and Y.-Q. Zhang, "Allocation of layer bandwidths and FECs for video multicast over wired and wireless networks," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 12, no. 12, Dec. 2002.

[126] T. Schierl, H. Schwarz, D. Marpe and T. Wiegand, "Wireless broadcasting using the scalable extension of H.264/AVC," in *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 884-887, July 2005.

[127] I. Amonou, N. Cammas, S. Kervadec, and S. Pateux, "Optimized rate-distortion extraction with quality layers in the scalable extention of H.264/AVC," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17, no. 9, pp. 1186-1193, Sep. 2007.

[128] H. Jiang, W. Zhuang, and X. Shen, "Cross-layer design for resource allocation in 3G wireless networks and beyond," *IEEE Communications Magazine*, vol. 43, no. 12, pp.120-126, Dec. 2005.

[129] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, Cambridge, MA: MIT Press, 1990.

[130] S. Winkler, *Digital video quality: vision models and metrics*, John Wiley & Sons Ltd, Mar. 2005.

[131] J. M. Pena, J. A. Lozano, and P. Larranaga, "An empirical comparison of four initialization methods for the K-means algorithm," *Pattern Recognition Letters*, vol. 20, pp. 1027-1040, 1999.