MULTIPLE DESCRIPTION CODING WITH PREDICTION COMPENSATION

by

Guoqian Sun

B.Sc., Southeast University, 2000

M.Sc., Southeast University, 2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTERS OF APPLIED SCIENCE
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
Spring 2009

APPROVAL

Name: Guoqian Sun

Degree: Masters of Applied Science

Title of thesis: MULTIPLE DESCRIPTION CODING WITH PREDICTION COMPENSATION

Examining Committee:

Dr. Mirza Faisal Beg
Assistant Professor, Engineering Science
Chair

Dr. Jie Liang
Assistant Professor, Engineering Science
Senior Supervisor

Dr. Daniel Lee
Associate Professor, Engineering Science
Supervisor

Dr. Sami Muhaidat
Assistant Professor, Engineering Science
Internal Examiner

Date Approved: January 16, 2009

Abstract

Multiple description coding (MDC) is an attractive diversity technique of combating transmission errors, where several compressed bit streams (descriptions) are generated, which can be transmitted via different paths. Judiciously designed redundancies are introduced in all bit streams such that the reconstruction quality degrades gracefully when some of them are lost.

A new multiple description coding paradigm is proposed in this dissertation, by combining the time domain lapped transform, block level source splitting, inter-description prediction, and coding of the prediction residual. The joint design of all system components is developed, and the asymptotic performance in the DPCM case for first-order Gauss-Markov sources is analyzed. Image coding results show that the proposed method can significantly outperform other methods in the literature. The effects of different transforms and the size of the prediction filter are also studied.

A number of generalizations of the proposed scheme is studied as well. First, the two dimensional Wiener filter is introduced into this scheme. These filters can improve the performance when the redundancy is low. Secondly, the proposed scheme is extended to three-description and four-description cases. The optimal design and image coding results are reported.

Acknowledgments

I am sincerely grateful to my senior supervisor, Dr. Jie Liang, for his supervision, advice, patience and encouragement through the course of my MASc studies. I am deeply indebted to Dr. Chao Tian and Dr. Chengjie Tu for their helps and suggestions on the MMDSQ and entropy coding. I also thank Dr. Mirza Faisal Beg, Dr. Daniel Lee and Dr. Sami Muhaidat for being my committee members.

I would like to thank my labmates, Upul Samarawickrama, Yue-meng Chen, and Jing Wang for their valuable advices and discussions on my research. I would also like to thank all of my friends who have made my time at Simon Fraser University enjoyable, especially Tong Jin who helped me out when I first came to Canada.

Finally, and above all, my heartfelt gratitude goes to my parents, who have always supported my studies and gave me everything in life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Image and video communications are ubiquitous in people's life. However, the storage and transmission of digital images and videos require much more resources and bandwidths than traditional voice and text messages. For examples, a low quality video communication with 352×288 resolution needs 5-10 Mbps bandwidth. The raw high definition TV 1080i30 signal requires a much higher bandwidth of 190 Mbps. With the increasing demand of video applications and the limited bandwidths, the video and image compression technology has gained more and more attentions. Most of the compression techniques can be categorized into two types, lossless and lossy compressions. The lossless compression allows a perfect reconstruction, but it has a theoretical maximum compression ratio, depending on the entropy of the data. The lossy compression allows some extend of quality degradation. It can adjust the quality level and the compression ratio accordingly for different applications. The objective of designing such a system is to find the best quality with a given compression ratio.

The internet has become one of the main communication methods. Millions of video

communications are transmitted over internet everyday. For example, more and more people watch IPTV and YouTube videos, and many companies use video webcast to announces their financial statement. But the internet is not a perfectly reliable transmission media, as it is packet based and could have packets loss. Multimedia applications are extremely sensitive to transmission errors, because of the extensive use of prediction and entropy coding in the compression algorithms. Therefore some kinds of error protection have to be introdcued to the multimedia applications.

One way to fight the packet loss is to retransmit the lost packets. But for real time communications, the re-transmission can introduce unpleasant delay. Sometimes the customers would prefer some quality degradations rather than excessive delay, especially in real-time applications. Another approach to alleviate the impact of packet loss is to add some redundancies into the bitstream, such that when some packets are lost, the decoder can still reconstruct an acceptable result. One way to achieve this is to use error correction coding. However, this increases the complexity of the encoder and the decoder, and may not suitable for some applications.

In this thesis, we are interested in another method, the multiple description coding (MDC), where the content is coded into several descriptions, which can be transmitted through different paths. If some descriptions are corrupted during the transition, the receiver can still reconstruct the source with other available descriptions. The more descriptions are available at the decoder, the better quality can be achieved. The robustness of the system is therefore improved by exploiting the diversity of the transfer paths.

Contrary to the conventional source coding which removes the redundancy from the source, the MDC tries to add back some controlled redundancy into the transmitted bitstream. Since compression is still a requirement, the design of MDC thus needs to strike a balance between the compression efficiency and robustness.

1.2 Main contributions

Most exisitng MDC algorithms are built on top of some conventional compression schemes. The most popular image compression scheme employes block transform coding, where the image are partitioned into small data blocks, which are then decorrelated by a block transform such as the Discrete Cosine Transform (DCT), followed by quantization and entropy coding. The DCT has many advantages for natural images, but at low bitrates, the discontinuities at the block boundaries cause the notorious blocking artifacts. To reduce the blocking artifacts, the wavelet transform and the lapped transform have been developed. In particular, a time-domain lapped transform(TDLT) is proposed in [46], which applies pre- and post-processing to the DCT. The TDLT keeps the block-based infrastructure, and maintains a good trade-off between the complexity and the performance. As a result, it has also been adopted by Microsoft Windows Media Video, the SMPTE VC-1 video coding standard, and the HD-DVD video coding format. It is also being standardized by the JPEG committee as a future image coding standard.

The time domain lapped transform also provides a flexible way to introduce redundancy to the transform coefficients. The amount of redundancy can be controlled by designing lapped transforms with different pre and postfilters. The redundancy can be used to predict the lost coefficients, when some data are missing during transmission. This feature was exploited to design an error resilient system in [25]. However, the prediction-only approach has poor performance at high rate, because of the presence of the prediction residual. In addition, the redundancy is only controlled by the transform. If the change of channel condition requires different redundancy, a different transform has to be used to encode the data, which increases the complexity of the system.

In this thesis, a new MDC framework is proposed, which is based on the time domain lapped transform and prediction compensation. In this scheme, the prefiltered source image is split into two parts. Each part is coded at a high-rate in one description. Each

description also encodes the prediction residual of the other description at a lower-rate. This allows us to control the redundancy of the system without changing the transform. The theoretical performance of this scheme is studied in the Differential Pulse Code Modulation (DPCM) case. We also implement this scheme in the lapped transform framework, and present the design of the corresponding optimal lapped transform. Experiment results show that the performance of the new technique can be significantly better than that of the two-stage modified multiple description scalar quantizer(MMDSQ) [43], which has the best result in the literature.

The prediction used in the scheme above is based on one-dimensional Wiener filter. Better prediction can improve the performance at low redundancy. Therefore, two-dimensional Wiener filter is also investigated in this thesis to enhance the performance of the system. In addition, we also generalize the basic scheme to more than two descritpions, and present the design and image coding results for three-description and four-description coding.

The results presented in this thesis has been published in [40, 24]. A journal paper was accepted in December 2008 and will appear in 2009 [39].

1.3 Thesis outline

Chapter 2 covers the fundamentals of the time domain lapped transform and multiple description coding. Chapter 3 introduces the proposed multiple description coding with prediction compensation. Chapter 4 investigates the application of the two dimensional Wiener filter in the proposed MDC scheme. Chapter 5 presents the design and the coding results for MDC with more than two descriptions. Finally Chapter 6 presents the conclusions and discusses some future works.

Chapter 2

Background

2.1 Time-Domain Lapped Transform

2.1.1 Conventional lapped transform

The discrete cosine transform (DCT) based compression algorithms are used in most image and video coding standards. This is because the DCT is near-optimal for smooth signals, and there are many fast algorithms for it. However, two problems exist in this kind of coding schemes. First, only spatial correlation inside the single block is considered and the correlation between neighboring blocks is not fully exploited. Second, since the blocks are non-overlapping, there may be discontinuities along the boundary regions of the blocks in the reconstruction, which will cause the notorious blocking artifacts, especially at low bit rates.

The lapped transform (LT) is one of the techniques developed to reduce the blocking artifacts in DCT based coding schemes. The LT is a linear transform which first partitions the input signal into overlapped blocks. Unlike the $M \times M$ block transform which maps M input sample to M transform coefficients, the LT is a $M \times L$ linear operator, where $L > M$. In one dimensional(1-D) case, the input signal $\mathbf{x}$ is grouped into overlapping sequences

$\mathbf{x}_m$ of length L, and there is an overlap of $L - M$ samples between neighboring $\mathbf{x}_m$. If the $M \times L$ transform matrix of the LT is written as $\mathbf{H}$, the corresponding transformed sequences $\mathbf{y}_m$ (length of M) is $\mathbf{y}_m = \mathbf{H}\mathbf{x}_m$, so the number of output coefficients is still equal to the number of input samples. That is, the transform is still critically sampled. Typically, L is chosen to be multiple of M, *i.e.*, $L = KM$, where K is called the overlap factor. Therefore each block has an overlap of $(K-1)M$ samples with each of the neighboring blocks.

At the decoder side, we have the $L \times m$ inverse transform matrix $\mathbf{F}$. The reconstructed signal sequences $\hat{\mathbf{x}}_m = \mathbf{F}\mathbf{y}_m$. $\hat{\mathbf{x}}_m$ is also overlapped with adjacent neighbors. Therefore to recover the original signal $\mathbf{x}$, the reconstructed sequences should be added together in a overlapping matter. For example, the last $L - M$ samples of current reconstructed sequence $\hat{\mathbf{x}}_m$ should be added with the first $L - M$ samples of the next reconstructed sequence $\hat{\mathbf{x}}_{m+1}$. For two-dimensional implementation, the transform is applied to each row and then each column separately or vice versa.

In general, more overlap in the lapped transform gives better performance. However, the computation cost also increases, and the quality will not improve too much after a certain value. Considering the trade-off between the complexity and the quality, most of the practical LT use the value $L = 2M$. The polyphase format of the LT can be written as [26]:

$$\mathbf{E}(z) = \frac{1}{2}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & z^{-1}\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}\mathbf{C} \triangleq \mathbf{G}(z)\mathbf{C}, \qquad (2.1)$$

where $\mathbf{C}$ is the M-points DCT matrix, and $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{V}$ is the only free variable in the formula. When the matrix $\mathbf{V}$ is orthogonal, the LT is called the Lapped Orthogonal Transform (LOT), otherwise it is called the Lapped Biorthogonal Transform (LBT). The design of the lapped transform is to find the optimal matrix $\mathbf{V}$ that optimizes some objective functions.

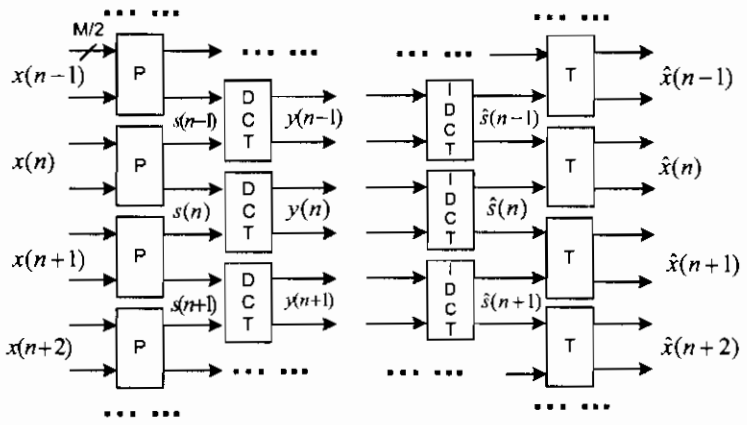


Figure 2.1: Forward and inverse time-domain lapped transforms (TDLT).

2.1.2 Time-domain lapped transform (TDLT)

The conventional lapped transform applies post processing to the DCT. One problem of this scheme is that existing DCT based scheme has to be redesigned to apply the LT. This can be difficult, especially in the hardware. A new structure of the lapped transform is developed in [46], where preprocessing of the DCT is applied. Therefore it is possible to keep the existing DCT based structure intact. Thus the lapped transform based codec can be obtained with minimal modification of the popular DCT based codec.

Fig. 2.1 shows the block diagram of the time-domain lapped transform (TDLT). At the encoder, an $M \times M$ prefilter $\mathbf{P}$ is employed at the boundary of two blocks (M is the block size). The M-point DCT $\mathbf{C}$ is then applied to each block, creating basis functions that cover two blocks. In the decoder, the inverse DCT and postfilter $\mathbf{T}$ at block boundaries are applied. Matrices $\mathbf{P}$ and $\mathbf{T}$ have the following structures to yield near-optimal linear-phase

lapped transform [46]:

$$\mathbf{P} = \mathbf{W}\ \mathrm{diag}\{\mathbf{I}, \mathbf{V}\}\mathbf{W}, \tag{2.2}$$

$$\mathbf{T} = \mathbf{P}^{-1} = \mathbf{W}\ \mathrm{diag}\{\mathbf{I}, \mathbf{V}^{-1}\}\mathbf{W}, \tag{2.3}$$

$$\mathbf{W} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & \mathbf{J} \\ \mathbf{J} & -\mathbf{I} \end{bmatrix}, \tag{2.4}$$

where $\mathbf{I}$ and $\mathbf{J}$ are $\frac{M}{2} \times \frac{M}{2}$ identity matrix and counter-identity matrix, respectively. Matrix $\mathbf{J}$ is obtained by flipping $\mathbf{I}$ horizontally or vertically ([26], pp. 149). The matrix $\mathbf{V}$ is an $\frac{M}{2} \times \frac{M}{2}$ invertible matrix that can be optimized for different purposes. In this thesis, the notation $\mathrm{diag}\{\mathbf{A}, \mathbf{B}\}$ denotes a block diagonal matrix with matrices $\mathbf{A}$ and $\mathbf{B}$ on the diagonal, and zeros elsewhere.

Denote $\mathbf{P}_0$ and $\mathbf{P}_1$ as the first and the last $M/2$ rows of the prefilter $\mathbf{P}$, and $\mathbf{T}_0$ and $\mathbf{T}_1$ as the first and the last $M/2$ columns of $\mathbf{T}$, respectively, *i.e.*,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_0^T & \mathbf{P}_1^T \end{bmatrix}^T, \tag{2.5}$$

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_0 & \mathbf{T}_1 \end{bmatrix}, \tag{2.6}$$

the $M \times 2M$ forward lapped transform $\mathbf{F}$ and $2M \times M$ inverse transform $\mathbf{G}$ can be written as

$$\mathbf{F} = \mathbf{C}\ \mathrm{diag}\{\mathbf{P}_1, \mathbf{P}_0\}. \tag{2.7}$$

$$\mathbf{G} = \mathrm{diag}\{\mathbf{T}_1, \mathbf{T}_0\}\mathbf{C}^T = \mathbf{T}_{21}\mathbf{C}^T. \tag{2.8}$$

The time-domain expression of the $M \times L$ transform can be represented by the $M \times M$

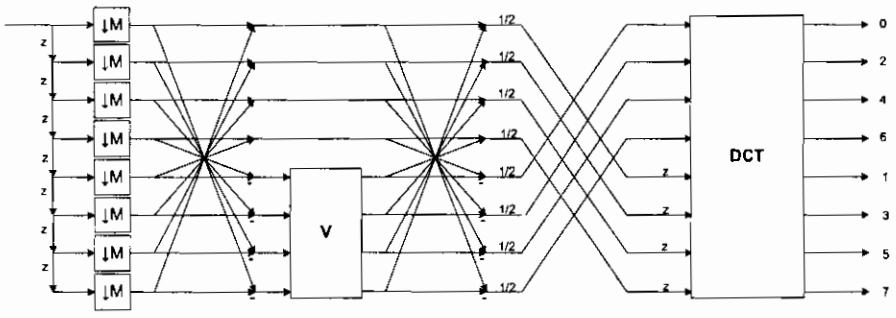


Figure 2.2: Polyphase structure of the time-domain lapped transform.

polyphase matrix as [46]:

$$\mathbf{E}(z) = \mathbf{C} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & z\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P}, \tag{2.9}$$

This is shown in Figure 2.1. Similarly, we can also represent the inverse transform by a $M \times M$ polyphase matrix.

2.1.3 Coding gain

To obtain the best performance of the system, one need to select the design objective, which depends on the applications of interest. For data compression, the coding gain is a very important factor for designing a transform. The coding gain is a measure of the reconstruction quality improvement of the codec over the Pulse Coding modulation (PCM) based scheme, with the same bitrate. The PCM coding is a very simple scheme, where the input signal is directly quantized to a specified bit rate. To get a higher coding gain, a transform should compact most of the energy into a few number of coefficients. In the TDLT, the prefilter acts as a flattering operator [46], which packs more energies to the low-frequency coefficients than the DCT.

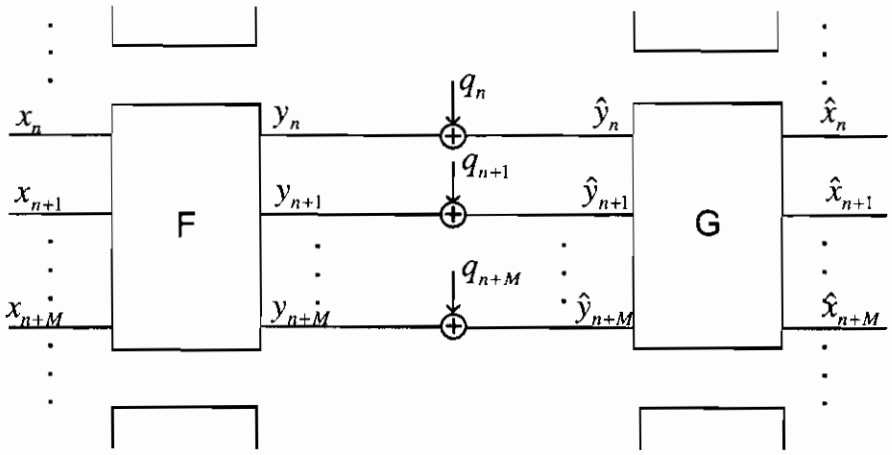


Figure 2.3: A general model for transform based coding system.

Since most of the transform are invertible, the distortion in the reconstructed image is contributed by the quantization noise. To derive the formula of the coding gain, we use the system in Figure 2.3 as an example, where $\mathbf{F}$ and $\mathbf{G}$ are block transform. For perfect reconstruction transforms, the forward and inverse transform matrices should satisfy $\mathbf{FG} = \mathbf{I}$.

We assume the input signal x_n is a stationary random process. The variance of the transformed signal y_n can be obtained from the transform matrix and the correlation matrix of the input signal. The variance $\sigma_{y_k}^2$ of signal y_k is the k-th diagonal element of matrix $\mathbf{FR}_{xx}\mathbf{F}^T$. By rate-distortion theory, the variance of the quantization noise, $\sigma_{q_k}^2$, is related to the $\sigma_{y_k}^2$ by

$$\sigma_{q_k}^2 = C2^{-2b_k}\sigma_{y_k}^2, \tag{2.10}$$

where b_k is the number of bits allocated to the k-th signal y_k. And C is a constant which depends on the statistic of input signal x_k. After the inverse transform, the variance of the reconstructed distortion is:

$$\sigma_{rec,k}^2 = \sigma_{q_k}^2 \|\mathbf{g}_k\|^2, \tag{2.11}$$

where $\mathbf{g}_k$ is the k-th column of the inverse transform matrix $\mathbf{G}$. Assuming the noise of different subbands are uncorrelated, the average distortion of each sample is

$$\sigma_{rec}^2 = \frac{1}{M}\sum_{k=0}^{M-1} C2^{-2b_k}\sigma_{y_k}^2 \|\mathbf{g}_k\|^2. \tag{2.12}$$

Let the average number of bits allocated to each channel be $b = \frac{1}{M}\sum_{k=0}^{M-1} b_k$. In the PCM system, the signal is directly quantized to bit b. Therefore, the quantization noise PCM is :

$$\sigma_{PCM}^2 = C2^{-2b}\sigma_x^2. \tag{2.13}$$

The coding gain of the given transform scheme over the PCM is thus

$$\gamma = \frac{\sigma_{PCM}^2}{\sigma_{rec}^2} = \frac{2^{-2b}\sigma_x^2}{\frac{1}{M}\sum_{k=0}^{M-1} 2^{-2b_k}\sigma_{y_k}^2 \|\mathbf{g}_k\|^2}. \tag{2.14}$$

In (2.14), the bit rate allocated to each subband is a free parameter. We need to select the best bit rate combination to obtain the highest coding gain for the given system. For example, we could allocated more bits to low frequency coefficients, which contain most information. The maximum coding gain is obtained, when $\sum_{k=0}^{M-1} 2^{-2b_k}\sigma_{y_k}^2 \|\mathbf{g}_k\|^2$ is minimized under the condition of $b = \frac{1}{M}\sum_{k=0}^{M-1} b_k$. We thus have the Lagrangian equation

$$L = \sum_{k=0}^{M-1} 2^{-2b_k}\sigma_{y_k}^2 \|\mathbf{g}_k\|^2 + \lambda(\frac{1}{M}\sum_{k=0}^{M-1} b_k - b). \tag{2.15}$$

The solution for (2.15) is

$$b_i = b + \frac{1}{2}\log_2 \frac{\sigma_{x_i}^2 \|\mathbf{g}_i\|^2}{\prod_{k=0}^{M-1}(\sigma_{y_k}^2 \|\mathbf{g}_k\|^2)^{\frac{1}{M}}}. \tag{2.16}$$

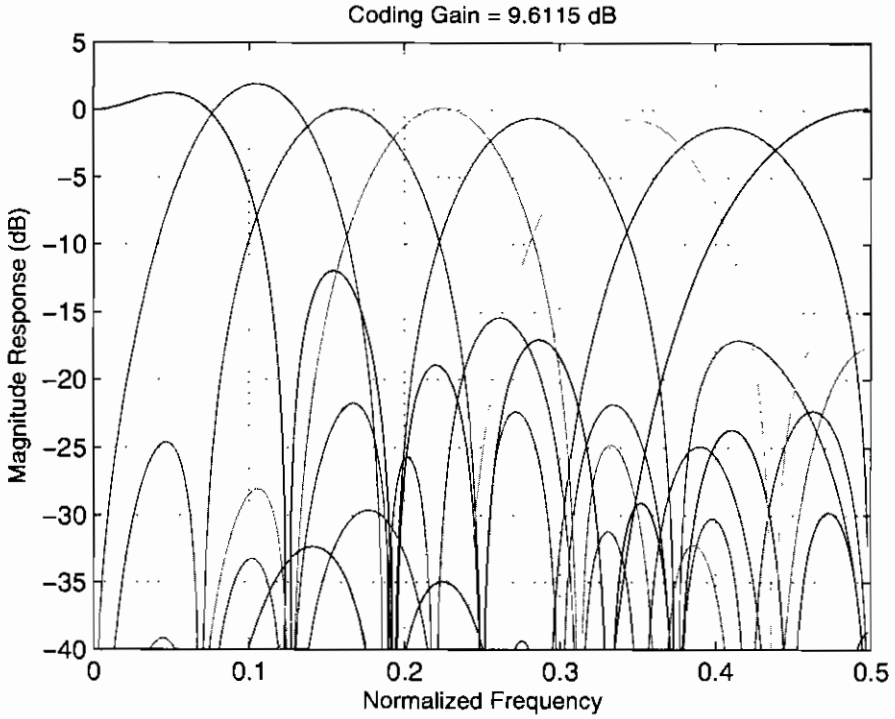


Figure 2.4: Design example: Biorthogonal 8×16 TDLT with coding gain 9.61 dB.

Plugging (2.16) into (2.14), we have the optimal coding as

$$\gamma = \frac{\sigma_x^2}{\prod_{k=0}^{M-1} (\sigma_{y_k}^2 \|\mathbf{g}_k\|^2)^{\frac{1}{M}}}. \tag{2.17}$$

When the transforms are orthogonal, $\|\mathbf{g}_k\| = 1$, so the denominator above reduces to the geometric mean of the subband variance.

When $\mathbf{F}$ and $\mathbf{G}$ are lapped transform, the coding gain still have the same formula. The difference is that the filter $\mathbf{g}_k$ has length L instead of M.

2.1.4 Design and Fast Implementations

In many compression applications, the goal of transform design is to find the one with the highest coding gain. Since there is no closed-form solution for our problem, we have to use numerical optimal functions to find out the best transform. Since matrix $\mathbf{V}$ is the only free parameter in the transform, the problem reduces to find the optimal matrix $\mathbf{V}$. It has been shown that the structure of the TDLT is close to optimal in the family of linear-phase filter banks [46]. For example, the optimized 8×16 TDLT gives a coding gain of 9.61 dB. The frequency response of the designed TDLT is shown in Figure 2.4. Compared to this, the ideal result in [1] is only 9.63 dB, but the structure and implementation of TDLT are much simpler.

As we know, any $N \times N$ orthogonal matrix can be factored into a cascade of $N(N-1)/2$ plane rotations and N sign parameters [49]. For the biorthogonal case, one can use $N(N-1)$ rotations and N diagonal entries according to the SVD factorization. This property can be used to find the fast implementation of the TDLT. For example, we can use the plane rotation operations to implement matrix $\mathbf{V}$ and the DCT. However, plane rotations still involve floating point multiplications, which are slow and undesired in DSP or VHDL implementation. To further reduce the cost, the lifting-based fast TDLT can be used, which is multiplier-free. Details of this solution can be found in [46]. Some approximations can also be applied in the lifting-based algorithm. Although the coding gain is not as high as the original structure, the system complexity can be further reduced.

2.1.5 Summary

The TDLT implements the lapped transform through time-domain pre-processing of DCT inputs and post-processing of the IDCT outputs. The pre- and postfilters are placed at the boundaries of DCT blocks. The overlap factor can be easily modified by change the size

of the pre- and postfilters. Moreover, we can have multi passes of the pre- and postfilters. With those features, the trade-off between the complexity and performance can be easily obtained. Since the DCT structure is kept intact, the original DCT-based software or hardware can be reused when the TDLT is applied. The TDLT provides better coding performance with a moderate increase of the computational complexity, when the lift-based fast TDLT is used.

This general framework allows a great degree of flexibilities. With different prefilters, the TDLT can easily adjust the coding performance, which is an important benefit for designing a multiple description coding system. Therefore, we chose TDLT as the basic framework for the technique proposed in this thesis.

2.2 Multiple Description Coding

2.2.1 Introduction

In multiple description (MD) coding, the source is encoded into several representations; each of them is called one description and can be transmitted over a different channel. Ideally, information of the original source can be distributed over all descriptions according to the capacities of different channels. At the receiver side, the quality improves steadily as the number of descriptions available increases. In order to gain robustness against the loss of descriptions, MD coding has to sacrifice some compression efficiencies, such that some redundancies can be introduced into the descriptions. The design questions in the MD coding include how to generate the descriptions, how to allocate the rates over different descriptions, and what are the optimal performance that can be achieved by the MD coding. These issues will be discussed next.

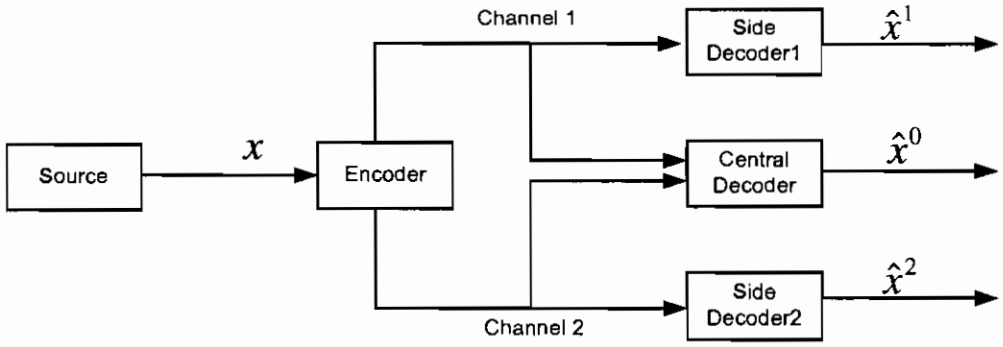


Figure 2.5: Typical 2 channel multiple description coding system.

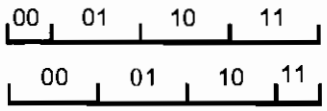


Figure 2.6: Example of the scalar quantizer for channel splitting

2.2.2 MD model

A typical two-channel multiple description coding scenario is shown in figure 2.5. The source signal passes through the MD encoder and is encoded into two descriptions, which are transmitted over Channel 1 and Channel 2 respectively. The two descriptions are somehow correlated. At the receiver side, there are three decoders, two side decoders and one central decoder. When one channel is broken and only one description is received, the receiver decodes the available description with the corresponding side decoder. When both descriptions are available, the central decoder is applied to obtain a better reconstructed signal.

One simple way for MD coding is to use specially designed scalar quantizers. Figure 2.6 shows an example by using two uniform quantizers for channel splitting. Each quantizer has an offset from each other. The signal is quantized into 2 bits for each channel.

The total bits used on the signal are 4 bits. When both channels are available, the decoded signal can achieve 3-bit accuracy, due to the shifted partitions of the two quantizers. Therefore each channel carries 0.5 bits per sample of extra information. In other words, the redundancy of the system is 0.5 bits per sample.

Two fundamental and interweaving difficulties in MDC design are how to introduce a controlled amount of redundancy into the descriptions which the decoder can conveniently utilize, and how to exploit the source correlation to facilitate MD encoding and decoding. The importance of these issues is corroborated by Shannon's comments on reliable communications ([37], pp. 75): *"Any redundancy in the source will usually help if it is utilized at the receiving point. In particular, If the source already has a certain redundancy ... a sizable fraction of the letters can be received incorrectly and still reconstructed by the context."* The "redundancy" in Shannon's paper includes both the additional redundancy introduced by the encoder and the original source correlation.

Many practical MDC schemes have been proposed. In [51], a multiple description scalar quantizer (MDSQ) is developed, which is asymptotically near optimal [52]. It is also shown in [52] that the product bound of the central and side distortions is constant for a given bit rate. This property has been widely used as a performance measure for MDC. The MDSQ has been used in, for example, [2, 36], together with DCT or the wavelet transform. One problem of MDSQ is that it requires special index assignment, which is difficult to design and implement. Its redundancy is also not easy to control.

In [43], a scalar quantizer-based modified MDSQ (MMDSQ) with the same asymptotical performance as the MDSQ is developed, in which two staggered scalar quantizers are used to generate the first layer of each description, respectively, similar to Fig. 2.6. Another scalar quantizer is used to further partition the joint bins of the first-layer quantizers, and the output of which is split into the two descriptions. The detail of this method is in the later section, since we will compare our proposed method mainly with MMDSQ.

Another large family of MDC schemes follows the source splitting approach pioneered by Jayant in [19, 20], where a speech signal is split into even and odd samples, and DPCM is used to encoded each description. If one description is lost, the missing data are predicted from their neighbors in the other description, using the correlation of the source. However, the prediction errors of the missing data are tied to the source correlation, which cannot be controlled. Therefore the overall distortion in this case is not satisfactory [19, 14]. In [18], DPCM is used before splitting, and the prediction in the DPCM is designed to preserve some source correlations. Therefore the redundancy between the descriptions can be adjusted, albeit in a very limited range. Although the method can reduce the inter-description prediction error, the remaining error still curbs the performance of the side decoder, especially at high rates.

In [22], the transform coefficients are split into two parts. Each part is quantized into one description. Each description also includes some redundant data by coarsely quantizing the other part, which helps the decoding when the other description is lost. The optimal redundancy rate allocation is studied. A similar approach is developed in [28] using the Set Partition in Hierarchical Trees (SPIHT) algorithm [35]. Recently this method is applied to the JPEG 2000 framework in [45] under the name of RD-MDC, in which each JPEG 2000 code-block is encoded at two rates, one in each description, and the rate allocation is determined by Lagrangian optimization. Some performance gains of the RD-MDC over [22, 28] are due to the more advanced entropy coding in JPEG 2000. In addition, to get balanced descriptions and optimal performance, the RD-MDC needs to classify all code-blocks into two subsets, such that any code-block in one subset has similar rate-distortion curve to another code-block in the other subset. This procedure is highly time-consuming. Another drawback of the RD-MDC is that its side decoder performance at low redundancies is also not satisfactory, because each description has too little information about the other description.

The pairwise correlating transform (PCT) [54] introduces another method of splitting the data. Different from [22, 28], the redundancy in [54] is controlled by a set of 2×2 correlating transforms. If one coefficient is lost, it is estimated from its counterpart in the other description. The PCT can achieve a lower redundancy range than the MDSQ, but it has worse performance at high rates [54]. It is shown in [55] that this is also caused by the residual of the linear prediction, similar to [19]. To resolve this problem, it is proposed in [55] to encode the prediction residual in each description, but no image coding result is reported. In [12], the PCT is generalized to introduce correlation among more than two coefficients. In [13, 23], the quantized frame expansion theory is developed to create oversampled MDC systems.

The PCT framework has some inherent drawbacks. First of all, although the PCT has good low redundancy performance in theory, its practical application could not achieve this, because the PCT can only be applied to coefficients with large variances relative to the quantization error [54]. Other coefficients are directly split into the two descriptions. In the side decoder, these low-variance coefficients are simply estimated as zero. Therefore at low redundancies, the side decoder performance will be very limited. Secondly, similar to [19], the PCT does not perform well at high redundancies because of the prediction residual. In this case the decoded image of the side decoder in [54] can be 2 dB lower than the MDSQ.

Thirdly, the PCT only uses the correlation it inserts between the two parts of a block, but does not exploit the rich correlation among neighboring blocks. Finally, the practical implementation of PCT is quite complicated. Given L coefficients, the PCT needs to pair the two coefficients with the k-th and $(L-k)$-th largest variances, and the optimal PCT depends on the coefficient variances. To estimate these variances, the method in [54] classifies all image blocks into four classes. Coefficient variances of each class are then calculated and sorted for PCT designs. In addition, existing entropy coding for single description coding

cannot be used for the PCT outputs due to the different statistics and block sizes.

A generalized PCT (GPCT) is proposed in [55], which is a hybrid method. At low redundancies it is simply the PCT. At high redundancies, in addition to PCT, each description also encodes the prediction residual of the other half of PCT outputs. Nevertheless, the analysis in [55] shows that the GPCT is still 3 dB away from the result that can be achieved by the MDSQ and the MMDSQ. Also, although the GPCT improves the PCT at high redundancies, it does not solve the other problems of the PCT. In fact, it further complicates the system, because another new transform and entropy coding have to be designed for the residual data, leading to a MDC system with three stages of transforms: the de-correlating transform (e.g., the DCT), the PCTs, and the transform for the residual. In addition, there has not been any practical application using the GPCT.

In addition to the DCT and the wavelet transform, the lapped transform [26] has also been utilized in MDC, originated from its application in error resilience and error concealment [16, 8]. In [17, 5, 6], the lapped orthogonal transform output is split at the block level. The redundancy is controlled by designing lapped transforms with different tradeoffs between compression efficiency and error resilience. When some descriptions are lost, the lost blocks are estimated by averaging neighboring blocks in [17]. In [5, 6], the missing areas are concealed by imposing a smoothness constraint. In [48], a new family of lapped transform called the time domain lapped transform [46] is used, which simplifies the design. The Wiener filter is applied in [25] to replace the simple average in estimating the lost blocks. Despite significant improvement over [17, 5, 6, 48], the performance in [25] is still below that of [43], due to the presence of the prediction error. Another common problem of [17, 5, 6, 48, 25] is that they cannot vary the redundancy for the given transform. From the point of view of introducing controlled redundancy, both the lapped transform approach and the PCT can be viewed as the generalization of the DPCM method in [18] to transform coding case.

As a summary, in [19, 20], the existing correlation in the source is cleverly utilized, but they cannot adjust the redundancy of the MDC system. Although the methods in [36, 43, 22, 28, 45, 54, 55] can control the amount of redundancy, they do not fully exploit the source correlation. The scheme in [18] uses the source correlation to provide a limited redundancy, but is based on DPCM, and suffers from the prediction error at high rates.

Given these limitations of existing methods, one of the motivations of this thesis is to develop a MDC scheme that can simultaneously achieve the following properties: taking full advantage of the rich correlation of the source in the MD encoding and decoding, providing effective control of the redundancy of the system, and achieving satisfactory performance at all rates and redundancies. Another goal of this thesis is to improve the performance of MD image coding by applying this scheme. Since MD image coding and MD video coding are two most important applications of MDC, and there are still many open problems in MDC theory, especially for more than two descriptions, new findings in the MD image coding can help the development of MD video coding.

The requirement of utilizing the source correlation in MD encoding and decoding suggests us to deviate from the traditional transform paradigm, for which the error resilient design of the lapped transform becomes a very suitable platform. In particular, we will use the time-domain lapped transform (TDLT) developed in [46].

2.2.3 The MD rate distortion region

For lossy data compression, the quality of the reconstructed data depends on the bit rate that is used to encode the source. The rate-distortion (R-D) theory is a branch of information theory that studies the achievable distortion at a given rate, or vice versa. The R-D bounds are usually difficult to compute except for a few simple situations, such as Gaussian sources. The definitions used here and more details of the rate distortion theory could be found in [7] and [3].

Several concepts need to be defined first. The n-length source sequence is noted as $x^{(n)} = \{x_1, x_2, \ldots x_n\}$; the reconstructed sequence is noted as $\hat{x}^{(n)} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n\}$. A distortion measure is a mapping from the set of source-reconstructed pairs into the set of non-negative real numbers. The most common one is the squared error:

$$d(x, \hat{x}) = (x - \hat{x})^2. \tag{2.18}$$

The average distortion between sequences $x^{(n)}$ and $\hat{x}^{(n)}$ is defined by

$$d(x^{(n)}, \hat{x}^{(n)}) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \tag{2.19}$$

Usually the input source is modeled as a sequence of independent, identically distributed, real random variables $X^{(n)} = \{X_1, X_2, \ldots, X_n\}$. The encoding function $f(.)$ is a mapping from all possible source sequences to an index in $\{1, 2, \ldots, 2^{nR}\}$, where R is the rate. The decoding function $g(.)$ is the mapping from an index in $\{1, 2, \ldots, 2^{nR}\}$ back to the reproduction sequence $\hat{X}^{(n)} = \{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n\}$. Those two functions is called a source code. The distortion associated with this code is the expected value

$$D = E[d(X^{(n)}, g(f(X^{(n)})))]. \tag{2.20}$$

A rate-distortion pair (R, D) is achievable if there exists a sequence of $(2^{nR}, n)$ rate distortion codes such that

$$\lim_{n \to \infty} E[d(X^{(n)}, g(f(X^{(n)})))] \leq D \tag{2.21}$$

The rate-distortion region for a source is the closure of the set of achievable rate distortion pairs.

In general, it is very difficult to find out the boundary of the rate distortion region from the definition. However, this is doable for some special cases. For example, a Gaussian

source with variance σ^2 has a distortion-rate function of

$$D(R) = \sigma^2 2^{-2R}. \tag{2.22}$$

In fact, this is the upper bound for the distortion-rate functions of all continuous-value sources. If a source has a probability density function $p(x)$ and variance σ^2, the distortion-rate function satisfies

$$\frac{1}{2\pi e} 2^{2h} 2^{-2R} \leq D(R) \leq \sigma^2 2^{-2R}, \tag{2.23}$$

where $h = -\int p(x) \log_2 p(x) dx$ is the differential entropy. From 2.23, we know the Gaussian sources are the most difficult to compress.

In the case of multiple description coding, the rate distortion theory is more complicated, and remains largely open, especially for more than two descriptions. For two-description case, the MD region is the closure of the set of achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$, where R_1 and R_2 are the rate of Description one and two respectively, D_1 and D_2 are the distortion of the side decoder 1 and decoder 2, and the D_0 is the distortion of the central decoder.

In [10], an achievable two-description rate-distortion region for independent and identically distributed sources was derived. The region was shown by Ozarow to be complete for memoryless Gaussian sources under the mean squared-error distortion measure [29]. For other memoryless sources, a set of outer and inner bounds was derived in [58], and these bounds were generalized to sources with memory in [57]. In particular, the outer bound is tight at high rates. For a memoryless Gaussian source with variance σ^2,the MD region should be:

$$D_i \geq \sigma^2 2^{-2R_i}, \text{ for } i = 1, 2 \tag{2.24}$$

$$D_i \geq \sigma^2 2^{-2(R_1+R_2)} . \gamma_D(R_1, R_2, D_1, D_2), \tag{2.25}$$

where $\gamma_D(R_1, R_2, D_1, D_2) = 1$ if $D_1 + D_2 \geq \sigma^2 + D_0$, otherwise

$$\gamma_D(R_1, R_2, D_1, D_2) = \frac{1}{1 - (\sqrt{(1-D_1)(1-D_2)} - \sqrt{D_1 D_2 - 2^{-2(R_1+R_2)}})^2}. \tag{2.26}$$

From these equations, we can know that when both side distortions are large, the central distortion can be very good. But if the side distortions are better, the central distortion will be sacrificed.

If both channels have the same condition, *i.e.*, the same probability of loss, we can design the MD system as $R_1 = R_2$, $D_1 = D_2$. If we assume the source has unit variance, it can be proved that the side distortion bound is [11]:

$$D_1 \geq min\left\{\frac{1}{2}\left[1 + D_0 - (1-D_0)\sqrt{1 - 2^{-2(R_1+R_2)}/D_0}\right], 1 - \sqrt{1 - 2^{-2(R_1+R_2)}/D_0}\right\} \tag{2.27}$$

If the base rate of D_0 is $r = R(D_0)$ and redundancy of the MD coding is $\rho = R_1 + R_2 - R(D_0)$, plug the basic rate distortion bound $D_0 \geq 2^{-2r}$ into (2.27), we have the side distortion bound:

$$D_1 \geq \begin{cases} \frac{1}{2}\left[1 + 2^{-2r} - (1-2^{-2r})\sqrt{1-2^{-2\rho}}\right], & \text{for } \rho \leq r - 1 + \log_2(1+2^{-2r}) \\ 1 - \sqrt{1-2^{-2\rho}}, & \text{for } \rho > r - 1 + \log_2(1+2^{-2r}). \end{cases} \tag{2.28}$$

Figure 2.7 shows the MD side distortion bound as the function of redundancy ratio, for unit basic rate, such that $r = 1$. The redundancy ratio ε is the ratio of redundancy rate ρ over the basic rate r:

$$\varepsilon = \frac{\rho}{r} = \frac{R_1 + R_2 - r}{r}. \tag{2.29}$$

From the figure 2.7, we can see that as $\varepsilon \to 0$, the slope of the curve is infinite. At that point, the idea MD system, which can achieve the distortion bound, uses all the bits to

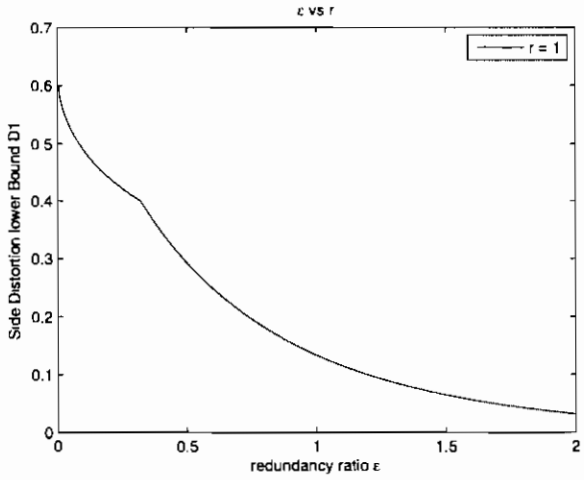


Figure 2.7: MD coding side distortion bound vs redundancy ratio with unit basic rate

improve the central distortion. The infinity slope means that a small additional bit can reduce the side distortion dramatically. As the redundancy ratio increases, the slope reduces. Therefore the efficiency of improving the side decoder decreases. When the total rate is fixed, how to distribute the bits between improving the central distortion and improving the side distortion becomes an important problem. In many cases, the linear combination of the central and side distortions is used as performance measure. The weights of the distortion could be determined by the probabilities of losing different descriptions.

2.2.4 The modified multiple description scalar quantizer (MMDSQ)

In multiple description coding, the specially designed MD scalar quantizer has gained more and more attentions. In [52], Vaishampayan presented the asymptotic analysis of the entropy-constrained multiple description scalar quantizer(MDSQ). In the balanced two-description case, it is shown that the distortion product of the central and side distortion of

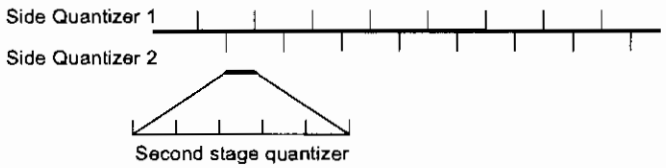


Figure 2.8: Modified multiple description scalar quantizer example

this quantizer with a uniform central quantizer is

$$D_0 D_1 \approx \left(\frac{2\pi e}{12}\right)^2 \frac{P_x^2}{4} 2^{-4R}, \tag{2.30}$$

where R is the rate of each description, $P_x = (2\pi e)^{-1} 2^{2h_p}$ is the entropy power of the source, and h_p is the differential entropy of the source.

The performance of the MDSQ is only 3.07 dB worse than the theoretical bound, which is the same as the gap of the single-description entropy-constrained scalar quantizer. When nonuniform central quantizers are used, this production can be further reduced by 0.4dB [44].

However, the side quantizers in MDSQ are not conventional quantizers. The cell does not consist of a continuous intervals; one side quantization cell is the union of several noncontiguous intervals. The MDSQ also involves a complex index assignment.

Recently, a modified multiple description scalar quantizer (MMDSQ) is introduced by Tian [43]. Instead of complex index assignment, the MMDSQ uses a simple two-stage scalar quantizer scheme, as shown in Fig. 2.8. This simple scheme still achieves an asymptotic performance that is identical to the entropy-constrained MDSQ.

The first stage of MMDSQ has two uniform side quntizers with staggered bins, which is similar to the simple example shown in Figure 2.6. Therefore the joint quantizer has a half size bin. In the second stage, the half size bin is divided evenly into a fixed number of bins to obtain the second stage quantizer. Each quantization value of the first stage side

quantizers is entropy-coded and put into the corresponding description. The indices from the second stage quantizer is also entropy-coded but split evenly between two descriptions. At the receiver side, if only one description is available, the corresponding side quantizer is utilized to decode the first stage information, whereas the information of the second stage is discarded. When both descriptions are received, the joint quantizer is used to get a coarse result, and then the decoder will use the second stage information to obtain a more accurate result.

The performance of the MMDSQ can be easily analyzed. The rate of the first stage of each description is R_1, and that of the second stage of each description is R_2. Assuming high bit rate coding, the side distortions of the two uniform side quantizers with entropy coding are [15]

$$D_1 \approx \frac{2\pi e}{12} P_x 2^{-2R_1}. \tag{2.31}$$

If we assume R_1 is high, the source distribution in each side quantization cell can be regarded as uniform. The number of bins in the second quantizer is $N = 2^{2R_2}$. Since the interval of bins of the joint quantizer is half of the side quantizer, and each of this interval is divided into N uniform bins, the central distortion of MMDSQ is

$$D_0 \approx \frac{1}{(2N)^2} D_1 = \frac{2\pi e}{12} \frac{P_x}{4} 2^{-2(R_1+2R_2)}. \tag{2.32}$$

Thus the production of the side and the central distortions is

$$D_0 D_1 \approx \left(\frac{2\pi e}{12}\right)^2 \frac{P_x^2}{4} 2^{-2(2R_1+2R_2)} = \left(\frac{2\pi e}{12}\right)^2 \frac{P_x^2}{4} 2^{-4R}, \tag{2.33}$$

where $R = R_1 + R_2$ is the rate of each description. This formula is identical to (2.30). This means that the MMDSQ has the same asymptotic performance as the MDSQ, but the complexity of the MMDSQ is much lower than the MDSQ.

The MMDSQ provides a more flexible tradeoff control mechanism. The two side quantizers controls the side distortions. The number of bins in the second stage quantizer controls how good the central reconstruction could be. The tradeoff between the side and central distortions can be easily controlled by changing the two quantizers. It is much simpler than the MDSQ which needs to reassign the indices to adjust the distortion tradeoff.

The combination of the MMDSQ, wavelet and Tarp filter outperforms other MDSQ-based image coding methods such as [36]. In fact, it represents the state of the art in MD image coding. However, the MMDSQ does not have satisfactory performance at low redundancy regime, which is a desired property of good MDC schemes ([54], pp. 365). Another problem of MMDSQ and MDSQ is that no attempt is made to utilize the source correlation for MDC purpose.

2.2.5 Summary

The multiple description coding has gained growing attentions recently, although it was proposed long time ago. The MD coding was invented for bad communication environment where the losing of partial information during transmission is common. The most typical circumstance for MD coding to be used is the IP/TCP network. The applications in which MD coding is useful are most likely real-time and have some quality degradation tolerance. The rate distortion for MD coding system is much more complicated than the normal rate distortion problems. To design a MD coding system, many things need to be considered. It requires a tradeoff between the side distortion and central distortion. Therefore, the flexility for adjusting the distortions is very important. Many techniques have been proposed for achieving MD system, and the MD scalar quantizer is a popular one, but it needs complicated index design. This problem is solved in the MMDSQ, which achieves the same performance as the MDSQ.

In addition to the quantizer, the transform can also be used for MD coding. In the

next few chapters, a new MD technique with the lapped transform is introduced and the performance is compared with MMDSQ.

Chapter 3

MDC with Prediction Compensation

3.1 Introduction

As mentioned before, an error resilient lapped transform is developed in [25], where the lost blocks are predicted by Wiener filter, which gives much better performance than the simple average scheme in [48]. However, this scheme cannot give satisfactory side decoder performance if used as a MDC method, because the prediction is not accurate enough. The problem is especially noticeable at high bit rates.

In this chapter, we develop a MDC scheme on top of the Wiener filter scheme in [25]. We first partition the image into two sub-images after the prefilter of the time-domain lapped transform. Each sub-image is coded as the base layer of one description. Each description also predicts the other description using Wiener filter, and encodes the prediction residual as an enhancement layer. This can improve the side decoder performance. This method has the same spirit as the method in [55]. However, the design and implementation of our method are both simpler than those in [54, 55], because we split the data at block level and only one transform and one Wiener filter need to be designed, whereas coefficient level splitting is used in [55], which calls for the design of a set of PCTs. Another advantage

of our method is that superior performance can be obtained without any modification of existing entropy coding algorithm.

The joint design of the Wiener filter and the lapped transform will be presented, including the optimal bit allocation between the base layer and the enhancement layer. The asymptotic performance of our method is analyzed by replacing the block transform with the DPCM. In this case, our method is only about 3 dB away from the theoretical bound given in [52, 57], when the input is assumed to be first-order Gauss-Markov sources. This makes it a simple but effective improvement over Jayant's method in [19]. When transform coding is used, our scheme is especially suitable for natural images, and experimental results show that at the same bit rate and central distortion, our method can outperform the method in [43] by more than 8 dB if only one description is received. We also study the performances of the DCT, the performance of the single description optimized TDLT in [46] and the effect of Wiener filter size in the proposed framework.

3.2 Problem Formulation and Optimal Design

In this section, we present the proposed MDC scheme and discuss its advantages over other methods. We then formulate the joint design of various components of the system, and derive the optimal solution.

3.2.1 Overview of the Proposed MDC Scheme and its Advantages

In this chapter, we modify the TDLT framework to generate two descriptions. Fig. 3.1 illustrates the encoding and decoding of one description by the proposed method. The other description is obtained similarly. In this chapter, we use $\mathbf{x}(k)$, $\mathbf{s}(k)$, and $\mathbf{y}(k)$ to denote the k-th block of prefilter input, DCT input, and DCT output. The reconstruction of a variable is denoted by the hat operator.

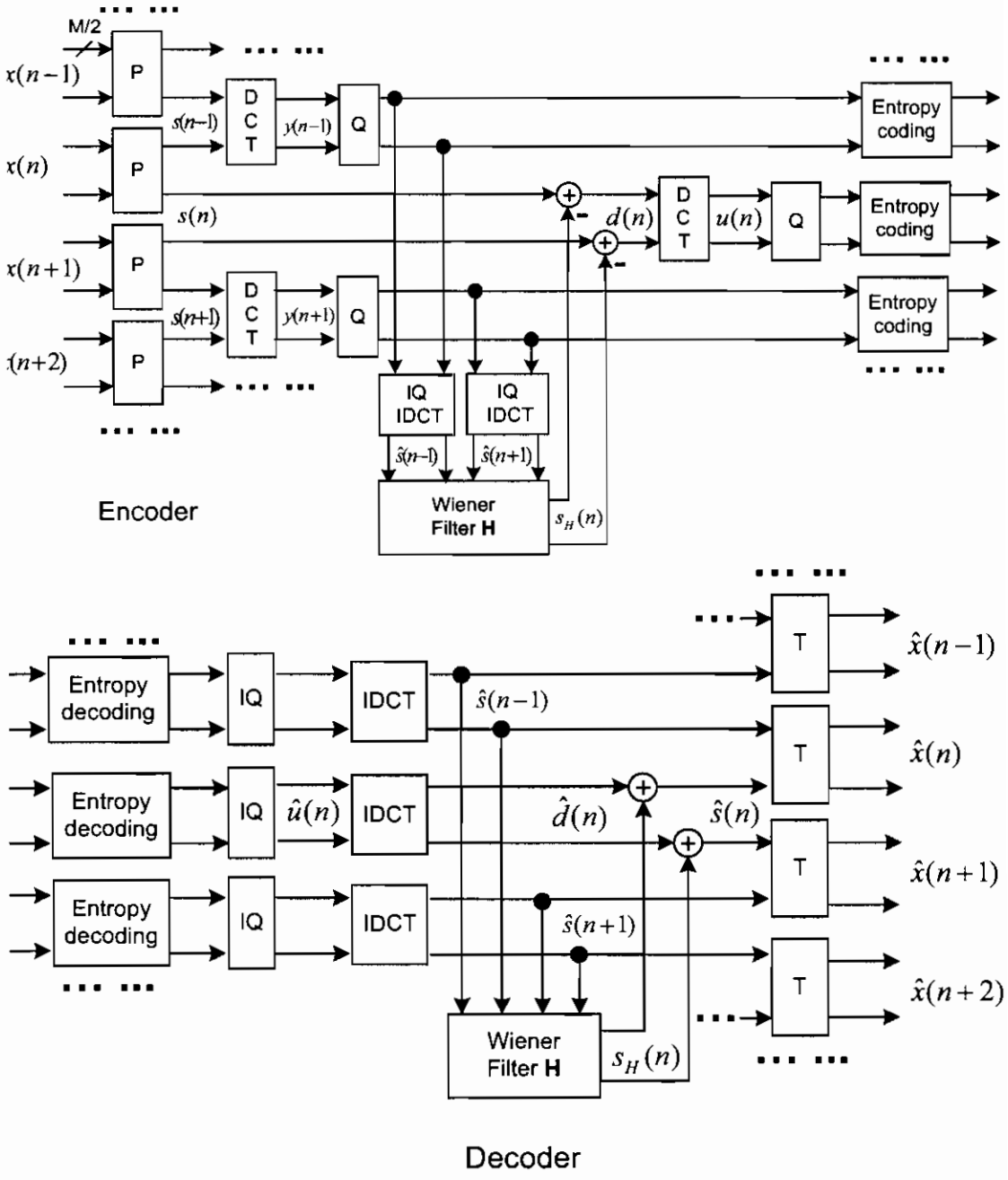


Figure 3.1: Block diagrams of encoding and decoding one description in the proposed method.

In Fig. 3.1, after prefiltering the input, the prefiltered blocks $\{\mathbf{s}(k)\}$ are split into even-indexed blocks and odd-indexed blocks. After DCT, quantization, and entropy coding, the two groups of blocks form the base layers of the two descriptions, respectively. Since the base layer only includes data of a half of the input, each description also contains an enhanced layer to help reconstructing the other half, if only one description is received. To fully exploit the correlation among neighboring blocks, in each description we use a Wiener filter $\mathbf{H}$ to predict a block using its two reconstructed neighboring blocks, as in [25]. More precisely, $\mathbf{s}(n)$ in Fig. 3.1 is predicted using the nearest N samples from $\hat{\mathbf{s}}(n-1)$ and N samples from $\hat{\mathbf{s}}(n+1)$, where N is chosen between 1 and M, and M is the block size. The formula of the Wiener filter is given in the section 3.2.2. Different from [25], in this chapter the prediction residuals $\{\mathbf{d}(k)\}$ are further DCT-transformed, quantized and entropy-coded to form the enhancement layer of each description.

At the decoder side, if only one description is received, the missing blocks are first estimated from the received base layer blocks by Wiener filter. The decoded enhanced layer blocks are then added to the estimation before postfiltering. When both descriptions are available, only the decoded base layers from the two descriptions are fed to the postfilter for reconstruction. The enhanced layer in each description is simply discarded. Therefore the enhanced layer is the redundancy introduced by our method, which can be easily controlled by adjusting the quantization step.

Our scheme enjoys various advantages over existing methods. Compared to MDSQ, MMDSQ, RD-MDC, PCT/GPCT, it uses Wiener filter based prediction to exploit the source correlation between neighboring blocks in creating MDC. The importance of utilizing the source correlation can be seen in Sec. 3.5, where even a simple system with DCT and linear interpolation based prediction can outperform other leading MDC methods in some cases. Compared with MDSQ and [19, 20, 18], our method offers a more flexible control of the redundancy. Our method is also superior to [22, 28, 45] by using predictive coding, as

it is well established that for correlated sources, predictive coding has better rate-distortion performance than direct encoding ([41], pp. 113), *i.e.*, it achieves lower distortion at the same bit rate.

The proposed scheme also avoids other problems of the PCT and GPCT. Firstly, prediction in the spatial domain allows the residual to be coded even at low redundancies, where important image edge information can still survive the large quantization step after the prediction and drastically improve the side decoder quality. In contrast, most coefficients in the PCT are split directly in this case, which are uncorrelated and cannot be predicted from each others, leading to inadequate performance. Secondly, we will show that at high redundancies, the theoretical performance of our method is as good as the GPCT. The practical performance of our method is even better, as it is very suitable for nonstationary signals can achieve similar or slightly better results than the MMDSQ and RD-MDC at high redundancies.

In addition, our scheme can be easily optimized and implemented, and good performance can be expected. For example, only one prefilter and one Wiener filter need to be designed instead of many PCTs. The block level splitting has much less impact on coding efficiency than coefficient level splitting. The prediction residual of the block Wiener filter is similar to the block-based motion estimation residual in video coding, and can be well decorrelated by the DCT, and there is no need for a new transform. Our experimental results also show that the entropy coding for the base layer can be applied to the enhanced layer directly with good performance. In addition, the postfilter can serve as a de-blocking filter to smoothen the artifacts and improves the visual quality, especially around the predictively coded blocks.

3.2.2 Wiener filter derivation

In this section, we derive the Wiener filter $\mathbf{H}$ in Fig. 3.1. In [25], all data in the two neighboring blocks are used to estimate a lost block. This needs a Wiener filter of size $M \times 2M$. Since the performance of the system is not sensitive to the size of the Wiener filter, in this chapter we use N ($1 \leq N \leq M$) boundary samples from each of the two neighboring blocks to estimate the lost block. The value of N is selected based on the application to obtain a desired trade-off between complexity and performance.

In Fig. 3.1, the prediction of $\mathbf{s}(n)$ is $\mathbf{s}_H(n) = \mathbf{H}\,\hat{\mathbf{s}}_{2,N}$, where $\mathbf{H}$ is the $M \times 2N$ prediction filter, and $\hat{\mathbf{s}}_{2,N}$ is a $2N \times 1$ vector containing $2N$ nearest neighboring samples next to $\mathbf{s}(n)$, N samples from $\hat{\mathbf{s}}(n-1)$ and N samples from $\hat{\mathbf{s}}(n+1)$. That is,

$$\hat{\mathbf{s}}_{2,N} = \begin{bmatrix} \hat{\mathbf{s}}_B^T(n-1) & \hat{\mathbf{s}}_B^T(n+1) \end{bmatrix}^T, \tag{3.1}$$

where

$$\begin{aligned} \hat{\mathbf{s}}_B(n-1) &= \begin{bmatrix} \hat{s}_{M-N}(n-1) & \ldots & \hat{s}_{M-1}(n-1) \end{bmatrix}^T, \\ \hat{\mathbf{s}}_B(n+1) &= \begin{bmatrix} \hat{s}_0(n+1) & \ldots & \hat{s}_{N-1}(n+1) \end{bmatrix}^T. \end{aligned} \tag{3.2}$$

The autocorrelation of the prediction residual is

$$\mathbf{R}_{\mathbf{dd}} = E\{(\mathbf{H}\hat{\mathbf{s}}_{2,N} - \mathbf{s}(n))(\mathbf{H}\hat{\mathbf{s}}_{2,N} - \mathbf{s}(n))^T\}, \tag{3.3}$$

The Wiener filter that minimizes the MSE $\text{trace}\{\mathbf{R}_{\mathbf{dd}}\}/M$ is $\mathbf{H} = \mathbf{R}_{\mathbf{s}\hat{\mathbf{s}}_{2,N}}\mathbf{R}^{-1}_{\hat{\mathbf{s}}_{2,N}\hat{\mathbf{s}}_{2,N}}$, where $\mathbf{R}_{\mathbf{s}\hat{\mathbf{s}}_{2,N}}$ is the correlation matrix between $\mathbf{s}(n)$ and $\hat{\mathbf{s}}_{2,N}$, and $\mathbf{R}_{\hat{\mathbf{s}}_{2,N}\hat{\mathbf{s}}_{2,N}}$ is the autocorrelation matrix of $\hat{\mathbf{s}}_{2,N}$. A high rate assumption is usually used to simplify the design ([25, 41], pp. 114). In this case, the quantization noise is ignored, and the Wiener filter can be

approximated as

$$\mathbf{H} = \mathbf{R}_{\mathbf{s}\mathbf{s}_{2,N}} \mathbf{R}^{-1}_{\mathbf{s}_{2,N}\mathbf{s}_{2,N}}. \tag{3.4}$$

The matrices involved can be obtained from the structure of the lapped transform in Fig. 2.1. Define

$$\begin{aligned} \mathbf{s}_3 &= \begin{bmatrix} \mathbf{s}^T(n-1) & \mathbf{s}^T(n) & \mathbf{s}^T(n+1) \end{bmatrix}^T, \\ \mathbf{x}_4 &= \begin{bmatrix} \mathbf{x}^T(n-1) & \mathbf{x}^T(n) & \mathbf{x}^T(n+1) & \mathbf{x}^T(n+2) \end{bmatrix}^T. \end{aligned} \tag{3.5}$$

As shown in Fig. 2.1, $\mathbf{s}_3 = \mathbf{P}_{34}\mathbf{x}_4$. Thus $\mathbf{R}_{\mathbf{s}_3\mathbf{s}_3} = \mathbf{P}_{34}\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}\mathbf{P}_{34}^T$, where

$$\mathbf{P}_{34} = \text{diag}\{\mathbf{P}_1, \mathbf{P}, \mathbf{P}, \mathbf{P}_0\}, \tag{3.6}$$

$\mathbf{P}_0$ and $\mathbf{P}_1$ are the first and the second $M/2$ rows of the prefilter $\mathbf{P}$, as given in (2.5), and $\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}$ is the autocorrelation matrix of $\mathbf{x}_4$. In this chapter, $\mathbf{R}_{\mathbf{x}_4\mathbf{x}_4}$ is obtained by assuming the input follows a first-order Gauss-Markov model with correlation coefficient of $r = 0.95$. Matrices $\mathbf{R}_{\mathbf{s}\mathbf{s}_{2,N}}$ and $\mathbf{R}_{\mathbf{s}_{2,N}\mathbf{s}_{2,N}}$ in (3.4) can then be obtained from the appropriate sub-matrices of $\mathbf{R}_{\mathbf{s}_3\mathbf{s}_3}$ according to their definitions.

From $\mathbf{u}(n) = \mathbf{C}\mathbf{d}(n) = \mathbf{C}(\mathbf{s}(n) - \mathbf{H}\,\hat{\mathbf{s}}_{2,N})$ and the Wiener filter $\mathbf{H}$ given in (3.4), we have

$$\mathbf{R}_{\mathbf{u}\mathbf{u}} = \mathbf{C}\{\mathbf{R}_{\mathbf{s}\mathbf{s}} - \mathbf{H}\mathbf{R}_{\mathbf{s}_{2,N}\mathbf{s}}\}\mathbf{C}^T. \tag{3.7}$$

In this chapter, this is used in Eq. (3.19) to obtain the distortion of the enhanced layer.

As in [25], we normalize the Wiener filter to have unit row sums. In addition, two $M \times N$ Wiener filters are used at the boundary to predict a block from only the top or the bottom neighboring block.

3.2.3 Joint Optimal Design of the System Components

Let R (in bits/pixel (bpp)) represent the overall bit rate of the two descriptions, *i.e.*, the ratio between the total bits of the two descriptions and the number of input samples. Let R_0 and R_1 denote the average bits for each base layer sample and each enhanced layer sample, respectively. The bit rate for each description is thus $\frac{1}{2}(R_0 + R_1)$ bpp/description, and the total rate is $R = R_0 + R_1$.

As shown below, given the target bit rate R, the probability p of losing one description, the matrix $\mathbf{V}$ in prefilter $\mathbf{P}$, and the value N for Wiener filter, we can find the closed-form expressions of the corresponding optimal Wiener filter $\mathbf{H}$ and bit allocation R_0 and R_1 that minimize the expected distortion at the receiver. To further find the optimal matrix $\mathbf{V}$ among all possible $\mathbf{V}$'s with the minimal expected distortion, an unconstrained numerical optimization program, such as the function *fminunc* in MATLAB, can be used, by treating all entries of $\mathbf{V}$ as unknown variables. Numerical optimization has to be used in this step because the objective function is a complicated nonlinear function of the entries of $\mathbf{V}$. Although global optimality is not guaranteed in general, this method has been widely used with satisfactory performance in filter bank optimization [17, 6, 48, 46, 25]. In this part, we give the derivation of the first step, *i.e.*, the optimal R_0 and R_1 for a given $\mathbf{V}$. The solution of the corresponding Wiener filter is given in the section 3.2.2.

Let D_0 and D_1 be the central distortion and side distortion, *i.e.*, the mean squared error (MSE) when the decoder receives two and one description, respectively. The expected distortion D is defined as

$$D = (1-p)^2 D_0 + 2p(1-p)D_1. \tag{3.8}$$

In the proposed MDC scheme, each description contains half base layer blocks and half enhanced layer blocks. Only base layer blocks are used if two descriptions are received. If

one of them is lost, half of the input is reconstructed via base layer and the other half is via enhanced layer. Therefore

$$\begin{aligned} D_0 &= D_{B,M}, \\ D_1 &= \frac{1}{2}(D_{B,M} + D_{E,M}), \end{aligned} \tag{3.9}$$

where $D_{B,M}$ and $D_{E,M}$ are the MSE caused by the subband quantization noise in base layer and prediction-compensated enhanced layer, respectively. Substituting into (3.8), we have

$$\begin{aligned} D &= (1-p)^2 D_{B,M} + p(1-p)(D_{B,M} + D_{E,M}) \\ &= (1-p) D_{B,M} + p(1-p) D_{E,M}. \end{aligned} \tag{3.10}$$

We first find the optimal expressions of $D_{B,M}$ and $D_{E,M}$ for given R_0 and R_1, under the optimal bit allocation within each block. Since the inverse lapped transform is generally not orthogonal, the reconstruction error $D_{B,M}$ or $D_{E,M}$ is the weighted combination of subband quantization noises, and the weighting parameters are the norms of the inverse transform filters. Let the quantization noise of $\mathbf{y}(k)$ be $\mathbf{q_y}(k)$. After the inverse TDLT, the reconstruction error becomes $\mathbf{G}\,\mathbf{q_y}(k)$. As usual, we assume the quantization noises of different subbands are uncorrelated. Therefore the MSE of the reconstruction is

$$D_{B,M} = \frac{1}{M} \sum_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma^2_{q_y(i)}, \tag{3.11}$$

where $\sigma^2_{q_y(i)}$ is the variance of the i-th entry of $\mathbf{q_y}(k)$, and $\mathbf{g}_i$ is the i-th column of $\mathbf{G}$. Under the assumptions of high rates and i.i.d. sources, $\sigma^2_{q_y(i)}$ can be written as ([41], pp. 108)

$$\sigma^2_{q_y(i)} = \epsilon\, \sigma^2_{y(i)} 2^{-2R_{0i}}, \tag{3.12}$$

where ϵ is a constant that depends on the input statistics and the quantization scheme.

R_{0i} is the bits allocated to the i-th entry of a base layer block, and $\frac{1}{M}\sum_{i=0}^{M-1} R_{0i} = R_0$. $\sigma^2_{y(i)}$ is the variance of the i-th entry of $\mathbf{y}(k)$, which is the i-th diagonal element of the autocorrelation matrix $\mathbf{R_{yy}}$:

$$\mathbf{R_{yy}} = \mathbf{F}\mathbf{R}_{\mathbf{x_2x_2}}\mathbf{F}^T, \tag{3.13}$$

and $\mathbf{x}_2 = \left[\mathbf{x}^T(k) \;\; \mathbf{x}^T(k+1)\right]^T$.

Upon optimal bit allocation of R_{0i} [21], the minimal value for (3.11) is given by

$$D_{B,M} = \epsilon \left(\prod_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma^2_{y(i)} \right)^{\frac{1}{M}} 2^{-2R_0} \triangleq \epsilon\, \sigma^2_{B,M} 2^{-2R_0}. \tag{3.14}$$

This is in fact the objective function of the single description coding. For block transforms, the minimum value of (3.14) is achieved by the Karhunen-Loève Transform (KLT). For lapped transforms and longer filter banks, there is no closed-form solution, but numerical optimization method can be used to find the solution that minimizes (3.14).

In single description coding, the coding gain in (2.17) is defined based on $\sigma^2_{B,M}$. When $M = 8$, the coding gain of the optimized TDLT in [46] is 9.62 dB for first-order Gaussian-Markov inputs with correlation $r = 0.95$. This is substantially higher than the 8.83 dB of the DCT.

We now look at $D_{E,M}$, the MSE caused by enhanced layer blocks in the side decoder. It can be seen from Fig. 3.1 that the prediction residual is

$$\mathbf{d}(n) = \mathbf{s}(n) - \mathbf{s}_H(n), \tag{3.15}$$

where $\mathbf{s}_H(n)$ is the Wiener filter-based prediction of $\mathbf{s}(n)$ from $\hat{\mathbf{s}}(n-1)$ and $\hat{\mathbf{s}}(n+1)$. The derivation of $\mathbf{s}_H(n)$ for the given prefilter is given in the section 3.2.2. At the decoder, the reconstruction of a predictively coded block $\mathbf{s}(n)$ is

$$\hat{\mathbf{s}}(n) = \hat{\mathbf{d}}(n) + \mathbf{s}_H(n), \tag{3.16}$$

where $\hat{\mathbf{d}}(n)$ is the reconstruction of $\mathbf{d}(n)$. From (3.15) and (3.16) we can get the relationship

$$\hat{\mathbf{s}}(n) - \mathbf{s}(n) = \hat{\mathbf{d}}(n) - \mathbf{d}(n). \tag{3.17}$$

In other words, the reconstruction error of $\mathbf{s}(n)$ is equal to that of the prediction residual $\mathbf{d}(n)$. This is indeed a property of any differential coding system ([41], pp. 113).

As in Fig. 3.1, let $\mathbf{u}(n)$ be the DCT transform of $\mathbf{d}(n)$, we have

$$\hat{\mathbf{d}}(n) - \mathbf{d}(n) = \mathbf{C}^T \mathbf{q_u}(n), \tag{3.18}$$

where $\mathbf{q_u}(n)$ is the quantization noise of $\mathbf{u}(n)$. After postfiltering, the reconstruction error becomes $\mathbf{T}_{21}\mathbf{C}^T\mathbf{q_u}(n) = \mathbf{G}\mathbf{q_u}(n)$, and its MSE is

$$D_{E,M} = \frac{1}{M} \sum_{i=0}^{M-1} \epsilon \, ||\mathbf{g}_i||^2 \sigma^2_{u(i)} 2^{-2R_{1i}}, \tag{3.19}$$

where R_{1i} is the bits allocated to the i-th entry of $\mathbf{u}(n)$. $\sigma^2_{u(i)}$ is the variance of the i-th entry of $\mathbf{u}(n)$, given by the i-th diagonal element of the autocorrelation matrix $\mathbf{R_{uu}}$. The expression of $\mathbf{R_{uu}}$ is given by Eq. (3.7) in the section 3.2.2.

Since (3.19) has the same format as (3.11), the derivation from (3.11) to (3.14) can be applied here, and the minimal value of $D_{E,M}$ after optimal bit allocation is therefore

$$D_{E,M} = \epsilon \left(\prod_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma^2_{u(i)} \right)^{\frac{1}{M}} 2^{-2R_1} \triangleq \epsilon \, \sigma^2_{E,M} 2^{-2R_1}, \tag{3.20}$$

The remaining bit allocation issue is to find the optimal R_0 and R_1 for the given $\mathbf{P}$ that minimize the expected distortion D in (3.10). Substituting (3.14) and (3.20) into (3.10), the

problem can be written as

$$\begin{aligned} &\arg \min_{R_0,R_1} (1-p)\epsilon\,\sigma_{B,M}^2 2^{-2R_0} + p(1-p)\epsilon\,\sigma_{E,M}^2 2^{-2R_1}, \\ &\text{s.t.} \quad R_0 + R_1 = R. \end{aligned} \tag{3.21}$$

This can be solved using the Lagrangian method by defining an objective function

$$(1-p)\epsilon\,\sigma_{B,M}^2 2^{-2R_0} + p(1-p)\epsilon\,\sigma_{E,M}^2 2^{-2R_1} + \lambda(R_0 + R_1 - R), \tag{3.22}$$

where λ is the Lagrangian multiplier. Taking derivatives of the objective function with respect to R_0 and R_1, and letting the results to be 0, we can find expressions of R_0 and R_1 in terms of λ. Substituting them into $R_0 + R_1 = R$, the optimal multiplier λ is found to be

$$\lambda = 2\ \ln 2 \sqrt{p}\,(1-p)\,\epsilon\,\sigma_{B,M}\sigma_{E,M} 2^{-R}, \tag{3.23}$$

from which the optimal R_0 and R_1 can be obtained. After considering the constraints of $R_0 \leq R$ and $R_1 \geq 0$, the optimal bit allocation can be found to be

$$\begin{aligned} R_0 &= \min\left(R, \frac{R}{2} + \frac{1}{4}\log_2 \frac{\sigma_{B,M}^2}{p\,\sigma_{E,M}^2}\right), \\ R_1 &= \max\left(0, \frac{R}{2} - \frac{1}{4}\log_2 \frac{\sigma_{B,M}^2}{p\,\sigma_{E,M}^2}\right). \end{aligned} \tag{3.24}$$

At high rates, *i.e.*, if R_1 is not forced to 0 in (3.24), substituting this into (3.14) and (3.20), we have

$$\begin{aligned} D_{B,M} &= \epsilon\,\sqrt{p}\,\sigma_{B,M}\sigma_{E,M} 2^{-R}, \\ D_{E,M} &= \epsilon\,\frac{1}{\sqrt{p}}\,\sigma_{B,M}\sigma_{E,M} 2^{-R} = D_{B,M}/p. \end{aligned} \tag{3.25}$$

In this case, the minimal objective function in (3.21) becomes

$$D_{min} = 2(1-p)D_{B,M} = 2\epsilon\sqrt{p}(1-p)\sigma_{B,M}\sigma_{E,M}2^{-R}. \tag{3.26}$$

Finally, plugging (3.25) into D_0 and D_1 in (3.9) yields the following distortion product D_0D_1:

$$D_0D_1 = \frac{1}{2}(1+p)\epsilon^2\sigma_{B,M}^2\sigma_{E,M}^2 2^{-2R}. \tag{3.27}$$

which will be further discussed in the DPCM case in Sec. 3.3.

The following remarks are in order.

Remark 1: Eq. (3.24) shows that more bits should be allocated to the prediction residual when the loss probability p is higher or when $\sigma_{E,M}^2$ is larger (the data are more difficult to predict). Notice that $R_1 = 0$ when $R < \frac{1}{2}\log_2\left(\frac{\sigma_{B,M}^2}{p\sigma_{E,M}^2}\right)$. In this case the method reduces to our previous approach in [25]. However, it should be emphasized that this threshold is derived based on the first-order Gauss-Markov model. For nonstationary signals like natural images, we will show in Sec. 3.5 that sending prediction residual is beneficial even at very low bit rates, because these bits are spent at regions with strong edges, and can thus significantly improve the reconstruction quality.

Remark 2: Eq. (3.26) shows that the optimal TDLT for the proposed MDC scheme needs to minimize $\sigma_{B,M}\sigma_{E,M}$, whereas the optimal single description transform should minimize $\sigma_{B,M}^2$ in (3.14). Therefore the optimal TDLT for the proposed MDC is different from the single description case, although the difference is not very much, as shown in Sec. 3.4.

Remark 3: Eq. (3.26) also shows that when the loss probability p changes, we always need to minimize $\sigma_{B,M}\sigma_{E,M}$. Therefore the optimal transform is independent of p. This is desired in practice since we only need to have one optimal transform for a large range of channel conditions. If R_1 is forced to 0, (3.26) will be invalid, and the optimal transform

will be a function of p. However, our optimization results in Sec. 3.4 show that the optimal transform is not sensitive to p.

Remark 4: The DCT can be considered as a special case of the TDLT when the prefilter $\mathbf{P} = \mathbf{I}$. In this case, the derivation above is still valid. The performance of the proposed method in the DCT case will be compared with the TDLT case in Sec. 3.4 and Sec. 3.5.

3.3 Asymptotic Performance of the Proposed Method

In this section, we analyze the asymptotic performance of our method by studying the DPCM case, which has the same performance as transform coding when the block size goes to infinity [21]. In the DPCM case, we split the data into even and odd samples, and use DPCM to encode each group. Each description also predicts the samples in the other group and encodes the prediction residual as the enhanced layer. Since the block size is $M = 1$, there is no block transform and lapped transform.

This special case of our method is indeed an improvement of Jayant's DPCM based method in [19, 20]. As discussed in previous chapter, Jayant's method splits the source at sample level and uses linear prediction between the two descriptions, but it does not encode the prediction residual. Therefore the performance of its side decoder is not satisfactory.

3.3.1 Optimization of the DPCM Case

Assume the input follows a first-order Gauss-Markov model with correlation coefficient r. After splitting, each part is still a first-order Gauss-Markov signal, but with correlation coefficient r^2. Let x_{n-1} and x_{n+1} be two consecutive samples in one description. If the DPCM is used in each description, the optimal prediction of x_{n+1} from $\hat{x}_{n-1}$ is given by

$\tilde{x}_{n+1} = r^2\hat{x}_{n-1}$. At high rates, the variance of the residual $e_{n+1} = x_{n+1} - \tilde{x}_{n+1}$ is

$$(1 - r^4)\sigma_x^2 + r^4\sigma_q^2 \approx (1 - r^4)\sigma_x^2 \triangleq \sigma_{B,1}^2, \tag{3.28}$$

where σ_q^2 is the quantization noise variance of x_{n-1} and is negligible at high rates.

Similar to (3.17) [41], the DPCM system also satisfies $\hat{x}_{n+1} - x_{n+1} = \hat{e}_{n+1} - e_{n+1}$. Therefore the reconstruction error fo $x(n+1)$ after DPCM with bit rate R_0 is

$$D_{B,1} = \epsilon\,\sigma_{B,1}^2 2^{-2R_0} = \epsilon\,(1 - r^4)\sigma_x^2 2^{-2R_0}. \tag{3.29}$$

This can be viewed as the counterpart for (3.14).

When one description is lost, each missing sample x_n is first predicted from $\hat{x}_{n-1}$ and $\hat{x}_{n+1}$, *i.e.*,

$$\tilde{x}_n = \mathbf{h}\,[\hat{x}_{n-1},\ \hat{x}_{n+1}]^T \triangleq \mathbf{h}\,\hat{\mathbf{x}}_2, \tag{3.30}$$

where the optimal solution for $\mathbf{h}$ is the Wiener filter. At high rates, it is approximately [20]

$$\mathbf{h} \approx R_{x_n\mathbf{x}_2}\,R_{\mathbf{x}_2\mathbf{x}_2}^{-1} = \frac{r}{1 + r^2}[1 \quad 1]. \tag{3.31}$$

In this case, the variance of the residual $x_n - \tilde{x}_n$ is

$$\sigma_{E,1}^2 = \frac{1 - r^2}{1 + r^2}\sigma_x^2 + \frac{2r^2}{(1 + r^2)^2}\sigma_{q0}^2 \approx \frac{1 - r^2}{1 + r^2}\sigma_x^2. \tag{3.32}$$

In Jayant's method, this error exists even at high rates. In our method, the prediction residual is further encoded at rate R_1 in the enhanced layer in each description, leading to a reduced distortion of

$$D_{E,1} = \epsilon\,\sigma_{E,1}^2 2^{-2R_1} = \epsilon\,\frac{1 - r^2}{1 + r^2}\sigma_x^2 2^{-2R_1}, \tag{3.33}$$

which is the counterpart for (3.20). Since (3.29) and (3.33) have the same format as (3.14) and (3.20), respectively, the bit allocation solution (3.24) to (3.21) can be applied here as well. From (3.28) and (3.32) we know that $\sigma_{B,1}^2/\sigma_{E,1}^2 = (1+r^2)^2$, so (3.24) becomes:

$$\begin{aligned} R_0 &= \min\left(R,\ \frac{1}{2}R + \frac{1}{4}\log_2\left(\frac{(1+r^2)^2}{p}\right)\right), \\ R_1 &= \max\left(0,\ \frac{1}{2}R - \frac{1}{4}\log_2\left(\frac{(1+r^2)^2}{p}\right)\right). \end{aligned} \tag{3.34}$$

3.3.2 Asymptotic Performance of the Proposed Method

It is shown in [52] (see also [43]) that under the high rate assumption, the product of the side distortion and the central distortion of a MDC scheme for a stationary source satisfies

$$D_0 D_1 \geq \frac{1}{4}\,\epsilon^2\, P_x^2 2^{-2R}, \tag{3.35}$$

where P_x is the entropy power of the source ([41], pp. 95). In the DPCM case of our method, substituting the bit allocation into (3.29), (3.33) and (3.9), we can get

$$D_0 D_1 = \frac{1}{2}\,\epsilon^2\,\sigma_{B,1}^2\,\sigma_{E,1}^2\,2^{-2R} = \frac{1}{2}\,\epsilon^2\,(1+p)(1-r^2)^2\sigma_x^4\,2^{-2R}. \tag{3.36}$$

This corresponds to (3.27) for $M = 1$. To gain more insights, notice that the variance of the innovation sequence of a first-order Gauss-Markov signal is $\sigma_w^2 = (1-r^2)\sigma_x^2$, which is also the entropy power of the signal, *i.e.*, $P_x = \sigma_w^2$. Therefore the distortion product in (3.36) can be written as

$$D_0 D_1 = \frac{1}{2}\,\epsilon^2\,(1+p) P_x^2\,2^{-2R}. \tag{3.37}$$

Comparing (3.37) and (3.35), the asymptotic performance of the proposed method is away from the theoretical bound by a factor of $2(1+p)$, or roughly 3 dB for small values of p. This is similar to the performance of the GPCT in [55] at high redundancies. However, it should

be noted that this result is for the first-order Gauss-Markov signal, which is stationary. We will show in Sec. 3.5 that for nonstationary signals such as natural images, the proposed method performs equally well as other MDC algorithms at high redundancies. For low redundancies, it significantly outperforms other methods, because the block-level prediction compensation is very suitable for nonstationary signals.

Fig. 3.2 compares the central, side and expected SNRs of Jayant's method and the DPCM case of the proposed method for $p = 0.1$. The input is assumed to be a first-order Gauss-Markov source with $r = 0.95$. The constant ϵ is chosen as 1. The figure shows that the side SNR of Jayant's method could not be improved at high rates, due to the existence of prediction error (3.32), which dominates the expected distortion at high rates. This problem is resolved in our method by encoding the prediction error. Thus both the side SNR and the expected SNR can be improved as the increase of the bit rate. However, this is achieved at the price of reduced central SNR. Fig. 3.2 also shows that when the bit rate is below about 2.6 bits/sample, no residual is encoded, and our method reduces to Jayant's method. However, we will show in Sec. 3.5 that for natural images, it is always helpful to encode the residuals.

3.4 Optimization Results

In this section, we show various optimized distortion products and filter coefficients for the proposed method. Table 3.1 summarizes the optimized distortion product $D_0 D_1$ under different configurations of the transform, block size M, description loss probability p, and value N of the Wiener filter. The input is assumed to be a first-order Gauss-Markov signal with $\sigma_x^2 = 1$ and $r = 0.95$. Similar to Fig. 3.2, $\epsilon = 1$ is selected. The bit rate is fixed at $R = 4$ bpp. The source codes to generate the results in this thesis can be downloaded from [27].

Three configurations of the proposed prediction compensated MDC algorithm are compared with the DPCM in Table 3.1. The first one jointly optimizes the prefilter and the

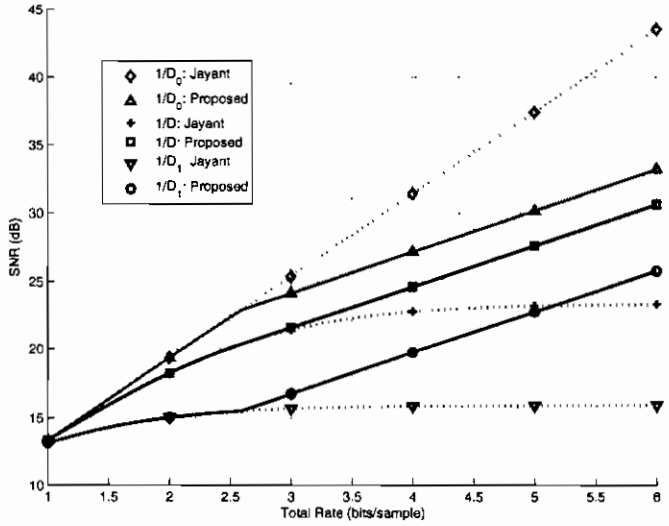


Figure 3.2: Comparison between the DPCM case of our method and Jayant's method (with $\sigma_x^2 = 1$, $r = 0.95$, $\epsilon = 1$).

Wiener filter to minimize the expected distortion in (3.10). We denote this method as multiple description lapped transform with prediction compensation (*MDLT-PC*). The second one is denoted as *TDLT-PC*, which uses the best filter for single description coding, *i.e.*, by optimizing (3.14). The last one is denoted as *DCT-PC*, which only uses the DCT, *i.e.*, no prefilter. In TDLT-PC and DCT-PC, Wiener filter and residual encoding are still used.

In Table 3.1, the performance of MDLT-PC already approaches the DPCM when $M = 16$. For $M = 8$, the distortion product of the three configurations is about 5%, 10% and 20% inferior to the DPCM, respectively. Reducing the size of the Wiener filter increases the distortion, but the change is less than 3%. For $M = 8$, even MDLT-PC with $N = 1$ has better performance than TDLT-PC with $N = 8$, showing the advantage of joint optimization. Another observation is that MDLT-PC with $M = 4$ is about 2.5% worse than DCT-PC with $M = 8$. Finally, it is shown in [25] that Wiener filter with $N = 1$ is optimal for DCT. In this case, it actually reduces to linear interpolation. Despite its simplicity, we will show in Sec.

Table 3.1: Distortion product D_0D_1 ($\times 10^{-5}$) of different configurations (with $\sigma_x^2 = 1$, $r = 0.95$, $\epsilon = 1$, and $R = 4$ bpp)

Transform	M	N	p = 0.01	p = 0.05	p = 0.1	p = 0.2
DPCM	1	1	1.88	1.95	2.04	2.23
MDLT-PC	16	16	1.89	1.97	2.06	2.25
	8	8	1.97	2.05	2.14	2.34
	8	2	2.00	2.08	2.18	2.38
	8	1	2.02	2.10	2.20	2.40
	4	4	2.29	2.38	2.49	2.72
TDLT-PC	8	8	2.06	2.14	2.24	2.45
	8	2	2.09	2.17	2.28	2.49
	8	1	2.11	2.19	2.30	2.50
DCT-PC	8	1	2.23	2.32	2.43	2.65

3.5 that the DCT-PC can still get better results than MMDSQ and RD-MDC in certain cases.

We show in Sec. 3.2 that the optimal transform is independent of p when $R_1 > 0$. Simulation results also show that the coding gain of the optimized transform does not change very much when R_1 is forced to 0. The optimal result is also not sensitive to R and N. For $M = 8$, when the bit rate R and the error probability p vary in a large range, the coding gain of the optimized TDLT only changes in a small region between 9.41 dB and 9.61 dB. This makes it possible to fix the transform and still achieves near optimal performance for all practical scenarios.

Two design examples will be used in the image coding in the next section. The first one is optimized for $M = 8$, $R = 1$ bpp, $p = 0.2$ and $N = 8$. The coding gain of the result is 9.53 dB, the product D_0D_1 is 0.00164, and the corresponding optimized matrix $\mathbf{V}$ in the TDLT prefilter is given by

$$\mathbf{V} = \begin{bmatrix} 0.8787 & 0.6591 & 0.2426 & 0.1521 \\ -0.5619 & 0.8044 & 0.5009 & 0.1444 \\ 0.1165 & -0.3914 & 0.9813 & 0.2933 \\ -0.0383 & 0.0129 & -0.1641 & 1.0875 \end{bmatrix}. \tag{3.38}$$

The second example is optimized for $M = 8$, $R = 1$ bpp, $p = 0.2$ and $N = 1$, with a coding gain of 9.54 dB and $D_0 D_1 = 0.00167$. The optimized matrix $\mathbf{V}$ is

$$\mathbf{V} = \begin{bmatrix} 0.9424 & 0.6840 & 0.1942 & 0.1046 \\ -0.5389 & 0.8658 & 0.5061 & 0.1037 \\ 0.1126 & -0.3688 & 1.0344 & 0.2778 \\ -0.0402 & -0.0001 & -0.1323 & 1.1183 \end{bmatrix}. \tag{3.39}$$

In this case, the corresponding 8×2 Wiener filter is given by

$$\begin{bmatrix} 0.67 & 0.63 & 0.59 & 0.54 & 0.46 & 0.41 & 0.37 & 0.33 \\ 0.33 & 0.37 & 0.41 & 0.46 & 0.54 & 0.59 & 0.63 & 0.67 \end{bmatrix}^T,$$

which can be easily implemented.

3.5 Image Coding Performance

In this section, we evaluate the performance of the proposed MDC method in the coding of natural images. Six 512×512 standard test images with very different characteristics are used. The block size M is selected to be 8. The two descriptions are generated by partitioning the transformed blocks in a checkerboard pattern such that the four immediate neighboring blocks of each block belong to the other description. All base layer blocks in each description are grouped together to form a 256×512 sub-image, which is then

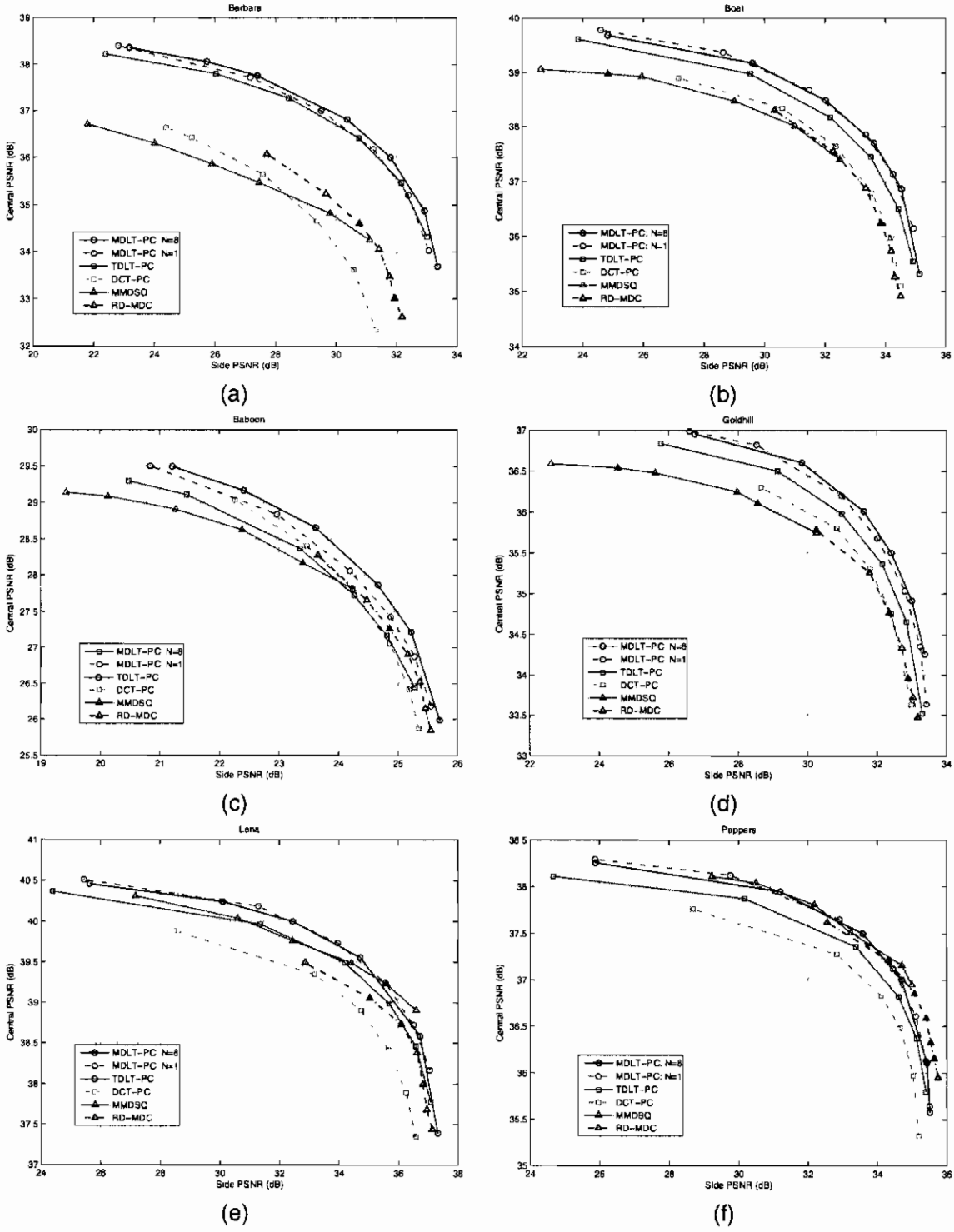


Figure 3.3: Trade-off between the central PSNR and the side PSNR for MDLT-PC, MMDSQ and RD-MDC at $R = 1$ bit/pixel. (a): Barbara; (b): Boat; (c): Baboon; (d): Goldhill; (e): Lena; (f): Peppers.

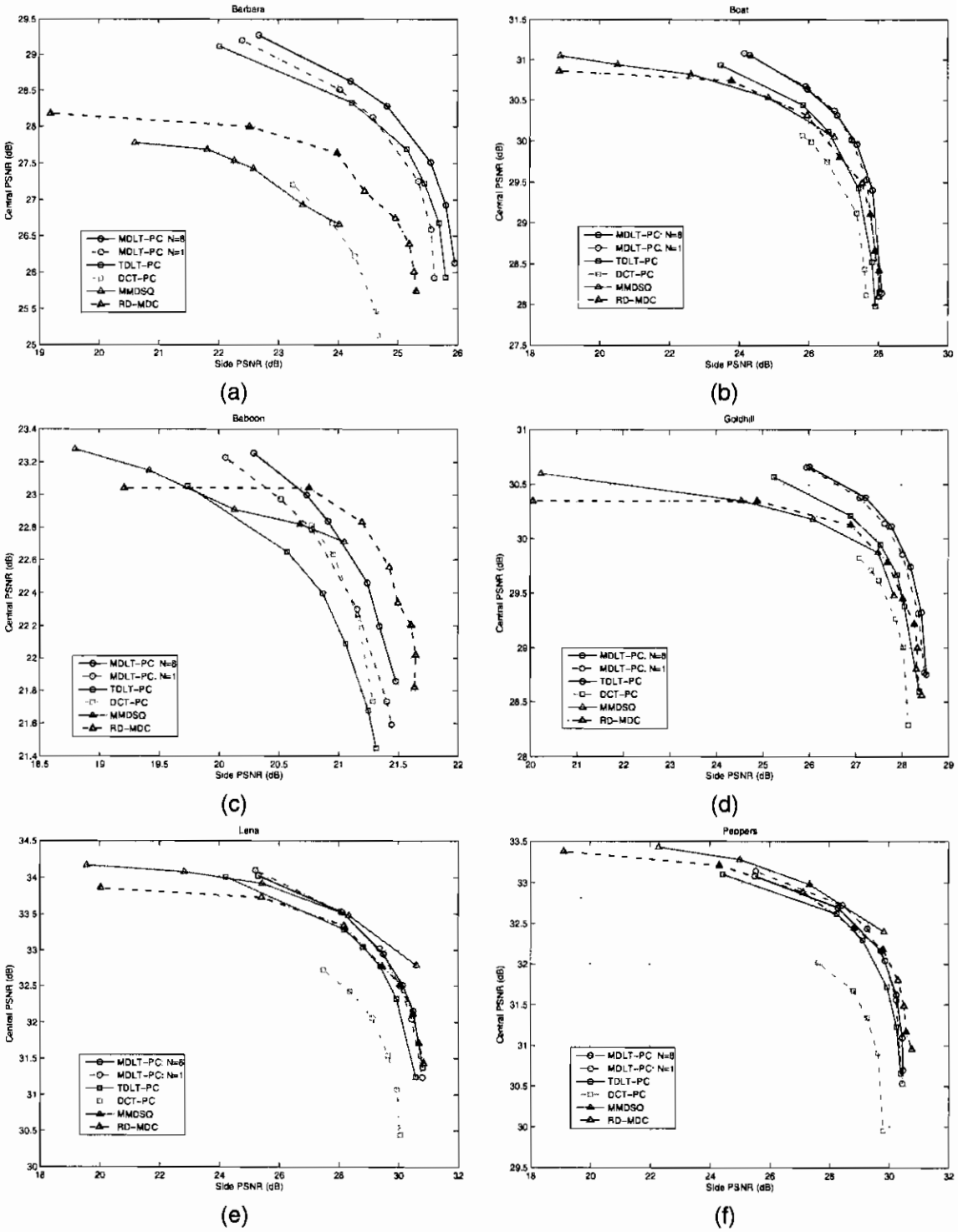


Figure 3.4: Trade-off between the central PSNR and the side PSNR for MDLT-PC, MMDSQ and RD-MDC at $R = 0.25$ bit/pixel. (a): Barbara; (b): Boat; (c): Baboon; (d): Goldhill; (e): Lena; (f): Peppers.

encoded by the embedded entropy coding in [47]. Similarly, all enhanced layer blocks in each description are also grouped together and encoded by the same entropy coding algorithm. The source codes and testing parameters for all results can be downloaded from [27].

The 1-D Wiener filter is applied to 2-D images in a separable manner. Firstly, each row of a block is Wiener-predicted using the same rows in the left and right neighboring blocks. Secondly, each column of the block is estimated using the same columns in the top and bottom neighboring blocks. After that, the average of the row and the column predictions is used as the final prediction of the 2-D block. The 2-D prediction residual is then calculated and encoded in the enhanced layer.

We first study the trade-off between the central and the side PSNRs. This is related to the distortion product $D_0 D_1$ and has been used as a performance measure in, for examples, [52, 36, 43, 45]. In our method, this can be easily achieved by varying the quantization steps of the two layers. In Fig. 3.3 and Fig. 3.4 the proposed method is compared with the wavelet and Tarp filter-based MMDSQ in [43], and the JPEG 2000-based RD-MDC in [45] (source code at [34]), which represent the state of the art in MDC. Four configurations of the proposed method are tested, namely, MDLT-PC with $N = 8$ and $N = 1$ (Eq. (3.38), (3.39)), TDLT-PC with $N = 8$ and DCT-PC with $N = 1$. The average of the two side PSNRs is used as the side PSNR in these figures. The overall bit rate R is chosen as 1 bpp and 0.25 bpp, respectively.

It can be seen that the performances of MMDSQ and RD-MDC are very similar in many cases. The MMDSQ performs better for smooth images like Lena and Peppers, whereas RD-MDC is better for images with more textures, such as Barbara.

Compared with MMDSQ and RD-MDC, our method outperforms in most cases of Barbara, Boat, Baboon, and Goldhill, many with large margins, especially at low redundancies. Given the same central PSNR, the side PSNR of our method can be more than 8 dB and 6

dB better for $R = 1$ bpp and $R = 0.25$ bpp, respectively. For smooth images like Lena and Peppers, our method can still achieve better performance in many low redundancy cases. At high redundancies, the three methods behave very similarly, especially for smooth images.

The results also show that the proposed method is not sensitive to the size of the Wiener filter. In most cases, the performance of the low-cost MDLT-PC with $N = 1$ is very similar to that of the MDLT-PC with $N = 8$, making it a good candidate for practical applications, due to its simpler Wiener filter.

When the single description optimized TDLT is used (TDLT-PC), the curves are lower than the jointly optimized MDLT-PC curves by less than 0.5 dB in all cases. If only the DCT is used (DCT-PC), the curves can be up to 2 dB lower, with more degradations at low rates. These relationships agree with Table 3.1 for Gauss-Markov sources. However, a surprising observation is that even the simple DCT-PC with linear interpolation can achieve similar or better performance than MMDSQ and RD-MDC at some high rate experiments. This underscores the importance of utilizing source correlation in MDC.

The first points of all curves of our method correspond to the case with $R_1 = 0$. In this case, the proposed method reduces to the prediction-only method in [25], and reasonable side PSNR is still obtained. In contrast, the side decoding performance of MMDSQ with low redundancies is quite poor, because most received bits are in the second layer and cannot be used. Similar defect also exists in the RD-MDC at low redundancies, because there is little information about half of the code-blocks.

Fig. 3.5 shows portions of some decoding results with one description. The three methods are compared at the same total bit rate and same central PSNR. Clearly, our method achieves significant improvement in both the PSNR and the visual quality.

In terms of complexity, the RD-MDC in [34] takes 34 seconds to encode an 512×512 image on a PC with 2.13GHz Intel Core 2 Duo CPU and 2GB memory, whereas both

MMDSQ and our method need less than one second. The decoding speeds of the three methods are all less than one second.

Finally, we compare our method with the PCT. The same condition as in Fig. 9 of [54] is used, where the central PSNR of image Lena is kept at 35.78 dB. To get this central PSNR, the JPEG-based single description coder in [54] needs a rate of $R^* = 0.60$ bpp, whereas the TDLT codec in [46] only needs 0.346 bpp, because of the improved transform and entropy coding.

Two side PSNRs of the PCT are reported in [54]: 27.94 dB at a total rate of $R = 0.688$ bpp and 29.63 dB at $R = 0.733$ bpp. The redundancy over R^* is 15% and 22%, respectively. To get the same side PSNRs, our MDLT-PC with $N = 8$ needs a rate of 0.379 and 0.400 bpp, and the redundancy is 9.6% and 15.7%, respectively. This shows that our method can achieve the same side PSNR with less redundancy than the PCT. The comparison is quite rough because of the difference in the codecs, but even if the PCT can be implemented in the TDLT framework, it is still difficult to surpass our method due to its various limitations, *e.g.*, it does not use any correlation between blocks, it cannot predict most of the coefficients at low redundancies, and its entropy coding will be compromised because of the coefficient-level split.

3.6 Conclusion

This chapter presents a MDC paradigm by integrating time-domain lapped transform, block level splitting, linear prediction and compensation. The method can fully utilize the source correlation while simultaneously provide effective redundancy control. Image coding results show that it outperforms existing methods in the literature significantly. Our source codes can be downloaded from [27].

The proposed method can be further improved. One possible solution is to improve the performance of the Wiener filter, as will be studied in the next chapter.

The method can also be generalized to create K descriptions, where $K > 2$. A direct generalization is to split the image into K sub-images after prefiltering. Each sub-image is coded as the base layer of one description. Each description also encodes the prediction residuals of all other sub-images as the second layer. At the decoder, if a block has a high-quality reconstruction via base layer coding, it is used directly. Otherwise, the average of all low-quality reconstructions of a block via enhanced layers is used as the final reconstruction. This will be studied in Chapter 5.

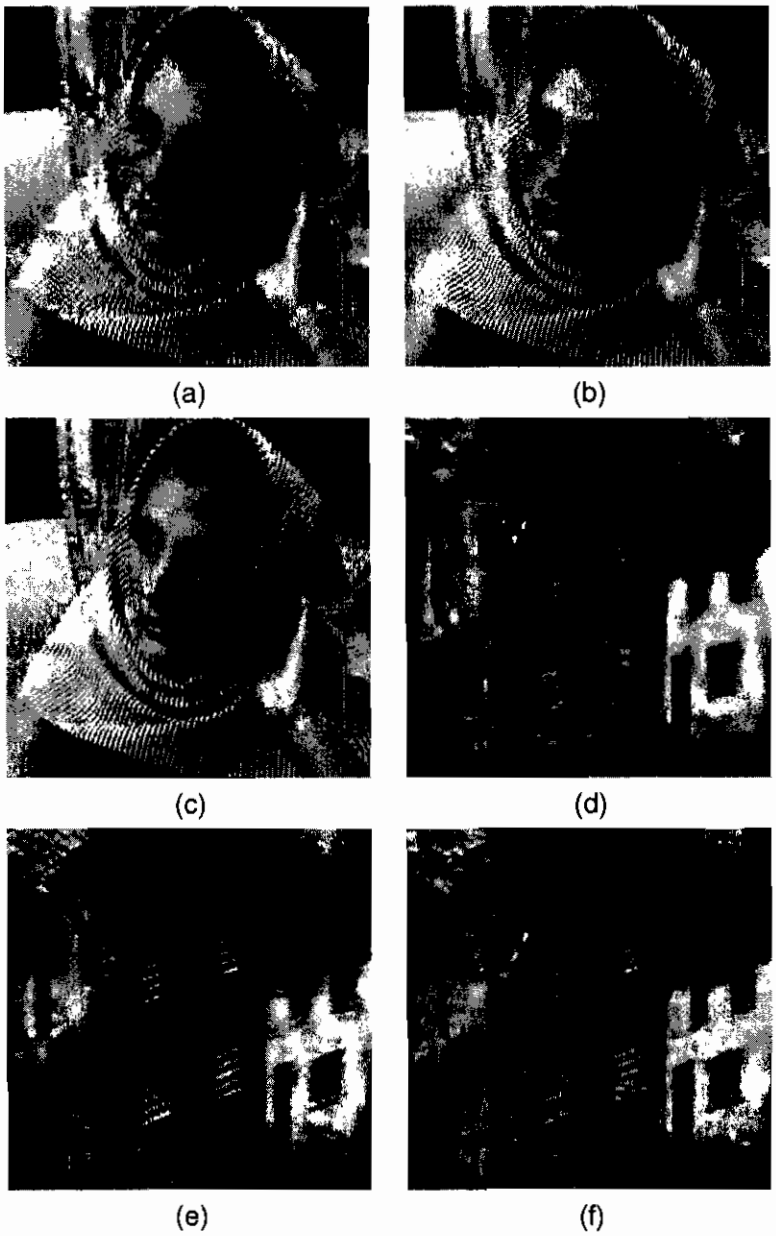


Figure 3.5: Side decoder results with total rate of 1 bpp in (a-c) and 0.25 bpp in (d-f). The central PSNR is included in the parentheses. (a): Barbara by MMDSQ: 24.52 dB (36.09 dB); (b): Barbara by RD-MDC: 27.60 dB (36.07 dB); (c): Barbara by MDLT-PC: 31.68 dB (36.07 dB); (d) Goldhill by MMDSQ: 24.58 dB (30.35 dB); (e) Goldhill by RD-MDC: 24.46 dB (30.35 dB); (f) Goldhill by MDLT-PC: 27.22 dB (30.35 dB).

Chapter 4

Improvement to MDLT-PC for image coding

4.1 Introduction

In this chapter, we investigate the techniques to improve the performance of the proposed multiple description codec on the image coding. At the side decoder, the reconstruction of the missing data involves two steps, the prediction from the first layer data of the received description, and the prediction residual from the second layer. Therefore if we can get a more accurate prediction of the lost data, the energy of the prediction residual will be reduced, and we will need less bits to encode the residual. The saved bits can be used to improve the quality of the base layer, which in turn will improve the overall rate-distortion performance of the system.

The Wiener filter method discussed in the previous chapter can be improved as follows. The design of the Wiener filter is based on one-dimensional (1-D) signal model, and the result is applied to two-dimensional (2-D) signal via the classic separate approach, i.e., each row of the lost block is first estimated from its horizontal neighbors. A vertical estimate

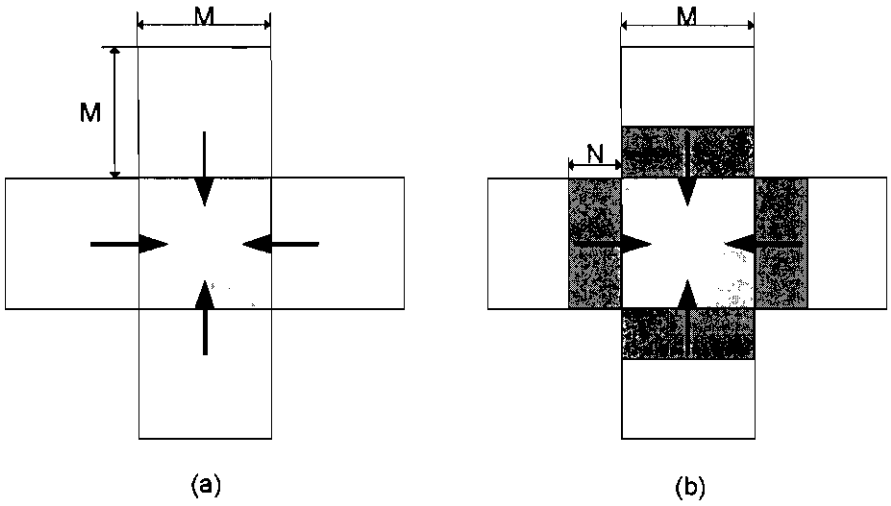


Figure 4.1: (a) Estimation of the lost block from four neighbors, (b) Estimation of the lost block from N layer of neighboring samples

is then obtained for each column, and the average of the two estimates is used as the final result. This simple treatment does not exploit the 2-D geometric structures of the input and can create some artifacts, especially near the edges. In this chapter, we will explore the feasibility of using two-dimensional Wiener filter to this problem.

4.2 Two dimensional prediction

4.2.1 2-D Wiener filter derivation

The descriptions of MDLT-PC are generated such that even and odd blocks are in different descriptions. Therefore in the two dimensional natural image, one block has its four direct neighboring blocks (top, bottom, left, right) in the other description. In other word, if one description is lost, each missing block can be predicted from four immediate neighboring

blocks. Figure 4.1(a) shows the estimation from the four neighboring blocks. If all neighboring data are used to estimate the lost block, the size of the corresponding 2-D Wiener filter will be $M^2 \times 4M^2$. Suppose $\mathbf{S}_0$ is the missing prefiltered block of size $M \times M$ to be predicted, and $\mathbf{S}_i, i = 1, \ldots, 4$ the top, bottom, left, and right neighbor of S_0, respectively. If we use $\vec{\mathbf{S}}_i, i = 0, \ldots, 4$ to represent the $M^2 \times 1$ vectors obtained by stacking the M columns of block $\mathbf{S}_i$ together, all coefficients of the 4-connection neighbors can be put into a $4M^2 \times 1$ vector as

$$\vec{\mathbf{S}}_{4C} \triangleq [\vec{\mathbf{S}}_1^T, \vec{\mathbf{S}}_2^T, \vec{\mathbf{S}}_3^T, \vec{\mathbf{S}}_4^T]^T \tag{4.1}$$

The 2-D Wiener filter is thus

$$\mathbf{H} = \mathbf{R}_{\vec{\mathbf{S}}_0 \vec{\mathbf{S}}_{4C}} \mathbf{R}^{-1}_{\vec{\mathbf{S}}_{4C} \vec{\mathbf{S}}_{4C}}, \tag{4.2}$$

where the $R_{\vec{\mathbf{S}}_0 \vec{\mathbf{S}}_{4C}}$ is the correlation matrix between $\vec{\mathbf{S}}_0$ and $\vec{\mathbf{S}}_{4C}$, and the $R_{\vec{\mathbf{S}}_{4C} \vec{\mathbf{S}}_{4C}}$ is the autocorrelation of the $\vec{\mathbf{S}}_{4C}$. The correlation matrices involved above can be obtained as the follows. Each $\mathbf{S}_i$ is a function of $2M \times 2M$ input samples $\mathbf{X}_i$. The relationship between $\mathbf{X}_i$ and $\mathbf{S}_i$ is thus

$$\mathbf{S}_i = \mathbf{P}_{12} \mathbf{X}_i \mathbf{P}_{12}^T, \tag{4.3}$$

Denote $\otimes$ the Kronecker product, equation (4.3) can be turned into the 1-D expression

$$\vec{\mathbf{S}}_i = (\mathbf{P}_{12} \otimes \mathbf{P}_{12}) \vec{\mathbf{X}}_i \triangleq \tilde{\mathbf{P}}_{12} \vec{\mathbf{X}}_i, \tag{4.4}$$

By the definition of $\vec{\mathbf{S}}_{4C}$ in (4.1), we have

$$\begin{aligned} \vec{\mathbf{S}}_{4C} &= \text{diag}\{\tilde{\mathbf{P}}_{12}, \tilde{\mathbf{P}}_{12}, \tilde{\mathbf{P}}_{12}, \tilde{\mathbf{P}}_{12}\} \left[\vec{\mathbf{X}}_1^T, \vec{\mathbf{X}}_2^T, \vec{\mathbf{X}}_3^T, \vec{\mathbf{X}}_4^T\right]^T \\ &\triangleq \tilde{\mathbf{P}}_{12,4} \vec{\mathbf{X}}_{4C}. \end{aligned} \tag{4.5}$$

From (4.4) and (4.5), we get

$$\begin{aligned}\mathbf{R}_{\vec{\mathbf{S}}_0\vec{\mathbf{S}}_{4C}} &= \tilde{\mathbf{P}}_{12}\mathbf{R}_{\vec{\mathbf{X}}_0\vec{\mathbf{X}}_{4C}}\tilde{\mathbf{P}}_{12,4}^T, \\ \mathbf{R}_{\vec{\mathbf{S}}_{4C}\vec{\mathbf{S}}_{4C}} &= \tilde{\mathbf{P}}_{12,4}\mathbf{R}_{\vec{\mathbf{X}}_{4C}\vec{\mathbf{X}}_{4C}}\tilde{\mathbf{P}}_{12,4}^T.\end{aligned} \tag{4.6}$$

Our experimental result shows that the performance is not sensitive to the size of the Wiener filter. Therefore, we can use less number of samples for prediction and reduce the computation cost. The reduced sized estimation is shown in Fig. 4.1(b). The downsized 2-D Wiener filter is a $M^2 \times 4MN$ matrix, where N is number of layers of the boundary samples from each of four neighboring blocks for estimation. $\vec{\mathbf{S}}_{4C,N}$ is the vector of Wiener filter input samples, which are the gray areas of the neighboring blocks in Fig. 4.1(b). It is a subset of $\vec{\mathbf{S}}_{4C}$. The downsized Wiener filter can be written as:

$$\mathbf{H}_N = \mathbf{R}_{\vec{\mathbf{S}}_0\vec{\mathbf{S}}_{N,4C}}\mathbf{R}^{-1}_{\vec{\mathbf{S}}_{N,4C}\vec{\mathbf{S}}_{N,4C}}, \tag{4.7}$$

The correlation matrix $\mathbf{R}_{\vec{\mathbf{S}}_0\vec{\mathbf{S}}_{4C,N}}$ and $\mathbf{R}_{\vec{\mathbf{S}}_{N,4C}\vec{\mathbf{S}}_{N,4C}}$ are the submatrices of $\mathbf{R}_{\vec{\mathbf{S}}_0\vec{\mathbf{S}}_{4C}}$ and $\mathbf{R}_{\vec{\mathbf{S}}_{4C}\vec{\mathbf{S}}_{4C}}$ respectively. Some column or row reordering operations may be required, depending on how the input vector is stacked.

4.2.2 Objective Function and Optimal Rate Allocation

Given the target bit rate R and the probability p of losing one description, our objective is to find the optimal prefilter $\mathbf{P}$ and the optimal bit allocation R_0 and R_1 that minimize the expected distortion. The expected distortion D is defined as

$$D = (1-p)^2 D_0 + 2p(1-p)D_1. \tag{4.8}$$

In the proposed MDC scheme, each description contains base layer blocks and the enhanced layer blocks. Only base layer blocks are used if two descriptions are received. If

one of them is lost, half of the image is reconstructed via base layer blocks and the other half is via prediction. Therefore D_0 and D_1 can be written as

$$\begin{aligned} D_0 &= D_{M,I}, \\ D_1 &= \frac{1}{2}(D_{M,I} + D_{M,P}), \end{aligned} \tag{4.9}$$

where $D_{M,I}$ and $D_{M,P}$ are the MSE contributed by base layer coded blocks and prediction-based blocks, respectively. The subscript M denotes the block size. Substituting into (4.8), we have

$$\begin{aligned} D &= (1-p)^2 D_{M,I} + p(1-p)(D_{M,I} + D_{M,P}) \\ &= (1-p)D_{M,I} + p(1-p)D_{M,P}. \end{aligned} \tag{4.10}$$

Let's note the prediction residual $M \times M$ matrix as $\mathbf{d}_0$ and the prediction residual $M \times M$ matrix after DCT as $\mathbf{t}_0$. We have

$$\vec{\mathbf{t}}_0 = \tilde{C}\vec{\mathbf{d}}_0 = \tilde{C}(\vec{\mathbf{S}}_0 - H\vec{\mathbf{S}}_{4C}), \tag{4.11}$$

where $\tilde{C} = C \otimes C$; $\vec{\mathbf{t}}_0$ and $\vec{\mathbf{d}}_0$ are the vectors which are stacked from $\mathbf{t}_0$ and $\mathbf{d}_0$. The encoder will quantize the prediction residual $\mathbf{t}_0$ to generate enhanced layer blocks. Therefore, at the decoder side, the distortion of the reconstructed losing block is introduced mainly by the quantization noise of the prediction residual if the total bit rate is high.

Let us note the $M \times M$ matrix, $\mathbf{e}_0$, as the quantization noise of the prediction residual blocks before inverse DCT. After the inverse DCT and post filter, the quantization noise is extended to the $2M \times 2M$ region. Because of the linearity of the transform and the additive of the quantization noise, the final distortion can be obtained by passing the quantization noise $\mathbf{e}_0$ through the synthesis filter. The synthesis filter, which includes the inverse DCT

and post filter, is a $2M \times M$ matrix, noted as G. The quantization noise in the output signal can be written as

$$\mathbf{v}_0 = G\mathbf{e}_0 G^T \tag{4.12}$$

Let

$$\tilde{G} = G \otimes G \tag{4.13}$$

The vector format of the quantization noise $\mathbf{v}_0$ can be written as:

$$\vec{\mathbf{v}}_0 = \tilde{G}\vec{\mathbf{e}}_0 \tag{4.14}$$

Therefore, the 2-D signal is written in form of 1-D vector, and the transform can also be regarded as 1-D transform. We can obtain the bit allocation and distortion in the same way as the 1-D case. The distortion for the reconstructed missing block is

$$D_{M,P\_area} = 2^{-2R_1} \prod_{i=0}^{M^2-1} (\sigma_{t_i}^2 ||g_i||^2)^{\frac{1}{M^2}} \tag{4.15}$$

Where, R_1 is the bit rate to transmit the enhanced layer blocks(prediction residuals); the $\sigma_{t_i}^2$ are the diagonal elements of the correlation matrix $R_{\vec{\mathbf{t}}_0\vec{\mathbf{t}}_0}$, which is

$$R_{\vec{\mathbf{t}}_0\vec{\mathbf{t}}_0} = \tilde{C} R_{\vec{\mathbf{d}}_0\vec{\mathbf{d}}_0} \tilde{C}^T. \tag{4.16}$$

$||g_i||^2, i = 1, \ldots, M^2$ is the norm of i-th columns the filter $\tilde{G}$. The distortion is calculated over a $2M \times 2M$ area. For the base layer block part, we will have the similar result

$$D_{M,I\_area} = 2^{-2R_0} \prod_{i=0}^{M^2-1} (\sigma_{y_i}^2 ||g_i||^2)^{\frac{1}{M^2}} \tag{4.17}$$

Where $\sigma^2_{y_i}$ is the diagonal element of the correlation matrix $R_{\vec{\mathbf{y}_0}\vec{\mathbf{y}_0}}$, which is

$$R_{\vec{\mathbf{y}_0}\vec{\mathbf{y}_0}} = \tilde{C} R_{\vec{\mathbf{s}_0}\vec{\mathbf{s}_0}} \tilde{C}^T. \tag{4.18}$$

Therefore, the distortions per pixel for the two types of blocks are

$$\begin{aligned} D_{M,P} &= \frac{1}{4M^2} 2^{-2R_1} \prod_{i=0}^{M^2-1} (\sigma^2_{t_i} ||g_i||^2)^{\frac{1}{M^2}} \triangleq \frac{1}{4M^2} 2^{-2R_1} \sigma^2_{M,P} \\ D_{M,I} &= \frac{1}{4M^2} 2^{-2R_0} \prod_{i=0}^{M^2-1} (\sigma^2_{y_i} ||g_i||^2)^{\frac{1}{M^2}} \triangleq \frac{1}{4M^2} 2^{-2R_0} \sigma^2_{M,I} \end{aligned} \tag{4.19}$$

To find the optimal R_0 and R_1 that minimize the expected distortion D, we substitute (4.19) into (4.10), and the problem can be simplified into

$$\begin{aligned} &\text{min } (1-p)\epsilon\, \sigma^2_{M,I} 2^{-2R_0} + p(1-p)\epsilon\, \sigma^2_{M,P} 2^{-2R_1}, \\ &\text{s.t.} \quad R_0 + R_1 = R. \end{aligned} \tag{4.20}$$

It can be solved by Lagrangian method. With the constrain of $R_0 \leq R$ and $R_1 \geq 0$, the optimal bit allocation can be found to be:

$$\begin{aligned} R_0 &= \min\left(R, \frac{R}{2} + \frac{1}{4}\log_2 \frac{\sigma_0^2}{p\sigma_1^2}\right) \\ R_1 &= \max\left(0, \frac{R}{2} - \frac{1}{4}\log_2 \frac{\sigma_0^2}{p\sigma_1^2}\right) \end{aligned} \tag{4.21}$$

4.3 2-D Statistical Models

The 2-D Wiener filter (4.2) requires the knowledge of the 2-D auto correlation matrix of the input. AR(1) model is the common model used for the natural images. There are two ways to generate the 2-D autocorrelation matrix by using AR(1), the separable model and the

isotropic model. In general, the isotropic model is a more accurate representation of the natural images, and we will use this model for the designs and experiments in this chapter. In the isotopic model, the correlation of two points, $r_{xx}(h,k)$, can be found to be

$$r_{xx}(h,k) = \sigma_x^2 \rho^{\sqrt{h^2+k^2}}, \tag{4.22}$$

where the h and k are the horizontal and vertical distant, respectively; the σ_x^2 is the variance of the input data, the ρ is the correlation coefficient. The value of ρ is chosen to be 0.95.

4.4 Optimization and Image Coding Performance

The optimization for the pre- and post-filters in lapped transform for 2-D Wiener filter is similar to one dimension case. There is no closed-form solution for the problem, and numerical optimization is used instead. The initial point of the optimization is the one obtained from the one-dimensional case. It turns out that the optimized transform using 2-D Wiener filter is very similar to the transform optimized for 1-D Wiener filter. This means we can use the same transform for both 1-D and 2-D cases. Therefore, we still use the transform given by (3.38) in this following experiments.

The testing images are 512 standard test images with different characteristics. The block size M is selected to be 8. All base layer blocks in each description are grouped together to form a 256×512 sub-image, which is then encoded by the embedded entropy coding in [47]. Similarly, all enhanced layer blocks in each description are also grouped together and encoded by the same entropy coding algorithm.

To implement the 2-D Wiener filter, all related samples are put into a vector by stacking the columns of the four neighboring blocks. So the input vector is a $4M^2 \times 1$ vector for the full size Wiener filter and $4MN \times 1$ for the reduced-complexity Wiener filter. The 2-D prediction residual is then calculated and encoded in the enhanced layer.

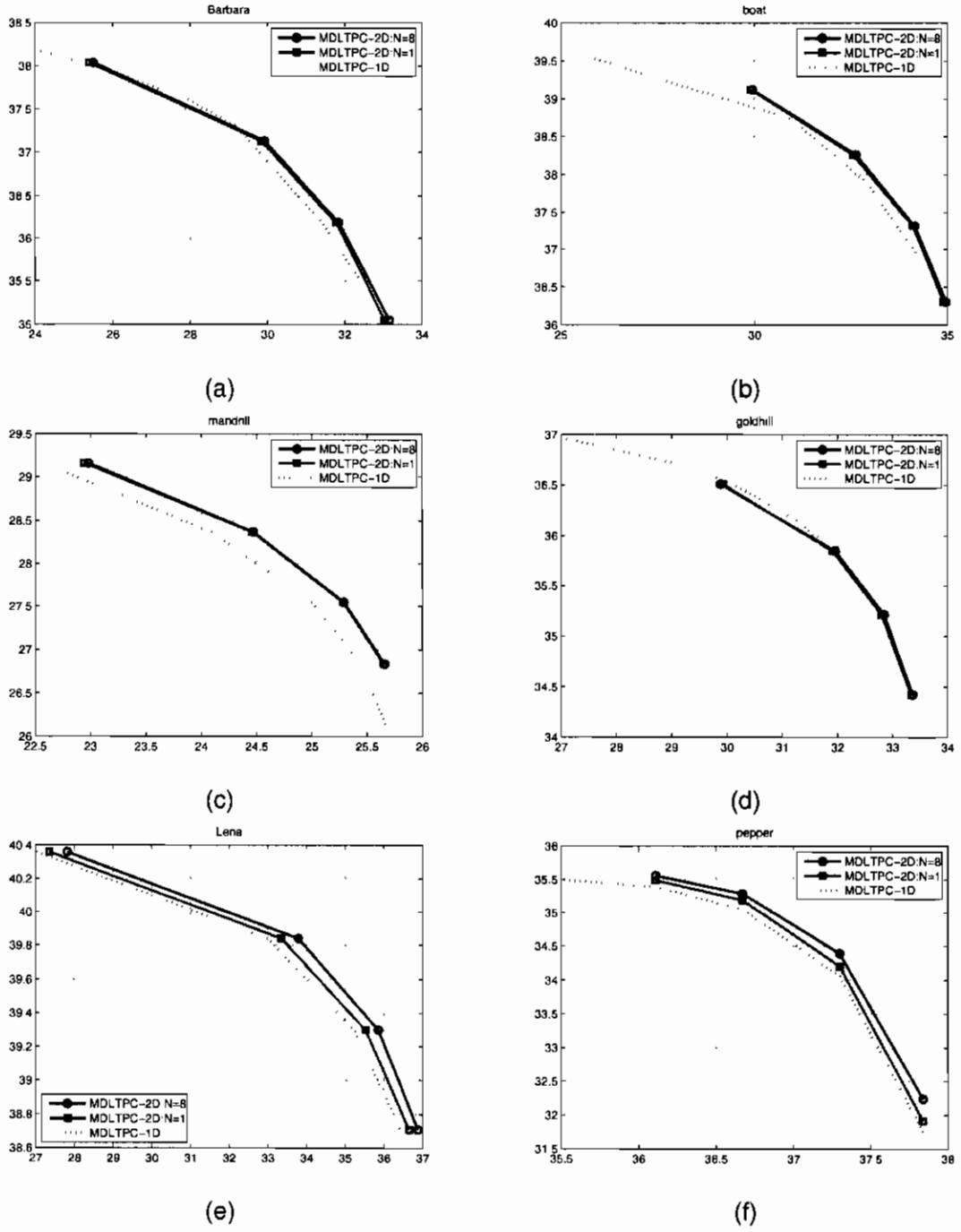


Figure 4.2: Trade-off between the central PSNR and the side PSNR for MDLT-PC with 2-D Wiener filter and 1-D Wiener filter at $R = 1$ bit/pixel. (a): Barbara; (b): Boat; (c): mandrill; (d): Goldhill; (e): Lena; (f): Peppers.

We first study the trade-off between the central and the side PSNRs. This trade off is achieved by varying the quantization steps of the two layers. In Fig. 4.2, we compares the proposed MDC with 1-D Wiener filter and 2-D Wiener filter. We have three configurations for the 2-D Wiener filter regarding its size and the adaptivity to the encoded images, namely, MDLTPC-2D with N = 8, and N = 1. The average of the two side PSNRs is used as the side PSNR in these figures. The overall bit rate R is chosen as 1 bpp.

Compared to the one-dimensional Wiener filter, the 2-D Wiener filter prediction does give some improvements. The gain is about 0.3 dB. In most of the cases, the performance of the 2-D Wiener filter with N = 1 is very similar to that of the 2-D Wiener filter with N = 8.

One interesting thing is that the improvement at the first point of all curves is the largest. This points corresponds to the case with $R_1 = 0$, which means that no enhanced layer data are transmitted. In this case, the system is degraded into the prediction-only mentioned in [24]. As the bit rate allocated to the enhanced layer increases, the improvement of both 2-D Wiener filter is reduced. This is because although the 2-D Wiener filter can reduce the MSE of the prediction, the improvement is quite limited, since there is no perfect prediction. As more bits are allocated to the enhanced layer, the quantization error drops much faster than the drop of the prediction residual. Therefore the improvement of the 2-D Wiener filter becomes less important.

4.5 Conclusion

This chapter discusses the techniques which can be used to improve the performance of the proposed MDC paradigm.By introducing the 2-D Wiener filter, we can improve the central and side distortion curve by 0.2 dB. However, as the enhanced layer gets more bits, the improvement of the 2-D prediction becomes less obvious. It should also be noticed that the computation cost of 2-D predictions are much higher than the 1-D case.

Chapter 5

MDLT-PC with More Descriptions

As discusses before, the multiple description coding can improve the robustness of the system to packet loss. However, MDC with more than two descriptions has not been well studied. In [12], it is proposed to obtain MDC with many channels by cascading various smaller pairwise correlating transforms. However, the complete solution is still unknown. A n-channel MDC is proposed in [30, 31], using the distributed source coding theory. A practical design example using scalar quantizer and channel code is reported in [32]. In [53], the MDC with individual and central receivers is studied. In [42], a sequential design of the MDSQ is proposed, which facilitates the design of MDSQ with three descriptions. Two such examples are developed in it. However, the sequential MDSQ method is difficult to be generalized to MDC with more than three descriptions.

In this chapter, we design the optimal lapped transform for the MDLT-PC with three descriptions and four descriptions, and demonstrate their performance in image coding. As will be shown later, the generation of four description coding is actually very similar to the two-description case, therefore we will focus on the design of three-description coding.

5.1 Overview of the 3-D System

To generate three descriptions, the prefiltered blocks are split into three subsets $\{\mathbf{s}_0(k) = \mathbf{s}(3k)\}$, $\{\mathbf{s}_1(k) = \mathbf{s}(3k+1)\}$, and $\{\mathbf{s}_2(k) = \mathbf{s}(3k+2)\}$. We call these blocks the *intra blocks* of each description. After DCT, quantization, and entropy coding, each subset of intra blocks form the base layers of one description. To improve the reconstruction quality when some descriptions are lost, in each description we also include the prediction residual of other subsets as the enhancement layer, usually at much lower bit rate than the base layer. The prediction is obtained by Wiener filter. In analogy to motion estimation and motion compensation in video coding, we refer these prediction residuals as the *inter blocks*.

At the decoder side, if all descriptions are received, the decoded intra blocks from all descriptions are combined. The postfilter is then applied to obtain the reconstructed signal. The inter block bits in each description are simply discarded. These data are therefore the redundant data of the MDC system. If some descriptions are lost, the missing blocks are first estimated from the received intra blocks by Wiener filter. The decoded inter blocks (residuals) are then added to the estimation before postfiltering.

To simplify the encoder and decoder structure, each description includes the prediction residual of all other blocks than are not intra-coded in that description. In other words, one third of the image blocks are intra-coded and the remaining two thirds are inter-coded in each description. As a result, if only two descriptions are received, each of them can generate a reconstruction for the lost subset. We take the average of the two reconstructions are the final reconstruction. This can reduce the MSE by one half.

The design of the system includes determining the optimal prefilter for the lapped transform, the optimal prediction filter for lost blocks, the optimal bit allocation between the intra blocks and inter blocks, as well as the bit allocation within a block. The objective is to minimize the expected distortion at the receiver.

We use R (in bits/pixel/description)) to represent the bit rate of each description. R_0 and

R_1 denote the average bits for each intra block pixel and each inter block pixel, respectively. Therefore if there are three descriptions, the bit rate for each description is $R = \frac{1}{3}(R_0+2R_1)$ bpp/description, or $3R = R_0 + 2R_1$.

As in the previous chapters, we use $\mathbf{x}(i)$, $\mathbf{s}(i)$, $\mathbf{y}(i)$ and $\mathbf{q_y}(i)$ to denote the i-th block of prefilter input, DCT input, DCT output and quantization noise of intra blocks, respectively. We also use $\mathbf{d}(i)$, $\mathbf{u}(i)$ and $\mathbf{q_u}(i)$ to denote the i-th inter block (prediction residual), its DCT output and quantization noise, respectively, as shown in Fig. 3.1. The reconstruction of a variable is denoted by the hat operator.

5.2 The Wiener Filter

To facilitate the R-D optimization of the problem, we use Wiener filter to estimate the blocks in other descriptions. Here we only give the formulas for the first description, where $\{\mathbf{s}(3k)\}$ are intra-coded. The formulas for other descriptions can be derived similarly. Define the two neighboring intra blocks as

$$\mathbf{s}_{I2} = \begin{bmatrix} \mathbf{s}^T(3k) & \mathbf{s}^T(3k+3) \end{bmatrix}^T, \tag{5.1}$$

The Wiener filters for estimating $\mathbf{s}(3k+1)$ and $\mathbf{s}(3k+1)$ from $\mathbf{s}_{I2}$ are

$$\begin{aligned} \mathbf{H_1} &= \mathbf{R}_{\mathbf{s}(3k+1)\mathbf{s}_{I2}}\mathbf{R}^{-1}_{\mathbf{s}_{I2}\mathbf{s}_{I2}}, \\ \mathbf{H_2} &= \mathbf{R}_{\mathbf{s}(3k+2)\mathbf{s}_{I2}}\mathbf{R}^{-1}_{\mathbf{s}_{I2}\mathbf{s}_{I2}}. \end{aligned} \tag{5.2}$$

As usual, the quantization noise is ignored in the Wiener filter. All matrices in the Wiener filters can be obtained by assuming an input signal model (AR(1) model is used in this paper) and considering the effect of the prefilter in the TDLT.

As in [25], we also normalize the Wiener filter so that all row sums are equal to 1. In addition, a special Wiener filter is used at the boundary to predict a missing block from only

one neighboring block.

To apply the Wiener filter in (5.2) to 2-D images, we first estimate each row of a lost block using the horizontal neighbors, and then estimate each column of the block using vertical neighbors. The average of the horizontal and vertical estimations is used as the final prediction.

5.2.1 Objective Function and Optimal Rate Allocation

Given the target bit rate R and the probability p of losing one description, our objective is to find the optimal prefilter $\mathbf{P}$ and the optimal bit allocation R_0 and R_1 for intra and inter blocks that minimize the expected distortion. We use D_i to denote the MSE when i descriptions are received. We also assume that all descriptions are balanced, *i.e.*, they have the same rate and distortion. The expected distortion D is therefore defined as

$$D = (1-p)^3 D_3 + 3p(1-p)^2 D_2 + 3p^2(1-p)D_1 + p^3\sigma_x^2. \tag{5.3}$$

Note that the scenario of losing all descriptions is also considered to enable fair comparison with the MDC with two descriptions in [39].

Let D_I and D_P be the MSE contributed by intra-coded blocks and prediction-based inter-coded blocks, respectively. The subscript M denotes the block size

When three descriptions are received, only intra-coded data are used in reconstruction. Thus $D_3 = D_I$. When only one description is received, one third of the signal is intra coded, whereas two thirds of the signal are inter-coded. Therefore the MSE is $D_1 = 1/3\,(D_I + 2D_P)$. If two descriptions are received, each of them can generate a reconstruction of the missing description. By averaging the two reconstructions, the MSE of the missing data can be halved. Therefore the overall MSE is $D_2 = 1/3\,(2D_I + 1/2D_P)$.

To find D_I, let the quantization noise of $\mathbf{y}(n)$ be $\mathbf{q_y}(n)$. After the inverse TDLT, the reconstruction error becomes $\mathbf{G}\mathbf{q_y}(n)$. As usual, we assume the quantization noises of

different subbands are uncorrelated. Therefore the MSE of the reconstruction is

$$D_I = \frac{1}{M} \sum_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma^2_{q_y(i)}, \tag{5.4}$$

where $\sigma^2_{q_y(i)}$ is the variance of the quantization noise of the i-th entry of $\mathbf{y}(n)$, and $\mathbf{g}_i$ is the i-th column of $\mathbf{G}$. At high rates, $\sigma^2_{q_y(i)}$ can be written as

$$\sigma^2_{q_y(i)} = \epsilon\, \sigma^2_{y(i)} 2^{-2R_{0i}}, \tag{5.5}$$

where ϵ is a constant that depends on the input statistics and the quantization scheme. For entropy constrained scalar quantization of Gaussian sources, we have $\epsilon = \pi e/6$. R_{0i} is the bits allocated to the i-th entry of an intra block, and the average intra block bit rate is $\frac{1}{M}\sum_{i=1}^{M} R_{0i} = R_0$. $\sigma^2_{y(i)}$ is the variance of the entry, which is the i-th diagonal element of the autocorrelation matrix $\mathbf{R_{yy}}$:

$$\mathbf{R_{yy}} = \mathbf{F}\mathbf{R}_{\mathbf{x_2x_2}}\mathbf{F}^T, \tag{5.6}$$

and $\mathbf{x}_2 = \left[\mathbf{x}^T(n)\ \ \mathbf{x}^T(n+1)\right]^T$.

Upon optimal bit allocation [21], the minimal value for (5.4) is given by

$$D_I = \epsilon \left(\prod_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma^2_{y(i)} \right)^{\frac{1}{M}} 2^{-2R_0} \triangleq \epsilon\, \sigma^2_{M,I} 2^{-2R_0}. \tag{5.7}$$

This is in fact the objective function of the single description coding (SDC). For block transforms, the minimum value of (5.7) is achieved by the KLT, of which the DCT is a close approximation, if the input follows an AR(1) model with strong correlation. For lapped transforms and longer filter banks, there is no closed form expression for the optimal solution, but numerical optimization method can be used to find the solution that minimizes (5.7).

If only one description is received, the error contributed by intra-coded blocks is the

same as above. To find D_P, the MSE caused by inter-coded blocks, we first write the reconstruction of $\mathbf{s}(3k+j)$ $(j=1,2)$ as

$$\hat{\mathbf{s}}(3k+j) = \hat{\mathbf{d}}(3k+j) + \mathbf{s}_{H_1}(3k+j), \tag{5.8}$$

where $\hat{\mathbf{d}}(3k+j)$ and $\mathbf{s}_{H_1}(3k+j)$ are the reconstruction of $\mathbf{d}(3k+j)$ and the prediction of $\mathbf{s}(3k+j)$ from $\hat{\mathbf{s}}(3k)$ and $\hat{\mathbf{s}}(3k+3)$, respectively. Since the prediction residual $\mathbf{d}(3k+j)$ at the encoder is given by

$$\mathbf{d}(3k+j) = \mathbf{s}(3k+j) - \mathbf{s}_{H_1}(3k+j), \tag{5.9}$$

the following relationship can be obtained from (5.8) and (5.9):

$$\hat{\mathbf{s}}(3k+j) - \mathbf{s}(3k+j) = \hat{\mathbf{d}}(3k+j) - \mathbf{d}(3k+j). \tag{5.10}$$

In other words, the reconstruction error of $\mathbf{s}(3k+j)$ is equal to that of the prediction residual $\mathbf{d}(3k+j)$. This is indeed a property of any differential coding system [41].

Since $\mathbf{d}(3k+j)$ and $\mathbf{u}(3k+j)$ are related by a DCT transform, we have

$$\hat{\mathbf{d}}(3k+j) - \mathbf{d}(3k+j) = \mathbf{C}^T\mathbf{q}_\mathbf{u}(3k+j). \tag{5.11}$$

After postfiltering, the reconstruction error becomes $\mathbf{T}_{21}\mathbf{C}^T\mathbf{q}_\mathbf{u}(3k+j) = \mathbf{G}\mathbf{q}_\mathbf{u}(3k+j)$, and its MSE is

$$D_P(3k+j) = \frac{1}{M}\sum_{i=0}^{M-1} \epsilon\,||\mathbf{g}_i||^2 \sigma^2_{u_i}(3k+j)2^{-2R_{1i}(3k+j)}. \tag{5.12}$$

where $\sigma^2_{u_i}(3k+j)$ is the i-th diagonal element of the autocorrelation matrix of $\mathbf{u}(3k+j)$ as given by

$$\mathbf{R}_{\mathbf{uu}}(3k+j) = \mathbf{C}\{\mathbf{R}_{\mathbf{ss}}(3k+j) - \mathbf{H}_j\mathbf{R}_{\mathbf{s}_{I2}\mathbf{s}(3k+j)}\}\mathbf{C}^T. \tag{5.13}$$

Since Eq. (5.12) has the same format as the reconstruction error caused by $\mathbf{q_y}(n)$ in (5.4), the derivation from (5.4) to (5.7) can be applied here, and the minimal value of $D_P(3k+j)$ after optimal bit allocation is therefore

$$D_P(3k+j) = \epsilon \left(\prod_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma_{u_i}^2(3k+j) \right)^{\frac{1}{M}} 2^{-2R_1}. \tag{5.14}$$

Since we assume the input signal a stationary random process, we can easily proved that $\sigma_{u_i}^2(3k+1) = \sigma_{u_i}^2(3k+2)$, thus j can be dropped.

$$D_P = \epsilon \left(\prod_{i=0}^{M-1} ||\mathbf{g}_i||^2 \sigma_{u_i}^2 \right)^{\frac{1}{M}} 2^{-2R_1} \triangleq \epsilon\, \sigma_{M,P}^2 2^{-2R_1}. \tag{5.15}$$

The expected distortion is thus

$$\begin{aligned} D = & (1-p)^3 D_I \\ & + 3p(1-p)^2 \frac{1}{3} \left(2D_I + \frac{1}{2} D_P \right) \\ & + 3p^2(1-p) \frac{1}{3} \left(D_I + 2D_P \right) \\ & + p^3 \sigma_x^2. \end{aligned} \tag{5.16}$$

After simplification, we get

$$D = (1-p)\epsilon \sigma_{M,I}^2 2^{-2R_0} + \frac{1}{2} p(1-p)(1+3p)\epsilon \sigma_{M,P}^2 2^{-2R_1} + p^3 \sigma_x^2. \tag{5.17}$$

The remaining bit allocation issue is to determine the optimal R_0 and R_1 that minimize the expected distortion D, subject to $R_0 + 2R_1 = 3R$.

$$\begin{aligned} & \text{min } (1-p)\epsilon\, \sigma_{M,I}^2 2^{-2R_0} + \frac{1}{2} p(1-p)(1+3p)\epsilon\, \sigma_{M,P}^2 2^{-2R_1}, \\ & \text{s.t.} \quad R_0 + 2R_1 = 3R. \end{aligned} \tag{5.18}$$

The problem can be solved by Lagrangian method. After considering the constraints of $R_0 \leq R$ and $R_1 \geq 0$, the optimal bit allocation can be found to be

$$R_0 = \min\left(R, R + \frac{1}{3}\log_2 \frac{4\sigma_{M,I}^2}{p(1+3p)\,\sigma_{M,P}^2}\right),$$
$$R_1 = \max\left(0, R - \frac{1}{6}\log_2 \frac{4\sigma_{M,I}^2}{p(1+3p)\,\sigma_{M,P}^2}\right). \tag{5.19}$$

The optimal bit allocation above is for a given prefilter $\mathbf{P}$. To further optimize $\mathbf{P}$, a unconstrained optimization program can be used to find the optimal prefilter $\mathbf{P}$ (more precisely the matrix $\mathbf{V}$ in the prefilter and the corresponding TDLT that minimize the expected distortion.

5.3 Implementation and Image Coding Results

Fig. 5.1(a) shows the partitions of image blocks into three descriptions. All blocks with the same index are grouped into one description as intra coded blocks, the rest of blocks are grouped as the inter coded blocks of that description. The 1-D Wiener filter is applied to the image, similar to the two descriptions case. The row prediction and column prediction are obtained first and the average of two predictions is used as the final prediction of the 2-D block.

The generation of four description coding is actually very similar to two-description coding. The block index assignment of 4-D MDC is shown in Fig. 5.1(b). The blocks with the same index are the intra coded blocks of corresponding description. The prediction process is slightly different from the two-description and three-description cases.

For example, in description 1, we first predict all blocks with index 2 from two vertical neighboring blocks with index 1. Next, all block with index 3 are predicted from two horizontal index-1 neighboring blocks. After that, the predicted blocks with indices 2 and 3 are

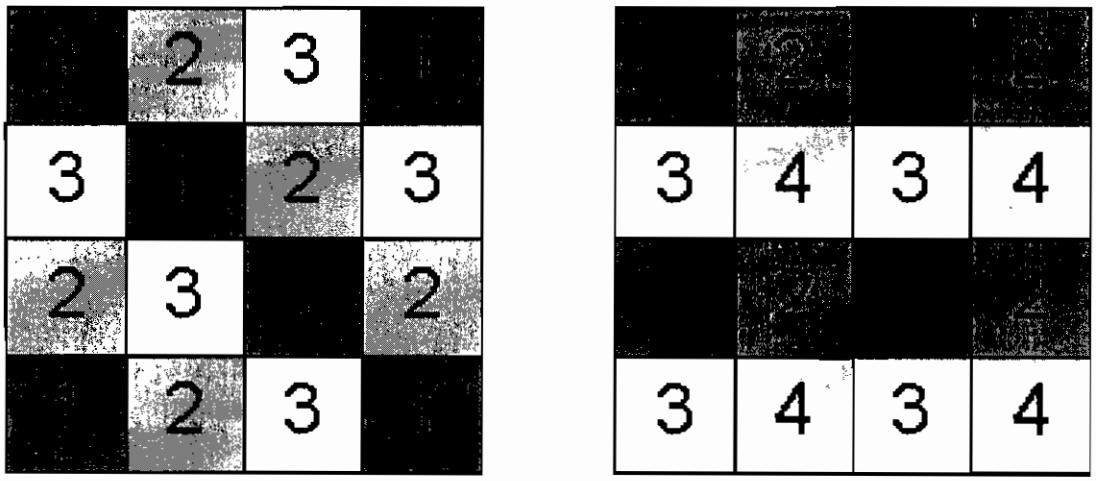


Figure 5.1: Block distribution for MDC with many descriptions. (a) Three descriptions (b) Four descriptions. ss

used to predict the blocks with index 4. In each case, the prediction only uses two neighboring blocks. Therefore the 1-D Wiener filter developed in Chapter 3 can be used directly. It is for this reason that we will apply the optimized lapped transform and Wiener filter for two-description coding to four-descriptions coding directly.

Fig. 5.2 summaries the decoded PSNRs of different images at $R = 1/N$ bpp for $N = 2$, 3, and 4, where N is the number of descriptions in the codec, and $D_{N,i}$ is the distortion of N-description codec when i descriptions are received, with i ranging from 1 to N. The redundancy is defined as $(N-1)R_1/R_0$, i.e., the ratio between the enhancement layer bit rate and the base layer bit rate. Some examples at 25% redundancy are shown in Fig. 5.3 and Fig. 5.4.

It can be seen that the results with all descriptions are similar in 2-description, 3-descriptions and 4-description cases. This suggests that the proposed 3-D and 4-D MDC

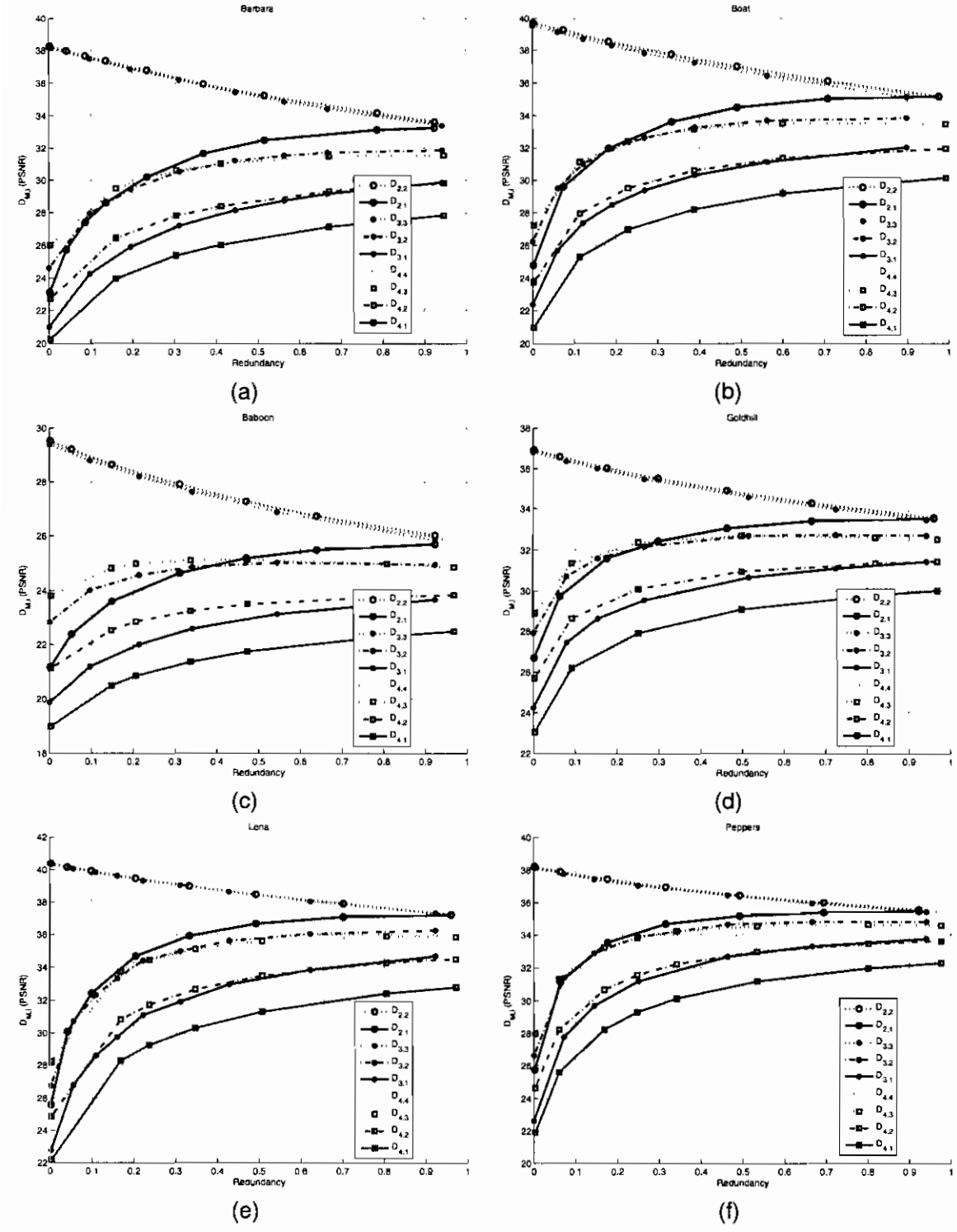


Figure 5.2: Image coding results with $R = 1/M$ bpp and $M = 2$, 3, and 4. (a) Barbara; (b) Boat; (c) Baboon; (d) Goldhill; (e) Lena; (f): Peppers.

do not introduce too much redundancy. As the number of descriptions increases, the codec provides more granularities . It also shows $D_{3,2}$ and $D_{4,3}$ are similar to $D_{2,1}$, whereas $D_{3,1}$ and $D_{4,2}$ are similar.

Although the coding efficiency of two-description coding is higher than three or four description case, its expected distortion can be worse at higher error probability. Fig. 5.5 shows the expected distortion of image Barbara at different error probability and bit allocation. The variance of the image is used as the distortion when all descriptions are lost. It can be seen that for $p = 0.01$, $M = 2$ gives the best result, whereas $M = 3$ and $M = 4$ gives the lowest distortion for $p = 0.05$ and $p = 0.15$, respectively. Therefore, having more descriptions is useful when the error probability is higher.

5.4 Conclusions

In this chapter, the prediction-compensated multiple description coding with many descriptions is studied. The performance of the proposed method in image coding is demonstrated, which shows that more descriptions can be useful when the error probability is higher.

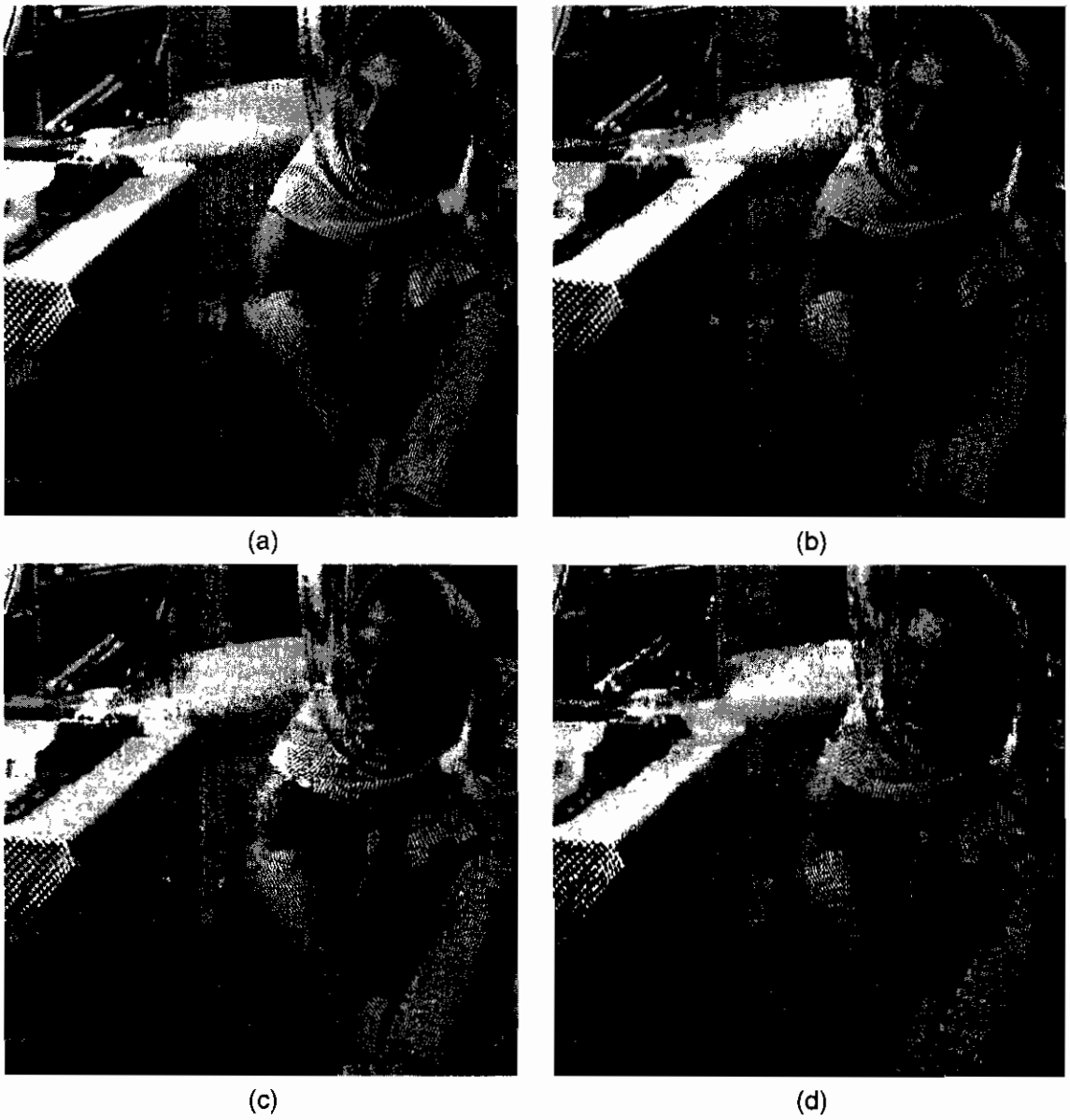
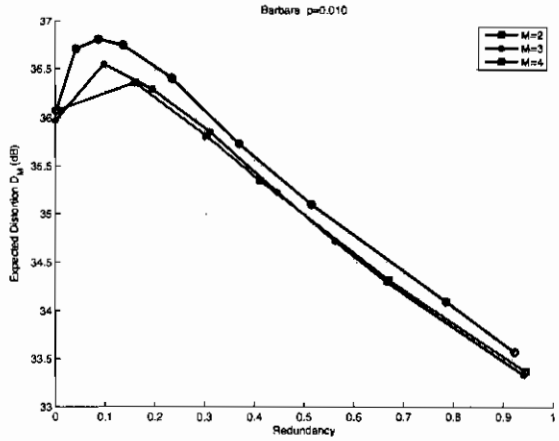
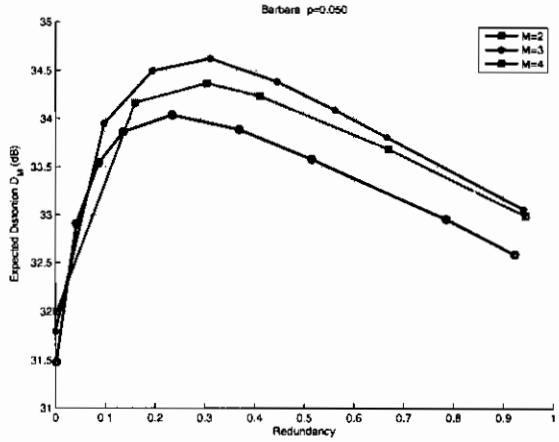


Figure 5.3: Reconstructed image Barbara at $R = 1/M$ bpp, 25% redundancy and different configurations. (a) $D_{3,3}$ (36.50 dB). (b) $D_{3,2}$ (30.01 dB). (c) $D_{3,1}$ (26.58 dB). (d) $D_{4,1}$ (24.93 dB).

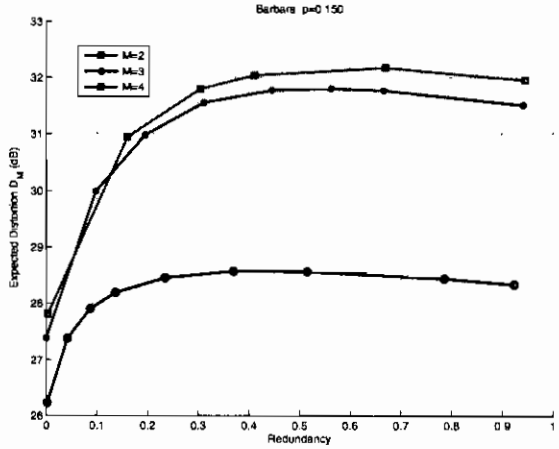
Figure 5.4: Reconstructed image Lena at $R = 1/M$ bpp, 25% redundancy and different configurations. (a) $D_{3,3}$ (39.25 dB). (b) $D_{3,2}$ (34.57 dB). (c) $D_{3,1}$ (31.36 dB). (d) $D_{4,1}$ (29.45 dB).



(a)



(b)



(c)

Figure 5.5: The expected distortion of the image Barbara with different number of descriptions and error probability. (a) $p = 0.01$. $N = 2$ has the best result. (b) $p = 0.05$. $M = 3$ has the best result. (c) $p = 0.15$. $N = 4$ has the best result.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we proposed a new multiple description coding scheme, by integrating the time domain lapped transform, block level splitting, linear prediction and prediction compensation.

The new scheme with two descriptions is studied first. The objective function for designing the proposed system is formulated and optimal result is obtained by numerical optimization. The asymptotic performance of the DPCM case of our method is analyzed and compared with Jayant's MDC method. The image coding experiment shows that our method can outperform the state-of-the-art method by up to 8 dB.

We next investigate the application of 2-D Wiener filter in the proposed scheme. The image coding results show that the performance can be improved when the redundancy is low. However, as the redundancy increases, the effect of the better prediction is less obvious.

Finally, we extend our work into more than two descriptions. In particular, the schemes with three descriptions and four description are developed. Image coding results show that

these cases can outperform the two-description case when the channel error probability is high.

The results presented in this thesis has been published in [40, 24]. A journal paper is also being reviewed [39].

6.2 Future Work

The proposed scheme can be further improved and applied to different areas.

6.2.1 Sequential Prediction

In the MDC with more than two descriptions in Chapter 5, each description uses one subset of data to predict all other subsets. This could be inefficient when the number of descriptions is large. Another method is to predict the other subsets sequentially and encode the corresponding prediction residuals. This should improve the rate-distortion performance of the method.

6.2.2 Entropy coding

The multiple description coding proposed here uses the same embedded entropy coding in [47] for both the base layer and the enhanced layer. However, due to the different properties of the base layer and enhanced layer data, different entropy codings should be used to maximize the performance. For example, the contexts for the arithmetic coding should be defined differently in the two layers.

6.2.3 Different transform coding

The proposed prediction compensation method can be applied to different transform codings, such as wavelet transform. The use of different transforms will pose new challenges

for the design of the prediction filter.

6.2.4 Applications in video coding

This thesis focuses on image coding. But the proposed method can be applied to video coding as well. Although it can be easily applied to the intra frames (I frame) in the video codec, the proposed method has not considered many issues in video coding, such as the error propagation. Special attentions should be made on these topics when the method is used in video coding.

Bibliography

[1] S.O. Aase and T.A.Ramstad. On the optimality of nonunitary filter banks in subband coders. *IEEE Trans. Image Proc.*, 4:1585–1591, Dec. 1995.

[2] J.-C. Batllo and V. A. Vaishampayan. Asymptotic performance of multiple description transform codes. *IEEE Trans. Inf. Theory*, 43(2):703–707, Mar. 1997.

[3] T. Berger. *Rate Distoriton Theory*. Prentice-Hall, 1971.

[4] J. Chen, C. Tian, and S. Diggavi. Multiple description coding for stationary and ergodic sources. *Proc. Data Compr. Conf.*, pages 73–82, Utah, Mar. 2007.

[5] D. Chung and Y. Wang. Multiple description image coding using signal decomposition and reconstruction based on lapped orthogonal transforms. *IEEE Trans. Circ. Syst. for Video Tech.*, 9(6):895–908, Sept. 1999.

[6] D. Chung and Y. Wang. Lapped orthogonal transform designed for error resilient image coding. *IEEE Trans. Circ. Syst. for Video Tech.*, 12(9):752–764, Sept. 2002.

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

[8] R. L. de Queiroz and K. R. Rao. Extended lapped transform in image coding. *IEEE Trans. Image Proc.*, 4(6):828–832, Jun. 1995.

[9] P. Dragotti, S. Servetto, and M. Vetterli. Optimal filter banks for multiple description coding: analysis and synthesis. *IEEE Trans. Inf. Theory*, 48(7):2036–2052, Nov. 2002.

[10] A. El Gamal and T. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, IT-28(6):851–857, Nov. 1982.

[11] V. Goyal. *Single and Multiple Description Transform Coding with Bases and Frames*. SIAM, 2002.

[12] V. Goyal and J. Kovacevic. Generalized multiple description coding with correlating transforms. *IEEE Trans. Inf. Theory*, 47(6):2199–2224, Sept. 2001.

[13] V. Goyal, J. Kovacevic, R. Arean, and M. Vetterli. Multiple description transform coding of images. *Proc. IEEE Conf. on Image Proc.*, 1:674–678, Oct. 1998.

[14] V. K. Goyal. Multiple description coding: compression meets the network. *IEEE Signal Process. Mag.*, 18(5):74–93, Sept. 2001.

[15] R.M. Gray and D.L.Neuhoff. Quantization. *IEEE Trans. Inf. Theory*, 44(6):2325–2383, Oct. 1998.

[16] P. Haskell and D. Messerschmitt. Reconstructing lost video data in a lapped orthogonal transform based coder. *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, 4:1985–1988, Apr. 1990.

[17] S. S. Hemami. Reconstruction-optimized lapped transforms for robust image transmission. *IEEE Trans. Circ. Syst. for Video Tech.*, 6(2):168–181, Apr. 1996.

[18] A. Ingle and V. A. Vaishampayan. DPCM system design for diversity systems with applications to packetized speech. *IEEE Trans. Speech and Audio Processing*, 3(1):48–58, Jan. 1995.

[19] N. S. Jayant. Subsampling of a DPCM speech channel to provide two self-contained half-rate channels. *Bell Syst. Tech. Jour.*, 60:501–509, Apr. 1981.

[20] N. S. Jayant and S. W. Christensen. Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure. *IEEE Trans. Commun.*, 29 (2):101–109, Feb. 1981.

[21] N. S. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video.* Prentice Hall, NJ, 1984.

[22] W. Jiang and A. Ortega. Multiple description coding via polyphase transform and selective quantization. *Proc. SPIE Conf. on Visual Comm. and Image Processing*, 3653:998–1008, Feb. 1999.

[23] J. Kovacevic, P. Dragotti, and V. Goyal. Filter bank frame expansions with erasures. *IEEE Trans. Inf. Theory*, 48(6):1439–1450, Jun. 2002.

[24] J. Liang, X. Li, G. Sun, and T. D. Tran. Two-dimensional Wiener filters for error resilient time domain lapped transform. *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, III:241–244, Toulouse, France, May 2006.

[25] J. Liang, C. Tu, L. Gan, T. D. Tran, and K.-K. Ma. Wiener filter-based error resilient time domain lapped transform. *IEEE Trans. Image Proc.*, 16(2):428–441, Feb. 2007.

[26] H. S. Malvar. *Signal Processing with Lapped Transforms.* Norwood, MA: Artech House, 1992.

[27] MDLT-PC source code: http://www.ensc.sfu.ca/~jiel/MDLTPC.html.

[28] A. C. Miguel, A. E. Mohr, and E. A. Riskin. SPIHT for generalized multiple description coding. *Proc. IEEE Conf. on Image Proc.*, 3:842–846, Oct. 1999.

[29] L. Ozarow. On a source coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59:1909–1921, Dec. 1980.

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

[31] R. Puri, S. Pradhan, and K. Ramchandran. n-channel symmetric multiple descriptions - Part II: an achievable rate-distortion region. *IEEE Trans. Inf. Theory*, 51(4):1377–1392, Apr. 2005.

[32] R. Puri, S. Pradhan, and K. Ramchandran. n-channel multiple descriptions: Theory and constructions. *Proc. Data Compr. Conf.*, pages 262–271, Snowbird, UT, Apr. 2002.

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

[34] RD-MDC source code: http://www.telematica.polito.it/sas-ipl.

[35] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partition in hierarchical trees. *IEEE Trans. Circ. Syst. for Video Tech.*, 6(3):243–250, Jun. 1996.

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

[37] C. E. Shannon and W. Weaver. *The mathematical theory of communication.* Univ. of Illinois Press, 1998.

[38] S. Srinivasan, P. Hsu, T. Holcomb, K. Mukerjee, S. Regunathan, B. Lin, J. Liang, M. Lee, and J. Ribas-Corbera. Windows media video 9: overview and applications. *Signal Processing: Image Communication*, 19(9):851–875, Oct. 2004.

[39] G. Sun, U. Samarawickrama, J. Liang, C. Tian, C. Tu, and T. D. Tran. Multiple description coding with prediction compensation. *IEEE Trans. Image Proc.*, Accepted, Dec. 2008.

[40] G. Sun, U. Samarawickrama, J. Liang, C. Tu, and T. D. Tran. Multiple description coding with prediction compensation. *Proc. IEEE ICIP*, vi:513–516, Sept. 2007.

[41] D. S. Taubman and M. W. Marcellin. *JPEG 2000: image compression fundamentals, standards, and practice.* Kluwer Academic, Boston, 2002.

[42] C. Tian and S. Hemami. Sequential design of multiple description scalar quantizers. *Proc. Data Compr. Conf.*, pages 32–41, Mar. 2004.

[43] C. Tian and S. S. Hemami. A new class of multiple description scalar quantizer and its application to image coding. *IEEE Signal Process. Lett.*, 12(4):329–332, Apr. 2005.

[44] C. Tian and S. S. Hemami. Universal multiple description scalar quantization: analysis and design. *IEEE Trans. Inf. Theory*, 50:2089–2102, Sept. 2004.

[45] T. Tillo, M. Grangetto, and G. Olmo. Multiple description image coding based on lagrangian rate allocation. *IEEE Trans. Image Proc.*, 16(3):673–683, Mar. 2007.

[46] T. D. Tran, J. Liang, and C. Tu. Lapped transform via time-domain pre- and post-processing. *IEEE Trans. Signal Proc.*, 51(6):1557–1571, Jun. 2003.

[47] C. Tu and T. D. Tran. Context based entropy coding of block transform coefficients for image compression. *IEEE Trans. Image Proc.*, 11(11):1271–1283, Nov. 2002.

[48] C. Tu, T. D. Tran, and J. Liang. Error resilient pre-/post-filtering for DCT-based block coding systems. *IEEE Trans. Image Proc.*, 15(1):30–39, Jan. 2006.

[49] P.P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall, 1993.

[50] V. Vaishampayan and J. Domaszewicz. Design of entropy-constrained multiple description scalar quantizers. *IEEE Trans. Inf. Theory*, 40(1):245–250, Jan. 1994.

[51] V. A. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Trans. Inf. Theory*, 39(3):821–834, May 1993.

[52] V. A. Vaishampayan and J.-C. Batllo. Asymptotic analysis of multiple description quantizers. *IEEE Trans. Inf. Theory*, 44(1):278–284, Jan. 1998.

[53] H. Wang and P. Viswanath. Vector gaussian multiple description for individual and central receivers. *IEEE Trans. Inf. Theory*, to appear, 2007.

[54] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman. Multiple description coding using pairwise correlating transforms. *IEEE Trans. Image Proc.*, 10(3):351–366, Mar. 2001.

[55] Y. Wang, A. R. Reibman, M. T. Orchard, and H. Jafarkhani. An improvement to multiple description transform coding. *IEEE Trans. Signal Proc.*, 50(11):2843–2854, Nov. 2002.

[56] X. Yang and K. Ramchandran. Optimal subband filter banks for multiple description coding. *IEEE Trans. Inf. Theory*, 46(7):2477–2490, Nov. 2000.

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

[58] R. Zamir. Gaussian codes and shannon bounds for multiple descriptions. *IEEE Trans. Inf. Theory*, 45(7):2629–2636, Nov. 1999.