ERROR-RESILIENT IMAGE AND VIDEO CODING OVER LOSSY NETWORKS

by

YILONG LIU

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
December 2005

ACKNOWLEDGEMENTS

During my doctoral studies at University of Texas, Arlington, I have had the opportunity to work with many excellent people whose support and encouragement have made this thesis possible. I would like to express my sincere gratitude to all of them.

First, I would like to thank my supervisor, Dr. Soontorn Oraintara, for his guidance and advice during my Ph.D. studies. His continuing support and encouragement have been instrumental on the completion of this research. It has been both a pleasure and an honor for me to work with him, an outstanding advisor – helpful, kind, insightful, and patient.

I would like to express my special thanks to Dr. K. R. Rao, Dr. Qilian Liang, Dr. Saibun Tjuatja, Dr. Zhou Wang, Dr. Hua-mei Chen, Dr. Venkat Devarajan, Dr. A. K. Fung and Dr. Hanli Liu, who served on my qualifying, proposal, and dissertation defense committees for taking the time to read my proposal or thesis and their valuable comments and suggestions.

I would like to thank my friends at the MSP lab, Yoshikazu Yokotani, Rohit Kumar, and many more, for the time we spent discussing our research works. And I would like to express my special thanks to Truong T. Nguyen, whose valuable advise and suggestions are helpful to our joint work.

I also want to thank my master degree supervisor, Dr. Xiqi Gao, who has led me to the interesting world of research.

Finally, I would like to express my deep gratitude to my family, who have encouraged and supported me throughout my doctorial studies.

November 10, 2005

ABSTRACT

ERROR-RESILIENT IMAGE AND VIDEO CODING
OVER LOSSY NETWORKS

Publication No. ______

YILONG LIU, Ph.D.

The University of Texas at Arlington, 2005

Supervising Professor: Soontorn Oraintara

The rapid growth of the Internet and increasing bandwidth in the past decade has heightened the interest in robust transmission of image and video signals over error-prone links. Recent literatures highlight the multiple description (MD) coding as a promising approach to solve this problem. MD coding is a source coding technique that provides a graceful quality degradation in the presence of packet losses during transmission. Topics of interest to construct MD coders with high coding efficiency and robustness are studied in this thesis.

To minimize the overall distortion of reconstructed image, we propose an edge-adaptive estimation algorithm that is able to capture the arbitrarily-oriented edges around the corrupted coefficients. It reconstructs the lost samples more accurately compared to the previously proposed methods in the literatures. A feature-oriented MD coding scheme is proposed by adopting this estimation algorithm to smartly identify the coefficients that are sensitive to packet loss. The redundancy is then inserted to each channel by jointly optimizing the subsets of sensitive coefficients and the related quantizers. The corrupted samples are reconstructed with the proposed estimation algorithm at the receiver.

By introducing the extra prediction loops to the correlating transform based MD video coder, the mismatching error, i.e. the drift, incurred by the one-channel reconstruction is decreased. However, the problem of effective redundancy allocation is introduced by the drift coding. An efficient algorithm is developed to solve this problem. The proposed approach optimally adjusts the source coding rates for the correlating transform coefficients and drifts in a coordinated fashion.

The proposed algorithms are implemented based on the image/video MD coding systems using DCT or wavelet transforms. The advantages of the introduced approaches to achieve high coding efficiency and error-resilience for transmission are demonstrated.

TABLE OF CONTENTS

Chapter

Appendix

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

There has been an increasing demand of multimedia communications in the past decade. This was inspired by several factors, including the proliferation of fast and affordable end-user communication devices, expansion of wired and wireless network infrastructure, and development of high performance compression algorithms and standards [75]. Typical multimedia applications combine sound, image, video and data (e.g. text), and large transmission bandwidth is required. To decrease the required transmission bandwidth, they have to be compressed prior to transmission. However, compressed multimedia bit streams are very sensitive to communication channel impairments such as bit errors and packet losses. Hence, additional mechanisms and strategies are necessary to reduce the effect of channel impairments and enable successful communication.

Transmission of image or video signals is the most challenging problem in multimedia communication because of the real-time constraint and large bandwidth. Figure 1.1 illustrates a typical system for image or video transmission over packet-based networks. Input image or video signal is first encoded by the source coder. The bit stream generated by the source coder is then manipulated into transmission data units called packets, and channel coding, such as forward error correction (FEC) codes, may be applied to each packet. Usually, source coder, packetizer and channel encoder are jointly combined into a single processing block. Packets are then sent through a wired network or wireless channel, to the receiver side. Transmission of packets over the network is controlled by transport layer network protocols, for instance transmission control protocol (TCP), user datagram protocol (UDP), realtime transport protocol (RTP) and real-time control protocol (RTCP) in case of Internet. These protocols provide end-to-end control of packet transmission between sender and receiver, transparent to the underlying physical network transmission technologies used [82].

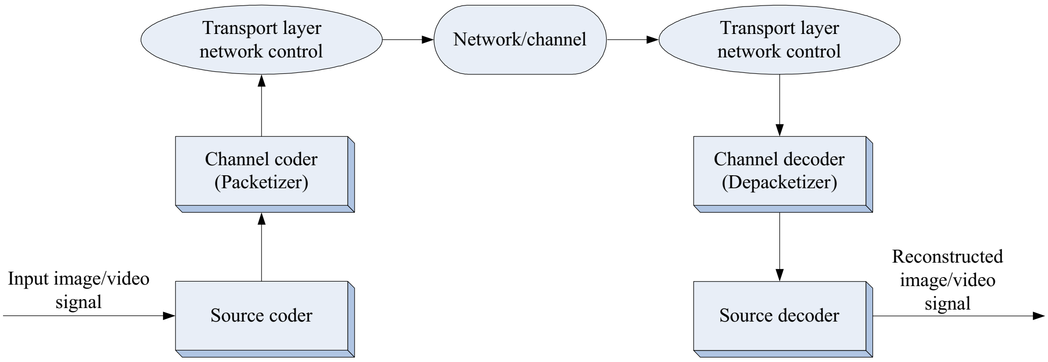


Figure 1.1. Diagram of packet-based image/video communication system.

Unfortunately, some of the packets may be dropped at intermediate nodes as they flow through the network towards the receiver, because of the bit errors, buffer overflow, excessive delay, transmission link/node failure, etc. Although channel decoder may be able to recover some missing packets or correct erroneously received bits, there are still errors that can not be corrected in this way left for the source decoder to handle due to the constraints of the ability of channel correction coding. For video signal, the error of the current frame may propagate to the following frames due to the prediction based coding strategy. The error may accumulate and result in an unacceptable quality. In this situation, a request will be sent to the sender to retransmit the corrupted piece of information.

1.1 Error Control for Robust Image/Video Transmission

Error control can be implemented at different levels in an image/video communication system as described in Fig.1.1. Prior works in this area can be broadly divided into three classes based on where in the system is the error control actually done [99]: transport-level control, encoder-based control and decoder-based control. Practical methods of error control often combine two or more of the approaches from different classes [100].

1.1.1 Error Control at Transport Level

Transport-level control executes error control at the transport level network protocols and exploits their functionalities. It is not required to modify the encoder and decoder, which makes these methods of this class relatively simple and easy to implement. However, capabilities of these methods are constrained by the current network infrastructure.

One of the well-known methods of traditional error control algorithms at this level is automatic repeat request (ARQ), where the receiver asks the sender to retransmit the data that were not received due to loss or excessive delay. This type of error control is used in TCP to enable lossless data delivery with unbounded delay. However, delay is a major issue in multimedia communications for both real-time and streaming applications. This makes ARQ not suitable for image/video transmission. Several solutions to this problem have been proposed [99]. For instance, one could limit the use of ARQ to the most important data, such as header information which typically accounts for a relatively small fraction of the bit stream, and not use it for other types of data. Another approach is the delay-constrained retransmission in which for each corrupted packet, the decoder estimates the arrival time of its retransmitted duplicate and requests retransmission of only those packets that are possible to arrive on time to be processed by real-time decoding.

Variant priorities can be also applied to different parts of the data to be transmitted, if the network can support transmission of packets with different priorities. When congestion occurs, low-priority packets are first discarded by the server, while high-priority packets are discarded only if the congestion persists. This can be effectively used in image/video communication by sending more important data in high-priority packets, and less important data in low-priority packets [70].

Forward error correction (FEC) can be also employed at the transport level. A (n, k) block code, e.g. Reed-Solomon code, can be used for each source data packet [64]. A group of $(n - k)$ FEC packets are created to protect the source data packets. Every packet contains one symbol from each of the (n, k) codewords, so the loss of one packet causes erasure of one symbol from each of the codewords, which can be corrected if $(n-k)$ is sufficiently large. For instance, a (n, k) Reed-Solomon code can correct up to $(n - k)$ erasures since it has the minimal distance of $(n - k + 1)$ [48]. It is suggested that the protection should be manipulated according to the importance of the data. However, this is not known at the transport level in general. Hence, approaches at the encoder level should be considered for the use of FEC.

1.1.2 Encoder-Based Error Control

At the transmitter, it is convenient for encoder to anticipate what effect will be incurred by the error in a particular bit, or the loss of a group of bits, at the decoder. Hence an appropriate level of protection can be assigned to different groups of source data accordingly. Unlike transport-level control, implementations of encoder-based error control methods require modifications of the encoder and sometimes decoder, but they can provide more flexibility compared to the transport-level methods.

Error propagation, caused by using the high-performance variable length code (VLC), is a major problem for the encoder-based error control approaches. This is because, in VLC, the lengths of individual codewords are different, and the decoder usually does not know the length of the variable length codeword until it is fully decoded. A missing or erroneous bit may cause the decoder to substitute one codeword for another, causing it to start decoding the following codeword from the wrong position in the bit stream. Hence, one bit error may cause incorrect decoding of a large block of bits.

A simple way of depressing error propagation is by insertion of re-synchronization markers [75]. These markers are blocks of bits which differ sufficiently and can be identified from the valid VLC codewords. Upon detecting a re-synchronization marker, the decoder can resume correct decoding of valid VLC codewords. For example, MPEG standards provide unique start markers for several types of coding units, such as GOP, frame, slice [1, 2]. The effects of error propagation can be further reduced by the use of reversible variable length code (RVLC) [105, 104]. These codes have the property that their codewords can be decoded in both forward and reverse directions [32]. When a bit error is detected, decoder can search for the next re-synchronization marker, and from there start decoding in reverse up to the location of the error.

For video coding systems, error propagation may occur in both spatial and temporal domains, due to predictive coding. If one block is erroneously decoded, the blocks predicted from it will also be erroneously decoded [75]. The effects on errors on compressed video streams depend on the type of the corrupted video parameter and the sensitivity of this parameter to errors. Hence improvements in error robustness of video streams could be achieved by separating the video data to two parts [86]. The motion data of each video packet, which is vulnerable to errors, is placed in the first partition, while the less sensitive *texture* data is placed in the second partition. The two partitions are separated by a re-synchronization code. This approach enables the decoder to restore the error-free motion data of a video packet when errors corrupt only the bits of the less sensitive texture data of the second partition. On the other hand, errors occurring in the second less sensitive partition can usually be successfully concealed, resulting in little visible distortion. As texture data makes up the majority of each packet, data partitioning allows errors to occur in a large part of the packet with relatively benign effects on video quality. Further improvement at the encoder side can be achieved by duplicating motion data, refreshing intra-coded macroblock/frame periodically or adaptively [75].

Scalable coding can also be used as a robustness tool when coupled with unequal error protection (UEP) [42, 66]. To provide graceful degradation of image/video quality with deteriorating channel conditions, more protection is emphasized on the layers with important information using some form of channel coding. Experiments of scalable coding show that individual layers differ in their contribution to the overall reconstructed signal quality. However, networks usually treat all data packets equally. Hence the UEP can be thought of as an adaptation mechanism that corrects this mismatch and maps layered bit streams to variant channels which treats the individual layer appropriately.

1.1.3 Error Concealment at the Decoder Side

Although the error control at transport level can correct some corrupted data, some part of the data may still not be available at the decoder in time for its scheduled playout, due to the constraint of real-time applications and the limit of FEC correction. Hence, error control mechanisms are also necessary at the decoder to handle the missing data. Fundamentally all decoder-based error control methods take advantage of the structure of image/video data, and capture the existing spatial and temporal correlation. Some form of interpolation is usually used to estimate the missing samples from the available neighbors. This process is also known as error concealment, which usually cannot guarantee exact reconstruction of missing data.

It is well known that images of natural scenes have predominantly low-frequency components, i.e., the values of spatial and temporally adjacent pixels vary smoothly, except in regions with sharp edges. In addition, the human eyes can tolerate more distortion to the high-frequency components than to the low-frequency components. These facts can be used to conceal the artifacts caused by transmission errors.

For video signal, one simple way to exploit the temporal correlation in video signals is by replacing a damaged macroblock with the spatially corresponding macroblock in the previous frame. This method is also known as motion-compensated temporal prediction. However, it can produce adverse visual artifacts in the presence of fast motion. Significant improvement can be obtained by replacing the damaged macroblock with the motion compensated block (i.e., the block specified by the motion vector of the damaged block). This method is very effective when combined with layered coding that includes all the motion information in the base layer [24]. Because of its simplicity, this method has been widely used. A problem with this approach is that it requires knowledge of the motion information, which may not be available in all circumstances. This can be solved by combining the error-concealment with a layered coder [39].

Another state-of-the-art error concealment approach, named as maximally smooth recovery, makes use of the smoothness property of most image and video signals through a constrained energy minimization approach [101]. The minimization is accomplished block by block. Specifically, to estimate the missing discrete cosine transform (DCT) coefficients in a block [65], the method minimizes a measure of spatial and temporal variation between adjacent pixels in this block and its spatially and temporally neighboring blocks, so that the resulting estimated signal is as smooth as possible. Wang et al. first used this approach to recover damaged blocks in still images coded adopting block-transform-based coders by making use of the spatial smoothness only [101]. Zhu et al. later extended this method to video coders using motion compensation and transform coding by adding the temporal smoothness measure [117].

An alternative of maximally smooth recovery is to use the projection onto convex sets (POCS) method. Sun and Kwok proposed to use this method to restore a damaged image block in a block transform coder [85]. The convex sets are derived by requiring the recovered block to have a limited bandwidth either isotropically (for a block in a smooth region) or along a particular direction (for a block containing a straight edge). Zeng and Liu extended similar idea with a geometry-structure-based error concealment to improve the interpolation directions [113].

Error-concealment techniques have also been developed for other objectives, including the reconstruction of lost coefficients in transform domain [33], recovery by making use of blockwise similarity within images [103], estimation of the corrupted coding mode and motion vectors [84], and fuzzy-logic based recovery of high-frequency components [44].

1.2 Motivation

In recent years, there has been a renewed interest in Multiple Description (MD) coding [26], primary reason for this is that MD coding is a form of joint source-channel coding that is well suited to transmission over channels such as lossy packet networks.

MD coding itself is not new. The first references on the topic include [63, 107] and date back to early 1980s. In MD coding, source encoder produces several descriptions of the source signal which are encoded independently and transmitted over several channels to the receiver. When decoded independently each description can provide a reasonable approximation to the original signal. However, when several descriptions are decoded jointly, the quality of the approximation is increased. This is due to the correlation that exists among descriptions which enables the decoder to estimate the missing descriptions from the received ones.

Designing an efficient MD coding system is an interesting topic in latest research. A number of approaches have been proposed for creating multiple descriptions [66, 13, 90, 37, 56, 28, 87]. As one of the popular joint source-channel coding methods, MD coding can be well designed and optimized through the interaction of encoding, decoding and UEP channel coding. This motivates our research interests to improve the current state-of-art MD coding techniques and design smarter coding schemes to process the variant data sources.

1.3 Contributions of the Thesis

This thesis makes the following contributions.

- First, we propose a framework to recover corrupted coefficients in wavelet-based MD coding systems. An edge-adaptive estimation algorithm is presented to exploit the local texture statistical characteristics for estimating the lost coefficients. Extensive experiment results show that impressive improvements on both objective and subjective measures can be achieved.
- Second, we introduce a novel MD coder, named as feature-oriented MDC. With the aid of edge-adaptive estimation algorithm [54], the proposed codec smartly analyzes the local statistics of the wavelet coefficients. The bit budget is then adaptively allocated to those coefficients that are sensitive to packet loss. By using the joint optimization of tree-pruning and quantizer selection in the rate-distortion sense, our MD scheme achieves high robustness and coding efficiency compared to the state-of-the-art MD coders.
- Third, we investigate the application of MD correlating transform based video coding. A joint optimization algorithm is proposed for the drift-free MD video coding scheme. The optimization is in the rate-distortion sense using Lagrangian relaxation method for optimum allocation of redundancy amongst the correlating

transform and the drift coding. The simulation results show that our algorithm achieves substantial gains over the conventional coder.

- Finally, two embedded image coding systems based on multiresolution directional filter banks (DFB) [60] are proposed. The first codec employs a multiscale filter bank (FB), which is a combination of the recently introduced non-uniform quincunx DFB (nuqDFB) at higher scales and the traditional wavelet filter bank at lower scales, to provide a sparse image representation. The coding algorithm then efficiently clusters the the significant coefficients using progressive morphological dilation [95]. The remaining scattered significant coefficients are coded with the aid of the Tarp filter [81]. The embedded coding approach is extended for the hybrid overcomplete pyramidal DFB (PDFB) in the second codec. Similar progressive morphological dilation based coding algorithms are implemented. Furthermore, the correlation exists among the neighboring directional subbands is adopted to reduce the high redundancy of the overcomplete transform. Both coders achieve better results than the current state-of-the-art wavelet based coders, such as JPEG2000, thanks to the efficient edge approximation provided by the directional decomposition.

1.4 Organization of The Thesis

The remainder of this thesis is organized as follows. Chapter 2 begins with the theoretical bounds of MD coding and an overview of conventional MDC approaches. Then several state-of-the-art MD coding schemes are discussed in details. In Chapter 3, the drawback of the data partitioning based MD coding algorithm is discussed. Two approaches are proposed to solve the problem. The edge-adaptive error concealment presented in this chapter is employed in Chapter 4, where we introduce a novel feature-oriented wavelet-based MD coding scheme for error-prone networks. The extensive experiments are carried out to verify the high resilience and efficiency of our codec. In Chapter 5, the problem of drift-free MD video coding is discussed, a joint optimization algorithm is proposed to achieve the optimal redundancy allocation between the central and side prediction loops. A novel multiresolution DFB based image coding algorithm is described in Appendix A, which uses the morphological dilation and Tarp filter to progressively encode the sparse significant coefficients of the DFB. Similar idea is extended to the coding algorithm using the pyramidal DFB. Context models are designed to exploit the inter-band and intra-band correlations of the overcomplete transform. The promising improvement of coding result is presented. Chapter 6 summarizes this thesis.

CHAPTER 2

REVIEW OF MULTIPLE DESCRIPTION CODING

2.1 Introduction

Recently, multiple description coding (MDC) has emerged as a promising approach to enhance the resilient coding and transmission for image and video signals. Typically, a multiple description (MD) coder is a source coding technique that provides a graceful quality degradation in the presence of packet losses during transmission. In contrast to a conventional media coder, which generates a single stream, a MD coder encodes a media source into two or more bitstreams as described in Fig. 2.1. These bitstreams, also called descriptions, are generated in such a way that each description can be independently decoded to produce a signal of basic quality. While receiving all the descriptions, the decoder can reconstruct a signal of improved quality. In order to achieve this goal, each description should contain a sufficient amount of information about the original source. That is, a certain amount of correlation has to be introduced among the descriptions. This correlation controls the media quality reconstructed from each description.

A primary reason for the increasing popularity of MDC is that it can provide adequate quality without requiring retransmission of any lost packets (unless the loss rate is very high). This has two important implications. First, it makes MDC particularly appealing for real-time interactive applications such as video phone and conference, for which retransmission is often not acceptable because it incurs overlong delay. Second, it simplifies the network design: no feedback or retransmission is necessary and all the packets can be treated equally. This is in contrast to layered coding (LC), which generates a base layer and one or more enhancement layers. One major drawback for LC is that

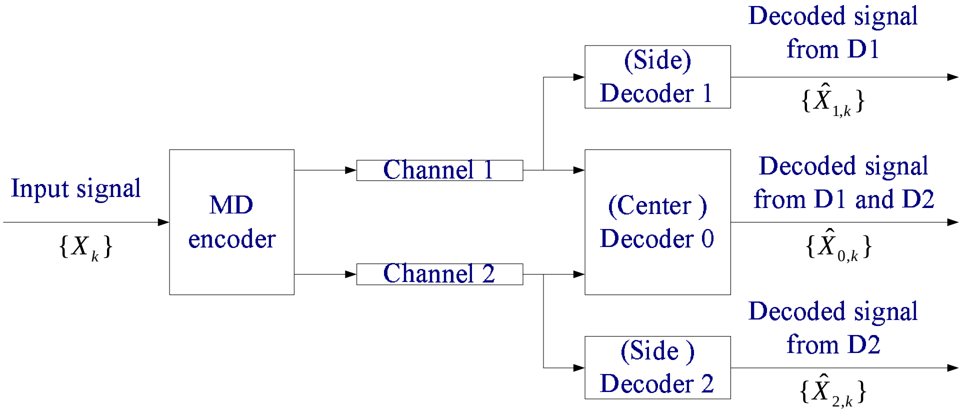


Figure 2.1. General structure for multiple description coding system.

the base layer must be delivered almost error free to guarantee a basic level of quality. This requires differential treatment of base-layer and enhancement-layer packets by the network and retransmission of lost base-layer packets, which may not always be feasible. In addition to error resilience, MDC also makes it easier for traffic dispersion and load balancing in the network, which can effectively relieve congestion and increase overall network utilization. The fact that substreams from MDC can be treated equally and independently makes the task of allocating packets onto transport paths much easier than for a conventional single description (SD) coder or a layered coder.

However, the benefits of MDC come at a price that the MD coder uses more bits to meet the same quality criterion as a conventional SD coder in the absence of any transmission errors. This excess rate or redundancy is inserted intentionally to make the bitstream more resilient to transmission errors. The primary objective in designing an MD coder is to minimize the redundancy (or the total rate) while meeting an end-to-end distortion requirement that takes into account transmission loss.

In the rest part of this chapter, Section 2.2 discusses the theoretical bounds of MD coding. In Section 2.3, the conventional MDC algorithms are reviewed. Several state-of-the-art MDC schemes are described in Sections 2.4 - 2.7 in detail. Section 2.8 summarizes the chapter.

2.2 Theoretical Bounds

The early history of MDC, [63, 14, 4, 115, 111], was focusing on obtaining information theoretical results for the achievable bounds. As depicted in Fig. 2.1, an encoder is given a sequence of source symbols $\{X_k\}_{k=1}^N$ to communicate to three receivers over two noiseless (or error-corrected) channels. The transmission rate over channel i is denoted by $R_i, i = 1, 2$, i.e. at most 2^{NR_i} symbols are used over channel i. Let $\{\hat{X}_{i,k}\}_{k=1}^N$ be the reconstruction sequence produced by decoder i, thus we have three distortions

$$D_i = \frac{1}{N}\sum_{k=1}^{N} E[\delta_i(X_k, \hat{X}_{i,k})], \quad i = 0, 1, 2, \tag{2.1}$$

where the $\delta_i(\cdot,\cdot)$'s are real-valued distortion measures, and D_i, are distortions associated with central and side decoders, respectively.

The theoretical problem is to determine the set of achievable values in the usual Shannon sense for the quintuple $(R_1, R_2, D_0, D_1, D_2)$. Specifically, $(r_1, r_2, d_0, d_1, d_2)$ is achievable for sufficiently large N such that

$$\begin{aligned} R_i &\leq r_i, \quad i = 1, 2; \\ D_i &\leq d_i, \quad i = 0, 1, 2. \end{aligned}$$

Since decoder i receives R_i bits and cannot have distortion less than $D(R_i)$, where $D(\cdot)$ is the rate-distortion function of the source. Hence the following bounds are obtained on the achievable region:

$$R_1 \geq R(D_1), \tag{2.2}$$

$$R_2 \geq R(D_2), \tag{2.3}$$

$$R_1 + R_2 \geq R(D_0). \tag{2.4}$$

where, $R(\cdot) = D^{-1}(\cdot)$.

Achieving equality simultaneously for the three bounds is impossible since optimal individual descriptions at rates R_1 and R_2 are similar to each other and hence redundant when combined. The distortion achieved by combined multiple descriptions cannot be optimal for the sum of rates of the descriptions. Although there have been some bounds obtained for various sources and distortion measures, the complete achievable region is known only for i.i.d. Gaussian source. El Gamal and Cover determined an achievable rate region for general memoryless sources [23]. Ozarow showed that this region is tight for the case of memoryless Gaussian source with variance σ^2. The achievable set of rates and mean-squared error distortions is the union of points satisfying [63]

$$D_1 \geq \sigma^2 \cdot 2^{-2R_1}, \tag{2.5}$$

$$D_2 \geq \sigma^2 \cdot 2^{-2R_2}, \tag{2.6}$$

$$D_0 \geq \frac{\sigma^2 \cdot 2^{-2(R_1+R_2)}}{1-(\sqrt{\Pi}-\sqrt{(\Delta)})^2}, \tag{2.7}$$

where

$$\Pi = (1 - D_1/\sigma^2)(1 - D_2/\sigma^2), \tag{2.8}$$

and

$$\Delta = (D_1 D_2/\sigma^4) - 2^{-2(R_1+R_2)}. \tag{2.9}$$

Ahlswede [4] studied the case of no excess rate (when there is equality in (2.4), and Zhang and Berger [115] considered the no excess marginal rate case (when there is equality in (2.2) and (2.3). They also showed by counterexample that in the excess rate case the achievable region of El Gamal and Cover is not tight [114]. More recently, Linder et al. [49] found a rate region for memoryless sources and locally quadratic distortion measure which is tight in the limit of small distortions (high bit rate). Finally, Zamir [111, 112] extended the Shannon bounds [8] to the MD case and showed that for a Gaussian source the outer bounds are asymptotically tight. The inverse of (2.5) - (2.7) are the following [49, 63]:

$$R_1 \geq \frac{1}{2} log(\frac{\sigma^2}{D_1}), \tag{2.10}$$

$$R_2 \geq \frac{1}{2} log(\frac{\sigma^2}{D_2}), \tag{2.11}$$

$$R_1 + R_2 \geq \frac{1}{2} log(\frac{\sigma^2}{D_1}) + \frac{1}{2} log(\frac{\sigma^2}{D_2}) + \delta, \tag{2.12}$$

where δ is defined by

$$\delta = \begin{cases} \frac{1}{2} log(\frac{1}{1-\rho^2}), & D_0 \leq D_0^{max} \\ 0, & D_0 \geq D_0^{max} \end{cases} \tag{2.13}$$

where

$$D_0^{max} = \frac{D_1 D_2}{D_1 + D_2 - (D_1 D_2 / \sigma^2)} \tag{2.14}$$

and

$$\begin{aligned} \rho &= -\frac{\sqrt{\Pi \epsilon_0^2 + \gamma} - \sqrt{\Pi \epsilon_0^2}}{(1-\epsilon_0)\sqrt{\epsilon_1 \epsilon_2}}, & (2.15) \\ \gamma &= (1-\epsilon_0)[(\epsilon_1 - \epsilon_0)(\epsilon_2 - \epsilon_0) + \epsilon_0 \epsilon_1 \epsilon_2 - \epsilon_0^2], & (2.16) \\ \Pi &= (1-\epsilon_1)(1-\epsilon_2), & (2.17) \\ \epsilon_i &= D_i / \sigma^2, \quad i = 0, 1, 2. & (2.18) \end{aligned}$$

Notice that δ depends on the three distortions (D_0, D_1, D_2) and on the variance σ^2. However, by rearranging (2.12), one can see the relationship between δ and the rates R_1, R_2 and interpret δ as the excess rate that is used to reduce the central distortion given the two side distortions or

$$\delta = R_1 - \frac{1}{2} log(\frac{\sigma^2}{D_1}) + R_2 - \frac{1}{2} log(\frac{\sigma^2}{D_2}). \tag{2.19}$$

If $\delta = 0$ then

$$R_1 = \frac{1}{2} log(\frac{\sigma^2}{D_1}) \quad \text{and} \quad R_2 = \frac{1}{2} log(\frac{\sigma^2}{D_2}). \tag{2.20}$$

That means that all the rates are used to minimize the side distortions and in this case D_0 is equal to its maximum value D_0^{max}. In particular, D_0 decreases from D_0^{max} to zero as δ increases from zero to infinity.

2.3 Overview of Conventional Multiple Description Coding

Based on the theoretical bounds, a number of practical MDC methods for image and video coding have been proposed during the past decade [26].

Vaishampayan proposed a simple and practical MDC scheme, known as multiple description scalar quantizers (MDSQ), to generate two sub-streams by producing two indices for each quantization level [90]. The index assignment is designed to be equivalent to a fine quantizer when both indices are received. However, a coarse quantizer can be used when only one index is received. More complicated quantizers, such as trellis-coded and vector quantizers, and entropy codes were later designed to improve the coding efficiency [91, 35, 93]. This MDSQ method was modified and applied to motion-compensated predictive video coders by using two independent prediction loops [92]. Lee et al. also proposed a two-stage multiple description video coder by adopting this MDSQ technique [45].

An alternative way of designing the MDC algorithm is to use correlation-inducing transforms [98, 27, 29]. In ideal source coding, the coefficients used to represent the signal should be least correlated to maximize the coding efficiency. However, under this paradigm it is very difficult to estimate the value of a lost coefficient from those that remain. To achieve robustness against coefficient losses, the transform coefficients are divided into multiple groups where the correlation within each group is minimized while inter-group correlation is tolerated. Wang proposed applying a pair-wise correlating transform (PCT) to each pair of uncorrelated variables obtained from the Karhunen-Loeve transform (KLT) [98]. Goyal et al. further generalized Wang's work to any number of variables, and coined the term generalized MDC (GMDC) [27]. GMDC was later applied to image coding with correlating transforms [29]. As compared to the MDSQ method, the PCT/GMDC has been shown to perform well at low rates [27]. Reibman et al. adopted this PCT method and proposed a multiple description video coder that uses motion-compensated predictions [72].

In addition to these two algorithms, many other MDC methods have also been proposed, including FEC-based MDC [67, 66], MDC using frame expansion [30], matching pursuits MDC [87, 58] and lapped orthogonal transform (LOT) based MDC [13].

Based on the way of generating descriptions, we categorize the MDC implementations into four classes:

A) MDC using redundant quantizers,

B) MDC by introducing controllable correlations,

C) MDC with partitioning, and

D) MDC through unequal forward error correction.

The typical MDC approaches for each class are discussed in the following sections.

2.4 Multiple Description Coding using Redundant Quantizers

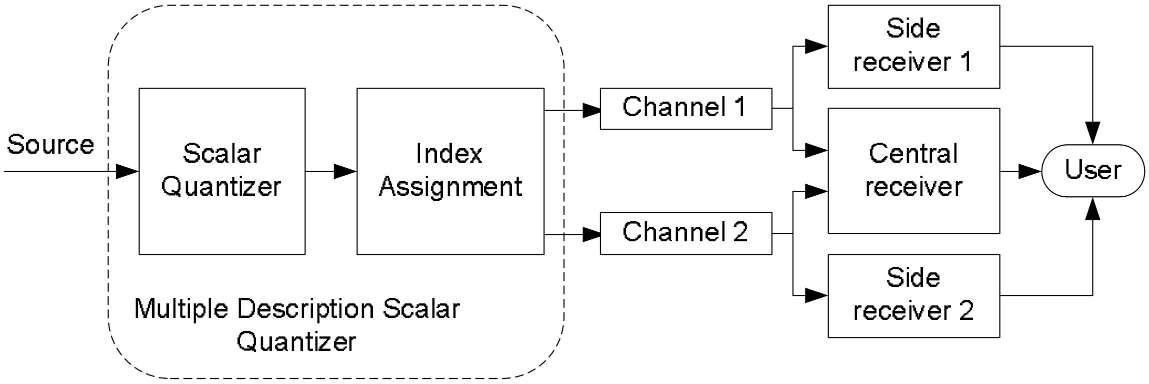


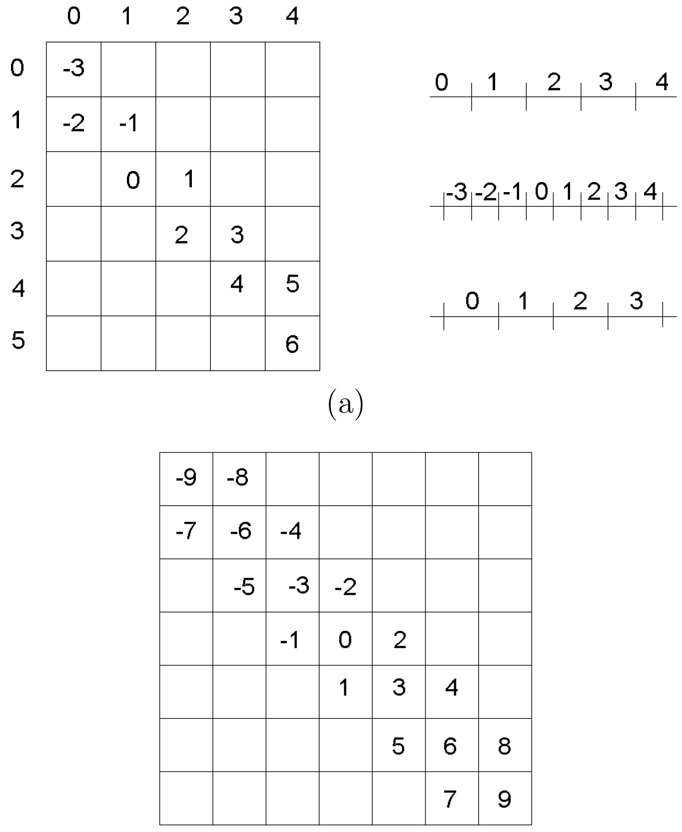
Figure 2.2. Basic architecture of multiple description scalar quantizer coding system.

In the first class pioneered by Vaishampayan, MDC is accomplished by using a MD scalar quantizer, which is simply designed with some remarkable asymptotic properties [90]. Fig. 2.2 illustrates the role of a MD quantizer in a two-channel diversity system.

In this approach, a MD qunatizer consists of two main components: a *scalar quantizer* (that maps continuous-valued random variables to points in a countable set), and an *index assignment* (that splits the information about each sample into two complementary and possibly redundant descriptions of the same sample). An index assignment is an injection I: $N \rightarrow N \times N$ (N is the set of natural numbers). When the scalar quantizer maps the source to a finite number of points, e.g. n, the map I can be thought of as a matrix of size $n \times n$, in which only n locations are occupied. Apparently, there exist $n!C_n^{n^2} = O(n^{3n})$ such distinct mappings. Two sample index assignments are shown in Fig. 2.3.

The problem of designing good index assignments was thoroughly studied in [90], using Ozarow's characterization of the MD rate-distortion region for the Guassian source as discussed in (2.5)-(2.7). Under the assumption of equal and high rates for both descriptions, and for a squared error distortion measure, a construction of a large class of index assignments is presented in [90], for which the exponential rate of decay of the mean squared error is exactly that predicted by Ozarow's result.

To further explain how the MDSQ works, let D_1 and D_2 be two descriptions that have been generated by two coarse quantizers. As shown in Fig. 2.4, S_{2k} and S_{2k+1}, denote the quantization bounds for D_1 and D_2 respectively. Typically, image coders based on uniform quantizers perform inverse quantization by mapping bins to the midpoint of their cell. With only one of the descriptions, we cannot decide the fine quantization interval, but we can restrict to a coarser quantization bound. For instance, if only D_1 is received, but D_2 is absent, the reconstruction level minimizing a mean square error distortion



	0	1	2	3	4
0	-3				
1	-2	-1			
2		0	1		
3			2	3	
4				4	5
5					6

(a)

-9	-8					
-7	-6	-4				
	-5	-3	-2			
		-1	0	2		
			1	3	4	
				5	6	8
					7	9

(b)

Figure 2.3. Two index assignments: (a) staggered quantization cells; (b) higher spread cells. The idea is that bins of the scalar quantizer are placed in a matrix, and then quantizer indices corresponding to row and column entries are sent over each channel. If both descriptions are available, the original quantization bin can be recovered; if not, the original quantization bin is known to be one of those in the received row/column.

metric corresponds to the midpoint of S_2 and S_4, i.e. S_3. When both descriptions are available, the reconstruction levels are S_3 and S_2 for D_1 and D_2, respectively. This further implies that the reconstruction level can be refined as L, the midpoint of S_2 and S_3.

Obviously, using the midpoint as the reconstruction level leads to remarkably poor performance when only one description is successfully received. To overcome this problem it is not necessary to reconstruct the midpoints of these quantization bounds, but their centroids instead. Using a Lagrangian formulation, a modified version of the generalized

Lloyd algorithm can be developed to design the quantizers [90], i.e. the partitions for the optimal encoder and the reconstruction levels for the optimum decoder. MD quantization using entropy constrained scalar quantizers can be found in [91].

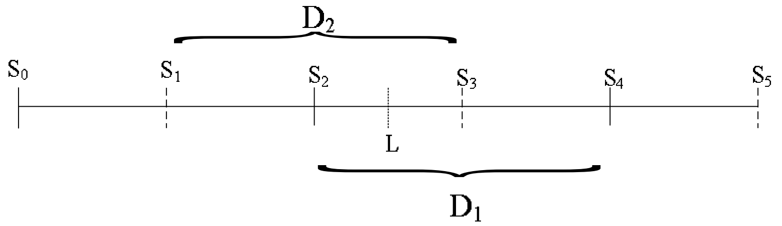


Figure 2.4. Reconstruction levels of the MD scalar quantizers.

Similar methods can be obtained by replacing the scalar quantizers with vector quantizers [91]. In the vector quantization case, a group of successive samples are blocked and the encoder produces the indices as the output of multiple vector quantizers. Basically each description induces a partition on the space. If MDs are received, then the signal is reconstructed as the centroid of the intersection of the cells. Other quantizer-based MD coding approaches include constrained vector quantizers [22], lattice vector quantizers [79] and trellis coded quantizers [35].

2.5 Multiple Description Correlating Transform

In the second category, pioneered by Wang, Orchard and Reibman, an MD correlating transform is constructed by separately quantizing and coding the coefficients of a linear block transform using P descriptions. The coefficients of the transform are scalar quantized and grouped into P descriptions as shown in Fig. 2.5. The block transform F is designed to introduce a controlled amount of correlation between pairs of transform coefficients. The KLT is the optimal transform to minimizes the expected mean-square error.

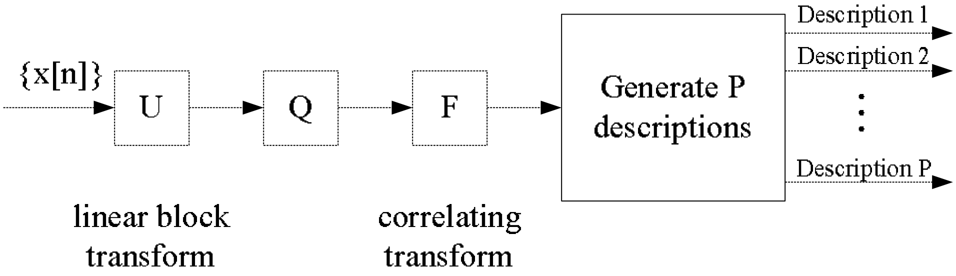


Figure 2.5. General structure of correlating transform based multiple description coding system.

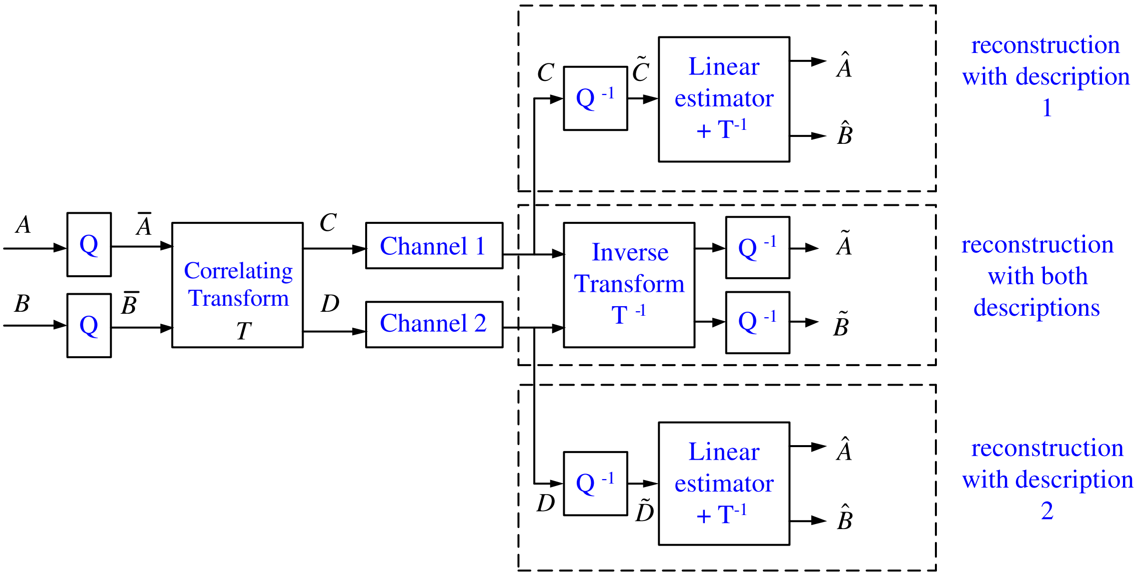


Figure 2.6. Diagram of the pairwise correlating transform based MD coding scheme.

The first introduction of the MDTC concept was for a two-dimensional case with $P = 2$, which is also known as pairwise correlating transform (PCT) [96]. The PCT is used to introduce controlled correlation for uncorrelated variables obtained from the KLT or DCT. Fig. 2.6 shows the block diagram of the PCT method, in which an input pair of Gaussian variables A and B are transformed into C and D, using

$$\begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{T} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}. \qquad (2.21)$$

In [62], Orchard et. al. presented an analysis to optimize the correlating transform to minimize the reconstruction error which was further extended in [96]. It is proved that the optimal PCT transform T is

$$T = \begin{bmatrix} \frac{1}{\sqrt{2\beta}} & \sqrt{\frac{\beta}{2}} \\ -\frac{1}{\sqrt{2\beta}} & \sqrt{\frac{\beta}{2}} \end{bmatrix}, \quad \beta = tan(\theta). \tag{2.22}$$

The transform parameter, β, or the rotation angle, θ, controls the correlation between A and B, which in turn controls the redundancy of the PCT coder. In this method, A and B are assumed to be two independent Gaussian variables with variances σ_A^2 and σ_B^2, respectively. If we parameterize the correlation of C and D by the angle ϕ, defined as $E\{CD\} = \sigma_C \sigma_D cos\phi$, then θ can be used to vary the correlation angle ϕ by $tan(\theta) = \frac{\sigma_A}{\sigma_B} tan(\frac{\phi}{2})$ [97].

Because the transform is, in general, nonorthogonal, quantizing A and B will lead to degraded quantization performance if the quantization is applied to the PCT coefficients. Therefore, A and B are first quantized to generate quantized indices $\overline{A}$ and $\overline{B}$. Then, the transformation is performed on $\overline{A}$ and $\overline{B}$ in the integer domain to produce correlation-induced coefficients C and D. Consequently, (2.21) should be changed to:

$$\begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{T} \begin{bmatrix} \overline{A} \\ \overline{B} \end{bmatrix}.$$

A lifting scheme has been proposed to study how to minimize quantization noise in lossless discrete transforms [47]. For an arbitrary transform, T, with determinant one, it can be factorized as

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (d-1)/b & 1 \end{bmatrix} \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ (a-1)/b & 1 \end{bmatrix}. \tag{2.23}$$

Hence, the lossless integer to integer transform can be implemented as the lifting structure shown in Figure 2.7.

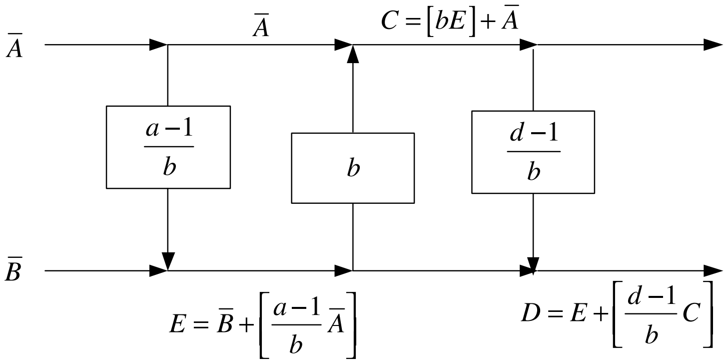


Figure 2.7. Lifting structure for invertible integer-to-integer transform.

The forward PCT with quantization stepsize Q is implemented as ($[\cdot]$ denotes rounding)

$$\begin{aligned} \overline{A} &= \left[\frac{A}{Q}\right], \overline{B} = \left[\frac{B}{Q}\right], \\ E &= \overline{B} + \left[\frac{a-1}{b}\overline{A}\right], \\ C &= [bE] + \overline{A}, \\ D &= E + \left[\frac{d-1}{b}C\right]. \end{aligned} \tag{2.24}$$

Then, C and D are individually entropy coded, and their resulting bitstreams are sent on two separate channels. Assuming both channels are received, the corresponding inverse transform is implemented as

$$\begin{aligned} E &= D + \left[\frac{1-d}{b}C\right], \\ \widehat{A} &= C - [bE], \\ \widehat{B} &= E + \left[\frac{1-a}{b}\widehat{A}\right]. \end{aligned} \tag{2.25}$$

It can be easily shown that the transform pair is reversible between $(\overline{A}, \overline{B})$ and (C, D), i.e. $\widehat{A} = \overline{A}$, $\widehat{B} = \overline{B}$.

If both descriptions are received intact at the destination, the inverse transform of T is applied to C and D to produce $\widehat{A}$ and $\widehat{B}$, which are then used to reconstruct the signal with full quality. If only one of the descriptions is received error-free, e.g. C, inverse quantization is first applied to yield $\widetilde{C}$, the dequantized C as shown in Figure 2.6. An optimal linear predictor is then used to estimate D from $\widetilde{C}$ by minimizing the mean squared error $E\left\{\left[D - D\left(\widetilde{C}\right)\right]^2\right\}$. The estimated D is obtained in the following form:

$$\widetilde{D} = -\frac{\sigma_A^2 - \sigma_B^2 tan^2(\theta)}{\sigma_A^2 + \sigma_B^2 tan^2(\theta)} \cdot \widetilde{C}. \tag{2.26}$$

By taking the inverse transform of T on $\widetilde{C}$ and $\widetilde{D}$, we can obtain the estimated A and B as follows:

$$\begin{bmatrix} \widetilde{A} \\ \widetilde{B} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{tan(\theta)}{2}} & -\sqrt{\frac{tan(\theta)}{2}} \\ \sqrt{\frac{1}{2tan(\theta)}} & \sqrt{\frac{1}{2tan(\theta)}} \end{bmatrix} \begin{bmatrix} \widetilde{C} \\ \widetilde{D} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_A^2 \sqrt{2tan(\theta)}}{\sigma_A^2 + tan^2(\theta)\sigma_B^2} \cdot \widetilde{C} \\ \frac{\sigma_B^2 tan(\theta)\sqrt{2tan(\theta)}}{\sigma_A^2 + tan^2(\theta)\sigma_B^2} \cdot \widetilde{C} \end{bmatrix}. \tag{2.27}$$

Finally, the decoder uses $\widetilde{A}$ and $\widetilde{B}$ to reconstruct the signal but with degraded quality.

Recent analysis on Gaussian distributed random variables [98] suggests that it is optimal to pair all variables according to their variances, with the kth largest paired with the $(N-k)$th largest if there are N variables. However, when the smaller variable's variance in a pair is too small relative to the quantization error, it is no longer meaningful to pair this variable. In this case, the two transformed coefficients' variances are very similar, so that the actual redundancy is larger than that predicted by theory, particularly for very low redundancies.

Therefore, for a given redundancy, the pairing should be applied only to variables having a large enough variance. Given variables, only the variables with large variances should be paired. The remaining variables with small variances can be simply split among the two descriptions. In the event that they are lost due to a channel failure, they can be simply estimated by their mean values. The estimation errors for these small variables are on the same order as the quantization error.

The extension of MDTC for more than two channels was presented by Goyal [28]. A solution for $P = 3$ was given by (2.28) for arbitrary a [25], when $\sigma_1 > \sigma_2 > \sigma_3$ and description loss probabilities are equal and small, where σ_i is the corresponding variance of the input random variable i of the MDTC.

$$\begin{bmatrix} a & -\frac{\sqrt{3}\sigma_1 a}{\sigma_2} & -\frac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2} \\ 2a & 0 & \frac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2} \\ a & \frac{\sqrt{3}\sigma_1 a}{\sigma_2} & -\frac{\sigma_2}{6\sqrt{3}\sigma_1^2 a^2} \end{bmatrix}. \tag{2.28}$$

A simple heuristic for designing systems with more than two channels is to cascade small transforms [28]. Just as the parallel use of two-by-two transforms gave a method for sending $2K$ variables over two channels, where K is a positive integer. A cascade combination of these transforms gives a method for sending 2^K variables over 2^K channels. The cascade structure simplifies the encoding, decoding, and design when compared to using a general $2^K \times 2^K$ transform.

2.6 Multiple Description Coding with Partitioning

The methods of Class C) are the most straightforward ways to generate MDs. Typically the descriptions are created by partitioning, duplicating or combining both strategies for the SD coefficients. In other words, the SD data is decomposed to multiple non-overlapping/overlapping subsets, in the spatial, temporal, or frequency domain. This method takes advantage of the fact that spatially or temporally adjacent image/video data samples are correlated. Thus, one description can be estimated from the others. One benefit of this approach is that no or less side information is needed. Another benefit is that it is easy to extend the algorithm to more than two descriptions by increasing the number of partitions. However, it is expected that the redundancy due to coding efficiency will increase rapidly as the number of descriptions grows.

The partitioning can be flexible based on how descriptions are generated. In [73], Ridge et. al. proposed the permuted smoothed description coding, which generates descriptions by partitioning DCT blocks into alternative odd/even stripes combined with maximum smoothing error concealment [9]. Boulgouris applied similar checkboard partitioning to wavelet subbands for each frame of a video sequence [9].

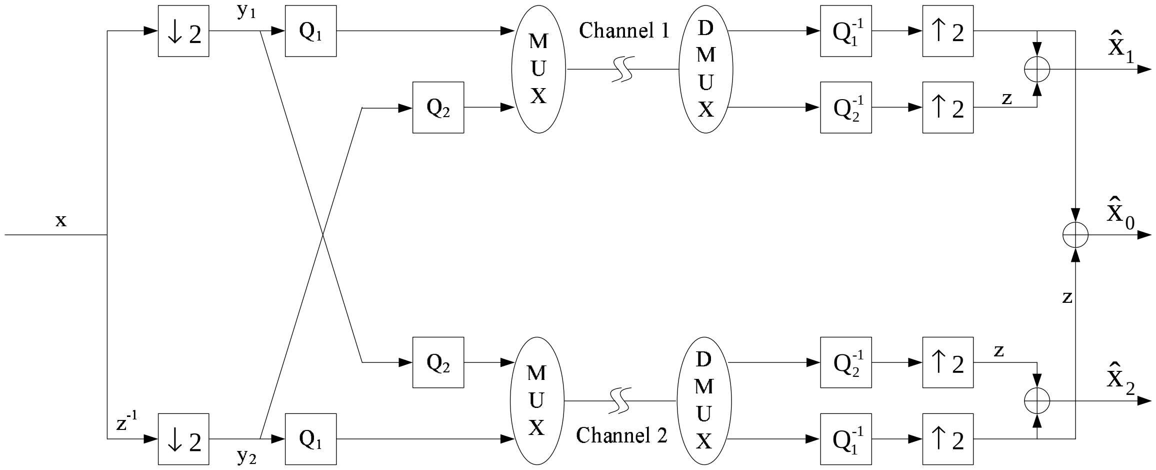


Figure 2.8. Diagram of polyphase-based MDC.

Jiang and Ortega proposed an MD coder by separating wavelet coefficients into polyphase components [37], which is well known as *Polyphase-based MDC*. As described in Fig. 2.8, the input signal x is first decomposed into two subsets y_1 and y_2 via a polyphase transform. Each of these two components is quantized independently by Q_1 and constitutes the primary part of the information for its corresponding channel. For reconstruction of the other channel in case of loss, each channel also carries information about the other channel, a coarsely quantized version by Q_2. Then quantized outputs from Q_1 and Q_2 are multiplexed together for transmission. At the receiver, if data from both channels arrive, fine quantized data of both polyphase components is then used for reconstruction. If one channel data is lost, one fine quantized polyphase and one coarsely quantized polyphase component are used for reconstruction.

This method explicitly separates the redundant information (coarsely quantized polyphase components) from the primary information (finely quantized polyphase components). It simplifies both the encoding and decoding processes compared to the MDTC system where a correlating transform has to be executed. The complexity is also decreased compared to the MDSQ system since both the polyphase transform and quantizers can be implemented easily. However, the way of partitioning by using polyphase is not efficient for exploring the inherit correlation of transform coefficients. Thus this method is not suitable for better error concealment. On the other hand, it is difficult to determine the optimal bit rate allocated to the redundant information at low bit rates. This actually motivate our proposed novel partitioning-based MDC algorithm presented in Chapter 4.

Many partitioning based approaches are also developed for MDC of video signals. In the video redundancy coding (VRC) algorithm [106], a video sequence is temporally down-sampled into two subsets, essentially with every other frame in each subset. The frames in each subset are coded into a description using an SD video encoder. At the decoder, if only one description is received, the missing frames can be estimated from the received frames.

Reibman et al. proposed the MD-split method [20], which uses the simplest possible algorithm for the MD encoder: duplication or alternation. Motion vectors and a varying number of low-frequency DCT coefficients are duplicated, while the remained high-frequency coefficients are alternatively split to one of the two descriptions. The number of coefficients to duplicate can be adapted easily based on varying source and channel statistics without explicitly informing the decoder.

In parallel, Comas et al. [21] proposed an algorithm with very similar properties. Again, coefficients are either duplicated or sent in just one of the two descriptions. However, one description of a block contains all coefficients while the other contains only low-frequency coefficients, i.e. one channel gets all coefficients and the other gets a subset of coefficients. By alternating on a frame or block basis for the two channels using this strategy, balanced distortion can be achieved for each channel.

The MD-Split coder duplicates the lower frequency coefficients only. Recently, Kim and Cho [41] extended this to allow any coefficients to be duplicated. Instead of operating on the DCT coefficients, the matching pursuits MD video coder proposed by Tang et al. [87] duplicates and alternates matching pursuits atoms to each description. Redundancy is controlled by adjusting the number of duplicated atoms.

2.7 Forward Error Correction Codes-Based Multiple Description Coding

Instead of designing the source encoder to yield multiple descriptions directly, one can apply unequal cross-packet FEC to different parts of a scalable bitstream. This method, pioneered in [67] is commonly known as MD-FEC, which converts the prioritized multiple-resolution bitstream into an N-packet un-prioritized MD packet stream using efficient erasure channel codes. Each description in the MD stream occupies an entire network packet, thus the terms "description" and "packet" are used interchangeably. Let $d(k)$ be the distortion when any k out of N descriptions are received.

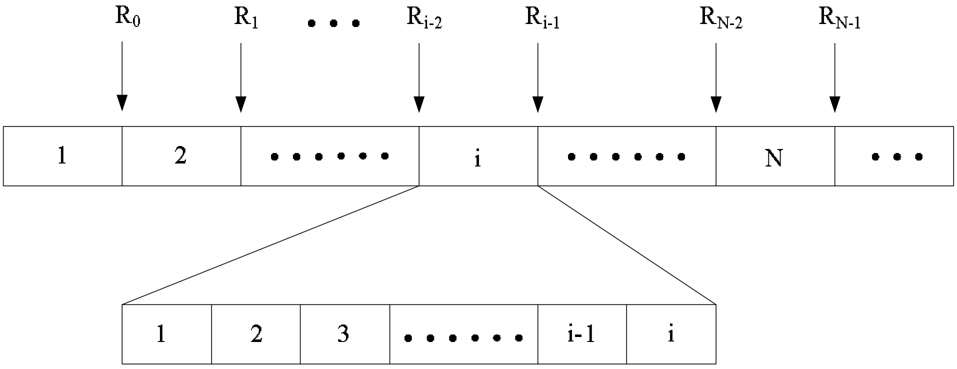


Figure 2.9. Progressive bitstream from the source coder partitioned into N layers.

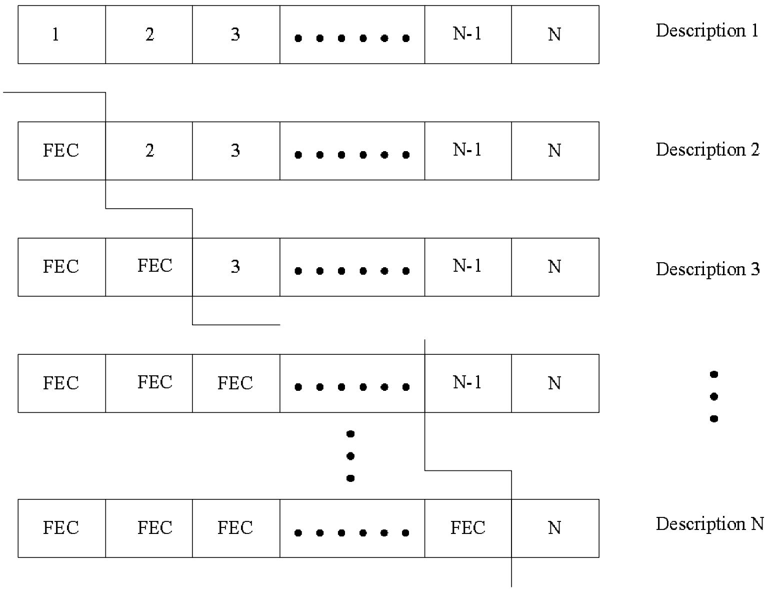


Figure 2.10. N-description generalized MD codes using FEC.

Given N and $d(k)$ and a progressive bitstream, the stream is marked at N positions as shown in Fig. 2.9 corresponding to the distortion levels $d(k)$ and is thus partitioned into N sections or resolution layers. The goal is to enable the ith layer to be decodable when i or more descriptions arrive at the receiver. This can be attained using the Reed-Solomon family of erasure-correction block codes. The ith layer is split into i equal parts, and the $(N, i, N-i+1)$ Reed Solomon code is applied to get the contribution from the ith layer to each of the N descriptions. The contributions from each of the N levels are then concatenated to form the N descriptions as described in Fig. 2.10. This packetization strategy provides the property that the more the number of packets received, the better the received quality.

Assume the channel model is characterized by $q(i)$, $i = -1, \ldots, N-1$, denoting the probability that $i+1$ out of N packets are delivered to the destination. From operational rate-distortion theory, it follows that distortion is a one-to-one function, $D(r)$, of the rate r. Hence determining $d(k)$ of order N corresponds to finding the rate partition $\{R = R_0, \ldots R_{N-1}\}$ of the bitstream. Let $E\{D\}$ be the expected distortion:

$$E\{D\} = q_{-1} \cdot E + \sum_{j=0}^{N-1} q_j \cdot D(R_j), \tag{2.29}$$

where E denotes the source variance, i.e. the distortion encountered when the source is represented by zero bits. Thus the total rate used R_t equals:

$$R_t = \frac{R_0}{1} \cdot N + \frac{R_1 - R_0}{2} \cdot N + \ldots + \frac{R_{N-1} - R_{N-2}}{N} \cdot N. \tag{2.30}$$

Hence, the optimization problem of FEC-based MDC is equivalent to finding R that minimizes $E\{D\}$ subject to [66]:

$$R_t \leq R^*, \tag{2.31}$$

$$R_0 \leq R_1 \leq \ldots \leq R_{N-2} \leq R_{N-1}, \tag{2.32}$$

$$R_i - R_{i-1} = k_i \cdot (i+1), \quad k_i \geq 0, \quad i = 1, \ldots, N-1. \tag{2.33}$$

A fast, nearly optimal algorithm to solve this optimization problem based on Lagrangian optimization has been discussed in [67]. Kim et. al. extended this method to 3-D SPIHT [40] coders for video coding [42]. In this approach, multiple-substream bitplane-wise UEP codes are packetized into different channels so as to generate unbalanced MDs. Thus, the important spatial and temporal information are more robust to packet loss.

2.8 Summary

This chapter reviews the state-of-the-art approaches of conventional MDC. We classify them into four categories and describe the typical implementation for each class. Amongst the variant research interests of MDC, improving quality of the reconstructed signal and realizing optimal redundancy control are the most interesting topics. In this thesis, the proposed algorithms are closely related to the MD coding approaches that use partitioning and correlating transform.

CHAPTER 3

RECONSTRUCTION FOR WAVELET-BASED MULTIPLE DESCRIPTION IMAGE CODING

3.1 Introduction

Recently, Bajic [5] proposed a novel MD wavelet coding scheme, namely domain-based MDC or data-partitioning based MDC (DP-MDC), which is created by optimally partitioning the wavelet transform coefficients. The method is bandwidth-efficient in the sense that no extra redundancy is introduced. Only the inherent spatial redundancy in the transform domain of the signal is employed to make the bitstream robust and enable lost descriptions to be estimated from the received ones. However, this MDC system only focuses on finding the best way to partition the transform coefficients but ignores efficiently adopting the inherent correlation within the signals in the reconstruction of the corrupted coefficients at the decoder.

In this chapter, we concentrate on improving the reconstruction of lost descriptions based on DP-MDC. Two algorithms are proposed. In the first algorithm, the corrupted wavelet coefficients are estimated with the aid of the wavelet Hidden Markov Tree (HMT) model [15]. Thus the cross-scale correlations can be exploited to reconstruct the lost coefficients. In the second algorithm, instead of only using the immediate available neighbors of the corrupted samples for reconstruction [54], we explore all the edge orientation correlations within a local window so that we can estimate the corrupted samples using the covariance method adaptively corresponding to the local texture statistical characteristics. As a result the lost descriptions can be more accurately restored. In this sense, our reconstruction algorithm is edge-adaptive to the local texture. Hence the proposed scheme is named as edge adaptive MD coder (EA-MDC).

The rest of this chapter is organized as follows. Section 3.2 reviews the DP-MDC scheme proposed in [5]. In Section 3.3, we briefly discuss the HMT model and propose a HMT based algorithm to reconstruct the corrupted descriptions for DP-MDC. Experimental results are compared to those of DP-MDC. In Section 3.4, the proposed EA-MDC framework is presented. We also evaluate the proposed coder from the sense of information theory and give the realization of fast implementation. Extensive experiment results are discussed. Finally, we make our conclusion in Section 3.5.

3.2 Domain Partitioning Based MD Coder

Successful estimation of missing signal coefficients relies on the autocorrelation of the signal. The autocorrelation is typically a decreasing function of distance. It is expected to get the best results if the missing coefficient is estimated from its immediate neighbors. Moreover, the distortion will be a non-increasing function of the size of the available neighbors surrounding a missing signal coefficient for typical estimation algorithms [5]. Hence, Bajic proposed an optimal way to partition the coefficients in transform domain.

Consider a signal f defined on the domain $\mathcal{D}$. In the case of digital images or video frames, the domain is typically a subset of the $\mathbf{Z}^2$ lattice. Suppose we wish to create descriptions of the signal. This would amount to constructing a collection of nonempty sets $\mathcal{C} = \{S_0, S_1, ..., S_{P-1}\}$ such that $\bigcup_{i=0}^{P-1} S_i = \mathcal{D}$. In the special case when $S_i \bigcap S_j = \emptyset, i \neq j$, $\mathcal{C}$ defines a *partition* of the domain $\mathcal{D}$. In MDC, S_i forms the *ith* description of the signal. Without introducing extra redundancy within the partition, the reconstruction distortion will depend on the ability of the decoder to estimate the missing samples from those that are available if each description has the same probability of being lost.

We define the *minimal intra-partition distance* of the *ith* partition as

$$d_{min}^{(i)} = \min_{x,y \in S_i; x \neq y} d(x,y), \tag{3.1}$$

where $d(x,y)$ is the the Euclidean distance between x and y. The *minimal distance* d_{min} is

$$d_{min} = \min_{i=0,1,...,P-1} d_{min}^{(i)}. \tag{3.2}$$

As d_{min} increases, the samples belonging to S_i get further away, meaning that more samples belonging to other subsets surround those from S_i. Hence, the neighborhood which can be used for their estimation (if S_i is lost) increases with d_{min}. Based on this, given the number of partitions P, the optimal partition C^* of the signal domain is

$$C^* = arg \max_{|C|=P} d_{min}(C), \tag{3.3}$$

where $|C|$ is the number of elements in C. A solution is given in [6] based on lattice partitioning. Figure 3.1 shows how Z^2 lattice can be partitioned into $P = 15$ partitions with $d_{min} = \sqrt{17}$. The pattern of dark dots (enclosed in the parallelogram, labelled 0 through 14) is repeated throughout the lattice, and then domain is cut out of the lattice. The label assigned to the point shows the index of the partition to which the point belongs. This labelling defines the *partitioning function* $p : Z^2 \rightarrow \{0, 1, ..., P-1\}$ which is 2-D periodic, it can be described by a 2×2 *periodicity matrix* $\mathbf{V} = [v1, v2]$ whose columns are the basis vectors for the partitioning. The number of partitions $P = |det(\mathbf{V})|$. For example, we have $\mathbf{V} = [4, 3; -1, 3]^T$ as in Fig. 3.1.

With the optimal domain partitioning, the transform coefficients are conveniently divided into several subsets, i.e. descriptions. An example, given in Fig. 3.2, illustrates a two-level wavelet decomposition of a 16×16 image, generating $P = 4$ descriptions. Numbers 0 through 3 indicate the partition to which the corresponding subband coefficient

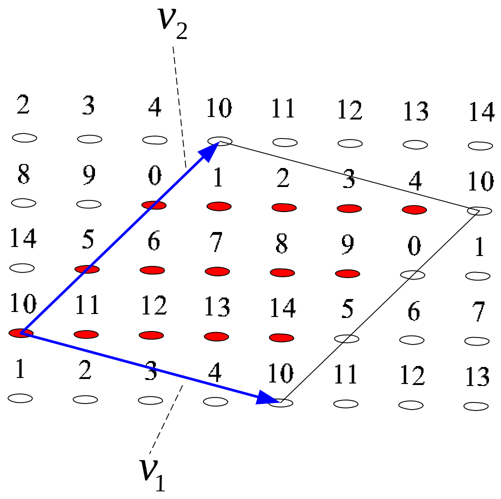
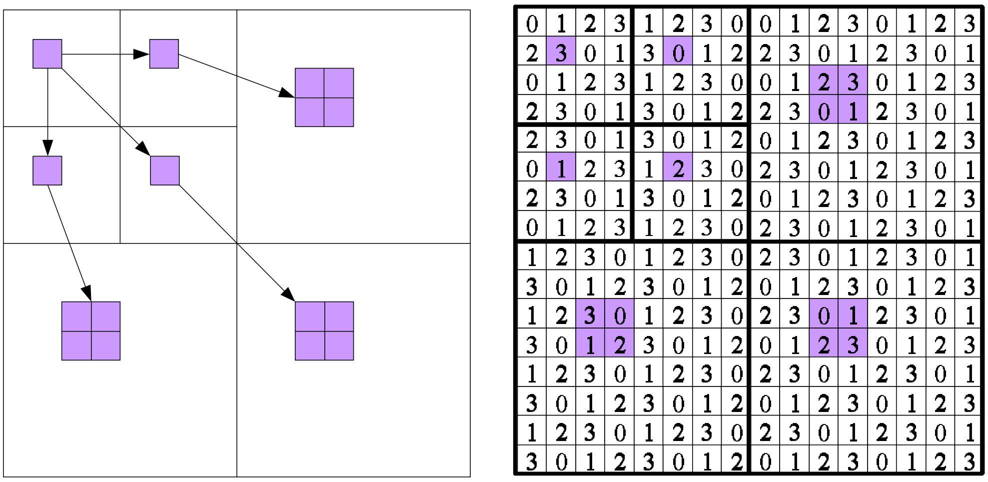


Figure 3.1. Example of optimal partitioning with $P = 15$.

belongs. In this case, the minimum intra-partition distance is maximized so that more neighbors from other descriptions are expected around the corrupted samples. Assume the subbands are ordered in a zig-zag manner from the lowest frequency band to the highest frequency band, and indexed as $k = 0, 1, ..., 6$. Given the partitioning scheme $p(x)$ for the lowest frequency band, the partition for each higher frequency band is obtained by applying modulo-shifted partitioning function, $p^{(k)}(x) = [p(x) + k]\ mod\ P$. It is apparent from Fig. 3.2 that each description contains the same number of coefficients from each tree and also the same number of coefficients from each subband. This makes the descriptions "equally important." After partitioning, each description is encoded independently from other descriptions.

3.3 HMT Model-Based Lost Description Reconstruction

Since the simple bi-linear interpolation only employs partial correlations inherited from the same subband, higher order correlations can be used within the spatial oriental tree. Motivated by this point, we proposed a lost description estimation algorithm based on wavelet hidden markov tree model to improve the DP-MDC.



(a) (b)

0	1	2	3	1	2	3	0	0	1	2	3	0	1	2	3
2	3	0	1	3	0	1	2	2	3	0	1	2	3	0	1
0	1	2	3	1	2	3	0	0	1	2	3	0	1	2	3
2	3	0	1	3	0	1	2	2	3	0	1	2	3	0	1
2	3	0	1	3	0	1	2	0	1	2	3	0	1	2	3
0	1	2	3	1	2	3	0	2	3	0	1	2	3	0	1
2	3	0	1	3	0	1	2	0	1	2	3	0	1	2	3
0	1	2	3	1	2	3	0	2	3	0	1	2	3	0	1
1	2	3	0	1	2	3	0	2	3	0	1	2	3	0	1
3	0	1	2	3	0	1	2	0	1	2	3	0	1	2	3
1	2	3	0	1	2	3	0	2	3	0	1	2	3	0	1
3	0	1	2	3	0	1	2	0	1	2	3	0	1	2	3
1	2	3	0	1	2	3	0	2	3	0	1	2	3	0	1
3	0	1	2	3	0	1	2	0	1	2	3	0	1	2	3
1	2	3	0	1	2	3	0	2	3	0	1	2	3	0	1
3	0	1	2	3	0	1	2	0	1	2	3	0	1	2	3

Figure 3.2. Example of generating descriptions of a two-level decomposition using optimal data partitioning: (a) the wavelet coefficients of a spatial orientation tree, (b) four descriptions generated for the subbands, the minimum intra-partition distance is 2 in this case.

3.3.1 Wavelet Hidden Markov Tree (HMT) Model

It is well known that the wavelet coefficients of a natural image have the following properties: [74]

(a). **Non-Gaussianity:** The wavelet coefficients have peaky, heavy-tailed marginal distributions.

(b). **Persistency:** Large/small values of wavelet coefficients tend to propagate through the scales of the quad-trees.

These properties give rise to joint wavelet statistics that are succinctly captured by the wavelet-domain hidden Markov tree model [15]. The HMT models the non-Gaussian marginal probability density function (pdf) as a two-component Gaussian mixture. The components are labelled by a hidden state signifying whether the coefficient is small or large. The Gaussian component corresponding to the small state has a relatively small variance, capturing the peakiness around the mean values, while the component corre-

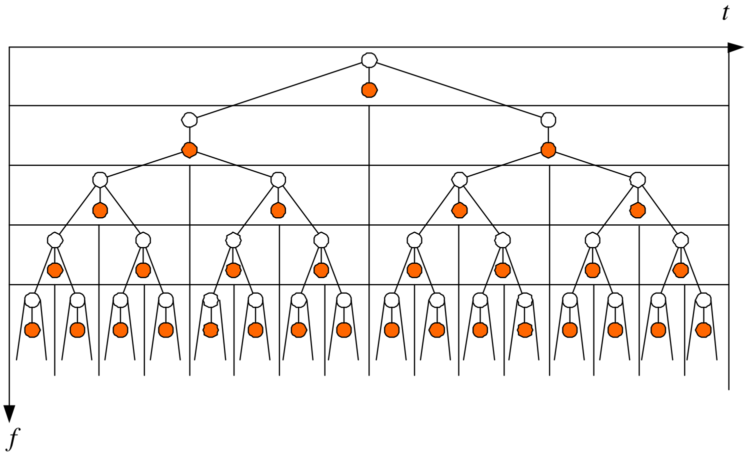


Figure 3.3. Hidden markov tree model for one-dimensional wavelet transform: every coefficient is modelled as a mixture with a hidden state variable, each dark node represents a continuous wavelet coefficient, connecting discrete nodes vertically across scale (solid links) yields the hidden markov tree model.

sponding to the large state has a relatively large variance, capturing the heavy tails (the shape of the heavy-tailed density in the region where large values are likely). The persistency of wavelet coefficient magnitudes across scale is modelled by linking these hidden states across scales in a Markov tree (see Fig. 3.3). A state transition matrix for each link quantifies statistically the degree of persistency of large/small coefficients. Given a set of training data (usually in the form of one or more observed images), maximum likelihood estimates of the mixture variances and transition matrices can be calculated using the Expectation-Maximization (EM) algorithm [15]. These parameter estimates yield a good approximation of the joint density function of the wavelet coefficients.

The form for the marginal distribution of a wavelet coefficient comes directly from the efficiency of the wavelet transform in representing real-world images: a few wavelet coefficients are large, but most are small. Gaussian mixture modelling runs as follows. Each wavelet coefficient is associated with an unobserved hidden state variable $S_i = m, m = 0, ..., M-1$, where M is the number of states for each node. If $M = 2$, then

S_0 and S_1 denote the states for coefficients with small and large variances. A state corresponds to a zero-mean, low-variance Gaussian (Note that we only consider the high-frequency subbands). Let

$$g(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}.$$

denote the Gaussian pdf, then for the wavelet coefficient w_i, we have

$$f(w_i|S_i = m) = g(w_i; 0, \sigma^2_{S_i=m}), \quad m = 0, 1, \tag{3.4}$$

corresponding to a zero-mean, low or high-variance Gaussian pdf for each wavelet coefficient in the highpass bands. The marginal pdf $f(w_i)$ is obtained by a convex combination of the conditional densities

$$f(w_i) = \sum_m p(S_i = m) g(w_i; 0, \sigma^2_{S_i=m}). \tag{3.5}$$

Since the persistency property implies a type of Markovian relationship among the wavelet states, with the probability of a wavelet coefficient being “large” or “small” affected only by the size of its parent. The HMT models the dependence as one order Markov: given the state of a wavelet coefficient S_i, the coefficient's ancestors and descendants are independent from each other.

The HMT captures dependency by using a probabilistic tree that connects the hidden state variable of a wavelet coefficient with the state variable of each of its children. This leads to the dependency graph having the same quad-tree topology as the wavelet coefficients (see Fig. 3.4). Each subband is represented with its own quad-tree; this assumes that the subbands are independent. Each parent child state-to-state link has a corresponding state transition matrix

$$A_i = \begin{bmatrix} \epsilon^{00}_{i,\rho(i)} & \epsilon^{01}_{i,\rho(i)} \\ \epsilon^{10}_{i,\rho(i)} & \epsilon^{11}_{i,\rho(i)} \end{bmatrix}, \tag{3.6}$$

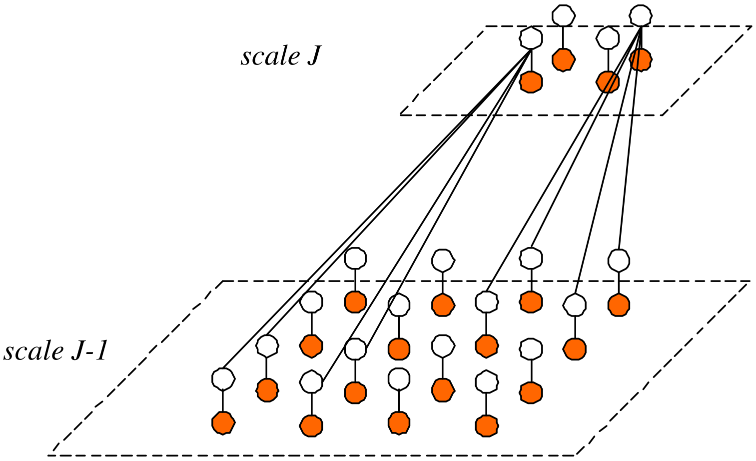


Figure 3.4. HMT for an image quadtree. Each parent hidden state is connected to its four child states. (Some parent-child connections are omitted for visual clarity.) The two fields of nodes depict the wavelet coefficients at scales J and $J-1$, respectively, and corresponding to 2-D wavelet basis functions with a specific spatial orientation.

where $S_{\rho(i)}$ is the parent state of node i, $\epsilon^{mr}_{i,\rho(i)} = p_{S_i|S_{\rho(i)}}\left[m|S_{\rho(i)}=r\right]$ represents the conditional probability that S_i is in state m given $S_{\rho(i)}$ is in state r.

With the non-Gaussianity and persistence characterized, the HMT model can be determined by the following parameters:

(a). the mixture variances $\sigma^2_{i,m}$;

(b). the state transition matrices A_i;

(c). the probability of a large state at the root node for each i in the coarsest scale p^L_i, where L is the number of decomposition levels.

Grouping these into a vector Θ, the HMT provides a parametric model for the joint pdf $f(w|\Theta)$ of the wavelet coefficients. In general, the variance and transition parameters can be different for each wavelet coefficient. However, this makes the model too complicated for some applications. For example, if there is only one observation of an n-pixel image, then we face with the impossible task of fitting parameters to data points. To reduce the

HMT complexity, we can make the simplifying assumption that each parameter is the same within a subband at each scale of the wavelet transform.

3.3.2 HMT Model Construction with Received Descriptions

Since partial descriptions are received at the decoder, the HMT model can be built using the available coefficients with EM algorithm [15]. Because the wavelet coefficients within a subband are modelled with the same two zero-mean Gaussians, the lost coefficients will follow the same model. Hence, a bilinear interpolation using the neighboring uncorrupted coefficients is applied to the lost samples for EM algorithm initialization.

The EM algorithm starts with an initial guess Θ^0 of the model parameters, then for each iteration l, $E_S\left[lnf(w,S|\Theta)|w,\Theta^l\right]$ is calculated. Finding this expectation, called the "**E step**", amounts to calculating the state probabilities $p(S_i = m|w,\Theta^l)$, for which the upward-downward algorithm is used [68]. The maximization, or "**M step**", consists of relatively simple, closed form updates of the parameters in Θ^l to obtain Θ^{l+1}. As $l \rightarrow \infty$, Θ^l approaches a local maximum of the likelihood function $f(w|\Theta)$.

The brief iterative structure of the model construction algorithm is listed as follows:

(a). Initialize equal probability for Θ^0, and set iteration counter $l = 0$.

(b). **E step:** Calculate $p(S|w,\Theta^l)$, which is the joint pdf for the hidden state variables (used in the maximization of $E_S\left[lnf(w,S|\Theta)|w,\Theta^l\right]$).

(c). **M step:** Set $\Theta^{l+1} = arg\ \max_\Theta E_S\left[lnf(w,S|\Theta)|w,\Theta^l\right]$.

(d). Set $l = l + 1$. If it converged, then stop the iteration; else, return to E step.

3.3.3 Lost Description Reconstruction with HMT Model

Since the initial estimation of the lost description is obtained by using bilinear interpolation from its available neighboring coefficients, the resulting estimation error is still a joint Gaussian random variable.

For convenience of illustration, let $w_{m,n}$ be the lost wavelet coefficient located at the position of (m,n) within a subband. Its neighboring four coefficients, $w_{m,n-1}$, $w_{m-1,n}$, $w_{m,n+1}$ and $w_{m+1,n}$, if available, are bilinearly interpolated to estimate $w_{m,n}$, which results in $\widehat{w}_{m,n}$. The corresponding estimation error is denoted as

$$\varepsilon = w_{m,n} - \widehat{w}_{m,n} = w_{m,n} - \frac{1}{4}(w_{m-1,n} + w_{m+1,n} + w_{m,n-1} + w_{m,n+1}).$$

Apparently, $E\{\varepsilon\} = 0$ according to the HMT model. Let $\alpha = E\{w_{m,n}w_{m\pm1,n}\}$, $\beta = E\{w_{m,n}w_{m,n\pm1}\}$, and $\gamma = E\{w_{m\pm1,n}w_{m,n\pm1}\}$. Thus we have

$$\begin{aligned} \sigma_\varepsilon^2 = E\{\varepsilon^2\} &= E\{[w_{m,n} - \frac{1}{4}(w_{m-1,n} + w_{m+1,n} + w_{m,n-1} + w_{m,n+1})]^2\} \\ &= \frac{5}{4}\sigma^2 - \frac{1}{2}\alpha - \frac{1}{2}\beta + \frac{3}{8}\gamma. \end{aligned} \tag{3.7}$$

Since a wavelet coefficient has two states corresponding to the mixture variances $\sigma_{S_i}^2$, Eq.(3.7) is adjusted to

$$\sigma_\varepsilon^2(i,m) = E\{\varepsilon_i^2|S_i = m\} = \frac{5}{4}\sigma_{i,m}^2 - \frac{1}{2}\alpha - \frac{1}{2}\beta + \frac{3}{8}\gamma, \tag{3.8}$$

where $\sigma_\varepsilon^2(i,m)$ denotes the estimation error variance for the ith wavelet coefficient within the current subband while $\sigma_{i,m}^2$ is the mixture variance retrieved from the HMT model.

Hence the initial estimation error is a zero-mean white Gaussian noise, the lost description estimation problem can be expressed in the wavelet domain as

$$w_i = y_i + \varepsilon_i, \tag{3.9}$$

where y_i is the wavelet coefficient of the signal to be reconstructed, ε_i denotes the noise, i.e. the initial estimation error. Using the Bayes estimator, the conditional mean estimate of y_i is [38]

$$E\left[Y_i|W_i = w_i, S_i = m\right] = \frac{\sigma_{i,m}^2}{\sigma_\varepsilon^2(i,m) + \sigma_{i,m}^2} w_i. \tag{3.10}$$

Considering the hidden state probabilities $p(S_i|w,\Theta)$, we obtain conditional mean estimates for y_i via the chain rule for conditional expectation

$$E\left[y_i|w,\Theta\right] = \sum_m p(S_i = m|w,\Theta) \times \frac{\sigma_{i,m}^2}{\sigma_\varepsilon^2(i,m) + \sigma_{i,m}^2} w_i. \qquad (3.11)$$

The final decoded signal is computed as the inverse wavelet transform of the coefficients including these estimates of the lost samples. Note that only the wavelet coefficients are processed. The original scaling coefficients are used in the inverse transform.

3.3.4 Experimental Results


(a) (b)

Figure 3.5. Lena image is compressed at 0.21 bpp for SDC. Four descriptions are generated and description 0 is assumed to be lost: (a) reconstructed image presented by DP-MDC, PSNR=26.81 dB; (b) reconstructed image using the proposed algorithm, PSNR=27.88 dB.

We carried out a set of experiments on 512×512 gray scale Lena and barbara images. The images are transformed with a 4-level wavelet decomposition using Daubichies 9/7


(a) (b)

Figure 3.6. Barbara image is compressed at 0.40 bpp for SDC. Description 0 is assumed to be lost among the four transmitted descriptions: (a) reconstructed image presented by DP-MDC, PSNR=24.22 dB; (b) reconstructed image using the proposed algorithm, PSNR=24.88 dB.

filters. Four descriptions are generated with optimal partitioning and independently coded. The LL band in the coarsest level is duplicated into each channel as redundancy. In each experiment one description is assumed to be corrupted, the reconstructed images are obtained from the estimated wavelet coefficients.

The PSNR performance of the proposed algorithm is compared with that of the DP-MDC. Figs. 3.5 - 3.6 show the reconstructed images for both methods when description 0 is assumed to be corrupted. Evidently, our algorithm consistently outperforms DP-MDC for both images with the same loss ratio. The gain over DP-MDC is significant, which can be up to 1.07 dB for Lena when the bit-rate of SDC is 0.21 bpp.

3.4 Edge Adaptive MD Coder

The HMT model based algorithm involves high computation loads when estimating the model from the received descriptions. Moreover, the ability of improving the reconstruction is constrained by the accuracy of the HMT model. Hence, it is not applicable for real-world image coding. In this section, we present an edge-adaptive (EA) error concealment technique that improves the corrupted coefficients reconstruction quality by fully taking advantage of the intra-band correlations.

3.4.1 Edge-Adaptive Error Concealment Algorithm

This section addresses the problem of estimating a lost coefficient X_0 from its available neighbors. The available neighbors could be either coefficients belonging to the successfully received descriptions, or previously recovered coefficients. To facilitate the discussion, we use the set $L_X = \{X_1, X_2, ..., X_L\}$, $L \geq 1$, also named as *local window*, to denote the available neighbors of X_0. The atomic problem in recovery is to obtain the optimal estimation of X_0 in the sense of maximizing $p(X_0|L_X)$.

Although the wavelet coefficients for natural images can not be modelled by a stationary Gaussian process, it can be viewed as locally stationary. Hence, the conditional probability distribution function $p(X_0|L_X)$ is mostly characterized by the second-order statistics (covariance). On the other hand, the Lth-order Markov property holds for image source, i.e. $p(X_0|X_1, X_2, ...) = p(X_0|X_1, X_2, ..., X_L)$. Therefore, the Maximum A Posterior (MAP) estimate $p(X_0|L_X)$ comes down to the linear minimum mean square error (MMSE) estimate problem of minimizing $\{E[(X_0 - \hat{X}_0)^2]\}$, where $\hat{X}_0$ is the linear estimation of X_0 based on the set of its immediate neighbors $N_X = \{X_1, X_2, ..., X_N\}$, with N ($N \leq L$) denoting the number of the immediate neighbors used in the linear estimate

$$\hat{X}_0 = \sum_{k=1}^{N} w_k X_k. \tag{3.12}$$

According to the classical Wiener filtering theory [36], the optimal weights, $\vec{w} = [w_1, w_2, ..., w_N]^T$ can be determined by the covariance of the Gaussian process

$$\vec{w} = (R_{XX})^{-1}\vec{r}_X, \tag{3.13}$$

where $\vec{r}_X = [r_1, r_2, ..., r_N]^T$, $r_k = cov(X_0 X_k)$, $1 \leq k \leq N$, and $R_{XX} = [R_{kl}]$, $R_{kl} = cov(X_k X_l)$, $1 \leq k, l \leq N$. The vector $\vec{w}$ in (3.13) is actually the projection of X_0 onto the subspace spanned by N_X in the least-square sense. Under the assumption of a locally stationary process, we can estimate R_{XX} based on the covariance [36] as,

$$\hat{R}_{XX} = C^T C, \qquad \hat{\vec{r}}_X = C^T \vec{l}, \tag{3.14}$$

where $\vec{l} = [X_1, X_2, ..., X_L]^T$ is an $L \times 1$ vector composed of all the available neighbors inside the local window L_X. C is an $L \times N$ matrix whose lth row contains the N immediate neighbors used to estimate X_l, and is given by

$$C = \begin{bmatrix} X_{1,1} & \ldots & X_{1,N} \\ \vdots & \ddots & \vdots \\ X_{L,1} & \ldots & X_{L,N} \end{bmatrix}. \tag{3.15}$$

By combining (3.13) and (3.14), the optimal weight $\vec{w}$ is obtained by

$$\vec{w} = (C^T C)^{-1}(C^T \vec{l}). \tag{3.16}$$

Image intensity field experiences a sharp transition across the edge orientation and is almost homogeneous along the edge orientation. These geometric constraints of edges hold in the wavelet domain, especially for high resolution subbands. The above covariance-based estimate can be viewed as an edge-adaptive error concealment method that is capable of tuning $\vec{w}$ along an arbitrarily-oriented edge across the local window. Therefore, important edge features can be well estimated. On the other hand, this method is also enhanced with the way to generate descriptions by DP-MDC. When only one description is lost, all the immediate neighbors of lost coefficients are available at the advantage of optimal partitioning. Hence, more accurate local texture information can be retrieved from the local window.

Intuitively the local neighbors can be divided into two classes: edge samples (close to edges) and non-edge samples (away from edges, i.e. samples in smooth areas). For the non-edge samples, the matrix $C^T C$ is often singular or close to singular, thus the solution of $\vec{w}$ is not unique but lies in the hyperplane $\sum_{k=1}^{N} w_k = 1$, which is approximately isotropic for any direction. In this case the solution of (3.16) can not be determined or may cause ill-posed estimation. Hence the simple bilinear interpolation instead of EA method is applied under this situation.

The elements of L_X and N_X include the available neighbors of X_0, which are the coefficients from the received descriptions and the corrupted coefficients that have been previously reconstructed. However, some neighbors of X_0 might not be available. Hence, it is necessary to check the validity of the elements in N_X and L_X. An element is said to be valid if it is either successfully received or previously reconstructed. As shown in Fig. 3.7, the valid L_X includes all the coefficients inside the $(2T+1) \times (2T+1)$ square window except for X_0 and the coefficients that have not been reconstructed yet. The valid N_X includes the immediate 8 coefficients of the successfully received descriptions. Accordingly, the related rows and elements for invalid coefficients will be removed from C and $\vec{l}$.

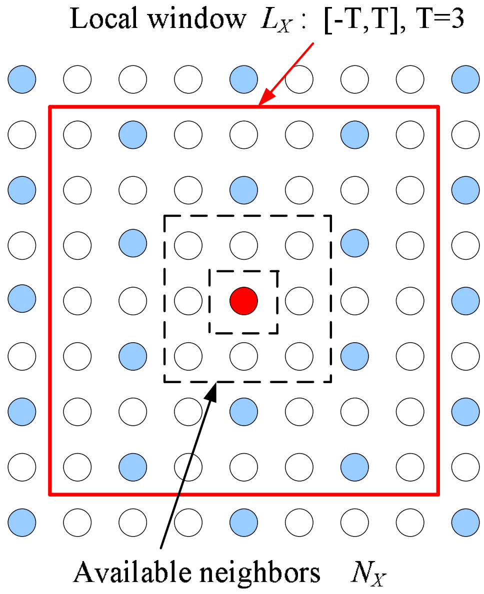


Figure 3.7. Definition of local window L_X and the set of neighbors N_X under the 4-description case described as in Fig. 3.2. The white points are coefficients of the successfully received descriptions, the gray points represent the corrupted coefficients from the lost description. The center dark point is the one to be estimated currently.

3.4.2 Evaluation for EA Algorithm.

Although wavelet transform nearly decorrelates images and can be viewed as an approximate KLT, significant dependencies still exist between wavelet coefficients [55]. Especially strong dependencies in the form of spatial clusters exist between wavelet coefficients inside each subband, which is well known as intra-band or intra-scale correlation [10]. Liu et al. [50] uses mutual information as defined in (3.17) to measure the dependencies within subbands:

$$\begin{aligned} I(X;Y) &= \int_x \int_y p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \\ &= E_{XY}\left[log \frac{p(x,y)}{p(x)p(y)} \right]. \end{aligned} \tag{3.17}$$

Let $I(X; N_X)$ denote the mutual information between X and its immediate neighbors N_X, thus N_X can provide information to X through a many-to-one mapping function $T = f(X, N_X)$, in the sense that

$$I(X; N_X) = I(X; T). \tag{3.18}$$

However, $I(X; N_X)$ is only the upper bound of intra-scale dependency one can use to estimate X. A function T can only approach $I(X; N_X)$. Hence a well designed function T will reach a tight bound so that $I(X; T) \rightarrow I(X; N_X)$. In other words, the intra-band correlation can be better adopted by a well-designed mapping function. The maximum likelihood (ML) algorithm is a natural way to optimize this problem. The EA algorithm is actually a Least-Square estimation approach, which is equivalent to the special case of ML with equal prior probability under the assumption that the local texture is stationary [46]. Therefore the EA algorithm is expected to hit a tighter bound of $I(X; T)$ compared to simple bilinear interpolation.

We test the images of Lena and Barbara (512×512) in order to evaluate the EA algorithm. The images are decomposed with four levels. The wavelet coefficients are partitioned into 4 descriptions without quantization as in Fig. 3.2. It is assumed that one of the descriptions is completely lost and all the other descriptions are successfully received. We calculate the mutual information as in [17] between the original coefficients and the reconstructed ones using bilinear interpolation (BI) and EA algorithm, respectively. Since the LL band is dominated by low frequency coefficients, the EA method does not show apparent improvement over BI. Hence only the high-pass subbands are considered here. Experimental results are listed in Tables 3.1. and 3.2, where the decomposition level index is in the ascending order from coarser to finer resolutions. Apparently, the EA algorithm achieves a larger mutual information compared to simple linear interpolation.

A better reconstruction is expected when more intra-scale dependencies are employed to estimate the lost coefficients.

Table 3.1. Mutual information for high-pass bands of Lena.

	Level	1	2	3	4
BI	HL	0.1808	0.2652	0.2191	0.0971
	LH	0.2331	0.2738	0.1896	0.0911
	HH	0.0947	0.1807	0.1462	0.0361
EA	HL	0.2272	0.3964	0.3497	0.2200
	LH	0.2854	0.2993	0.2354	0.1436
	HH	0.1444	0.2818	0.1936	0.0382

Table 3.2. Mutual information for high-pass bands of Barbara.

	Level	1	2	3	4
BI	HL	0.1553	0.2076	0.2478	0.3274
	LH	0.0833	0.1773	0.1707	0.2585
	HH	0.0867	0.1280	0.2868	0.2295
EA	HL	0.2702	0.4115	0.5235	0.6523
	LH	0.1024	0.2725	0.3712	0.3376
	HH	0.1000	0.2478	0.5286	0.2708

3.4.3 Fast Implementation for EA Algorithm

The drawback with the EA algorithm is its extremely high computational complexity. The main bottleneck lies in the computation of the covariance matrix $\hat{R}_{XX}$ by (3.14). To facilitate the application of the EA algorithm, we provide a low-complexity approach to efficiently speed up the calculation without performance sacrifice.

Since the local windows are overlapped with each other, partial row elements are shared by the resulting matrix C as defined in (3.15). This information should be considered while executing the EA algorithm instead of repeating the burdensome calculation.

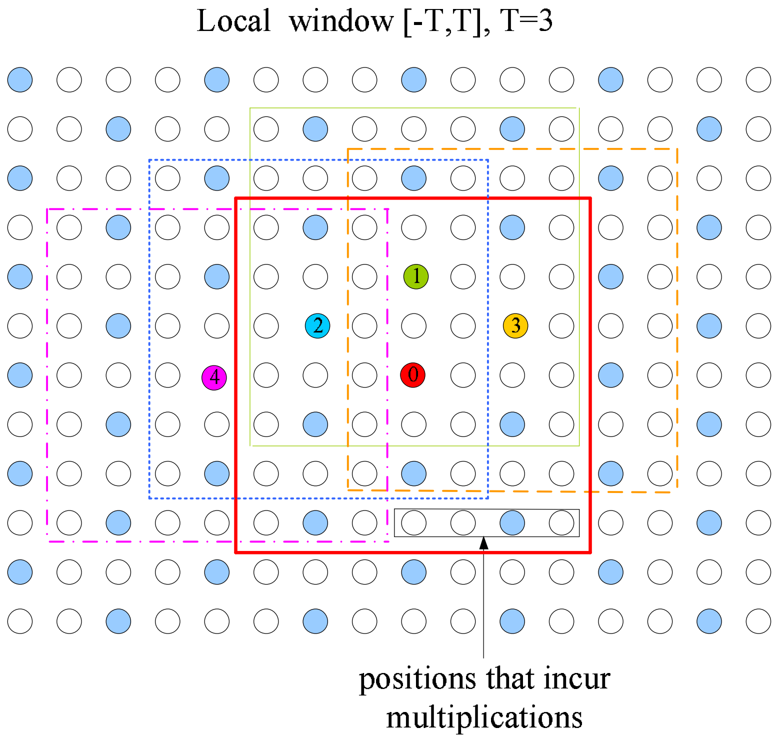


Figure 3.8. Overlapped local windows for a 4-description case. Points 0 through 4 represent the current, top, left-top, right-top and left corrupted coefficients, respectively.

Fig. 3.8 illustrates the overlapped local windows for a 4-description case, where we use indices 0 through 4 to denote the current, top, left-top, right-top and left corrupted coefficients, respectively. The local windows of coefficients 1 to 4 share elements with the local window of coefficient 0, thus only the 4 points at the right-bottom corner will introduce unknown information for estimating coefficient 0. Each element inside a local window is related to one row of C. We rewrite (3.15) as $C = [C_0 ... C_m ... C_L]^T$, where C_m

denotes the *mth* row of C. Thus $R_{XX} = C^TC = \sum_m C_m^T C_m$. If the matrices, $C_m^T C_m$, related to the elements inside the overlapped area are pre-stored, we can spare a lot of multiplications. As can be seen from Fig. 3.8, only few elements at the right-bottom corner are involved in the multiplications.

Let L_{X_0} denotes the local windows for the corrupted coefficient X_0 to be reconstructed currently, $R_{X_0} = C_0^T C_0$ be the covariance matrix for estimating X_0, and L_{X_j}, $j = 1, 2, ..., M$ denote the local windows used for neighboring corrupted coefficients that have been reconstructed previously, where M is the number of local windows that overlap with L_{X_0}. $C_{i,j}$, $i = 1, 2, ..., L$ denote the *ith* row of C for L_{X_j}. The idea of reducing complexity of the EA algorithm is to retrieve the matrices $R_{i,j} = C_{i,j}^T C_{i,j}$ for L_{X_j} that have been previously calculated and stored in the memory, so that R_{X_0} can be obtained by

$$R_{X_0} = \sum_j \sum_i R_{i,j} + \sum_m C_{0,m}^T C_{0,m}, \tag{3.19}$$

where $C_{0,m}$ denotes the rows of C_0 not available from the memory, which introduces multiplications in computing R_{X_0}. The fast implementation is summarized as follows. A vector $\vec{M}_R$ is used to represent the validity of each row in C_0.

Fast implementation for computing R_{X_0}.

- *Set $R_{X_0} = [0]_{N\times N}$, $\vec{M}_R = \vec{0}$.*
- *For j=1 to M do*
  - *Find the overlapping area between L_{X_0} and L_{X_j}. Each element in the overlapping area is mapped to a corresponding entry m of $\vec{M}_R$.*
  - *If $\vec{M}_R(m) = 0$, set $\vec{M}_R(m) = 1$, retrieve the $R_{i,j}$ related to that element, $R_{X_0} = R_{X_0} + R_{i,j}$.*

- *For m=1 to M do*

  If $\vec{M}_R(m) = 0$, $R_{X_0} = R_{X_0} + C_{0,m}^T C_{0,m}$.

3.4.4 Simulation Results

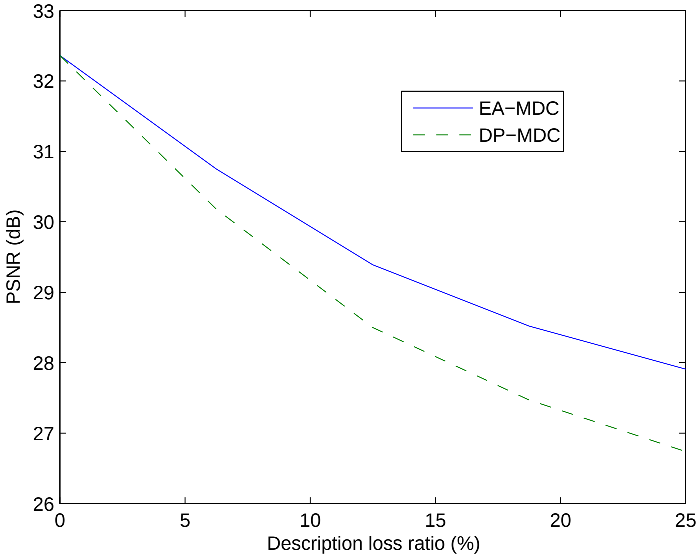
We compare our algorithm against the DP-MDC [5] for robust image transmission. The MD image coder is based on the wavelet codec from [18] for both methods. A set of embedded quantizers is designed for each subband and coefficients are quantized in a layered manner, as described in [89]. The wavelet coefficients are partitioned into the descriptions as described in DP-MDC for each subband. Each description is independently quantized and encoded using a context-based adaptive arithmetic coder. Four level decompositions using 9/7 filters are applied to the images. For both methods, the LL band is duplicated for each channel. For clarity, we represent the proposed coding scheme as EA-MDC.

In DP-MDC, missing coefficients are bilinearly interpolated from available immediate neighbors for the high-pass subbands, whereas the EA algorithm is used in our MDC scheme.

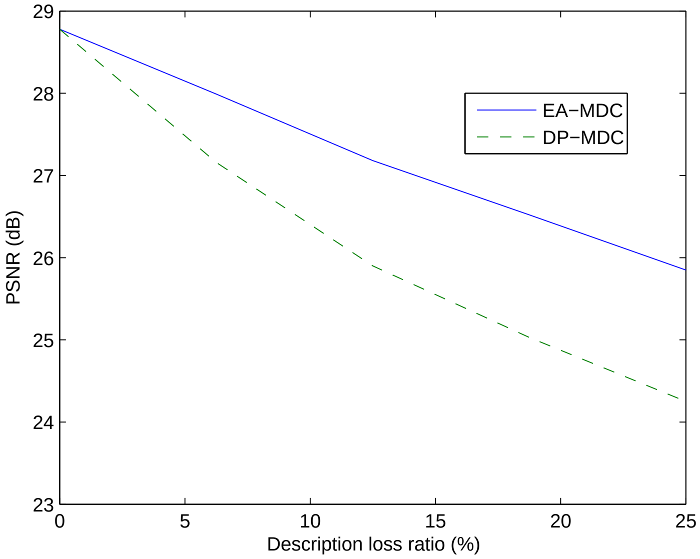
Table 3.3. Performance comparison in PSNR (dB) for Lena.

loss ratio (%)	0	6.25	12.5	18.75	25
DP-MDC	32.36	30.18	28.50	27.47	26.74
EA-MDC	32.36	30.75	29.39	28.52	27.91

We carried out a set of experiments on the 512×512 Lena and Barbara images. In accordance with [5], the bitrates of the conventional single description coding are fixed at 0.21 bpp and 0.4 bpp for Lena and Barbara, respectively. We encode the images into 16



(a)



(b)

Figure 3.9. Average PSNRs of different MDC algorithms for : (a) Lena; (b) Barbara.

(a) (b)

Figure 3.10. Sample image reconstruction for Lena with descriptions loss ratio of 25% (Description 0 is assumed to be lost): (a) image reconstructed by DP-MDC, PSNR = 26.81 dB; (b) image reconstructed by EA-MDC, PSNR = 28.00 dB.

descriptions. In each experiment we fix the number of lost descriptions and completely remove a set of descriptions according to a given description loss ratio.

Table 3.3 includes the average PSNR comparison results of Lena between DP-MDC and our algorithm. The PSNR performance improvement over DP-MDC ranges from 0.6 to 1.2 dB under different description loss ratios. Similar experimental result for Barbara is compared in Table 3.4. Because of the rich textures in Barbara, higher improvement, ranging from 0.8 to 1.6 dB is achieved by using our EA algorithm. Average PSNR results are also plotted in Fig. 3.9 which clearly shows remarkable gains of EA-MDC over DP-MDC, especially for images with a lot of high frequency components.

It appears that the visual quality is also improved. Figs. 3.10 and 3.11 respectively compare the Lena and Barbara images reconstructed by both methods with 25% descrip-

(a) (b)

Figure 3.11. Sample image reconstruction for Barbara with descriptions loss ratio of 25% (Description 2 is assumed to be lost): (a) image reconstructed by DP-MDC, PSNR = 24.30 dB; (b) image reconstructed by EA-MDC, PSNR = 25.97 dB.

(a) (b)

Figure 3.12. Comparison of the zoomed portion in Fig. 3.10.

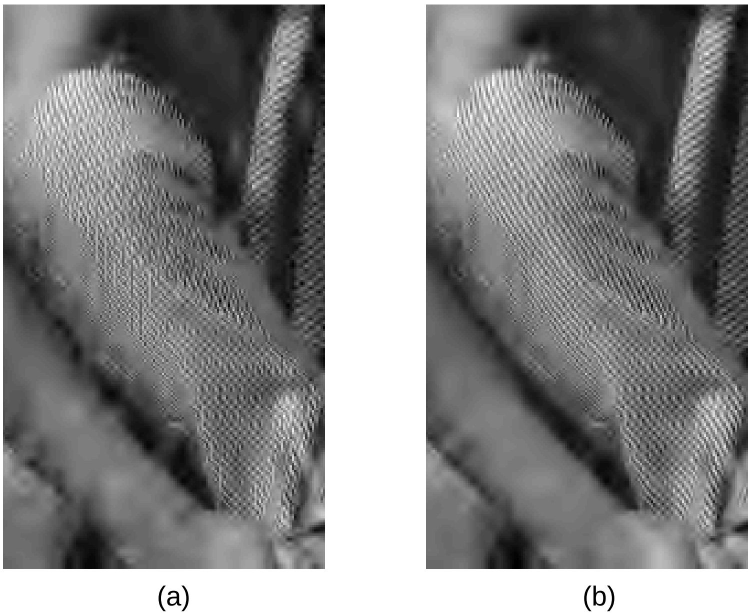


Figure 3.13. Comparison of the zoomed portion in Fig. 3.11.

Table 3.4. Performance comparison in PSNR (dB) for Barbara.

loss ratio (%)	0	6.25	12.5	18.75	25
DP-MDC	28.78	27.16	25.90	25.03	24.25
EA-MDC	28.78	27.99	27.18	26.52	25.85

tion lost. Noticeable improvement can be found around the edges as shown in Figs. 3.12 and 3.13, for instance, the hair, the eyes and the rim of the hat in Lena, and the table cloth, the right arm and the trousers texture in Barbara.

Since the human visual system is highly adapted to extract structural information from the viewing field, it follows that a measure of structural information change can provide a good approximation to perceived image distortion. Hence, a new measure, structural similarity (SSIM) index [102], is used to quantify the visibility of distortions for the reconstructed images. Fig. 3.14 compares the SSIM index maps of the reconstructed

Barbara images in Fig. 3.11. The higher the SSIM index is, the better reconstruction can be expected [102]. It is evident that the proposed algorithm achieves higher SSIM index values in the regions containing large quantities of high-frequency components compared to DP-MDC. This further verifies that our algorithm is suitable for recovering the geometry structural information within the corrupted images.

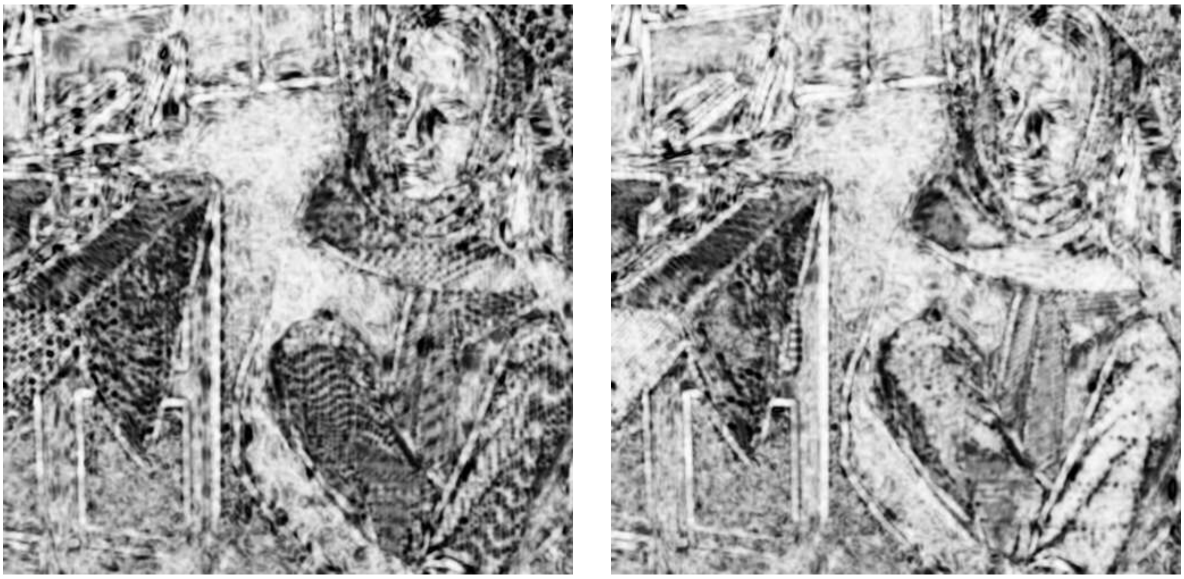

(a) (b)

Figure 3.14. SSIM index maps comparison for the reconstructed images in Fig. 3.11: (a) DP-MDC, (b) EA-MDC.

3.5 Summary

In this chapter, two algorithms are presented to improve the state-of-the-art domain-based MD coder. Based on the wavelet HMT model, the cross-band dependency is exploited when estimating the lost descriptions. By adjusting the bilinear interpolated corrupted coefficients with the aid of the HMT model, the reconstructed image quality can be improved. Instead of depending on the HMT model, another scheme, named as

EA-MDC, is proposed. It successfully retrieves the correlations along all the orientations between the corrupted coefficient and its available neighbors inside a local window by using the covariance method. Hence the weights of linear estimation are well manipulated to emphasize the directions along the important high-frequency components. With the ease of adapting estimation to the dominant edge orientations, EA-MDC can effectively reconstruct the lost descriptions by taking full advantages of the intra-scale dependency within the wavelet subbands. The simulation results show that improvement can be obtained in terms of average PSNR of up to 1.6 dB. Furthermore, the subjective quality of EA-MDC is superior to that of DP-MDC and the HMT model based method since large quantities of high-frequency components are well reconstructed.

CHAPTER 4

FEATURE-ORIENTED MULTIPLE DESCRIPTION WAVELET BASED IMAGE CODING

4.1 Introduction

Recently, research interests for error-tolerant coders are concentrated on wavelet-based MD coding. A MDSQ based wavelet algorithm is presented in [78] by optimally selecting the number of diagonals and the quantization steps of the MD scalar quantizer. Kim [42, 43] proposed to assign different UEP codes to each bit-plane of the SPIHT-coded MD bitstreams [76]. Franchi et. al [57] extended the polyphase down-sampling technique to create more flexibility in MD video coding. Different strategies are used to introduce an amount of redundancy among the descriptions in such approaches. However, all the strategies are designed for the sense of general data source without the ability to smartly adjust the inserted redundancy for the benefit of specified image source, i.e. all the candidate redundant data are equally weighted to be introduced with a predetermined strategy. This causes inefficient redundancy allocation amongst the variantly distributed data of a specified image. When certain packets are lost or a channel transmission fails, only partial redundancy can significantly contribute to recover the corrupted contents.

In this chapter, we propose a novel wavelet-based MD coder, named as feature-oriented MDC (FO-MDC). Instead of designing a complicated quantizer to code the source data or adding weighted channel protection for descriptions, we fully exploit the potential of the redundancy that can contribute to the lost packets reconstruction. The wavelet coefficients of each subband are partitioned into two subsets by maximizing the minimum intra-partition distance [6], so that more correlations between partitions can be used for data estimation. The regions (of the wavelet coefficients) sensitive to packet loss are identified with the aid of estimation using an error concealment technique. These identified regions hold the coefficients that really require to be well protected. Hence, we manipulate the rate budget to allocate as many bits as possible to these coefficients. This is realized by joint optimization of tree-structured pruning and quantizer selecting. Consequently, the reconstruction of corrupted packets can greatly benefit from the inserted redundancy at a given target rate.

The rest of this chapter is organized as follows. Section 4.2 describes the motivation and formulates the problem. Section 4.3 discusses the FO-MDC system and the proposed coding algorithms. Numerical results of the proposed MD coder are presented in Section 4.4. Concluding remarks are given in Section 4.5.

4.2 Problem Formulation

4.2.1 Motivation

MD coding is a technique that can tolerate data loss by benefiting from the estimation with an amount of introduced redundancy. A simple MD coder based on data partitioning, such as polyphase based MDC [37], is shown in Fig. 4.1. The source data x, either in spatial or transform domain, are partitioned into two subsets, y_1 and y_2. The data set y_i is coded at a higher rate for channel i as the *primary* description and is also coded at a lower rate for the other channel as the *redundant* description. Note that in this framework, the primary descriptions are generated by data partitioning, i.e. no overlapping exists between the primary descriptions of different channels. Hence the primary description of one channel can also be considered as the candidate for the other channel.

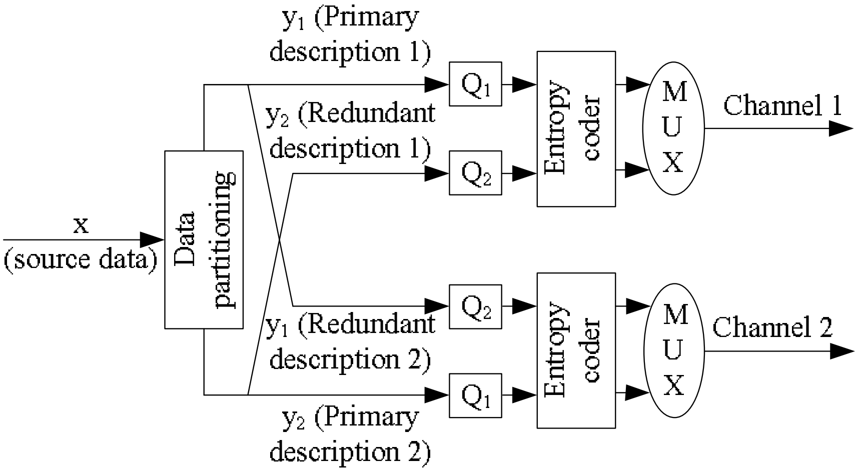


Figure 4.1. A prototype of a simple MD coder based on data partitioning. The input source data x are partitioned into two descriptions, y_1 and y_2, defined as the primary descriptions. They are coded with a set of finer quantizers, Q_1, at a higher rate. At the same time, each of them is taken as the candidate of the redundant part, defined as the redundant description, for the other channel, which is either not coded or coded with a set of coarser quantizers, Q_2 at a lower rate .

Fig. 4.1 illustrates the above MD coder. The selection of coding strategy for the primary description is actually a problem of SDC, i.e. the conventional image coding, which is not the focus of this work. The interests of our MD scheme are twofold. First, in what way the data of the redundant description should be selected to benefit the lost data estimation? Second, how can we efficiently code the set of selected data so as to satisfy the rate constraint for the redundant description?

Wavelet image decomposition provides a hierarchical data structure for representing images. It can be thought of as a tree-structured set of coefficients, with each coefficient corresponding to a spatial region in the image. A wavelet spatial orientation tree is well known as the set of coefficients from different bands that represent the same spatial region. If the spatial region associated with any node of the tree has minor energy, removing the nodes related to the region from the set of data to be coded will not significantly affect the reconstruction performance. Pruning these nodes actually decrease the coding cost and improve the rate-distortion (RD) performance. On the other hand, strong inherited correlations exist within some spatial regions of the subbands, especially the ones corresponding to slow varying textures of the images. As a result, it is much convenient for these regions to be estimated if partial coefficients of them are not available at the receiver.

Therefore, only those coefficients, belonging to the regions with rich edges, that are much difficult to be estimated, may significantly contribute to the distortion when they are not available at the decoder. These coefficients are actually the candidates to be considered for coding in the redundant description. In summary, the underlying theme of our feature-oriented MD coder is a combination of estimation (error concealment) and a pruned tree-structured zerotree quantization to exploit the spatial characterization.

4.2.2 Primary Description Generation

The efficiency of estimating redundant descriptions depends on how the primary descriptions are generated. As described in Fig. 4.1, the primary description of a wavelet-based MD coder results from the data partitioning of the transform coefficients. Thus no extra information other than the inherited correlation between partitions can be used to reconstruct the lost data. The amount of inherited correlation that may be exploited depends on the way of data partitioning. The more correlation there is, the better reconstruction performance can be expected. Hence, it is important to design an efficient partitioning approach for generating primary descriptions.

As discussed in the previous chapter, successful estimation of missing coefficients relies on the autocorrelation of the signal. The autocorrelation is typically a decreasing function of distance. Better results can be expected if the missing coefficient is estimated from its immediate neighbors. The distortion is also a non-increasing function of the size of the available neighbors surrounding a missing coefficient for typical estimation algorithms. Hence we adopt the optimal partitioning algorithm proposed in [5] to generate the non-overlapped partitions for each subband. Each partition is then formed as a primary description then.

Interestingly, the optimal partitioning for the case of two descriptions boils down into a well-known checkboard pattern that is employed by our MD coder in the remaining sections of this chapter.

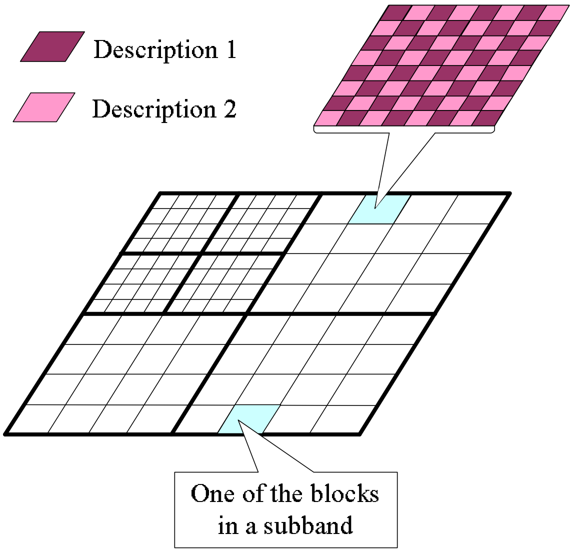


Figure 4.2. Primary description generation with block access for the case of two-channel MDC. Each subband is divided into square blocks. The block size can be variant for different subbands. Every block is then partitioned into two subsets, each subset is associated with one of the two primary descriptions.

In the FO-MDC algorithm described in Section 4.3, the rate-distortion optimization is executed based on the unit of a block instead of a pixel since the pixel-wise optimization may cause much higher computation loads and more overhead bits. Hence, each subband is divided into a group of blocks, the block size may vary for different subbands. Then each block is further partitioned into two subsets as in Fig. 4.2 with each subset belonging to one of the primary descriptions. In this way the wavelet coefficients are separated into two primary descriptions in the manner of block access.

4.2.3 Data Estimation

For one channel, if only the quantized primary description is coded and transmitted, i.e. no redundancy is introduced, all the coefficients of the redundant description are assumed to be lost and need to be estimated at the receiver. Estimating lost coefficients from the primary description is naturally related to error concealment techniques. The technique of edge-adaptive error concealment (EA-EC) proposed in the previous chapter can be conveniently adopted to estimate the redundant description. This algorithm can efficiently capture the geometric constraints of edges hold in the wavelet domain. On the other hand, the simple bilinear interpolation is employed in the low-low (LL) band for its simplicity since the LL band is dominated by coefficients with low activities. Therefore, the redundant description can be well estimated with the primary description based on the intra-band dependency.

4.2.4 Tree Formulation of the Redundant Description and Problem Statement

In this section, we describe how to construct a tree structure of the redundant description for one of the two channels. Since the two channels are assumed to be equally characterized, we define the notations for channel 1 in the rest part of this chapter, while these notations are also applicable to the other channel.

Let S^p and S^r denote the primary description and the redundant description of channel 1, respectively. $\hat{S}^r$ denotes the quantized S^r. At a given target rate for the redundancy, the FO-MDC tries to insert only a subset of $\hat{S}^r$ into channel 1. By executing EA-EC technique described above with the quantized primary description, $\hat{S}^p$, an estimated version of the redundant description, $\tilde{S}^r$, can be obtained. Based on $\tilde{S}^r$, it is convenient to recognize those coefficients of S^r that are not sensitive to packet loss by pruning a wavelet coefficient tree constructed from $\hat{S}^r$ in the rate-distortion sense. It is natural to throw away these identified coefficients from $\hat{S}^r$ and reconstruct them according to $\tilde{S}^r$ at the receiver.

Let T be the set of all trees grown from the nodes belonging to S^r in the LL band. Let P be the set of all subtrees extracted from T. Consider a node $n \in S^r$. Let T_n be the tree grown from n to its full depth. Let P_n be any subtree of T_n, and O_n be the set of the direct descendants of n. For each wavelet coefficient w_j of node n, there are two candidates, the quantized $\hat{w}_j$ and the estimated $\tilde{w}_j$ which can be used to recover w_j. Define their corresponding distortions as D_n^q and D_n^e, respectively. In the framework of FO-MDC, we seek to minimize the distortion subject to a rate constraint. Let $D^r(q, P)$ and $R^r(q, P)$, respectively, be the distortion and rate for S^r with the quantization step $q \in Q^r$ and the choice of P. Hence the problem can be formulated for the redundant description as

$$\min_{q \in Q^r, P \subseteq T} D^r(q, P) \quad \text{subject to} \quad R^r(q, P) \leq R_T^r, \tag{4.1}$$

where R_T^r is the target rate for S^r.

In words, our goal is to find an optimal combination between the subset of coefficients (nodes) in S^r that are sensitive to packet loss and the optimal uniform scalar quantizer, by which these sensitive coefficients are quantized before the entropy coding. The insensitive coefficients are excluded from the bit stream and are estimated at the receiver.

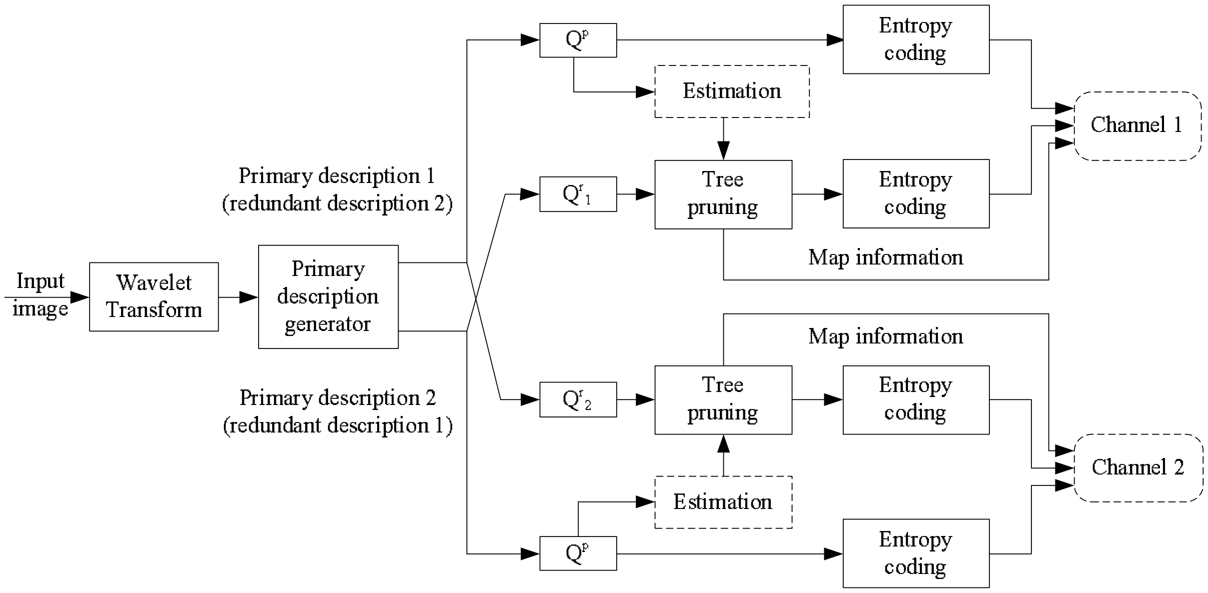


Figure 4.3. Diagram of FO-MDC encoder.

4.3 Feature-Oriented Multiple Description Coding

4.3.1 System Outline

In the FO-MDC framework described in Fig. 4.3, the wavelet coefficients are partitioned into two primary descriptions. At the same time, each of these primary descriptions also serves as the redundant description for the other channel. Let $R_T = 2(R_T^p + R_T^r)$, be the target rate for both channels, where R_T^p and R_T^r ($R_T^r \leq R_T^p$) denote the target rates for the primary description S^p and the redundant description S^r per channel. Let Q^p be the quantizer for both primary descriptions S^p's. Q_i^r represents the quantizer for the redundant description S^r of the ith channel, i =1, 2. The rest of the development in this section is for channel 1 while the other channel can be done in the same manner. Assume R^p and R^r denote the actual rate of S^p and S^r achieved by the MD coder. The primary description is independently coded by varying the quantization stepsize to minimize the distortion subject to the constraint $R^p \leq R_T^p$, which is equivalent to a SDC optimization problem [55, 69]. The EA-EC technique is applied to the quantized $\hat{S}^p$ so as to estimate S^r. The estimated $\tilde{S}^r$ is used to find an optimal combination of quantization step, $q \in Q_1^r$, and a pruned spatial tree of S^r. The resulting map for tree-pruning is sent as side information.

The constrained optimization problem of (4.1) can be converted to an unconstrained formulation using the well-known Lagrange multiplier method. It can be shown [80] that the solution to (4.1) is identical to that of the equivalent unconstrained problem with $R^r(q,P) = R_T^r$:

$$\min_{q \in Q_1^r, P \subseteq T} \left\{ J^r(q,P) = D^r(q,P) + \lambda R^r(q,P) \right\}, \tag{4.2}$$

where $J^r(q,P)$ is the Lagrangian cost including both rate and distortion. The solution of (4.2) finds an appropriate operating point that resides on the convex hull of the rate-distortion function through sweeping λ from 0 to ∞ [69]. We approximate the solution of (4.2) by solving

$$\min_{q \in Q^r} \left\{ \min_{P \subseteq T} \left[D^r(q,P) + \lambda R^r(q,P) \right] \right\}. \tag{4.3}$$

For fixed values of λ and q, minimizing (4.3) involves optimal tree-pruning to find the best P. It can be written as

$$\min_{P \subseteq T} \left\{ D^r(q,P) + \lambda [R_{coef}^r(q,P) + R_{side}^r(q,P)] \right\}, \tag{4.4}$$

where $R_{coef}^r(q,P)$ and $R_{side}^r(q,P)$ are the rates for encoding survival coefficients and the map information of tree-pruning, respectively. Since $R_{coef}^r(q,P)$ and $R_{side}^r(q,P)$ are correlated, we solve the problem by a two-stage algorithm. In the first stage, $R_{side}^r(q,P)$ is assumed to be 0. Thus the optimal P is found through the tree-pruning algorithm (see Section 4.3.2), i.e.

$$P_1^* = arg \min_{P \subseteq T} \left\{ D^r(q,P) + \lambda R_{coef}^r(q,P) \right\}. \tag{4.5}$$

The side information is coded with the resulting P_1^* in the second stage. It may be further adjusted, if necessary, to obtain the final appropriate pruned tree P_2^* that reflects the global constraint, $R^r_{coef}(q,P) + R^r_{side}(q,P) \leq R^r_T$. This is described in Section 4.3.3. Consequently, we can achieve a local optimal point on the convex rate-distortion curve for each choice of (q,λ). A fast iterative algorithm is developed in Section 4.3.4 to search for the best stepsize $q \in Q^r_1$ and the associated slope λ.

4.3.2 Tree-pruning for fixed q and λ

With q and λ fixed, the following *Algorithm 1* is used to find the best $P \subseteq T$. The spatial trees of T formulated from the quantized redundant description $\hat{S}^r$ are scanned from the finest to the coarsest resolution. Each node n can be labelled as a *zerotree root*, an *isolated zerotree root*, or a *valued node*. A zerotree root is a node that the whole subtree T_n is carved out. All the coefficients of T_n including the node n are reconstructed from the estimated redundant description $\tilde{S}^r$. An isolated zerotree root is a node where the node itself is carved out while at least one of its descendants remains. A valued node identifies a node with nonzero quantized coefficients.

Let k be the iteration count. At each node n, we assign a value to the node as:

$$m_n^{(k)} = \begin{cases} 0 & \text{if } n \text{ is a zerotree root,} \\ 1 & \text{if } n \text{ is an isolated zerotree root,} \\ 2 & \text{if } n \text{ is a valued node,} \end{cases}$$

where $m_n^{(k)}$ refers to the zerotree map of node n at the kth iteration of the algorithm. Let $P^{(k)}$ denote the best subtree obtained after k iterations with $P^{(0)} = T$. D^q_n is the distortion from quantizing node n. At each iteration, the probability model of T is updated with the survived nodes using

$$p_d^{(k)} = n_d / N^P,$$

where N^P refers to the number of surviving coefficients in $\hat{S}^r$. Let d be any quantization bin resulting from quantizing S^r using q, and n_d be the number of coefficients that are quantized to the interval specified by d. Thus $p_d^{(k)}$ is the probability estimate at the kth iteration for bin d. With the updated model, the bitrate of each node is approximated by the theoretical first-order entropy, which can be approached very closely by applying adaptive arithmetic coding. The resulting bitrate is denoted as $R_n^{(k)}$.

When the node n is identified as an isolated zerotree node, a cost of the estimation error D_n^e will be incurred by removing node n. Similarly, it results in a cost of $D^e(T_n)$ with the subtree T_n being pruned if node n is determined as a zerotree node, where $D^e(T_n) = \sum_{i \in T_n} D_i^e$. Let J_n^* and $J^*(T_n)$ be the minimum cost associated with node n and T_n, respectively. Note that $J^*(T_n) = J_n^*$ for the finest resolution. The pruning algorithm is executed in a bottom-up way from the finest resolution ($L = N - 1$) to the coarsest resolution ($L = 0$), where N denotes the level of decomposition and L is the level index.

Algorithm 1

1) Set $k \leftarrow 0$, $L \leftarrow N - 1$ and $P^{(0)} \leftarrow T$.

Set all $m_n^{(0)} \leftarrow 2$, i.e. all the nodes are initialized as valued nodes.

2) Update probability model based on the quantization indices associated with $P^{(k-1)}$ if $k \geq 1$. Estimate $R_n^{(k)}$ for each survived node n with the first-order entropy.

3) For each node n at level L, determine whether to carve out either the whole subtree T_n or the node n only. Note that each node is composed of half coefficients of a square block as described in Section 4.2.4.

- (*Valued node*): If $D_n^e \leq D_n^q + \lambda R_n^{(k)}$, then $J_n^* \leftarrow D_n^e$, $m_n^{(k)} \leftarrow 1$; else $J_n^* \leftarrow [D_n^q + \lambda R_n^{(k)}]$,

  where

$$D_n^e = \sum_{j \in \text{node}\, n} (w_j - \tilde{w}_j)^2, \quad D_n^q = \sum_{j \in \text{node}\, n} (w_j - \hat{w}_j)^2.$$

- (*Isolated zerotree root*): When $m_n^{(k)} = 1$ (node n has been recognized as an isolated zerotree root either in the previous or current iteration), we have to determine whether it can be further refined as a zerotree root.

  If

  $$\sum_{i \in O_n} D^e(T_i) \leq \sum_{i \in O_n} J^*(T_i), \tag{4.6}$$

  then

  $$m_n^{(k)} \leftarrow 0, \quad J^*(T_n) \leftarrow D^e(T_n),$$

  else

  $$J^*(T_n) \leftarrow \left[D_n^e + \sum_{i \in O_n} J^*(T_i) \right]. \tag{4.7}$$

- (*Zerotree root*):

  If

  $$D^e(T_n) \leq \left[D_n^q + \lambda R_n^{(k)} \right] + \sum_{i \in O_n} J^*(T_i), \tag{4.8}$$

  then

  $$m_n^{(k)} \leftarrow 0, \quad J^*(T_n) \leftarrow D^e(T_n),$$

  else

  $$J^*(T_n) \leftarrow \left[D_n^q + \lambda R_n^{(k)} + \sum_{i \in O_n} J^*(T_i) \right]. \tag{4.9}$$

4) Set $L \leftarrow L - 1$, go to *Step 3)* if $L \geq 0$.

5) Carve out the nodes with $m^{(k)} = 0$ or 1. If $P^{(k)} \neq P^{(k-1)}$, i.e. some nodes are pruned, then $k \leftarrow k + 1$, go back to *Step 2)* to update the probability model and iterate again. Otherwise stop the iteration, and set the tree-pruning map as $m_n \leftarrow m_n^{(k)}$, $n \in T$. The resulting $P^{(k)}$ is declared as the local optimal pruned trees P_1^* for the choice of (q, λ).

It should be mentioned that a similar tree-pruning approach, known as SFQ, is proposed in [109]. Our algorithm differs from SFQ on three aspects. First, SFQ tries to throw away a subset of wavelet coefficients that have minor energy. There is no further attempt to consider the correlations existing within different regions with variant distributions, while these correlations are well exploited by *Algorithm 1* with the aid of estimation. As a result, more nodes are expected to be pruned in the rate-distortion sense. Second, the isolated zerotree roots are ignored by SFQ, while they can be processed with our algorithm to further spare quite a lot of bits, especially when the nodes of the low-pass bands can be well estimated without coding. Third, our tree structure is constructed based on the block access of the redundant description for each channel; more bits can be spared for coding the map information compared to the pixel-access based tree structure of SFQ.

Algorithm 1 is guaranteed to converge to a local optimal solution. Similar to [109], we can easily prove it as follows. Assume $J\{P^{(k)}\}$ denotes the cost of $P^{(k)}$. Let $J_n^{q,(k)} = D_n^q + \lambda R_n^{(k)}$ be the coding cost of node n. Since $P^{(k+1)} \subseteq P^{(k)}$, $J\{P^{(k)}\}$ can be expressed as

$$\begin{aligned} J\{P^{(k)}\} &= \sum_{n \in P^{(k)}} J_n^{q,(k)} + \sum_{i \in [T-P^{(k)}]} D_i^e \\ &= \sum_{n \in P^{(k+1)}} J_n^{q,(k)} + \sum_{i \in [T-P^{(k)}]} D_i^e \sum_{i \in [P^{(k)}-P^{(k+1)}]} J_i^{q,(k)}, \end{aligned} \tag{4.10}$$

where, $[T-P^{(k)}]$ denotes the subset of nodes that are pruned from $P^{(k)}$, and $[P^{(k)}-P^{(k+1)}]$ refers to the subset of nodes pruned at the $(k+1)th$ iteration. Similarly,

$$\begin{aligned} J\{P^{(k+1)}\} &= \sum_{n \in P^{(k+1)}} J_n^{q,(k+1)} + \sum_{i \in [T-P^{(k+1)}]} D_i^e \\ &= \sum_{n \in P^{(k+1)}} J_n^{q,(k+1)} + \sum_{i \in [T-P^{(k)}]} D_i^e \sum_{i \in [P^{(k)}-P^{(k+1)}]} D_i^e. \end{aligned} \tag{4.11}$$

Because of the pruning,

$$\sum_{i\in[P^{(k)}-P^{(k+1)}]} D_i^e \leq \sum_{i\in[P^{(k)}-P^{(k+1)}]} J_i^{q,(k)}.$$

Comparing (4.10) with (4.11), we have

$$\begin{aligned} J\{P^{(k)}\} - J\{P^{(k+1)}\} &\geq \sum_{n\in P^{(k+1)}} \left[J_n^{q,(k)} - J_n^{q,(k+1)}\right] \\ &= \lambda \sum_{n\in P^{(k+1)}} \left[R_n^{(k)} - R_n^{(k+1)}\right]. \end{aligned}$$

The probability of the quantization index associated with node n is always increasing with the iteration of pruning since the number of surviving indices decreases monotonously. Thus $R_n^{(k)} \geq R_n^{(k+1)}$ based on the first-order entropy approximation. This completes the proof by $J\{P^{(k)}\} \geq J\{P^{(k+1)}\}$. According to the experiments, *Algorithm 1* usually converges within 5 iterations.

4.3.3 Side information coding and adjustment

The tree-pruning map is deeply coupled with the choice of (q, λ). The rate for coding map as the side information is considered in the second stage to satisfy the rate constraint. The block access and the zerotree structure significantly reduce the number of bits required to code the map information. Many bits are spared for those nodes if their parents are identified as zerotree roots. We organized the valid m_n (i.e. m_n is not a descendant of a zerotree root) into a vector $\vec{m}$ for each subband. An adaptive arithmetic coding is applied to the resulting $\vec{m}$. In this way, the side information can be coded with a very low bitrate that will not significantly affect the total rate.

It is possible to vary the choice of (q, λ) (discussed in Section 4.3.4) so as to satisfy $|R_{coef}^r(q, P) + R_{side}^r(q, P) - R_T^r| \leq \varepsilon$ at a high R_T^r for some small value ε. In this situation, the final pruned tree, P_2^*, remains the same as P_1^*. However the rate of side information $R_{side}^r(q, P)$ is comparable with $R_{coef}^r(q, P)$ when R_T^r is very small. It is more difficult to meet the rate constraint by only varying (q, λ). Therefore, it is necessary to further prune P_1^* so that the resulting total rate can approach R_T^r. This is realized by *Algorithm 2* that progressively prunes the nodes of P_1^* with high coding costs in a bottom-up way until the rate constraint is satisfied. Let R_{coef}^r and R_{side}^r be the rate of coefficients and side information resulting from P_1^*, respectively.

Algorithm 2

```
1) Set $L \leftarrow N - 1$ and $P_2^* \leftarrow P_1^*$.
2) Update probability model. Estimate bitrate for each surviving node by the first-order entropy.
3) Let $J_n^* = D_n + \lambda R_n$ be the minimum cost of node $n$ obtained from Algorithm 1, where $D_n$ and $R_n$ refer to the corresponding distortion and rate. For each node $n$ with $m_n = 2$ at level $L$, order $\triangle_n = R_n / D_n$ in descending value and put these nodes in a list (assume the length of the list is $M_L$).
4) For $j = 0$ to $M_L - 1$ loop
     • $R_{coef}^r \leftarrow R_{coef}^r - R_j$.
     • If $|R_{coef}^r + R_{side}^r - R_T^r| \leq \varepsilon$, stop the loop and go to Step 6).
5) Update tree-pruning map and recalculate $R_{side}^r$.
   If $|R_{data}^r + R_{side}^r - R_T^r| \leq \varepsilon$, go to Step 6),
   else $L \leftarrow L - 1$. Go back to Step 2) if $L \geq 0$.
6) Carve out the nodes associated with those pruned subtrees from $P_2^*$. Update and encode the tree-pruning map.
```

4.3.4 Determine the best pair of (q, λ)

With each fixed (q, λ), an optimally pruned tree can be obtained as described above. To satisfy the rate constraint, we have to search for the optimal pair (q^*, λ^*) in a finite admissible set located on the RD curve. Let $R = f(q)$ be the function specifying the relationship between q and R for S^r without tree-pruning. Since $R_T^r \leq R_T/2$, the set of admissible q is constrained in the range of (q_{min}, q_{max}), where $q_{min} = f^{-1}(R_T/2)$ and $q_{max} = f^{-1}(R_T^r)$.

It is burdensome and not necessary to generate the set of operating points that includes all the potential candidates. Instead, we construct the RD curve with a few operating points and maintain it in an operational way during the iteration in the algorithm. Initially, a list of operating points, $L_{RD} = \{(R_i^r, D_i^r),\ 0 \leq i \leq M\}$, are computed for S^r by bisectionally splitting the interval $[q_{min}, q_{max}]$ with $R_0^r = R_T^r$ and $R_M^r = R_T/2$, where M is a small integer. The corresponding slopes λ_i can be approximated by [69]

$$\lambda_i \approx \left| \frac{D_i^r - D_{i-1}^r}{R_i^r - R_{i-1}^r} \right| + \delta, \tag{4.12}$$

where δ is an arbitrarily small positive number added to make sure that the smallest rate is picked if λ_i is a singular slope value. A fast search for (q^*, λ^*) is described in the following algorithm by iteratively shrinking the interval $[q_l, q_h]$ that contains the optimal stepsize q^*. The RD curve, i.e. L_{RD}, is updated after each iteration. The points that are not located on the convex hull of the RD function are removed from L_{RD}. In this way, heavy computation loads are avoided by tolerating an approximation of λ.

Algorithm 3

1) Set $k \leftarrow 0$, $q^{(0)} \leftarrow q_{max}$, $q_l \leftarrow q_{min}$ and $q_h \leftarrow q_{max}$. Construct L_{RD} with a few operating points as discussed above.

2) Quantize coefficients of S^r with $q^{(k)}$. Update L_{RD} by inserting the new operating point. Compute $\lambda^{(k)}$ with L_{RD} using (4.12).

3) Obtain the optimal subtree P_1^* by pruning the tree of S^r with $(q^{(k)}, \lambda^{(k)})$ according to *Algorithm 1*.

- If $|R^r - R^r_T| \le \varepsilon$, stop the iteration, and go to *Step 5)*.
- If $|q^{(k)} - q^{(k-1)}| \le \varepsilon$ (the potential of varying (q, λ) has been fully exploited; we resort to further tree-pruning), stop the iteration, and go to *Step 4)*.
- If $R^r > R^r_T$, set $q_l \leftarrow q^{(k)}$; else set $q_h \leftarrow q^{(k)}$.

  $k \leftarrow k+1$, $q^{(k)} \leftarrow \frac{1}{2}[q_l + q_h]$, go back to *Step 2)*.

4) Obtain P_2^* by adjusting P_1^* according to *Algorithm 2*.

5) Code the side information and the surviving quantized coefficients according to the tree map.

4.4 Simulation Results

In this section, we compare the proposed FO-MDC with the state-of-the-art MDSQ scheme proposed in [78]. A four-level wavelet decomposition is used with the 10-18 Daubechies wavelet. All the subbands share the same uniform scalar quantizer. The test images are the 512 × 512 gray-scale *Lena* and *Barbara*.

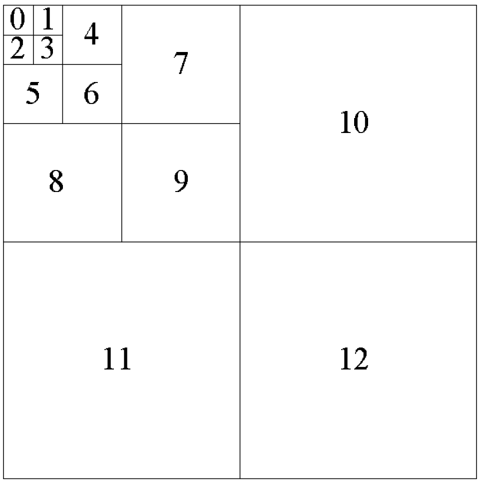
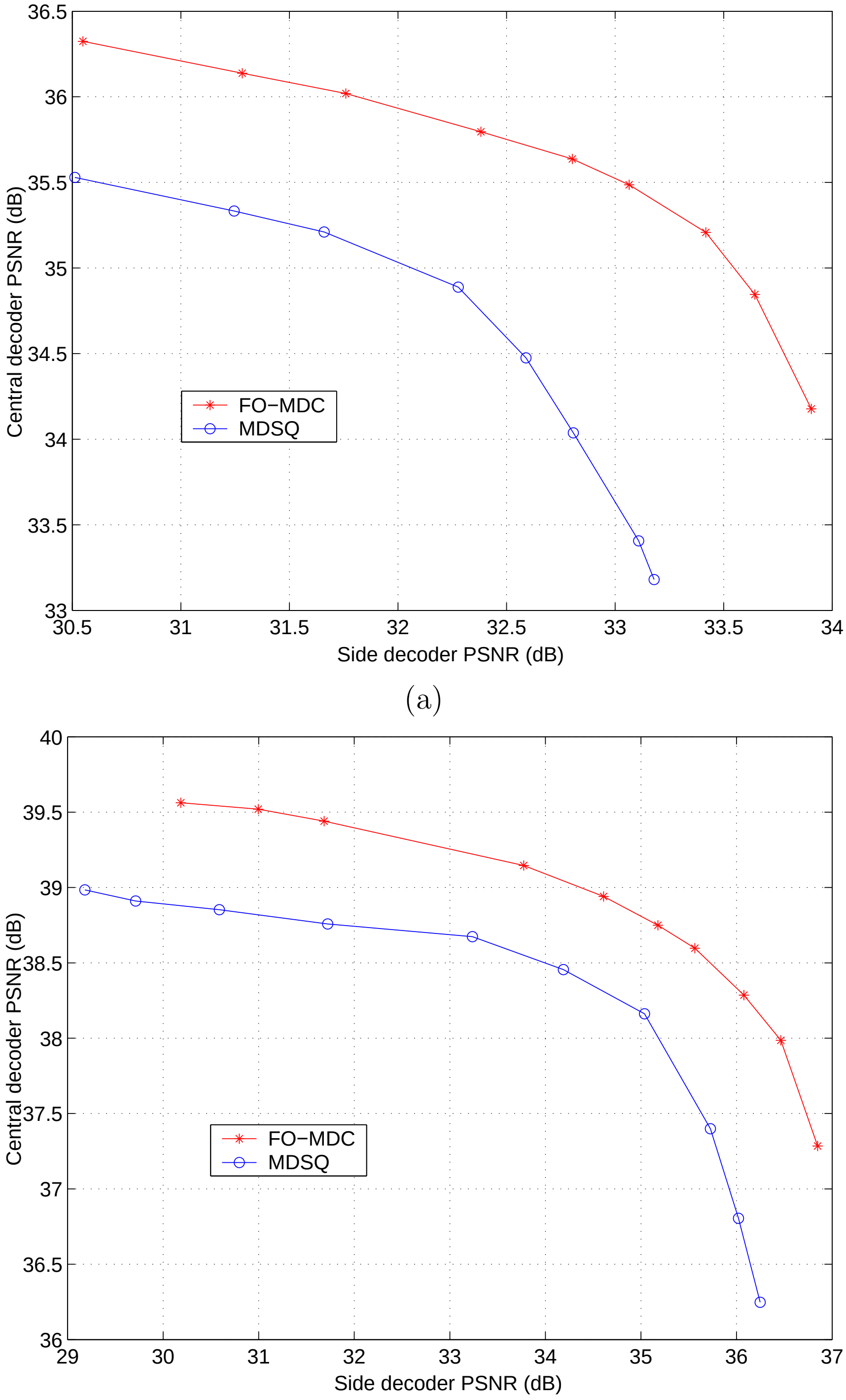


Figure 4.4. Subband indices for a four-level wavelet decomposition.

In our FO-MDC, each subband is partitioned into two primary descriptions. The target rate for both channels, $R_T = 2(R_T^p + R_T^r)$, is allocated to their S^p and S^r. Let R_k^p and R_k^r, $k = 1$ or 2, denote the actual rate of S^p and S^r obtained by a MD coder of channel k. Thus the transmitted data of channel k is composed of the bitstreams with the rates of R_k^p and R_k^r, respectively, where $R_k^p \leq R_T^p$, and $R_k^r \leq R_T^r$. This explicit way of bit budget allocation makes it more clear and easier for redundancy control, whereas, it is much difficult to achieve an accurate redundancy control for MDSQ- and PCT-based MD coding approaches which employ an implicit way to vary the introduced redundancy [90, 98].

We further divide every subband into a group of blocks. The block size depends on the frequency bands. We emphasize the low-frequency bands with smaller block size for their significant contributions to distortion resulting from packet loss. In our four-level decomposition shown in Fig. 4.4, subbands 0-3 have a block size of 2×2, subbands 4-6 have a block size of 4×4, and the rest subbands have a block size of 8×8. Then we apply the error concealment to estimate S^r based on S^p for each channel. It is observed in the experiments that the estimation has minor improvement at high-pass bands. Hence we actually execute the error concealment for subbands 0-6 only and use zero as the estimation for the rest high-pass bands. In this way, the heavy computation loads incurred by error concealment are significantly decreased with minor degradation.

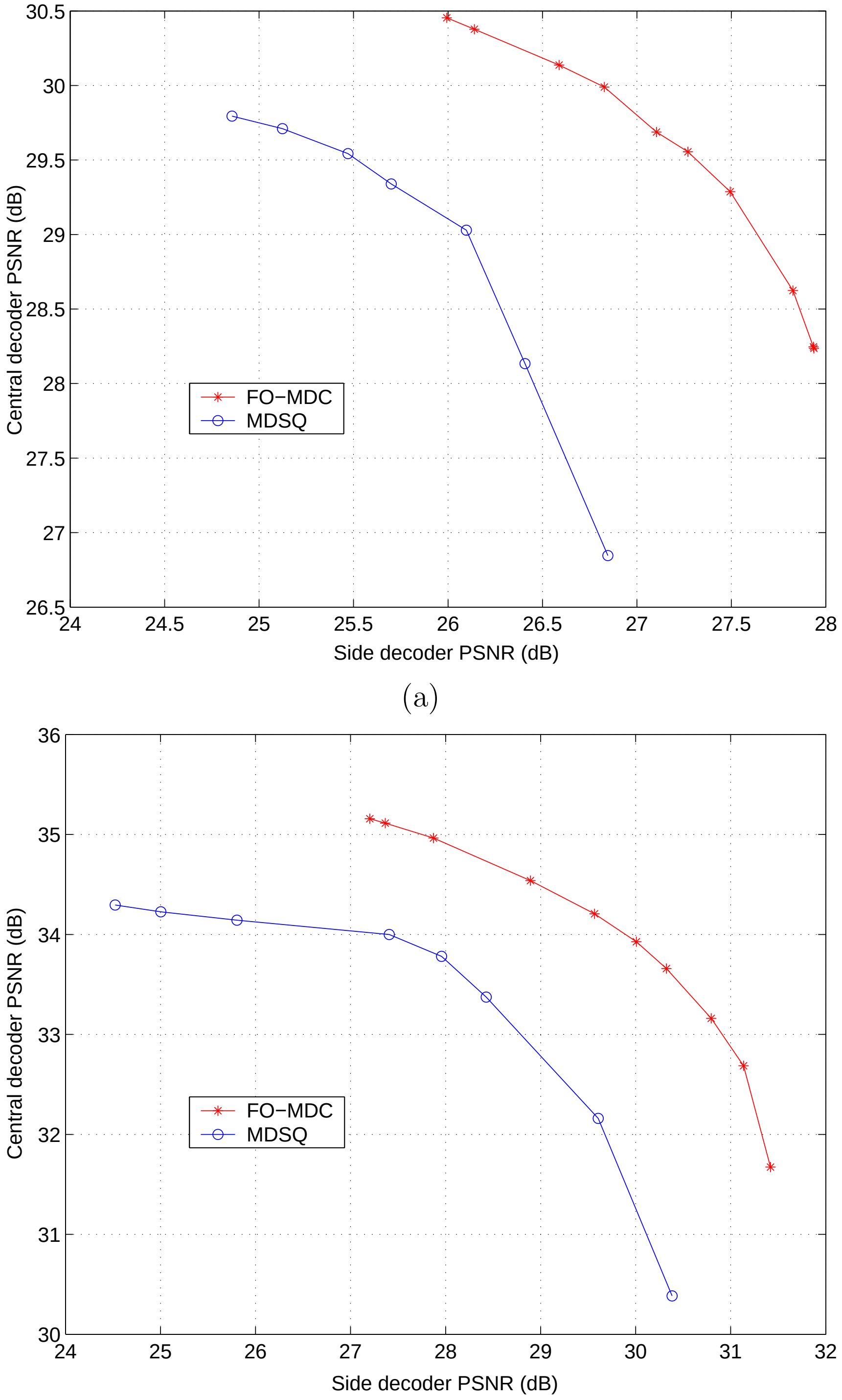
Fig. 4.5 compares the performance attained by our FO-MDC against MDSQ. Two target bitrates, $R_T = 0.5$ and 1.0 bpp, are tested. The RD curves are plotted with two-channel (central) reconstruction PSNR versus the average one-channel (side) reconstruction PSNR. This experiment is to verify the ability to achieve reasonable tradeoff points between the central and side decoders with variant redundancies. For a good MD coder, it is expected to achieve higher PSNR from the central (side) decoder with the side (central) PSNR fixed.



(a)

(b)

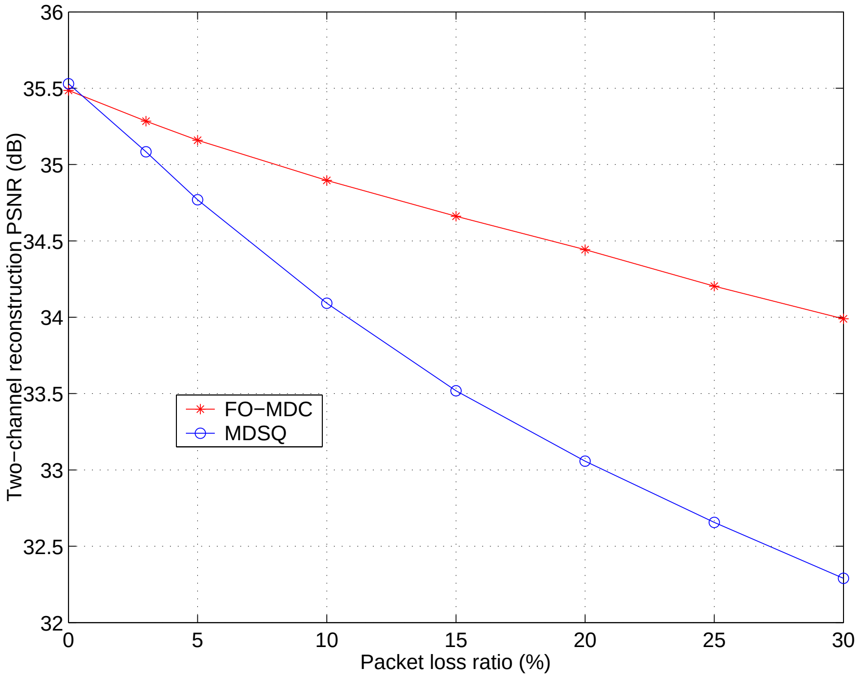
Figure 4.5. Performance comparison of Lena between FO-MDC and MDSQ with the target bit rates: (a) $R_T = 0.5$ bpp, (b) $R_T = 1.0$ bpp.



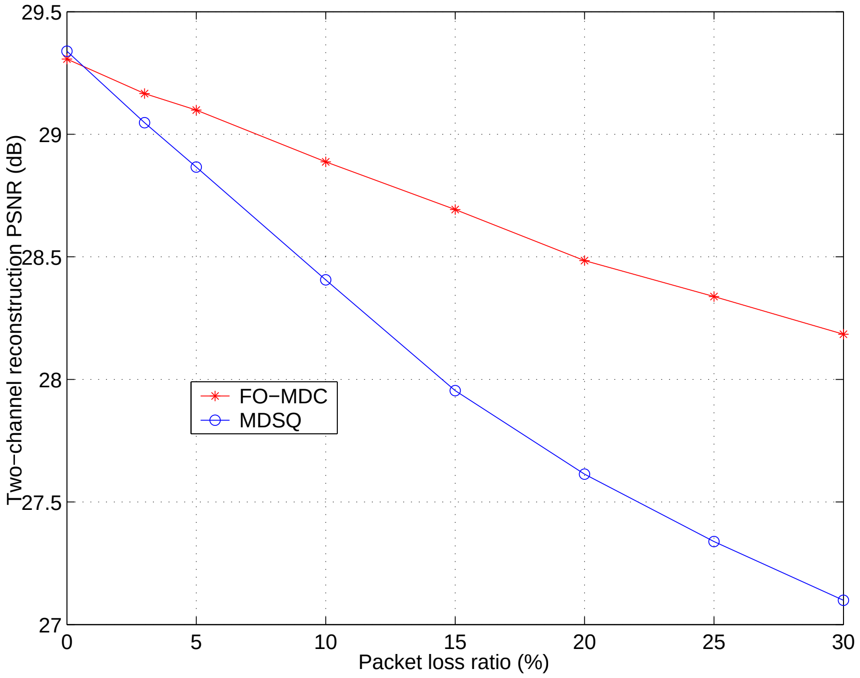
(a)

(b)

Figure 4.6. Performance comparison of Barbara between FO-MDC and MDSQ with the target bit rates: (a) $R_T = 0.5$ bpp, (b) $R_T = 1.0$ bpp.



(a)



(b)

Figure 4.7. Experimental results with variant packet loss ratios. We explicitly set $R_T = 0.5$ bpp, $R_T^p = 0.2$ bpp, $R_T^r = 0.05$ bpp for FO-MDC. Correspondingly we choose the MDSQ coder with the parameters that can achieve the same central PSNR at the same target bitrate of 0.5 bpp. (a) Two-channel reconstruction performance comparison of Lena, (b) Two-channel reconstruction performance comparison of Barbara.

Evidently, FO-MDC achieves consistent improvement over MDSQ at all testing points. With the same side reconstruction distortion, the performance gap of the central decoder between these two schemes is significant, ranging from 0.6 - 2 dB. It should be noted that the improvement of FO-MDC is remarkable even through a very simple quantizer is used for all the subbands. As reported in [37, 16], even larger gains can be expected if variant quantizers or superior bit-plane coding techniques are employed, which is to be considered in our future work.

It is observed that FO-MDC is more robust at low bit rates. Since it allocates the available bits to the coefficients sensitive to data corruption as many as possible, whereas, MDSQ equally distributes the available bits among all the coefficients. We also notice that FO-MDC can easily achieves more tradeoff points than MDSQ because of the flexible rate-distortion optimizer described in Section 4.3. Consequently, the resulting RD curve is more smooth compared to that of MDSQ which is obtained by varying the number of diagonals in the index assignment matrix [78].

Fig. 4.5 compares the performance attained by our FO-MDC against MDSQ. Two target bitrates, $R_T = 0.5$ and 1.0 bpp, are tested. The RD curves are plotted with two-channel (central) reconstruction PSNR versus the average one-channel (side) reconstruction PSNR. This experiment is to verify the ability to achieve reasonable tradeoff points between the central and side decoders with variant redundancies. For a good MD coder, it is expected to achieve higher PSNR from the central (side) decoder with the side (central) PSNR fixed.

Evidently, FO-MDC achieves consistent improvement over MDSQ at all testing points. With the same side reconstruction distortion, the performance gap of the central decoder between these two schemes is significant, ranging from 0.6 to 2 dB. It should be noted that the improvement of FO-MDC is remarkable even through a very simple quantizer is used for all the subbands. As reported in [37, 16], even larger gains can be expected if variant quantizers or superior bit-plane coding techniques are employed.

It is observed that FO-MDC is more robust at low bit rates, since it allocates the available bits to the coefficients sensitive to data corruption as many as possible, whereas, MDSQ equally distributes the available bits among all the coefficients. We also notice that FO-MDC can easily achieves more tradeoff points than MDSQ because of the flexible rate-distortion optimizer described in Section 4.3. Consequently, the resulting RD curve is more smooth compared to that of MDSQ which is obtained by varying the number of diagonals in the index assignment matrix [78].

Fig. 4.6 describes the performance for Barbara. An even larger gap can be found in FO-MDC over MDSQ. Since Barbara is an image with rich textures and edge information, our estimation can capture those high frequency features successfully. A lot of coefficients can be well estimated by the EA-EC technique. Thus we spared more bits for better encoding the remained coefficients that are sensitive to packet loss. Therefore, a larger performance gap is expected using our FO-MDC.

In the real world, the probability that a channel is completely blocked during the process of transmission is very low. Hence it is reasonable to assume that a certain percentage of the total packets are corrupted in the simulation. We simply take each block as a packet, i.e. packets are formulated with variable sizes. A uniformly distributed random sequence that contains the indices of the corrupted packets (blocks) are independently generated for each channel. The probability that both channels failed is kept less than 1%. The survived packets from both channels are combined for the two-channel reconstruction. This experiment is tested with the packet loss ratios of 3%, 5%, 10%, 15%, 20% and 30%. For each specified packet loss ratio, we generate 20 corrupted packet sequences and average the resulting PSNR's.

The target bit rate R_T is fixed at 0.5 bpp. R_T^p and R_T^r are set at 0.2 bpp and 0.05 bpp, respectively. For a fair comparison, we select the MDSQ coder with the parameters that can achieve the same central PSNR at 0.5 bpp. In the experiment, the corrupted packet is recovered with the quantized mean value of the subband if both channels fail on that packet. The performances under variant packet loss ratios are presented in Fig. 4.7. Both systems achieve the same PSNR if no packet gets lost. However, the performance of MDSQ degrades rapidly with the packet loss ratio increasing; the drop of PSNR can be up to 3.2 dB. On the other hand, FO-MDC proves to be much more robust. The PSNR degradation is less than 1.5 dB even for a high loss ratio of 30%.

We also execute the subjective test at low bitrates. The Barbara image is coded with $R_T = 0.35$ bpp. We choose 0.1375 bpp and 0.0375 bpp for R_T^p and R_T^r, respectively. The corresponding MDSQ parameters are adjusted to attain the same central PSNR to make sure of fair comparison. The reconstructed images with 30% packet loss are shown in Figs. 4.8-4.9. Compared with MDSQ, FO-MDC achieves a better reconstruction of the textures and edges with the same corrupted packets.

4.5 Summary

In this work, we have proposed a novel multiple description coding system, which features high coding efficiency, robustness and asymptotical tradeoff points. Unlike most of the recent developments, our approach is based on the observation that only partial samples of the specified image source are highly sensitive to data corruption. Hence it is reasonable to emphasize these samples with high priorities instead of equally dealing with all the data. An adaptive redundancy allocator is well designed by jointly considering the data estimation and the rate-distortion optimization. As a result, the sensitive samples are well protected with more coding bits while the remaining data can still be reconstructed at acceptable qualities with the convenience of error concealment technique.

A significant improvement is obtained by the FO-MDC compared to the state-of-the-art MDSQ coder in both objective and subjective experiments. Furthermore, we attain much less degradation with the variation of packet loss ratio, which shows a strong robustness to packet-erasure channels. Since only a simple uniform quantizer is used in our system, it is expected that there is still room for further improvement by employing “smarter” coding techniques. It is also our interest to extend the basic idea of FO-MDC to the case with more than two descriptions in the near future.

(a)

(b)

Figure 4.8. Two-channel reconstructions for Barbara with packet loss ratio of 30% at 0.35 bpp (i.e. 0.175 bpp/channel). (a) MDSQ, PSNR = 25.87 dB, (b) FO-MDC, PSNR = 26.67 dB.

(a)

(b)

Figure 4.9. Comparison of the zoomed region in Fig. 4.8: (a) MDSQ, (b) FO-MDC.

CHAPTER 5

OPTIMAL MULTIPLE DESCRIPTION VIDEO CODING BASED ON PAIRWISE CORRELATING TRANSFORM

5.1 Introduction

Recently, Wang et al. developed a MD encoding scheme that uses multiple description correlating transform (MDTC) to introduce a controlled amount of correlating (and hence, redundancy) between the two bitstreams to improve the quality when only one description is received [98]. This general framework has been applied to image coding. It yields acceptable images from a one-channel reconstruction with only small amount of redundancy. Reibman adopted this MDTC method and proposed a MD video coder that uses motion-compensated predictions [72].

However, MD video coding is more than simply applying a MD image coder to prediction errors. A mismatching problem occurs when doing motion compensations with the reference frames reconstructed by only one description. Reibman proposed a drift-free MD video coding scheme by introducing two additional side loops besides the original center loop so that the mismatching error can be decreased. However, new redundancy arising from the side loops has to be considered within the proposed scheme. It is necessary to efficiently allocate the redundancy while still achieving the best rate-distortion performance.

The rest of this chapter is organized as follows. Section 5.2 briefly reviews the MDTC and illustrates the challenge of applying MDTC to MD video coding system. The drift-free MDTC based coding scheme and the new problems incurred are discussed. In Section 5.3, the measurement of redundancy rate-distortion is described. The proposed optimization approach is presented in Section 5.4. Simulation results are reported in Section 5.5. Section 5.6 summarizes this work.

5.2 MDTC Based MD Video Coding

5.2.1 Multiple Description Transform Coding with Two Channels

As mentioned in Chapter 2, MDTC with two channels uses a pairwise correlating transform to each pair of uncorrelated variables obtained from the KLT or DCT as described in Fig. 2.6. Assume A and B are two independent Gaussian variables with variances σ_A^2 and σ_B^2 respectively. They are paired and transformed into C and D using

$$\begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{T} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{cot(\theta)}{2}} & \sqrt{\frac{tan(\theta)}{2}} \\ -\sqrt{\frac{cot(\theta)}{2}} & \sqrt{\frac{tan(\theta)}{2}} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}. \tag{5.1}$$

The transform rotation angle θ, $\theta \in (0, \frac{\pi}{2})$, controls the correlation between C and D, which changes the redundancy of the MDTC coder. Suppose the variances of C and D are denoted as σ_C^2 and σ_D^2, respectively. If the dependency between C and D is parameterized by the angle ϕ, described as $E\{CD\} = \sigma_C \sigma_D cos\phi$, then the θ can be used to vary ϕ according to $tan(\theta) = \frac{\sigma_A}{\sigma_B} tan(\frac{\phi}{2})$ [97]. The value of θ adjusts the projections on the bases of MDTC, which actually distributes the source energy to the outputs. In order to capture most principal components of the source, the basis should be skewed toward the variable that has larger variance [28]. Since the transform is nonorthogonal except $\theta = \frac{\pi}{4}$, quantizing C and D may lead to degradation. Hence the transform is performed on the quantized indices of A and B, denoted as $\overline{A}$ and $\overline{B}$. An invertible integer transform is implemented to generate integer indices, thus (5.1) should be changed to $[C, D]^T = \mathbf{T_{int}}[\overline{A}, \overline{B}]^T$, where $\mathbf{T_{int}}$ denotes the integer correlating transform following the method proposed in [110].

If only one description, e.g., C, is received as shown in the top block of Fig. 2.6, inverse quantization is first applied to yield $\widetilde{C}$, then an optimal linear predictor is used to estimate D from $\widetilde{C}$ by minimizing the MSE.

$$\widehat{D} = -\frac{\sigma_A^2 - \sigma_B^2 tan^2(\theta)}{\sigma_A^2 + \sigma_B^2 tan^2(\theta)} \cdot \widetilde{C} \tag{5.2}$$

where $\widehat{D}$ is the estimate of D. Take inverse transform on $\widetilde{C}$ and $\widehat{D}$, the estimated $\widehat{A}$ and $\widehat{B}$ can be obtained as follows:

$$\begin{aligned} \begin{bmatrix} \widehat{A} \\ \widehat{B} \end{bmatrix} &= \begin{bmatrix} \sqrt{\frac{tan(\theta)}{2}} & -\sqrt{\frac{tan(\theta)}{2}} \\ \sqrt{\frac{cot(\theta)}{2}} & \sqrt{\frac{cot(\theta)}{2}} \end{bmatrix} \begin{bmatrix} \widetilde{C} \\ \widehat{D} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma_A^2 \sqrt{2tan(\theta)}}{\sigma_A^2 + tan^2(\theta)\sigma_B^2} \cdot \widetilde{C} \\ \frac{\sigma_B^2 tan(\theta)\sqrt{2tan(\theta)}}{\sigma_A^2 + tan^2(\theta)\sigma_B^2} \cdot \widetilde{C} \end{bmatrix} \end{aligned} \tag{5.3}$$

The reconstruction when only $\overline{D}$ is received follows from symmetry.

5.2.2 Video Coding Using MDTC

The key challenge in developing a MD approach to video coding lies in the coding of prediction errors. The difficulty arises from the variety of different predictions that might be used at the decoder of a MD system. If the data from both channels are received, the best predictor would be formed from information on both channels. If the data from either single channel is received, two other predictors would be formed. Without motion, it is possible to design the information on the two channels to impose a known relationship between the two-channel predictor and the two one-channel predictors. However, when motion compensation is used, no such design is known. Consequently, three distinct prediction error signals might be available at the decoder. If at any time the decoder uses a predictor whose corresponding prediction error is not available, a mismatch condition exists between the encoding and decoding loops. Mismatch errors are never corrected until the encoding and decoding loops are cleared by an intra-coded frame.

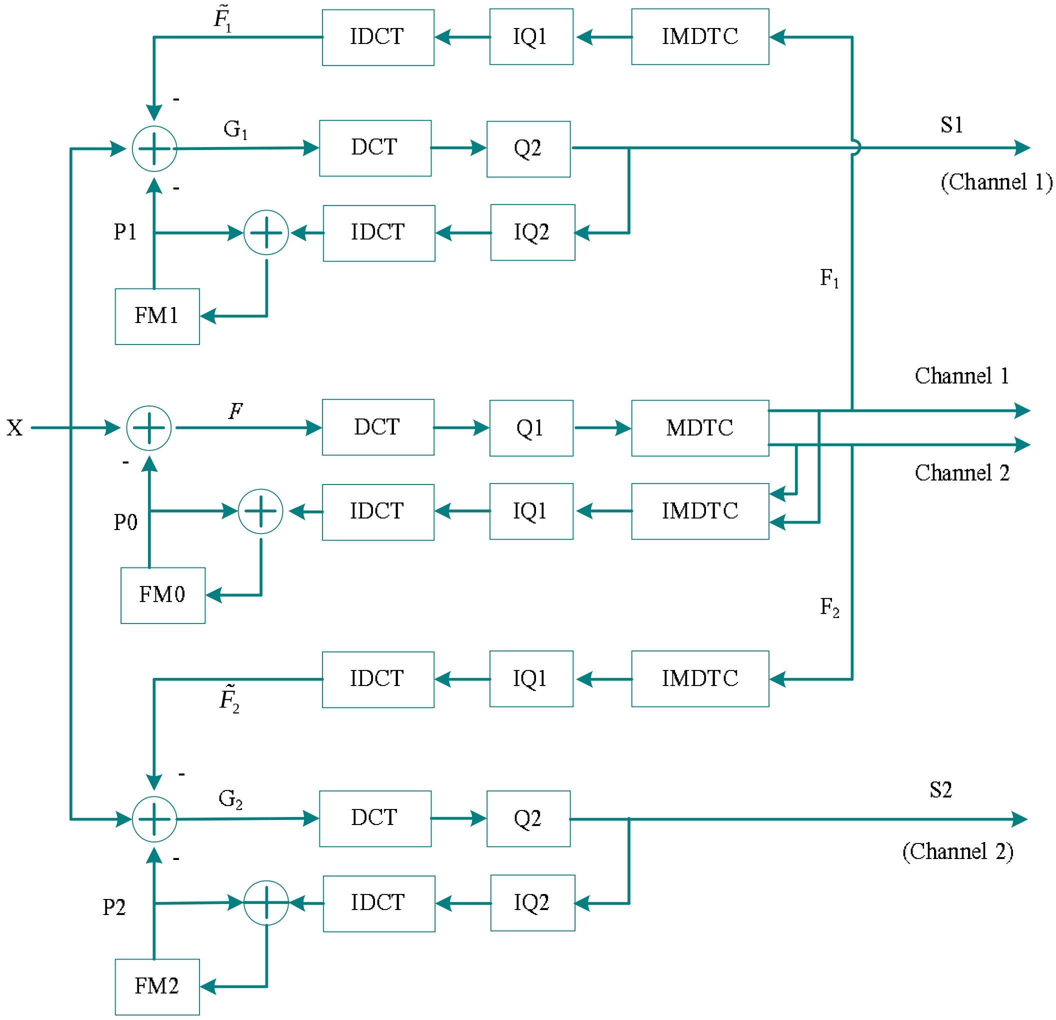


Figure 5.1. Framework for MDTC based video coding system.

In general, there are two sources for distortion in a MD video coder. One source of distortion is the quantization of prediction errors. This is common between a single description (SD) and an MD video coder; although the MD coder may have more than one prediction loop. The second source of distortion is the mismatch between the prediction loops at the encoder and decoder.

Reibman et al. uses three separate prediction paths at the encoder [71], to mimic the three possible scenarios at the decoder: both descriptions received or either of the two descriptions received. Specifically, the encoder has three frame buffers, storing the previously reconstructed frames from both descriptions ($\psi_{0,k-1}$), description one ($\psi_{1,k-1}$), and description two ($\psi_{2,k-1}$), where k represents the current frame time. For each block X_k, the encoder generates a predicted block $P_{i,k}, i = 0, 1, 2$ based on the motion vector and the previous frame $\psi_{i,k-1}$. $P_{0,k}$ is referred as the central prediction, and $P_{1,k}$ and $P_{2,k}$ are taken as the side predictions. Fig. 5.1 describes the framework of the MDTC based MD video coding scheme, where F_i and $G_{i,k}, k = 0, 1, 2$ represents the prediction error and mismatch, respectively. Note that in Fig. 5.1, the time dependency, indicated by the subscript k, is omitted for clarity. Since the mismatch is decreased greatly by coding G_i, this MD codec is also named as a *drift-free MD video coding system.*

The prediction error assuming both descriptions are available ($F = X - P_0$) is coded into two descriptions F_1 and F_2 using a MDTC based coder. If both descriptions, F_1 and F_2, are available to the decoder, it can recover F, the reconstructed prediction error from both descriptions. If the decoder receives only description i, it generates $\widetilde{F}_i$. In the absence of any additional information, the reconstruction from description i alone will be $P_i + \widetilde{F}_i$. However, to reduce the future mismatch between the prediction at the encoder and decoder, the encoder also generates and codes the side prediction error $G_i = X - P_i - \widetilde{F}_i$, transmitted as the side information S_i.

Note that in this framework, the total redundancy ρ includes two parts: (1) the redundancy introduced by MDTC, denoted as ρ_1, which is controlled by varying the transform parameter, i.e. the rotation angle θ; (2) the bits used for $G_{i,k}$, which are primarily redundancy since a typical decoder will not use them when both descriptions are received. This portion of the total redundancy, ρ_2, can be controlled directly by varying the quantization accuracy of G_i, i.e. Q_2.

Therefore, a new problem of allocating redundancy between the correlated MDTC coefficients and the purely redundant side information is incurred. Heuristic redundancy allocation is applied in [72]. The redundancy is allocated to ρ_1 at first, then the rest portion of redundancy is allocated to ρ_2 when the target correlation is getting higher. This method is simple and easy for implementation. However, it loses the property of optimal rate-distortion design. It is very important to smartly allocate the redundancy between ρ_1 and ρ_2 so that the optimal rate-distortion performance can be achieved. Motivated by this point, we proposed an optimization approach to solve the problem.

5.3 Redundancy Rate Distortion (RRD)

All approaches to MDC involve creating redundancy in the bitstreams transmitted over different channels. The *redundancy rate distortion* (RRD) function, proposed by Wang [98], is used as a measure of the efficiency for a MDC coder.

Given the average channel rate across both channels, R, a MDC coder attempts to jointly minimize two distortion measures, D_0 (distortion of the two-channel reconstruction) and $\overline{D}_1$ (average distortion of the one-channel reconstruction given equi-probable loss of either channel).

Conventionally, the coder minimizes D_0 for a given rate R of the standard source coder, i.e. SDC. The performance of such a coder is characterized by its rate-distortion function for a given source. Intuitively, redundancy is the extra bit rate compared to the SDC for the purpose of decreasing $\overline{D}_1$. More precisely, the redundancy in coding a source at two-channel distortion D_0, is the difference, $\rho = R - R^*$, between the per-variable transmitted bit rate $R = (R_1 + R_2)/2$ and the lowest rate R^* needed by any SDC coder to achieve the same two-channel distortion D_0, where R_1 and R_2 denote the rates for channel 1 and channel 2, respectively. The purpose of introducing redundancy in a representation is to lower the average one-channel distortion $\overline{D}_1$, and the function $\rho(\overline{D}_1; D_0)$ is used to represent the relationship between ρ and $\overline{D}_1$. This function describes how many bits of redundancy per variable are required by a coder to achieve a desired $\overline{D}_1$ at a given two-channel distortion D_0. Similar to RD curves, the RRD curve characterizes the performance of a specific MD coder on the source.

5.4 Redundancy Rate-Distortion Optimization

Since the redundancy and the distortion associated with the descriptions are incurred by the MDTC and the side information, they can be represented by the function of rotation angle for MDTC, θ, and the quantization parameter for the side loop, QP_2 ($Q_2 = 2QP_2$). To achieve a sophisticated rate-distortion control, we construct the optimization problem at the slice level [34].

Given the target redundancy, the optimal rotation angles and side information quantization parameters for the current frame can be determined by solving the constrained optimization problem,

$$min \sum_i \overline{D}(\theta_i, QP_{2,i}) \tag{5.4}$$

Subject to

$$\sum_i \rho(\theta_i, QP_{2,i}) \leq \rho^*, 0 \leq \rho^* \leq 1 \tag{5.5}$$

where, θ_i and $QP_{2,i}$ are the rotation angle and quantization parameter, respectively, used for the ith slice, $\overline{D}(\theta_i, QP_{2,i})$ denotes the associated average one-channel reconstruction distortion,

$$\overline{D}(\theta_i, QP_{2,i}) = \frac{1}{2}[D_{1,i}(\theta_i, QP_{2,i}) + D_{2,i}(\theta_i, QP_{2,i})], \tag{5.6}$$

and ρ^* is the target redundancy. $D_{k,i}(\theta_i, QP_{2,i}), k = 1, 2$ is the MSE between the original i^{th} slice and the one reconstructed from the estimated lost description using (5.3) and the dequantized drift errors for description k. $\rho(\theta_i, QP_{2,i})$ denotes the redundancy function for the i^{th} slice, which is computed as:

$$\rho(\theta_i, QP_{2,i}) = [R_1(\theta_i, QP_{2,i}) + R_2(\theta_i, QP_{2,i}) - R_i]/R_i$$

R_i is the bit rate of the ith slice under SDC scheme, while $R_k, k = 1, 2$ represents the bit rate for description k of the ith slice.

To solve this problem, we construct a cost function based on RRD framework with MSE as the distortion criterion. Thus, the optimal θ_i and $QP_{2,i}$ can be determined by:

$$\begin{aligned} min \Bigg\{ \sum_i \overline{D}(\theta_i, QP_{2,i}) + \lambda \Bigg[\sum_i R_1(\theta_i, QP_{2,i}) \\ + \sum_i R_2(\theta_i, QP_{2,i}) - (1+\rho^*) \sum_i R_i \Bigg] \Bigg\} \end{aligned} \tag{5.7}$$

Since redundancy and distortion measures over the slices within a frame are additive, (5.7) can be converted to a simpler equivalent unconstrained problem as in (5.8).

$$\begin{aligned} (\mathbf{\Theta}, \mathbf{QP_2}) = arg \min_{\Theta, QP_2} \Bigg\{ \sum_i \Bigg[\overline{D}(\theta_i, QP_{2,i}) \\ + \lambda_i \Big(R_1(\theta_i, QP_{2,i}) + R_2(\theta_i, QP_{2,i}) - (1+\rho^*) R_i \Big) \Bigg] \Bigg\} \end{aligned} \tag{5.8}$$

Where, $\mathbf{\Theta} = \{\theta_1, ..., \theta_M\}$ and $\mathbf{QP_2} = \{QP_{2,1}, ..., QP_{2,M}\}$, M is the number of slices within one frame.

Although (5.8) relaxes the budget constraints, it is still burdensome to pursue the operating points on the convex hull of the RRD curve. Hence we propose a fast algorithm to approximate this procedure for determining the best operating point of each slice. It is optimal to introduce more redundancy between the correlating transform coefficients and then allocate the remained redundancy to the mismatching errors incurred by linear prediction [97]. Therefore we first determine the optimal operating point by fixing the $QP_{2,i}$ as the maximum quantization parameter allowed. A set of candidate θ's, $\Theta = \{\theta_j \in (0, \frac{\pi}{2}); 1 \leq j \leq n\}$, are considered for generating the operating points. By using the bisection algorithm [69], the optimal value of rotation angle, θ_i^*, can be determined by solving

$$\begin{aligned}\theta_i^* = arg\min_{\theta_i} \Bigg\{ \Big[& \overline{D}(\theta_i, QP_{2,max}) + \lambda_i \Big(R_1(\theta_i, QP_{2,max}) \\ & + R_2(\theta_i, QP_{2,max}) - (1+\rho^*)R_i \Big) \Big] \Bigg\}\end{aligned} \tag{5.9}$$

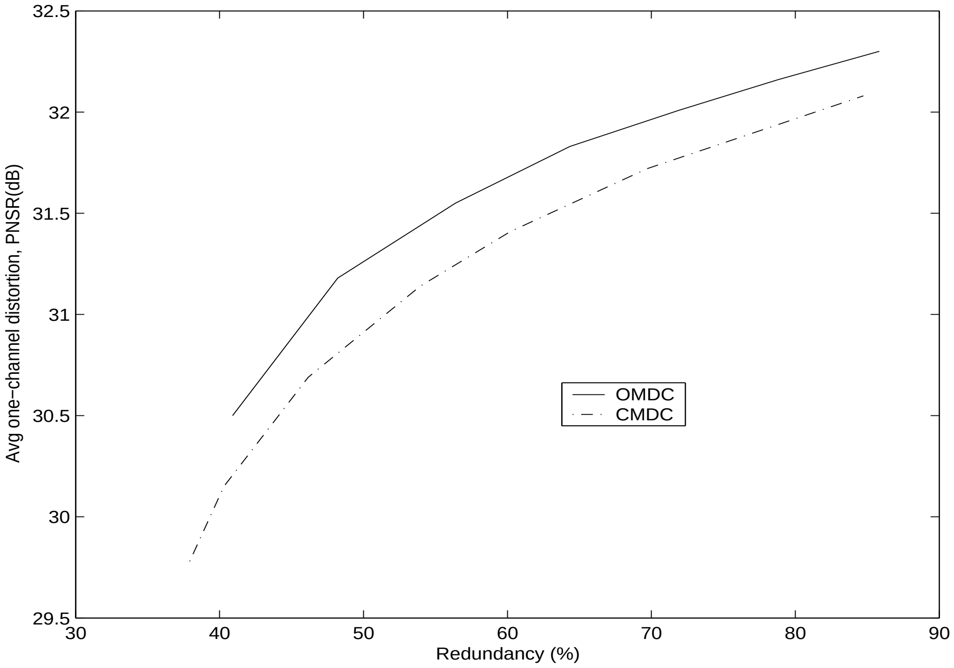
Note that θ_i^* is the rotation angle which achieves the optimal rate-distortion performance while still does not exceed the target redundancy ρ^*. After we obtain θ_i^*, a local rate-distortion curve can be built by varying $QP_{2,i}$.

$$\begin{aligned}QP_{2,i}^* = arg\min_{QP_{2,i}} \Bigg\{ \Big[& \overline{D}(\theta_i^*, QP_{2,i}) + \lambda_i \Big(R_1(\theta_i^*, QP_{2,i}) \\ & + R_2(\theta_i^*, QP_{2,i}) - (1+\rho^*)R_i \Big) \Big] \Bigg\}\end{aligned} \tag{5.10}$$

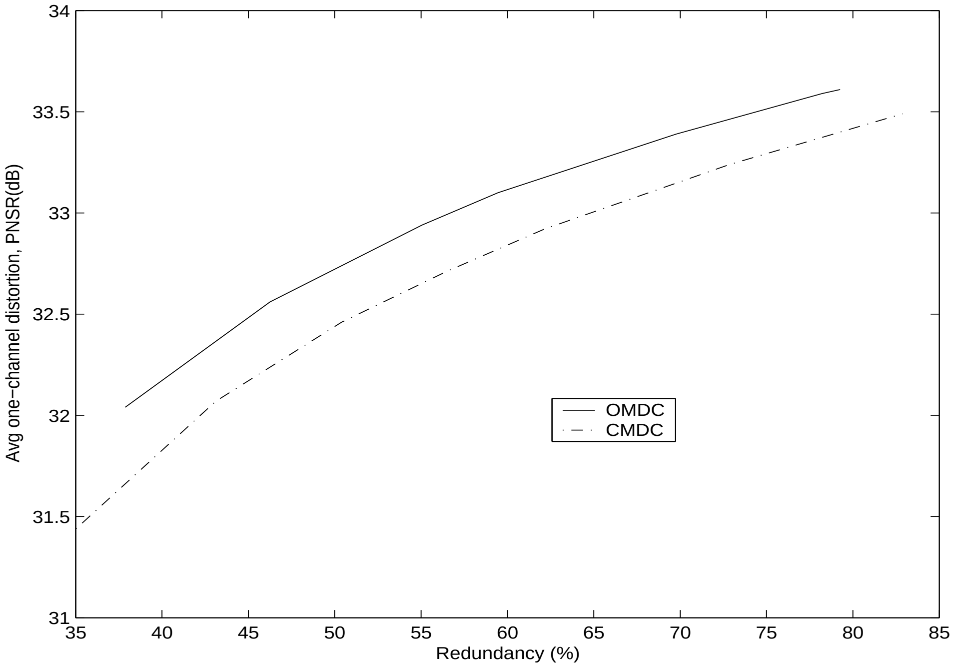
Similarly, we can determine the optimal $QP_{2,i}^*$ with bisection algorithm.

5.5 Simulation Results

We examine the performance of our proposed rate-distortion optimized MDC algorithm (OMDC) with the conventional non-optimized MDC method (CMDC) proposed in [72], in which a heuristic method is used to allocate the redundancy. Two sequences, “Foreman” and “Hall monitor” (both in CIF,30Hz), are tested in the simulations. All the comparisons are made by assuming that one entire description is lost. Distortion is measured as the average PSNR (dB) over the reconstructed frames using the survived description. The slice is simply formed by a row of macroblocks. Quantization

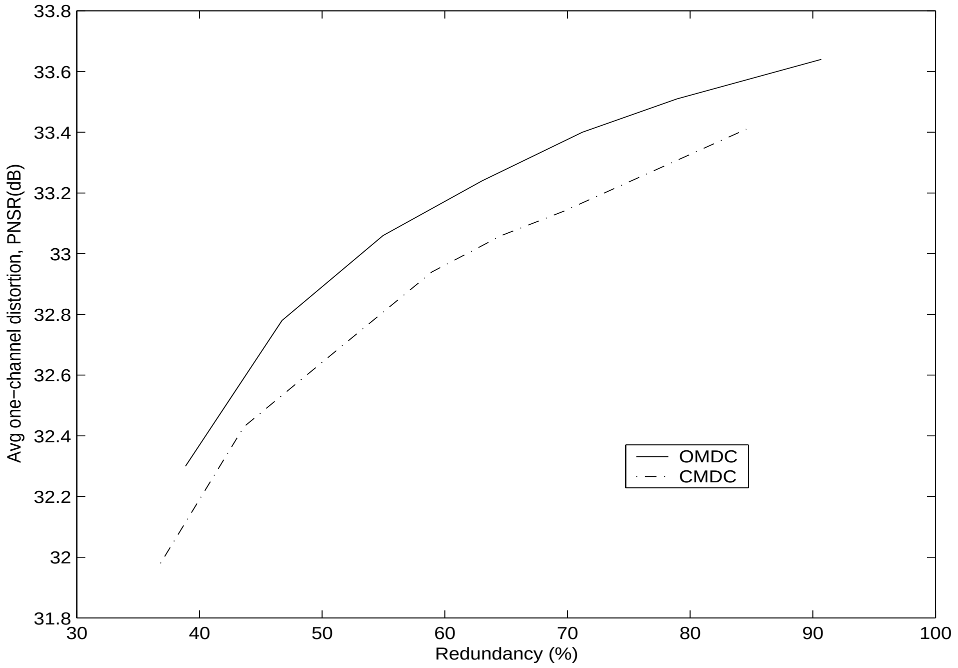


(a)

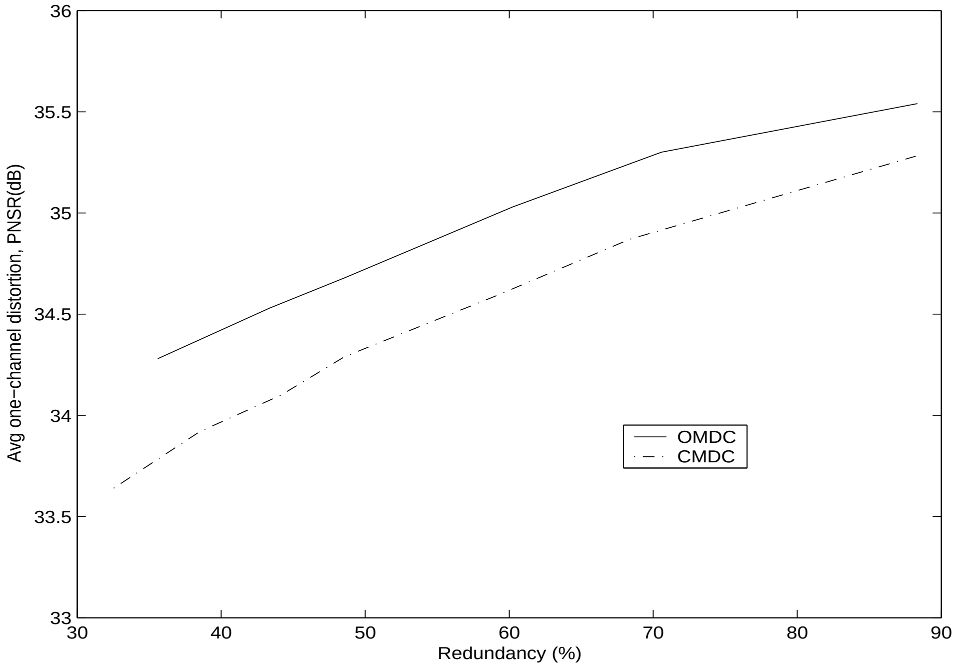


(b)

Figure 5.2. Rate-distortion performance comparison (Foreman, CIF, 30 Hz): (a) SDC target bit rate is 243 Kbps, two-channel reconstruction PSNR is 32.76 dB, (b) SDC target bit rate is 408 Kbps, two-channel reconstruction PSNR is 34.48 dB.



(a)



(b)

Figure 5.3. Rate-distortion performance comparison (Hall monitor, CIF, 30 Hz): (a) SDC target bit rate is 111 Kbps, two-channel reconstruction PSNR is 33.86 dB, (b) SDC target bit rate is 215 Kbps, two-channel reconstruction PSNR is 35.80 dB.

parameters are maintained to be the same within one slice because of our optimization strategy. The descriptions generated by the OMDC are compliant to the syntax of H.263. Motion vectors and head information are duplicated into both descriptions. For simplicity, the number of the coefficients paired in Y, Cb and Cr are 16, 4 and 4, respectively [72]. The remained coefficients are alternatively split to one of the descriptions. We use $\Theta = \{\frac{j\pi}{32}; 1 \leq j \leq 15\}$ as the candidate θ_i's for (5.9).

In the experiments, a target bit rate for the SDC encoder is selected. The descriptions are optimally generated by using the proposed OMDC algorithm under a given target redundancy ρ^*. Only the first frame is intra-coded, and all the remaining frames are inter-coded. The average luminance PSNR across time is recorded under certain redundancy which is expressed in terms of the percentage over the reference luminance bit rate, i.e. the bit rate of SDC. By varying ρ^*, the RRD curve can be obtained.

Fig. 5.2 shows the RRD curves obtained for “Foreman” at the bit-rates of 243 kbps and 408 kbps. It is evident that the proposed ODMC consistently outperforms CMDC by about 0.2 dB with the same redundancy at low bit rate (243 kbps). When the target bit-rate increases, the improvement is also increased, the average gain is about 0.3 dB. This is because that more redundancy is allocated to the side loop drift with a high bit-rate budget. While at the low bit-rate, most of the bits have to be used to code the correlating transform coefficients since the estimation can work better with fewer transmitted information. Similar improvement can be found from “Hall monitor” as described in Fig. 5.3, which has the gains over CMDC varying from 0.2 to 0.5 dB.

5.6 Summary

In this chapter we proposed an efficient redundancy rate-distortion optimization approach for improving the performance of the popular MDTC based drift-free MD video coding scheme [53]. The optimization is implemented on each slice within one frame. A fast sub-optimal operating points pursuit method is applied to simplify the procedure. It is shown through simulation results that the presented approach improves the conventional MDC scheme by 0.2-0.5 dB with the same redundancy over variant bit rates.

CHAPTER 6

CONCLUSIONS

This thesis introduces three novel error-resilient algorithms based on the multiple description coding techniques for image and video transmission over error-prone networks.

First, we introduced an edge-adaptive error concealment technique for improving the state-of-the-art domain-based MD coding scheme. The intra-band dependency is fully exploited in this algorithm by successfully retrieving the correlations along all the orientations between a corrupted coefficient and its available neighbors inside a local window using the covariance method. Hence, the important high-frequency components can be well recovered by manipulating the weights of linear estimation. With the ease of adapting estimation to the dominant edge orientations, the corrupted coefficients of the lost descriptions can be effectively reconstructed within the wavelet subbands. The simulation results show that the presented error concealment technique achieves an excellent performance on both objective and subjective measures.

Second, a novel feature-oriented MD coding system is presented for its high coding efficiency and robustness. Instead of equally dealing with the wavelet coefficients, the proposed algorithm smartly analyzes the local statistics of each subband and identifies those samples that are highly sensitive to data corruption for the specified image source. An adaptive redundancy allocator is well designed by jointly considering the data estimation and the rate-distortion optimization. As a result, the sensitive samples are well protected with more coding bits while the remaining data can still be reconstructed at acceptable qualities with the convenience of error concealment technique. Extensive experiments are carried out to test the efficiency and resilience of our codec. Simulation results show that significant improvement over the well-known MDSQ coder is achieved in both objective and subjective experiments. Furthermore, much less degradation is attained by our codec with the variation of packet loss ratios, which shows a strong robustness to packet-erasure channels.

Finally, we proposed an efficient redundancy rate-distortion optimization approach to improve the performance of the conventional MDTC based drift-free video coding scheme. By using the Lagrangian relaxation method, the optimization is implemented on each slice of a frame to achieve the best trade-off among the redundancy allocation of the correlating transform coefficients and the side-loop drifts. The proposed algorithm outperforms the conventional codec with the same redundancy over variant bit rates.

APPENDIX A

IMAGE CODING BASED ON MULTIRESOLUTION DIRECTIONAL FILTER BANK

A.1 Introduction

In the past two decades, wavelets and filter banks (FBs) have gained considerable interest in signal processing, partly due to the ability of the wavelet functions and their associated regular FBs to optimally represent one-dimensional piecewise smooth signals. However, the separable wavelets are not effective in capturing line discontinuities since they cannot take advantage of the geometrical regularity of image structures. Imagetransitions such as edges are expensive to represent through wavelets. Therefore, integrating the geometric regularity in the image representation is a key challenge to improve the performance of the current image coders.

Recently, Candes and Donoho constructed the curvelet transform [11], and proved that it is an essentially optimal representation of continuous two-variable function except at regularity along $\boldsymbol{C}^2$. The nonlinear approximation of a function f, $f_M^{(c)}$, is reconstructed by M curvelet coefficients, which have an asymptotic decay rate $\|f - f_M^{(c)}\|^2 \leq CM^{-2}(\log_2 M)^3$. This decay rate of the approximation error is a huge improvement compared to those by the wavelet functions, or Fourier basis, which are $O(M^{-1})$ and $O(M^{-1/2})$, respectively [55].

Since the space of smooth functions with singularities along $\boldsymbol{C}^2$ curves is similar to natural images with regions of continuous intensity value and discontinuous along smooth curves (edges), there is a strong motivation for finding similar transforms in the discrete domain [94]. Note that there are already several proposed methods to implement the digital curvelet transform [19, 83], but they all are redundant.

In this appendix, two novel image coding schemes are proposed based on the multiresolution directional filter banks (DFB) [51, 52]. The first work applies a non-uniform quincunx directional filter bank (nuqDFB) proposed in [59] to image decomposition. Based on the study of joint statistical characteristics of coefficients of nuqDFB, we propose an embedded coder for low bit-rate image coding using the nuqDFB. The morphological operation is employed to find the clusters of significant positions in each bit-plane. Contexts are designed to encode these large values witharithmetic coding. The remained sparse regions full of large quantities of zeros are coded with the aid of probability modeling using a tarp filter [81]. In the second work, we propose a novel embedded image coding scheme with an overcomplete pyramidal directional filter bank (PDFB). The morphological operation is employed progressively to identify clusters of significant coefficients in each bit-plane. Context-based arithmetic coding is used to encode these significant coefficients. We design the context models so that the intra-band and inter-band correlations of the overcomplete PDFB can be well exploited. Experimental results show that the proposed coding algorithms outperform the current state-of-the-art wavelet based coders, such as SPIHT and JPEG2000, for images with directional features.

The rest of the appendix is organized as follows. Section A.2 presents the non-uniform quincunx DFB and the fast implementation with separable structure. Then we discuss the proposed nuqDFB based coding scheme including the algorithms for coding clustered significant coefficients. The coding results are compared to the state-of-the-art wavelet image coders. In Section A.3, the coding scheme based on PDFB is described. The embedding passes and the context based arithmetic coding are discussed. Experiments are carried out to justify the coding efficiency of the proposed coders. Section A.4 summarizes this Chapter.

A.2 Image Coding Based on Non-Uniform Quincunx DFB

A.2.1 The non-uniform quincunx DFB

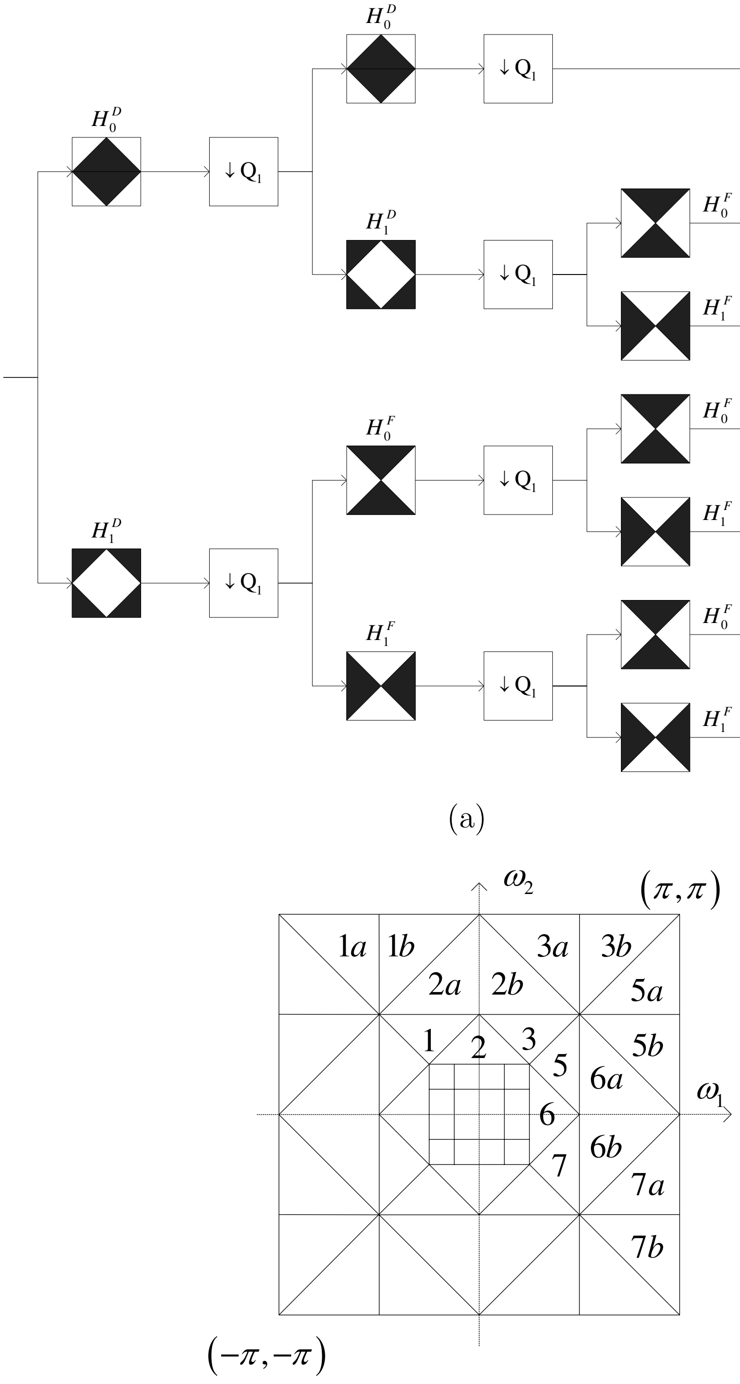
The nuqDFB is a member of a family of multiresolution directional filter banks (DFBs) introduced in [59]. The nuqDFB is a multiresolution FB, whose support of the lowpass filter is the same as that of the wavelet FB. But instead of three highpass filters like the traditional separable FB, the nuqDFB has six (or twelve) directional highpass subbands. Other principal characteristics of the nuqDFB are: (i) it can be implemented by a very efficient structure, with overall complexity equal to a separable FB; (ii) at the lowpass branch, a different multiresolutional FB can be used to cascade to obtain a (hybrid) multiscale image decomposition; (iii) it is quite simple to design since it can be constructed by using only two-channel diamond FBs as shown in Fig. A.1(a). The detailed construction of the multiresolution DFB and the nuqDFB is referred to [59, 60].

In this paper, the image is decomposed by the nuqDFB at the first two highest resolutions, and the separable wavelet FB is used at the lower scales. The partitioning of the discrete frequency plane is depicted in Fig. A.1(b).

A.2.2 Image coding using hybrid nuqDFB and wavelet FB

Since the number of significant coefficients of the nuqDFB is less than that of the wavelet, recognizing the statistical properties of transform coefficients is a crucial task in design of high-performance nuqDFB based coders. An explicit way to solve this problem is to classify the coefficients of each subband into two subsets that represent the insignificant and significant coefficients. Thus the quantization and estimated models can be adapted to each subset independently.

A potential approach to exploit this behavior has been introduced with the morphological operation for wavelet coding [77]. Based on the observation that clusters tend to grow both in spatial and in frequency domains, the previously detected significant coefficients are used as seeds for the search of new significant ones. The clustering trend of significant coefficients also exists in the nuqDFB bands. This suggests using a morphological dilation to identify the significant coefficients in the nuqDFB subbands before the coding step.



(a)

(b)

Figure A.1. The nuqDFB employed in image coding. (a) The tree structure to implement the nuqDFB and (b) The partitioning of the frequency plane by a hybrid nuqDFB and wavelet decomposition..

A.2.2.1 Progressive morphological dilation

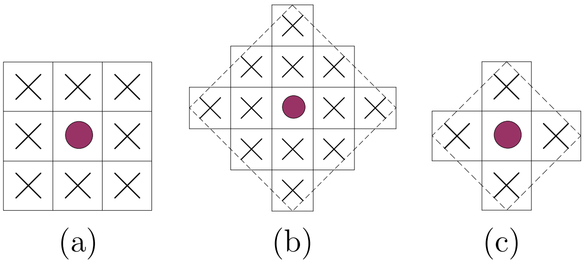


Figure A.2. Structuring elements used in the progressive morphological dilation: (a) SE_1, (b) SE_2, (c) SE_3.

Erosion and dilation are two morphological operators frequently adopted [95]. Let I be a binary-valued image where dilation will be applied. Dilation of a given set $A \subseteq I$ with set $B \subseteq I$ is defined by

$$A \oplus B = \cup_{b \in B} A_b, \tag{A.1}$$

where B is a binary-valued array called a *structuring element* (SE), A_b denotes the translation of A to a point b [31]. The dilation operation produces an enlarged set, $A \oplus B$, which can also be written as $A \cup (A \oplus B \setminus A)$, where $(A \oplus B \setminus A)$ represents the set of new points obtained by dilation. If A is the set of previously detected significant points, the points in set $(A \oplus B \setminus A)$ have a much higher probability to be recognized as significant.

Organizing and representing each nuqDFB subband as irregular shaped clusters of significant coefficients provide an efficient way for encoding. Exploiting cross-scale dependency can further improve the detection accuracy of significant coefficients. Moreover, shaping the clustering boundaries with less cost should also be considered with adaptive structuring element [116]. Hence, three different SE's, as shown in Fig. A.2, are adopted in our coding scheme for variant steps of dilation.

A progressive embedded coding algorithm is proposed to code the significant coefficients based on morphological dilation. Three passes are performed at each bit-plane:

1) *Significance detection* (SD) pass: The intra-band dependency is exploited in this pass. The square SE_1 in Fig. A.2(a) is adopted to detect new significant positions based on significant neighbors in the previous bit-planes.
2) *Cross-band prediction* (CBP) pass: Based on the fact that the inter-band correlation exists between the successive decomposition levels, a significant cluster in a children subband can be predicted by that in the parent subband. The diamond SE_2 in Fig. A.2(b) is employed to dilate around the associated children positions corresponding to each significant position in the parent subband.
3) *Boundary shaping* (BS) pass: Typically, on the boundaries of a large cluster, there are a few scattered significant coefficients located in small isolated clusters. It is difficult to forecast the dimension of these isolated clusters. Hence adaptive dilation is expected to search around a cluster. The smaller rood SE_3 in Fig. A.2(c) is adopted for dilation on the previously formed cluster boundaries which are identified by the insignificant positions. The boundary extension is adaptively controlled based on the occurrence of new significant positions detected. It stops until the recursive dilation results in no more new significant positions.

The entropy coding relies upon the use of classical context adaptive arithmetic coding to efficiently represent a collection of binary symbols. The probability models used by the arithmetic coder involve different contexts. A context for the coefficients to be coded is determined according to the significance status of its neighbors. Further details can be found in [88].

Although most significant coefficients in a bit-plane can be recognized by these three passes with the morphological dilation, there are still some scattered significant coefficients remained undetected. Zhong et. al. suggested to explicitly coding these scattered positions by transmitting their coordinates [116]. However, the efficiency depends on the image source and the SE of dilation. Since these scattered significant positions are located in the regions dominated by zeros, the probability is highly skewed. Hence a single context model based Tarp filter entropy coder [81] is employed to process the remaining scattered significant coefficients including the large quantity of zeros.

A.2.2.2 Tarp-filter-based coding for scattered significant coefficients

The Tarp filter technique was first introduced in [81]. The basic idea of this technique is to estimate the probability models with a causal filter that can drive a non-adaptive arithmetic coder to binary image compression.

Consider a long series of Bernoulli random variables whose probability of being one varies smoothly. Such an assumption is generally true for filter bank coefficients, especially in regions where there is little activity. The probability of the next variable being one is small and highly correlated with how far away the other nonzero variables are. The probability of the predicted variable being one can be assumed to be proportional to an exponential decay of the distance between the current variable and the others. A simple first-order recursive filter can be used to estimate the probability of the next variable being one:

$$p(t+1) = \alpha p(t) + (1-\alpha)v(t), \tag{A.2}$$

where $p(t)$ is the estimate of the probability of position t being one, $v(t) = 0$ or 1, is the observed value at position t, and α is a parameter which controls the speed of the probability estimate adaptation to the source data. It is easy to show that $p(t)$ is the output of convolving $v(t)$ and $f(t)$, where $f(t)$ has the frequency response of $H(w) = \frac{1-\alpha}{1-\alpha e^{-jw}}$.

Simard et. al. applied this 1-D filter to 2-D binary images by three 1-D filtering steps, which results in the Tarp filter. The first filter goes from left to right; the second filter goes from right to left, and is executed after each full line has been processed; the third filter goes from top to bottom for each column [81]. The processing strictly follows a raster scanning order and only causal information is employed to estimate the probability.

Within this work, we apply the Tarp-filter-based algorithm to the remaining regions after the three passes discussed above, i.e. the tarp filter is only used to estimate the probability of the scattered significant coefficients outside the clusters for each bit-plane. This process is also defined as the *scattered significant coefficient detection* (SSD) pass.

The filtering is still executed in the raster scan order. If a coefficient has been coded by the previous passes, it should be skipped without scanning. However, it conflicts with the raster scan encoding order since this "observed" value can not be used in the probability estimate, thus the Tarp filtering can not operate properly. To solve this, instead of the real values of the skipped coefficients, a cumulated probability estimate $p_c^d(k)$ is used for these coded positions, and is calculated as:

$$p_c^d(k) = n(k)_{sig}/N(k), \tag{A.3}$$

where $n(k)$ denotes the number of significant coefficients of the kth subband that have already been coded, and $N(k)$ is the total number of coefficients in subband k. By doing this, the Tarp filtering can work appropriately without imperiling the accuracy of probability estimation.

A.2.2.3 Algorithm summary

In this section, we summarize the coding algorithm. Five passes are used in this approach. The produced progressive bitstream can be truncated at any pass. Let $w_{ij}(k)$

be the coefficient with the coordinate (i, j) relative to the upper-left corner of subband k. $LSC(k)$ and $LIC(k)$ represent the lists of significant and insignificant coefficients, respectively. $D(SE_m)$, $m = 1, 2, 3$, denotes the operation of morphological dilation with the structuring element specified by SE_m. $V[w_{ij}(k)|SE_m] = w_{ij}(k) \oplus SE_m \backslash w_{ij}(k)$, defines the vicinity of $w_{ij}(k)$ generalized by the dilation with SE_m. Let $P(k)$ be the parent of subband k and $C[w_{ij}(k)]$ denote the children of $w_{ij}(k)$.

The embedding coding algorithm is described as follows.

1) (Initialization): Decompose the image with the nuqDFB. Find the maximum number of bit-planes, M, and set $n \leftarrow M$.
2) (*SD pass*): If $n = M$, go to step 5).

   Apply $D(SE_1)$ to each entry of $LSP(k)$. Encode all $w \in V[w_{ij}(k)|SE_1]$ that have not been scanned with arithmetic coding; if w is significant, encode its sign, and add it to the end of $LSC(k)$; else attach it to the end of $LIC(k)$.
3) (*CBP pass*): If $k = 0$, i.e. the LL band, go to step 4).

   For the significant coefficients of $P(k)$, apply $D(SE_2)$ to each entry of $C[w_{ij}(P(k))]$. Similar to *SD pass*, encode each newly scanned coefficient. Sign coding is applied if it is significant. Add these coefficients to the end of either $LSC(k)$ or $LIC(k)$ depending on whether they are significant or not.
4) (*BS pass*): Apply $D(SE_3)$ to each entry of $LIC(k)$. If a new significant coefficient is found, the recursive dilation is implemented around it until no more new significant coefficient is detected. Encode the scanned coefficients and update $LSC(k)$ and $LIC(k)$ correspondingly.
5) (*SSD pass*): Use Tarp-filter-based non-adaptive arithmetic coding to encode the remaining coefficients that have not been visited. Update $LSC(k)$ correspondingly.

6) (*Magnitude refinement pass*): If $n = M$, go to step 7).

   Encode the *nth* bit of those significant coefficients recognized by processing previous bit-planes. We employ the same context model as in JPEG2000 for this pass.

7) (New bit-plane preparation):

   Empty all LIC's, set $n \leftarrow n - 1$, and go back to step 2).

A.2.2.4 Simulation results

For the sake of comparison, the proposed nuqDFB-based algorithm is compared with three state-of-the-art wavelet based coding schemes: SPIHT [76] with arithmetic coding, JPEG2000 [3] and the TCE coder [12] which was reported as the best embedding Tarp filter based codec.

Three 512 × 512 grayscale images, Lena, Barbara and Fingerprint, are tested. A 5-level decomposition is used for SPIHT, TCE and JPEG2000 with the 9/7 Daubechies wavelet filters. On the other hand, a 5-level nuqDFB decomposition is applied in the proposed scheme, where the three coarser levels are decomposed using 9/7 filters and the two finer levels are decomposed into 12 bands each using nuqDFB as described in Section A.3.1.

Table A.1. Performance comparison (PSNR [dB]) for Barbara.

Rate(bpp)	SPIHT	TCE	JPEG2000	nuqDFB
0.10	24.24	24.23	24.66	25.54
0.15	25.63	25.90	25.93	26.95
0.20	26.63	26.24	27.31	27.93
0.25	27.56	27.88	28.36	29.21
0.30	28.54	28.55	29.24	29.89
0.40	30.09	30.40	30.83	31.27
0.50	31.38	31.82	32.26	32.83

Table A.2. Performance comparison (PSNR [dB]) for Fingerprint.

Rate(bpp)	SPIHT	TCE	JPEG2000	nuqDFB
0.10	21.44	21.15	21.50	22.11
0.15	22.94	23.33	23.31	24.04
0.20	24.19	24.35	24.53	25.01
0.25	25.17	24.83	25.41	26.10
0.30	26.02	25.85	26.33	27.00
0.40	27.59	27.86	27.96	28.49
0.50	28.80	28.65	29.06	29.67

Table A.3. Performance comparison (PSNR [dB]) for Lena.

Rate(bpp)	SPIHT	TCE	JPEG2000	nuqDFB
0.10	30.17	29.50	29.87	30.17
0.15	31.87	31.90	31.69	32.01
0.20	33.12	33.13	32.97	33.33
0.25	34.09	34.17	34.13	34.40
0.30	34.93	34.94	34.75	35.13
0.40	36.22	36.21	36.09	36.30
0.50	37.20	37.16	37.28	37.37

(a) (b)

Figure A.3. Reconstructed images of barbara at 0.25 bpp using (a) JPEG2000, PSNR = 28.36 dB, (b) nuqDFB, PSNR = 29.21 dB..

Tables A.4 - A.5 show the performance of different systems in terms of PSNR. The performances are calculated by truncating the embedded bitstreams at variant rates during decoding. In the comparison, the proposed algorithm consistently outperforms the other three schemes for all the images. The improvements are remarkable especially for rich-edgy images such as Barbara and Fingerprint. Moreover, the proposed algorithm is more efficient at very low bit rates since the nuqDFB can well represent the geometrical structures even with very few coefficients. For instance, the nuqDFB-based coder gains 1.02 dB in PSNR over JPEG2000 at 0.15 bpp for Barbara image. The reconstructed images of Barbara at 0.25 bpp using JPEG2000 and nuqDFB, respectively, are shown in Fig. A.3.

A.3 Image coding using PDFB

A.3.1 The reduced-aliasing PDFB

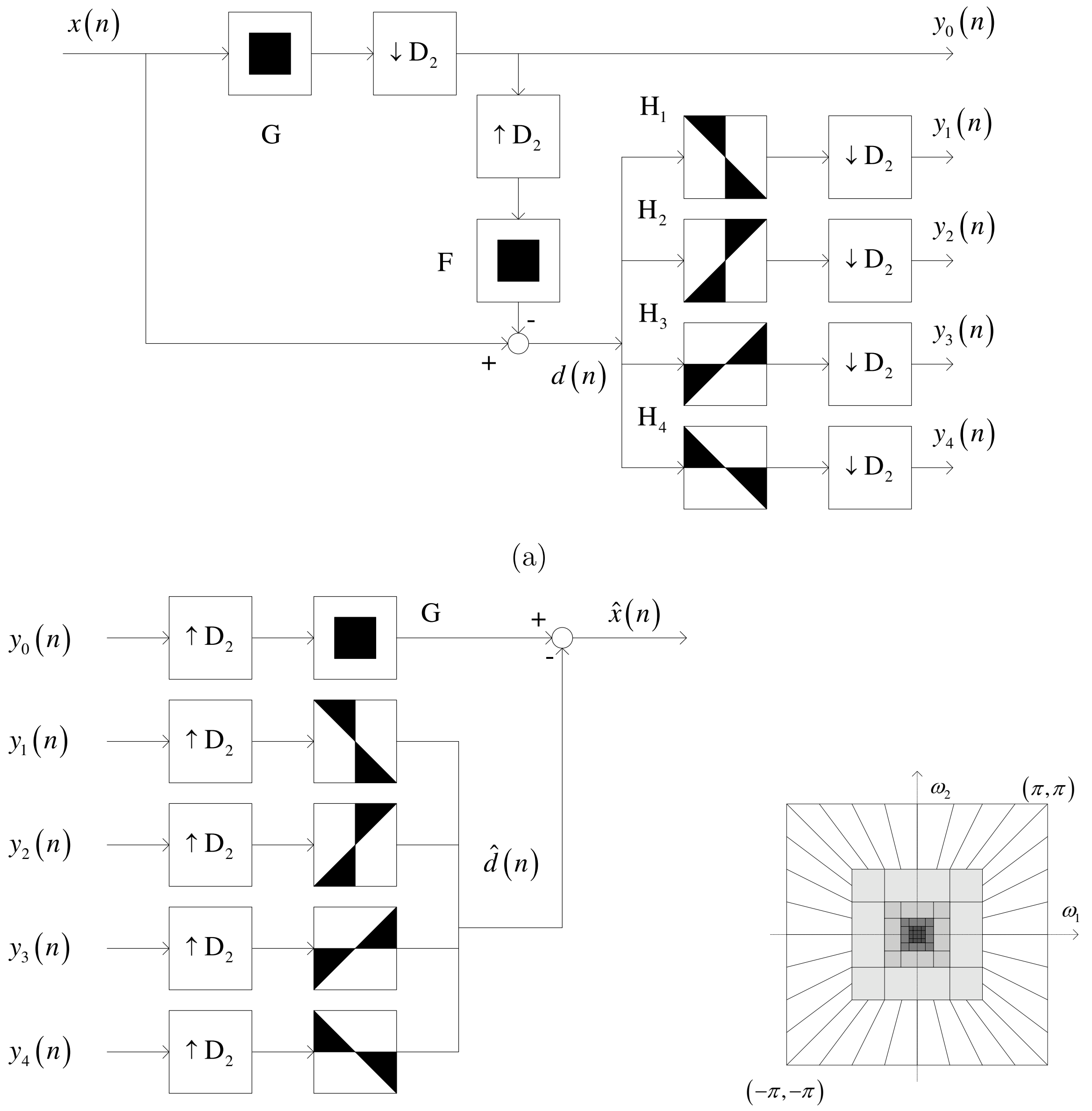
A PDFB consisting of a Laplacian pyramid and a four-channel DFB (Fig. A.4(a)) is equivalent to an overcomplete five-channel FB with a decimation matrix $\boldsymbol{D}_2 = 2\boldsymbol{I}$, ($\boldsymbol{I}$ is the identity matrix). The overall directional filters can be expressed in a closed-form formula from the Laplacian and the DFB filters [61]. It is shown that the aliasing occurs in these filters is from iterations in the Laplacian pyramid. It can be reduced or removed by requiring the two filters of the pyramid to satisfy the Nyquist property. This condition means that the frequency responses of the two lowpass filters in the pyramid ($F(\boldsymbol{\omega})$ and $G(\boldsymbol{\omega})$ in Figs. A.4(a) and (b)) should have the passband regions (including the transition bands) limited in $[-\pi/2, \pi/2]^2$. $F(\boldsymbol{\omega})$ and $G(\boldsymbol{\omega})$ in this paper are designed to have transition bands in the region $0.3\pi < |\omega_i| < 0.6\pi$.

Fig. A.4(c) depicts the frequency supports of the image decomposition, in which the PDFB is used at the highest resolution, and the DWT is used in the next four resolutions. To illustrate the construction, the analysis and synthesis sides of a PDFB with four directional filters are described in Fig. A.4(a) and (b), respectively. Note that the conventional DFB is implemented by a binary tree of two-channel FBs [7], and the four filters $H_i(\boldsymbol{z})$ in Fig. A.4(a) represent the equivalent directional filters of the DFB tree.

A.3.2 Image coding using hybrid overcomplete PDFB and wavelet FB

As discussed in Section A.2, recognizing the statistical properties of transform coefficients is also a crucial task in design of high-performance PDFB based coders. Similar morphological dilation is employed to help identifying the significant coefficients of PDFB. In this way, the intra-band and cross-scale correlations can be well exploited. On the other hand, although the overcomplete transform introduces more coefficients to be coded, strong correlations exist between neighboring directional subbands that can be adopted to improve the probability modelling. Hence contexts should be designed to reflect the local statistics among the directional subbands.

Besides the coding passes introduced in Section A.2.2, a new coding pass, *Neighboring correlation prediction pass* (NCP), is designed in PDFB based codec. This pass is designed to capture the redundancy within the overcomplete directional subbands. If the two spatial filters have small angle difference in their principal directions, their corresponding decimated subbands exhibit significant dependency between those coefficients at the same positions relative to their upper-left corners. Hence, the distribution of significant coefficients of each directional subband is highly correlated to their neighboring subbands (*cousin subbands*). Those coefficients associated with the same positions that have been recognized as significant by either one of the two cousin subbands are coded. SE_1 is employed to implement the dilation around the resulting new significant coefficients.



(a)

(b)

(c)

Figure A.4. (a) The analysis side of a PDFB; the DFB has four directional subbands and (b) the synthesis PDFB corresponds to analysis PDFB in (c). (c) The partitioning of the frequency plane by a hybrid PDFB and wavelet decomposition.

A.3.2.1 Context modeling

According to the conditional probability theory, context modeling can exploit the inter-symbol redundancy by switching between different probability models [108]. In embedded coding systems, the coding is conducted on a series of significance maps that correspond to a set of decreasing thresholds. After each pass of coding, all coefficients are quantized to specific values and can be used as context information. Thus, a significant context template, as described in Fig. A.5, is defined to exploit the coefficients that have been coded by the previous passes or bit-planes.

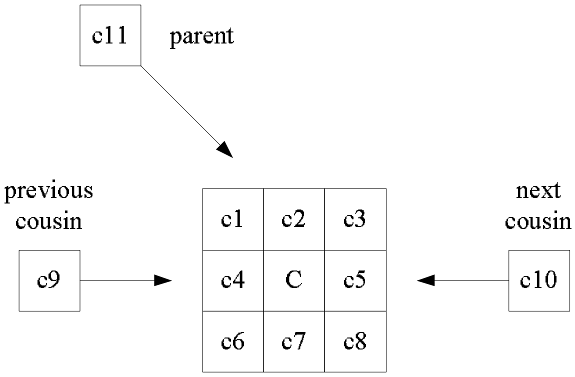


Figure A.5. Context modeling template..

In Fig. A.5, C is the current symbol that needs to be coded. $c_1, ..., c_{11}$ are reconstructed values up to the current coding pass of the corresponding coefficients. We define a series of reconstruction matrices $\mathbf{Y}_s^n$, in which each element is the reconstructed value of the corresponding coefficient, n indicates the order of coding passes, s denotes the subband index. Assume C is the (i, j)th element of the binary significance map in pass n of subband s. Thus the intra-band correlation is derived from the 8 neighboring coefficients,

$$c_m = y_s^{n'}(i + a, j + b),\ a, b = \text{-1, 0, or 1},\ m = 1, ..., 8. \tag{A.4}$$

where $n' = n-1$ or n since the morphological dilation does not guarantee a raster scan order for the coefficients that have been coded. The coefficients of parent and cousin subbands are also used to exploit the inter-band dependency

$$\begin{aligned} c_9 &= y^n_{s-1}(N_c(i), N_c(j)), \\ c_{10} &= y^{n-1}_{s+1}(N_c(i), N_c(j)), \\ c_{11} &= y^n_{P(s)}(P_c(i), P_c(j)), \end{aligned} \tag{A.5}$$

where $P(\cdot)$ specifies the parent of band s, $N_c(\cdot)$ and $P_c(\cdot)$ are functions for the cousin and parent subbands, respectively, that specify the coordinates corresponding to the positions in the current subband. Considering the weak correlation among the cousin subbands for wavelet, c_9 and c_{10} are not used for coding wavelet coefficients in the simulation. Thus the NCP pass is executed only for those directional subbands.

A.3.2.2 Algorithm summary

In this section, we summarize the coding algorithm. Six passes are designed in this approach. The produced progressive bitstream can be truncated at any pass. Let $w_{ij}(k)$ be the coefficient with the coordinate (i, j) relative to the upper-left corner of subband k. $LSC(k)$ and $LIC(k)$ represent the lists of significant and insignificant coefficients, respectively. $D(SE_m)$, $m = 1, 2, 3$, denotes the operation of morphological dilation with the structuring element specified by SE_m. $V[w_{ij}(k)|SE_m] = w_{ij}(k) \oplus SE_m \setminus w_{ij}(k)$, defines the vicinity of $w_{ij}(k)$ generated by dilation with SE_m. Let $P(k)$ and $N(k)$ be the parent and cousin bands of subband k, respectively. $C[w_{ij}(k)]$ and $N[w_{ij}(k)]$ denote the corresponding coefficients in $P(k)$ and $N(k)$ associated with $w_{ij}(k)$.

The PDFB-based coding algorithm is summarized as follows.

1) (*Initialization*): Decompose the image with PDFB. Find the maximum number of bit-planes M and set $n \leftarrow M$.

2) (*SD pass*): If $n = M$, go to step 6).

   Apply $D(SE_1)$ to each entry of $LSP(k)$. Encode all $w \in V[w_{ij}(k)|SE_1]$ that have not been scanned with arithmetic coding, if w is insignificant, attach it to the end of $LIC(k)$; else encode the sign of w, and add it to the end of $LSC(k)$.

3) (*CBP pass*): If $k = 0$, i.e. the LL band, go to step 5).

   For the significant coefficients of $P(k)$, apply $D(SE_2)$ to each entry of $C[w_{ij}(P(k))]$. Similar to *SD pass*, encode each newly scanned coefficient and add these coefficients to the end of either $LSC(k)$ or $LIC(k)$ depending on whether they are significant or not.

4) (*NCP pass*): If subband k is a wavelet band, go to step 5).

   For the significant coefficients of $N(k)$, apply $D(SE_1)$ to each entry of $N[w_{ij}(N(k))]$. Encode each scanned coefficient. Add the coefficient to the end of $LSC(k)$ if it is found to be a new significant coefficient; otherwise append it to $LIC(k)$.

5) (*BS pass*): Apply $D(SE_3)$ to each entry of $LIC(k)$. If a new significant coefficient is found, the recursive dilation is implemented around it until no more new significant coefficient is detected. Encode the scanned coefficients and update $LSC(k)$ and $LIC(k)$ correspondingly.

6) (*SSD pass*): The remaining coefficients are scanned and coded in raster order. However, if one significant coefficient is found, apply $D(SE_3)$ at once, then continue to scan the following coefficients that have not been coded. Update $LSC(k)$ correspondingly.

7) (*MR pass*): If $n = M$, go to step 8);

   Encode the *nth* bit of those significant coefficients recognized by previous bit-planes.

8) (*New bit-plane*): Empty all LIC's, set $n \leftarrow n - 1$, and go back to step 2).

A.3.2.3 Simulation results

For the sake of comparison, the proposed PDFB-based algorithm is compared with three state-of-the-art wavelet based coding schemes: SPIHT [76] with arithmetic coding, JPEG2000 [3] and the embedded coder of morphological representation of wavelet data (MRWD) [77]. Two 512×512 grayscale images, Barbara and Lena, are tested. A 5-level decomposition is used for SPIHT, MRWD and JPEG2000 with 9/7 Daubechies wavelet filters. On the other hand, a 5-level PDFB decomposition, as described in Fig. A.4(a), is applied in the proposed scheme, where the four coarser levels are decomposed using 9/7 filters and the level with the finest resolution is decomposed into 16 directional subbands using PDFB.

Table A.4. Performance comparison (PSNR [dB]) for Barbara.

Rate(bpp)	SPIHT	MRWD	JPEG2000	PDFB
0.10	24.24	24.15	24.66	25.24
0.15	25.63	25.32	25.93	26.74
0.20	26.63	26.86	27.31	27.84
0.25	27.56	27.51	28.36	28.94
0.30	28.54	28.16	29.24	29.86
0.40	30.09	30.18	30.83	30.97
0.50	31.38	31.31	32.26	32.42

Table A.4 shows the performance of Barbara with different systems in terms of PSNR. The performances are calculated by truncating the embedded bitstreams at different rates during decoding. In the comparison, the proposed algorithm consistently outperforms all the other three schemes. Although the overcomplete transform is employed in the proposed scheme, the improvements are still remarkable for rich-edgy images. This is because of that the PDFB can represent the geometrical regularity of image

Table A.5. Performance comparison (PSNR [dB]) for Lena.

Rate(bpp)	SPIHT	MRWD	JPEG2000	PDFB
0.10	30.17	30.18	29.87	29.86
0.15	31.87	31.55	31.69	31.87
0.20	33.12	33.14	32.97	33.07
0.25	34.09	33.90	34.13	33.98
0.30	34.93	34.57	34.75	34.91
0.40	36.22	36.17	36.09	35.96
0.50	37.20	37.01	37.22	36.78

structures with fewer coefficients and smaller magnitudes. Hence, the proposed algorithm is more efficient at very low bit rates when a few of significant coefficients are actually used for reconstruction. For instance, the PDFB-based coder gains 0.81 dB in PSNR over JPEG2000 at 0.15 bpp for Barbara image. Fig. A.6 shows the zoomed portion of the reconstructed Barbara images at 0.15 bpp for JPEG2000 and the proposed coder. On the other hand, the PDFB-based codec exhibits competitive performance for smooth images, such as Lena. As shown in Table A.5, the performance of our codec is comparable to that of the other three coders, and the performance difference is getting smaller as the bit-rate decreases.

A.4 Summary

In this appendix, two novel embedded coders have been proposed based on hybrid multiresolution DFB. The concentrated significant coefficients in the directional subbands are identified with the aid of progressive morphological dilation. The cluster boundaries are adaptively shaped to detect the scattered significant coefficients. We design the context models so as to take advantage of the the intra-band and inter-band statistical dependencies. The highly optimized rate-distortion embedding performance is achieved

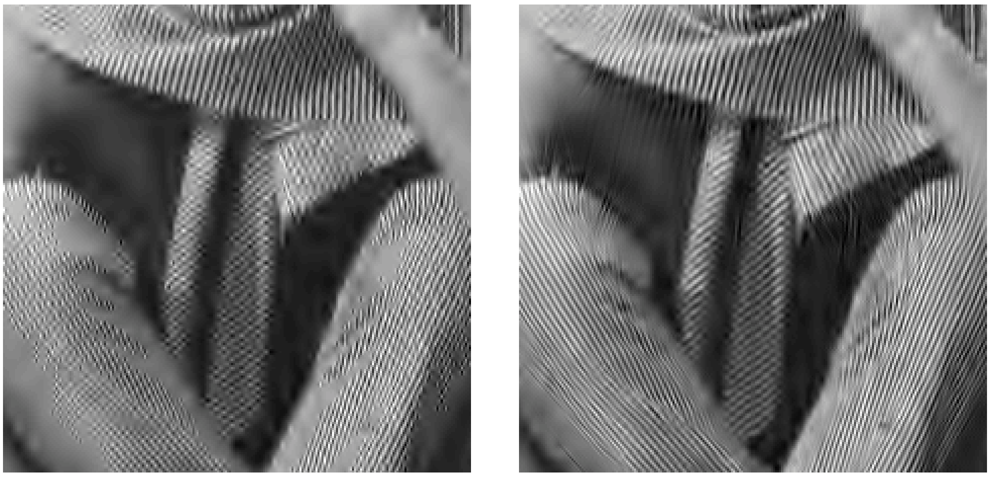

(a) (b)

Figure A.6. Reconstructed Barbara images at 0.15 bpp using (a) JPEG2000, PSNR = 25.93 dB and (b) PDFB, PSNR = 26.74 dB..

with the progressive coding structure. Experimental results justify that the proposed coding schemes are superior to the state-of-the-art wavelet-based image coders.

APPENDIX B

ABBREVIATION LIST

ARQ	Automatic Repeat Request.
BI	Bilinear Interpolation.
DCT	Discrete Cosine Transform.
DFB	Directional Filter Bank.
DP-MDC	Data-partitioning based Multiple Description Coding.
EA-EC	Edge-Adaptive Error Concealment.
EM	Expectation-Maximization.
FB	Filter Bank.
FEC	Forward Error Correction.
FO-MDC	Feature-oriented Multiple Description Coding.
HMT	Hidden Markov Tree.
KLT	Karhunen-Loeve Transform.
LC	Layered Coding.
LOT	Lapped Orthogonal Transform.
MD	Multiple Description.
MDC	Multiple Description Coding.
MDSQ	Multiple Description Scalar Quantizer.
MDTC	Multiple Description Correlating Transform.
ML	Maximum Likelihood.
nuqDFB	non-uniform quincunx Directional Filter Bank.
PCT	Pairwise Correlating Transform.
PDFB	Pyramidal Directional Filter Bank.
POCS	Projection Onto Convex Sets.
PSNR	Peak Signal to Noise Ratio.
RD	Rate-Distortion.
RRD	Redundancy Rate Distortion.

RTCP	Real-Time Control Protocol.
RTP	Realtime Transport Protocol.
RVLC	Reversible Variable Length Code.
SDC	Single Description Coding.
SFQ	Spatial-Frequency Quantization.
SPIHT	Set Partitioning in Hierarchical Trees.
SSIM	Structural similarity.
TCE	Tarp-filter-based system which utilizes Classification of coefficients to achieve Embedding.
TCP	Transmission Control Protocol.
UEP	Unequal Error Protection.
UDP	User Datagram Protocol.
VLC	Variable Length Code.
VRC	Video Redundancy Coding.

REFERENCES

[1] Generic coding of moving pictures and associated audio information- part 2: Video. *ISO/IEC Committee Draft 13 818-2, Information Technology.*

[2] Motion pictures expert group-overview of the mpeg-4 standard. *ISO/IEC JTC1/SC29/WG11 N2459.*

[3] T. Acharya and P.S. Tsai. *JPEG2000 Standard for Image Compression.* John Wiley and Sons, Inc., 2005.

[4] R. Ahlswede. The rate-distortion region of a binary source for multiple description without excess rate. *IEEE Trans. Inform. Theory*, 31:721–726, Nov. 1985.

[5] Ivan V. Bajic and John W. Woods. Domain-based multiple description coding of images and video. *IEEE Trans. Image Processing*, 12(10):1211–1225, Oct. 2003.

[6] Ivan V. Bajic and John W. Woods. Maximum minimal distance partitioning of the $\mathcal{Z}^2$ lattice. *IEEE Trans. Inform.Theory*, 49(4):981–992, April 2003.

[7] Roberto H. Bamberger and Mark J. T. Smith. A filter bank for the directional decomposition of images: theory and design. *IEEE Transactions on Signal Processing*, 40(7):882 –893, April 1992.

[8] T. Berger. *Rate distortion theory.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[9] N.V. Boulgouris, K.E. Zachariadis, A. Leontaris, and M.G. Strintzis. Drift-free multiple description doding of video. In *Multimedia Signal Processing*, Cannes, France, Oct. 2001.

[10] R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Trans. Image Processing*, 8:1688–1701, Dec. 1999.

[11] Emmanuel J. Candès and David L. Donoho. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In L. L. Schumaker et al., editor, *Curves and Surfaces.* Vanderbilt University Press, Nashville, TN, 1999.

[12] T. Chao and S. S. Hemami. An embedded image coding system based on tarp filter with classification. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, volume 3, pages 49–52, May 2004.

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

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

[15] Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46(4):886–902, Apr. 1998.

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

[17] G.A. Darbellay and I. Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Trans. Information Theory*, 45(4):1315–1321, May 1999.

[18] G. Davis and J. Danskin. Wavelet image compression construction kit version 0.3. Technical report, 1997.

[19] Minh N. Do and Martin Vetterli. Pyramidal directional filter banks and curvelets. In *Proc. of IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, October 2001.

[20] A. R. Reibman et al. Multiple description video coding using rate-distortion splitting. In *IEEE Int. Conf. Image Processing*, volume 1, pages 978–981, 2001.

[21] D. Comas et al. Rate-distortion optimization in a robust video transmission based on unbalanced multiple description coding. In *Proc. IEEE Int. Workshop Multimedia Signal Processing*, pages 581–586, 2001.

[22] M. Fleming and M. Effros. Generalized multiple description vector quantization. In *Proc. IEEE Conf. on Data Compression*, pages 3–12, Snowbird, UT, Mar. 1999.

[23] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inform. Theory*, IT-28:851–857, Nov. 1982.

[24] M. Ghanbari. Cell-loss concealment in atm video codes. *IEEE Trans. Circuits Syst. Video Technol.*, 3:238–247, June 1993.

[25] V. K. Goyal. *Beyond Traditional Tranform Coding.* PhD thesis, University of California, Berkeley, 1998.

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

[27] V. K. Goyal and J. Kovacevic. Optimal multiple description transform coding of gaussian vectors. In *Proc. Data Compression Conference (DCC'98)*, pages 388–397, 1998.

[28] V. K. Goyal and J. Kovacevic. Generalized multiple description transform coding with correlating transforms. *IEEE Trans. Inform. Theory*, 47:2199–2224, Sept. 2001.

[29] V. K. Goyal, J. Kovacevic, R. Arean, and M. Vetterli. Multiple description transform coding of images. In *Proc. IEEE Int. Conf. Image Processing(ICIP'98)*, Oct. 1998.

[30] V. K. Goyal, J. Kovacevic, and M. Vetterli. Multiple description transform coding: robustness to erasures using tight frame expansions. In *IEEE Int. Symp. Inform. Theory*, page 408, Aug. 1998.

[31] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision.* Reading, MA: Addison-Wesley, 1992.

[32] S.S. Hemami. Robust image transmission using resynchronizing variable-length codes and error concealment. *IEEE Trans. on Selected Areas in Communications*, 18(6):927–939, June 2000.

[33] S.S. Hemami and T.H.-Y. Meng. Transform coded image reconstruction exploiting interblock correlation. *IEEE Trans. Image Processing*, 4:1023–1027, July 1995.

[34] ITU-T. Video coding for low bit-rate communication. *ITU-T Rec. H.263, Version 1: Nov. 1995, Version 2: Jan. 1998, Version 3: Nov. 2000*, 1993.

[35] H. Jafarkhani and V. Tarokh. Multiple description trellis-coded quantization. *IEEE Trans. Commun.*, 47(6):799–803, June 1999.

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

[37] W. Jiang and A. Ortega. Multiple description coding via polyphase transform and selective quantization. In *Proc. VCIP 99*, pages 998–1008, Feb. 1999.

[38] Steven M. Kay. *Fundamentals of statistical signal processing: estimation theory.* Prenice hall PTR, Upper Saddle River, NJ, 1993.

[39] L.H. Kieu and K.N. Ngan. Cell-loss concealment techniques for layered video codecs in an atm network. *IEEE Trans. Image Processing*, 3:666–677, Sept. 1994.

[40] B. J. Kim, Z. Xiong, and W. A. Pearlman. Low bit-rate scalable video coding with 3-dimentional set partitioning in hierarchical trees (3-d spiht). *IEEE Trans. Circuits Sys. Video Technol.*, 10(8):1374–1387, Dec. 2000.

[41] I.K. Kim and N.I. Cho. Error resilient video coding using optimal multiple description of dct coefficients. In *Proc. IEEE Int. Conf. Image Processing*, volume 3, pages 663–666, 2003.

[42] Joohee Kim, R.M. Mersereau, and Y. Altunbasak. Error-resilient image and video transmission over the internet using unequal error protection. *IEEE Trans. Image Processing*, 12(2):121–131, Feb. 2003.

[43] Joohee Kim, R.M. Mersereau, and Y. Altunbasak. A multiple-substream unequal error-protection and error-concealment algorithm for spiht-coded video bitstreams. *IEEE Trans. Image Processing*, 13(12):1547–1553, Dec. 2004.

[44] X. Lee, Y.-Q. Zhang, and A. Leon-Garcia. Information loss recovery for block-based image coding techniquesa fuzzy logic approach. *IEEE Trans. Image Processing*, 4:259–273, Mar. 1995.

[45] Y.-C Lee, Y. Altunbasak, and R. M. Mersereau. A two-stage multiple description video coder with drift-preventing motion compensated prediction. In *Proc. IEEE International Conference on Image Processing*, pages 557–560, Sept. 2002.

[46] X. Li. *Edge directed statistical inference and its applications to image processing.* PhD thesis, Princeton University, 2000.

[47] X. Li, B. Tao, and M. T. Orchard. On implementing transforms from integers to integers. In *Proc. IEEE Int. Conf. Image Processing*, volume 1, pages 881–885, Oct. 1998.

[48] S. Lin and D.J. Costello. *Error control coding.* Prentice-Hall, 1983.

[49] T. Linder, R. Zamir, and K. Zeger. The multiple description rate region for high resolution source coding. In *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 1998.

[50] J. Liu and P. Moulin. Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. *IEEE Trans. Circuits Sys. Video Technol.*, 10(11):1647–1658, Nov. 2001.

[51] Y. Liu, T. T. Nguyen, and S. Oraintara. Image coding using quincunx directional filter bank. *Sumbitted to International Symposium on Circuits and Systems (ISCAS'06)*, Sept. 2005.

[52] Y. Liu, T. T. Nguyen, and S. Oraintara. Low bit-rate image coding based on pyramidal directional filter banks. *Sumbitted to International Conference on Acoustics, Speech, and Signal Processing (ICASSP'06)*, Sept. 2005.

[53] Y. Liu and S. Oraintara. Drift-free multiple description video coding with redundancy rate-distortion optimization. In *International Symposium on Circuits and Systems (ISCAS'05)*, pages 4034–4037, May 2005.

[54] Y. Liu and S. Oraintara. Multiple description coding with edge-adaptive reconstruction. *WSEAS Trans. on Information Science and Applications*, 2(3):335–342, Mar. 2005.

[55] Stephane Mallat. *A wavelet tour of signal processing.* Academic press, London, UK, 1999.

[56] S. Mehrotra and P.A. Chou. On optimal frame expansions for multiple description quantization. In *IEEE Int. Symp. Inform. Theory*, page 176, June 2000.

[57] R. Motwani and C. Guillemot. Tree-structured oversampled filterbanks as joint source-channel codes: application to image transmission over erasure channels. *IEEE Trans. Signal Processing*, 52(9):2584–2599, Sept. 2004.

[58] T. Nguyen and A. Zakhor. Matching pursuits based multiple description video coding for lossy environments. In *Proc. International Conference on Image Processing*, volume 1, pages 57–60, Sept. 2003.

[59] Truong Trong Nguyen and Soontorn Oraintara. A class of multiresolution directional filter bank. *Submitted to the IEEE Transaction on Signal Processing*, May 2005.

[60] Truong Trong Nguyen and Soontorn Oraintara. Multiresolution direction filter banks: theory, design and application. *IEEE Transaction on Signal Processing*, 53(10):3895–3905, Oct 2005.

[61] Truong Trong Nguyen and Soontorn Oraintara. On the aliasing effect of the contourlet filter banks. *Submitted to the International Symposium on Circuits and Systems (ISCAS'06)*, May 2006.

[62] M. T. Orchard, Y. Wang, V. A. Vaishampayan, and A. R. Reibman. Redundant rate-distortion analysis of multiple description coding. In *Proc. IEEE Int. Conf. Image Processing*, Santa Barbara, CA., Oct. 1997.

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

[64] V. Parthasarathy, J.W. Modestino, and K.S. Vastola. Design of a transport coding scheme for high quality video over atm networks. *IEEE Trans. Circuits Syst. Video Technol.*, 7(2):358–376, Apr. 1997.

[65] William K. Pratt. *Digital Image Processing.* John Wiley and Sons Ltd, 2001.

[66] R. Puri, K. W. Lee, K. Ramchandran, and V. Bharghavan. An integrated source transcoding and congestion control paradigm for video streaming in th internet. *IEEE Trans. Mutimedia*, 3(1):18–32, Mar 2001.

[67] R. Puri and K. Ramchandran. Multiple description source coding through forward error correction codes. *in Proc. 33rd Asilomar Conf. Signals, System Comp.*, 1:342–346, 1999.

[68] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–285, Feb. 1989.

[69] K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Trans. Image Processing*, 2:160–175, Apr. 1993.

[70] K.R. Rao and Z.S. Bojkovic. *Packet video communications over ATM networks.* Prenice hall, 2000.

[71] A. Reibman, H. Jafarkhani, Y. Wang, M. Orchard, and R. Puri. Multiple description coding for video using motion compensated prediction. *IEEE Int. Conf. Image Processing, Kobe, Japan*, Oct. 1999.

[72] A. R. Reibman, H. Jafarkhani, Y. Wang, and M. T. Orchard. Multiple description video coding using motion-compensated temporal prediction. *IEEE Trans. Circuit Syst. Video Technol.*, 12(3):193–204, Mar. 2002.

[73] J. Ridge, F. W. Ware, and J. D. Gibson. Permuted smoothed descriptions and refinement coding for images. *IEEE J. Select. Areas Commun.*, 18(6):915–926, June 2000.

[74] Justin K. Romberg, Hyeokho Choi, and Richard G. Baraniuk. Bayesian tree-structred image modeling using wavelet-domain hidden markov models. *IEEE Trans. Image Processing*, 10(7):1056–1068, July 2001.

[75] Abdul H. Sadka. *Compressed Video Communications.* John Wiley and Sons Ltd, 2002.

[76] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. Circuits Syst. Video Technol.*, 6:23–50, June 1996.

[77] S. D. Servetto, K. Ramchandran, and M. T. Orchard. Image coding based on a morphological representation of wavelet data. *IEEE Transaction on Image Processing*, 8(9):1161–1174, Sept. 1999.

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

[79] S. D. Servetto, V.A. Vaishampayan, and N.J.A. Sloane. Multiple description lattice vector quantization. In *Proc. IEEE Conf. on Data Compression*, pages 13–22, Snowbird, UT, Mar. 1999.

[80] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. Acoust., Speech, Signal Processing*, 36:1445–1453, Sept. 1988.

[81] P. Simard, D. Steinkraus, and H. Malvar. On-line adaptation in image coding with a 2-d tarp filter. In *Proc. IEEE International Conference on Data Compression Conference*, pages 23–32, Apr. 2002.

[82] J.D. Spragins, J.L. Hammond, and K. Pawlikowski. *Telecommunications: Protocols and design*. Addison-Wesley, 1991.

[83] Jean-Luc Starck, Emmanuel J. Candès, and David L. Donoho. The curvelet transform for image denoising. *IEEE Transaction on Image Processing*, 11(6):670–684, June 2002.

[84] H. Sun, K. Challapali, and J. Zdepski. Error concealment in digital simulcast ad-hdtv decoder. *IEEE Trans. Consumer Electron.*, 38:108–117, Aug. 1992.

[85] H. Sun and W. Kwok. Concealment of damaged block transform coded images using projections onto convex sets. *IEEE Trans. Image Processing*, 4:470–477, Apr. 1995.

[86] R. Talluri. Error resilient video coding in the mpeg-4 standard. *IEEE Commun. Mag.*, pages 112–119, June 1998.

[87] X. Tang and A. Zakhor. Matching pursuits multiple description coding for wireless video. *IEEE Trans. Circuits Syst. Video Technol.*, 12(6):566–575, June 2002.

[88] D. Taubman. High performance scalable image compression with ebcot. *IEEE Trans. Image Processing*, 9:1158–1170, July 2000.

[89] D. Taubman and A. Zakhor. Multirate 3-d subband coding of video. *IEEE Trans. Image Processing*, 3:572–588, Sept. 1994.

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

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

[92] V. A. Vaishampayan and S. John. Interframe balanced multiple description video compression. In *Proc. Packet Video 99*, pages 812–816, Apr. 1999.

[93] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto. Multiple-description vector quantization with lattice codebook: design and analysis. *IEEE Trans. Inform. Theory*, 47(5):1718–1734, July 2001.

[94] Martin Vetterli. Wavelets, approximation and compression. *IEEE Signal Processing Magazine*, 18(5):59–73, 2001.

[95] L. Vincent. Morphological grayscale reconstruction in image analysis: Applications and effective algorithms. *IEEE Transaction on Image Processing*, 2:176–201, Apr. 1993.

[96] Y. Wang, M. T. Orchard, and A. R. Reibman. Multiple description image coding for noisy channels by pairing transform coefficients. In *Proc. Workshop on Multimedia Signal Processing*, pages 419–424, Princeton, NJ., June 1997.

[97] Y. Wang, M. T. Orchard, and A. R. Reibman. Optimal pairwise correlating transform for multiple description coding. In *IEEE Int. Conf. Image Processing*, Chicago, IL, 1998.

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

[99] Y. Wang, J. Ostermann, and Y.-Q. Zhang. *Video processing and communication.* Prentice-Hall, 2002.

[100] Y. Wang and Q.-F. Zhu. Error control and concealment for video communication. *Proceedings of the IEEE*, 86(5):974–997, May 1998.

[101] Y. Wang, Q.-F. Zhu, and L. Shaw. Maximally smooth image recovery in transform coding. *IEEE Trans. Commun.*, 41:1544C1551, Oct. 1993.

[102] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, Apr. 2004.

[103] Z. Wang, Y. Yu, and D. Zhang. Best neighborhood matching: An information loss restoration technique for block-based image coding system. *IEEE Trans. Image Processing*, 7(7):1056–1061, July 1998.

[104] Jennifer L. H. Webb. Efficient table access for reversible variable-length decoding. *IEEE Trans. Circuits Syst. Video Technol.*, 11(8):981–985, Aug. 2001.

[105] J. Wen and J.D. Villasenor. A class of reversible variable length codes for robust image and video coding. In *Proc. IEEE Int. Conf. Image Processing*, volume 2, pages 65–68, Oct. 1997.

[106] S. Wenger. Video redundancy coding in h.263+. In *Audio-Visual Services Over Packet Networks Conf.*, Aberdeen, U.K., 1997.

[107] J.K. Wolf, A.D. Wyner, and J. Ziv. Source coding for multiple descriptions. *Bell Syst. Tech. J.*, 59:1417–1426, Oct. 1980.

[108] X. Wu. High-order context modeling and embedded conditional entropy coding of wavelet coefficients for image compression. In *Proc. 31st Asilomar Conf. Signals, Systems, Computers*, pages 1378–1382, Nov. 1997.

[109] Z. Xiong, K. Ramchandran, and M.T. Orchard. Space-frequency quantization for wavelet image coding. *IEEE Trans. Image Processing*, 6(5):677–693, May 1997.

[110] X.Li, B.Tao, and M.T.Orchard. On implementing transforms from integers to integers. In *IEEE Int. Conf. Image Processing (ICIP'98)*, pages 881–885, June 1998.

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

[112] R. Zamir. Shannon-type bounds for multiple descriptions of a stationary source. In *J. Comb., Inform. Syst. Sci.*, Dec. 2000.

[113] W. Zeng and B. Liu. Geometric-structure-based error concealment with novel applications in block-based low-bit-rate coding. *IEEE Trans, Circuits and Systems for Video Technology*, 9(4):648–665, June 1999.

[114] Z. Zhang and T. Berger. New results in binary multiple descriptions. *IEEE Trans. Inform. Theory*, IT-33:502C521, July 1987.

[115] Z. Zhang and T. Berger. Multiple description source coding with no excess maginial rate. *IEEE Trans. Inform. Theory*, 41(2):349–357, March 1995.

[116] J. M. Zhong, C. H. Leung, and Y. Y. Tang. Wavelet image coding based on significance extraction using morphological operations. In *Proc. Inst. Elect. Eng.– Vis. Image Signal Processing*, volume 146, pages 206–210, 1999.

[117] Q.-F. Zhu, Y. Wang, and L. Shaw. Coding and cell loss recovery for dct-based packet video. *IEEE Trans. Circuits Syst. Video Technol.*, 3:248–258, June 1993.

BIOGRAPHICAL STATEMENT

Yilong Liu was born in JiangSu, China. He received the B.Eng. and the M.Eng. degrees in Electrical Engineering from Southeast University in Nanjing, China, in 1998 and 2001, respectively. Starting from the spring of 2002, he continued his education at the University of Texas, Arlington, where he conducted research on signal processing. During his Ph.D studies, he worked as an intern at the Nokia research center, Dallas, from June 2004 till December 2004. His research interests are in the area of image/video processing, multiple description coding and motion estimation. He is also a student member of the IEEE.