Issues in multiple description coding : algorithm design, analysis and applications

Liu, Ming Lei

2010

Liu, M. L. (2010). Issues in multiple description coding : algorithm design, analysis and applications. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/20858

https://doi.org/10.32657/10356/20858

Issues in Multiple Description Coding: Algorithm Design, Analysis and Applications

LIU MINGLEI

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

Acknowledgement

First and foremost, I offer my sincerest gratitude to my supervisor, Dr. Zhu Ce, for his professional guidance, invaluable support and kindly encouragement during my Ph. D. study as well as his constructive comments on this thesis.

I owe many thanks to Dr. Wu Xiaolin for valuable discussions. I would like to thank Mr. Huang Xiang, Dr. Jan Ostergaard and Mr. Lin Chunyu for their kindly help.

I am also grateful to my friends and colleagues at the Media Technology Lab for the stimulating and friendly atmosphere I enjoyed much. Especially, I thank Mr. Zhang Yixuan and Ms. Xu Yuanyuan for their collaboration in the last years, for their support and help in solving problems.

This thesis is dedicated to my parents, my wife, my daughter and my elder brother. Their love and inspiration are always with me.

Summary

In the modern packet-switched networks, an information source can be transmitted from multiple servers to multiple clients via different paths. However, current networks do not provide a guaranteed performance in terms of bandwidth or delay and therefore the desired quality of service may not be achieved. Especially for some real-time services, such as voice over Internet (VoiP) and video conferencing which require high bandwidth, low delay and low packet-loss rates, the problem of reliable delivering is a significant issue.

Multiple description coding (MDC) is a promising technique to support error resilient multimedia communications with low delay over lossy networks such as the Internet and wireless networks. In an MDC implementation, an encoder generates M descriptions (or streams) for one source. The streams are transmitted from one or multiple servers to a receiver over separate channels (paths) respectively, where each channel may have its own constraints. At the receiver, depending on the number of descriptions received correctly, different reconstruction qualities will be obtained. Due to source diversity and path diversity, MDC can adapt to the network conditions much better than single description coding, especially in real-time communications where retransmission may not be viable.

While the design, analysis and applications of MDC have been widely studied, this thesis addresses MDC from a source coding point of view. Specifically,

three issues are considered, namely, **i)** index assignment design and analysis for multiple description lattice vector quantization, **ii)** two-stage multiple description quantization and its application in image coding, and **iii)** a general multiple description video coding framework based on hierarchical B pictures.

Firstly (in **Chapter 3**), we investigate the design of symmetric entropy-constrained multiple description lattice vector quantization (MDLVQ). We show that with the lattice partition, the complicated MDLVQ index assignment design can be reduced to resemble the transportation problem in operations research. Both greedy and general algorithms are developed to pursue optimality of the index assignment for M-description ($M > 2$) LVQ taking into account computational complexity. Under the high-resolution assumption, the proposed schemes show similar results in terms of optimality with lower complexity compared against other relevant techniques. We also derive an asymptotical closed-form expression of side distortions, which is considered to be a comparable analytical result as opposed to other known asymptotical results.

Secondly (in **Chapter 4**), we consider enhancing the coding efficiency of two-stage two-description quantization, which features an efficient and flexible mechanism to control the central-side distortion tradeoff. We show that the second stage refinement information, i.e. the quantized residual errors, can be used to further reduce the side distortions as well as improving the central distortion. Based on this observation, we propose two enhancement schemes using two-stage two-description scalar quantization to make the product of central and side distortions closer to the rate-distortion bound of two-description coding under the high-resolution assumption. Meanwhile, we propose a practical two-stage two-description image coding based on trellis coded quantization, which is comparable

(or usually superior) to state-of-the-art two-description image coding schemes.

Thirdly (in **Chapter 5**), a general multiple description video coding (MDVC) framework based on hierarchical B pictures is proposed. By distributing key pictures amongst separate descriptions, a stable MDVC platform is generated that enables flexible tradeoffs between central and side distortions. This proposed MDVC framework is H.264/AVC-compliant for each temporal scalable description. Certain existing temporal-splitting MDVC techniques can be considered as reduced cases of the proposed structure.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction to Multiple Description Coding

1.1 Background

TELECOMMUNICATION services have pervaded our every-day lives in recent decades. Simultaneously, ubiquitous sensors, devices and networks are paving the way towards an information society in which multimedia data (textual data, audio/speech, image, video and animation) are distributed throughout the physical environment as reliable and relevant communication services for people. Along with the greatly decreasing cost of acquisition and storage devices for multimedia data, rapid developments have taken place in multimedia representation, processing and communications. On the other hand, compared with the traditional point-to-point communications via a fixed route, communications in the heterogeneous communication infrastructure of modern packet-switched networks are more complicated. For example, an information source may be transmitted from

multiple servers to multiple clients via different paths. Specifically, in a best-effort network such as the Internet, all clients receive best-effort services, i.e., they may obtain unspecified and variable bit rates and delivery times depending on the prevailing traffic load. The characteristics of a best-effort network imply that it cannot provide a guaranteed performance in terms of bandwidth or delay. The problem of reliability is more serious for some real-time services through the Internet, such as voice over Internet (VoiP), video conferencing and streaming multimedia, which requires high bandwidth, low delay and low packet-loss rates.

Furthermore, traditional source representations for point-to-point communications are not sufficiently powerful in such a diversity-based network, where the server diversity and path diversity enable more efficient use of network bandwidths for combating network congestion and channel errors. A more effective source representation should utilize the diversity property of the network to combat the errorprone channels. In real-time applications, this new representation is required to improve transmission reliability without relying on standby links.

Multiple description coding (MDC) is a promising technique which supports error resilient multimedia communications over lossy channels such as the Internet and wireless networks. In an MDC implementation, an encoder generates M descriptions (or streams) for one source. The streams are transmitted from one or multiple servers to a receiver over separate channels (paths) respectively, where each channel may have its own constraints. At the receiver, depending on the number of descriptions received correctly, different reconstruction qualities will be obtained. To be more specific, if only k out of a total of M streams are received, the reconstruction quality associated with a so-called k-description side distortion is expected to be acceptable, and an incremental improvement can be achieved

with a reduced side distortion if more streams are received. When all streams are available, the best reconstruction quality is obtained, corresponding to the smallest central distortion. Exploiting the source diversity and path diversity, MDC can adapt to the network conditions much better than single description coding, especially in real-time communications where retransmission is not always viable.

Source coding performance depends intimately on the content of the original signal. Some basic techniques for single-description based source coding include prediction, quantization, frequency transformation and entropy coding. MDC can also be built from these basic techniques along with channel codes for joint source and channel coding. Layered coding (LC) is another kind of source coding techniques for erasure resilient, scalable transmission. In contrast to a conventional media coder that generates a single stream, LC and MDC encode a media source into two or more sub-streams. In the structure of LC, one base layer stream and several enhancement layer streams are generated. The base layer can be decoded individually to provide a basic quality of the source while the enhancement layers are mainly used to refine the quality of the source based on the base layer. In the case of base layer being corrupted, the enhancement layers become useless, even if they are received correctly. Therefore, the base layer is the most important and is usually protected using error correction techniques such as automatic repeat request (ARQ) or forward error correction (FEC) codes.

The comparison of MDC and LC in different scenarios have been carefully studied in recent years. Reibman et al. analyzed and compared MDC with LC+FEC on both a binary symmetric channel and a random erasure channel in [6]. The rate-distortion performance of MDC vs. LC+ARQ for video coding

over wireless networks with multiple path routing is discussed in [7]. Scenarios of MDC+ARQ/FEC and LC+ARQ/FEC are carefully compared by using two different video codes in multi-path environments at various packet loss probabilities in [8]. A good summary for comparison of MDC and LC can be found in [9]. Based on above literatures, some conclusions for MDC and LC can be summarized as follows. Although MDC and LC have some aspects in common, such as the sub-streams representation and the intrinsic scalability, they may differ in the capability of error resilience and the dependence on a feed-back channel. Generally speaking, MDC outperforms LC in the networks with no feedback or with very stringent delay constraints. However, if the networks or the applications support prioritized transmission or error control, LC could be better than MDC. MDC and LC perform similarly at low error probabilities when they are both combined with error protection coding, while MDC generally shows better results than LC at high error probabilities. Recently, to achieve both robustness to unreliable channels and adaptivity to client bandwidth heterogeneity and network congestion, layered multiple description coding has been developed [10–12], which has the advantages of both layered coding and multiple description coding. As has been stated, MDC is an efficient error-resilient encoding technique. To the best of the author's knowledge, MD coding has been successfully applied to some commercial softwares such as 'Skype'[i].

[i] 'Skype' is a software application that allows users to make voice calls over the Internet. Additional features of this software include instant messaging, file transfer and video conference. The homepage of 'Skype' is http://www.skype.com/intl/en/.

1.2 Existing Problems

While the design, analysis and applications of MDC have been widely studied in recent years, many challenging problems remain open. In this thesis, we address the MD problem from a source coding point of view. Specifically, we carefully examine some major issues in MDC. They are described below.

Firstly, it is known that multiple description scalar quantization (MDSQ) [13] is a simple and flexible scheme to generate two representations for a symbol. The tradeoff between the central distortion and side distortion can be controlled in MDSQ through index assignment design. Although it is easy to extend the formalism of MDSQ to vectors, the index assignment problem becomes much more difficult because the code vectors cannot be naturally ordered. Moreover, the encoding complexity increases substantially with dimension. Multiple description lattice vector quantization (multiple description LVQ or MDLVQ) [14, 15] is an elegant technique which helps to avoid these difficulties. It is known that under the high-resolution assumption and with a sufficiently large vector dimension, the information theoretic rate-distortion bounds for two-description coding can be approximated by the two-description LVQ for memoryless Gaussian source. The index assignment design for two-description LVQ has been intensively studied by many researchers with interesting results. While for a general case involving M ($M > 2$) descriptions, the problems in terms of optimality and complexity in the MDLVQ index assignment design, which will be investigated in the thesis, are still challenging.

Secondly, compared with the one-stage multiple description quantization (MDQ), the two-stage MDQ [16–18] provides a more efficient and flexible mechanism to control the central-side distortion tradeoff. A typical two-stage MDQ is two-

stage MDSQ with uniform scalar quantizers. However, the asymptotic analysis shows that the distortion product of the central and side distortions for such two-stage MDSQ is 3.066 dB away from the rate-distortion bound for memoryless Gaussian source. We will show that this gap can be reduced by enhancing the structure of two-stage MDQ. In addition, the applications of different two-stage MDQ schemes in image coding have also shown very good rate-distortion results. When applying to the image coding in the range of medium/low bit rates, the two-stage MDQ needs to be adapted since high-resolution analytical results on Gaussian source may not be valid.

Thirdly, in the video coding application, many multiple description video coding (MDVC) algorithms are developed based on pre-processing, e.g., sub-sampling in spatial or temporal domain [2, 19, 20]. The major advantage of the pre-processing based MDC methods is that the substantive encoding process is untouched, thus facilitating all the existing standard codecs of high coding efficiency to be conveniently incorporated in coding each description with little or no change. However, the performance of these temporal or spatial subsampling based MDVC algorithms highly depends on how to perform the splitting as well as how to estimate missing descriptions for a better reconstruction. The intra-description redundancy minimization and good inter-description redundancy control are the major concerns in the design of a good MDVC algorithm, which will be discussed in the thesis.

1.3 Outline and Contributions

An overview of the multiple description coding is presented in Chapter 2. The main contributions of this thesis are presented in Chapter 3 (multiple description lattice vector quantization), Chapter 4 (two-stage multiple description quantization and its application in image coding) and Chapter 5 (multiple description video coding based on hierarchical B pictures). The final conclusion is drawn in Chapter 6 along with several future research topics.

In **Chapter 3**, we investigate the index assignment design of symmetric entropy-constrained multiple description lattice vector quantization. We consider a fine lattice containing clean similar sublattices with S-similarity. Due to the S-similarity of the sublattices, an M-fraction lattice can be used to regularly partition the fine lattice with smaller Voronoi cells than a sublattice does. With the partition, the MDLVQ index assignment design can be boiled down to the transportation problem in operations research. Both greedy and general algorithms are developed to pursue optimality of the index assignment. Under the high-resolution assumption, we compare the proposed schemes with other relevant techniques in terms of optimality and complexity. We also derive an asymptotical closed-form expression of side distortions, which is considered to be a comparable or better analytical result as opposed to other known asymptotical results. Simulation results on coding different sources of Gaussian, speech and image are presented to validate the effectiveness of the proposed schemes.

In **Chapter 4**, we consider enhancing coding performance of two-stage two-description quantization. We propose two enhancement schemes to improve the coding efficiency. First, an enhancement scheme based on two-stage multiple description scalar quantization is proposed to make the product of central and

side distortions closer to the rate-distortion bound of multiple description coding under the high-resolution assumption. We show analytically that the second stage refinement information, i.e. the quantized residual errors, can be used to further reduce the side distortions apart from the central distortion, which is substantiated with a memoryless Gaussian source. Second, we discuss the case that a good lattice quantizer is used to replace the scalar quantizer in the second stage for reducing the central distortion further. Meanwhile, we propose a practical two-stage multiple description image coding based on trellis coded quantization, which is comparable (often superior) to state-of-the-art MD image coding schemes.

In **Chapter 5**, a general multiple description video coding framework based on hierarchical B pictures is proposed. Two or more descriptions are generated by employing the hierarchical B pictures of H.264/AVC scalable extension, where temporal-level-based key pictures are selected in a staggered way among different descriptions. Based on this hierarchical and staggered structure, inter-description redundancy control is studied to achieve a central/side-distortion-rate tradeoff. Moreover, to better exploit multiple complementary descriptions, a linear combination based on received descriptions is employed to optimize decoding results. This proposed MDVC framework is H.264/AVC-compliant for each temporal scalable description. Some existing temporal-splitting MDVC techniques can be considered as a reduced case in the proposed structure. Experimental results validate the effectiveness of the proposed design for MDVC.

Chapter 2

Literature Review

2.1 Introduction

MULTIPLE description coding (MDC) is a source coding scheme designed to combat transmission errors, especially applicable in non-prioritized lossy networks where retransmission may fail to work. A multiple description encoder generates two or more streams (known as descriptions) for one source. These streams are transmitted over separate channels, and each can be individually decoded with a certain level of fidelity[i]. Since these streams are also designed to complement each other, incremental reconstruction quality is achievable with more received descriptions. In MDC, side distortion and central distortion are defined to measure reconstruction quality when part and all descriptions are received respectively. In contrast with traditional single description coding which

[i]In the multiple description problem one description (or stream) is transmitted over one channel. Therefore we do not strictly distinguish 'M-channel' from 'M-description' in this thesis. More specifically, the expression of 'M-channel multiple description coding' is equivalent to 'M-description coding'.

aims to achieve optimal rate-distortion performance, MDC design needs to consider both central and side distortions given a total bit rate. However, it is known that the minimization of central and side distortions conflict each other, and their trade-off is controlled by redundancy among the different descriptions. Generally speaking, more redundancy will favor side distortion but result in less improvement of central distortion over side distortion. Therefore MDC must strike a balance between minimizing the central and side distortions for a given rate, requiring a good design for the redundancy control.

In this chapter, we first provide a brief overview of single and multiple description rate-distortion theory. Next we review the existing MD methods with applications to audio/speech, image and video sources. The MDC application to watermarking for multimedia security is also reviewed in brief.

2.2 Single Description Rate-Distortion Theory

Information theory is the foundation of modern digital communications. In [21] and [22], Shannon first introduced the fundamental concepts of information theory and established well known channel coding and source coding theorems. Channel coding is a viable approach to reduce information rate through the channel and provide the increased reliability. Channel coding theorems define the channel capacity, which is the maximum information which can be transmitted without error per channel used. Source coding, i.e. lossless or lossy data compression, is the process of encoding information using fewer bits than an unencoded representation would use, through employing specific encoding schemes. Source coding theorems give the rate limitations of data compression. In the

lossless case, the lowest rate is the entropy of the source. In the lossy case, the rate-distortion function bounds the lowest rate needed for a given distortion. A fundamental problem of rate-distortion theory is to describe the required rate R to encode a source X at a prescribed distortion (fidelity) level D.

Assume that a source X produces a sequence of independent identically distributed (i.i.d.) random variables (or letters) $X^L = \{X_i\},\ i = 1,\ \cdots,\ L$. Denote $\hat{X}$ as the reproduction of X. Similarly, let x, $\hat{x}$ and x^L be realizations of X, $\hat{X}$ and X^L, respectively. An encoder f maps an L-dimensional source vector $x^L = (x_1, \cdots, x_L)$ to an index (or codeword) $f(x^L) \in \{1, \cdots, 2^{LR}\}$, where R is the bit rate per symbol. A decoder g maps an index from $\{1, \cdots, 2^{LR}\}$ to a reproduction sequence $\hat{x}^L = g(f(x^L)) = (\hat{x}_1, \cdots, \hat{x}_L)$.

Denote $d(x, \hat{x})$ as the non-negative distortion measure (or fidelity criterion) to evaluate the approximation between the source x and its reproduction $\hat{x}$. The squared-error measure $d(x, \hat{x}) = (x - \hat{x})^2$, or the squared-error single-letter fidelity criterion, is widely used due to its simplicity and its relationship to the least square prediction. Under the squared-error measure, the distortion between sequences x^L and $\hat{x}^L$ is defined as

$$d(x^L, \hat{x}^L) = \frac{1}{L}\sum_{i=1}^{L} d(x_i, \hat{x}_i) = \frac{1}{L}\sum_{i=1}^{L} (x_i - \hat{x}_i)^2 \tag{2.2.1}$$

The distortion associated with the source code (f, g) is defined as the expected distortion between X^L and $\hat{X}^L$, that is

$$D = E[d(X^L, \hat{X}^L)] = E[d\left(X^L, g(f(X^L))\right)] \tag{2.2.2}$$

A rate-distortion pair (R, D) is said to be achievable if there exits a source

code (f, g) with the rate R and the distortion D such that [23]

$$\lim_{L\to\infty} E[d\left(X^L, g(f(X^L))\right)] \leq D \tag{2.2.3}$$

The rate-distortion region for a source is the closure of the set of achievable rate-distortion pairs (R, D). The rate-distortion function $R(D)$ is the infimum of all rates R for a given D in the rate-distortion region. In dual, the distortion-rate function $D(R)$ is the infimum of all distortions D for a given R in the rate-distortion region [23].

The information rate-distortion function $R(D)$ for a source X with distortion $d(x, \hat{x})$ is defined by [23]

$$R(D) = \min_{p(\hat{x}|x):\sum p(x)p(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X}) \tag{2.2.4}$$

where the mutual information $I(X;\hat{X})$ is defined as

$$I(X;\hat{X}) \triangleq \sum_{x,\hat{x}} p(x,\hat{x}) \log \frac{p(x,\hat{x})}{p(x)p(\hat{x})} \tag{2.2.5}$$

Generally, it is difficult to obtain the rate-distortion region. While for a Gaussian memoryless i.i.d. source with zero mean and variance σ^2, the rate-distortion function with squared-error measure is given by

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases} \tag{2.2.6}$$

and the inverse distortion-rate function is

$$D(R) = \sigma^2 2^{-2R} \tag{2.2.7}$$

The most difficult source to compress is Gaussian source, which requires the greatest number of bits to achieve the same distortion among different sources with the same variance. For a memoryless continuous-valued source with variance σ^2 and differential entropy $h(p)$, its distortion-rate function under squared-error criteria is bounded by [22] [24]

$$\frac{1}{2\pi e} 2^{2h(p)} 2^{-2R} \leq D(R) \leq \sigma^2 2^{-2R} \tag{2.2.8}$$

where the differential entropy $h(p)$ is determined by the source probability density function (pdf) $p(x)$ [23]

$$h(p) \triangleq \int p(x) \log_2 p(x) dx \tag{2.2.9}$$

For a memoryless Gaussian sources with variance σ^2, $h(p)$ is given by

$$h(p) = \frac{1}{2} \log(2\pi e \sigma^2) \tag{2.2.10}$$

For a memoryless source other than memoryless Gaussian source, the Shannon lower bound $\frac{1}{2\pi e} 2^{2h(p)} 2^{-2R}$ is generally strictly less than $D(R)$ of any $R > 0$. The bound becomes asymptotically tight in the high rate limit [24] [25]

$$\lim_{R \to \infty} D(R) = \frac{1}{2\pi e} 2^{2h(p)} 2^{-2R} \tag{2.2.11}$$

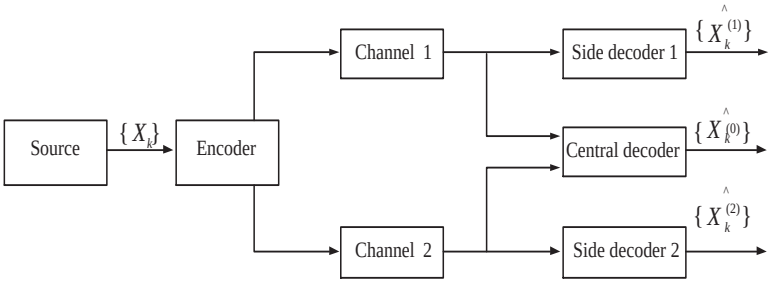


Figure 2.1: Two-channel multiple description coding system.

2.3 Multiple Description Rate-Distortion Region

The multiple description problem is concerned with lossy encoding of information for transmission over an unreliable M-channel communication system. The problem is to design an multiple description system given rate constraints for each channel and minimize the distortions based on the received information of any subsets of channels. Each description in MDC is often described as a packet in the packet-switched network, where individual packets are either untouched or totally lost. The classical MDC involves two descriptions with three receivers as shown in Fig. 2.1. Denote R_{s1} as the rate for description 1 and R_{s2} for description 2. The distortion observed at the receiver depends on which descriptions arrive. If both descriptions are available, the central distortion D_c is lower than if only one description is received (D_{s1} or D_{s2}). The structure of 2-description coding can be easily extended to a general case involves M ($M \geq 2$) descriptions.

2.3.1 Two-Description Rate-Distortion Results

El Gamal and Cover [26] constructed an achievable rate region (denoted as the EGC region) with a memoryless source and a certain fidelity criterion based on Shannon's random coding approach. Ozarow [27] has shown that the EGC region

is the rate-distortion region for the case of a memoryless Gaussian source with respect to the squared-error measure. In [28] and [29], it is shown that the inner bounds[ii] in [26] are also tight for general sources and distortion measures in the no excess sum rate case (i.e. $D_c = D(R_{s1} + R_{s2})$). Nevertheless, Zhang and Berger [29] also showed that in the excess sum rate case, the EGC region is not always tight for the binary memoryless source under the Hamming distortion measure.

Outer bounds[iii] for the binary symmetric source and Hamming distortion have also been obtained in [31], [32] and [33]. Zamir [34] [35] obtained both inner and outer bounds for smooth stationary sources and the squared-error fidelity criterion, where the bounds become tight at high resolution. Linder, Zamir and Zeger [36] have obtained high-resolution bounds for smooth sources and locally quadratic distortion measures. Recently, Feng and Effros [37] and Lastras-Montrio and Castelli [38] have shown the outer bounds for arbitrary memoryless sources and squared-error distortion measure.

Until now the entire achievable MD rate-distortion region is only known for two-description coding with quadratic Gaussian source[iv] [26] [27]. This region consists of the convex hull of the set of achievable quintuples $(R_{s1}, R_{s2}, D_{s1}, D_{s2}, D_c)$.

[ii] In the MD problem, an inner bound is a set of achievable rate-distortion points for a specific source and fidelity criterion [30].

[iii] In the MD problem, an outer bound is a set of rate-distortion points for a specific source and fidelity criterion. No points outside this bound can be reached [30].

[iv] The quadratic Gaussian case refers to the case of squared-error fidelity criterion and the memoryless Gaussian source [30].

Specifically, the rates satisfy [27] [39]

$$\begin{aligned} R_{s1} &\geq R(D_{s1}) = \tfrac{1}{2}\log_2\left(\tfrac{\sigma^2}{D_{s1}}\right) \\ R_{s2} &\geq R(D_{s2}) = \tfrac{1}{2}\log_2\left(\tfrac{\sigma^2}{D_{s2}}\right) \\ R_{s1}+R_{s2} &\geq R(D_c) + \tfrac{1}{2}\log_2\delta(D_{s1},D_{s2},D_c) \\ R(D_c) &= \tfrac{1}{2}\log_2\left(\tfrac{\sigma^2}{D_c}\right) \end{aligned} \tag{2.3.12}$$

where σ^2 denotes the source variance and $\delta(\cdot) = 1$ is given by [39]

$$\delta(D_{s1},D_{s2},D_c) = \begin{cases} 1, & D_c < D_{s1}+D_{s2}-\sigma^2 \\ \frac{\sigma^2 D_c}{D_{s1}D_{s2}}, & D_c > \left(\frac{1}{D_{s1}}+\frac{1}{D_{s2}}-\frac{1}{\sigma^2}\right)^{-1} \\ \frac{(\sigma^2-D_c)^2}{(\sigma^2-D_c)^2-\left(\sqrt{(\sigma^2-D_{s1})(\sigma^2-D_{s2})}-\sqrt{(D_{s1}-D_c)(D_{s2}-D_c)}\right)^2}, & o.w. \end{cases} \tag{2.3.13}$$

and the distortions satisfy [27]

$$\begin{aligned} D_{s1} &\geq \sigma^2 2^{-2R_{s1}} \\ D_{s2} &\geq \sigma^2 2^{-2R_{s2}} \\ D_c &\geq \tfrac{\sigma^2 2^{-2(R_{s1}+R_{s2})}}{1-(\sqrt{\Pi}-\sqrt{\Delta})^2} \end{aligned} \tag{2.3.14}$$

where $\Pi = (1-\frac{D_{s1}}{\sigma^2})(1-\frac{D_{s2}}{\sigma^2})$ and $\Delta = (\frac{D_{s1}D_{s2}}{\sigma^4}) - 2^{-2(R_{s1}+R_{s2})}$.

Vaishampayan et al. [40] [41] have shown that for the symmetric two-channel MDC under the high-resolution assumption, if the side distortion $D_s (= D_{s1} = D_{s2})$ satisfies

$$D_s = \sigma^2 b 2^{-2R_s(1-a)} \tag{2.3.15}$$

for $b \geq 1$ and $0 < a < 1$, then the central distortion is lower bounded by

$$D_c \geq \frac{\sigma^2}{4b} 2^{-2R_s(1+a)} \tag{2.3.16}$$

which leads to the distortion product $D_c D_s$ being

$$D_c D_s \geq \frac{\sigma^4}{4} 2^{-4R_s} \tag{2.3.17}$$

Equality in (2.3.17) can be achieved when $D_c \ll D_s$. However, for small ratios of D_s/D_c, it is not possible to achieve equality in (2.3.17). Instead, at the high-resolution assumption, the more general but less used distortion product is also achievable [41]

$$D_c D_s = \frac{\sigma^4}{4} \cdot \frac{1}{1 - D_c/D_s} 2^{-4R_s} \tag{2.3.18}$$

2.3.2 M-Description Rate-Distortion Results

Recently, an achievable MD rate-distortion region with a complicated form for more than two descriptions was obtained by Venkataramani, Kramer and Goyal [42] [43] for arbitrary memoryless sources and single-letter fidelity criterions. This region becomes simpler in the quadratic Gaussian case and outer bounds were given in [43]. The region presented in [43] describes an asymmetric MD rate-distortion region and the achievable region meets this outer bound for certain symmetric cases. The construction of this region depends on forming layers of conditional random codebooks. Pradhan, Puri and Ramchandran have shown in [44–48] that the conditional codebooks can be replaced with universal codebooks by exploiting results on distributed source coding. A code constructed in this

way is named as source-channel erasure code (SCEC). Pradhan et al. [47] [48] also showed an achievable rate-distortion region for the M-channel MD problem. It is an inner bound which is conjectured to be tight for the quadratic Gaussian case [48]. The key ideas behind the achievable region obtained by Pradhan et al. are well explained in [47] and [48] and we will here present some of their results. A good summary for MD rate-distortion theory can also be found in Otergaard's thesis [30].

Now we describe the achievable M-description rate-distortion region for the quadratic Gaussian case in [47] and [48]. For the (M, k) source-channel erasure codes, the rate of each description is given by [47]

$$R_s = \frac{1}{2}\log_2\left(\frac{k+\sigma_q^2(1+(k-1)\rho_q)}{\sigma_q^2(1-\rho_q)}\right)^{1/k}\left(\frac{1-\rho_q}{1+(M-1)\rho_q}\right)^{1/M} \tag{2.3.19}$$

The quantization error variance σ_q^2 can be obtained from (2.3.19)

$$\sigma_q^2 = k\left((1-\rho_q)2^{2kR_s}\left(\frac{1+(M-1)\rho_q}{1-\rho_q}\right)^{k/M} - (1+(k-1)\rho_q)\right)^{-1} \tag{2.3.20}$$

The performance of (M, k) source-channel erasure codes can be discussed in three different situations, i.e. according to the amount of correlation ρ_q introduced in the quantization noise. We only show the results in the following. More details can be found in [47].

Independent quantization noise: $\rho_q = 0$

The quantization noise is i.i.d., i.e. $\rho_q = 0$. Assuming that the quantization noise is normalized such that $\sigma_q^2 = k/(2^{2kR_s} - 1)$, then expressions for the

distortion are as follows

$$\begin{cases} D_r = \frac{\sigma_q^2}{\sigma_q^2+r} = \frac{k}{r\cdot 2^{2kR_s-(r-k)}}, & k \le r \le M \\ D_n = 1, & 0 \le n < k \end{cases} \tag{2.3.21}$$

where D_r (or D_n) denotes the distortion when receiving r (or n) out of M descriptions.

Correlated quantization noise: $\rho_q = \rho_q^*$

If the rate-distortion point can be on the rate-distortion function when M description are received (i.e. when $k = M$), then the amount of correlation ρ_q^* needed is given by

$$\rho_q^* = -\frac{2^{2MR_s}-1}{(M-1)2^{2MR_s}+1} \approx -\frac{1}{M-1} \tag{2.3.22}$$

This leads to the following performance

$$D_r = 1 - \frac{r}{M}(1-2^{-2MR_s}), \qquad 0 \le r \le M \tag{2.3.23}$$

and $D^{(M,M)} = 2^{-2MR_s}$.

Correlated quantization noise: $\rho_q^* < \rho_q < 0$

With a varying degree of correlation the performance is given by

$$\begin{cases} D_r = \frac{\sigma_q^2(1+(r-1)\rho_q)}{\sigma_q^2(1+(r-1)\rho_q)+r}, & k \le r \le M \\ D_n = 1, & 0 \le n < k \end{cases} \tag{2.3.24}$$

2.4 MDC Design Methods

The previous section described some known information theoretic bounds. Now we present the practical MD schemes in this section. Existing MD schemes can be divided into four categories: quantization based, subsampling based, transform based and source-channel erasure codes based. For further details on many existing MD techniques we refer the reader to the excellent survey articles by Goyal [49] and Wang [50]. In the following, we review these four categories separately. Later, we will also classify these MDC methods in anther perspective.

2.4.1 Quantization Based MDC

Quantization based schemes include scalar quantization [13, 16, 51–54], trellis coded quantization [41, 55, 56] and vector quantization [14, 15, 39, 57–77]. In this thesis we mainly introduce multiple description scalar quantization (MDSQ) and one of the vector quantization methods—-multiple description lattice vector quantization (MDLVQ).

MDSQ

Vaishampayan [13] developed the MDSQ technique for communication systems that uses diversity to overcome channel impairments. Given a fixed rate, 2-channel MDSQ is comprised of an encoder and three decoders. The encoder α_0 generates a pair of quantization indices $(i_1,\ i_2)$ for each scalar sample x, and three decoders reconstruct from $(i_1,\ i_2)$, i_1 and i_2, respectively. In MDSQ, α_0 is decomposed in two steps: $\alpha_0 = \alpha(Q(x))$, where Q is a regular quantizer and the index assignment function α maps the index generated by Q to a pair of

	1	2	3	4	5	6	7	8
1	1	3						
2	2	4	5					
3		6	7	9				
4			8	10	11			
5				12	13	15		
6					14	16	17	
7						18	19	21
8							20	22

Figure 2.2: An example of MDSQ index assignment matrix.

indices $(i_1,\ i_2)$. α should be invertible so that the central decoder can recover the output of Q. The index assignment matrix can represent α^{-1} visually. Fig. 2.2 shows an example of the index assignment matrix. In this matrix, the cells of the quantizer Q, taken in increasing values of x, are numbered from '1' to '22'. One number corresponds to a pair of indices. For instance, the number '11' in the table corresponds to the index pair (4, 5). The quality of the side distortions is represented by the ranges of values in any row or column.

MDLVQ

MDLVQ is an effective technique to generate two or more representations for a symbol, and is an elegant technique for extending the formalism of MDSQ to vectors. Symmetric two-description LVQ (2DLVQ) was developed in [14] and [15] by Servetto, Vaishampayan, and Sloane (known as the SVS technique) for two balanced (symmetric) channels, whereas asymmetric multiple description lattice vector quantization (AMDLVQ) was developed for possibly unbalanced (asym-

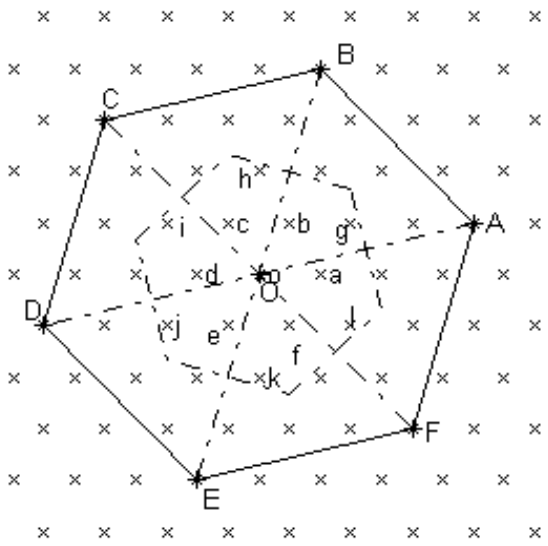


Figure 2.3: An example of 2DLVQ index assignment using A_2 lattice and its sublattice with index number 13. Fine lattice points are marked by x and sublattice points are marked by *. Seven sublattice points used in the index assignment are labeled by 'O' and 'A' to 'F', respectively. Thirteen fine lattice points, which lie in the Voronoi cell of the sublattice point 'O', are labeled by 'o' and 'a' to 'l', respectively.

metric) channels [60]. Later this asymmetric construction provided in [60] was improved in the case of two descriptions and further extended to the case of an arbitrary number of descriptions in [30]. For a given fine lattice Λ[v] and similar sublattice Λ_s[vi], the SVS technique maps each point λ in the fine lattice Λ to a pair of sublattice points $(\lambda_1,\ \lambda_2)$, where the key is to design such a mapping (also known as labeling function or index assignment) α to minimize the side distortion. In Fig. 2.3, an example of the 2DLVQ index assignment design is given based on A_2 lattice and its sublattice with index number 13. The index assignment table (or labeling function table) for Fig. 2.3 is shown in Table 2.1. It has been known

[v] An L-dimensional lattice is a discrete set of equidistantly spaced points in the L-dimensional Euclidean vector space $\Re^L$.

[vi] Sublattice Λ_s is also a lattice, which is a subset of the elements of fine lattice Λ with $\Lambda \subseteq \Lambda$. Furthermore, sublattice Λ_s is similar to Λ if it can be obtained by scaling and rotating (and possibly reflecting) the lattice Λ so that all points of Λ coincide with points of Λ [78].

Table 2.1: 2DLVQ index assignment table for A_2 lattice and its sublattice with index number 13.

lattice point λ	ordered index pair	lattice point λ	ordered index pair
o	(O, O)	–	-
a	(O, A)	d	(O, D)
b	(O, B)	e	(O, E)
c	(O, F)	f	(O, C)
g	(B, F)	j	(E, C)
h	(C, A)	k	(F, D)
i	(D, B)	l	(A, E)

in [15] that under the high-resolution assumption and large lattice vector quantizer dimension, the information theoretic two-description rate-distortion bounds can be achieved by the 2DLVQ. Three or more description LVQ designs are more complex, which were studied in [1, 30, 70, 74, 75, 77, 79].

2.4.2 Subsampling Based MDC

MD subsampling decomposes the original signal into subsets, either in the spatial, temporal, or frequency domain [3, 4, 80–88] where each subset corresponds to a different description. These MD subsampling algorithms are widely used in video coding and take advantage of the fact that spatially or temporally adjacent data samples are correlated. In this way, one description can be estimated from the other. Representative algorithms include spatial pixel interleaving applied either to image samples [3, 4, 87], or motion vectors [84], temporal frame interleaving [2, 19, 20, 74, 81, 89], and transform coefficient interleaving [82, 83, 90]. Optimal partitioning strategies are considered in [88].

2.4.3 Transform Coding Based MDC

Transform based approaches include correlating transforms [91–95], overcomplete expansions and filterbanks [96–100]. In MD transform coding a pairwise correlating transform or generalized transforms are applied on the source symbols and introduce a controlled amount of redundancy directly at the source level. Within each description, coefficients should be uncorrelated for maximum coding efficiency. At the decoder, missing coefficients can be estimated from the received description. Representative algorithms include [101] and [92].

Overcomplete expansions and filterbanks are very similar to a block channel code and attempt to alleviate the 'cliff effect' in channel coding, which will cause a sharp drop-off in reconstruction quality. The basic idea is as follows. Suppose the source produces a vector L-dimensional vector $x \in \Re^L$. x is left-multiplied by a rectangular matrix $F \in \Re^{N \times L}$ $(N > L)$ to produce N transform coefficients. These coefficients are scalar quantized and partitioned into M sets to form M descriptions. Please refer to [49] and [100] for details.

2.4.4 Source-Channel Erasure Codes Based MDC

Instead of designing the source encoder to yield multiple descriptions directly, one can apply unequal forward error correction (FEC) to different parts of a scalable bitstream. This FEC based MD method is pioneered in [102] and [103]. Maximum distance separable (M, k) erasure codes are used to construct multiple independent bitstreams under a joint source-channel coding framework. The information symbols can be recovered if any k description symbols are correctly received. The redundancy is controlled by changing k [104]. Inspired by the

concept of maximum-distance separable (MDS) erasure channel codes, an MD framework based on source-channel erasure codes (SCEC) was presented in [44–48], where random binning ideas [23] are applied to soften the 'cliff effect' caused by FEC.[vii] This SCEC based MD framework subsumes the previous FEC based MD approaches as a special case.

2.4.5 MDC Classification: An Alternative Perspective

To gain a better understanding of these different algorithms, we classify the various MDC methods in the following way. The source may be split into multiple streams (descriptions) in different stages at the encoder side which include pre-processing stage, encoding stage and post-processing stage. Accordingly, all the MDC methods can be classified depending on in which stage the one-to-multiple mapping occurs. In pre-processing based MDC, original source is split into multiple subsources before encoding and then these subsources are encoded separately to generate multiple descriptions. Typical examples are MDC based on subsampling in temporal and spatial domains [2, 19, 20]. For encoding based MDC, the one-to-multiple mapping is performed by dedicated coding techniques such as the above mentioned MDSQ/MDLVQ and MDTC. The post-processing based MDC realizes the one-to-multiple mapping by transforming the encoded bit stream to multiple streams in the compressed domain, which can simplify the MDC application for the media data in the compressed format. The FEC based MDC for embedded compression bitstream [104] is an example. Fig. 2.4 illustrates these three classes of MDC schemes. More sophisticated MDC structures can be achieved with any combination of these three basic types.

[vii] Decoding fails if less than k descriptions are received in FEC, which causes a sharp drop-off in reconstruction quality named as the 'cliff effect'.

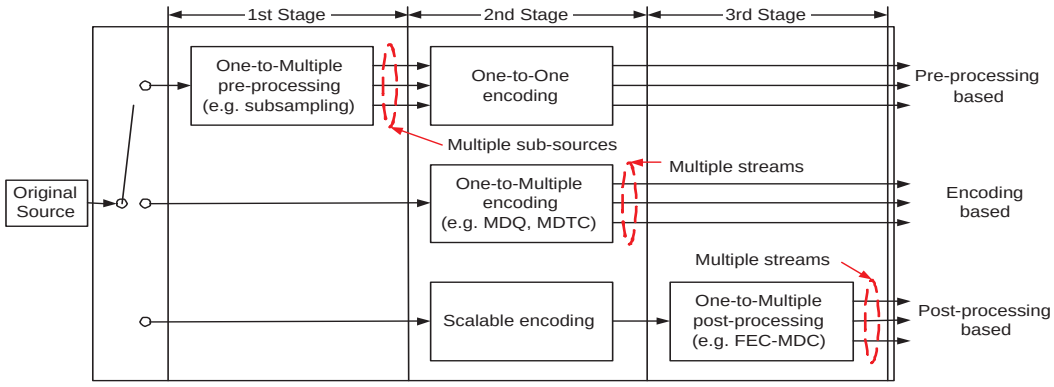


Figure 2.4: Three different stages in encoder to generate multiple coded descriptions.

2.5 Applications

2.5.1 Speech/Audio Coding

The original motivation for MD coding comes from speech coding for the telephone network. The first technique uses odd- and even-numbered samples separation as in [105]. Other techniques use prediction [80], perceptual models [106], and repetition with optimized bit allocations [107]. [108] described a perceptual audio coder with correlating transforms. More recent review on techniques for MD speech coding can be found in [109].

2.5.2 Image Coding

A number of papers have investigated MD image coding. The correlating transform method was introduced in the context of image coding in [95] and [101]. Other transform based techniques are given in [82] and [110]. MD scalar quantizers are introduced in a wavelet image coder in [111]. Two-stage MD quantization based image coder can be found in [17, 18, 112, 113], which features an efficient

and flexible mechanism to control the central-side distortion tradeoff. In addition, in the two-stage structure, well-designed image coders (such as JPEG2000, JPEG) can be utilized with little changes. In some other literatures, unequal error protection is applied to progressively compressed images by using efficient progressive image coders. Examples include [104,114,115] based on FEC. In [116] the application of unequal error protection is combined with additional channel coding for wireless channels. Quantized frames are applied to images in [117] and [96].

2.5.3 Video Coding

To avoid long delays caused by retransmission or long channel codes, MD coding can be used for video transmission in the network. Several MD coding techniques have been proposed for video. In [118], a codec is proposed with the addition of MD protection of the most significant DCT coefficients in the macro blocks. A technique for MD coding of motion vectors is presented in [119]. An MD video coder can also be built on top of the successful block-based motion-compensated prediction framework [81, 86, 119–123]. In designing such an MD video coder, the main problem is how to control the mismatch between the reference frames used at the encoder and decoder when only one description is received at the decoder. One way to avoid such a mismatch is to have two independent prediction loops, each based on the reconstructed frames from one description. For example, in [121] and [123], MD video coders utilize MD transform coding and three separate prediction paths at the encoder to simulate possible scenarios at the decoder. In [124] and [125], the compression technique and transport protocols are joint designed to achieve significant improvements. Meanwhile, many

MD video coding algorithms are subsampling based, which take advantage of the assumption that spatially or temporally adjacent video data samples are highly correlated. Thus, one description can be estimated from the others. A typically simple and efficient pre-processing scheme for two-description video coding is to split the odd and even pictures of a video sequence into two separate streams, each of which can be encoded using a certain coding scheme such as MPEG-1/2/4 or H.261/263/264. Some of such standard-compliant MDVC schemes are developed in [3, 4, 9, 19, 20, 89, 126, 127] based on the similar temporal subsampling. Another kind of standard-compliant MDVC schemes is redundant slice based method in [5] which generates fully standard-compliant descriptions, where the goal is achieved by exploiting the redundant slice coding option available in the H.264 standard. MDVC methods based on spatial subsampling were developed in [2]. There also exist some MDVC algorithms based on frequency domain subsampling. One example known as domain-based MDVC was proposed by Bajic and Woods [88]. Van der Schaar and Turaga developed the multiple description scalable video coding (MDSVC) scheme [128] based on the motion compensated temporal filtering (MCTF) [129], where high frequency frames are grouped into two descriptions and missing frames are estimated using motion vectors in the two descriptions. In [130], after preforming MCTF, 2-D spatial wavelet transform is employed to achieve spatial scalability and embedded bit-plane coding is used to achieve SNR scalability. Different descriptions are produced by coding different code blocks in different rates. In [12], an FEC-based two-dimensional layered multiple description coding is developed.

Since in many cases the descriptions are transmitted through packet based network environments with burst error and/or packet losses, the performance of MDC over different(wireless or wire) channels have been carefully studied in

recent years. The rate-distortion performance of MDC for video coding over wireless networks with multiple path routing is discussed in [7,131,132]. Scenarios of MDC+ARQ/FEC are carefully studied and compared with LC+ARQ/FEC by using two different video codes in multi-path environments at various packet loss probabilities in [8]. A good summary for comparison of MDC and LC can be found in [9]. Recently, layered multiple description coding has been developed in [10–12]. Taking advantages of both layered coding and multiple description coding, layered multiple description coding tries to achieve both robustness to unreliable channels and adaptivity to client bandwidth heterogeneity and network congestion.

2.5.4 Watermarking

As one of the solutions for multimedia security, digital watermarking [133] implementations can provide multimedia security applications with copyright protection, tamper proofing and assessment, broadcast monitoring and fingerprinting, etc.. Many multimedia services are provided in wireless communication networking, thus error resilience of watermarks over wireless channels should be considered. In this regard, multiple description coding is a good choice among error resilient encoding techniques[viii], and may protect the embedded watermark or watermarked signal against different attacks. It is natural to combine these two techniques into one framework–multiple description watermarking (MDWM) framework. There are three different signals before transmission, named host signal, watermark, and watermarked signal, respectively. Depending on what signal

[viii] We have shown in Chapter 1 that multiple description coding is an efficient error-resilient technique. It outperforms other error-resilient coding, such as layered coding, at high error probabilities or in the network with no feedback or very stringent delay constraints.

will use MD coding, various multiple description watermarking frameworks can be classified in the following way: MDC for host signal [134], MDC for watermark [135], and MDC for watermarked signal [136] [137]. In [138], three types of MDWM schemes are studied based on the literatures published in the recent years.

2.6 Conclusion

In this chapter, single and multiple description rate-distortion theory have been briefly described. Some existing MDC design methods have been classified and reviewed. Since numerous MD coders have been proposed for coding multimedia or applied to multimedia security, the literature covering these aspects have also been briefly discussed here. In the following chapters, we will analyze some issues of multiple description coding that involve design methods with performance analysis and applications.

Chapter 3

M-Description Lattice Vector Quantization: Index Assignment and Analysis

3.1 Introduction

As introduced in Chapter 2, multiple description coding can be considered as a source coding scheme to combat transmission errors, especially applicable in non-prioritized lossy networks. In an MDC implementation, an encoder generates M descriptions (or streams) for one source. The streams are transmitted over M separate channels respectively, where each channel may have its own constraints. At the decoder, depending on the number of descriptions received correctly, different reconstruction quality will be obtained. To be more specific, if only k out of a total of M streams are received, the reconstruction quality associated with a so-called k-description side distortion is expected to be acceptable,

and an incremental improvement can be achieved with a reduced side distortion if more streams are received. As all streams are available, the best reconstruction quality is obtained corresponding to the smallest central distortion.

Multiple description lattice vector quantization (MDLVQ) is an effective MDC technique to generate two or more representations for a symbol. Symmetric MDLVQ was introduced in [14] and [15] by Servetto, Vaishampayan and Sloane (known as the SVS technique) for two balanced (symmetric) channels, whereas asymmetric multiple description lattice vector quantization (AMDLVQ) was developed for possibly unbalanced (asymmetric) channels [60]. Later this asymmetric construction provided in [60] was improved for the case of two descriptions and further extended to the case of an arbitrary number of descriptions in [30]. For a given fine lattice Λ and similar sublattice Λ_s with index number[i] N, the SVS technique maps each point in the fine lattice to a pair of sublattice points, where the key is to design such a mapping α (also equivalently known as labeling function or index assignment) that minimizes the 1-description (averaged) side distortion D_1. In M-description LVQ where $M > 2$, each fine lattice point is indexed by M sublattice points which construct an M-tuple, given a fine lattice and a similar sublattice. The goal of this index assignment is to minimize k-description (averaged) side distortion or overall expected distortion. For a finite index number N, minimizing side distortions with different k may lead to different index assignment solutions, while a consistent asymptotical solution can be achieved as $N \to \infty$ [70]. It is noted that the MDLVQ design can also be accomplished without using index assignment [71] [39], which will not be discussed here, and our focus in this chapter is the index assignment for MDLVQ design.

[i]Index number N is defined as the number of elements (points) of Λ in each Voronoi cell of Λ_s.

In [70], the analytical expressions for the central and side distortions were derived for symmetric MDLVQ under a high-resolution assumption[ii], where the index assignment problem was approached as a linear assignment problem. Huang and Wu developed a greedy labeling algorithm for MDLVQ in [79]. Given a fine lattice Λ and a sublattice Λ_s, the main idea is to partition the fine lattice based on its M-fraction lattice $\Lambda_{\frac{s}{M}}$[iii], and then to label the fine lattice points in each Voronoi cell of the M-fraction lattice separately. The advantage of this method is that the M-fraction lattice delicately classifies the fine lattice points by its Voronoi cells and all the fine lattice points in a same Voronoi cell can be treated equally, thus reducing the mapping complexity significantly. The feasibility of using M-fraction lattice partition is due to the S-similarity[iv] of the sublattice Λ_s. This greedy algorithm is shown to be optimal in minimizing side distortion for 2-description LVQ (2DLVQ) using many commonly used lattices and the closed-form analytical solution of the side distortion is presented in [79]. However, the optimality does not hold for more than 2 descriptions. Later the authors further developed a local adjustment algorithm [1] to address the optimality problem in the greedy algorithm, which, as we will discuss in the following section, still cannot accomplish optimality in the general case. In our recent work [74] [75] we developed a 3-description LVQ (3DLVQ) greedy algorithm based on sublattice partition, which mainly deals with the index assignment design for small

[ii] High-resolution quantization theory is built on the assumption that the number of quantization cells is very large [139]. High-resolution quantization theory [140] [141] is concerned with the performance of vector quantizers with a given vector dimension and provides analytical approximations to the performance of globally optimal quantizers. These approximations become tight as the rate increases. High resolution quantization theory has been widely used to identify key characteristics of well-designed codes and to analyze the performance of structured vector quantizers. It is noted that no such analyzes are known at arbitrary (low) rates [141].

[iii] M-fraction lattice $\Lambda_{\frac{s}{M}}$ is defined as $\Lambda_{\frac{s}{M}} \equiv \frac{1}{M}\Lambda_s$ [1].

[iv] A sublattice Λ_s is said to be S-similar to the fine lattice Λ, if Λ_s can be generated by scaling and rotating Λ around any point $G_2 \in \Lambda_{\frac{s}{2}}$. Many commonly used lattices have S-similar sublattices, such as A_2, Z, Z^2, Z^L ($L = 4k$), and Z^L (L odd) [1].

index numbers based on A_2 lattice. In that method, we consider partitioning the sublattices by exploiting the lattice symmetry and then sorting the sublattice points based on their distances to a partitioned region. These ordered sublattice points are used as candidate elements to construct 3-tuples for index assignment of 3DLVQ[v]. Besides the research efforts on the index assignment design and theoretical analysis mentioned above, 2DLVQ has also been applied to image coding [142] which shows better coding performance compared with its scalar counterpart—multiple description scalar quantization (MDSQ) based image coding [111].

In this chapter, we further investigate the optimality and time complexity of index assignment in a general case, for commonly used lattices and higher dimensionalities. We firstly find that the greedy algorithm in [79] and the linear assignment based algorithm in [70] cannot always achieve optimal index assignment for certain finite index numbers. We consider a greedy and a general index assignment scheme for MDLVQ. Borrowing the idea of M-fraction lattice partition in [79], we show that the index assignment problem can be boiled down to a transportation problem in operations research, and then develop two index assignment schemes based on two different transportation models to address the limitation of the algorithms in [70] and [79]. In the proposed schemes, M-tuple construction and assignment are performed orderly for minimizing a certain distortion for clean similar sublattices with finite index number N. As an example, we consider 3DLVQ with applications in Gaussian and real-world signal sources (e.g. speech and image) for experimental validation. The major contributions of this chapter consist of the following aspects. First, we formulate the index assignment problem as transportation problems as common in operations research.

[v]The main work in [74, 75] can be considered as a special case of our latest work presented in this chapter, therefore we do not bring them into the thesis any more.

Second, we propose two algorithms to solve the formulated transportation problems, respectively. Particularly, a greedy algorithm is developed to reduce the complexity of index assignment optimization, which would be otherwise much more computational intensive by the linear assignment method for a higher dimensional lattice. We conjecture that the greedy scheme also produces an optimal index assignment in terms of minimizing a certain side distortion, which is identified in our experimental results, although a rigorous proof is not available at the moment. Third, following our index assignment design, we make an asymptotical analysis under the high-resolution assumption, and obtain a closed-form expression of k-description side distortion, which is a comparable analytical result as those obtained in [79] and [70].

This chapter is organized as follows. In Section 3.2, some preliminaries on lattice are introduced and the existing MDLVQ index assignment schemes are discussed. In Section 3.3, the index assignment problem is formulated as two different transportation models, two proposed index assignment schemes are thereby developed to solve the transportation problems, respectively. We also investigate the asymptotical performance of side distortions based on our scheme. Numerical results are presented in Section 3.4. The chapter is concluded in Section 3.5.

3.2 Review and Discussions

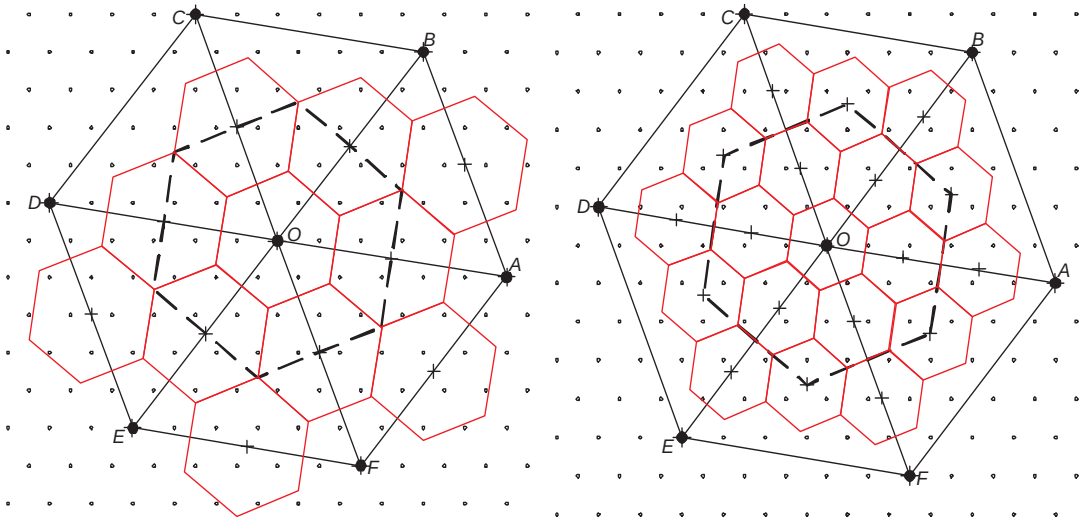
3.2.1 Preliminaries

A real L-dimensional lattice Λ is a discrete set of points in the L-dimensional Euclidean space $\Re^L$. Let Λ and Λ_s be an L-dimensional lattice and a sublattice

respectively, $\Lambda_s \subseteq \Lambda$. Denote G' as the generator matrix for Λ and G'_s for Λ_s. Λ_s is geometrically similar to Λ if $G'_s = cG'U$, where c is a scalar and U is an orthogonal matrix of determinant 1. The index number N of the similar sublattice Λ_s is equal to c^L. Λ_s is a clean geometrically similar sublattice if no lattice point lies on the boundary of the Voronoi cells of Λ_s. In addition, the M-fraction lattice ($M \geq 2$) is defined as $\Lambda_{\frac{s}{M}} \equiv \frac{1}{M}\Lambda_s$ and partitions the space into smaller Voronoi cells than Λ_s does. A sublattice Λ_s is said to be S-similar to Λ, if Λ_s can be generated by scaling and rotating Λ around any point $G_2 \in \Lambda_{\frac{s}{2}}$. Three properties regarding the M-fraction lattice and S-similarity of sublattice are as follows. Suppose the sublattice Λ_s is S-similar, then (i) Λ_s is symmetrical with respect to (w.r.t.) any point in $\Lambda_{\frac{s}{2}}$, (ii) the centroid of any M-tuple with elements being the points of Λ_s should be in $\Lambda_{\frac{s}{M}}$ and $\Lambda_{\frac{s}{M}}$ consists only of these centroids and (iii) if Λ_s is clean similar, $\Lambda_{\frac{s}{M}}$ is also clean. The proofs for the above properties can be found in [1]. In this chapter, only those lattices which have S-similar sublattices are considered. As an example, Fig. 3.1 shows an A_2 lattice and its sublattice with index number 31, where M-fraction lattices $\Lambda_{\frac{s}{2}}$ and $\Lambda_{\frac{s}{3}}$ are also plotted in (a) and (b), respectively.

3.2.2 Index Assignment Design for MDLVQ

In this chapter, we consider the l_2-norm (normalized per dimension) defined by $\|a\|^2 = \langle a, a\rangle$, where the inner product is given by $\langle a, b\rangle \triangleq \frac{1}{L}\sum_{i=0}^{L-1} a_i b_i$. For a given input vector $x \in R^L$, it is firstly mapped to a fine lattice point of Λ denoted as λ, $\lambda = Q(x) = \arg\min_{z\in\Lambda} \|x - z\|$. In order to generate M descriptions for λ, one clean similar sublattice Λ_s with index number N is predefined. An index assignment mapping $\alpha : \Lambda \to \Lambda_s^M$, is designed to minimize a certain (side) distortions, where



(a) Λ, Λ_s and $\Lambda_{\frac{s}{2}}$ (b) Λ, Λ_s and $\Lambda_{\frac{s}{3}}$

Figure 3.1: A_2 lattice Λ, clean similar sublattice Λ_s with index number 31 and M-fraction lattice $\Lambda_{\frac{s}{M}}$. Points of Λ, Λ_s and $\Lambda_{\frac{s}{M}}$ are marked with $\cdot$, $\bullet$ and $+$, respectively. The dashed hexagon is the boundary of Voronoi cell of sublattice point 'O'. The small hexagons represent Voronoi cells of M-fraction lattice points.

α is a bijection. In this way, an ordered M-tuple $\overrightarrow{\lambda_1\lambda_2\cdots\lambda_M}$ with $\lambda_i \in \Lambda_s$ $(1 \leq i \leq M)$, is generated to label a fine lattice point. We need to determine the index assignment (mapping) in a certain region, which is generally a discrete Voronoi cell $V_0(O)$ of a sublattice Λ',[vi] and map each fine lattice point λ of Λ in $V_0(O)$ to M ordered sublattice points of Λ_s. Finally the index assignment in $V_0(O)$ is extended to the entire lattice using shift-invariance property to complete the MDLVQ index assignment design.

3.2.3 Different Kinds of Distortions and Rate

The central distortion, k-description (averaged) side distortions, overall expected distortion and rate in MDLVQ are summarized as follows [15], [70].

[vi] In [70] a product lattice Λ_π is selected as Λ', which has N^2 fine lattice points and N sublattice points in $V_0(O)$. In [79] and [74], Λ' is identical to Λ_s and only N fine lattice points and one sublattice point are in $V_0(O)$.

1) Central Distortion

At the central decoder where all M descriptions (i.e. M-tuple in MDLVQ) are received, the input vector x is reconstructed to λ with the inverse mapping α^{-1} and the central distortion is given by

$$d_c = \sum_{\lambda \in \Lambda} \int_{V(\lambda)} \|x - \lambda\|^2 p_X(x) dx \tag{3.2.1}$$

where $V(\lambda)$ is the Voronoi cell of the fine lattice point λ. Based on the high-resolution assumption for lattice quantizers [15], the central distortion can be obtained as

$$d_c \approx G(\Lambda) v^{2/L} \tag{3.2.2}$$

where $G(\Lambda)$ is the dimensionless normalized second moment of inertia and v is the volume of Voronoi cell of Λ [15].

2) k-description (Averaged) Side Distortion

Denote $SPSD(M, \lambda)$ as the sum of pairwise squared distances (SPSD) between elements of the M-tuple for the fine lattice point λ and G_M as the centroid of this tuple. Given an M-tuple $\overrightarrow{\lambda_1 \lambda_2 \cdots \lambda_M}$ assigned to λ, then

$$SPSD(M, \lambda) = \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \|\lambda_i - \lambda_j\|^2 \tag{3.2.3}$$

Assume that M descriptions are symmetric[vii] or balanced. When only k out of M descriptions are received, the input vector x is reconstructed as the average

[vii] Symmetrical descriptions are achievable through time sharing, although the existence of an exact symmetry is not guaranteed for finite N.

of the k received descriptions. The corresponding k-description (averaged) side distortion D_k is given by [70]

$$\begin{aligned} D_k &\approx d_c + \sum_{\lambda \in V_0(O)} \left(\tfrac{M-k}{M^2 k(M-1)} SPSD(M,\lambda) + \|\lambda - G_M\|^2 \right) P(\lambda) \\ &\approx d_c + \tfrac{1}{N} \sum_{\lambda \in V_0(O)} \left(\tfrac{M-k}{M^2 k(M-1)} SPSD(M,\lambda) + \|\lambda - G_M\|^2 \right) \end{aligned} \tag{3.2.4}$$

where $P(\lambda) = \Pr(Q(x) = \lambda) \approx \frac{1}{N}$ with N fine lattice points in $V_0(O)$ [70].

3) Overall Expected Distortion

Assuming that M descriptions are independent with the same loss probability p_l, the overall expected distortion is given by

$$D_{exp} = \sum_{k=0}^{M} \binom{M}{k} (1-p_l)^k p_l^{M-k} D_k \tag{3.2.5}$$

where D_M is the central distortion d_c and D_0 is the distortion with no description received given by $D_0 \approx E[\|x\|^2]$. Generally, $D_M (= d_c)$ is a constant when Λ is given.

4) Rate

Under the high-resolution assumption, the entropy per dimension for each description R_s is given by [15]

$$R_s \approx h(x) - \frac{1}{L}\log_2(v_s) = h(x) - \frac{1}{L}\log_2(Nv) \tag{3.2.6}$$

where $h(x)$ is the differential (continuous) entropy of the source x, and v_s is the volume of a Voronoi cell of Λ_s. Thus the total entropy R_{total} for M descriptions is $M \cdot R_s$.

3.2.4 Index Assignment with Different Distortions Minimization

1) Index Assignment by Minimizing D_k

Since d_c is a constant for a given Λ, MDLVQ index assignment design for minimizing the k-description (averaged) side distortion D_k in (3.2.4) is to minimize its equivalent

$$\tilde{D}_k = \frac{1}{N} \sum_{\lambda \in V_0(O)} \left(\frac{M-k}{M^2 k(M-1)} SPSD(M, \lambda) + \|\lambda - G_M\|^2 \right) \tag{3.2.7}$$

The same guiding principle for minimizing $\tilde{D}_k$ has been obtained in [70] and [74]. That is, an M-tuple combination of sublattice points is used to represent a fine lattice point based on the principle that the SPSD value of the M-tuple needs to be as small as possible while the centroid of the M-tuple is as close as possible to the fine lattice point. To achieve an optimal index assignment, different strategies have been developed in [70] and [79].

In [70], Ostergaard et al used linear assignment [143] to tackle the index assignment optimization. It is shown in [70] that under the high-resolution assumption, the first term SPSD in (3.2.7) is dominant as $N \to \infty$, which implies that those M-tuples with the smallest SPSD values should be selected in linear assignment. However, the search of tuples with the smallest SPSD values also encounters such a problem that a search region has to be determined which should contain an enough number of sublattice points for M-tuple construction. If the search region $\tilde{V}$ is too large, too many sublattice points are included in $\tilde{V}$. These sublattice points will generate many unnecessary tuples which lead to a higher computational complexity. Ostergaard et al [70] introduced a lower-bound value

for the search region and pointed out that this lower-bound region should be enlarged by a dimensionless expansion factor to guarantee the optimal solution, where the lower-bound value and the expansion factor are discussed under the assumption of $N \to \infty$. However, although the first term is dominant as $N \to \infty$, it is not always true for a finite index number N. A 3DLVQ example is given in Fig. 3.2, based on a Z^2 lattice with a sublattice Λ_s of index number $N = 9$, where $\Lambda_{\frac{s}{3}}$ happens to lap over the fine lattice Λ. In the figure, for instance, $\overrightarrow{OOA}$ is an ordered 3-tuple assigned to the fine lattice point a. It can be shown easily that using $\overrightarrow{OOB}$ instead of $\overrightarrow{OAO}$ to label the fine lattice point b can achieve a smaller distortion D_2, based on (3.2.7). However, according to the tuple selecting rule in [70], $\overrightarrow{OAO}$ with a smaller SPSD value will be used for the labeling of b. Therefore, to achieve optimal index assignment, both terms should be taken into account for a small index number. On the other hand, computational complexity of the linear assignment method used in [70] is very high, especially when a high dimension lattice is used for MDLVQ, which will be discussed in subsection 3.3.1.

Rather than minimizing the first term of (3.2.7) in [70], Huang and Wu [79] developed a greedy algorithm to prioritize minimizing the second term $\|\lambda - G_M\|^2$. Firstly, the fine lattice Λ is partitioned by the M-fraction lattice $\Lambda_{\frac{s}{M}}$. To label the fine lattice points inside the Voronoi cell $V(G_M)$, where $G_M \in \Lambda_{\frac{s}{M}}$, $|\Lambda \cap V(G_M)|$ ordered M-tuples with the same centroid G_M and the smallest SPSD values are selected. This method is a divide-and-conquer approach, which enables each $V(G_M)$ to be labeled separately. In this way, only a small number of fine lattice points are considered at one time, thus simplifying the optimization. Moreover, fine lattice points in each $V(G_M)$ are treated equally in the labeling step, so that the selected tuples can be arbitrarily assigned to the fine lattice points in $V(G_M)$. Consequently, the computational complexity of this index assignment method is

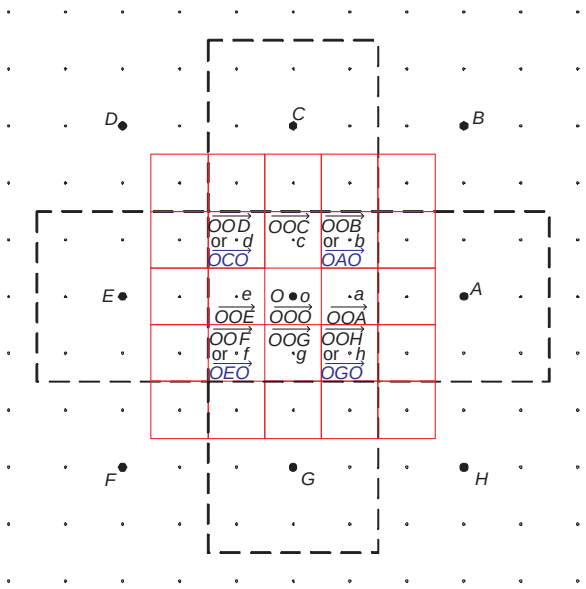


Figure 3.2: 3DLVQ index assignment using Z^2 lattice Λ and sublattice Λ_s with $N = 9$. Points of Λ and Λ_s are marked with $\cdot$ and $\bullet$, respectively. Nine fine lattice points, 'o' and 'a' to 'h' in the Voronoi cell of sublattice 'O' are labeled by ordered 3-tuples.

substantially lessened. This approach is proved to be optimal for 2DLVQ index assignment given some conditions [1], since in 2DLVQ both terms of (3.2.7) can be minimized simultaneously and independently[viii]. However, this simultaneity is not always held in MDLVQ of more than two descriptions. As mentioned in [70], the first term of $\tilde{D}_k$ is dominant, and the M-tuples selected by minimizing the second term are generally not optimal in minimizing the first term, which makes the greedy algorithm not optimal. Later Huang developed a local adjustment algorithm in [1] to try to rectify this problem, which does work for some cases but fails in other cases. A counter-example will be discussed later (shown in Fig. 3.4).

[viii] The first condition is that a sublattice is centric. The sublattice Λ_s is said to be centric, if each sublattice Voronoi cell contains the N nearest fine lattice points to the center of this cell. This condition is not a necessary condition, but it is useful for the proof of optimization. The second condition is that the sublattice Λ_s is S-similar to the fine lattice Λ. With these two conditions, it is interesting to find that the selected N 2-tuples to minimize the second term of (3.2.7) happen to be those N 2-tuples for minimizing the first term. Therefore these N 2-tuples can in turn minimize the overall distortion of (3.2.7). The proof of this optimization problem is complicated which includes 4 properties and 3 lemmas, therefore I will not involve the proof in this thesis. Please refer to Huang's thesis [1] for details.

2) Index Assignment by Minimizing D_{exp}

An index assignment in terms of minimizing the overall expected distortion D_{exp} shown in (3.2.5) is discussed here, given the description loss possibility p_l and the entropy R_s per description. Again, $D_M = d_c$ is a constant for a given Λ due to (4.2.11), and D_0 is independent of the index assignment, both of which can be skipped for the index assignment design. Optimal MDLVQ index assignment w.r.t. the expected distortion boils down to minimizing

$$\tilde{D}_{exp} \equiv \sum_{\lambda \in V_0(O)} \left(\eta_1 SPSD(M, \lambda) + \eta_2 \left\| \lambda - G_M \right\|^2 \right) \frac{1}{N} \tag{3.2.8}$$

where

$$\eta_1 = \sum_{k=1}^{M-1} \begin{pmatrix} M \\ k \end{pmatrix} (1-p_l)^k p_l^{M-k} \frac{M-k}{M^2 k(M-1)} \tag{3.2.9}$$

and

$$\eta_2 = \sum_{k=1}^{M-1} \begin{pmatrix} M \\ k \end{pmatrix} (1-p_l)^k p_l^{M-k} \tag{3.2.10}$$

Comparing (3.2.8) against (3.2.7), we can see that the minimization of $\tilde{D}_{exp}$ is the same as minimizing $\tilde{D}_k$ except that p_l is in the weight coefficients of both terms. Ostergaard et al [70] show that optimal index assignment is independent of the description loss probability p_l. It is true in the asymptotical analysis, since $\left\| \lambda - G_M \right\|^2$ can be skipped as $N \to \infty$. Nevertheless, for a finite and small index number, both terms of $\tilde{D}_{exp}$ need to be considered in the minimization and the optimal index assignment for minimizing $\tilde{D}_{exp}$ is thereby related to p_l.

3.3 Proposed MDLVQ Index Assignment

From the above discussions we can see that it is more advisable to study the index assignment design for a finite and infinite index number N separately. To pursue an optimal index assignment w.r.t. $\tilde{D}_k$ or $\tilde{D}_{exp}$ for a finite index number, we consider a design of the index assignment in the following way.

A discrete Voronoi cell $V_0(O)$ of sublattice Λ_s is selected and N fine lattice points in the cell will be assigned by N M-tuples. A sufficient number of candidate M-tuples with the smallest SPSD values needs to be identified and the number of these tuples should be more than N considering that the first N M-tuples with the smallest SPSD values may not always be the optimal choice as discussed above. The final selection of N M-tuples will be made by taking into account the second term of $\|\lambda - G_M\|^2$ in $\tilde{D}_k$ or $\tilde{D}_{exp}$.

In this section we analyze the index assignment design by minimizing $\tilde{D}_k$ as an example for a finite index number, and an analysis by minimizing $\tilde{D}_{exp}$ can be made similarly. In the case of infinite index number (i.e. $N \to \infty$), we will show that with the proposed scheme, the k-description (averaged) side distortion can be expressed by the normalized second moment of a sphere.

We start our design from the representation of the SPSD term in $\tilde{D}_k$. Considering an M-tuple $\overrightarrow{\lambda_1\lambda_2\cdots\lambda_M}$ with the centroid G_M, which is assigned to the

fine lattice point λ, the SPSD value of the M-tuple is

$$\begin{aligned}
&SPSD(M,\lambda) = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\|\lambda_i-\lambda_j\|^2 \\
&= \sum_{j=2}^{M}\|\lambda_1-\lambda_j\|^2 + \sum_{i=2}^{M-1}\sum_{j=i+1}^{M}\|\lambda_i-\lambda_j\|^2 \\
&= \begin{pmatrix} M-1 \\ 1 \end{pmatrix} \tfrac{1}{(M-1)^2}\sum_{i=2}^{M-1}\sum_{j=i+1}^{M}\|\lambda_i-\lambda_j\|^2 \\
&+ \begin{pmatrix} M-1 \\ 1 \end{pmatrix}\|\lambda_1-G_{M-1}\|^2 + \sum_{i=2}^{M-1}\sum_{j=i+1}^{M}\|\lambda_i-\lambda_j\|^2 \\
&= (M-1)\|\lambda_1-G_{M-1}\|^2 + \tfrac{M}{M-1}\sum_{i=2}^{M-1}\sum_{j=i+1}^{M}\|\lambda_i-\lambda_j\|^2 \\
&= (M-1)R_M^2 + \tfrac{M}{M-1}r_M^2
\end{aligned} \tag{3.3.11}$$

where G_{M-1} is the centroid of the $(M-1)$-tuple $\overrightarrow{\lambda_2\lambda_3\cdots\lambda_M}$, and G_{M-1} always belongs to $\Lambda_{\frac{s}{M-1}}$ since Λ_s is S-similar. R_M is the distance from λ_1 to G_{M-1}, and r_M^2 is the SPSD value of $\overrightarrow{\lambda_2\lambda_3\cdots\lambda_M}$.

Without loss of generality, we first consider those M-tuples with their first elements being the sublattice point λ_1. We denote the fine lattice points labeled by these M-tuples as the set T_{λ_1}. Supposing $SPSD_M^{\max} = \max\{SPSD(M,\lambda), \lambda \in T_{\lambda_1}\}$, then R_M^2 may take value from $[0,(R_M^{\max})^2]$, where

$$(R_M^{\max})^2 = \frac{1}{M-1}SPSD_M^{\max} \tag{3.3.12}$$

Similarly, r_M^2 may lie in the range of $[0,(r_M^{\max})^2]$ with

$$(r_M^{\max})^2 = \frac{M-1}{M}SPSD_M^{\max} \tag{3.3.13}$$

Since r_M and R_M are both nonnegative, $R_M^{\max}$ can be achieved when $r_M = 0$, and

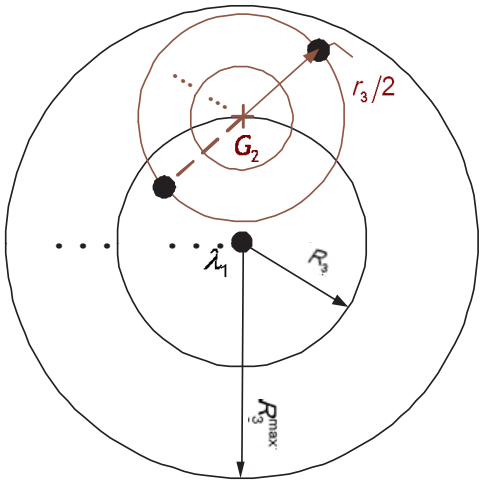


Figure 3.3: Relationship between R_M and r_M in (3.3.11) for a given $SPSD_3^{\max}$ as $M = 3$. Lattice points of Λ_s and $\Lambda_{\frac{s}{2}}$ are marked with $\bullet$ and $+$, respectively.

vice versa. When $SPSD(M, \lambda)$ in (3.3.11) reaches $SPSD_M^{\max}$, we have

$$SPSD_M^{\max} = (M-1)R_M^2 + \frac{M}{M-1}r_M^2 \tag{3.3.14}$$

We can further express r_M^2 as a function of R_M^2, which is given by

$$r_M^2 = \frac{(M-1)^2}{M}\left(\frac{SPSD_M^{\max}}{M-1} - R_M^2\right) = \frac{(M-1)^2}{M}((R_M^{\max})^2 - R_M^2) \tag{3.3.15}$$

It can be seen that a larger R_3 leads to a smaller r_3, and vice versa. An example is shown in Fig. 3.3 to illustrate the relationship between R_3 and r_3 in 3DLVQ, where R_3 is the distance between λ_1 and G_2, and r_3 is the length of an edge with two endpoints being sublattice points λ_2 and λ_3. $SPSD_3^{\max}$ (or $R_3^{\max}$) value is assumed to be fixed in the figure. Note that in 2DLVQ ($M = 2$), r_2^2 becomes zero regardless of R_2.

3.3.1 Index Assignment for a Finite Index Number

1) Constructing and Sorting of M-tuples

The first task is to construct the tuples of the set T_{λ_1}. Let $S(\lambda_1, R_M)$ be the sphere of radius R_M centered at λ_1 based on l_2-norm. It is known [78] that the number of $(M-1)$-fraction lattice points of $\Lambda_{\frac{s}{M-1}}$ lying on the surface of $S(\lambda_1, R_M)$ is given by the coefficients of the Theta series. We now present an approach to construct the M-tuples with the SPSD value no more than $SPSD_M^{\max}$.

Consider a given G_{M-1} on the surface of $S(\lambda_1, R_M)$. All the $(M-1)$-tuples centered at G_{M-1} with their SPSD values no more than r_M^2 are selected[ix]. By adding λ_1 in the front of these selected $(M-1)$-tuples, we can construct all the M-tuples with the SPSD values no more than $SPSD_M^{\max}$ for the given G_{M-1}. In the following we consider an example of 3DLVQ shown in Fig. 3.3 for the 3-tuple construction. First we construct $S(G_2, r_3/2)$, i.e. a sphere centered at G_2 with radius of $r_3/2$. If λ_2 and λ_3 are exactly the endpoints of a diameter line in this sphere, then the SPSD value of $\overrightarrow{\lambda_1\lambda_2\lambda_3}$ and $\overrightarrow{\lambda_1\lambda_3\lambda_2}$ can achieve $SPSD_3^{\max}$. In this way, we construct a pair of 3-tuples with the SPSD value being $SPSD_3^{\max}$, given the first element λ_1 and G_2. Depending on the number of pairs of sublattice points lying on the surface of $S(G_2, r_3/2)$, we can find the same number of 3-tuple pairs with the same SPSD value of $SPSD_3^{\max}$. Similarly, we can also obtain many concentric spheres centered at G_2 with radius less than $r_3/2$. Then sublattice point pairs on the surfaces of these smaller spheres together with λ_1 can construct 3-tuples with their SPSD values smaller than $SPSD_3^{\max}$.

We would like to highlight that the M-tuple construction mentioned above

[ix]In this sense, r_M^2 is the maximum SPSD value of the selected $(M-1)$-tuples, given $SPSD_M^{\max}$ and G_{M-1}. We can let $SPSD_{M-1}^{\max} = r_M^2$, where r_M^2 is determined by $SPSD_M^{\max}$ and G_{M-1} (or R_M). This is used in the derivation of (A.0.9) in Appendix A.

is based on $SPSD_M^{\max}$, λ_1 and G_{M-1}. We have constructed all the M-tuples corresponding to a particular λ_1 and G_{M-1}, with their SPSD values no more than $SPSD_M^{\max}$. Furthermore, if the location of G_{M-1} changes[x], i.e., R_M and r_M change accordingly, it is possible for us to construct all the M-tuples with the first element being λ_1 and SPSD values no more than $SPSD_M^{\max}$. Then we can use the shift-invariance property of the lattice to find equivalent M-tuples with their centroids lying on/in $V_0(O)$[xi]. If we further allow $SPSD_M^{\max}$ to increase one by one, more new M-tuples are generated. Note that in the above construction process, the tuples are generated in order of SPSD due to the gradual increment of $SPSD_M^{\max}$ in the process.

2) Selecting and Mapping of M-tuples for Minimizing $\tilde{D}_k$

In this subsection, we develop two different algorithms to select N M-tuples from those obtained in the above step and then map them to fine lattice points for minimizing $\tilde{D}_k$.

Suppose there are m M-fraction lattice points on/in $V_0(O)$ in Λ_s, denoted as τ_j[xii] with $j = 1$ to m. We use N_j to represent the number of fine lattice points in the Voronoi cell $V(\tau_j)$ in $\Lambda_{\frac{s}{M}}$, and define a state s associated with a set of N M-tuples, T_j is the number of M-tuples centered at τ_j in the state s, and $N = \sum_{j=1}^{m} T_j = \sum_{j=1}^{m} N_j$.

There are two useful extreme states. The first extreme state $s^{(1)}$ corresponds

[x] G_{M-1} is always inside the sphere $S(\lambda_1, R_M^{\max})$.

[xi] As mentioned earlier, the M-fraction lattice points are the centroids of M-tuples. For example, in Fig. 3.1(a) there are six 2-fraction lattice points (marked with +) on the boundary of $V_0(O)$ and one in it, i.e. totally seven 2-fraction lattice points on/in $V_0(O)$ for $M = 2$. Similarly, there are thirteen 3-fraction lattice points on/in $V_0(O)$ in Fig. 3.1(b).

[xii] Note that in the previous discussion for constructing M-tuples, we use G_M ($G_M \in \Lambda_{\frac{s}{M}}$) to represent the centroid of an M-tuple which labels a fine lattice point in T_{λ_1}, i.e. a tuple with the first element being λ_1. In this subsection we use τ_j ($\tau_j \in \Lambda_{\frac{s}{M}}$) to represent the j-th M-fraction lattice point on/in $V_0(O)$.

to the N M-tuples with the smallest SPSD values. Therefore $s^{(1)}$ will minimize the first term $\sum_{\lambda \in V_0(O)} SPSD(M, \lambda)$ in $\tilde{D}_k$. Note that $s^{(1)}$ may be not unique, and the state minimizing $\sum_{j=1}^{m} |T_j - N_j|$ is selected. In [70], such a selected $s^{(1)}$ is argued to be the final selection to achieve the optimal index assignment IA^* corresponding to minimum distortion $\tilde{D}_k^*$. If $N_j = T_j$ holds for any j, then the selected $s^{(1)}$ can also minimize the second term $\sum_{\lambda \in V_0(O)} \|\lambda - G_M\|^2$ in $\tilde{D}_k$. Under this condition, $s^{(1)}$ can therefore be confirmed as an optimal selection for index assignment. However, for many finite index numbers, the condition of $N_j = T_j$ for any j may not hold, which may make the optimality invalid. In the second extreme state $s^{(2)}$, the selection of N M-tuples is made based on each individual τ_j. For each τ_j, the first N_j M-tuples centered at τ_j with the smallest SPSD values are selected, thus $T_j = N_j$ holds for any j. Such a $s^{(2)}$ can guarantee minimizing the second term of $\tilde{D}_k$. In fact, $s^{(2)}$ is the solution obtained in [79]. However, the first term of $\tilde{D}_k$ may not be minimized in the state. In the case that the (N_1+1)-th ordered M-tuple centered at τ_1 (for notational simplicity we use 'the tuple N_1+1' in the following) has a smaller SPSD value than the N_2-th ordered M-tuple centered at τ_2 (i.e. 'the tuple N_2'), the tuple N_2 will be selected instead of the tuple N_1+1 for the $s^{(2)}$ to make $T_j = N_j$ hold for $j = 1,\ 2$. In view of the problems for the two states, which either minimize the first term or the second term of $\tilde{D}_k$, we wonder if there exist any other states between $s^{(1)}$ and $s^{(2)}$ to achieve a smaller $\tilde{D}_k$. To address this question, we consider generating more M-tuples, where a total of T'_j ($T'_j \geq N_j$ for each j) tuples centered at τ_j are constructed. In this way, the optimal solution can be obtained by selecting N out of T' ($= \sum_{j=1}^{m} T'_j$) tuples, i.e., totally $C_{T'}^{N}$ possible states. A large $C_{T'}^{N}$ will lead to time consuming exhaustive search. Now we develop two algorithms to simplify the search problem. The first algorithm is a greedy algorithm, and the second is

a general algorithm.

a) Greedy Algorithm

We set $s^{(1)}$ as the initial state, i.e. $s_0 = s^{(1)}$. Note that if the tuple T_j (centered at τ_j) is selected by one state, then the first $T_j - 1$ tuples with smaller SPSD values (centered at τ_j) are necessarily included in the state. For a new state, we can simply replace one tuple in s_0 by another one beyond s_0, and need to determine which tuple replaced for searching a smaller distortion.

First we consider that the tuple T_1 of s_0 is replaced for searching candidate selected from the tuple $T_2 + 1$, $T_3 + 1, \cdots$, and $T_m + 1$ one by one. Then $m - 1$ new states will be generated. Similarly we replace each tuple T_j ($j = 2$ to m) of s_0 in turn, and $m - 1$ new states are obtained for each j. In this way, a total of $m \cdot (m-1)$ new states are produced. The state with the lowest $\tilde{D}_k$ value is denoted as s_1 with the subscript '1' representing the 1st iteration. It can be seen that s_1 is obtained with a new tuple replacement from s_0 by performing $m \cdot (m-1)$ trials in the iteration. If s_1 results in a smaller $\tilde{D}_k$ than s_0, we continue the second iteration based on the new state s_1 with the same processing as in the first iteration. The iteration continues until the newly generated state s_l cannot produce smaller $\tilde{D}_k$ compared with its previous state s_{l-1}. It can be seen that the search of optimal state is done based on decreasing $\tilde{D}_k$ along its negative gradient in each iteration, leading to a greedy algorithm formulated in Table 3.1. As can be seen from the table, we need to obtain the minimum $\tilde{D}_k(ts_{ij})$ value for a given trial state ts_{ij}, which specifically involves the optimal mapping between N fine lattice points and N selected tuples. In the following we will show that this distortion minimization problem defined as Dist-Min procedure in Table 3.1, can be reduced to a Hitchcock transportation problem [144].

Table 3.1: Proposed Greedy Algorithm

$s_0 = s^{(1)}$;
$[IA(s_0), \tilde{D}_k(s_0)]$ = Dist-Min (input=s_0);
$s^* = s_0$; $IA^* = IA(s_0)$; $\tilde{D}_k^* = \tilde{D}_k(s_0)$; $l = 0$;
DO {
  Calculate each T_j based on s_l (=s^*);
  FOR $j = 1 : m$
    FOR $i = 1 : m$ and $i \neq j$
      Replace the tuple T_j of s_l by the tuple $T_i + 1$ to generate a new trial state ts_{ij};
      Update each T_j value based on the centroid distribution of tuples in ts_{ij};
      $[IA(ts_{ij}), \tilde{D}_k(ts_{ij})]$ = Dist-Min (input=ts_{ij});
    END
  END
  Let s_l be the state which achieves the smallest $\tilde{D}_k$ among all the ts_{ij} trial states;
  IF $\tilde{D}_k(s_l) \leq \tilde{D}_k(s_{l-1})$
    $s^* = s_l$; $IA^* = IA(s_l)$; $\tilde{D}_k^* = \tilde{D}_k(s_l)$; $l = l + 1$;
  ELSE
    Stop and exit;
  END IF } WHILE ($l < N$)

For a given trial state ts_{ij}, the first term of $\tilde{D}_k$ shown in (3.2.7) is fixed. Therefore we only consider the second term $\sum_{\lambda \in V_0(O)} \|\lambda - G_M\|^2$ for minimizing $\tilde{D}_k$. Minimizing the second term can be translated into a typical balanced transportation problem with N sources and m destinations, which is shown to be solvable in polynomial time [145]. Specifically, this transportation problem can be summarized as follows.

We consider N fine lattice points as N sources and each source can provide only one product. We also take m M-fraction lattice points τ_j as m destinations, each with the capacity of receiving T_j products. T_j ($j = 1$ to m) is determined by the input state ts_{ij}. The squared distance from the i-th ($1 \leq i \leq N$) fine lattice point to τ_j ($1 \leq j \leq m$) is considered to be the transport cost from source i to destination j. The optimization of the transportation problem is to find the minimum total cost of transporting all the N products from N sources to m

Table 3.2: Dist-min Procedure

STEP1: Input a state s containing N M-tuples.
STEP2: Determine T_j ($j = 1$ to m) for s.
STEP3: Solve (3.3.16) using T_j obtained above.
STEP4: The solution is translated back to index assignment result $IA(s)$.
STEP5: Calculate $\tilde{D}_k(s)$ by using $IA(s)$. Output $IA(s)$ and $\tilde{D}_k(s)$.

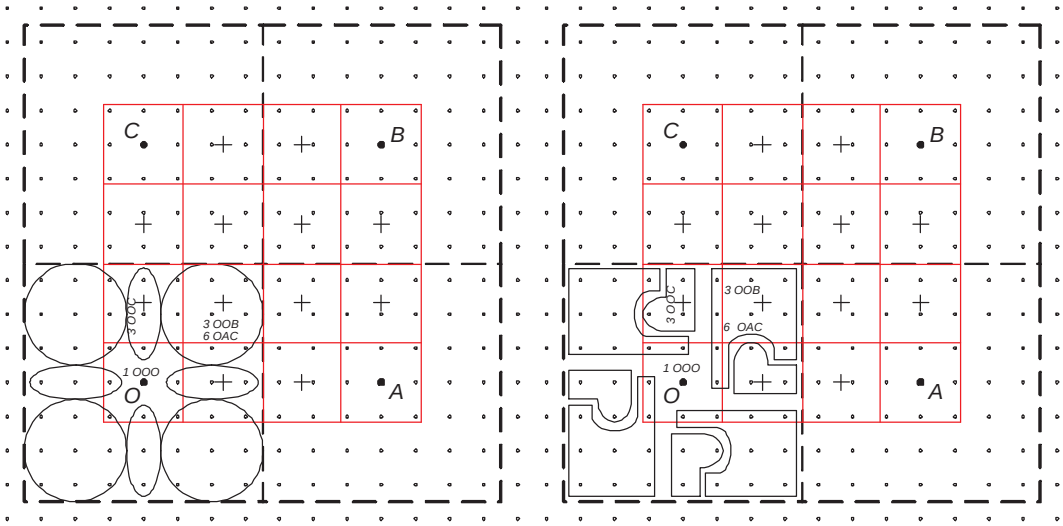
destinations with capacity constraint for each destination, which is equivalent to minimizing the second term of (3.2.7). Therefore, we have

$$\begin{aligned} &\text{Minimize} \quad \sum_{i=1}^{N}\sum_{j=1}^{m} c_{ij}x_{ij}, \quad \text{subject to} \\ &\sum_{i=1}^{N} x_{ij} = T_j, \quad j = 1, 2\cdot\cdot, m; \\ &\sum_{j=1}^{m} x_{ij} = 1, \quad i = 1, 2\cdot\cdot, N; \\ &x_{ij} = 0 \text{ or } 1 \end{aligned} \tag{3.3.16}$$

where c_{ij} is the squared distance between the i-th fine lattice point and τ_j, i.e., $\|\lambda - G_M\|^2$. A procedure called Dist-Min for solving the minimization problem is presented in Table 3.2. Note that in STEP3 of the table, the solution to (3.3.16) will specify the T_j sources for each destination j ($j = 1, \cdots, m$). Then, these T_j fine lattice points (i.e. T_j sources in STEP3) are arbitrarily mapped by T_j M-tuples centered at τ_j to obtain $IA(s)$ in STEP4.

Remark 1: If we initialize the first input s_0 to be $s^{(2)}$, the proposed greedy algorithm is a generalized scheme of Huang's local adjustment algorithm [1], and the method of [1] can be considered as a special case of the proposed greedy algorithm. According to [1], the initial state is $s^{(2)}$ and adjustment is only made between two spatially adjacent (or neighboring) Voronoi cells of the M-fraction lattice $\Lambda_{\frac{s}{M}}$. Suppose that adjustment is made between two neighbor Voronoi cells $V(\tau_i)$ and $V(\tau_j)$, and the tuple $N_i + 1$ (centered at τ_i) will replace the tuple N_j

(centered at τ_j) if the first has a smaller SPSD value than the second, based on the adjustment rule. This replacement is simple but it is not clear whether the loser tuple (the tuple N_j) should be discarded or reserved for further checking. Another problem for Huang's local adjustment algorithm [1] is the priority of replacement when more cells of $\Lambda_{\frac{s}{M}}$ need to make adjustment with the same cell. Our Dist-Min procedure solves the transportation problem considering all the Voronoi cells for adjustment without the neighboring constraint in Huang's local adjustment algorithm [1]. In this sense, our algorithm is a global adjustment algorithm. To show a comparison of our algorithm against the local adjustment algorithm, we consider a 3-description index assignment based on Z^2 lattice and its sublattice with index number of 49. Fig. 3.4 illustrates the index assignments by the two algorithms. From the figure we can see that although all the 3-tuples used for labeling are the same (i.e. using the same state) for both algorithms, the final assignments are very different, where the same groups of 3-tuples are assigned to different fine lattice points inside different enclosed areas. For example, in Fig. 3.4(b), ordered tuples consisting of O, O, C ($\overrightarrow{OOC}$, $\overrightarrow{OCO}$ and $\overrightarrow{COO}$) are used to label three fine lattice points in the same Voronoi cell (a small square region in the figure) of 3-fraction lattice according to Huang's local adjustment rule, while one of them is labeled to a fine lattice point outside the cell in our method shown in Fig. 3.4(a). Since in both schemes the same 49 3-tuples are used for index assignment, the contributions of the first term in D_1 are the same for both schemes. The difference between two 1-description side distortions of two schemes comes from the second term. As discussed in previous greedy algorithm, the minimization of the second term can be formulated to a balanced transportation model. We can evaluate the distortion performance from a view of geometry in the figure. It would be the best choice if each destination τ_j (i.e. 3-fraction lattice



(a) Proposed greedy algorithm (b) Local adjustment algorithm

Figure 3.4: Comparison of 3-description index assignments obtained by the proposed greedy algorithm and Huang's local adjustment algorithm [1]. Z^2 lattice with index number $N = 49$ is applied and the index assignments are designed in terms of minimizing 1-description side distortion D_1. Points of Λ, Λ_s and $\Lambda_{\frac{s}{3}}$ are marked with $\cdot$, $\bullet$ and $+$, respectively. The small squares indicate the Voronoi cell of 3-fraction lattice. The label '6 OAC' in the figure refers to six ordered 3-tuples consisting of sublattice points O, A and C, and the same for others like '3 OOB', '3 OOA', '3 OOC' and '1 OOO'. These tuples are used to label the fine lattice points enclosed in a circle or an oval in (a) based on the proposed index assignment scheme, while they are for fine lattice points inside other shapes of enclosed curves shown in (b) according to Huang's local adjustment method.

point '+' in the figure) can obtain T_j products from its nearest T_j sources (i.e. fine lattice points '$\cdot$'). It can be shown geometrically and mathematically that the index assignment shown in (a) by the proposed greedy algorithm can achieve a smaller 1-description (averaged) side distortion D_1, compared with that shown in (b) by Huang's method [1].

b) General Algorithm

As discussed above, we have constructed T' candidate tuples for labeling N fine lattice points in the proposed greedy algorithm, and known that the optimal solution is achievable if we traverse all the possible $C_{T'}^{N}$ states formed by

these T' candidate tuples. This can be translated to an unbalanced transportation problem with T' sources and N destinations. Specifically, the unbalanced transportation problem can be modeled as follows.

There are T' sources and each of them can provide only one product, while there are N destinations and each demands only one product, $T' > N$ (more supplying than demanding). If the i-th ($1 \leq i \leq T'$) tuple is assigned to the j-th ($1 \leq j \leq N$) fine lattice point, then the obtained k-description side distortion for this tuple can be considered as the transport cost from source i to destination j. The optimization of the transportation problem is to find the minimum total cost of transporting the N-selected products from T' sources to N destinations with capacity constraint for each destination. We add a virtual destination which needs $(T' - N)$ units, so that the transportation problem becomes a balanced one shown in the following formula.

$$\begin{aligned}
&\text{Minimize} \sum_{i=1}^{T'} \sum_{j=1}^{N+1} d_{ij} x_{ij}, \quad \text{subject to} \\
&\sum_{\mathrm{j}=1}^{N+1} x_{ij} = 1, \ i = 1, 2 \cdot \cdot, T'; \\
&\sum_{i=1}^{T'} x_{ij} = 1, \ j = 1, 2 \cdot \cdot, N; \\
&\sum_{i=1}^{T'} x_{ij} = T' - N, \ j = N+1; \quad x_{ij} = 0 \text{ or } 1; \\
&d_{ij} = 0, \ j = N+1;
\end{aligned} \tag{3.3.17}$$

where d_{ij} is the k-description side distortion with the i-th tuple labeling the j-th fine lattice point. All the values of d_{ij} can be calculated offline. The solution to (3.3.17) can produce an optimal index assignment result.

Remark 2: To minimize $\tilde{D}_{exp}$ for a given finite index number N, the optimal index assignment will generally change with the description loss possibility p_l. It is

interesting to find that there still exist some index numbers for which the index assignment independent of p_l, such as $N = 13$ or 73 of A_2 lattice in 3DLVQ, because the two terms of $\tilde{D}_k$ can be minimized simultaneously with the same index assignment for $k < 3$ [70] [79].

3) Time Complexity Comparison

In [70], a linear assignment problem is considered to assign a set of M-tuples to the N^2 fine lattice points for the index assignment. To solve a N-point linear assignment problem, the well-known Hungarian method [146] can be used, which has complexity of cubic order $O(N^3)$. Therefore the complexity of the N^2-point assignment problem in $V_0(O)$ is on the order of $O(N^6)$. If the method in [147] is used, the complexity can be reduced to $O(N^5)$. Note that if we select $V_0(O)$ to be the Voronoi cell of sublattice Λ_s, instead of the product sublattice Λ_π, only N fine lattice points need to be labeled. Then the complexities based on the Hungarian method and the method in [147] are reduced to $O(N^3)$ and $O(N^{2.5})$, respectively. The greedy index assignment algorithm in [79] further reduces the complexity greatly to a linear time complexity of $O(N)$.

In our greedy algorithm, the STEP3 in the Dist-Min procedure shown in Table 3.2 solves the transportation problem with N sources and m destinations in $O(mN^2 \log N + m^2 N \log^2 N)$ [145]. [148] further simplifies the complexity to $O(m^2 N \log^2 N)$, and m here is a small number which can be estimated by M^L. Therefore the Dist-Min procedure can be finally solved in $O(N \log^2 N)$ for a limited lattice dimensionality L. The time complexity for each iteration in the proposed greedy algorithm will be $O(m^4 N \log^2 N)$, in view that at most $m(m-1)$ states are tested in one iteration. In the worst case that $N - 1$ iterations are needed, the overall time complexity for the proposed greedy algorithm is of the

order of $O(N^2 \log^2 N)$ for a small m. For the proposed general algorithm, more states are tested and its complexity is higher than the proposed greedy algorithm. Since $T' \leq mN$, the worst running time for the general algorithm is thereby of order $O\left(m^2N^3 \log(mN) + mN^3 \log^2(mN)\right)$, according to the results in [145].

Remark 3: Note that the time complexities for the schemes in [70] and [79] are presented assuming the awareness of the input state (e.g., N M-tuples). Therefore it is fair to compare the schemes in [70] and [79] with the proposed Dist-Min procedure, which also assumes an input state is known. From the above analysis, we know that the Dist-Min procedure achieves a lower time complexity $O(N \log^2 N)$ than the linear assignment algorithm [70] that has the complexity of $O(N^{2.5})$. The greedy algorithm in [79] has the lowest complexity with $O(N)$. As our proposed scheme searches much more different states for optimality, the overall complexity for our greedy scheme will reach $O(N^2 \log^2 N)$ in the worst case (totally $m(m-1)N$ states are searched), which is still lower than the linear assignment complexity $O(N^{2.5})$ for a sufficient large N. We conjecture that the proposed greedy scheme produces an optimal index assignment, which is supported by our experimental results, although no rigorous proof can be provided at the moment. We can see that the proposed greedy algorithm can find an optimal index assignment with a moderate complexity, while the proposed general algorithm achieves an optimal index assignment with higher complexity. The linear assignment based method in [70] and the greedy algorithm in [79] can take the description loss probability into account without increasing the computational complexity, since the choice of state for index assignment is independent of the description loss probability. The input state is $s^{(1)}$ in [70] and $s^{(2)}$ in [79]. However for some finite index numbers, these two schemes cannot achieve the optimal solution in terms of minimizing k-description (averaged) side distortion or over-

all expected distortion, which has been shown in our earlier discussions and will be further seen from the 3DLVQ coding results on the Gaussian source in the experimental section.

3.3.2 Asymptotical Performance

Now we make asymptotical analysis of the k-description (averaged) side distortion as well as the overall expected distortion D_{exp} for the optimal index assignment. Based on a closed-form expression of D_{exp}, optimal volume v_{opt} of the fine lattice as well as optimal index number N_{opt} can be determined by minimizing the D_{exp}.

1) Closed-form Expression of D_k

For a finite index number N, we can increase $SPSD_M^{\max}$ in step to find appropriate candidates of M-tuples with the help of Theta series [78]. However, for infinite N we have to estimate the number of M-tuples with their SPSD values no larger than $SPSD_M^{\max}$, where three reasonable assumptions used in [70] and [79] are also employed. The first assumption is that as $N \to \infty$ and $v_s \to 0$, each Voronoi cell of $\Lambda_{\frac{s}{M}}$ contains approximately N/M^L fine lattice points [79]. In the second assumption, the number of lattice points of $\Lambda_{\frac{s}{i}}$ $(1 \leq i \leq M)$ in a given sphere S approximately equals to $\mathrm{Vol}(S)/v_{\frac{s}{i}}$ with $v_{\frac{s}{i}}$ being the volume of Voronoi cell of $\Lambda_{\frac{s}{i}}$ [70]. In the third assumption, in view that the SPSD term in $\tilde{D}_k$ is dominant when $N \to \infty$ and $M < \infty$ [70], we only consider the SPSD term in $\tilde{D}_k$, where the N M-tuples with the smallest SPSD values are selected. We show in Appendix A that given the number of descriptions M $(M \geq 2)$, the index

number N can be expressed as a function of $R_M^{\max}$, that is

$$N \approx \omega_L^{M-1} \cdot \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \prod_{i=2}^{M} \beta_i \cdot (R_M^{\max})^{(M-1)L} \tag{3.3.18}$$

where ω_L is the volume of the L-dimensional unit sphere [15], and β_M is given in (3.3.19). N can be further expressed as a function of $SPSD_M^{\max}$ by substituting (3.3.12) into (3.3.18).

$$\beta_M = \begin{cases} 1, & M = 2 \\ \frac{(M-1)^{(M-3)L}}{(M^2-2M)^{(M-2)L/2}} \cdot \sum_{n=0}^{\frac{(M-2)L}{2}} \binom{\frac{(M-2)L}{2}}{n} (-1)^n \frac{L}{L+2n}, & M \geq 3 \end{cases} \tag{3.3.19}$$

With the help of (3.3.18) we obtain the sum of SPSD values for the N fine lattice points

$$\sum_{\lambda \in V_0(O)} SPSD(M,\lambda) \approx G_L \cdot (L+2) \left(\frac{1}{(M-1)!}\right)^{\frac{2}{M-1}} \cdot \frac{\gamma_M}{\prod_{i=2}^{M} \beta_i^{\frac{2}{(M-1)L}+1}} \cdot N \cdot N^{\frac{M}{M-1}\cdot\frac{2}{L}} v^{\frac{2}{L}} \tag{3.3.20}$$

where G_L is the dimensionless normalized second moment of an L-dimensional sphere, and γ_M is given by

$$\gamma_M = \begin{cases} \frac{1}{L+2}, & M = 2 \\ \varphi_3 + \psi_3, & M = 3 \\ \varphi_M \cdot \prod_{i=2}^{M-1} \beta_i + \gamma_{M-1}\psi_M, & M \geq 4 \end{cases} \tag{3.3.21}$$

with φ_M given in (3.3.22) and ψ_M in (3.3.23). The proof of (3.3.20) is presented

in Appendix B.

$$\varphi_M = \begin{cases} 1, & M = 2 \\ \frac{(M-1)^{1+(M-3)L}}{(M^2-2M)^{(M-2)L/2}} \cdot \sum_{n=0}^{\frac{(M-2)L}{2}} \binom{\frac{(M-2)L}{2}}{n} \frac{(-1)^n}{L+2n+2}, & M \geq 3 \end{cases} \tag{3.3.22}$$

$$\psi_M = \begin{cases} \frac{3-1}{3^{L/2}(L+2)} \cdot \sum_{n=0}^{\frac{L+2}{2}} \binom{\frac{L+2}{2}}{n} \frac{(-1)^n}{L+2n}(1+O(1))^{-\frac{2}{L}}, & M = 3 \\ \frac{(M-1)^{1+(M-3)L}}{(M-2)^{1+(M-2)L/2}M^{(M-2)L/2}} \cdot \sum_{n=0}^{\frac{(M-2)L+2}{2}} \binom{\frac{(M-2)L+2}{2}}{n} \frac{L\cdot(-1)^n}{L+2n+2}, & M \geq 4 \end{cases} \tag{3.3.23}$$

Substituting (3.3.20) into (3.2.4) and skipping the term of $\sum_{\lambda \in V_0(O)} \|\lambda - G_M\|^2$ which is negligible compared with the term of $\sum_{\lambda \in V_0(O)} SPSD(M, \lambda)$ under the third assumption [70], we can obtain the closed-form expression for D_k as

$$\begin{aligned} D_k &\approx d_c + \tfrac{1}{N} \cdot \tfrac{M-k}{M^2k(M-1)} \sum_{\lambda \in V_0(O)} SPSD(M, \lambda) \\ &\approx G(\Lambda) v^{2/L} \\ &+ G_L \cdot (L+2) \cdot \tfrac{M-k}{M^2k(M-1)} \cdot \left(\tfrac{1}{(M-1)!}\right)^{\frac{2}{M-1}} \cdot \frac{\gamma_M}{\prod_{i=2}^{M} \beta_{iL}^{\frac{2}{(M-1)L}+1}} \cdot N^{\frac{M}{M-1}\cdot\frac{2}{L}} v^{\frac{2}{L}} \end{aligned} \tag{3.3.24}$$

It can be seen that D_k can be expressed through the normalized second moment of a sphere G_L, which is a comparable analytical result as those in [70] and [79].

2) Optimal v and N for Minimizing D_{exp}

We now derive expressions of the optimal v and N. From (3.2.6) we have

$$Nv = 2^{L(h(X)-R_s)} \triangleq \rho \tag{3.3.25}$$

where ρ is a constant. Substituting (3.3.20) into (3.2.5) we obtain a closed-form expression of the overall expected distortion as

$$\begin{aligned} D_{exp} &= \sum_{k=0}^{M} \begin{pmatrix} M \\ k \end{pmatrix} (1-p_l)^k p_l^{M-k} D_k \\ &= (1-p_l^M) d_c + \eta G_L N^{\frac{M}{M-1}\frac{2}{L}} v^{\frac{2}{L}} + p_l^M E[\|X\|^2] \end{aligned} \tag{3.3.26}$$

where

$$\eta = \eta_1 (L+2) \left(\frac{1}{(M-1)!} \right)^{\frac{2}{M-1}} \cdot \frac{\gamma_M}{\prod_{i=2}^{M} \beta_i^{\frac{2}{(M-1)L}+1}}$$

Differentiating D_{exp} w.r.t. v and equating to zero yields optimal v as

$$v_{opt} = \rho \cdot \left(\frac{\eta G_L}{(1-p_l^M) G(\Lambda)(M-1)} \right)^{\frac{L(M-1)}{2M}} \tag{3.3.27}$$

Substituting the v_{opt} into (3.3.25), we obtain optimal N as

$$N_{opt} = \left(\frac{\eta G_L}{(1-p_l^M) G(\Lambda)(M-1)} \right)^{-\frac{L(M-1)}{2M}} \tag{3.3.28}$$

As an example, we consider an A_2 lattice for 3DLVQ and show the performance of optimal N (for minimizing D_{exp}) versus p_l in Fig. 3.5. The $N - p_l$ performance of 2DLVQ is also plotted as a reference. It can be seen that the

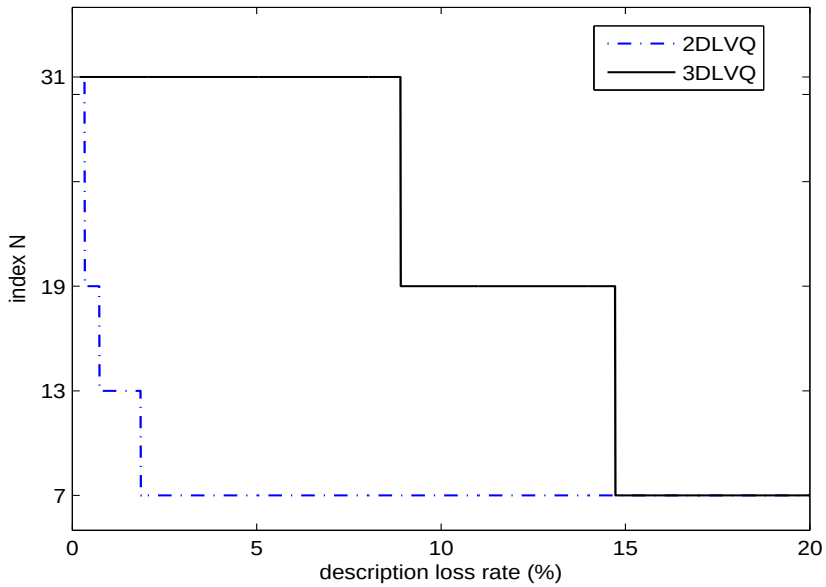


Figure 3.5: Optimal index number N versus description loss probability for 2DLVQ and 3DLVQ based on A_2 lattice.

optimal N can be considered as a monotonic function of the description loss probability p_l for a given M. A larger p_l will lead to a smaller N. Note that we do not show the optimal N larger than 31, since as $p_l \to 0$, the optimal $N \to \infty$.

3.4 Experimental Results

To validate the effectiveness of the proposed index assignment schemes, we consider encoding Gaussian, speech and image sources with the proposed schemes and the two schemes in [70] and [79] for comparison. 3DLVQ with an A_2 lattice is used as an example. Note that our index assignment design can perform well to other lattices (which have M-fraction lattices) and other number of descriptions, since we do not specify the type of lattice and the number of descriptions in our analysis.

3.4.1 Gaussian Source

We quantize a two-dimensional zero-mean and unit variance Gaussian source with the different 3DLVQ schemes. To make a fair comparison, the experimental setting is the same as in [70], which includes 2×10^6 testing vectors with entropy of 5 bits/dimension for each description. Index numbers of 31, 37, 43 and 61 are used for testing. Description balance is achieved by time sharing through a suitable mixing of labels. The central distortions D_c for all the schemes are the same for a given index number since the same fine lattice is used for all the schemes. Table 3.3 shows the comparison of 1-description (averaged) side distortion D_1 among the different schemes, while the results for 2-description (averaged) side distortion D_2 are listed in Table 3.4, where "Theo" and "Exp" represent the theoretical and experimental results, respectively. All the theoretical results in both tables are obtained from (3.2.4). In [70], the state $s^{(1)}$ mentioned in Subsection 3.3.1 is used to calculate both D_1 and D_2 based on (3.2.4). Similarly, $s^{(2)}$ is used for the scheme of [79]. [70] also provides a group of theoretical results for D_1 and D_2 based on the asymptotic k-description (averaged) side distortion formula therein, which is claimed to be exact for infinite dimension ($L \to \infty$) lattices. In our experiments with an A_2 lattice ($L = 2$), the asymptotic approximation in [70] results in a larger gap between experimental results and theoretical values than using (3.2.4) in this chapter. The experimental results for the method in [70] are extracted from [70]. Due to the randomness in generating a Gaussian source, the experimental results for the same index assignment may vary with negligible differences. For instance, in the case of $N = 31$ in Table 3.3, the experimental result of [70] is a little different from ours with the same theoretical result. In the proposed algorithms, the states will change with different N and D_k. The proposed greedy algorithm is found to produce the same index assignment as the

Table 3.3: Comparison of 1-description (averaged) side distortion D_1 (in dB) at 5 bits/dim/description

	Ref. [79]		Ref. [70]		Proposed	
N	Theo.	Exp.	Theo.	Exp.	Theo.	Exp.
31	-25.0943	-25.0901	-25.6725	-25.6729	-25.6725	-25.6767
37	-24.4183	-24.4211	-25.0345	-24.5521	-25.0345	-24.9808
43	-24.4120	-24.4098	-24.5879	-24.5647	-24.5879	-24.5817
61	-23.7976	-23.7941	-23.8737	-23.8757	-23.8737	-23.8724

Table 3.4: Comparison of 2-description (averaged) side distortion D_2 (in dB) at 5 bits/dim/description

	Ref. [79]		Ref. [70]		Proposed	
N	Theo.	Exp.	Theo.	Exp.	Theo.	Exp.
31	-30.3626	-30.3642	-30.7325	-30.7335	-30.7325	-30.7347
37	-29.9443	-29.9250	-30.3386	-29.8850	-30.3386	-30.3195
43	-29.8073	-29.8081	-29.9412	-29.9159	-29.9412	-29.9305
61	-29.3122	-29.3096	-29.2884	-29.2992	-29.3122	-29.3096

proposed general algorithm in all the cases we tested. Therefore we only present the results of the proposed greedy algorithm.

From Table 3.3 and Table 3.4 it can be seen that in the case of index number N being 31, 37 or 43, the scheme of [70] outperforms that of [79] both in D_1 and D_2. However, the scheme of [79] achieves a smaller theoretical and experimental D_2 than the method of [70] when $N = 61$. It implies that using only one state (e.g., $s^{(1)}$ in [70] or $s^{(2)}$ in [79]) cannot always achieve minimum D_1 and D_2 simultaneously for some finite index numbers. If we consider expected distortion $D_{exp} = (1-p_l)^3 d_c + 3p_l(1-p_l)^2 D_2 + 3p_l^2(1-p_l)D_1 + p_l^3 D_0$, it can be further inferred that D_{exp} is dependent on the description loss probability p_l for a given N, because no state can minimize D_1 and D_2 simultaneously for some index numbers, which makes the minimization of D_{exp} be dependent on p_l.

Our proposed schemes can always achieve the best D_1 and D_2 for all the index numbers in both tables. The reason is that the proposed algorithms are applied

Table 3.5: Comparison of 1-description (averaged) side distortion D_1 (in dB) at 2 bits/dim/description

	Ref. [79]		Ref. [70]		Proposed	
N	Theo.	Exp.	Theo.	Exp.	Theo.	Exp.
31	-7.0325	-7.0304	-7.6107	-7.6115	-7.6107	-7.6115
43	-6.3502	-6.3522	-6.5261	-6.5152	-6.5261	-6.5152
61	-5.7358	-5.7450	-5.8119	-5.8001	-5.8119	-5.8001

Table 3.6: Comparison of 2-description (averaged) side distortion D_2 (in dB) at 2 bits/dim/description

	Ref. [79]		Ref. [70]		Proposed	
N	Theo.	Exp.	Theo.	Exp.	Theo.	Exp.
31	-12.3008	-12.3156	-12.6707	-12.6733	-12.6707	-12.6733
43	-11.7455	-11.7676	-11.8794	-11.8640	-11.8794	-11.8640
61	-11.2504	-11.2689	-11.2266	-11.2285	-11.2504	-11.2689

to jointly minimizing both terms of a certain D_k, where more states (including $s^{(1)}$ and $s^{(2)}$) are considered for the best choice. Note that in this case, the local adjustment method in [1] can also obtain the same results as the proposed methods. However, as we have discussed in Section 3.3, the local adjustment method does not always lead to optimal solution, e.g., in the example shown in Fig. 3.4.

For performance test at a lower rate, we also quantize the Gaussian source at the entropy of 2 bits/dimension for each description. Table 3.5 shows the comparison of 1-description side distortion D_1 among the different schemes, whilst the results for 2-description side distortion D_2 are listed in Table 3.6. From the tables, we can also see that neither of the schemes in [70] and [79] can always achieve the minimum distortion for the three testing index numbers. For instance, in Table 3.5 the D_1 values for the method of [79] are worse for all the three index numbers, while in Table 3.6 the scheme of [70] has a larger D_2 value in the case of $N = 61$. In contrast, our proposed scheme always achieves minimum D_1 and D_2 in all the cases.

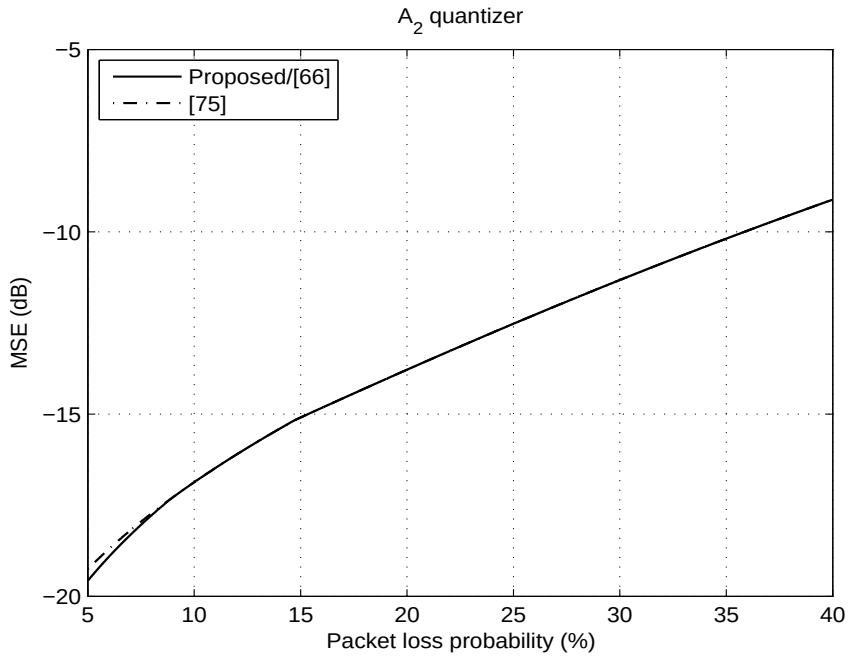


Figure 3.6: Distortion as a function of packet-loss probability for the A_2 quantizer by using three different 3DLVQ. The target entropy is 2 bits/dimension/description.

We now show the expected distortion as a function of the packet-loss probability for three different 3DLVQ systems in Fig. 3.6. We sweep the packet-loss probability p_l in the range of $[0.05, 0.4]$[xiii] in steps of $1/10000$. For each p_l we measure the distortion for all admissible index numbers and use the index number which gives the lowest distortion. We can see from the figure that the curve of [79] gets worse when p_l is close to 5%, it is because that $N = 31$ is selected as the optimal index number in [79], while side distortions by using $N = 31$ are worse than those of proposed scheme and [70]. Please also refer to Tables 3.5 and 3.6 for details.

[xiii] According to [70], in the case of $p_l < 5\%$, 2DLVQ is better than 3DLVQ, and in the case of $p_l > 40\%$, the index number $N = 1$ for 3DLVQ. Therefore we only choose 5% to 40% for testing.

3.4.2 Speech Source

The testing 23600 16-bit samples of speech are from one female speaker sampled at 8 kHz, and the different 3DLVQ schemes are applied to the original speech signal for comparison. The SNR results are summarized in Table 3.7, where $\text{SNR} = 10\log_{10}(\sigma_x^2/MSE)$, and σ_x^2 is the variance of the source sample x. $\text{SNR}_{i,j}$ is the 2-description side distortion when the i-th and the j-th descriptions are obtained and SNR_i is the side distortion value when only the i-th description is received. The total bit rate (after adaptive arithmetic coding of the indexes) for each row in the table is 51.92, 52.87 and 53.24 kbps, respectively. For $N = 7$, the three schemes produce the same index assignment thus leading to the same experimental results. For $N = 31$, the proposed scheme and the scheme in [70] can achieve smaller side distortions as opposed to the scheme in [79], with the same central distortion.

3.4.3 Image Source

We also compare the proposed scheme against the schemes in [79] and [70] for image coding. The 512×512 8-bit "Boat" image and 256×256 8-bit "Peppers" image are coded at the total bit rate of 1 bit per pixel (bpp). Five-level wavelet decomposition using the 9/7-tap filter is applied to the testing images. Two neighboring wavelet coefficients in the same subband are formed as an input vector. After the 3DLVQ encoding, the produced indexes are coded by adaptive arithmetic coding. Two index numbers of 7 and 31 are tested. For the different index numbers, we apply different quantization parameters in the subbands for the LVQ coding to achieve the same total bit rate of 1 bpp. Note that coding optimization or bit allocation for quantization parameter adaption in different

Table 3.7: SNR comparison of the different 3DLVQ schemes for speech coding.

	N	SNR_{all}	$\text{SNR}_{1,2}$	$\text{SNR}_{2,3}$	$\text{SNR}_{1,3}$	SNR_1	SNR_2	SNR_3
all	7	25.3152	19.3553	19.2735	19.3313	14.9678	14.9324	15.0189
proposed/ [70]	31	31.1733	18.4929	18.4914	18.4724	13.4769	13.4633	13.4227
[79]	31	31.1733	18.1767	18.1534	18.1863	12.9556	12.9542	12.9140

Table 3.8: PSNR comparison of the different 3DLVQ schemes in coding "Boat" at 1 bpp (total rate).

	N	PSNR_{all}	$\text{PSNR}_{1,2}$	$\text{PSNR}_{2,3}$	$\text{PSNR}_{1,3}$	PSNR_1	PSNR_2	PSNR_3
all	7	32.98	29.67	29.65	29.65	25.99	26.02	26.00
proposed/ [70]	31	34.04	24.95	24.95	24.97	20.73	20.73	20.74
[79]	31	34.04	24.86	24.86	24.86	20.54	20.56	20.55

Table 3.9: PSNR comparison of the different 3DLVQ schemes in coding "Peppers" at 1 bpp (total rate).

	N	PSNR_{all}	$\text{PSNR}_{1,2}$	$\text{PSNR}_{2,3}$	$\text{PSNR}_{1,3}$	PSNR_1	PSNR_2	PSNR_3
all	7	32.38	28.51	28.60	28.51	24.78	24.74	24.76
proposed/ [70]	31	34.17	24.56	24.64	24.65	20.37	20.36	20.31
[79]	31	34.17	24.46	24.42	24.54	20.08	20.13	20.11

subbands can be applied to the encoding process as in [142] and [111]. For simplicity, such coding optimization is not considered, since our focus is to evaluate the different 3DLVQ schemes in terms of quantization performance with the same quantization parameter for all the subbands in all the schemes for a given index number.

Table 3.8 and Table 3.9 show the reconstruction results for "Boat" and "Peppers", respectively. From both tables we can see that by changing the index number N, a tradeoff between central and side distortions associated with different levels of redundancy is shown. Again, similar to the results of Gaussian and speech source coding, both the proposed scheme and the scheme in [70] achieve smaller side distortions than the scheme in [79], with the same central distortion. The subjective quality of reconstruction images for all the schemes using $N = 7$ is shown in Fig. 3.7.

In practice, MDLVQ coding is normally applied with a small index value [142], e.g., 7 to 31 for A_2 lattice. With a larger index number, the side coding performance will drop significantly with very poor side distortion results which may not make practical sense. Therefore, we only tested the index numbers no more than 31 for speech and image sources using the A_2 lattice based 3DLVQ. If a higher dimensional lattice or more descriptions based LVQ is considered, a larger index number can be used.

3.5 Conclusion

In this chapter, we have investigated the existing MDLVQ index assignment schemes, specifically, the schemes in [70], [79] and [1]. To address non-optimality or high complexity problems in these schemes, we have proposed two new MDLVQ index assignment algorithms with moderate complexity. The proposed methods take advantage of M-fraction lattice for index assignment design, which can help construct sublattice tuples easily and simplify the design of labeling function. We have then translated the index assignment optimization to transportation problems in operations research in the development of the two new algorithms, one being a greedy algorithm and the other a general algorithm. In particular, the proposed greedy algorithm aims to reduce computational complexity without losing optimality, which has exhibited lower complexity than the linear assignment method [70], thus making it a good option for a high dimension lattice based MDLVQ design. We conjecture that the proposed greedy algorithm can provide an optimal index assignment, which produces the same results as the optimal general algorithm shown in our experiments of coding different sources. Furthermore, we have made an asymptotical analysis of the k-description (averaged)

side distortion and obtained its closed-form expression based on the proposed index assignment scheme. We have validated the proposed schemes in the case of 3DLVQ coding of different sources.

(a) Original image

(b) Central PSNR=34.17 dB by all

(c) $PSNR_{1,2}$=24.46 dB by [79]

(d) $PSNR_{1,2}$=24.56 dB by proposed/ [70]

(e) $PSNR_1$=20.08 dB by [79]

(f) $PSNR_1$=20.37 dB by proposed/ [70]

Figure 3.7: Subjective quality comparison of the three different 3DLVQ schemes in coding "peppers" with index number $N = 31$.

Chapter 4

Two-Stage Multiple Description Quantization and Its Application in Image Coding

4.1 Introduction

The MD problem formulated in [31] usually deals with two descriptions given a source and a distortion measure. As shown in Fig. 2.1, a multiple description encoder generates two descriptions for one source. Both descriptions are transmitted over separate channels respectively, and each can be individually decoded with a certain level of fidelity corresponding to a side distortion. Since these two descriptions are also designed to complement each other, the incremental reconstruction quality associated with a so-called central distortion is achievable with both received descriptions. In [26], an achievable rate region with a memoryless source and a certain fidelity criterion was constructed based on

Shannon's random coding approach. Ozarow [27] has shown that the achievable rate region derived in [26] is the rate distortion region for the case of a memoryless Gaussian source with respect to the squared-error distortion criterion. In [40] and [34], it is shown that the product of central and side distortions is bounded by a constant given a fixed rate, which can be used as the information theoretic bound for the quantizers.

As a practical solution to the balanced two-channel MD problem, MD scalar quantization (MDSQ) was first proposed in [13] and [54]. The asymptotic analysis in [40] and [41] shows that the distortion product of the central and side distortions for entropy-constrained MDSQ (ECMDSQ) with a uniform central quantizer is 3.066 dB away from the rate-distortion bound (or distortion product bound) for memoryless Gaussian source. Recently, balanced two-stage multiple description quantization (MDQ) methods were discussed in [16], [17] and [18]. In the first stage, two descriptions are generated using two quantizers. In the second stage, a refinement is generated by encoding the estimation error of the source from the first stage. The two-stage MDQ features an efficient and flexible mechanism to control the central-side distortion tradeoff. In [16], independent dithering in the first stage and dependent dithering (corresponding to staggering) as a possible improvement were discussed briefly with respect to the ratio of side to central distortion, rather than the distortion product. Though common random dither leads to a mathematically precise analysis, schemes based on random dither are often difficult to use in practical applications [140]. Tian and Hemami [17] developed a modified MDSQ (MMDSQ) which achieves the same performance as ECMDSQ but through a simpler manner.

In recent years, many two-stage MD image coding schemes have been devel-

oped [17,18,112,113]. Besides the advantage of the flexible mechanism to control the central-side distortion tradeoff due to the two-stage structure, another advantage is that some existing efficient single description image coding schemes, such as image coding based on set partitioning in hierarchical trees (SPIHT) [149], JPEG2000 [150], tarp filter with classification based image coding (TCE) [151], trellis coded quantization (TCQ) based image coding [152], can be adopted with little change in the two-stage MD image coding. Lin et al [18] applied two-stage MDQ in image coding based on spatial diversity and trellis coded quantization, which has obtained better results than most of one-stage and two-stage image coding schemes.

In this chapter, we first revisit the problem on further reducing granular distortion in the balanced two-stage MDSQ with large side-to-central distortion ratios under the high-resolution assumption. We show that the refinement information at the second stage can be used to improve side coding performance more or less. Based on this observation, we enhance a specific two-stage MDSQ, i.e. MMDSQ in [17], in terms of distortion product and present asymptotic expressions of the side and central distortions for a memoryless Gaussian source. Then we show that the distortion product can be further reduced when a good lattice instead of the scalar quantizer as the refinement quantizer in the second stage. Finally, we propose a practical two-stage multiple description image coding based on trellis coded quantization, and validate it by experimental results.

4.2 Asymptotic Analysis of a General Two-stage MDSQ

4.2.1 General Two-stage MDSQ

In a general baseline two-stage MDSQ shown in Fig. 4.1, two descriptions are generated as follows. First, quantization indices from each side quantizer are entropy coded into a bitstream of each description. Second, the indices from the second stage are split evenly between two descriptions and again entropy coded separately.

In the first stage, denote Q_i $(i = 1, 2)$ as the scalar quantizer used at description i. Assume x_i is the side reconstruction value of an original input sample x at the first stage, then

$$x_1 = Q_1^{-1}\left(Q_1(x)\right) = x + e_1 \tag{4.2.1}$$

$$x_2 = Q_2^{-1}\left(Q_2(x)\right) = x + e_2 \tag{4.2.2}$$

where e_i is the quantization error by the quantizer Q_i. Denote Q_J as the joint quantizer in the first stage. x_c is the reconstruction value by using Q_J, which is just the average of x_1 and x_2, then

$$x_c = Q_J^{-1}\left(Q_J(x_1, x_2)\right) = \frac{x_1 + x_2}{2} = x + e_c \tag{4.2.3}$$

where the residual error e_c for x_c is given by

$$e_c = x_c - x = \frac{x_1 + x_2}{2} - x = \frac{e_1 + e_2}{2} \tag{4.2.4}$$

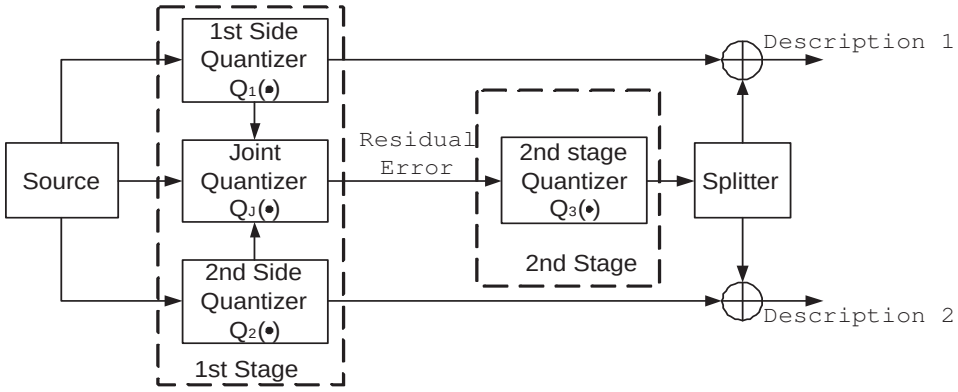


Figure 4.1: Structure of a general two-stage MDSQ.

In the second stage, Q_3 is used as the scalar quantizer for quantizing the residual error e_c from the first stage. Thus the reconstructed counterpart of e_c is

$$e'_c = Q_3^{-1}\left(Q_3(e_c)\right) = e_c + \delta'_c = \frac{e_1 + e_2}{2} + \delta'_c \tag{4.2.5}$$

where δ'_c is the quantization error due to $Q_3(\cdot)$. The second stage is to refine the reconstruction results x_1, x_2 and x_c obtained in the first stage. The final central reconstruction of x'_c, which is the value after refining x_c, is given by

$$x'_c = x_c - e'_c = (x + e_c) - (e_c + \delta'_c) = x - \delta'_c \tag{4.2.6}$$

Similarly, the final side reconstructions of x'_1 and x'_2 by taking into account the refinement information e'_c are

$$x'_1 = x_1 - e'_c = x_1 - e_c - \delta'_c = x + e_1 - e_c - \delta'_c \tag{4.2.7}$$

$$x'_2 = x_2 - e'_c = x_2 - e_c - \delta'_c = x + e_2 - e_c - \delta'_c \tag{4.2.8}$$

4.2.2 Asymptotic Distortion Analysis

For the balanced two-stage MDSQ, we denote the rates in the first and second stage as R_1 and R_2 per description, respectively. The side distortion d_{s1} and d_{s2} in the first stage are given by

$$d_{s1} = E[(x - x_1)^2] = E[e_1^2] = d_s \tag{4.2.9}$$

$$d_{s2} = E[(x - x_2)^2] = E[e_2^2] = d_s \tag{4.2.10}$$

and the central distortion d_c in the first stage is

$$\begin{aligned} d_c &= E[(x - x_c)^2] = E[e_c^2] = E[(\tfrac{e_1+e_2}{2})^2] \\ &= \tfrac{1}{4} \cdot E[e_1^2] + \tfrac{1}{4} \cdot E[e_2^2] + \tfrac{1}{2} \cdot E[e_1 e_2] \\ &= \tfrac{1}{2} \cdot (1 + \rho) \cdot d_s \end{aligned} \tag{4.2.11}$$

where ρ is the correlation coefficient between e_1 and e_2.

Now we consider using e_c' obtained in the second stage to refine d_c, d_{s1} and d_{s2}. The central distortion after the refinement, d_c', is given by

$$d_c' = E[(x - x_c')^2] = E[(\delta_c')^2] < E[e_c^2] = d_c \tag{4.2.12}$$

It is known that when R_1 is sufficiently high, the side quantizer bin in the first stage is sufficiently small such that the probability density function (pdf) in each single bin is almost uniform [140]. Under this assumption and in the case of two side quantizers with staggered bins being used in the first stage, d_c' can be

approximated as [17]

$$d_c' \approx d_c/N^2 \tag{4.2.13}$$

where N is the number of fine bins to divide the joint quantizer Q_J for the second stage quantizer.

As shown in Fig. 4.1, quantized residual errors are split in half into two descriptions. Therefore only half of the values in description 1 are refined and half are unchanged. In this sense, the side distortion after refinement for description 1 is

$$\begin{aligned}
d_{s1}' &= \tfrac{1}{2}\cdot E[(x-x_1)^2] + \tfrac{1}{2}\cdot E[(x-x_1')^2] \\
&= \tfrac{1}{2}\cdot E[e_1^2] + \tfrac{1}{2}\cdot E[(e_1-e_c-\delta_c')^2] \\
&= \tfrac{1}{2}\cdot E[e_1^2] + \tfrac{1}{2}\cdot E[(e_1-e_c)^2 - 2\cdot(e_1-e_c)\cdot\delta_c' + (\delta_c')^2] \\
&= \tfrac{1}{2}\cdot E[e_1^2] + \tfrac{1}{2}\cdot E[\big(\tfrac{e_1-e_2}{2}\big)^2 - 2\cdot\tfrac{e_1-e_2}{2}\cdot\delta_c' + (\delta_c')^2] \\
&= \tfrac{1}{2}\cdot d_s + \tfrac{1}{2}\cdot E[\big(\tfrac{e_1-e_2}{2}\big)^2] - E[\tfrac{e_1-e_2}{2}\cdot\delta_c'] + \tfrac{1}{2}\cdot d_c' \\
&= \tfrac{1}{2}\cdot d_s + \tfrac{1}{2}\cdot\big(\tfrac{1}{4}\cdot 2\cdot d_s - \tfrac{1}{4}\cdot 2\cdot\rho\cdot d_s\big) - E[\tfrac{e_1-e_2}{2}\cdot\delta_c'] + \tfrac{1}{2}\cdot d_c' \\
&= \tfrac{3-\rho}{4}\cdot d_s - \tfrac{1}{2}\cdot(\rho_1-\rho_2)\sqrt{d_s d_c'} + \tfrac{1}{2}\cdot d_c' \\
&\approx \tfrac{3-\rho}{4}\cdot d_s
\end{aligned} \tag{4.2.14}$$

where ρ_i $(i=1,2)$ is the correlation coefficient between e_i and d_c'. The approximation in the last step in (4.2.14) is made by assuming $d_s >> d_c'$ (or equivalently a sufficiently large R_2), which is normally considered in the asymptotic analysis [41] [16]. We consider two staggered scalar quantizers in the first stage and then the term $E[\frac{e_1-e_2}{2}\cdot\delta_c'] = E[\frac{e_1-e_2}{2}]\cdot E[\delta_c'] = 0$ in (4.2.14), where e_1-e_2 is a constant Δ or $-\Delta$ (Δ is the bin length in the joint quantizer Q_J) due to the staggering structure and $E[\delta_c'] = 0$. Similarly, the side distortion after the refinement

for the second description can be obtained as $d'_{s2} = d'_{s1}$.

From (4.2.12), (4.2.13) and (4.2.14) it can be seen that e'_c can be employed to reduce both central and side distortions in view that $(1+\rho)/2 \leq 1$ and $(3-\rho)/4 \leq 1$ for any ρ $(-1 \leq \rho \leq 1)$. The side distortion improvement increases as ρ changes from -1 to $+1$ whereas the coding gain in central distortion decreases with the increasing ρ. Specifically, side distortion can achieve a gain of $-10 \cdot \log_{10}(\frac{3-\rho}{4})$ which ranges from 1.25 dB at $\rho = 0$ to 3 dB at $\rho = 1$. Such two-stage MDQ with a non-negative ρ can be found in [16] and [18], where independent dithered quantization and spatial diversity based quantization are used in the first stage, respectively. On the other hand, central distortion benefits more with a negative ρ. In [16] and [39], it is suggested that a negative ρ may lead to a smaller distortion product, which motivates us to consider a two-stage structure with a negative ρ. Specifically, the modified multiple description scalar quantization (MMSDQ) [17] with a negative $\rho = -1/2$ will be examined in the next section.

4.3 Enhancing MMDSQ

In the following, we will present a scheme to enhance the coding performance for the MMDSQ shown in Fig. 4.2. Under the assumption of large side-to-central distortion ratios, a coding gain of 0.58 dB in the side distortion can be obtained by enhancing side decoding with the refinement information from the second stage.

In the MMDSQ, the first stage consists of two uniform side scalar quantizers Q_1 and Q_2 with staggered bins. The joint quantizer Q_J formed by these two staggered quantizers has a smaller half-size bin. As shown in Fig. 4.2, we denote 2Δ as the length of bins in Q_1 or Q_2 and Δ as the length of bins in Q_J. The

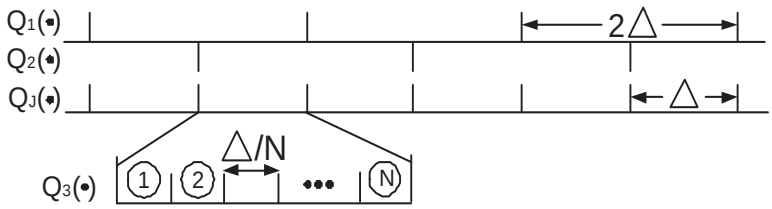


Figure 4.2: Structure of the MMDSQ.

second stage scalar quantizer Q_3 uniformly divides the joint quantizer Q_J in the first stage into N finer bins corresponding to the length of Δ/N. In the MMDSQ, the refinement information at the second stage is considered to be useful only for the central reconstruction and split evenly between two descriptions after entropy coding. We propose applying the refinement information to enhancing the side coding performance as analyzed in Subsection 4.2.2. Unlike in the MMDSQ, the refinement information in the proposed scheme is split evenly followed by entropy coding, thus the entropy-coded bitstream in each description can be decoded separately. There is a negligible rate penalty when two separate entropy coders are used at the second stage, since the refinement information is approximately uniformly distributed.

4.3.1 Asymptotic Distortion Analysis

Now we investigate the asymptotic rate-distortion expressions for a memoryless Gaussian source using the mean squared error measure. Again, we denote the rates in the first and second stage as R_1 and R_2 per description, respectively, and then $N = 2^{2R_2}$. When R_1 is high, the side-distortion-rate function for the uniform side quantizer Q_1 and Q_2 can be expressed as [140]

$$d_s \approx \frac{(2\Delta)^2}{12} = \frac{1}{12} 2^{2h(p)} 2^{-2R_1} = \frac{\pi e}{6} \cdot 2^{-2R_1} \tag{4.3.15}$$

where the differential entropy of the source, $h(p)$, is $\log_2 \sqrt{2\pi e}$ for a memoryless unit variance Gaussian source (please also refer to (2.2.10)). Since the joint quantizer Q_J reduces the first-stage bin by half, we have

$$d_c \approx \frac{\Delta^2}{12} = \frac{1}{4} \cdot d_s = \frac{1}{4} \cdot \frac{\pi e}{6} \cdot 2^{-2R_1} \tag{4.3.16}$$

The second stage quantizer Q_3 divides the bin of Q_J into finer uniform bins (with length of Δ/N). The central distortion after the refinement, d'_c, is reduced as [17]

$$d'_c \approx \frac{(\Delta/N)^2}{12} = \frac{1}{N^2} \cdot d_c = \frac{1}{4} \cdot \frac{\pi e}{6} \cdot 2^{-2(R_1+2R_2)} \tag{4.3.17}$$

We now consider the side decoding enhancement. Without loss of generality, we assume an input sample x is quantized to x_1 by Q_1 and the bin '2' by Q_3, as shown in Fig. 4.3. At the side decoder 1, only x_1 and the finer bin '2' are known. Since the information from Q_2 is absent, the side decoder 1 cannot determine whether the bin '2' is in the left bin of Q_J centered at $x_c^{(L)}$ or the right bin centered at $x_c^{(R)}$. Therefore in the original MMDSQ, the information of bin '2' from Q_3 is not used in the side decoding, where the side decoding result is simply treated to be x_1 from Q_1, resulting in the side distortion of $(2\Delta)^2/12 = \Delta^2/3$. Let us re-examine the side decoding at the side decoder 1 given a 2Δ interval centered at x_1 and the finer bin '2'. Denote $x'^{(L)}_c$ ($x'^{(R)}_c$) as the centroid of the finer bin '2' on the left (right) in the figure. It is known that the distance between the finer bins is Δ, and then $x'^{(R)}_c - x'^{(L)}_c = \Delta$. The bin index '2' suggests that x can only fall in one of the two bins labeled by '2', which means the central reconstruction value x'_c for x is either $x'^{(L)}_c$ or $x'^{(R)}_c$ with equal probability. Therefore, it would be better to use the average $x'_1 = (x'^{(L)}_c + x'^{(R)}_c)/2$ to reconstruct x for the side decoding in the sense of minimizing squared error. In this way, the side distortion can be

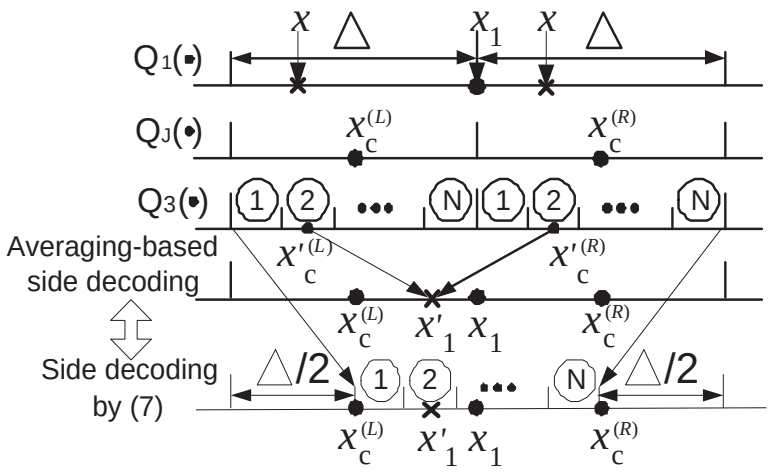


Figure 4.3: Side decoding refinement in the MMDSQ.

approximated as $d_c' + (\Delta/2)^2$ ($< \Delta^2/3$). From the figure it can be seen that the averaging process is in fact to shift the central reconstruction value towards x_1 by a displacement of $\Delta/2$.

We can also directly apply (4.2.7) for the side decoding refinement. Compared with (4.2.6) for the central reconstruction, the side decoding refinement in (4.2.7) is to obtain x_1' by substituting the base x_c ($x_c^{(L)}$ or $x_c^{(R)}$ in Fig. 4.3) with x_1, which is equivalent to shifting towards x_1 with a displacement of $\Delta/2$, since $\left|x_1 - x_c^{(L)}\right| = \left|x_1 - x_c^{(R)}\right| = \Delta/2$. With the geometrical interpretation also shown in Fig. 4.3, we can see that the averaging-based side decoding is exactly the same as the side decoding refinement as shown in (4.2.7).

With the proposed enhancement for side decoding, side distortion can be reduced to $d_c' + (\Delta/2)^2$ from $\Delta^2/3$ in the original MMDSQ. Note that the refinement information is split into two descriptions, and only half of the samples in each

description can be refined. Therefore the side distortion for each description is

$$\begin{aligned} d_s' &= \tfrac{1}{2}\cdot\tfrac{(2\Delta)^2}{12} + \tfrac{1}{2}\cdot\left(d_c' + \left(\tfrac{\Delta}{2}\right)^2\right) = \tfrac{7}{8}\cdot d_s + \tfrac{1}{2}\cdot d_c' \\ &= \tfrac{7}{8}\cdot\tfrac{\pi e}{6}\cdot 2^{-2R_1} + \tfrac{1}{2}\cdot\tfrac{1}{4}\cdot\tfrac{\pi e}{6}\cdot 2^{-2(R_1+2R_2)} \\ &\approx \tfrac{7}{8}\cdot\tfrac{\pi e}{6}\cdot 2^{-2R_1} \end{aligned} \tag{4.3.18}$$

where the approximation in the last step is made by assuming $d_s >> d_c'$ (or a sufficiently large R_2). (4.3.18) is consistent with the result by (4.2.14) in the case of MMDSQ where $\rho = -1/2$. Compared with the side distortion in the original MMDSQ shown in (4.3.15), enhancement can decrease the side distortion by $10\cdot\log_{10}\left(\frac{\pi e/6}{7\pi e/48}\right) = 0.58$ dB, which will be verified in the following numerical evaluations. With the reduced side distortion, the final distortion product becomes

$$d_c' d_s' = \tfrac{7}{8}\cdot\tfrac{1}{4}\cdot\left(\tfrac{\pi e}{6}\right)^2\cdot 2^{-4R} = \tfrac{7\cdot(\pi e)^2}{1152}\cdot 2^{-4R} \tag{4.3.19}$$

Compared with the theoretical distortion product bound (or the rate-distortion bound) for a unit-variance memoryless Gaussian source given in [34], i.e. $d_c^* d_s^* = \frac{1}{4}\cdot 2^{-4R}$ shown in (2.3.17) with $\sigma^2 = 1$, the distortion product in (4.3.19) is $10\cdot\log_{10}\left(\frac{7(\pi e)^2/1152}{1/4}\right) = 2.486$ dB away from the bound.

Remark: Although a simple two-stage structure is applied, the side decoding enhanced MMDSQ outperforms the MDQ framework with a complex structure using scalar quantizers in [39]. According to [39], the MDQ can be constructed by successive scalar quantization along with quantization splitting, which is reported to achieve the distortion product of 2.596 dB away from the distortion product bound, that is, 0.11 dB poorer than the product in (4.3.19) based on our side decoding enhancement. We would also like to point out that under

the high-resolution assumption and large lattice vector quantizer dimension, the information theoretic rate-distortion MD bounds can be achieved by the two-description LVQ [15]. Nonetheless, two-stage MDQ still shows its superiority in practical MD applications for some real-world signal sources.

4.3.2 Numerical Evaluations

A memoryless unit-variance zero-mean Gaussian source with $2 \cdot 10^6$ samples was used in the test. Both analytical and numerical results of coding performance are presented in Fig. 4.4 to compare the proposed MMDSQ enhancement against the (original) MMDSQ at a rate of 4 bits/sample/channel. The theoretical rate-distortion bound in [27] is also plotted in the same figure as a reference. Note that analytical results for MMDSQ are plotted based on (4.3.17) and (4.3.15), while (4.3.17) and (4.3.18) are used to produce the analytical results of the proposed side decoding enhanced MMDSQ. Numerical results for both schemes are indicated by the points 'o' for the proposed scheme, and '+' for the MMDSQ, respectively. From Fig. 4.4 it can be seen that all the numerical points at a high rate closely follow the analytical results, which shows the proposed scheme outperforms the MMDSQ in both analytical and numerical results. Fig. 4.5 illustrates the numerical results of coding performance for both schemes at rates of 2 and 3 bits/sample/channel. We can see that although the performance of both schemes deteriorates at lower rates, the proposed MMDSQ enhancement still achieves an improvement (0.42 dB to 0.58 dB gain in side distortion) compared with the MMDSQ.

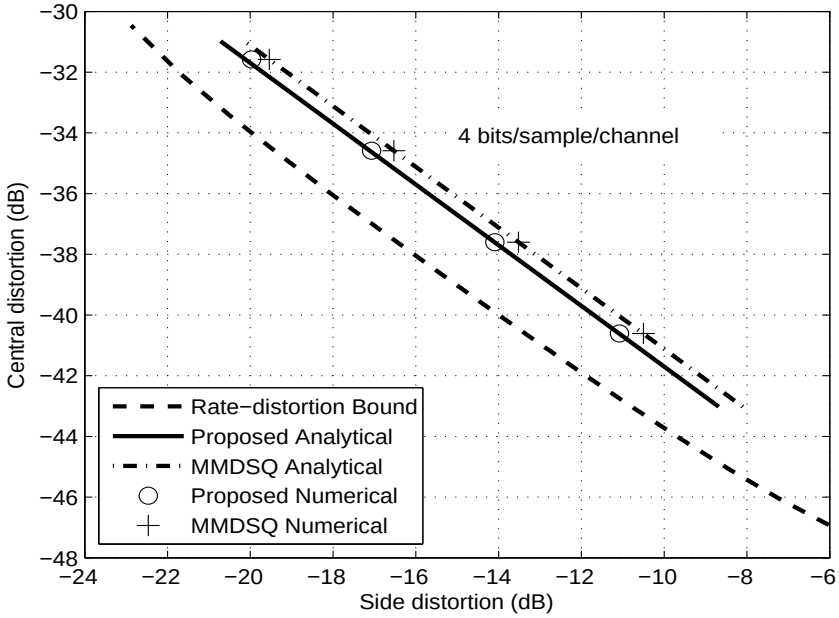


Figure 4.4: Central-side distortion comparison at 4 bits/sample/channel.

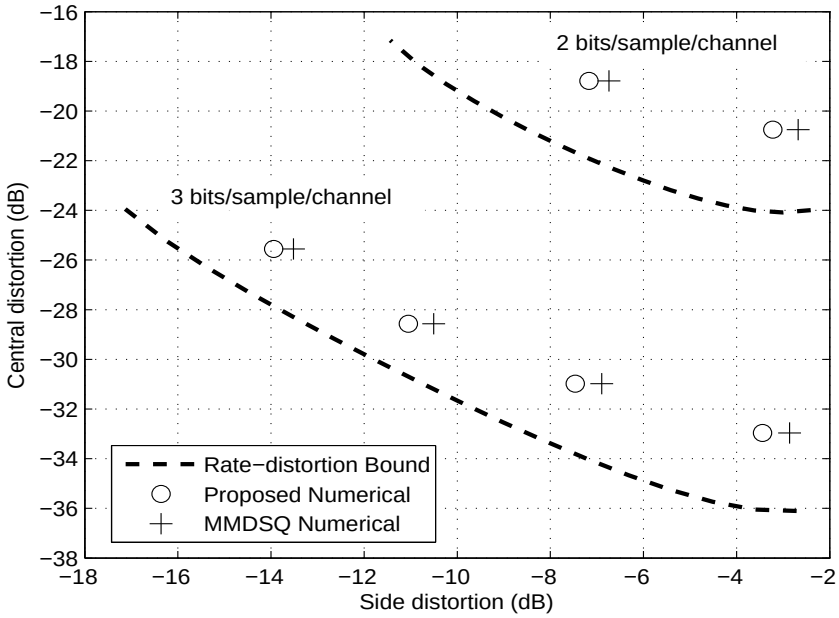


Figure 4.5: Central-side distortion comparison at 2 and 3 bits/sample/channel.

4.4 L-dimensional Lattice Quantization for Q_3

Apart from the proposed side decoding enhancement for the two-stage MMDSQ, central distortion can also be reduced if an L-dimensional lattice is used for Q_3 instead of a simple scalar quantizer under the assumption of $d_s >> d_c'$ (or a sufficiently large R_2). This change will not affect the side decoding enhancement discussed in the above subsection. The reasons are as follows. We have assumed that R_2 is sufficiently high, thus the Voronoi cell of lattice for Q_3 is sufficiently small. Denote $G(\Lambda)$ as the dimensionless normalized second moment of inertia for lattice Λ, and then d_c' can be approximated as [153]

$$d_c' \approx G(\Lambda) \cdot 2^{-4R_2} \cdot \frac{d_c}{1/12} = G(\Lambda) \cdot \frac{\pi e}{2} \cdot 2^{-2(R_1+2R_2)} \tag{4.4.20}$$

without considering boundary effects of lattice quantization. As the lattice dimensionality L increases, it has been shown [153] that there are good lattices with $G(\Lambda) \to \frac{1}{2\pi e}$, thus making $d_c' \to \frac{1}{4} \cdot 2^{-2(R_1+2R_2)}$. On the other hand, in the derivations of (4.2.14) and (4.3.18), we have shown that the terms containing d_c' can be ignored, which leads to an expression of d_s' without taking d_c' into account.

It can be seen that the distortion product can be reduced to $d_c' d_s' \to \frac{7\pi e}{192} \cdot 2^{-4R}$. This value is $10 \cdot \log_{10}\left(\frac{7(\pi e)^2/1152}{1/4}\right) = 0.953$ dB away from the distortion product bound and $10 \cdot \log_{10}\left(\frac{7\pi e/192}{7(\pi e)^2/1152}\right) = 1.533$ dB better than the product in (4.3.19) based on scalar quantization. In addition, compared with the two-stage lattice-based MDQ scheme in [153] where Q_1 and Q_2 are Z^L lattices and Q_3 is a good lattice (in terms of the second moment of inertia), our enhanced two-stage MDQ with the same Q_3 still exhibits a coding gain of 0.58 dB for the asymptotic distortion product due to the side decoding enhancement.

4.5 Proposed Two-Stage MD Image Coding Based on TCQ

4.5.1 Structure of the Proposed Scheme

In [18] a TCQ based two-stage MD image coding scheme has been developed which shows better results than many other two-stage MD image coding schemes (please refer to [18] or the part of experimental results in this chapter). At the first stage of the scheme, an input image is wavelet-transformed and coded by quadtree classification and trellis coded quantization (QTCQ) [152] (i.e. Q_1 in Fig. 4.2) to produce stream 1. To generate stream 2 by Q_2, the original image is first rotated by 180 degrees and then goes through the same coding process as for the first description[i]. Stream 2 is further decoded to obtain an inverted reconstruction. The inverted reconstructed image is contra-rotated in spatial domain and again wavelet-transformed without QTCQ. In this way, wavelet coefficients in both descriptions can be directly averaged by Q_J in Fig. 4.2. A smaller central distortion can be obtained through Q_J^{-1} than side distortion through Q_1^{-1} or Q_2^{-1} at the first stage. At the second stage, the wavelet residuals (between the original wavelet coefficients and the averaged coefficients) are split evenly and coded by QTCQ separately through Q_3. These two parts of residuals are appended to the two existing bitstreams, i.e. streams 1 and 2, respectively. At the decoder side, when both descriptions are received, all the information from the both stages is

[i]It is known that quantizing the signal and its elementary transformation version with vector quantizer (VQ) can form diversity [154]. In view that the wavelet coefficients of the original image and that of the rotated image are different, this diversity can also be considered as transform diversity. The image can be rotated by different angles (such as 90, 180 and 270 degrees) before wavelet transform to generate different descriptions. The quantization style of TCQ is similar to that of VQ, therefore the method in [18] can be considered as a diversity-based MDC.

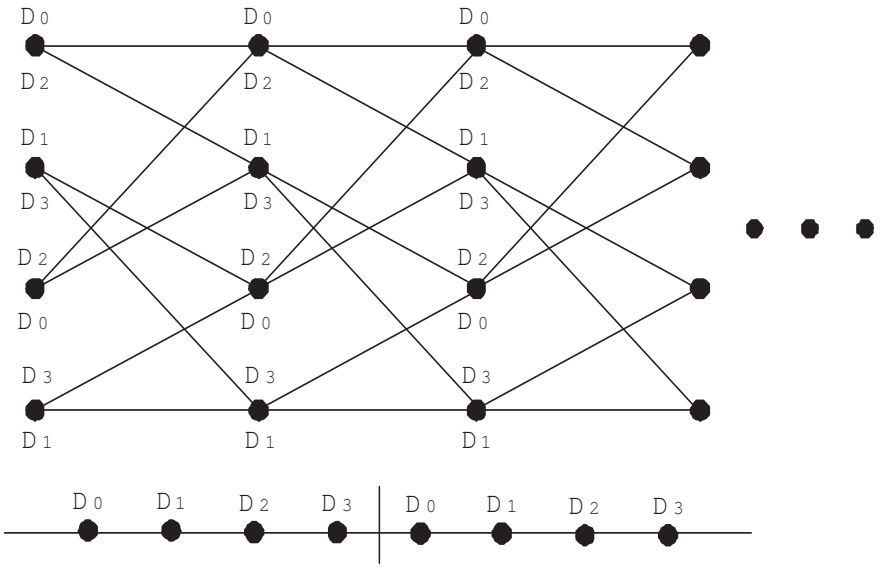


Figure 4.6: A four-state trellis with subset labeling.

used for the central reconstruction at the central decoder. If only one stream and its appended residuals are available, the side decoder will reconstruct a preliminary side reconstruction first, and then enhance it by using the appended residuals for a better side reconstruction quality. The contra-rotation operation has to be applied to stream 2 at Q_2, central decoder and the second side decoder.

One important advantage of the scheme in [18] is the use of QTCQ for encoding. QTCQ has shown its superior performance in image coding [152], which uses the quadtree structure to classify wavelet coefficients and applies TCQ to quantize the classified wavelet coefficients followed by arithmetic coding. An example of a four-state trellis with corresponding codebook and partition (for 2 bits/sample) is shown in Fig. 4.6. How the trellis coded quantization works can be found in [155]. On the other hand, using QTCQ can also help enhance the diversity because the quantization style of TCQ is similar to that of VQ. However, since the averaging and enhancing operations have to be done in wavelet domain in the method of [18], repeating rotation, transform/inverse transform

are required. In the following we propose a simple two-stage scheme which deals with these operations without rotation.

In the proposed two-stage scheme, a wavelet transform is first applied to the original image. At the first stage, all the wavelet coefficients are increased by a constant C followed by QTCQ to generate stream 1. Similarly all the wavelet coefficients are reduced by the same constant C followed by QTCQ to generate stream 2. By subtracting the averaged coefficients of streams 1 and 2 from the original ones, the residual errors of wavelet coefficients are generated by Q_J, which are for further quantizing by Q_3 in the second stage. At the second stage, the residual wavelet coefficients are split evenly to two parts and coded by QTCQ separately, which are appended to the two existing bitstreams. At the decoder side, when both descriptions are received, all the information from the both stages is used for the central reconstruction at the central decoder. If only one stream and its appended residual wavelet coefficients (after quantization) are available, the side decoder will subtract (for stream 1) or add (for stream 2) C to reconstruct a preliminary side reconstruction first, and then enhance (or refine) it by using received residual errors for a better side reconstruction quality. The constant C added or subtracted in the proposed scheme is related to the quantization step size in the QTCQ coder. In the single description QTCQ coding [152], an appropriate TCQ codebook step size Δ has to be selected. Empirical tests show that using $\Delta = 0.7 \cdot T$ for each classes provides most of the TCQ granular gain [152], where $T(> 0)$ is the quality factor for classification. Based on our experiments, the constant C taking a value of 0.5Δ can achieve a good peak signal-to-noise ratio (PSNR) performance in the proposed scheme.

At medium and low bit rates, the deadzone quantization process in the image

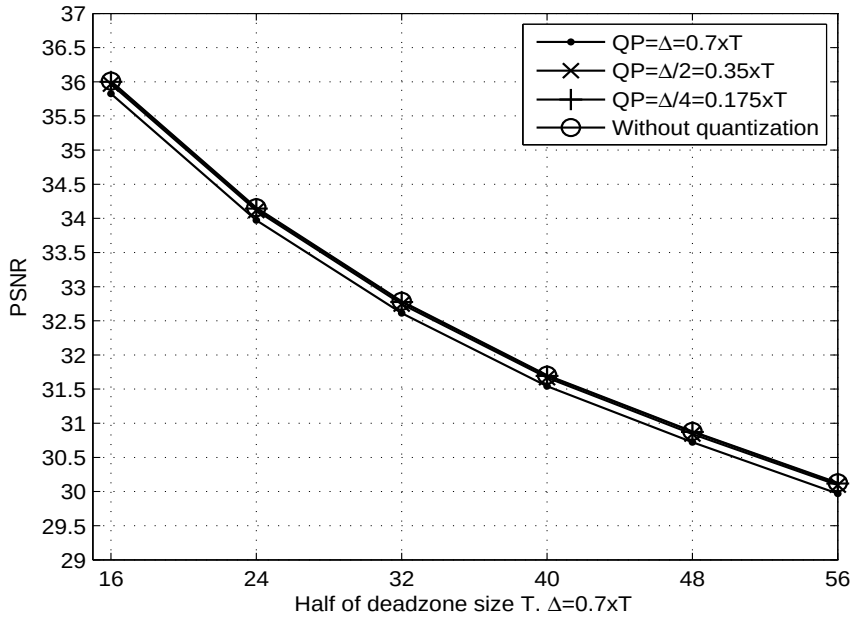


Figure 4.7: The effect of deadzone on PSNR in coding 'Lena' image.

coding will affect (or even dominate) the central-side PSNR performance. Fig. 4.7 simulates the effect of deadzone in a single description image coding. The 512×512 8-bit 'Lena' image is tested. In the figure, the wavelet coefficients inside the interval $[-T, T]$ are quantized to zero, which is a deadzone process. Then all the non-zero wavelet coefficients are quantized with different quantization parameters QP[ii]. The best achieved PSNR value is upper bounded by the PSNR values without quantization process as shown in the figure. It can be seen that the possible improvement in terms of PSNR is very limited (up to 0.3 dB) through refining no-zero coefficients. It is not surprising that most of the information of the image is lost due to the deadzone except for some coefficients with large magnitudes. Even allocating more bits to these large-magnitude coefficients still does not help improve performance significantly.

[ii] As shown above, for the single description QTCQ coding, $QP = \Delta = 0.7 \cdot T$ [152].

In two-stage MD image coding schemes, if the equivalent deadzone of Q_J is designed to be smaller than deadzones of quantizers Q_1 and Q_2, then the central PSNR at the first stage may be 0.5-1.5 dB better than the side PSNR values at the first stage, based on our simulations. It seems that a smaller deadzone of Q_J leads to a better utilization of bits allocated to the first stage, which further results in better central-side PSNR performance for the whole two-stage scheme. Specifically, in the SPIHT-based [112] and JPEG2000-based [113] schemes, Q_1 and Q_2 are the same as each other, which leads to a Q_J without any contribution to reducing the central distortion. Thus the deadzone of Q_J is equivalent to that of Q_1 or Q_2. Although the mismatch disappears (due to $Q_1 = Q_2$) when the refinement information (the residual image) is used to refine stream 1 or 2 at the second stage, the overall central-side PSNR performance of the scheme still drops, as shown in the following experimental part. In addition, taking the same Q_1 and Q_2 will result in a worse distortion product in the high-resolution analysis, since the correlation coefficient ρ will become 1. In the (MMDSQ+TCE)-based image coding scheme [17], a smaller equivalent deadzone for Q_J is obtained through staggering, and the overall performance is shown to be better than those of [112] and [113] in most cases. In Lin's scheme [18], although it is difficult to show an equivalent deadzone geometrically, an efficient first stage structure brings the best overall central-side PSNR performance. The equivalent deadzone of the proposed scheme is similar to the (MMDSQ+TCE)-based image coding scheme [17], which can be easily determined to be $2T - \Delta$.

4.5.2 Experimental Results

We compare the proposed two-stage image coder against several other two-stage MD image coders in this section. Specifically, we give the performance of modified MDSQ based on a Tarp filter image coder with classification for embedding (MMDSQ+TCE) [17] by Tian, the JPEG2000-based MD coder [113] by Tillo, SPIHT-based MD coder [112] by Miguel and TCQ-based MD coder [18] by Lin. All four methods (including ours) make use of the wavelet transform in their coding scheme. A six-level wavelet decomposition is applied with the Daubechies 9/7 filters and 8-state TCQ is used in the the proposed scheme. The tested bit rate is 0.5 bpp per description. The central and side PSNR performance for 'Lena' (of size 512×512) is shown in Fig. 4.8.

It can be seen from the figure that for all the two-stage schemes, the tradeoff points can be obtained conveniently and flexibly by changing the bit allocation in the two stages. In most schemes, the redundancy added can change from 5% to 50% compared with the single description image coding. JPEG2000-based and SPIHT-based coders can even achieve a redundancy as low as 1%. Furthermore, we can see that as side PSNR increases with more redundancy introduced, the central PSNR will drop dramatically to be close to the side PSNR for all the schemes. The proposed scheme achieves similar central-side distortion performance with low redundancy, compared with Lin's TCQ-based MD scheme [18], but getting worse as redundancy becomes high[iii]. There are some reasons for the performance deterioration of our coding. The first one is due to shifting of

[iii] To compare the coding performance of two schemes in Fig. 4.8, we can first fix the same central PSNR on the figure and then compare their side PSNR values. If a scheme can always achieve better side PSNRs for different central PSNRs, we say this scheme is better than the other. Meanwhile, a good MD image coding design should achieve flexible tradeoff between central and side distortions.

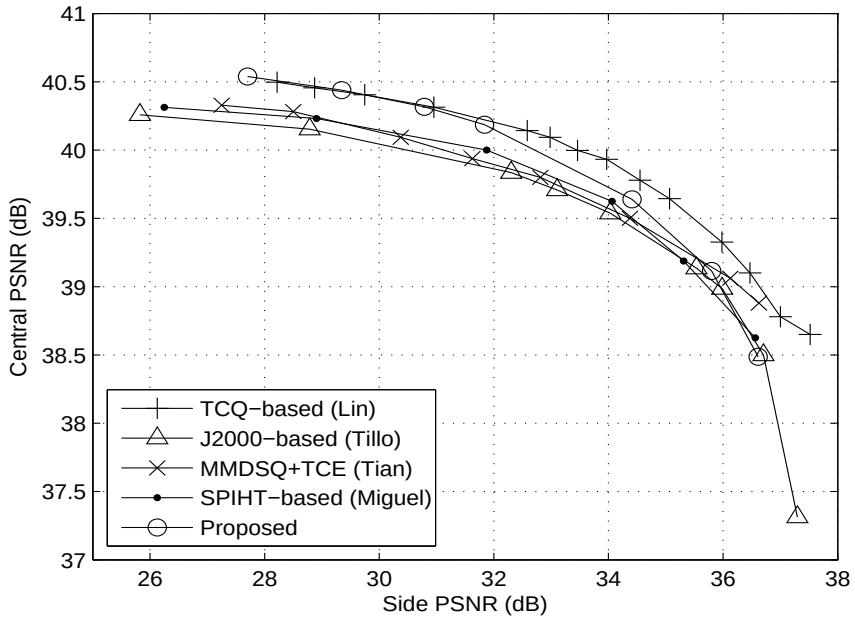


Figure 4.8: Comparison of central-side PSNR curves at 0.5 bpp/description for 'Lena'.

wavelet coefficients by a constant, which will lead to a slightly worse side PSNR than Lin's scheme. The second reason is that the equivalent deadzone of Lin's scheme leads to a better staggering gain for the central PSNR at the first stage. The proposed scheme outperforms other schemes in the case of low and medium redundancy. Since all the above two-stage MD image coders outperform most of state-of-the-art one-stage MD coders (e.g. [111], [95] and [142]), we do not present the results for these one-stage MD coders any more.

4.6 Conclusion

We have presented a scheme to make the product of central and side distortions for the two-stage MDSQ closer to the multiple description rate-distortion bound. The proposed scheme is to use the refinement information generated at the second

stage to reduce the side distortion, where an improvement of 0.58 dB coding gain in the side distortion has been shown compared with the two-stage MDSQ without refinement. With the scheme, a simple structure of two-stage MDSQ can achieve an asymptotic distortion product with a smaller gap from the rate-distortion bound, compared with other relevant MDSQ schemes. In addition, if a good lattice is used in the refinement stage instead of a scalar quantization for reducing central distortion, a better asymptotic distortion product with only 0.953 dB from the rate distortion bound can be obtained. Finally, we have proposed a practical two-stage multiple description image coding based on trellis coded quantization, which is comparable (or usually superior) to the state-of-the-art MD image coding schemes.

Chapter 5

Multiple Description Video Coding Based on Hierarchical B Pictures

5.1 Introduction

In contrast with the traditional single description coding which aims to achieve optimal rate-distortion performance, MDC design needs to consider both central and side distortions given a total bit rate. However, it is known that the minimization of central and side distortions conflicts with each other, and their trade-off is controlled by redundancy among different descriptions (termed as inter-description redundancy in this chapter). Generally speaking, more inter-description redundancy will favor side distortion but result in less improvement of central distortion over side distortion. Therefore MDC attempts to strike a balance between minimizing the central and side distortions for a given rate, which

requires a good design for the inter-description redundancy control in addition to minimizing the intra-description redundancy in each description.

The major advantage of pre-processing based MDC methods (reviewed in Chapter 2) is that the substantial encoding process is untouched, thus facilitating all the existing standard codecs of high efficiency to be conveniently applied in coding each description with little or no change. In multiple description video coding (MDVC), most algorithms are based on pre-processing schemes, e.g., subsampling in spatial or temporal domains [2, 19, 20]. These subsampling based MDVC algorithms take advantage of the assumption that spatially or temporally adjacent video data samples are highly correlated. Thus, one description can be estimated from the others. A typically simple and efficient pre-processing scheme for 2-description video coding is to split the odd and even pictures of a video sequence into two separate streams, each of which can be encoded using a certain coding scheme such as MPEG-1/2/4 or H.261/263/264. Some of such standard-compliant MDVC schemes were developed in [19] and [20] based on the similar temporal subsampling. An MDVC method based on spatial subsampling was developed in [2]. In view that the correlation among neighboring pixels is reduced after subsampling thus reducing the coding efficiency for each description, the authors in [2] introduced the DCT-domain zero padding technique before subsampling to alleviate the problem of correlation reduction due to subsampling. However, the performance of these temporal and spatial subsampling based MDVC algorithms highly depends on how good the estimation of missing description is. In our view, this kind of "hard" splitting of neighboring frames or pixels into different descriptions may not be a good way at high rate in terms of achieving intra-description redundancy minimization and good inter-description redundancy control. In [3], the authors developed slice-group based

MDVC (SGMDVC) based on H.264. Slice group is a coding tool provided in H.264. The picture is divided into slice groups and it can be further divided into slices in scan order. There are totally 7 types of macroblock allocation for slice group, i.e. slice group map types, where a "dispersed" slice group map partitions two slice groups A and B like a checkerboard. This kind of partition is helpful for error concealment. The basic idea of SGMDVC is that, in each description, only one slice group is finely encoded, and the other slice group is encoded coarsely to keep basic information. The performance of SGMDVC transmitted in the packet loss network can be found in [4].

There also exist some MDVC algorithms based on frequency domain subsampling. One example known as domain-based MDVC was proposed by Bajic and Woods [88]. In the domain-based MDVC, wavelet coefficients are partitioned into maximally separated sets to produce multiple bitstreams. At the decoder side, simple error concealment methods can be employed to produce good estimates of lost data. However, this scheme fails to well exploit the temporal correlation. Van der Schaar and Turaga developed the multiple description scalable video coding (MDSVC) scheme [128] based on the motion compensated temporal filtering (MCTF) [129], where high frequency frames are grouped into two descriptions and missing frames are estimated using motion vectors in the two descriptions. In [130], after preforming MCTF, 2-D spatial wavelet transform is employed to achieve spatial scalability and embedded bitplane coding is used to achieve SNR scalability. Different descriptions are produced by coding different codeblocks in different rates.

In this chapter, we employ H.264/AVC scalable extension of hierarchical B pictures [156] for a pre-processing based MDVC, where different temporal-level

key pictures are selected in the coding of different descriptions. In this way, different descriptions may contain varying amounts of information of each frame, which corresponds to inter-description redundancy. We then present a simple scheme to control the amount of redundancy. In view of varying and complementary representations of each frame in different descriptions, we propose a scheme to optimize the central decoder for better reconstruction. Consequently, a general MDVC framework can be developed with high coding efficiency. The one-description decoding is compatible with the H.264/AVC scalable extension decoder. The temporal scalability with the hierarchical structure for each description is also obtained.

This chapter is organized as follows. In Section 5.2, the structure of hierarchical B pictures is firstly reviewed. In Section 5.3 and Section 5.4, we analyze the 2-description and more-description video coding based on the hierarchical B pictures, respectively. Experimental results of the proposed MDVC compared with some other schemes are shown in Section 5.5. Finally, this chapter is concluded in Section 5.6.

5.2 A Brief Review of Hierarchical B Pictures

In the hierarchical B pictures based video coding, some pictures are regularly (or irregularly) selected from the original sequence as key pictures. A typical regular hierarchical prediction structure of 4 dyadic hierarchy levels [157] is depicted in Fig. 5.1, where the key pictures can be I pictures or P pictures. A key picture and all pictures that are temporally located between the current key picture and the previous key picture construct a group of pictures (GOP). The remaining

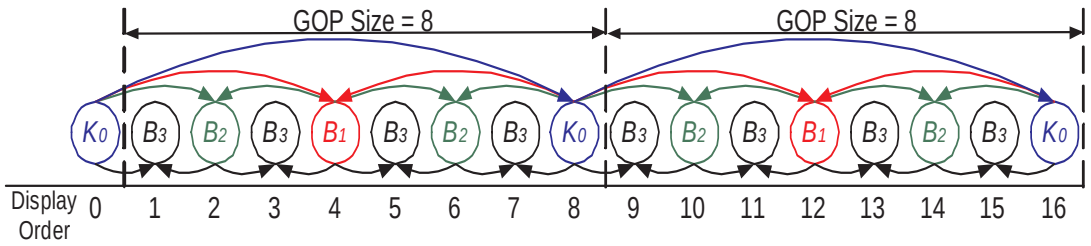


Figure 5.1: The hierarchical coding structure with 4 temporal levels of picture, K_0, B_1, B_2 and B_3. K_0 represents the key picture and B_1 to B_3 are the pictures at decreasing lower temporal level.

pictures in a GOP are hierarchically predicted and coded using the bi-predictive (B) slice syntax of H.264/AVC in the following way. Key pictures in the highest level (denoted as K_0, corresponding to level 0) are encoded with the highest fidelity which are used as references for motion-compensated prediction (MCP) of all other pictures. Hierarchical prediction structures can be used for supporting several levels of temporal scalability. It is required that only pictures of a higher or the same temporal level are employed as references for MCP. Then, the sequence of key pictures represents the lowest supported temporal resolution, and this temporal resolution can be refined by adding the temporal refinement pictures of lower temporal levels. The coding structure in the figure is "KBBBBBBBKB...". To simplify the notation in this chapter, we denote "KBBBBBBBKB..." as KB7 referring to 1 key picture followed by 7 B pictures corresponding to 4 levels.

5.3 Two-Description Video Coding with Hierarchical B Pictures

The structure of the hierarchical B pictures was proposed in the scalable extension of H.264/AVC to provide temporal scalability. However we find that it

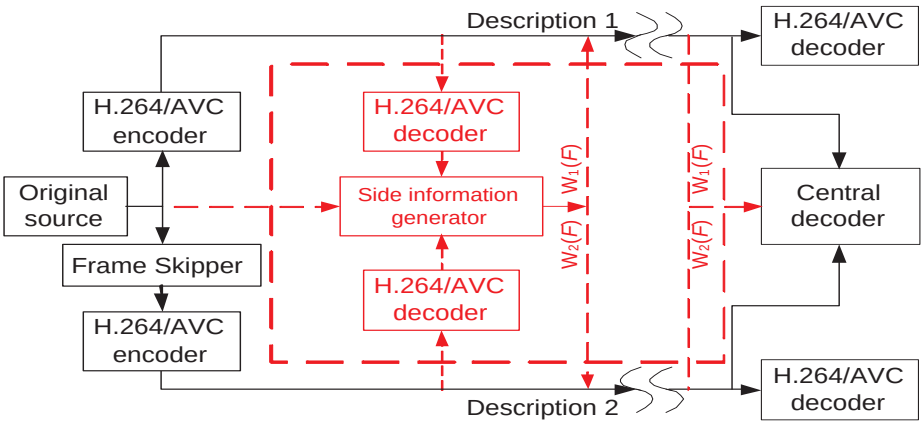


Figure 5.2: Block diagram of the proposed 2-description video coding. The dashed part (optional) in the figure is for central decoder optimization.

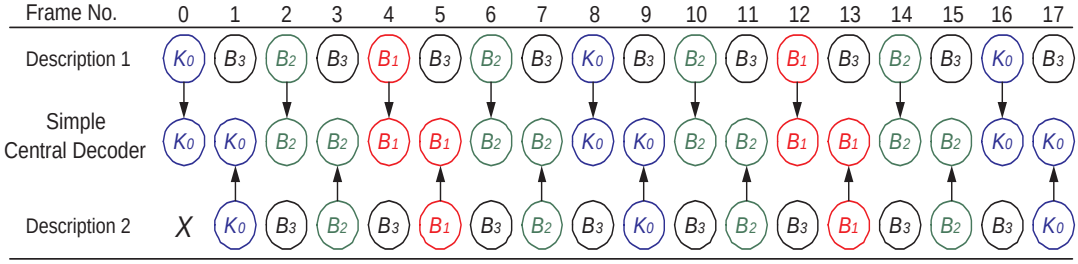


Figure 5.3: Hierarchical B Pictures based 2-description video coding (KB7 coding structure) with a simple central decoder. X means the frame is skipped.

can be employed to develop MDVC effectively. 2-description video coding using the hierarchical B pictures based H.264 codec can be obtained in this way. For different descriptions, different key pictures are selected and each description is encoded separately to produce a temporal scalable bitstream. That is, the difference between two descriptions lies in the selection of key pictures. Fig. 5.2 shows the block diagram of the proposed 2-description video coding. Different description can be achieved by skipping the first k frames of the video sequence. In the case of 2-description video coding, we choose $k = 1$. Fig. 5.3 shows an example of 2-description video coding with 4 dyadic hierarchy levels (of coding structure KB7) for each description. The first key picture in description 1 is frame 0, while that in description 2 is frame 1. In this way the key pictures

in the two descriptions are staggered. For example, in one GOP, description 1 includes frame 0 as K_0 frame at the highest level 0, frame 4 as B_1 frame at the second highest level 1, frame 2 and 6 as B_2 frames at second-lowest level 2 and frame 1, 3, 5, 7 as B_3 frames at the lowest level 3. In contrast, in description 2, frame 1 is encoded at the highest level 0, frame 5 at the second-highest level 1, frame 3, 7 at the second-lowest level 2 and 2, 4, 6, 8 at the lowest level 3. For the H.264/AVC codec in each description, the coding performance of the hierarchical prediction structure is dependent on the quantization parameters (QPs) at different levels, where a larger QP generally leads to a lower reconstruction fidelity. Thus our proposed 2-description video coding based on the staggered and hierarchical structure can make each frame have different fidelities in different descriptions, such that the descriptions are supplementary each other, while each can be decoded independently without any modifications to the H.264/AVC codec. It can be seen that the proposed structure is also a pre-processing based 2-description video coding in that the two streams are produced before the actual encoding is performed. In this structure, the two streams are generated by duplicating the original sequence and then coded by H.264/AVC scalable coding with staggered key pictures coding structure in the two streams. Alternatively, we can interpret the pre-processing as producing the two streams with one being the original sequence and the other being obtained by skipping the first frame in the original source. Then the following H.264/AVC encoding with the same key picture coding structure is applied to the two staggered sequences. In this way, two staggered sequences are also obtained. With the structure, we will consider in the following controlling the inter-description redundancy between the two descriptions for a balance between central and side distortion given a rate.

5.3.1 Inter-description Redundancy Analysis and Control

From above we know that each frame is represented in each description with different fidelity at different levels. The inter-description redundancy hence can be considered as inter-level redundancy between two different temporal levels. It is reasonable to treat the bit budget for the lower fidelity representation as the redundancy, which can be controlled by the quantization parameter and its motion information in each frame. The higher the inter-level redundancy, the better performance for side decoding results at the cost of performance loss of central decoding.

To strike a good tradeoff between central and side distortions, QPs for all the levels in both descriptions should be determined properly. The default QP selection strategy in the scalable extension of H.264/AVC [158] focuses on reconstructing each single description as good as possible at a given rate, which generally results in high inter-description redundancy. Therefore a more effective and flexible redundancy control scheme is desired to balance the central and side distortion at a given rate. To maintain a good rate-distortion performance, we consider only modifying the QP value at the lowest level for different amounts of redundancy. The QP value can be changed in the range from the default value to the maximum allowable value of 51. To further reduce the redundancy, we can totally discard residuals by only keeping motion vectors (MVs), or even skip some lowest-level pictures in the extreme case. Note that in the extreme case of skipping the pictures at the lowest level, the proposed 2-description video coding is degraded to an equivalence of even/odd frame splitting based 2-description video coding. We find that although MVs consume some bit budget, they are helpful in reconstructing good motion compensated frames, especially in the video se-

quences containing high motion. At a low bit rate, the bit cost for MVs cannot be neglected due to the limited rate budget and totally skipping the pictures seems to be reasonable. At a relatively high bit rate, the bit cost for MVs is trivial. Therefore a preferable scheme at a high rate for inter-level redundancy control is to increase quantization parameter from the high level to the low level, where the lowest-level B pictures are most coarsely quantized with only few quantized residua preserved with MVs kept. In this way, the trade-off between central and side distortion for a rate can be flexibly adjusted according to the application requirement, which is verified in our simulations in Section 5.5.

5.3.2 Central Decoder Design and Optimization

The simplest central decoder for the 2-description video coding can be obtained in such a way that a high-level fidelity is always selected for reconstructing a frame, as shown in Fig. 5.3. However, this selection completely discards the lower fidelity B pictures which may contain some useful information not in the higher fidelity frames. Consider the decoding of an $m \times n$ block c in the frame 4 in Fig. 5.3. Two reconstructions for the block c can be obtained from frame 0 and 8 in description 1, and frame 3 and 5 in description 2, respectively. Note that $\hat{c}_1$ has higher fidelity and $\hat{c}_2$ has lower fidelity. Although $\hat{c}_2$ is coded with lower fidelity, it is obtained from its most neighboring frames. Therefore, it is possible to combine $\hat{c}_1$ and $\hat{c}_2$ to generate a better reconstruction for block c. For simplicity, consider a linear combination for block c

$$\hat{c} = w_1\hat{c}_1 + w_2\hat{c}_2 \quad \text{with} \quad w_1 + w_2 = 1 \tag{5.3.1}$$

where w_1 and w_2 are the weights for $\hat{c}_1$ and $\hat{c}_2$ respectively. Intuitively, with this combined technique, the reconstructed frame is obtained from 4 reference pictures, which may more likely give better reconstruction quality. Furthermore, from the point of estimation theory, the optimal linear combination of blocks is always better (at least no worse) than using any individual block to estimate another block. In the extreme case with $w_1 = 1$ or $w_2 = 1$, the estimator becomes one individual block.

The central distortion for the block c using the combination can be calculated by

$$d_0 = \|c - w_1\hat{c}_1 - w_2\hat{c}_2\|^2 \tag{5.3.2}$$

The optimal weight can be obtained by minimizing d_0 with respect to w_1 or w_2 ($w_1 + w_2 = 1$). Then we have

$$\frac{\partial d_0}{\partial w_1} = \frac{\partial \|c - w_1\hat{c}_1 - w_2\hat{c}_2\|^2}{\partial w_1} = 0, \tag{5.3.3}$$

where c, $\hat{c}_1$ and $\hat{c}_2$ are $mn \times 1$ vectors rearranged from $m \times n$ blocks. We can obtain that the optimal weight

$$w_1 = \frac{(\hat{c}_1 - \hat{c}_2)^T(c - \hat{c}_2)}{\|\hat{c}_1 - \hat{c}_2\|^2}. \tag{5.3.4}$$

Note that the weights w_1 and w_2 are obtained in the encoder side and then transmitted to the decoder as side information. In practice, to reduce the bit cost on these weights, we consider quantizing the weights with a quantization step T. These weights can be further entropy coded to save more bits. Our experimental results show that quantizing the weights by $T = 1/8$ produces reconstruction

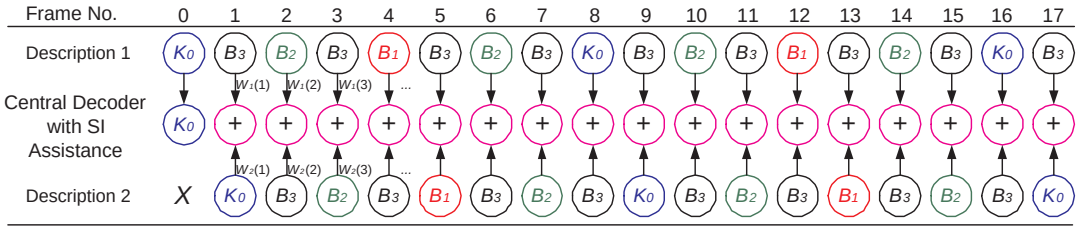


Figure 5.4: Hierarchical B Pictures based 2-description video coding (KB7 coding structure) with the optimized central decoding based on linear combination.

distortions very close to that based on the optimal w_1 without quantization, while yielding a small amount of overhead. For example, for the CIF sequence "Football" (30 fps), the total overhead for coding the block-based side information ranges from 26.8 kbps to 33.3 kbps over the total bit rate from 940 kbps to 2.9 Mbps, when $m = n = 16$ and $T = 1/8$, while the central PSNR value increases significantly, up to 0.75 dB, which will be shown in the experimental section. Note that the weights for each frame are compressed by arithmetic coding separately. It can be seen that the side information only account for 2.85% in the total bit rate of 940 kbps while the percentage further drops to 1.1% in the total rate of 2.9 Mbps. The optimized central decoding using linear combination is shown in Fig. 5.4.

5.4 More-Description Video Coding Using Hierarchical B Pictures

The proposed 2-description video coding scheme can also be easily extended to more than 2 descriptions, where the key pictures in different descriptions are staggered in a way similar to in the 2-description case. As an example, a 4-description video coding will be considered. The problem of high inter-description

redundancy is more noticeable in more-description video coding, compared with 2-description video coding.

5.4.1 Four-description Video Coding

Fig. 5.5 shows the structure for 4-description video coding, where a GOP size of 8 (corresponding to KB7 coding structure) is adopted. Note that GOP size is always selected to be no less than the total number of descriptions to avoid producing same descriptions based on the staggered structure. It can be seen that from frame 3 in the four descriptions, there are always two B_3 representations for each frame. The representations at the same level are similar, corresponding to a high level of repetition close to duplication.

We introduce a term of "similar description" to facilitate the following discussion. Two descriptions are considered to be similar for a level if they share the same frames coded at the level. First, we only consider all the B_3 representations in the 4 descriptions in Fig. 5.5, which are more explicitly and neatly shown in Fig. 5.6. It can be seen that from frame 3 onwards there are two similar-description sets, where the two odd-numbered descriptions are in set 1 and the other two even-numbered descriptions in set 2. Similarly, if temporal level 2 is considered, 4 similar-description sets can be formed, with each set containing only one description.

The redundancy control method of coarsely quantizing and coding the lower level in 2-description video coding may not be effective in reducing the high redundancy in more-description case. Therefore we consider a redundancy control strategy based on temporal-level skipping. A simple example is to skip all the frames at the lowest temporal level. That is, all the B_3 frames at temporal level

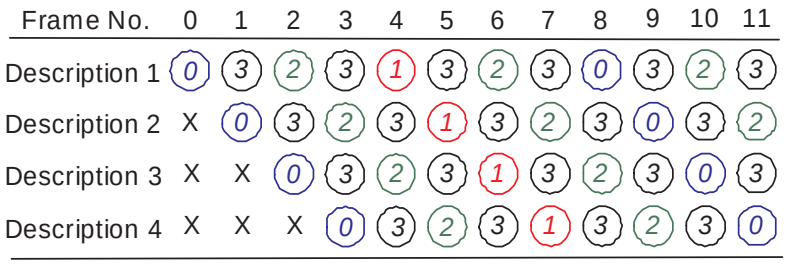


Figure 5.5: Encoding structure of the proposed 4-description video coding with KB7 coding structure. K_0 and B_1 to B_3 are represented by their subscripts for simplicity.

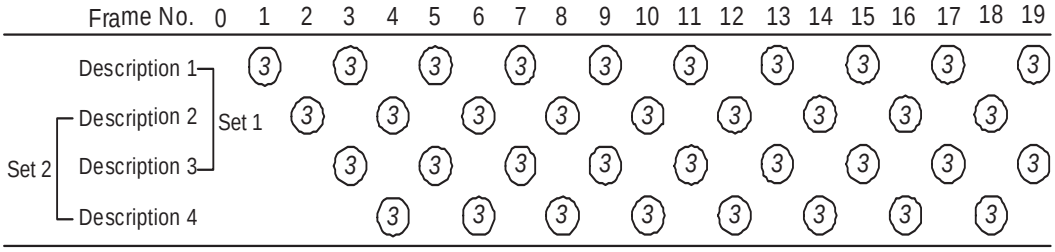


Figure 5.6: Similar descriptions for temporal level 3 in the proposed 4-description video coding.

3 are skipped in Fig. 5.5. We know that there are 2 similar-description sets for the level 3. At the decoder, if m descriptions are received all from the same similar-description set, we need to interpolate the skipped frames at the decoder side.

Assume two descriptions are received in the following discussion. For instance, if descriptions 2 and 4 are received which are from the same set, all the even-numbered frames are skipped and need to be interpolated or concealed by the received odd frames. In contrast, if the two received descriptions are from 2 similar-description sets (e.g. descriptions 1 and 2), it can be seen from Fig. 5.5, that the combination of received descriptions contains different levels of representation for all the frames, thus requiring no frame interpolation. This is desirable without the uncertainty of frame interpolation result. Generally, frame

Table 5.1: Possibility requiring frame interpolation or not in the case of skipping temporal level 3 as m (≥ 2) descriptions are received in 4-description video coding.

	Interpolation required	Interpolation not required
k	$2C_2^0C_2^k/C_4^k$	$(C_4^k - 2C_2^0C_2^k)/C_4^k$
2	2/6	4/6
≥ 3	0	1

interpolation techniques perform well for the sequences containing low or regular motion, but poorly for the sequences with high or irregular motion. More computations are also required in the decoder side for frame interpolation. Therefore, the 4-description video coding design needs to consider striking a good balance between reducing redundancy and lowering the possibility of frame interpolation. Table 5.1 shows the possibility of frame interpolation for different k (≥ 2) values when all the frames at temporal level 3 are skipped in 4-description video coding. It can be seen that no frame interpolation is required as k increases to 3 and above.

We can also see that for $k = 2$, skipping all the level 3 frames may produce unbalanced k-description reconstruction results depending on which k descriptions are received and whether frame interpolation is required or not. The reason for the unbalance is that skipping all the level 3 frames will make the descriptions in the same similar-description set suffer by losing the same frames at level 3, thus demanding frame interpolation if only descriptions from the same set are received. Therefore, to achieve a balanced k-description reconstruction in terms of the same number of frames requiring interpolation (or no interpolation), we need to wisely remove some selected frames rather than indiscriminatingly skipping all the level 3 frames. Two examples to skip part of the level 3 frames for designing a balanced 4-description video coding are shown in Fig. 5.7. The first way is to skip the frames which are crossed out as "1st" periodically, produc-

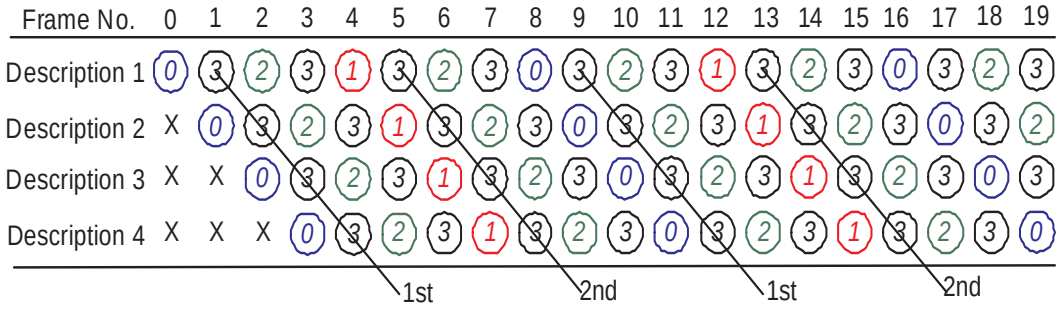


Figure 5.7: Balanced 4-description video coding designs: two examples shown by skipping the frames progressively according to the skipping orders (1st and 2nd) labeled in the figure.

ing balanced k-description reconstruction results for any $k \geq 1$. If the frames crossed out as "2nd" are further removed, the second way can be obtained to yield balanced k-description reconstruction results for $k \geq 1$. From the figure we can see that quarter and half of level 3 frames are removed in the two balanced design examples respectively, resulting in more redundancy compared with skipping all the level 3 frames. There are also other frame skipping variants for the 4-description video coding design, such as skipping even higher levels of frames. Note that the frame skipping is normally done from lower to higher temporal levels without affecting the decoding of other remained frames. We would like to highlight that we can design a 4-description video coding based on different (conflicting) requirements and priorities, where reducing redundancy and maintaining k-description based reconstruction balance are conflicting. Design complexity is another consideration. Taking these factors into consideration, we think skipping all the level 3 frames is a good scheme in terms of reducing redundancy as well as design simplicity, which is verified in the experimental section, although it may lead to unbalanced k-description-based reconstruction for $k = 2$.

Remark 1: Multiple description coding normally requires each description to be individually decodable with good enough (at least acceptable) reconstruction

quality. When considering M-description coding for $M \geq 4$, the possibility of all descriptions being lost except one is very low. For M-description video coding, if we suppose the loss probability for each description is p_l, then the probability of all but one description being lost is $M(1-p_l)p_l^{M-1}$. The probability is no more than 2.56% for $K = 4$ and $p_l \leq 20\%$. In this sense, the requirement of reconstruction quality for each description being individually good may be inefficient for the MDC design, in view of the very small possibility introduced and the much more inter-description redundancy. Therefore, it is reasonable to consider at least a small number of k descriptions out of the total M descriptions being received when M is large, e.g. $M = 4$ or more. In this chapter we take $k \geq 2$ (i.e., at least 2 descriptions received) in studying the 4-description video coding.

Remark 2: 8DVC or even more-description video coding can be constructed in a similar way as that for 4-description video coding. However, there would be higher redundancy in the more descriptions, which makes 8DVC or more-description video coding much less practical. Due to page limit, we will not discuss it here.

5.4.2 k-description Based Decoder Optimization

We give the general expression of the decoder of k out of M descriptions being received ($M \geq k \geq 2$) as

$$\hat{c} = \sum_{t=1}^{k} w_t \hat{c}_t \text{ with } \sum_{t=1}^{k} w_t = 1 \tag{5.4.5}$$

These k weights for the k representations of block c are determined at the

encoder by

$$\{w_1, \ldots, w_k\} = \arg \min_{w_1,\ldots,w_k} \{\|c - \hat{c}\|^2\} \tag{5.4.6}$$

In view that a large number of weights (totally C_M^k groups of weights) need to be calculated and transmitted, the computational complexity and bit consumption will be high. The optimal weights may also be calculated by taking the partial derivatives with respect to w_i and solving the obtained equations with k unknowns. In the experiment, we use numerical method to estimate w_i. To simplify the numerical calculation of optimal weights in (5.4.6) as well as reduce the bit budget for these weights, we assume frame-based weighting (rather than block-based), with $0 \leq w_t \leq 1$. If we further consider that the weights for those representations at the same level are equal to each other, then we can further reduce bit consumption for these frame-based weights. Finally, the generated weights are quantized to 32 levels and entropy coded to save more bits. In this way computational complexity and bit consumption for the weights are significantly reduced.

5.5 Experimental Results

The extension of H.264/AVC reference software, JSVM7 [158], was used for coding each description in our proposed MDVC scheme. The mode of hierarchical B pictures was selected in the reference software. In the following we only encoded the first key frame as I frame and the following key frames as P frames,unless otherwise specified.

5.5.1 Results of Two-description Video Coding

On-off Channel Scenario

In this scenario, we assume that an individual description will be either error-free or totally lost. We consider two settings of quantization parameter (QP) for the proposed scheme. The first is to use the default QPs selection in H26.4/AVC for encoding each description. In the second setting we modify the QP for the lowest level pictures as the maximum allowable value 51 while remaining the default QPs for all the other levels.

We firstly compare the proposed 2-description video coding against zero padding based 2-description video coding in [2]. Fig. 5.8 compares the rate-distortion performance for the video sequence "Paris" (CIF, 30fps) by using the two methods. In the proposed 2-description video coding, the KB1 and KB7 coding structures are applied by using default QP setting. From the figure it can be seen that the proposed 2-description video coding method outperforms Wang's method in [2] remarkably. The reason behind may lie in several aspects. The first is that our 2-description video coding employs B picture in the coding while the 2-description video coding method in [2] is based on H.264/AVC codec with "IPPP..." coding structure. The other reason for the poorer performance of Wang's method in [2] is due to its frame splitting scheme. Frame splitting in spatial domain results in frame resolution reduction, which in turn decreases the temporal correlation in each description, thus compromising the coding efficiency. This point of view can be testified by better performance of the curve using "IPPP..." coding structure without B pictures than the curve by [2]. Our 2-description video coding does not change the temporal correlation for coding each description and achieves better rate-distortion performance. Thirdly, due to the staggered coding

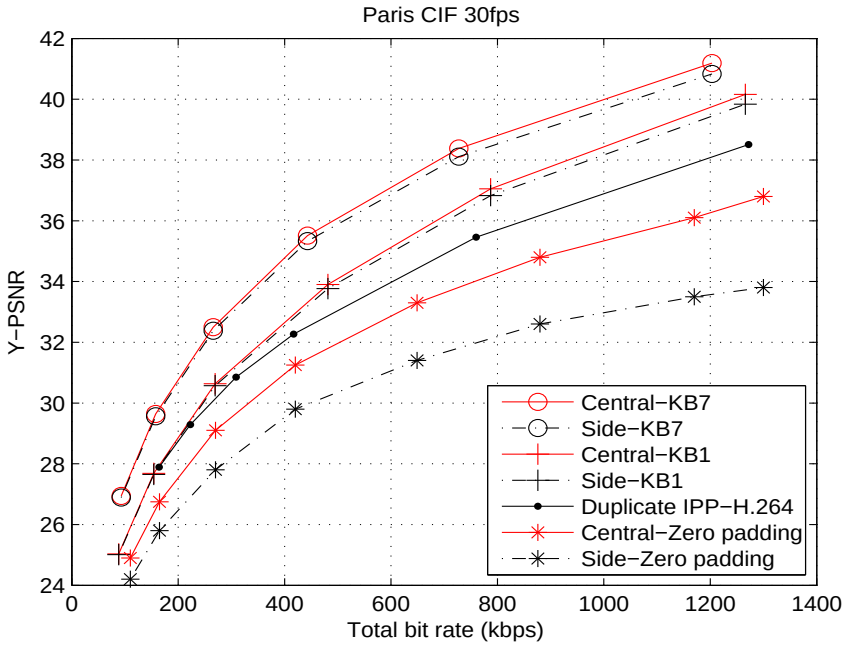


Figure 5.8: Comparison of the proposed 2-description video coding against Wang's 2-description video coding based on zero padding in [2].

structure to reduce redundancy, the central reconstruction is with smaller distortion illustrated in the figure, even though the central decoding optimization is not applied. We also compare the proposed 2-description video coding against the slice group based 2-description video coding (SGMDVC) in [3]. H.264/AVC codec with "IPPP..." coding structure enabling the slice group mode is used for SGMDVC. The results for "Paris" are shown in Fig. 5.9. Since employing B picture in the coding, the proposed 2-description video coding with KB1 coding structure outperforms SGMDVC. In addition, it is the use of the slice group in SGMDVC, which impairs the compression efficiency at the two side encoders more or less, although three MD loops have been applied. Note that SGMDVC is flexible for redundancy control and can achieve much low redundancy (less than 10% as shown in [3]).

In our proposed 2-description video coding scheme, different coding structures

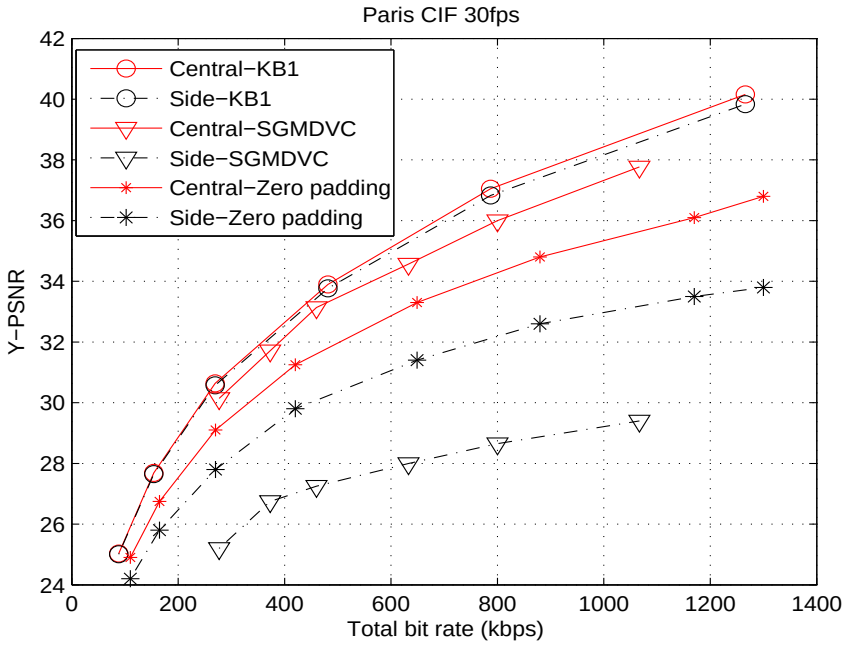


Figure 5.9: Comparison of the proposed 2-description video coding against Wang's slice group based 2-description video coding (SGMDVC) in [3].

may lead to varying rate-distortion performance. We compare several coding structures with different GOP sizes, that is KB1, KB3 and KB7. We would like to point out that, the odd/even frame splitting based 2-description video coding in [20], where the MV information for all the even/odd frames are also included in each description, is equivalent to our KB1 coding structure with MVs of the B pictures being kept. Fig. 5.10 depicts the rate-distortion curves for the sequence "Foreman" (QCIF, 30fps), where default QP setting and the simple central decoder shown in Fig. 5.3 are applied. We can see that a larger GOP size tends to achieve better central and side reconstruction qualities over different rates.

Next we compare the two QP settings, the default QP setting and the modified setting, for the tradeoff between central and side distortion. The first 102 pictures of the "Football" sequence are encoded. Fig. 5.11 demonstrates the

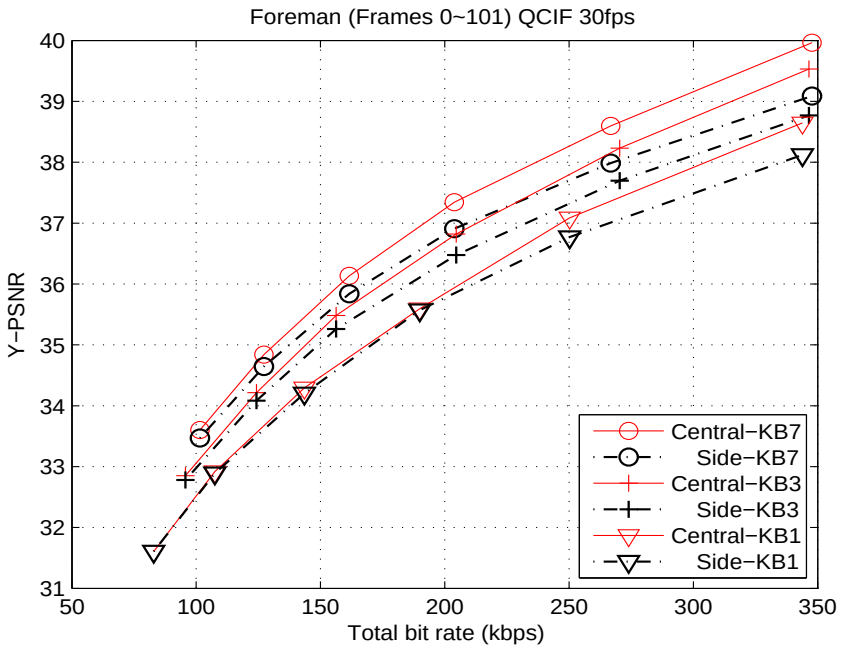


Figure 5.10: Comparison of different coding structures in the proposed 2-description video coding with default QP setting.

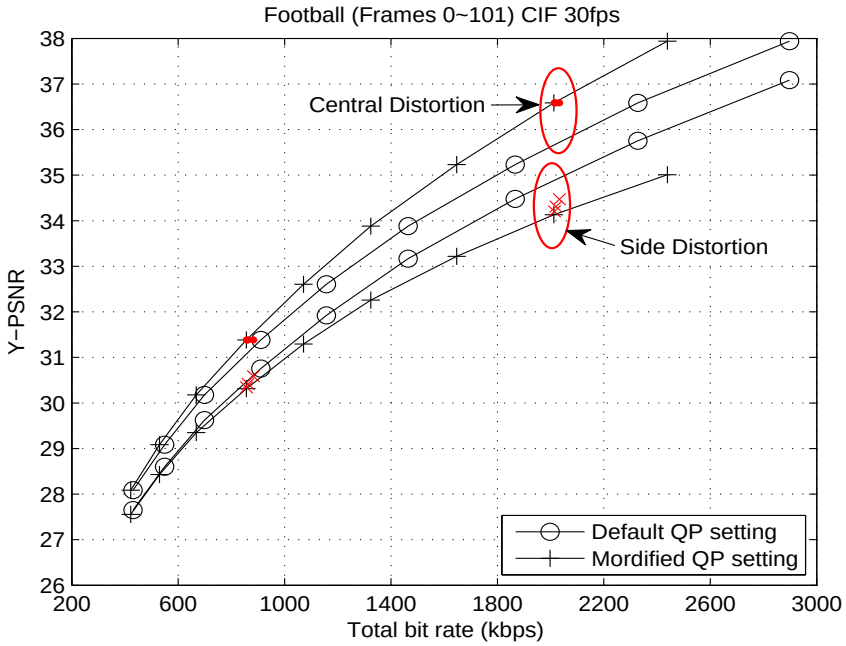
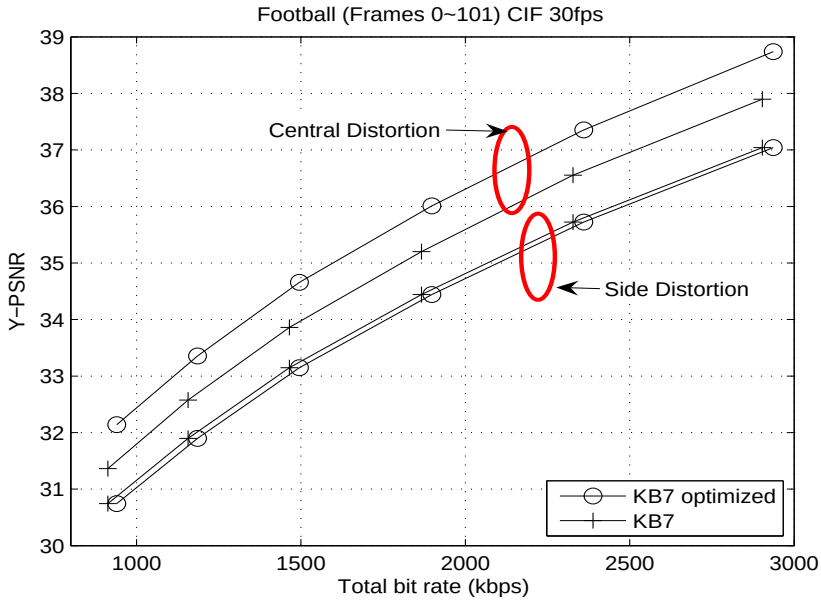


Figure 5.11: The proposed 2-description video coding rate-distortion comparison with different QP settings using KB7 coding structure.

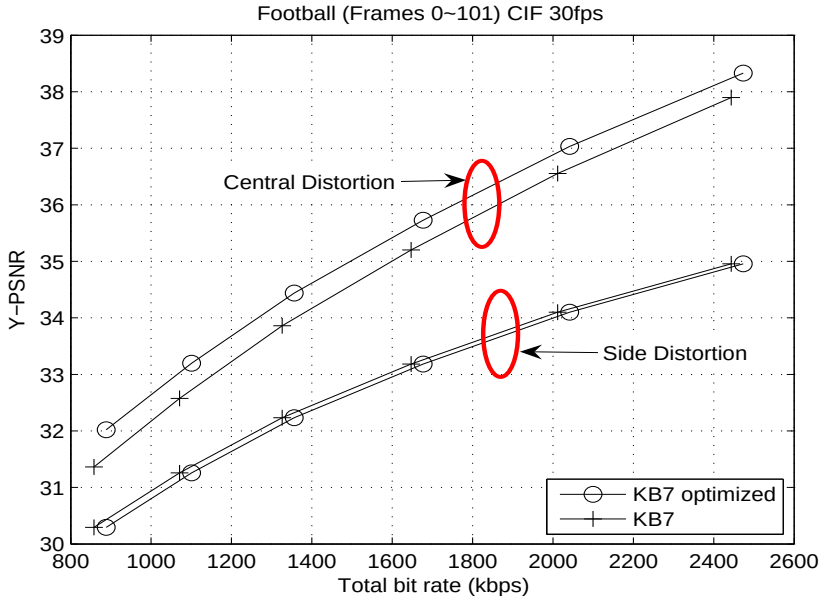
rate-distortion performance for the central and side decoders based on the two settings. It can be seen that with the modified QP setting the central distortion is improved at the cost of side distortion, which means that the tradeoff between central and side distortion can be achieved by modifying the QP parameters to adjust the inter-description redundancy. For those QP values between the two settings, the rate-distortion results and redundancy are also in between. Please refer to the separate data points (marked by '·' for central distortions and 'x' for side distortions) in the figure, which are obtained by changing QP for the lowest level pictures to 48, 45 and 42, respectively. We can see that all the separate points lie between the curves of default QP setting and modified QP setting.

Then we test the proposed central decoder optimization scheme discussed in Subsection 5.3.2. The first 102 pictures of the sequence "Football" are encoded with and without central decoding optimization, respectively. The weights for 16×16 blocks are compressed with arithmetic coding. Fig. 5.12(a) shows the rate-distortion performance with default QP setting. Up to 0.75 dB improvement in PSNR is achieved at the central decoder with the assistance of the optimized linear estimation in the case of KB7 coding structure, while the degraded side-distortion-rate performance due to the side information is almost negligible. Fig. 5.12(b) shows the similar results (but reduced improvement) when the modified QP setting is applied. The reason for the improvement reduction is that the modified QP setting provides poor picture quality at the lowest level, which in turn degrades the linear combination results. Therefore, we can see that the optimized central decoding can always achieve better performance more or less with the linear estimation.

Finally, we discuss the bit rate redundancy for the "Football" sequence coded



(a) Default QP setting



(b) Modified QP setting

Figure 5.12: Performance comparison of the proposed 2-description video coding with and w/o central decoding optimization.

by the proposed 2-description video coding with KB7 coding structure. Fig. 5.13 shows the performance comparison of the proposed 2-description video coding

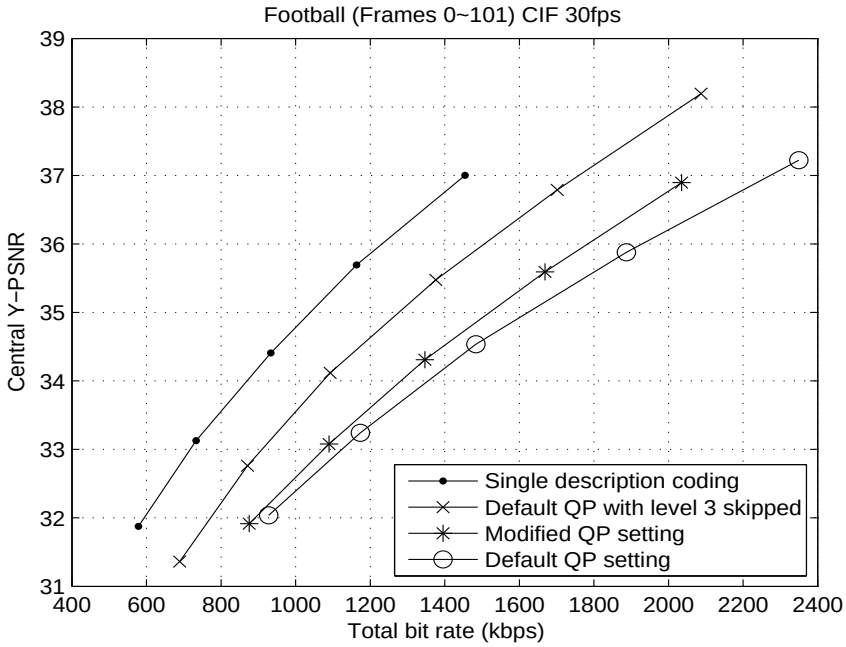


Figure 5.13: Comparison of the proposed 2-description video coding with the single description coding. Central decoding optimization is considered.

Table 5.2: Bit rate redundancy ($\rho(d)$) in 2-description video coding (Football)

	Default QP		Modified QP		Default QP (level 3 skipped)	
$\text{PSNR}_{central}$	$\rho(d)$	PSNR_{side}	$\rho(d)$	PSNR_{side}	$\rho(d)$	PSNR_{side}
33 dB	56.9%	31.67 dB	49.7%	31.19 dB	26.8%	27.34 dB
35 dB	56.4%	33.59 dB	46.2%	32.74 dB	22.9%	28.31 dB
37 dB	56.4%	35.30 dB	41.8%	34.18 dB	21.0%	29.35 dB

with different settings against the single description coding, from which we can estimate the bit redundancy for a certain central PSNR. Table 5.2 shows the estimated redundancy for different QP settings, in which the redundancy $\rho(d)$ for 2-description video coding is defined as a function of central distortion d by

$$\rho(d) = \frac{R_{2D}(d) - R_{single}(d)}{R_{single}(d)} \cdot 100\%, \tag{5.5.7}$$

where $R_{2D}(d)$ is the total bit rate for achieving central distortion d in 2-description video coding and $R_{single}(d)$ is the rate for single description coding with the same

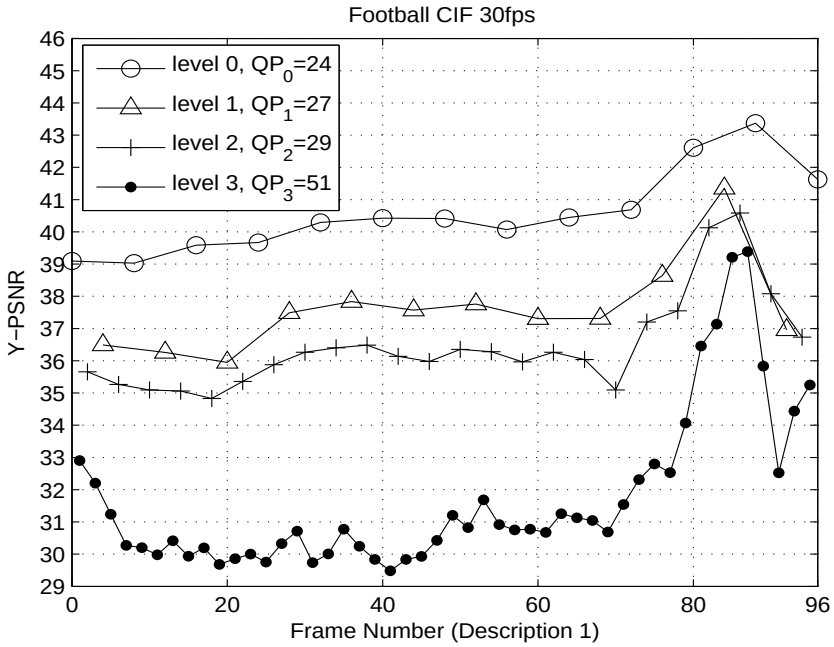


Figure 5.14: Frame-based PSNR for the sequence "Football" by using modified QP setting with $QP_0 = 24$. The default $QP_3 = 30$.

coding structure to achieve the same reconstruction quality of distortion d. From the table, we can find that: (1) based on the default QP setting inter-description redundancy is the highest with good side distortion; (2) with the default QP but skipping level 3, the redundancy is the lowest at the cost of the side distortion; (3) using the modified QP can achieve a good trade off between the above two cases. One can change the QP for level 3 by another value (instead of using 51), to achieve different amounts of redundancy. However, one problem of using the modified QP setting is the quality fluctuation in each description, which may further result in low subjective quality. Fig. 5.14 gives frame-based Y-PSNR for the sequence in description 1. It can be seen that the average reconstructed PSNR values of pictures are descending from the highest level to the lowest level due to the increasing QPs from the highest level to the lowest level. The value of QP_3 in the modified QP setting is 51. The reconstructed video in each descrip-

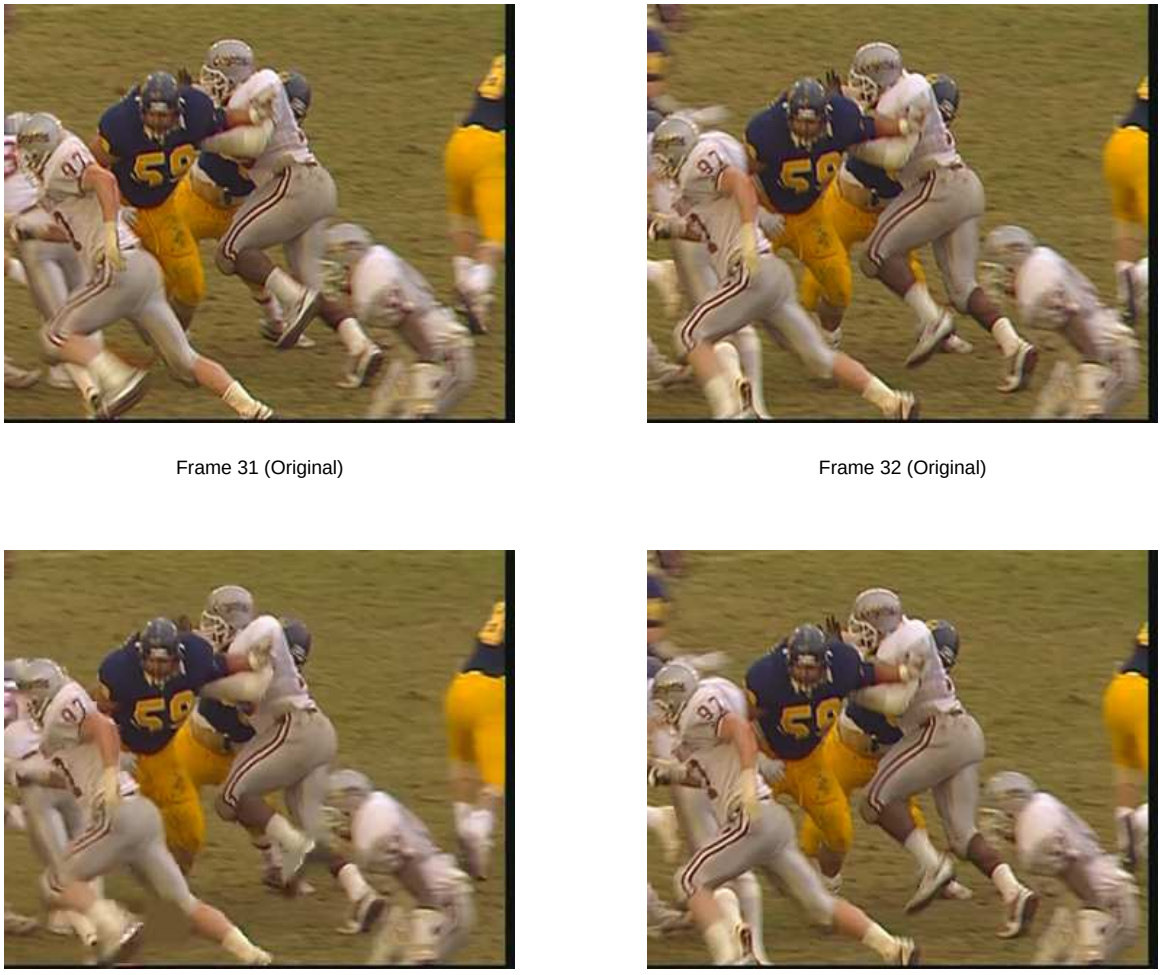


Figure 5.15: Visual comparison of frame 31 and 32 of the sequence "Football".

tion appears to be temporally smooth without much annoying temporal pumping artifacts. In Fig. 5.15, we show the visual qualities of two neighboring frames, frame 31 from level 3 and frame 32 from level 0, which correspond to the maximum PSNR difference of around 10 dB in our experiments. It can be seen from the pictures that the subjective quality appears to be acceptable.

Packet Loss Scenario

In this scenario, we consider the multiple description video transmission over packet-loss networks. Both descriptions are received at the decoder side, but each may contain errors.

We first compare the proposed 2-description video coding against the SG-MDVC method [4], which studied the performance of SGMDVC (proposed in [3]) transmitted over the packet-loss channels. The video sequence "Paris" (CIF, 30fps) is used for testing. To make a fair comparison, our experimental setting is similar to the setting in [4]. The video coding standard H.264/AVC is used as the basic coder. Packets are generated with 22 macro-blocks (MB) per slice, i.e., per packet. Simulations are run 50 times for each packet loss probability value. Fixed frame rate (30 fps) and "IBPBP..." coding structure with the group of size 16 are used in our scheme with default QP settings. Note that in [3], the coding parameters such as the redundancy and the coding rate are not specified. To achieve the same central PSNR of 37.5 dB without loss as in [4], we tested different QP values and the final total bit rate in our scheme is around 1500 kbps. According to [3], this bit rate value is comparable with the case of adding 40% redundancy in SGMDVC. In [4], three sets of encoders and decoders are communicating with each other in real time through windows programming at the decoder side, and the error concealment strategy based on slice groups are used. In our scheme, we do not use the slice group mode and the decoding strategy is a little different from that in [4]. At the side decoder, each description is first concealed by using the default error concealment method in H.264/AVC. Central reconstruction will proceed after side decoding. At the central decoder, there are several scenarios to reconstruct the current MB. In the case of both MBs from side decoders being

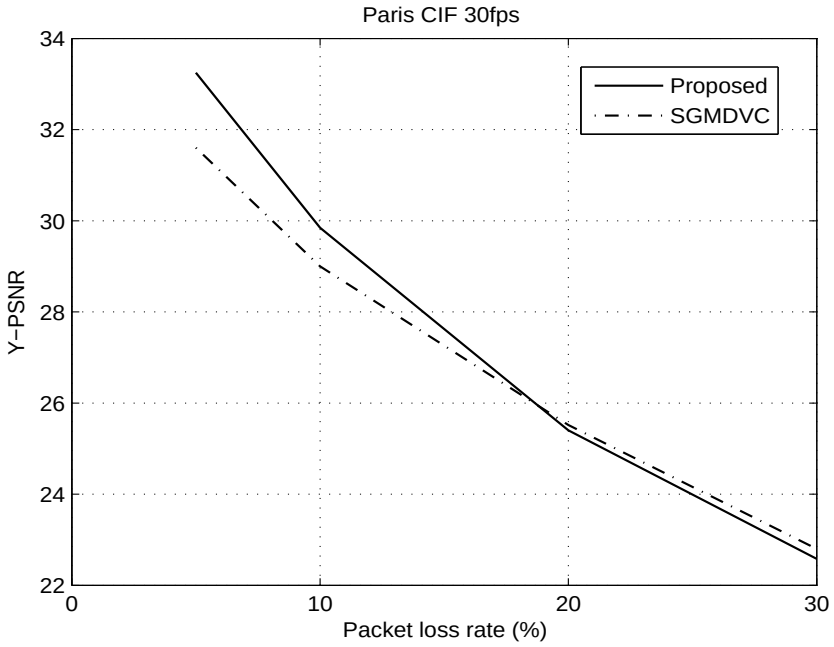


Figure 5.16: Comparison of the proposed 2-description video coding and SGMDVC in [4] for various packet loss probabilities. Y-PSNR of "Paris" without loss: 37.5dB for both schemes.

damaged, we just average the two repaired reconstruction blocks. If two MBs from side decoders are correctly reconstructed, we select the one with a smaller quantization parameter as the final reconstruction. If only one side reconstruction is impaired, we select the other one for decoding. The reconstruction quality in terms of PSNR versus packet loss probability for both schemes is shown in Fig. 5.16. It can be seen from the figure that our proposed scheme achieves better reconstruction quality than SGMDVC at small packet loss probabilities, but getting worse when the packet loss probability is larger than 20%. The possible reason is that the slice-group based error concealment works and brings better results with a larger packet loss probability. And our simple decoding strategy is very simple and may be improved to alleviate error propagations with a large packet loss.

It is interesting to evaluate the proposed coding performance in the packet loss network for a given packet loss probability, which requires to determine an appropriate trad-off between central and side distortions by minimizing the expected average distortion. An analysis on this issue has been shown in [5] by Tillo et al for redundant slice based 2-description video coding. In Appendix C, we show that the bit allocation for optimizing the expected average distortion in our method is similar to Tillo's method in [5]. Therefore we follow the bit allocation strategy developed in [5]. Now we compare the proposed 2-description video coding against the method in [5] considering the packet loss network with the packet loss probability being 5%[i]. The video sequence "Foreman" (QCIF, 15fps) is used for testing. In our scheme, the video coding standard H.264/AVC is used as the basic coder. Packets are generated with 33 macro-blocks (MB) per slice, i.e., per packet. Simulations are run 30 times for the packet loss probability 5% by using the rapid transport protocol (RTP) [159] simulator in H.264, which can be used to simulate packet loss on RTP files. Fixed frame rate (15 fps) and "IBPBP..." coding structure with the group of size 30 are used in our scheme with three different QP settings, which are the default QP setting, the modified QP setting and the optimized QP setting. The above experimental settings are similar to those in [5] except that the "IPPP..." coding structure with group of size 50 and the Bernoulli model for packet loss pattern are used in [5]. Fig. 5.17 shows the average Y-PSNR versus total bit rate by the proposed scheme and the scheme in [5]. From the figure we can see that the default QP and modified QP settings are not the best solutions for packet loss scenarios in our scheme. To minimize the expected distortion (here is the average Y-PSNR of the video sequence) for a particular packet loss probability, the optimal bit allocation strategy

[i]The optimal bit allocation scheme becomes accurate with a small packet loss probability p_l. In [5] only the case of $p_l \leq 10\%$ is considered.

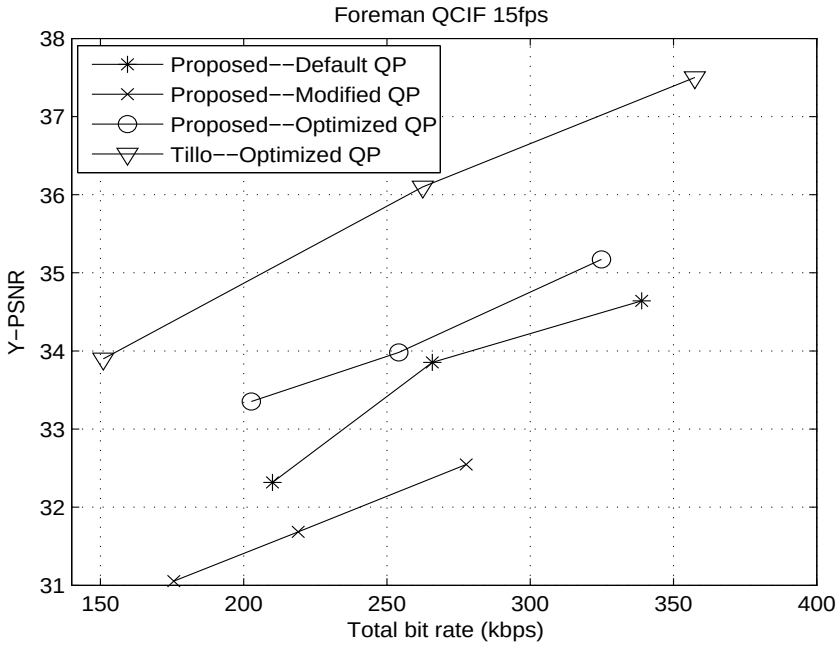


Figure 5.17: Expected distortion comparison of the proposed 2-description video coding and Tillo's method in [4] with the packet loss probability being 5%.

can efficiently allocate the redundancy to achieve a better expected distortion. Based on the same bit allocation method, our scheme is 1.5-2 dB worse than the method in [5]. The main reason is that we use a very simple decoding strategy as mentioned in the previous experiment. In this decoding strategy, two side coders are not communicated with each other and B slices from description 2 (or 1) are not used for error concealment in description 1 (or 2), which may lead to worse error propagations. We would like to improve our proposed decoder in our future work.

5.5.2 Results of Four-description Video Coding

In this part, we mainly check the efficacy of the temporal-level-based frame skipping scheme in our proposed 4-description video coding framework. As discussed

in Subsection 5.4.1, three ways of frame skipping in 4-description video coding are shown, i.e., skipping all the level 3 frames, skipping half and quarter of such frames. We can expect rationally that skipping more level 3 frames tends to eliminate more redundancy. We compare two cases in the proposed 4-description video coding framework, i.e., skipping versus keeping all the B_3 frames at temporal level 3 in each description, in terms of side distortion-rate performance. In both cases, frame-based weighting is used and transmitted as side information for decoding optimization. The sequence "Football" (CIF, 30fps) is tested with KB7 coding structure.

We assume $\text{PSNR}_{i,j}$ as the PSNR value obtained by jointly decoding descriptions i and j, and $\text{PSNR}_{4D}^{(2)}$ is the average 2-description based side PSNR value, which can be calculated by

$$\text{PSNR}_{4D}^{(2)} = \sum_{i=1}^{3} \sum_{j=i+1}^{4} \text{PSNR}_{i,j}/6 \tag{5.5.8}$$

It is reasonable to use the average for represent 2-description side PSNR when each $\text{PSNR}_{i,j}$ value is comparable. Similarly, the $\text{PSNR}_{i,j,k}$ is the PSNR value obtained by jointly decoding descriptions i j and k, and the average 3-description based side PSNR can be obtained as

$$\text{PSNR}_{4D}^{(3)} = (\text{PSNR}_{1,2,3} + \text{PSNR}_{1,2,4} + \text{PSNR}_{1,3,4} + \text{PSNR}_{2,3,4})/4 \tag{5.5.9}$$

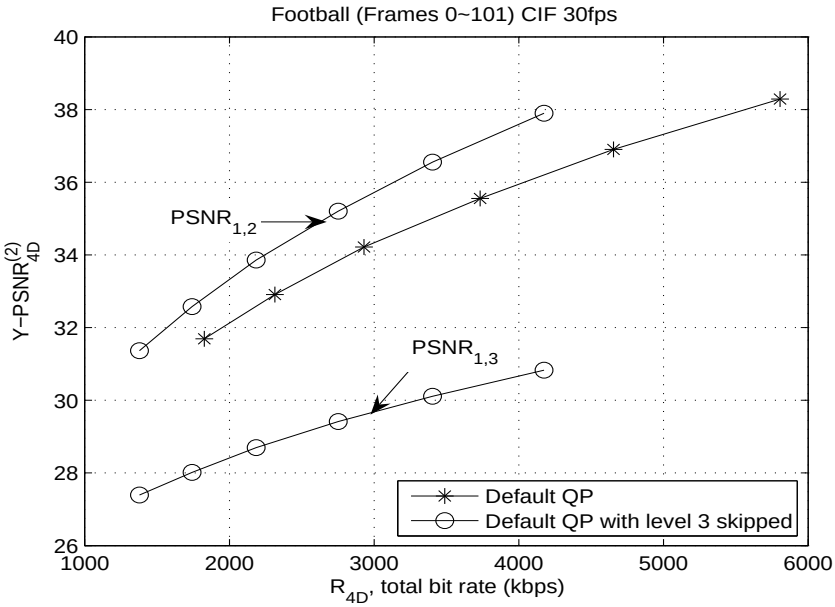
Assume two out of four descriptions are received correctly. There are C_4^2 $(= 6)$ different combinations for any two descriptions available, of which four combinations do not require frame interpolation and the other two need frame interpolation considering the case of frame skipping (corresponding to $k = 2$ in Table 5.1).

The frame interpolation technique in [160] is adopted for the high-motion "Football" sequence if the 1st and 3rd descriptions or the 2nd and 4th descriptions shown in Fig. 5.5 are received correctly. Fig. 5.18(a) plots 2-description based side distortions for the skipping and no-skipping schemes. Considering that the two cases with interpolation produce noticeably lower PSNR values than the other four cases without interpolation, the average $\text{PSNR}_{4D}^{(2)}$ in (5.5.8) may not be an appropriate measurement, we instead show the curves of $\text{PSNR}_{1,2}$ and $\text{PSNR}_{1,3}$ separately. When three or more descriptions are received ($k \geq 3$), no interpolation is required and thus the uncertainty for the frame interpolation result vanishes. As expected, due to the reduced redundancy the frame skipping method achieves better average rate-distortion performance than the no-skipping scheme for $k = 3$, shown in Fig. 5.18(b). The central PSNR ($k = 4$) with the skipping scheme necessarily outperforms the no-skipping one, which is shown in Fig. 5.19 used for redundancy estimation in the following.

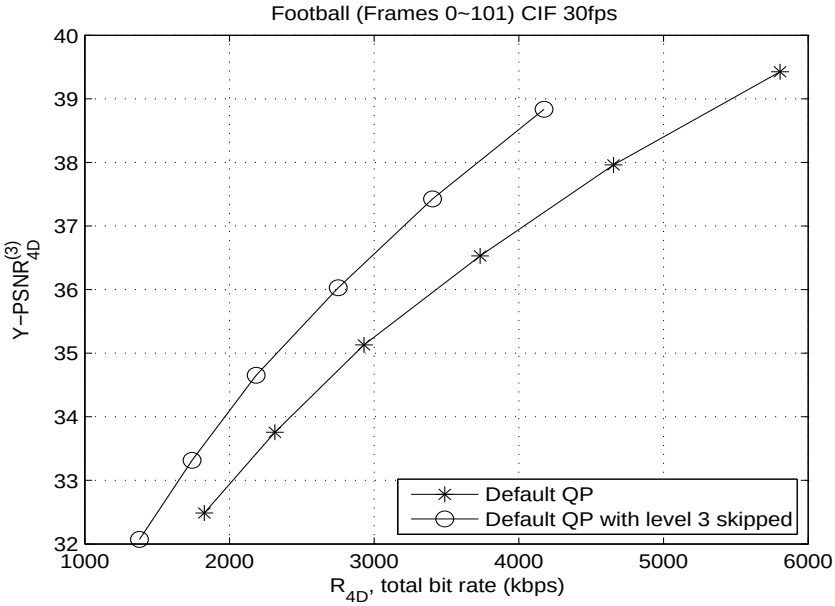
The redundancy $\rho_{4D}(d)$ in the case of all the 4 descriptions being received can be obtained as a function of central distortion d by

$$\rho_{4D}(d) = \frac{R_{4D}(d) - R_{single}(d)}{R_{single}(d)} \times 100\% \tag{5.5.10}$$

In Fig. 5.19, we show the rate-distortion curves of the proposed 4-description video coding with different skipping options and the single description coding. For the 4-description video coding, three curves are drawn, one without skipping any frames, the other skipping level 3, and the third one skipping both level 2 and level 3. For a given central PSNR value, we can estimate $\rho_{4D}(d)$ from the figure. For example, if the central PSNR is 35 dB, the redundancy is about 156% for the case of using default QP without frame skipping, while the redundancy can be



(a) 2-description based side distortion



(b) 3-description based side distortion

Figure 5.18: m-description based rate-distortion performance ($m = 2, 3$) in the proposed 4-description video coding on "Football".

reduced to 99% and 46%, when one level and two levels are skipped, respectively.

We can see that within the 4-description video coding framework, the redundancy

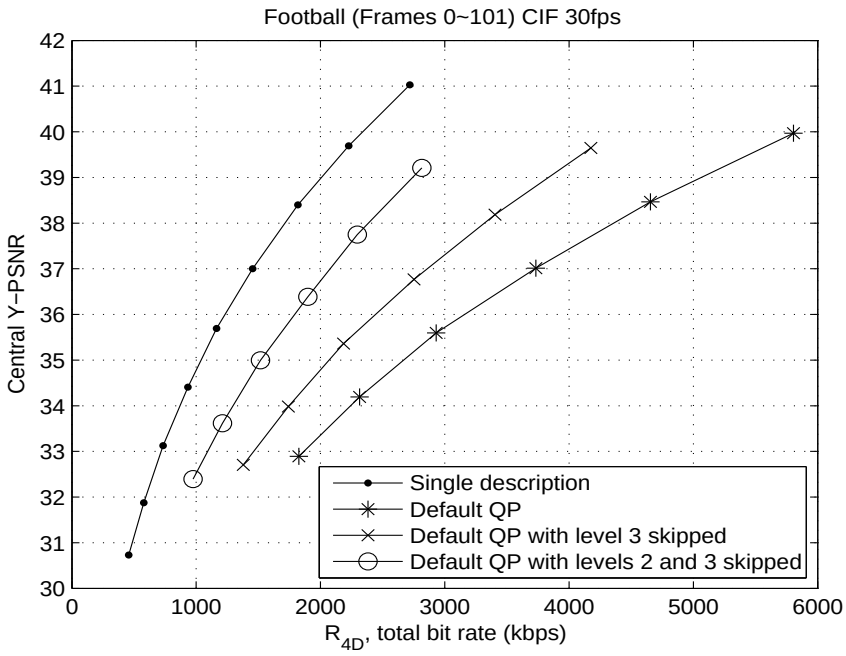


Figure 5.19: Comparison of central distortion of the proposed 4-description video coding with the distortion of single description coding.

can be adjusted with different frame skipping options.

In conclusion, frame skipping in the proposed 4-description video coding structure can effectively reduce redundancy on one hand, but may degrade the rate-distortion performance on the other hand when frame interpolation is required. For different requirements, the frame skipping scheme can be adapted accordingly within the proposed framework, which may produce other tradeoff results. It can be expected that some alternative frame skipping options in the proposed 4-description video coding framework, such as skipping half or quarter of the frames at the lowest temporal level (partially skipping) shown in Fig. 5.7 or even some frames in the higher levels, may achieve different tradeoffs between frame skipping case (all B_3) and no-skipping case discussed above.

5.6 Conclusion

In this chapter, we have proposed a general framework, hierarchical B pictures based multiple description video coding, which includes the odd/even frame splitting method as its simplest special case. Key frames are selected to be staggered in two or more descriptions to facilitate the better central decoding. The inter-description redundancy based on the proposed MDVC structure has been investigated. Specifically, 2-description video coding and 4-description video coding have been examined, for which some frame skipping ways are discussed. Based on the proposed MDVC framework, frame skipping can be adapted to control redundancy and k-description based side decoding quality. In view of complementary representations of each frame in different descriptions, a linear combination of the representations is considered to further improve the reconstruction quality at the central decoder or k-description based side decoders. Experimental results have demonstrated the effectiveness of our design.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Multiple description coding is a promising technique for supporting error resilient multimedia communications over lossy networks such as the Internet and wireless networks. Due to the source diversity and path diversity, MDC can adapt to varying network conditions much better than single description coding, especially in real-time communications where retransmission is not always viable.

The design, analysis and applications of MDC have been widely studied in recent years. In this thesis, we have addressed some important MDC issues on algorithm design, analysis and applications from a source coding point of view, which are summarized as follows.

In **Chapter 3**, we have shown that the existing index assignment algorithms of MDLVQ [79] [70] cannot always achieve an optimal index assignment for a finite

index number. To address non-optimality or high complexity problems in these schemes, we have proposed both greedy and general MDLVQ index assignment algorithms to pursue an optimal solution. The proposed methods take advantage of M-fraction lattice for index assignment design, which can help construct sub-lattice tuples easily. It is a 'divide-and-conquer' scheme to simplify the design of labeling function. We have shown that the optimal index assignment problem can be reduced to the transportation problem in operations research. In particular, the proposed greedy algorithm aims to reduce computational complexity without losing optimality, thus making it a good option for a high dimension lattice based MDLVQ. We conjecture that the proposed greedy algorithm provides an optimal index assignment. Furthermore, we have made an asymptotical analysis of the side distortions and obtained the closed-form expression of side distortions based on the proposed index assignment schemes, which is considered to be a comparable analytical result compared with those obtained in [79] and [70].

In **Chapter 4**, we have presented two enhancement schemes to make the product of central and side distortions for the two-stage MDQ closer to the multiple description rate-distortion bound. The first one is to use the refinement information generated at the second stage to reduce the side distortion, where an improvement of 0.58 dB coding gain has been shown. Moreover, central distortion can be further reduced if a good lattice quantizer is used to replace the scalar quantizer in the second stage. With the two improvements, a simple structure of two-stage MDQ can achieve an asymptotic distortion product with a smaller gap from the rate-distortion bound, compared with other relevant MDQ schemes. The two-stage MDC schemes have achieved very good performance in image applications, and we have also proposed a practical two-stage multiple description image coding based on trellis coded quantization, which is comparable (or usually

superior) to existing MD image coding schemes.

In **Chapter 5**, we have proposed a general framework, hierarchical B pictures based multiple description video coding. It is a pre-processing based scheme and includes the odd/even frame splitting method as its simplest special case. We have investigated the inter-description redundancy based on the proposed MDVC structure and developed an effective strategy to control the redundancy. In view of complementary representations of each frame in different descriptions, a linear combination of the representations is employed to improve the reconstruction quality at the central decoder or k-description based side decoders.

6.2 Recommendations for Future Work

Some possible directions of future research are listed as follows.

6.2.1 Source-Channel Erasure Codes with Lattice Codebooks for Multiple Description Coding

It has been shown in [47] and [48] as well as in Chapter 2 that a subset of the rate-distortion region of the symmetric M-channel multiple description coding problem can be achieved by employing (M, k) source-channel erasure codes (SCEC). The construction of the SCEC in [47] and [48] uses source coding with side information as well as random codebooks. In [161], Ostergaard et al. replaced the random codebooks of $(3, k)$ SCEC by structured (lattice) codebooks. With the closed-form expressions of central and side distortions for 3DLVQ shown in [70] and under high-resolution assumptions in certain cases, the SCEC with

lattice codebooks improves the achievable rate-distortion region over the traditional $(3, k)$ SCEC in [47] and [48]. In Chapter 3, we have also derived a better closed-form expression of side distortion for 3DLVQ, which leads to a smaller side distortion and may improve the achievable rate-distortion region over the lattice codebook based $(3, k)$ SCEC in [161].

6.2.2 Joint Optimal Multi-path Routing and Rate Control for MDC

When combined with appropriate channel coding schemes such as FEC or SCEC, MD coding with multi-path transport can significantly improve the reconstruction quality over the conventional single description single-path coding schemes. An interesting issue, the joint multi-path routing and rate control for multiple description coding in wireless networks, could be studied. Relevant work can be found in [162]. There might be several possible research problems for this part. The first problem is how to select paths to optimize the expected reconstruction quality. Since searching for paths to minimize the expected end-to-end reconstruction distortion by simultaneously taking into account the priority and bandwidth of paths is a highly complicated problem, some simplified and efficient greedy solutions could be developed. The second problem is how to explore an rate control strategy in order to adapt to channel conditions. Jointly designing multi-path routing and rate control may provide an important methodology for high-quality real-time applications over wireless networks.

6.2.3 Multiple Description Watermarking

Digital watermarking is a useful and powerful tool for multimedia security such as copyright protection, tamper proofing and assessment, broadcast monitoring and fingerprinting. Many multimedia services are provided through wire/wireless communications, where transmission errors are present. Robustness of watermarking is a major concern during the transmission of watermarked multimedia messages. As we have known that MDC is a promising method for robust transmission of information over non-prioritized and unpredictable networks. It is also interesting to find that MDC can be applied in multimedia watermarking. Recently several applications of MDC in watermarking, known as multiple description watermarking, have been developed which have exhibited some advantages. In these researches, MDC can not only be used to protect the watermark, host signal or watermarked host signal, but also be used as an attack method. We have also done some preliminary work and presented a summary of multiple description watermarking in [138]. More advanced multiple description watermarking schemes could be further investigated.

Appendix A

Proof of (3.3.18) in Chapter 3

We prove that in the asymptotical case of $N \to \infty$ and $v_s \to 0$, the index number N can be expressed as a function of $R_M^{\max}$, given the number of descriptions M. That is

$$N \approx \omega_L^{M-1} \cdot \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \prod_{i=2}^{M} \beta_i \cdot (R_M^{\max})^{(M-1)L}.$$

Proof:

We denote $f(M, r_{M+1}^2)$ as the total number of M-tuples assigned in $V_0(O)$ with their SPSD values no larger than r_{M+1}^2. These M-tuples are obtained from the M-tuples in T_{λ_1} through shifting. We assume that there are approximately $f(M, r_{M+1}^2)/M^L$ fine lattice points in each Voronoi cell of $\Lambda_{\frac{s}{M}}$ as $v_{\frac{s}{M}} \to 0$. Accordingly, there are approximately $f(M, r_{M+1}^2)/M^L$ M-tuples with their SPSD values no larger than r_{M+1}^2 assigned in each Voronoi cell of $\Lambda_{\frac{s}{M}}$. Moreover, it is assumed that the center of each Voronoi cell of $\Lambda_{\frac{s}{M}}$ is just the common centroid of all the $f(M, r_{M+1}^2)/M^L$ M-tuples assigned to it[i].

[i]We will show in (A.0.2) that the last assumption is reasonable for $M = 2$. We argue that

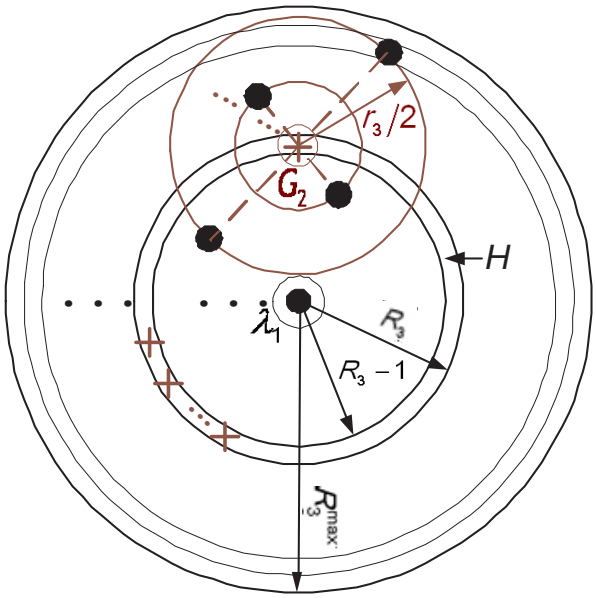


Figure A.1: 2-fraction lattice points in a hollow region H and sublattice points in a sphere $S(G_2, r_3/2)$.

Consider the L-dimensional hollow sphere $H = S(\lambda_1, R_M) - S(\lambda_1, R_M - 1)$. The number of sublattice points of $\Lambda_{\frac{s}{i}}$ in H is $\left|H \cap \Lambda_{\frac{s}{i}}\right|$ and asymptotically as $v_{\frac{s}{i}} \to 0$

$$\left|H \cap \Lambda_{\frac{s}{i}}\right| \approx \mathrm{Vol}(H)/v_{\frac{s}{i}} \approx \omega_L(R_M^L - (R_M - 1)^L)/v_{\frac{s}{i}} \tag{A.0.1}$$

Based on the above preliminaries, we prove (3.3.18) in the following. We start with $M = 2$. It is straightforward to obtain that $f(2, r_3^2) = \frac{\omega_L}{v_s} \cdot (R_2^{\max})^L$, where $SPSD_2^{\max} = (R_2^{\max})^2 = r_3^2$.

We consider the case of $M = 3$ next. Fig. A.1 shows a modified case of Fig. 3.3, which helps to illustrate the proof. We have known that the centroid of any two sublattice points of Λ_s is always lying on $\Lambda_{\frac{s}{2}}$. We consider a hollow region H corresponding to a particular R_3. Given a 2-fraction lattice point G_2

it is still acceptable for $M \geq 3$, although no rigorous proof is available at the moment.

(marked with + in the figure) in H, those 2-tuples with their centroids being G_2 and their SPSD values no larger than r_3^2 are selected. The number of these 2-tuples[ii] is given by $\frac{\omega_L}{v_s}\left(\sqrt{\frac{(R_3^{\max})^2-R_3^2}{3}}\right)^L$, which is equivalent to the number of sublattice points of Λ_s in a sphere $S(G_2, \frac{r_3}{2})$ with $\frac{r_3}{2} = \sqrt{\frac{(R_3^{\max})^2-R_3^2}{3}}$. All the sublattice points of Λ_s in the sphere are symmetric about G_2 [1]. We can easily obtain that

$$\frac{f(2,r_3^2)}{(3-1)^L} = \frac{\omega_L}{v_s}\left(\sqrt{\frac{(R_3^{\max})^2-R_3^2}{3}}\right)^L \tag{A.0.2}$$

where the expression on the left side is the average number of tuples with SPSD values no more than r_3^2 in each Voronoi cell of $\Lambda_{\frac{s}{2}}$. The above process is to calculate the number of 3-tuples for a given G_2. This kind of G_2 points in H is given by $\omega_L(R_3^L-(R_3-1)^L)/v_{\frac{s}{2}}$ (i.e. let $M=3$ in (A.0.1)). In view that G_2 can also lie in different hollow regions, where different hollow regions correspond to different R_3 values, the total number of 3-tuples with their SPSD values no larger than $SPSD_3^{\max}(=r_4^2)$ is given by[iii]

$$\begin{aligned}
f(3,r_4^2) &= \sum_{R_3=1}^{R_3^{\max}} \frac{\omega_L}{v_{\frac{s}{2}}}\left(R_3^L-(R_3-1)^L\right)\frac{f(2,r_3^2)}{(3-1)^L} \\
&= \sum_{R_3=1}^{R_3^{\max}} \frac{\omega_L}{v_{\frac{s}{2}}}\left(R_3^L-(R_3-1)^L\right)\frac{\omega_L}{v_s}\left(\sqrt{\frac{(R_3^{\max})^2-R_3^2}{3}}\right)^L \\
&\overset{(a)}{\approx} \frac{\omega_L^2}{v_{\frac{s}{2}}v_s}\sum_{R_3=1}^{R_3^{\max}} LR_3^{L-1}((R_3^{\max})^2-R_3^2)^{\frac{L}{2}}\left(\tfrac{1}{3}\right)^{\frac{L}{2}} \\
&\overset{(b)}{=} \frac{\omega_L^2}{v_{\frac{s}{2}}v_s}\cdot\left(\tfrac{1}{3}\right)^{\frac{L}{2}}\sum_{n=0}^{\frac{L}{2}}\binom{\frac{L}{2}}{n}(-1)^n\frac{L}{L+2n}\cdot(R_3^{\max})^{2L} \\
&= \frac{\omega_L^2}{v_{\frac{s}{2}}v_s}\cdot\beta_3\cdot(R_3^{\max})^{2L}
\end{aligned} \tag{A.0.3}$$

[ii] In Section 3.3.1 we have discussed how to generate these 2-tuples based on G_2 and $SPSD_2^{\max}$ in details.

[iii] In this asymptotical analysis we assume that all the G_2 points within a given H share the same distance R_M from λ_1. The error due to this assumption is negligible, according to [70].

where (a) follows by using

$$(r-1)^L = r^L - Lr^{L-1} + O(r^{L-2}) \tag{A.0.4}$$

and (b) follows by using

$$\begin{aligned}
&\sum_{R=1}^{R^{\max}} R^{L-1}\left((R^{\max})^2 - R^2\right)^{xL} \\
&= \sum_{R=1}^{R^{\max}} \left(R^{\frac{L-1}{xL}} (R^{\max})^2 - R^{\frac{L-1+2xL}{xL}} \right)^{xL} \\
&\overset{(c)}{=} \sum_{R=1}^{R^{\max}} \sum_{n=0}^{xL} \binom{xL}{n} (R^{\frac{L-1}{xL}} (R^{\max})^2)^{xL-n} (-R^{\frac{L-1+2xL}{xL}})^n \\
&= \sum_{n=0}^{xL} \binom{xL}{n} (-1)^n (R^{\max})^{2xL-2n} \sum_{R=1}^{R^{\max}} R^{L+2n-1} \\
&\overset{(d)}{=} \frac{1}{L+2n} \sum_{n=0}^{xL} \binom{xL}{n} (-1)^n (R^{\max})^{2xL+L}
\end{aligned} \tag{A.0.5}$$

In (A.0.5), (c) follows by use of the binomial series expansion [163]

$$(x+y)^k = \sum_{n=0}^{k} \binom{k}{n} x^{k-n} y^n, \tag{A.0.6}$$

and (d) is obtained by

$$\sum_{r=1}^{z} r^L = \frac{1}{L+1} z^{L+1} + O(r^L) \tag{A.0.7}$$

Based on mathematical induction, we assume the statement is true for $(M-1)$

descriptions with $(M-1) \geq 3$, i.e.,

$$f(M-1, r_M^2) = \omega_L^{M-2} \left(\prod_{i=1}^{M-2} v_{\frac{s}{i}}^{-1} \right) \left(\prod_{i=2}^{M-1} \beta_i \right) (R_{M-1}^{\max})^{(M-2)L} \tag{A.0.8}$$

Then in the M-description case,

$$\begin{aligned}
f(M, r_{M+1}^2) &= \sum_{R_M=1}^{R_M^{\max}} \frac{\omega_L \left(R_M^L - (R_M-1)^L \right)}{v_{\frac{s}{M-1}}} \frac{f(M-1, r_M^2)}{(M-1)^L} \\
&\overset{(a)}{\approx} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M-1} \beta_i \cdot \sum_{R_M=1}^{R_M^{\max}} \frac{R_M^L - (R_M-1)^L}{(M-1)^L} (R_{M-1}^{\max})^{(M-2)L} \\
&\overset{(b)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M-1} \beta_i \sum_{R_M=1}^{R_M^{\max}} \frac{L R_M^{L-1}}{(M-1)^L} (R_{M-1}^{\max})^{(M-2)L} \\
&= \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M-1} \beta_i \sum_{R_M=1}^{R_M^{\max}} \frac{L R_M^{L-1}}{(M-1)^L} \cdot \left(\frac{SPSD_{M-1}^{\max}}{M-2} \right)^{\frac{M-2}{2}L} \\
&= \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M-1} \beta_i \cdot \sum_{R_M=1}^{R_M^{\max}} \frac{L R_M^{L-1}}{(M-1)^L} \left(\frac{r_M^2}{M-2} \right)^{\frac{M-2}{2}L} \\
&\overset{(c)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M-1} \beta_i \cdot \sum_{R_M=1}^{R_M^{\max}} \frac{L R_M^{L-1}}{(M-1)^L} \left(\frac{(M-1)^2((R_M^{\max})^2 - R_M^2)}{M(M-2)} \right)^{\frac{M-2}{2}L} \\
&\overset{(d)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \prod_{i=2}^{M} \beta_i \cdot (R_M^{\max})^{(M-1)L}
\end{aligned} \tag{A.0.9}$$

where (a) to (c) follows by use of (A.0.8), (A.0.4) and (3.3.15), respectively. (d) is obtained by using (3.3.19) and (A.0.5).

The number of M-tuples $f(M, r_{M+1}^2)$ should be sufficiently large to label N fine lattice points. Therefore in the M-description case, let r_{M+1}^2 $(= SPSD_3^{\max})$ large enough to make $f(M, r_{M+1}^2) = N$, then (3.3.18) is proved.

Appendix B

Proof of (3.3.20) in Chapter 3

By using the results obtained in Appendix I, we prove that in the asymptotical case of $N \to \infty$ and $v_s \to 0$, the sum of SPSD values of N M-tuples for assigning N fine lattice points in $V_0(O)$ is given by

$$\sum_{\lambda \in V_0(O)} SPSD(M,\lambda) \approx G_L \cdot (L+2) \left(\frac{1}{(M-1)!}\right)^{\frac{2}{M-1}} \cdot \frac{\gamma_M^{\frac{2}{(M-1)L}+1}}{\prod_{i=2}^{M} \beta_i} \cdot N \cdot N^{\frac{M}{M-1}\cdot\frac{2}{L}} v^{\frac{2}{L}}$$

Proof:

Denote $T_{\lambda_1}(r_{M+1}^2)$ as a set of fine lattice points with volume equal to $f(M, r_{M+1}^2)$, where $f(M, r_{M+1}^2)$ is the total number of M-tuples assigned in $V_0(O)$ with their SPSD values no larger than r_{M+1}^2. Each fine lattice point in $T_{\lambda_1}(r_{M+1}^2)$ is labeled by an M-tuple with the first element being λ_1 and SPSD value no larger than r_{M+1}^2. If $N = f(M, r_{M+1}^2)$, then the side distortion sum for the N fine lattice points in T_{λ_1} is equal to the sum for those in $V_0(O)$, which will be used in the following analysis. The sum of SPSD values of N M-tuples assigned to $V_0(O)$ is

given by

$$\begin{aligned}\sum_{\lambda\in V_0(O)} SPSD(M,\lambda) &= \sum_{\lambda\in T_{\lambda_1}(r^2_{M+1})} SPSD(M,\lambda)\\ &= \sum_{\lambda\in T_{\lambda_1}(r^2_{M+1})} \left((M-1)R_M^2 + \tfrac{M}{M-1}r_M^2\right)\end{aligned} \tag{B.0.1}$$

We have assumed previously that all the sublattice points of $\Lambda_{\frac{M}{i}}$ within the hollow sphere H are at the exact same distance (i.e. R_M) from λ_1. For 2DLVQ ($M = 2$), r_2^2 is always zero, and the sum of SPSD values of N 2-tuples with $SPSD_2^{\max} = r_3^2$ is

$$\begin{aligned}\sum_{\lambda\in V_0(O)} SPSD(2,\lambda) &= \sum_{\lambda\in T_{\lambda_1(r_3^2)}} SPSD(2,\lambda)\\ &\approx \sum_{R_2=1}^{R_2^{\max}} \tfrac{\omega_L}{v_s}\left(R_2^L - (R_2-1)^L\right)\cdot R_2^2\cdot\tfrac{1}{L}\\ &\approx \tfrac{\omega_L}{v_s}\cdot\tfrac{1}{L+2}(R_2^{\max})^{L+2}\\ &= \tfrac{\omega_L}{v_s}\cdot\gamma_2\cdot(R_2^{\max})^{L+2}\end{aligned} \tag{B.0.2}$$

For the case of $M \geq 3$, we need to deduce the two terms of (B.0.1) separately. Let us firstly consider the first term. In each Voronoi cell of $\Lambda_{\frac{s}{M-1}}$, there are approximately $\frac{f(M-1,r_M^2)}{(M-1)^L}$ $(M-1)$-tuples with their SPSD values no larger than

r_M^2. Thus, for $M \geq 3$

$$
\begin{aligned}
&\sum_{\lambda \in T_{\lambda_1}(r_{M+1}^2)} (M-1)R_M^2 \\
&\approx \sum_{R_M=1}^{R_M^{\max}} \frac{\omega_L}{v_{\frac{s}{M-1}}} \left(R_M^L - (R_M-1)^L\right) \cdot \frac{f(M-1, r_M^2)}{(M-1)^L} \cdot (M-1) \cdot R_M^2 \cdot \frac{1}{L} \\
&\overset{(a)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \prod_{i=2}^{M-1} \beta_i \cdot \frac{1}{(M-1)^{L-1}} \\
&\cdot \frac{1}{L} \sum_{R_M=1}^{R_M^{\max}} \left(R_M^L - (R_M-1)^L\right) R_M^2 \cdot (R_{M-1}^{\max})^{(M-2)L} \\
&\overset{(b)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \prod_{i=2}^{M-1} \beta_i \cdot \frac{(M-1)^{1+(M-3)L}}{(M^2-2M)^{(M-2)L/2}} \\
&\cdot \sum_{n=0}^{\frac{(M-2)L}{2}} \binom{\frac{(M-2)L}{2}}{n} \frac{(-1)^n}{L+2n+2} \cdot (R_M^{\max})^{(M-1)L+2} \\
&\overset{(c)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \prod_{i=2}^{M-1} \beta_i \cdot \varphi_M \cdot (R_M^{\max})^{(M-1)L+2}
\end{aligned}
\tag{B.0.3}
$$

where (a) follows by use of (3.3.18), and the proof of (b) is similar to that of (A.0.9) (steps (b) and (c)). (c) is obtained by using (3.3.22).

The deduction for the second term is more complicated. Consider the second term when $M = 3$. We have assumed in Appendix I that there are approximately $\frac{f(2,r_3^2)}{(3-1)^L}$ 2-tuples with their SPSD values no larger than r_3^2 in each Voronoi cell of $\Lambda_{\frac{s}{2}}$. The SPSD sum of these 2-tuples in the Voronoi cell of G_2 can be estimated by $\frac{1}{(3-1)^L} \cdot \sum_{\lambda \in T_{\lambda_1}(r_3^2)} SPSD(2, \lambda)$. Therefore the sum of SPSD values for N 3-tuples

with $SPSD_3^{\max}$ equal to r_4^2 is

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_4^2)} \tfrac{3}{3-1} r_3^2 \\
&\approx \tfrac{3}{3-1} \cdot \sum_{R_3=1}^{R_3^{\max}} \tfrac{\omega_L}{v_{\frac{s}{2}}} \left(R_3^L - (R_3-1)^L\right) \cdot \left(\sum_{\lambda\in T_{\lambda_1}(r_3^2)} SPSD(2,\lambda) \right) \cdot \tfrac{1}{(3-1)^L} \\
&\overset{(a)}{=} \omega_L^2 \prod_{i=1}^{2} v_{\frac{s}{i}}^{-1} \cdot \tfrac{1}{L+2} \cdot \tfrac{3-1}{3^{L/2}} \cdot \sum_{n=0}^{\frac{L}{2}+1} \binom{\frac{L}{2}+1}{n} \tfrac{(-1)^n}{L+2n} \cdot (R_3^{\max})^{2L+2} \\
&\overset{(b)}{=} \omega_L^2 \prod_{i=1}^{2} v_{\frac{s}{i}}^{-1} \cdot \psi_3 \cdot (R_3^{\max})^{2L+2}
\end{aligned} \tag{B.0.4}$$

where (a) follows by use of (B.0.2), (3.3.12), $r_3^2 = SPSD_2^{\max}$, (3.3.15) and (A.0.5) in order. (b) follows by (3.3.23). Thus

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_4^2)} SPSD(3,\lambda) \\
&= \sum_{\lambda\in T_{\lambda_1}(r_4^2)} (3-1)R_3^2 + \sum_{\lambda\in T_{\lambda_1}(r_4^2)} \tfrac{3}{3-1} r_3^2 \\
&\approx \omega_L^2 \prod_{i=1}^{2} v_{\frac{s}{i}}^{-1} \cdot (\varphi_3 + \psi_3) \cdot (R_3^{\max})^{2L+2} \\
&= \omega_L^2 \prod_{i=1}^{2} v_{\frac{s}{i}}^{-1} \cdot \gamma_3 \cdot (R_3^{\max})^{2L+2}
\end{aligned} \tag{B.0.5}$$

In the following, we use mathematical induction to prove the case of $M > 3$. As $M = 4$,

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_5^2)} \tfrac{4}{4-1} r_4^2 \\
&\approx \tfrac{4}{4-1} \cdot \sum_{R_4=1}^{R_4^{\max}} \tfrac{\omega_L}{v_{\frac{s}{3}}} \left(R_4^L - (R_4-1)^L\right) \cdot \tfrac{1}{(4-1)^L} \sum_{\lambda\in T_{\lambda_1}(r_4^2)} SPSD(3,\lambda) \\
&\approx \tfrac{4}{(4-1)^{L+1}} \cdot \tfrac{\omega_L}{v_{\frac{s}{3}}} \sum_{R_4=1}^{R_4^{\max}} L R_4^L \cdot \omega_L^2 \prod_{i=1}^{2} v_{\frac{s}{i}}^{-1} \cdot \gamma_3 \cdot \left(\tfrac{(4-1)^2}{4\cdot(3-1)} ((R_4^{\max})^2 - R_4^2) \right)^{L+1} \\
&\approx \omega_L^3 \prod_{i=1}^{3} v_{\frac{s}{i}}^{-1} \cdot \gamma_3 \cdot \psi_4 \cdot (R_4^{\max})^{3L+2}
\end{aligned} \tag{B.0.6}$$

and

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_5^2)} SPSD(4,\lambda) \\
&\approx \omega_L^3 \prod_{i=1}^{3} v_{\frac{s}{i}}^{-1}\cdot(\varphi_4\beta_3+\gamma_3\psi_4)\cdot(R_4^{\max})^{3L+2} \\
&= \omega_L^3 \prod_{i=1}^{3} v_{\frac{s}{i}}^{-1}\cdot\gamma_4\cdot(R_4^{\max})^{3L+2}
\end{aligned} \tag{B.0.7}$$

Now suppose the following equation for the sum of SPSD values in the $(M-1)$-description case is true (we include the case of $\beta_2 = 1$ in the following equation.)

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_M^2)} SPSD(M-1,\lambda) \\
&\approx \omega_L^{M-2} \prod_{i=1}^{M-2} v_{\frac{s}{i}}^{-1}\cdot(\varphi_{M-1}\prod_{i=2}^{M-2}\beta_i+\gamma_{M-2}\psi_{M-1})\cdot(R_{M-1}^{\max})^{(M-2)L+2} \\
&= \omega_L^{M-2} \prod_{i=1}^{M-2} v_{\frac{s}{i}}^{-1}\cdot\gamma_{M-1}\cdot(R_{M-1}^{\max})^{(M-2)L+2}
\end{aligned} \tag{B.0.8}$$

Therefore the second term for M-description case is

$$\begin{aligned}
&\sum_{\lambda\in T_{\lambda_1}(r_{M+1}^2)} \tfrac{M}{M-1} r_M^2 \\
&\approx \tfrac{M}{M-1}\cdot\sum_{R_M=1}^{R_M^{\max}} \tfrac{\omega_L}{v_{\frac{s}{M-1}}}\left(R_M^L-(R_M-1)^L\right) \\
&\cdot\left(\sum_{\lambda\in T_{\lambda_1}(r_M^2)} SPSD(M-1,\lambda)\right)\tfrac{1}{(M-1)^L} \\
&\approx \tfrac{M}{(M-1)^{L+1}}\cdot\tfrac{\omega_L}{v_{\frac{s}{M-1}}}\sum_{R_M=1}^{R_M^{\max}} LR_M^L \\
&\cdot\left(\omega_L^{M-2}\prod_{i=1}^{M-2} v_{\frac{s}{i}}^{-1}\cdot\gamma_{M-1}\cdot(R_{M-1}^{\max})^{(M-2)L+2}\right) \\
&\approx \tfrac{M}{(M-1)^{L+1}}\cdot\omega_L^{M-1}\cdot\prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1}\cdot\gamma_{M-1}\cdot\sum_{R_M=1}^{R_M^{\max}} LR_M^L \\
&\cdot\left(\tfrac{(M-1)^2}{M\cdot(M-2)}((R_M^{\max})^2-R_M^2)\right)^{\frac{(M-2)L+2}{2}} \\
&\approx \omega_L^{M-1}\prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1}\cdot\gamma_{M-1}\cdot\psi_M\cdot(R_M^{\max})^{(M-1)L+2}
\end{aligned} \tag{B.0.9}$$

If we let $f(M, r_{M+1}^2) = N$, the sum of SPSD values with these N M-tuples is given by

$$\begin{aligned}
&\sum_{\lambda \in V_0(O)} SPSD(M,\lambda) = \sum_{\lambda \in T_{\lambda_1}} SPSD(M,\lambda) \\
&\approx \omega_L^{M-1} \cdot \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} (\varphi_M \prod_{i=2}^{M-1} \beta_i + \gamma_{M-1}\psi_M)(R_M^{\max})^{(M-1)L+2} \\
&\overset{(a)}{=} \omega_L^{M-1} \prod_{i=1}^{M-1} v_{\frac{s}{i}}^{-1} \cdot \gamma_M \cdot \left(N \cdot \omega_L^{1-M} \prod_{i=1}^{M-1} v_{\frac{s}{i}} \cdot \prod_{i=2}^{M} \beta_i^{-1} \right)^{\frac{(M-1)L+2}{(M-1)L}} \\
&\overset{(b)}{=} \omega_L^{\frac{-2}{L}} \left(\frac{(Nv)^{M-1}}{((M-1)!)^L} \right)^{\frac{2}{(M-1)L}} \cdot \frac{\gamma_M}{\frac{2}{\prod_{i=2}^{M} \beta_i^{\overline{(M-1)L}+1}}} \cdot N^{\frac{(M-1)L+2}{(M-1)L}} \\
&\overset{(c)}{=} G_L \cdot (L+2) \left(\frac{1}{(M-1)!} \right)^{\frac{2}{M-1}} \cdot \frac{\gamma_M}{\frac{2}{\prod_{i=2}^{M} \beta_i^{\overline{(M-1)L}+1}}} \cdot N \cdot N^{\frac{M}{M-1} \cdot \frac{2}{L}} v^{\frac{2}{L}}
\end{aligned} \tag{B.0.10}$$

where (a) follows by using (3.3.18), (b) follows by $v_{\frac{s}{i}} = vNi^L$, and (c) follows by use of (B.0.11) (see [15])

$$\omega_L = G_L^{-2/L}(L+2)^{-2/L} \tag{B.0.11}$$

and G_L is the normalized second moment of an L-dimensional sphere. This completes the proof.

Appendix C

Bit allocation problem for the proposed 2-description video coding in Chapter 5

In this appendix, we show that the proposed optimal bit allocation for minimizing the expected average distortion for a given packet loss probability p_l is similar to the bit allocation strategy proposed in [5].

In [5], the redundant slice based 2-description video coding is shown in Fig. C.1. Primary slices are used to code the primary picture, which are associated to a single description decoding procedure. On the other hand, redundant slices represent an alternative representation of a picture. Once some blocks of the primary picture cannot be correctly decoded due to transmission losses, whereas the redundant slice is correctly decoded, the decoder shall replace the damaged blocks of primary picture with the corresponding ones of the redundant slice. The obtained H.264 bit stream with redundant slices can be used to form two

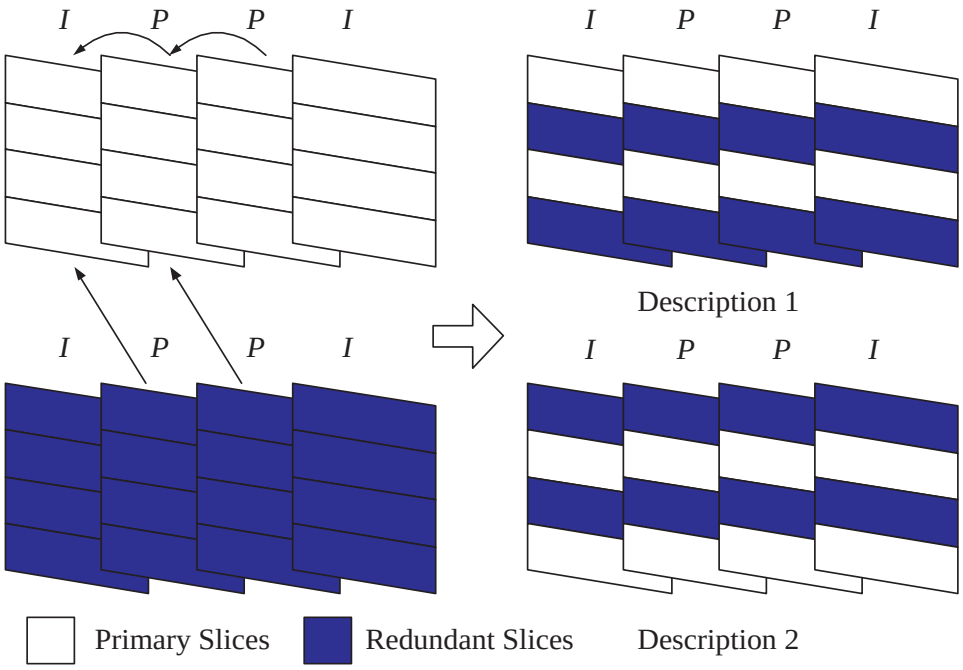


Figure C.1: Redundant slice based 2-description video coding in [5].

balanced descriptions of the original video sequence by simply rearranging the compressed slices. Based on this structure, an effective allocation strategy was developed to tune the redundancy (bit cost on coding redundant slices) to limit the decoding drift.

In [5], the optimal redundancy allocation is shown as a function of the transmission conditions and the amount of drift distortion based on the results of [164–167], which investigated the rate distortion analysis on single description video coding. In the following, we consider the expected distortion of GOP caused by the loss of the k-th slice in the frame $F(k)$ with the packet loss probability p_l, then it can be evaluated as

$$d_k = (1 - p_l)d_{p,k} + p_l(1 - p_l)d_{t,k} + p_l^2 d_{0,k} \tag{C.0.1}$$

where $d_{p,k}$ is the distortion due to the primary slice k, $d_{t,k}$ is the distortion caused

by the loss of the slice k, and $d_{0,k}$ is the distortion when both representations of slice k are not available. Nevertheless, if p_l is small, the last term of $p_l^2 d_{0,k}$ in (C.0.1) can be ignored. Therefore, the expected distortion can be simplified to

$$\tilde{d}_k \approx (1-p_l)d_{p,k} + p_l(1-p_l)d_{t,k} \tag{C.0.2}$$

It needs a lengthy derivation in [5] to show that $d_{t,k}$ can be approximated as

$$d_{t,k} \approx d_{r,k} \sum_{n=1}^{N-F(k)} f[n] = d_{r,k}\phi_{F(k)} \tag{C.0.3}$$

where $\phi_{F(k)} = \sum_{n=1}^{N-F(k)} f[n]$, and $f[n]$ (with $f[0]=1$) is the power transfer function to model the distortion attenuation experienced at distance n from the mismatch.

The total expected distortion of the GOP (with the size being N) can be evaluated by summing all the slice contributions, that is

$$\begin{aligned}
\tilde{D}_{\text{exp}} &= \sum_k \tilde{d}_k \\
&= \sum_{i=1}^{N} \sum_{k:F(k)=i} (1-p_l)d_{p,k} + p_l(1-p_l)d_{t,k} \\
&= (1-p_l) \sum_{i=1}^{N} \sum_{k:F(k)=i} (d_{p,k} + p_l\phi_i d_{r,k}) \\
&= (1-p_l)\left(\sum_{i=1}^{N} D_{p,i} + p_l \sum_{i=1}^{N} D_{r,i}\phi_i \right)
\end{aligned} \tag{C.0.4}$$

where $D_{p,i}$ is the distortion of the redundant representation of picture i and $D_{r,i}$ is the distortion of the primary representation of picture i. Further denote $R_{p,i}$ and $R_{r,i}$ as the bit rates for $D_{p,i}$ and $D_{r,i}$, respectively, then the bit allocation

optimization problem can be formulated as

$$\begin{aligned} &\min \tilde{D}_{\text{exp}} \\ &\text{subject to} \\ &\sum_{i=1}^{N} (R_{p,i} + R_{r,i}) = R_{GOP} \end{aligned} \tag{C.0.5}$$

where R_{GOP} is the total bit rate of the GOP. (C.0.5) can be solved by using standard Lagrangian approach. We can see from [5] when the following two conditions hold, $\tilde{D}_{\text{exp}}$ can be minimized:

$$\frac{\partial D_p}{\partial R_p} = p_l \phi_1 \frac{\partial D_{r,1}}{\partial R_{r,1}} \tag{C.0.6}$$

$$\frac{\partial D_{r,i}}{\partial R_{r,i}} = \frac{\phi_1}{\phi_i} \frac{\partial D_{r,1}}{\partial R_{r,1}} \tag{C.0.7}$$

Based on (C.0.6) and (C.0.7), an optimal redundancy allocation strategy can be accomplished by using a closed-loop rate-control technique, which can partition the rate budget between the primary and redundant slices. In [5], a simple but effective open-loop approach based on the optimal selection of the quantization parameter is adopted. Specifically, the following R-D approximation [168] in H.264 is used

$$\frac{\partial D}{\partial R} = -0.85 \cdot 2^{\frac{QP-12}{3}} \tag{C.0.8}$$

From (C.0.6), (C.0.7) and (C.0.8) we obtain the optimized QP values for primary

and redundant pictures

$$QP_p = QP_{r,1} + 3\log(p_l\phi_1) \tag{C.0.9}$$

$$QP_{r,i} = QP_{r,1} + 3\log(\tfrac{\phi_1}{\phi_i}) \tag{C.0.10}$$

where $QP_{r,i}$ is the QP for the i-th redundant pictures. In [5], the power transfer function $f[n]$ can be approximated as $f[n] = e^{-\alpha n}$ (with $\alpha = 0.1$ in this thesis), thus ϕ_i in (C.0.10) comes to be

$$\phi_i = \tfrac{1-e^{-\alpha(N-i+1)}}{1-e^{-\alpha}} \tag{C.0.11}$$

with $i \in [1, N]$. Once the QP value for primary pictures is given, the QP values for redundant pictures can be obtained from (C.0.9) and (C.0.10). Note that all the above derivations together with detailed proofs can be found in [5].

Now we consider our proposed 2-description video coding with the KB1 coding structure in Fig. C.2. From the decoder side, we can see that the proposed scheme has no significant difference compared with Tillo's scheme in [5] except the reference frames and prediction orders. Here the B frames can be considered as redundant frames. Therefore it is reasonable to apply the QP optimization strategy developed in [5] to our proposed scheme. In some sense, the methods in [5] and [3] together with our method can be in the same category. Specifically, in the methods of [5] and [3] redundancy is added based on slice or slice groups, while in our method redundancy is added based on frames. Since one frame can be split to several slices, it is possible to generate two different descriptions for one frame by quantizing different slices or slice groups with different QPs, as shown

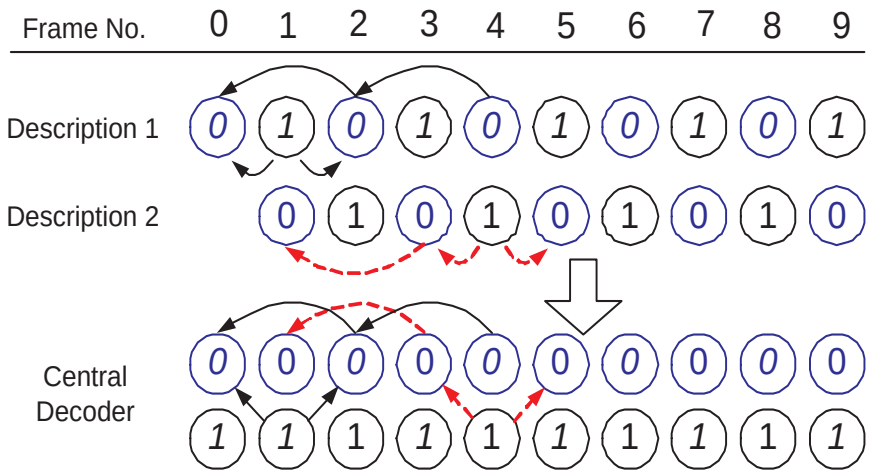


Figure C.2: An equivalent representation of the proposed 2-description video coding with KB1 coding structure.

in [5] and [3]. One can imagine that if the slice size in the methods of [5] and [3] is enforced to change to the frame size, two group of slices (which correspond to two descriptions) cannot be generated for a separate frame any more. In this case, a pair of frames instead of one frame has to be considered as a minimum unit in the MDC design. Our staggered MDC method, where two neighbor frames are considered, can be a good solution towards this change.

The main idea in [5] and [3] is to use the H.264 codec to generate an H.264 single description stream. Then a redundant representation of each coded slice, which may be redundant slice in [5] or coarsely quantized slice groups in [3], is encoded. The advantage of this idea is that when the channel is error free, the central decoding performance is only determined by the H.264 single description stream. That is, the central PSNR can be known before decoding. The side PSNR is changed according to the amount of redundancy added. While in our scheme, the idea is to generate a good side PSNR first, and the central PSNR depends on the staggering gain of two descriptions. One limitation of Tillo's method in [5] is that the redundant pictures cannot be used with B frames in

H.264. Wang's method in [3] also meets some optimization problems to generate coarsely quantized representations of slices as B pictures are used, since the whole B picture has been already coarsely quantized, compared with I and P pictures. In our method, different hierarchical B picture levels can be adopted in the MDC design. In this thesis, we have only considered the bit allocation optimization with the KB1 coding structure. To extend the optimization process to the coding structure with higher hierarchical levels (e.g. using KB7 coding structure) will be considered as future work.

Appendix D

Glossary of Symbols and Terms

Table D.1: List of Acronyms

ABBREVIATIONS	FULL EXPRESSIONS
ARQ	Automatic repeat-request
DCT	Discrete cosine transform
GOP	Group of pictures
LC	Layered coding
ECMDSQ	Entropy constrained multiple description scalar quantization
FEC	Forward error correction
MB	Micro block
MCTF	Motion compensated temporal filtering
MDC	Multiple description coding
MDLVQ	Multiple description lattice vector quantization
MDSQ	Multiple description scalar quantization
MMDSQ	Modified multiple description scalar quantization
MDVC	Multiple description video coding
PSNR	Peak signal to noise ratio
QTCQ	Quadtree classification and trellis coded quantization
QP	Quantization parameter
SPSD	Sum of pairwise squared distances
TCE	Tarp filter with classification
TCQ	Trellis coded quantization
2DLVQ/3DLVQ	2/3-description lattice vector quantization
2DVC/4DVC/8DVC	2/4/8-description video coding

Table D.2: List of Symbols

SYMBOL	DESCRIPTION
L	Number of dimensions
M	Number of descriptions
N	Index number of sublattices Λ_s (Chapter 3) or number of bins (Chapter 4)
X	Scalar random process or L-dimensional random vector
x	L-dimensional vector (realization of X)
$D(R)$	Distortion-rate function
$R(D)$	Rate-distortion function
D_k	k-description (averaged) side distortion
D_{exp}	Overall expected distortion
$h(\cdot)$	Differential entropy
$I(\cdot;\cdot)$	Mutual information
$E[\cdot]$	Statistical expectation operator
$Q(\cdot)$	Quantization operator
$Q^{-1}(\cdot)$	Dequantization operator
$O(\cdot)$	Order of computational or time complexity
Λ	Lattice (or fine lattice)
Λ_s	Sublattice
$\Lambda_{s/i}$	i-fraction lattice of sublattice Λ_s
Λ_π	Product lattice
λ	Lattice point of fine lattice Λ
λ_i	Sublattice point of sublattice Λ_s
$\lambda_{s/i}$	Lattice point of i-fraction lattice $\Lambda_{s/i}$
$G(\Lambda)$	Dimensionless normalized second moment of lattice Λ
G_L	Dimensionless normalized second moment of L-sphere
V_0	Voronoi cell of Λ_s
v	Volume of Voronoi cell of lattice Λ
v_s	Volume of Voronoi cell of sublattice Λ_s
w_L	Volume of unit L-sphere
A_2	Hexagonal two-dimensional lattice
D_4	Four dimensional (checker board) lattice
Z^1	Scalar lattice (uniform lattice)
Z^2	Square lattice
Z^L	Hypercubic lattice
p_l	Description or packet loss probability

Publications

Journal Papers

[1] Minglei Liu and Ce Zhu, "Index Assignment for 3-description Lattice Vector Quantization Based on A_2 Lattice", *Elsevier Signal Processing*, 88(11), pp. 2754-2763, Nov. 2008.

[2] Ce Zhu and Minglei Liu, "Multiple Description Video Coding Based on Hierarchical B Pictures", *IEEE Transactions on Circuits and Systems for Video Technology*, 19(4), pp. 511-521, Apr. 2009.

[3] Minglei Liu and Ce Zhu, "Enhancing Two-Stage Multiple Description Scalar Quantization", *IEEE Signal Processing Letters*, 16(4), pp. 253-256, Apr. 2009.

[4] Minglei Liu and Ce Zhu, "M-Description Lattice Vector Quantization: Index Assignment and Analysis", *IEEE Transactions on Signal Processing*, 57(6), pp. 2258-2274, Jun. 2009.

[5] Wei Xiang, Ce Zhu, Chee Kheong Siew, Yuanyuan Xu and Minglei Liu, "FEC-Based Two-Dimensional Layered Multiple Description Coding for Error Resilient H.264/SVC Video Transmission", *IEEE Transactions on Circuits and Systems for Video Technology*, 19(12), pp. 1730-1738, Dec. 2009.

Book Chapter

[1] Minglei Liu and Ce Zhu, "Multiple Description Coding with Application in Multimedia Watermarking", in *Handbook of Research on Secure Multimedia Distribution* edited by Shiguo Lian and Yan Zhang, IGI Global Publisher, USA, 2009.

Conference Papers

[1] Minglei Liu, Ce Zhu and Xiaolin Wu, "Index Assignment Design for Three-Description Lattice Vector Quantization", *IEEE International Symposium on Circuits and Systems*, 21-24 May 2006, pp. 3101-3104, Island of Kos, Greece.

[2] Minglei Liu and Ce Zhu, "Multiple Description Video Coding Using Hierarchical B Pictures", *IEEE International Conference on Multimedia and Expo*, 2-5 July 2007, pp. 1367-1370, Beijing, China.

Bibliography

[1] X. Huang, "Multiple description lattice vector quantization," (M.A.Sc. Thesis, Dept. of Electrical & Computer Engineering, McMaster University, Canada), June 2006.

[2] D. Wang, N. Canagarajah, D. Redmill, and D. Bull, "Multiple description video coding based on zero padding," in *Proc. IEEE International Symposium on Circuits and Systems 2004 (ISCAS'04)*, vol. 2, (Vancouver, Canada), pp. 205–208, May 2004.

[3] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding with three motion compensation loops," in *Proc. IEEE International Symposium on Circuits and Systems 2005 (ISCAS'05)*, (Kobe, Japan), pp. 960–963, May 2005.

[4] D. Wang, N. Canagarajah, D. Agrafiotis, and D. Bull, "Error concealment for slice group based multiple description video coding," in *Proc. IEEE International Conference on Image Processing 2005 (ICIP'05)*, pp. 769–772, Sept. 2005.

[5] T. Tillo, M. Grangetto, and G. Olmo, "Redundant slice optimal allocation for h.264 multiple description coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18(1), pp. 59–70, Jan. 2008.

[6] A. R. Reibman, H. Jafarkhani, M. T. Orchard, and Y. Wang, "Performance of multiple description coders on a real channel," in *Proc. IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'99)*, vol. 5, pp. 2415–2418, Mar. 1999.

[7] Y. Wang, S. Panwar, S. Lin, and S. Mao, "Wireless video transport using path diversity: multiple description vs layered coding," in *Proc. IEEE International Conference on Image Processing 2002 (ICIP'02)*, vol. 1, pp. 21–24, Sept. 2002.

[8] Y. C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded vs. multiple description coded video over error-prone networks," *Elsevier Signal Processing: Image Communication*, vol. 18(5), pp. 337–356, May 2003.

[9] A. Vitali, A. Borneo, M. Fumagalli, and R. Rinaldo, "Video over IP using standard-compatible multiple description coding: an IETF proposal," in *Proc. Workshop Packet video*, vol. 7(5), (Hangzhou, China), pp. 668–676, Apr. 2006.

[10] A. Chou, J. Wang, and N. Padmanabhan, "Layered multiple description coding," in *Proc. Workshop Packet video*, 2003.

[11] V. Stankovic and Z. Xiong, "Packet erasure protection for multicasting," in *Proc. IEEE Data Compression Conference 2004 (DCC'04)*, pp. 142–151, Mar. 2004.

[12] W. Xiang, C. Zhu, C. K. Siew, Y. Xu, and M. Liu, "FEC-based two-dimensional layered multiple description coding for error resilient H.264/SVC video transmission," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 19(12), pp. 1730–1738, Dec. 2009.

[13] V. A. Vaishampayan, "Design of multiple description scalar quantizer," *IEEE Trans. Information Theory*, vol. 39, pp. 821–834, May 1993.

[14] S. D. Servetto, V. A. Vaishampayan, and N. J. A. Sloane, "Multiple description lattice vector quantization," in *Proc. IEEE Data Compression Conference 1999 (DCC'99)*, (Snowbird, UT), pp. 13–22, Mar. 1999.

[15] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Information Theory*, vol. 47, pp. 1718–1734, July 2001.

[16] Y. Frank-Dayan and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Trans. Information Theory*, vol. 48(1), pp. 192–204, Jan. 2002.

[17] C. Tian and S. S. Hemami, "A new class of multiple description scalar quantizer and its application to image coding," *IEEE Signal Processing Letters*, vol. 12(4), pp. 329–332, Apr. 2005.

[18] C. Lin, Y. Zhao, and C. Zhu, "Two-stage diversity-based multiple description image coding," *IEEE Signal Processing Letters*, vol. 15, pp. 837–840, 2008.

[19] J. G. Apostolopoulos, "Error-resilient video compression through the use of multiple states," in *Proc. IEEE International Conference on Image Processing 2000 (ICIP'00)*, vol. 3, pp. 352–355, Sept. 2000.

[20] G. Zhang and R. L. Stevenson, "Efficient error recovery for multiple description video coding," in *Proc. IEEE International Conference on Image Processing 2004 (ICIP'04)*, vol. 2, pp. 829–832, Oct. 2004.

[21] C. E. Shannon, "A mathematical theory of communication," *The Bell Syst. Tch. J.*, pp. 27:379–423; 623–656, Jul. and Oct. 1948.

[22] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Conv. Rec.*, vol. 7, pp. 142–163, 1959.

[23] T. Cover and J. Thomas, "Elements of information theory," *NY: Harper and Row*, 1991.

[24] T. Berger, "Rate distortion theory: A mathematical basis for data compression," *Englewood Cliffs: Prentice-Hall*, 1971.

[25] T. Linder and R. Zamir, "On the asymptotic tightness of the shannon lower bound," *IEEE Trans. Information Theory*, vol. 40(6), pp. 2026–2031, Nov. 1994.

[26] A. A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Information Theory*, vol. 28(6), pp. 851–857, Nov. 1982.

[27] L. H. Ozarow, "On a source coding problem with two channels and three receivers," *The Bell Syst. Tch. J.*, vol. 59, pp. 1909–1921, Dec. 1980.

[28] R. Ahlswede, "The rate-distortion region for multiple description without excess rate," *IEEE Trans. Information Theory*, vol. 31, pp. 721–726, Nov. 1985.

[29] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Information Theory*, vol. 33, pp. 502–521, July 1987.

[30] J. Otergaard, "Multiple description lattice vector quantization," (Ph.D. Thesis, Delft University of Technology, Netherlands), June 2007.

[31] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *The Bell Syst. Tch. J.*, vol. 59, pp. 1417–1426, Oct. 1980.

[32] H. Witsenhausen, "On source networks with minimal breakdown degradation," *The Bell Syst. Tch. J.*, vol. 59(6), pp. 1083–1087, Jul.-Aug. 1980.

[33] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Information Theory*, vol. 41(2), pp. 349–357, Mar. 1995.

[34] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Information Theory*, vol. 45(7), pp. 2629–2636, Nov. 1999.

[35] R. Zamir, "Shannon type bounds for multiple descriptions of a stationary source," *Journal of Combinatorics, Information and System Sciences*, pp. 1–15, Dec. 2000.

[36] T. Linder, R. Zamir, and K. Zeger, "The multiple description rate region for high resolution source coding," in *Proc. IEEE Data Compression Conference 1998 (DCC'98)*, (Snowbird, UT), pp. 149–158, Mar.-Apr. 1998.

[37] H. Feng and M. Effros, "On the rate loss of multiple description source codes," *IEEE Trans. Information Theory*, vol. 51(2), pp. 671–683, Feb. 2005.

[38] L. A. Lastras-Montano and V. Castelli, "Near sufficiency of random coding for two descriptions," *IEEE Trans. Information Theory*, vol. 52(2), pp. 681–695, Feb. 2006.

[39] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. Information Theory*, vol. 52, pp. 5197–5217, Dec. 2006.

[40] V. A. Vaishampayan and J. C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. Information Theory*, vol. 44(1), pp. 278–284, Jan. 1998.

[41] V. A. Vaishampayan, A. R. Calderbank, and J.-C. Batllo, "On reducing granular distortion in multiple description quantization," in *Proc. IEEE International Symposium on Information Theory 1998*, p. 98, Aug. 1998.

[42] R. Venkataramani, G. Kramer, and V. K. Goyal, "Bounds on the achievable region for certain multiple description coding problems," in *Proc. IEEE Int. Symp. Information Theory*, (Washington, DC), p. 148, June 2001.

[43] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Information Theory*, vol. 49(9), pp. 2106–2114, Sept. 2003.

[44] S. S. Pradhan, R. Puri, and K. Ramchandran, "(n, k) source-channel erasure codes: Can parity bits also refine quality?," in *Conference on Information Sciences and Systems*, Mar. 2001.

[45] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel multiple descriptions: Theory and constructions," in *Proc. IEEE Data Compression Conference 2002 (DCC'02)*, pp. 262–271, Mar. 2002.

[46] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions: New rate regions," in *Proc. IEEE Int. Symp. Information Theory*, p. 93, June 2002.

[47] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions part I: (n, k) source-channel erasure codes," *IEEE Trans. Information Theory*, vol. 50(1), pp. 47–61, Jan. 2004.

[48] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions- part ii: An achievable rate-distortion region," *IEEE Trans. Information Theory*, vol. 51(4), pp. 1377–1392, Apr. 2005.

[49] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, pp. 74–93, Sept. 2001.

[50] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, pp. 57–70, Jan. 2005.

[51] T. Y. Berger-Wolf and E. M. Reingold, "Index assignment for multichannel communication under failure," *IEEE Trans. Information Theory*, vol. 48(10), pp. 2656–2668, Oct. 2002.

[52] C. Tian and S. S. Hemami, "Sequential design of multiple description scalar quantizers," in *Proc. IEEE Data Compression Conference 2004 (DCC'04)*, pp. 32–41, Mar. 2004.

[53] C. Tian and S. S. Hemami, "Universal multiple description scalar quantizer: analysis and design," *IEEE Trans. Information Theory*, vol. 50, pp. 2089–2102, Sept. 2004.

[54] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple-description scalar quantizers," *IEEE Trans. Information Theory*, vol. 40(1), pp. 245–250, Jan. 1994.

[55] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," *IEEE Trans. Communications*, vol. 47, pp. 799–803, June 1999.

[56] X. Wang and M. T. Orchard, "Multiple description coding using trellis coded quantization," in *Proc. IEEE International Conference on Image Processing 2000 (ICIP'00)*, vol. 1, pp. 391–394, Sept. 2000.

[57] J. Cardinal, "Entropy-constrained index assignments for multiple description quantizers," *IEEE Trans. Signal Processing*, vol. 52(1), pp. 265–270, Jan. 2004.

[58] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "A new class of universal multiple description lattice quantizers," in *Proc. IEEE Int. Symp. Information Theory*, pp. 1803–1807, Sept. 2005.

[59] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Design of asymmetric multiple description lattice vector quantizers," in *Proc. IEEE Data Compression Conference 2000 (DCC'00)*, pp. 490–499, Mar. 2000.

[60] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Information Theory*, vol. 48(1), pp. 174–191, Jan. 2002.

[61] M. Fleming and M. Effros, "Generalized multiple description vector quantization," in *Proc. IEEE Data Compression Conference 1999 (DCC'99)*, (Snowbird, UT), p. 3, Mar. 1999.

[62] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Information Theory*, vol. 50(8), pp. 1584–1604, Aug. 2004.

[63] N. Gortz and P. Leelapornchai, "Optimization of the index assignments for multiple description vector quantizers," *IEEE Trans. Communications*, vol. 51(3), pp. 336–340, Mar. 2003.

[64] V. K. Goyal, J. A. Kelner, and J. Kovacevic, "Multiple description vector quantization with a coarse lattice," *IEEE Trans. Information Theory*, vol. 48, pp. 781–788, Mar. 2002.

[65] J. A. Kelner, V. K. Goyal, and J. Kovacevic, "Multiple description lattice vector quantization: variations and extensions," in *Proc. IEEE Data Compression Conference 2000 (DCC'00)*, pp. 480–489, Mar. 2000.

[66] P. Koulgi, S. L. Regunathan, and K. Rose, "Multiple description quantization by deterministic annealing," *IEEE Trans. Information Theory*, vol. 49(8), pp. 2067–2075, Aug. 2003.

[67] J. Ostergaard, R. Heusdens, and J. Jensen, "n-channel asymmetric multiple description lattice vector quantization," in *Proc. IEEE Int. Symp. Information Theory*, pp. 1793–1797, Sept. 2005.

[68] J. Ostergaard, J. Jensen, and R. Heusdens, "Entropy constrained multiple description lattice vector quantization," in *Proc. IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'04)*, vol. 4, (Montreal, QC, Canada), pp. 601–604, May 2004.

[69] J. Ostergaard, J. Jensen, and R. Heusdens, "n-channel symmetric multiple-description lattice vector quantization," in *Proc. IEEE Data Compression Conference 2005 (DCC'05)*, pp. 378–387, Mar. 2005.

[70] J. Ostergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Information Theory*, vol. 52, pp. 1956–1973, May 2006.

[71] J. Otergaard and R. Zamir, "Multiple-description coding by dithered delta-sigma quantization," in *Proc. IEEE Data Compression Conference 2007 (DCC'07)*, pp. 63–72, Mar. 2007.

[72] C. Tian and S. S. Hemami, "Optimality and suboptimality of multiple description vector quantization with a lattice codebook," *IEEE Trans. Information Theory*, vol. 50(10), pp. 2458–2470, Oct. 2004.

[73] P. Yahampath, "On index assignment and the design of multiple description quantizers," in *Proc. IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'04)*, vol. 4, (Montreal, QC, Canada), pp. 597–600, May 2004.

[74] M. Liu, C. Zhu, and X. Wu, "Index assignment design for three-description lattice vector quantization," in *Proc. IEEE International Symposium on Circuits and Systems 2006 (ISCAS'06)*, pp. 3101–3104, May 2006.

[75] M. Liu and C. Zhu, "Index assignment for 3-description lattice vector quantization based on A_2 lattice," *Elsevier Signal Processing*, vol. 88(11), pp. 2754–2763, Nov. 2008.

[76] M. Liu and C. Zhu, "Enhancing two-stage multiple description scalar quantization," *IEEE Signal Processing Letters*, vol. 16(4), pp. 253–256, Apr. 2009.

[77] M. Liu and C. Zhu, "M-description lattice vector quantization: Index assignment and analysis," *IEEE Trans. Signal Processing*, vol. 57(6), pp. 2258–2274, June 2009.

[78] J. H. Conway and N. J. A. Sloane, "Sphere packings, lattices and groups," (3rd ed. New York: Springer-Verlag), 1999.

[79] X. Huang and X. Wu, "Optimal index assignment for multiple description lattice vector quantization," in *Proc. IEEE Data Compression Conference 2006 (DCC'06)*, pp. 272–281, Mar. 2006.

[80] A. Ingle and V. A. Vaishampayan, "DPCM system design for diversity systems with applications to packetized speech," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 48–57, Jan. 1995.

[81] S. Wenger, "Video redundancy coding in H.263+," in *presented at the Audio-Visual Services Over Packet Networks Conf.*, (Aberdeen, U.K.), 1997.

[82] D.-M. Chung and Y. Wang, "Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 9(6), pp. 895–908, Sept. 1999.

[83] D.-M. Chung and Y. Wang, "Lapped orthogonal transforms designed for error resilient image coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 752–764, Sept. 2002.

[84] C. Kim and S. Lee, "Multiple description coding of motion fields for robust video transmission," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, pp. 999–1010, Sept. 2001.

[85] J. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. Visual Communications Image Processing*, pp. 392–409, 2001.

[86] Y. Wang and S. Lin, "Error-resilient video coding using multiple description motion compensation," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 438–452, June 2002.

[87] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over ip," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, pp. 321–334, Mar. 2005.

[88] I. V. Bajic and J. W. Woods, "Domain-based multiple description coding of images and video," *IEEE Trans. Image Processing*, vol. 12, pp. 1211–1225, Oct. 2003.

[89] C. Zhu and M. Liu, "Multiple description video coding based on hierarchical B pictures," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 19(4), pp. 511–521, Apr. 2009.

[90] S. Cho and W. A. Pearlman, "A full-featured, error resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, pp. 157–170, Mar. 2002.

[91] V. K. Goyal, J. Kovacevic, and J. Kelner, "Quantized frame expansions with erasures," *Journal of Appl. and Comput. Harmonic Analysis*, vol. 10(3), pp. 203–233, May 2001.

[92] V. K. Goyal and J. Kovacevic, "Generalized multiple descriptions coding with correlating transforms," *IEEE Trans. Information Theory*, vol. 47(6), pp. 2199–2224, Sept. 2001.

[93] V. K. Goyal, J. Kovacevic, and M. Vetterli, "Multiple description transform coding: Robustness to erasures using tight frame expansions," in *Proc. IEEE International Symposium on Information Theory 1998*, p. 408, Aug. 1998.

[94] M. T. Orchard, Y. Wang, V. A. Vaishampayan, and A. R. Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms," in *Proc. IEEE International Conference on Image Processing 1997 (ICIP'97)*, vol. 1, pp. 608–611, 1997.

[95] Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by pairing transform coefficients," in *Proc. IEEE Workshop on Multimedia Signal Processing*, (Princeton, NJ), pp. 419–424, June 1997.

[96] P. A. Chou, S. Mehrotra, and A. Wang, "Multiple description decoding of overcomplete expansions using projections onto convex sets," in *Proc. IEEE Data Compression Conference 1999 (DCC'99)*, (Snowbird, UT), pp. 72–81, Mar. 1999.

[97] V. K. Goyal, J. Kovacevic, and M. Vetterli, "Quantized frame expansions as source channel codes for erasure channels," in *Proc. IEEE Data Compression Conference 1999 (DCC'99)*, (Snowbird, UT), pp. 326–355, Mar. 1999.

[98] R. Balan, I. Daubechies, and V. Vaishampayan, "The analysis and design of windowed fourier frame based multiple description source coding schemes," *IEEE Trans. Information Theory*, vol. 46(7), pp. 2491–2536, Nov. 2000.

[99] P. L. Dragotti, J. Kovacevic, and V. K. Goyal, "Quantized oversampled filter banks with erasures," in *Proc. IEEE Data Compression Conference 2001 (DCC'01)*, (Snowbird, UT), pp. 173–182, Mar. 2001.

[100] J. Kovacevic, P. L. Dragotti, and V. K. Goyal, "Filter bank frame expansions with erasures," *IEEE Trans. Information Theory*, vol. 48(6), pp. 1439–1450, June 2002.

[101] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Processing*, vol. 10, pp. 351–366, Mar. 2001.

[102] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Information Theory*, vol. 42(6), pp. 1737–1744, Nov. 1996.

[103] G. Davis and J. Danskin, "Joint source and channel coding for image transmission over lossy packet networks," in *Proc. SPIE Conf. Wavelet Applicat. Digital Image Process XIX*, vol. 2847, pp. 376–387, 1996.

[104] R. Puri and K. Ramchandran, "Multiple description source coding using forward error correction codes," in *Proc. 33rd Asilomar Conf. on Signals, Systems and Computers 1999*, vol. 1, (Pacific Grove, CA), pp. 342–346, Oct. 1999.

[105] N. S. Jayant, "Subsampling of a dpcm speech channel to provide two 'self-contained' half-rate channels," *Bell Syst. Tech. J.*, vol. 60(4), pp. 501–509, Apr. 1981.

[106] G. Kubin and W. B. Kleijn, "Multiple-description coding (mdc) of speech with an invertible auditory model," in *Proc. IEEE Workshop Speech Coding*, (Porvoo, Finland), pp. 81–83, Jun. 1999.

[107] W. Jiang and A. Ortega, "Multiple description speech coding for robust communication over lossy packet networks," in *Proc. IEEE International Conference on Multimedia & Expo 2000 (ICME'00)*, vol. 1, (New York, NY), pp. 444–447, Jul.-Aug. 2000.

[108] R. Arean, J. Kovacevic, and V. Goyal, "Multiple description perceptual audio coding with correlating transformsh," *IEEE Trans. Speech Audio Processing*, vol. 8(2), pp. 140–145, Mar. 2000.

[109] J. Balam and J. D. Gibson, "Multiple descriptions and path diversity for voice communications over wireless mesh networks," *IEEE Trans. Multimedia*, vol. 9(5), pp. 1703–1088, Aug. 2007.

[110] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. SPIE Conf. Visual Commun. and Image Processing*, vol. 3653, (San Jose, CA), pp. 998–1008, 1999.

[111] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Processing*, vol. 9(5), pp. 813–826, May 2000.

[112] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proc. IEEE International Conference on Image Processing 1999 (ICIP'99)*, vol. 3, (Kobe, Japan), pp. 842–846, Oct. 1999.

[113] T. Tillo and G. Olmo, "A novel multiple description coding scheme compatible with the JPEG2000 decoder," *IEEE Signal Processing Letters*, vol. 11(11), pp. 908–911, Nov. 2004.

[114] A. E. Mohr, E. A. Riskin, and R. E. Ladner, "Generalized multiple description coding through unequal loss protection," in *Proc. IEEE International Conference on Image Processing 1999 (ICIP'99)*, vol. 1, (Kobe, Japan), pp. 411–415, Oct. 1999.

[115] N. Varnica, M. Fleming, and M. Effros, "Multi-resolution adaptation of the SPIHT algorithm for multiple description," in *Proc. IEEE Data Compression Conference 2000 (DCC'00)*, (Snowbird, UT), pp. 303–312, Mar. 2000.

[116] D. G. Sachs, A. Raghavan, and K. Ramchandran, "Wireless image transmission using multiple-description based concatenated codes," in *Proc. SPIE Image Video Processing*, vol. 3794, (San Jose, CA), pp. 300–311, Jan. 2000.

[117] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli, "Multiple description transform coding of images," in *Proc. IEEE International Conference on Image Processing 1998 (ICIP'98)*, vol. 1, (Chicago, IL), pp. 674–678, Oct. 1998.

[118] W. S. Lee, M. R. Pickering, M. R. Frater, and J. F. Arnold, "A robust codec for transmission of very low bit-rate video over channels with bursty errors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, pp. 1403–1412, Dec. 2000.

[119] C.-S. Kim and S.-U. Lee, "Multiple description motion coding algorithm for robust video transmission," in *Proc. IEEE International Symposium on Circuits and Systems 2004 (ISCAS'04)*, vol. 4, (Lausanne, Switzerland), pp. 717–720, May 2000.

[120] V. A. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proc. IEEE International Conference on Image Processing 1999 (ICIP'99)*, vol. 3, (Kobe, Japan), pp. 24–28, Oct. 1999.

[121] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," in *Proc. IEEE International Conference on Image Processing 1999 (ICIP'99)*, vol. 3, (Kobe, Japan), pp. 837–841, Oct. 1999.

[122] A. Reibman, H. Jafarkhani, Y. Wang, and M. Orchard, "Multiple description video using rate-distortion splitting," in *Proc. IEEE International Conference on Image Processing 2001 (ICIP'01)*, vol. 1, pp. 978–981, Oct. 2001.

[123] A. R. Reibman, H. Jafarkhani, Y. Wang, M. T. Orchard, and R. Puri, "Multiple description coding for video using motion compensated prediction," *IEEE Trans. Circuits and Systems for Video Technology*, pp. 193–204, Mar. 2002.

[124] S. D. Servetto and K. Nahrstedt, "Video streaming over the public internet: Multiple description codes and adaptive transport protocols," in *Proc. IEEE International Conference on Image Processing 1999 (ICIP'99)*, vol. 3, (Kobe, Japan), pp. 85–89, Oct. 1999.

[125] S. D. Servetto and K. Nahrstedt, "Broadcast-quality video over IP," *IEEE Trans. Multimedia*, vol. 3, pp. 162–173, Mar. 2001.

[126] D. Wang, N. Canagarajah, and D. Bull, "Slice group based multiple description video coding using motion vector estimation," in *Proc. IEEE International Conference on Image Processing 2004 (ICIP'04)*, pp. 3237–3240, Oct. 2004.

[127] M. Liu and C. Zhu, "Multiple description video coding using hierarchical B pictures," in *Proc. IEEE ICME'07*, (Beijing, China), pp. 1367–1370, July 2007.

[128] M. Van der Schaar and D. S. Turaga, "Multiple description scalable coding using wavelet-based motion compensated temporal filtering," in *Proc. IEEE International Conference on Image Processing 2003 (ICIP'03)*, vol. 2, pp. 489–492, Sept. 2003.

[129] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, pp. 559–571, Sept. 1994.

[130] E. Akyol, A. M. Tekalp, and M. R. Civanlar, "Scalable multiple description video coding with flexible number of descriptions," in *Proc. IEEE International Conference on Image Processing 2005 (ICIP'05)*, vol. 3, pp. 712–715, Sept. 2005.

[131] M. Tesanovic, D. Bull, A. Doufexi, and A. Nix, "Enhanced MIMO wireless coomunication using multiple-description coding," *Elsevier Signal Processing: Image Communication*, vol. 23(4), pp. 325–336, Apr. 2008.

[132] T. Schierl, K. Ganger, C. Hellge, T. Wiegand, and T. Stockhammer, "SVC-based multi-source streaming for robust video transmission in mobile ad hoc networks," *IEEE Wireless Communications Magazine*, vol. 13(5), pp. 96–103, Oct. 2006.

[133] D. Kundur, "Watermarking with diversity: Insights and implications," *IEEE Multimedia*, vol. 8(4), pp. 46–52, 2001.

[134] R. Chandramouli, B. Graubard, and C. Richmond, "A multiple description framework for oblivious watermarking," in *Proc. SPIE Security and Watermarking for Multimedia Contents*, vol. 3, pp. 585–593, 2001.

[135] Y. Hsia, C. Chang, and J. Liao, "Multiple-description coding for robust image watermarking," in *Proc. IEEE International Conference on Image Processing 2004 (ICIP'04)*, vol. 4, pp. 2163–2166, Oct. 2004.

[136] S. Chu, Y. Hsin, H. Huang, K. Huang, and J. Pan, "Multiple description watermarking for lossy network," in *Proc. IEEE International Symposium on Circuits and Systems 2004 (ISCAS'04)*, (Vancouver, Canada), pp. 3990–3993, May 2004.

[137] M. Day, S. Lee, and I. Jou, "Robust multirate lattice quantization index modulation watermarking resilient to multiple description transmission channel," *SPIE Optical Engineering*, vol. 46(3), p. 037005, 2007.

[138] M. Liu and C. Zhu, "Multiple description coding with application in multimedia watermarking," (USA: IGI Global), pp. 495–516, 2009.

[139] R. M. Gray, "Source coding theory," *Kluwer Academic Publishers*, 1990.

[140] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Information Theory*, vol. 44(6), pp. 2325–2383, Oct. 1998.

[141] T. Linder and R. Zamir, "High-resolution rate-distortion theory," in *Proc. IEEE Workshop Information Theory and Networking*, p. 31, Jun. 1999.

[142] H. Bai, C. Zhu, and Y. Zhao, "Optimized multiple description lattice vector quantization for wavelet image coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17(7), pp. 912–917, July 2007.

[143] D. B. West, "Introdcution to graph theory," (Englewood Cliffs, NY: Prentice-Hall), 2001.

[144] F. S. Hillier and G. J. Lieberman, "Introduction to operations research," (7th ed. New York: McGraw-Hill), pp. 354–355, 2001.

[145] J. Orlin, "A faster strongly polynomial minimum cost algorithm," (Proc. 20th ACM STOC), pp. 377–387, 1988.

[146] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quaterly*, vol. 2, pp. 83–97, 1955.

[147] J. Hopcroft and R. Karp, "An $O(n^{5/2})$ algorithm for maximum matchings in bipartite graphs," *SIAM Journal of Computing*, vol. 2(4), pp. 225–231, 1973.

[148] T. Tokuyama and J. Nakano, "Efficient algorithms for the hitchcock transportation problem," *SIAM Journal of Computing*, vol. 24, pp. 563–578, June 1995.

[149] A. Said and W. A. Pearlman, "A new fast and efficient codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, pp. 243–250, June 1996.

[150] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG2000 still image compression standard," *IEEE Signal Processing Mag.*, vol. 18, pp. 36–58, Sept. 2001.

[151] C. Tian and S. S. Hemami, "An embedded image coding system based on tarp filter with classification," in *Proc. IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'04)*, vol. 3, (Montreal, QC, Canada), pp. iii–49–52, May 2004.

[152] B. Banister and T. R. Fischer, "Quadtree classification and TCQ image coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11(1), pp. 3–8, Jan. 2001.

[153] C. Tian and S. S. Hemami, "Staggered lattices in multiple description quantization," in *Proc. IEEE Data Compression Conference 2005 (DCC'05)*, pp. 398–407, Mar. 2005.

[154] X. Zhong and B.-H. Juang, "Multiple description speech coding with diversities," in *Proc. IEEE Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP'02)*, vol. 1, (Atlanta, USA), pp. 177–180, 2002.

[155] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Communications*, vol. 38, pp. 82–93, Jan. 1990.

[156] ITU-T Rec. & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," 2005.

[157] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *Proc. IEEE International Conference on Multimedia & Expo 2006 (ICME'06)*, pp. 1929 – 1932, July 2006.

[158] MPEG Committee & ISO/IEC JTC 1/SC 29/WG 11 N8243, "JSVM 7 software," 2006.

[159] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," *Internet Engineering Task Force—-RFC 1889*, Jan. 1996.

[160] H. C. Bergmann, "Motion adaptive frame interpolation," in *Proc. International Zurich Seminar on Digital Communications*, vol. D, pp. 21–25, Mar. 1984.

[161] J. Ostergaard, R. Heusdens, and J. Jensen, "Source-channel erasure codes with lattice codebooks for multiple description coding," in *Proc. IEEE Int. Symp. Information Theory*, pp. 2324–2328, July 2006.

[162] G. Xie, M. N. S. Swamy, and M. O. Ahmad, "Joint optimal multipath routing and rate control for multidescription coded video streaming in Ad Hoc networks," *IEEE Trans. Multimedia*, vol. 10(8), pp. 1687–1697, Dec. 2008.

[163] R. L. Graham, D. E. Knuth, and O.Patashnik, "Concrete mathematics," (2nd ed. Reading, MA: Addison-Wesley), 1994.

[164] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," in *IEEE J. Sel. Areas Commun.*, vol. 18(6), pp. 966–976, Jun. 2000.

[165] R. Zhang, S. L. Regunathan, and K. Rose, "End-to-end distortion estimation for RD-based robust delivery of pre-compressed video," in *Proc. 35th Asilomar Confe. Signals, Syst. Comput.*, pp. 210–214, Nov. 2001.

[166] Z. He, H. Cai, and C. W. Chang, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12(6), pp. 511–523, June 2002.

[167] Y. Wang, Z. Wu, and J. M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16(6), pp. 716–732, June 2006.

[168] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13(7), pp. 688–703, July 2003.